Universidad de Granada

FACULTAD DE PSICOLOGÍA

Departamento de Psicología Social

y Metodología de las Ciencias del Comportamiento

VIABILIDAD DE LA ENSEÑANZA DE LA INFERENCIA BAYESIANA EN EL ANÁLISIS DE DATOS EN PSICOLOGÍA

TESIS DOCTORAL

Presentada por: M. Carmen Díaz Batanero

Directora: E. Inmaculada de la Fuente Solana

Granada, 2007

La investigación presentada en esta tesis doctoral ha sido financiada por la Beca de Formación de Profesorado Universitario del Ministerio de Educación y Ciencia AP2003- 5130 y los proyectos SEJ2004-00789 y SEJ2006-13009/psic.

VIABILIDAD DE LA ENSEÑANZA DE LA INFERENCIA BAYESIANA EN EL ANÁLISIS DE DATOS EN PSICOLOGÍA

Tesis Doctoral presentada por
Dña. M. CARMEN DÍAZ BATANERO dentro del
programa *"Psicología Social: Aplicaciones y Métodos"*
para aspirar al grado de Doctora, Mención Europea
por la Universidad de Granada,
dirigida por la Catedrática de Universidad
Dña. E. INMACULADA DE LA FUENTE SOLANA

Granada, 30 de Enero de 2007

Fdo. M. Carmen Díaz Batanero

Dra. Dña. E. Inmaculada de la Fuente Solana, catedrática de Metodología de las Ciencias del Comportamiento de la Universidad de Granada, como directora de la tesis presentada para aspirar al grado de doctora por Dña. M. Carmen Díaz Batanero

HACE CONSTAR:

Que la tesis, "*Viabilidad de la enseñanza de inferencia bayesiana en el análisis de datos en psicología"* realizada por M. Carmen Díaz Batanero, reúne las condiciones científicas y académicas necesarias para su presentación.

Fdo. E. Inmaculada de la Fuente Solana

Agradecimientos

Quisiera dedicar la finalización de esta Tesis Doctoral a todas aquellas personas que me han acompañado y facilitado su apoyo, consejo y ánimo a lo largo de este proceso, sin las cuales no hubiera sido posible lograr este objetivo.

En primer lugar agradecer a mi directora de tesis, no sólo por ofrecerme sus valiosos conocimientos y experiencia profesional, sino también por animarme y alentarme en cada una de las fases de la investigación.

Hago extensivos estos agradecimientos a los profesores del Departamento de Psicología Social y Metodología de las Ciencias del Comportamiento por facilitarme la integración en el Departamento y hacerme sentir una compañera más durante estos años de desarrollo de mi beca.

A todos y cada uno de los investigadores que han colaborado en los diversos paneles de expertos, por el interés que han puesto en esta investigación y las valiosas sugerencias aportadas.

Igualmente, a los alumnos de Psicología de las Universidades de Granada, Jaén, Murcia y Huelva por su colaboración en la recogida de datos, en especial al grupo de Primero de Psicología de la Universidad de Granada 2004-2005 por su valiosa colaboración. Gracias por vuestro entusiasmo y buena disposición.

Un recuerdo especial a mis padres por su eterna entrega y capacidad para mantener la ilusión por una meta alcanzable y porque han sido una indudable referencia y guía durante estos años. Espero continuar sus pasos.

Tampoco puedo dejar de agradecer a mis hermanos y cuñadas, porque aún desde la distancia he recibido incondicional apoyo y cariño. En este mundo global, las relaciones no se miden por la distancia física sino por la cercanía emocional.

A todos los amigos y compañeros que en algún momento han sufrido los efectos de escribir una tesis, por su apoyo y escucha en los momentos de estrés.

Por último, a ti Alberto, por tu incomparable mezcla de paciencia, comprensión, cariño y sentido del humor. Confío en poder acompañarte en tus proyectos futuros tal y como tu lo has hecho conmigo. Seguimos caminando juntos.

Desde estas páginas, un recuerdo muy especial para todos y todas.

¡Gracias!

A mis padres

VIABILIDAD DE LA ENSEÑANZA DE LA INFERENCIA BAYESIANA EN EL ANÁLISIS DE DATOS EN PSICOLOGÍA

ÍNDICE

INDICE DE TABLAS

INDICE DE FIGURAS

INTRODUCCIÓN

Hoy día se presenta una situación paradójica en la aplicación de la inferencia estadística, tanto en Psicología, como en otras ciencias empíricas: Mientras que el uso incorrecto o insuficiente de los métodos estadísticos se denuncia desde las sociedades profesionales, no se observan cambios ni en la metodología de investigación ni en la enseñanza de los métodos estadísticos dentro de los planes de estudio de grado y postgrado.

En esta Tesis nos hemos centrado en esta problemática, y más particularmente, en las posibilidades brindadas por la inferencia bayesiana, cuyo estudio abordamos desde diferentes puntos de vista:

1. La reflexión sobre las prácticas actuales relacionadas con los métodos estadísticos en el campo de la psicología, los errores denunciados y las posibles aportaciones de la inferencia bayesiana. Este análisis se aborda desde el punto de vista filosófico (diferentes filosofías de la inferencia) y psicológico (errores comunes y causas de los mismos.
2. El estudio de algunas de las posibilidades brindadas por los métodos bayesianos, dentro de la psicometría, al estimar los diferentes indicadores utilizados en la teoría clásica de los tests. Estas posibilidades se contemplan desde el punto de vista teórico y práctico, aplicándolo en el proceso de elaboración de un cuestionario de evaluación del razonamiento sobre probabilidad condicional (RPC), cuya construcción se justifica dentro de la Memoria.
3. La viabilidad de la enseñanza de estos métodos, al menos desde un punto de vista elemental, en los cursos de grado en la licenciatura de Psicología. A este efecto se diseña un material didáctico que tiene en cuenta los análisis anteriores, las investigaciones sobre didáctica de la estadística y las características de nuestros estudiantes. Este material es experimentado con una metodología de enseñanza en pequeños grupos (4 grupos con un total de 78 estudiantes), proporcionando datos del aprendizaje de los estudiantes al final de la experiencia.

La Memoria se organiza en 7 capítulos y 9 estudios interrelacionados:

El Capítulo 1 recoge los fundamentos sobre filosofía de la inferencia, aportaciones de los métodos bayesianos a la mejora de la práctica metodológica e investigaciones sobre comprensión de la probabilidad condicional condicional, finalizando con la justificación de la necesidad de construcción del cuestionario RPC.

El segundo Capítulo recoge los objetivos e hipótesis de la investigación, resumiendo sus etapas y la forma cómo se relacionan.

Seguidamente (Capítulo 3), se realiza una interpretación bayesiana de la Teoría Clásica de los Tests desde el punto de vista teórico y se analizan las implicaciones de esta interpretación sobre la estimación de diversas características de los ítems y globales.

El Capítulo 4 describe el proceso de construcción del cuestionario RPC, organizado en cuatro estudios: definición semántica de la variable (Estudio 1), selección de ítems para el instrumento (Estudios 2.1 y 2.2), prueba y revisión del instrumento piloto (Estudios 3 y 4).

En el Capítulo 5 se recogen las aportaciones a la fiabilidad y validez del instrumento (Estudios 5 y 6). Tanto en la construcción como en la validación del cuestionario se complementan los métodos clásicos con la aplicación de los métodos bayesianos analizados en el Capítulo 3.

Una vez finalizada la construcción de este instrumento, se presenta el diseño y validación del material para la enseñanza de inferencia bayesiana en primeros cursos de psicología. El material se apoya en el estudio de evaluación del razonamiento condicional de los estudiantes llevado a cabo con el cuestionario RPC (Estudio 7), así como en los fundamentos filosóficos, metodológicos y didácticos expuestos en el Capítulo 1. El diseño del material se recoge en el estudio 8.1 y su evaluación en el estudio 8.2. Finalmente, en el estudio 9 se muestra la interrelación de los resultados de evaluación del aprendizaje de la inferencia bayesiana y el razonamiento condicional, justificando de este modo la construcción previa del cuestionario RPC.

Pensamos que esta investigación realiza diferentes tipos de aportaciones a la Metodología de las Ciencias del Comportamiento:

- Análisis de la Teoría Clásica de los Tests desde el punto de vista bayesiano, señalando sus posibles aportaciones y poniendo a punto subrutinas de cálculo para algunas características psicométricas habituales.

- Aplicación de estos procedimientos en el proceso de construcción de un cuestionario y en el análisis de los resultados de la evaluación llevada a cabo con dicho instrumento.
- Síntesis de la problemática filosófica, psicológica y didáctica relacionada con la aplicación y enseñanza de la inferencia estadística y en particular de los métodos bayesianos.
- Puesta a punto de un material factible para la introducción de conceptos elementales de inferencia bayesiana a estudiantes de psicología y evaluación del aprendizaje logrado.
- Apoyo empírico a nuestra hipótesis sobre la importancia de la formación del estudiante sobre probabilidad condicional, que sugiere la conveniencia de reforzar la enseñanza de este tema.

En consecuencia, pensamos que esta tesis abre nuevas perspectivas de investigación en la Metodología de las Ciencias del Comportamiento, tanto desde el punto de vista estrictamente metodológico (la puesta a punto y aplicación de métodos de investigación), como desde el punto de vista didáctico. Resultados parciales de cada una de las etapas anteriores se han publicado en diversas revistas y congresos internacionales[1].

En el actual proceso de convergencia al Espacio Europeo de Educación Superior, no sólo es viable, sino necesario que los profesores del área de conocimiento realicen investigaciones sobre la enseñanza de los métodos de investigación, incluyendo temas no tradicionales. Sólo por medio de la investigación sistemática en la enseñanza se podrá mejorar la práctica docente y con ello contribuir a la mejora de la aplicación de los métodos en la investigación. Se espera, por tanto, nuevos estudios que continúen el iniciado en esta Memoria.

[1] Estas publicaciones se citan a lo largo de la memoria e incluyen trabajos en Metodología de las Ciencias del Comportamiento, Biaix, Educación Matemática, Epsilon, Suma, Uno, ICOTS.7, CIBEM V, 38 Journées de la Societé Française de Statistique, IX Simposio Nacional Colombiano de Estadística, Simposio de SEIEM y Congreso de Metodología de Encuestas.

CAPÍTULO 1.

FUNDAMENTOS

1. INTRODUCCIÓN

En este Capítulo se analizan los fundamentos de la investigación, que se organizan en la forma siguiente:

En primer lugar se describe la problemática del uso de la estadística en las ciencias empíricas, describiendo las soluciones aportadas por las diversas escuelas de inferencia estadística al problema de la inducción, los errores comunes en la aplicación del contraste de hipótesis y posibles causas psicológicas de los mismos, y se resume la actual controversia sobre la práctica estadística en psicología. A continuación se analizan la filosofía y características metodológicas del método bayesiano y sus posibles aportaciones para mejorar esta práctica. Todo ello nos permitirá argumentar la necesidad de introducir este tema en la formación estadística de los investigadores en el campo de psicología y de diseñar y evaluar procesos formativos sobre esta materia, lo que será el objetivo del Capítulo 6 de esta Memoria (Estudios 8 y 9).

El último bloque se dedica a mostrar la importancia del concepto de probabilidad condicional, cuya comprensión deficiente está en la base de muchos de los errores y aplicaciones incorrectas denunciadas. En consecuencia, se justifica la necesidad de llevar a cabo una evaluación del grado de comprensión y errores de los estudiantes de Psicología respecto a dicho concepto, con el fin de fundamentar el diseño de enseñanza llevado a cabo en el estudio 9. El análisis de la investigación sobre probabilidad condicional muestra también la necesidad de construir un instrumento de evaluación, objetivo que se abordará en los Capítulos 3 y 4 (Estudios 1 a 6).

2. PROBLEMÁTICA DEL USO DE LA ESTADÍSTICA EN LA INVESTIGACIÓN EMPÍRICA

En esta sección se justifica brevemente la importancia de la inferencia estadística en la investigación empírica y se comparan las características de las distintas metodologías actuales de inferencia. Seguidamente se describen los errores más frecuentes en la comprensión y aplicación de conceptos inferenciales y se analizan las críticas actuales al contraste de hipótesis, así como las sugerencias de sustituir este procedimiento por otros métodos estadísticos alternativos.

2.1. LA JUSTIFICACIÓN DEL RAZONAMIENTO INDUCTIVO

La problemática filosófica de la inferencia estadística se relaciona con la naturaleza incierta del conocimiento empírico y, por tanto, de las teorías científicas obtenidas a partir del mismo, es decir, con la dificultad de justificar el razonamiento inductivo y sus conclusiones (Black, 1979; Cabria, 1994; Vallecillos, 1994; Ares, 1999; Borges, San Luis, Sánchez, y Cañadas, 2001; de la Fuente y Díaz, 2004).

En el proceso de construcción de nuevo conocimiento[1] hay dos tipos de razonamientos básicos, que a veces se combinan entre sí:

- En el razonamiento *deductivo* la conclusión está contenida en las premisas, por lo que, si partimos de un conocimiento verdadero, obtenemos otro conocimiento también cierto.
- Por el contrario en el método *inductivo* la conclusión es más amplia que las premisas, siendo por tanto posible que la conclusión esté equivocada, aunque las premisas sean verdaderas.

El razonamiento inductivo es el más común en las ciencias empíricas y es difícil de justificar, lo que explica que, a veces, una teoría científica se admita como cierta durante un cierto periodo histórico y luego se abandone, cuando se encuentra evidencia en contra de la misma. La contestación a la pregunta de qué es el conocimiento y con qué metodo podemos lograrlo ha sido un reto a lo largo de los años, aunque la lógica de la investigación científica es relativamente reciente (Rivadulla, 1991; Cabriá, 1994). La búsqueda de criterios para dar validez al razonamiento inductivo es muy antigua, e incluso preocupó a Aristóteles, aunque Hume fue el primero que planteó con claridad el problema de *invalidez de la inducción:* la incompatibilidad entre el principio fundamental del empirismo (todas las teorías científicas debieran obtenerse a partir de la observación de la realidad) y el de invalidez de la inducción (las conclusiones obtenidas a partir de la inducción no tienen validez lógica). Si la deducción es válida a partir de premisas que se suponen ciertas, esto no ocurre en la inducción. Muchos filósofos han

[1] En toda comunidad científica hay un conjunto de proposiciones admitidas como verdaderas, que se amplia progresivamente, produciendo nuevo conocimiento mediante un proceso de "inferencia" o paso de proposiciones ya conocidas y admitidas como ciertas (premisas) a un nuevo conocimiento o enunciado (conclusión) (Rivadulla, 1991).

tratado de resolver este problema, sin llegar a una solución (Black, 1979; Vallecillos, 1994).

Un autor que ha tenido una fuerte influencia en el debate sobre el método inductivo es Popper, quien propuso que una cierta teoría puede racionalmente considerarse como cierta frente a otras con las que se halla en competencia, si, a pesar de nuestros intentos, no conseguimos refutarla. Popper (1967) sugirió poner a prueba las hipótesis científicas, mediante experimentos u observaciones y comparar los patrones deducidos de la teoría con los datos obtenidos. Si los datos apoyan nuestra teoría, ésta recibe una confirmación provisional, aunque los datos futuros podrían contradecirla. En cambio si los datos del experimento se apartan del patrón esperado, la teoría es refutada. Popper sugiere que el rechazo y la aceptación de las teorías no tienen el mismo estatuto lógico (Popper, 1967; Rivadulla, 1991; Cabriá, 1994).

Una objeción a la propuesta de Popper es que los datos nunca se ajustan exactamente a lo que las teorías predicen y por ello, nunca estamos seguros al "confirmar" las teorías, ya que los modelos teóricos son siempre aproximaciones de la realidad, que es demasiado compleja. La comunidad científica, será la que proponga un criterio para decidir cuando el resultado de un experimento confirma la hipótesis, la contradice o son ambiguos[2]. Otro problema es que para establecer una teoría científica habría que compararla con todas las hipótesis rivales plausibles, pero éstas no se suelen conocer de antemano (Vallecillos, 1994).

Soluciones estadísticas al problema de la inducción

Todas estas ideas de Popper tienen una gran influencia en el desarrollo de la inferencia estadística, en la que algunos autores tratan de apoyar el razonamiento inductivo, recurriendo a la probabilidad. Ya que no se puede llegar a la verdad cierta en un razonamiento inductivo, se trata de calcular la probabilidad de que sea cierta, es decir, la "verdad probable".

Popper diferencia la probabilidad de un suceso y la probabilidad de una hipótesis (Ridadulla, 1991; Batanero, 2000):

[2] Cabria (1994) diferencia entre convalidación y contrastación de datos. La convalidación confronta las conclusiones obtenidas de unos datos con las que se deducen de nuevos datos, mientras que la contrastación confronta los datos con teorías previas.

- La probabilidad de un suceso hace referencia a su frecuencia, porque el suceso unas veces ocurre y otras no.
- La probabilidad de una hipótesis no tiene sentido en inferencia clásica, porque la hipótesis es cierta o falsa. No es posible obtener la probabilidad objetiva de una hipótesis, pero sí podemos asignar a la hipótesis una probabilidad subjetiva o grado de creencia personal dentro del marco de la inferencia bayesiana (Gingerenzer, 1993; Lecoutre, 1999).

En este último caso podremos diferenciar dos usos del concepto de probabilidad de una hipótesis:

- Probabilidad inicial, antes de recoger datos en experimentos basados en la hipótesis.
- Probabilidad final, es decir, una probabilidad condicional, obtenida a partir de los datos empíricos.

A continuación analizamos los soluciones aportadas al problema de la inducción por Fisher, Neyman y Pearson y la escuela Bayesiana.

2.2. METODOLOGÍA DE LOS TESTS DE SIGNIFICACIÓN DE FISHER

Hay dos concepciones sobre los contrastes de hipótesis dentro de la estadística frecuencial. Aunque el procedimiento de cálculo sea muy similar, la diferencia se encuentra en el razonamiento subyacente (Rivadulla, 1991; Cabriá, 1994; Moore, 1995):

- La concepción de Fisher, basada en los tests de significación.
- La concepción de Neyman y Pearson, donde se entiende el contraste como una regla de decisión entre dos hipótesis.

Un *test de significación* es para Fisher un procedimiento que permite rechazar una hipótesis, con un cierto nivel de significación[3]. Se trata de comprobar si una afirmación sobre una propiedad de la población es apoyada o no por la información obtenida en una muestra de dicha población. Fisher, en su libro *The design of experiments,* publicado en

[3] Fisher consideró los tests de significación como la forma más general de inferencia, que no requerían una justificación probabilística.

1935 introduce su teoría de las pruebas de significación, que se centra sobre la hipótesis nula. Para él, la prueba de significación evalúa la *"fuerza de la evidencia"* en contra de dicha hipótesis. En resumen, se aplica el siguiente razonamiento (Fisher, 1979/1956):

Se quiere comprobar si una cierta hipótesis H_o (hipótesis nula) es o no cierta. Se escoge un cierto suceso *S* de un experimento aleatorio asociado con H_o, del cual se sabe que, si H_o es cierta, hay una probabilidad *p* muy pequeña de que se verifique (menor que un cierto *nivel de significación* que convencionalmente se establece como 0'05). Realizado el experimento ocurre precisamente *S.* Hay dos posibles conclusiones:

- Bien la hipótesis H_o era cierta y ha ocurrido *S*, a pesar de su baja probabilidad.
- Bien la hipótesis H_o era falsa.

Generalmente el experimento consiste en tomar una muestra de la población sobre la que se realiza el contraste y calcular un estadístico, cuya distribución depende del valor supuesto del parámetro implicado. El estadístico establece una medida de discrepancia entre los datos y la hipótesis y, en caso de que se cumpla la hipótesis, esta medida de discrepancia define una distribución, al variar los datos aleatoriamente (Cabriá, 1994; Batanero, 2000). Un test de significación efectúa una división entre los posibles valores de este estadístico en dos clases: resultados estadísticamente significativos[4] (para los cuales se rechaza la hipótesis) y no estadísticamente significativos (Benzecri, 1983; Ridavulla, 1991; Moore, 1997a).

Aunque, aparentemente este procedimiento parece de tipo inductivo, el razonamiento es deductivo. Más precisamente, se trata de un razonamiento a partir de una distribución conocida (la del estadístico en todas las posibles muestras de la población) hacia uno de sus valores (la muestra que tomamos para el experimento). La distribución en el muestreo es conocida y a partir de ella se calcula la probabilidad del valor particular obtenido, lo cual es un razonamiento deductivo[5].

[4] La idea clave es que el suceso es inusual dada la hipótesis (Lindley, 1993)

[5] El test de significación de Fisher es un procedimiento deductivo, aunque esto puede ser difícil de comprender y hasta el mismo Fisher estuvo confuso sobre ello en el comienzo de sus estudios (Rivadulla, 1991; Vallecillos, 1994).

Los aspectos diferenciales de la lógica del contraste según Fisher son:

- El investigador determina lo que es un "suceso improbable" según su juicio subjetivo y su experiencia.
- El objetivo del test de significación es falsar la hipótesis.
- No nos preocupamos por una hipótesis alternativa concreta o el error asociado a la misma (De la Fuente y Díaz, 2003).

2.3. LOS CONTRASTES DE HIPÓTESIS DE NEYMAN Y PEARSON

Neyman y Pearson conceptualizan el contraste de hipótesis como un proceso de decisión que permite elegir entre una hipótesis dada H_0 y otra hipótesis alternativa H_1 (Benzecri, 1983; Valera, Sánchez y Marín, 2000). Se contemplan dos posibles decisiones respecto a H_0: rechazar esta hipótesis, asumiendo que es falsa y aceptando la alternativa, o abstenerse de esa acción. Al tomar una de estas decisiones sobre las hipótesis a partir de los resultados del contraste se pueden cometer dos tipos de error (Ríos, 1967; Nortes Checa, 1993; Moore, 1997a; Peña y Romo, 1997):

- *Error tipo I:* Rechazar una hipótesis nula que de hecho sea verdadera. Este es el error que, desde el punto de vista estadístico, se ha considerado más grave. Para evitarlo, se suele establecer un criterio de prueba que asegura que la probabilidad de cometer este tipo de error sea menor que un número α preestablecido o *nivel de significación*.
- *Error tipo II:* aceptar la hipótesis nula que de hecho es falsa. Beta es la probabilidad de cometer este tipo de error y el complemento de beta $(1 - \beta)$ sería la *potencia* del contraste. Mientras que α es un número preestablecido, β es variable, porque su valor depende de cual es el valor del parámetro (generalmente desconocido).

Una vez definidas las hipótesis nula y alternativa y fijada la probabilidad de cometer error tipo I, se elige el contraste de mayor potencia (Ríos, 1967; Nortes Checa, 1993). Calculado el estadístico, se toma la decisión de rechazar o no rechazar la hipótesis nula, comparando el valor-p con el nivel de significación o, equivalentemente, comparando el valor del estadístico calculado con el valor crítico (Vallecillos, 1994).

Los conceptos diferenciales de esta concepción del contraste de hipótesis son (De la Fuente y Díaz, 2003):

- El contraste proporciona un criterio para decidir entre una de las dos hipótesis.
- Se reconocen los errores tipo II.
- Se da una interpretación a la hipótesis alternativa.

Varias son las dificultades relacionadas con la idea de contraste de hipótesis (Vallecillos, 1994, 1996, 1999; Vallecillos y Batanero, 1996; Batanero y Díaz, 2006):

- La gran cantidad de conceptos matemáticos implicados y sus relaciones complejas (hipótesis nula/ alternativa, estadístico/ parámetro, región de aceptación/ rechazo, error tipo I y II, potencia/ nivel de significación, distribución en la población y distribución muestral, etc).
- La existencia de riesgo en la decisión tomada, puesto que siempre se puede cometer alguno de los tipos de error;
- Las probabilidades de error son *probabilidades iniciales* y no *finales*. Por tanto, no es posible calcular inductivamente la probabilidad de la hipótesis a partir de los datos, ya que el procedimiento de Neyman y Pearson es un procedimiento deductivo. Los autores niegan la posibilidad de usar el concepto de credibilidad de una hipótesis en función de los datos en el marco frecuentista (Rivadulla, 1991; Cabriá, 1994; Lecoutre, 1996, 1999).
- La relación entre las probabilidades de error que hace que, al aumentar una disminuya la otra, si se mantiene constante el tamaño de la muestra.

2.4. INFERENCIA BAYESIANA

El primer tratamiento matemático de la inferencia inductiva la da *el teorema de Bayes*, que permite calcular las *probabilidades finales*, a partir del conocimiento de las *probabilidades iniciales* y de los datos obtenidos experimentalmente (Rivadulla, 1991: Bolstad, 2004). En su forma más simple, este teorema se expresa en la forma siguiente (Serrano, 2003):

Tenemos un suceso *B* (los datos) y queremos saber si ha sido producido por una de las causas A_1 A_2, ... A_n (una serie de hipótesis científicas rivales; son las posibles causas de *B*). Se conocen las probabilidades $P(A_1)$, $P(A_2)$, ... $P(A_n)$, es decir la probabilidad inicial de cada una de las hipótesis rivales, así como las probabilidades $P(B/A_1)$, $P(B/A_2)$, $P(B/A_n)$ o verosimilitud de obtener los datos *B* dependiendo de cual de las hipótesis es cierta. Entonces, la probabilidad $P(A_i /B)$ (probabilidad final de que la hipótesis *Ai* sea la verdadera, una vez que hemos obtenido los datos *B)* viene dada por la siguiente expresión:

$$P(Ai / B) = \frac{P(B / Ai)P(Ai)}{\sum_{i-1}^{n} P(B / Ai)P(Ai)}$$

La escuela Bayesiana[6] postula que el teorema de Bayes (y toda la inferencia bayesiana que se obtiene generalizándolo y desarrollándolo) es un instrumento adecuado para obtener un conocimiento inductivo, pues las probabilidades iniciales pueden ser transformadas en probabilidades finales a la luz de los sucesos observados (Box y Tiao, 1992; Berry, 1995; Bernard, 1998).

Comparación de la inferencia clásica y Bayesiana

El teorema de Bayes contiene la esencia del método inductivo, porque podría aplicarse sucesivamente en nuevos experimentos, tomando en el segundo experimento como probabilidades iniciales las probabilidades finales que se obtuvieron en el primer experimento y así sucesivamente (Benzecri, 1983; Rivadulla, 1991; Cabriá, 1994; Rindskopf, 1997)[7].

La inferencia frecuencial al tratar de estimar un parámetro θ a partir de unos datos *D,* considera este parámetro fijo pero desconocido y sólo usa y da sentido a las probabilidades iniciales $P(D/\theta)$ que se interpretan en forma frecuencial (lo que ocurriría si se repitiese el experimento indefinidamente). Dentro de esta tendencia, Rouanet (1998a) diferencia entre frecuentismo radical y moderado.

- En el frecuentismo radical, liderado por Neyman y Pearson, se hace una interpretación literal de la probabilidad en términos de "muestreo repetido en una

[6] Este enfoque se va introduciendo gradualmente en Psicología (De la Fuente, García y de la Fuente, 2002).

[7] Se habla así de "proceso de aprendizaje" (Box y Tiao, 1992).

misma población"[8]. El contraste de hipótesis no pretende evaluar la probabilidad de una hipótesis, sino rechazarla o no. Según este enfoque el nivel de significación ha de elegirse antes de hacer el test, de acuerdo con un "razonamiento deductivo".

- En el frecuentismo moderado, siguiendo a Fisher, se busca la evaluación de la evidencia en favor de las hipótesis y no se considera razonable que un investigador mantenga siempre el mismo nivel de significación fijo en todos sus contrastes; aunque esto es lo que hoy día se hace en la práctica.

La inferencia bayesiana, por el contrario, considera θ como aleatorio y lo asocia a una distribución de probabilidades $p(\theta)$ epistémicas que conjuga con los datos a partir de las verosimilitudes para hallar las probabilidades finales P(θ/D), "*las inferencias bayesianas contienen su propia medida interna de la situación relacionada únicamente con la situación actual, sin necesidad de acudir a repeticiones de esa situacion mediante distribuciones en el muestreo"* (Cabría, 1994, p. 216).

Rouanet (1998a) diferencia también entre bayesianos moderados y radicales:

- Los radicales, como por ejemplo Savage, insisten en el carácter subjetivo de la probabilidad. Esta tendencia ha dominado mucho tiempo y eso hizo que se confunda "estadística bayesiana" y "estadística subjetiva", provocando el rechazo de algunos investigadores al método bayesiano.
- Los moderados, como Jeffreys, tratan de diseñar técnicas bayesianas "objetivas", investigando distribuciones iniciales que hoy se conocen como "no informativas"[9], por ejemplo distribuciones conjugadas, de modo que tanto la distribución inicial como final pertenezcan a la misma familia de funciones (Cabriá, 1994; Lecoutre, 2006).

Una crítica al método bayesiano es que, si no se tiene un conocimiento del fenómeno, no es posible asignar las probabilidades iniciales y por tanto no es posible aplicar el teorema de Bayes. Este problema se suele soslayar, suponiendo que, en

---

[8] Esto asegura que a la larga la proporción de hipótesis nulas ciertas rechazadas sea igual al nivel de significación.

[9] Es importante resaltar que nunca estamos en estado de "ignorancia completa", sino que distribución a priori no informativa representa el conocimiento del parámetro antes que un experimento particular se lleve a cabo (Box y Tiao, 1992, p. 25).

ausencia de información inicial, las probabilidades iniciales tienen una distribución uniforme, lo cual es considerado por algunos autores poco correcto, porque en muchas situaciones no conocemos la distribución inicial, pero sabemos claramente que no es uniforme (Rivadulla, 1991).

Otra posibilidad sería asignar subjetivamente las probabilidades iniciales, pero esto llevaría a que dos investigadores con los mismos datos obtuviesen unas probabilidades finales diferentes (Pruzek, 1997; McLean, 2001). Sin embargo la aplicación sucesiva del teorema de Bayes llevaría a ir corrigiendo poco a poco, en los nuevos experimentos el posible sesgo introducido en la asignación inicial de probabilidades (Box y Tiao, 1992; Lindley, 1993). Esto lleva a otro problema filosófico, porque esto es admitir la existencia de probabilidad "inicial" objetivas hacia las cuales se irían acercando los cálculos y esto es precisamente contrario a la filosofía bayesiana radical, donde todas la probabilidades son subjetivas, aunque es admisible en el bayesianismo moderado (Rivadulla, 1991).

En definitiva, el método de Bayes no permite calcular ni revisar "probabilidades iniciales/ objetivas de la hipótesis" sino las "probabilidades iniciales subjetivas" que estableció cada investigador[10]. En este sentido subjetivo la inferencia bayesiana apoya el método inductivo, modificado en el sentido de buscar, no la veracidad de una hipótesis (probabilidad objetiva), sino nuestra creencia en la misma (probabilidad subjetiva) y traslada el problema de justificación de la inducción al debate sobre el carácter objetivo/ subjetivo de la probabilidad (Rivadulla, 1991).

2.5. VEROSIMILITUD Y METODO FIDUCIAL

Los enfoques frecuentistas están basados en la idea de muestreo aleatorio, cuyo ejemplo más simple es la extracción de fichas de dos colores en una urna. Cuando conocemos la proporción p de fichas de uno de los colores en la urna, el cálculo de probabilidades nos permite deducir la distribución en el muestreo de cualquier estadístico asociado a una muestra aleatoria de bolas; y la probabilidad de la proporción muestral viene dada por $p^{k}(1-p)^{n-k}$.

[10] Es decir su grado personal de creencia en la veracidad de la hipótesis.

Rouanet (1998a) nos sugiere considerar el caso en que p es desconocido y hemos obtenido una muestra de bolas de tamaño n. Estamos en una situación de inferencia y el modelo probabilístico define para cada posible valor de p una distribución muestral. La expresión $L(p)=p^{k}(1-p)^{n-k}$ considerada como función de p nos da la *verosimilitud* de cada posible valor de p para la muestra observada. Buscando el máximo de esta función, encontramos el valor $p=k/n$[11]. Es decir, podríamos tomar como estimador del parámetro aquél valor, entre los posibles, que nos de la máxima verosimilitud para la muestra particular. La función de verosimilitud proporciona una solución al problema de la inferencia inductiva, pero no nos da la distribución de probabilidad para los posibles valores de p (Rivadulla, 1991). Sin embargo tiene en cuenta sólo los valores ocurridos, no los posibles, por tanto no transfiere ningún elemento de probabilidad (Cabriá, 2004).

Fisher, que siempre se preocupó por mantener una visión frecuencial de la probabilidad desarrolló el método llamado "fiducial" basado en la función de verosimilitud. Este método trata de evitar las probabilidades iniciales de las hipótesis (como la estadística frecuencial), pero produce probabilidades finales de las hipótesis, dados los datos (como la inferencia bayesiana). Puesto que en algunos casos las distribuciones fiduciales de Fisher coinciden con las distribuciones bayesianas finales, de modo que se podría considerar que Fisher fue un bayesiano moderado (Rivadulla, 1991).

2.6. LA LÓGICA HÍBRIDA Y PRINCIPALES ERRORES EN EL USO Y APLICACIÓN DE LA INFERENCIA

A menudo cuando se enseña o se usa la inferencia estadística se mezclan conceptos e interpretaciones de Fisher y Neyman - Pearson, incluyendo a veces también conceptos relacionados con la inferencia bayesiana (Gingerenzer et al., 1989; Gingerenzer, 1993; Valera, Sánchez y Marín, 2000). Fisher y Neyman – Pearson tuvieron diferentes interpretaciones de los contrastes estadísticos, incluyendo la forma en que se deben determinar los niveles de significación y la interpretación de un resultado significativo.

Neyman y Pearson dieron una interpretación frecuencial a los tests (en el contexto de muestreo repetido), mientras que Fisher dio una interpretación epistémica (referido a

[11] Una extensión de esta idea es usar la razón de verosimilitudes $L(p_1)/L(p_2)$ para medir el apoyo relativo que dan los

un solo experimento) al resultado significativo (Gigerenzer, 1993; Falk y Greenbaum, 1995). Mientras que Neyman y Pearson piden que el nivel de significación se fije antes del experimento y se mantenga constante para una serie de tests, en la filosofía de Fisher el investigador puede fijar un nivel de significación diferente en distintas ocasiones o incluso no fijarlo.

La disputa entre estos autores se ha ocultado en las aplicaciones de la inferencia estadística en Psicología y otras ciencias experimentales, donde se ha asumido una única solución para la inferencia, que es la que se aplica actualmente (Oakes, 1986; Chow, 1996)[12]. Mientras que Neyman y Pearson idearon el contraste de hipótesis como un proceso de decisión entre dos hipótesis, a veces se emplea este procedimiento simplemente para proporcionar evidencia en contra de una única hipótesis (es decir, no se especifica bien la alternativa), lo que es más propio de la filosofía de Fisher. Otras características tomadas de Neyman – Pearson son que H_0 es la hipótesis de no diferencia, y que el nivel de significación α debe escogerse antes de analizar los datos y debe mantenerse constante.

De Fisher conservamos la sugerencia de que la inferencia se basa en una probabilidad condicional; la probabilidad de obtener los datos supuesta cierta H_0, y que H_0 y H_1 son mutuamente exclusivas y complementarias (Chow, 1996). Deberíamos añadir que algunos investigadores suelen dar una interpretación bayesiana a los resultados de los contrastes de hipótesis (clásicos), a pesar de que el enfoque de la estadística Bayesiana es muy diferente de las teorías tanto de Fisher como de Neyman y Pearson (Gingerenzer et al., 1989; Gingerenzer, 1993; Lecoutre, 1999, 2006). Esta mezcla genera muchos errores que describimos en lo que sigue.

2.6.1. CONCEPCIONES SOBRE EL NIVEL DE SIGNIFICACIÓN Y EL VALOR-P

Sobre estos dos conceptos se han descrito muchas interpretaciones erróneas, que podemos clasificar en los siguientes apartados (Batanero, 2000):

---

datos al valor p_1 en relación con el valor p_2 (Cabriá, 1994).

12 Esta estrategia mixta es denominada marco ortodoxo (Oakes, 1986) o lógica híbrida (Gigerenzer, 1993).y es transmitida o al menos inducida en algunos libros de texto que no separan claramente las metodologías de Fisher y Neyman-Pearson.

Intercambio de los términos en la probabilidad condicional

Uno de los errores que se encuentran a menudo entre estudiantes e investigadores concierne al nivel de significación α (probabilidad de rechazar la hipótesis nula siendo cierta). Concretamente, se cambian los términos de la probabilidad condicional, interpretándose como la probabilidad de que la hipótesis nula sea cierta, habiendo tomado la decisión de rechazarla (Falk, 1986a; Cohen, 1994; Falk y Greenbaum, 1995; Chow, 1996; Ares, 1999). Birnbaum (1982), por ejemplo, informó que sus estudiantes encontraban razonable la siguiente definición: "*Un nivel de significación del 5% indica que, en promedio, 5 de cada 100 veces que rechacemos la hipótesis nula estaremos equivocados"*. [13]

Vallecillos (1994) planteó los siguientes ítems a una muestra de 436 estudiantes universitarios de diferentes especialidades (estadística, medicina, psicología, ingeniería y empresariales) que habían estudiado el tema:

> Ítem 1: Un nivel de significación del 5% significa que, en promedio 5 de cada 100 veces que rechacemos la hipótesis nula estaremos equivocados (verdadero /falso). Justifica tu respuesta.
>
> Ítem 2: Un nivel de significación del 5% significa que, en promedio, 5 de cada 100 veces que la hipótesis nula es cierta la rechazaremos (verdadero / falso). Justifica tu respuesta.

En el ítem 2 se presenta una interpretación frecuencial del nivel de significación (y es correcto), mientras que en el ítem 1 se han intercambiado los dos sucesos que definen la probabilidad condicional (y es incorrecto). Sin embargo, sólo el 32% de los estudiantes de la investigación de Vallecillos (1994) dio una respuesta correcta al ítem 1 y el 54% dio una respuesta correcta al ítem 2. De 135 estudiantes que justificaron su respuesta, el 41% dio un argumento correcto en los dos ítems. Un error común en todos los grupos de estudiantes fue el intercambio de los términos de la probabilidad condicional, juzgando, por tanto correcto, el ítem 1 y falso el ítem 2. Vallecillos y Batanero (1996) vieron que el hecho de intercambiar los términos de la probabilidad

[13] Resultados semejantes se encuentran en estudios hechos a investigadores (Pollard y Richardson, 1987; Lecoutre Lecoutre y Poitevineau, 2001).

condicional se daba incluso en estudiantes que eran capaces de discriminar entre una probabilidad condicional y su inversa.

Cohen (1994) indica que la interpretación errónea del valor-p como probabilidad de que la hipótesis nula sea falsa es casi universal y que, además, esta interpretación viene acompañada del error de suponer que su complementario es la probabilidad de que la siguiente replicación del experimento tendrá éxito. La explicación dada por Cohen (1994) es que el test de significación *"no nos dice lo que queremos saber, y queremos saber tanto lo que queremos saber, que, en nuestra desesperación, creemos que lo dice. Lo que queremos saber es: Dado estos datos, ¿cuál es la probabilidad de que H_0 sea cierta? Pero, como la mayoría sabemos, lo que nos dice es: Dado que H_0 es cierta, ¿cual es la probabilidad de estos datos (o más extremos)?*" (p. 997).

Interpretación de resultados significativos

La interpretación incorrecta del nivel de significación se une normalmente a la confusión entre significación estadística y significación práctica (Pollard y Richardson, 1987; Abelson, 1997; Lecoutre, Lecoutre y Poitevineau, 2001). Un resultado significativo implica para Fisher que los datos proporcionan evidencia en contra de la hipótesis nula, mientras que para Neyman y Pearson sólo establece la frecuencia relativa de veces que rechazaríamos la hipótesis nula cierta a la larga (Error Tipo I). La significación práctica implica significación estadística más un efecto experimental suficientemente elevado (Gigerenzer, 1993; Cohen, 1994). Sin embargo, podemos obtener una significación estadística con un pequeño efecto experimental, siempre que tomemos una muestra grande, lo que no suele ser percibido por los investigadores, como muestra Lecoutre (1999) en su análisis de las prácticas y actitudes de los investigadores en psicología hacia el contraste de hipótesis.

Otra interpretación errónea del valor-p es pensar que este valor es la probabilidad de que el resultado se deba al azar. Pero cuando en el contraste de hipótesis rechazamos la hipótesis nula, no podemos inferir la existencia de una causa particular en el experimento a partir de un resultado (Batanero, 2000; Batanero y Díaz, 2006)[14].

[14] Por ejemplo, si la diferencia entre un grupo experimental y otro control es significativa, la diferencia puede ser debida a un tratamiento particular, pero también puede ser que el grupo experimental estuviese formado por sujetos más cualificados. El diseño experimental trata por eso de controlar las variables extrañas.

Similarmente, uno de los errores más extendidos es pensar que el rechazo de la hipótesis nula proporciona un apoyo experimental a la hipótesis alternativa y en particular al efecto de la variable manipulada por el investigador (Birnmaum, 1982; Falk, 1986a; Granaas, 2002). Es un error de lógica elemental, puesto que si *A* implica no *B*, no es cierto que *B* implique no *A* (Cohen, 1994).

Comparaciones múltiples

Moses (1992) advierte de otro error que consiste en creer en la conservación del valor del nivel de significación cuando se realizan contrastes consecutivos en el mismo conjunto de datos. El significado del nivel de significación es que, si llevamos a cabo 100 comparaciones sobre el mismo conjunto de datos y usamos en todos ellos el nivel de significación 0'05, habrá que esperar que 5 de las 100 pruebas sean significativas por azar, incluso cuando la hipótesis nula en cada una sea cierta. Esto dificulta la interpretación de los resultados significativos en el caso que se hayan hecho muchas comparaciones.

White (1980) cita como ejemplos: comparar todos los pares de medias con el test de la t (en lugar de emplear el análisis de varianza) y usar repetidamente el análisis de varianza de un factor para estimar el efecto de un conjunto de factores (en vez del análisis factorial de varianza).

Elección de valores particulares para el nivel de significación

Asimismo se piensa que hay una justificación matemática para dar un valor de niveles de significación 0'05 y 0'01. Esto es un error, ya que podemos tomar cualquier valor arbitrario y aunque se recomienda anotar el valor exacto del valor-p, los valores anteriores se usan casi de forma universal. Skipper, Guenter y Nass (1970) sugirieron que esto trae como consecuencia la diferenciación de los resultados de investigación, provocando a menudo que un trabajo que no se ajuste a esos valores no se publique y llaman la atención sobre las posibles implicaciones en los problemas investigados. A veces, si la potencia del contraste es baja y el error Tipo II es importante, sería preferible una probabilidad mayor de Error Tipo I, pero los investigadores no son conscientes de ello.

Creencia en la replicación

Una consecuencia de la interpretación incorrecta del nivel de significación como probabilidad final de la hipótesis (una interpretación bayesiana) es la creencia en la replicación de los resultados que ya fue descrita por Kahneman, Slovic y Tversky (1982) cuando hablan de la "creencia en la ley de los pequeños números"[15]. También se asume incorrectamente que a mayor significación estadística, los resultados son más replicables (Pascual, García y Frías, 2000; Gingerenzer, 2003).

2.6.2. CONFUSIÓN SOBRE LOS DIVERSOS NIVELES DE HIPÓTESIS Y EL PLANTEAMIENTO DE HIPÓTESIS

Investigaciones como la de Vallecillos (1994, 1996) y libros como el de Chow (1996) sugieren que también se confunden los papeles de las hipótesis nula y alternativa, e incluso la hipótesis estadística alternativa y la hipótesis de investigación. Chow (1996) diferencia diversas hipótesis implicadas en los distintos niveles de abstracción de la investigación experimental orientada a la confirmación de teorías:

- *Hipótesis substantiva*: es la explicación teórica que nos planteamos acerca del fenómeno de estudio. Suele hacer referencia a un constructo teórico inobservable (inteligencia, actitud, etc.), por lo que su estudio directo no es posible. Para poder investigar la hipótesis substantiva debemos deducir algunas implicaciones observables de la misma.
- *Hipótesis de investigación*: es una deducción observable de la hipótesis substantiva (por ejemplo, el rendimiento de los estudiantes se puede deducir de su inteligencia). Tratamos de encontrar apoyo para la hipótesis sustantiva a través de la hipótesis de investigación.
- *Hipótesis experimental*: En muchas situaciones la hipótesis de investigación es todavía demasiado ambigua y necesitamos especificarla mejor para poder estudiarla. Así llegamos a la hipótesis experimental, que se define en términos de variables independientes y dependientes bien definidas[16].

[15] Consistente en pensar que las características de una población se han de reproducir incluso en una muestra pequeña y tener una confianza excesiva en los primeros resultados.

[16] Por ejemplo, comparar las puntuaciones en tests de inteligencia en diferentes grupos de sujetos.

- *Hipótesis estadística alternativa*: es una consecuencia de la hipótesis experimental a nivel estadístico. Hace referencia a una población de sujetos, descrita mediante un modelo matemático que se especifica por uno o varios parámetros[17]
- *Hipótesis estadística nula*: el complemento lógico de la hipótesis alternativa. En estadística trabajamos con la hipótesis nula de no diferencia. Que especifica que no hay diferencias entre los grupos control y experimental.

Aunque la teoría estadística se ocupa del último nivel de hipótesis, alumnos e investigadores confunden todos estos niveles y cuando encuentran un resultado significativo, lo interpretan en relación a la hipótesis de investigación o incluso a la substantiva (aunque el resultado solo se refiere a las hipótesis nula y alternativa). Un resultado significativo siempre depende del instrumento particular y el resto de los componentes del proceso inferencial, del cual es responsable el marco teórico de la investigación (Chow, 1996, Batanero, 2000).

Otra práctica errónea es fijar las hipótesis después de recoger los datos. Esto no sería problema si el contraste se realizase en una nueva muestra, pero si la hipótesis se establece a posteriori el nivel de significación real podría no corresponder al nivel fijado[18].

2.6.3. TAMAÑO MUESTRAL Y LA HEURÍSTICA DE REPRESENTATIVIDAD

Otros errores relacionados con conceptos inferenciales se han descrito en la investigación sobre heurísticas y sesgos llevada a cabo por Kahneman, Tversky y sus colaboradores[19]. La teoría estadística nos indica que el valor esperado del estadístico de la muestra es el del parámetro en la población y no depende del tamaño de la muestra, aunque la varianza de la muestra cambia proporcionalmente a su tamaño, cambiando entonces las probabilidades de los sucesos. La comprensión de la inferencia implica el

---

[17] Por ejemplo, coeficientes de inteligencia en alumnos de una cierta edad, que se supone siguen una cierta distribución normal, definida por su media y desviación típica.

[18] Ya que, por ejemplo, en 5% de las muestras tomadas de la población se rechazaría la hipótesis, de ser cierta, es posible que nuestra muestra particular sea una entre este 5 por ciento (Selvin, 1970).

[19] Las heurísticas se describen como *"mecanismos por los que reducimos la incertidumbre que produce nuestra limitación para enfrentarnos con la complejidad de estímulos ambientales"* (Pérez Echeverría, 1990, p. 51).

equilibrio adecuado entre las ideas de *representatividad y la variabilidad muestral* (Moses, 1992).

En la heurística de la representatividad (Kahneman, Slovic y Tversky, 1982) se enfatiza sólo la representatividad que sirve de base al cálculo de probabilidades de sucesos. Se prescinde del tamaño de la muestra, y de la variabilidad del muestreo, produciéndose una confianza indebida en las pequeñas muestras. Estos autores hablan de la existencia de una "ley de los pequeños números", en la que pequeñas muestras serían representativas en todas sus características estadísticas de las poblaciones de donde proceden. Este error puede tener importantes consecuencias de cara a la investigación experimental, ya que los científicos que creen en la "ley de los pequeños números" sobreestiman la potencia de sus métodos estadísticos, subestiman la amplitud de sus intervalos de confianza y tienen unas expectativas injustificadas en la replicabilidad de experimentos realizados con pequeñas muestras (Tversky y Kahneman, 1982d).

Tampoco se aprecia la relación entre tamaño muestral, efecto y significación estadística (Monterde-Bort, Pascual, Frías, en prensa). Así, se supone más fiable para un mismo valor del nivel de significación un estudio realizado con una muestra grande, cuando de hecho, el tamaño del efecto ha de ser mayor para producir la misma significación en un estudio pequeño (Abelson, 1997)[20].

En algunos casos no se diferencian los diferentes niveles de concreción de un mismo concepto. En la estadística descriptiva la unidad de análisis es una observación (una persona, un objeto) y se calcula la media de una muestra de tales objetos. En inferencia, el interés se centra en la media teórica o esperanza matemática de la población de la que ha sido tomada la muestra dada. Se considera tal muestra como una observación de otra población diferente, la población de todas las posibles muestras de tamaño similar al dado, que podrían extraerse de la población de referencia[21]. Esto supone una gran dificultad conceptual y muchos estudiantes e investigadores confunden las diferentes variables implicadas y sus medias (Schuyten, 1991).

[20] Pero en pequeñas muestras una relación importante puede no ser significativa y cualquier relación insignificante puede resultar significativa si se aumenta suficientemente la muestra (Kish, 1970).

[21] Se cambia, en consecuencia, la unidad de análisis, que es ahora la muestra, y consideramos la media de la muestra como variable aleatoria, estudiando su media y distribución.

2.6.4. OTROS ERRORES

Entre otros muchos posibles errores posibles White (1980) describe los debidos a una incorrecta identificación de la población en estudio, elección de tamaño insuficiente de la muestra, interpretar resultados de muestras no aleatorias como si el muestreo hubiese sido aleatorio e incorrecta asignación de sujetos a tratamientos en diseños aleatorizados. Todos estos errores inciden en los resultados e incorrecta interpretación de la estadística (Wilkinson, 1999).

Un ejemplo típico es no especificar con claridad la población objeto de estudio. En los estudios descriptivos no se precisa el uso de la inferencia y en los casos prácticos de muestreo aleatorio repetido (como en control de calidad) la pertinencia de la inferencia es clara. Sin embargo lo habitual es tomar una única muestra y generalizar a una población (universo hipotético) que no está claramente definida; se trataría de la población que se obtendría al repetir ilimitadamente la investigación en las mismas condiciones temporales, culturales, sociales, y cognitivas (Hagod, 1970).

El uso de un contraste sin ninguna información adicional es poco menos que inútil, como por ejemplo, dar únicamente la lista de coeficientes de correlación significativos marcados con asteriscos, sin indicar el valor del coeficiente, aunque esta es una práctica frecuente (Abelson, 1997) y es innecesaria en el caso de estudios descriptivos (Lipset, Trow y Coleman, 1970).

Por otro lado, en muchas áreas de investigación, bien debido a la dificultad de tomar muestras aleatorias o al tipo de variables analizadas es muy difícil alcanzar las condiciones exigidas para aplicar los contrastes en forma correcta (Selvin, 1970; Borges y Sanchez Bruno, 2004). Por ejemplo, al aplicar un test de Chi cuadrado para tratar de estudiar la posible asociación entre dos variables en una tabla de contingencia, los supuestos podrían ser violados, bien por no modificar las hipótesis previas a la luz de los datos bien por haber reagrupado convenientemente filas o columnas en la tabla para que aparezca una cierta asociación, o por realizar la prueba en tablas binarias cuando se tiene un conjunto de datos multivariante, en que una relación podría implicar más de una de dichas variables (Lipset, Trow y Coleman, 1970). Otra dificultad al aplicar los contrastes de hipótesis en las ciencias humanas es el control experimental. Puesto que no es posible controlar todas las variables relevantes, la estadística sugiere la necesidad de aleatorización, que garantiza la existencia de técnicas para medir la probabilidad de que un efecto haya sido producido aleatoriamente[22].

2.7. FACTORES PSICOLÓGICOS QUE CONTRIBUYEN A LA PREVALENCIA DE ERRORES

La prevalencia de los errores descritos se ha justificado con base en diferentes mecanismos psicológicos. Una primera explicación del comportamiento mecánico en los procedimientos de contraste de hipótesis[23], es la creencia infundada en que la inferencia estadística da una solución algorítmica al problema de la inferencia inductiva (Oakes, 1986; Gingerenzer et al., 1989; Falk y Greenbaum, 1995).

Gingerenzer (1993) usa una metáfora tomada de la teoría psicoanalítica para analizar la situación actual en la práctica de los tests de hipótesis. Compara las características de Neyman-Pearson en dicha práctica con el *superego* del razonamiento estadístico, que prescribe lo que debería hacerse y no da libertad a los investigadores (especificando hipótesis, nivel de significación y potencia antes de recoger los datos; e interpretando la probabilidad de error en el contexto de muestreo repetido). Los componentes tomados de Fisher se comparan al *ego* del razonamiento estadístico, conveniente para los investigadores, que quieren llevar a cabo su investigación y publicar sus trabajos, incluso a costa de determinar el nivel de significación después del experimento, establecer una hipótesis alternativa difusa o no establecerla antes de coger los datos, e interpretar la probabilidad de error como la probabilidad de error en un experimento particular. El tercer componente en el comportamiento del investigador descrito por Gingerenzer (1993) es el deseo bayesiano de asignar probabilidades a las hipótesis con base en los datos de investigación (el *id* de la lógica híbrida). Cuando encontramos un resultado significativo nos preguntamos si este resultado puede ser debido al azar, pero lo interpretamos como consecuencia directa de nuestra manipulación experimental[24].

Falk y Greenbaum (1995) describen la *ilusión de la prueba probabilística por*

---

[22] En la práctica esta situación no se da. Por un lado, el número de observaciones dificulta que se pueda controlar simultáneamente todas las variables de interés. Por otro lado, quizás no todas las variables relevantes hayan sido medidas y algunas variables “confundidas” podrían no ser controladas (Selvin, 1970).

[23] Consistente en rechazar la hipótesis nula siempre que alfa sea inferior a 0’05 sin tener en cuenta otros elementos.

[24] El conflicto entre estos tres componentes psicológicos explica, según Gigerenzer (1993) los usos incorrectos de la inferencia estadística y la institucionalización del nivel de significación como medida de la calidad de la investigación en revistas científicas.

contradicción o *ilusión de alcanzar la improbabilidad*, que sería la creencia de que la hipótesis nula se vuelve improbable cuando se obtiene un resultado significativo[25]. Esta creencia resulta de una generalización abusiva del razonamiento en lógica deductiva, donde una contradicción entre la premisa y su deducción prueba la falsedad de la premisa de partida. Sin embargo, en razonamiento estadístico es erróneo pensar que cuando se obtienen datos cuya probabilidad condicional bajo una hipótesis es baja, la hipótesis condicionante sea improbable (Gigerenzer, 1993; Cohen, 1994).

Otra explicación dada por Falk (1986) es la *falacia de la condicional transpuesta* o dificultad en diferenciar las dos direcciones de las probabilidades condicionales $P(A/B)$ y $P(B/A)$. Además, la expresión "Probabilidad de Error Tipo I" es desafortunada y no debería usarse. Aunque el nivel de significación $\alpha = P(\text{rechazar } H_0 \,/\, H_0 \text{ cierta})$ es una probabilidad condicional bien definida, la expresión "Error Tipo I" no está redactada como una condicional, ni indica cual de las dos combinaciones posibles de los sucesos que intervienen $P(\text{rechazar } H_0 \,/\, H_0 \text{ cierta})$ o $P(H_0 \text{ cierta} \,/\, H_0 \text{ rechazada})$ se refiere. En consecuencia, cuando rechazamos H_0 y nos preguntamos por el tipo de error que podríamos cometer, el concepto "Error Tipo I" nos viene a la mente en forma inmediata ya que la distinción crucial entre las dos direcciones opuestas de la probabilidad condicional ha sido difuminada. Esto nos lleva a interpretar el nivel de significación como la probabilidad de la conjunción de dos sucesos "la hipótesis nula es cierta" y "la hipótesis nula es rechazada" (Menon, 1993).

Por último, Falk y Greenbaum (1995) también apuntan a la sensación subjetiva de que los contrastes estadísticos garantizan la objetividad como factor que provoca el que se sigan utilizando a pesar de las críticas recibidas. Muchos usuarios de la estadística no tienen una preparación matemática fuerte, por lo que, la complicación de los cálculos y razonamientos matemáticos y la visión de las matemáticas como una ciencia exacta les lleva a creer que como resultado se obtiene una solución “determinista”, es decir, una “prueba matemática” de la hipótesis implicada.

[25] Creer que se elimina el azar y se minimiza su incertidumbre cuando se obtiene un resultado significativo en un contraste de hipótesis.

2.8. CONTROVERSIA SOBRE EL USO DE LA INFERENCIA

La obtención de resultados estadísticamente significativos (interpretados incorrectamente como aquellos que hacen la hipótesis nula muy improbable) se convirtió en un requisito frecuente para que los trabajos fuesen aceptados en las revistas o congresos científicos en la década de los 70, independientemente de la significación práctica de sus resultados (Frías, Pascual y García, 2000)[26]. En otros casos los investigadores no se han preocupado de calcular intervalos de confianza para los valores medios de las variables o dar información sobre los efectos producidos ((Falk y Greenbaum, 1995) u olvidan el estudio de la potencia (Valera, Sánchez. y Marín, 2000).

Todo ello, junto con los errores descritos ha ocasionado fuertes críticas al uso y abuso de la inferencia por parte de muchos investigadores desde hace bastante años (véase por ejemplo, Morrison y Henkel, 1979), y recientemente con nuevas energías (Harlow, Mulaik y Steiger, 1997; Valera y Sánchez, 1997; Borges, San Luis, Sánchez, y Cañadas, 2001; De la fuente y Díaz, 2003)[27]. En este apartado detallamos las principales críticas que reciben los contrastes estadísticos. Algunas de estas críticas se refieren al procedimiento matemático en sí y no están bien justificadas, sino que indican que las dificultades de comprensión de los contrastes de hipótesis alcanzan a veces a los mismos críticos. Otras se refieren al uso que se hace del procedimiento o a la interpretación de sus resultados y están completamente justificadas. A continuación analizamos los dos tipos de críticas.

2.8.1. PRINCIPALES CRÍTICAS AL CONTRASTE DE HIPÓTESIS

Críticas al procedimiento matemático

Crítica 1. Las investigaciones experimentales siempre se hacen esperando un efecto de una determinada magnitud (que se desconoce), pero la hipótesis nula se plantea estableciendo que no hay diferencia entre dos poblaciones, o que no hay efecto. Con este planteamiento nos encontramos con el siguiente problema: la hipótesis nula se

[26] La comprobación de la significación estadística fue la técnica por excelencia del análisis de datos (Monterde-Bort, Pascual, Frías, en prensa).

[27] Incluso después de publicadas las recomendaciones de la APA algunos autores sugieren que no se ha hecho bastante y que los editores del campo de la psicología necesitan ahora tener en cuenta estas recomendaciones a la hora de aceptar los trabajos para publicación (Borges, San Luis, Sánchez, y Cañadas, 2001; Fidler, 2002).

plantea a priori como falsa y su falsedad realmente nunca se pone en duda (Bakan, 1970; Falk y Greenbaum, 2005). La única cuestión es si el investigador ha recogido suficientes datos para que el resultado salga estadísticamente significativo[28]. Sería inválido usar el contraste de hipótesis, puesto que a priori pensamos que la hipótesis nula es falsa, mientras el procedimiento se construye en el supuesto de que sea cierta.

Esta crítica al procedimiento no está bien fundada, porque el hecho de que la hipótesis nula sea falsa no afecta a la lógica del contraste, ya que lo que afirmamos en el contraste es que un resultado significativo es improbable en el caso de que la hipótesis nula sea cierta. Esto es una propiedad matemática de la distribución muestral del estadístico que no tiene nada que ver con la certeza o falsedad de la hipótesis nula (Chow, 1996).[29] La lógica del test de hipótesis se refiere a un modelo ideal que representa la situación experimental (Estes, 1997) y el test proporciona evidencia sobre si el modelo basado en la hipótesis nula representa bien o no a la situación. También podría usarse una hipótesis nula que represente un valor significativo del parámetro, diferente de cero si se sospecha de un tamaño del efecto dado (Granaas, 2002).

Crítica 2. El significado de un resultado no significativo no es claro. No se entiende por qué cuando obtenemos un resultado significativo se rechaza la hipótesis nula, pero cuando obtenemos un resultado no significativo no se acepta (sólo se dice que no se ha podido rechazarla). Algunos investigadores piensan que no se debiera usar el contraste de hipótesis porque no se trata igualmente a las hipótesis nula y alternativa (Chow, 1996).

En el contraste de hipótesis, las hipótesis nula y alternativa, así como los resultados significativos y no significativos no tienen un papel simétrico, puesto que el razonamiento es parecido a la prueba por contradicción. En una prueba por contradicción, se hace el siguiente razonamiento (Vallecillos, 1994):

Si A es cierta, B es falsa

Entonces si B es cierta, A es falsa

[28] Todo el razonamiento del contraste de hipótesis parte del supuesto de que la hipótesis nula es cierta y a partir de este supuesto determinamos la región crítica y tomamos la decisión final.

Si B ocurre, podemos concluir que A no es posible, pero cuando no ocurre B no se deduce que A sea necesariamente cierta; hay una asimetría entre las consecuencias de B y no B. Lo mismo ocurre en el contraste de hipótesis, aunque aquí el resultado siempre está sometido a la posibilidad de error (Cohen, 1994). Razonamos en la siguiente forma:

Si A es cierta, B es muy poco probable
Entonces si B ocurre, se rechaza A
Si B no ocurre, no se obtiene conclusión

En consecuencia esta crítica es infundada. Neyman y Pearson introdujeron el concepto de Error tipo II y potencia, para tener en cuenta los posibles errores relacionados con la aceptación de la hipótesis nula. La consideración de los dos tipos de error parece pertinente en contextos tales como el control de calidad, pero ha sido discutida por el mismo Fisher en el terreno de la inferencia estadística (Gigerenzer, 1993; Chow, 1996).

Críticas al uso actual del procedimiento

Crítica 3. La práctica de entender que cuando se rechaza la hipótesis nula, se obtiene un apoyo a la hipótesis de investigación (alternativa) es inadecuada. La hipótesis alternativa va ligada con la hipótesis de interés en el estudio. Sin embargo no indica la magnitud de la diferencia, por lo que la hipótesis estadística no informa sobre la significación práctica de los datos (Hager, 2000; Borges, San Luis, Sánchez y Cañadas, 2001).

Esta crítica es razonada, pero se refiere no al contraste en sí, sino al uso que se hace de él. Los investigadores confunden los diferentes niveles de hipótesis implicados al hacer un contraste que hemos descrito anteriormente. Por eso, al rechazar una hipótesis alternativa, piensan que han corroborado la "hipótesis de investigación" o incluso la "hipótesis substantiva". Lo único que indica un resultado significativo es que el efecto encontrado sería muy poco probable si la hipótesis nula fuese cierta. Los contrastes deben complementarse con la estimación del efecto, proporcionando intervalos de confianza para los mismos o el análisis de la potencia, especialmente si se llega a un

[29] Es decir nos basamos en una probabilidad condicional del tipo P(A/B). Aunque B sea falsa en un caso particular, la

resultado no significativo (Thompson, 1996; Valera y Sánchez, 1997; Pascual, García y Frías, 2000; Vacha-Haase, 2001).

> *Crítica 4.* El nivel de significación se establece con un criterio arbitrario, por lo que unos mismos datos podrían ser estadísticamente significativos para un nivel de significación y no serlo para otro (Skipper, Guenter y Nash, 1970).

Los investigadores que hacen esta crítica tienen razón, ya que el uso de la estadística siempre tiene un componente subjetivo. Este componente subjetivo no implica, sin embargo, que el procedimiento sea inválido o inútil. Además, aunque en la práctica actual se sugiere elegir el nivel de significación antes de recoger los datos para ser más objetivos, si se sigue el enfoque de Fisher, sería posible también usar el valor-p exacto para rechazar las hipótesis nulas a diferentes niveles de significación. Para elegir un nivel de significación se debería tener en cuenta la importancia práctica de cada tipo de error y la plausibilidad de las alternativas (Labovitz, 1970), así como la potencia del test, ya que en uno de gran potencia una pequeña diferencia puede ser significativa.

> *Crítica 5.* Aunque los errores tipo I y tipo II estén inversamente relacionados, muchos investigadores prestan excesiva atención al error tipo I e ignoran el error tipo II (Cohen, 1994).

Esta crítica está justificada, pero aunque las probabilidades de los dos tipos de error están inversamente relacionadas, hay una diferencia fundamental entre ellas. Mientras que la probabilidad α es una constante que puede elegirse antes de realizar el experimento, la probabilidad β es función del verdadero valor del efecto, que el investigador no conoce. Para resolver este problema, en el análisis de la potencia suponemos diferentes valores posibles del parámetro y calculamos la probabilidad β para estos diferentes valores (Valera, Sánchez, Marín y Velandrino, 1998). Este cálculo es necesario cuando se aplica la inferencia siguiendo la filosofía de Neyman y Pearson, es decir, se orienta a decidir entre dos hipótesis diferentes (nula y alternativa). En los casos de experimentos orientados a apoyar hipótesis substantivas teóricas, los dos tipos de error no juegan el mismo papel, siendo más importante el error tipo I (es decir, cuando seguimos la filosofía de Fisher) (Chow, 1996).

---

probabilidad P(A/B) no cambia.

Crítica 6. La significación estadística no informa de la probabilidad de que la hipótesis sea cierta, ni del verdadero valor del parámetro, ni del tamaño del efecto. Esto puede producir situaciones en que rechazar una hipótesis nula no proporcione información nueva, ya que lo único que nos dice es que hay un efecto, pero no en que dirección o de qué magnitud (Falk y Greenbaum, 1995).

También esta crítica es razonable, porque sólo en la inferencia bayesiana es donde podemos calcular probabilidades finales de la hipótesis y donde podremos dar probabilidades de que el efecto tenga un cierto tamaño, aunque hay que aclarar que ahora hablamos de probabilidades subjetivas, es decir de grados de creencia personal. En la inferencia clásica estas probabilidades finales no tienen sentido (Lecoutre, 1999; Rivadulla, 19991).

En consecuencia este segundo conjunto de críticas al uso que se hace del contraste están justificadas y lo que indican es que los contrastes de hipótesis no proporcionan toda la información que busca el investigador. Por tanto habrá que complementarlos o sustituirlos en algunas ocasiones por otros métodos, en los que podría estar incluida la inferencia bayesiana (De la Fuente, Díaz y Cañadas, 2005).

2.8.2. RECOMENDACIONES DE LA AMERICAN PSYCHOLOGICAL ASSOCIATION

Un primer ejemplo del interés de la Psicología por este debate lo tenemos en el número monográfico dedicado al tema en *Psychological Science* (Vol. 8, n.1, Enero, 1997)[30]. En este número Hunter (1997, p. 3) indica que "*El uso actual del test de significación es un desastre. Mientras que la mayoría de investigadores creen falsamente que el test de significación tiene una tasa de error del 5%, los estudios empíricos muestran que la tasa de error medio en psicología es del 60%- 12 veces mayor que lo que cree el investigador. La tasa de error para la inferencia usando el test de significación es mayor que la tasa de error si se usara el lanzamiento de una moneda para sustituir el estudio empírico*".

[30] El mismo editor (Estes,1997) reconoce que la obtención de resultados significativos fue uno de los criterios para una primera selección de trabajos, aunque no el único, primando el tipo de estudios y su calidad. Sugiere no imponer

El *Board of Scientific Affairs* de la *American Psychological Associaton* creó la *Task Force on Statistical Inference* proporcionó recomendaciones al respecto, tratando de clarificar los puntos controvertidos que rodean la aplicación de la estadística incluyendo los test de significación y sugerir posibles alternativas (Wilkinson, 1999). Recomendaron revisar la sección de estadística del Manual de Publicaciones de la APA y publicaron un artículo como forma de iniciar la discusión en el campo sobre los cambios en las prácticas actuales de análisis de datos y elaboración de informes. Este comité y algunos autores han hecho varias recomendaciones sobre muestreo, diseño, medición, instrumentos y procedimientos y resultados, que resumimos a continuación (Borges, San Luis, Sánchez, y Cañadas, 2001; Fiedler, 2002). Desgraciadamente, la efectividad de estas propuestas ha sido escasa hasta el momento (Monterde-Bort, Pascual, Frías, en prensa)

Recomendaciones generales

Se enfatiza la importancia del marco teórico como guía de la investigación y la necesidad de suficientes estudios de tipo exploratorio, antes de avanzar en estudios orientados a la confirmación de teorías[31]. Alertan también de que los potentes programas de cálculo pueden llevar a mal uso de los mismos, como publicar trabajos sin comprender cómo se hacen los cálculos o qué significan (Wilkinson, 1999).

Población y muestra

Puesto que todo estudio depende de las características de la población y muestra, estas deben ser definidas con claridad, así como el procedimiento de muestreo o la forma de asignar sujetos a los tratamientos (White, 1980; Wilkinson, 1999). También se recomienda controlar las variables relevantes cuando no ha habido diseño aleatorio y en caso que esto no sea posible, usar el término "grupo control" con precaución. (Wilkinson, 1999).

---

o prohibirse ningún método, sino crear una atmósfera que favorezca la innovación y la adaptación de la metodología al avance de la ciencia.

[31] Una recomendación general es no usar la estadística en forma mecánica (Kish, 1970, Cohen, 1994).

Métodos descriptivos y análisis exploratorio de datos

Se sugiere mayor uso del *análisis exploratorio de datos,* para visualizar y comprender mejor el conjunto particular de datos (Cohen, 1994). Se pide una mejor descripción de los datos, incluyendo los datos faltantes, asegurándose que los resultados publicados no se producen debido a anomalías o valores atípicos, en la selección de los datos (Wilkinson, 1999).

Intervalos de confianza

Se propone presentar los resultados de la investigación junto con alguna medida de la variabilidad en el muestreo, tal como el error estándar o intervalos de confinanza (Valera, Sánchez y Marín, 2000; Finch, Cumming y Thomason, 2001; Fidler, 2002). Su escaso uso se debe al hecho de que usualmente son demasiado amplios, pero se pueden mejorar aumentando la fiabilidad de los procedimientos y el tamaño de las nuestras (Cohen, 1994). Los intervalos de confianza no dan una tasa de errores de aplicación tan elevada y los estudiantes de psicología los comprenden mejor que el test de hipótesis. (Hunter, 1997)

Puesto que es extraño tratar los datos como aleatorios una vez recogidos, Rouanet (1998b) advierte del peligro de olvidar la interpretación frecuencial a los intervalos de confianza[32] y da una interpretación bayesiana, que es más natural[33]. En este sentido, los intervalos de confianza no son una solución al contraste, puesto que, aunque dan una idea del signo del efecto, están sujetos a los mismos errores de interpretación que los contrastes y los investigadores pueden interpretar los coeficientes de confianza (que son probabilidades iniciales) como probabilidades finales (Harris, 1997). De hecho un intervalo del 95% de confianza que no incluya el valor cero equivale al test de hipótesis de significación 5% (Abelson, 1997).

Uso del contraste de hipótesis

Es difícil imaginar una situación en que sea preferible una decisión dicotómica (aceptar-rechazar) que informar sobre el valor-p o el intervalo de confianza. Los contrastes de diferencia o de efecto nulo suelen ser débiles y con frecuencia triviales. El investigador debiera probar

---

[32] (En cien muestras tomadas al azar de la población un tanto por ciento de intervalos cubren al parámetro, pero la muestra particular no sabemos si lo contiene o no).

[33] (la probabilidad de que el parámetro está en el intervalo particular que hemos obtenido es un valor dado).

hipótesis sobre diferencias de un tamaño dado, basado en un cierto modelo teórico (Kish, 1970, Frías, Pascual y García, 2002). Se recomienda las estimaciones de los efectos y los intervalos de confianza y no usar nunca la expresión "aceptar la hipótesis nula" (Wilkinson, 1999 Monterde-Bort, Pascual y Frías, en presna). Por otro lado, hay casos en que un test puede ser muy valioso, por ejemplo, para comprobar la bondad de ajuste de los datos a un modelo (Harlow, 1997) o usos de test usuales como el de la T para decidir que la dirección de un efecto no era la que esperábamos (Abelson, 1997).

Harris (1997) cree que los tests de significación todavía pueden ser útiles en psicología, pero sólo si se pasa de la lógica de solamente dos alternativas hacia un test de tres alternativas: hipótesis nula de no diferencia, hipótesis de efecto positivo e hipótesis de efecto negativo[34]. Este autor considera tres tipos de errores:

- *Error Tipo I,* rechazar la Hipótesis nula cuando es cierta.
- *Error Tipo II,* no rechazar la Hipótesis nula cuando es falsa
- *Error Tipo III,* rechazar la Hipótesis nula en dirección equivocada. Esto ocurre cuando hemos supuesto una determinada dirección para el efecto y pensamos que el efecto tiene signo contrario.

Potencia, efectos y tamaño de muestra

Una posibilidad es incluir el análisis de la potencia, aunque su enseñanza no es fácil en las condiciones actuales de los cursos universitarios (Hunter, 1997) Se debe proporcionar información sobre el tamaño de muestra y por qué se eligió dicho tamaño, informar sobre el tamaño de los efectos (Frías, Pascual y García, 2000; Vacha-Haase, 2001), hipótesis y procedimientos analíticos que llevan al análisis de la potencia (Cohen, 1994; Valera, Sánchez, Marín. y Velandrino, 1998; Valera, Sánchez y Marín, 2000, Cañadas, Borges, Sánchez y San Luis, 2000)[35].

Puesto que el análisis de potencia es mejor si se hace antes del estudio, se debiera explicar el criterio elegido para determinar el tamaño del efecto (por ejemplo si se usó información de trabajos anteriores) (Wilkinson, 1999).

[34] Es decir, indica que hay que tener en cuenta la dirección posible del efecto antes de coger los datos.

[35] Algunos investigadores dan mayor importancia al valor del nivel de significación que al tamaño del efecto; incluso es el unido dato que se ofrece (Selvin, 1970).

Comparaciones múltiples

Se sugiere emplear procedimientos especiales (por ejemplo el test de Bonferroni) para tratar las situaciones en que tenemos que hacer varias comparaciones en la misma muestra (Moses, 1992; Wilkinson, 1999).

Replicación

Finalmente, debido al carácter inductivo de nuestras ciencias, no tenemos más remedio que apoyarnos en la replicación de experimentos (Cohen, 1994). La medida de los efectos permite acumular conocimiento de diversos estudios, así como la discusión social de los trabajos de investigación (Abelson, 1997). La replicación es una práctica muy positiva que debe fomentarse (Harlow, 1997; Monterde-Bort; Pascual y Frías, en prensa).

2.9. CONCLUSIONES

Los debates anteriores no sólo se refieren a la filosofía de la inferencia. Ito (1999), sugiere tres tipos de controversia respecto a los contrastes estadísticos, que están relacionados entre sí:

(a) La discusión entre los estadísticos entre las escuelas de Fisher, Neyman-Pearson y el enfoque Bayesiano que todavía no están resueltas.

(b) La controversia en la aplicación de la estadística, donde, en la práctica el contraste de significación es una mezcla informal de los tres enfoques y los editores de revistas, así como sociedades profesionales están sugiriendo cambios en las políticas de publicación científica, respecto a los métodos estadísticos (Lecoutre, 1999).

(c) La controversia en la enseñanza acerca de cuándo, como y con qué profundidad deberíamos enseñar la inferencia estadística.

Aunque tanto en este tema como en el resto de la inferencia *"la decisión sobre tests de significación o procedimientos alternativos debe dejarse al investigador individual, e*n lo que sigue se analiza con más detalle la inferencia bayesiana, desde los tres puntos de vista anteriores. Todo ello con el fin de fundamentar la propuesta didáctica que se evalúa en los estudios 9 y 10, estudios que se justifican ya que: "*El principal foco para los esfuerzos hacia la mejora del diseño de la investigación y el análisis de datos en psicología y otras ciencias del comportamiento debe ser mejorar la preparación matemática y científica de los estudiantes que se preparan para ser investigadores en estos campos"* (Lecoutre, 1999, p. 18).

Dedicaremos la sección 3 a analizar con más detalle las recomendaciones sobre el uso de métodos bayesianos.

3. APORTACIONES DE LOS MÉTODOS BAYESIANOS A LA MEJORA DE LA PRÁCTICA METODOLÓGICA

En los últimos años se sugiere que la inferencia bayesiana proporciona una respuesta más ajustada a los problemas de inducción en ciencias empíricas y que la interpretación de los conceptos bayesianos es más intuitiva para el investigador Se recomienda introducir los métodos bayesianos en Psicología, en especial los que siguen las corrientes más objetivas dentro de este paradigma, utilizando las ideas de Fisher[36] de inferencia fiducial que tienen gran interés teórico y práctico (Edwards, Lindman y Savage; 1963; Rozeboom; 1970; Lindley, 1993; Pruzek, 1997 y Rindskopf, 1997; Lecoutre, 1999; 2006)[37].

En lo que sigue aportaremos los siguientes argumentos a favor de la metodología bayesiana:

a) La metodología bayesiana es adecuada para aprender inductivamente de la experiencia;
b) Su aplicación no contiene mayor subjetividad que la de otros métodos estadísticos;
c) Proporciona la información que realmente interesa a los investigadores y
d) Existe software estadístico que facilita la aplicación de esta metodología.

[36] El mismo Fisher argumentó en contra de dar una interpretación frecuencial al nivel de significación y estuvo interesado en la probabilidad final de una hipótesis, lo que se evidencia no solo por su trabajo sobre inferencia fiducial sino porque usó métodos bayesianos en sus últimos años (Rivadulla, 1991; Lecoutre, 1999).

[37] Una serie de factores parecen haber influido en que los métodos bayesianos elementales no se incluyan en los cursos iniciales de estadística: tradición, convicciones filosóficas, dificultad de aprendizaje, escasez de software, resistencia de las revistas tienen una resistencia a publicar trabajos basados en métodos bayesianos. Aparte del filosófico, todos estos argumentos son rebatidos por Lecoutre (1999).

Finalmente razonaremos que los conceptos básicos de esta metodología son asequibles para nuestros alumnos, si se realiza el necesario esfuerzo didáctico y justificamos el interés de los estudios 9 y 10 dirigidos a diseñar y experimentar material para la enseñanza de inferencia bayesiana a estudiantes de psicología (De la Fuente, Díaz y Cañadas, 2005).

3.1. SOBRE LA SUBJETIVIDAD EN LOS MÉTODOS BAYESIANOS

El teorema de Bayes permite transformar las probabilidades iniciales (antes de realizar un experimento) de varias causas, una vez se observan sus consecuencias, en probabilidades finales, qu incorporan la información de los datos observados. Las probabilidades de tales causas podrían entonces revisarse y pierden de este modo el carácter objetivo que les asigna la concepción frecuencial[38]. (Cabriá, 1994; Hacking, 1975). Keynes, Ramsey y de Finetti describen las probabilidades como grados de creencia personal, basadas en el conocimiento y experiencia de la persona que las asigna sobre el suceso dado (Batanero y Díaz, en prensa). Para ellos la probabilidad de un suceso siempre está condicionada por un cierto sistema de conocimientos y puede ser, por tanto, diferente, para distintas personas **(**Rindskopf, 1997).

Una dificultad del enfoque subjetivo fue hallar una regla para asignar valores numéricos a las probabilidades, de forma que expresen los grados de creencia personal. Ramsey (1926) y de Finetti (1937) deducen una teoría de decisión consistente, que permite separar las creencias de las preferencias, a partir de un sistema de apuestas, e inferir los valores de las probabilidades subjetivas. En el enfoque subjetivo, ya no es necesaria la repetición del experimento en las mismas condiciones, para dar sentido a la probabilidad y ello amplía el campo de aplicación, en particular al estudio de decisiones en economía, diagnóstico médico y otras aplicaciones. En la actualidad la escuela bayesiana aplica probabilidades a todo tipo de sucesos inciertos, aunque la controversia sobre el estatuto científico de las probabilidades subjetivas continúa (Rivadulla, 1991; Batanero y Díaz, en prensa).

Puesto que el investigador especifica la distribución inicial, el enfoque bayesiano

[38] De Finetti (1974) llega a decir que "la probabilidad no existe" puesto que su estimación depende de demasiados factores. Para él suponer que la probabilidad tiene una existencia objetiva supone confundir modelo y realidad (Batanero, Henry y Parzysz, 2005).

tiene en cuenta la perspectiva del investigador, su conocimiento del problema. Ello puede llevar a que distintos investigadores obtengan diferentes resultados de los mismos datos, o bien que, si no se tiene una idea clara de la distribución inicial al comienzo del análisis, cambiemos los modelos e interpretaciones a lo largo del mismo. Pero, dejar a los datos "hablar por si mismos" es usual en la modelización, donde se asume que los modelos son útiles para describir los datos, pero no son exactamente iguales a los datos y es posible cambiar de modelo a lo largo del análisis (Pruzek, 1997; McLean, 2001).

Además, la influencia de la distribución inicial, depende del tamaño muestral y los posibles sesgos iniciales se corrigen en sucesivos experimentos, ya que el peso se vuelca sobre las verosimilitudes y se aumenta progresivamente el tamaño de la muestra (Lindley, 1993; Lecoutre, 1993, 2006).

Por otro lado, mientras algunos defensores del método bayesiano insisten en el carácter subjetivo de la probabilidad, otros tratan de diseñar técnicas bayesianas "objetivas", investigando distribuciones iniciales "no informativas" (Rozeboom, 1970, Rouanet, 1998a). Se recomienda también repetir el análisis con diferentes distribuciones iniciales e informar de las diferencias obtenidas en la distribución final (Zhu y Lu, 2004) y normalizar los procedimientos, usando distribuciones conjugadas, de modo que tanto la distribución inicial como la final pertenezcan a la misma familia de funciones (Cabriá, 1994; de la Fuente y cols., 2002).

Tampoco los métodos frecuenciales escapan de la subjetividad[39]: el nivel de significación se establece arbitrariamente, por lo que unos mismos datos podrían ser estadísticamente significativos o no dependiendo del nivel de significación elegido. (Skipper, Guenter y Nash, 1970)[40]. La definición semántica o sintáctica de las variables, escalas de medición, pruebas de significación utilizadas, son otros ejemplos de las elecciones subjetivas del investigador.

Lo cierto es que la subjetividad no implica arbitrariedad, es inevitable en las ciencias sociales debido a la aleatoriedad inherente en sus variables y tiene un papel importante en la investigación científica (Ayçaguera y Benavides, 2003). La comunidad científica acepta los diferentes hallazgos, sin que ello implique que cada científico deba

---

[39] Más aún, la subjetividad es inevitable en la parte más importante de la inferencia, que es la interpretación de los resultados (Ayçaguera y Benavides, 2003).

[40] La interpretación de la significación estadística deja de tener sentido cuando el tamaño de la muestra es tan grande que cualquier diferencia detectada, por pequeña que sea, permitirá rechazar la hipótesis nula.

comprobarlos por si mismo, lo que sería impracticable, sino estableciendo otros criterios –metodológicos, plausibilidad o prestigio del autor (Matthews, 1998).

3.2. INFERENCIA BAYESIANA Y APRENDIZAJE INDUCTIVO

La visión subjetiva amplia las aplicaciones de la probabilidad, no siendo ya un requisito la repetición de una experiencia en las mismas condiciones. Gradualmente, se desarrolla la distinción entre *probabilidad frecuencial,* empíricamente accesible a través de la frecuencia y *probabilidad epistémica* o grado de creencia en la ocurrencia de un suceso en un experimento simple (Rouanet, 1998a) mientras se conforman dos escuelas de inferencia.

Al estimar un parámetro θ a partir de unos datos y, la inferencia frecuencial lo considera constante y desconocido, no asignándole una probabilidad. La inferencia bayesiana, por el contrario, considera el parámetro θ como aleatorio y le asocia a una distribución inical de probabilidades $p(\theta)$, de carácter epistémico. Más específicamente, siendo $y = (y_1,..., y_n)$ un conjunto de datos, cuya distribución de probabilidad $p(y/\theta)$ depende de varios parámetros $\theta=(\theta_1,... \theta_k)$ aleatorios que tienen su propia distribución de probabilidad $p(\theta)$, entonces la distribución condicional de θ dados los datos observados y viene dada por el teorema de Bayes :

$$p(\theta / y) = \frac{p(y / \theta) p(\theta)}{p(y)}$$

Siendo $p(y) = \sum p(y/\theta)$, donde la suma se extiende a través de todo el rango admisible de valores de θ (Box y Tiao, 1992).

La *distribución inicial.* $p(\theta)$ recoge toda la información sobre θ sin conocer los datos. La *función de verosimilitud* $p(y/\theta)$ nos da la probabilidad de obtener y en función de θ y la *distribución final* $p(\theta/y)$ recoge toda la información sobre θ una vez conocidos los datos. Este método podría aplicarse sucesivamente en nuevos experimentos, tomando como probabilidades iniciales del segundo experimento las probabilidades finales obtenidas en el primero y así sucesivamente. Hablamos así de "proceso de aprendizaje" (Box y Tiao, 1992).

Aunque la distribución final p(θ/y) coincide a veces con la distribución muestral, su interpretación es muy diferente[41]. Mientras el área en un intervalo de la distribución muestral es la probabilidad de que el estadístico de una muestra aleatoria esté en dicho intervalo, el área en un intervalo en la distribución final es la probabilidad de que el parámetro esté en el intervalo en la muestra particular obtenida (Rindskopf, 1997; Bolstad, 2004).

Esta forma de razonar, opuesta a la inferencia clásica que no usa la información previa sobre el suceso de interés (Ayçaguera y Suárez 1995), es sin embargo, muy cercana al razonamiento científico y constituye el primer tratamiento matemático de la inferencia inductiva (Rivadulla, 1991). El teorema de Bayes (y la inferencia bayesiana que se obtiene generalizándolo) permiten obtener conocimiento inductivo, pues las probabilidades iniciales se transforman en probabilidades finales a la luz de los sucesos observados (Berry, 1995; Bernard, 1998).

3.3. APORTACIONES DE LOS MÉTODOS BAYESIANOS A LA NECESIDADES DEL INVESTIGADOR

Varios trabajos sugieren que la inferencia bayesiana proporciona una respuesta más ajustada que la dada por la inferencia clásica a las necesidades del investigador (Lindley, 1993; Lecoutre, 1999; 2006).

En primer lugar, la estadística bayesiana conserva el sentido de la probabilidad en el mismo del lenguaje ordinario: *medida condicional de incertidumbre* asociada a la ocurrencia de un suceso, cuando se asumen ciertos supuestos (Bernardo, 2003). Esta es la interpretación intuitiva- aunque incorrecta dada a las probabilidades frecuenciales asociadas al test de hipótesis (Gingerenzer, 1993, Rouanet, 1998a, Lecoutre, 2006)[42]. En consecuencia, la interpretación bayesiana de la inferencia parece ser más sencilla y natural que la frecuencial (Pruzek, 1997), además de proporcionar una base de toma de decisiones coherentes en situaciones de incertidumbre (Western, 1999).

[41] "*Las inferencias bayesianas contienen su propia medida interna de la situación relacionada únicamente con la situación actual, sin necesidad de acudir a repeticiones de esa situacion mediante distribuciones en el muestreo*" (Cabría, 1994, p. 216).

[42] Estos errores se producen tanto en estudiantes como en profesionales de la psicología (Falk, 1986; Lecoutre, Lecoutre y Poitevineau, 2001; Haller y Krauss, 2002).

La inferencia bayesiana proporciona un método totalmente general, debido a que su aplicación no requiere un tipo particular de distribución (Bernardo, 2003). Lindley (1993) compara con un ejemplo el contraste de hipótesis según la concepción de Fisher con un enfoque bayesiano y saca las siguientes conclusiones:

- El enfoque de Fisher arroja resultados significativos más "fácilmente" que el enfoque bayesiano, pudiendo sugerir un efecto cuando en realidad no lo haya.
- El análisis bayesiano proporciona al investigador lo que pide, la probabilidad (subjetiva) final de que la hipótesis sea cierta, o la probabilidad de la magnitud el efecto estudiado, mientras que el nivel de significación da la probabilidad (objetiva e inicial) de un suceso, bajo la condición de que la hipótesis nula sea cierta, aunque como hemos dicho puede que esta hipótesis no sea cierta.
- El enfoque bayesiano tiene en cuenta la perspectiva del investigador, su conocimiento del problema, ya que el investigador tiene que especificar la distribución inicial. Esto a menudo se interpreta como crítica a este enfoque, pues se piensa que diferentes investigadores pueden obtener distintos resultados de los mismos datos. Sin embargo, se ha comprobado que el método bayesiano corregirá en sucesivos experimentos los posibles sesgos iniciales, ya que el peso se vuelca sobre las verosimilitudes.
- El método bayesiano es comparativo. Compara la probabilidad del suceso observado bajo la hipótesis nula y bajo diferentes hipótesis alternativa, mientras que el enfoque de Fisher sólo tiene en cuenta la hipótesis nula.

En lo que sigue analizamos como el método bayesiano responde a algunas de las necesidades de los investigadores.

Intensidad del efecto

Una de las recomendaciones en que coinciden diversos autores para complementar los tests de hipótesis es estudiar la magnitud del efecto. Pero una estimación puntual es insuficiente, puesto que no tiene en cuenta el error muestral (Poitevineau, 1998).

Un estudio de potencia sería recomendable para evitar conclusiones erróneas al concluir la ausencia de efecto cuando el resultado es no significativo (Cohen, 1990), pero el cálculo de la potencia no depende del valor observado del estadístico en la muestra y por tanto no es pertinente para interpretar un resultado particular, recogidos los datos (Falk y Greenbaum, 1995). Los intervalos de confianza tienen la misma interpretación frecuencial que los contrastes, ya que sólo nos indican la proporción de intervalos calculados de la misma población con tamaño de muestra dado que cubrirían el valor del parámetro, pero no nos indica si el intervalo calculado lo cubre o no (Cumming, Williams, y Fidler, 2004).

El estudio del efecto y su magnitud aparece en forma natural en los métodos bayesianos, que consideran el parámetro como variable aleatoria. La probabilidad de que dicho parámetro tome un cierto valor puede calcularse en la distribución final, pudiendo, por ejemplo utilizar frases como "la probabilidad de que el efecto supere tal magnitud es igual a 0,25". El intervalo de credibilidad proporciona los límites en los que el parámetro está incluido con una cierta probabilidad (Poitevineau, 1998; Bolstad, 2004; Lecoutre, 2006).

Contraste de hipótesis

El p-valor no tiene utilidad para el investigador, ya que es la probabilidad de que obtuviésemos unos datos más extremos que los obtenidos, si repitiésemos innumerablemente el experimento y la hipótesis fuese cierta (Matthews 1998)[43].

Interpretar el rechazo de la hipótesis nula como apoyo a la hipótesis de investigación (alternativa) no es adecuado[44], por lo que la hipótesis estadística no informa sobre la significación práctica de los datos (Hager, 2000)[45]. La significación estadística no informa sobre la probabilidad de que la hipótesis sea cierta, ni sobre el verdadero valor del parámetro o el tamaño del efecto. En muchas situaciones rechazar una hipótesis nula no proporciona información nueva, ya que lo único que se deduce es que hay un efecto, pero no su dirección o magnitud (Falk y Greenbaum, 1995; Lecoutre, 1999).

Por el contrario en la inferencia bayesiana se calculan probabilidades finales de la hipótesis y probabilidades de que el efecto tenga un cierto tamaño (Lindley, 1993). El método bayesiano además compara la probabilidad del suceso observado bajo la

[43] Ningún investigador está interesado en repetir indefinidamente el mismo experimento; el fin de la investigación científica no es adoptar una decisión acerca de la veracidad de la hipótesis sino ajustar nuestro grado de creencia en la hipótesis que está siendo contrastada (Rozeboom, 1970).

[44] Ya que un resultado significativo no indica la magnitud del efecto.

hipótesis nula y bajo diferentes hipótesis alternativas (Lindley, 1993). Por otro lado el enfoque bayesiano resulta mucho más natural que el frecuencial,, en situaciones, como los ensayos de bioequivalencia, donde el interés se centra en verificar la hipótesis nula, (Molinero, 2002).

Probabilidades predictivas y replicación

Interpretar la significación estadística como seguridad en la replicabilidad de los datos no tiene base estadística (Falk, 1986; Gingerenzer, 1994; Cohen, 1994; Falk y Greenbaum, 1995; Pascual, García y Frías, 2000). Puesto que la significación estadística no es evidencia de que la hipótesis de investigación sea cierta; ni proporciona su probabilidad, no hay base para estudiar la replicación ni evidencias verificables de replicación (Sohn, 1998).

En el método bayesiano podemos calcular la probabilidad de un suceso futuro, mediante la distribución predictiva dada en el denominador de la fórmula de Bayes[46] Esto permite estudiar la posibilidad de replicación de nuestros resultados o estimar el tamaño necesario de muestra para que un futuro estudio sea concluyente (Lecoutre, 1996). Por supuesto, siempre que se cumplan los requisitos de precisión de los datos, adecuación de procedimientos y competencia de los investigadores (Sohn, 1998). Entendida correctamente, la replicabilidad se relaciona con la fiabilidad y consistencia de los datos, y la única forma de comprobarla es mediante sucesivos contrastes empíricos (Pascual, García y Frías, 2000).

Uso de la información previa

Mientras que los métodos frecuenciales consideran cada muestra como completamente nueva y no incorporan la información de estudios previos al actual, en el marco bayesiano es concebible una secuencia de experimentos articulados donde la información de cada uno se va incorporando al siguiente (Box y Tiao, 1991; Pruzek, 1997) e incluso reconoce la posibilidad de diferentes opiniones o conocimientos (Lindley, 1993; Bolstad, 2004).

Aunque es posible usar la inferencia bayesiana cuando no se tiene información

[45] A pesar de ello, numerosos investigadores continúan interpretando un resultado estadísticamente significativo como prueba de la hipótesis alternativa (Finch, Cumming, y Thomason, 2001).

previa sobre el parámetro, la característica más interesante es usar probabilidades iniciales "informativas" cuando sea posible, incluso para investigar el efecto de diferentes distribuciones iniciales supuestas. La idea central del enfoque bayesiano radica en actualizar los conocimientos probabilísticos que tenemos de un fenómeno, en función de la información disponible, por lo que resulta de gran interés en el meta-análisis, que combina información sobre diferentes estudios (Molinero, 2002).

3.4. SOFTWARE PARA EL CALCULO BAYESIANO

Un requisito para introducir los métodos de análisis de datos es la disponibilidad de programas de cálculo que faciliten su aplicación. En los últimos años son varios los investigadores que están desarrollando diversos logiciales, por lo que este enfoque se va introduciendo gradualmente en Psicología (De la Fuente y cols., 2002). Por ejemplo, Albert (1996) publicó una serie de subrutinas de Minitab para el análisis bayesiano elemental, que se pueden descargar desde la página de Internet del autor (http://bayes.bgsu.edu/).

First Bayes (http://www.shef.ac.uk/~st1ao/1b.html) es un paquete preparado en la Universidad de Sheffield para la enseñanza de conceptos elementales Bayesianos, compatible con Windows, que se distribuye gratuitamente. Admite diferentes familias de distribuciones y calcula probabilidades iniciales finales y predictivas resultantes del análisis de modelos uniparamétricos, análisis de varianza de una vía y regresión (Lawrence, 2003).

Asimismo, el programa PAC (Lecoutre, 1996) permite analizar datos de diseños experimentales generales, incorporando métodos de comparación de medias y análisis de varianza univariante o multivariante, incluyendo medidas repetidas y covariables. El programa incorpora análisis frecuencial y bayesiano, con distribuciones iniciales no informativas y conjugadas. Ha sido desarrollado por un equipo de investigación que trata de incorporar el análisis bayesiano en los métodos estadísticos más frecuentemente empleados en psicología.

Para análisis más complejos Buggs (Bayesian inference Using Gibbs Sampling) es un software flexible interactivo y compatible con Windows, que permite cálculos

[46] Es decir, la media ponderada de la función de verosimilitud, ponderada por las probabilidades iniciales (Berry, 1995).

complejos bayesianos, basados en simulación (ver información en http://www.mrc-bsu.cam.ac.uk/bugs/). Ofrece facilidades en línea, tales como tutoriales, grupos de usuarios y ejemplos.

El programa BACC (Bayesian Analysis Computation and Communication), desarrollado a partir de un proyecto financiado por la National Science Foundation en Estados Unidos, se va actualizando y ofrece recursos para el cálculo bayesiano, disponibles libres de costo. El énfasis se pone en la combinación de modelos y el desarrollo de distribuciones predictivas. Hay versiones disponibles para Matlab, S-PLUS y R. E en versiones Windows, Unix y Linux (http://www2.cirano. qc.ca/ ~bacc/index.html).

Otros programas de cálculo bayesiano, algunos específicos y otros generales se van desarrollando y poniendo en Internet. Una lista de estos recursos puede verse en: http://www.mas.ncl.ac.uk/~ndjw1/bookmarks/Stats/Software-Statistical_computing/ Bayesian_software/.

3.5. PROBLEMÁTICA DIDÁCTICA ASOCIADA A LA ENSEÑANZA DE MÉTODOS BAYESIANOS

Una característica bien conocida de la inferencia bayesiana es que permite reinterpretar muchos procedimientos frecuentistas, y estas interpretaciones sirven para ver claramente las interpretaciones erróneas de los tests de significación (Lecoutre, 1996, 1999; Bosltad, 2004)[47].

Más aún la posibilidad de investigar interactivamente las distribuciones finales con software de tipo visual da una simplicidad conceptual a los conceptos bayesianos y facilita su comprensión. Un ejemplo es el software desarrollado por Lecoutre (1999)[48], con la finalidad de facilitar la aplicación y enseñanza de estos métodos.

La introducción de una nueva metodología en ciencias sociales requerirá de su difusión y comprensión por los posibles usuarios, es decir, del grado hasta el cuál

[47] Por ejemplo, la distribución final de la media de una distribución normal $N(\mu,\sigma)$ con desviación típica conocida es también normal , siendo n el tamaño de la muestra. Por ello los límites del intervalo de credibilidad para la media coinciden con los del intervalo de confianza frecuencial, aunque la interpretación es diferente. El intervalo de credibilidad nos da los límites en los que se encuentra la media de la población con una cierta probabilidad; esta es precisamente la interpretación errónea que se suele dar al intervalo de confianza.

seamos capaces de transmitir sus principales ideas en los cursos de estadística aplicada. Nuestro último punto es estudiar la posibilidad de enseñanza de este tema a alumnos de psicología o ciencias sociales.

3.5.1. INVESTIGACIONES SOBRE ENSEÑANZA DE CONCEPTOS BAYESIANOS

La enseñanza de conceptos básicos sobre inferencia bayesiana se recomienda cada vez con mayor frecuencia por algunos profesores de estadística (Stangl, 1998, Lecoutre, 1999, 2006, Albert, 2002; Bolstad, 2002), quienes han incorporado los métodos bayesianos a su enseñanza y sugieren que su comprensión por parte de los estudiantes es mayor que la alcanzada respecto a la estadística frecuencial. Por otro lado, las situaciones es la que se dispone de información previa en la toma de decisiones son cada día más frecuentes y hay software disponible para poder realizar el análisis. Las revistas estadísticas de impacto contienen hoy dia un porcentaje apreciable de métodos bayesianos, pero este hecho no se traduce en un incremento similar de la enseñanza de los mismos, privando a los futuros profesionales de una herramienta que les sería fundamental (Bernardo, 2006).

Iglesias y cols. (2000) sugieren la necesidad de que en los cursos de pregrado se introduzcan algunas ideas de inferencia bayesiana, junto con los de inferencia clásica, enfoque que ya se presentaba en algunos libros clásicos, como el de De Groot (1988) y sugiriendo que los temas de inferencia en estos cursos debieran ser, al menos, los siguientes:

- Conceptos básicos fundamentales: población, parámetro, muestra estadístico, función de verosimilitud, distribución inicial y final.
- Estimación puntual: Métodos clásicos y bayesianos
- Estimación por intervalos: intervalos de confianza y de credibilidad
- Contrastes de hipótesis. Tests clásicos y bayesianos, problemas de decisión múltiple.

En este sentido encontramos un número creciente de textos cuya comprensión no

[48] Quien considera que la inferencia bayesiana es hoy día fácil de aplicar y enseñar. Buscar las formas de facilitar su enseñanza e interpretación es un gran desafío actual.

requiere grandes conocimientos matemáticos y en los que los elementos básicos de inferencia bayesiana se presentan contextualizados en ejemplos interesantes y familiares para los alumnos (por ejemplo Berry, 1995 o Albert y Rossman, 2001). Estos materiales pueden complementarse con las muchas referencias que explican en forma sencilla los rudimentos de inferencia bayesiana (ej., Ayçaguera y Benavides, 2003; Ayaguera y Suárez, 1995).

Por ejemplo Albert (1995) presenta ejemplos y una metodología sencilla para introducir la inferencia bayesiana para una proporción en el caso discreto. Reproducimos la Tabla 1.1, donde compara las ventajas de las aproximaciones clásica o Bayesiana a la enseñanza: En la sección que sigue razonaremos que también en la inferencia clásica es necesario el estudio de la probabilidad condicional y que el teorema de bayes puede aprenderse con mayor facilidad si se utilizan recursos didácticos adecuados.

Tabla 1.1. Ventajas e inconvenientes de la enseñanza de inferencia clásica o bayesiana (Albert, 1995)

	Tradicional	Bayes
Ventajas	Procedimientos familiares incluidos en el software habitual Procedimientos automáticos, fáciles de usar Procedimientos "objetivos"	Interpretación natural Las interpretaciones son condicionales respecto a los datos observados Nos enfocamos en la interpretación de los parámetros La regla de Bayes da una regla general para actualizar las probabilidades
Inconvenientes	Hay que estudiar las distribuciones muestrales Enfoque excesivo en los procedimientos más que en los conceptos	Hay que estudiar la probabilidad condicional Hay que enseñar la asignación de probabilidades iniciales Hay que enseñar el Teorema de Bayes

Por otro lado, podemos encontrar en Internet también recursos didácticos que facilitan el aprendizaje de estos conceptos (Díaz y de la Fuente, en prensa); tales como applets que permiten visualizar el teorema de Bayes o las distribuciones de probabilidad, o bien realizar los cálculos de distribuciones finales, para la inferencia de medias y proporciones con distribución inicial discreta o continua (ver, por ejemplo, la página de Jim Albert, http://bayes.bgsu.edu/).

3.5.2. JUSTIFICACIÓN DE UN ESTUDIO EMPÍRICO SOBRE VIABILIDAD DE LA ENSEÑANZA DE MÉTODOS BAYESIANOS ELEMENTALES EN PSICOLOGÍA

La mayoría de los autores citados en este y el anterior apartado han incorporado los métodos bayesianos a su enseñanza y han indicado en sus escritos que los alumnos parecen comprender mejor la inferencia bayesiana que la clásica. Encontramos también descripciones de experimentos concretos de enseñanza y sugerencias sobre la forma de llevarlos a cabo (Stangl, 1998; Albert, 2002, Boldstad, 2002). Somos conscientes, sin embargo, de que esta posición es aún controvertida, incluso por educadores estadísticos prestigiosos quienes indican que la inferencia bayesiana se apoya demasiado en la probabilidad condicional, un concepto difícil para los estudiante (e.g. Moore, 1997b y c).

Por otro lado, el aprendizaje de la inferencia bayesiana no sólo se apoya en la probabilidad condicional – también requerida en la inferencia clásica-, sino que precisa conocimiento sobre la lógica de la inferencia científica y su diferencia respecto a la prueba por contradicción. Pero, durante los últimos 50 años se ha mostrado la dificultad de comprender y aplicar esta lógica en el caso de la estadística inferencial clásica (Harlow, Mulaik & Steiger, 1997; Lecoutre, Lecoutre & Poitevineau, 2001). Como se expuso en las secciones anteriores, los investigadores dan interpretaciones bayesianas (incorrectas) a conceptos como el nivel de significación, valor p o intervalo de confianza. No es, en consecuencia, evidente que estos errores no se produzcan en el aprendizaje de la inferencia bayesiana, puesto que la investigación empírica sobre el aprendizaje real de los estudiantes en contextos de enseñanza naturales (dentro de los cursos de estadística) es aún muy incipiente.

Iversen (1998) lleva a cabo una evaluación de la percepción de los estudiantes sobre la inferencia bayesiana, en comparación con la clásica en una pequeña muestras de estudiantes. En su investigación los estudiantes parecen interpretar correctamente los intervalos de credibilidad, pero se muestran confusos sobre el hecho de que distintas distribuciones iniciales puedan proporcionar diferentes distribuciones finales con los mismos datos, incluso cuando esta diferencia se corrija poco a poco en nuevos experimentos. También resultó problemático comprender la diferencia entre la concepción de parámetro en inferencia clásica (como constante) y bayesiana (como variable aleatoria). Puesto que la evaluación se basa en la opinión de los estudiantes (encuesta de opinión) y no en tareas estadísticas propuestas a los mismos, los resultados podrían variar en una prueba de evaluación convencional.

Albert (2000) realiza una aproximación a la enseñanza de la inferencia sobre una proporción, evaluando la comprensión de los estudiantes por medio de un proyecto. Indica que los estudiantes tienen dificultar en realizar asignaciones de las distribuciones iniciales, y en interpretar los intervalos de credibilidad. Sin embargo, examina sólo 12 proyectos y estos fueron realizados por estudiantes trabajando en grupo, por lo que es difícil apreciar la comprensión individual adquirida por cada estudiante.

Basándonos en todos estos antecedentes, abordaremos en el Capitulo 6 (estudios 9 y 10) el diseño y evaluación de un material elemental con la finalidad de ser utilizado en la enseñanza del teorema de Bayes e inferencia sobre media y proporción, en modalidad presencial o a distancia. Consta de material escrito para el alumno, ejercicios resueltos, actividades y ejercicios de autoevaluación, así como subrutinas Excel de ayuda al cálculo. Siguiendo la sugerencia de Iglesias y cols. (2000), nos hemos concentrado más en transmitir conocimientos conceptuales que procedimentales (puestos que éstos vienen resueltos por el software) y enfatizando el lenguaje de comunicación con los alumnos, al contextualizar los conceptos en áreas de aplicación a la psicología o de su experiencia cotidiana.

4. RAZONAMIENTO SOBRE PROBABILIDAD CONDICIONAL Y SU IMPORTANCIA EN LA COMPRENSIÓN DE CONCEPTOS BAYESIANOS ELEMENTALES

4.1. IMPORTANCIA DEL RAZONAMIENTO CONDICIONAL EN LA COMPRENSIÓN DE CONCEPTOS BAYESIANOS ELEMENTALES

El análisis realizado en las secciones anteriores indica que la inferencia estadística es altamente compleja y su enseñanza requiere un tiempo suficiente de maduración. Los diversos procedimientos (tests de significación, contrastes de hipótesis, etc.) requieren la comprensión previa de muchos conceptos estadísticos y probabilísticos, tales como los de variable estadística y aleatoria, distribución, población y muestra, probabilidad, estadístico y parámetro (Vallecillos, 1994; 1999). Muchos estudiantes de Psicología carecen del conocimiento matemático y lógico suficiente para comprender con profudidad los métodos estadísticos; por ello los esfuerzos para mejorar la práctica de la estadística en Psicología deberían empezar con la mejora de la enseñanza (Estes, 1997, p. 19).

Uno de los conceptos más fundamentales en la comprensión de la inferencia, tanto en su perspectiva clásica, como bayesiana es el de probabilidad condicional, que permite incorporar cambios en nuestro grado de creencia sobre los sucesos aleatorios a medida que obtenemos nueva información (Heitele, 1975; Falk 1989; Tarr y Lannin, 2005). La probabilidad condicional está en la base de muchos conceptos inferenciales, incluyendo los que tienen mayor dificultad de comprensión por parte de los investigadores (Batanero, Díaz y De la Fuente, en prensa):

- Con relación al test de significación, interviene en la definición del valor-p, nivel de significación, distribución muestral del estadístico, región crítica, así como en la aplicación de la lógica del contraste, definición de errores tipo I y II, potencia y región de aceptación.
- En el método fiducial lo usamos para definir la función de verosimilitud.
- En la inferencia Bayesiana, tanto las probabilidades finales como las verosimilitudes se definen en término de la probabilidad condicional. Es también el núcleo del teorema de Bayes y del concepto de región de credibilidad.
- También en la metodología de intervalos de confianza frecuenciales, tanto los conceptos de nivel de confianza, como la construcción del intervalo se basan en la probabilidad condicional.

Además, la probabilidad condicional interviene también en el estudio de la correlación y regresión, que se definen basándose en este concepto y que se extienden posteriormente a muchos procedimientos, tales como los modelos lineales, el análisis factorial o discriminante, la asociación en variables cualitativas, etc.

Por otro lado muchas de las interpretaciones erróneas de estos conceptos descritas en la Sección 2.6 se deben al intercambio de términos en una probabilidad condicional, la confusión entre probabilidad condicional y simple o a no comprender un razonamiento basado en una lógica condicional. Aparentemente la probabilidad condicional y los conceptos ligados a ella no debieran tener una gran dificultad para los alumnos universitarios, ya que no requieren cálculos complejos ni conocimientos profundos de análisis matemático o álgebra. Sin embargo, la investigación psicológica y didáctica ha mostrado que dichas dificultades existen, incluso en alumnos con buena preparación matemática y que se producen en el nivel de resolución de problemas (Díaz y de la Fuente, 2005b; Tarr y Lannin, 2005; Batanero, Díaz y De la Fuente, en prensa).

Pensamos que no se ha prestado la debida atención a este concepto en la enseñanza universitaria, ya que se supone estudiado en secundaria. Además, el lenguaje y notación usadas en el caso de enseñarse no es siempre el más sencillo[49] e incluso algunos autores sugieren que es posible la enseñanza de la inferencia sin utilizar la probabilidad (Moore, 1997a y b).

Todo ello será analizado en las siguientes secciones. El objetivo principal es recabar información sobre los principales errores relacionados con el concepto de probabilidad condicional que han sido investigados, y justificar la necesidad de elaborar un instrumento comprensivo de evaluación. Al mismo tiempo, se recopilarán una serie de ítems empleados en dichas investigaciones que serán la base de su construcción (Estudios 1 a 6). El instrumento será empleado para evaluar las dificultades de los estudiantes de psicología (Estudio 7) y tenerlas en cuenta en el diseño de un material de enseñanza para iniciar la introducción de algunas ideas de inferencia bayesiana (que se aborda en los Estudios 8 y 9).

4.2. ERRORES EN EL RAZONAMIENTO SOBRE PROBABILIDAD CONDICIONAL

4.2.1. RELACIÓN ENTRE INDEPENDENCIA Y PROBABILIDAD CONDICIONAL

Un primer grupo de trabajos relaciona el concepto de probabilidad condicional con la idea de independencia, ligada al anterior concepto, tanto por su definición, como por

[49] *La noción de probabilidad condicional es un instrumento básico de la teoría de probabilidades y, por desgracia, su gran simplicidad se ve a veces oscurecida por una terminología singularmente inadecuada*" (Feller, 1973, pg. 127)

su significado intuitivo. Las dificultades aparecen al comprender la independencia de sucesos de distintos experimentos, porque no se identifica con claridad la intersección de los sucesos en el espacio muestral producto (Sánchez, 1996; Truran y Truran, 1996). Tampoco es sencillo identificar si dos variables son independientes o están asociadas en situaciones experimentales (Kelly y Zwiers, 1986; Ortega, 1991; Ortega, Montes y Ortega, 1992; Estepa, 1993)[50]. Otras dificultades se relacionan con conceptos filosóficos y matices psicológicos que los estudiantes asocian a la probabilidad condicional. A continuación se describen algunas de estas dificultades, incluyendo algunos de los ítems empleados en las mismas, que luego entrarán (modificados) a formar parte del cuestionario.

Concepciones y obstáculos

Maury (1986) utiliza situaciones elementales, como el problema 1 para analizar las concepciones de 374 estudiantes del curso preuniversitario y últimos cursos de Bachillerato, respecto a la probabilidad condicional.

Problema 1. Se tiene una bolsa con 1 bola azul y 2 bolas rojas. Se saca una bola al azar. Sin reponerla otra vez en la bolsa se saca una segunda bola. ¿Qué suceso es más probable?

- *Sacar dos bolas rojas.*
- *Sacar primero una bola roja y luego una azul.*

Propone cuatro problemas diferentes cambiando el contexto (con y sin reemplazamiento; contexto de bolas en urnas y ruletas). También utiliza dos tipos de vocabulario (técnico y cotidiano) al plantear las preguntas. La tasa de éxito observada es muy baja (sólo la cuarta parte de los alumnos produce respuestas correctas), comparada con otros problemas de probabilidad simple, para los que obtuvo un 60 % de aciertos. Supone que la dificultad no está ligada, en sí, a la noción de independencia, sino que el hecho de que los dos sucesos (azul/ rojo) sean no equiprobables introduce un distractor que aumenta la dificultad de las tareas. Esta suposición se confirma en otro experimento (Maury, 1985; 1986) en la que plantea a 290 alumnos de entre 13 y 16 años la siguiente

[50] La investigación sobre asociación en tablas de contingencia muestra nuestra escasa capacidad para identificar si las variables son o no independientes (Pérez Echeverría, 1990).

pregunta:

Problema 2. Una persona lanza 3 veces la misma moneda, obteniendo en este orden los resultados siguientes: cara, cruz, cruz

Lanza la moneda una cuarta vez.

¿Piensas que en el cuarto resultado es más probable la cara o la cruz?

Obtiene unos buenos resultados, que se aproximan a 70 % de éxitos. El problema se plantea en cuatro variantes, cambiando la forma de la pregunta final y la autora observa un éxito mayor cuando en la pregunta se presenta un listado de todos los sucesos posibles del espacio muestral, que cuando no se listan.

Sánchez (1996) pasa un cuestionario de probabilidad a 88 profesores de matemáticas quienes estaban participando en un programa de actualización de profesores. El cuestionario incluía dos preguntas de independencia, una de las cuales era la siguiente:

Problema 3. Se extrae una carta al azar de una baraja americana: sea A el suceso "se extrae un trébol" y B el suceso "se extrae una reina" ¿Los sucesos A y B son independientes? Argumentar.

Sólo 44 profesores hicieron intentos sistemáticos por resolver el problema; de éstos, 39 llegaron a una respuesta, pero sólo 4 lo hicieron correctamente, utilizando la regla del producto. En las respuestas incorrectas, así como en algunas entrevistas, encuentra dos tipos de concepciones:

a) Identificar eventos independientes con eventos excluyentes: "*no son independientes porque hay la reina de tréboles*"[51]. La definición formal de independencia hace intervenir la regla del producto, en que aparece la operación de intersección, así como en la definición de sucesos mutuamente excluyentes. Matemáticamente decimos que dos sucesos son independientes cuando la aparición /no aparición de uno de ellos no proporciona información sobre la ocurrencia del otro, por lo que precisamente dos sucesos excluyentes serían dependientes, ya que si uno ocurre, el otro es imposible.

b) Creer que sólo se puede aplicar la idea de independencia a sucesiones de experiencias, es decir a experimentos simples no simultáneos en el tiempo: *"si extraemos una carta para verificar el evento A y se vuelve a colocar en la baraja para verificar el evento B entonces A y B son independientes. Si se extrae la carta para verificar A y no se regresa, entonces A y B no son independientes"*.

Este estudio demuestra la persistencia de esas concepciones, y el olvido total de la regla del producto o la definición de independencia estocástica, en sujetos que, en algún momento, habían realizado un curso de probabilidad. Los resultados se confirman en Sánchez y Hernández (2003).

Totohasina (1992) propone problemas de probabilidad condicional a 67alumnos del curso preuniversitario, que habían estudiado la probabilidad pero no la probabilidad condicional, variando el formato (porcentajes o frecuencias absolutas), obteniendo, aproximadamente el 60% respuestas correctas. Estudia los tipos de representaciones usados espontáneamente por los alumnos para resolver el problema, que fueron: diagrama en árbol, representación mediante tabla de doble entrada o representación rectangular. Encuentra las siguientes dificultades:

- Interpretación no probabilística del enunciado, haciendo uso sólo de las frecuencias absolutas de casos, pero no de los porcentajes.
- No restringir el espacio muestral al calcular las probabilidades, lo que implica una comprensión deficiente de la idea de probabilidad condicional.
- Confusión entre la probabilidad conjunta y la probabilidad condicional: *"tenemos un 25% de bolas rojas en el conjunto y de ellas el 45% tienen la marca M; luego (25/100) x (45/100)= 11,25% de las bolas rojas tienen la marca"*.

4.2.2. CONDICIONAMIENTO Y CAUSACIÓN

La causalidad es un concepto científico, filosófico y psicológico complejo, pero también un concepto naturalmente aceptado por las personas ya que construimos nuestro conocimiento del mundo sobre la base de relaciones de causa y efecto entre

[51] Error ya descrito por Kelly y Zwiers (1986), quienes sugieren que puede ser debido a la imprecisión del lenguaje ordinario, en que "independiente" puede significar, a veces, separado.

diferentes sucesos (Kelly y Zwiers, 1986). Una interpretación errónea relacionada con el concepto de probabilidad condicional es establecer una relación de causalidad[52].

Razonamiento causal y diagnóstico

Desde el punto de vista de la probabilidad, si un suceso *A* es la causa estricta de un suceso *B*, siempre que suceda *A*, sucederá *B,* por lo que la $P(B/ A) = 1$. La relación causal estricta es difícil de hallar en el mundo real y hablamos de relación de causa débil cuando al suceder A cambia la probabilidad de que ocurra *B*. Es decir $P(B/A) \neq P(B)$, por lo cual una relación de causalidad implica una dependencia de tipo estadístico.

Sin embargo si $P(A/ B) = 1$, *B* no tiene que ser la causa de *A* ni la existencia de dependencia entre dos sucesos *A* y *B* implica que uno sea causa del otro[53]. Desde el punto de vista psicológico, la persona que evalúa una probabilidad condicional *P(A/B)* puede percibir dos relaciones muy diferentes entre *A* (suceso evaluado) y *B* (suceso condicionante), dependiendo del contexto (Tversky y Kahneman, 1982a).

> *Problema 4 ¿Cuál de los siguientes sucesos es más probable? a) Que una niña tenga los ojos azules si su madre tiene los ojos azules. b) Que una madre tenga los ojos azules si su hija tiene los ojos azules. c) Los dos sucesos son igual de probables" (Pollatsek y cols., 1987).*

Aunque matemáticamente los dos enunciados son equivalentes, desde un punto de vista psicológico no son percibidos como idénticos por las personas, ya que la creencia en las relaciones causales es más fuertes que en las relaciones diagnósticas. Tversky y Kahneman (1982a)[54] explican este hallazgo con la existencia de un sesgo causal cuando las personas se enfrentan con tareas relacionadas con la probabilidad condicional.

- Si dentro del contexto se percibe que *B* (suceso condicionante) es una causa de *A*, la persona establecerá entre *A* y *B* una *relación causal*. En la primera opción, *A* sería

[52] Hay muchos ejemplos de relaciones condicionales no causales (hacer fuego es condicional a la presencia de oxigeno, pero éste no causa el fuego).

[53] La existencia de una relación condicional indica que una relación causal es posible, pero no segura. Una asociación estadística entre variables puede ser debida a otras variables intervinientes o incluso ser espúrea y no implica relación causal.

[54] Tversky y Kahneman (1982a) encontraron que las personas encontraban más probable que "una niña tenga los ojos azules si su madre tiene los ojos azules" que "una madre tenga los ojos azules si su hija tiene los ojos azules", aunque casi la mitad de los sujetos de su estudio respondieron correctamente "ambos sucesos son igual de probables".

que la niña tenga los ojos azules, *B* que la madre tenga los ojos azules (causa). Por tanto, *P(A/ B)* es una relación causal. En este caso el estudiante tendrá que realizar un razonamiento causal, estimando el efecto dado cierto conocimiento de las causas.

- Si dentro del contexto se percibe *A* como una causa de *B,* la persona establecerá entre *A* y *B* una *relación diagnóstica*. En este caso *A* sería que la madre tenga los ojos azules (causa) y *B* que la niña tenga los ojos azules, por tanto *P(A/ B*), es una relación diagnóstica, estimando la causa, dado el conocimiento del efecto.

Condicionamiento, temporalidad y causación

La relación de causalidad también se asocia con la secuencia temporal. Falk (1986, 1989) propuso el problema 5 a sus estudiantes y encontró que, mientras no tenían dificultad para resolver la primera parte, muchos eran incapaces de dar una solución a la segunda y responden diciendo que para la primera extracción, no afecta el resultado en la segunda extracción. En otros casos dan como respuesta ½, teniendo en cuenta sólo la composición de la urna y sin utilizar el dato del resultado posterior. Las respuestas de los estudiantes no consideran la evidencia ocurrida después del suceso que juzgamos, y refleja un razonamiento causal que se conoce como *falacia del eje temporal* y consiste en suponer que el suceso condicionante en la probabilidad condicional ha de preceder temporalmente al condicionado.

Problema 5. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin reemplazamiento.

- *¿Cuál es la probabilidad de extraer una bola negra en segundo lugar, habiendo extraído una bola negra en primer lugar?* $P(N_2/ N_1)$
- *¿Cuál es la probabilidad de extraer una bola negra en primer lugar, sabiendo que hemos extraído una bola negra en segundo lugar?* $P(N_1/ N_2)$*”*

En el problema, el resultado en la segunda extracción depende causalmente del resultado en la primera extracción, pero no al contrario. Sin embargo, el resultado en la segunda extracción influye en la estimación del resultado en la primera, puesto que al saber que en la segunda extracción ha salido una bola negra, ha habido una reducción del espacio muestral y $P(N_1/ N_2)$ sería un tercio, igual que $P(N_2/ N_1)$ (Falk, 1989; Truran y Truran, 1996).

En la primera parte del problema la inferencia causal es una situación natural y compatible con el eje temporal, pero en la segunda situación se pide hacer una inferencia inversa, lo que requiere un razonamiento probabilístico que es indiferente al orden temporal, lo que puede crear dificultades psicológicas[55].

Gras y Totohasina (1995) usaron entrevistas y preguntas directas[56] a una muestra de 75 estudiantes de secundaria (17-18 años), después de haber recibido una enseñanza experimental sobre probabilidad condicional e identificaron tres tipos de concepciones erróneas sobre la probabilidad condicional en estudiantes:

- En la *concepción cronológica,* los estudiantes interpretan la probabilidad condicional *P(A/B)* como una relación temporal, donde el evento condicionante *B* siempre precede al *suceso A*.
- En la *concepción causal*, los estudiantes interpretan la probabilidad condicional *P(A/B)* como una relación causal implícita, donde el suceso condicionante *B* es la causa y *A* la consecuencia.
- En la *concepción cardinal,* los estudiantes interpretan la probabilidad condicional *P(A/B)* como la proporción $\frac{\text{Card}\,(A \cap B)}{\text{Card}(B)}$, que es correcta en el caso de un espacio muestral finito equiprobable. Aun así, cuando se maneja un espacio muestral continuo o las probabilidades de los diferentes sucesos no son iguales esta concepción lleva a un obstáculo en el aprendizaje. Otros estudiantes interpretan *P(A/B)* como la proporción $\frac{\text{Card}(A)}{\text{Card}(B)}$, que es siempre falso.

Los autores sugieren que el origen de las dos primeras concepciones erróneas es un obstáculo cognitivo, mientras que la concepción cardinal es inducida por la enseñanza.

---

[55] La falacia del eje temporal puede afectar la resolución de problemas bayesianos, donde la actualización de las probabilidades de sucesos anteriores a la luz de los resultados de otros posteriores juega un papel tan importante.

[56] Usaron las siguientes preguntas:

A Para calcular la probabilidad condicional P(A/B), ¿Debe ocurrir el suceso B cronológicamente antes que el suceso A? Si__ No__ No sabe__

B. Para calcular la probabilidad condicional P(A/B), ¿Debemos asumir que el suceso B es la causa y el suceso A el

4.2.3. INTERCAMBIO DE SUCESOS EN LA PROBABILIDAD CONDICIONAL

Muchos estudiantes no discriminan adecuadamente entre las dos direcciones de la probabilidad condicional *P(A/ B)* y *P(B/ A)* (Falk, 1986, 1989; Rossman y Short, 1995; Tarr y Lannin, 2005)[57]. Pollatsek y cols. (1987) encontraron los siguientes resultados en el siguiente problema: (entre paréntesis el porcentaje de sujetos que marcaron la opción)

Problema 6. ¿Cuál de los siguientes sucesos es más probable?

a) *Que un taxi azul sea correctamente identificado por la noche como un taxi azul* (24%)
b) *Que un taxi identificado por la noche como azul sea realmente un taxi azul* (7%)
c) *Los dos sucesos son igual de probables"* (69%)

Estepa (1993) encontró resultados similares en la interpretación de tablas de contingencia por parte de estudiantes, confundiendo "porcentaje de fumadores que contraen cáncer de pulmón" con "porcentaje de personas con cáncer de pulmón que fuman".

Esta confusión se extiende al contexto de la interpretación de α en los test de significación. El nivel de significación α se define como la probabilidad condicional de obtener un *resultado R* en la región de rechazo cuando la hipótesis nula H_o es cierta $P(R/ H_o)$. Sin embargo cuando un test resulta significativo (lo que quiere decir que *R* ha ocurrido) y alguien pregunta por la probabilidad de haber cometido un error (la probabilidad de que H_0 sea cierta) a menudo se contesta con α. En esta situación se estaría confundiendo $P(R/ H_o)$ con $P(H_o/ R)$ (Birnbaum, 1982; Vallecillos, 1994, 1999; Falk y Greenbaum, 1995; Lecoutre, Lecoutre y Poitevineau, 2001).

4.2.4. CONFUSIÓN ENTRE PROBABILIDAD CONDICIONAL Y CONJUNTA

Tversky y Kahneman (1982d) describieron la *falacia de la conjunción* como la

---

efecto o la consecuencia de B? Si__ No__ No sé__

[57] Este error también se ha visto en contextos médicos confundiendo la probabilidad de tener una enfermedad dado positivo el test de diagnóstico con la probabilidad de un resultado positivo en el test de diagnóstico, dado que se tiene la enfermedad (Eddy, 1982).

creencia de que es más probable la intersección de dos sucesos que uno de ellos. La "falacia" contradice los resultados fundamentales de la teoría de la probabilidad, según el cual la probabilidad de un suceso que es subconjunto de otro ha de ser menor o igual a la probabilidad del suceso que lo contiene (Fiedler, 1988; Agnoli, 1989; Ojeda, 1995; Hertwig y Gingerenzer, 1999; Stolarz-Fantino, Fantino, Zizzo y Wen, 2003).

La regla del producto se viola cuando los juicios de probabilidad se basan en heurísticas tales como la representatividad y disponibilidad. La falacia de la conjunción aparece cuando la tarea carece de un diseño en que la resolución del problema recuerde otros ejemplos que ayuden a resolverlo (Agnoli, 1989).Un ejemplo es el problema 7, propuesto por Tversky y Kahneman (1982d), tanto a estudiantes como a personas de diferentes edades y ocupaciones:

Problema 7. ¿Qué porcentaje de los hombres entrevistados tienen tanto 55 años y además han tenido uno o más ataques al corazón?"

"¿Qué porcentaje de los hombres entrevistados ha tenido uno o más ataques al corazón?"

"De los hombres entrevistados con más de 55 años, ¿qué porcentaje ha tenido uno o más ataques al corazón?"

Según Tversky y Kahneman (1982d), la falacia de la conjunción es resultado de: a) considerar la conjunción como más representativa de la población generadora que cada evento separado (*heurística de la representatividad*[58]) o b) de que la conjunción hace que los sujetos recuerden o imaginen más ejemplos de una categoría o modelo más restringido. Los autores probaron que la falacia de la conjunción se mantiene incluso cuando el diseño de la investigación evita los posibles problemas lingüísticos que dan lugar a la confusión entre distintos tipos de probabilidades.

Fiedler (1988) considera que este error se debe a la interpretación errónea de la palabra "probabilidad" en el enunciado de las preguntas y que la enseñanza de la probabilidad condicional y conjunta simultáneamente es razón de que los alumnos interpreten la probabilidad de intersección de los sucesos como la probabilidad de un suceso condicional. Kahneman y Tversky niegan esta posibilidad, puesto que si el estudiante falla en la regla de la conjunción no podemos esperar que use una regla más compleja. El cálculo de la probabilidad condicional se realiza mediante un cociente entre una intersección y una probabilidad simple, por lo que es más compleja que el cálculo de probabilidad conjunta (Ojeda, 1995).

La falacia de la conjunción es un efecto fuerte, resistente incluso al refuerzo verbal y monetario por las respuestas correctas. El efecto ocurre, tanto con una descripción del contexto, como sin ella y cuando la conjunción consiste en dos enunciados simples. Sin embargo, los enunciados que representan la conjunción de tres sucesos se juzgan correctamente como menos probables que la conjunción de dos (Stolarz-Fantino, Fantino, Zizzo y Wen, 2003).

4.2.5. INFLUENCIA DEL LENGUAJE Y FORMATO

Un problema en las investigaciones sobre probabilidad es la necesidad de utilizar un lenguaje preciso para describir las situaciones aleatorias (Falk, 1986, 18989; Ojeda, 1995). En general, el lenguaje cotidiano es más flexible que el lenguaje analítico o lógico – deductivo (Maury, 1985; Lonjedo y Huerta, 2004, 2005). Una posible explicación para la prevalencia del error de intercambio del suceso condicionante y condicionado es que el lenguaje ordinario, que es el que usamos en el enunciado de los problemas de probabilidad condicional no tiene la suficiente precisión y es por ello ambiguo (Falk, 1989; Pollatsek, Well, Konold y Hardiman, 1987)[59].

Un ejemplo son los enunciados que usan la conjunción "y"que pueden llevar a confundir la probabilidad conjunta y la probabilidad condicional (Einhorn y Hogarth, 1986). Tanto los estudiantes de esta investigación como los de Ojeda (1995) y Lonjedo y Huerta (2004, 2005) encuentran las preguntas sobre intersección de sucesos fueron más difíciles que las de probabilidad condicional y la mitad de los sujetos interpretan la intersección como condicionamiento.

---

[58] La intersección de sucesos es donde hay más contraste entre representatividad y probabilidad, ya que cuanto más especificamos un suceso, su probabilidad es menor mientras que su representatividad es mayor.

[59] Cuando escribimos una probabilidad condicional usando la notación matemática es claro cuál es el suceso condicionante y cuál el condicionado, pero en el lenguaje ordinario la probabilidad condicional (tener cáncer si se es fumador) y su inversa (ser fumador si se tiene cáncer) no siempre se distinguen claramente entre si o de la probabilidad conjunta (ser fumador y tener cáncer).

Pollatesk y cols. (1987) comprobaron mediante dos experimentos que estudiantes sin una preparación estadística fuerte eran capaces de comprender superficialmente el concepto de probabilidad condicional y su direccionalidad, aunque hay algunos factores que pueden interferir con esta habilidad, como la forma en que están expresados los enunciados, el conocimiento del contexto y la confusión entre las nociones de sucesos independientes y sucesos equiprobables. En su experimento incluyeron dos tipos de problemas: problemas probabilísticos y problemas porcentuales, como los problemas 8 y 9. En los problemas, donde el sesgo causal se hacía presente (como el problema madre – hija), la versión porcentual conseguía mayor número de respuestas correctas.

Problema 8. "Indica cuál de los siguientes sucesos es más probable:

a) Que una niña tenga los ojos azules si su madre tiene los ojos azules
b) Que una madre tenga los ojos azules si su hija tiene los ojos azules
c) Los dos sucesos son igual de probables"

Problema 9. "Indica cuál de los siguientes porcentajes es mayor:

a) El porcentaje de madres con los ojos azules que tienen hijas con ojos azules
b) El porcentaje de hijas con los ojos azules que tienen madres con ojos azules
c) Los dos porcentajes son iguales"

4.3. ERRORES EN EL RAZONAMIENTO BAYESIANO

4.3.1. SITUACIONES SINCRÓNICAS Y DIACRÓNICAS

Se pueden distinguir dos tipos de situaciones relacionadas con la probabilidad condicional (Falk, 1986; Gras y Totohasina, 1995; Ojeda, 1995):

- Situaciones sincrónicas: son situaciones estáticas, en las que no subyace una secuencia de experimentos, sino que los experimentos aleatorios se realizan simultáneamente. En el problema 10 (Falk ,1986) la mayor parte de los estudiantes dan como respuesta ½ porque sólo perciben un experimento y suponen que la tarjeta con las dos caras rojas no es relevante para calcular la probabilidad condicional.

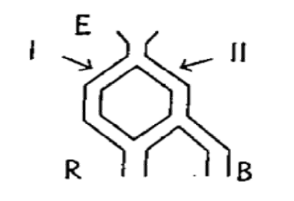
Problema 10. En una urna hay tres tarjetas: una es azul por ambos lados, otra es roja por ambos lados y la última es por un lado azul y por otro roja. Seleccionamos una tarjeta al azar y la ponemos sobre la mesa tal y como la sacamos de la urna. La cara que vemos es roja, ¿cuál es la probabilidad de que la otra cara sea roja también?

- Situaciones diacrónicas: son situaciones en las que hay una clara secuencia temporal, donde se realizan un experimento detrás de otro, como en el ejemplo siguiente:

Problema 11. En una urna hay dos bolas blancas y dos bolas negras. Tomamos una bola blanca de la urna y sin reemplazarla tomamos una segunda bola al azar de la urna. ¿Cuál es la probabilidad de que la segunda bola sea blanca?

Al calcular una probabilidad condicional, es clave que cambiemos el espacio muestral al suceso condicionante. Las situaciones diacrónicas dificultan especialmente el cambio de espacio muestral. Ojeda (1995) plantea el problema 11 a 255 alumnos de Bachillerato, después de estudiar probabilidad condicional. Aunque tanto el problema 8 como el 11, tienen la misma estructura matemática, el 11 se percibe más fácilmente como una secuencia de experimentos[60].

Problema 12. Una bola se suelta en E. Si sale por R, ¿qué es más probable, que haya seguido el canal I, o que se haya ido por el canal II?



4.3.2. RESOLUCIÓN DE PROBLEMAS BAYESIANOS

Los problemas bayesianos fueron investigados por Tversky y Kahneman (1982c) como parte de su trabajo sobre la heurística de representatividad y la falacia de las tasas base. En estos problemas los riesgos y probabilidades se suelen definir en términos de frecuencias relativas y los estudios psicológicos muestran estimaciones incorrectas de estos riesgos (Eddy, 1982; Totohasina, 1992; Teigen, Brun y Frydenlund, 1999; Díaz y De la Fuente, en prensa).

En el problema 13 (Tversky y Kahneman, 1982b) la respuesta típica es 0'80 (estimación que coincide con la fiabilidad del testigo)[61]; otro un grupo importante de sujetos elige 0'50, Al resolver el problema mediante el teorema de Bayes se obtiene que la probabilidad de que el taxi implicado sea Azul es de 0'41. En general hay tres tipos de informaciones relevantes a la predicción estadística: a) las tasas base o información inicial; b) la evidencia específica del caso individual (lo que dijo el testigo); c) la precisión esperada de la predicción (el número de aciertos). Una regla fundamental en estadística es que la precisión esperada está controlada por el peso de la evidencia y la información inicial, pero los sujetos olvidan esta regla.

Problema 13. Un taxi se vio implicado en un accidente nocturno con choque y huida posterior. Hay dos compañías de taxis en la ciudad, la Verde y la Azul. El 85% de los taxis de la ciudad son Verde y el 15% Azul. Un testigo identificó el taxi como Azul. El tribunal comprobó la fiabilidad del testigo bajo las mismas circunstancias que había la noche del accidente y llegó a la conclusión de que el testigo identificaba correctamente cada uno de los colores en el 80% de las ocasiones y fallaba en el 20%.¿Cuál es la probabilidad de que el taxi implicado en el accidente fuera en efecto Azul?

Uno de los factores que incide en la falacia de las tasas base es el de causalidad, ya que si el problema está expresado en términos causales, se reduce el número de respuestas erróneas. También la especificidad mejora la respuesta (Bar – Hillel, 1987).

Totohasina (1992) analiza las estrategias que llevan a respuestas correctas en los problemas bayesianos encontrando las siguientes (problema 5):

- Cambio de referencial: "*En el total de cajas azules o rojas 45/4 +45=225/4 tienen la marca M. En 36 cajas rojas 45/4 llevan la marca. La probabilidad de que la caja con marca sea roja es P(D)= (45/4)/ (225/4)= 1/5".*

---

[60] La proporción de respuestas correctas sube del 25% al 40%, aunque los alumnos todavía siguen con dificultades y no reducen convenientemente el espacio muestral al resolver el problema.

[61] De este error se deduce que las personas ignoran las tasas base.

- Cálculo de cocientes de porcentajes, lo que implica, en la práctica la fórmula de Bayes: "*Porcentaje de bolas con la marca: (45/199)x(25/100)+(60/100)x(75/100) = 56.25%. Por tanto, probabilidad de que la caja con marca sea roja = (45/100)x(25/100) /56.25/100=20 %*".

Después de realizar un experimento de enseñanza de la probabilidad condicional, en el que no se introduce formalmente el teorema de Bayes, aunque se plantean y resuelven problemas de probabilidad inversa basándose en árboles y tablas de doble entrada, propone a los alumnos el siguiente problema:

Problema 14. Una persona se somete a una prueba para detectar una cierta enfermedad. Si la persona está enferma, el test será positivo con 96% de certeza. Si la persona está sana el test será negativo con 94% de certeza. Se sabe también que 1 de cada 100 personas de esta edad está enferma.
a) El test resulta positivo. ¿Cuál es la probabilidad de que la persona esté enferma?
b) El test resulta negativo. ¿Cuál es la probabilidad de que la persona esté sana?

Este problema tiene la misma estructura matemática que el problema 4, aunque el aspecto de causalidad está más marcado. El dato "1 persona de 100" podría inducir el uso de la concepción cardinalista. El autor examina los tipos de diagramas en árbol utilizados por los alumnos. De un total de 65, 26 construyen un árbol directo marcando el aspecto secuencial de los experimentos (cálculo de la probabilidad directa) y 9 llegan al árbol inverso. Otros 7 alumnos usan el árbol, pero no contemplan el aspecto secuencial y no llegan a asignar correctamente las probabilidades. Sólo 9 alumnos llegan a la solución correcta de los problemas.

Para tratar de explicar las dificultades en éste y otros experimentos similares, analiza los procedimientos de los alumnos. El autor supone que los alumnos pueden encontrarse con dificultades en función del tipo de representación elegida para resolver el problema, que les es dado en formato verbal[62]. Otra dificultad es la necesidad de invertir condición y condicionado en los problemas tipo Bayes, ya que los alumnos con frecuencia

[62] Al pasar, por ejemplo, a una tabla de doble entrada, se dificulta la percepción de la naturaleza secuencial de algunos problemas, porque lo que queda más visible es la intersección de los dos sucesos y puede llevar a los alumnos a confundir la probabilidad condicional y la conjunta.

confunden el papel de estos dos sucesos en una probabilidad condicional y por tanto confunden una probabilidad condicional con su inversa. Totohasina (1992) sugiere que este obstáculo de "reversibilidad" también aparece en el aprendizaje de otras nociones matemáticas, por ejemplo al aprender el concepto de primitiva de una función, partiendo del de derivada (Gras y Totohasina, 1995)[63].

4.4. EXPERIMENTOS DE ENSEÑANZA DE LA PROBABILIDAD CONDICIONAL

Diferentes marcos teóricos en psicología han tratado de explicar los errores y falacias en el razonamiento en situación de incertidumbre y tienen diversas implicaciones respecto a la forma en cómo debe llevarse a cabo la enseñanza del tema (Sedlmeier, 1999). A continuación describimos los experimentos de enseñanza llevados a cabo desde estos enfoques y las conclusiones obtenidas.

4.4.1. PROGRAMA DE HEURÍSTICAS Y SESGOS

Este programa se basa en la idea de racionalidad acotada, e interpreta las falacias como resultado de un proceso cognitivo que lleva a una conclusión incorrecta, bien por usar un modelo inapropiado de la situación o por falta de estructuras cognitivas específicas (Kahneman, Slovic y Tversky, 1982; Pérez Echeverría, 1990). En este paradigma, las heurísticas y sesgos se suponen resistentes a la enseñanza. Se comparan las ilusiones cognitivas a las ilusiones ópticas que no pueden ser controladas con facilidad (Tversky y Kahneman, 1982b).

La instrucción específica no ha tenido efecto sobre la falacia de la conjunción en experimentos basados en este marco teórico y los sesgos y falacias descritos por los autores se han encontrado en sujetos con alta preparación matemática (Tversky y Kahneman, 1982d). Estos resultados parecen explicar el escaso éxito de algunos experimentos de enseñanza llevados a cabo desde la matemática o estadística, cuyos autores en general habían aceptado el marco teórico de Kahneman y sus colaboradores.

[63] Un árbol es más efectivo para resolver problemas de probabilidad condicional, sobre todo cuando se refiere a un problema diacrónico (dirigido en el tiempo). Pero puede reforzar las concepciones causalista o cronologista en los alumnos que las manifiestan.

4.4.2. ADQUISICIÓN DE REGLAS ABSTRACTAS

Otra perspectiva es considerar que podemos adquirir un razonamiento estadístico correcto y que desarrollamos reglas abstractas intuitivas, por ejemplo, la Ley de los Grandes Números. Aplicaríamos estas reglas para resolver los problemas cotidianos, siempre que encontrásemos en las situaciones ciertas claves que nos permitan reconocerlas como aleatoria (Nisbett y Ross, 1980). La enseñanza podría mejorar nuestro razonamiento estadístico natural, que se adquiere por la experiencia repetida en resolución de problemas. Por tanto, estas reglas podrían enseñarse mediante ejercitación, aunque el aprendizaje en concreto de la probabilidad condicional requiere, además, enseñanza de la lógica condicional (Sedlemeier, 1999). La enseñanza debiera hacerse comenzando por las reglas abstractas y luego aplicándolas a la resolución de problemas. Los experimentos realizados bajo este paradigma no aportan resultados concluyentes.

4.4.3. FRECUENCIAS NATURALES Y ALGORITMOS ADAPTATIVOS

Gigerenzer (1994) sugiere que la dificultad en la resolución de problemas referidos al teorema de Bayes desaparece cuando las preguntas se plantean en términos de frecuencias. Analiza el experimento de Hamerton (1973) que da el siguiente problema a sus sujetos, quienes dan en promedio una probabilidad de 85% como respuesta (cercana a la tasa base del 90%):

Problema 15: Un dispositivo sirve para identificar una cierta enfermedad. Si alguien está enfermo, hay un 90% de posibilidades de que la prueba sea positiva. Si no está enfermo hay todavía un 1% de posibilidades de que la prueba sea positiva. Aproximadamente el 1% de la población está enferma. Smith pasa la prueba y resulta positiva. La probabilidad de que tenga la enfermedad es:

Gigerenzer (1994) describe otras investigaciones en las que se cambia el enunciado, y en las que se observa un aumento notable de los razonamientos de tipo Bayesiano. El nuevo enunciado es el siguiente:

Problema 16. Uno de cada 1000 americanos tiene una enfermedad. Una prueba ha sido diseñada para detectarla. Cada vez que la persona está enferma, el test da positivo, pero también da positivo en 50 de cada 1000 personas sanas. ¿Cuantas personas de entre aquellas a las que el test dio positivo tendrán realmente la enfermedad?

Gigerenzer sugiere que nuestra mente está mejor equipada para resolver problemas bayesianos cuando la información y las preguntas se dan en términos de frecuencias[64]. En el ejemplo, la persona razona en la siguiente forma: "*En 1000 americanos hay 999 sanos y uno enfermo. Al pasar el test 50 (aproximadamente) de los 999 sanos darán positivos y también el sujeto enfermo. Luego la probabilidad de que el sujeto esté enfermo si el test es positivo es 1/51 porque solo hay un enfermo entre los 51 a los que el test dio positivo*".

Esta teoría parece confirmar el efecto de la instrucción cuando se aplica. En el caso particular de la falacia de la conjunción, Fiedler (1988) encontró una reducción considerable del número de respuestas incorrectas al plantear las preguntas en formato de frecuencias, aunque un experimento organizado para hacer comprender a los sujetos que la probabilidad de la intersección ha de ser menor que la de cada suceso, no tuvo efecto significativo. Esto mismo es comprobado por Ojeda (1995), para el caso de problemas de Bayes, puesto que alrededor del 45 % de estudiantes de secundaria (255 alumnos en la muestra) fueron capaces de resolver el siguiente problema, que requiere un razonamiento bayesiano:

Problema 17. En una ciudad hay 60 hombres y 40 mujeres por cada 100 habitantes. La mitad de los hombres y una tercera parte de las mujeres fuman. Si se selecciona al azar un fumador, ¿qué es más probable, que sea hombre o mujer?

Efecto del uso de representaciones en la mejora del aprendizaje

El término "probabilidad" se interpreta en formas muy variadas en los problemas, mientras que la expresión "frecuencias" conlleva un significado más ajustado al

[64] Llama a este formato *frecuencias naturales* porque se asemeja más a la forma en que recogemos información de las frecuencias de sucesos aleatorios en una situación de *muestreo natural* a lo largo de nuestra experiencia (por ejemplo un médico en su consulta).

matemático (Hertwig y Gigerenzer, 1999)[65]. El método de enseñanza se basa en enseñar a los estudiantes a traducir los enunciados de los problemas bayesianos de un formato probabilístico a otro frecuencial, que les resulta más familiar. Esta estrategia resulta ser mucho más efectiva que la simple enseñanza de reglas abstractas para resolver los problemas (Kurzenhauser y Hoffrage, 2002).

Martignon y Waisner (2002) plantean el uso de una representación especial en forma de diagrama en árbol, junto con las frecuencias naturales para enseñar la resolución de problemas que involucren el Teorema de Bayes. Estos autores indican que, mientras que los problemas de probabilidad que los estudiantes encuentran en la vida diaria son concretos y numéricos, los instrumentos de cálculo que les presentamos (por ejemplo el teorema de Bayes) son altamente formalizados. Los alumnos adquieren reglas nemotécnicas de aplicación que olvidan fácilmente.

Las representaciones juegan un papel primordial en la resolución de problemas matemáticos (Kaput, 1991). Los sistemas de representación son, para este autor, entidades abstractas compartidas que se usan para organizar la información mediante determinadas reglas sintácticas. En este proceso se comienza identificando los sucesos a que se refiere la pregunta del problema y se les denota. A continuación se les organiza para operar con ellos, por ejemplo, sustituyéndolos en una fórmula y finalmente se opera para obtener la solución.

Gigerenzer y Hoffrage (1995) indican que una representación adecuada de los problemas facilita el cálculo y produce soluciones acertadas a los problemas tratados. Apoyándose en esta recomendación, Martignon y Wassner (2002) sugieren el uso de un esquema similar al de la Figura 1.1 para representar los datos del problema y como ayuda en su resolución[66].

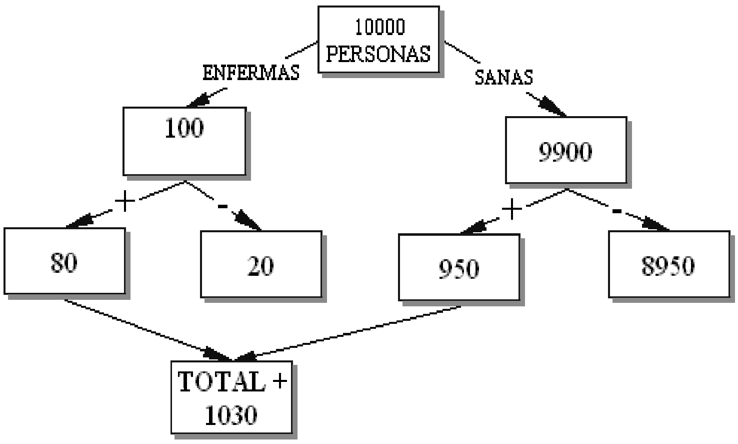
Los autores indican que este éxito se debe a la estrecha relación entre esta representación y la forma inductiva en que procesamos la información en las tareas bayesianas. En primer lugar se divide la muestra (1000 personas en el ejemplo) en función de la tasa base (1 de cada 100) obteniendo una división en dos grupos (10 enfermos y 990 sanos) lo que hace imposible que en el resto del problema olvidemos la

[65] Nuestra mente no está equipada para trabajar con las reglas probabilísticas, ya que estas reglas no tienen en cuenta el contexto del problema, que si es considerado por la mente humana al realizar inferencias pragmáticas y semánticas en situaciones de incertidumbre.

[66] En sus ensayos con estudiantes de los últimos cursos de secundaria, alcanzan casi un 80 % de éxitos con este método.

tasa base. La división en el primer nivel del árbol produce una división binaria (sujetos con y sin la condición). Seguidamente se incluye la información dada por los condicionales y se llega al tercer nivel del árbol que consiste en una segmentación produciendo cuatro sucesos intersección.

Figura 1.1. Representación mediante árbol y frecuencias naturales de un problema de Bayes



A partir de aquí, se suman los sucesos que correspondan a la condición nueva (test positivo) que se identifican fácilmente y se obtiene la cuarta rama que es una nueva división dicotómica de la muestra, para la nueva condición (test positivo o negativo). En este momento es posible resolver el problema de una forma muy sencilla simplemente aplicando la regla de Laplace, puesto que los casos favorables (intersección de enfermos con prueba positiva) y posibles (prueba positiva) se identifican claramente en el diagrama.

Usando la regla matemática de Bayes se realizan las mismas operaciones, pero en un orden diferente, es decir, siguiendo un proceso deductivo, y aunque la mente humana está equipada para ambos tipos de procesos, el primero es más intuitivo y natural.

4.5. JUSTIFICACIÓN DE LA NECESIDAD DE UN CUESTIONARIO COMPRENSIVO DE EVALUACIÓN DEL RAZONAMIENTO SOBRE PROBABILDIDAD CONDICIONAL

Como se analiza en la Sección 4.1, la probabilidad condicional es un concepto básico en inferencia Bayesiana, donde interviene en la definición de las probabilidades finales y verosimilitudes, así como en el Teorema de Bayes, los intervalos de credibilidad y contrastes Es también fundamental en inferencia frecuencial, ya que las distribuciones de los estadísticos en el muestreo, el nivel de significación y potencia en un contraste de hipótesis. Las distribuciones marginales y sus momentos, así como las rectas de regresión, son otros conceptos, estadísticos apoyados en la probabilidad condicional. Es, por tanto, necesario asegurar que el alumno adquiere un conocimiento y competencia básica en probabilidad condicional, antes de avanzar en el estudio de la estadística.

No obstante, la investigación en Educación y Psicología revisada en los apartados anteriores muestra que este no es un tema sencillo, porque tiene una amplia variedad de matices y los alumnos lo asocian con la problemática de la causalidad y temporalidad (Totohasina, 1992),, teniendo dificultad en la percepción de los experimentos compuestos, en situaciones sincrónicas (Totohasina, 1992, Gras y Totohasina, 1995). Se confunde independencia y exclusión (Ojeda, 1995, Sánchez, 1996), se cambian los términos de la probabilidad condicional (Falk, 1986, 1989),, se confunde ésta con la probabilidad conjunta(Lomjedo y Huerta, 2004, 2005) y se asigna a la probabilidad conjunta un valor mayor que a la probabilidad simple, violando las reglas lógicas del cálculo de probabilidades (Eddy, 1982; Tversky y Kahneman (1982 d).,Fiedler, 1988)..

Muchos de estos errores se presentan en estudiantes de Bachillerato o Universidad: Las tablas 1.3 y 1.3 recogen una clasificación de las investigaciones analizadas, según el tipo de problemas que han propuesto a los sujetos de estudio (tabla 3.2) o las propiedades y definiciones cuya comprensión han evaluado (tabla 3.1).

Hay que hacer notar que dichas investigaciones en su mayor parte se han centrado en la comprensión aislada de propiedades específicas, o en la capacidad de cálculo de ciertos tipos muy localizados de problemas[67]. En la mayor parte de los casos la evaluación de la comprensión no tiene en cuenta la enseñanza del tema, o bien las muestras de estudiantes participantes en los estudios son pequeñas y es difícil generalizar los resultados.

No se ha tratado tampoco de relacionar los diferentes sesgos, para ver si los presentan los mismos sujetos, ni de analizar si el conocimiento formal del tema y la capacidad de resolución de problemas complejos, como los que involucran la probabilidad total y Bayes tienen ina influencia en la superación de los errores descritos. La constatación de este hecho sugirió la conveniencia de llevar a cabo un estudio riguroso de evaluación del razonamiento sobre probabilidad condicional de los estudiantes de psicología, después de estudiado el tema, tanto en lo que se refiere al conocimiento formal y la capacidad de resolución de problemas, como en relación a la presencia de los sesgos de razonamiento descritos.

Tabla 1.2. Comprensión de propiedades evaluadas en diferentes investigaciones

Contenidos	Investigaciones que lo han analizado
Restricción del espacio muestral en la probabilidad condicional	Totohasina (1992), Tarr y Jones (1997), Tarr y Lannin. (2005)
Diferencia entre P(A/B) y P(B/A)	Eddy, (1982), Birnbaum (1982), Falk (1986), Pollatsek y cols. (1987), Tversky y Kahneman (1982b), Vallecillos (1994)
$P(A \cap B) < P(B)$	Eddy (1982); Tversky y Kahneman (1982 d). Fiedler (1988); Stolarz-Fantino, Fantino, Zizzo y Wen, 2003
Tiempo y probabilidad condicional	Falk (1989), Totohasina (1992); Gras y Totohasina (1995)
Independencia	Sánchez (1996); Truran y Truran (1996)
Dependencia	Sánchez (1996)
Diferencia dependencia- causalidad	Tversky y Kahneman (1982a); Pollatsek y cols., 1987, Falk, 1989; Totohasina (1992); Gras y Totohasina (1995)
Sucesos mutuamente excluyentes; diferencia con independencia	Sánchez (1996), Kelly y Zwiers (1986)
Probabilidad total	Implícitamente las investigaciones que proponen problemas de Bayes
Teorema de Bayes	Hamerton (1973), Falk (1986), Sánchez (1996), Pollatesk y cols. (1987) Fiedler (1988) Gigerenzer (1994) Ojeda (1995) Martignon y Waisner (2002)
Lectura de frecuencias condicionales y conjuntas en tablas de contingencia	Batanero y cols (1996)

De ello se dedujo la necesidad de construir un instrumento válido y fiable de evaluación, adecuado a los alumnos de psicología, cuya construcción esté basada en el estudio previo de la literatura, la enseñanza de análisis de datos en psicología y en principios metodológicos rigurosos. El instrumento sería potencialmente útil para evaluación de otros estudiantes de cursos de estadísticas en ciencias sociales y educación y para fines de investigación. Permitiría también analizar si los sesgos de razonamiento (falacia de la conjunción, falacia de la condicional transpuesta, falacia del eje de tiempo, etc.) están o no relacionados con el grado de conocimiento formal y

[67] Por ejemplo, Tarr y Jones (1997) se centran en problemas de extracción de urnas con y sin reemplazamiento. Sánchez (1996) en preguntas relacionadas con el concepto de independencia.

capacidad de resolución de problemas.

Tabla 1.3. Tipos de problemas evaluados evaluadas en diferentes investigaciones

Tipos de problema	Investigaciones que lo tratan
1. Calcular una probabilidad condicional en un experimento de muestreo con reposición (regla de Laplace)	Maury (1986), Tarr y Jones (1997), Tarr y Lannin. (2005)
2. Calcular una probabilidad condicional en un experimento de muestreo sin reposición (regla de Laplace)	Maury (1986), Tarr y Jones (1997), Lonjedo y Huerta (2004, 2005), Tarr y Lannin. (2005)
3. Calcular una probabilidad compuesta haciendo uso de la regla del producto en un caso de sucesos dependientes	Maury (1986), Fiedler (1988), Agnoli (1989)
4. Calcular una probabilidad compuesta haciendo uso de la regla del producto en un caso de sucesos independientes	Maury (1986), Sánchez y Hernández (2003)
5. Determinar si dos sucesos son dependientes o independientes	Maury (1986), Sánchez (1996)
6. Diferenciar entre sucesos mutuamente excluyentes y sucesos independientes	Sánchez (1996), Kelly y Zwiers (1986)
7. Resolver problemas bayesianos	Eddy (1982), Totohasina (1992) Hamerton (1973) Gigerenzer (1994) Ojeda (1995) Martignon y Wassner (2002)
8. Resolver problemas en situaciones sincrónicas y diacrónicas	Falk (1986), Sánchez (1996), Ojeda (1995)
9. Resolver problemas en formato probabilístico y formato frecuencial	Hamerton (1973) Pollatsek y cols. (1987) Fiedler (1988) Gigerenzer (1994) Ojeda (1995) Martignon y Wassner (2002)

En los capítulos 4 y 5 se aborda la construcción de dicho instrumento, partiendo del análisis de los textos utilizados en la enseñanza de estadística en psicología y el estudio previo de la literatura, diversas pruebas piloto, juicio de expertos, análisis de fiabilidad y validez. En el Capítulo 6 se lleva a cabo un estudio detallado de evaluación en una muestra de 414 estudiantes de psicología después de estudiar el tema, lo que permitirá diagnosticar las principales dificultades y errores que se mantienen después de la enseñanza (estudio 7) y tenerlos en cuenta en el diseño y evaluación del material para enseñanza de inferencia bayesiana (estudios 8 y 9).

CAPITULO 2.
OBJETIVOS Y ETAPAS DE LA INVESTIGACIÓN

1. INTRODUCCIÓN

La síntesis realizada en el Capítulo 1 sobre la problemática del uso incorrecto de la estadística sugirió la necesidad de introducir nuevas herramientas metodológicas en el campo de la Psicología y otras ciencias sociales, para poder superar la situación descrita. Esta introducción debe fundamentarse en un doble tipo de investigación:

- Investigación metodológica, consistente en la reformulación, adaptación y aplicación de los nuevos métodos a problemas específicos de investigación en dichas ciencias.
- Investigación didáctica, que apoye la enseñanza de los citados métodos a los futuros profesionales, pues sólo mediante la formación de los licenciados es posible el cambio metodológico.

Respecto al primer tipo, en esta Memoria se realiza una reformulación bayesiana de la Teoría Clásica de los Tests (TCT), analizando algunas posibles implicaciones de este cambio de perspectiva sobre los métodos de estimación de las características psicométricas de los tests y los ítems que los componen. Asimismo se ponen a punto algunos programas de cálculo que permiten realizar de forma sencilla la estimación de dichas características desde una perspectiva bayesiana. Todo ello es objeto del Capítulo 3.

Se aborda también una investigación didáctica relacionada con la enseñanza de métodos elementales de inferencia bayesiana a estudiantes de primer curso de Psicología[1]. Para ello se ponen a punto los materiales necesarios, se realiza un experimento de enseñanza y se evalúa la efectividad de la propuesta en una serie de estudios relacionados que se incluyen en el Capitulo 6.

Debido a la importancia del concepto de probabilidad condicional dentro de la inferencia bayesiana[2], se considera necesario partir de un estudio previo de evaluación

[1] En De la Fuente, Díaz, y Cañadas. (2005) argumentamos la necesidad de llevar a cabo esta investigación, basándonos en la viabilidad de aplicación, posibilidad de comprensión por parte de los estudiantes y ventajas en diferentes situaciones respecto a los métodos frecuenciales.

[2] Importancia que se deduce del estudio teórico llevado a cabo en el Capítulo 1, Sección 4.

del razonamiento condicional en estudiantes de psicología e incorporar posibles sesgos y dificultades para tenerlos en cuenta en la enseñanza planificada[3].

Dicho estudio de evaluación requiere la construcción previa de un cuestionario, que se denominó RPC (Razonamiento sobre Probabilidad Condicional), ya que no se encontró uno adecuado en las investigaciones previas[4]. El cuestionario RPC será construido según un riguroso proceso metodológico, que se detalla en los Capítulos 4 y 5. También se completa la investigación metodológica, aplicando los métodos bayesianos y programas de cálculo puesto a puntos en el Capítulo 3 en dicha construcción.

A continuación se resume la investigación, presentando sus objetivos e hipótesis previas y describiendo la forma en que ha sido estructurada.

2. OBJETIVOS Y SU INTERÉS

En consecuencia con lo expuesto, se pueden identificar cuatro objetivos principales del trabajo:

- *Objetivo 1. Reformular la Teoría Clásica de los Tests (TCT) desde un punto de vista bayesiano, analizando las implicaciones de este cambio de perspectiva sobre los métodos de estimación de algunas características psicométricas de los ítems y del cuestionario en su conjunto.*

En el Capítulo 3, partiendo de un análisis de los objetivos y métodos en inferencia bayesiana[5] se analizarán las implicaciones de la perspectiva bayesiana sobre la estimación de puntuaciones medias y diferencia de puntuaciones medias en el test, índices de dificultad y discriminación. Para cada uno de ellos se estudiarán los casos de distribución inicial informativa y no informativa, comparando los resultados con la estimación frecuencial y preparando programas que permitan realizar los cálculos.

---

[3] Ya que si los estudiantes tuviesen dificultades generalizadas de comprensión de la probabilidad condicional y teorema de Bayes, sería muy difícil intentar con ellos la enseñanza de métodos de inferencia bayesiana.

[4] Dicho estudio se recoge en la sección 4 del capítulo 1.

[5] (siguiendo entre otros los textos de Box y Tiao, 1992; Bernard, 1998; Berry, 1995; Bernardo, 2003; Bolstad, 2004).

Este estudio tiene utilidad en la construcción, adaptación y validación de otros cuestionarios, especialmente en las situaciones en la que se disponga de información inicial de las características psicométricas de los mismos.

- *Objetivo 2. Aplicar el análisis anterior en el proceso de construcción de un cuestionario, comparando los resultados de la estimación clásica y bayesiana de algunas de las características psicométricas citadas.*

La construcción del cuestionario RPC, partirá de la definición semántica de la variable objeto de medición mediante análisis de contenido de textos de estadística dirigidos a estudiantes de psicología[6]. El proceso de construcción se recoge en el Capítulo 4 y seguirá las normas psicométricas recientes (APA, AERA y NCME, 1999). Se incluirán pruebas pilotos de ítems, juicio de expertos para fijar el contenido y seleccionar los ítems, pruebas piloto y nuevo juicio de expertos para mejorar la redacción de los ítems.

Asimismo se realizará un proceso de validación y estudio de fiabilidad con una muestra de estudiantes de psicología de diversas universidades españolas (Capítulo 6). Todo este proceso se complementará con la aplicación de métodos bayesianos en la estimación de diversas características psicométricas[7]. El cuestionario RPC podrá servir en el futuro, tanto para la evaluación en procesos de instrucción, como para usos en investigación didáctica.

- *Objetivo 3. Realizar un estudio de evaluación del razonamiento sobre probabilidad condicional de los estudiantes de psicología con el cuestionario RPC para decidir sobre la viabilidad de la enseñanza de la inferencia bayesiana a estos estudiantes*

Este objetivo se aborda en el estudio 7, donde se aplicará el cuestionario RPC a una muestra amplia de estudiantes de Psicología de primer curso de diferentes universidades. Se analizarán las respuestas obtenidas de la pasación del cuestionario desde diversos puntos de vista.

---

[6] El análisis de los textos se ha presentado en un grupo de trabajo sobre didáctica de la probabilidad en el Congreso Iberoamericano de Investigación Matemática (Díaz. y Batanero, 2005).

[7] Resultados parciales de la construcción del cuestionario se publicaron en Díaz (2004), Díaz y de la Fuente (2005c; 2006).

Se pretende comprobar que los estudiantes tienen un conocimiento suficiente para abordar la enseñanza de la inferencia bayesiana, y tener en cuenta los posibles sesgos de razonamiento y errores detectados para posteriormente diseñar el material de enseñanza[8].

- *Objetivo 4. Preparar materiales didácticos para la enseñanza de la inferencia bayesiana a estudiantes de psicología que se apoyen en el estudio de evaluación previo y evaluar su efectividad.*

Una vez analizados los conocimientos sobre probabilidad condicional que tienen los estudiantes de Psicología podremos saber en qué punto debe iniciarse nuestra enseñanza. El material didáctico será elaborado a partir del análisis de diversos materiales de estadística bayesiana y será revisado un grupo de expertos en la materia en cuanto a aspectos de contenido incluido como de aplicabilidad a la población objetivo de la enseñanza. Asimismo se prepararán una serie de instrumentos de evaluación de los conocimientos adquiridos en cada sesión y globalmente, que corresponda a los principales objetivos y contenidos de aprendizaje.

Para valorar si la enseñanza de este material es factible, se experimentará en un grupo de estudiantes y se evaluarán los resultados de la enseñanza (estudios 8 y 9).

3. HIPÓTESIS INICIALES

Son diversas las hipótesis que nos planteamos. Podemos agruparlas en: hipótesis relativas al estudio de evaluación e hipótesis relativas el estudio didáctico.

3.1. HIPÓTESIS DEL ESTUDIO DE EVALUACIÓN

Hipótesis 1. *Los alumnos en Psicología adquieren suficientes conocimientos lógico – matemáticos sobre probabilidad condicional, condición necesaria para abordar el estudio de la inferencia bayesiana[9].*

---

[8] Resultados del estudio de evaluación se han publicado en Díaz (2005), Díaz y de la Fuente (2005c; en prensa) y Díaz, de la Fuente. y Batanero (2004b).

[9] En el capítulo 1, sección 4.5 se justifica suficientemente la importancia de la comprensión de la probabilidad condicional para abordar el aprendizaje de la inferencia bayesiana.

Los estudios de Tarr y Jones (1997) indican una progresión en la comprensión de la probabilidad condicional en alumnos de secundaria hasta alcanzar el nivel más alto en su taxonomía, especialmente en aquellos que han recibido instrucción sobre el tema[10]. Sin embargo, dichos estudios se han centrado en aspectos puntuales o problemas simples.

La mayoría de los estudiantes de Psicología ha recibido algún tipo de enseñanza de probabilidad simple y condicional en secundaria, por tanto se considera que un segundo contacto con el tema en la asignatura "Análisis de datos" producirá el efecto citado por estos autores y se considera necesario comprobar si este aprendizaje se extiende a una comprensión más global y resolución de problemas más complejos.

Hipótesis 2. *Determinados sesgos de tipo psicológicos son resistentes a la enseñanza tradicional que se realiza en las universidades, por tanto, los alumnos, incluso después de estudiar el tema de probabilidad condicional, continúan presentando dichos sesgos.*

La prevalencia de los sesgos relacionados con la probabilidad condicional ha sido mostrada por Maury (1986), Falk (1986; 1989), Pollatesk et al. (1987), Totohasina (1992), Gingerenzer (1994) y Sánchez (1996) entre otros. Pero estos autores han estudiado estos sesgos aisladamente, sin tratar de relacionarlos y el efecto de la instrucción no ha sido estudiado adecuadamente.

La hipótesis es que los sesgos se mantienen después de la enseñanza, incluso en alumnos con buena comprensión lógico – matemática, porque la enseñanza actual no confronta a los estudiantes con dichos sesgos[11].

3.2. HIPÓTESIS DEL ESTUDIO DIDÁCTICO

Hipótesis 3. *En un tiempo limitado de enseñanza es posible conseguir un aprendizaje de conocimientos elementales sobre inferencia bayesiana, dentro de una dinámica habitual de clase, con estudiantes de primer año de Psicología.*

---

[10] Resultados similares sobre el efecto de la instrucción son citados por Ojeda (1995), Sedlemeier (1999) y Martignon y Wassner (2002).

[11] Nissbet y Ross (1980) indican esta confrontación como paso necesario para que las personas se hagan conscientes de los sesgos y los superen.

Esta hipótesis se apoya en la existencia de diversos materiales didácticos (Albert y Rossman, 2001; Berry, 1995), así como escritos de algunos autores (Lecoutre, 1996; 1999 y 2006; Bolstad, 2002) que indican la posibilidad de introducir estos temas a estudiantes de primeros cursos universitarios. Los citados trabajos, sin embargo, no informan con detalle qué porcentaje de objetivos se alcanzan, ni en cuales en concreto presentan dificultad. Sería necesario también adaptar los materiales didácticos en cuanto al lenguaje, en cuanto a los ejemplos utilizados y el tiempo disponible[12].

Hipótesis 4. *Las principales dificultades que se esperan encontrar en los estudiantes, tras la instrucción, no serán específicas de la inferencia bayesiana, sino que serán de tipo lógico o estarán relacionadas con otros temas.*

Esta hipótesis se apoya en el hecho de que las ideas básicas en inferencia bayesiana son sencillas y se basan especialmente en el teorema de Bayes y la sucesiva aplicación del mismo. Pero la aplicación de estos métodos requiere el conocimiento de otros conceptos previos de estadística descriptiva, variable aleatoria y probabilidad, cuyo conocimiento insuficiente pudiera provocar algunos fallos.

Hipótesis 5. *Se espera una correlación entre los resultados de los estudiantes en el cuestionario RPC y la evaluación del aprendizaje de inferencia bayesiana.*

Esta hipótesis es consecuencia inmediata de la anterior. Se espera una correlación estadísticamente significativa entre las puntuaciones en ambos instrumentos de evaluación, así como entre ítems relacionados con el teorema de Bayes. La comprobación de esta hipótesis aportaría, además, nuevas evidencias de validez predictiva al cuestionario RPC, al utilizar la prueba de Aprendizaje de Inferencia Bayesiana (AIB) como criterio externo.

4. ETAPAS DE LA INVESTIGACIÓN

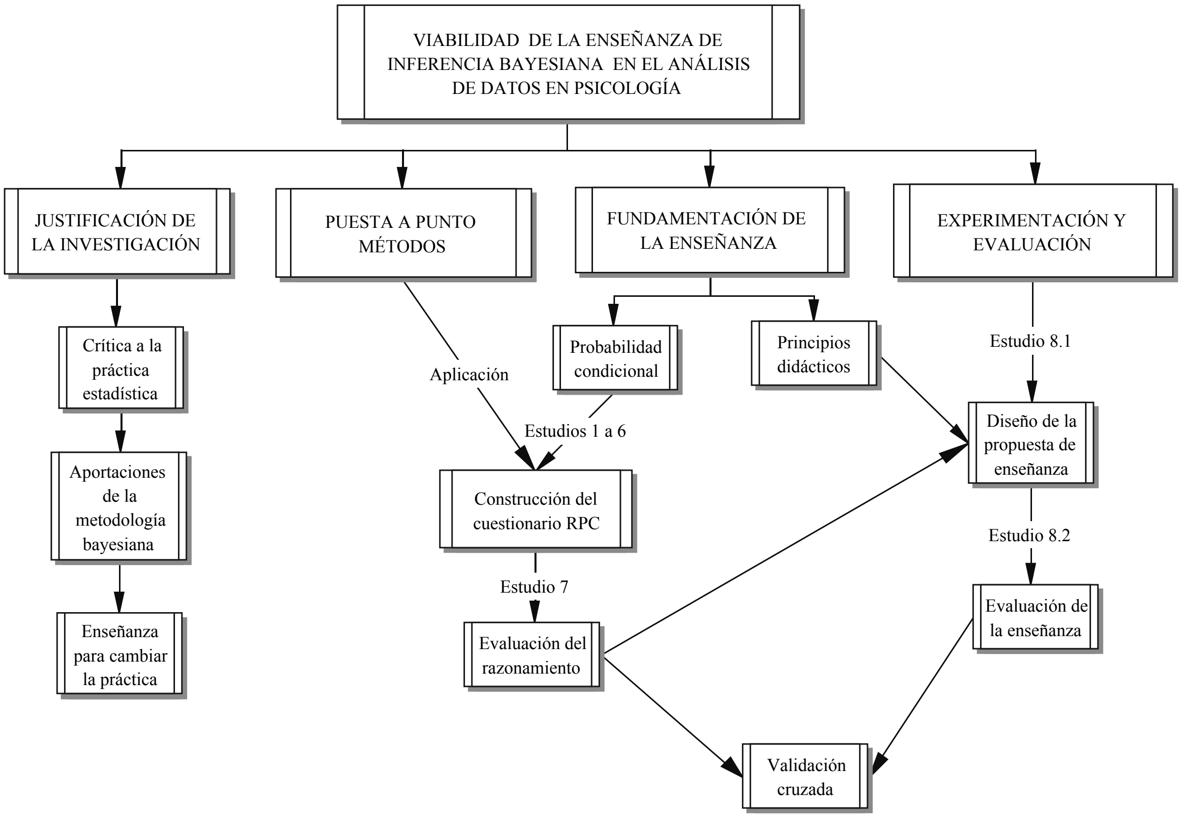
La investigación se ha organizado en una parte teórica (justificación y puesta a punto de métodos) y una parte empírica. Esta segunda parte se divide a su vez en

[12] Los materiales citados están en ingles y francés, son excesivamente largos y presentan ejemplos no contextualizados en el campo de la Psicología.

fundamentación, que incluye la construcción del cuestionario RPC y un estudio de evaluación del razonamiento probabilístico condicional en los estudiantes y experimentación /evaluación de una propuesta de enseñanza. Se organiza en una serie de Estudios que se relacionan entre sí como se muestra en el esquema de la figura 2.1.

Seguidamente se describen brevemente estas etapas y sus características.

Figura 2.1. Esquema de las etapas de la investigación



4.1. ESTUDIOS DE TIPO TEÓRICO Y DE SÍNTESIS

El primer bloque teórico lo compone la Justificación de la investigación, que se ha presentado en el Capítulo 1 y consta de tres partes esenciales:

- La síntesis de las actuales críticas a la práctica estadística, y el análisis de las razones filosóficas y psicológicas de la situación citada (Sección 2 del Capítulo 1).
- El análisis de las razones que justifican la inferencia bayesiana como una alternativa interesante y razonable, tanto desde el punto de vista metodológico como didáctico

(Sección 3 del Capítulo 1) y la defensa de la necesidad de cambio de la enseñanza para mejorar la práctica de los métodos de investigación (Sección 3.5.2, Capítulo 1).

- Justificación de la importancia de la probabilidad condicional en la comprensión de la inferencia y de la necesidad de construcción de un cuestionario comprensivo de evaluación del razonamiento sobre probabilidad condicional (Sección 4, Capítulo 1).

Un segundo bloque consiste en una reinterpretación de la Teoría Clásica de los Tests desde un punto de vista bayesiano (Capítulo 3), conservando los supuestos esenciales de esta teoría, pero considerando los parámetros del Tests e Ítems como variables aleatorias. Como consecuencia se aplican algunos métodos bayesianos de estimación a la determinación de las distribuciones finales de tales parámetros, tanto en el caso de distribución inicial no informativa como informativa. Se comparan los resultados de ambos métodos y se ponen a punto subrutinas de cálculo para aplicar cómodamente los métodos bayesianos a la estimación de estos parámetros. Dichos resultados y programas se aplicarán a lo largo del trabajo, en el proceso de construcción y validación del cuestionario RPC, así como en los procesos de evaluación descritos en el Capítulo 6.

4.2. ELABORACIÓN DE UN CUESTIONARIO Y ESTUDIO DE EVALUACIÓN

Una parte importante de la fundamentación de la propuesta de enseñanza se basa en un estudio de evaluación del razonamiento sobre probabilidad condicional de los estudiantes de psicología, siendo para ello preciso la construcción del cuestionario RPC, en cuyo proceso de construcción, se aplica, asimismo los resultados de los métodos bayesianos analizados en el Capítulo 3.

La construcción del cuestionario RPC se organiza mediante 4 estudios consecutivos, que se resumen en la Tabla 2:

- Estudio 1. Definición semántica de la variable objeto de medición para lo cual se realizó un análisis de contenido de los libros de texto usados en la asignaturas "Análisis de datos" en las facultades de Psicología españolas.
- Estudio 2. Se compone de dos subestudios:
    - Estudio 2.1. Elaboración del banco inicial de ítems y ensayos piloto de ítems con alumnos de Psicología

Tabla 2.1. Métodos empleados en los diferentes estudios que componen la elaboración del cuestionario

Estudios	Descripción	Muestras	Material	Análisis	Resultados
1	Especificación del contenido de la variable	20 libros, elegidos por consulta a 31 universidades Todos los citados por al menos 4 universidades	Capítulos de probabilidad condicional en libros de texto	Análisis de contenido	Primera tabla de especificación de la variable
2	Construcción del cuestionario RPC				
2.1.	Elaboración de un banco inicial de ítems y ensayos de ítems	Muestras de entre 50 y 120 alumnos de psicología	50 ítems, seleccionados de la literatura	Estimación clásica y bayesiana de índices de dificultad; Respuestas a los distractores; resúmenes estadísticos y prueba de normalidad de la distribución de índices de dificultad	Características psicométricas de los ítems
2.2.	Juicio de expertos n.1	10 expertos	50 ítems, seleccionados de la literatura Cuestionario a expertos 1	Puntuación en contenidos Puntuación en concordancia Ítem-contenido Concordancia entre jueces	Cuestionario piloto Ítems seleccionados según grado de acuerdo jueces y características psicométricas
3.	Prueba piloto del cuestionario RPC	57 alumnos psicología 37 alumnos matemáticas	Cuestionario piloto (dos versiones)	Estimación clásica y bayesiana de índices de dificultad e índices discriminación como correlación ítem-escala ydiferencia grupos; resúmenes estadísticos y prueba de normalidad de la distribución de índices de dificultad. Fiabilidad (Alfa; codificación), coeficientes generalizabilidad Validez contenido (análisis contenido; juicio expertos)	Características psicométricas índices Primera aproximación fiabilidad Primera aproximación validez (contenido)
4.	Revisión del instrumento piloto mediante juicio de expertos		Cuestionario a expertos 2 3 versiones diferentes para cada ítem del cuestionario piloto	Puntuación a grupos de ítems (según propiedades formales) Concordancia entre jueces (pruebas no paramétricas)	Cuestionario RPC final Ítems seleccionados según grado de acuerdo jueces

- Estudio 2.2. Juicio de expertos (investigadores en didáctica de la probabilidad) para validar la tabla de especificaciones de la variable y seleccionar los ítems del cuestionario. Con ello se hace una primera justificación de la validez de contenido del cuestionario.

- Estudio 3. Prueba piloto del cuestionario, estimando características psicométricas de los ítems y aportando una primera estimación de la fiabilidad de consistencia interna.
- Estudio 4. Revisión del instrumento mediante juicio de expertos (metodólogos). Con ello se completa la construcción del cuestionario.

El proceso de validación del cuestionario RPC se compone de los estudios 5 y 6 (Tabla 2.2)

- El estudio 5 aborda la estimación de la fiabilidad mediante los métodos de consistencia interna, prueba repetida y generalizabilidad.
- Estudio 6. Aborda la validación del cuestionario con un estudio de validez referida a criterio y validez de constructo.

Tabla 2.2. Métodos empleados en los diferentes estudios que componen la validación del cuestionario

Estudios	Descripción	Muestras	Material	Análisis	Resultados
5.	Fiabilidad del cuestionario RPC	591 alumnos 110 alumnos para retest	Cuestionario RPC	Fiabilidad consistencia interna (Alfa), Theta de Carmines; Omega de Heise y Bohrnsted: Fiabilidad prueba repetida (estimación clásica y bayesiana) Generalizabilidad	Evidencias de fiabilidad del cuestionario RPC
6.	Validación del cuestionario RPC	208 alumnos después de la enseñanza 177 alumnos antes de enseñanza (validez criterio); 591 sujetos (validez constructo)	Cuestionario RPC	Comparación medias (clásica y bayesiana) Análisis discriminante Análisis factorial	Evidencias de validez del cuestionario RPC: Validez discriminante Validez de constructo Validez predictiva

4.3. ESTUDIOS DIDÁCTICOS

- La última parte del trabajo empírico consiste en la elaboración de un material de enseñanza fundamentado en todos los estudios anteriores y su evaluación. Consta de los siguientes estudios (Tabla 2.3):

Tabla 2.3. Métodos empleados en los diferentes estudios que componen la elaboración de la propuesta didáctica

Estudios	Descripción	Muestras	Material	Análisis	Resultados
7	Evaluación de la comprensión de la probabilidad condicional en alumnos de psicología	414 alumnos después de la enseñanza	Cuestionario RPC	Análisis de respuestas Estudios correlacionales; estimación clásica y bayesiana de índices de dificultad	Determinación de errores más frecuentes y su interrelación
8	Evaluación de una propuesta de enseñanza de conceptos elementales de inferencia bayesiana en psicología				
8.1	Diseño y revisión de la propuesta didáctica	10 jueces	Material didáctico Programas de cálculo	Juicio de expertos Grado de acuerdo jueces Comentarios cualitativos para mejora material	Propuesta didáctica revisada (versión presencial y a distancia) Programas de cálculo para el alumno
8.2.	Evaluación del aprendizaje de conceptos bayesianos en una experiencia didáctica	78 alumnos	Cuestionarios de autoevaluación Pruebas abiertas de resolución de problemas Evaluación final (AIB)	Estimación clásica y bayesiana de índices de dificultad en pruebas parciales y evaluación final Análisis implicativo de ítems	Índices de dificultad para los diferentes ítems de evaluación Interrelación entre variables
9.	Interrelación entre razonamiento condicional y aprendizaje de inferencia bayesiana	57 alumnos	Cuestionario RPC AIB	Estudio de correlacionales paramétricos y no paramétricas	Evidencias de validez predictiva Confirmar la importancia del razonamiento condicional en el aprendizaje de inferencia bayesiana

- Estudio 7. Evaluación del razonamiento sobre probabilidad condicional en una muestra de estudiantes de psicología de diversas universidades, una vez estudiado el tema y categorización de las principales dificultades encontradas para tenerlas en cuenta en la elaboración del material.

- Estudio 8. Se compone a su vez de dos partes:
  - Estudio 8.1. Elaboración del material didáctico que incluye apuntes para los estudiantes (versión presencia y a distancia), ejemplos y ejercicios, actividades de autoevaluación, problemas abiertos para resolver y programas auxiliares de cálculo.
  - Estudio 8.2. Experimentación de este material en un grupo de estudiantes con un total de 12 horas de enseñanza, análisis de las tareas realizadas por los alumnos durante la experiencia y de los resultados finales de una prueba de evaluación.
- Estudio 9. El objeto de este estudio es correlacionar las puntuaciones de los alumnos participantes en la prueba de Aprendizaje de Inferencia Bayesiana (AIB) y el cuestionario RPC con la finalidad de aportar nuevas evidencias de validez para dicho cuestionario.

Todos estos estudios y sus resultados se detallan en los capítulos que siguen.

CAPITULO 3.

UNA APROXIMACIÓN BAYESIANA A LA TEORÍA CLÁSICA DE LA MEDICIÓN

1.INTRODUCCIÓN

En este capítulo se realiza un estudio teórico de la teoría clásica de los tests desde un punto de vista bayesiano, con la finalidad de aplicar los resultados obtenidos en los diferentes estudios empíricos que se presentan en esta Memoria. Ello permitirá mejorar desde varios puntos de vista las inferencias frecuenciales realizadas sobre las características psicométricas del cuestionario RPC, cuya construcción y proceso de validación se describen en los capítulos 4 y 5 así como en el estudio 7 de evaluación llevado a cabo con dicho cuestionario.

En lo que sigue se analizan con detalle las implicaciones de la incorporación de la metodología bayesiana en la estimación de puntuaciones medias y diferencias de puntuaciones medias, índices de dificultad y discriminación. Todos estos parámetros característicos de un test, que en la teoría clásica se contemplan como constantes[1], son aquí analizados como variables aleatorias con una distribución inicial de probabilidades, lo que es típico del enfoque bayesiano (Rouanet, 1998b; Lecoutre, 1999; 2006).

Todos estos puntos son abordados en la parte teórica del estudio. Asimismo se desarrollan algunos programas de cálculo en Excel[2] necesarios para poder aplicar el estudio teórico al análisis de los datos del cuestionario RPC. Cada uno de estos programas se preparó en forma de tabla en la que se introdujesen simultáneamente los datos requeridos para todos los ítems del cuestionario, que son obtenidos mediante el Paquete estadístico SPSS u otro similar. Los programas pueden ser adaptados con facilidad al análisis de otros cuestionarios con mayor o menor número de ítems, sin más que variar el número de filas[3].

A continuación se realiza una reformulación del modelo lineal en la teoría clásica de los test desde un punto de vista bayesiano y se describen los objetivos y métodos básicos de estimación en inferencia bayesiana. Seguidamente se aplica lo anterior al estudio de diversas características psicométricas del test y su estimación, diferenciando

[1] Esto es típico de la inferencia frecuencial

[2] Se decidió usar Excel, porque está disponible en los ordenadores personales, y proporciona suficientes herramientas para programar los cálculos necesarios.

[3] Se puede aumentar el número de filas copiando una de las filas y repitiendo la acción de pegar; puesto que las celdas constantes han sido programadas para que se mantengan fijas. Los programas están protegido, pero el autor puede facilitar una adaptación para otros cuestionarios.

en cada caso los supuestos de distribución inicial no informativa e informativa, describiendo también los programas de cálculo elaborados.

2. FORMULACIÓN BAYESIANA DEL MODELO LINEAL EN LA TEORÍA CLÁSICA DE LA MEDICIÓN

2.1. SUPUESTOS BÁSICOS DE LA TEORÍA CLÁSICA

El modelo lineal clásico de la teoría de los test (Thorndike, 1989; Muñiz, 1994; Martínez Arias, 1995; Meliá, 2001) fue formulado por Spearman (1904). En dicho modelo, la puntación empírica X obtenida para un sujeto en el total de una prueba es una variable aleatoria suma de dos componentes: la puntuación verdadera (V) del sujeto en ese test o prueba, que se supone constante, y el error de medida (e). El modelo hace las siguientes hipótesis (Muñiz, 1994):

$$X=V+e$$
$$E(X)=V$$

En las expresiones anteriores $E(X)$ es la esperanza matemática o media de la variable aleatoria X. Este supuesto define la puntuación verdadera, es decir, la puntuación media que se obtendría si aplicásemos repetida e indefinidamente la prueba al sujeto. Se supone también que e es una variable aleatoria que sigue una distribución normal de media cero (es decir, los errores positivos y negativos se compensan) (Carmines y Zeller, 1979). Fijada una persona i, su puntuación verdadera en la prueba sería constante V_i; pero su puntuación observada X_i y el error de medida e_i serían variables aleatorias. Se asumen los siguientes puntos (Martínez Arias, 1995):

1. $E(e_i)=0$, tanto para la población de personas medidas, como para infinitas repeticiones de la prueba en una persona. Se supone que los errores siguen una distribución normal.
2. $\rho(V,e)=0; \rho(e_i,e_j)=0$. Se asume que el error de medida no está correlacionado con la puntuación verdadera (el supuesto es más amplio; se asume que el error es independiente de la puntuación empírica obtenida) y los errores de medida de diferentes sujetos son también independientes.

De estos supuestos se derivan las siguientes igualdades:

$$E\,(X)=\mu_X=E\,(V)=\mu_V$$

$$\sigma_X^2=\sigma_V^2+\sigma_e^2$$

$$\rho_{XV}^2=\frac{\sigma_V^2}{\sigma_X^2}=1-\frac{\sigma_e^2}{\sigma_X^2}=1-\rho_{Xe}^2$$

2.2. OBJETIVOS Y MÉTODOS DE ESTIMACIÓN EN INFERENCIA BAYESIANA

La hipótesis básica de la inferencia bayesiana es considerar los parámetros a estimar como variables aleatorias, y el objetivo de la inferencia es actualizar la distribución inicial de dichos parámetros para llegar a una distribución final, que depende, tanto de la distribución inicial, como de los datos observados. Esta distribución final sería la esencia del método bayesiano de estimación, pues contiene toda la información del parámetro, condicional con los datos observados (Lecoutre, 1999; 2006; Lee, 2004; O'Hagan y Forster, 2004).

Dado un conjunto de datos de una muestra $y' = (y_1,..., y_n)$, cuya distribución de probabilidad $p(y/\theta)$ depende de uno o varios parámetros $\theta = (\theta_1,..., \theta_k)$, la inferencia bayesiana asume que θ tiene su propia distribución de probabilidad $p(\theta)$. Entonces la distribución condicional de θ dados los datos observados y viene dada por (Box y Tiao, 1992):

$$p(\theta/y)=\frac{p(y/\theta)p(\theta)}{p(y)}$$ [4]

siendo $p(y)=\sum p(y/\theta)p(\theta)$ a través de todo el rango admisible de valores de θ. Simplificadamente, esta expresión también se denota por:

$$p(\theta/y)=\propto p(y/\theta)p(\theta)$$

En la expresión anterior $\propto$ indica una relación de proporcionalidad, es decir, la distribución final es proporcional al producto de la distribución inicial por la

[4] Esta expresión no es otra cosa que el teorema de Bayes.

verosimilitud, siendo el coeficiente de proporcionalidad el numerador del Teorema de Bayes (Box y Tiao, 2002; Serrano, 2003; Lee, 2004).

En esta expresión, *p(θ)* es la *distribución inicial de θ* que indica el conocimiento sobre θ antes de tomar los datos y se puede usar para representar tanto el conocimiento previo como el desconocimiento relativo previo. *p(y/θ)* es la *función de verosimilitud de θ dado y* a través de la cual los datos **y** modifican el conocimiento inicial de *θ*. Desde el punto de vista subjetivo *p(y/θ)* mide el grado de creencia del investigador en el valor *y*, teniendo en cuenta la distribución inicial especificada. No es, por tanto una hipótesis sobre el modelo que genera los datos, sino una expresión explícita del investigador respecto a su creencia sobre dicho modelo (O'Hagan y Forster, 2004).

Finalmente, *p(θ/y)* contiene el conocimiento sobre *θ* una vez conocidos los datos y se denomina *distribución final de θ*[5]. Mientras diversos investigadores podrían tener desacuerdos sobre la distribución inicial, ya que esta expresa la creencia de cada uno antes de recoger los datos, hay un acuerdo básico respecto a la función de verosimilitud, es decir a la función que liga el parámetro con los datos (Lee, 2004).

El teorema de Bayes proporciona una formulación matemática de cómo el conocimiento previo puede combinarse con el conocimiento nuevo (Box y Tiao, 1992) y permite actualizar continuamente la información sobre un conjunto de parámetros θ a medida que se toman más datos[6]. Al aplicar sucesivamente el teorema, la distribución final del paso anterior se convierte en distribución inicial en cada paso: Al ser independientes los datos *z* e *y* se tiene (Lee, 2004):

$$p(\theta / z, y) = \propto p(z / \theta) p(\theta / y) p(\theta) = \propto p(\theta / z) p(\theta / y)$$

La aplicación sistemática del teorema de Bayes y de estos principios constituyen el método principal de la inferencia bayesiana, cuyo objetivo básico es actualizar la distribución inicial de los parámetros (Bernardo, 2003). Es también lo suficientemente

[5] El teorema de Bayes calcula distribución final de θ, que es proporcional al producto de la distribución inicial de θ a y la verosimilitud de θ dado *y*. El denominador en la fórmula asegura que la distribución a posteriori sume uno.

[6] El teorema de Bayes describe el proceso de aprendizaje y muestra cómo el conocimiento sobre el estado del problema de estudio, representado por *θ*, es modificado continuamente a medida que disponemos de nuevos datos (Berry, 1995; Bernardo, 2003).

flexible para reflejar la complejidad de la práctica científica, por ejemplo, para llevar a cabo procedimientos de estimación y contraste (Bolstad, 2004; Lecoutre, 2006).

Una diferencia fundamental con la inferencia clásica será el carácter subjetivo[7] (y no frecuencial) de las probabilidades. Por tanto no se plantea el problema de muestreo repetido ni requiere del concepto de distribución muestral[8]. Las probabilidades subjetivas pueden definirse para cualquier proposición, mientras que la probabilidad frecuencial se define sólo para sucesos en un espacio muestral (O'Hagan y Forster, 2004). Asimismo, el método bayesiano hace uso de toda la información previa disponible, mientras que en inferencia clásica esta información no se tiene en cuenta (Bernardo y Smith, 2004; Lecoutre, 2006).

No hay una única forma de elegir la distribución inicial, que condiciona los resultados de la inferencia y este hecho supuso una fuerte crítica el método bayesiano. En este sentido la idea de Jeffreys de usar distribuciones no informativas proporciona un posible criterio al comenzar a aplicar estos métodos, y es generalmente aceptada[9] (Robert, 2001). Esta idea será la que aplicaremos a lo largo del trabajo, partiendo de distribuciones iniciales no informativas la primera vez que aplicamos el método y actualizando estas distribuciones iniciales en nuevas aplicaciones, con los resultados del paso anterior.

La distribución final es la esencia de la estimación bayesiana. A la pregunta, una vez visto los datos, ¿qué sabemos del parámetro? la respuesta es dar la distribución final, puesto que esta distribución sintetiza toda la información sobre el parámetro, recogidos los datos y contiene todas las inferencias que puedan hacerse del mismo (Bernardo, 2003; O'Hagan y Forster, 2004). El estimador puntual óptimo del parámetro será la esperanza del mismo en la distribución final, puesto que minimiza el error cuadrático esperado (O'Hagan y Forster, 2004).

La distribución final también permitirá calcular probabilidades de que los parámetros se encuentren en intervalos de valores dados (intervalos de credibilidad), así

[7] Al poderse revisar la probabilidades y distribuciones en función de nueva información se pierde el carácter objetivo de las mismas que es asumido por las concepciones clásica, lógica y de propensión. La probabilidad es siempre condicional sobre un sistema de creencias y conocimientos y puede ser diferente para distintas personas (Batanero y Díaz, en prensa).

[8] En la mayoría de problemas de inferencia no es realista pensar en una sucesión de repeticiones del experimento, por tanto en una interpretación frecuencial del valor p o el coeficiente de confianza (O'Hagan y Forster, 2004).

[9] Con excepción de los bayesianos radicales (Robert, 2001).

como calcular probabilidades de que ciertas hipótesis sean verdaderas o falsas. La inferencia bayesiana básica sobre una hipótesis es calcular su probabilidad final, al contrario que en inferencia clásica, donde las hipótesis se aceptan o rechazan, lo cual no es una inferencia, sino una decisión (O'Hagan y Forster, 2004).

La distribución predictiva o marginal

$$p(y) = \int p(y/\theta)p(\theta)d\theta$$

representa las predicciones sobre futuros valores de y que tienen en cuenta la incertidumbre sobre el valor del parámetro θ así como la incertidumbre residual sobre y cuando se conoce θ (Bernardo, 2003; Lee, 2004). Esta distribución permite calcular probabilidades de futuros valores de la variable (probabilidad predictiva). Ninguna de estas probabilidades puede calcularse en inferencia clásica (Lecoutre, 1999; 2006; Bolstad, 2004).

2.3. FORMULACIÓN BAYESIANA DEL MODELO

En una formulación bayesiana consistente del modelo clásico de medición, se respetarán los supuestos básicos del mismo. La principal diferencia se introducirá al considerar los parámetros del modelo como variables aleatorias, con su distribución inicial y final. Hecho este supuesto, la estimación de dichos parámetros se llevará a cabo desde una metodología bayesiana, siguiendo los procedimientos y objetivos de la misma.

En consecuencia, la puntación empírica X obtenida para un sujeto en el total de una prueba es una variable aleatoria suma de dos componentes: la puntuación verdadera (V) del sujeto en ese test o prueba, y el error de medida (e). La puntuación verdadera se considera ahora como una variable aleatoria con una distribución normal inicial[10]. Consecuentemente, aceptamos los siguientes supuestos:

$$X=V+e$$

[10] El supuesto de normalidad es razonable, en cuanto la puntuación total es suma de las puntuaciones en los ítems y los teoremas de límite aseguran una distribución normal aproximada (Alvarado, 2004).

En esta expresión e es una variable aleatoria que sigue una distribución normal de media cero. Fijada una persona i, su puntuación verdadera en la prueba V_i, así como su puntuación observada X_i y el error de medida e_i serían variables aleatorias.

Asumimos además los siguientes supuestos de la TCT (Thorndike 1989; Muñiz, 1994; Martínez Arias, 1995):

3. E $(e_i)=0$. Se supone que los errores siguen una distribución normal[11].
4. $\rho(V,e)=0$. Asumimos que el error de medida es independiente de la puntuación verdadera (por tanto el error no está correlacionado con la puntuación empírica obtenida).

De estos supuestos se derivan las siguientes igualdades, similares a las de la TCT, puesto que se aplican cuando V es una variable aleatoria:

$$E\,(X)=E\,(V)$$

$$\sigma_X^2=\sigma_V^2+\sigma_e^2$$

$$\rho_{XV}^2=\frac{\sigma_V^2}{\sigma_X^2}=1-\frac{\sigma_e^2}{\sigma_X^2}=1-\rho_{Xe}^2$$

3. PUNTUACIÓN MEDIA Y DIFERENCIA DE PUNTUACIONES MEDIAS

La inferencia bayesiana permite hacer estimaciones de la media poblacional, o de la diferencia de dos medias[12] que serán diferentes, según se conozcan o no las desviaciones típicas de las distribuciones iniciales.

3.1. DISTRIBUCIÓN INICIAL NO INFORMATIVA

Es el caso es que no se tiene o no se quiere utilizar información previa sobre el valor de la media, o equivalentemente, se supone que la distribución inicial de la media

[11] No nos planteamos la distribución muestral, o el contexto de muestreo repetido; además, puesto que la suma de variables aleatorias normales es normal, asumiendo una distribución inicial normal para la puntuación verdadera, la puntuación observada seguirá una distribución normal (Rios, 1967; Peña, 1986).

[12] Situación que interesa para comparar dos grupos, por ejemplo, en los estudios de validez discriminante del instrumento.

es uniforme en un cierto rango de valores. Para hallar la distribución final de la puntuación media se presentarán dos casos;

- Se conoce la desviación típica σ_i de la distribución inicial de la media.

En este caso, para una distribución inicial uniforme la *distribución final* de la media es normal $N(\bar{x}, \sigma/\sqrt{n})$, siendo $\bar{x}$ el valor obtenido de la media muestral. La expresión $Z = \dfrac{\mu_f - \bar{x}}{\sigma / \sqrt{n}}$ sigue una distribución normal $N(0,1)$ (Berry, 1995; Bolstad, 2004).

El estimador puntual de la media de la distribución final μ_f sería la media muestral $\bar{x}$ de los datos. El intervalo de credibilidad para un coeficiente de credibilidad α viene dado por: $(\bar{x} - Z_{1-(1-\alpha)/2}\sigma / \sqrt{n}; \bar{x} + Z_{1-(1-\alpha)/2}\sigma / \sqrt{n})$, siendo Z un percentil de la distribución normal tipificada.

- Si σ (desviación típica poblacional) se desconoce, podemos usar *s,* la estimación insesgada de la desviación típica (raíz de la cuasivarianza muestral).

La variable tipificada $T = \dfrac{\mu_f - \bar{x}}{s / \sqrt{n}}$ sigue una distribución *T*[13] con *n-1* grados de libertad, siendo *n* el tamaño de la muestra de datos (Bolstad, 2004). El intervalo de credibilidad para un coeficiente de credibilidad α viene dado por: $(\bar{x} - T_{1-(1-\alpha)/2}s / \sqrt{n}; \bar{x} + T_{1-(1-\alpha)/2}s / \sqrt{n})$, siendo *T* un percentil de la distribución *T* con *n-1* grados de libertad. Para muestras de tamaño suficiente puede aproximarse mediante la distribución normal.

3.2. DISTRIBUCIÓN INICIAL NO INFORMATIVA

Cuando la distribución inicial para la media de la población sigue una curva normal $N(\mu_i, \sigma_i)$ y se conoce la desviación típica σ_i de la distribución inicial de la media, la

[13] La distribución *T* proporciona intervalos de credibilidad más amplios, porque tiene en cuenta también la variabilidad de la estimación de la varianza de la distribución inicial (Bolstad, 2004).

distribución final también sigue la distribución normal $N(\mu_f, \sigma_f)$. Los valores de la media y desviación típica de la distribución final vienen dados por las fórmulas siguientes:

$$(1) \qquad \mu_f = \frac{\frac{n\bar{x}}{s^2} + \frac{\mu_i}{\sigma_i^2}}{\frac{n}{s^2} + \frac{1}{\sigma_0^2}} \qquad\qquad \sigma_f = \frac{1}{\sqrt{n/s^2 + 1/\sigma_i^2}}$$

En las expresiones anteriores n es el tamaño de la muestra, $\bar{x}$ y s la media y desviación típica de la muestra. Para el caso de que no se conozca la desviación típica σ_i de la distribución inicial de la media, ésta se estima a partir de la raíz cuadrada de la cuasivarianza muestral s. Las fórmulas anteriores de la media y desviación típica de la distribución final se conservan, pero ahora la distribución será T con $n-1$ grados de libertad, siendo n el tamaño de muestra, que se podrá aproximar por la normal para tamaño suficiente de muestra (Bolstad, 2004).

El intervalo de credibilidad para un coeficiente de credibilidad α viene dado por: $(\bar{x} - Z_{1-(1-\alpha)/2}\sigma_f/\sqrt{n}; \bar{x} + Z_{1-(1-\alpha)/2}\sigma_f/\sqrt{n})$, siendo Z un percentil de la distribución normal tipificada cuando se conoce la desviación típica de la distribución inicial y por $(\bar{x} - T_{1-(1-\alpha)/2}s_f/\sqrt{n}; \bar{x} + T_{1-(1-\alpha)/2}s_f/\sqrt{n})$, siendo T un percentil de la distribución T con $n-1$ grados de libertad cuando no se conoce.

Comparación de las estimaciones clásica y bayesiana

Mientras que $\bar{x}$ es un estimador insesgado de mínima varianza de la media poblacional, μ_f es un estimador sesgado. El sesgo viene dado por la expresión $\mu_f = \frac{\sigma_i^2}{ns^2 + \sigma_i^2}(\tilde{x} - \mu_i)$ (Bolstad, 2004), es decir, disminuye con la diferencia entre la media de la distribución inicial y la media de la muestra y también con las varianzas y tamaño de muestra. También decrece con el valor de la desviación típica en la distribución inicial.

Por otro lado la varianza del estimador es menor que la de la media muestral puesto que $\sigma^2{}_f = \frac{1}{n/s^2 + 1/\sigma_i^2} = \frac{s^2}{n} x \frac{n/s^2}{n/s^2 + 1/\sigma_i^2}$; por tanto es menor que $\frac{s^2}{n}$ al estar multiplicado por una cantidad menor que 1.

La distribución final que se usaría para calcular el intervalo de credibilidad nos da una probabilidad epistémica (referida a nuestra muestra concreta) de que el valor verdadero del parámetro en la población se encuentre en los límites dados, lo que es una probabilidad bayesiana. El intervalo de confianza no indica entre qué valores se encuentra el verdadero valor del parámetro; sólo indica que si hubiese tomado cien muestras de la población, el parámetro estará incluido en 95 intervalos. Sería una probabilidad objetiva frecuencial. Mientras el intervalo de credibilidad permite dar la probabilidad de que la media se encuentre en un intervalo de valores cuyos extremos son fijos, el intervalo de confianza, el valor de la media es fijo y los extremos de los intervalos son variables aleatorias (Bernard, 1998). Consideraciones semejantes se pueden hacer para el resto de parámetros.

3.3. DIFERENCIA DE MEDIAS

Otra estimación muy necesaria en la construcción o análisis de datos de un cuestionario es la de diferencia de medias para, por ejemplo, comparar dos grupos o en los estudios de discriminación del cuestionario. La situación más común es la comparación de muestras independientes[14], donde se pueden encontrar diferentes casos Trataremos únicamente el caso de distribuciones iniciales informativas, ya que el caso no informativo se puede incluir en el anterior[15].

Puesto que las dos muestras tienen distribuciones independientes, si las distribuciones iniciales son independientes, lo serán también las distribuciones finales en cualquiera de los casos se discuten a continuación. Las medias y varianzas de las distribuciones finales se calculan mediante las expresiones dadas en (1).

Caso 1. Varianzas iguales y conocidas

La media y varianza de la diferencia de puntuaciones en la distribución final vendrán dadas por:

---

[14] Además, la comparación de dos muestras relacionadas se podría reducir a la estimación de una sola media, trabajando con las diferencias de valores en cada caso, por lo que podrían incluir en el caso anterior, reduciendo las dos variables a una (diferencias).

[15] Para el caso de distribuciones no informativas ya sabemos obtener la distribución final de la media. A partir de ella se obtiene la distribución de la diferencia de medias y los intervalos de credibilidad del mismo modo que se hace en el caso informativo.

$$\mu_d^f = \mu_1^f - \mu_2^f$$
$$\sigma_d^{2f} = \sigma_1^{2f} + \sigma_2^{2f}$$

que coincidirían con la media y varianzas de la distribución muestral en caso de distribución inicial no informativa[16]. De donde el intervalo de credibilidad de la diferencia de medias para un coeficiente de credibilidad α sería:

$$(\mu_1^f - \mu_2^f \pm Z_{1-(1-\alpha)/2}\sqrt{\sigma_1^{2f} + \sigma_2^{2f}})$$

Caso 2. Varianzas distintas conocidas

Cuando se usan muestras independientes, y distribuciones iniciales independientes en las dos muestras, también serán independientes las distribuciones finales. Las medias y varianzas de la distribución final es de nuevo:

$$\mu_d^f = \mu_1^f - \mu_2^f$$
$$\sigma_d^{2f} = \sigma_1^{2f} + \sigma_2^{2f}$$

De donde el intervalo de credibilidad de la diferencia de medias para un coeficiente de credibilidad α coincide con el caso anterior[17] y sería:

$$(\mu_1^f - \mu_2^f \pm Z_{1-(1-\alpha)/2}\sqrt{\sigma_1^{2f} + \sigma_2^{2f}})$$

Caso 3. No se conocen las varianzas

En este caso, cada una de las varianzas debe ser estimada a partir de los datos de la muestra (utilizando las cuasivarianzas muestrales $s_1^2; s_2^2$). Esto aumenta la incertidumbre de la estimación, por lo que se usará la distribución *T*, en lugar de la normal (Box y Tiao, 1992). Los grados de libertad vienen dados por la fórmula de Satterhwaite: (redondeada al número entero más cercano) (Bolstad, 2004):

$$\upsilon = \frac{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}{\dfrac{(s_1^2/n_1)^2}{n_1+1} + \dfrac{(s_2^2/n_2)^2}{n_2+1}}$$

[16] Aunque con diferente interpretación.

[17] Con la salvedad de que las varianzas iniciales son diferentes.

El intervalo de credibilidad aproximado vendría dado por:

$$(\mu_1^f - \mu_2^f \pm T_{1-(1-\alpha)/2}\sqrt{\sigma_1^{2f} + \sigma_2^{2f}})$$

donde las medias y varianzas de la distribuciones finales se calculan mediante las expresiones (1), pero estimando las varianzas iniciales mediante las cuasivarianzas muestrales y usando la fórmula de Satterhwaite para calcular los grados de libertad. Se aproximará a la distribución normal cuando el tamaño de muestra sea suficientemente grande.

Para el caso de distribución inicial no informativa, esta expresión se transforma en:

$$(\mu_1^f - \mu_2^f \pm T_{1-(1-\alpha)/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}})$$

coincidiendo con el intervalo de confianza frecuencial, aunque con diferente interpretación.

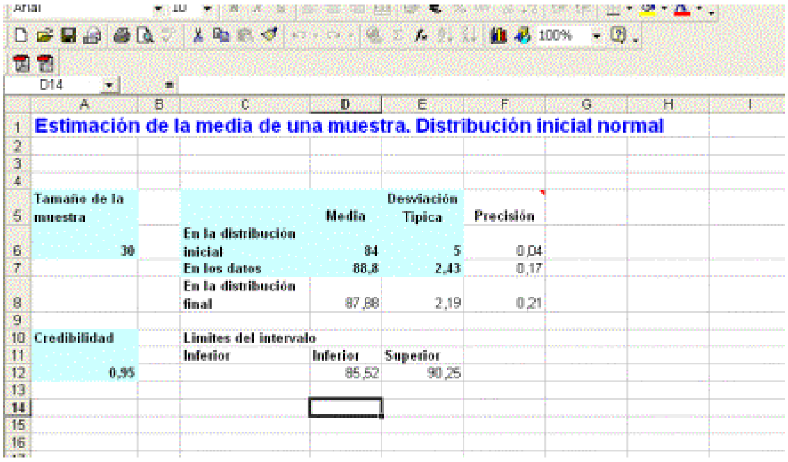
3.4. PROGRAMAS DE CÁLCULO

Para realizar los cálculos correspondientes a la estimación de una media se preparó un programa Excel (Ver Figura 3.1), utilizando las fórmulas de cálculo que se han descrito en los apartados anteriores. Los datos requeridos son los siguientes: Tamaño de muestra, medias y desviaciones típicas en los datos y sus estimaciones en la muestra y coeficiente de credibilidad.

El programa proporciona los parámetros en la distribución final, e intervalos de confianza y credibilidad para cada uno de los ítems. El programa puede ser utilizado para el caso de distribución inicial informativa. Para el caso de distribución inicial no informativa se puede emplear los programas de cálculo de intervalos de confianza, por ejemplo en SPSS, aunque dando una interpretación bayesiana. Al aplicar el método en pasos sucesivos, los parámetros de la distribución inicial en cada paso, serán los resultantes en la distribución final en el paso anterior

Los cálculos pueden aplicarse para estimar la puntuación media del cuestionario, o para estimar las puntuaciones medias en los ítems, en subconjuntos de ítems, submuestras de estudiantes, etc.

Figura 3.1. Estimación de la puntuación media

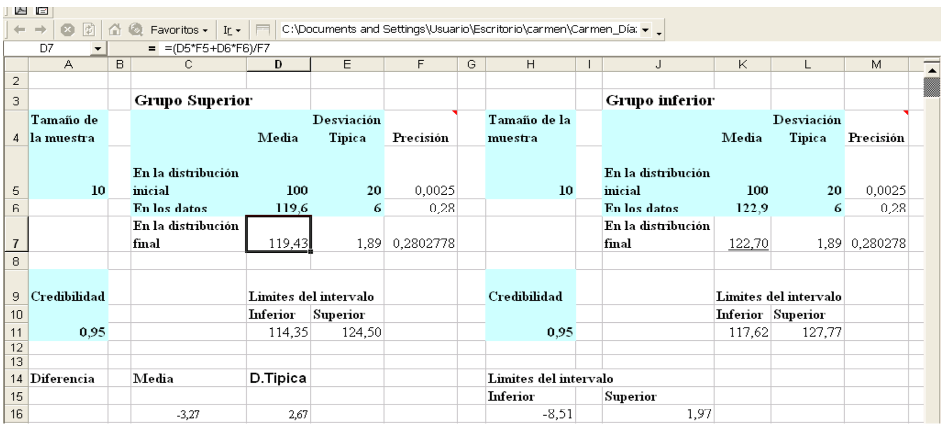


	A	B	C	D	E	F	G	H	I
1	Estimación de la media de una muestra. Distribución inicial normal								
2									
3									
4									
5	Tamaño de la muestra			Media	Desviación Tipica	Precisión			
6	30		En la distribución inicial	84	5	0,04			
7			En los datos	88,8	2,43	0,17			
8			En la distribución final	87,88	2,19	0,21			
9									
10	Credibilidad		Limites del intervalo						
11			Inferior	Inferior	Superior				
12	0,95			85,52	90,25				
13									
14									
15									
16									

Para realizar los cálculos correspondientes a la estimación de la diferencia de medias se preparó un programa Excel (Ver Figura 3.2), utilizando las fórmulas de cálculo que se han descrito en los apartados anteriores. Se programó sólo el caso de varianzas desconocidas y desiguales que es el más habitual; el caso de varianzas iguales es poco probable al comparar dos grupos, pero se podría también estimar utilizando el mismo programa.

Los datos requeridos son los siguientes: Tamaño de muestra, medias y desviaciones típicas en los datos y sus estimaciones en la muestra para cada uno de los grupos y coeficiente de credibilidad. El programa proporciona los parámetros en la distribución final, e intervalos de confianza y credibilidad para las medias de cada muestra. Al aplicar el método en pasos sucesivos, los parámetros de la distribución inicial en cada paso, serán los resultantes en la distribución final en el paso anterior

Figura 3.2. Estimación de la diferencia de medias



D7 = =(D5*F5+D6*F6)/F7

	A	B	C	D	E	F	G	H	I	J	K	L	M
2													
3			Grupo Superior							Grupo inferior			
4	Tamaño de la muestra			Media	Desviación Tipica	Precisión		Tamaño de la muestra			Media	Desviación Tipica	Precisión
5	10		En la distribución inicial	100	20	0,0025		10		En la distribución inicial	100	20	0,0025
6			En los datos	119,6	6	0,28				En los datos	122,9	6	0,28
7			En la distribución final	119,43	1,89	0,2802778				En la distribución final	122,70	1,89	0,280278
8													
9	Credibilidad			Limites del intervalo				Credibilidad			Limites del intervalo		
10				Inferior	Superior						Inferior	Superior	
11	0,95			114,35	124,50			0,95			117,62	127,77	
12													
13													
14	Diferencia		Media	D.Tipica				Limites del intervalo					
15								Inferior		Superior			
16			-3,27	2,67				-8,51		1,97			

ESTIMACIÓN DE ÍNDICES DE DIFICULTAD

En el análisis de ítems, se estudian las propiedades que influyen en las del cuestionario, siendo los índices más relevantes los de dificultad y discriminación. El índice de dificultad se define como la proporción p de sujetos que acertarán el ítem, entre todos los que tratan de resolverlo en una cierta población (Thorndike, 1991; Martínez Arias, 1995). Sería un parámetro de esta población y se estima mediante la proporción $\hat{p}$ de sujetos que lo responden en una muestra representativa de la población. La variable *x:* "número de estudiantes que responde correctamente en una muestra de n casos a un ítem con índice de dificultad p" sigue una distribución binomial *B (n, p),* cuya distribución de probabilidad (que depende de p y n) es:

$$p(x/p,n) = \binom{n}{x} p^x (1-p)^{n-x}$$

Mientras que en la inferencia clásica la proporción p se considera constante, en la inferencia bayesiana el índice de dificultad p sería una variable aleatoria. Fijados x y n, la expresión anterior es función de p *(verosimilitud),* que viene dada por una probabilidad condicional (Bernard, 1998):

$$l(p/x,n) = \binom{n}{x} p^x (1-p)^{n-x}$$

Dentro de las posibles distribuciones de probabilidad continuas que pueden describir la distribución a priori de una proporción, se recomienda tomar la familia de distribuciones *Beta*, que vienen definidas por dos parámetros, a y b (Berry, 1995; Serrano, 2003)[18]. Dada una función beta de probabilidad a priori *Be(a, b)* para una proporción, si en una nueva muestra se observan e éxitos y f fracasos, la función de probabilidad a posteriori sería beta *Be(a + e, b + f),* es decir, si a los parámetros a y b de la función Beta se interpretan como número de éxitos y fracasos, la distribución a priori incrementa su número de éxitos con los éxitos encontrados en la muestra e igual ocurre

[18] La ventaja de usar una distribución a priori tipo Beta para la proporción es porque ésta es *conjugada* con la función de verosimilitud de la distribución binomial, es decir, produce una distribución a posteriori de la misma familia de distribuciones (Gelman, Carlin, Stern y Rubin, 2003).

con el número de fracasos (Serrano, 2003).

La distribución Beta puede tomar formas muy diferentes, dependiendo de los valores *a* y *b*, por lo que es posible encontrar una distribución beta que exprese diferentes creencias sobre las probabilidades iniciales de la proporción. Además produce una distribución final Beta (familia de distribuciones conjugadas). El valor medio es *a/(a+b)* y para el mismo valor medio, la mayor dispersión depende de la suma total de *a+b* (tamaño de la muestra). Se pueden tener dos casos diferentes, según esta distribución sea o no informativa.

4.1. DISTRIBUCIÓN INICIAL NO INFORMATIVA

Como distribución inicial no informativa, se podría utilizar cualquier distribución Beta con *a=b*, es decir una distribución uniforme del parámetro *p*[19] (Berry, 1995; Lecoutre, 1996). En nuestro estudio se utilizará como distribución inicial *Be(0,5, 0,5)* como recomiendan entre otros Lecoutre (1996) y Serrano (2003). La distribución final del índice de dificultad, viene dada par la distribución *Be(0,5+p, 0,5+q),* siendo *p* y *q* el número de respuestas correctas e incorrectas en el ítem[20]. El intervalo de credibilidad se calcula mediante la siguiente expresión,

$$\left[\beta^{-1}_{0{,}5+\mathrm{p},\,0{,}5+\mathrm{q}}(\alpha/2) - \beta^{-1}_{0{,}5+\mathrm{p},\,0{,}5+\mathrm{q}}(1-\alpha/2)\right]$$

donde $\beta^{-1}_{\mathrm{a,b}}$ es la función de distribución inversa de la distribución *Be (a,b)* y α el coeficiente de credibilidad.

4.2. DISTRIBUCIÓN INICIAL INFORMATIVA

En este caso, se elegiría una distribución Beta que represente la información previa. Si se dispone de datos de una estimación previa del índice de dificultad en la misma población siguiendo la recomendación de Serrano (2003) se podría considerar la razón

[19] Ausencia de información (porque los aciertos serían iguales a los fracasos).

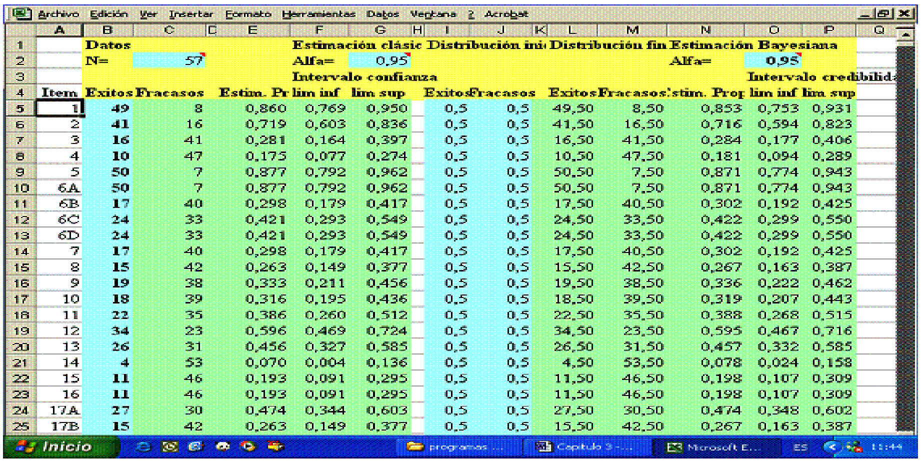
[20] La ventaja de utilizar una distribución inicial Beta para la estimación de la proporción, es que esta distribución es reproductiva respecto a sus parámetros.

de ocurrencia de respuestas correctas /incorrectas. Es decir, siguiendo la metodología bayesiana, se utilizaría como distribución inicial, la final obtenida en la primera prueba.

4.3. PROGRAMA DE CÁLCULO

Para realizar los cálculos se preparó un programa Excel (Ver Figura 3.3), utilizando las fórmulas de cálculo que se han descrito en los apartados anteriores. Los datos requeridos son los siguientes: Tamaño de muestra y número de acierto en cada ítem; coeficientes de confianza y credibilidad (que pueden ser iguales o diferentes); número de éxitos y fracasos por ítems en la distribución inicial.

Figura 3.3. Estimación de índices de dificultad



	Datos		Estimación clásic			Distribución ini		Distribución fin		Estimación Bayesiana		
	N=	57		Alfa=	0,95					Alfa=	0,95	
				Intervalo confianza							Intervalo credibilid	
Item	Exitos	Fracasos	Estim. Pr	lim inf	lim sup	Exitos	Fracasos	Exitos	Fracasos	stim. Prop	lim inf	lim sup
1	49	8	0,860	0,769	0,950	0,5	0,5	49,50	8,50	0,853	0,753	0,931
2	41	16	0,719	0,603	0,836	0,5	0,5	41,50	16,50	0,716	0,594	0,823
3	16	41	0,281	0,164	0,397	0,5	0,5	16,50	41,50	0,284	0,177	0,406
4	10	47	0,175	0,077	0,274	0,5	0,5	10,50	47,50	0,181	0,094	0,289
5	50	7	0,877	0,792	0,962	0,5	0,5	50,50	7,50	0,871	0,774	0,943
6A	50	7	0,877	0,792	0,962	0,5	0,5	50,50	7,50	0,871	0,774	0,943
6B	17	40	0,298	0,179	0,417	0,5	0,5	17,50	40,50	0,302	0,192	0,425
6C	24	33	0,421	0,293	0,549	0,5	0,5	24,50	33,50	0,422	0,299	0,550
6D	24	33	0,421	0,293	0,549	0,5	0,5	24,50	33,50	0,422	0,299	0,550
7	17	40	0,298	0,179	0,417	0,5	0,5	17,50	40,50	0,302	0,192	0,425
8	15	42	0,263	0,149	0,377	0,5	0,5	15,50	42,50	0,267	0,163	0,387
9	19	38	0,333	0,211	0,456	0,5	0,5	19,50	38,50	0,336	0,222	0,462
10	18	39	0,316	0,195	0,436	0,5	0,5	18,50	39,50	0,319	0,207	0,443
11	22	35	0,386	0,260	0,512	0,5	0,5	22,50	35,50	0,388	0,268	0,515
12	34	23	0,596	0,469	0,724	0,5	0,5	34,50	23,50	0,595	0,467	0,716
13	26	31	0,456	0,327	0,585	0,5	0,5	26,50	31,50	0,457	0,332	0,585
14	4	53	0,070	0,004	0,136	0,5	0,5	4,50	53,50	0,078	0,024	0,158
15	11	46	0,193	0,091	0,295	0,5	0,5	11,50	46,50	0,198	0,107	0,309
16	11	46	0,193	0,091	0,295	0,5	0,5	11,50	46,50	0,198	0,107	0,309
17A	27	30	0,474	0,344	0,603	0,5	0,5	27,50	30,50	0,474	0,348	0,602
17B	15	42	0,263	0,149	0,377	0,5	0,5	15,50	42,50	0,267	0,163	0,387

El programa proporciona los parámetros (éxitos y fracasos) en la distribución final, estimaciones de la proporción, intervalos de confianza y credibilidad para cada uno de los ítems.

El programa puede ser utilizado, tanto para el caso de distribución inicial no informativa (en cuyo caso se aconseja dar valores iguales y próximos a cero de los parámetros de la distribución, por ejemplo 0,5) o informativa. Para este segundo supuesto se pueden utilizar otros parámetros en la distribución inicial. Al aplicar el método en pasos sucesivos, los parámetros de la distribución inicial en cada paso, serán los resultantes en la distribución final en el paso anterior.

5. ESTIMACIÓN DE INDICES DE DISCRIMINACIÓN

Una primera aproximación al estudio de los índices de discriminación es analizar la diferencia de proporciones de aciertos al ítem en dos grupos de alumnos, que se diferencian por su competencia (Thorndike, 1989; Martínez Arias, 1995; Barbero, 2003). Estos dos grupos se pueden determinar de forma diversa:

- Alumnos que puntúan por encima y debajo de la mediana
- Alumnos que puntúan en el 27% superior e inferior
- Alumnos con y sin instrucción..

En cualquiera de estos supuestos, denominemos p_s, p_i los índices de dificultad en los grupos de conocimiento que hipotéticamente han de dar un resultado superior e inferior en el cuestionario. En teoría clásica estos parámetros serían constantes desconocidas en sus respectivas poblaciones y la estimación puntual del índice de discriminación vendría dada por:

$$d = \widehat{p}_s - \widehat{p}_i$$

siendo $\widehat{p}_s$ $\widehat{p}_i$ los estimadores puntuales de p_s y p_i respectivamente. Este índice en teoría clásica sería, asimismo, constante, y se podrían calcular intervalos de confianza con la fórmula habitual para diferencia de proporciones, que viene implementada en la mayor parte de paquetes estadísticos. Asimismo se podría utilizar la aproximación normal para muestras de tamaño suficiente, mediante la expresión:

$$(p_s - p_i) \pm \left(Z^{-1}{}_{1-\alpha/2} \sqrt{\frac{p_s(1-p_s)}{n_s} + \frac{p_s(1-p_s)}{n_s}} \right)$$

siendo Z la variable normal tipificada y α el coeficiente de confianza.

En la interpretación bayesiana las proporciones anteriores y su diferencia serían variables aleatorias. Las funciones de verosimilitud del número de éxitos en cada una de las dos muestras vienen determinadas por las expresiones siguientes:

$$l(p_s / x_s, n_s) = \binom{n_s}{x_s} p_s^{x_s} (1 - p_s)^{n_s - x_s} \quad l(p_i / x_i, n_{\underset{i}{s}}) = \binom{n_i}{x_i} p_i^{x_i} (1 - p_i)^{n_i - x_i}$$

La función de verosimilitud conjunta sería el producto de las dos anteriores, al ser las muestras independientes.

Si la distribución inicial de cada una de las proporciones $p_{s,}$ p_i se toma de la familia de distribuciones Beta, se obtendrán distribuciones finales Beta para la proporción en cada una de las poblaciones, y al ser las variables independientes la distribución final conjunta de la variable doble ($p_{s,}\, p_i$) será producto de dos distribuciones obtenidas.

Para muestras de tamaño suficiente, se podrá utilizar la distribución normal para aproximar los intervalos de credibilidad (Berry, 1995). Se estudiarñan los dos casos posibles, en lo que sigue.

5.1. DISTRIBUCIÓN INICIAL NO INFORMATIVA

En este caso, se utilizan distribuciones iniciales uniformes, por ejemplo *Be(1,1).* Si se ha obtenido e_s éxitos y f_s fracasos en el grupo superior y e_i éxitos y f_i fracasos en el grupo inferior, respectivamente, y las muestras son independientes, entonces p_s *y* p_i *s*on independientes con distribuciones finales *Be(*$e_{s;}$*+1,* $f_{s;}$*+1), Be(*$e_{i;}$*+1,,* $b_{i;}$*+1)* respectivamente. Por lo tanto los estimadores respectivos de las proporciones serán:

$$\hat{p}_s = \frac{e_s + 1_s}{2 + e_s + f_s} \qquad \hat{p}_i = \frac{e_i + 1}{e_i + f + 2_i} \quad \text{(Albert, 1995; 1996)}$$

5.2. DISTRIBUCIÓN INICIAL INFORMATIVA

Si la distribución inicial de p_s es *Be(*a_s*,* b_s*)*y la distribución inicial de p_i *Be(*a_i*,* b_i*),* habiendo obtenido e_s éxitos y f_s fracasos en el grupo superior y e_i éxitos y f_i fracasos en el grupo inferior, respectivamente, y las muestras son independientes, entonces p_s *y* p_i son independientes con distribuciones finales *Be(*$a_{s,}$*+* $e_{s;}$ b_{s+} f_s*), Be(*a_i *+*$e_{i,}$*,* b_{i+} e_i*)* respectivamente. Por lo tanto los estimadores respectivos de las proporciones serán:

$$\hat{p}_s = \frac{a_s + e_s}{a_s + b_s + e_s + f_s} \qquad \hat{p}_i = \frac{a_i + e_i}{a_i + b_i + e_i + f_i} \quad \text{(Albert, 1996)}$$

5.3. APROXIMACIÓN NORMAL DE LA DISTRIBUCIÓN DE LAS DIFERENCIAS

En cualquiera de los casos anteriores, sea las distribuciones finales de las poblaciones p_s *Be(a'$_s$, b'$_s$)* y la distribución inicial de p_i *Be(a'$_i$, b'$_i$)* que vendrán dada por algunas de las fórmulas anteriores y serán independientes, al serlo las distribuciones iniciales (Bolstad, 2004). Siguiendo a Berry (1995), los estimadores de las medias de la distribuciones finales serán:

$$\widehat{p}_s = \frac{a'_s}{a'_s + b'_s} \qquad \widehat{p}_i = \frac{a'_i}{a'_i + b_i'}$$

Los estimadores de las desviaciones típicas de las distribuciones finales serían (Bolstad, 2004):

$$\partial_s = \sqrt{\frac{\widehat{p}_s \widehat{q}_s}{n_s + 1}} \qquad \partial_i = \sqrt{\frac{\widehat{p}_i \widehat{q}_i}{n_i + 1}}$$

En consecuencia, la diferencia de proporciones tendrá una distribución normal aproximada $N(\widehat{p}_s - \widehat{p}_i, \sqrt{\partial_s^2 + \partial_i^2})$, por lo que el intervalo aproximado de credibilidad vendrá dado por:

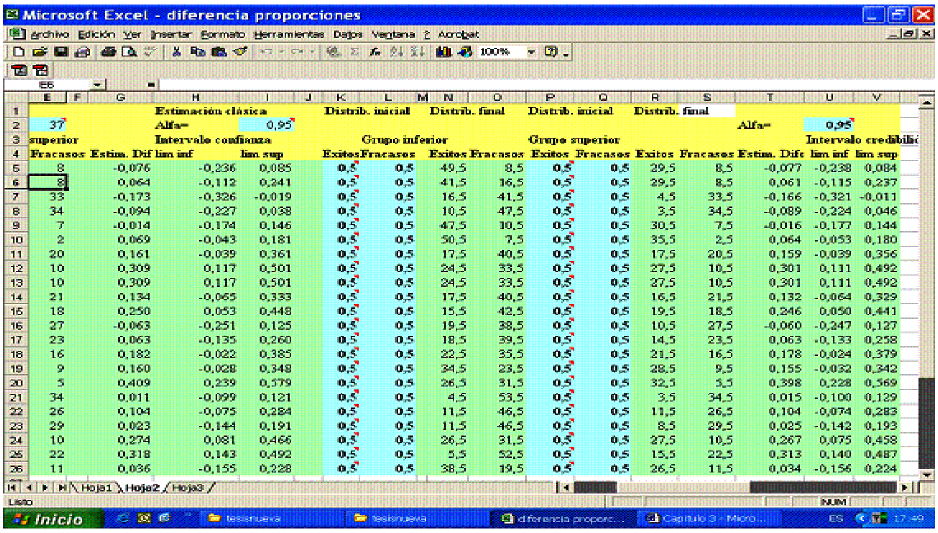
$$\widehat{p}_s - \widehat{p}_i \pm Z_{1-(1-\alpha)/2}^{-1} \sqrt{\partial_s^2 + \partial_i^2}$$

siendo *Z* la distribución normal N (0,1).

5.4. PROGRAMA DE CÁLCULO

Para realizar los cálculos se preparó un programa Excel (Ver Figura 3.4), utilizando las fórmulas de cálculo que se han descrito en los apartados anteriores. Los datos requeridos son los siguientes: Tamaño de muestra y número de acierto en cada ítem y para cada uno de los grupos (datos) coeficientes de confianza y credibilidad (que pueden ser iguales o diferentes); número de éxitos y fracasos por ítems en la distribución inicial para cada uno de los grupos.

Figura 3.4. Estimación de índices de discriminación



	E	G	H	I	K	L	N	O	P	Q	R	S	T	U	V
1			Estimación clásica		Distrib. inicial		Distrib. final		Distrib. inicial		Distrib. final				
2	37		Alfa=	0,95									Alfa=	0,95	
3	superior		Intervalo confianza			Grupo inferior			Grupo superior					Intervalo credibilid	
4	Fracasos	Estim. Dif	lim inf	lim sup	Exitos	Fracasos	Exitos	Fracasos	Exitos	Fracasos	Exitos	Fracasos	Estim. Dif	lim inf	lim sup
5	8	-0,076	-0,236	0,085	0,5	0,5	49,5	8,5	0,5	0,5	29,5	8,5	-0,077	-0,238	0,084
6	8	0,064	-0,112	0,241	0,5	0,5	41,5	16,5	0,5	0,5	29,5	8,5	0,061	-0,115	0,237
7	33	-0,173	-0,326	-0,019	0,5	0,5	16,5	41,5	0,5	0,5	4,5	33,5	-0,166	-0,321	-0,011
8	34	-0,094	-0,227	0,038	0,5	0,5	10,5	47,5	0,5	0,5	3,5	34,5	-0,089	-0,224	0,046
9	7	-0,014	-0,174	0,146	0,5	0,5	47,5	10,5	0,5	0,5	30,5	7,5	-0,016	-0,177	0,144
10	2	0,069	-0,043	0,181	0,5	0,5	50,5	7,5	0,5	0,5	35,5	2,5	0,064	-0,053	0,180
11	20	0,161	-0,039	0,361	0,5	0,5	17,5	40,5	0,5	0,5	17,5	20,5	0,159	-0,039	0,356
12	10	0,309	0,117	0,501	0,5	0,5	24,5	33,5	0,5	0,5	27,5	10,5	0,301	0,111	0,492
13	10	0,309	0,117	0,501	0,5	0,5	24,5	33,5	0,5	0,5	27,5	10,5	0,301	0,111	0,492
14	21	0,134	-0,065	0,333	0,5	0,5	17,5	40,5	0,5	0,5	16,5	21,5	0,132	-0,064	0,329
15	18	0,250	0,053	0,448	0,5	0,5	15,5	42,5	0,5	0,5	19,5	18,5	0,246	0,050	0,441
16	27	-0,063	-0,251	0,125	0,5	0,5	19,5	38,5	0,5	0,5	10,5	27,5	-0,060	-0,247	0,127
17	23	0,063	-0,135	0,260	0,5	0,5	18,5	39,5	0,5	0,5	14,5	23,5	0,063	-0,133	0,258
18	16	0,182	-0,022	0,385	0,5	0,5	22,5	35,5	0,5	0,5	21,5	16,5	0,178	-0,024	0,379
19	9	0,160	-0,028	0,348	0,5	0,5	34,5	23,5	0,5	0,5	28,5	9,5	0,155	-0,032	0,342
20	5	0,409	0,239	0,579	0,5	0,5	26,5	31,5	0,5	0,5	32,5	5,5	0,398	0,228	0,569
21	34	0,011	-0,099	0,121	0,5	0,5	4,5	53,5	0,5	0,5	3,5	34,5	0,015	-0,100	0,129
22	26	0,104	-0,075	0,284	0,5	0,5	11,5	46,5	0,5	0,5	11,5	26,5	0,104	-0,074	0,283
23	29	0,023	-0,144	0,191	0,5	0,5	11,5	46,5	0,5	0,5	8,5	29,5	0,025	-0,142	0,193
24	10	0,274	0,081	0,466	0,5	0,5	26,5	31,5	0,5	0,5	27,5	10,5	0,267	0,075	0,458
25	22	0,318	0,143	0,492	0,5	0,5	5,5	52,5	0,5	0,5	15,5	22,5	0,313	0,140	0,487
26	11	0,036	-0,155	0,228	0,5	0,5	38,5	19,5	0,5	0,5	26,5	11,5	0,034	-0,156	0,224

El programa proporciona los parámetros (éxitos y fracasos) en la distribución final para cada uno de los dos grupos, estimaciones de la diferencia de proporciones, intervalos de confianza y credibilidad para cada uno de los ítems.

El programa puede ser utilizado, tanto para el caso de distribución inicial no informativa (en cuyo caso se aconseja dar valores iguales y próximos a cero de los parámetros de la distribución, por ejemplo 0,5) o informativa. Para este segundo supuesto se pueden utilizar otros parámetros en la distribución inicial.

Al aplicar el método en pasos sucesivos, los parámetros de la distribución inicial en cada paso, serán los resultantes en la distribución final en el paso anterior.

6. ESTIMACIÓN DE COEFICIENTES DE FIABILIDAD Y CORRELACIÓN

Se entiende por *fiabilidad* de un test, un cuestionario u otro instrumento de medida la estabilidad de las puntuaciones que proporciona si se administra en repetidas ocasiones al mismo grupo de personas[21].

El *coeficiente de fiabilidad* es un indicador de la fiabilidad teórica de las puntuaciones observadas, en el sentido de proporcionar un valor numérico que indica el *grado de confianza* que podíamos tener en dichas puntuaciones como estimadores de las

[21] La medida siempre produce un cierto error aleatorio, pero dos medidas del mismo fenómeno sobre un mismo individuo suelen ser consistentes. La fiabilidad es esta tendencia a la consistencia o precisión del instrumento en la población medida (Carmines y Zeller, 1979; Bisquerra, 1989).

puntuaciones verdaderas de los sujetos (Martínez Arias, 1995; Barbero, 2003). Se define como la correlación entre las puntuaciones *X* y *X'* obtenidas por el sujeto en un test cuando se le pasa dos veces sucesivas. Si no hubiese errores de medida, las puntuaciones coincidirían y la correlación sería perfecta, por lo que este coeficiente sería igual a 1 (Carmines y Zeller, 1979; Muñiz, 1994).

Este coeficiente de fiabilidad tiene un carácter teórico y en la TCT se supone que es un parámetro constante del test, que debe ser estimado por algún procedimiento empírico, a través de las respuestas de un grupo de sujetos a un conjunto de ítems. Hay diversos procedimientos para el cálculo del estimador del coeficiente de fiabilidad, algunos de los cuáles se basan en la estimación del coeficiente de correlación entre dos pasaciones del cuestionario a los sujetos o de dos formas equivalentes del cuestionario (Carmines y Zeller, 1979; Meliá, 2001; Barbero, 2003; Díaz, Batanero y Cobo, 2003). Son las siguientes.

- *Test- retest* : En este método se administra el mismo test dos veces a las mismas personas y se calcula el coeficiente de correlación entre las puntuaciones obtenidas en las dos ocasiones. Intenta medir el porcentaje de variabilidad debido a las fuentes que contribuyen a que un sujeto tenga diferente puntuación en aplicaciones repetidas de la "misma prueba": temporales, ambientales, estado, sentimientos... y da una medida de la estabilidad del rasgo en las personas durante el periodo de tiempo dado. Por ello se denomina *coeficiente de estabilidad*.
- *Formas paralelas*: Este método de estimación requiere construir dos formas paralelas del test que son administradas al mismo grupo de sujetos dejando entre ambas administraciones un periodo de tiempo breve para evitar cambios en los sujetos, aunque suficiente para evitar la fatiga. Las componentes de variabilidad que se desean medir por este método son las debidas a fluctuaciones de disposición de unos ítems respecto a otros o ligeros cambios en los ítems. Indica la estabilidad de variaciones de la misma prueba o de ítems que han sido diseñados para medir las mismas funciones. El coeficiente obtenido se denomina *coeficiente de equivalencia*.
- *Test- retest con formas paralelas*: Es una combinación de los dos procedimientos anteriores. Se administra una forma del test, se deja transcurrir un periodo de tiempo semejante al del procedimiento Test- retest y se administra la forma paralela

del test a la misma muestra de sujetos. El coeficiente que se obtiene se denomina *coeficiente de estabilidad y equivalencia*.

- *Dos mitades del test* : Una vez pasado el test, se divide en dos mitades y se calcula la puntuación del sujeto en cada mitad. Al dividir el test se debe conseguir que las dos mitades sean equivalentes, algo que en ocasiones puede resultar difícil. Se pueden dividir aleatoriamente, según la dificultad o según el contenido y la dificultad (el procedimiento más aconsejable). Se calcula la correlación entre las puntuaciones obtenidas en las dos mitades del test, con un procedimiento de corrección que puede ser el de Spearman- Brown:

$$S = \frac{2r}{1+r}$$

siendo r el coeficiente de correlación empírico entre las dos mitades del test.

En todos los casos anteriores, el coeficiente de correlación implicado se supone un parámetro constante en la TCT y la finalidad de la estimación es proporcionar una estimación puntual o un intervalo de confianza para el parámetro.

En una interpretación bayesiana de la TCT la correlación (entre dos pasaciones del test, entre la pasación de dos formas equivalentes pasadas simultáneamente o no o entre las dos mitades del test) será una variable aleatoria con una distribución inicial. El objetivo de la estimación será actualizar esta distribución, para llegar a la distribución final, a partir de los datos de la muestra.

Estudiaremos, en lo que sigue, de una forma general el método de estimación del coeficiente de correlación, que podremos adaptar a cualquiera de los casos anteriores, para proporcionar estimaciones de los diversos coeficientes de fiabilidad descritos.

6.1. DISTRIBUCIÓN FINAL DEL COEFICIENTE DE CORRELACIÓN

Dado un conjunto de pares de observaciones (x_1, y_1), (x_1, y_1), ... (x_1, y_1), por ejemplo el conjunto de pares de puntuaciones de una muestra de sujetos en dos pasaciones, dos formas equivalentes o dos mitades de un test, se asume que estas observaciones provienen de una variable aleatoria bidimensional (X, Y) con distribución normal bivariante[22]. Asumimos que las medias, varianzas y correlación de las puntuaciones vienen dadas por:

$$\begin{aligned}&E(X)=\mu;Var(X)=\sigma^2\\&E(Y)=\eta;Var(X)=\varphi^2\\&\rho(X,Y)=\rho\end{aligned}$$

Se supone calculada las medias $\overline{x}$ e $\overline{y}$ y la correlación r en el conjunto de observaciones (datos). Puede demostrarse (Lee, 2004) que en el caso de usar distribuciones iniciales no informativas para las medias y varianzas de *X e Y,* y teniendo una distribución inicial del coeficiente de correlación $p(\rho)$, una aproximación razonable de la distribución final del coeficiente de correlación viene dada por:

$$p(\rho/(x,y)) \propto p(\rho)\frac{\left(1-\rho^2\right)^{(n-1)/2}}{\left(1-\rho r\right)^{n-3/2}}$$

6.2. APROXIMACIÓN NORMAL. CASO DE DISTRIBUCIÓN NO INFORMATIVA

Sustituyendo $\rho=\tanh\xi;r=\tanh z$ se obtiene una nueva aproximación, esta vez mediante la distribución normal:

$$\xi \sim N(z,1/n)$$

Esta aproximación[23] podría usarse para encontrar intervalos de credibilidad para la tangente hiperbólica del coeficiente de correlación y, a partir de estos intervalos, deshaciendo el cambio de variable, podrían encontrarse los correspondientes al coeficiente de correlación[24].

Particularizando, según los casos, se podrían encontrar intervalos de credibilidad

---

[22] Este supuesto es natural, puesto que se ha asumido que la puntuación observada en un test sigue una distribución normal.

[23] Este conjunto de transformaciones fue sugerido por Fisher en su trabajo "On the probable error of a coefficient of correlation deduced from a small sample", publicado en 1971 en la revista Metro (n.1, pp. 3-32).

[24] Box y Tiao (1992) incluyen en la Sección 8.4.8 la comparación de los valores exactos y aproximados para un ejemplo en un pequeño conjunto de datos.

para los coeficientes de fiabilidad descritos anteriormente. Nótese que el resultado anterior puede también aplicarse en el proceso de construcción de cuestionarios en dos pasos diferentes:

- *Estimación de índices de discriminación,* cuando éstos se entienden como correlación simple o correlación corregida de la puntuación de un ítem y la puntuación total en el test (supuesto que usamos el coeficiente de correlación de Pearson).
- *Estudio de validez de criterio,* cuando usamos como indicador la correlación de Pearson de la puntuación en el test con la puntuación en un criterio externo.

6.3. APROXIMACIÓN NORMAL. CASO DEDISTRIBUCIÓN INICIAL INFORMATIVA

La distribución final del coeficiente de correlación se obtiene fácilmente, cuando la distribución inicial tuviese la forma $p(\rho) = \left(1-\rho^2\right)^c$ que incluye a la distribución uniforme (cuando $c=0$). Sin embargo, resulta más sencillo transformar toda la información que se posea sobre el coeficiente de correlación haciendo uso de la transformación de Fisher $\xi = \tanh^{-1}\rho$ porque de este modo se puede usar los resultados que ya se conocen sobre la distribución normal (Lee, 2004).

Supongamos, por ejemplo, que en la primera ocasión se observa un coeficiente de correlación r_1 en una muestra de tamaño n_1, lo que lleva a una distribución final para el coeficiente de correlación $N(tanh^{-1}\ r_{1},\ 1/n_1)$. En una segunda ocasión se observa se observa un coeficiente de correlación r_2 en una muestra de tamaño n_2. Al tomar la primera observación como distribución inicial, podemos aplicar lo que se conocer sobre estimación de la media de la distribución normal. Por tanto, para estimar $tanh^{-1}\ r$ tendremos una distribución final normal, cuyas media y varianza vienen dadas por las siguientes expresiones.

$$Varianza = \frac{1}{n_1 + n_2}$$

$$Media = Varianza(n_1 \tanh^{-1} r_1 + n_1 \tanh^{-1} r_1)$$

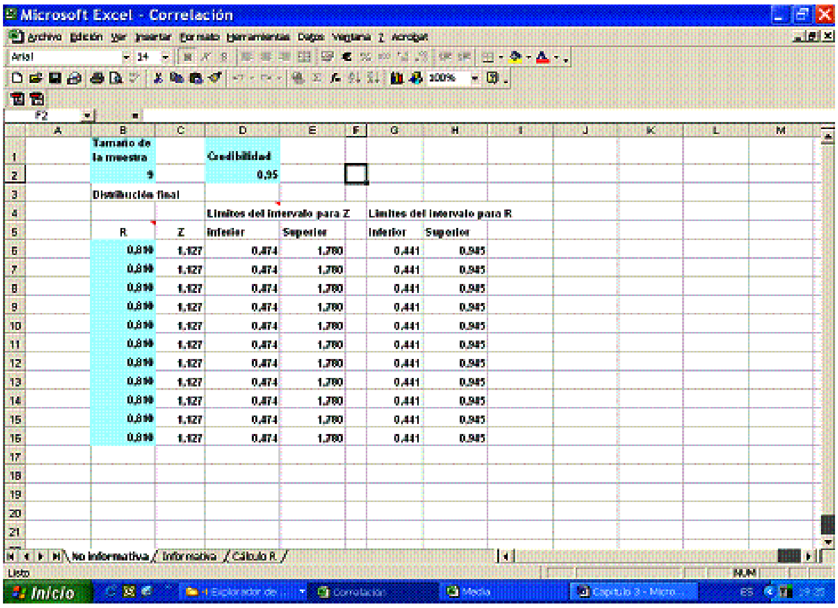
De nuevo se aplicaría esta transformación para obtener un intervalo de credibilidad del arco tangente hiperbólica para el coeficiente de correlación, y, deshaciendo la transformación, se obtiene el intervalo de credibilidad para el coeficiente (Lee, 2004).

6.4. PROGRAMA DE CÁLCULO

Para realizar los cálculos se preparó un programa Excel (Ver Figura 3.5), utilizando las fórmulas de cálculo que se han descrito en los apartados anteriores. Se ha diferenciado (en diferentes hojas del fichero Excel) los casos de distribución inicial informativa y no informativa.

Para la distribución inicial no informativa (Figura 3.5) los datos requeridos son los siguientes: Tamaño de muestra, coeficientes de correlación observados, coeficiente de credibilidad. Se admite el cálculo simultáneo de diferentes coeficientes de correlación sobre la misma muestra (por ejemplo para los estudios de índices de discriminación).

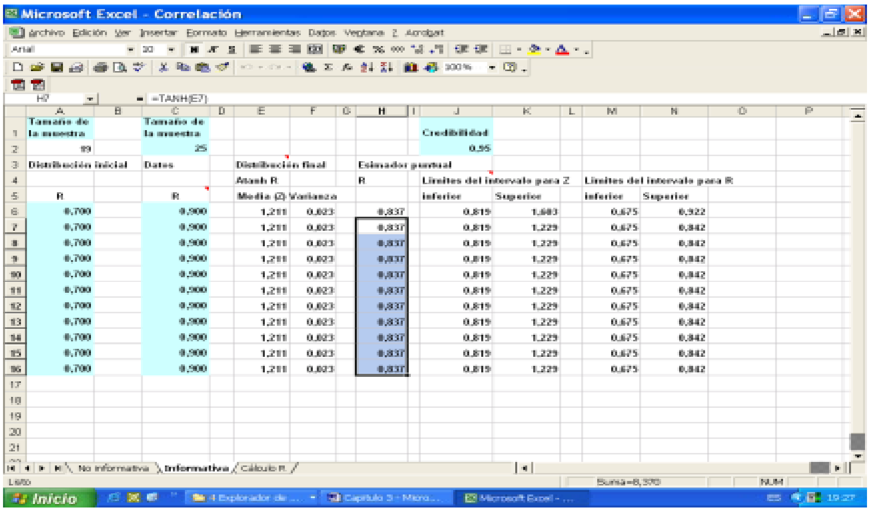
Figura 3.5. Estimación de coeficientes de correlación. Caso no informativo



El programa calcula los intervalos de credibilidad para el arco tangente hiperbólica del coeficiente de correlación y para el coeficiente de correlación.

Para la distribución inicial informativa (Figura 3.6) los datos requeridos son los siguientes: Tamaño de muestra en la distribución inicial y en los datos, coeficientes de correlación observados y valor medio en la distribución inicial, coeficientes de confianza y credibilidad (que pueden ser iguales o diferentes). Se admite el cálculo simultáneo de diferentes coeficientes de correlación sobre la misma muestra (por ejemplo para los estudios de índices de discriminación).

Figura 3.5. Estimación de coeficientes de correlación. Caso informativo



El programa calcula la estimación puntual de *R* en la distribución final, así como los intervalos de credibilidad para el arco tangente hiperbólica del coeficiente de correlación y para el coeficiente.

7. DISCUSIÓN

En este Capítulo se analizan algunas implicaciones de considerar, bajo un punto de vista bayesiano, la teoría clásica de los tests. Respetando el resto de los supuestos de dicha teoría, se introduce la modificación de considerar los parámetros del tests variables aleatorias, con una distribución inicial.

Ello ha permitido aplicar algunos resultados de estimación en inferencia bayesiana a la estimación de la puntuación media del test, diferencia de puntuaciones medias (entre dos grupos de estudiantes), índices de dificultad de los ítems, índices de discriminación (como diferencia de proporciones y como correlación con la puntuación total), coeficiente de fiabilidad test-restest, formas paralelas, test-retest con formas paralelas y dos mitades, así como otros casos de aplicación de la estimación de medias, proporciones, sus diferencias y coeficiente de correlación, en estudios de validación.

Para cada uno de dichos parámetros se han examinado los casos de distribución inicial no informativa e informativa, pues la metodología bayesiana sugiere comenzar con una distribución no informativa en la primera aplicación, para usar, posteriormente la distribución final como distribución inicial informativa en las siguientes aplicaciones. También se pusieron a punto subrutinas Excel que permiten aplicar los anteriores resultados en forma sencilla.

Todo ello permitirá realizar una estimación bayesiana de los principales parámetros del test y los ítems, en el proceso de construcción del cuestionario RPC y en el estudio posterior de evaluación de los conocimientos sobre probabilidad condicional en los estudiantes de psicología (Capítulos 4 a 6).

Para el caso de distribución inicial no informativa, los resultados (intervalos de credibilidad y confianza) obtenidos por las aproximaciones bayesiana y clásica serán muy similares[25], aunque la interpretación es completamente diferente, pues el intervalo de credibilidad proporciona al investigador lo que desea: los límites en los que espera encontrar al parámetro con una cierta probabilidad (epistémica) (Lecoutre, 1996, 1999). Desde el punto de vista teórico, el marco de la probabilidad subjetiva es más razonable, puesto que de ordinario nos encontramos interesados en una población dada de sujetos a los que va dirigido el instrumento y para la cual se estiman sus parámetros (Bernardo, 2003; O'Hagan y Forster, 2004).

Cuando es posible especificar una distribución inicial informativa, el método bayesiano proporciona resultados más precisos, es decir, intervalos más estrechos cuando se igualan el tamaño de muestra y los coeficientes (credibilidad- confianza). El estimador obtenido es sesgado, pero el sesgo se orienta hacia los valores obtenidos de los parámetros en las anteriores aplicaciones del cuestionario, es decir tiene en cuenta el conocimiento previo del investigador. Además, este sesgo va desapareciendo, a medida que aumenta el tamaño de muestra, por lo tanto, en las sucesivas aplicaciones del cuestionario. Puesto que en la construcción de un instrumento de evaluación se realizan pruebas repetidas del mismo, es más adecuado usar la información que se genera en

---

[25] Una pequeña diferencia aparece, debido a diferentes motivos que se han discutido en el capítulo: Uso de la distribución *Beta (0,5, 0,5)* en el caso de la proporción, uso de resultados exactos, en lugar de la aproximación normal, etc.).

cada paso, para informar el siguiente, en lugar de suponer en cada prueba que no se conoce nada de las características de los ítems o del test en su conjunto.

En resumen, pensamos que la metodología bayesiana abre unas posibilidades interesantes en el proceso de construcción de cuestionarios, en cuanto permite incorporar progresivamente la información que se obtiene en diferentes pruebas del instrumento, ayudando al investigador a aprender de su experiencia[26]. Asimismo, en estudios de evaluación, los investigadores podrían utilizar el conocimiento de los parámetros del test y los ítems obtenidos en aplicaciones previas del cuestionario a poblaciones de estudiantes semejantes a las que participan en el estudio, para mejorar su precisión.

[26] Otra razón para la elección bayesiana es que ésta presenta una metodología unificada para la inferencia. Al modelizar los parámetros como variables aleatorias con una distribución, permiten incorporar, tanto la información inicial, como la imprecisión de dicha información (Robert, 2001).

CAPITULO 4.
CONSTRUCCIÓN Y REVISIÓN DEL CUESTIONARIO DE RAZONAMIENTO SOBRE PROBABILIDAD CONDICIONAL (RPC)

1. INTRODUCCIÓN

La problemática del mal uso de la estadística en las ciencias empíricas y la importancia del concepto de probabilidad condicional dentro de la inferencia bayesiana, discutidas en el Capítulo 1 sugirieron la necesidad de asegurar su correcta comprensión en la instrucción de profesionales e investigadores en Psicología, antes de abordar la enseñanza del tema. Por otro lado, el estudio de las investigaciones previas descrito en el Capitulo 2 indicó la conveniencia de elaborar un instrumento comprensivo de evaluación, lo que requería una serie de estudios que se abordan en este capítulo:

- *Estudio 1.* Especificación del contenido de la variable objeto de medición. Se basó, en primer lugar, en la revisión bibliográfica de las investigaciones previas, descrita en el capítulo anterior, que permitió conocer los errores relacionados con este concepto que han sido investigados, y recopilar algunos ítems usados en dichas investigaciones. Una definición más precisa de la variable se fundamentó en el análisis de contenido de una muestra de 21 libros de texto de estadística aplicada, elegida entre los que se usan en la licenciatura psicología en las asignaturas de análisis de datos.

- Estudio 2. *Construcción de la versión piloto del cuestionario RPC*, que a su vez consta de dos sub-estudios: Estudio 2.1: *Pruebas piloto de ítems*. Un total de 49 ítems adaptados de investigaciones previas se probaron en muestras de entre 49 y 117 estudiantes de primer curso de psicología, para determinar su índice de dificultad y la distribución de respuestas a los distractores. En el estudio 2.2 se hizo la *selección final de unidades de contenido* en la definición semántica de la variable y los *ítems* finales del cuestionario que cubrirían dicho contenido *a partir de juicio de expertos*.

- Estudio 3. Una vez que se dispuso de una primera versión del instrumento, se llevó a cabo una *prueba piloto* del cuestionario RPC con nuevas muestras de estudiantes, para obtener una primera determinación de sus características psicométricas.

- Estudio 4. Se *revisó el instrumento piloto* a partir de juicio de expertos para mejorar su legibilidad y disminuir la dificultad de algunos ítems, finalizando, de este modo, el proceso de construcción del cuestionario.

Antes de abordar estos estudios, clasificamos y describimos los objetivos del cuestionario RPC. Estos objetivos, junto con el método de elaboración se ajustaron a las normas APA, AERA y NCME (1985, 1999) para este tipo de trabajos.

2. OBJETIVOS Y CLASIFICACIÓN DEL INSTRUMENTO

Aunque la probabilidad condicional es un concepto básico, cuya comprensión correcta se requiere antes de tratar de mejorar la formación metodológica en cualquiera de las escuelas de inferencia, en la actualidad no se presta una atención adecuada a este concepto en los cursos de iniciación a la inferencia. Por el contrario, hay una tendencia reciente a "enseñar estadística sin probabilidad" (Spatz, 1993; Moore, 1997 a y b) o con una base mínima de probabilidad, propugnada desde libros – por otro lado excelentes- que han tenido una amplia difusión en los cursos universitarios de estadística aplicada (por ejemplo, Moore, 1995)[1].

Es paradójico que se denuncie la mala práctica de la inferencia estadística y se insista en enseñarla descuidando la base probabilística que la sustenta. Es por ello que el *objetivo inicial* en la construcción del instrumento es *contribuir a aportar alguna información sobre el grado de comprensión y las dificultades específicas que los estudiantes de psicología tienen respecto a la probabilidad condicional, como base para una acción diagnóstica y correctiva de los errores, en caso de encontrarlos*.

Fijado este objetivo, los primeros pasos se encaminaron a tratar de encontrar un instrumento de evaluación adecuado y estudiar las investigaciones previas relacionadas, que se describen en el capítulo 2. Desafortunadamente, no se encontró un cuestionario comprensivo que evaluase adecuadamente la comprensión de la probabilidad condicional.

En consecuencia, se decidió iniciar la construcción y sus características fueron determinadas por los objetivos de la investigación y la naturaleza de las inferencias que se

[1] Esta tendencia se ha visto sustentada por la dificultad reconocida que la probabilidad tiene para muchos estudiantes (Shaughnessy, 1977; 1992; Díaz y De la Fuente, 2005b; Tarr y Lannin, 2005).

van a extraer de las puntuaciones obtenidas a partir de su aplicación a una muestra de sujetos[2].

Al tratar de evaluar la comprensión sobre la probabilidad condicional, se tuvo en cuenta que es un constructo inobservable (León y Montero, 2002), por lo que sus características deben ser inferidas de las respuestas de los alumnos. Se trata de un instrumento de medición ya que, por medio de las preguntas planteadas a los encuestados, proporciona una estimación de conocimientos y capacidades de los sujetos, que no son accesibles por simple observación o encuesta (Dane, 1990, Barbero, 1993). Se trata de establecer un conjunto de reglas para asignar a los sujetos números que representan cantidades de atributos (Castro Posada, 2001)

Entre la clasificación de objetivos en la construcción de cuestionarios de Millman y Greene (1989) esta investigación se sitúa en un *dominio curricular*[3], por tanto se partió de un análisis sistemático del concepto, contemplado sus diferentes matices. Aunque el instrumento de medida podría ser aplicado a otras poblaciones de alumnos, se enfocó principalmente en los alumnos de primer curso de Psicología, para poder tomar decisiones sobre la viabilidad de introducir algunas ideas básicas de inferencia Bayesiana a estos alumnos

El instrumento permitirá tomar decisiones sobre la ejecución de los estudiantes particulares con referencia a un dominio curricular concreto, la probabilidad condicional. Un segundo objetivo fue diferenciar, a partir de las puntuaciones en el cuestionario, a los alumnos que tengan una buena comprensión del concepto de probabilidad condicional de los que no la tengan para, en su caso, proporcionar clases de apoyo o una instrucción específica a estos últimos en probabilidad condicional, requisito previo para poder enseñar inferencia estadística, tanto desde el enfoque clásico como bayesiano.

Estos objetivos se relacionan con posibles decisiones relativas al nivel de *instrucción,* que dependen de si un sujeto o grupo de sujetos domina o no ciertas destrezas predeterminadas (Martínez Arias, 1995). También se contemplan posibles decisiones de *diagnóstico,* para identificar las dificultades o errores particulares en el rendimiento de un sujeto o grupo de sujetos, con el objetivo de llevar a cabo una intervención.

---

[2] Una delimitación clara de los objetivos proporciona un marco global para las especificaciones de la variable, así como para el desarrollo, prueba y revisión de los ítems (Sax, 1989).

[3] Conjunto de habilidades y conocimientos que se desarrollan como resultado de la instrucción deliberada sobre un cierto contenido curricular (Millman y Greene, 1989).

Clasificación del instrumento

Se trató de construir un instrumento de *medida centrada en el sujeto* (Millman y Greene, 1989), orientado a localizar a los alumnos en diferentes puntos de un continuo teórico "comprensión de la probabilidad condicional", en función de sus respuestas al cuestionario. Las interpretaciones hechas a partir de las respuestas se refieren a un *dominio específico*, es decir, harán referencia a lo que los sujetos hacen o son capaces de hacer y sus conocimientos y errores sobre el mismo (*test referido a criterio*).

Por tanto, no se realizarán comparaciones de los resultados con respecto a un grupo normativo (Sax, 1989). Se trataría de un cuestionario de *potencia*, ya que el tiempo, aunque controlado, no determinaría el resultado, sino que las diferencias en la puntuación serian debidas a la calidad de su ejecución y conocimiento (Martínez Arias, 1995; APA, AERA y NCME, 1999). Se trataría de un test psicométrico puesto que evalúa las respuestas en forma cuantitativa y se refiere a un rasgo diferenciado del sujeto. Tendría una doble finalidad: investigación y diagnóstico. Se trataría de un test de *ejecución máxima* (en cada ítem el sujeto pone en funcionamiento su capacidad), rendimiento académico y su construcción se basó en la teoría clásica de los tests (López Feal, 1986; Muñiz, 1994).

Respecto al dilema entre precisión y amplitud de contenido, se llegó a un compromiso, en función al tiempo disponible. Mientras que un test unidimensional es muy fiable, podría tener menor validez o cubrir una gama muy estrecha de contenidos. Por ello se tomó un contenido más amplio, incluso cuando la fiabilidad podría ser menor. Esto está relacionado con la dimensionalidad del contenido del instrumento que se refiere a la homogeneidad o heterogeneidad teórica del contenido (Millan y Greene, 1989). Debido a la variedad de tipos de problemas y propiedades relacionados con la probabilidad condicional, se esperó que los resultados empíricos mostrasen más de una dimensión al reflejar el contenido del dominio.

3. ESTUDIO 1. ESPECIFICACIÓN DEL CONTENIDO DE LA VARIABLE OBJETO DE MEDICIÓN

3.1. INTRODUCCIÓN

En este estudio se abordó la definición semántica de la variable objeto de medición, partiendo de la revisión bibliográfica de las investigaciones previas, descrita en el Capítulo 1, que permitió conocer los errores investigados en relación a este concepto, y recopilar algunos ítems usados en dichas investigaciones.

La probabilidad condicional puede definirse con diversos grados de formalización, que varían según el tipo de alumnos a que va dirigida la enseñanza, como en diferentes textos. El estudio se llevó a cabo a partir del análisis de contenido de una muestra de libros de texto de estadística aplicada, elegidos entre los que se usan en la licenciatura psicología para explicar el curso de análisis de datos.

A continuación se describen los fundamentos de definición de la variable, muestra de libros de texto utilizada, metodología seguida y resultados del análisis. Finalmente se presenta la primera delimitación de la definición sintáctica y semántica de la variable y la elaboración de la tabla de especificaciones, que fue posteriormente depurada a partir de juicio de expertos en el estudio 4.

3.2. FUNDAMENTOS DE LA DEFINICIÓN DE LA VARIABLE OBJETO DE MEDICIÓN

El cuestionario RPC se orientó a medir el constructo "comprensión del concepto de probabilidad condicional"[4]. Se tuvo que hacer frente a las siguientes dificultades:

- No hay una única aproximación a su medida.
- Esta medida debía estar basada en muestras limitadas de tareas y conductas; por tanto siempre hay un error, ligado al muestreo.
- El origen y unidad de medida no está bien definido (Martínez Arias, 1995).

[4] Como cualquier constructo psicológico, la comprensión de la probabilidad condicional caracteriza los comportamientos de los individuos y permite explicar patrones de comportamiento. La dificultad de su evaluación se deriva de que sólo pueden ser medido indirectamente, y está sujeto al cambio (Osterlind, 1989).

La definición del constructo se realizó a dos niveles:

- *Definición semántica:* Se especificó los comportamientos observables o reglas de correspondencia entre el constructo y la conducta.
- *Definición sintáctica:* Se describieron las relaciones lógicas del constructo con otras variables.

Además de la descripción cuidadosa de las áreas de contenido que serían evaluadas, la construcción del instrumento se basó en el análisis de los procesos mentales que serían evaluados y su importancia relativa (Osterlind, 1989; APA, AERA y NCME, 1999). Entre los diferentes sistemas jerárquicos para categorizar las operaciones cognitivas útiles al fijar las especificaciones de un instrumento se eligió la taxonomía de Bloom (1956) por ser la más conocida (Sax, 1980; Martínez Arias, 1995). Esta taxonomía consta de los siguientes niveles:

1. *Conocimiento:* Recuerdo de hechos específicos o generales, método, procesos, patrones o estructura. Desde el punto de vista de la evaluación, los ítems implicarán poco más que recuerdo y las alteraciones del material presentado durante la instrucción serán pequeñas. El proceso psicológico que se enfatiza es la memorización. Se requiere alguna relación en el sentido de que un ítem de conocimiento requiere la organización o reorganización de un problema para obtener las claves de información y conocimiento del individuo.
2. *Comprensión:* A su nivel más bajo, traslado, interpretación o extrapolación de un concepto en forma algo diferente del originalmente presentado. El sujeto puede hacer uso del material o idea comunicado sin relacionarlo necesariamente con otra materia o sin ver sus implicaciones finales.
3. *Aplicación:* Uso de abstracciones en situaciones particulares y concretas. Las abstracciones pueden ser ideas generales, reglas o procedimientos o métodos generales, principios técnicos o teorías que deben recordarse y aplicarse.
4. *Análisis:* Descomponer la comunicación en sus elementos o partes constituyentes de modo que la jerarquía de ideas y relaciones entre ideas se expresen y hagan explícitas. Tales análisis tratan de clarificar la comunicación, para indicar cómo se organiza y la forma en que produce su efecto, así como su fundamento y disposición.
5. *Síntesis:* Poner juntos los elementos y partes para formar un todo. Esto implica el

trabajo con las partes, su disposición y colocación hasta constituir un patrón o estructura que no se veía anteriormente con claridad.

6. *Evaluación:* Juicios acerca del valor del material y métodos para un propósito dado. Juicios cuantitativos y cualitativos acerca de la extensión hasta la cual el material y método satisface unos criterios. Uso de estándares de evaluación. Los criterios pueden ser determinados por el estudiante o serles dados.

Se utilizaron los niveles dos a cinco en la anterior taxonomía, aplicados a la comprensión de la probabilidad condicional. Para dotar de una mayor objetividad a la definición de la variable, se realizó un análisis de contenido de una muestra de libros de texto que utilizan los alumnos de Psicología en las asignaturas de análisis de datos[5].

La definición semántica se fundamentó también en la revisión bibliográfica, llevada a cabo, de las investigaciones previas sobre probabilidad condicional, realizada mediante un procedimiento cíclico. Se comenzó analizando algunas tesis doctorales sobre didáctica de la probabilidad y estadística del Departamento de Didáctica de las Matemáticas de la Universidad de Granada, entre ellas algunas específicas sobre la probabilidad condicional (Maury, 1986; Totohasina, 1992; Ojeda, 1995; Sánchez, 1996). La lectura de estas tesis facilitó el realizar una primera versión de los antecedentes y la localización de artículos relacionados con el tema. Progresivamente se completó la revisión, realizado una búsqueda en las fuentes documentales que se citan a continuación:

- Revistas específicas completas de Educación Estadística, desde su aparición: *Journal of Statistics Education*, http://www.amstat.org/publications/jse/, *Statistics Education Research Journal*, http://www.stat.auckland.ac.nz/~iase /publications,*Teaching Statistics*.
- Actas de conferencias internacionales sobre Educación Estadística y Educación Matemática: Conferencias Internacionales sobre Enseñanza de la Estadística: Grey (1982); Davidson y Swift (1988); National Organising Committee (1994), Pereira-Mendoza y cols. (1998); Phillips (2002); Rossman y Chance (2006). Conferencias temáticas sobre educación estadística: Garfield y Burrill (1997); Batanero (2001), Beltrán (2001), Curtis (2002), Engel (2003), Burrill y Camdem (2006).

[5] Para Selander (1990), la mejor forma de conseguir una adecuada perspectiva sobre la enseñanza de un cierto tema es analizar el mismo tema en distintos libros.

- Actas de las conferencias anuales del PME (Psychology of Mathematics Education) en el periodo 1990-2002 y Actas del Grupo de Trabajo sobre Estadística en los Congresos Internacionales de Educación Matemática ICME (Phillips, 1996; Starkings, 2000).
- Libros o capítulos de libros que contienen revisiones de la investigación en educación estadística o sobre razonamiento estadístico: Morrison y Henkel (1970); Nisbett y Ross (1980); Kahneman, Slovic y Tversky (1982); Pérez Echeverría (1990); Scholz (1991); Shaughnessy (1992, 2006, en prensa); Shaughnessy, Garfield, y Greer (1996); Gal y Garfield (1997); Harlow, Mulaik y Steiger (1997); Sedlmeier (1999), Ben-Zvi y Garfield (2004), Jones (2005).
- Bases de datos documentales[6]: Se han hecho consultas a bases de datos temáticas, con las palabras claves "educación", "enseñanza", "aprendizaje" combinadas con los términos "Probabilidad", "probabilidad condicional" e "independencia", entre otras ProQuest, Education Journals y ProQuest Psychology Journals, introduciendo las palabras clave: "conjunction fallacy", "bayesian reasoning" y "conditional probability".

3.3. MÉTODO

3.3.1. MUESTRA

En la elección de la muestra de documentos para llevar a cabo el análisis de contenido se decidieron las fuentes, número de documentos y contenidos a analizar (Visauta, 1989). Las fuentes fueron libros de texto destinados a alumnos de Psicología, incluyendo textos de orientación clásica y bayesiana, ya que el cuestionario se destinaba principalmente a alumnos de esta especialidad, y para tener en cuenta los contenidos de probabilidad condicional requeridos en los dos enfoques de inferencia.

La población posible de textos entre los cuales seleccionar la muestra, se decidió recurriendo a los libros de texto recomendados en las asignaturas de análisis de datos en las universidades españolas. Se realizó un muestreo polietápico (Kish, 1970; Azorín, y

---

6 Una de las bases utilizadas es Mathdi (Mathematics Didacktik) que incluye resúmenes de los trabajos publicados en prácticamente todas las revistas publicadas en didáctica de las matemáticas y un buen número de revistas de educación y psicología que publican algunos artículos relacionados con las matemáticas.

Sánchez-Crespo, 1986). Para la primera etapa se listaron las 31 universidades españolas en las que se imparte la licenciatura de Psicología, que constituyeron la población objetivo y tratando de recopilar los programas de las asignaturas de análisis de datos de cada una de estas universidades, junto con la lista de libros recomendados en la bibliografía. En la mayoría de los casos, esta información estaba disponible en las páginas web de los departamentos encargados de impartir esta asignatura en las diferentes universidades. En los casos en que no había disponible un programa actualizado de la asignatura con este método, se contactó con los profesores de las asignaturas, mediante correo electrónico, solicitándoles esta información. Mediante este procedimiento se consiguió la bibliografía recomendada de 23 de estas universidades para dicha asignatura, lo que constituye la población alcanzada en el estudio.

Tabla 4.1. Lista de libros recomendados en al menos 4 Universidades

Libro	Nº citas	
Botella, J., León, G. y San Martín, R. (1993). *Análisis de datos en psicología I*. Madrid: Pirámide	20	A
Pardo, A. y San Martín, R. (1994). *Análisis de datos en psicología II*. Madrid. Pirámide.	15	N
Amón, J. (1984). *Estadística para psicólogos I*. Madrid: Pirámide.	14	N
Amón, J. (1987). *Estadística para psicólogos II*. Estadística inferencial. Madrid: Pirámide.	14	A
Glass y Stanley (1974). *Métodos estadísticos aplicados a las ciencias sociales*. México: Prentice Hill.	7	A
San Martín, R. y Pardo, A. (1989). *Psicoestadística. Contrastes paramétricos y no paramétricos*. Madrid: Pirámide.	7	N
San Martín, R., Espinosa, L. y Fernández, L. (1987). *Psicoestadística descriptiva*. Madrid: Pirámide.	7	N
Peña, D. y Romo, J. (1997). *Introducción a la estadística para las ciencias sociales*. Madrid: McGraw-Hill.	6	A
Merino, J.M., Moreno, E., Padilla, M., Rodríguez Miñon, P. y Villarino, A. (2002). *Análisis de datos en psicología 1*. Madrid: UNED.	6	A
Botella, B. y Barriopedro, M.I. (1991). *Problemas y ejercicios de psicoestadística*. Madrid: Pirámide.	5	A
Macía, A., Lubin, P. y Rubio, P. (2000). *Psicología matemática II*. Madrid: UNED.	5	N
MacRae, S. (1995). *Modelos y métodos para las ciencias del comportamiento*. Barcelona: Ariel.	5	N
Padilla, M., Merino, J.M. y Pardo, A. (1986). *Psicología matemática I: Ejercicios resueltos*. Madrid: UNED.	5	A
San Martín, R. y cols. (1987). *Psicoestadística: Estimación y contraste*. Madrid: Pirámide	5	A
Peña, D. (1986). *Estadística: Modelos y métodos I. Fundamentos*. Madrid: Alianza Uiversidad Textos.	4	A
Wonnacot, T.H. y Wonnacot, R.J. (1991). *Estadística básica práctica*. México: Limusa.	4	A
Abelson, R.P. (1998). *La estadística razonada: Reglas y principios*. Barcelona: Paidós.	4	N
Cuadras, C.M., Echevarría B., Mateo, J. y Sánchez, P. (1984). *Fundamentos de estadística. Aplicación a las ciencias humanas*. Madrid: Promociones Publicaciones Universitarias.	4	A
Nortes Checa, A. (1993). *Estadística teórica y aplicada*. Barcelona: PPU.	4	A
Spiegel, M.R. (1991). *Estadística.* México: Mc Graw Hill.	4	A
Jáñez, L. (1989). *Fundamentos de psicología matemática*. Madrid: Pirámide.	4	N

N: No incluye el tema de probabilidad condicional, no procede. A: Analizado

En la segunda etapa del muestreo, la lista de referencias en cada uno de los programas se introdujo en una hoja de cálculo Excel y se clasificó alfabéticamente, obteniéndose un total de 149 libros diferentes, junto con su frecuencia de aparición. Una primera revisión de esta bibliografía permitió clasificarla en dos tipos de textos:

1. Libros de texto de estadística aplicada (en total, 79 libros diferentes).
2. Libros de ampliación sobre temas relacionados, tales como, historia o filosofía de la estadística, estadística descriptiva, el diseño de experimentos, análisis de varianza, la metodología de investigación, el muestreo o el uso del software (en total 70 libros). Por su temática, ninguno de estos libros incluye específicamente el tema de la probabilidad condicional.

La muestra empleada en el trabajo se eligió de entre los 79 libros de la categoría 1, y fue intencional (Azorín y Sánchez- Crespo, 1989) pues se siguieron ciertos criterios en su selección. En primer lugar se seleccionaron todos aquellos libros que habían sido citados por al menos cuatro profesores (recomendados en al menos el 18% de los programas para la asignatura de análisis de datos) resultando un total de 20 libros que se presentan en la tabla 4.1.

Se analizaron estos libros, clasificándolos en dos grupos, según el enfoque con que se aborde la inferencia:

1. Libros en los que reduce al mínimo o se suprime la probabilidad, pasando directamente de la estadística descriptiva al estudio de la inferencia. En todo caso, se introduce brevemente la definición de probabilidad y en algunos casos las distribuciones binomial, normal, T u otras ideas sobre la distribución normal y otras distribuciones de probabilidad. Estos libros, aunque examinados inicialmente, no se han incluido en el análisis de contenido, puesto que no contienen el tema de la probabilidad condicional. Son los que en la tabla 4.1 aparecen con la letra N en la última columna
2. Textos que contienen un capítulo explícito de probabilidad. Estos textos son aquellos en los que con mayor o menor extensión incluyen la enseñanza de la probabilidad condicional y se incluyeron en este estudio. Aparecen en la tabla 4.1 con la letra A.

Tabla 4.2. Documentos analizados

	Documento
Enfoque clásico	1. Botella, J., León, O. G. y San Martín, R. (1993). *Análisis de datos en psicología I.* Madrid: Pirámide.
	2. Amón, J. (1987). *Estadística para psicólogos II. Estadística inferencial.* Madrid: Pirámide.
	3. Glass, G. V. y Stanley, J. C. (1974). *Métodos estadísticos aplicados a las ciencias sociales*. México: Prentice Hall.
	4. Peña, D. y Romo, J. (1997). *Introducción a la estadística para las ciencias sociales.* Madrid: McGraw-Hill. 2ª Edición
	5. Merino, J. M., Moreno, E., Padilla, M., Rodríguez Miñon, P. y Villarino, A. (2002). *Análisis de datos en psicología 1*. Madrid: UNED.
	6. Botella, B. y Barriopedro, M. I. (1991). *Problemas y ejercicios de psicoestadística*. Madrid: Pirámide.
	7. Padilla, M., Merino, J. M. y Pardo, A. (1986). *Psicología matemática I: Ejercicios resueltos*. Madrid: UNED.
	8. San Martín, R. y cols. (1987). *Psicoestadística: Estimación y contraste*. Madrid: Pirámide
	9. Peña, D. (1986). *Estadística: Modelos y métodos I. Fundamentos*. Madrid: Alianza Universidad Textos.
	10. Wonnacot, T. H. y Wonnacot, R. J. (1991). *Estadística básica práctica*. México: Limusa.
	11. Cuadras, C. M., Echevarría B., Mateo, J. y Sánchez, P. (1984). *Fundamentos de estadística. Aplicación a las ciencias humanas.* Madrid: Promociones Publicaciones Universitarias.
	12. Nortes Checa, A. (1993). *Estadística teórica y aplicada.* Barcelona: PPU.
	13. Spiegel, M. R. (1991). *Estadística*. México: Mc Graw Hill.
Enfoque bayesiano	14. Sacerdote, A. y Balima, G. (En preparación). *Estadística bayesiana*. Buenos Aires: Universidad de Buenos Aires. On line: http://www.fi.uba.ar/materias/6109/libro/cap-1.pdf
	15. Berry, D. A. (1995). *Basic statistics: A bayesian perspective*. Belmont: Wadsworth.
	16. Bernardo, J. M. (1981). *Bioestadística. Una perspectiva bayesiana*. Barcelona: Vicens-Vives.
	17. Serrano Angulo. J. (2003). *Iniciación a la estadística bayesiana*. Madrid: La Muralla.
	18. Albert, J. H. y Rossman, A. (2001). *Workshop Statistics. Discovery with data*. A bayesian approach. Bowling Green, OH: Key College Publishing.

La muestra incluyó libros de texto publicados durante el periodo 1984 – 2003. En concreto, los libros y documentos analizados se presentan en la tabla 4.2 donde se diferencia los libros que presentan la inferencia con un enfoque clásico y aquellos que la presentan con un enfoque bayesiano.

La muestra se completó con un número reducido de libros de texto de estadística aplicada, dirigidos a alumnos de primeros cursos de estadística y con orientación bayesiana, puesto que el objetivo es recoger los contenidos de probabilidad condicional que sean útiles a los dos tipos de inferencia. Esta orientación y el tipo de alumnos a que va dirigido se señalan explícitamente en el texto por sus autores, en la introducción. También se deduce de la lectura del mismo, del nivel de formalización reducido, y de los contenidos explícitos de inferencia bayesiana. Estos textos fueron elegidos intencionalmente, puesto que sólo uno de los textos elegidos en la segunda etapa tiene algunas nociones de inferencia bayesiana[7].

3.3.2. MATERIAL Y PROCEDIMIENTO

El análisis de contenido, utilizado para la definición de la variable, es un tipo de medición aplicado a un texto, que se basa en la idea de que las unidades del texto pueden clasificarse en un número reducido de categorías (Weber, 1985). Sirve para efectuar inferencias mediante la identificación sistemática y objetiva de las características específicas de un texto (Ghiglione y Matalón, 1989). Su objetivo final es la búsqueda del significado, cuya percepción depende de la existencia de las señales y de las características de los significantes. *"Es un proceso complejo, seguramente el que más esfuerzo intelectual requiere de entre todas las técnicas de análisis de datos y es uno de los pocos campos de los comprendidos en las etapas finales del proceso de investigación en la que el investigador desempeña un papel individual y creativo"* (Fox, 1981, p. 704).

No hay una única forma de efectuar un análisis de contenido (Weber, 1985). En este estudio, se han seguido las técnicas lógico-semánticas del llamado análisis de contenido temático, que son las más frecuentes y típicas. En ella se recurren a la lógica, y al conocimiento previo sobre el tema adquirido a través de la revisión bibliográfica, para resumir el contenido del texto, definir categorías y verificar la validez de las mismas. Llevan al analista a actuar principalmente como verificador, clasificador y eventualmente estadístico. Sólo considera la presencia de términos o conceptos, independientemente de las relaciones entre ellos (Ghiglione y Matalón, 1989).

Además se restringió al análisis de contenido cualitativo, en donde simplemente se trata de verificar la presencia – ausencia de una cierta característica, sin llegar a cuantificarla (Visauta, 1989). Es un análisis directo (en la terminología de este mismo autor) puesto que se ciñe estrictamente al contenido de la unidad de análisis sin ir más allá de lo que ésta contiene.

Una vez elegidos los textos a analizar, se describe el proceso seguido hasta llegar a los datos básicos[8]. Se comenzó eligiendo la unidad de contenido a analizar, y elaborando a

[7] Se trata del libro de Jesús Amón (1987), *Estadística para psicólogos*, volumen 2.

continuación un conjunto de categorías, junto con un fundamento lógico que sirva para asignar las unidades a estas categorías (Fox, 1981). Siendo un análisis de datos cualitativo, el proceso se realizó en varias etapas dividiendo el texto en unidades, que fueron clasificadas en categorías y a continuación ensambladas y relacionadas para conseguir un todo coherente (Goetz y Lecompte, 1988)[9]. El proceso se dividió en tres etapas: la reducción de datos, el análisis de datos y la obtención y verificación de conclusiones (Delgado y Gutiérrez, 1994; Gil Flores, 1994; Huberman y Miles, 1994).

La primera operación fue la separación de segmentos o unidades de análisis, en varios niveles[10] Los libros de texto analizados se estructuraron en capítulos o unidades primarias. Siguiendo este criterio, una vez seleccionados los libros, una primera lectura de los mismos permitió seleccionar los capítulos que contuviesen nociones de probabilidad o inferencia, que han sido las *unidades primarias* de análisis.

La división de las unidades primarias puede realizarse siguiendo diversos criterios, como espaciales o temporales, aunque es más frecuente hacerlo en función del tema abordado (Goetz y Lecompte, 1988; Ghiglione y Matalón, 1989; Delgado y Gutiérrez, 1994). En el estudio se fijaron, los temas de probabilidad simple y conjunta, experimento compuesto, teorema de la probabilidad total y Bayes, dependencia e independencia... Para cada uno de estos contenidos, extraídos de la revisión bibliográfica, se identificaron en cada uno de los capítulos *unidades secundarias* de análisis que fueron los párrafos que cumplían algunos de los siguientes criterios: a) contienen las definiciones explícitas de algunos de los conceptos descritos; b) contienen la descripción de algunas de sus propiedades; c) unidades que proponen problemas o ejercicios que requieren el uso implícito o explícito de los mismos. Estos tipos de unidades se eligieron porque se quería evaluar tanto el conocimiento conceptual (definiciones y propiedades) como procedimental (problemas y ejercicios).

En estas unidades secundarias se analizó el texto, siguiendo los siguientes pasos:

1. Se hizo una lectura analítica de los textos, recogiendo todos los aspectos relacionados

---

[8] El dato es inseparable del proceso que lo registra y del modo en que se comunica, ya que el investigador no puede captar la realidad, sino lo que registra son las elaboraciones que hace de la misma. En este proceso intervienen la elaboración conceptual y el lenguaje en que se expresa la información, que es el resultado del modo en el que el investigador se acerca al problema e interacciona con el objeto de su estudio (Gil Flores, 1994).

[9] Un texto se divide en componentes organizados de acuerdo a ciertos patrones. El análisis deberá necesariamente tener en cuenta esta estructura. Se accede, al sentido del discurso partiendo del significado de todos sus componentes y de la interrelación de éstos (Weber, 1985).

[10] Esta segmentación de los datos en unidades relevantes y significativas es considerada como una de las prácticas más características del análisis de datos cualitativos.

con el contenido analizado. Se examinaron los datos para encontrar en ellos determinados componentes temáticos que permitan clasificarlos en una u otra categoría de contenido (Gil Flores, 1994).

2. Se clasificaron dichos párrafos en función de las diversas propiedades matemáticas a las que se hacía referencia, es decir, en función de los contenidos conceptuales y procedimentales resaltados explícitamente o precisados implícitamente, cuando se tratase de un ejercicio propuesto al alumno[11].
3. Se compararon los diversos párrafos clasificados en un mismo apartado para cada categoría, depurándose la clasificación. La categorización permitió agrupar conceptualmente las unidades que correspondían a un mismo tópico[12].
4. Se elaboraron tablas, recogiendo los contenidos incluidos en los diferentes textos (tanto conceptuales como procedimentales). Esta transformación y ordenación de los datos permitió presentarlos en modo abarcable y operativo para la extracción de conclusiones. Las tablas utilizadas, que pueden considerarse como matrices de incidencia, permiten comprobar si en cada uno de los textos, se presenta o no una cierta categoría.
5. Se elaboró el informe de análisis incluyendo ejemplos de las diferentes categorías, elaboradas mediante un proceso interpretativo guiado por el análisis epistemológico de los conceptos intervinientes y por los resultados de la investigación psicológica y educativa sobre el razonamiento estocástico y el aprendizaje.

Fiabilidad y validez

Las tablas resúmenes de datos fueron revisadas dos veces por la autora del trabajo y posteriormente por otro investigador, para comprobar la ausencia o presencia en los libros de texto de determinadas categorías, procediéndose a la corrección de los escasos desacuerdos con la codificación original. Este procedimiento asegura una adecuada fiabilidad en el proceso de codificación de los datos a partir de las categorías definidas[13].

---

[11] El guión para el análisis de este material curricular estuvo constituido por la lista de contenidos conceptuales y procedimentales que se identificaron en el estudio de las investigaciones previas (capítulo 2) y se fue completando inductivamente durante el análisis de los textos (Martínez Bonafé, 1995).

[12] El proceso de categorización conlleva una decisión que es cómo asociar cada unidad a una determinada categoría "Una categoría soporta un significado o tipo de significado" (Gil Flores, 1994, p. 47).

[13] La fiabilidad del análisis de contenido depende del procedimiento de codificación y de la ambigüedad de las categorías (Ghiglione y Matalón, 1989). Metodológicamente se estima calculando el porcentaje de veces que dos codificaciones independientes coinciden cuando codifican el mismo material (Fox, 1981).

La validez se justifica mediante al juicio externo de expertos para valorar la pertinencia de las categorías elaboradas (especificaciones de contenidos) a la finalidad del instrumento, desechándose aquellas categorías en que no se alcanzó acuerdo en este proceso[14]. Este juicio de expertos se lleva a cabo en el estudio 2.2.

3.3.3. ANÁLISIS

En lo que sigue se presentan los resultados del análisis de contenido, organizados en dos apartados:

1. *Conocimiento conceptual presentado en los libros*. Se refiere a la definición, propiedades y relaciones con otros conceptos. Se organiza en los siguientes apartados: Probabilidad condicional y sus propiedades; dependencia, independencia e intercambiabilidad; y probabilidad total/ teorema de Bayes. Además se analizan los recursos didácticos empleados en los libros para introducir este conocimiento conceptual.
2. *Conocimiento procedimental*, que se refiere al conocimiento de los algoritmos y procedimientos que llevan a la solución de los problemas así como de los símbolos y sintaxis necesarios para aplicarlos. Este conocimiento procedimental queda definido por los problemas propuestos, cada uno de los cuales requiere un algoritmo que el alumno debe conocer para resolverlo e implica el conocimiento del lenguaje matemático.

3.3.3.1. CONOCIMIENTO CONCEPTUAL

El conocimiento conceptual, se delimitó examinando las definiciones, propiedades y relaciones con otros conceptos presentados explícitamente en la muestra de libros de texto. También se consideró que un texto incluye una propiedad o definición cuando, aunque no se refiere explícitamente a ella, la describe indirectamente o la usa explícitamente en la resolución de un problema. En lo que sigue se describen las categorías encontradas, incluyendo ejemplos que permitan clarificar el significado dado a estas categorías.

[14] Respecto a la validez, Fox (1981) indica que el código debe tener una validez inmediata en el sentido de que los temas niveles y las categorías deben tener una relación directa con la finalidad para la que se han creado.

Probabilidad condicional

Se encontró la definición de probabilidad condicional en la mayor parte de los libros analizados. Algunos libros definen intuitivamente la probabilidad condicional *P(A/B)* de un suceso *A* dado otro suceso *B* como la probabilidad de que ocurra *A* sabiendo que *B* se ha verificado. Desde un punto de vista más formal se define mediante la expresión siguiente:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \text{ siempre que } P(B) > 0.$$

Por ejemplo, Peña (1986) define la frecuencia relativa de *A* condicionada a la ocurrencia de *B* *"considerando únicamente los casos en los que aparece B y viendo en cuántos de esos casos ocurre el suceso A; es, por tanto, igual a la frecuencia de ocurrencia conjunta de A y B, partida por el número de veces que ha ocurrido B"* (p. 71) y de esta definición deduce el concepto de probabilidad condicional, usando la siguiente notación:

$$\text{“} P(A/B) = \frac{P(AB)}{P(B)} \text{”},$$

indicando *"donde AB representa el suceso ocurrencia conjunta de A y B y suponemos P(B)>0"* (Peña, 1986, p. 76).En esta definición está implícita la restricción del espacio muestral. Se hace notar que no usa explícitamente el símbolo de la intersección de sucesos para la probabilidad conjunta, aunque el autor resalta explícitamente la diferencia entre probabilidad conjunta y probabilidad condicionada.

Propiedades de la probabilidad condicional

Además de la definición de probabilidad condicional, los diferentes libros incluyen algunas de sus propiedades, demostradas o sólo enunciadas. Peña (1986) hace hincapié en las siguientes propiedades:

- Es importante diferenciar entre $P(A \cap B)$ y *P(A/B)*
- Una probabilidad conjunta $P(A \cap B)$ es siempre menor que las probabilidades simples *P(A)* y *P(B)*.
- Una probabilidad condicionada *P(A/B)* puede ser mayor, menor o igual que *P(A)*.

- El espacio muestral en la probabilidad condicional *P(A/B)* queda restringido a *B*.

Nortes (1993) indica específicamente y demuestra que una probabilidad condicional cumple los tres axiomas de la probabilidad: la probabilidad condicional es siempre un número entre 0 y 1; la probabilidad condicional del suceso seguro es igual a 1 y el axioma de la unión.

Cuadras y cols. (1984) también resaltan que la probabilidad condicional implica una reducción del espacio muestral. Estos autores distinguen entre *P(A)* probabilidad libre o incondicionada, donde el espacio muestral es *Ω (E)* y *P(A/B)* probabilidad condicionada, donde el espacio muestral es *B*. *B* es el suceso condicionante o lo que Cuadras y cols. llaman "información". Sacerdote y Balima (en preparación) resalta el hecho que "condicionar" implica un "saber" y también recuerdan que nunca podría aparecer en una fórmula una suma o diferencia de probabilidades condicionadas a diferentes sucesos, dado que en este caso las probabilidades estarían definidas sobre distintos espacios muestrales.

Independencia

Un concepto relacionado en estos libros con la probabilidad condicional es el de independencia que, aunque matemáticamente puede deducirse de la regla del producto: *A y B son independientes si y sólo si P (A∩B) =P (A) x P(B),* también se relaciona con el de probabilidad condicional[15]. Cuadras y cols. (1984) afirman que esta propiedad es condición necesaria y suficiente para que dos sucesos sean independientes. Esta regla del producto se puede extender a más sucesos independientes:

$$P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$$

Los autores de los textos analizados han definido el concepto de independencia de diversas formas:

- *"Dos sucesos son independientes si la verificación de uno no altera la probabilidad del otro P(A/B) =P(A)."* (Botella, León y San Martín, 1993, p. 282).
- *"Dos sucesos A y B son independientes si el conocimiento de la ocurrencia de uno no modifica la probabilidad de aparición del otro"* (Peña y Romo, 1997).

[15] Ya que dos sucesos son independientes si la probabilidad de uno de ellos no cambia al condicionarlo por el otro (definición intuitiva, "a priori" de independencia (Maury, 1996).

- *"A es independiente de B si P(A/B)=P(A)"* (Nortes Checa, 1993, p. 220).

Según Peña (1986), la independencia entre sucesos, aunque se puede prever en algunos casos, generalmente debe determinarse experimentalmente. Cuadras y cols. (1984) deducen algunas propiedades más: Si A y B son dos sucesos independientes, también son independientes $\overline{A}$ y B; A y $\overline{B}$; $\overline{A}$ y $\overline{B}$ (p. 116).

Dependencia

Del concepto de independencia se deduce el concepto de sucesos dependientes. Cuadras y cols. (1984) lo definen de esta forma: "*Si P(A/B) ≠ P(A), la presencia de B altera la probabilidad de A y se dice que el suceso A es estocásticamente dependiente del suceso B. Si P(A/B) > P(A), la presencia del suceso B favorece al suceso A. Si P(A/B) < P(A), la presencia del suceso B desfavorece al suceso A."* (p.114).

Un concepto que, aunque no está directamente relacionado con los dos anteriores, se suele confundir con ellos es el de sucesos *mutuamente excluyentes*, que son dos sucesos que no pueden ocurrir simultáneamente. La probabilidad conjunta de dos sucesos mutuamente excluyentes es igual a cero[16].

Otro punto que los alumnos confunden es la relación entre dependencia y causalidad. En Sacerdote y Balima (en preparación) se discute explícitamente esta diferencia, indicando que "*la estadística no se ocupa de las relaciones causa-efecto;... la estadística puede ayudar a orientar una ciencia que estudia la naturaleza hacia una explicación causal, pero no a justificarla"* (p. 16).

También resalta que *"el tiempo no interviene en la estadística, aunque para algunos modelos será el factor ordenador de los resultados y el orden puede interesar"* (p. 16), este comentario lo hace para justificar que es posible calcular tanto una probabilidad condicional directa (respecto al tiempo) como su inversa.

Intercambiabilidad

En algunos de los textos de enfoque bayesiano se incluye este concepto, como hipótesis no tan restrictiva como la independencia. Por ejemplo, en Berry (1996, p. 134) se

[16] Precisamente, dos sucesos mutuamente excluyentes son dependientes, ya que la ocurrencia de uno hace que el otro no pueda ocurrir.

presenta este concepto en la forma siguiente:

"Dos experimentos son intercambiables si se cumple lo siguiente:

1. *Los posibles resultados son los mismos en los dos;*
2. *La probabilidad de cada resultado de uno es la misma que en el otro;*
3. *La probabilidad condicional para el segundo experimento, dados los resultados del primero son las mismas que las probabilidades condicionales del primero, dado el segundo".*

Regla del producto

Una propiedad relacionada con las anteriores es la regla del producto, que permite calcular las probabilidades en experimentos compuestos. Algunos de estos textos, como Nortes (1993), diferencian entre sucesos independientes y dependientes para la aplicación de la regla del producto, proporcionando explícitamente las fórmulas, o bien ejemplos, en los dos casos.

Para dos sucesos independientes, la regla del producto se define en la fórmula siguiente: $P\ (A \cap B) = P(A)\ x\ P(B)$. Botella y cols. (1993) introducen el teorema del producto del siguiente modo: "*La probabilidad de verificación simultánea de dos sucesos independientes es igual al producto de sus probabilidades simples"* (p. 283). Para más de dos sucesos independientes, se iría generalizando la regla (por ejemplo, para tres sucesos independientes: $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$).

$$P\ (A_1 \cap A_2 \cap \ldots \cap A_k) = P(A)\ x\ P(A_2)\ x\ \ldots\ x\ P(A_k)$$

Para dos sucesos dependientes, la probabilidad de que ocurran A y B (sucesivas o simultáneas) es igual a la probabilidad de que ocurra A por la probabilidad de que ocurra B condicionando de A, es decir, según expresa Nortes Checa (1993, p. 219):

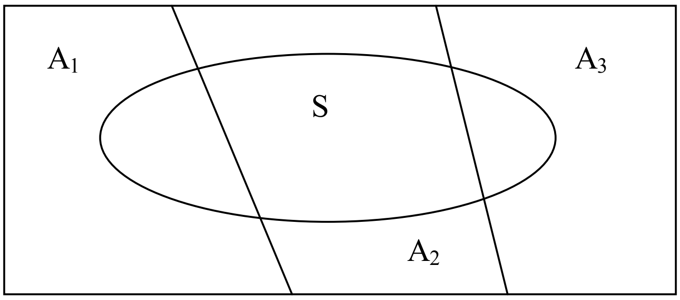
"$P\ (A \cap B) = P\ (A)\ x\ P(B/A)$

En el caso de más de dos sucesos dependientes, la regla del producto sería:

$$P(A \cap B \cap C) = P(A) \times P(B / A) \times P(C / A \cap B)\text{''.}$$

Ley de la probabilidad total

Introducido el experimento compuesto, se suele pasar al teorema de la probabilidad total. Diversos autores hacen uso de diagramas como el siguiente para ejemplificar esta situación (Cuadras y cols, 1984, p. 122).



Dados *n* sucesos $A_1, A_2, ..., A_n$ incompatibles o mutuamente excluyentes, tales que su unión de el espacio muestral *E* y conocidas las probabilidades $P(A_1), P(A_2), ... P(A_n)$ así como un suceso *B* del que conocemos su probabilidad condicionada $P(B/A_1), P(B/A_2), P(B/A_n)$, a cada uno de los sucesos A_i, la ley de probabilidad total permite conocer *P(B)* (Nortes Checa, 1993, p. 223).

$$P(B) = \sum_{i=1}^{n} P(A_i) \times P(B / A_i)$$

En otros casos sólo se presenta implícitamente como parte del estudio del Teorema de Bayes (es el caso de Albert y Rossman, 2001).

Probabilidad inversa y Teorema de Bayes

Cuadras y cols. (1984, p. 122) definen los sucesos $\{A_1, A_2, ..., A_n\}$ con $P(A_i) > 0$ como una partición del espacio muestral *E*. A estos sucesos los denominan *"causas"* y al suceso *B* "*efecto". B* es un suceso cualquiera que si ocurre, lo hace conjuntamente con uno de los sucesos de la partición $\{A_1, A_2, ..., A_n\}$.

Sabiendo que ha ocurrido *S*, podemos calcular la probabilidad de que proceda de la causa A_i haciendo uso del teorema de Bayes[17].

$$P(A_i / B) = \frac{P(A_i) \times P(B / A_i)}{\sum_{i=1}^{n} P(A_i) \times P(B / A_i)}$$

Como resumen, presentamos en la tabla 4.3 los contenidos que, en relación a la probabilidad condicional, se presentan en los documentos analizados. Hacemos notar que

[17] Cuadras y cols. (1984, p.125) hablan específicamente de probabilidades a priori P(Ai), probabilidades a posteriori P(Ai/B) y verosimilitudes P(B/Ai).

lo más común es incluir la definición de probabilidad condicional y conjunta, dependencia e independencia y luego se observa una variación en las propiedades que incluyen los libros.

Tabla 4.3. Conceptos y propiedades en los libros analizados

	1. Botella,Leon	2. Amón	3. Glass Stanley	4. Peña Romo	5. Merino	6. Botella Barropiedro	7. Padilla Merino	8. San Martín	9. Peña	10. Wonnacot	11. Cuadras, Echeverria	12. Nortes Checa	13. Spiegel	14. Sacerdote	15. Berry	16. Bernardo	17. Serrano	18. Albert
Definición prob. condicional	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
P(B) >0 para poder definir *P(A/B)*		x					x			x	x	x					x	
Axiomas en la probabilidad condicional								x				x			x	x		
Restricción del espacio muestral en la probabilidad condicional	x	x		x						x					x			
No se puede sumar o restar probabilidades condicionadas a diferentes sucesos														x				
Regla del producto $P(A \cap B)$	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x
Diferencia entre *P(A/B)* y $P(A \cap B)$											x	x						
$P(A \cap B) < P(A)$	x										x							
Tiempo y probabilidad condicional														x	x			
Experimentos intercambiables															x	x		
Independencia	x	x	x	x			x	x	x	x	x	x	x		x	x	x	x
Dependencia	x						x			x	x	x	x		x		x	
Diferencia dependencia- causalidad										x				x				x
Sucesos mutuamente excluyentes; diferencia con independencia		x															x	
Probabilidad total				x	x	x	x	x				x		x	x	x	x	
Teorema de Bayes		x		x	x	x	x	x	x	x		x		x	x	x	x	x

El teorema de la probabilidad total y teorema de Bayes no siempre se incluyen en los libros con enfoque de inferencia clásica, posiblemente porque este teorema no jugará un papel tan importante como el que desempeña en la inferencia bayesiana. En este último caso, el Teorema de Bayes siempre se incluye y generalmente el de la probabilidad total, aunque a veces está solo implícito. El concepto de intercambiabilidad ni siquiera se incluye en todos los textos de enfoque bayesiano, posiblemente porque los que hemos elegido tienen un nivel muy elemental y están dirigidos a alumnos con conocimientos no muy fuertes en matemáticas.

Destacamos que son pocas las propiedades específicamente resaltadas en los textos, aunque los estudiantes tienen bastantes problemas con ellas.

Los errores relacionados con la inferencia estadística podrían evitarse con una correcta comprensión de la probabilidad condicional y sus diversas propiedades. Se incluye aquí también el teorema de Bayes que permite profundizar en la diferencia entre una probabilidad y su inversa. En consecuencia, en la definición semántica de la variable se recogerán los contenidos resumidos en la tabla 4.3.

3.3.3.2. CONOCIMIENTO PROCEDIMENTAL

Se analizaron también los tipos de problemas encontrados, que se clasificaron en las siguientes categorías:

1. Calcular una probabilidad condicional, dentro de un único experimento, como en el siguiente ejemplo, tomado de Nortes (1993, p. 218): *"Considerando el lanzamiento de un dado y siendo A={sacar impar} y B={3,4}, la probabilidad P(A/B) de sacar número impar en el lanzamiento de un dado habiendo sacado antes 3 o 4 es ..."*

2. Calcular una probabilidad condicional en un experimento compuesto, en el contexto de muestreo con reposición (regla de Laplace), como en el siguiente ejemplo, tomado de Spiegel (1991, p. 138): *"De una baraja de 52 naipes mezclados al azar se sacan dos naipes. Hallar la probabilidad de que ambos sean ases si la primera extraída se devuelve a la baraja".*

3. Calcular una probabilidad condicional en un experimento compuesto, en el contexto de muestreo sin reposición (regla de Laplace), por ejemplo en el siguiente ejercicio (Spiegel, 1991, p. 130). *"Una caja contiene 3 bolas blancas y 2 bolas negras. Sea E1 el suceso "la primera bola extraída es negra" y E2 el suceso "la segunda bola extraída es negra". Las bolas no se devuelven a la caja..."*

4. Calcular una probabilidad condicional a partir de probabilidades conjuntas y simples $(P(A/B) = P(A \cap B)/P(B)$, siempre que $P(B) > 0)$. Un ejemplo es el siguiente problema resuelto de Botella, León y San Martín (1993, p.179):

	Mujer	Varón	Total
Hasta 10 años de experiencia	4	3	7
Más de 10 años de experiencia	2	1	3
Total	6	4	10

"Como puede apreciarse en el grupo total hay 3 de 10 que tienen más de 10 años de experiencia, pero dentro de las mujeres son 2 de 6 las que cumplen esa condición. La probabilidad del suceso tal como la hemos descrito será:

$$\frac{Numerodecasosfavorables}{Numerodecasosposibles} = \frac{2}{6} = 0,33\text{ ''}$$

5. Calcular una probabilidad compuesta haciendo uso de la regla del producto en un caso de sucesos dependientes, por ejemplo: "*En una bolsa se tienen 4 bolas blancas y 2 bolas negras; otra contiene 3 bolas blancas y 5 bolas negras. Si se saca una bola de cada bolsa, hallar la probabilidad de que ambas sean blancas"* (Spiegel, 1991, p. 139).

6. Calcular una probabilidad compuesta haciendo uso de la regla del producto en un caso de sucesos independientes. "*En una clase hay 100 personas. Calcular la probabilidad de que al menos dos tengan el mismo cumpleaños"* (Peña, 1986, p. 77).

7. Determinar si dos sucesos son dependientes o independientes: *"En un grupo de 1000 sujetos se pasó un test de inteligencia y se midió su rendimiento, clasificando a cada sujeto según fuese superior o inferior a la media en inteligencia y según fuese o no apto en la prueba de rendimiento académico. Se encontraron 500 sujetos aptos de los que 300 tenían inteligencia superior. De los 400 con inteligencia superior 300 resultaron aptos. ¿Son los sucesos A: "ser superior a la media en inteligencia" y B "ser apto en rendimiento" independientes?"* (Botella, León y San Martín, 1993, p. 283).

8. Diferenciar entre sucesos mutuamente excluyentes y sucesos independientes. Por ejemplo, "*¿Puede ser independientes dos sucesos mutuamente excluyentes que tienen probabilidad no nula?"* (Peña, 1986, p.77).

9. Calcular la probabilidad total. Todos los problemas relacionados con el teorema de Bayes implica la resolución de un problema de probabilidad total. Además, algunos libros

incluyen explícitamente este tipo de problemas.

10. Resolver problemas bayesianos. Por ejemplo (Bernardo, 1981, p. 77):*"Un test médico diseñado para la diagnosis de una enfermedad detecta dicha enfermedad en el 90% de los pacientes que la padecen y que son sometidos al mismo, pero da también positivo en el 2% de personas que sin padecerla son sometidos al test. Sin embargo, al pasar el test a una población determinada el porcentaje de contestaciones positivas del test correspondientes a personas no afectas es nada menos que el 50%. Explicar este hecho en función de la proporción de personas no afectadas por la enfermedad en la población estudiada".*

11. Resolver problemas de probabilidad condicional y conjunta en experimentos compuestos de diferentes experimentos simples. Por ejemplo:*"En una universidad terminan la carrera el 5% de Arquitectura, el 10% de Ciencias y el 50% de Letras (sólo hay estas carreras). Eligiendo un estudiante al azar se pide: a) Probabilidad de que sea de Arquitectura y haya terminado la carrera; b) Nos dice que ha terminado la carrera. Probabilidad de que sea de Arquitectura"*(Cuadras y cols., 1984, p. 137).

12. Estimar probabilidades condicionales y conjuntas mediante simulación. Estos problemas sólo los hemos encontrado en los dos libros con enfoque basado en el ordenador. Por ejemplo: *"Considera el ejemplo de prueba sanguínea discutido en el capítulo. Supón que haces tres pruebas y son todas positivas. ¿Tienes una alta probabilidad de tener la enfermedad? Usa Minitab en los cálculos"* (Albert y Rossman, 2001, p. 34); siguen una serie de instrucciones.

En la tabla 4.4 se resumen los tipos de problemas encontrados. Estos tipos serán tenidos en cuenta en la elaboración y selección de ítems del cuestionario.

Tabla 4.4. Tipos de problemas encontrados en los documentos analizados

	1. Botella,Leon	2. Amón	3. Glass Stanley	4. Peña Romo	5. Merino	6. Botella Barropiedro	7. Padilla Merino	8. San Martín	9. Peña	10. Wonnacot	11. Cuadras, Echeverria	12. Nortes Checa	13. Spiegel	14. Sacerdote	15. Berry	16. Bernardo	17. Serrano	18. Albtert
Calcular una probabilidad condicional, dentro de un único experimento	x			x						x	x			x		x		
Calcular una probabilidad condicional en un experimento de muestreo con reposición	x		x	x		x		x	x	x		x		x	x			x
Calcular una probabilidad condicional en un experimento de muestreo sin reposición	x		x	x		x		x	x	x		x	x	x	x			
Calcular una probabilidad condicional a partir de probabilidades conjuntas y simples										x					x			x
Calcular una probabilidad compuesta haciendo uso de la regla del producto en un caso de sucesos dependientes		x		x	x	x	x		x		x	x		x				x
Calcular una probabilidad compuesta haciendo uso de la regla del producto en un caso de sucesos independientes	x	x	x	x	x		x		x	x	x	x		x				x
Determinar si dos sucesos son dependientes o independientes	x	x		x			x		x	x			x		x			x
Diferenciar entre sucesos mutuamente excluyentes y sucesos independientes		x							x									
Calcular la probabilidad total		x		x		x			x	x	x			x	x			x
Resolver problemas bayesianos		x		x					x	x	x			x	x	x	x	x
Resolver problemas de probabilidad condicional y conjunta en experimentos compuestos de diferentes experimentos simples	x	x		x	x			x	x	x	x		x	x	x			x
Estimar probabilidades condicionales y conjuntas mediante simulación																	x	x

3.4. RESULTADOS. TABLA DE ESPECIFICACIONES

El análisis realizado fue un primer paso para elaborar la tabla de especificaciones previa a la construcción de nuestro cuestionario, en el que se tuvo en cuenta los anteriores contenidos, tanto a nivel conceptual (tabla 4.3) como procedimental (tabla 4.4). Se identificaron las siguientes conductas observables que son indicadores de la comprensión del concepto:

Conocimiento conceptual

1. Ser capaz de dar una definición correcta de la probabilidad condicional. La definición de probabilidad condicional aparece prácticamente en todos los libros de texto.
2. Reconocer que la probabilidad de $P(B) > 0$ para poder definir $P(B/A)$. Aparece explícitamente en seis de los libros de texto analizados.
3. Reconocer que una probabilidad condicional cumple los axiomas. Se presenta explícitamente en cuatro de los libros de texto analizados.
4. Reconocer que la probabilidad condicional supone una restricción del espacio muestral. Se incluye explícitamente en tres de los libros de texto analizados.
5. Recordar la regla del producto. Aparece prácticamente en todos los libros de texto.
6. Distinguir una probabilidad condicional de su inversa (Eddy, 1982; Falk, 1986; Pollatsek y cols. 1987).
7. Distinguir probabilidad simple, condicional y probabilidad conjunta (Estepa, 1993; Einhorn y Hogarth, 1986; Totohashina, 1992) y se contempla en dos libros de texto.
8. Comprender que una probabilidad conjunta es menor que una probabilidad simple (Tversky y Kahneman, 1982b). Se incluye en dos libros de texto.
9. Reconocer que puede condicionarse sucesos con otros que son posteriores en el tiempo. Encontrado en dos libros; se tiene también en cuenta en Falk, (1986); Gras y Totohasina, (1995).
10. Distinguir una situación condicional de una situación causal o diagnóstica (Falk, 1986; Pollatsek y cols., 1987; Gras y Totohasina, 1995); presentado en tres libros de texto.
11. Distinguir sucesos dependientes, independientes y mutuamente excluyentes (Kelly y Zwiers, 1986; Sánchez, 1996). Se incluye en varios libros como ejercicios.
12. Recordar el teorema de la probabilidad total (aparece en una parte importante de los libros y está implícito en las investigaciones que han propuesto este tipo de problemas).
13. Recordar el teorema de Bayes (aparece en una parte importante de los libros y está implícito en las investigaciones que han propuesto este tipo de problemas).

Conocimiento procedimental

14. Calcular una probabilidad condicional dentro de un único experimento (seis libros).
15. Resolver correctamente problemas sobre probabilidad condicional en un contexto de muestreo con reposición (casi todos los libros y Ojeda, 1995; Maury, 1996).
16. Resolver correctamente problemas sobre probabilidad condicional en un contexto de muestreo con reposición (casi todos los libros y Ojeda, 1995; Maury, 1996).
17. Resolver correctamente problemas sobre probabilidad condicional a partir de probabilidades conjuntas y simples (tres libros).
18. Resolver correctamente problemas condicionales cuando se invierte el eje de tiempo (Falk, 1986; Gras y Totohasina, 1995).
19. Resolver correctamente problemas en situaciones sincrónicas (Falk, 1986; Sánchez, 1996).
20. Resolver correctamente problemas en situaciones diacrónicas (Falk, 1986; Sánchez, 1996).
21. Calcular una probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos independientes (doce libros; Maury, 1986).
22. Calcular una probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos dependientes (diez libros; Maury, 1986).
23. Aplicar correctamente el cálculo de la probabilidad condicional en situaciones de sucesos múltiples (regla de la probabilidad total) (Totohasina, 1992). Aparece a nivel teórico o práctico en muchos libros de texto.
24. Aplicar correctamente el cálculo de la probabilidad condicional en situaciones de probabilidad inversa (regla de Bayes) (Totohasina, 1992). Aparece a nivel teórico o práctico en la mayoría de los libros de texto. La identificación correcta de las tasas bases en problemas de probabilidad Bayes se considera en Tversky y Kahneman (1982).

Además de la descripción cuidadosa de las áreas de contenido que serán evaluadas, se describieron los procesos mentales que serían evaluados[18], teniendo en cuenta en el estudio los cuatro niveles superiores de la taxonomía de Bloom (1956). La mayoría de los

[18] Aunque el análisis del currículum – complementado por el de las investigaciones previas – es importante para establecer el contenido del instrumento, no es suficiente para producir buenos cuestionarios y buenos ítems (Osterlind, 1989).

conocimientos conceptuales se trabajarán a nivel de comprensión y aplicación, puesto que la intención no es simplemente pedir a los estudiantes que recuerden una definición o propiedad, sino que se pedirá algo más que la simple memorización y el material se les presentará en forma diferente a cómo les fue explicado. La intención es presentar ítems en los que el alumno deba usar las abstracciones generales en situaciones concretas en las que tendrá que aplicarla para dar la respuesta correcta.

Por otro lado, en los contenidos procedimentales la finalidad es trabajar a nivel de análisis y síntesis, puesto que se propondrán al alumno situaciones problemáticas en las que deberá descomponer el enunciado del problema para determinar los datos básicos e incógnitas, y establecer una jerarquía de ideas y relaciones entre datos e incógnitas (análisis). Además habrá que idear un plan de solución y llevarlo a cabo poniendo juntos los elementos hasta llegar a una solución que no se veía anteriormente con claridad (síntesis).

Puesto que se trataba de un test orientado a criterio y con la finalidad de incluir en el cuestionario ítems que representen los objetivos claramente delimitados es desarrolló una tabla de especificaciones (tabla 4.5) con tres elementos básicos: Contenido, procesos e importancia de los mismos (Osterlind, 1989). La tabla permite describir de forma completa y fácil la estructura y contenido del constructo que pretendemos medir, indicando los comportamientos que lo evidencian (definición semántica) y las dimensiones en que se agrupan.

En resumen, se siguieron todos los pasos recomendados por Martínez Arias (1995) al constructor de un cuestionario a la hora de especificar el contenido de un cuestionario: Análisis de contenido, revisión de las investigaciones publicadas, observaciones directas, juicios de expertos y análisis de objetivos instruccionales o programas de intervención.

Tabla 4.5. Tabla previa de especificaciones del cuestionario

	Principales áreas del contenido	Comprensión	Aplicación	Análisis	Síntesis
Conocimiento conceptual	1. Definición de la probabilidad condicional	x	x		
	2. Reconocer que la probabilidad de P(B) >0 para poder definir P(B/A)	x	x		
	3. Reconocer que una probabilidad condicional cumple los axiomas.	x	x		
	4. Reconocer que la probabilidad condicional supone una restricción del espacio muestral	x	x		
	5. Recordar la regla del producto	x	x		
	6. Distinguir probabilidad conjunta, condicional y simple	x	x		
	7. Distinguir probabilidad condicional con inversa	x	x		
	8. Probabilidad conjunta menor que probabilidad simple	x	x		
	9. Puede condicionarse con un suceso posterior en el tiempo	x	x		
	10. Distinguir situación condicional, causal y diagnóstica	x	x		
	11. Distinguir sucesos independientes, dependientes y mutuamente excluyentes	x	x		
	12. Recordar el Teorema de la Probabilidad Total	x	x		
	13. Recordar el Teorema de Bayes	x	x		
Conocimiento procedimental	14. Resolver correctamente problemas de probabilidad condicional dentro de un único experimento			x	x
	15. Resolver correctamente problemas de probabilidad condicional en un contexto de muestreo con reposición			x	x
	16. Resolver correctamente problemas de probabilidad condicional en un contexto de muestreo sin reposición			x	x
	17. Resolver correctamente problemas de probabilidad condicional a partir de probabilidades conjuntas y simples			x	x
	18. Resolver correctamente problemas de probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos independientes			x	x
	19. Resolver correctamente problemas de probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos dependientes			x	x
	20. Resolver correctamente problemas condicionales cuando se invierte el eje de tiempo			x	x
	21. Resolver correctamente problemas en situaciones sincrónicas			x	x
	22. Resolver correctamente problemas en situaciones diacrónicas			x	x
	23. Aplicar correctamente el cálculo de la probabilidad condicional en situaciones de sucesos múltiples (regla de la probabilidad total)			x	x
	24. Aplicar correctamente el cálculo de la probabilidad condicional en situaciones de probabilidad inversa (regla de Bayes)			x	x

Tabla 4.6 Tabla definitiva de especificaciones del cuestionario

	Contenido	Comprensión	Aplicación	Análisis	Síntesis
Conocimiento conceptual	1. Definición de la probabilidad condicional	x	x		
	2. Reconocer que la probabilidad de P(A) >0 para poder definir P(B/A)	x			
	3. Reconocer que una probabilidad condicional cumple los axiomas.	x			
	4. Reconocer que la probabilidad condicional supone una restricción del espacio muestral	x	x		
	5. Distinguir probabilidad condicional con inversa	x	x		
	6. Distinguir probabilidad conjunta, condicional y simple	x	x		
	7. Probabilidad conjunta menor que probabilidad simple	x	x		
	8. Distinguir sucesos independientes, dependientes y mutuamente excluyentes	x	x		
Conocimiento procedimental	9. Calcular una probabilidad condicional dentro de un único experimento			x	x
	10. Resolver correctamente problemas de probabilidad condicional en un contexto de muestreo con reposición			x	x
	11. Resolver correctamente problemas de probabilidad condicional en un contexto de muestreo sin reposición			x	x
	12. Resolver correctamente problemas de probabilidad condicional a partir de probabilidades conjuntas y simples			x	x
	13. Resolver correctamente problemas condicionales cuando se invierte el eje de tiempo			x	x
	14. Distinguir situación condicional, causal y diagnóstica			x	x
	15. Resolver correctamente problemas en situaciones diacrónicas			x	x
	16. Resolver correctamente problemas en situaciones sincrónicas			x	x
	17. Resolver correctamente problemas de probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos independientes			x	x
	18. Resolver correctamente problemas de probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos dependientes			x	x
	19. Aplicar correctamente el cálculo de la probabilidad condicional en situaciones de sucesos múltiples (regla de la probabilidad total)			x	x
	20. Aplicar correctamente el cálculo de la probabilidad condicional en situaciones de probabilidad inversa (regla de Bayes)			x	x

Posteriormente se ha simplificado obteniendo la tabla 4.6 final de especificaciones, puesto que algunos contenidos están englobados en otros:

- Recordar la regla del producto (ya que está implícito en el cálculo de probabilidades compuestas en experimentos dependientes e independientes).

- Reconocer que puede condicionarse sucesos con otros que son posteriores en el tiempo (ya que está implícito en la resolución correcta de problemas condicionales cuando se invierte el eje de tiempos).
- Recordar el teorema de la probabilidad total (ya que está implícito en la solución de problemas relacionados con este teorema).
- Recordar el teorema de Bayes (ya que está implícito en la solución de problemas relacionados con este teorema).

3.5. DISCUSIÓN DEL ESTUDIO 1

Este estudio es el primero de una serie que se orienta a la construcción de un cuestionario de evaluación del razonamiento de los estudiantes de psicología sobre la probabilidad condicional (cuestionario RPC).

El análisis de contenido del tema de probabilidad condicional en una muestra de libros de texto de estadística aplicada, recomendados por al menos cuatro Universidades españolas para las asignaturas de análisis de datos permitió un acercamiento a los contenidos principales de la enseñanza del tema en estas Facultades. Se hace notar que la muestra se eligió entre las referencias aportadas por dos tercios de las Universidades lo que le da una gran validez. Asimismo se incluyeron en la muestra algunos libros de orientación bayesiana (hasta totalizar 18 libros en el análisis) para considerar los dos acercamientos posibles a la inferencia y la enseñanza de la probabilidad condicional en los dos enfoques.

El análisis de contenido, realizado de una manera metódica y rigurosa, junto con la presentación de ejemplos en las diferentes categorías de análisis aporta la traducibilidad y comparabilidad del estudio (Goetz y Lecompte, 1988) necesaria en el análisis cualitativo.

Como resultado se obtuvo una primera tabla de especificaciones del contenido, que incluye el conocimiento conceptual y procedimental sobre el tema, y atiende también a la evaluación de sesgos descritos en la investigación sobre la psicología del razonamiento en probabilidad condicional descrita en el Capítulo 1. Se tiene, asimismo en cuenta los niveles de aplicación y análisis y niveles inferiores en la taxonomía de Bloom. Esta tabla de especificaciones será depurada posteriormente mediante juicio de expertos (estudio 2.2) y servirá asimismo para elegir los ítems del cuestionario, en forma que se cubran dichas

especificaciones (estudios 2.1 y 2.2). Todo este proceso es un primer paso en la justificación de una adecuada *validez de contenido* del instrumento.
Resultados del análisis de contenido se han publicado en Batanero y Díaz (2005) y Díaz y Batanero (2005).

4. ESTUDIO 2. CONSTRUCCIÓN DE LA VERSIÓN PILOTO DEL CUESTIONARIO RPC

4.1. INTRODUCCIÓN

Preparada la tabla de especificaciones, se procedió a elaborar un banco inicial de ítems que, en su conjunto, sirviesen para evaluar los contenidos previstos. En esta sección se describe el proceso de elaboración de este banco inicial de ítems y selección posterior de aquellos que finalmente constituirían el cuestionario piloto, proceso que se llevó a cabo siguiendo los dos sistemas sugeridos en Osterlind (1989):

- Se realizó una *valoración empírica*, para lo cual los ítems se administraron a muestras de sujetos, determinando a partir de sus respuestas una serie de indicadores estadísticos de los ítems.
- Se utilizó el *juicio de expertos*, para valorar las unidades de contenido previamente definidas y su grado de acuerdo sobre si cada uno de los ítems cubría el contenido para el cual había sido diseñado. Para ello se identificaron las personas que habrían de colaborar en esta tarea, se preparó un cuestionario de valoración de contenidos e ítems que se envió a dichos expertos, y se analizaron las respuestas obtenidas.

A continuación se presentan cada uno de estos estudios y sus resultados.

4.2. ESTUDIO 2.1. ELABORACIÓN DE UN BANCO INICIAL DE ÍTEMS Y ENSAYOS DE ÍTEMS

En este estudio se aborda la elaboración de un banco inicial de ítems que serviría para evaluar el contenido semántico de la variable objeto de medición definida en el Estudio 1. Se procuró que fuese lo suficientemente amplio para permitir una selección de entre ellos, mediante un proceso objetivo, de aquellos que finalmente constituirían el cuestionario RPC. Asimismo se abordan las pruebas empíricas de estos ítems en muestras de estudiantes de psicología similares a aquellos a los que iba destinado el instrumento.

En lo que sigue se describe las muestras participantes en las pruebas, el método seguido en la elaboración de los ítems iniciales, y resultados obtenidos de las pruebas.

4.2.1. MÉTODO

Para formar un banco de ítems útiles en la construcción del cuestionario se revisaron diversas investigaciones sobre la probabilidad condicional y se recopilaron los ítems utilizados en las mismas, junto con información sobre el tipo de estudiantes que tomó parte en su evaluación y sobre las características psicométricas, en caso de que la información estuviese disponible[19]. La revisión bibliográfica siguió el método descrito en el estudio 1.

Cada uno de los ítems usados en la investigación previa se analizó, para determinar los contenidos evaluados, comparándolos con la tabla de especificaciones de la variable presentada como resultado del estudio 1, eligiendo sólo aquellos ítems relacionados con alguna de las especificaciones de dicha tabla para formar nuestro conjunto inicial. Mediante un proceso iterativo[20] se comparó el conjunto de ítems seleccionado con la tabla de especificaciones, completando el conjunto progresivamente, según encontrábamos otras investigaciones cuyas tareas pudiesen utilizarse, directamente, o con pequeñas modificaciones de formato o redacción.

También se tuvo en cuenta los razonamientos erróneos descritos en la investigación, como, por ejemplo, confusión entre condicionamiento y causación (Falk, 1979, 1989), falacia del eje de tiempo (Totohasina, 1992; Gras y Totohasina, 1995) o falacias de tasas base (Bar-Hillel, 1983; 1987; Koehler, 1996) para la elaboración de distractores en los ítems de opciones múltiples. Si en alguna de las especificaciones no fue posible encontrar ítems o tareas adecuadas en las investigaciones previas, se tomaron ejercicios propuestos en los libros de texto examinados en el Estudio 1. Finalmente se elaboraron algunos

[19] Esta es una estrategia común al adaptar cuestionarios a poblaciones diferentes de aquellas en que originalmente se usaron los ítems (López Feal, 1986).

[20] Como todo análisis cualitativo, fue un proceso iterativo, en lugar de efectuarse al final de la investigación (Huberman y Miles, 1994).

nuevos ítems por no haber encontrado ninguno adecuado con los dos procedimientos anteriores. El objetivo de este proceso iterativo fue representar adecuadamente las especificaciones de contenido.

El banco de ítems inicial, obtenido con el proceso anterior, se modificó, progresivamente, mejorando la redacción en los ítems, teniendo en cuenta las siguientes recomendaciones (APA, AERA y NCME, 1985, 1999; Osterlind, 1989):

- Se separó con claridad el núcleo del ítem y las respuestas alternativas.
- Se redactó el núcleo en forma interrogativa.
- Se cuidó la precisión y claridad en el lenguaje.

Finalmente se llegó a un conjunto de 49 ítems, que cubrían la tabla de especificaciones del contenido, en forma que hubiese al menos dos ítems para cada unidad de contenido. La decisión final de cuáles ítems se incluirían en el cuestionario, se tomó a partir de algunas pruebas sobre su dificultad y legibilidad en los dos niveles de análisis sugeridos por Thorndike (1989).

1. Un primer nivel de análisis, se basó en la administración en condiciones similares a las que se utilizarían en la forma final del cuestionario, para obtener información valiosa tanto de los ítems, como de dificultades, legibilidad y ajuste al tiempo previsto. Observando el comportamiento de los alumnos mientras responden los ítems (dudas, exceso de respuestas en blanco...) se extrajeron indicadores de confusión en algunos ítems y se mejoró su redacción.
2. En un segundo nivel de análisis se sometieron a prueba todos los ítems, obteniendo sus índices de dificultad, y frecuencias de respuestas a los diferentes distractores. Los ítems fueron divididos en cuatro cuestionarios, con objeto de que el número total de ítems a completar en una sesión por un mismo grupo de alumnos no fuera excesivamente largo, de modo que tuviesen tiempo suficiente para responder. En el Anexo 2 se incluyen los resultados detallados.

4.2.1.1. SUJETOS

Al tratarse de un cuestionario de habilidad y ejecución, las características más importantes de la población fueron sus capacidades cognitivas y conocimientos. Puesto que los estudiantes que acceden a Psicología no siguen un itinerario específico en secundaria, su formación matemática previa es dispar. Debieran haber estudiado durante la secundaria el tema de la probabilidad condicional, aunque algunos alumnos tienen una formación estadística más completa que otros. Su nivel de comprensión lectora es alto. El cuestionario sería utilizado una vez que los alumnos hubiesen estudiado el tema de la probabilidad condicional en la asignatura de análisis de datos.

Los grupos utilizados pueden considerarse muestras representativas –aunque intencionales (Ghiglione y Matalón, 1991) – del grupo objeto de estudio en la investigación final. Participaron en las pruebas pre – piloto dos grupos diferentes de alumnos, de primer año en la Licenciatura de Psicología (en total 157 alumnos, con una nota media de acceso en selectividad de 7,49). Los porcentajes de estudiantes con diferentes tipos de bachillerato fueron los siguientes: Ciencias 27,6%, humanidades 72,4%. Se trata de sujetos voluntarios, como es frecuente en investigaciones en ciencias sociales (Hernández, Fernández y Baptista, 1998).

Cada uno de estos grupos completó el cuestionario en dos sesiones de clase. Se dispuso de dos versiones que fueron alternadas en cada grupo, de modo que la mitad de alumnos completó uno de los cuestionarios y la otra mitad el otro (entregados aleatoriamente), con la finalidad de que no se copiasen y de poder probar simultáneamente el mayor número de ítems.

En ambas pasaciones se pidió a los alumnos que colaboraran con nuestro trabajo y se les animó a tratar de responder todas las preguntas. Los alumnos colaboraron con interés y el tiempo fue suficiente.

4.2.1.2. MATERIAL

Al redactar el grupo inicial de ítems se tuvo en cuenta tanto los objetivos de la evaluación y la población a la que se iba a administrar el cuestionario, como las limitaciones de tiempo disponible (Fox, 1981; Barbero, 2003). En lo que sigue describimos estos requisitos y la forma en que se han tenido en cuenta (Millman y Greene, 1989; AERA, APA y NCME, 1999):

- *Tipo de administración*: por razones de eficiencia el cuestionario sería administrado de forma grupal. Para que la administración fuese eficaz, el cuestionario incluiría instrucciones claras y fáciles de seguir[21].
- *Limitaciones temporales en la administració*n: Se hizo un balance entre el número suficiente de ítems (y su dificultad) para cubrir el contenido y el tiempo disponible para cumplimentarlo (contando además con el posible cansancio de los estudiantes que respondan al cuestionario). En este caso, aún siendo posible utilizar más tiempo, no se consideró oportuno excederse demasiado de una hora de administración. Se limitó el número de ítems y se proporcionó un tiempo de hora y media para evitar que algunos ítems quedasen sin contestar por falta de tiempo.

El número final de ítems en el cuestionario[22] habría de cubrir adecuadamente el contenido y asegurar una fiabilidad satisfactoria (Millan y Greene, 1989), teniendo en cuenta la restricción de su posible longitud. Consecuentemente, se formó un conjunto de ítems inicial suficiente para que después de todas las pruebas, quedase un número de ítems que permitiese completar el cuestionario. Puesto que se dio un peso igual a los diferentes contenidos, el objetivo fue elaborar un cuestionario con al menos dos ítems por contenido. Teniendo en cuenta que algunos ítems podrían evaluar más de un contenido, se trató de llegar a un cuestionario de entre 20 y 25 ítems, que pudiese ser completado en el espacio aproximado de una hora y media. El número de ítems para las pruebas piloto habría de ser mayor que el definitivo.

En consecuencia, el material se compuso de un total de 49 ítems que en su conjunto evaluaban las unidades de contenido de la definición semántica de la variable definidas en el Estudio 1 (2-3 ítems para cada unidad de contenido). En la tabla 4.7 presentamos los ítems utilizados para cada unidad de contenido.

[21] Dado que la interacción entre el administrador del cuestionario y las personas que responden a él es menor que en administraciones individuales.

[22] Para elegir el número de ítems debemos tener en cuenta la amplitud del contenido. Su división o no en componentes, la estructura homogénea o heterogénea del contenido, la forma de aplicación y características de los sujetos (López Feal, 1986).

Tabla 4.7. Contenidos e Ítems sometidos a prueba

Contenido 1: Definición de la probabilidad condicional.

Ítem 1. Explica con tus propias palabras la diferencia entre una probabilidad simple y una probabilidad condicional.

Ítem 2. ¿Qué quiere decir la expresión "la probabilidad condicional de *A* dado *B* es 1/4"?
a) En la cuarta parte de los experimentos obtenemos *A* y *B* simultáneamente
b) *A* ocurre la cuarta parte de las veces en que ocurre *B*
c) *B* ocurre la cuarta parte de las veces en que ocurre *A*
d) *A* o *B* ocurren la cuarta parte de las veces

Contenido 2: Reconocer que *P(B)* >0 para poder definir *P(B/A)*

Ítem 3. ¿Cuál de las siguientes condiciones es necesaria para poder definir la probabilidad condicional *P(A/B)*?
a) *P(A)* > 0
b) *P(B)* > 0
c) *A* y *B* sean independientes
d) No hace falta ninguna de estas condiciones

Ítem 4. Para poder definir la probabilidad condicional *P(A/B)* es necesario que *P(B)* > 0
Verdadero___ Falso ___

Contenido 3: Reconocer que una probabilidad condicional cumple los axiomas de probabilidad

Ítem 5. Sea *A* el suceso "Pedro instala una alarma antirrobo en su casa". Sea *B* el suceso "La casa de Pedro es robada antes de fin de año". Sean $\overline{A}$ y $\overline{B}$ las negaciones de *A* y *B* respectivamente. ¿Cuál de las siguientes afirmaciones es correcta?
a) $\boldsymbol{P(A/B) + P(A/\overline{B}) = P(A)}$
b) $P(A \cap B) > P(A)$
c) $P(A/B) = P(A) \times P(B)$
d) $P(A/B) = P(A \cap B) / P(A)$

Ítem 6. Sea *A* el suceso "Pedro instala una alarma antirrobo en su casa". Sea *B* el suceso "La casa de Pedro es robada antes de fin de año". Sean $\overline{A}$ y $\overline{B}$ las negaciones de *A* y *B* respectivamente. ¿Cuál de las siguientes afirmaciones es incorrecta?
a) $P(A/B) \leq 1$ $P(A/B) + P(A/\overline{B}) = P(A)$
b) $P(A/A) = 1$
c) $\boldsymbol{P(A/B) = P(A \cap B) / P(A)}$

Contenido 4: Reconocer que la probabilidad condicional supone una restricción del espacio muestral

Ítem 7. ¿Cuál es el espacio muestral de una probabilidad condicionada *P(A/B)*?
a) El mismo que el de la probabilidad simple
b) El suceso condicionante
c) El suceso condicionado
d) La intersección de los dos sucesos

Ítem 8. El espacio muestral de una probabilidad condicionada *P(A/B)* es el suceso condicionante *B*
Verdadero___ Falso _____

Contenido 5: Distinguir una probabilidad condicional y su inversa: Diferenciar *P(A/B)* y *P(B/A)*

Ítem 9. En un centro médico se han observado a 250 personas para estudiar si el hábito de fumar tiene relación con los trastornos bronquiales, obteniéndose los siguientes resultados:

	Fuma	No Fuma	Total
Padece trastornos	90	60	150
No padece trastornos	60	40	100
Total	150	100	250

Si escogemos al azar una de estas personas:
. ¿Cuál es la probabilidad de que fume si sabemos que tiene trastornos? ___________
. ¿Cuál es la probabilidad de que tenga trastornos, si sabemos que fuma? ___________

Ítem 10. En un barrio de Granada se ha realizado una entrevista a un grupo de hombres, obteniendo los siguientes resultados:

	menos de 55 años	más de 55 años	Total
Ha sufrido un ataque al corazón	29	75	104
No ha sufrido ataque	401	275	676
Total	430	350	780

Si elegimos al azar una de estas personas:
a) Dado que la persona escogida tiene más de 55 años, ¿cuál es la probabilidad de que haya tenido un ataque al corazón?
b) Dado que la persona escogida ha tenido un ataque al corazón, ¿cuál es la probabilidad de que tenga más de 55 años?

Ítem 11. ¿En cuál de las siguientes predicciones tendrás mayor seguridad o confianza?
a) Una persona adicta a la heroína también fuma marihuana
b) Una persona que fuma marihuana es también adicta a la heroína
c) Los dos sucesos son iguales de probables

Contenido 6: Distinguir probabilidad conjunta, condicional y simple

Ítem 12. En un centro médico se han observado a 250 personas para estudiar si el hábito de fumar tiene relación con los trastornos bronquiales, obteniéndose los siguientes resultados:

	Fuma	No Fuma	Total
Padece trastornos	90	60	150
No padece trastornos	60	40	100
Total	150	100	250

Si escogemos al azar una de estas personas:
I. ¿Cuál es la probabilidad de tener trastornos bronquiales?
II. ¿Cuál es la probabilidad de que tenga trastornos y fume?
III. ¿Cuál es la probabilidad de que fume si sabemos que tiene trastornos?

Ítem 13. En un barrio de Granada se ha realizado una entrevista a un grupo de hombres, obteniendo los siguientes resultados:

	menos de 55 años	más de 55 años	Total
Ha sufrido un ataque al corazón	29	75	104
No ha sufrido ataque	401	275	676
Total	430	350	780

Si elegimos al azar una de estas personas:
a) ¿Cuál es la probabilidad de que tenga más de 55 años y además haya tenido uno o más ataques al corazón?
b) ¿Cuál es la probabilidad de que haya tenido un ataque al corazón?
c) Dado que la persona escogida tiene más de 55 años, ¿cuál es la probabilidad de que haya tenido un ataque al corazón?

Contenido 7: Reconocer que la probabilidad conjunta *P(A y B)* ha de ser menor o igual que la probabilidad simple *P(A)*

Ítem 14. Supón que Carlos Ferrero alcanza la final de Roland Garros en 2004. ¿Cuál de los siguientes sucesos consideras más probable?
a) Carlos Ferrero pierde el primer set
b) Carlos Ferrero pierde el primer set pero gana el partido
c) Los dos sucesos son iguales de probables

Ítem 15. ¿Cuál de los siguientes sucesos consideras más probable?
a) Aznar mandará más tropas a Irak y aumentará el presupuesto de becas
b) Aznar aumentará el presupuesto de becas.
c) Los dos sucesos son igual de probables

Contenido 8: Distinguir sucesos independientes, dependientes y mutuamente excluyentes

Ítem 16. Se extrae una carta al azar de una baraja americana: sea *A* el suceso "se extrae un trébol" y *B* el suceso "se extrae una reina" ¿Los sucesos *A* y *B* son independientes?
a) Sí, en todos los casos.
b) No son independientes porque en la baraja hay una reina de tréboles.
c) Sólo si sacamos primero una carta para ver si es reina y se vuelve a colocar en la baraja y luego sacamos una segunda carta y para mirar si es trébol.
d) No, porque P(reina de trébol)= P(reina) x P(trébol)

Ítem 17. ¿Qué quiere decir el término "mutuamente excluyentes" cuando se usa en referencia a la probabilidad?

a) La ocurrencia de un suceso no afecta la ocurrencia del otro suceso
b) La probabilidad de la intersección es el producto de las probabilidades simples
c) Si un suceso ocurre, el otro no puede ocurrir
d) A y B son correctas

Ítem 18. En un grupo de mil sujetos se pasó un test de inteligencia y se midió su rendimiento, clasificando a cada sujeto según fuese superior o inferior a la media en inteligencia, y según fuese apto o no apto en la prueba de rendimiento académico. Se encontraron 500 aptos, de los que 300 tenían inteligencia superior. De los 400 con inteligencia superior, 300 resultaron aptos. Extraemos un sujeto al azar y definimos los sucesos, *A*: "Ser superior a la media en inteligencia", y *B*: "Ser apto en rendimiento". ¿Son los sucesos *A* y *B* independientes?

Contenido 9: Calcular una probabilidad condicional dentro de un único experimento

Ítem 19. Cogemos un número al azar menor que 1000. Sabiendo que es múltiplo de cinco. ¿Cuál es la probabilidad de que sea múltiplo de cuatro?

a) 1/20 **b) 1/4** c) 1/5 d) 4/5

Ítem 20. Hemos lanzado dos dados y sabemos que el producto de los dos números obtenidos ha sido 12 ¿Cual es la probabilidad de que ninguno de los dos números sea un 6?

Contenido 10: Calcular la probabilidad condicional en un contexto de muestreo con reposición

Ítem 21. Una persona lanza un dado y anota si saca un número par o impar. Estos son los resultados:
par, impar, impar, par, par, impar, par, par, par, par, impar, impar, par, par, par Lanza el dado de nuevo
¿Cuál es la probabilidad de sacar un número par en la siguiente tirada?

Ítem 22. Una persona lanza 3 veces la misma moneda, obteniendo en este orden los resultados siguientes: cara, cruz, cruz. Lanza la moneda una cuarta vez. ¿Qué resultado piensas que es más probable que salga a continuación?

a) Cara b) Cruz **c) Los dos sucesos son iguales de probables**.

Contenido 11: Calcular la probabilidad condicional en un contexto de muestreo sin reposición

Ítem 23. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin reemplazamiento.

- ¿Cuál es la probabilidad de extraer una bola negra en segundo lugar, habiendo extraído una bola negra en primer lugar? $P\ (N_2/\ N_1)$

a) **1/ 3** b) 1/ 2 c) 1/ 6 d) 1/ 4

Ítem 24. Una caja tiene cuatro focos, de los cuales dos son defectuosos. Se sacan dos al azar, uno tras otro sin reemplazamiento. Si el primer foco fue defectuoso, ¿qué es más probable?

a) que el segundo sea defectuoso
b) que el segundo no sea defectuoso
c) los dos sucesos son iguales de probables

Contenido 12: Calcular una probabilidad condicional a partir de probabilidades conjuntas y simples: Aplicando directamente la definición $P(A/B) = P(A \cap B)/P(B)$

Ítem 25. La probabilidad de que una mujer de 40 años tenga una mamografía positiva es el 10.3 %. La probabilidad de que una mujer de 40 años tenga cáncer de pecho y una mamografía positiva es de 0.8%. Una mujer de 40 años se realiza una mamografía y resulta positiva. ¿Cuál es la probabilidad de que realmente tenga cáncer?

a**) 0,8/10,3 = 0,0776, probabilidad de 7'76%**
b) 10,3 x 0,8 = 8,24, probabilidad del 8'24%
c) 0,8 %

Ítem 26. 103 de cada 1000 mujeres de 40 años tienen una mamografía positiva. 8 de cada 1000 mujeres de 40 años tienen cáncer de pecho y obtienen una mamografía positiva. En una muestra representativa de mujeres de 40 años que se realizan una mamografía y resulta positiva. ¿Cuántas de ellas esperas que realmente tengan cáncer?

a) 7,7 de cada 100 b) 8,24 de cada 100 c) 0,8 de cada 100

Ítem 27. En un país el 30 % de las personas tiene los ojos claros y el 15 % tiene los ojos claros y el pelo rubio. Si en este país cogemos una persona rubia al azar, ¿Cuál es la probabilidad de que tenga los ojos oscuros?

Contenido 13: Resolver correctamente problemas condicionales cuando se invierte el eje de tiempo

Ítem 28. ¿Cuál de los siguientes sucesos es más probable?

a) Que una niña tenga los ojos azules si su madre tiene los ojos azules
b) Que una madre tenga los ojos azules si su hija tiene los ojos azules
c) Los dos sucesos son iguales de probables
d) La probabilidad de que la madre tenga los ojos azules dado que su hija tiene los ojos azules no puede calcularse ya que es anterior en el tiempo.

Ítem 29. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin reemplazamiento.

- ¿Cuál es la probabilidad de extraer una bola negra en primer lugar, sabiendo que hemos extraído una bola negra en segundo lugar? $P(N_1/N_2)$

a) 1/ 3 b) No se puede calcular c) 1/ 6 d) 1/ 2

Contenido 14: Distinguir situación condicional, causal y diagnóstica. El suceso condicionante puede o no ser causa del suceso condicionado

Ítem 30. ¿Cuál de los siguientes sucesos es más probable?

a) Un divorcio es resultado de un matrimonio infeliz
b) Un matrimonio infeliz lleva a un divorcio
c) Los dos enunciados son iguales de probables.

Ítem 31. ¿En cuál de las siguientes predicciones tendrás mayor seguridad?

a) Predecir que una persona que está enferma, tiene fiebre.
b) **Predecir que una persona que tiene fiebre está enferma.**
c) Tendría igual seguridad en ambas predicciones.

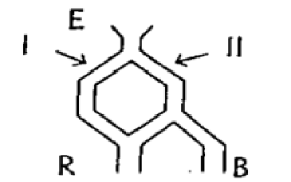
Ítem 32. Un test diagnóstico de cáncer fue administrado a todos los residentes en una gran ciudad. Un resultado positivo en el test es indicativo de cáncer y un resultado negativo es indicativo de ausencia de cáncer. ¿En cuál de las siguientes predicciones tienes más confianza?

a) Predecir que una persona tiene cáncer si ha dado positivo en el test de diagnóstico.
b) Predecir un resultado positivo en el test de diagnóstico si la persona tiene cáncer.
c) Tengo la misma confianza en ambas predicciones.

Contenido 15: Resolver correctamente problemas en situaciones diacrónicas (experimentos simples consecutivos)

Ítem 33. Una bola se suelta en *E*. Si sale por *R*, ¿Cuál es la probabilidad de que haya pasado por el canal 1?

a) 0,50
b) 0,66
c) 0,33
d) No se puede calcular



Ítem 34. Una caja tiene cuatro focos, de los cuales dos son defectuosos. Se sacan dos al azar, uno tras otro sin reemplazamiento.

- Si el segundo foco es defectuoso, ¿qué es lo más probable?

a) que el primero sea defectuoso
b) que el primero no sea defectuoso
c) los dos sucesos son iguales de probables
d) No se puede saber

Contenido 16: Resolver correctamente problemas en situaciones sincrónicas (experimentos simples simultáneos)

Ítem 35. En una urna hay tres tarjetas: una es azul por ambos lados, otra es roja por ambos lados y la última es por un lado azul y por otro roja. Seleccionamos una tarjeta al azar y la ponemos sobre la mesa tal y como la sacamos de la urna. La cara que vemos es roja, ¿cuál es la probabilidad de que la otra cara sea roja también?

a) **2/ 3 b)** 1/ 2 c) 3/4 d) 1/3 x 1/2

Ítem 36. En un grupo de mil sujetos se pasó un test de inteligencia y se midió su rendimiento, clasificando a cada sujeto según fuese superior o inferior a la media en inteligencia, y según fuese o no reflexivo. Se encontraron 500 sujetos reflexivos, de los que 300 tenían inteligencia superior. ¿Cuál es la probabilidad de que un sujeto reflexivo sea de inteligencia superior?

Ítem 37. En una encuesta publicada en un periódico se informa que el 91% de los habitantes de una ciudad mienten usualmente y de ellos el 36% mienten sobre cosas importantes. Si nos encontramos con una persona al azar de esta ciudad, ¿cuál es la probabilidad de que mienta sobre cosas importantes?

Contenido 17: Calcular una probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos independientes

Ítem 38. Un grupo de alumnos de un colegio se examina de Matemáticas e Inglés. La proporción de alumnos que aprueban Matemáticas es del 80% y la proporción de alumnos que aprueba Inglés es del 70%. Suponiendo que las notas en cada una de las asignaturas es independiente, ¿Cuál es la probabilidad de que un alumno escogido al azar haya aprobado ambas asignaturas?

Ítem 39 Se lanza cuatro veces un dado. ¿Cuál es la probabilidad de no obtener ningún uno?

Ítem 40. En una secuencia de tres dígitos, cada uno de los dígitos es seleccionado al azar del conjunto {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}. ¿Cuál es la probabilidad de que de este proceso resulten tres números diferentes?

Contenido 18: Calcular una probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos dependientes

Ítem 41. Se tiene una bolsa con 1 bola azul y 2 bolas rojas. Se saca una bola al azar. Sin reponerla otra vez en la bolsa se saca una segunda bola. ¿Cuál suceso es más probable?

a) Sacar dos bolas rojas
b) Sacar primero una bola roja y luego una azul.
c) Los dos sucesos son iguales de probables
d) Sacar primero una bola azul y luego una roja

Ítem 42. Un hombre, por accidente, mezcla dos baterías gastadas con tres baterías nuevas del mismo modelo. Con ayuda de un voltímetro va probando una batería tras otra hasta encontrar las dos baterías gastadas. ¿Cuál es la probabilidad de que tenga que probar sólo dos veces para encontrar las baterías gastadas?

Ítem 43. De cada 1000 niños varones nacidos vivos 746 llegan a cumplir 65 años. La probabilidad de que una persona que acaba de cumplir 65 muera en los cinco años siguientes es 0'160. ¿Cuál es la probabilidad de que un niño varón alcance su 70 cumpleaños?

Contenido 19: Aplicar correctamente la regla de la probabilidad total en un problema

Ítem 44. En una escuela hay en marcha dos programas educativos, A y B. El 60% de los estudiantes están en el programa A y el 40% en el programa B. En el programa A, el 50% son chicos y en el programa B, el 25% son chicos. Encontramos al azar un estudiante ¿cuál es la probabilidad de que sea chico?

Ítem 45. En una ciudad hay 60 hombres y 40 mujeres por cada 100 habitantes. 50 de cada 100 hombres y 25 de cada 100 mujeres fuman. Si cogemos 200 personas al azar, ¿cuántas de ellas fumarán?

Ítem 46. Al azar se seleccionan dos de cuatro cartas numeradas 1, 2, 3 y 4, una después de otra. Los números de las cartas se suman y si la suma es par, se gana un premio. Qué es más probable, ¿obtener el premio o no?

Contenido 20: Aplicar correctamente el teorema de Bayes en un problema

Ítem 47. Un taxi se vio implicado en un accidente nocturno con choque y huida posterior. Hay dos compañías de taxis en la ciudad, la Verde y la Azul. El 85% de los taxis de la ciudad son Verde y el 15% Azul. Un testigo identificó el taxi como Azul. El tribunal comprobó la fiabilidad del testigo bajo las mismas circunstancias que había la noche del accidente y llegó a la conclusión de que el testigo identificaba correctamente cada uno de los colores en el 80% de las ocasiones y fallaba en el 20%. ¿Cuál es la probabilidad de que el taxi implicado en el accidente fuera en efecto Azul?

a) 80 % **b) 48 %** c) 15% d) (15/100) x (80/100)

Ítem 48. Una fábrica dispone de dos máquinas M1 y M2 que fabrican bolas. La máquina M1 fabrica el 40 % de las bolas y la M2 el 60%. El 5% de las bolas fabricadas por M1 y el 1% de las fabricadas por M2 son defectuosas. Tomamos una bola al azar que resulta ser defectuosa. ¿Cuál es la probabilidad de que haya sido fabricada por M1?

Ítem 49. Se sabe que 100 de cada 10000 personas está enferma. De cada 100 personas enfermas, el test será positivo en 80 de ellas. De cada 9900 personas sanas el test saldrá positivo en 950 de ellas. El test resulta positivo en 103 personas. ¿Cuantas de estas personas están realmente enfermas?

Cada uno de los ítems se consideró como una unidad de medida, compuesto de un estímulo y una forma prescriptiva de respuesta, cuya finalidad era inferir la capacidad del examinado en el constructo, proporcionando datos cuantificables sobre la persona que lo completa (Osterlind, 1989). En la selección de ítems se siguieron los siguientes criterios:

- Alto grado de congruencia entre el ítem particular y el contenido que se pretende evaluar, que es un aspecto clave para poder lograr la validez del cuestionario. Se aseguraría en el estudio 2.2 mediante la técnica de juicio de expertos.
- Minimizar en lo posible la contribución de cada ítem al error de medida de la puntuación en el test. Para ello se harían pruebas de los ítems con objeto de revisar los que fuesen inadecuados en las pruebas (por ejemplo, pasando de opciones múltiples a respuesta abierta) y descartando otros en los que se encontrase excesiva dificultad. Posteriormente se estudiaría la contribución de cada ítem al error de medida, como parte del estudio de la fiabilidad del instrumento.
- Los ítems cumplirían los criterios de independencia local (la repuesta a uno no depende de la respuesta a otro) y unidimensionalidad (Muñiz, 1994), pues todos se refieren al constructo razonamiento sobre probabilidad condicional.

Formato de los ítems

"Las decisiones sobre el tipo de ítem a usar se deben apoyar en la conceptualización de los componentes específicos del dominio de rendimiento o habilidad, tal como se refleja en las especificaciones del contenido. Estas especificaciones podrían sugerir más de un tipo particular de ítem" (Millman y Greene, 1989, p. 143)[23]. En consecuencia, se determinaron:

- Los elementos estructurales de los ítems necesarios para elicitar las habilidades cognitivas y procesos identificados en la tabla de especificaciones del test.
- El tipo de ítems que incluyen estos elementos estructurales.
- De entre esos distintos tipos de ítems, se eligió el formato con base en la economía, precisión y ajuste a la población (Millman y Greene, 1989).

[23] Mientras algunos cuestionarios contienen sólo ítems abiertos o cerrados, otros contienen ítems de los dos tipos (Hernández, Fernández y Baptista, 1998).

Se utilizaron dos tipos de formato en el cuestionario[24]. Dado que nos encontramos ante un test de ejecución óptima (cuestionario de rendimiento) una parte de los ítems se presentaron en formato de *elección múltiple*[25] que permitían asignar fácilmente puntuaciones a los sujetos e identificar patrones de respuesta erróneos, puesto que en las investigaciones previas donde se tomaron dichos ítems ha descrito detalladamente los razonamientos de los estudiantes. Esto permitió predecir las dificultades y errores más frecuentes y tenerlas en cuenta en distintos distractores. Por otro lado, este formato de ítem era muy familiar para la población de estudiantes que respondería al cuestionario, que suelen ser evaluados con cuestionarios formados por ítems de elección múltiple[26].

El número usual de respuestas en estos ítems de opciones múltiples se varió entre tres y cuatro, prefiriendo un mayor número de ítems con menos alternativas para aumentar la fiabilidad (Martínez Arias, 1995). Se decidió no forzar una alternativa más, cuando esta no fuese natural, ni emplear respuestas tales como "ninguno de los anteriores" o "todos los anteriores" (Osterlind, 1989; AERA, APA y NCME, 1999).

También se incluyeron algunos *ítems de respuesta construida por el alumno* y más particularmente problemas simples sobre probabilidad condicional. Estos ítems permitirían un acceso no sólo a la solución final, sino a los procedimientos empleados en su solución, y dar un criterio de calificación parcial (y no simplemente dicotómico), es decir dar una puntuación a la respuesta basada en su grado de completitud y precisión. El *crédito parcial* es deseable en muchas situaciones de evaluación en las que se desea conocer el conocimiento parcial de los examinados. Su principal ventaja es que puede contribuir a aumentar la fiabilidad de un instrumento de longitud restringida (Millman y Greene, 1989).

Nivel de dificultad

Al tratarse de un cuestionario de referencia a un dominio, no fue necesario especificar por adelantado el nivel de dificultad de los ítems, ya que el contenido está muy concretamente especificado por el dominio. Interesaba, por tanto, incluir ítems concretos,

---

[24] "Algunos instrumentos pueden contener varios formatos diferentes. Un ejemplo de cuestionario que usa una variedad de formatos en el Alabama English Proficiency Test" (Osterlind, 1989, p. 29).

[25] Este tipo de ítem se incluye los que Osterlind (1989) denomina como de respuesta construida, donde se proporciona al examinado la respuesta correcta, junto con otras alternativas y se pide al examinado elegir una de las opciones.

[26] Además este tipo de ítems se puede adaptar a una gran variedad de objetivos educativos, permite una interpretación precisa de la respuesta que puede constituir evidencia para la validez de contenido del cuestionario y su formato es flexible, aunque sofisticado, y sencillo de usar (Osterlind, 1989).

incluso aunque fuesen difíciles, puesto que se quería evaluar la dificultad real de conceptos muy determinados. En este estudio la dificultad del ítem se consideró una descripción de la capacidad del examinado, más que una característica prescrita del ítem (Millman y Greene, 1989).

4.2.2. ANÁLISIS

Pasados los cuestionarios, los datos se codificaron y analizaron usando el paquete estadístico SPSS. El análisis se redujo a tablas de frecuencias y porcentajes de respuestas en las diferentes categorías, no presentando estudio de puntuaciones globales o análisis de fiabilidad. La finalidad de esta recogida de datos fue tener un criterio complementario (además del juicio de los expertos) para seleccionar nuestros ítems, de forma que abarquen el contenido, tengan una dificultad variada y sirvan para detectar los sesgos y errores descritos en la investigación.

Además de usar ítems de opciones múltiples, se incluyeron algunos ítems de respuesta abierta con la finalidad de aplicar crédito parcial a la hora de calificarlos. Es decir, hemos otorgado diferente puntuación en el mismo ítem a los sujetos según su grado de corrección, dando una puntuación menor a la solución incompleta o parcial (pero correcta) de dichas tareas. En los casos en que se ha tomado esta decisión, se describe en el anexo 2 los criterios seguidos para puntuar los diferentes ítems. Algunos ítems se incluyeron en los dos cuestionarios A y B, por lo que la muestra resultante es mayor que para el resto de los ítems.

Puesto que el índice de dificultad es la proporción de alumnos que superan el ítem se calcularon adicionalmente los intervalos de confianza de dicha proporción utilizando la aproximación normal, ya que las dos muestras contaban con más de 30 elementos. En los ítems de respuestas abiertas además de la respuesta totalmente correcta (máxima puntuación), también se consideraron correctas (a efecto del cálculo del índice de dificultad) las respuestas con razonamiento e identificación de los datos correctos y ligeros errores de cálculo.

Una aportación metodológica es el uso de inferencia bayesiana para la estimación de los índices de dificultad. Esta estimación se ha realizado con distribución a priori no informativa, puesto que no se disponía de información previa sobre la dificultad de los ítems para los alumnos de psicología (Lecoutre, 1996; Bernard, 1998). El intervalo de credibilidad se calcula mediante la siguiente expresión:

$$\left[\beta^{-1}_{0,5+e,\,0,5+f}(\alpha/2) - \beta^{-1}_{0,5+e,\,0,5+f}(1-\alpha/2)\right]$$

donde es la $\beta^{-1}_{a,b}$ función de distribución inversa de la distribución *Be(a,b), e y f* los éxitos y fracasos en la muestra y α el coeficiente de credibilidad.

Asimismo, se usó la metodología bayesiana para estimar la media de la distribución de índices de dificultad, supuesta la normalidad de dicha distribución, que es contrastada mediante la prueba de Kolmogorov Smirnov. Puesto que no se conoce σ (desviación típica poblacional) se usa *s,* la estimación insesgada de la desviación típica (raíz de la cuasivarianza muestral). En el caso de distribución inicial no informativa (pues no se tiene información previa, la variable tipiticada $T = \frac{\mu_f - \bar{x}}{s/\sqrt{n}}$ sigue una distribución *T*[27] con *n-1* grados de libertad, siendo *n* el tamaño de la muestra de datos (Bolstad, 2004).

En consecuencia, se calcula el intervalo de credibilidad para un coeficiente de credibilidad α viene dado por $(\bar{x} - T_{1-(1-\alpha)/2} s/\sqrt{n}; \bar{x} + T_{1-(1-\alpha)/2} s/\sqrt{n})$, siendo *T* un percentil de la distribución *T* con *n-1* grados de libertad.

4.3.3. RESULTADOS

En el anexo 2 se presenta el análisis detallado de los resultados obtenidos en cada ítem. En la tabla 4.8 resumimos los porcentajes de respuestas correctas y resultados en los distractores. Se observa que en algunos ítems uno de los distractores fue dominante (elegido por la mayoría de los alumnos, más del 45% de los alumnos) o bien no fue elegido, con lo cual estos ítems se desecharían o bien, sería necesario revisar la redacción de las alternativas.

[27] La distribución T proporciona intervalos de credibilidad más amplios, porque tiene en cuenta también la variabilidad de la estimación de la varianza de la distribución inicial (Bolstad, 2004).

Tabla 4.8. Porcentaje de respuestas correctas y observaciones respecto a los distractores

Ítem	N	Índice de dificultad	Observaciones	Ítem	n	Índice de dificultad	Observaciones
1	49	75,5		26	52	25,3	
2	52	71,2	No eligen el distractor d	27	52	15,4	
3	49	24,5		28	157	33,1	Domina el distractor b
4	52	73,1		29	81	30,9	Domina el distractor b
5	157	24,8	Domina el distractor c	30	157	26,8	Domina el distractor c
6	49	36,7		31	157	91,1	
7	52	15,4	Domina el distractor d	32	157	31,8	
8	49	36,7		33	52	13,5	Domina el distractor a
9	52	a) 65,4; b) 46,2		34	76	40,8	No eligen el distractor a
10	49	a) 46,9; b) 38,8		35	157	48,4	
11	157	47,1	Domina el distractor c	36	157	8,0	
12	52	a) 76,9;b) 44,2; c) 65,4		37	49	55,1	
13	49	a) 24,5; b) 71,4; c) 46,9		38	52	21,2	
14	81	37,0	Domina el distractor c	39	52	9,6	
15	76	27,6	Domina el distractor c	40	49	0,0	
16	157	16,6		41	157	70,1	
17	49	75,5		42	49	40,8	
18	52	5,7		43	52	15,4	
19	52	48,1		44	76	39,5	
20	49	28,6		45	81	43,2	
21	49	73,5		46	49	14,3	
22	157	91,7		47	157	14,6	Domina el distractor a
23	81	56,8		48	52	40,4	
24	76	59,2		49	49	8,1	
25	81	33.8					

En las tablas 4.9 y 4.10 se presentan la estimación clásica y bayesiana de los índices de dificultad, en el segundo caso, utilizando una distribución inicial no informativa. Se utilizó como distribución inicial *Be(0,5, 0,5)* como recomiendan entre otros Lecoutre (1996) y Serrano (2003). La distribución final del índice de dificultad, viene dada par la distribución *Be(0,5+p, 0,5+q),* siendo *p* y *q* el número de respuestas correctas e incorrectas en el ítem. Los resultados de las dos estimaciones son muy parecidos, explicándose la diferencia por los valores (0,5; 0,5) incluidos en la distribución inicial, y porque en el cálculo del intervalo de credibilidad se ha usado la distribución exacta Beta, mientras que el cálculo de intervalos de confianza se hace mediante la aproximación normal. La interpretación, como se indica en varios apartados de la memoria es muy diferente, pues el intervalo de credibilidad indica el rango de valores en que se encuentra el índice de dificultad, con una probabilidad (epistémica) del 95%.

Tabla 4.9. Estimación clásica y bayesiana (distribución no informativa) de los índices de dificultad

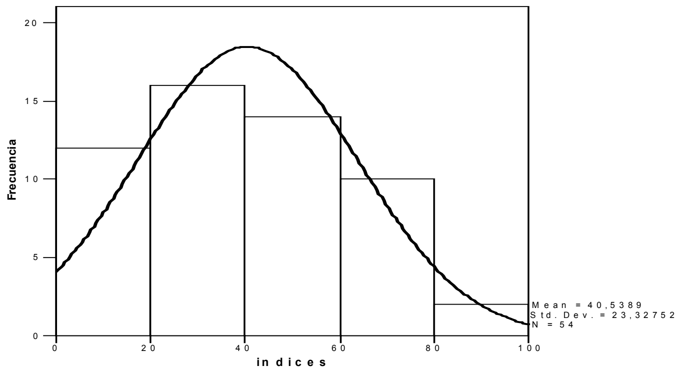
			Intervalo confianza 95%		Intervalo credibilidad 95%	
Ítem	n	Índice dificultad				
1	49	75,5	0,635	0,876	0,622	0,859
2	52	71,2	0,588	0,835	0,664	0,882
3	49	24,5	0,124	0,365	0,141	0,378
4	52	73,1	0,610	0,851	0,600	0,837
5	157	24,8	0,181	0,316	0,186	0,320
6	49	36,7	0,232	0,502	0,243	0,507
7	52	15,4	0,056	0,252	0,076	0,269
8	49	36,7	0,232	0,502	0,243	0,507
9a	52	65,4	0,525	0,783	0,519	0,772
9b	52	46,2	0,326	0,597	0,331	0,596
10a	49	46,9	0,330	0,609	0,335	0,608
10b	49	38,8	0,251	0,524	0,261	0,527
11	157	47,1	0,393	0,549	0,394	0,549
12a	52	76,9	0,655	0,884	0,642	0,867
12b	52	44,2	0,307	0,577	0,313	0,577
12c	52	65,4	0,525	0,783	0,519	0,772
13a	49	24,5	0,124	0,365	0,141	0,378
13b	49	71,4	0,588	0,841	0,578	0,826
13c	49	46,9	0,330	0,609	0,335	0,608
14	81	37,0	0,265	0,476	0,271	0,479
15	76	27,6	0,176	0,377	0,185	0,384
16	157	16,6	0,107	0,224	0,114	0,230
17	49	75,5	0,635	0,876	0,622	0,859
18	52	5,7	-0,006	0,121	0,017	0,146
19	52	48,1	0,345	0,617	0,349	0,615
20	49	28,6	0,159	0,412	0,174	0,422
21	49	73,5	0,611	0,858	0,600	0,842
22	157	91,7	0,874	0,960	0,866	0,953
23	81	56,8	0,460	0,676	0,459	0,672
24	76	59,2	0,482	0,703	0,480	0,697
25	81	33.8	0,231	0,436	0,238	0,440

Con objeto de aplicar el supuesto de normalidad en la estimación del valor medio de la distribución de índices de dificultad, se harán varias pruebas. En las figuras 4.1 y 4.2 se representan gráficamente los índices de dificultad obtenidos en las pruebas empíricas de ítems, donde se aprecia una distribución simétrica, aproximadamente normal, concentrada alrededor de una dificultad media del 40%. El 50% de valores centrales tiene una dificultad entre el 26 y el 66%.

Tabla 4.10. Estimación clásica y bayesiana (distribución no informativa) de los índices de dificultad

Ítem	n	Índice dificultad	Intervalo confianza 95%		Intervalo credibilidad 95$	
26	52	25,3	0,132	0,368	0,148	0,379
27	52	15,4	0,056	0,252	0,076	0,269
28	157	33,1	0,258	0,405	0,261	0,407
29	81	30,9	0,208	0,409	0,216	0,415
30	157	26,8	0,198	0,337	0,203	0,341
31	157	91,1	0,866	0,955	0,859	0,948
32	157	31,8	0,246	0,391	0,249	0,394
33	52	13,5	0,042	0,227	0,062	0,246
34	76	40,8	0,297	0,518	0,303	0,520
35	157	48,4	0,406	0,562	0,407	0,562
36	157	8,0	0,040	0,126	0,047	0,134
37	49	55,1	0,412	0,690	0,412	0,684
38	52	21,2	0,101	0,323	0,118	0,336
39	52	9,6	0,016	0,176	0,038	0,198
40	49	0,0	0,000	0,000	0,000	0,050
41	157	70,1	0,629	0,772	0,626	0,768
42	49	40,8	0,271	0,546	0,279	0,548
43	52	15,4	0,056	0,252	0,076	0,269
44	76	39,5	0,285	0,505	0,290	0,507
45	81	43,2	0,320	0,534	0,324	0,535
46	49	14,3	0,045	0,241	0,066	0,260
47	157	14,6	0,091	0,202	0,098	0,208
48	52	40,4	0,270	0,537	0,279	0,539
49	49	8,1	0,005	0,158	0,028	0,182

Figura 4.1. Histograma de índices de dificultad



En la tabla 4.11 se presentan los estadísticos descriptivos de los índices de dificultad, donde se observa la similitud de las medidas de posición central. La ausencia de valores atípicos, así como los coeficientes de asimetría y curtosis de nuevo indican la normalidad, que es confirmada en la prueba de Kolmogorov Smirnov (tabla 4.12). Ello nos autoriza también a aplicar la estimación bayesiana para el caso de normalidad que se incluye en la tabla 4.11, para el caso de distribución inicial no informativa. También en este caso la estimación clásica y bayesiana coinciden aproximadamente en valor pero tienen diferente interpretación. La pequeña diferencia se debe a que en nuestro programa hemos usado la cuasivarianza como estimador de la varianza y la distribución T exacta, en lugar de usar la aproximación normal, proporcionada por la subrutina SPSS.

Figura 4.2. Gráfico Q-Q y gráfico de caja de índices de dificultad

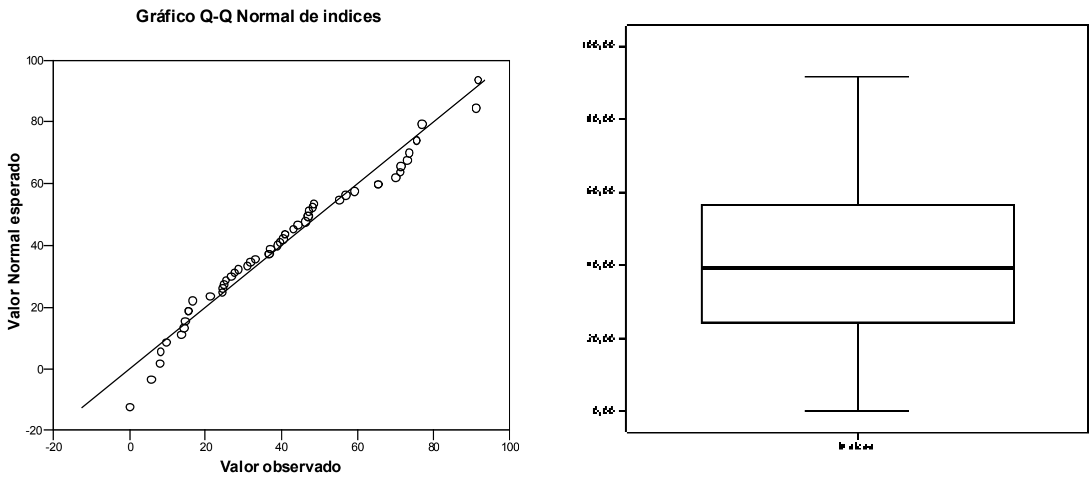


Tabla 4.11. Estadísticos Descriptivos de los índices de dificultad

		Estadístico	Error típ.
Media		40,5389	3,17447
Intervalo de credibilidad para la media al 95%	Límite inferior	33,351	
	Límite superior	47,726	
Intervalo de confianza para la media al 95%	Límite inferior	34,1717	
	Límite superior	46,9061	
Media recortada al 5%		39,9418	
Mediana		39,1500	
Varianza		544,173	
Desv. típ.		23,32752	
Amplitud intercuartil		33,88	
Asimetría		0,374	0,325
Curtosis		-0,695	0,639

Tabla 4.12. Prueba de Kolmogorov-Smirnov para una muestra

N		54
Parámetros normales(a,b)	Media	40,5389
	Desviación típica	23,32752
Diferencias más extremas	Absoluta	0,090
	Positiva	0,090
	Negativa	-0,083
Z de Kolmogorov-Smirnov		0,663
Sig. asintót. (bilateral)		0,771

a La distribución de contraste es la Normal.
b Se han calculado a partir de los datos.

Todos estos datos se tendrán en cuenta en la selección final de los ítems de la prueba, en base a su dificultad.

4.3. ESTUDIO 2.2. SELECCIÓN DE ÍTEMS PARA EL INSTRUMENTO PILOTO A PARTIR DE JUICIO DE EXPERTOS

En la valoración mediante juicio se requirió que algunas personas evaluasen los ítems respecto a los siguientes criterios (Millman y Greene, 1989; Thorndike, 1989):

- Comunicabilidad: Redacción correcta, claridad y consistencia. Tanto el director del trabajo como tres compañeros de uno de los cursos de doctorado, que son profesores de estadística, leyeron las versiones previas. También se pidió a los estudiantes que han participado en las pruebas que indicaran las posibles dificultades de comprensión.
- Nivel adecuado de dificultad. En el cuestionario se incluyó una gama de dificultades, deducidas de los datos en las prueba pre-piloto de ítems.
- La adecuación a las especificaciones del test. Este aspecto fue valorado por expertos.

4.3.1. MÉTODO

4.3.1.1. SUJETOS

Una vez preparada la tabla de especificaciones del cuestionario y recopilados los ítems que a nuestro juicio cubrían los objetivos propuestos, se llevó a cabo una fase de valoración por medio de expertos externos[28]. Millman y Greene (1989) indican que el "experto" lo define el propósito del instrumento y que el grupo elegido de expertos ha de representar una diversidad relevante de capacidades y puntos de vista. Las muestras de expertos son frecuentes en los estudios cualitativos y exploratorios para generar la materia prima en el diseño de un cuestionario (Hernández, Fernández y Baptista, 1998). En este trabajo, fueron seleccionados en base a su conocimiento experto de probabilidades y más

[28] La selección de ítems por muestreo racional, llevado a cabo a partir del juicio de expertos es uno de los criterios fundamentales usados en la construcción de instrumentos (López Feal, 1986). Lo que se persigue es la validación del contenido del cuestionario y usar la información recogida de los expertos para seleccionar los ítems que finalmente constituyan en cuestionario.

particularmente de probabilidad condicional, así como a la experiencia de investigación sobre el tema.

Para identificarlos, se comenzó a partir de la búsqueda bibliográfica descrita en el capítulo 1, preparando una lista de investigadores que hubiesen publicado en los últimos 10 años artículos en revistas o congresos internacionales de prestigio o tesis doctorales directamente relacionadas con la enseñanza de la probabilidad condicional o inferencia estadística. Es decir, la lista inicial se componía de los investigadores que han sido citados en las referencias utilizadas en el análisis de investigaciones previas. De entre ellos, un total de 23 podían leer la lengua castellana, cuestión fundamental, puesto que los ítems están realizados en dicha lengua. Se trataba de un total de 11 investigadores españoles y 12 de otros países (Brasil, Portugal, Venezuela, México, Colombia). Todos ellos eran localizables por correo electrónico, puesto que sus direcciones eran accesibles en Internet, a través de sus Departamentos o de listas de miembros de Sociedades Internacionales.

Se decidió invitar a participar a un total de 10 expertos, de los cuáles 9 aceptaron. Se eligieron de modo que se cubriesen diversos países (España, México, Venezuela y Brasil) y con mayoría de investigadores españoles (cubriendo 4 universidades españolas y otras cuatro extranjeras).

La muestra de expertos incluyó siete matemáticos y dos estadísticos, todos ellos profesores universitarios de matemáticas y estadística. De este grupo de profesores dos poseían un grado de Master o equivalente y siete de doctorado. Seis de ellos realizan su investigación en el campo de la probabilidad, uno en la idea de la independencia, uno en variable aleatoria y otro la asociación. La edad abarca desde 40 a 55 años. Todos los participantes contaban con al menos 10 años de experiencia investigadora.

4.3.1.2. MATERIAL

Una vez finalizada la tabla de especificaciones y seleccionado el conjunto inicial de ítems, se preparó un cuestionario que se recoge en el Anexo 1 para ser completado por estos expertos con un objetivo doble, todo ello para aportar evidencias de validez de contenido (Messick, 1989, 1995, 1998; Martínez Arias, 1995).

- Establecer un consenso sobre la tabla de especificaciones del instrumento, decidiendo cuales especificaciones del contenido eran relevantes para los propósitos del instrumento. De este modo se reforzarían los resultados obtenidos del análisis de contenido de los libros de texto.

- Establecer un consenso de opiniones de los expertos sobre cómo cada ítem particular se ajusta bien para evaluar el contenido específico para el cuál ha sido diseñado, que sirva como base para elegir los ítems definitivos.

El cuestionario comenzaba con una carta de presentación, explicando los fines de la investigación y requiriendo la ayuda del experto. A continuación se incluían unas instrucciones sobre la forma de completarlo (véase la primera parte de la tabla 4.13). Seguidamente, para cada una de las unidades de contenido, se presentaba su descripción, seguida de los ítems que servían para evaluarla y una tabla de datos que el experto debía completar evaluando dos aspectos (ver ejemplo, en tabla 4.13), ambos de acuerdo a una escala de cinco puntos, en la que los valores extremos toman el siguiente significado:

1. Nada: bajo grado de relevancia o relevancia incierta.

5. Mucho: grado de relevancia alto.

La decisión de usar una valoración en una escala (1-5) se tomó teniendo en cuenta las recomendaciones de Osterlind (1989), quien opina que hay que pedir algo más que el acuerdo/ desacuerdo a los jueces, pero sugiere usar únicamente las categorías anteriores, ya que una discriminación mayor tendría poca utilidad práctica y puede complicar innecesariamente el proceso de alcanzar el acuerdo.

El primer punto a valorar fue la relevancia de la unidad de contenido para la comprensión de la probabilidad condicional. Para llevar a cabo esta tarea se proporcionó, al final del cuestionario a los expertos la tabla de especificaciones de contenido. Asimismo se les pidió añadiesen otros posibles contenidos no incluidos en la tabla de especificaciones, si así lo consideraban necesario.

En segundo lugar se pidió a los expertos su grado de acuerdo respecto al emparejamiento de cada uno de los ítems con su contenido principal, siguiendo la misma escala (1 a 5). Se pidió a los jueces que evaluasen el emparejamiento previamente hecho entre especificaciones de contenido e ítems[29]. Por tanto, se indicó a los jueces cuáles ítems

[29] Uno de los dos métodos principales que sugiere Osterlind (1989) para recoger la opinión de los expertos.

se pensaba evaluaban cada contenido específicos y el papel de los jueces fue confirmar o rechazar esta opinión[30].

Tabla 4.13. Ejemplo de pregunta del cuestionario de expertos

Cuestionario para Expertos

A continuación presentamos una lista de contenidos que consideramos relevantes para evaluar la comprensión del tema de la probabilidad condicional. Junto a cada contenido, presentamos una lista de ítems de evaluación.

Requerimos su colaboración para evaluar:

- El grado en que el contenido propuesto es relevante para la comprensión de la probabilidad condicional.
- El grado en que cada ítem es adecuado para evaluar la comprensión del contenido específico propuesto.

Le agradeceríamos marque para cada uno de estos dos aspectos su opinión en la escala 1 a 5, donde 1 indica: nada relevante y 5 muy relevante.

Contenido 1: Definición de la probabilidad condicional.

Ítem 1. Explica con tus propias palabras la diferencia entre una probabilidad simple y una probabilidad condicional.

Ítem 2. ¿Qué quiere decir la expresión "la probabilidad condicional de A dado B es 1/4"?
a) En la cuarta parte de los experimentos obtenemos A y B simultáneamente
b) A ocurre la cuarta parte de las veces en que ocurre B
c) B ocurre la cuarta parte de las veces en que ocurre A
d) A o B ocurren la cuarta parte de las veces

	1: Nada	2	3	4	5: Mucho
El contenido "Definición de la probabilidad condicional" es relevante					
El ítem 1 es adecuado para este contenido					
El ítem 2 es adecuado para este contenido					

Puesto que la finalidad era asegurar que se evaluaran todas las especificaciones de la definición semántica de la variable, se seleccionaron al menos dos ítems que evaluasen cada contenido principal, (aunque, adicionalmente, cada ítem evalúa otros contenidos secundarios). Una vez elegidos los ítems, en base a su contenido principal, se realizara un análisis teórico del contenido evaluado, que se presenta en el estudio 6.

[30] Se eligió este método frente a la posibilidad de pedir a los jueces que libremente emparejen los ítems a los diferentes contenidos ya que el número de objetivos e ítems diferentes era alto.

Por otro lado, frente a la opción de dividir el banco inicial de ítems en varios grupos que fuesen evaluados por diferentes grupos de expertos, se prefirió dar a todos los expertos el conjunto completo de ítems y objetivos.

4.3.2. ANÁLISIS

Para resumir el grado de emparejamiento que todos los jueces dan al contenido (o al ítem), uno de los procedimientos habituales es utilizar como resumen global la media o la mediana (Martínez Arias, 1995). En este estudio se han calculado ambos, pero se ha usado la media para la selección, ya que tiene en cuenta los valores extremos, por lo que si hay alguna discrepancia fuerte, se tiene en cuenta. También se tuvo en cuenta la desviación típica, en caso de igualdad de medias. De esta forma, se eligió un criterio más estricto para la selección de los contenidos.

4.3.3. RESULTADOS

En la tabla 4.14 se presentan la media, mediana y desviación típica de las puntuaciones asignadas por los expertos a cada uno de los contenidos de la tabla de especificaciones del cuestionario.

Se aceptaron todos los contenidos que recibieron una alta valoración (por encima de 4 puntos), que fueron todos, excepto los contenidos 2 y 3, que fueron considerados algo menos adecuados, en relación con el tipo de alumnos a los que va dirigido el cuestionario.

En estos contenidos hubo desacuerdo. Aunque la mayoría de los expertos los consideraron importantes, algunos jueces los puntuaron con muy baja puntuación, sugiriendo que eran demasiado abstractos para los alumnos a quienes iba dirigida la prueba. Por ejemplo, el experto 8 indica: "*Definitivamente considero irrelevante que a alumnos de la áreas de humanidades y sociales se le evalúe este contenido por exigirles un nivel de formalización. Es mejor dar énfasis a otro tipo de habilidades y contenidos*". En consecuencia, se decidió eliminar los contenidos 2 y 3.

Tabla. 4.14. Frecuencia de asignación de acuerdo (1-5) por los expertos a cada contenido, media, mediana y desviación típica

	Contenido	1	2	3	4	5	$\bar{x}$	med	s
Conocimiento conceptual	1. Definición de la probabilidad condicional					9	5,00	5	0,00
	2. Reconocer que la probabilidad de P(A) >0 para poder definir P(B/A)	1	2		1	5	3,78	5	1,64
	3. Reconocer que una probabilidad condicional cumple los axiomas.		2	1	2	4	3,89	4	1,27
	4. Reconocer que la probabilidad condicional supone una restricción del espacio muestral				1	8	4,89	5	0,33
	5. Distinguir probabilidad condicional con inversa					9	5,00	5	0,00
	6. Distinguir probabilidad conjunta, condicional y simple				1	8	4,89	5	0,33
	7. Probabilidad conjunta menor que probabilidad simple				3	6	4,67	5	0,50
	8. Distinguir sucesos independientes, dependientes y mutuamente excluyentes				1	8	4,89	5	0,33
Conocimiento procedimental	9. Calcular una probabilidad condicional dentro de un único experimento			1	1	7	4,67	5	0,71
	10. Resolver correctamente problemas de probabilidad condicional en un contexto de muestreo con reposición				4	5	4,56	5	0,53
	11. Resolver correctamente problemas de probabilidad condicional en un contexto de muestreo sin reposición					9	5,00	5	0,00
	12. Resolver correctamente problemas de probabilidad condicional a partir de probabilidades conjuntas y simples				1	8	4,89	5	0,33
	13. Resolver correctamente problemas condicionales cuando se invierte el eje de tiempo				5	4	4,44	4	0,53
	14. Distinguir situación condicional, causal y diagnóstica				7	2	4,22	4	0,44
	15. Resolver correctamente problemas en situaciones diacrónicas	1			2	6	4,33	5	1,32
	16. Resolver correctamente problemas en situaciones sincrónicas			1	1	7	4,67	5	0,71
	17. Resolver correctamente problemas de probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos independientes				1	8	4,89	5	0,33
	18. Resolver correctamente problemas de probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos dependientes				2	7	4,78	5	0,44
	19. Aplicar correctamente el cálculo de la probabilidad condicional en situaciones de sucesos múltiples (regla de la probabilidad total)				1	8	4,89	5	0,33
	20. Aplicar correctamente el cálculo de la probabilidad condicional en situaciones de probabilidad inversa (regla de Bayes)				1	8	4,89	5	0,33

Por otro lado, algunos jueces mencionaron contenidos adicionales, que fueron los siguientes:

- *Reconocer el lenguaje de la probabilidad condicional* (experto 1), refiriéndose a la notación. Este experto incluso sugirió un ítem para valorar este contenido. Se pensó que es posible valorar el uso del lenguaje a partir de las respuestas de los alumnos a los ítems de respuesta abierta, puesto que en el ensayo de estos ítems los alumnos utilizaron las notaciones de la probabilidad simple, compuesta y condicional, tanto correcta, como incorrectamente.

Tabla 4.15 Frecuencia de asignación de acuerdo (1-5) por los expertos a cada ítem respecto a su contenido principal, media, mediana y desviación típica

Ítem	1	2	3	4	5	Media	Mediana	Desv. típ.
I1		1		4	4	4,22	4	0,972
I2		1	1	1	6	4,33	5	1,118
I3	**2**		**1**	**3**	**3**	**3,56**	**4**	**1,590**
I4	**1**	**2**	**5**		**1**	**2,78**	**3**	**1,093**
I5	**1**	**4**	**2**	**1**	**1**	**2,67**	**2**	**1,225**
I6	**1**	**2**	**1**	**2**	**3**	**3,44**	**4**	**1,509**
I7		**1**	**3**	**2**	**3**	**3,78**	**4**	**1,093**
I8		**3**	**5**	**1**		**2,78**	**3**	**0,667**
I9	1	1		1	6	4,11	5	1,537
I10				1	8	4,89	5	0,333
I11	**2**		**1**	**2**	**4**	**3,67**	**4**	**1,658**
I12				1	8	4,89	5	0,333
I13			1	2	6	4,56	5	0,726
I14			3	2	4	4,11	4	0,928
I15				5	4	4,44	4	0,527
I16			2	4	3	4,11	4	0,782
I17	**1**	**1**	**1**	**4**	**2**	**3,56**	**4**	**1,333**
I18				2	7	4,78	5	0,441
I19		**2**	**2**	**2**	**3**	**3,67**	**4**	**1,225**
I20		1	1	2	5	4,22	5	1,093
I21			1	5	3	4,22	4	0,667
I22			1	5	3	4,22	4	0,667
I23				3	6	4,67	5	0,500
I24				3	6	4,67	5	0,500
I25			1	4	4	4,33	4	0,707
I26				4	5	4,56	5	0,527
I27	1			2	6	4,33	5	1,323
I28		2		3	4	4,00	4	1,225
I29			1	4	4	4,33	4	0,707
I30	**1**		**3**	**3**	**2**	**3,56**	**4**	**1,236**
I31			2	3	4	4,22	4	0,833
I32			1	4	4	4,33	4	0,707
I33	1			3	5	4,22	5	1,302
I34	1			5	3	4,00	4	1,225
I35		**1**	**3**	**1**	**4**	**3,89**	**4**	**1,167**
I36	1		1	3	4	4,00	4	1,414
I37		1	1	2	5	4,22	5	1,093
I38				2	7	4,78	5	0,441
I39			1	3	5	4,44	5	0,726
I40			3	2	4	4,11	4	0,928
I41			1	4	4	4,33	4	0,707
I42			1	2	6	4,56	5	0,726
I43			2	1	6	4,44	5	0,882
I44				3	6	4,67	5	0,500
I45				3	6	4,67	5	0,500
I46		**2**	**1**	**4**	**2**	**3,67**	**4**	**1,118**
I47			1	3	5	4,44	5	0,726
I48				5	4	4,44	4	0,527
I49				3	6	4,67	5	0,500

- *La probabilidad condicional obliga a hacer una nueva asignación de probabilidades* (experto 3). De hecho este contenido está implícito en el 4 *"Reconocer que una probabilidad condicional supone una restricción del espacio muestral*".
- *Uso de diagrama de árbol o tablas para resolver problemas de probabilidad condicional* (experto 5). También este contenido se evalúa mediante los ítems de respuesta abierta.

En consecuencia se decidió conservar todos los contenidos, excepto los 2 y 3 y no añadir más, puesto que los sugeridos por los expertos estaban contemplados implícitamente en las tareas.

En la tabla 4.15 se presentan los resultados del grado de acuerdo que los expertos han juzgado entre cada uno de los ítems y su contenido principal. Los resultados indicaron, en general, ítems bien valorados, con puntuaciones superiores a 4 puntos.

Los casos en que la puntuación media se encuentra por debajo de esta puntuación se refieren, bien a ítems relacionados con contenidos considerados demasiado abstractos para los alumnos (contenidos 2 y 3 a los que corresponden los ítems 3, 4, 5 y 6 todos los cuales reciben puntuación media por debajo de 4), bien a que la redacción de los ítems o su contexto se consideraba demasiado formal para los alumnos.

Puesto que para cada contenido había al menos dos ítems, se pudo elegir el mejor valorado. Se mejoró también la redacción o el contexto en caso de que algún experto lo sugiriese.

4.4. SELECCIÓN DE ÍTEMS Y COMPOSICIÓN DEL PRIMER INSTRUMENTO PILOTO

Finalizadas las fases anteriores, se seleccionaron los ítems que compondrían el cuestionario piloto, con el siguiente proceso:

- Primeramente se desecharon aquellas especificaciones de contenido en las que el grado de acuerdo sobre su relevancia no fuese suficientemente elevado y consensuado. La relevancia se valoró mediante la puntuación media dada al contenido por los nueve expertos (a mayor valor medio mayor relevancia) y el grado de acuerdo por la desviación típica (a menor valor, mayor acuerdo).

- Para aquellos contenidos con alto grado de acuerdo respecto a su relevancia se examinaron los ítems.
- Se desecharon los ítems menos valorados por los expertos en cuanto a su emparejamiento con el contenido (puntuación media menor de 4 en las escala 1-5).
- De entre aquellos ítems bien valoraros en cuando a su emparejamiento con el contenido (puntuación superior a 4), se eligió el que hubiese presentado menor índice de dificultad en las pruebas pre-piloto de ítems.

Los restantes ítems quedaron reservados, para el caso de que los resultados del cuestionario piloto hiciesen necesarios aumentar el número de ítems para un cierto contenido o sustituir algunos de los ítems previamente seleccionados. Hacemos notar, que, además del contenido principal, cada ítem permite evaluar la comprensión de otros contenidos secundarios. En el estudio 6 se analizarán las posibles respuestas a cada uno de los ítems seleccionados para formar el cuestionario piloto, y los contenidos secundarios evaluados por cada uno de ellos. Esto permitirá aportar evidencias de validez de contenido del instrumento.

A continuación se analizan para cada uno de los contenidos los resultados de los ítems diseñados para evaluarlo, así como los criterios seguidos para seleccionar los ítems que compondrán el instrumento piloto.

Contenido 1: Definición de la probabilidad condicional

Este contenido fue valorado muy positivamente ($\bar{x}$ = 5,00, s = 0), (véase tabla 4.14). Puesto que en el ítem 2 una de las alternativas no fue elegida por los alumnos, se decidió conservar el ítem 1 para la prueba piloto.

Tabla 4.16. Resultados en los ítems relacionados con el contenido 1

Ítem	Valoración expertos							Índice		Observaciones
	1	2	3	4	5	Media	Desv. típ.	dificultad	n	
I1		1		4	4	4,22	0,972	0,755	49	
I2		1	1	1	6	4,33	1,118	0,712	52	Alternativa d no se eligió

Contenidos 2 y 3: Reconocer que *P(A)* > 0 para poder definir *P (B/A)* y reconocer que una probabilidad condicional cumple los axiomas de probabilidad

Estos dos contenidos se descartaron, dado que la valoración media de los expertos de los mismos fue es menor que 4 (contenido 2, $\bar{x} = 3{,}78$, $s = 1{,}64$; contenido 3, $\bar{x} = 3{,}89$, $s = 1{,}27$). La razón fue considerarlo demasiado formal para la población de sujetos a la que va dirigido el cuestionario: *"Definitivamente considero irrelevante que a alumnos de las áreas de humanidades y sociales se le evalúe este contenido por exigirles un nivel de formalización. Es mejor dar énfasis a otro tipo de habilidades y contenidos"* (experto 8).

Contenido 4: Reconocer que la probabilidad condicional supone una restricción del espacio muestral

Tabla 4.17. Resultados en los ítems relacionados con el contenido 4

Ítem	Valoración expertos							Índice dificultad	Observaciones	
	1	2	3	4	5	Media	Desv. típ.		n	
I7		1	3	2	3	3,78	1,093	0,154	52	
I8		3	5	1		2,78	0,667	0,367	49	

Este contenido, tuvo una valoración media de los expertos de $\bar{x} = 4{,}89$ y $s = 0{,}33$. Se descartaron ambos ítems, que se consideraron muy formales: *"El contenido creo que es muy importante. Ahora bien, quizá los ítems sean demasiado memorísticos o poco indicadores de que realmente se entiende. En mi opinión debería construirse un ítem en el que fuera necesario especificar el subespacio y calcular las probabilidades de cada uno de los sucesos elementales en este subespacio"* (experto 6). *"Este contenido se me hace especialmente importante porque fomenta la noción intuitiva de probabilidad condicional, independientemente de su expresión formal. Pero no me parece que los ítems sean correctos, sugeriría un problema de probabilidad condicional donde se tuvieran que elegir entre distintos espacios maestrales posibles expresados en los incisos"* (experto 8). Uno de los expertos (experto 3) sugirió el siguiente ítem que se incluyó en el cuestionario piloto:

Ítem. Define el espacio muestral de los siguientes experimentos aleatorios:
a) El sexo de los hijos de las familias con tres descendientes.
b) El sexo de las familias de tres descendientes en los que dos son varones

Contenido 5: Distinguir una probabilidad condicional y su inversa

Este es otro de los contenidos más valorados ($\bar{x}$ = 5, s = 0). De los tres ítems, el 11 fue mal valorado por los expertos, por lo que se descartó. Se conservó el ítem 10, dado que en el ítem 9 la condicional y la inversa toman el mismo valor, hecho comentado por los expertos: *"Como las respuestas son idénticas no podemos estar seguros si ha distinguido bien P(A/B) de P(B/A). Creo que se deberían cambiar los números."* (experto 3).

Tabla 4.18. Resultados en los ítems relacionados con el contenido 5

Ítem	Valoración expertos							Índice dificultad		Observaciones
	1	2	3	4	5	Media	Desv. típ.		n	
I9	1	1		1	6	4,11	1,537	a) 0,654	52	En este ejemplo, la condicional y su inversa es el mismo número.
								b) 0,462	52	
I10				1	8	4,89	0,333	a) 0,469	49	
								b) 0,388	49	
I11	2		1	2	4	3,67	1,658	0,471	157	

Contenido 6: Distinguir probabilidad conjunta, condicional y simple

En este contenido, (valoración de los expertos $\bar{x}$ = 4,89, s = 0,33) ambos ítems están bien valorados. Se seleccionó el 13, que, aunque un poco más difícil, va ligado al ítem 10 elegido para el contenido anterior.

Tabla 4.19. Resultados en los ítems relacionados con el contenido 6

Ítem	Valoración expertos							Índice dificultad		Observaciones
	1	2	3	4	5	Media	Desv. típ.		n	
I12				1	8	4,89	0,333	a) 0,769	52	
								b) 0,442	52	
								c) 0,654	52	
I13			1	2	6	4,56	0,726	a) 0,245	49	
								b) 0,714	49	
								c) 0,469	49	

Contenido 7: Reconocer que la probabilidad conjunta *P(A y B)* ha de ser menor o igual que la probabilidad simple *P(A)*

Este contenido tuvo valoración media $\bar{x}$ = 4,67 y s = 0,50. Los expertos valoran algo mejor el ítem 15, aunque el ítem 14 resulta más fácil. Se eligió el ítem 14, dado que el contexto político del ítem 15 puede ocasionar un juicio subjetivo, como señala uno de los expertos: *"Los contextos políticos suelen tener fuertes teoría previas y preferencias/ rechazos subjetivas."*

Tabla 4.20. Resultados en los ítems relacionados con el contenido 7

Ítem	Valoración expertos							Índice dificultad	n	Observaciones
	1	2	3	4	5	Media	Desv. típ.			
I14			3	2	4	4,11	0,928	0,370	81	
I15				5	4	4,44	0,527	0,276	76	

Contenido 8: Distinguir sucesos independientes, dependientes y mutuamente excluyentes

La valoración media de los expertos de este contenido fue de $\bar{x}$ = 4,89 y s =0,33. El ítem 17 queda descartado por los jueces y conservamos el ítem 16, dado que el ítem 18 presenta un índice de dificultad demasiado alto. Además las alternativas del ítem 16 quedan adecuadamente distribuidas.

Tabla 4.21. Resultados en los ítems relacionados con el contenido 8

Ítem	Valoración expertos							Índice dificultad	n	Observaciones
	1	2	3	4	5	Media	Desv. típ.			
I16			2	4	3	4,11	0,782	0,166	157	
I17	1	1	1	4	2	3,56	1,333	0,755	49	
I18				2	7	4,78	0,441	0,057	52	

Contenido 9: Calcular una probabilidad condicional dentro de un único experimento

Este contenido fue valorado por los expertos con una puntuación media de $\bar{x}$ = 4,67 y s = 0,71. Incluye los ítems 19 y 20. Los expertos descartan el ítem 19, conservamos el ítem 20 para el cuestionario piloto.

Tabla 4.22. Resultados en los ítems relacionados con el contenido 9

Ítem	Valoración expertos							Índice dificultad	n	Observaciones
	1	2	3	4	5	Media	Desv. típ.			
I19		2	2	2	3	3,67	1,225	0,481	49	
I20		1	1	2	5	4,22	1,093	0,286	52	

Contenido 10: Calcular la probabilidad condicional en un contexto de muestreo con reposición.

Los expertos valoran este contenido con una puntuación media de $\bar{x} = 4{,}56$ y $s = 0{,}53$. Los dos ítems son valorados positivamente por los expertos. Se elige el ítem 21, dado que el 22 tiene un índice de dificultad demasiado bajo.

Tabla 4.23. Resultados en los ítems relacionados con el contenido 10

Ítem	Valoración expertos							Índice dificultad	Observaciones	
	1	2	3	4	5	Media	Desv. típ.		n	
I21			1	5	3	4,22	0,667	0,735	48	
I22			1	5	3	4,22	0,667	0,917	157	

Contenido 11: Calcular la probabilidad condicional en un contexto de muestreo sin reposición.

La puntuación media de la valoración de los expertos para este contenido fue una de las más altas, $\bar{x} = 5$, $s = 0$. Ambos ítems son valorados positivamente por los expertos y ambos tienen un índice de dificultad similar. Escogemos el ítem 23 dado que en el ítem 24 hay una alternativa que no es escogida.

Tabla 4.24. Resultados en los ítems relacionados con el contenido 11

Ítem	Valoración expertos							Índice dificultad	Observaciones	
	1	2	3	4	5	Media	Desv. típ.		n	
I23				3	6	4,67	0,500	0,568	81	
I24				3	6	4,67	0,500	0,592	76	Hay una alternativa poco elegida

Contenido 12: Calcular una probabilidad condicional a partir de probabilidades conjuntas y simples.

Este contenido tiene una valoración media de los expertos de $\bar{x} = 4{,}89$ y $s = 0{,}33$. Los tres ítems fueron valorados positivamente por los expertos. Eliminamos los ítems 26 y 27 que han resultado demasiado difíciles.

Tabla 4.25. Resultados en los ítems relacionados con el contenido 12

Ítem	Valoración expertos							Índice dificultad		Observaciones
	1	2	3	4	5	Media	Desv. típ.		n	
I25			1	4	4	4,33	0,707	0,338	81	
I26				4	5	4,56	0,527	0,253	52	
I27	1			2	6	4,33	1,323	0,154	52	

Contenido 13: Resolver correctamente problemas condicionales cuando se invierte el eje de tiempo

Este contenido es valorado por los expertos con una puntuación media de $\bar{x}$ = 4,44 y s = 0,53. Los expertos valoran positivamente ambos ítems, aunque algo mejor el 29, que se mantiene, dado que además es la continuación del ítem 23 que se eligió para un contenido anterior.

Tabla 4.26. Resultados en los ítems relacionados con el contenido 13

Ítem	Valoración expertos							Índice dificultad		Observaciones
	1	2	3	4	5	Media	Desv. típ.		n	
I28		2		3	4	4,00	1,225	0,331	157	
I29			1	4	4	4,33	0,707	0,309	81	

Contenido 14: Distinguir situación condicional, causal y diagnóstica

Este contenido tuvo una valoración media de los expertos de $\bar{x}$ = 4,22 y s =0,44. Incluye los ítems 30, 31 y 32. El ítem 30 es descartado por los expertos: *"El contenido se me hace valioso. El inconveniente de los ítems es que pueden ser contestados por sentido común, pero no por eso saber distinguir realmente entre una situación condicional, causal y diagnóstica"* (experto 8). Se conservó el ítem 32, dado que el 31 es demasiado fácil.

Tabla 4.27. Resultados en los ítems relacionados con el contenido 14

Ítem	Valoración expertos							Índice dificultad		Observaciones
	1	2	3	4	5	Media	Desv. típ.		n	
I30	1		3	3	2	3,56	1,236	0,268	157	
I31			2	3	4	4,22	0,833	0,911	157	
I32			1	4	4	4,33	0,707	0,318	157	

Contenido 15: Resolver correctamente problemas en situaciones diacrónicas

Este contenido tiene una valoración media de los expertos de $\bar{x}$ = 4,33 y s = 1,32. Los expertos valoran más positivamente el ítem 33. Aunque resulta más difícil, se elige el ítem

33, dado que el 34 tiene una alternativa que no se escoge y además el ítem 33 permite dar un valor numérico.

Tabla 4.28. Resultados en los ítems relacionados con el contenido 15

Ítem	Valoración expertos							Índice dificultad	Observaciones	
	1	2	3	4	5	Media	Desv. típ.		n	
I33	1			3	5	4,22	1,302	0,135	52	
I34	1			5	3	4,00	1,225	0,408	76	Alternativa no elegida

Contenido 16: Resolver correctamente problemas en situaciones sincrónicas

Los expertos han valorado este contenido favorablemente con una puntuación media de $\bar{x}$ = 4,67 y s = 0,71. Los expertos descartan el ítem 35, que es contraintuitivo: *"Este es otro de los problemas contraintuitivos que aparece en la literatura. Al ser contraintuitivo, los sujetos pueden cometer errores por este hecho y, en consecuencia no se mide lo que se pretende."* El ítem 36 resultó demasiado difícil, posiblemente porque el enunciado no resultó comprensible. Se mantiene el ítem 37 que presenta un índice de dificultad adecuado.

Tabla 4.29. Resultados en los ítems relacionados con el contenido 16

Ítem	Valoración expertos							Índice dificultad	Observaciones	
	1	2	3	4	5	Media	Desv. típ.		n	
I35		1	3	1	4	3,89	1,167	0,484	157	
I36	1		1	2	4	4,00	1,414	0,080	157	Gran proporción respuestas en blanco
I37		1	1	2	5	4,22	1,093	0,551	49	

Contenido 17: Calcular una probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos independientes

Los expertos han valorado este contenido con una puntuación de $\bar{x}$ = 4,89 y s = 0,33. Se toma el ítem 38 dado que es el que mejor valoran los expertos y el que mejores resultados tiene en la pasación a los alumnos.

Tabla 4.30. Resultados en los ítems relacionados con el contenido 17

Ítem	Valoración expertos							Índice dificultad	Observaciones	
	1	2	3	4	5	Media	Desv. típ.		n	
I38				2	7	4,78	0,441	0,212	52	
I39			1	3	5	4,44	0,726	0,096	52	
I40			3	2	4	4,11	0,928	0,000	49	14'3% de alumnos identifica el problema como p. compuesta, pero no lo resuelven

Contenido 18: Calcular una probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos dependientes

Los expertos han valorado este contenido con una puntuación media de $\bar{x}$ = 4,78 y s = 0,44. Los tres ítems fueron valorados positivamente. Escogemos el ítem 41 dado que el que mejores resultados ha obtenido en las pasaciones.

Tabla 4.31. Resultados en los ítems relacionados con el contenido 18

Ítem	Valoración expertos							Índice dificultad	Observaciones
	1	2	3	4	5	Media	Desv. típ.		n
I41			1	4	4	4,33	0,707	0,701	157
I42			1	2	6	4,56	0,726	0,408	49
I43			2	1	6	4,44	0,882	0,154	52

Contenido 19: Aplicar correctamente la regla de la probabilidad total en un problema

La puntuación media que han dado los expertos a este contenido es de $\bar{x}$ = 4,89 y s = 0,33. El ítem 46 fue descartado por los expertos y el 44 resulta algo difícil. Escogemos el ítem 45 que, además, está en versión frecuencial.

Tabla 4.32. Resultados en los ítems relacionados con el contenido 19

Ítem	Valoración expertos							Índice dificultad	Observaciones
	1	2	3	4	5	Media	Desv. típ.		n
I44				3	6	4,67	0,500	0,395	76
I45				3	6	4,67	0,500	0,432	81
I46		2	1	4	2	3,67	1,118	0,143	49

Contenido 20: Aplicar correctamente el Teorema de Bayes en un problema

Los expertos han valorado este contenido con una puntuación de $\bar{x}$ = 4,89 y s = 0,33. Los tres ítems son valorados positivamente por los expertos. Descartamos el ítem 49 que resulta muy difícil. Decidimos escoger un ítem de más para el contenido, dado que se estima muy importante este contenido para nuestros objetivos. Se tomaron los ítems 47 y 48, dado que dan mejores resultados.

Tabla 4.33. Resultados en los ítems relacionados con el contenido 20

Ítem	Valoración expertos							Índice dificultad	Observaciones
	1	2	3	4	5	Media	Desv. típ.		n
I47			1	3	5	4,44	0,726	0,146	157
I48				5	4	4,44	0,527	0,404	52
I49				3	6	4,67	0,500	0,081	49

5. ESTUDIO 3. PRUEBA PILOTO DEL CUESTIONARIO RPC

5.1. INTRODUCCIÓN

Una vez elegidos los ítems que se incluirían en el instrumento piloto, se procedió a la realización de pruebas del mismo, con objeto de obtener información empírica sobre las características de esta primera versión del instrumento y comprobar que era útil para los objetivos pretendidos. Al mismo tiempo se deseaba analizar sus limitaciones, para identificar aquellos puntos en que sería necesario continuar el trabajo para mejorarlo.

Por otro lado, dado que más adelante (Estudio 8) se pretendía evaluar las dificultades de los alumnos de psicología en la probabilidad condicional, se consideró necesario obtener unos primeros datos, con carácter exploratorio. A partir de esta información se podrían obtener unas primeras indicaciones de la dificultad y discriminación de los ítems, así como determinar algunas categorías de respuestas en los ítems de respuesta abierta.

El objetivo del Estudio 3 fue realizar unas pruebas del cuestionario piloto, aún siendo conscientes de que éste no era un instrumento definitivo y que sería necesario mejorarlo.

A continuación se describen las muestras participantes, material sometido a prueba, análisis realizados y resultados.[31]

5.2. MÉTODO

5.2.1. SUJETOS

Las pruebas piloto se llevaron a cabo con dos muestras de estudiantes universitarios. La primera de ella estuvo formada por 37 alumnos que cursaban 5º año de la Licenciatura de Matemáticas en la especialidad de Metodología. Se eligió a estos alumnos con objeto de usar los datos en el estudio de la discriminación de los ítems, ya que este grupo tiene una alta preparación en probabilidad y porque el disponer de este

[31] Resultados parciales se han publicado en Díaz (2004); Batanero y Díaz (2005); Díaz y de la Fuente (2005d).

grupo de alumnos permitiría comparar si los errores más frecuentes en alumnos de psicología se repetían en alumnos con alta preparación matemática. La muestra estaba formada por 20 mujeres (54%) y 17 hombres (46%) con una nota de acceso media en selectividad de 7,04 (desviación típica 1,04). El cuestionario se completó como una actividad dentro de este curso, indicando a los estudiantes que se tendría en cuenta en su evaluación final.

La segunda muestra estuvo formada por 57 estudiantes de psicología de primer curso, alumnos de la asignatura "Análisis de Datos en Psicología" de los grupos de tarde, cuyas características son similares a los alumnos participantes en las pruebas pre-pilotos de ítems y de aquellos a quienes finalmente va dirigido el cuestionario, con 42 mujeres (74%) y 15 hombres (26%). Este grupo de alumnos había ya estudiado el tema de probabilidad condicional, aunque su preparación no es tan alta como la de los alumnos de matemáticas, quienes han seguido varios cursos de estadística en su licenciatura. La nota media de acceso en selectividad de este grupo fue 6,97 (desviación típica 0,76) muy similar a la del grupo de alumnos de matemáticas.

5.2.2. MATERIAL

El material utilizado es la versión piloto del cuestionario RPC, elaborada con el proceso descrito en los estudios 1 y 2.

Una vez seleccionados los ítems que formarían parte de cuestionario piloto, se prepararon dos versiones del mismo, en los que la única diferencia es el orden de colocación de las preguntas, que se varió con objeto de evitar que las preguntas finales quedasen con menor respuesta por efecto del cansancio. Además, al colocar las preguntas en el cuestionario, se alternaron preguntas sencillas y difíciles, para evitar que los alumnos se desanimasen si se ponían las preguntas difíciles al principio o que, por el contrario, abandonasen la prueba al llegar a las más difíciles. En la tabla 4.34 se presenta una de las dos versiones.

Tabla 4.34. Cuestionario piloto

Lee atentamente los enunciados y elabora o selecciona la respuesta que consideres correcta.

Ítem 1. Explica con tus propias palabras la diferencia entre una probabilidad simple y una probabilidad condicional.

Ítem 2. Define el espacio muestral de los siguientes experimentos aleatorios:

a) Observar el género (masculino/ femenino) de los hijos de las familias con tres descendientes.
b) Observar el género (masculino/ femenino) de los hijos de las familias con tres descendientes en los que dos son varones.

Ítem 3. Un taxi se vio implicado en un accidente nocturno con choque y huida posterior. Hay dos compañías de taxis en la ciudad, la Verde y la Azul. El 85% de los taxis de la ciudad son Verde y el 15% Azul. Un testigo identificó el taxi como Azul. El tribunal comprobó la fiabilidad del testigo bajo las mismas circunstancias que había la noche del accidente y llegó a la conclusión de que el testigo identificaba correctamente cada uno de los colores en el 80% de las ocasiones y fallaba en el 20%. ¿Cuál es la probabilidad de que el taxi implicado en el accidente fuera en efecto Azul?

☐ a) 80 %
☐ b) 15%
☐ c) (15/100) x (80/100)
☐ d) 41 %

Ítem 4. Se extrae una carta al azar de una baraja española (40 cartas con números del 1 al 7, sota caballo y rey): sea A el suceso "se extrae una carta de oros " y B el suceso "se extrae un rey" ¿Los sucesos A y B son independientes?

☐ a) No son independientes porque en la baraja hay un rey de oros.
☐ b) Sólo si sacamos primero una carta para ver si es rey y se vuelve a colocar en la baraja. Luego sacamos una segunda carta para mirar si es oros.
☐ c) No, porque P(rey de oros)= P(rey) x P(oros)
☐ d) Sí, en todos los casos

Ítem 5. Una caja tiene cuatro focos, de los cuales dos son defectuosos. Se sacan dos al azar, uno tras otro sin reemplazamiento. Si el primer foco fue defectuoso, ¿qué es más probable?

☐ a) que el segundo sea defectuoso
☐ b) que el segundo no sea defectuoso
☐ c) los dos sucesos son iguales de probables

Ítem 6. En un barrio de Granada se ha realizado una entrevista a un grupo de hombres, obteniendo los siguientes resultados:

	menos de 55 años	más de 55 años	Total
Ha sufrido un ataque al corazón	29	75	104
No ha sufrido ataque	401	275	676
Total	430	350	780

Si elegimos al azar una de estas personas:

a) ¿Cuál es la probabilidad de que haya tenido un ataque al corazón? ___________________
b) ¿Cuál es la probabilidad de que tenga más de 55 años y además haya tenido uno o más ataques al corazón? ____
c) Dado que la persona escogida tiene más de 55 años, ¿cuál es la probabilidad de que haya tenido un ataque al corazón? ____
d) Dado que la persona escogida ha tenido un ataque al corazón, ¿cuál es la probabilidad de que tenga más de 55 años? ____

Ítem 7. La probabilidad de que una mujer de 40 años tenga una mamografía positiva es el 10.3 %. La probabilidad de que una mujer de 40 años tenga cáncer de pecho y una mamografía positiva es de 0.8%. Una mujer de 40 años se realiza una mamografía y resulta positiva. ¿Cuál es la probabilidad de que realmente tenga cáncer?

- ☐ a) $\frac{0,8}{10,3} = 0,0776$, probabilidad de 7,76%
- ☐ b) $10,3 \times 0,8 = 8,24$, probabilidad del 8,24%
- ☐ c) 0,8 %

Ítem 8. Hemos lanzado dos dados y sabemos que el producto de los dos números obtenidos ha sido 12 ¿Cual es la probabilidad de que ninguno de los dos números sea un 6?

Ítem 9. Supón que Carlos Ferrero alcanza la final de Roland Garros en 2004. Para ganar el partido hay que ganar tres sets. ¿Cuál de los siguientes sucesos consideras más probable?

- ☐ a) Carlos Ferrero pierde el primer set.
- ☐ b) Carlos Ferrero pierde el primer set pero gana el partido.
- ☐ c) Los dos sucesos son iguales de probables.

Ítem 10. Un test diagnóstico de cáncer fue administrado a todos los residentes de una gran ciudad. Un resultado positivo en el test es indicativo de cáncer y un resultado negativo es indicativo de ausencia de cáncer. ¿En cuál de las siguientes predicciones tienes más confianza?

- ☐ a) Predecir que una persona tiene cáncer si ha dado positivo en el test de diagnóstico.
- ☐ b) Predecir un resultado positivo en el test de diagnóstico si la persona tiene cáncer.
- ☐ c) Tengo la misma confianza en ambas predicciones.

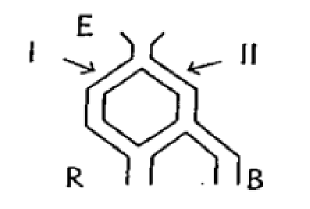
Ítem 11. En una ciudad hay 60 hombres y 40 mujeres por cada 100 habitantes. 50 de cada 100 hombres y 25 de cada 100 mujeres fuman. Si cogemos 200 personas al azar, ¿cuántas de estas 200 personas fumarán?

Ítem 12. Una persona lanza un dado y anota si saca un número par o impar. Estos son los resultados al lanzarlo 15 veces: par, impar, impar, par, par, impar, par, par, par, par, impar, impar, par, par, par
Lanza el dado de nuevo ¿Cuál es la probabilidad de sacar un número par en esta nueva tirada?

Ítem 13. Un grupo de alumnos de un colegio se examina de Matemáticas e Inglés. La proporción de alumnos que aprueban Matemáticas es del 80% y la proporción de alumnos que aprueba Inglés es del 70%. Suponiendo que las notas en cada una de las asignaturas es independiente, ¿Cuál es la probabilidad de que un alumno escogido al azar haya aprobado ambas asignaturas?

Ítem 14. Una bola se suelta por la entrada E. Si sale por R, ¿Cuál es la probabilidad de que haya pasado por el canal 1?

- ☐ a) 0,50
- ☐ b) 0,33
- ☐ c) 0,66
- ☐ d) No se puede calcular



Ítem 15. En una encuesta publicada en un periódico se informa que el 91% de los habitantes de una ciudad mienten usualmente y de ellos el 36% mienten sobre cosas importantes. Si elegimos al azar a una persona de esta ciudad, ¿cuál es la probabilidad de que mienta sobre cosas importantes?

Ítem 16. Una fábrica dispone de dos máquinas M1 y M2 que fabrican bolas. La máquina M1 fabrica el 40 % de las bolas y la M2 el 60%. El 5% de las bolas fabricadas por M1 y el 1% de las fabricadas por M2 son defectuosas. Tomamos una bola al azar que resulta ser defectuosa. ¿Cuál es la probabilidad de que haya sido fabricada por M1?

Ítem 17. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin reemplazamiento. ¿Cuál es la probabilidad de extraer una bola negra en segundo lugar, habiendo extraído una bola negra en primer lugar? $P\ (N_2/\ N_1)$

- [ ] a) 1/ 2
- [ ] b) 1/ 6
- [ ] c) 1/ 3
- [ ] d) 1/ 4

¿Cuál es la probabilidad de extraer una bola negra en primer lugar, sabiendo que hemos extraído una bola negra en segundo lugar? $P\ (N_1/\ N_2)$

- [ ] a) 1/ 3
- [ ] b) No se puede calcular
- [ ] c) 1/ 6
- [ ] d) 1/ 2

Ítem 18. Se tiene una bolsa con 1 bola azul y 2 bolas rojas. Se saca una bola al azar. Sin reponerla otra vez en la bolsa se saca una segunda bola. ¿Cuál suceso es más probable?

- [ ] a) Sacar dos bolas rojas
- [ ] b) Sacar primero una bola roja y luego una azul.
- [ ] c) Los dos sucesos son iguales de probables

5.3. ANÁLISIS

Una vez recogidos los datos, se codificaron y se analizaron utilizando el paquete estadístico SPSS y algunas subrutinas de cálculo bayesiano que se programaron en Excel. En los ítems de respuestas abiertas se usó el crédito parcial[32] puesto que la respuesta del alumno nos permitió el acceso no sólo a la solución final, sino a los procedimientos empleados en su solución, y posibles errores. En consecuencia se usó un criterio de calificación parcial (y no simplemente dicotómico) es decir dar una puntuación a la respuesta basada en su grado de completitud y precisión.

El análisis de los datos constó de las siguientes partes: análisis de ítems y aproximación a la fiabilidad, generalizabilidad y validez.

Análisis de ítems

Se analizaron, en primer lugar, las características psicométricas de los ítems, dificultad y discriminación siguiendo la teoría clásica de los tests (Muñiz, 1994; Martínez Arias, 1995; Barbero, 2003).

[32] El *crédito parcial* es deseable en muchas situaciones de evaluación en las que se desea conocer el conocimiento parcial de los examinados. Su principal ventaja es que puede contribuir a aumentar la fiabilidad de un instrumento de longitud restringida, como es en nuestro caso (Millman y Greene, 1989).

Puesto que el índice de dificultad es la proporción de alumnos que superan el ítem se calcularon adicionalmente los intervalos de confianza de dicha proporción utilizando la aproximación normal, ya que las dos muestras contaban con más de 30 elementos. . En los ítems de respuestas abiertas, además de la respuesta totalmente correcta (máxima puntuación), también se consideraron correctas (a efecto del cálculo del índice de dificultad) las respuestas con razonamiento e identificación de los datos correctos y ligeros errores de cálculo.

Una aportación metodológica es el uso de inferencia bayesiana para mejorar la estimación de los índices de dificultad y discriminación. Esta estimación se ha realizado con dos tipos de distribución a priori (Bernard, 1998):

- Distribución a priori no informativa, es decir, sin tener en cuenta la información previa (Lecoutre, 1996). En este caso, el intervalo de credibilidad se calcula mediante la siguiente expresión,

$$\left[\beta^{-1}_{0,5+p,\,0,5+q}(\alpha/2) - \beta^{-1}_{0,5+p,\,0,5+q}(1-\alpha/2)\right]$$

donde $\beta^{-1}_{a,b}$ es la función de distribución inversa de la distribución *Beta (a,b), e y f* los éxitos y fracasos en la muestra y α el coeficiente de credibilidad.

- Utilizando resultados del ítem en las pruebas pre-piloto, siguiendo uno de los métodos recomendados en Serrano (2003), que consiste en tener en cuenta los éxitos y fracasos en la prueba anterior para definir la distribución a priori. En este caso, el inter*valo de credibilidad se calcula mediante la siguiente expresión, donde ahora e' y f'* son los éxitos y fracasos en la distribución a priori.

$$\left[\beta^{-1}_{0,5+e+e',\,0,5+f+f'}(\alpha/2) - \beta^{-1}_{0,5+e+e',\,0,5+f+f'}(1-\alpha/2)\right]$$

En ambos casos se utilizan las funciones beta como distribuciones a priori, que permiten obtener distribuciones a posteriori conjugadas (Berry, 1995; Gelman, Carlin, Stern y Rubin, 2003). Se calcularon los intervalos de credibilidad (Corroyer y Wolf, 2003; Lee, 2003; Bolstad, 2004) y se compararon con los intervalos de confianza obtenidos en la aproximación clásica. Todo el procedimiento se justifica con mayor detalle en el Estudio 7.

También se han calculado los índices de discriminación para cada uno de los ítems por dos sistemas diferentes (Muñiz, 1994):

1. Como correlación entre el ítem y la puntuación total del cuestionario.
2. Mediante el estudio de la diferencia en proporción de aciertos al ítem en los estudiantes de psicología respecto a un grupo de mayor preparación (estudiantes de matemáticas, con alta formación en matemáticas y estadística), aplicando el test usual de diferencia de proporciones.

Para los dos casos, se han realizado estimaciones clásicas y bayesianas. Las estimaciones clásicas vienen dadas por intervalos de confianza de la diferencia de proporciones. Puesto que la muestra tiene tamaño suficiente se usó la aproximación normal:

$$(p_s - p_i) \pm \left(Z^{-1}{}_{1-\alpha/2} \sqrt{\frac{p_s(1-p_s)}{n_s} + \frac{p_s(1-p_s)}{n_s}} \right)$$

siendo Z la variable normal tipificada y α el coeficiente de confianza

Las estimaciones bayesianas sólo se calculan para el caso de distribución inicial no informativa, pues es la primera vez que se calculan estos índices y no se tenía información previa disponible. Siguiendo a Berry (1995), los estimadores de las medias de la distribuciones finales serán:

$$\hat{p}_s = \frac{a'_s}{a'_s + b'_s} \qquad \hat{p}_i = \frac{a'_i}{a'_i + b_i'}$$

$$\hat{p}^{+}{}_s = \frac{a'_s + 1}{a'_s + b'_s + 1} \qquad \hat{p}^{+}{}_i = \frac{a'_i + 1}{a'_i + b_i' + 1}$$

Los estimadores de las desviaciones típicas de las distribuciones finales:

$$\partial_s = \sqrt{\hat{p}_s(\hat{p}^{+}{}_s - \hat{p}_s)} \qquad \partial_i = \sqrt{\hat{p}_i(\hat{p}^{+}{}_i - \hat{p}_i)}$$

En consecuencia, la diferencia de proporciones tendrá una distribución normal aproximada $N(\widehat{p}_s - \widehat{p}_i, \sqrt{t_s^2 + t_i^2})$, por lo que el intervalo aproximado de credibilidad vendrá dado por:

$$\widehat{p}_s - \widehat{p}_i \pm Z_{1-\alpha/2}^{-1}\sqrt{\partial_s^2 + \partial_i^2}$$

siendo *Z* la distribución normal *N (0,1)* y tomando las distribuciones iniciales como *B (0,5, 0,5)* en ambos grupos, en consistencia con el caso de una proporción.

Para la estimación bayesiana del índice de discriminación como correlación ítem total, se usa la transformación de Fischer (Lee, 2002), sustituyendo $\rho = \tanh \xi; r = \tanh z$, de donde se obtiene una aproximación de la distribución final de dicha transformada, mediante la distribución normal:

$$\xi \sim N(z, 1/n)$$

Esta aproximación se usa para calcular los intervalos de credibilidad para la tangente hiperbólica del coeficiente de correlación y a partir de estos intervalos, deshaciendo el cambio de variable se determinan los correspondientes al coeficiente de correlación.

El análisis de ítems se completa con el estudio de la distribución de los índices de dificultad, representando diversos gráficos y estadísticos descriptivos, así como las estimaciones clásica y bayesiana de la media de dicha distribución. En este último caso, se usa la distribución inicial no informativa

Los valores de la media y desviación típica de la distribución final vienen dados por las fórmulas siguientes:

$$(1) \qquad \mu_f = \frac{\frac{n\bar{x}}{s^2} + \frac{\mu_i}{\sigma_i^2}}{\frac{n}{s^2} + \frac{1}{\sigma_0^2}} \qquad\qquad \sigma_f = \frac{1}{\sqrt{n/s^2 + 1/\sigma_i^2}}$$

En las expresiones anteriores *n* es el tamaño de la muestra, $\bar{x}$ y *s* la media y desviación típica de la muestra. La distribución será *T* con *n-1* grados de libertad, siendo *n* el tamaño de muestra (Bolstad, 2004).

El intervalo de credibilidad para un coeficiente de credibilidad α viene dado por:

$(\bar{x} - T_{1-(1-\alpha)/2} s_f / \sqrt{n}; \bar{x} + T_{1-(1-\alpha)/2} s_f / \sqrt{n})$, siendo *T* un percentil de la distribución *T* con *n-1* grados de libertad.

Aproximación a la fiabilidad

Puesto que se está en un proceso de medida en educación, el objetivo es realizar inferencias sobre conceptos abstractos a partir de indicadores empíricos, más precisamente, relacionar los conocimientos de los alumnos sobre un concepto con sus respuestas a los ítems del cuestionario (Carmines y Zeller, 1979; Thorndike, 1989; Meliá, 2001).

La medida siempre produce un cierto error aleatorio, pero dos medidas del mismo fenómeno sobre un mismo individuo suelen ser consistentes, siendo la fiabilidad la tendencia a la consistencia o precisión del instrumento en la población medida (Bisquerra, 1989; Barbero, 1993; Melía, 2001)[33]. Una vez realizado el análisis de ítems, se buscaron algunas evidencias de la fiabilidad, siendo conscientes que, al estar todavía en la fase piloto, estas evidencias serían provisionales y deberían ser completadas en el futuro. Se partió para ello de la teoría clásica de los tests (Muñiz, 1994; Martínez Arias, 1995).

Entre los diversos procedimientos para el cálculo del estimador del coeficiente de fiabilidad se tomó el coeficiente Alfa de Cronbach (que se reduce al de Kuder-Richardson para ítems dicotómicos) (Carmines y Zeller, 1979; Meliá, 2001; Díaz, Batanero y Cobo, 2003)[34]. Se calculó este coeficiente a partir de la muestra total de estudiantes de psicología (*n*=57) mediante el programa correspondiente del paquete estadístico SPSS, analizando los estadísticos de cada ítem si se suprime del instrumento y estudiando el efecto sobre el coeficiente de ir suprimiendo sucesivamente los ítems que presentan peores resultados.

[33] En Teoría Clásica de Tests se define como correlación entre las puntuaciones verdadera y observada (Martínez Arias, 1995).

[34] Entre otras propiedades: a) refleja el grado en el que covarían los ítems que constituyen el test; b) es el valor medio de todos los que se obtendrían con el método de las dos mitades si se utilizasen todas las combinaciones de ítems; c) es cota inferior de la que se obtendría por el método de la prueba repetida; y d) estima la *fiabilidad en el acto* (Martínez Arias, 1995).

Fiabilidad de la codificación de los datos

Se analizó también la fiabilidad de la codificación[35] de las respuestas abiertas mediante porcentajes de acuerdo en la codificación (Fox, 1981; León y Montero, 2002). Para llevar a cabo esta estimación dos personas distintas aplicaron el procedimiento de codificación y posteriormente se calcularon los índices.

El principal problema podría venir dado en los ocho ítems con respuestas abiertas, en los cuáles había que otorgar una puntuación, en función de la completitud de la respuesta. Para estimar la fiabilidad con la cual se llevaba a cabo este proceso de codificación se pidió a un segundo investigador, que codificase los ítems de respuesta abierta, siguiendo el esquema dado con el investigador y se llevó a cabo un estudio de la consistencia por dos procedimientos. Se calculó el porcentaje de coincidencias (la misma codificación exacta) de los dos investigadores en cada una de las preguntas y en el global de los datos.

Aproximación a la validez

Se entiende por validez un concepto unitario, considerando sus diferentes formas (validez de constructo, contenido, criterio...) como diferentes maneras de recoger evidencias de una única noción de validez (Messick, 1989, 1995, 1998; Martínez Arias, 1995; Melía, 2001). Asimismo entendemos la validez en relación al uso que se dará al cuestionario y la interpretación de sus puntuaciones[36].

La validación de un cuestionario es un proceso complejo de recogida y documentación de evidencia (Martínez Arias, 1995; APA, AERA y NCME, 1985; 1999; Carmona, 2004)[37]. En el estudio piloto se inició dicho proceso[38] que se completará en el Estudio 6.

En primer lugar se inició el estudio de la validez de contenido para asegurar que el instrumento recoge una muestra representativa de los contenidos que se pretenden evaluar (Carmines y Zeller, 1979; Morales, 1988; Messick, 1989, 1995, 1998). Para ello se hizo una planificación cuidadosa de los ítems que se incluirán en el mismo y un

[35] El concepto de fiabilidad también se aplica al análisis de datos, donde se refiere a la exactitud, estabilidad y precisión (Fox, 1981).

[36] No se valida un test sino una inferencia o interpretación (Osterlind, 1989).

[37] La importancia de la validez se resalta también en los Standards for Educational and Psychological Testing (APA, AERA, NCTM, 1999).

[38] *"Extensión por la cual las medidas son útiles para tomar decisiones relevantes a un propósito dado"* (Sax, 1989, p. 292).

estudio de cómo estos ítems contribuyen a la medida del constructo subyacente. Para salvar la dificultad de determinar qué es un muestreo adecuado de ítems en el dominio particular, se describió de antemano el dominio, sus dimensiones, facetas y objetivos en el Estudio 1 (Martínez Arias, 1995).

También se ha usado una metodología adecuada para revisar la congruencia entre los ítems y las especificaciones (Osterlind, 1989)[39]. Como es usual en este tipo de validación se usó el juicio de expertos, que emparejaron los ítems con objetivos y resumiendo la información numéricamente. Esta información se presentó en el Estudio 2.2.

5.4. RESULTADOS

Recogidos los datos, se codificaron y se analizaron utilizando el paquete estadístico SPSS (versión 12.0) y algunas subrutinas programadas en Excel. A continuación se presenta el informe psicométrico estructurado en los siguientes apartados:

- Índices de dificultad de los ítems.
- Índices de discriminación.
- Estudio de las características globales del cuestionario (estimación de la fiabilidad por el método de consistencia interna y evidencias de validez).

5.4.1. INDICES DE DIFICULTAD

En la tabla 4.35 se presentan en los índices de dificultad en la muestra de estudiantes de psicología, junto con los intervalos de confianza correspondientes, cuya amplitud se explica por el tamaño reducido de la muestra.

Se observa que algunos ítems (ítem 4, 14, 15, 16) son todavía muy difíciles para estos alumnos, que ya habían estudiado la probabilidad condicional, mientras que en otros se obtienen resultados razonables o incluso muy buenos (ítems 1, 2, 5, 6a, 18). En el resto de los ítems se obtienen resultados moderados.

[39] El establecimiento de la validez mediante la determinación de la correspondencia entre ítems y objetivos se aplica especialmente a los tests de rendimiento (Sax, 1985), como es el caso del cuestionario RPC.

Tabla 4.35. Índices de dificultad e intervalos de confianza en estudiantes de psicología (*n*=57)

	Índice	I. confianza 95%	
I1	0,860	0,7667	0,9526
I2	0,720	0,5990	0,8396
I3	0,281	0,1604	0,4010
I4	0,175	0,0736	0,2773
I5	0,825	0,7227	0,9264
I6A	0,877	0,7893	0,9651
I6B	0,298	0,1758	0,4207
I6C	0,421	0,2889	0,5532
I6D	0,421	0,2889	0,5532
I7	0,298	0,1758	0,4207
I8	0,263	0,1453	0,3810
I9	0,333	0,2071	0,4595
I10	0,316	0,1914	0,4402
I11	0,386	0,2556	0,5163
I12	0,596	0,4652	0,7278
I13	0,456	0,3228	0,5895
I14	0,070	-0,0071	0,1124
I15	0,193	0,0873	0,2986
I16	0,193	0,0873	0,2986
I7A	0,474	0,3400	0,6073
I7B	0,263	0,1453	0,3810
I18	0,719	0,5990	0,8396

Figura 4.3. Distribución de índices

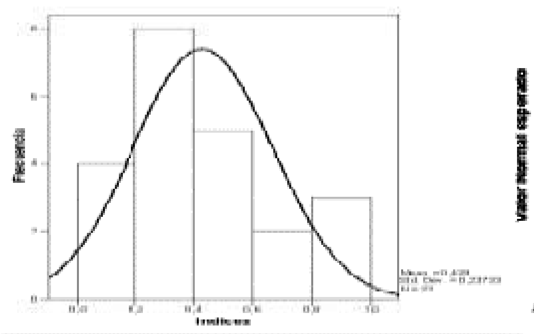


Figura 4.4. Gráfico Q-Q de índices dificultad

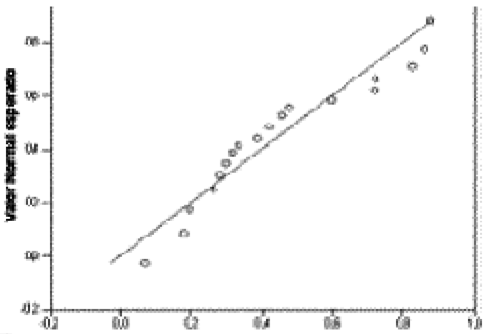


Tabla 4.36. Estadísticos descriptivos de la distribución de índices de dificultad

		Estadístico	Error típ.
Media		0,4290	0,0506
Intervalo de confianza para la media al 95%	Límite inferior	0,3238	
	Límite superior	0,5342	
Intervalo credibilidad para la media al 95%	Límite inferior	0,3238	
	Límite superior	0,5342	
Mediana		0,3595	
Varianza		0,056	
Desv. típ.		0,2373	
Amplitud intercuartil		0,360	
Asimetría		0,696	0,491
Curtosis		-0,599	0,953

En las figuras 4.3 y 4.4 se presenta la distribución de estos índices, que abarcan todo el posible rango de variación, con algo más de concentración en el intervalo 0,2- 0,4 y una normalidad aproximada, como se pone también de manifiesto en los estadísticos y prueba de Kolmogorov Smirnov (tablas 4.36 y 4.37).

Tabla 4.37. Prueba de Kolmogorov-Smirnov para una muestra (22 índices)

Parámetros normales(a,b)	Media	0,429
	Desviación típica	0,237
Diferencias más extremas	Absoluta	0,157
	Positiva	0,157
	Negativa	-0,116
Z de Kolmogorov-Smirnov		0,737
Sig. asintót. (bilateral)		0,649

Estimación de índices de dificultad mediante métodos bayesianos

En la tabla 4.38 se presentan los estimadores bayesianos e intervalos de credibilidad del 95% para los índices de dificultad. Se utiliza una distribución inicial no informativa *Be (0,5, 0,5)* por lo que la estimación del índice varía ligeramente respecto a la clásica.

Tabla 4.38. Estimación Bayesiana con Distribución a priori no informativa

Ítem	Aciertos/ fallos en prueba piloto	Estimación bayesiana del índice de dificultad	
	(n=57)	Valor medio	Intervalo de credibilidad (95%)
I1	(49,8)	0,853	0,753-0,931
I2	(41,16)	0,716	0,594-0,823
I3	(16,41)	0,284	0,177-0,406
I4	(10, 47)	0,171	0,094-0,289
I5	(47,10)	0,819	0,711-0,906
I6A	(50,7)	0,871	0,774-0,943
I6B	(17,40)	0,299	0,192-0,425
I6C	(24, 33)	0,422	0,264-0,550
I6D	(24,33)	0,422	0,264-0,550
I7	(17,40)	0,302	0,192-0,425
I8	(15,42)	0,267	0,163-0,387
I9	(19,38)	0,336	0,222-0,462
I10	(18,39)	0,319	0,195-0,443
I11	(22,35)	0,389	0,268-0,515
I12	(34,23)	0,595	0,467-0,717
I13	(26,31)	0,457	0,332-0,585
I14	(4,53)	0,070	0,024-0,158
15	(11,46)	0,198	0,107-0,309
I16	(11,46)	0,199	0,107-0,309
I17A	(27,30)	0,474	0,348-0,602
I17B	(15,42)	0,213	0,116-0,330
I18	(41,16)	0,716	0,594-0,823

No obstante, la variación, así como la de los intervalos es muy pequeña; los resultados clásicos y bayesianos coinciden para este caso cuando la distribución inicial es no informativa (Lecoutre, 1996).

La prueba del instrumento piloto en estudiantes de psicología fue realizada con muestras de estudiantes del mismo curso y universidad que los estudiantes de psicología que colaboraron en las pruebas pilotos de ítems. Se cuenta, por tanto, para esta muestra, con una información previa sobre la posible dificultad de estos ítems, por lo que se podrían aplicar técnicas bayesianas de estimación de proporciones para mejorar la estimación obtenida de los índices de dificultad. Se usaron los datos de las pruebas pilotos de ítems (anexo 2) para construir la distribución a priori del índice de dificultad de dichos ítems presentamos en la tabla 4.39.

Tabla. 4.39. Estimación Bayesiana con Distribución a priori informativa

Ítem	Índice observado en prueba piloto (n=57)	(Éxitos/ fallos) en Prueba prepiloto	Estimación bayesiana del índice de dificultad	
			Valor medio	Intervalo de credibilidad (95%)
I1	0,860	(37, 12)	0,808	0,729-0,877
I2	0,720	No probado	---	---
I3	0,281	(23, 134)	0,185	0,135-0,238
I4	0,175	(26, 131)	0,170	0,123-0,223
I5	0,825	(45, 31)	0,690	0,610-0,765
I6A	0,877	(23, 26)	0,688	0,596-0,771
I6B	0,298	(19, 30)	0,341	0,261-0,433
I6C	0,421	(12, 37)	0,342	0,255-0,433
I6D	0,421	(35, 14)	0,556	0,462-0,649
I7	0,298	(17, 40)	0,300	0,220-0,386
I8	0,263	(14, 35)	0,276	0,196-0,364
I9	0,333	(30, 51)	0,356	0,279-0,437
I10	0,316	(50, 107)	0,319	0,258-0,382
I11	0,386	(35, 46)	0,414	0,333-0,496
I12	0,596	(36, 13)	0,659	0,567-0,745
I13	0,456	(11,41)	0,341	0,256-0,432
I14	0,07	(17,59)	0,160	0,104-0,227
I15	0,193	(11,26)	0,237	0,157-0,327
I16	0,193	(21,31)	0,295	0,214-0,384
I17A	0,474	(46,35)	0,529	0,446-0,611
I17B	0,26	(25, 66)	0,252	0,185-0,325
I18	0,72	(110,47)	0,715	0,654-0,772

Al compararla con la 4.38 se ve no sólo que la amplitud de los intervalos de credibilidad disminuye sino que hay ligeros cambios en el índice de dificultad, puesto que se tiene en cuenta la información previa. Se obtiene una estimación más precisa.

5.4.2. INDICES DE DISCRIMINACIÓN

También se ha calculado los índices de discriminación (tabla 4.38) como correlación corregida entre el ítem y la puntuación total del cuestionario, calculada a partir del programa escalas en el paquete estadístico SPSS. Todos los ítems tienen correlación positiva.

En algunos ítems la correlación no llega a ser estadísticamente significativa e incluso en otros es prácticamente inexistente. Esto puede ser debido, por un lado al pequeño tamaño de muestra y también a que los sesgos sobre razonamiento condicional aparecen incluso en los alumnos con buenos resultados en el total del cuestionario. No obstante, esto es un punto a mejorar en futuras versiones del cuestionario.

Tabla 4.40. Índices de discriminación (n = 57)

	Correlación corregida con total	Intervalo confianza 95%		Intervalo credibilidad 95%	
		Límite inferior	Límite superior	Límite inferior	Límite superior
I1	***0,687	0,53	0,81	0,525	0,801
I2	**0,297	0,04	0,52	0,047	0,512
I3	*0,202	-0,06	0,44	-0,055	0,434
I4	0,036	-0,23	0,29	-0,220	0,287
I5	0,107	-0,16	0,36	-0,151	0,351
I6A	*0,205	-0,06	0,44	-0,052	0,436
I6B	*0,215	-0,05	0,46	-0,041	0,445
I6C	**0,371	0,12	0,58	0,129	0,571
I6D	**0,345	0,09	0,56	0,100	0,551
I7	0,065	-0,20	0,32	-0,192	0,314
I8	***0,526	0,31	0,70	0,314	0,688
I9	0,046	-0,22	0,31	-0,210	0,296
I10	0,101	-0,17	0,36	-0,157	0,346
I11	***0,633	0,45	0,77	0,452	0,764
I12	***0,677	0,51	0,80	0,511	0,794
I13	***0,734	0,59	0,83	0,590	0,833
I14	0,048	-0,22	0,31	-0,208	0,298
I15	***0,568	0,37	0,73	0,367	0,718
I16	***0,490	0,27	0,66	0,270	0,662
I17A	**0,330	0,08	0,55	0,083	0,539
I17B	0,033	-0,23	0,29	-0,223	0,285
I18	0,005	-0,26	0,27	-0,249	0,259

* Significativa al nivel 0,05; ** Significativa al nivel 0,01; *** Significativa al nivel 0,001(unilateral).

Otra forma empleada para el cálculo de los índices de discriminación ha sido el estudio de la diferencia en proporción de aciertos al ítem en los estudiantes de psicología respecto a un grupo de mayor preparación (estudiantes de matemáticas, con alta formación en matemáticas y estadística) (tabla 4.41). Se aplica el test usual de diferencia de proporciones. En este caso la discriminación no es tan clara, pues son pocos los ítems que muestran una diferencia estadísticamente significativa y en algunos casos esta diferencia favorece al grupo con menos preparación teórica (psicólogos).

La explicación de este hecho es que los matemáticos han tenido una preparación muy formal, consistente sólo en la parte matemática, sin tener tan en cuenta las aplicaciones y posibles aspectos psicológicos de la probabilidad condicional por lo que han fallado en muchos ítems referidos a sesgos de razonamiento.

Tabla 4.41. Índices de discriminación, valores tipificados e intervalos confianza 95%

	n = 37	*n* =57	Valor diferencia	Valor tipificado	Lim Inf	Lim Sup
I1	0,784	0,860	-0,076	-0,93	-0,24	0,08
I2	0,784	0,720	0,064	0,71	-0,11	0,24
I3	0,108	0,281	-0,173	-2,21	-0,33	-0,02
I4	0,081	0,175	-0,094	-1,39	-0,23	0,04
I5	0,811	0,825	-0,014	-0,17	-0,17	0,15
I6A	0,946	0,877	0,069	1,21	-0,04	0,18
I6B	0,459	0,298	0,161	1,58	-0,04	0,36
I6C	0,730	0,421	***0,309	3,15	0,12	0,50
I6D	0,730	0,421	***0,309	3,15	0,12	0,50
I7	0,432	0,298	0,134	1,32	-0,06	0,33
I8	0,513	0,263	**0,25	2,48	0,05	0,45
I9	0,270	0,333	-0,065	-0,66	-0,25	0,13
I10	0,378	0,316	0,062	0,62	-0,14	0,26
I11	0,567	0,386	*0,181	1,74	-0,02	0,38
I12	0,757	0,596	0,161	1,68	-0,03	0,35
I13	0,865	0,456	***0,409	4,72	0,24	0,58
I14	0,081	0,07	0,011	0,20	-0,10	0,12
I15	0,297	0,193	0,104	1,14	-0,08	0,28
I16	0,216	0,193	0,023	0,27	-0,10	0,15
I17A	0,703	0,474	**0,229	2,29	0,05	0,41
I17B	0,135	0,263	-0,128	-1,58	-0,29	0,03
I18	0,757	0,719	0,038	0,41	-0,29	0,37

* Significativa al nivel 0,05; ** Significativa al nivel 0,01;*** Significativa al nivel 0,001(unilateral)

Figura 4.5. Comparación ítem a ítem

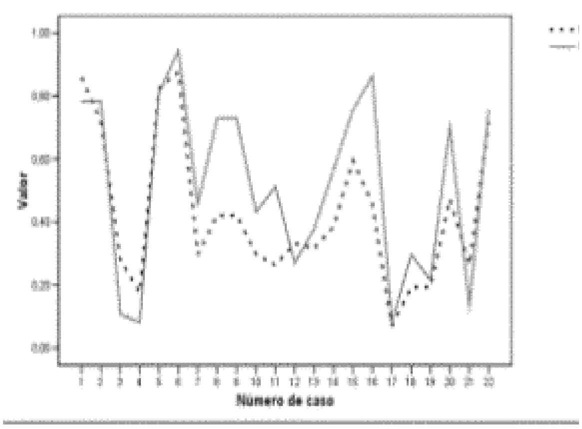


Figura 4.6. Comparación distribución de ítems

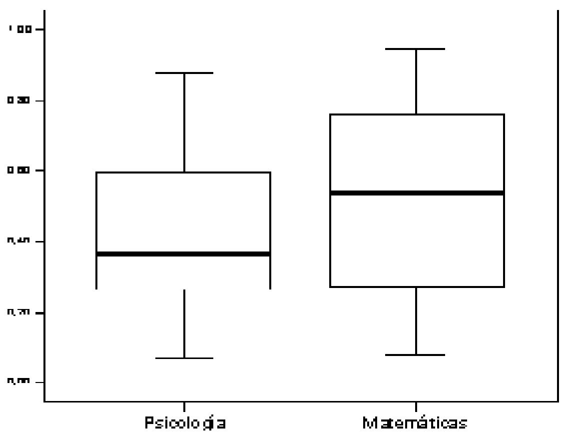


Tabla 4.42. Índices de discriminación e intervalos credibilidad 95%, distribución inicial no informativa

	n = 37	n =57	Valor diferencia	Lim Inf	Lim Sup
I1	0,784	0,860	-0,076	-0,24	0,08
I2	0,784	0,720	0,064	-0,11	0,24
I3	0,108	0,281	-0,173	-0,33	-0,02
I4	0,081	0,175	-0,094	-0,23	0,04
I5	0,811	0,825	-0,014	-0,17	0,15
I6A	0,946	0,877	0,069	-0,04	0,18
I6B	0,459	0,298	0,161	-0,04	0,36
I6C	0,730	0,421	***0,309	0,12	0,50
I6D	0,730	0,421	***0,309	0,12	0,50
I7	0,432	0,298	0,134	-0,06	0,33
I8	0,513	0,263	**0,25	0,05	0,45
I9	0,270	0,333	-0,065	-0,25	0,13
I10	0,378	0,316	0,062	-0,14	0,26
I11	0,567	0,386	*0,181	-0,02	0,38
I12	0,757	0,596	0,161	-0,03	0,35
I13	0,865	0,456	***0,409	0,24	0,58
I14	0,081	0,07	0,011	-0,10	0,12
I15	0,297	0,193	0,104	-0,08	0,28
I16	0,216	0,193	0,023	-0,10	0,15
I17A	0,703	0,474	**0,229	0,05	0,41
I17B	0,135	0,263	-0,128	-0,29	0,03
I18	0,757	0,719	0,038	-0,29	0,37

No obstante, al comparar las distribuciones de los índices de dificultad en las dos muestras, se ve una ligera diferencia a favor de los matemáticos en el valor medio de todos estos índices (véase tabla 4.41 y 4.42 donde se comparan estas distribuciones). Resultados similares se obtienen en las comparaciones gráficas, ítem a ítem o de toda la distribución (figuras 4.5 y 4.6).

Tabla 4.43. Estadísticos de muestras onadas

	Media	N	Desviación típ.	Error típ. de la media
Psicología	0,4290	22	0,2373	0,0506
Matemáticas	0,5184	22	0,2845	0,0606

Tabla 4.44. Prueba de muestras independientes /Psicología - Matemáticas

Media	Desviación típ.	Error típ. de la media	Intervalo de confianza diferencia (95%)		t	gl	Sig. (bilateral)
			Inferior	Superior			
-0,08936	0,1528	0,0325	-0,1571	-0,0215	-2,74	21	0,012

Tabla 4.45. Estimación bayesiana de índices de discriminación, e intervalos credibilidad 95%

	n = 37	n =57	Valor diferencia	Lim Inf	Lim Sup
I1	0,784	0,860	-0,077	-0,286	0,132
I2	0,784	0,720	0,061	-0,168	0,290
I3	0,108	0,281	-0,166	-0,368	0,036
I4	0,081	0,175	-0,089	-0,265	0,087
I5	0,811	0,825	-0,016	-0,225	0,193
I6A	0,946	0,877	0,081	-0,076	0,237
I6B	0,459	0,298	0,159	-0,098	0,416
I6C	0,730	0,421	0,301	0,053	0,549
I6D	0,730	0,421	0,301	0,053	0,549
I7	0,432	0,298	0,123	-0,130	0,376
I8	0,513	0,263	0,237	-0,015	0,489
I9	0,270	0,333	-0,086	-0,325	0,153
I10	0,378	0,316	0,072	-0,186	0,329
I11	0,567	0,386	0,178	-0,084	0,440
I12	0,757	0,596	0,155	-0,088	0,398
I13	0,865	0,456	0,398	0,177	0,620
I14	0,081	0,070	0,015	-0,135	0,164
I15	0,297	0,193	0,104	-0,128	0,336
I16	0,216	0,193	0,025	-0,192	0,243
I17A	0,703	0,474	0,223	-0,030	0,476
I17B	0,135	0,263	-0,123	-0,330	0,085
I18	0,757	0,719	0,034	-0,200	0,269

En la tabla 4.45 se presentan los estimadores bayesianos e intervalos de credibilidad del 95% para los índices de discriminación. Se utiliza una distribución inicial no informativa *Be(0.5; 0,5)* por lo que la estimación del índice varía ligeramente respecto a la clásica. No obstante la variación, así como la de los intervalos es muy pequeña; como se sabe los resultados clásicos y bayesianos coinciden para este caso cuando la distribución inicial es no informativa (Lecoutre, 1996; Lee, 2003; Bolstad, 2004).

5.4.3. FIABILIDAD

Una vez realizado el análisis de ítems, se ha tratado de hacer una estimación de la fiabilidad y de buscar algunas evidencias validez de la prueba, siendo conscientes que, al estar todavía en las fases iniciales de la construcción, estas evidencias son sólo provisionales y deberán ser completadas en el futuro.

Fiabilidad de consistencia interna

Se ha obtenido el coeficiente de consistencia interna alfa de Cronbach que se basa en el análisis relativo de la varianza total del cuestionario y de las varianzas de los ítems particulares. En el caso de este cuestionario, puesto que la prueba es heterogénea no hay que esperar un índice de consistencia interna muy alto. Se ha calculado este coeficiente a partir de la muestra total de estudiantes (n = 94) y con la muestra de psicología (n=57) y de matemáticas (n = 37) independientemente. Para el cálculo del coeficiente alfa se ha utilizado el programa correspondiente del paquete estadístico SPSS. Siendo conscientes de que, como señala Martínez Arias (1995) a mayor tamaño de muestra el coeficiente obtenido será más preciso, se repetirá el cálculo con nuevas muestras de sujetos en el Estudio 5. No obstante, se cumplen las otras condiciones requeridas por la autora para el cálculo del coeficiente de fiabilidad: representatividad de la muestra, independencia de las respuestas a distintos ítems y administración idéntica al uso pretendido del test. En la tabla 4.46 se muestran los resultados obtenidos con el total de ítems del cuestionario.

El valor obtenido con el total de la muestra es de 0,7738, que corresponde a una correlación entre la puntuación observada y la puntuación verdadera de 0,879. Se considera que el valor es razonable, dado que se trata de un cuestionario piloto que puede ser mejorado. Al calcular separadamente el coeficiente de fiabilidad para ambos grupos se obtuvo un valor de α = 0,6043 para el grupo de psicología (n = 57) y un valor α de 0,5604 en el grupo de matemáticas (n =37).

Se observa que el coeficiente de fiabilidad para los grupos separados tiene un valor moderado (0,6043 y 0,5604), que corresponde a una correlación entre puntuación verdadera y observada igual a 0,776 y 0,748 respectivamente. Estos resultados pueden ser debidos al hecho de que el cuestionario evalúa un conjunto de conocimientos muy amplio, como se ha visto en el análisis a priori del cuestionario. Estos valores moderados son frecuentes en cuestionarios orientados a la evaluación de conocimientos complejos, como se ha mostrado en Díaz, Batanero y Cobo (2003) y por lo tanto no es unidimensional.

A pesar de que todos estos conocimientos están relacionados con la probabilidad condicional, algunos de ellos requieren la puesta en relación de dicho concepto con otros, como la independencia, probabilidad conjunta o simple. Más aún, la literatura psicológica indica la existencia de sesgos persistentes, como la falacia del eje de tiempo o la falacia de la conjunción, incluso en sujetos que conozcan el tema.

Tabla 4.46. Resultados del análisis de fiabilidad con el total de la muestra

	Media si se omite el ítem	Varianza si se omite el ítem	Correlación corregida ítem- total	Alfa si se omite el ítem
I1	11,8298	25,8417	0,6870	0,7330
I2	12,6383	36,8785	0,2973	0,7671
I3	13,2553	40,1062	-0,2026	0,7861
I4	13,3298	39,2557	-0,0360	0,7790
I5	12,6489	38,5528	0,1076	0,7749
I6A	12,5638	38,3776	0,2058	0,7719
I6B	13,1064	37,7090	0,2153	0,7711
I6C	12,9255	36,7148	0,3713	0,7640
I6D	12,9255	36,8654	0,3458	0,7652
I7	13,1170	38,5991	0,0654	0,7774
I8	12,7340	31,8102	0,5262	0,7487
I9	13,1170	38,5130	0,0466	0,7803
I10	13,1277	38,3921	0,1017	0,7759
I11	12,5319	30,2086	0,6338	0,7369
I12	12,4787	32,5533	0,6777	0,7391
I13	12,4894	31,5644	0,7340	0,7326
I14	13,3830	38,9700	0,0487	0,7759
I15	12,9574	34,0842	0,5689	0,7493
I16	12,7766	34,4549	0,4907	0,7540
I17A	12,9043	36,9692	0,3301	0,7659
I17B	13,2553	39,2244	-0,0338	0,7801
I18	12,7340	38,9930	0,0052	0,7793

Reliability Coefficients
N of Cases = 94,0 N of Items = 22
Alpha = 0,7738

Finalmente, como se ha indicado, el tamaño de la muestra es limitado y el cuestionario sería posteriormente mejorado, por lo que se pensó que con todos estos condicionantes, los valores obtenidos no eran excesivamente bajos. Por ejemplo Morales (1988, pg. 249) indica que no hay regla fija para cuál debe ser el valor mínimo del coeficiente y que diversos autores clásicos han dado distintas reglas, desde 0,70, propuesto por Nunnally hasta 0,50 propuesto por Guilford. Carmines y Zeller (1979) proponen 0,8, al utilizar el coeficiente alfa; Santisteban (1990) indica, como limite general, 0,50. También se observa que el valor del coeficiente Alfa aumentaría al quitar algunos ítems del cuestionario. Se estudia esta posibilidad en la siguiente fase del trabajo.

Fiabilidad de codificación

Cinco de las ocho preguntas tuvieron exactamente la misma codificación en todos los casos y en general el número de acuerdos fue mayor del 91 % en todos los ítems, excepto en el primero (definición de la probabilidad condicional). En todo caso el porcentaje total de acuerdos es superior al 96% lo que deja menos de un 4% de margen de error en la codificación de las preguntas abiertas.

Tabla 4.47. Porcentaje de coincidencias

	% coincidencias
A1 – B1	93,8
A2 – B2	91,4
A8 – B8	91,6
A11 – B11	100
A12 – B12	100
A13 – B13	100
A15 – B15	100
A16 – B16	100
Global	96,7

5.4.4. APROXIMACIÓN A LA VALIDEZ

Ya se hicieron algunas consideraciones sobre la validez al realizar la definición de la variable en el Estudio 1. Entre los tipos de validación que se suele recomendar en la bibliografía se han tenido en cuenta, en las fases realizadas la validez de contenido que se describe en lo que sigue.

5.4.4.1. VALIDEZ DE CONTENIDO

Se considera justificada por el grado en que el contenido del instrumento refleja dominio en forma satisfactoria (Carmines y Zeller, 1979)[40]. Es decir, la adecuación de los ítems de un test como muestra de un universo más amplio de ítems representativos del contenido (Martínez Arias, 1995). Siguiendo a esta autora la validación del contenido se ha hecho mediante el examen sistemático del contenido del test para probar su representatividad y relevancia. Se han que los ítems del test son relevantes para el uso que se dará a las puntuaciones y representativos del contenido que se quiere evaluar, representando sus características esenciales. En la construcción del cuestionario se proporcionaron tres tipos de evidencia de validez de contenido.

[40] Grado en el que los *procesos* empleados por los sujetos para llegar a la respuesta son típicos de los procesos subyacentes a las respuestas del dominio (APA, AERA, NCTM, 1985).

Definición objetiva del contenido semántico de la variable

La definición semántica de la variable (Estudio 1) se basó en un estudio objetivo y exhaustivo de libros de texto realizados a partir del cual se construyeron las especificaciones del contenido. Esto contribuiría a recoger evidencias de *validez curricular*, es decir, el grado en que el cuestionario es relevante para los objetivos del currículo, formalmente descrito (Martínez Arias, 1995).

Juicio de expertos

En la validación de contenido se ha usado el juicio de expertos, definiendo previamente el universo de observaciones admisibles, identificando expertos en el campo, pidiendo a los expertos que emparejen ítems con objetivos y resumiendo la información numéricamente[41]. En consecuencia, se ha llevado a cabo un primer estudio que proporciona evidencias de validez de contenido del instrumento, mediante los siguientes indicadores:

- Se llevó a cabo un juicio de expertos en el que participaron 9 especialistas en el tema (cuyo conocimiento sobre el mismo se justifica).
- La relevancia de las especificaciones de contenido final se justifica por la alta valoración y el alto grado de acuerdo en los contenidos finalmente seleccionados para construir el instrumento. La representatividad se muestra en el hecho de que los escasos contenidos adicionales (lenguaje de probabilidad, diagramas en árbol) sugeridos cada uno de ellos por un sólo experto, están de hecho implícitos en los ítems seleccionados para el instrumento, en cuanto es posible evaluar su conocimiento por parte de los alumnos en los ítems de respuesta abierta.
- También se han proporcionado evidencias de la relevancia y representatividad de los ítems mediante la alta valoración y grado de acuerdo que los jueces han concedido a la idoneidad de los ítems finalmente seleccionados para evaluar los contenidos especificados.

Justificación mediante análisis a priori del cuestionario

Otro tipo de evidencia de la validez de contenido del cuestionario la proporciona el análisis de las posibles respuestas correctas e incorrectas a los diferentes ítems que

[41] En APA, AERA, NCME (1985) se considera que la validez de contenido se puede justificar mediante juicio lógico sobre la correspondencia que existe entre el rasgo del aprendizaje del evaluado y lo que se incluye en la prueba, recurriendo a expertos para valorar la adecuación de cada ítem al rasgo a evaluar.

componen el cuestionario, identificando los conceptos y propiedades necesarios para su resolución, así como los posibles errores conceptuales que llevarían a una solución errónea y comprobando que todo ello cubre la definición semántica de la variable.

Se analizaron los ítems en el orden en que aparecen en el primer cuestionario piloto, ya que el segundo contiene los mismos ítems, aunque colocados en orden inverso.

Análisis del ítem 1

Este ítem permite valorar la comprensión del contenido 1, "*definición de la probabilidad condicional"* y también detectar errores en dicha comprensión, por ejemplo, la confusión con una probabilidad simple o con una probabilidad compuesta. Por tanto el contenido *"distinguir probabilidad conjunta, condicional y simple* sería secundario. La *"definición de probabilidad condicional"* sería un contenido secundario en todos los ítems en que se pide esta probabilidad, puesto que es necesario conocerla para poderla calcular.

Análisis del ítem 2

Este ítem no ha sido probado en las pruebas iniciales y ha sido incluido a sugerencia de uno de los expertos, para valorar el contenido 4: "*Reconocer que la probabilidad condicional supone una restricción del espacio muestral"*. Como contenido secundario *"resolver correctamente problemas en contexto de muestreo con reposición"*, ya que el experimento se puede comparar con extraer al azar bolas entre dos posibles que representan el género de los hijos.

La respuesta correcta requiere la enumeración de todos los elementos del espacio muestral completo (parte a) y reducido por la condición dada (parte b), que serían los siguientes:

E= {(vvv), (vvm), (vmv), (mvv), (vmm), (mvm), (mmv), (mmm)}

E'= {(vvm), (vmv), (mvv)}

Una posible dificultad es que los alumnos no lleguen a completar la enumeración, faltándoles algunos de los casos. También podría ocurrir que los alumnos entiendan la pregunta b) como "al menos dos hijos son varones" y, por tanto, incluyan el caso de tres varones en el espacio muestral reducido.

Puesto que el interés es únicamente ver si los alumnos realizan una reducción correcta del espacio muestral, se calificará como correctos los casos en que se consideren 2 o "al menos 2" varones. También se aceptarán como parcialmente correctos aquellos casos en que no se complete todo el espacio muestral *E* (por ejemplo, si no se considera el orden), pero se reduzca correctamente el espacio *E'* a partir de *E*.

Análisis del ítem 3

Para resolver este problema, el alumno debe aplicar el teorema de Bayes:

$$P(M_1/D) = \frac{P(D \cap M_1)}{P(D)} = \frac{P(D/M_1)P(M_1)}{P(D)}$$

y, para ello, previamente calcular P(D) mediante la fórmula de la probabilidad total:

$$P(D) = P(D/M_1)P(M_1) + P(D/M_2)P(M_2)$$

Por tanto, este ítem (47 en la prueba de ítems) tiene como contenido primario *"Aplicar correctamente el cálculo de probabilidad condicional en situaciones de probabilidad inversa (Teorema de Bayes)".* Como contenidos secundarios aparecen "*Aplicar correctamente la regla de la probabilidad total", "resolver problemas de probabilidad condicional a partir de probabilidades conjuntas y simples", "resolver correctamente problemas condicionales cuando se invierte el eje de tiempos"* y *"resolver correctamente problemas en situaciones sincrónicas".*

Análisis del ítem 4

Este ítem corresponde al contenido primario *"distinguir sucesos dependientes, independientes y mutuamente excluyentes".* El alumno debe también "*resolver correctamente problemas de probabilidad compuesta haciendo uso de la regla del producto en experimentos independientes"* para reconocer que el distractor d es incorrecto. Los distractores tratan de detectar diferentes errores descritos en la literatura de investigación sobre probabilidad condicional:

- En el distractor b se trata de detectar la posible confusión entre sucesos excluyentes y sucesos independientes.
- En el distractor c se trata de detectar la posible creencia errónea que sólo pueden ser independientes los sucesos de experimentos diacrónicos.
- En el distractor d se presenta la definición correcta de la regla del producto, junto con la afirmación incorrecta que esta regla no se cumple en el caso de sucesos independientes. Precisamente en este caso se cumple porque los sucesos son

independientes y se trata de ver si el alumno detecta el error en la afirmación presentada en el distractor.

Análisis del ítem 5

Para resolver este problema hay que caer en la cuenta de que al extraer un foco defectuoso en la primera extracción (y no reponerlo), en la segunda extracción hay más probabilidades de que salga un foco normal. Este ítem evalúa como contenido primario "*resolver correctamente problemas en contexto de muestreo sin reposición"* y secundario *"resolver correctamente problemas de probabilidad condicional en situaciones diacrónicas",* así como *"distinguir sucesos dependientes e independientes.*

Análisis del ítem 6

Este ítem evalúa la capacidad de lectura de datos en una tabla de contingencia y la comprensión de la diferencia entre diversos tipos de probabilidad. En la tabla, se proporcionan las frecuencias absolutas dobles (casillas interiores), marginales por filas y columnas (márgenes de la tabla) y total de casos. A partir de estos datos se pide al alumno calcular las siguientes probabilidades:

- Probabilidad simple (pregunta a). El alumno debe calcular esta probabilidad dividiendo la frecuencia marginal por el número total de casos.
- Probabilidad condicional y diferenciación con su inversa (pregunta c) y (pregunta d). El alumno debe calcular estas probabilidades dividiendo la frecuencia doble por la marginal correspondiente a la condición pedida.
- Probabilidad conjunta (pregunta b). El alumno debe calcular esta probabilidad dividiendo la frecuencia doble por el total de casos.

Los errores que a priori se pueden detectar con este ítem son los siguientes:

- Confusión de una probabilidad simple con una compuesta.
- Confusión de una probabilidad simple con una condicional.
- Confusión de una probabilidad condicional con una compuesta.
- Confusión de una probabilidad condicional con su inversa.

Es un ítem cuyos contenidos primarios son *"distinguir probabilidad condicional con* *su inversa"*. El contenido *"distinguir probabilidad conjunta, condicional y simple"* aparece como secundario en todos los problemas en que se pide calcular una de estas probabilidades.

Análisis del ítem 7

Para resolver este ítem el alumno debe identificar los datos del problema: $P(+)=0,103$; $P(C\cap+)=0,008$. Debe deducir que el problema pide una probabilidad condicional $P(C/+)$ y recordar la fórmula para calcularla: $P(C/+)=\frac{0,8}{10,3}=0,0776$.

Por tanto, el contenido primario es "*resolver correctamente problemas de probabilidad condicional a partir de probabilidades conjuntas y simples"* y el secundario *"resolver correctamente problemas de probabilidad condicional en situaciones sincrónicas"*

Si el alumno elige el distractor b, está aplicando la fórmula del producto, en lugar de dividir las probabilidades, por tanto no recuerda la fórmula de la probabilidad condicional. Si el alumno elige el distractor c confunde probabilidad conjunta y condicional, por tanto otro contenido secundario sería *"confundir probabilidad conjunta, condicional y simple"*.

Análisis del ítem 8

Este ítem corresponde al contenido primario "*calcular la probabilidad condicional en un sólo experimento"* puesto que nos centramos en el experimento "elegir al azar el producto de dos números". Para resolver correctamente el problema, el alumno debe ser capaz de enumerar el espacio muestral producto obtenido al lanzar dos dados (36 casos posibles), identificando los casos en que el producto es 12.

1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

Para calcular la probabilidad pedida, el alumno debe reconocer que "*la probabilidad condicional supone una restricción del espacio muestral",* ya que sólo debe tomar como casos posibles {(2, 6) (3,4), (4,3) y (6,2)}. La probabilidad pedida es igual a ½ ya que hay dos casos favorables.

Análisis del ítem 9

Este ítem corresponde al contenido primario *"reconocer que una probabilidad conjunta ha de ser menor o igual que la probabilidad simple"* y como secundario *"distinguir probabilidad conjunta, condicional y simple"* ya que el alumno ha de interpretar los distractores a) y b) respectivamente como probabilidades simples y conjuntas. Se incluye en el cuestionario con el propósito específico de detectar la *falacia de la conjunción*, que consiste en asignar un valor mayor a una probabilidad conjunta que a la probabilidad simple de uno de los sucesos que interviene en la anterior.

Análisis del ítem 10

Este ítem cubre el contenido primario *"distinguir situación condicional, causal y diagnóstica"* y como secundario "*distinguir entre probabilidad condicional y su inversa"* (distractor b). Cuando se responde el distractor c el alumno podría confundir una probabilidad condicional con su inversa. Cuando se responde a) se manifiesta confusión entre situación condicional y diagnóstica.

Análisis del ítem 11

Aplicando la fórmula de la probabilidad total:

$$P\,(F) = P(H\ y\ F) + P(M\ y\ F) = P(H)\ P(F/H) + P(M)\ P(F/M) = 0{,}6 \times 0{,}5 + 0{,}4 \times 0{,}25 = 0{,}4.$$

El número esperado de personas que fuman, de un total de 200 es 200 x 0,4 = 80.

Para resolver este problema, el alumno podría utilizar un diagrama de doble entrada como el siguiente, donde trabajamos con datos enteros y hay que aplicar una proporción para calcular los datos de las celdas:

	Fuma	N fuma	
Hombres	30	30	60
Mujeres	10	30	40
	40	60	100

Si de 100 personas, fuman 40, de 200 personas, fumarán 80.

Este ítem evalúa como contenido primario "*aplicar correctamente la regla de la probabilidad total"* y como secundarios *"calcular la probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos dependientes"*, y *"resolver* *correctamente problemas en situaciones sincrónicas".*

Análisis del ítem 12

Este ítem corresponde al contenido primario "*resolver correctamente problemas en contexto de muestreo con reposición"* y como secundario *"resolver correctamente problemas de probabilidad condicional en situaciones diacrónicas",* así como *"distinguir sucesos dependientes e independientes".* Se requiere que el alumno comprenda que las ocurrencias anteriores no afectan la probabilidad de un suceso cuando los experimentos son independientes, como es el caso del muestreo sin reposición. Algunos alumnos podrían presentar la llamada *"falacia del jugador"* en el caso de que consideren que para el cálculo de la probabilidad pedida se debe tener en cuenta los resultados anteriores.

Análisis del ítem 13

Para resolver este ítem hay que hacer uso de la regla del producto, teniendo en cuenta que son sucesos independientes:

$$P\ (M \cap I) = P(M) \text{ x } P(I) = 0{,}8 \text{ x } 0{,}7 = 0{,}56$$

Este ítem evalúa como contenido principal *"calcular la probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos independientes",* y como secundario "*distinguir sucesos dependientes e independientes"*

Análisis del ítem 14

Para resolver este ítem, el alumno tiene que darse cuenta de que si la bola pasa por el canal II puede caer por R o por B, y si pasa por I, solo puede caer por R. El espacio muestral de este experimento sería: *{(IR), (IIR), (IIB)}* siendo *P (IR)* = 1/2 y *P (IIR)* = 1/4 y *P(IIB)* = 1/4.

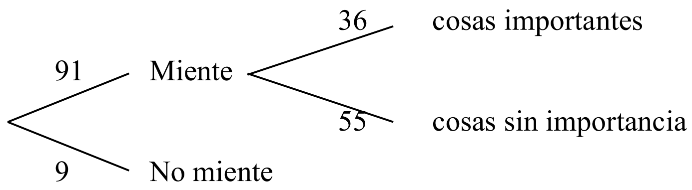
$$P(I/R) = P(IR)/\ P(R) = (1/2)/(3/4) = 2/3$$

Otro razonamiento es que a *R* llegan el doble de bolas desde *I* que desde *II*. Luego, *P (I/R)* = 2/3. El primer distractor indica que no se tienen en cuenta las bolas que caen por B, no se tiene en cuenta el condicionamiento por un suceso posterior. El distractor d indica que el alumno considera equiprobables los tres caminos, no diferenciando la probabilidad simple de la compuesta.

Este ítem evalúa como contenido primario el contenido "*resolver correctamente problemas de probabilidad condicional en situaciones diacrónicas"* y como secundarios, los contenidos *"resolver correctamente problemas condicionales cuando se invierte el eje del tiempo", "distinguir entre situación condicional y causal"* y *"resolver correctamente problemas de probabilidad condicional a partir de probabilidades conjuntas y simples.*

Análisis del ítem 15

Este ítem pide calcular una probabilidad conjunta en una situación sincrónica a partir de la probabilidad simple y condicional. El alumno podría utilizar un diagrama en árbol como el siguiente



$P\ (M \cap I) = P\ (M)\ x\ P\ (I/M)$ = 0,91 x 0,36 = 0,327

Este ítem evalúa como contenido primario *"resolver correctamente problemas de probabilidad condicional en situaciones sincrónicas"* y como secundario *"resolver correctamente problemas de probabilidad condicional a partir de probabilidades compuestas y simples"* y *"calcular una probabilidad condicional dentro de un único experimento".*

Análisis del ítem 16

Para resolver este problema, el alumno debe aplicar el teorema de Bayes

$$P(M_1/D) = \frac{P(D \cap M_1)}{P(D)} = \frac{P(D/M_1)P(M_1)}{P(D)}$$

y, para ello, previamente calcular P(D) mediante la fórmula de la probabilidad total

$$P(D) = P(D/M_1)P(M_1) + P(D/M_2)P(M_2)$$

Por tanto, este ítem tiene como contenido primario o principal *"aplicar correctamente el cálculo de probabilidad condicional en situaciones de probabilidad* *inversa (Teorema de Bayes)"*. Como contenido secundarios aparecen "a*plicar correctamente la regla de la probabilidad total", "resolver problemas de probabilidad condicional a partir de probabilidades conjuntas y simples", "resolver correctamente problemas condicionales cuando se invierte el eje de tiempos"* y *"resolver correctamente problemas en situaciones sincrónicas"*.

Análisis del ítem 17

Este ítem corresponde al contenido primario "*resolver correctamente problemas en contexto de muestreo sin reposición"* y secundario *"resolver correctamente problemas de probabilidad condicional en situaciones diacrónicas",* así como *"distinguir sucesos dependientes e independientes"*. Puesto que el muestreo es sin reemplazamiento, el alumno ha de tener en cuenta la composición de la urna una vez sacada la bola negra, es decir, el hecho de que *la probabilidad condicional supone una restricción del espacio muestral.* Una vez sacada una bola negra de la urna se tiene una bola negra y dos blancas, por lo que la probabilidad de obtener una bola negra es igual a 1/3. Se puede esperar los siguientes errores:

- Confundir muestreo con y sin reposición, considerando que la extracción de la bola negra no modifica el espacio muestral y, por tanto, la probabilidad es igual a 1/2 (distractor b).
- Confundir probabilidad condicional y conjunta (distractor c) aplicando la regla del producto (1/2) x (1/3)=1/6, por lo que aparece como contenido secundario "*distinguir probabilidad conjunta, condicional y simple"*
- Confundir probabilidad condicional y simple (distractor d) considerando la probabilidad simple de obtener una bola negra en la primera extracción (1/4) que corresponde al contenido secundario anterior.

Para resolver la segunda parte de este ítem el alumno tiene que comprender la diferencia entre muestreo con reposición y sin reposición. Además tiene que darse cuenta de que si en segundo lugar, hay una bola negra, ésta queda eliminada en la primera extracción, por lo que el espacio muestral de la primera extracción queda restringido. El alumno tiene que aceptar que es posible condicionar un suceso con otro que ocurra posteriormente (el distractor d evalúa la falacia del eje temporal). El distractor c evalúa la confusión entre probabilidad condicional, conjunta y simple.

Este ítem tiene como contenido primario *"resolver correctamente problemas de probabilidad condicional cuando se invierte el eje de tiempo" y* secundario *"distinguir situación condicional, causal y diagnostica (apartado b)*, "*resolver correctamente problemas en contexto de muestreo sin reposición" y "reconocer que la probabilidad condicional supone una restricción del espacio muestral" y "distinguir probabilidad condicional, conjunta y simple".*

Análisis del ítem 18

El espacio muestral de este experimento sería: *{(AR), (RR), (RA)}* siendo los tres sucesos equiprobables. El alumno puede confundir muestreo con y sin reposición y elegir el distractor a. También puede hacer una mala enumeración del espacio muestral y elegir el distractor b.

Este ítem evalúa como contenido primario *"calcular la probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos dependientes",* y como contenidos secundarios los siguientes contenidos: "*distinguir sucesos dependientes e independientes"*, *"resolver correctamente problemas de muestreo sin reposición"* y *"reconocer que una probabilidad condicional supone una restricción del espacio muestral".*

En la tabla 4.48 se muestran los contenidos que abarca el cuestionario piloto, diferenciando para cada ítem los contenidos primarios y secundarios abarcados. Se observa que todos los contenidos aparecen dos o más veces excepto el contenido "*probabilidad conjunta menor que probabilidad simple"* y todos los contenidos aparecen una vez como contenido principal.

Tabla 4.48. Contenidos primarios y secundarios evaluados por los ítems

	Contenido	Primario	Secundario
Conocimiento conceptual	1. Definición de la probabilidad condicional	1	Todos los que se pide la probabilidad condicional
	2. Reconocer que la probabilidad de *P(A)* >0 para poder definir *P(B/A)*	rechazado por los expertos	
	3. Reconocer que una probabilidad condicional cumple los axiomas.	rechazado por los expertos	
	4. Reconocer que la probabilidad condicional supone una restricción del espacio muestral	2	8, 17
	5. Distinguir probabilidad condicional con inversa	6	10
	6. Distinguir probabilidad conjunta, condicional y simple	6	Todos los que se pide la probabilidad condicional
	7. Probabilidad conjunta menor que probabilidad simple	9	
	8. Distinguir sucesos independientes, dependientes y mutuamente excluyentes	4	5, 12,13, 17, 18
C. Procedimental	9. Calcular una probabilidad condicional dentro de un único experimento	8	15, 16
	10. Resolver correctamente problemas de probabilidad condicional en un contexto de muestreo con reposición	12	18
	11. Resolver correctamente problemas de probabilidad condicional en un contexto de muestreo sin reposición	5	17
	12. Resolver correctamente problemas de probabilidad condicional a partir de probabilidades conjuntas y simples	7	14, 15
	13. Resolver correctamente problemas condicionales cuando se invierte el eje de tiempo	17	3, 14, 16
	14. Distinguir situación condicional, causal y diagnóstica	10	14, 17
	15. Resolver correctamente problemas en situaciones diacrónicas	14	5, 12, 17
	16. Resolver correctamente problemas en situaciones sincrónicas	15	3, 7, 11, 16
	17. Resolver correctamente problemas de probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos independientes	13	4
	18. Resolver correctamente problemas de probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos dependientes	18	11
	19. Aplicar correctamente el cálculo de la probabilidad condicional en situaciones de sucesos múltiples (regla de la probabilidad total)	11	3. 16
	20. Aplicar correctamente el cálculo de la probabilidad condicional en situaciones de probabilidad inversa (regla de Bayes)	3, 16	

5.5. DISCUSIÓN DEL ESTUDIO 3

El objetivo del Estudio 3 fue realizar unas pruebas del cuestionario piloto, aún siendo conscientes de que éste no era un instrumento definitivo y que sería necesario mejorarlo.

Recogidos datos de una muestra de 57 estudiantes de psicología se analizaron los índices de dificultad y discriminación de los ítems, mediante diversos procedimientos. La distribución de los ítems de dificultad, aproximadamente normal, cubrió un rango amplio de dificultad que permitiría alcanzar el objetivo de evaluar los diversos grados de conocimiento de los estudiantes respecto a la probabilidad condicional, e indicó su adecuación para el estudio definitivo. Señalamos, no obstante que algunos ítems resultaron excesivamente difíciles, por lo que se decidió proceder a una nueva revisión de su redacción y presentación sin variar el contenido. Esto será motivo del Estudio 4 en el que se recurre a un nuevo juicio de expertos para mejorar, en general los ítems del cuestionario, en especial los que fueron demasiado difíciles.

Aunque la mayoría de los ítems muestran una buena discriminación, medida mediante la correlación corregida ítem- total, se observa la falta de discriminación en ítems que recogen sesgos clásicos de razonamiento condicional, que no son tenidos en cuenta en la enseñanza. Resultados similares se producen al analizar la correlación de estos ítems como diferencia de índices de dificultad entre dos grupos, uno de los cuales (estudiantes de matemáticas) tiene una mayor preparación en el tema. La principal conclusión que se obtiene es didáctica, más que metodológica y es la necesidad de reformar la enseñanza de la probabilidad condicional, para tener en cuenta los resultados de la investigación psicológica. Otra consecuencia es el carácter multidimensional del cuestionario, que habrá que tener en cuenta en el futuro estudio de su validez y fiabilidad.

Globalmente, tanto desde el punto de vista de la puntuación total en el cuestionario, como de la media de la distribución de índices de dificultad, se observa una discriminación entre dos grupos de diferente preparación a favor del grupo superior, lo que potencialmente apoya la *validez discriminante* del cuestionario. No obstante, se repetirá el estudio de discriminación en el cuestionario final, utilizando tan sólo estudiantes de psicología con y sin instrucción.

Una aportación es la puesta a punto y aplicación de las estimaciones bayesianas de los índices de dificultad y discriminación con distribuciones iniciales no informativas, que se desarrollan a nivel teórico en el Capítulo 3.

En el caso de los índices de dificultad, también se ponen a punto el estudio con distribuciones iniciales informativas, usando para definirlas los resultados de los diversos ítems en las pruebas prepiloto (Estudio 2.1). Ello permite una mayor precisión de las estimaciones logradas. Las distribuciones finales de los índices obtenidas mediante este procedimiento podrían también usarse para contrastar hipótesis sobre los mismos o estudiar la posibilidad de futuras replicaciones de los resultados.

En este estudio también se inició la primera aproximación a la fiabilidad y validez del instrumento. Respecto a la fiabilidad, se obtiene un valor suficiente del coeficiente

Alfa, a pesar del tamaño limitado de la muestra y del carácter multidimentsional del constructo evaluado. Respecto a la validez, se realiza una primera aproximación a la *validez de contenido*, que se justifica, tanto por la definición semántica objetiva de la variable, llevada a cabo en el Estudio 1, como por la validación de dicha definición y de la congruencia de los ítems utilizados para evaluar las diversas unidades de contenido llevada a cabo mediante juicio de expertos en el Estudio 2.2, como por un análisis teórico detallado del contenido de los ítems finales del cuestionario que se realiza en el Estudio 4.

Con todo ello, se pasa al estudio 4 en el que, mediante nuevo juicio de expertos se mejoran los ítems del cuestionario sin variar el contenido, finalizando de este modo la construcción del cuestionario RPC.

6. ESTUDIO 4. REVISIÓN DEL INSTRUMENTO PILOTO MEDIANTE JUICIO DE EXPERTOS

6.1. INTRODUCCIÓN

Una vez finalizadas las pruebas piloto del cuestionario, descritas en el Estudio 3 se decidió llevar a cabo una nueva revisión de los ítems que lo componían. Se decidió mantener el contenido de los ítems, puesto que éste cubría adecuadamente las unidades de contenido determinadas en la definición semántica de la variable, por lo que el cuestionario piloto ofrecía amplias garantías de validez de contenido, como se justificó en el Estudio 3.

La finalidad del Estudio 4 es tratar de disminuir la dificultad de los ítems del cuestionario, con el fin de mejorar sus características psicométricas; en particular la de aquellos ítems que resultaron demasiado difíciles en el cuestionario piloto.

Para lograrlo se decidió realizar un segundo juicio de expertos que sirviera para mejorar la redacción de los ítems, evitar posibles ambigüedades en su enunciado, así como interacciones contextuales con el contenido, y detectar cualquier otro aspecto de la redacción que pudiese mejorarse.

Es decir, este segundo juicio de expertos tenía como finalidad principal mejorar la adecuación de los ítems en cuanto a criterios metodológicos, tales como: forma de redacción de los ítems y presentación de las alternativas, En lo que sigue describimos el método, resultados y conclusiones.

6.2. MÉTODO

El estudio se basó en la identificación de una muestra de expertos en metodología que pudiesen colaborar en el estudio, a los cuáles se mandó un cuestionario, preparado para este estudio. Para cada uno de los ítems del cuestionario piloto RPC se ofreció a los expertos tres versiones diferentes, elaboradas con los criterios que describimos a continuación y pidiendo que los ordenaran, en función de su adecuación metodológica a los fines de nuestro estudio. También se les pidió que añadiesen cuantas sugerencias contribuyeran a mejorar la redacción de los ítems.

Las tres versiones elaboradas de cada uno de los ítems del cuestionario RPC tuvieron en cuenta, en primer lugar, las sugerencias recibidas por escrito de los propios estudiantes que participaron en las pruebas piloto, a los cuales se les pidió su opinión sobre posibles puntos que no comprendieron bien o sobre mejoras en la redacción.

También se tuvieron en cuenta las recomendaciones sobre elaboración de ítems de pruebas citadas por diversos autores (López Feal, 1986; Osterlind, 1989; Thorndike, 1991).

6.2.1. SUJETOS

En este segundo juicio de expertos y siguiendo los criterios de Millman y Greene (1989), fueron seleccionados con base en su conocimiento expertos en metodología y más particularmente de elaboración de tests. Las muestras de expertos son frecuentes en los estudios cualitativos y exploratorios para generar la materia prima en el diseño de un cuestionario (Hernández, Fernández y Baptista, 1998).

Para identificarlos, se preparó una lista de investigadores en el área de metodología que hubieran investigado en la elaboración de cuestionarios durante los últimos 10 años y que hubiesen realizado investigaciones o hubiesen impartido docencia relacionada con la probabilidad condicional, pues era fundamental que dominasen bien el contenido evaluado.

Localizamos un total de 28 investigadores que podían leer la lengua castellana, cuestión fundamental, puesto que los ítems están realizados en dicha lengua. De entre ellos, se decidió invitar a la mitad, elegidos mediante sorteo aleatorio. Se trataba de un total de diez investigadores españoles y cuatro de otros países (Colombia, Estados Unidos, México). Todos ellos eran localizables por correo electrónico, puesto que sus direcciones eran accesibles en Internet, a través de sus Departamentos o de listas de miembros de Sociedades Internacionales. De ellos doce aceptaron participar.

La nuestra incluyó nueve investigadores españoles, uno de Estados Unidos y dos colombianos, cubriendo siete universidades españolas y cuatro extranjeras. Nuestra muestra de expertos incluyó siete profesores titulares del Área de Metodología de Investigación en Ciencias del Comportamiento, un profesor del área de Métodos de Investigación y Diseños en Educación, tres Estadísticos y un profesor de Didáctica de las Matemáticas. Todos ellos habían impartido cursos de doctorado de Metodología de Investigación y eran doctores. La edad abarca desde 38 a 56 años. Todos los participantes contaban con al menos 10 años de experiencia investigadora y habían elaborado instrumentos de medición.

6.2.2. MATERIAL

A partir del conjunto de ítems que compondrá el cuestionario final, elaboramos para cada ítem otras dos formas alternativas adicionales de redacción, además de la utilidad en la prueba piloto. Las tres formas de un mismo ítem evalúan el mismo contenido relacionado con la probabilidad condicional y bien tienen el mismo contexto o muy similar. Las variaciones se producen, bien en la redacción o presentación del enunciado, bien porque se proporciona un ejemplo de respuesta.

Con las tres formas alternativas de cada ítem preparamos un cuestionario que se recoge en el Anexo 5 para ser completado por estos expertos y que es similar a la muestra incluida en la Tabla 4.49. Como se observa en la muestra, para cada conjunto de tres ítems relativos al mismo contenido se pide al experto que ordene los tres ítems según su adecuación metodológica.

En dicha valoración deberían emplearse siguientes criterios (Millman y Greene, 1989): Comunicabilidad, redacción correcta, claridad y consistencia, facilidad de respuesta. En esta ordenación 1 indica el ítem más adecuado y 3 el menos adecuado.

Tabla 4.49. Muestra del material que componen el cuestionario entregado a los expertos

Contenido 1. Definición de la probabilidad condicional

Ítem 1a. Explica con tus propias palabras la diferencia entre una probabilidad simple y una probabilidad condicional.

Ítem 1b. Define probabilidad simple y probabilidad condicional, dando un ejemplo de cada una de ellas.

Ítem 1c. Define probabilidad simple y probabilidad condicional.

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

1. ¿Por qué considera que el ítem elegido en primer lugar es el más adecuado?

2. Si lo considera necesario, incluya algunas sugerencias para mejorar el enunciado del ítem elegido en primer lugar (más adecuado), sin variar su contenido:

Contenido 2: Reconocer que la probabilidad condicional supone una restricción del espacio muestral

Ítem 2a. Define el espacio muestral de los siguientes experimentos aleatorios:
a. Observar el género (varón/ mujer) de los hijos de las familias con tres descendientes, teniendo en cuenta el orden de nacimiento.
b. Observar el género (varón/ mujer) de los hijos de las familias con tres descendientes en los que dos o tres son varones, teniendo en cuenta el orden de nacimiento.

Ítem 2b. Define el espacio muestral de los siguientes experimentos aleatorios:
a. Observar el género (varón/ mujer) de los hijos de las familias con tres descendientes.
b. Observar el género (varón/ mujer) de los hijos de las familias con tres descendientes en los que dos o tres son varones.

Ítem 2c. Complete el espacio muestral de los siguientes experimentos aleatorios:
a. Observar el género (varón/ mujer) de los hijos de las familias con tres descendientes (ejemplo MVM,...).
b. Observar el género (varón/ mujer) de los hijos de las familias con tres descendientes en los que dos o tres son varones (ejemplo VVM,...).

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

¿Por qué considera que el ítem elegido en primer lugar es el más adecuado?

Si lo considera necesario, incluya algunas sugerencias para mejorar el enunciado del ítem elegido en primer lugar (más adecuado), sin variar su contenido:

En segundo lugar se pidió a los expertos que sugirieran mejoras en la redacción del ítem que consideraran más adecuado, pero sin variar el contenido, proporcionándoles un espacio para esta función. El objetivo era establecer un consenso de opiniones de los expertos sobre cómo cada ítem particular era más adecuado metodológicamente, seleccionando el ítem más adecuado para el cuestionario.

6.3. ANÁLISIS Y RESULTADOS

Una vez recogidos los cuestionarios completados por los expertos que participaron en el segundo juicio, se analizó la información recogida. Para cada uno de los ítems, los datos constituyen una ordenación de tres variables (versiones a, b y c de los ítems) según una escala ordinal que toma los valores 1,2,3. En consecuencia se recogen datos ordinales para un conjunto de 18 ítems por 3 versiones diferentes y 12 valoraciones ordinales de cada uno. En consecuencia, realizamos un estudio no paramétrico de los datos, utilizando el programa SPSS. En la Tabla 4.50 y 4.51 se presentan los estadísticos descriptivos de rangos.

Tabla 4.50. Estadísticos descriptivos de rangos asignados a las tres versiones de los ítems 1 a 9

			Media	Desviación típica	Percentiles		
ITEM		Rango promedio			25	50 (Mediana)	75
1	a	1,75	1,750	0,753	1,00	2,00	2,00
	b	**1,42**	**1,500**	**0,522**	**1,00**	**1,50**	**2,00**
	c	2,83	2,750	0,621	3,00	3,00	3,00
2	a	2,00	2,000	0,894	1,00	2,00	3,00
	b	2,55	2,363	0,674	2,00	2,00	3,00
	c	**1,45**	**1,636**	**0,809**	**1,00**	**1,00**	**2,00**
3	a	2,09	2,000	0,894	1,00	2,00	3,00
	b	**1,55**	**1,636**	**0,674**	**1,00**	**2,00**	**2,00**
	c	2,36	2,181	0,873	1,00	2,00	3,00
4	a	2,70	2,500	0,849	1,75	3,00	3,00
	b	**1,55**	**1,600**	**0,516**	**1,00**	**2,00**	**2,00**
	c	1,75	1,700	0,823	1,00	1,50	2,25
5	a	2,27	2,090	0,539	2,00	2,00	2,00
	b	**1,23**	**1,272**	**0,646**	**1,00**	**1,00**	**1,00**
	c	2,50	2,363	0,809	2,00	3,00	3,00
6	**a**	**1,73**	**1,818**	**0,873**	**1,00**	**2,00**	**3,00**
	b	2,00	2,000	0,774	1,00	2,00	3,00
	c	2,27	2,181	0,873	1,00	2,00	3,00
7	a	2,68	2,545	0,687	2,00	3,00	3,00
	b	**1,32**	**1,454**	**0,687**	**1,00**	**1,00**	**2,00**
	c	2,00	2,000	0,774	1,00	2,00	3,00
8	a	2,00	2,000	0,774	1,00	2,00	3,00
	b	2,27	2,181	0,750	2,00	2,00	3,00
	c	**1,73**	**1,818**	**0,981**	**1,00**	**1,00**	**3,00**
9	**a**	**1,58**	**1,750**	**0,866**	**1,00**	**1,50**	**2,75**
	b	1,96	2,000	0,852	1,00	2,00	3,00
	c	2,46	2,333	0,778	2,00	2,50	3,00

Tabla 4.51. Estadísticos descriptivos de rangos asignados a las tres versiones de los ítems 10 a 18

			Media	Desviación típica	Percentiles		
ITEM		Rango promedio			25	50 (Mediana)	75
10	a	2,00	2,000	0,774	1,00	2,00	3,00
	b	2,41	2,272	0,646	2,00	2,00	3,00
	c	**1,59**	**1,727**	**1,009**	**1,00**	**1,00**	**3,00**
11	a	1,65	1,600	0,516	1,00	2,00	2,00
	b	**1,50**	**1,400**	**0,699**	**1,00**	**1,00**	**2,00**
	c	2,85	2,500	0,707	2,00	3,00	3,00
12	a	2,82	2,727	0,646	3,00	3,00	3,00
	b	**1,36**	**1,454**	**0,522**	**1,00**	**1,00**	**2,00**
	c	1,82	1,818	0,750	1,00	2,00	2,00
13	**a**	**1,64**	**1,636**	**0,924**	**1,00**	**1,00**	**3,00**
	b	1,73	1,818	0,404	2,00	2,00	2,00
	c	2,64	2,545	0,820	2,00	3,00	3,00
14	a	1,75	1,800	0,788	1,00	2,00	2,25
	b	2,65	2,500	0,849	1,75	3,00	3,00
	c	**1,60**	**1,700**	**0,674**	**1,00**	**2,00**	**2,00**
15	**a**	**1,50**	**1,636**	**0,674**	**1,00**	**2,00**	**2,00**
	b	2,64	2,454	0,820	2,00	3,00	3,00
	c	1,86	1,909	0,831	1,00	2,00	3,00
16	**a**	**1,32**	**1,363**	**0,809**	**1,00**	**1,00**	**1,00**
	b	2,14	2,090	0,301	2,00	2,00	2,00
	c	2,55	2,545	0,820	2,00	3,00	3,00
17	a	1,86	1,818	0,750	1,00	2,00	2,00
	b	**1,73**	**1,727**	**0,646**	**1,00**	**2,00**	**2,00**
	c	2,41	2,181	0,981	1,00	3,00	3,00
18	**a**	**1,42**	**1,583**	**0,514**	**1,00**	**2,00**	**2,00**
	b	2,13	2,083	0,792	1,25	2,00	3,00
	c	2,46	2,333	0,984	1,00	3,00	3,00

En dichas tablas se han resaltado en negrita aquellas versiones en cada uno de los ítems que presentan mejor valoración, en cuando a menor rango medio en la ordenación dada (lo que indica mayor valoración) y menores valores de mediana y percentiles.

En caso de igualdad de medianas se optó por la versión con menor rango medio y percentiles y en caso de igualdad de todos estos estadísticos, por aquellos con mayor acuerdo, dado por la menor dispersión (recorrido intercuartílico y desviación típica).

Asimismo, se llevó a cabo, para cada ítem una prueba no paramétrica de comparación de muestras relacionadas, considerando que cada experto da un rango a cada una de tres versiones a, b, y c del ítem. Presentamos en la tabla 4.52 los resultados de las pruebas de comparación de muestras relacionadas de Friedman y W de Kendall. Observamos que, excepto un caso, las pruebas fueron estadísticamente significativas, lo que indica una clara preferencia de los expertos por el ítem tomado finalmente para formar parte del cuestionario.

Tabla 4.52. Resultados de pruebas no paramétricas

Ítem	W de Kendall(a)	Chi-cuadrado (Friedman)	Sig. asintót.
1	0,775	18,588	0,000
2	0,545	12,000	0,002
3	0,424	9,333	0,009
4	0,604	12,080	0,002
5	0,654	14,387	0,001
6	0,273	6,000	0,050
7	0,682	15,000	0,001
8	0,218	4,800	0,091
9	0,440	10,571	0,005
10	0,351	7,714	0,021
11	0,730	14,600	0,001
12	0,761	16,750	0,000
13	0,434	9,548	0,008
14	0,614	12,286	0,002
15	0,593	13,040	0,001
16	0,477	10,500	0,005
17	0,382	8,400	0,015
18	0,485	11,643	0,003

Generalmente esta preferencia estuvo motivada en razones, que se analizaron y categorizaron, según las categorías mostradas en la tabla 4.53.

Tabla 4.53. Resultados de pruebas no paramétricas

Tipo de comentario	Frecuencia	Porcentaje
Mejor redacción (terminología o formulación de preguntas mas adecuada)	16	22
Formulación mas simple o clara	14	19
Existencia de elementos que crean ambigüedad o confusión	10	14
Uso de porcentajes o fracciones	10	14
Facilita la resolución	7	10
Uso de ejemplos	6	8
Representa mejor el contenido	5	7
Evita respuestas memorísticas	3	5
Mejor distractor	2	3
TOTAL	73	

Por ejemplo, las razones dadas para preferir el enunciado a del ítem 1 fueron las siguientes:

- *"En los enunciados alternativos se solicita una definición, la cual puede ser entendida como una definición formal. Desde mi punto de vista, esto puede ser inconveniente puesto que es posible que algunos alumnos hayan memorizado alguna definición que no comprendan"* (Experto 3).
- *"Porque genera más iniciativa por parte del que responde y ello genera más variabilidad en el momento del análisis del contenido de la respuesta. Además es*

una definición por comparación y puede dar más alternativas" (Experto 6).

- *"En los enunciados alternativos se solicita una definición, la cual puede ser entendida como una definición formal. Desde mi punto de vista, esto puede ser inconveniente puesto que es posible que algunos alumnos hayan memorizado alguna definición que no comprendan".* (Experto 8)
- *"Porque da al estudiante mayor libertad para expresar lo que realmente sabe. No lo "encasilla" en definiciones"* (Experto 9).

Razones semejantes se incluyeron en el resto de los ítems. Los expertos también indicaron algunas sugerencias para mejorar la redacción, todas las cuales se tuvieron en cuenta para preparar la versión final del cuestionario.

6.4. DISCUSIÓN DEL ESTUDIO 4

El Estudio 4 estuvo orientado a mejorar la redacción de los ítems que componen el cuestionario RPC sin variar su contenido ni el contexto empleado. Con ello se deseaba aumentar la facilidad de los ítems, en particular algunos que fueron especialmente difíciles en la prueba piloto del cuestionario.

Se contó con la colaboración de doce expertos en metodología de investigación con experiencia en la impartición de cursos de psicometría o en la construcción del cuestionario, a los cuales se mandó un cuestionario. En dicho cuestionario se presentaron tres versiones alternativas de cada ítem para ordenar, de acuerdo a criterios metodológicos.

El estudio descriptivo de rangos y contrastes de Friedman y W de Kendall para comparación de muestras relacionadas, permitió elegir una versión de cada uno de los ítems. Las sugerencias de los expertos en cuanto a mejora de la redacción también se tuvieron en cuenta.

Con esto se finaliza la construcción del cuestionario RPC, cuya fiabilidad y validez de analizarán con muestras de mayor tamaño en los estudios 5 y 6.

CAPÍTULO 5.
VALIDEZ Y FIABILIDAD DEL CUESTIONARIO RPC

1. INTRODUCCIÓN

Una vez finalizada la construcción del cuestionario RPC (estudios 1 a 4) según el proceso descrito en el Capítulo 4, se aborda en este capítulo el análisis del error de medida y su control. Puesto que, como en todo proceso de medida en educación, el objetivo es realizar inferencias sobre conceptos abstractos a partir de indicadores empíricos,[1] se está sujeto a errores sistemáticos (sesgos) y aleatorios (Carmines y Zeller, 1979; Thorndike, 1989; Meliá, 2001). Los análisis de validez y fiabilidad, incluidos en este capítulo atienden a estas dos dimensiones.

El objetivo del estudio 5 es proporcionar evidencias del control del error aleatorio, desde una doble vertiente:

- Desde la teoría clásica de los tests (Muñiz, 1994; Martínez Arias, 1995; Barbero, 2003), se estimará el coeficiente de fiabilidad de consistencia interna (primera parte del estudio 5.1), así como los coeficientes θ y Ω basados en los resultados del análisis factorial que se presenta en el estudio 6.2 y de la prueba repetida (estudio 5.2).
- Además en la segunda parte del estudio 5.1 se hace una aproximación desde la teoría de la Generalizabilidad (López Feal, 1987; Feldt y Brennan, 1991; Martínez Arias, 1995), calculando un coeficiente de generalizabilidad, tanto para ítems como para sujetos, atendiendo, de este modo, a estas dos posibles fuentes de error en el muestreo.

Por otro lado, en el estudio 6 se completa la validación del instrumento iniciada en el estudio 3 mediante la justificación de validez de contenido, con el análisis de la validez de criterio y de constructo.

Con los resultados obtenidos, se justifica la utilización del cuestionario RPC para realizar una evaluación de los razonamientos de estudiantes de psicología sobre la probabilidad condicional, que se presenta en el capítulo 6, y cuyos resultados se usan en el diseño de la experiencia de enseñanza descrita en el capítulo 7.

[1] Más precisamente, relacionar los conocimientos de los alumnos sobre un concepto con sus respuestas a los ítems del cuestionario.

2. ESTUDIO 5. FIABILIDAD DEL CUESTIONARIO RPC

2.1. ESTUDIO 5.1. FIABILIDAD DE CONSISTENCIA INTERNA Y GENERALIZABILIDAD

El estudio 5.1 estuvo orientado a proporcionar evidencias del grado de control del error aleatorio en el instrumento. Estas evidencias se proporcionan, tanto desde la Teoría Clásica de los Tests, realizando una estimación de la fiabilidad de consistencia interna, como desde la Teoría de la Generalizabilidad. En lo que sigue se describe el método y resultados de este estudio.

2.1.1. MÉTODO

2.1.1.1. SUJETOS

Los sujetos que participaron en este estudio son 590 alumnos de Psicología de las Universidades de Granada (cuatro grupos de estudiantes; n=307), Murcia (dos grupos, n=106), Jaén (dos grupos, n=101) y Huelva (un grupo, n=76). Todos los participantes cursaban primero de Psicología y eran alumnos de la asignatura Análisis de Datos. La mayoría de los alumnos tienen una edad de 18 o 19 años.

2.1.1.2. MATERIAL

El material utilizado es el cuestionario RPC en su versión final, que se compone de los ítems seleccionados en el último juicio de expertos (Estudio 4). De este cuestionario se elaboraron dos versiones (A y B) que difieren sólo en el orden de presentación de los ítems, con el fin de evitar que los alumnos copien la respuesta y que los ítems finales tengan mayor proporción de no respuesta debido al posible cansancio. En la Tabla 5.1 se presenta el cuestionario.

Tabla 5.1. Cuestionario RPC

Ítem 1. Define probabilidad simple y probabilidad condicional, dando un ejemplo de cada una de ellas.

Ítem 2. Completa el espacio muestral de los siguientes experimentos aleatorios:

Observar el género (hombre/mujer) de los hijos de las familias con tres descendientes (ejemplo MHM,...).

Observar el género (hombre/mujer) de los hijos de las familias con tres descendientes en los que dos o más son hombres.

Ítem 3. Un taxi se vio implicado en un accidente nocturno con choque y huida posterior. Hay dos compañías de taxis en la ciudad, la Verde y la Azul. El 85% de los taxis de la ciudad son Verdes y el 15% Azules. Hubo un testigo del accidente. El tribunal comprobó la fiabilidad del testigo bajo las mismas circunstancias que había la noche del accidente y llegó a la conclusión de que el testigo identificaba correctamente cada uno de los colores en el 80% de las ocasiones y fallaba en el 20%. Sabiendo que el testigo identificó el taxi como azul, ¿Cuál es la probabilidad de que el taxi implicado en el accidente fuera en efecto Azul?

☐ a) 80/100
☐ b) 15 /100
☐ c) (15/100) x (80/100)
☐ d) $\frac{15\times 80}{85\times 20+15\times 80}$

Ítem 4. Se extrae una carta al azar de una baraja española (40 cartas con cuatro palos: oros, bastos, espadas y copas. Cada palo tiene los números del 1 al 7, sota caballo y rey). Sea *A* el suceso "se extrae una carta de oros" y *B* el suceso "se extrae un rey". ¿Los sucesos *A* y *B* son independientes?

☐ a) No son independientes, porque en la baraja hay un rey de oros.
☐ b) Sólo si sacamos primero una carta para ver si es rey y se vuelve a colocar en la baraja y luego sacamos una segunda carta para mirar si es oros.
☐ c) Si, porque P(rey de oros)= P(rey) x P(oros),
☐ d) No, porque P(rey / oros) $\neq$ P(rey).

Ítem 5. Una caja tiene cuatro focos, de los cuales dos son defectuosos. Se sacan dos al azar, uno tras otro sin reemplazamiento. Si el primer foco fue defectuoso, entonces:

☐ a) Es mas probable que el segundo sea defectuoso;
☐ b) Es más probable que el segundo no sea defectuoso;
☐ La probabilidad de que el segundo sea defectuoso es igual a la probabilidad de que no lo sea.

Ítem 6. En una población se ha realizado una entrevista a un grupo de hombres, obteniendo los siguientes resultados:

	menos de 55 años	más de 55 años	Total
Ha sufrido un ataque al corazón	29	75	104
No ha sufrido ataque	401	275	676
Total	430	350	780

Si elegimos al azar una de estas personas:

¿Cuál es la probabilidad de que haya tenido un ataque al corazón?

¿Cuál es la probabilidad de que tenga más de 55 años y además haya tenido un ataque al corazón?

Sabiendo que la persona escogida tiene más de 55 años, ¿Cuál es la probabilidad de que haya tenido un ataque al corazón?

Sabiendo que la persona escogida ha tenido un ataque al corazón, ¿Cuál es la probabilidad de que tenga más de 55 años?

Ítem 7. La probabilidad de que una mujer tenga una mamografía positiva es el 10.3%. La probabilidad de que una mujer tenga cáncer de pecho y una mamografía positiva es de 0.8%. Una mujer se realiza una mamografía y resulta positiva. ¿Cuál es la probabilidad de que realmente tenga cáncer?

☐ a) $\frac{0,8}{10,3} = 0,0776$, probabilidad de 7,76%

☐ b) 10,3 x 0,8 = 8,24; probabilidad del 8,24%

☐ c) 0,8 %

Ítem 8. Hemos lanzado dos dados y sabemos que el producto de los dos números obtenidos ha sido 12 ¿Cual es la probabilidad de que ninguno de los dos números sea un 6? (Se diferencia si un número ha aparecido en un dado o en otro).

Ítem 9. Supón que una estrella del tenis alcanza la final de Roland Garros en 2005. Para ganar el partido hay que ganar tres sets de cinco. ¿Cuál de los siguientes sucesos consideras más probable?

☐ a) El jugador pierde el primer set.

☐ b) El jugador pierde el primer set pero gana el partido.

☐ c) Los dos sucesos son iguales de probables.

Ítem 10. Un test diagnóstico de cáncer fue administrado a todos los residentes de una gran ciudad en la que hay pocos casos de cáncer. Un resultado positivo en el test es indicativo de cáncer y un resultado negativo es indicativo de ausencia de cáncer. ¿Qué te parece más probable?

☐ a) Que una persona tenga cáncer si ha dado positivo en el test de diagnóstico.

☐ b) Que un test de diagnóstico resulte positivo si la persona tiene cáncer.

☐ c) Los dos sucesos tienen la misma probabilidad.

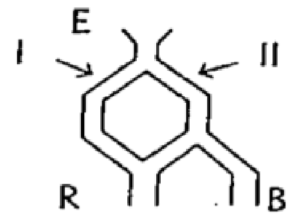
Ítem 11. En una ciudad el 60 % son hombres y el 40 % mujeres. El 50 % de los hombres y el 25 % de las mujeres fuman. Si se escoge una persona al azar, ¿Cuál es la probabilidad de que sea fumadora?

Ítem 12. Una persona lanza un dado y anota si saca un número par o impar. Se trata de un dado no sesgado (es decir todos los números tienen la misma probabilidad). Estos son los resultados al lanzarlo 15 veces:

par, impar, impar, par, par, impar, par, par, par, par, impar, impar, par, par, par

Lanza el dado de nuevo ¿Cuál es la probabilidad de sacar un número par en esta nueva tirada?

Ítem 13. Un grupo de alumnos de un colegio se examina de matemáticas e inglés. La proporción de alumnos que aprueban matemáticas es del 80% y la proporción de alumnos que aprueba inglés es del 70%. Suponiendo que las notas en cada una de las asignaturas son independientes, ¿Cuál es la probabilidad de que un alumno escogido al azar haya aprobado ambas asignaturas?



Ítem 14. Una bola se suelta por la entrada E. Si sale por R, ¿Cuál es la probabilidad de que haya pasado por el canal 1?

☐ a) 1/2

☐ b) 1/3

☐ c) 2/3

☐ d) No se puede calcular

Ítem 15. En una encuesta publicada en un periódico se informa que el 91% de los habitantes de una ciudad mienten usualmente y, de ellos, el 36% mienten sobre cosas importantes. Si elegimos al azar a una persona de esta ciudad, ¿Cuál es la probabilidad de que mienta sobre cosas importantes?

Ítem 16. Una fábrica dispone de dos máquinas M1 y M2 que fabrican bolas. La máquina M1 fabrica el 40 % de las bolas y la M2 el 60%. El 5% de las bolas fabricadas por M1 y el 1% de las fabricadas por M2 son defectuosas. Tomamos una bola al azar que resulta ser defectuosa. ¿Cuál es la probabilidad de que haya sido fabricada por M1?

Ítem 17. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin devolver la primera a la urna. ¿Cuál es la probabilidad de extraer una bola negra en segundo lugar, habiendo extraído una bola negra en primer lugar? $P\ (N_2/\ N_1)$

- ☐ a) 1/ 2
- ☐ b) 1/ 6
- ☐ c) 1/ 3
- ☐ d) 1/ 4

¿Cuál es la probabilidad de haber extraído una bola negra en primer lugar, sabiendo que hemos extraído una bola negra en segundo lugar? $P\ (N_1/\ N_2)$

- ☐ a) 1/ 3
- ☐ b) No se puede calcular
- ☐ c) 1/ 6
- ☐ d) 1/ 2

Ítem 18. Se tiene una bolsa con 1 bola azul y 2 bolas rojas. Se saca una bola al azar. Sin reponerla otra vez en la bolsa se saca una segunda bola. ¿Qué suceso es más probable?

- ☐ a) Sacar dos bolas rojas.
- ☐ b) Sacar primero una bola roja y luego una azul.
- ☐ c) Los dos sucesos son iguales de probables.

2.1.1.3. PROCEDIMIENTO

Una vez completada la versión final del cuestionario, se elaboraron dos formas alternativas, alterando el orden de las preguntas. La primera versión es la que se presenta en el Anexo 6 y la segunda contiene exactamente los mismos ítems con el orden invertido. En cada uno de los grupos de alumnos participantes se repartieron de forma aleatoria cada una de las dos formas del cuestionario, de modo que aproximadamente la mitad de cada grupo pasó una versión. De este modo se trató de evitar, tanto la posibilidad de copia, como que los últimos ítems tuviesen una tendencia a menor número de respuestas.

Una vez repartidos los cuestionarios se pidió a los alumnos que completaran los datos solicitados al principio del cuestionario y que leyeran las instrucciones escritas, no dejando que ninguno de los alumnos formulara ninguna pregunta a las personas presentes durante la administración. De esta forma se aseguró que todos los alumnos recibían las mismas instrucciones y explicaciones acerca de la prueba (las instrucciones escritas).

El cuestionario fue completado exactamente en las mismas condiciones en todos los grupos; durante una de las sesiones de la asignatura de análisis de datos. Los estudiantes fueron avisados con anterioridad que se les daría un cuestionario sobre probabilidad condicional, pidiéndoles que estudiasen el tema. La participación en los dos grupos fue voluntaria, motivándose al tener una posible mejora en la puntuación si se obtenía un cierto número de respuestas correctas. Se dejó un tiempo de dos horas, aunque prácticamente todos finalizaron en hora y media; en algunos casos llegaron a hora y tres cuartos en la cumplimentación. El periodo académico fue el segundo semestre, no demasiado cerca del periodo de exámenes.

2.1.2. ANÁLISIS

2.1.2.1. ESTUDIO DE FIABILIDAD

En este estudio se entenderá la fiabilidad como tendencia a la consistencia o precisión del instrumento en la población medida (Bisquerra, 1989; Melía, 2001; Carmona, 2004)[2]. Se trata de evaluar el grado de control de los errores de carácter aleatorio, para que no interfieran de forma significativa en las predicciones que se hagan a partir de las puntuaciones obtenidas (Barbero, 2003)[3].

Entre los diversos procedimientos para el cálculo del estimador del coeficiente de fiabilidad se tomó el coeficiente Alfa de Cronbach (que se reduce al de Kuder-Richardson para ítems dicotómicos) (Carmines y Zeller, 1979; López Féal, 1986; Morales, 1992; Meliá, 2001) por varias propiedades: a) reflejar el grado en el que covarían los ítems que constituyen el test; b) ser el valor medio de todos los que se obtendrían con el método de las dos mitades si se utilizasen todas las combinaciones de ítems; c) ser cota inferior de la que se obtendría por el método de la prueba repetida; y d) estimar la *fiabilidad en el acto* (Martínez Arias, 1995; Meliá, 2001).

Se calculó este coeficiente a partir de la muestra de estudiantes de psicología (n = 591) mediante el programa correspondiente del paquete estadístico SPSS, analizando los estadísticos de cada ítem si se suprime del instrumento y estudiando el efecto sobre el coeficiente de ir suprimiendo sucesivamente los ítems que presentan peores resultados.

---

[2] En Teoría Clásica de Tests se define como correlación entre las puntuaciones verdadera y observada (Martínez Arias, 1995).

[3] Los errores aleatorios son debidos a la variabilidad del material experimental y pueden disminuir con el tamaño de la muestra o mediante la precisión o fiabilidad del instrumento (Carmines y Zeller, 1979).

También se calcularon los siguientes coeficientes, basados en el análisis factorial que se presenta en el Estudio 6 (Barbero, 2003):

- Theta de Carmines[4]: $\theta = \frac{n}{n-1}\left(1 - \frac{1}{\lambda_1}\right)$, donde n es el número de ítems y λ el primer autovalor en el análisis factorial.
- Omega de Heise y Bohrnsted: $\Omega = 1 - \frac{n - \sum h_j^2}{n + 2\sum r_{jh}}$, que expresa la relación entre comunalidades y correlaciones.

2.1.2.2. ESTUDIO DE GENERALIZABILIDAD

Se utilizó la teoría de la generalizabilidad que extiende la teoría clásica de la medición, y permite, por medio del análisis de varianza, analizar diferentes fuentes de error en un proceso de medida (Feldt y Brennan, 1991)[5]. Un concepto central es del de "faceta" que se asocia a todas las condiciones similares de medida posibles de un tipo dado[6]:

Siguiendo a López Feal (1986, p. 512) se consideró que "*una faceta puede ser un conjunto de ítems, un conjunto de métodos particulares de recogida de datos, un conjunto de sistemas escolares, un conjunto de evaluadores, etc.*" Cuando el objeto de estudio son las personas, la fiabilidad del estudio se maximiza cuando la varianza entre sujetos aumenta y es grande respecto a la de los ítems, que son homogéneos entre si. Pero, si el objeto de estudio es medir el grado en que los diferentes objetivos educativos se alcanzan, como en este caso, la varianza de los ítems será grande respecto a la de los sujetos y se podría obtener una generalizabilidad grande –referida a los ítems como objeto de estudio (López- Feal, 1986).

Este es precisamente el caso en que de esta investigación; puesto que se investiga la comprensión de un concepto matemático avanzado. Para alcanzar una validez alta, se incluyó un conjunto de ítems muy variados en cuanto a su dificultad y a los componentes de comprensión del concepto que evalúan. Las limitaciones del tiempo requerido de administración del cuestionario restringieron el número de ítems para evitar el cansancio

[4] Carmines y Zéller (1979) definen este coeficiente para cuestionarios no unidimensionales, como es el RPC.

[5] Las fuentes de error pueden ser los mismos sujetos, las preguntas o las condiciones en que se aplican los instrumentos de medida. López Feal (1986) sugiere que esta teoría proporciona un potente instrumento en aquellos casos en que los sujetos no son necesariamente el objeto central del estudio.

[6] El concepto puntuación universo reemplaza el de puntuación verdadera asociado a la teoría clásica de los tests. Brennan (1983) diferencia entre población que se refiere a los objetos de medida y universo de observaciones admisibles que se refiere a las condiciones de medida y puede incluir varias facetas. La puntuación universo representa el valor esperado de la puntuación del sujeto en un universo de observaciones admisibles.

de los alumnos al completar la prueba.

Ello hizo razonable la variabilidad entre los distintos ítems y se decidió aplicar la teoría de la generalizabilidad, considerando las facetas de ítems y sujetos, siguiendo la misma aproximación de otras investigaciones sobre comprensión de conceptos matemáticos avanzados (ej. Navarro- Pelayo, 1994) y nuestro propio trabajo anterior en otros cuestionarios (Díaz, Batanero y Cobo, 2003).

Se tuvo en cuenta dos etapas en relación con la obtención de coeficientes ligados a la teoría de la generalizabilidad (López Feal, 1986; Feldt y Brennan, 1991; Martínez Arias, 1995):

1. Estudio G. Estimación de los componentes de la varianza

En el estudio G o de generalizabilidad, se estimaron los componentes diferenciados de la varianza que afectan a las medidas[7]. Se consideró el diseño cruzado de una faceta (ítems), donde cada sujeto pasa una sola vez la prueba (es decir el conjunto de condiciones o ítems). La puntuación media X_{pi} de una persona p bajo la condición i se descompone en la forma siguiente (Martínez Arias, 1995):

$$X_{pi} = \mu + (\mu_p - \mu) + (\mu_i - \mu) + (X_{pi} - \mu_p + \mu_i - \mu) \qquad (1)$$

siendo μ la media de todos los sujetos en la prueba; $\alpha_p = (\mu_p - \mu)$ el efecto debido a la persona, $\beta_i = (\mu_i - \mu)$ el efecto de la condición i (en nuestro caso del ítem i) y $(\alpha\beta)_e = (X_{pi} - \mu_p + \mu_i - \mu)$ la interacción entre persona- condición.

Puesto que la ecuación (1) define un modelo lineal de efectos aleatorios, se usó el análisis de varianza para estimar cada uno de estos componentes de la varianza. Como cada persona pasa una sola vez la prueba, la interacción se confunde con el error aleatorio (Martínez Arias, 1995).

Todos los componentes del modelo, excepto μ son variables aleatorias independientes. En el estudio G se estimaron las diferentes varianzas:

- Varianza debida a los sujetos σ^2_s o varianza de diferenciación de todas las personas del universo de personas entre sí.

[7] Estos componentes son estimadores de las varianzas reales (parámetros) asociados al universo de observaciones admisibles (Brennan, 1983).

- Varianza debida a los ítems σ^2_i, refleja los errores asociados con los diferentes niveles de dificultad de los ítems.
- Varianza residual σ^2_r o del error aleatorio de medida, que en este tipo de diseño está confundido con la interacción ítem- persona.

2. Estudio D. Generalización a un universo de ítems

En el estudio D o de decisión se eligió una versión modificada del diseño del estudio G[8] estimando su precisión mediante el cálculo de un coeficiente de generalizabilidad. Sus resultados se pueden usar para tomar decisiones sobre cómo se podría mejorar un instrumento específico de medida[9]. Se siguieron los siguientes pasos:

1. Identificar las facetas fija y aleatoria del diseño.
2. Calcular los componentes de la varianza de la parte aleatoria. Cada componente es la suma del correspondiente al diseño aleatorio (estimado en el estudio G) más el componente de interacción dividido por el número de condiciones de la faceta.

El coeficiente de generalizabilidad se define con el cociente (1),

$$G = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2} \tag{1}$$

es decir como cociente entre la varianza de las puntuaciones verdaderas de la prueba y la varianza observada, que es suma de la varianza verdadera más la varianza debida al error aleatorio.

La varianza de error depende de cómo se define el universo de puntuaciones verdaderas y en el análisis de generalizabilidad se consideran ciertas fuentes como parte de la varianza de error en unas condiciones y otras fuentes en otras (Thorndike, 1989)[10].

[8] Un concepto importante el en estudio D es el de *universo (o universos) de generalización,* al que se quiere generalizar con un instrumento de medida específico (Brennan, 1983).

[9] En los estudios G se supone que se está trabajando un modelo de análisis de varianza de efectos aleatorios, pero, en la práctica es difícil obtener muestras aleatorias de los niveles de los factores de los correspondientes universos. Por ello, en el estudio D una faceta se considera fija, cuando el conjunto de condiciones incluidas en el diseño recoge todas las condiciones de interés, es decir, no se trata de generalizar el estudio más allá de dichos niveles (por tanto, se trabajaría con un modelo de análisis de varianza de efectos fijos) (Martínez Arias, 1995).

[10] En este trabajo se diferenciaron dos fuentes para el error aleatorio y se calcularon dos coeficientes de generalizabilidad: la generalizabilidad a otros alumnos de la misma prueba y la generalizabilidad si pasáramos otros problemas similares a los incluidos en la prueba a los mismos alumnos.

2.1.3. RESULTADOS

2.1.3.1. ESTUDIO DE FIABILIDAD DE CONSISTENCIA INTERNA

Presentamos a continuación los resultados del estudio de fiabilidad por el método de consistencia interna, calculados mediante el programa Escalas de SPSS.

El valor obtenido (0,797) corresponde a una correlación entre la puntuación observada y la puntuación verdadera de 0,889 e indica una consistencia bastante alta (Meliá, 2001). En todo caso, Morales (1988, p. 249) indica que no hay regla fija para cuál debe ser el valor mínimo del coeficiente y que diversos autores clásicos han dado distintas reglas, desde 0,70, propuesto por Nunnally hasta 0,50 propuesto por Guilford. Santisteban (1990) indica, como limite general, 0,50. Es importante recordar que en la valoración de este coeficiente interviene la complejidad con que se ha definido el constructo[11].

Tabla 5.2. Estadísticos total-elemento de la prueba RPC (*n*=590)

	Media de la escala si se elimina el ítem	Varianza de la escala si se elimina el ítem	Correlación elemento-total corregida	Correlación múltiple al cuadrado	Alfa de Cronbach si se elimina el ítem
I1	16,15	38,799	0,451	0,244	0,792
I2	16,51	44,698	0,458	0,261	0,781
I3	17,47	48,287	0,264	0,111	0,792
I4	17,65	49,123	0,165	0,158	0,796
I5	17,09	48,887	0,264	0,117	0,793
I6a	17,09	48,185	0,401	0,259	0,789
I6b	17,41	46,921	0,466	0,315	0,785
I6c	17,40	47,057	0,447	0,406	0,785
I6d	17,42	47,218	0,421	0,376	0,786
I7	17,55	49,546	0,156	0,098	0,798
I8	17,11	45,212	0,378	0,206	0,786
I9	17,69	48,948	0,208	0,119	0,794
I10	17,60	50,129	0,002	0,048	0,801
I11	16,63	43,039	0,646	0,497	0,770
I12	16,58	43,603	0,561	0,367	0,775
I13	16,81	43,060	0,572	0,420	0,773
I14	17,81	49,554	0,154	0,097	0,796
I15	16,73	42,606	0,624	0,504	0,770
I16	15,76	36,453	0,600	0,498	0,773
I17a	17,22	49,030	0,177	0,098	0,795
I17b	17,64	50,185	-0,004	0,112	0,801
I18	17,21	48,895	0,202	0,083	0,794

[11] Morales (1988) también señala que constructos de definición compleja (como es el RPC) requieren ítems más diferenciados, menos relacionados, por la que la consistencia interna será menor (p. 305).

La correlación corregida ítem- test (Tabla 5.3) ha oscilado desde valores 0,154 a 0,624 y fueron estadísticamente significativas excepto en los ítems 10 y 17b que se refieren a sesgos en el razonamiento sobre probabilidad condicional que, aparentemente no mejoran con la instrucción (Sedlemeier, 1999), lo que explica la falta de correlación con el resto de la escala. La supresión de estos dos elementos elevaría algo la fiabilidad. La correlación media entre ítems (Tabla 5.4) es también bastante elevada, pero se decidió conservarlos por estar interesados en la evaluación de estos sesgos en los estudiantes de psicología.

Tabla 5.3. Estadísticos de fiabilidad de la prueba RPC (*n*=590)

Alfa de Cronbach	Alfa de Cronbach basada en los elementos tipificados	N de elementos
0,796	0,797	22

Tabla 5.4. Estadísticos de resumen de los elementos

	Media	Mínimo	Máximo	Rango	Máximo/ mínimo	Varianza
Medias de los elementos	0,815	0,120	2,169	2,049	18,056	0,291
Varianzas de los elementos	0,550	0,106	2,569	2,463	24,263	0,463
Covarianzas inter-elementos	0,083	-0,040	0,882	0,922	-22,146	0,018
Correlaciones inter-elementos	0,144	-0,090	0,591	0,681	-6,588	0,018

Se realizó la prueba de no aditividad de Tukey (tabla 5.5) sobre el supuesto de que no exista una interacción multiplicativa entre los elementos. Al ser estadísticamente significativo el resultado se asume la aditividad[12].

Tabla 5.5. ANOVA con la prueba de Friedman y la prueba de no aditividad de Tukey

			Suma de cuadrados	gl	Media cuadrática	Chi-cuadrado de Friedman	Sig.
Inter-personas			1350,734	590	2,289		
Intra-personas	Inter-elementos		3613,588	21	172,076	367,981	0,000
	Residual	No aditividad	1010,569(a)	1	1010,569	2617,453	0,000
		Equilibrio	4783,252	12389	0,386		
		Total	5793,821	12390	0,468		
	Total		9407,409	12411	0,758		
Total			10758,143	13001	0,827		

Media global =0,82
Estimación de Tukey de la potencia a la que es necesario elevar las observaciones para conseguir la aditividad = -0,337

La prueba T-cuadrado de Hotelling genera un contraste multivariado sobre la hipótesis nula de que todos los elementos de la escala tienen la misma media. Como era de esperar

[12] La presencia de un componente de variación de no aditividad significativo puede ser indicativo de la existencia de relaciones proporcionales o multiplicativas entre los elementos y los casos. También reduce la precisión de las pruebas de significación de otros efectos aumentando de un modo excesivo la varianza error estimada.

el resultado es estadísticamente significativo, pues las medias son diferentes, ya que se obtuvo una gama variada de índices de dificultad.

Tabla 5.6. Prueba T cuadrado de Hotelling

T-cuadrado de Hotelling	F	gl1	gl2	Sig.
3773,615	173,605	21	570	0,000

Tabla 5.7. Coeficiente de correlación intraclase

	Correlación intraclase	Intervalo de confianza 95%		Prueba F con valor verdadero 0			
		Límite inferior	Límite superior	Valor	gl1	gl2	Sig.
Medidas individuales	0,138	0,121	0,157	4,515	590,0	12390	0,000
Medidas promedio	0,779	0,752	0,804	4,515	590,0	12390	0,000

El coeficiente de correlación intraclase[13] proporcionó un alto valor promedio, indicador de la fiabilidad de una sola medida y fue significativo tanto para los valores promedios como para las medidas individuales, aunque para éstas el valor fue menor.

Tabla 5.8. Estadísticos de fiabilidad basados en la división en dos mitades para el cuestionario RPC

Lambda	1	0,770
	2	0,839
	3	0,796
	4	0,783
	5	0,840
	6	0,848

También se calcularon los coeficientes de fiabilidad Lambda de Guttman (Martínez Arias, 1995) de los que el Lambda 3 es equivalente al coeficiente Alfa. Observamos que los valores obtenidos de los coeficientes Lambda son todos mayores que 0,76, lo que indica una consistencia bastante alta entre las dos mitades del cuestionario.

Coeficientes basados en el análisis factorial

A partir de los resultados del análisis factorial exploratorio que se presenta en el estudio 6.2 se calcularon los coeficientes de fiabilidad basados en el análisis factorial (Barbero, 2003):

[13] Medidas sobre la consistencia o sobre el acuerdo de los valores entre los propios casos (Barbero, 2003).

1. El coeficiente theta[14], bastante alto, que, junto con el hecho de que el primer factor explicó mucho mayor porcentaje de varianza que los siguientes, y que la mayoría de los ítems tenían pesos apreciables en el mismo, antes de la rotación, indica una cierta unidimensionalidad del cuestionario[15]. Este coeficiente tiene en cuenta los pesos de los ítems en el primer factor (Morales, 1988).

$$\theta = \frac{n}{n-1}\left(1-\frac{1}{\lambda_1}\right)=\frac{22}{21}\left(1-\frac{1}{4,609}\right)=0,82$$

2. El coeficiente Omega dio todavía un valor superior. Como es de esperar se cumplió la relación: $\alpha < \theta < \Omega$.

$$\Omega = 1-\frac{n-\sum h_j^2}{n+2\sum r_{jh}} = 1-\frac{22-15,273}{22+2\times 62,526}=0,896$$

2.1.3.2. ESTUDIO DE GENERALIZABILIDAD

Seguidamente se analizó la posibilidad de generalización de los datos obtenidos en la aplicación del cuestionario, tanto al universo de personas, como al universo de condiciones (ítems).

Estudio G. Estimación de los componentes de la varianza

Para estimar estos componentes de la varianza se utilizó la tabla de análisis de varianza (Tabla 5.9) generada por el análisis de escalas del programa SPSS, y el modelo de estimación de Dunn y Clarck (1987) para el análisis de varianza de medidas repetidas.

Tabla 5.9. Análisis de varianza de medidas repetidas para un diseño de una faceta

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-personas		1350,734	590	2,289		
Intra-personas	Inter-elementos	3613,588	21	172,076	367,981	0,000
	Residual(a)	5793,821	12390	0,468		
	Total	9407,409	12411	0,758		
Total		10758,143	13001	0,827		

Media global = 0,82

[14] Se trata del coeficiente α cuando las varianzas verdaderas (suma de las covarianzas) y total se refieren a las puntuaciones factoriales derivadas del primer factor común, antes de las rotaciones; este coeficiente sintetiza la información aportada por el primer factor (Morales, 1988).

[15] No completa, pues el primer autovalor sólo explicó un 20% de varianza y unos pocos ítems (relativos a sesgos) no aportan ponderaciones al mismo. Se discutirá esto con más detalle en el estudio de validez.

De esta tabla se obtuvo los cuadrados medios entre sujetos, entre los diferentes ítems y residual, así como sus grados de libertad.

CM_r = 0,468 es un estimador de σ^2_r

CM_p = 2,289 es un estimador de $b\sigma^2_p + \sigma^2_r$, siendo *b* el número de ítems

CM_i = 172,076 es un estimador de $a\sigma^2_i + \sigma^2_r$ siendo *a* el número de sujetos

Tabla 5.10. Estimaciones de los componentes de la varianza

Fuente de variación	Componentes varianza	% Varianza total
Sujetos	0,087	10,3
Ítems	0,291	34,4
Residual	0,468	56,3
Total	0,845	100

Despejando en las anteriores ecuaciones se obtuvo las estimaciones de componentes de varianza presentadas en la Tabla 5.10. La variabilidad total explicada por los ítems fue mucho mayor que la explicada por los sujetos, lo que confirma el hecho de que el coeficiente de fiabilidad α de Cronbach (α=0,796) aunque alto, no se haya aproximado más a la unidad, debido a la hetereogeneidad entre los diversos ítems.

Estudio D. Generalización a un universo de ítems

Se fijó, en primer lugar, como faceta las personas tomando como universo de generalización el conjunto de todos los ítems posibles asociados a la variable medida. En consecuencia, las decisiones que se tomasen a partir de la puntuación de un sujeto particular estarían basadas en la puntuación total sobre los 20 ítems del cuestionario. El universo de generalización sería el universo de tests aleatoriamente paralelos de la misma longitud[16].

La fórmula (1) del coeficiente de generalizabilidad se transforma en este caso en la expresión (2)

$$(2) \qquad G_i = \frac{\sigma_s^2}{\sigma_s^2 + \frac{\sigma_r^2}{22}} = 0{,}799$$

[16] Cualquier otra medida del examinado se tomaría a partir de la puntuación en otra muestra del mismo número de ítems, que serían *aleatoriamente paralelos*.(López Féal, 1986).

Se obtuvo un valor próximo al del coeficiente Alfa, lo cual es lógico, puesto que el coeficiente de generalizabilidad a otros ítems coincide con la fiabilidad de consistencia interna, ya que la única fuente de variación es la debida a variabilidad entre ítems. Las pequeñas diferencias son debidas a redondeos en los cálculos[17].

Este coeficiente mide la generalizabilidad de los resultados si a los mismos alumnos les pasáramos otra prueba del mismo número de ítems, elegidos aleatoriamente entre el universo de ítems. Debido a la complejidad conceptual del constructo "probabilidad condicional" no sería lógico esperar unas posibilidades de generalización excesivas si se cambian los ítems.

Estudio D. Generalización a un universo de alumnos

En segundo lugar se consideró la población de alumnos como universo de generalización y la faceta ítems como fija. Esta elección no es usual, pero la teoría de la generalizabilidad nos lo permite; *"La teoría de generalizabilidad no define per se una varianza de la población universo; es más bien el investigador quien la define mediante una consideración activa de la naturaleza de su proceso de medición"* (Brennan, 1983, p. 123). "*Desde una perspectiva general, cualquier grupo de condiciones (personas, ítems, codificadores, ocasiones, clases, etc.) pueden ser el objeto de medida"* (p. 31)[18].

Teniendo en cuenta estas consideraciones, y siguiendo otros trabajos relacionados con conceptos matemáticos avanzados (Navarro-Pelayo, 2004) se intercambió el papel de ítems y sujetos, considerando los ítems como objeto de medida. La finalidad sería en este caso evaluar la dificultad de cada ítem particular en el estudio para un universo de estudiantes (estudiantes de psicología) que constituiría el universo de generalización.

La "puntuación" (valor medio) de cada ítem en el universo de generalización vendría dada por el valor medio si se pasase el ítem a toda la población de estudiantes, que vendría estimada por el valor medio obtenido en la muestra. La fórmula (1) del coeficiente de generalizabilidad se transforma en este caso en la expresión (3):

---

[17] La igualdad es sólo de tipo algebraico y no conceptual: Mientras que el coeficiente Alfa mide la correlación entre la puntuación observada y verdadera, la generalizabilidad mide la posible generalización a otros tests aleatoriamente paralelos en el universo de generalización (Brennan, 1983).

[18] Incluso incluye ejemplos de posibles situaciones en que las personas puedan verse como una faceta en el universo de observaciones admisibles: *"Usualmente las personas se ven como el objeto de medida en las aplicaciones específicas de la teoría de la generalizabilidad; sin embargo la teoría pos si no dicta el papel que juega la población (o universo)de personas"* y, citando un contexto de evaluación donde el objeto de medida son las clases participantes, *"en estos casos sería razonable ver el universo de observaciones admisibles teniendo personas cruzadas con ítems y las clases siendo el objeto de medida"* (p.31).

$$(3) \qquad G_i = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_i^2 / 37} = 0,987$$

Se obtuvo un valor muy alto, cuando se considera los alumnos como universo posible de generalización, fijando los ítems de la prueba, es decir, considerando los ítems como faceta fija en el diseño. Por supuesto, en la hipótesis de que se conserven las características sociológicas y educativas de los alumnos a los que se pasa la prueba.

2.2. ESTUDIO 5.2. FIABILIDAD DE PRUEBA REPETIDA

El objetivo del estudio 5.2 fue realizar una estimación de la fiabilidad del cuestionario RPC mediante el procedimiento de prueba repetida, es decir, evaluar la estabilidad de las puntuaciones en aplicaciones repetidas del instrumento a los mismos sujetos (Barbero, 2003).

El cálculo empírico de un coeficiente de estabilidad requiere de dos conjuntos de medidas de los mismos sujetos para de esta forma, correlacionar ambas medidas, es decir, la estabilidad temporal de las mediciones si las condiciones o el concepto no cambian[19].

La operacionalización de estos dos conjuntos de medidas es diversa. Entre los posibles procedimientos se eligió el procedimiento test - retest o de prueba repetida.

2.2.1. MÉTODO

2.2.1.1. SUJETOS

Los sujetos participantes en este estudio fueron 112 estudiantes de Psicología de primer curso, que realizan la asignatura de Análisis de Datos y que formaban parte del grupo de estudiantes de la Universidad de Granada, en el estudio anterior. Pertenecían a dos de los grupos de estudiantes de dicha Universidad que fueron seleccionados aleatoriamente entre los cuatro grupos de esta Universidad participantes en el estudio. Dentro de estos grupos se pidieron estudiantes voluntarios para pasar la segunda prueba.

[19] Puesto que se consideró el cuestionario RPC es fiable, la administración del cuestionario en dos ocasiones separadas por un intervalo breve debía dar lugar a resultados similares.

Entre los 206 estudiantes que participaron en la primera pasación del cuestionario, se ofrecieron 112 para la segunda prueba. Al comparar la puntuación media de estos dos grupos de estudiantes en la prueba inicial del cuestionario se obtuvo valores similares ($\bar{x}$ =19,88 y $\bar{x}$ =20,13) en los sujetos que pasaron y no pasaron el retest, no resultando la diferencia significativa en la prueba *T* de diferencia de medias independientes (p=0,12).

2.2.1.2. MATERIAL Y PROCEDIMIENTO

El material utilizado es el cuestionario RPC, que fue presentado en la sección 2.1.1.2 de este capítulo. En cada uno de los grupos seleccionados para el estudio se administró el cuestionario dos veces, una vez estudiado el tema de probabilidad condicional, y con un intervalo de 4 semanas. Este intervalo se consideró suficiente para que no hubiese un cambio en el constructo medido (no hubo exámenes intermedios ni lecciones sobre el tema) y para que no se produjese un efecto de recuerdo de las respuestas anteriores.

En cada una de las administraciones se repartió, de forma aleatoria, cada una de las dos formas del cuestionario, de modo que aproximadamente la mitad, contestasen cada versión. Una vez repartidos los cuestionarios se les pidió a los alumnos que completaran los datos solicitados al principio del cuestionario y que leyeran las instrucciones escritas, no dejando que ninguno de los alumnos formulara ninguna pregunta a las personas presentes durante la administración. De esta forma se aseguró que todos los alumnos recibían las mismas instrucciones y explicaciones acerca de la prueba (las instrucciones escritas).

Se les dejó tiempo suficiente para que cumplimentaran la prueba, una hora y media, habiendo finalizado todos los alumnos en el tiempo asignado.

Durante la primera aplicación, a los alumnos participantes se les avisó que iban a formar parte de un experimento que constaba de dos partes, y se les citó para la segunda prueba, pero no se les informó de en qué consistiría la segunda parte.

2.2.2. ANÁLISIS

Para estimar la fiabilidad test – retest[20] se han seguido diversos procedimientos: En primer lugar se estudiaron las correlaciones de Pearson y el Coeficiente Rho de correlación de Spearman entre las puntuaciones totales y las puntuaciones en cada ítem en las dos pasaciones. Seguidamente se llevó a cabo un estudio de fiabilidad de formas paralelas con el programa escalas de SPSS, considerando las dos pasaciones del test, como formas paralelas. Los resultados se presentan a continuación.

Se incluyen también estimaciones bayesianas de los coeficientes de correlación entre ítems y puntuación total, que intervienen en la estimación de fiabilidad de la prueba repetida y fiabilidad de formas paralelas (correlación entre formas). Para ello se usa la transformación de Fischer (Lee, 2002), sustituyendo $\rho = \tanh \xi; r = \tanh z$, de donde se obtiene una aproximación de la distribución final de dicha transformada, mediante la distribución normal:

$$\xi \sim N(z, 1/n)$$

Esta aproximación se usa para calcular los intervalos de credibilidad para la tangente hiperbólica del coeficiente de correlación y a partir de estos intervalos, deshaciendo el cambio de variable se determinan los correspondientes al coeficiente de correlación.

2.2.3. RESULTADOS

Correlaciones entre partes

Al analizar las correlaciones entre las puntuaciones totales, se observa un alto valor del coeficiente de correlación, tanto de Pearson como de Spearman, lo que indica una gran estabilidad del cuestionario; bastante superior al mínimo de 0,7 requerido por Carmines y Zéller (1979) para la fiabilidad test- retest.

---

[20] La fiabilidad test-retest es una de las formas más rigurosas de medir la fiabilidad porque evalúa la estabilidad de la medida en el tiempo (Carmines y Zéller, 1979).

Tabla 5.11. Correlaciones entre las dos pasaciones para cada uno de los ítems (n=112)

	Correlación de Pearson	Sig. (bilateral)
i1	0,691(**)	0,000
i2	0,664(**)	0,000
i3	0,521(**)	0,000
i4	0,440(**)	0,000
i5	0,349(**)	0,000
I6a	0,683(**)	0,000
I6b	0,599(**)	0,000
I6c	0,515(**)	0,000
I6d	0,392(**)	0,000
I7	0,473(**)	0,000
I8	0,729(**)	0,000
I9	0,447(**)	0,000
I10	0,291(**)	0,002
I11	0,577(**)	0,000
I12	0,651(**)	0,000
I13	0,568(**)	0,000
I14	0,551(**)	0,000
I15	0,555(**)	0,000
I16	0,561(**)	0,000
I17a	0,360(**)	0,000
I17b	0,427(**)	0,000
I18	0,299(**)	0,001

** La correlación es significativa al nivel 0,01 (bilateral).

Tabla 5.12. Intervalo de confianza y credibilidad de los coeficientes de correlación para cada ítem con distribución inicial no informativa (n=112)

	Intervalo confianza 95%		Intervalo credibilidad 95%	
	Extremo inferior	Extremo superior	Extremo inferior	Extremo superior
i1	0,58	0,78	0,581	0,776
i2	0,55	0,77	0,547	0,755
i3	0,38	0,65	0,374	0,643
i4	0,28	0,58	0,279	0,577
i5	0,18	0,50	0,177	0,500
I6a	0,58	0,78	0,571	0,770
I6b	0,49	0,71	0,467	0,705
I6c	0,37	0,65	0,366	0,638
I6d	0,23	0,54	0,225	0,537
I7	0,32	0,61	0,317	0,604
I8	0,64	0,81	0,630	0,805
I9	0,29	0,59	0,287	0,582
I10	0,11	0,46	0,114	0,450
I11	0,44	0,70	0,440	0,687
I12	0,53	0,75	0,531	0,745
I13	0,43	0,68	0,430	0,680
I14	0,41	0,67	0,409	0,667
I15	0,42	0,67	0,414	0,670
I16	0,43	0,68	0,421	0,675
I17a	0,19	0,51	0,189	0,510
I17b	0,27	0,57	0,265	0,566
I18	0,05	0,40	0,123	0,457

Tabla 5.13. Correlaciones entre las dos pasaciones para la puntuación total en el cuestionario (*n*=112)

		Correlación	Sig. (bilateral)
Rho de Spearman		0,871(**)	0,000
Pearson		0,861(**)	0,000
Intervalo confianza 95%	Lim. Inferior	0,810	
	Lim. Superior	0,910	
Intervalo credibilidad 95%	Lim. Inferior	0,805	
	Lim. Superior	0,902	

** La correlación es significativa al nivel 0,01 (bilateral).

Asimismo resultaron altamente significativas y de intensidad moderada o alta las correlaciones entre ítems pasados en test y retest. Todo ello asegura una alta estabilidad del cuestionario, medido por la fiabilidad test-retest.

Fiabilidad entre formas paralelas

Los valores de los coeficientes de fiabilidad de formas paralelas (consideradas las dos pasaciones como formas paralelas del test) arrojan valores muy altos, siendo también suficientemente altos y superiores a los mínimos requeridos por los diversos autores citados con anterioridad.

Tabla 5.14. Estadísticos de fiabilidad de formas paralelas

Alfa de Cronbach	Parte 1	Valor	0,781
		N de elementos	22(a)
	Parte 2	Valor	0,772
		N de elementos	22(b)
	N total de elementos		44
Correlación entre formas			0,864
Intervalo confianza 95%	Lim Inferior		0,810
	Lim Superior		0,910
Intervalo credibilidad 95%	Lim Inferior		0,809
	Lim Superior		0,904
Coeficiente de Spearman-Brown	Longitud igual		0,916
	Longitud desigual		0,916
Dos mitades de Guttman			0,914

Los estadísticos de las dos versiones de la escala, tanto en su puntuación total, como en lo que se refiere a los elementos, sus covarianzas y correlaciones son muy similares, aportando otro dato a la estabilidad de la respuesta en pasaciones sucesivas del cuestionario.

Tabla 5.15. Estadísticos de la escala para cada una de las formas

	Media	Varianza	Desviación típica	N de elementos
Parte 1	20,8964	25,655	5,06505	22(a)
Parte 2	21,0179	23,360	4,72864	22(b)
Ambas partes	41,2143	88,458	9,40522	44

Tabla 5.16. Estadísticos de resumen de los elementos

		Media	Mínimo	Máximo	Varianza	Nº elementos
Medias de los elementos	Parte 1	0,928	0,143	2,973	0,507	22(a)
	Parte 2	0,945	0,098	2,982	0,514	22(b)
	Ambas partes	0,937	0,098	2,982	0,499	44
Varianzas de los elementos	Parte 1	0,475	0,067	2,521	0,352	22(a)
	Parte 2	0,472	0,059	2,503	0,336	22(b)
	Ambas partes	0,473	0,059	2,521	0,336	44
Covarianzas inter-elementos	Parte 1	0,133	0,104	0,576	0,007	22(a)
	Parte 2	0,136	0,191	0,516	0,006	22(b)
	Ambas partes	0,136	0,191	0,708	0,011	44
Correlaciones inter-elementos	Parte 1	0,165	0,108	0,405	0,017	22(a)
	Parte 2	0,169	0,105	0,409	0,017	22(b)
	Ambas partes	0,171	0,.115	0,429	0,021	44

Finalmente los coeficientes de Guttman para el conjunto de las dos escalas en los sujetos que la pasaron (n=112) dan de nuevo valores muy altos, que se acercan o superan a 0,9, indicando una gran estabilidad de las respuestas de los estudiantes en ocasiones sucesivas.

2.3. DISCUSIÓN DEL ESTUDIO 5

El objetivo del estudio 5 (dividido en dos partes) fue aportar evidencias de fiabilidad del cuestionario RPC, lo cual se llevó a cabo con diferentes métodos y perspectivas teóricas:

1. En el estudio 5.1 llevado a cabo sobre una muestra de 591 sujetos proporcionó las siguientes estimaciones:
   - Un valor del coeficiente Alfa = 0,797 que indica una consistencia bastante alta, teniendo en cuenta la complejidad del constructo *comprensión de la probabilidad condicional*. Los coeficientes Lambda varían en un rango de 0,77 a 0,83.
   - Se calcularon dos coeficientes basados en los resultados del análisis factorial que se presentará en el estudio 6, obteniendo valores $\theta = 0,82$ y $\Omega = 0,896$ que, además de una alta fiabilidad, serán indicadores posteriormente de una cierta dimensionalidad y se utilizarán en el estudio de la validez de constructo.

- Se aplicó la teoría de generalizabilidad, lo que permitió calcular dos coeficientes; uno de los cuales (generalizabilidad a otros ítems) coincide con el coeficiente Alfa y el segundo de los cuales, $G_i = 0{,}987$ muy elevado indica unas altas posibilidades de generalización de los resultados cuando la misma prueba se pase a otros alumnos del mismo contexto socio-educativo.

2. Una submuestra de 112 estudiantes a los que se pasó dos veces el cuestionario en condiciones muy similares sirvió para dar las siguientes estimaciones de la fiabilidad test- retest:
   - Se obtuvo altos valores de los coeficientes de correlación de Pearson (0,871) y Spearman (0,861) para las puntuaciones totales en las dos pasaciones del cuestionario, que fueron acompañadas de correlaciones entre ítems respectivos estadísticamente significativa y de intensidad alta o moderada.
   - El análisis de fiabilidad por el método formas equivalentes, consideradas las dos pasaciones del cuestionario como formas equivalentes arroja unos valores del Coeficiente de Spearman-Brown = 0,916; y Guttman = 0,914 y correlación entre formas 0,864. Asimismo se obtienen estadísticos muy similares tanto para los totales como para los ítems y sus intecorrelaciones y covarianzas en las dos pasaciones.
   - Se completa el análisis con la estimación clásica y bayesiana de los coeficientes de correlación que aparecen en el estudio de fiabilidad de prueba repetida y de consistencia interna, incluyendo dichos coeficientes de fiabilidad.

Todas estas aportaciones sugieren una alta fiabilidad del cuestionario RPC. Resultados parciales se publicaron en Díaz y de la Fuente (2005a; 2006).

3. ESTUDIO 6. VALIDACIÓN DEL CUESTIONARIO RPC

El objetivo del estudio 6, dividido en varios sub-estudios, fue completar el proceso de validación del cuestionario RPC, que se inició con el análisis de la validez de contenido en el estudio 3[21]. Como se indicó entonces, la validación es un proceso por el cual se aportan evidencias que apoyen la interpretación propuesta de los datos recogidos

[21] En consecuencia no se hace una separación entre validez de contenido y constructo sino que la primera se incluye como parte de la valides de constructo (Pedhazur y Schmelkin, 1991).

mediante la prueba (Carmona, 2004) y ha de estar basadas tanto en los procesos de respuesta, como en las consecuencias de la evaluación (AERA/ APA/ NCME, 1999).

En esta investigación se asume una visión unificada del concepto de validez propuesto por Messick (1989; 1995; 1998)[22] y recogido en los citados estándares (APA, AERA, NCME, 1999). Según estos estándares, el proceso de validación se concibe como el análisis de la significación de las puntuaciones de los instrumentos de medida expresada en términos de los conceptos psicológicos asumidos en su medición. En lo que sigue, se completará la acumulación de evidencias de validez iniciada en el estudio 3, mediante el análisis de la validez de criterio y constructo.

La aproximación a la validez se realiza mediante dos métodos complementarios:

a) Determinación de la capacidad del instrumento para discriminar entre grupos de estudiantes de psicología antes y después de la enseñanza del tema de probabilidad condicional (estudio 6.1).
b) Análisis de la estructura de las respuestas al cuestionario y comparación con la estructura supuesta del constructo (estudio 6.2).

En lo que sigue se describen la metodología y resultados de estos estudios, finalizando con la discusión de las evidencias de validez acumuladas para el cuestionario RPC[23].

3.1. ESTUDIO 6.1. VALIDEZ REFERIDA A CRITERIO

Para aportar evidencias de validez referida a criterio[24], se estudiaron las relaciones del constructo “comprensión de la probabilidad condicional” con otros constructos, operacionalizándolos mediante estudios de las correlaciones[25] del test con otras medidas o criterios (Martínez Arias, 1995).

---

[22] La validez de constructo es el concepto unificador que integra las consideraciones de validez de contenido y de criterio en un marco común para probar hipótesis acerca de relaciones teóricamente relevantes» (Messick, 1989).

[23] Concordamos con Messick (1989) que la validez es una cuestión de grado y el proceso de validación no es definitivo, pudiéndose aportar nuevas evidencias de validez en el futuro.

[24] Correlación con otras medidas o criterios, tomadas simultáneamente (concurrente) o posteriormente (predictiva) (Morales, 1988). Indica la eficacia del test en la predicción de algún tipo de comportamiento del sujeto en situaciones específicas (Santisteban, 1990).

[25] Entendiendo aquí correlación en sentido amplio, como “asociación”, medida bien con coeficientes de correlación, diferencias o discriminación.

Un primer criterio usado fue el estudio previo del tema, puesto que la dificultad del constructo "razonamiento sobre probabilidad condicional", analizada en la definición semántica de la variable (estudio 1) hace difícil un desarrollo espontáneo de este constructo en los estudiantes.

Dicho criterio se operacionalizó, mediante una variable dicotómica, consistente en haber tenido o no instrucción específica en el tema el mismo curso en que se pasó el cuestionario[26]. En esta investigación se está interesado en que el instrumento discrimine estos dos tipos de estudiantes, es decir se trataría de validez discriminante (Santisteban, 1990)[27]. Si el cuestionario tiene esta característica se obtendría otro indicio acumulativo de validez[28]. El indicador elegido cumple los criterios propuestos por Thorndike (1989): es relevante, libre de sesgos, fiable y accesible. Podría considerarse dentro del método de validación de *grupos contrastados* (Martínez Arias, 1995)[29].

Para establecer la relación de la puntuación en RPC con dicha variable, se analizó la capacidad de discriminación del cuestionario respecto a grupos de estudiantes similares antes y después del estudio del tema de probabilidad condicional. Puesto que la medida del criterio (si el sujeto tuvo o no instrucción) se toma simultáneamente a la del constructo, se trata de una validez concurrente[30] (Barbero, 2003). Además, la finalidad es hacer un diagnóstico (sobre los conocimientos de los sujetos).

En lo que sigue se describe la muestra, procedimiento, análisis y resultados de esta comparación.

3.1.1. SUJETOS

Para que las diferencias entre grupos proporcionen evidencia concluyente sobre la validez de la prueba, los grupos han de ser similares en el resto de atributos (Thorndike, 1991).

---

[26] Auque algunos alumnos podrían haber estudiado probabilidad condicional en Bachillerato, la profundidad e intensidad de esta posible instrucción no es comparable a la que se hace en el curso de Análisis de datos, que comprende las unidades semánticas incluidas en la definición semántica de la variable. Además, el tiempo (2 años) transcurrido predice el olvido del tema.

[27] El interés no es sólo conocer si el test discrimina o no entre los dos grupos de sujetos sino explicar la naturaleza de las diferencias.

[28] Santisteban (1990, p. 198) denomina validez discriminante a esta característica.

[29] Esta autora indica que en ocasiones esta validación se aplica cuando la teoría sugiere un cambio en el constructo frente a ciertas intervenciones experimentales; en nuestro caso sería la instrucción.

[30] Medida de las relaciones del test con otras medidas que evalúan el mismo constructo (Anastasi y Urbina, 1998).

En consecuencia se eligieron dos muestras de estudiantes de primer año de psicología que cursaban la asignatura de análisis de datos. Los estudiantes cursaban sus estudios de primer año de psicología en las Universidades de Granada, Huelva, Jaén y Murcia. La mayoría de los alumnos tienen una edad de 18-19 años. La principal diferencia entre los grupos es que el primero de ellos (grupo con instrucción, n = 208) habían tenido enseñanza del tema específico de probabilidad condicional y teorema de Bayes en el primer cuatrimestre, y lo habían estudiado como parte del examen parcial, realizado en Febrero, antes de tomar los datos. En el segundo grupo (grupo sin instrucción, n = 177) el tema todavía no se había explicado, pues correspondía al segundo cuatrimestre. No obstante todos los alumnos habían ya estudiado la probabilidad simple.

3.1.2. MATERIAL Y MÉTODO

El material puesto a prueba fue el cuestionario RPC y el procedimiento de recogida de datos fue similar al ya descrito en el análisis de fiabilidad (estudio 5); es decir, se dispuso de dos formas equivalentes del cuestionario, distribuidas aleatoriamente a los alumnos participantes. Las instrucciones para cumplimentarlo, tiempo disponible e instrucciones fueron iguales a las ya descritas en las secciones anteriores.

Recogidos los datos se codificaron con el mismo criterio y se revisaron para asegurar la fiabilidad. Al calcular el coeficiente de fiabilidad de Cronbach para el total de las dos muestras se obtuvo un valor α = 0,804; mientras que el valor para el primer grupo fue α= 0,753 y para el segundo α= 0,762, valores que consideramos suficientes para llevar a cabo la comparación.

3.1.3. ANÁLISIS

Se comparó en primer lugar la diferencia de puntuación total media en los dos grupos de estudiantes, mediante el estudio de los estadísticos descriptivos, intervalos de confianza y credibilidad y contraste T de diferencia de medias independientes.

Se completó con un análisis discriminante[31], tomando como variable dependiente el criterio (haber tenido o no instrucción en probabilidad condicional) y como conjunto de variables dependientes las puntuaciones tipificadas de los ítems, para dar igual peso a todos ellos. Se dio igual probabilidad inicial a los dos grupos (puesto que sus tamaños son aproximadamente iguales). Se incluyeron simultáneamente todas las variables en la ecuación, pues la finalidad era la discriminación del total del cuestionario (no de un subconjunto de ítems).

Para este análisis discriminante se calculan las funciones discriminantes y probabilidad de clasificación correcta. También se realizan análisis univariados de las variables y cálculo de estadísticos en cada uno de los grupos.

3.1.4. RESULTADOS

Distribución del número de ítems correctos por grupo

En primer lugar se analiza el número total de ítems completamente correctos en el cuestionario en los dos grupos. Como se observa en la tabla 5.17, el grupo de alumnos con instrucción tuvo en promedio 4,3 puntos más y un valor similar del error típico en los dos grupos, lo que indica una primera evidencia de discriminación del cuestionario entre grupos.

Tabla 5.17. Número de ítems totalmente correctos por grupo

Grupo	*N*	Media	Desviación típ.	Error típ. de la media
Sin Instrucción	177	7,8927	3,7302	0,2804
Con instrucción	208	12,2404	3,8947	0,2700

Estos resultados se confirman en la Tabla 5.18 con los resultados de la prueba Levine de igualdad de varianzas, que no nos permite rechazar la hipótesis de igualdad de varianzas. La prueba *T* de diferencia de medias, realizada para el supuesto de varianzas iguales, así como con el intervalo de confianza de la diferencia de medias en los dos grupos permite rechazar la hipótesis de igualdad de puntuaciones medias e indica por tanto un poder discriminante del cuestionario.

[31] Barbero (2003) indica que el análisis discriminante es el método más usado en el estudio de validez de criterio, cuando el criterio es dicotómico.

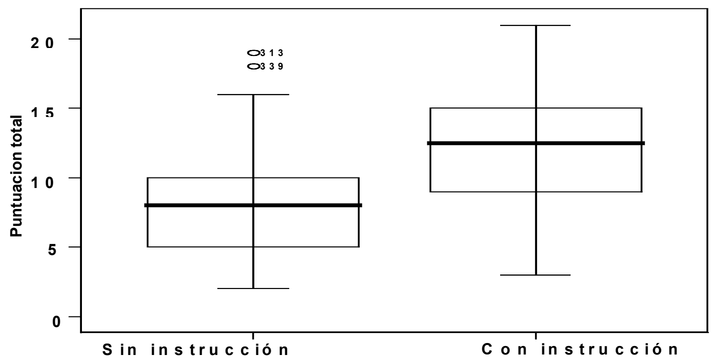
Tabla 5.18. Prueba de muestras independientes

Prueba de Levene		Prueba T para la igualdad de medias						
F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típico	Intervalo de confianza (95%) diferencia	
							Inferior	Superior
3,030	0,083	-11,13	383	9,000	-4,34773	0,39064	-5,1158	-3,5797

Tabla 5.19. Estadísticos descriptivos de la puntuación total por grupos

	Sin instrucción (n=177)		Con instrucción(n=208)	
	Estadístico	Error típ.	Estadístico	Error típ.
Media	7,893	0,280	12,240	0,27005
Límite inferior intervalo confianza 95%	7,339		11,708	
Límite superior Intervalo confianza 95%	8,446		12,773	
Límite inferior intervalo credibilidad 95%	7,263		11,582	
Límite superior Intervalo credibilidad 95%	8,523		12,898	
Media recortada al 5%	7,694		12,229	
Mediana	8,000		12,500	
Varianza	13,915		15,169	
Desv. típ.	3,730		3,894	
Mínimo	2,000		3,000	
Máximo	19,000		21,000	
Rango	17,000		18,000	
Amplitud intercuartil	5,000		6,000	
Asimetría	0,752	0,183	-0,051	0,169
Curtosis	0,219	0,363	-0,853	0,336

Gráfico 5.1. Gráfico de la caja para la puntuación total



Se completa el estudio de la puntuación total con los estadísticos descriptivos por grupo (Tabla 5.19) y la representación gráfica de cajas (Figura 5.1) que muestra los valores claramente superiores de mediana y cuartiles en el grupo con instrucción. Por tanto la diferencia se produce no sólo a nivel de promedios, sino que una distribución supera a la otra en todos sus percentiles.

Discriminación entre grupos

Seguidamente se trató de determinar si existe discriminación sobre el conjunto de ítems del cuestionario RPC considerado como vector, y no sólo en cuanto a la puntuación total en la prueba. Es decir se trata diferenciar los alumnos que tuvieron o no instrucción específica sobre la probabilidad condicional y teorema de Bayes, en el conjunto de ítems. Para ello se llevó a cabo un análisis discriminante, tomando como variable dependiente (dicotómica) el grupo de pertenencia del estudiante y como variables independientes las respuestas (correcta/ incorrecta) en los ítems[32].

En primer lugar se comprobó la aplicabilidad del método. Una de las hipótesis en análisis discriminante (Afifi y Clark, 1990) es la igualdad aproximada de las matrices de covarianzas en los grupos. Esta hipótesis se contrastó mediante la prueba *M* de Box, cuyos datos se recogen en la Tabla 5.20 y donde se obtuvo un valor estadísticamente significativo por el elevado tamaño de muestra, aunque pequeño (*F* aproximadamente igual a 2,1). Se decidió continuar el análisis puesto que el tamaño elevado de muestro concede robustez al método (Cuadras, 1981).

Tabla 5.20. Prueba de Box sobre la igualdad de las matrices de covarianza

Grupo	Rango	Logaritmo del determinante	M de Box	F Aprox.	gl1	gl2
Sin instrucción	22	-38,373	583,306	2,108		
Con instrucción	22	-40,406			253	4243395
Intra-grupos combinada	22	-37,949		Sig.	0,001	

En la Tabla 5.21 se presentan las medias y desviaciones típicas de los ítems (índices de dificultad) en los dos grupos y las pruebas *F* de diferencia de medias. Observamos que se obtiene una diferencia estadísticamente significativa en todos los ítems, excepto en tres y la diferencia es siempre a favor del grupo con instrucción, incluso en los casos en que no fue estadísticamente significativa.

Los ítems que mejor discriminan los dos grupos fueron los siguientes: i12, i11, i15, i13, i16, todos ellos ítems de resolución de problemas con respuesta abierta. También discriminaron bien las distintas partes de los ítems i6 (cálculo de probabilidades a partir de la tabla doble), i1 (definición), i3 (falacia de tasas base) e i4 (independencia). No discriminan los ítems i9 (falacia de la condicional transpuesta) i10 (falacia de la conjunción) y i14 (falacia del eje de tiempos).

[32] En sentido geométrico se trata de hallar un subespacio de dimensión reducida en que la proyección de los sujetos se separe convenientemente y el cociente de la inercia entre grupos se maximice respecto a l inercia dentro de los grupos (Santisteban, 1990).

Tabla 5.21. Estadísticos de grupo

	Media	Desv. típ.	Media	Desv. típ.	Lambda de Wilks	F	Sig.
	Sin instrucción (n=177)		Con instrucción (n=206)				
i1	0,17	0,376	0,47	0,500	0,898	43,429	0,000
i2	0,56	0,698	0,67	0,472	0,991	3,300	0,050
i3	0,33	0,471	0,53	0,500	0,957	17,120	0,000
i4	0,23	0,423	0,41	0,494	0,963	14,776	0,000
i5	0,77	0,419	0,89	0,308	0,973	10,454	0,001
i6a	0,67	0,471	0,94	0,234	0,878	53,075	0,000
i6b	0,29	0,454	0,63	0,483	0,880	52,010	0,000
i6c	0,35	0,478	0,69	0,465	0,887	49,015	0,000
i6d	0,37	0,485	0,70	0,461	0,895	45,134	0,000
i7	0,37	0,485	0,48	0,501	0,989	4,172	0,042
i8	0,20	0,404	0,33	0,472	0,979	8,066	0,005
i9	0,21	0,408	0,24	0,428	0,999	0,535	0,465
i10	0,35	0,478	0,35	0,478	1,000	0,000	0,989
i11	0,18	0,386	0,55	0,498	0,854	65,318	0,000
i12	0,34	0,477	0,75	0,431	0,830	78,523	0,000
i13	0,25	0,437	0,56	0,497	0,903	41,067	0,000
i14	0,08	0,271	0,13	0,332	0,994	2,165	0,142
i15	0,24	0,427	0,59	0,493	0,873	55,787	0,000
i16	0,20	0,754	0,49	0,501	0,949	20,627	0,000
i17a	0,72	0,449	0,81	0,395	0,990	3,863	0,050
i17b	0,37	0,485	0,25	0,434	0,982	6,879	0,009
i18	0,62	0,486	0,76	0,425	0,976	9,463	0,002

Tabla 5.22. Resumen de las funciones canónicas discriminantes. Autovalores

Autovalor	% de varianza	% acumulado	Correlación canónica
0,584(a)	100,0	100,0	0,697

Se ha empleado la primera función discriminante canónica en el análisis.

Mediante la prueba Lambda de Wilks se determinó el poder discriminante del conjunto de variables, puesto que se obtuvo un valor cercano a cero, que se traduce en un alto valor de Chi cuadrado y significación elevada (Cuadras, 1981; Afifi y Clark, 1990). Asimismo, la correlación canónica (0,697) indica un alto poder discriminante y es un indicador de validez de criterio del cuestionario RPC.

Tabla 5.23. Lambda de Wilks

Lambda de Wilks	Chi-cuadrado	gl	Sig.
0,631	171,117	22	0,000

Tabla 5.24. Funciones en los centroides de los grupos

Grupo	Función
Sin instrucción	-0,826
Con instrucción	0,703

En la tabla 5.25 presentamos los coeficientes estandarizados de las funciones discriminantes canónicas[33] (columna 1), coeficientes de la matriz de estructura (columna 2), coeficientes no tipificados de la función canónica discriminante de los grupos (columna 3) y coeficientes de la función de clasificación en los grupos[34], cuyos centroides se presentan en la tabla 5.24.

Tabla 5.25. Resultados del análisis discriminante

Ítem	Coeficiente	Estructura	Función canónica discriminante	Coeficientes de la función de clasificación	
				Sin instrucción	Con instrucción
i1	0,291	0,592	0,651	-0,632	0,364
i2	-0,361	0,540	-0,616	0,490	-0,452
i3	0,032	0,499	0,066	0,620	0,721
i4	0,207	0,487	0,447	1,135	1,819
i5	0,017	0,482	0,048	4,898	4,971
i6a	0,281	0,468	0,774	4,407	5,591
i6b	0,214	0,449	0,456	-0,597	0,101
i6c	0,061	0,441	0,131	0,229	0,429
i6d	0,156	0,428	0,331	0,261	0,767
i7	0,136	0,304	0,276	1,146	1,568
i8	0,048	0,277	0,108	-0,619	-0,454
i9	-0,167	0,257	-0,398	0,127	-0,482
i10	0,063	0,216	0,131	1,740	1,940
i11	0,272	0,206	0,603	-0,739	0,183
i12	0,293	0,190	0,647	-0,066	0,923
i13	0,055	-0,175	0,117	-0,460	-0,281
i14	0,031	0,137	0,100	-0,194	-0,040
i15	0,163	0,131	0,352	-0,622	-0,083
i16	0,056	0,121	0,088	0,054	0,189
i17a	-0,013	0,098	-0,031	3,441	3,394
i17b	-0,186	0,049	-0,405	1,212	0,592
i18	0,074	0,001	0,163	3,006	3,254
(Constante)			-2,061	-7,049	-10,108

La utilidad de la discriminación dependerá de la probabilidad de error en la clasificación, que se produce en base al cálculo de la probabilidad de pertenencia de cada sujeto a su grupo (Peña, 2002). Las probabilidades de clasificación correcta se calculan teniendo en cuenta los sujetos correcta e incorrectamente clasificados y no requieren de la normalidad multivariante (Peña, 2002).

Los resultados de la clasificación (Tabla 5.26) indican un buen poder discriminante del cuestionario respecto a estudiantes que habían recibido o no enseñanza específica sobre probabilidad condicional en el curso, ya que se obtiene un 82% de clasificaciones correctas. Esta alta probabilidad de clasificación correcta constituye una nueva

[33] Proyección de las variables en la dirección que maximice la separación de los centroides; cuyo cálculo no requiere la normalidad multivariante (Peña, 2002).

[34] Esta función permite calcular la distancia de Mahalanobis de cada sujeto al centroide de los dos grupos y clasificarlo en aquél respecto al cuál su distancia sea mínima (Peña, 2002).

evidencia de validez de criterio del cuestionario RPC.

Tabla 5.26. Resultados de la clasificación

		Grupo	Grupo de pertenencia pronosticado		Total
			Sin instrucción	Con instrucción	
Original	Recuento	Sin instrucción	151	26	177
		Con instrucción	42	166	208
	%	Sin instrucción	85,31	14,69	100,0
		Con instrucción	20,19	79,81	100,0

Clasificados correctamente el 82,34% de los casos agrupados originales.

En resumen, el cuestionario RPC discrimina adecuadamente a los estudiantes de psicología que tuvieron o no instrucción específica en probabilidad condicional en el mismo curso en que se pasó el cuestionario, tanto a nivel de número total de respuestas correctas, como mediante el vector de respuestas a los ítems. Individualmente todos los ítems menos tres (relacionados con sesgos psicológicos) también discriminan los dos grupos de estudiantes.

3.2. ESTUDIO 6.2. VALIDACIÓN DE CONSTRUCTO

Mediante la validación de constructo se trató de comprobar las hipótesis teóricas sobre la estructura del constructo que fueron especificadas en la definición sintáctica del constructo (Muñiz, 2004), llevada a cabo en el estudio 3 (capítulo 4).

Entre las técnicas posibles para analizar la validez de constructo la más usada es el análisis factorial o el análisis de componentes principales (Martínez Arias, 1995). Seguidamente se analiza la estructura factorial del cuestionario, para comprobar la dimensionalidad latente e interpretar los factores en términos de las dimensiones supuestas en el cuestionario[35].

3.2.1. SUJETOS

Los sujetos participantes fueron los mismos que los del estudio de fiabilidad de consistencia interna (Estudio 5.1), es decir 591 estudiantes de las Universidades de Granada, Jaén, Murcia y Huelva. Se remite a dicho estudio para la descripción de la muestra.

[35] El test sería validado mediante el análisis de su consistencia interna (Martínez Arias, 1995).

3.2.2. MATERIAL Y MÉTODO

El material puesto a prueba fue el cuestionario RPC y procedimiento de recogida de datos fue similar al ya descrito en el análisis de fiabilidad (estudio 5); es decir, se dispuso de dos formas equivalentes del cuestionario, distribuidas aleatoriamente a los alumnos participantes. Las instrucciones para cumplimentarlo, tiempo disponible e instrucciones fueron iguales a las ya descritas en las secciones anteriores.

3.2.3. ANALISIS

Para comprobar la validez factorial, se realizó un análisis factorial exploratorio del conjunto de respuestas a los ítems del cuestionario RPC para examinar la estructura de las puntuaciones al cuestionario y detectar las fuentes de variación en las medidas observadas (Tabachnick y Fidell, 2001). Se trataba de determinar si se reproducían las dimensiones planteadas teóricamente[36]. Al ser la primera vez que se aplica el análisis factorial, se prefirió una aproximación exploratoria y no confirmatoria. Se trataba de confirmar, por un lado, la existencia de un constructo subyacente que agrupase a la mayor parte de los ítems y que explicase el razonamiento lógico matemático sobre la probabilidad condicional. Al mismo tiempo se esperaba un constructo multidimensional en que algunos de los factores explicasen los sesgos denunciados en la investigación psicológica sobre el razonamiento en probabilidad condicional.

La extracción de factores se llevó a cabo mediante el método de componentes principales; con objeto de obtener factores estadísticamente independientes y de máxima variabilidad al tiempo que no se deforma la estructura de los datos (Cuadras, 1981). Este método parte de una estimación inicial más alta de las comunalidades (Martínez Arias, 1995).

Como método de rotación se usó el Varimax, método ortogonal, que conserva las comunalidades y la suma de porcentajes de varianza explicados por los factores (Afifi y Clark, 1990). Está orientado a maximizar la varianza de los factores.

[36] La idea central es distinguir entre variables observadas medidas por el investigador y variables latentes o factores que explican las anteriores.

3.2.4. RESULTADOS

Antes de aplicar el método, se comprobó la unidad experimental de las variables (todos son ítems referidos al mismo constructo) y número de casos necesarios (se recomiendan al menos 10 por variable; en nuestro caso tenemos aproximadamente tres veces esta cantidad) (Cuadras, 1981).

Se contrastó la hipótesis de que los elementos de la matriz de correlaciones fuera de la diagonal principal son diferentes de cero, mediante la prueba de esfericidad de Barlett. Se obtuvo un valor altamente significativo, lo que indica la existencia de correlaciones altas entre las variables.

Tabla 5.27. KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		0,841
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	2312,960
	gl	231
	Sig.	0,000

5.28. Comunalidades

	Inicial	Extracción
I1. Definición	1,000	0,458
I2. Espacio muestral	1,000	0,407
I3. Teorema Bayes (tasas base)	1,000	0,437
I4. Independencia	1,000	0,527
I5. Probabilidad condicional sin reemplazamiento	1,000	0,488
I6a. Probabilidad simple (tabla)	1,000	0,474
I6b. Probabilidad compuesta (tabla)	1,000	0,529
I6c. Probabilidad condicional (tabla)	1,000	0,670
I6d. Probabilidad condicional (tabla)	1,000	0,698
I7. Probabilidad condicional (de simple y compuesta)	1,000	0,644
I8. Probabilidad condicional; experimento simple	1,000	0,550
I9. Falacia conjunción	1,000	0,513
I10. Condicional transpuesta	1,000	0,639
I11. Probabilidad total	1,000	0,628
I12. Probabilidad condicional; con reemplazamiento	1,000	0,476
I13. Probabilidad compuesta independencia	1,000	0,556
I14. Eje temporal	1,000	0,566
I15. Probabilidad compuesta dependencia	1,000	0,634
I16.Teorema de Bayes	1,000	0,653
I17a. Probabilidad condicional sin reemplazamiento	1,000	0,657
I17b. Eje temporal, diacrónico	1,000	0,633
I18. Probabilidad condicional, sin reemplazamiento	1,000	0,440

Método de extracción: Análisis de Componentes principales.

La prueba de esfericidad de Barlett estadísticamente significativa permite rechazar la hipótesis de que la matriz de correlaciones tenga un problema de multicolinearidad o se reduzca a la matriz identidad (Peña, 2002). El valor del determinante de la matriz de correlaciones fue 0,014, que es otro comprobante de altas correlaciones entre ítems[37]. La prueba de Kaiser-Meyer-Olkin dio un valor mayor que 0,6 recomendado para poder llevar a cabo un análisis factorial con garantías. Todo ello garantizó la aplicabilidad del método.

En el análisis de componentes principales la comunalidad toma como valor inicial 1. En la tabla 5.28 se presentan las comunalidades obtenidas que oscilan entre 0,407 (ítem 2, espacio muestral) y 0,698 (cálculo de probabilidad condicional a partir de una tabla de doble entrada) lo que indica la variabilidad de cada ítem que es explicada por el conjunto de factores retenidos[38] (Afifi y Clark, 1990). Esto indica que cada ítem tiene una parte específica fuerte, en especial alguno de ellos.

Tabla 5.29 Varianza total explicada por cada uno de los factores extraidos

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	4,903	22,288	22,288	4,903	22,288	22,288	4,903	22,288	22,288
2	1,366	6,211	28,499	1,366	6,211	28,499	1,366	6,211	28,499
3	1,303	5,921	34,420	1,303	5,921	34,420	1,303	5,921	34,420
4	1,195	5,432	39,852	1,195	5,432	39,852	1,195	5,432	39,852
5	1,113	5,060	44,912	1,113	5,060	44,912	1,113	5,060	44,912
6	1,069	4,857	49,769	1,069	4,857	49,769	1,069	4,857	49,769
7	1,031	4,684	54,453	1,031	4,684	54,453	1,031	4,684	54,453
8	0,994	4,520	58,365						
9	0,951	4,323	62,688						
10	0,909	4,132	66,820						
11	0,896	4,074	70,894						
12	0,805	3,658	74,552						
13	0,789	3,588	78,140						
14	0,694	3,153	81,293						
15	0,679	3,087	84,380						
16	0,626	2,845	87,225						
17	0,585	2,661	89,886						
18	0,544	2,474	92,360						
19	0,510	2,320	94,680						
20	0,427	1,941	96,621						
21	0,383	1,739	98,360						
22	0,361	1,640	100,000						

Se usaron múltiples métodos para determinar cuántos factores extraer. La extracción inicial obtuvo 7 factores[39] con autovalor mayor que 1, que explicaron el 54,5% de la

[37] Un valor alto del determinante indicaría que las variables están incorrelacionadas (Peña, 2002).

[38] Pues los factores únicos sólo afectan a cada variable.

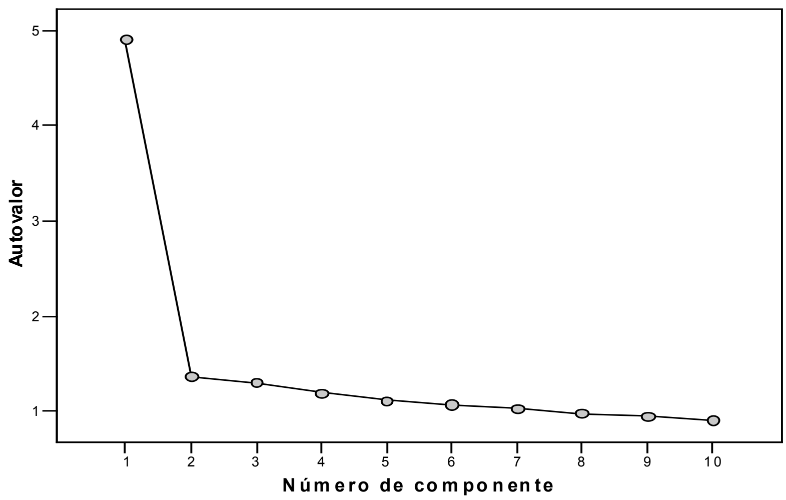
[39] Morales (1988) indica que si todos o casi todos los ítems tienen un peso alrededor o mayor a 0,3 en el primer factor es un indicio de que los ítems se refieren al mismo constructo.

varianza total (tabla 5.29)[40]. También se sigue la sugerencia de Peña (2002) de que el número máximo de factores a extraer ha de ser menor a la mitad del número inicial de variables menos 1.

El primer factor explica un 22,3% de la varianza, mientras que los siguientes entre el 4,6 y el 6,3 cada uno, lo que indica la importancia relativa del primer factor. Este porcentaje proporciona una evidencia de validez del constructo, en cuanto que no hay acuerdo sobre el porcentaje de varianza mínima que debe explicar el primer factor[41] pero si que este porcentaje debe ser claramente superior al explicado por los restantes (Morales, 1988).

En el gráfico de sedimentación (scree plot) (figura 5.2) se observa de nuevo que la mayor varianza es debida al primer factor (que interpretaríamos como razonamiento sobre probabilidad condicional). Por otro lado el punto de inflexión se produce en el quinto factor, lo que sería también un criterio para decidir el número de factores a retener, pues a partir de él los autovalores apenas cambian de magnitud (Martínez Arias, 1995).

Figura 5.2. Gráfico de sedimentación



Incluimos la matriz no rotada de componentes, donde las variables se presentan ordenadas según la importancia relativa de su contribución al primer factor.

---

[40] Los factores cuyo autovalor es menor que la unidad suelen descartarse.

[41] Carmines y Zeller (1979) proponen el 40 %, pero Morales, a partir de la revisión de varios trabajos sugiere el 20% e indica que esto puede ser compatible a valores del coeficiente de fiabilidad cercanos o superiores a 0,8 como es nuestro caso.

Tabla 5.30. Matriz no rotada de componentes

	Componente						
	1	2	3	4	5	6	7
I15. Probabilidad compuesta dependencia	0,72	-0,31	0,00	-0,04	-0,04	0,07	0,09
I11. Probabilidad total	0,68	-0,33	-0,04	0,03	-0,10	0,18	0,09
I13. Probabilidad compuesta independencia	0,67	-0,31	0,00	0,08	0,02	0,00	0,03
I16.Teorema de Bayes	0,67	-0,35	-0,10	-0,20	0,02	0,12	0,12
I12. Probabilidad condicional; con reemplazamiento	0,64	-0,01	0,04	0,10	0,02	-0,07	-0,10
I6b. Probabilidad compuesta (tabla)	0,57	0,31	-0,17	-0,11	-0,10	0,09	0,17
I2. Espacio muestral	0,55	-0,08	0,00	0,11	-0,18	-0,17	-0,15
I6c. Probabilidad condicional (tabla)	0,54	0,50	-0,22	-0,14	-0,03	0,22	-0,03
I6a. Probabilidad simple (tabla)	0,51	0,41	-0,18	0,02	-0,04	-0,04	0,03
I1. Definición	0,50	-0,04	0,04	0,12	-0,10	-0,19	-0,38
I8. Probabilidad condicional; experimento simple	0,44	0,10	0,20	0,32	0,05	-0,25	-0,31
I6d. Probabilidad condicional (tabla)	0,48	0,55	-0,10	-0,23	0,17	0,12	-0,14
I17b. Eje temporal, diacrócnico	-0,05	0,13	0,67	-0,24	-0,22	0,07	-0,21
I14. Eje temporal	0,15	0,04	0,61	0,06	-0,17	0,34	0,14
I10. Condicional transpuesta	-0,01	0,16	0,19	0,63	0,28	0,23	-0,15
I17a. Probabilidad condicional sin reemplazamiento	0,19	0,01	0,35	-0,51	0,11	-0,36	-0,20
I18. Probabilidad condicional, sin reemplazamiento	0,25	0,00	0,12	0,12	-0,57	0,17	0,04
I4. Independencia	0,19	-0,11	0,20	-0,11	0,54	0,31	-0,21
I3. Teorema Bayes (tasas base)	0,31	-0,18	0,10	-0,13	0,43	0,13	0,18
I5. Probabilidad condicional sin reemplazamiento	0,34	-0,01	0,03	0,09	0,14	-0,55	0,14
I7. Probabilidad condicional (de simple y compuesta)	0,08	0,21	0,35	-0,11	0,04	-0,20	0,56
I9. Falacia conjunción	0,27	0,20	0,21	0,38	0,15	-0,14	0,41

Incluso antes de la rotación se observa que diferentes variables contribuyen a diversos factores (altas y bajas correlaciones con ellos). Pero también se observa que la mayoría de los ítems contribuyen al primer factor y un número apreciable de ellos tiene un peso importante en el mismo, lo cual contribuye una nueva evidencia de existencia del constructo subyacente (Morales, 1988). No obstante, puesto que el método de extracción trata de maximizar la varianza de los primeros factores, puede considerarse que la matriz factorial no rotada contiene factores no puros, en el sentido de que cada factor explica su varianza y parte del siguiente, por lo que se aconseja hacer una rotación para facilitar la interpretación (Peña, 2002).

Con objeto de obtener una estructura mas simple[42] se realizó una rotación Varimax, que maximiza la varianza de los coeficientes que definen los efectos de cada factor sobre las variables observadas (Peña, 2002). Tras la rotación (ver tabla 5.31) la estructura se clarifica, ya que las variables con correlaciones negativas prácticamente desaparecen o los valores de la correlación son muy pequeños (por debajo de 0,3), mientras que las correlaciones positivas fuertes en cada factor se mantienen o incluso

[42] La finalidad de la rotación es facilitar la interpretación de los factores, asociando éstos a bloques de variables interrelacionadas.

crecen[43].

Tabla 5.30. Matriz de componentes rotados

	Componente						
	1	2	3	4	5	6	7
I16.Teorema de Bayes	0,76	0,17	0,03	-0,04	0,00	0,15	0,12
I11. Probabilidad total	0,76	0,14	0,13	0,06	-0,01	-0,06	0,00
I15. Probabilidad compuesta dependencia	0,75	0,16	0,19	0,04	0,06	0,05	0,06
I13. Probabilidad compuesta independencia	0,67	0,09	0,29	-0,02	0,07	-0,02	0,08
I12. Probabilidad condicional; con reemplazamiento	0,43	0,26	0,42	0,03	0,09	-0,01	0,07
I6c. Probabilidad condicional (tabla)	0,17	0,79	0,06	0,02	-0,03	-0,03	0,01
I6d. Probabilidad condicional (tabla)	0,02	0,77	0,15	0,04	-0,01	0,09	0,22
I6b. Probabilidad compuesta (tabla)	0,32	0,61	0,03	-0,01	0,15	0,02	-0,10
I6a. Probabilidad simple (tabla)	0,15	0,61	0,21	-0,07	0,14	-0,03	-0,09
I8. Probabilidad condicional; experimento simple	0,10	0,12	0,67	0,08	0,09	-0,09	0,05
I1. Definición	0,27	0,15	0,59	0,02	-0,12	0,06	-0,04
I2. Espacio muestral	0,40	0,15	0,45	0,01	0,01	0,05	-0,14
I17b. Eje temporal, diacrócnico	-0,17	-0,04	0,12	0,71	-0,02	0,27	0,03
I14. Eje temporal	0,15	-0,01	-0,04	0,70	0,18	-0,14	0,04
I7. Probabilidad condicional (de simple y compuesta)	-0,02	0,07	-0,15	0,19	0,66	0,17	-0,01
I9. Falacia conjunción	0,09	0,11	0,16	0,04	0,62	-0,29	0,02
I5. Probabilidad condicional sin reemplazamiento	0,14	0,01	0,39	-0,25	0,44	0,20	-0,01
I17a. Probabilidad condicional sin reemplazamiento	-0,01	0,04	0,26	0,18	0,08	0,66	0,23
I10. Condicional transpuesta	-0,17	-0,02	0,27	0,13	0,06	-0,65	0,25
I4. Independencia	0,14	0,05	0,05	0,13	-0,12	-0,06	0,68
I3. Teorema Bayes (tasas base)	0,34	0,05	-0,07	-0,02	0,19	0,03	0,48
I18. Probabilidad condicional, sin reemplazamiento	0,27	0,11	0,06	0,35	-0,05	-0,11	-0,46

La rotación ha convergido en 13 iteraciones.

Para facilitar la interpretación, se repiten los resultados en la Tabla 5.32, suprimiendo las correlaciones menores a 0,3 con lo que se ve más fácilmente la estructura. Para elegir la matriz final se siguieron los criterios (Martínez Arias, 1995):

- Tomar sólo factores que sean interpretables para el investigador (Afifi y Clark, 1990).
- Cada fila de la matriz rotada tenga al menos un cero; es decir para cada variable debe haber al menos un factor que no contribuya a su varianza.
- Para cada factor, habrá un conjunto de variables cuyas saturaciones se aproximen a cero (Afifi y Clark, 1990).
- Si hay cuatro o más factores comunes, hay un conjunto de variables que tienen saturaciones próximas a cero en las dos columnas (Martinez Arias, 1995).
- Cada factor debe tener peso importante de al menos dos variables, pues de otro modo sería un factor específico (Afifi y Clark, 1990). Los factores se interpretan a continuación.

[43] La rotación ortogonal no cambia la relación entre las variables entre sí, sólo la cercanía de las variables a cada factor (Afifi y Clark, 1990).

Tabla 5.31. Matriz de componentes rotados simplificada

	Componente						
	1	2	3	4	5	6	7
I16.Teorema de Bayes	0,76						
I11. Probabilidad total	0,76						
I15. Probabilidad compuesta dependencia	0,75						
I13. Probabilidad compuesta independencia	0,67						
I12. Probabilidad condicional; con reemplazamiento	0,43		0,42				
I6c. Probabilidad condicional (tabla)		0,79					
I6d. Probabilidad condicional (tabla)		0,77					
I6b. Probabilidad compuesta (tabla)	0,32	0,61					
I6a. Probabilidad simple (tabla)		0,61					
I8. Probabilidad condicional; experimento simple			0,67				
d I1. Definición			0,59				
I2. Espacio muestral	0,40		0,45				
I17b. Eje temporal, diacrócnico				0,71			
I14. Eje temporal				0,70			
I7. Probabilidad condicional (de simple y compuesta)					0,66		
I9. Falacia conjunción					0,62		
I5. Probabilidad condicional sin reemplazamiento			0,39		0,44		
I17a. Probabilidad condicional sin reemplazamiento						0,66	
I10. Condicional transpuesta						-0,65	
I4. Independencia							0,68
I3. Teorema Bayes (tasas base)	0,34						0,48
I18. Probabilidad condicional, sin reemplazamiento				0,35			-0,46

El primer factor explica el 22,28% de la varianza. Este alto porcentaje de varianza explicado refleja la importancia del primer factor, al cual contribuyen la mayoría de los ítems de respuesta abierta. En particular el problema relacionado con el Teorema de Bayes, presenta la contribución mayor, seguido del problema de probabilidad total y los problemas de probabilidad compuesta. Todos estos problemas requieren un proceso de resolución de al menos dos pasos, de los que el primero es el cálculo de una probabilidad condicional, que se utiliza en los siguientes pasos (ej. regla del producto). Podemos interpretar este factor como la *habilidad de resolver problemas de probabilidad condicional complejos*.

El segundo factor explica un 6,2% de la varianza. El cálculo de una probabilidad simple, probabilidad conjunta y probabilidad condicional a partir de una tabla de doble entrada (ítem 6) aparece en este componente separado, probablemente porque el *formato de la tarea* afectara a la ejecución, un hecho que también fue informado por Ojeda (1996) y Gingerenzer (1994) entre otros investigadores.

Un tercer factor (5,92% de la varianza) muestra la relación entre la definición, el espacio muestral y el cálculo de probabilidades condicionales en *situaciones con y sin reemplazamiento*; esto es el nivel 4 en la clasificación de Tarr y Jones (1997).

El resto de factores sugieren que los diferentes sesgos que afectan al razonamiento condicional no aparecen relacionados con la ejecución matemática. Cada uno de los sesgos (*condicional transpuesta*, *falacia del eje temporal*, *falacia de la conjunción, confusión entre independencia* y *mutua exclusividad*, así como las situaciones sincrónicas) aparecen además separados del resto. En algunos casos algunos aparecen opuestos o relacionados con algunas unidades semánticas del razonamiento probabilístico condicional. Por ejemplo, la independencia está ligada a la falacia de las tasas base (donde los sujetos tienen que juzgar si os sucesos son independientes o no) y opuesta a la idea de dependencia.

En consecuencia se confirma la hipótesis previa de existencia de un constructo subyacente (definido por el peso comparativamente alto del primer factor antes de la rotación) que se subdivide en diferentes componentes, todos ellos relacionados con el razonamiento de tipo lógico matemático en probabilidad condicional. Al mismo tiempo aparecen otra serie de factores cada uno de ellos identificativo de sesgos particulares en este razonamiento, lo que apunta a nuestra hipótesis de la estructura multidimensional del cuestionario RPC.

3.3. DISCUSIÓN DEL ESTUDIO 6

El estudio 6 en su conjunto se orientó a proporcionar evidencias de la validez de criterio y constructo, que unidas a la validez de contenido discutida en el Estudio 3 y posteriormente completada en el estudio 9 con nuevas aportaciones a la validez de criterio contribuyan a la validación del cuestionario RPC.

El estudio 6.1 estuvo orientado a mostrar evidencias de la validez de criterio del instrumento, eligiendo un criterio dicotómico (haber tenido o no instrucción específica en el curso en que se pasó el cuestionario RPC). Para ello se compararon los resultados en dos grupos de estudiantes equivalentes, salvo en la variable criterio.

El análisis del número total de ítems correctamente resuelto en cada grupo dio valores estadísticamente significativos de la diferencia de medias, medianas y cuantiles, mostrando un desplazamiento de toda la distribución (gráfico de cajas), pero conservando valores similares de los estadísticos de dispersión.

Asimismo, el análisis discriminante mostró una correlación canónica de 0,69 y una probabilidad de clasificación correcta en los grupos del 82% Todos los ítems resultaron discriminativos (diferencias estadísticamente significativas en los grupos), excepto tres relacionados con sesgos de razonamiento en probabilidad condicional. Ello nos indicó que posiblemente dichos sesgos constituyan uno o varios factores diferenciados en el estudio de la validez factorial que se aborda en el estudio 6.2.

En consecuencia, el estudio 6.1 proporciona evidencias de la validez de criterio del cuestionario RPC.

El análisis factorial realizado en el estudio 6.2 sirvió para identificar 7 factores que explicaron en su conjunto un 54% de la varianza y con autovalores mayores que 1. El primer factor tuvo un peso considerablemente mayor que los restantes, aunque sólo explica el 22,3% de la varianza, lo que indica el carácter no unidimensional del cuestionario.

Se interpretaron los tres primeros factores como componentes específicos del razonamiento matemático en probabilidad condicional: el tercero de ellos corresponden al nivel más alto de razonamiento en probabilidad condicional identificado por Tarr y Jones (1997), quienes en su investigación sólo se centraron en escolares. Por ello es posible que no se completase su tipología de razonamientos y pudiese considerar niveles superiores a los considerados por los autores.

Por ejemplo, el primer factor en nuestro estudio contempla con importantes contribuciones la mayor parte de los problemas que requieren dos o más pasos, lo hemos denominado *capacidad de resolución de problemas complejos de probabilidad condicional,* que incluiría, por ejemplo, los problemas de probabilidad compuesta en situaciones con y sin reemplazamiento, problemas de probabilidad total y Bayes. El segundo factor se interpreta como *efecto del formato* ya que los datos se dan en frecuencias naturales y mediante una tabla doble[44].

[44] Tarr y Jones (1997) no estudiaron el efecto del formato.

El resto de los factores corresponde cada uno a los sesgos descritos en razonamiento sobre probabilidad condicional, que aparecen independientes del razonamiento lógico matemático.

Todo ello aporta evidencias de validez de constructo, puesto que, por un lado, los tres primeros factores darían cuenta de la dimensión lógico- matemática del constructo "razonamiento en probabilidad condicional" y el resto de los sesgos de razonamiento que tienen base psicológica, según se hipotetizó en la definición semántica de la variable.

Estos resultados se vieron confirmados posteriormente mediante análisis cluster (Díaz, de la Fuente y Bacelar-Nicolau, en prensa).

CAPITULO 6.
DISEÑO Y VALIDACIÓN DE MATERIAL PARA LA ENSEÑANZA DE CONCEPTOS BÁSICOS DE INFERENCIA BAYESIANA

1. INTRODUCCIÓN

En este Capítulo se describe el diseño de materiales didácticos destinados a la enseñanza de inferencia bayesiana a los estudiantes de primer curso de Psicología (estudio 8.1), así como la validación de dicho material en un experimento llevado a cabo con 78 estudiantes de primer año de dicha licenciatura (estudio 8.2).

El diseño del material de enseñanza se apoyó, entre otros puntos, en un estudio de evaluación de los conocimientos adquiridos por los estudiantes de psicología y sus sesgos de razonamiento sobre la probabilidad condicional, que fue llevado a cabo con el cuestionario RPC, una vez finalizada su validación (estudio 7).

En lo que sigue se describen estos tres estudios y sus resultados.

2. ESTUDIO 7. EVALUACIÓN DE LA COMPRENSIÓN DE LA PROBABILIDAD CONDICIONAL EN ALUMNOS DE PSICOLOGÍA

2.1. INTRODUCCIÓN

El estudio 7 estuvo orientado a evaluar los conocimientos adquiridos por los estudiantes de Psicología sobre probabilidad condicional, una vez estudiado el tema, como parte de la asignatura de Análisis de Datos. La finalidad era comprobar que los estudiantes tenían un conocimiento suficiente para abordar la enseñanza de la inferencia bayesiana, y tener en cuenta los posibles sesgos de razonamiento y errores detectados en el diseño del material de enseñanza.

Para ello se administró el cuestionario RPC, una vez finalizada su validación a una muestra de 414 alumnos después de haber estudiado la probabilidad condicional y se analizaron las respuestas obtenidas desde diversos puntos de vista. En lo que sigue se describen la metodología seguida y los resultados

2.2. SUJETOS

La muestra utilizada en este estudio fueron 414 estudiantes de Psicología de primer curso de las universidades de Granada (cuatro grupos de estudiantes; n = 307), y Murcia (dos grupos; n = 106). Todos los participantes cursaban Primero de Psicología y eran alumnos de la Asignatura de Análisis de Datos. La mayoría de los alumnos tienen una edad de 18 o 19 años. Predominaron las mujeres, situación que es habitual en las Facultades de Psicología. Los estudiantes habían estudiado el tema durante dos semanas, aproximadamente un mes antes que se pasase el cuestionario. Se realizó la toma de datos después del examen parcial que incluía el tema, para asegurarse que los alumnos lo habían estudiado. Los alumnos fueron avisados del contenido de la prueba con suficiente tiempo para prepararse para ella. Todos ellos colaboraron voluntariamente y con interés en la investigación.

2.3. MATERIAL Y MÉTODO

El instrumento de evaluación utilizado fue el cuestionario RPC y el procedimiento de recogida de datos fue similar al ya descrito en el análisis de fiabilidad (estudios 5) y los estudios de validez (estudio 6); es decir, se dispuso de dos formas equivalentes del cuestionario, distribuidas aleatoriamente a los alumnos participantes. Las instrucciones para cumplimentarlo, tiempo disponible e instrucciones fueron iguales a las ya descritas en las secciones que tratan de los estudios de validez y fiabilidad.

2.4. ANÁLISIS

Se realizó un estudio descriptivo de los resultados del estudio de evaluación, categorizando, en primer lugar, las respuestas abiertas, según su nivel de completitud, para darles una puntuación creciente. Para cada ítem y subítem se calcularon las frecuencias y porcentajes de cada respuesta.

En el análisis factorial llevado a cabo como parte de la validación del cuestionario (estudio 6), se identificaron una serie de factores diferenciados en el constructo que se agruparon en dos ejes principales: a) ítems que evalúan los conocimientos de tipo lógico-matemático y b) ítems con sesgos de razonamiento sobre probabilidad condicional[1].

Los resultados del estudio descriptivo se clasifican respecto a los dos componentes

[1] Martínez Arias (1995) indica que los factores identificados en el análisis factorial pueden posteriormente usarse para describir la composición factorial de un test.

citados del razonamiento sobre probabilidad condicional y dentro de ellos, respecto a los factores incluidos en los mismos. Para cada conjunto de ítems dentro de un mismo factor se calcularon las correlaciones de Pearson y no paramétricas.

Se estimaron también los índices de dificultad de los ítems, tanto desde una perspectiva frecuencial como bayesiana, realizando los cálculos en este caso en dos supuestos: a) distribución inicial no informativa; b) distribución inicial informativa, tomando en este último caso como distribución inicial la distribución final obtenida en el estudio piloto del cuestionario[2] (estudio 3). Los procedimientos de estimación y programas de cálculo son los descritos en el Capítulo 3.

2.5. RESULTADOS SOBRE CONOCIMIENTOS LÓGICO-MATEMÁTICOS

Dentro del componente lógico-matemático de describieron tres factores diferenciados, el primero de los cuales (*resolución de problemas complejos de probabilidad condicional*) explicó la mayor proporción de varianza en el análisis factorial. Asimismo se identificaron otros dos factores que denominamos *discriminación entre diversos tipos de probabilidades* y *conocimiento conceptual del tema*.

A continuación se describen con detalle los resultados en cada uno de los ítems que configuran los citados factores.

2.5.1. RESOLUCIÓN DE PROBLEMAS COMPLEJOS DE PROBABILIDAD CONDICIONAL

En primer lugar se presentan las correlaciones bivariadas de Pearson, Kendall y Spearman entre los ítems que contribuyeron específicamente a configurar el primer factor, todos ellos relacionados con resolución de problemas abiertos que involucran al menos dos pasos, uno de los cuales implica el cálculo de una probabilidad condicional. Se observan unas correlaciones altas o moderadas entre todos los pares de ítems, lo que

[2] Se sigue así la metodología bayesiana, donde en cada nueva aplicación del método se toma como distribución inicial la distribución final obtenida en la aplicación anterior. En nuestro caso es lógico asumir que la distribución del índice de dificultad se aproximará la obtenida en la prueba piloto del cuestionario, pues se trata de la misma población de estudiante y las variaciones a los ítems fueron mínimas.

refuerza la interpretación dada al factor en el Estudio 6.

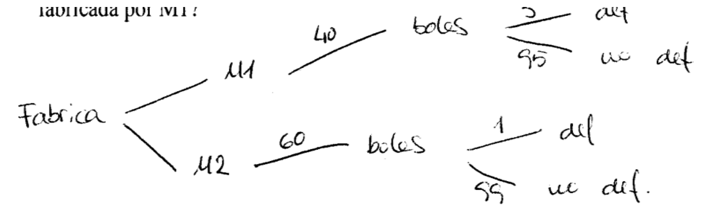
Tabla 6.1. Correlaciones entre ítems que definen el Factor 1

Item		I16. Bayes	I11 P. total	I15 P. compuesta dependencia	I13 P. compuesta independencia	I12. P. Condicional Reemplazamiento
I16.	Correlación de Pearson	1,000	0,584(**)	0,591(**)	0,472(**)	0,344(**)
	Tau_b de Kendall	1,000	0,532(**)	0,523(**)	0,412(**)	0,316(**)
	Rho de Spearman	1,000	0,600(**)	0,594(**)	0,474(**)	0,361(**)
I11	Correlación de Pearson	0,584(**)	1,000	0,565(**)	0,466(**)	0,429(**)
	Tau_b de Kendall	0,532(**)	1,000	0,527(**)	0,418(**)	0,390(**)
	Rho de Spearman	0,600(**)	1,000	0,566(**)	0,457(**)	0,425(**)
I15	Correlación de Pearson	0,591(**)	0,565(**)	1,000	0,559(**)	0,399(**)
	Tau_b de Kendall	0,523(**)	0,527(**)	1,000	0,504(**)	0,361(**)
	Rho de Spearman	0,594(**)	0,566(**)	1,000	0,550(**)	0,393(**)
I13	Correlación de Pearson	0,472(**)	0,466(**)	0,559(**)	1,000	0,428(**)
	Tau_b de Kendall	0,412(**)	0,418(**)	0,504(**)	1,000	0,390(**)
	Rho de Spearman	0,474(**)	0,457(**)	0,550(**)	1,000	0,425(**)
I12.	Correlación de Pearson	0,344(**)	0,429(**)	0,399(**)	0,428(**)	1,000
	Tau_b de Kendall	0,316(**)	0,390(**)	0,361(**)	0,390(**)	1,000
	Rho de Spearman	0,361(**)	0,425(**)	0,393(**)	0,425(**)	1,000

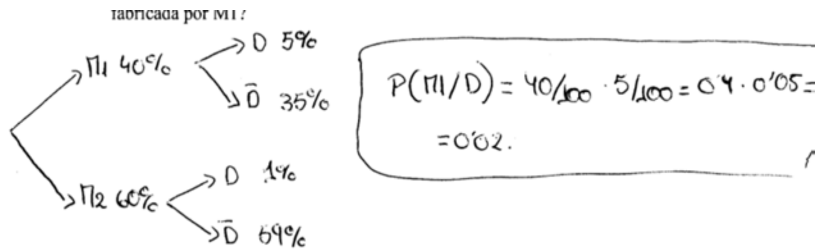
** La correlación es significativa al nivel 0,001 (bilateral).

El ítem que mayor peso tuvo en el factor es el 16 que evalúa la capacidad de resolución de problemas que involucran el teorema de Bayes. Este ítem es el más complejo, puesto que el alumno ha de aplicar el cálculo de probabilidad condicional y compuesta, probabilidad total y probabilidad inversa. Se puntuó siguiendo el siguiente criterio:

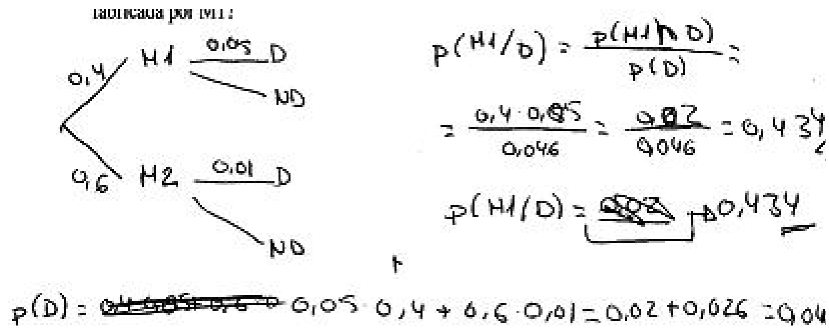
0. No se responde o se responde incorrectamente, no llegándose a la identificación correcta de todos los datos del problema.
1. Se construye un diagrama en árbol adecuado, como en el caso siguiente, en que el alumno no finaliza el problema, aunque identifica los datos pertinentes y construye el diagrama, completando incluso datos faltantes. Ello requiere la identificación correcta del espacio muestral en el experimento producto, la asignación de probabilidades a cada uno de los sucesos que lo componen y la aplicación del axioma de la unión.



2. Diagrama en árbol correcto e identifica el problema como de probabilidad condicional, como en el caso siguiente en que el alumno identifica el problema como de probabilidad condicional, sin aplicar la regla de Bayes. Además de las destrezas del caso anterior, el estudiante muestra que diferencia el concepto y usa la notación de probabilidad condicional, diferenciando una probabilidad condicional de su transpuesta.



3. Calcula correctamente la probabilidad total. Además de los pasos anteriores, el alumno identifica el problema como de probabilidad condicional y llega al cálculo correcto del denominador de la fórmula, aplicando el cálculo de probabilidades compuestas y el axioma de la unión, pero hace un error en el numerador de la regla de Bayes.



4. Resuelve correctamente, como en el caso siguiente, donde el alumno finalmente calcula correctamente el numerador de la fórmula de Bayes.

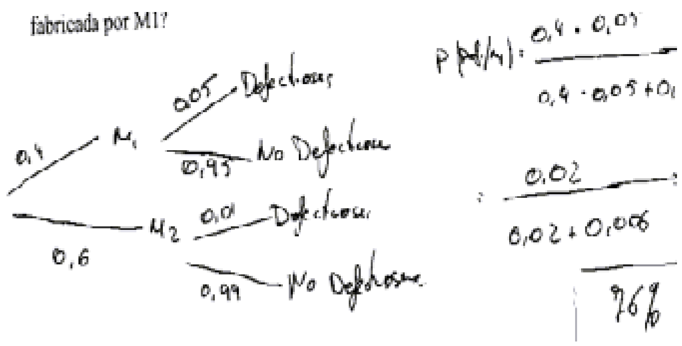


Tabla 6.2. Frecuencias (y porcentajes) de respuestas en el ítem 16 (n=414)

	Frecuencia	Porcentaje	Porcentaje acumulado
0	57	13,8	13,8
1	50	12,1	25,8
2	78	18,8	44,7
3	42	10,1	54,8
4	187	45,2	100,0

Los resultados de la tabla 6.2 sugieren una dificultad moderada, aunque el 45 % de los estudiantes resolvió correctamente el problema[3], y un 55% llegó al menos al cálculo correcto de la probabilidad total. Estas dos respuestas se tuvieron en cuenta como correctas para el índice de dificultad. Son pocos los estudiantes que no llegan a completar el diagrama en árbol correctamente[4]. Hay que tener también en cuenta que el problema se presentó en formato de probabilidad, mientras que Gigerenzer (1994) sugiere que el formato de frecuencias sería más sencillo.

El segundo ítem en orden de importancia en este factor fue el **ítem 11**, que evalúa la capacidad de resolver problemas sobre el teorema de la probabilidad total, y se ha puntuado siguiendo el siguiente criterio:

0. No se responde, o se responde incorrectamente, no llegándose a identificar correctamente los datos del problema: Ejemplo *"130"*.
1. Identifica los datos o construye un diagrama de árbol, pero hace algún error de menor importancia. Ello requiere la identificación correcta del espacio muestral en el experimento producto, la asignación de probabilidades a cada uno de los sucesos que lo componen y la aplicación del axioma de la unión. Por ejemplo el caso siguiente, donde el alumno comienza bien, pero luego olvida que la proporción de hombres y mujeres en la población es distinta: *"60 hombres y 40 mujeres";* $50\text{x}\frac{60}{100}=30;25\text{x}\frac{40}{100}=70$ *; 200 personas; 100 hombres fuman y 50 mujeres fuman; hay 150 personas que fuman de 200".*

[3] Este porcentaje supera al obtenido por Totohasina (1992) con estudiantes del curso preuniversitario (25% de respuestas correctas). Hay que tener en cuenta que éste autor no introdujo formalmente el teorema de Bayes en la enseñanza, aunque si planteó problemas de probabilidad condicional inversa; como éste que fue tomado de dicho autor.

[4] El porcentaje supera bastante al de la investigación de Totohasina (1992), donde apenas la mitad de los estudiantes completan correctamente el árbol.

2. Calcula la probabilidad total de personas que fuman. Además de los pasos anteriores, el alumno identifica el problema como de probabilidad condicional y llega al cálculo correcto del denominador de la fórmula, aplicando el cálculo de probabilidades compuestas y el axioma de la unión. Por ejemplo, el caso siguiente: *"P(f)= 60x50+40x25=0,3+0,1=0,4"*.

Los resultados (tabla 6.3) indican que la mayoría de los alumnos contestan correctamente al ítem 11[5]. En el cálculo del índice de dificultad se consideran correctas las respuestas puntuadas con 1 y 2 puntos.

Tabla 6.3. Frecuencias (y porcentajes) de respuestas en el ítem 11 (*n*=414)

	Frecuencia	Porcentaje	Porcentaje acumulado
0	50	12,1	12,1
1	97	23,4	35,5
2	267	64,5	100,0

El **ítem 15,** tercero en importancia en el factor pide calcular una probabilidad conjunta en el caso de sucesos dependientes. Se ha puntuado siguiendo el siguiente criterio:

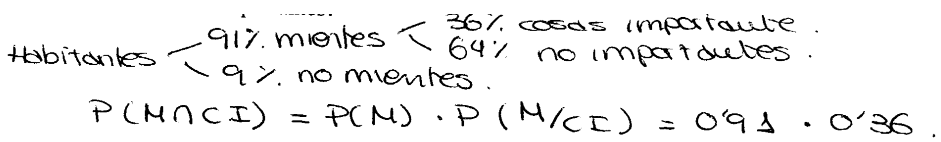
0. No se responde o se responde incorrectamente, no llegando a la identificación de todos los datos, por ejemplo *"55%"*.
1. Identifican los datos o realizan el diagrama en árbol correctamente, como en el caso siguiente en que identifica incorrectamente el problema como de probabilidad condicional, en lugar de como probabilidad conjunta[6].

$$M = 0,91; \bar{M} = 0,09; P(I/M) = 0,36; P(NI/M) = 0,64$$
$$P(M/I) = \frac{I/M}{M} = \frac{0,36}{0,91} = 0,40$$

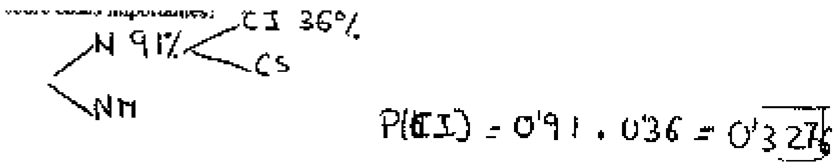
2. Resuelven el problema correctamente empleando la notación adecuada, como en el caso siguiente en que el alumno emplea tanto un diagrama en árbol como la notación correcta y fórmula de la probabilidad condicional, identificando correctamente los datos del problema y aplicando la regla del producto.

[5] El porcentaje es mucho mayor que el de Ojeda (1995) quien obtuvo un 44% de respuestas correctas con estudiantes de secundaria.

[6] Ya hemos señalado repetidamente que los estudiantes confunden las probabilidades simples con conjuntas o compuestas, especialmente en problemas dados en forma verbal (Ojeda, 1995).



En otros casos llegan a la probabilidad pedida pero sin emplear la fórmula explícitamente:



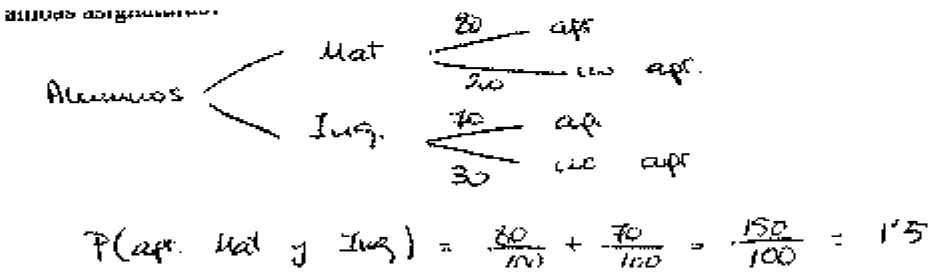
En la tabla 6.4 presentamos los resultados del ítem 15, donde la mayoría de los alumnos resuelve correctamente el problema[7].

Tabla 6.4. Frecuencias (y porcentajes) de respuestas en el ítem 15 (n=414)

	Frecuencia	Porcentaje	Porcentaje acumulado
0	92	22,2	22,2
1	72	17,4	39,6
2	250	60,4	100,0

El **ítem 13,** pide a los alumnos calcular una probabilidad conjunta en el caso de sucesos independientes, y se puntuó siguiendo el siguiente criterio:

0. No se responde o se responde incorrectamente, no llegando a identificar los datos.
1. Construye correctamente el árbol o deja indicada la fórmula, pero no resuelve el problema, como en el siguiente caso en que el alumno construye el diagrama en árbol, pero considera incorrectamente que los sucesos aprobar matemáticas e inglés son incompatibles[8]. Obtiene además una probabilidad mayor que uno y no se da cuenta del problema.



[7] En la investigación de Ojeda (1995) con 255 estudiantes de secundaria el porcentaje de alumnos que resuelve este mismo problema es el 45% En la investigación de Maury (1985), también con estudiantes de secundaria sólo el 25% resuelve correctamente los problemas de probabilidad compuesta para sucesos dependientes.

[8] Confusión entre independencia y muta exclusividad, señalada por Sánchez (1996).

2. Resuelve correctamente. Por ejemplo en el caso siguiente el alumno identifica los datos y la probabilidad pedida y aplica correctamente la fórmula del producto para el caso de sucesos independientes.

$$P(A_1 \cap A_2) = P(A_1) \times P(A_2) = 0,8 \times 0,7$$

En la tabla 6.5 presentamos los resultados del **ítem 13**, donde más de la mitad de los alumnos dan la respuesta correcta, siendo de nuevo los datos superiores a los de otras investigaciones (Maury, 1985; Ojeda, 1995).

Tabla 6.5. Frecuencias (y porcentajes) de respuestas en el ítem 13 (n=414)

	Frecuencia	Porcentaje	Porcentaje acumulado
0	112	27,1	27,1
1	84	20,3	47,3
2	218	52,7	100,0

El **ítem 12**, sobre cálculo de probabilidad condicional en situación de muestreo con reemplazamiento que es el ítem con carga menor en el primer factor, se ha puntuado siguiendo el siguiente criterio:

0. No se responde, o se responde incorrectamente. Por ejemplo, si se indica que el caso siguiente tiene mayor probabilidad de ser un valor impar, argumentando que ya hay muchos valores pares en la secuencia. En este caso se tiene una concepción incorrecta de independencia, ya que se supone que la probabilidad depende de los lanzamientos anteriores[9].
1. Se da una estimación frecuencial de la probabilidad. Corresponde al caso en que el alumno calcula la probabilidad a partir de la frecuencia relativa obtenida en los ensayos que se describen en el ítem, es decir, da cómo valor de la probabilidad el valor 10/15 (hay un total de 10 pares en los quince lanzamientos). En este caso el alumno podría considerar que estamos trabajando con un dado sesgado.
2. Se calcula correctamente la probabilidad, dando el valor 1/2. Los alumnos que proporcionan esta respuesta razonan correctamente que la probabilidad del siguiente lanzamiento no depende de los resultados en los anteriores.

En la tabla 6.6 presentamos los resultados del ítem 12, donde la mayoría da una

[9] Es decir, se razona de acuerdo a la heurística de la representatividad (Kahneman, Slovic y Tversky, 1982), esperando que una sucesión corta de ensayos tenga que reproducir necesariamente el proceso que la genera.

respuesta correcta[10] aunque un 11,8% diera una estimación frecuencial de la probabilidad.

Tabla 6.6. Frecuencias (y porcentajes) de respuestas en el ítem 12 (*n*=414)

	Frecuencia	Porcentaje	Porcentaje acumulado
0	78	18,8	18,8
1	49	11,8	30,7
2	287	69,3	100,0

En resumen, los resultados sobre los ítems que configuran el primer factor indican una buena capacidad de resolución de problemas complejos de probabilidad condicional, puesto que los ítems que configuran este factor son resueltos correctamente por la mitad o más de los estudiantes, llegando en algunos casos a proporciones del 70%. Un dato a destacar es que la mayor parte de estudiantes identifica los datos del problema y construye correctamente el diagrama en árbol, lo que sin duda influye en los buenos resultados en la resolución de los problemas[11].

2.5.2. DISCRIMINACIÓN ENTRE DIFERENTES TIPOS DE PROBABILIDAD Y LECTURA DE TABLAS

En la tabla 6.7 se presentan las correlaciones bivariadas de Pearson, Kendall y Spearman entre los ítems que configuran el segundo factor y que están todos relacionados con la lectura de tablas[12]. En las cuatro preguntas que componen el ítem 6, hay que aplicar la regla de Laplace, deduciendo los datos necesarios de la tabla de doble entrada. Se observan unas correlaciones moderadas entre todos los pares de ítems. Las correlaciones más altas se presentan entre el ítem 6c y 6d, lo que es razonable ya que ambos ítems piden una probabilidad condicionada.

Tabla 6.7. Correlaciones entre ítems que definen el Factor

[10] En este ítem Maury (1986) obtuvo un 70% de respuestas correctas con alumnos del curso preuniversitario, por tanto nuestros resultados reproducen los suyos.

[11] Para Totohasina (1992) el uso de un árbol es el recurso más efectivo para resolver problemas de probabilidad condicional, sobre todo cuando se refiere a un problema diacrónico (dirigido en el tiempo). El mayor porcentaje de estudiantes en nuestra investigación que usan correctamente el diagrama en árbol, respecto al de aquél autor explica los mejores resultados en resolución de problemas.

[12] Lonjedo y Huerta (2004; 2005) señalan la importancia del formato de presentación de las tareas sobre los problemas de probabilidad condicional.

		tabla a	tabla b	tabla c	tabla d
tabla a	Correlación de Pearson	1,000	0,389(**)	0,327(**)	0,314(**)
	Tau_b de Kendall	1,000	0,389(**)	0,327(**)	0,314(**)
	Rho de Spearman	1,000	0,389(**)	0,327(**)	0,314(**)
tabla b	Correlación de Pearson	0,389(**)	1,000	0,383(**)	0,320(**)
	Tau_b de Kendall	0,389(**)	1,000	0,383(**)	0,320(**)
	Rho de Spearman	0,389(**)	1,000	0,383(**)	0,320(**)
tabla c	Correlación de Pearson	0,327(**)	0,383(**)	1,000	0,550(**)
	Tau_b de Kendall	0,327(**)	0,383(**)	1,000	0,550(**)
	Rho de Spearman	0,327(**)	0,383(**)	1,000	0,550(**)
tabla d	Correlación de Pearson	0,314(**)	0,320(**)	0,550(**)	1,000
	Tau_b de Kendall	0,314(**)	0,320(**)	0,550(**)	1,000
	Rho de Spearman	0,314(**)	0,320(**)	0,550(**)	1,000

** La correlación es significativa al nivel 0,001 (bilateral).

En la tabla 6.8 se presentan los resultados de los diferentes apartados del **ítem 6**.

Tabla 6.8. Frecuencias (y porcentajes) de respuestas en ítem 6 (*n*=414)

	a. $P(A)$	b. $P(A \cap B)$	c. $P(A/B)$	d. $P(B/A)$
Incorrecto	23 (5,5)	138 (33,3)	132 (31,3)	133 (32,1)
Correcto	378 (91,3)	261 (66,0)	259 (62,6)	241 (58,2)
Blanco	13 (3,2)	15 (3,7)	23 (5.5)	40 (9,5)

El apartado a resulta bastante sencillo, así como cálculo de una probabilidad conjunta. Mayor dificultad aparece en el cálculo de probabilidades condicionales y distinción entre una probabilidad condicional y su inversa, donde el porcentaje de fallos se acerca al 40%. Los resultados son bastante similares a los obtenidos en nuestros estudios previos con versiones piloto del cuestionario (Díaz y de la Fuente, 2005), así como con otras muestras de futuros profesores de educación primaria y secundaria (Estrada y Díaz, 2006).

Los principales errores fueron la confusión de probabilidades condicionales y conjuntas ya destacados por Einhorn y Hogarth (1986), Totohasina (1992) y Ojeda (1995), que se producen en un 27,1% de los estudiantes[13]. También aparecen en casos aislados los siguientes errores no descritos en la literatura:

- Confusión de un suceso y su complementario, respondiendo, por ejemplo, dar la solución 275/676 ($P(A / \text{no } B)$), cuando se pregunta $P(A/B)$.
- Confusión de probabilidades con casos posibles (frecuencias absolutas), dando, por ejemplo, el valor 75 para $P(A \cap B)$.

[13] Esta misma confusión en la interpretación de tablas de contingencia aparece en el 20% de los estudiantes del curso preuniversitario en la investigación de Batanero y cols. (1996).

- Obtención de probabilidades mayores que la unidad; como responder *P(A∩B)=780/75,* invirtiendo en este caso la fórmula de Laplace.
- Confusión de la unión y la intersección, dando la solución $\mathrm{P}(\mathrm{A} \cap B)=\frac{104}{780}+\frac{350}{780}-\frac{75}{780}$. Hacemos notar que estos alumnos (tres casos) han notado que los sucesos A y B no son mutuamente excluyentes y han aplicado la fórmula correcta de la unión de sucesos para estos casos, pero han dado a la palabra "y" del enunciado un significado de unión, que es frecuente en el lenguaje cotidiano, pero no coincide en este caso con el significado matemático.
- Suponer independencia en los datos, aunque la dependencia es patente en la tabla. Estos alumnos calculan la probabilidad de la intersección, como producto de las dos probabilidades *P(A)* y *P(B)*, calculadas estas correctamente, dando la solución $\mathrm{P}(\mathrm{A} \cap B)=\frac{104}{780} x \frac{350}{780}$.
- Dar el cociente *P(A)/ P(B)* cuando se pide la probabilidad condicional (responder 350/104; en este caso el alumno recuerda que la probabilidad condicional viene dada por un cociente, pero hay un conflicto consistente en confundir el numerador de dicho cociente.

En resumen, estos resultados sugiere la necesidad de atender a la capacidad de lectura de datos en tablas dobles, que será necesaria en el estudio de la inferencia bayesiana, donde se organiza el cálculo precisamente mediante tablas de Bayes.

2.5.3. COMPRENSIÓN CONCEPTUAL DE LA PROBABILIDAD CONDICIONAL

Los ítems que definieron el tercer factor en el análisis factorial presentado en el estudio de validación fueron los siguientes: Definición del concepto (ítem 1), determinación del espacio muestral (ítem 2), cálculo de probabilidad condicional en situaciones con y sin reemplazamiento (ítem 12 y 5) y cálculo de una probabilidad condicional en experimento simple (ítem 8) utilizando directamente la definición.

Consecuentemente, este factor se correspondía con las capacidades definidas para el nivel 4 en la categorización definida por Tarr y Jones (1997) y se interpretó como *conocimiento conceptual de la probabilidad condicional*. En lo que sigue se presentan los resultados en los ítems citados.

Tabla 6.9. Correlaciones entre ítems que definen el Factor 3

		Ítem1. Definición	Ítem 2. E. muestral	Ítem 5. Sin reemplaz.	Ítem 8. Exp. simple	Ítem 12. Con reemplaz.
Ítem 1	Correlación Pearson	1,000	0,449(***)	0,291(***)	0,241(**)	0,217(**)
	Tau_b de Kendall	1,000	0,421(***)	0,254(***)	0,240(**)	0,177(**)
	Rho de Spearman	1,000	0,420(***)	0,287(***)	0,251(**)	0,200(**)
Item 2	Correlación Pearson	0,449(**)	1,000	0,311(***)	0,221(***)	0,239(**)
	Tau_b de Kendall	0,421(***)	1,000	0,302(***)	0,214(***)	0,214(**)
	Rho de Spearman	0,420(***)	1,000	0,312(***)	0,220(***)	0,230(**)
Item 5	Correlación Pearson	0,291(***)	0,311(***)	1,000	0,231(***)	0,208(**)
	Tau_b de Kendall	0,254(***)	0,302(***)	1,000	0,231(***)	0,185(**)
	Rho de Spearman	0,287(***)	0,312(***)	1,000	0,231(***)	0,191(**)
Item 8	Correlación Pearson	0,241(**)	0,221(***)	0,231(***)	1,000	0,291(**)
	Tau_b de Kendall	0,240(**)	0,214(***)	0,231(***)	1,000	0,261(**)
	Rho de Spearman	0,251(**)	0,220(***)	0,231(***)	1,000	0,286(**)
Ítem 12	Correlación Pearson	0,217(**)	0,239(**)	0,208(**)	0,291(**)	1,000
	Tau_b de Kendall	0,177(**)	0,214(**)	0,185(**)	0,261(**)	1,000
	Rho de Spearman	0,200(**)	0,230(**)	0,191(**)	0,286(**)	1,000

** La correlación es significativa al nivel 0,01 (bilateral).
*** La correlación es significativa al nivel 0,001 (bilateral).

En la tabla 6.9 se presentan las correlaciones bivariadas de Pearson, Kendall y Spearman entre los ítems que contribuyeron específicamente a configurar el tercer factor. Se observan unas correlaciones moderadas, pero estadísticamente significativas entre todos los pares de ítems, lo que apoya su interpretación como factor.

El **ítem 1** que evalúa la definición de la probabilidad condicional, se puntuó siguiendo el siguiente criterio:

0. Respuesta totalmente incorrecta o no da respuesta. La incorrección se produce por confusión con probabilidad conjunta o simple.
1. Define correcta, pero imprecisamente una de las probabilidades pedidas, por ejemplo: *"En la probabilidad condicional, para que se dé un suceso, se tiene que dar otro".* Este alumno se da cuenta que intervienen dos sucesos, pero la respuesta no es totalmente correcta porque matemáticamente podemos definir la probabilidad

condicional, independientemente de que el suceso ocurra o no[14].

2. Define correcta, y precisamente una de las probabilidades pedidas.
3. Define correcta, pero imprecisamente las dos probabilidades pedidas, como en el caso siguiente: *"Probabilidad simple: aquella en la que hay un sólo elemento y en la probabilidad condicional intervienen dos sucesos"*. La respuesta es imprecisa, porque en la probabilidad conjunta también intervienen dos sucesos. Otro ejemplo es el siguiente: *"La probabilidad simple es la probabilidad de que ocurra una variable y la condicional que ocurra sabiendo que ha ocurrido otra que la condiciona"*. Es impreciso porque se refiere a variables y no a sucesos.
4. Define correcta y de manera precisa las dos probabilidades pedidas. Por ejemplo: *"Probabilidad simple P(A), calcular la probabilidad de un sólo suceso. Probabilidad condicional* $P(A/B) = \frac{P(A \cap B)}{P(B)}$ *calcular A si se da B.*

Los resultados (Tabla 6.10) sugieren que una parte importante de los alumnos define correctamente las dos probabilidades o al menos una (respuestas puntadas con tres puntos). Aún así la tercera parte de los alumnos no da respuesta o tiene imprecisiones en una o las dos definiciones.

Tabla 6.10. Frecuencias (y porcentajes) de respuestas en el ítem 1 (*n*=414)

	Frecuencia	Porcentaje	Porcentaje acumulado
0	119	28,7	28,7
1	28	6,8	35,5
2	90	21,7	57,2
3	50	12,1	69,3
4	127	30,7	100,0

En el **ítem 2** que evalúa la comprensión de que la probabilidad condicional supone una restricción del espacio muestral[15] se ha puntuado según el siguiente criterio:

0. No responde o responde incorrectamente. Un ejemplo de respuesta incorrecta a la primera pregunta sería la siguiente: *(MF, MF, MF)*.

[14] Un ejemplo es el contraste de hipótesis, donde se definen los dos tipos de error (probabilidad de que se rechace la hipótesis siendo cierta y probabilidad de que se acepte la hipótesis siendo falsa). Estas dos probabilidades se condicionan por sucesos incompatibles (la hipótesis es cierta o falsa) y, sin embargo, las dos probabilidades están definidas.

[15] El espacio muestral es una de las ideas estocásticas fundamentales (Heitele, 1970) que permitió el desarrollo del cálculo de probabilidades. La comprensión de esta idea es un paso fundamental para adquirir el concepto según Tarr y Jones (1997).

1. Falta más de un suceso del espacio muestral o no tiene en cuenta el orden, como en el siguiente ejemplo de respuesta a la primera pregunta: *{(MMM), (MMF), (MFF), (FFF)}*, pero restringe correctamente el espacio muestral *{(MMF), (MMM)}*.
2. Responde correctamente. Un ejemplo de respuesta correcta es la siguiente: *Género de los hijos de las familias con tres descendientes {(MMM), (MMF), (MFM), (FMM), (FFM), (FMF), (FFM), (FFF)} y género de los hijos de las familias con tres descendientes en los que dos son varones {(MMM), (MMF), (MFM), (FMM)}.*

Dado que nuestro interés es comprobar que los alumnos restringen el espacio muestral, se consideraron correctas las respuestas puntuadas con 2 puntos y 1 punto.

Tabla 6.11. Frecuencias (y porcentajes) de respuestas en el ítem 2 (*n*=414)

	Frecuencia	Porcentaje	Porcentaje acumulado
0	62	15,0	15,0
1	82	19,8	34,8
2	270	65,2	100,0

La mayor parte de alumnos restringe correctamente el espacio muestral. Los errores más frecuentes fueron los siguientes: *"{masculino, femenino}"* (dar el espacio muestral de un experimento simple) y no tener en cuenta el orden de nacimiento[16]. También se han dado errores al faltar sucesos del espacio muestral.

Situaciones con y sin reemplazamiento

Tabla 6.12. Frecuencias (y porcentajes) de respuestas en el ítem 5

Una caja tiene cuatro focos, de los cuales dos son defectuosos. Se sacan dos al azar, uno tras otro sin reemplazamiento. Si el primer foco fue defectuoso, entonces:			
		Frecuencia	Porcentaje
a)	Es mas probable que el segundo sea defectuoso	4	1,0
b)	**Es más probable que el segundo no sea defectuoso**	**367**	**88,6**
c)	La probabilidad de que el segundo sea defectuoso es igual a la probabilidad de que no lo sea	42	10,1
d)	Blanco	1	0,2

En la tabla 6.15 presentamos los resultados del **ítem 5.** La respuesta correcta a este ítem es la alternativa b). Este ítem ha resultado sencillo, por lo que los alumnos

[16] Para muchos estudiantes es difícil decidir si se requiere tener o no en cuenta el orden en el cálculo de probabilidades de sucesos compuestos. Ello se debe a dificultad de razonamiento combinatorio (Maury, 1986).

discriminan las situaciones con reemplazamiento presentando un alto porcentaje de respuestas correctas. La alternativa a) fue poco escogida.

El **ítem 8** que evalúa el cálculo de una probabilidad condicional en una situación sincrónica y sin reemplazamiento, se ha puntuado siguiendo el siguiente criterio:

0. No se responde, o se responde incorrectamente.
1. Identifica los casos favorables o posibles, pero no resuelve el problema. Un ejemplo es el siguiente en que el alumno ha identificado los casos favorables, pero no identifica el problema como de probabilidad condicional, sino que lo confunde con otro de probabilidad simple. Por ello, encuentra la probabilidad de cada caso, aplicando la regla del producto (correctamente) y suma las probabilidades (aplicando el axioma de la unión). El alumno muestra un razonamiento probabilístico bueno, pero no en probabilidad condicional.

$$3 \cdot 4 = 12 \qquad P(3_1 \cap 4_2) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}.$$
$$4 \cdot 3 = 12 \qquad P(4_1 \cap 3_2) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}. \qquad \frac{2}{36}$$

Otro ejemplo, en el caso siguiente, donde el alumno identifica los casos posibles, pero falla al aplicar la regla de Laplace.

"Sucesos posibles
Dado 1 Dado 2
3 x 4 =12
4 x 3=12
2 x 6=12
6 x 2=12
P= 6/8"

2. Resuelve el problema correctamente bien por la regla de Laplace, bien por la fórmula del producto. Por ejemplo, en el caso siguiente el alumno resuelve el problema por la regla de Laplace: *"{(2,6), (3,4), (6,2), (4,3)}; 2/4=1/2= 0,5"*.

En la tabla 6.13 se presentan los resultados del ítem 8. El 44,9% de los alumnos han fallado en este ítem, que según Tarr y Jones (1997) indicaría que no han alcanzado el nivel superior de razonamiento en probabilidad condicional en su esquema, al no diferenciar situaciones con y sin reemplazamiento.

Tabla 6.13. Frecuencias (y porcentajes) de respuestas en el ítem 8 (n=414)

	Frecuencia	Porcentaje	Porcentaje acumulado
0	186	44,9	44,9
1	86	20,8	65,7
2	142	34,3	100,0

2.6. SESGOS EN EL RAZONAMIENTO SOBRE PROBABILIDAD CONDICIONAL

Los tres factores anteriores correspondían a capacidades de tipo lógico-matemático y la mitad o más de los estudiantes (llegando en algunos ítems al 90%) resuelven correctamente los ítems de los mismos. Resaltamos que los ítems incluyen resolución de problemas complejos, que superan el nivel superior de razonamiento en probabilidad condicional en la categoría de Tarr y Jones (1997). Esto apoya nuestra idea de que los estudiantes tienen suficientes conocimientos de probabilidad condicional para abordar el estudio de la inferencia bayesiana. En lo que sigue se analiza el resto de los factores, que se interpretaron como sesgos en el razonamiento sobre probabilidad condicional, no relacionados con el conocimiento lógico-matemático.

2.6.1. FALACIA DEL EJE DE TIEMPO Y CAUSACIÓN

Tabla. 6.14. Correlaciones entre ítems que definen el Factor 4

		Ítem 14. eje temporal 1	Ítem 17b. eje temporal 2	Ítem 18. sin reempl 3
Ítem 14	Correlación de Pearson	1,000	0,257(**)	0,091(*)
	Tau_b de Kendall	1,000	0,257(**)	0,091(*)
	Rho de Spearman	1,000	0,257(**)	0,091(*)
Ítem 17b	Correlación de Pearson	0,257(**)	1,000	0,027
	Tau_b de Kendall	0,257(**)	1,000	0,027
	Rho de Spearman	0,257(**)	1,000	0,027
Ítem 18	Correlación de Pearson	0,091(*)	0,027	1,000
	Tau_b de Kendall	0,091(*)	0,027	1,000
	Rho de Spearman	0,091(*)	0,027	1,000

** La correlación es significativa al nivel 0,01 (bilateral).
* La correlación es significante al nivel 0,05 (bilateral).

El primero de ellos se relaciona con la falacia del eje de tiempo. Presentamos las correlaciones bivariadas de Pearson, Kendall y Spearman entre los ítems que contribuyeron específicamente a configurar el cuarto factor. El ítem 18, que contribuía con bajo peso al factor no correlaciona con el resto de los ítems. Sin embargo el ítem 14 y 17b, que evalúan la falacia del eje temporal si presentan una correlación significativa. Se confirma que esta falacia no guarda relación con el resto de razonamientos.

En la tabla 6.15 presentamos los resultados del **ítem 17b**. Aproximadamente la cuarta parte da la respuesta correcta (alternativa a), mientras otra cuarta parte presentan la falacia del eje temporal[17] y el sesgo de equiprobabilidad lo presenta un 36,5%[18].

Tabla 6.15. Frecuencias (y porcentajes) de respuestas en el ítem 17b (*n*=414)

¿Cuál es la probabilidad de haber extraído una bola negra en primer lugar, sabiendo que hemos extraído una bola negra en segundo lugar? P (N_1/ N_2)		
	Frecuencia	Porcentaje
a) 1/ 3	**99**	**23,9**
b) No se puede calcular	103	24,9
c) 1/ 6	38	9,2
d) 1/ 2	151	36,5
Blanco	23	5,6

En la tabla 6.16 presentamos los resultados del **ítem 14**. La respuesta correcta es la alternativa c). Este ítem ha resultado especialmente difícil. Tan sólo un 9% de alumnos lo han resuelto. La mayoría de los alumnos han escogido la alternativa a) que indica que no tienen en cuenta todos los sucesos del espacio muestral[19]. Han sido pocos los alumnos que han escogido la alternativa d), en contraste con el ítem anterior, por lo que en general no son conscientes de la falacia del eje temporal[20].

Tabla 6.16. Frecuencias (y porcentajes) de respuestas en el ítem 14 (*n*=414)

Una bola se suelta por la entrada E. Si sale por R, ¿Cuál es la probabilidad de que haya pasado por el canal 1?		
	Frecuencia	Porcentaje
a) 1/2	318	76,8
b) 1/3	37	8,9
c) 2/3	**41**	**9,9**
d) No se puede calcular	7	1,7
Blanco	11	2,7

[17] Concepción cronologista de la probabilidad condicional, según Gras y Totohasina (1995), consistente en entender que sólo podemos condicionar por un suceso que es anterior en el tiempo al condicionado.

[18] Los resultados reproducen los obtenidos por Falk (1986) ya que sus estudiantes también eligen mayoritariamente estas dos opciones.

[19] La restricción incorrecta del espacio muestral al calcular las probabilidades implica una comprensión deficiente de la idea de condicionamiento (Totohasina, 1992).

[20] Contrastan estos resultados con lo de Ojeda (1995) quien obtiene mucho mejores resultados con este ítem que con los anteriores. La autora presentó el ítem en términos de frecuencias, en lugar de probabilidades, lo que podría explicar la discrepancia de resultados.

En la tabla 6.17 presentamos los resultados del **ítem 18.** La alternativa correcta fue elegida por la mayoría[21]. El error más frecuente fue confundir el muestreo con y sin reposición (alternativa b), confirmándose de nuevo que algunos estudiantes no alcanzan el nivel superior de razonamiento en la clasificación de Tarr y Jones (1997).

Tabla 6.17. Frecuencias (y porcentajes) de respuestas en el ítem 18

Se tiene una bolsa con 1 bola azul y 2 bolas rojas. Se saca una bola al azar. Sin reponerla otra vez en la bolsa se saca una segunda bola. ¿Qué suceso es más probable?		
	Frecuencia	Porcentaje
a) Sacar dos bolas rojas.	39	9,4
b) Sacar primero una bola roja y luego una azul.	56	13,5
c) Los dos sucesos son iguales de probables	**313**	**75,6**
Blanco	6	1,4

2.6.2. FALACIA DE LA CONJUNCIÓN

El quinto factor estuvo definido sólo por tres ítems, uno de ellos relacionado con la falacia de la conjunción, que no puntuó en ninguno de los factores anteriores.

Tabla 6.18. Correlaciones entre ítems que definen el Factor 5

		Ítem 5. Sin reemplaz	Ítem 7. Calculo P. condicional	Ítem 9. F. conjunción
Ítem 5	Correlación de Pearson	1,000	0,030	0,131(**)
	Tau_b de Kendall	1,000	0,030	0,131(**)
	Rho de Spearman	1,000	0,030	0,131(**)
Ítem 7	Correlación de Pearson	0,030	1,000	0,100(*)
	Tau_b de Kendall	0,030	1,000	0,100(*)
	Rho de Spearman	0,030	1,000	0,100(*)
Ítem 9	Correlación de Pearson	0,131(**)	0,100(*)	1,000
	Tau_b de Kendall	0,131(**)	0,100(*)	1,000
	Rho de Spearman	0,131(**)	0,100(*)	1,000

** La correlación es significativa al nivel 0,01 (bilateral).
* La correlación es significante al nivel 0,05 (bilateral).

Presentamos las correlaciones bivariadas de Pearson, Kendall y Spearman entre los ítems que contribuyeron específicamente a configurar el quinto factor. Se muestran correlaciones bajas, lo que apoya la idea de falta de relación de la falacia de la conjunción con el resto de los ítems. Comentaremos sólo los ítems 7 y 9, pues el 5 se analizó anteriormente.

En la tabla 6.19 presentamos los resultados del **ítem 9.** La respuesta correcta es la

[21] Resultados mucho mejores que los de Maury (1986) de quien se tomó el ítem, pues la autora solo obtiene 25% de respuestas correctas, aunque sus estudiantes son menores (Bachillerato y Preuniversitario).

alternativa a). El 24,9% de los alumnos ha acertado el ítem. Los alumnos han elegido mayoritariamente la alternativa c), presentando el sesgo de equiprobabilidad. Por tanto, la falacia de la conjunción no aparece en nuestro trabajo con tanta frecuencia como en las investigaciones de Tversky y Kahneman (1982b).

Tabla 6.19. Frecuencias (y porcentajes) de respuestas en el ítem 9 (*n*=414)

Supón que una estrella del tenis alcanza la final de Roland Garros en 2005. Para ganar el partido hay que ganar tres sets de cinco. ¿Cuál de los siguientes sucesos consideras más probable?		
	Frecuencia	Porcentaje
a) El jugador pierde el primer set.	**103**	**24,9**
b) El jugador pierde el primer set pero gana el partido.	39	9,4
c) Los dos sucesos son iguales de probables.	256	61,8
Blanco	16	3,9

En la tabla 6.20 se presentan los resultados del **ítem 7**, donde los alumnos podrían haber utilizado inconscientemente la falacia de la conjunción, pues la opción correcta tiene una probabilidad muy baja. La alternativa correcta, la a) es escogida por un 33,8% de alumnos. Hay un porcentaje similar de estudiantes que confunden la probabilidad condicional con la simple o conjunta.

Tabla 6.20. Frecuencias (y porcentajes) de respuestas en el ítem 7 (*n*=414)

La probabilidad de que una mujer tenga una mamografía positiva es el 10,3%. La probabilidad de que una mujer tenga cáncer de pecho y una mamografía positiva es de 0,8%. Una mujer se realiza una mamografía y resulta positiva. ¿Cuál es la probabilidad de que realmente tenga cáncer?		
	Frecuencia	Porcentaje
a) $\frac{0,8}{10,3} = 0,0776$ **, probabilidad de 7,76%**	**140**	**33,8**
b) $10,3 \times 0,8 = 8,24$, probabilidad del 8,24%	127	30,7
c) 0’8 %	135	32,6
Blanco	12	2,9

2.6.3. FALACIA DE LA CONDICIONAL TRANSPUESTA

El ítem 10 aparece prácticamente aislado en el sexto factor, relacionado solamente con el 17a. Se presentan en la tabla 6.21 las correlaciones bivariadas de Pearson, Kendall y Spearman entre estos dos ítems, que además de ser negativa es prácticamente cero, aunque significativa. Comentamos a continuación estos ítems.

Tabla 6.21. Correlaciones entre ítems que definen el Factor 6

		Ítem 10. condicional trans	Ítem 17a. Sin reemplaz
Ítem 10	Correlación de Pearson	1,000	-0,090(*)
	Tau_b de Kendall	1,000	-0,090(*)
	Rho de Spearman	1,000	-0,090(*)
Ítem 17a	Correlación de Pearson	-0,090(*)	1,000
	Tau_b de Kendall	-0,090(*)	1,000
	Rho de Spearman	-0,090(*)	1,000

* La correlación es significante al nivel 0,05 (bilateral).

En la tabla 6.22 presentamos los resultados del **ítem 10.** La respuesta correcta es la alternativa b). Hay una mayor elección de la alternativa c), que evalúa la confusión entre una probabilidad condicional y su transpuesta[22] y alrededor de un tercio del grupo que escoge la respuesta correcta[23]. En consecuencia la confusión entre una probabilidad condicional y su inversa, al resolver problemas que involucran estas probabilidades está ampliamente extendida en los estudiantes[24].

Tabla 6.22. Frecuencias (y porcentajes) de respuestas en el ítem 10 (*n*=414)

Un test diagnóstico de cáncer fue administrado a todos los residentes de una gran ciudad en la que hay pocos casos de cáncer. Un resultado positivo en el test es indicativo de cáncer y un resultado negativo es indicativo de ausencia de cáncer. ¿Qué te parece más probable?	Frecuencia	Porcentaje
a) Que una persona tenga cáncer si ha dado positivo en el test de diagnóstico.	24	5,8
b) Que un test de diagnóstico resulte positivo si la persona tiene cáncer.	**133**	**32,1**
c) Los dos sucesos tienen la misma probabilidad.	245	59,2
Blanco	12	2,9

En la tabla 6.23 presentamos los resultados del **ítem 17a**. La respuesta correcta es la alternativa c). La mayoría de los alumnos (68,8%) han respondido correctamente y en este sentido el ítem se opone al anterior, lo que puede explicar la pequeña correlación negativa entre ambos. El error más frecuente en ambos grupos es la elección de la alternativa b), donde los alumnos han confundido la probabilidad condicional y conjunta aplicando la regla del producto (1/2) x (1/3)=1/6.

[22] Falk (1986) denomina a este error *falacia de la condicional transpuesta.*

[23] Los resultados son algo mejores que Pollatsek y cols. (1987); el 69% de los sujetos de su estudio tienen este sesgo.

[24] Esta confusión también apareció, aunque con menor frecuencia en la lectura de tablas dobles (ítem 4c y 6d).

Tabla 6.23. Frecuencias (y porcentajes) de respuestas en el ítem 17a (n=414)

Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin devolver la primera a la urna. ¿Cuál es la probabilidad de extraer una bola negra en segundo lugar, habiendo extraído una bola negra en primer lugar? $P\ (N_2/\ N_1)$		
	Frecuencia	Porcentaje
a) 1/ 2	23	5,6
b) 1/ 6	70	16,9
c) 1/ 3	**285**	**68,8**
d) 1/ 4	30	7,2
Blanco	6	1,4

2.6.4. INDEPENDENCIA Y FALACIA DE LAS TASAS BASE

Presentamos las correlaciones bivariadas de Pearson, Kendall y Spearman entre los ítems que contribuyeron específicamente a configurar el ultimo factor retenido que agrupa el ítem 4 (independencia), 3 (falacia de tasas base), con una pequeña contribución del 8. Hemos interpretado este factor como relación entre independencia y falacia de las tasas base. Se presenta en lo que sigue los dos primeros ítems, puesto que el 8 se analizó anteriormente.

Tabla. 6.24. Correlaciones entre ítems que definen el Factor 7

		Ítem 4. Independencia	Ítem 3. Bayes (tasas base)	Ítem 18. Sin reemplaz
Ítem 4	Correlación de Pearson	1,000	0,129(**)	-0,028
	Tau_b de Kendall	1,000	0,129(**)	-0,028
	Rho de Spearman	1,000	0,129(**)	-0,028
Ítem 3	Correlación de Pearson	0,129(**)	1,000	0,032
	Tau_b de Kendall	0,129(**)	1,000	0,032
	Rho de Spearman	0,129(**)	1,000	0,032
Ítem 18	Correlación de Pearson	-0,028	0,032	1,000
	Tau_b de Kendall	-0,028	0,032	1,000
	Rho de Spearman	-0,028	0,032	1,000

** La correlación es significativa al nivel 0,01 (bilateral).

En la tabla 6.25 presentamos los resultados del **ítem 4** que evalúa el concepto de independencia. La respuesta correcta es la alternativa c). Un 29% de alumnos responden correctamente al ítem[25]. El error más frecuente entre los alumnos es la confusión con el concepto de mutua exclusividad[26]. El distractor b) evalúa el error de suponer que el concepto de independencia sólo se puede aplicar en experimentos

[25] Sánchez (1996) obtiene un 40% de respuestas correctas en licenciados de matemáticas que se preparaban para ser profesores de matemáticas.

[26] Kelly y Zwiers (1986) sugieren que el error puede ser debido a la imprecisión del lenguaje ordinario, en donde independiente también se puede interpretar como “separado”.

diacrónicos. La conclusión es que la mayor parte de los estudiantes presenta uno de los dos sesgos anteriores.

Tabla 6.25. Frecuencias (y porcentajes) de respuestas en el ítem 4 (*n*=414)

Se extrae una carta al azar de una baraja española (40 cartas con cuatro palos: oros, bastos, espadas y copas. Cada palo tiene los números del 1 al 7, sota caballo y rey). Sea A el suceso "se extrae una carta de oros" y B el suceso "se extrae un rey". ¿Los sucesos A y B son independientes?	Frecuencia	Porcentaje
a) No son independientes, porque en la baraja hay un rey de oros.	118	28,5
b) Sólo si sacamos primero una carta para ver si es rey y se vuelve a colocar en la baraja y luego sacamos una segunda para ver si es oros.	61	14,7
c) Si, porque P(rey de oros)= P(rey) x P(oros).	**120**	**29,0**
d) No, porque P(rey / oros) $\neq$ P(rey).	82	19,8
Blanco	33	8,0

Tabla 6.26. Frecuencias (y porcentajes) de respuestas en el ítem 3 (*n*=414)

Un taxi se vio implicado en un accidente nocturno. Hay dos compañías de taxis en la ciudad, la Verde y la Azul. El 85% de los taxis de la ciudad son Verdes y el 15% Azules. Hubo un testigo del accidente. El tribunal comprobó la fiabilidad del testigo bajo las mismas circunstancias que había la noche del accidente y llegó a la conclusión de que el testigo identificaba correctamente cada uno de los colores en el 80% de las ocasiones y fallaba en el 20%. Sabiendo que el testigo identificó el taxi como azul, ¿Cuál es la probabilidad de que el taxi implicado en el accidente fuera en efecto Azul?	Frecuencia	Porcentaje
a) 80/100	33	8,0
b) 15 /100	29	7,0
c) (15/100) X (80/100)	121	29,2
d) $\frac{15\times 80}{85\times 20+15\times 80}$	**206**	**49,8**
Blanco	25	6,0

En la tabla 6.26 presentamos los resultados del **ítem 3**. La respuesta correcta es la alternativa d). En este ítem, un 49,8% de alumnos que responden correctamente[27]. El distractor más escogido, después de la opción correcta es el c) (confusión de la probabilidad condicional y conjunta). Un 15% de alumnos (respuestas a) y b)) presentan la falacia de las tasas base.

En consecuencia se aprecia un contraste entre la dificultad de estos ítems relacionados con el sesgo en el razonamiento, en relación con los de primer tipo que recogen la capacidad lógico matemática. Todo ello se tendrá en cuenta en la elaboración del material de enseñanza.

[27] Mejores resultados que los obtenidos en las investigaciones de Tversky y Kahneman (1982), posiblemente debido al hecho de haber incluido las fórmulas del teorema de Bayes en el enunciado de los discractores.

2.7. INDICES DE DIFICULTAD

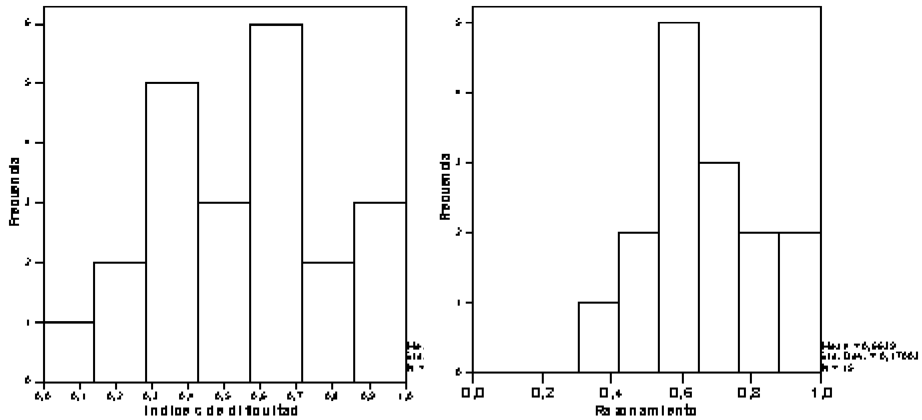
El análisis detallado de ítems se completa a continuación con la estimación de los índices de dificultad de los ítems del cuestionario RPC, desde la perspectiva clásica y bayesiana. En la tabla 6.27 se presentan los índices de dificultad e intervalos de confianza al 95%, se observa gran variación en estos índices (del 9% al 91% de respuestas correctas), aunque el valor medio del índice de dificultad (figura 6.1) supera el 50%, lo que indica que más de la mitad de los estudiantes, en promedio resuelven los ítems.

Los resultados son aún mejores cuando se restringe el análisis a los ítems que configuraron los tres primeros factores (razonamiento lógico matemático); como se observa en el segundo gráfico de la figura 6.1 el rango se reduce (siendo el mínimo superior a 0,3) y la media aumenta al 66% de estudiantes que resuelve correctamente cada ítem.

Tabla 6.27. Índices de dificultad e intervalos de confianza de la versión final del cuestionario en estudiantes de psicología (*n* = 414)

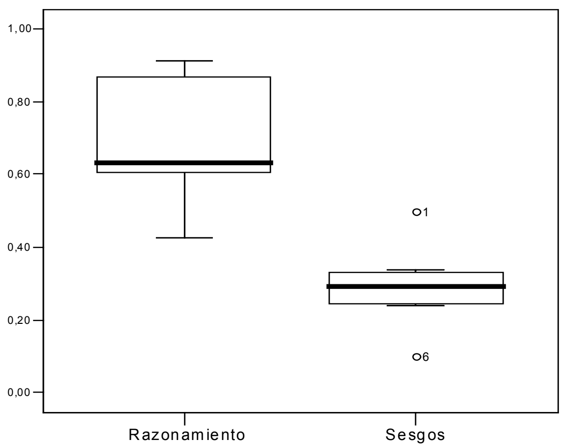
	Índice	I. confianza 95%	
I1	0,428	0,36	0,50
I2	0,850	0,80	0,90
I3	0,498	0,43	0,57
I4	0,290	0,23	0,35
I5	0,886	0,84	0,93
I6A	0,913	0,87	0,95
I6B	0,630	0,56	0,70
I6C	0,626	0,56	0,69
I6D	0,582	0,51	0,65
I7	0,338	0,27	0,40
I8	0,343	0,28	0,41
I9	0,249	0,19	0,31
I10	0,321	0,26	0,38
I11	0,879	0,83	0,92
I12	0,693	0,63	0,76
I13	0,527	0,46	0,60
I14	0,099	0,06	0,14
I15	0,604	0,54	0,67
I16	0,553	0,49	0,62
I7A	0,688	0,62	0,75
I7B	0,239	0,18	0,30
I18	0,756	0,70	0,81

Figura 6.1. Histograma para los índices de dificultad del total de ítems (a) y de los ítems de razonamiento matemático (b)



La diferencia de dificultad entre estos ítems y los que evalúan sesgos descritos en la literatura psicológica, se notable en la figura 6.2, donde la mediana supera en más del doble en los primeros ítems a los segundos[28].

Figura 6.2. Índices de dificultad en los ítems de razonamiento matemático y los que evalúan sesgos



En las tablas 6.28 y 6.29 se presenta las estimaciones bayesianas de los índices de dificultad, para el caso no informativo e informativo. Se hace notar que el fin principal de la inferencia bayesiana es obtener las distribuciones finales de los índices, que serían distribuciones Beta[29] *Be(e,f)*, siendo *e* y *f* los éxitos y fracasos dados en la distribución final. Esta distribución final permitiría calcular probabilidades de que los índices de dificultad se encontrasen en intervalos especificados[30]. No obstante, para comparar con

[28] También los cuartiles del primer grupo de ítems son claramente superiores a los del segundo grupo.

[29] La distribución muestral de la proporción en la estimación frecuencial sería Normal.

[30] En la estimación frecuencial no podemos calcular estas probabilidades.

la estimación frecuencial, se han añadido las estimaciones puntuales y los intervalos de credibilidad del 95% para los índices.

Tabla 6.28. Estimación bayesiana de los índices de dificultad con distribución inicial no informativa

	Prueba evaluación (Datos, n=414)		Distribución inicial		distribución Final		Estimación Índice	Intervalo Credibilidad 95%	
Ítem	Éxitos	Fracasos	Éxitos	Fracasos	Éxitos	Fracasos		lim inf	lim sup
1	177	237	0,5	0,5	177,5	237,5	0,428	0,381	0,476
2	352	62	0,5	0,5	352,5	62,5	0,849	0,813	0,882
3	206	208	0,5	0,5	206,5	208,5	0,498	0,450	0,546
4	120	294	0,5	0,5	120,5	294,5	0,290	0,248	0,335
5	367	47	0,5	0,5	367,5	47,5	0,886	0,853	0,914
6A	378	36	0,5	0,5	378,5	36,5	0,912	0,883	0,937
6B	261	153	0,5	0,5	261,5	153,5	0,630	0,583	0,676
6C	259	155	0,5	0,5	259,5	155,5	0,625	0,578	0,671
6D	241	173	0,5	0,5	241,5	173,5	0,582	0,534	0,629
7	140	274	0,5	0,5	140,5	274,5	0,339	0,294	0,385
8	142	272	0,5	0,5	142,5	272,5	0,343	0,298	0,390
9	103	311	0,5	0,5	103,5	311,5	0,249	0,209	0,292
10	133	281	0,5	0,5	133,5	281,5	0,322	0,278	0,367
11	364	50	0,5	0,5	364,5	50,5	0,878	0,845	0,908
12	287	127	0,5	0,5	287,5	127,5	0,693	0,648	0,736
13	218	196	0,5	0,5	218,5	196,5	0,527	0,478	0,574
14	41	373	0,5	0,5	41,5	373,5	0,100	0,073	0,131
15	250	164	0,5	0,5	250,5	164,5	0,604	0,556	0,650
16	229	185	0,5	0,5	229,5	185,5	0,553	0,55	0,601
17A	285	129	0,5	0,5	285,5	129,5	0,688	0,643	0,732
17B	99	315	0,5	0,5	99,5	315,5	0,240	0,200	0,282
18	313	101	0,5	0,5	313,5	101,5	0,755	0,713	0,796

En el caso de distribución inicial informativa se tomó como distribución inicial la definida por el número de aciertos y fracasos en el estudio piloto del cuestionario. Se obtienen ligeras variaciones en el estimador puntual de la proporción e intervalos de credibilidad más precisos. No obstante, puesto que el tamaño de la muestra en los datos es mucho mayor que la del estudio piloto, la mayor influencia en la distribución final viene dada por los datos (y no por la distribución inicial).

Hay que tener también en cuenta que la principal finalidad de la estimación bayesiana es la de obtener la distribución final, a partir de la cuál podría obtenerse otras probabilidades.

Tabla. 6.29. Estimación bayesiana de los índices de dificultad con distribución inicial informativa

	Prueba evaluación (Datos, n=414)		Distribución inicial		Distribución Final		Estimación Índice	Intervalo Credibilidad 95%	
Ítem	Éxitos	Fracasos	Éxitos	Fracasos	Éxitos	Fracasos		lim inf	lim sup
1	177	237	49	8	226	245	0,480	0,435	0,525
2	352	62	41	16	393	78	0,834	0,800	0,867
3	206	208	16	41	222	249	0,471	0,426	0,516
4	120	294	10	47	130	341	0,276	0,237	0,317
5	367	47	47	10	414	57	0,879	0,848	0,907
6A	378	36	50	7	428	43	0,909	0,881	0,933
6B	261	153	17	40	278	193	0,590	0,546	0,634
6C	259	155	24	33	283	188	0,601	0,556	0,645
6D	241	173	24	33	265	206	0,563	0,518	0,607
7	140	274	17	40	157	314	0,333	0,291	0,377
8	142	272	15	42	157	314	0,333	0,291	0,377
9	103	311	19	38	122	349	0,259	0,220	0,299
10	133	281	18	39	151	320	0,321	0,279	0,363
11	364	50	22	35	386	85	0,820	0,784	0,853
12	287	127	34	23	321	150	0,682	0,639	0,723
13	218	196	26	31	244	227	0,518	0,473	0,563
14	41	373	4	53	45	426	0,096	0,071	0,124
15	250	164	11	46	261	210	0,554	0,59	0,599
16	229	185	11	46	240	231	0,510	0,464	0,555
17A	285	129	27	30	312	159	0,662	0,619	0,704
17B	99	315	15	42	114	357	0,242	0,204	0,282
18	313	101	41	16	354	117	0,752	0,712	0,790

2.8. DISCUSIÓN DEL ESTUDIO 7

En el estudio 7 se realizó una evaluación del razonamiento sobre probabilidad condicional en una muestra de 414 estudiantes de psicología, de dos universidades diferentes, después de que hubieran estudiado el tema.

El análisis detallado de las respuestas a los ítems, agrupados según los factores identificados en el análisis factorial (estudio 6) indica dos tendencias principales:

1. Los estudiantes muestran una comprensión razonable del concepto y una buena capacidad de resolver problemas de probabilidad condicional, incluyendo el teorema de Bayes. Esta conclusión se obtiene del análisis de los porcentajes de respuestas correctas en los ítems que configuran los tres primeros factores.
2. Al mismo tiempo se observa la existencia generalizada de sesgos en el razonamiento sobre probabilidad condicional: Falacia de la condicional transpuesta, falacia del eje de tiempo, concepciones incorrectas de la independencia y en menor medida, falacia de tasas base.

El primer punto apoya la viabilidad de intentar una enseñanza de conceptos bayesianos elementales a los estudiantes de psicología, por tanto da validez al experimento de enseñanza que se describe y analiza en los estudios 8 y 9. El segundo punto sirvió para identificar sesgos que cabía esperar en los estudiantes participantes en dicho experimento y que se tuvieron en cuenta en el diseño del material didáctico. Resultados parciales del estudio de evaluación se han publicado en Díaz, de la Fuente y Batanero (2004b); Díaz, y de la Fuente (2005b; 2006; en prensa).

3. ESTUDIO 8. EVALUACIÓN DE UNA PROPUESTA DE ENSEÑANZA DE CONCEPTOS ELEMENTALES DE INFERENCIA BAYESIANA EN PSICOLOGÍA

3.1. INTRODUCCIÓN

Realizada la evaluación del razonamiento condicional de los estudiantes de psicología mediante el cuestionario RPC (estudio 7), se pasó a elaborar y experimentar una propuesta de enseñanza de conceptos elementales de inferencia bayesiana, que tuviese en cuenta las principales dificultades mostradas en el estudio de evaluación citado, el análisis de la inferencia bayesiana elemental, llevado a cabo en el capítulo 1 así como una serie de principios didácticos que se detallan en la sección 3.2.1.

El objetivo del estudio 8 fue llevar a cabo un experimento de enseñanza de inferencia bayesiana elemental a estudiantes de primer año de psicología y obtener información empírica sobre la comprensión lograda por los estudiantes participantes sobre los conceptos enseñados. Más específicamente, se pretendió evaluar, hasta que punto es posible para los estudiantes de psicología comprender la lógica de la inferencia bayesiana y cuáles son las dificultades usuales que surgen en el aprendizaje de la misma.

Para lograr este objetivo fue necesario preparar un material de enseñanza adecuado a estos alumnos y al tiempo disponible, puesto que los textos en castellano de inferencia bayesiana (por ejemplo, Serrano Angulo, 2003) no están específicamente centrados en el campo de la psicología y tampoco se encontraron textos de inferencia bayesiana en castellano con un enfoque práctico y basado en la tecnología. Asimismo hubo que preparar una serie de instrumentos de evaluación de los conocimientos adquiridos en cada sesión y globalmente, que correspondiese a los principales objetivos y contenidos de aprendizaje.

A continuación se describe este estudio, que se ha dividido en dos partes: diseño de la propuesta didáctica (estudio 8.1) y experimentación y evaluación de la misma (estudio 8.2).

3.2. ESTUDIO 8.1. DISEÑO DE LA PROPUESTA DIDÁCTICA

El diseño de una experiencia de enseñanza de un contenido específico de estadística puede encuadrarse en el campo de la Didáctica de los Métodos de Investigación y más específicamente, en la Didáctica de la Estadística, campo que actualmente es reconocido científicamente por el Instituto Internacional de Estadística[31]. Más aún, dentro de los congresos internacionales sobre esta materia (International Conferences on Teaching Statistics, ICOTS, iniciados en 1982), se contempla una sesión específica de trabajo sobre Enseñanza de la Estadística en Psicología y Ciencias Sociales[32].

Dentro de este campo, el diseño de una propuesta concreta de enseñanza para unos estudiantes particulares en contexto dado entra dentro de los *estudios curriculares,* que debe tener en cuenta cuatro grandes cuestiones: *¿Qué enseñar?, ¿cuándo enseñar? ¿cómo enseñar?*, y *¿qué, cómo y cuándo evaluar?* (Bolívar, 1998), que se trata de responder en lo que sigue.

Para el diseño de la enseñanza, se siguió el método utilizado en otros estudios sobre enseñanza de la inferencia estadística a nivel universitario (Delmas, Garfield y Chance, 1998; Tauber, 2001; Alvarado, 2004). Se comenzó por un análisis conceptual del tema a enseñar en una muestra amplia de libros de texto universitarios con enfoque bayesiano (Bernardo, 1981; 2003; Box y Tiao, 1992; Robert, 2001; Gelman, Carlin, Stern y Rubin, 2003; Serrano, 2003; Bolstad, 2004; O'Hagan y Forster, 2004). Seguidamente se analizaron otros orientados especialmente a la enseñanza a estudiantes que no cuentan con una base matemática excesivamente amplia (Albert, 1996; Berry, 1996; Lecoutre, 1996; Albert, y Rossman, 2001), recopilando en ellos ejemplos, ejercicios y recursos

[31] Una de las secciones de este instituto es la International Association for Statistical Education, http://www.stat.auckland.ac.nz/~iase/, que organiza congresos internacionales y cuenta con su propia revista de investigación *Statistics Education Research Journal.*

[32] Se ha presentado un trabajo invitado sobre nuestro estudio de evaluación (Díaz y de la Fuente, 2006) en dicha sesión en ICOTS-7: http://www.maths.otago.ac.nz/icots7/.

que pudieran servir de base para nuestro diseño.

Con todo este bagaje y teniendo en cuenta las restricciones de tiempo, así como los principios metodológicos que exponemos a continuación, se definieron los objetivos, seleccionaron los contenidos que cubrirían estos objetivos, se elaboró un material de apoyo para los alumnos, tanto escrito, como software, y las diferentes pruebas de evaluación de los conocimientos adquiridos.

El material de enseñanza se diseñó en dos versiones. La primera de ellas estaba destinada a la docencia presencial, alternando sesiones en aula ordinaria, con otras en el laboratorio de informática. Adicionalmente se elaboró una versión del material para ser utilizada en modalidad de enseñanza virtual, bien como complemento de la enseñanza presencial mencionada anteriormente o bien como única forma de enseñanza.

3.2.1. PRINCIPIOS METODOLÓGICOS Y DIDÁCTICOS

Los principios tenidos en cuenta en la elaboración del material y el diseño de la enseñanza se basaron en las conclusiones obtenidas del estudio de la problemática didáctica analizada en la Sección 3.5 del Capítulo 1 y tuvieron en cuenta, tanto las sugerencias obtenidas en investigaciones previas sobre la enseñanza de métodos bayesianos (Sección 3.5.1 de dicho capítulo), como tres principios que se resumen a continuación: Aprendizaje constructivo, contextualización y uso de la tecnología como recurso didáctico. Asimismo se tomaron en cuenta las conclusiones obtenidas del estudio 7, donde se llevó a cabo una evaluación del razonamiento condicional en los estudiantes de psicología y se describieron las principales dificultades observadas sobre este razonamiento.

Constructivismo y aprendizaje

Las teorías constructivistas[33] (Ernest, 1994; 1998) suponen que el aprendizaje de las matemáticas se produce cuando el alumno se enfrenta a la resolución de un problema y encuentra por si mismo la solución, asumiendo que el conocimiento no es recibido pasivamente por el sujeto cognitivo sino activamente. Describen la comprensión del

[33] El constructivismo emerge como el principal paradigma de investigación en psicología de la educación matemática, a partir de la influencia de Piaget: La metáfora subyacente concibe al sujeto cognitivo como organismo sujeto a evolución, que se adapta al entorno (Piaget; 1979).

sujeto como la construcción de estructuras mentales, y se habla de "reestructuración", o "cambio conceptual". El proceso es recursivo y por ello los "bloques constructivos" de la compresión son ellos mismos producto de actos previos de construcción (Godino, 2005).

En el diseño de la propuesta didáctica se ha tenido en cuenta estas teorías, y más concretamente la teoría de situaciones didácticas de Brousseau (1997), que considera un triple carácter de las matemáticas: Sistema conceptual organizado, lenguaje y actividad de resolución de problemas. El principio de aproximación comunicacional a la cognición visualiza la comunicación no sólo como ayuda al pensamiento sino que concibe el pensamiento como un caso especial de comunicación. El aprendizaje matemático significa dominar un discurso que sea reconocido como matemático por interlocutores expertos (Sfard, 2001).

Siguiendo estos principios se planificó un sistema organizado de conceptos, propiedades y procedimientos, en forma que se mostrase sus interrelaciones a los alumnos. Se organizaron para la enseñanza *situaciones de acción*, donde el alumno ha de resolver por si mismo problemas; *situaciones de comunicación*, donde el alumno ha de usar el lenguaje matemático y verbalizar sus conocimientos para resolver la prueba de autoevaluación; y *situaciones de validación* o prueba. Las *situaciones de institucionalización* se producen en la corrección por parte del profesor de las actividades y en la presentación de la parte teórica.

También nos basamos en la perspectiva interaccionista[34], que asume que el profesor y los estudiantes constituyen interactivamente la cultura del aula, las convenciones y convenios tanto en lo relativo al contenido de la disciplina, como en las regularidades sociales, que emergen interactivamente, y el proceso de comunicación se apoya en la negociación y los significados compartidos (Bauersfeld, 1995). Por ello hemos tratado de ampliar las posibilidades de comunicación entre la profesora y los estudiantes y de estos entre sí dentro de las sesiones.

[34] El constructivismo social considera al sujeto individual y el dominio social que lo rodea están indisolublemente interconectados.

Contextualización

Se ha tratado de mostrar al alumno la utilidad de los métodos bayesianos en su tarea profesional y vida cotidiana. Como sugiere Moore (1991, p. 14) "*la estadística es la ciencia de los datos. Con más precisión, el objeto de la estadística es el razonamiento a partir de datos empíricos. Los datos no son números, sino números en un contexto"*; el desinterés del alumno se debe con frecuencia a la presentación de los conceptos aislados de las aplicaciones y de la cultura estadística (Carvalho, 2005).

Para el diseño de la propuesta de enseñanza se buscaron situaciones próximas al interés del alumno, tanto en los ejemplos presentados, como en los ejercicios propuestos o resueltos. Entre otra situaciones se presentaron: Pruebas médicas y de diagnóstico; evaluación de un programa de intervención; contextos de práctica profesional (depresión post parto, incidencia de la hipertensión en mujeres, estudio demográfico sobre proporción de mayores); temas de interés general (predicción del sexo de un bebé, número de hijos por mujer, distribución del peso medio) y contextos de la vida cotidiana (consumo y procesos de producción, seguros de accidente, alarmas, estudio de intención de voto y encuesta de satisfacción).

Se tomó también en cuenta el modelo de *razonamiento estadístico*[35] de Wild y Pfannkuch (1999), que contiene cinco componentes fundamentales:

- *Reconocer la necesidad de los datos* para resolver problemas y tomar decisiones.
- *Transnumeración:* o creación de conocimiento a partir del cambio de representación en los datos. Se tuvo en cuenta tres tipos de transnumeración: (1) a partir de la medida que "captura" las cualidades o características del mundo real; (2) al pasar de los datos brutos a representaciones tabulares o gráficas para extraer sentido de los mismos; (3) al comunicar este significado que surge de los datos, en forma que sea comprensible a los estudiantes.
- *Percepción de la variación:* La recogida adecuada de datos y los juicios correctos a partir de los mismos requieren la comprensión de la variación que hay y se transmite en los datos, así como de la incertidumbre originada por la variación no explicada. El material se plantea con el fin de mostrar a los estudiantes la forma de hacer predicciones, buscar explicaciones y causas de la variación y aprender del contexto.

---

[35] Razonamiento específico que reconoce la variabilidad presente en múltiples fenómenos y utiliza métodos de análisis, reducción y control de esta variabilidad, para tomar decisiones o hacer predicciones.

- *Razonamiento con modelos estadísticos:* Cualquier útil estadístico, incluido el teorema de Bayes puede contemplarse como modelo, puesto que es una forma de representar la realidad. Se enfatizó las diferencias entre el modelo y los datos, relacionando, al mismo tiempo, el modelo con los datos.
- *Integración de la estadística y el contexto*: Es también un componente esencial del razonamiento estadístico que se resaltó en el material preparado.

Tecnología como recurso didáctico

Se tuvo en cuenta el papel esencial dado a la tecnología en la enseñanza y aprendizaje de la estadística, reconocido, por ejemplo en la IASE Round Table Conference, celebrada en 1996 (Garfield y Burrill, 1997), la IASE Satellite Conference cebrada en Berlín en 2003 (Engel, 2003) y los congresos ICOTS (Rossman y Chance, 2006). Se planificó utilizar la tecnología como auxiliar de cálculo, preparando una colección de programas en Excel que se pusieron a disposición de los alumnos, permitiendo de este modo facilitar la comprensión de los conceptos, puesto que los medios técnicos de cálculo para poder aplicarlos son dados por el software (Biehler, 1997a y b). Estos recursos se facilitaron, tanto en los laboratorios de informática (clase presencial) como en una página de Internet (que permitía la enseñanza a distancia[36] y el estudio complementario de los alumnos en su casa).

Frente a otros posibles programas, se eligió Excel para desarrollar estas aplicaciones, porque está disponible en la mayoría de los ordenadores personales de los estudiantes y porque proporciona los recursos suficientes para nuestro trabajo. El uso de estos programas en la resolución de las actividades acortaría el tiempo necesario para la resolución de las actividades, permitiendo un mayor número y variedad de actividades en el tiempo disponible. Además sería una fuente de motivación, ya que los alumnos siempre se interesan por aprender los recursos informáticos.

Por otro lado, se preparó un foro de discusión que permite a los estudiantes conectar con el profesor para resolver sus dudas o discutir las dudas con sus compañeros para una futura versión a distancia de la asignatura.

[36] El e-learning facilita a los estudiantes la consulta de los textos y ejemplos, así como plantear sus preguntas en cualquier momento, siempre que los materiales cuenten con suficiente flexibilidad, interactividad e interfaz sencilla para el usuario (Aydýnlý, Härdle, y Rön, 2003).

Razonamiento condicional

En la elaboración de la propuesta de enseñanza se tuvo también en cuenta las principales dificultades constatadas en el razonamiento condicional de los alumnos en el estudio de evaluación descrito en los capítulos anteriores. En concreto, se tuvieron en cuenta las siguientes consideraciones:

- Dificultad en la diferenciación de una probabilidad condicional de otra conjunta y una probabilidad condicional de su transpuesta (Pollatsek y cols., 1987; Eddy, 1992; Ojeda, 1995). Se puso mucho cuidado en el lenguaje utilizado, y se incluyeron recursos como tablas de doble entrada para tratar de paliar este problema. Asimismo se planificaron actividades en que los alumnos debían identificar una probabilidad condicional y su inversa a partir del enunciado verbal de los problemas.
- Dificultad en discriminación de relaciones causales y diagnósticas (Tversky y Kahneman, 1982a). Se eligieron específicamente ejemplos de contextos médicos y pruebas diagnostica, insistiendo en la diferenciación de estos dos tipos de situaciones.
- Falacia temporal (Falk, 1986; Gras y Totohasina, 1995). Se resaltó la relevancia de sucesos que tienen lugar en un instante posterior para la reducción del espacio muestral en un experimento, y por tanto para la definición de probabilidades condicionales. En particular esta situación se produce en la actualización de probabilidades iniciales con los nuevos datos.
- Lectura de tablas (Lonjedo y Huerta, 2004; 2005). Hemos elegido el formato tipo tabla propuesto por Albert y Rossman (2003) para la organización de los datos y resolución de este tipo de problemas para tratar de salvar así la dificultad de los alumnos en la identificación de los datos, la construcción del diagrama en árbol y el cálculo de las probabilidades finales (Gras y Totohasina, 1995).
- Razonamiento y resolución de problemas bayesianos (Totohasina, 1992; Gigerenzer, 1994; Ojeda, 1995). Asimismo este tipo de disposición de datos favorece la generalización necesaria para aplicar el teorema de Bayes en los temas siguientes.

3.2.2. OBJETIVOS DE APRENDIZAJE

La propuesta se planificó para un total de 20 horas, de las cuales 15 serían de docencia y trabajo presencial de los alumnos y las otras 5 de trabajo personal, consistente en la resolución de actividades prácticas que los alumnos debían entregar al principio de la siguiente sesión. Con estas restricciones, y en base al análisis conceptual previo fijamos los siguientes objetivos de aprendizaje:

1. Diferenciar entre probabilidades iniciales y finales y verosimilitudes.
2. Analizar el teorema de Bayes como herramienta para transformar probabilidades iniciales en finales.
3. Organizar los datos en una tabla de forma que se facilite el cálculo, bien manual o con recursos informáticos (hoja Excel, applet) para calcular las probabilidades finales a partir de las probabilidades iniciales y de las verosimilitudes.
4. Visualizar los parámetros como variables aleatorias con una distribución de probabilidad asociada.
5. Identificar valores de la proporción o media poblacional y asignarles probabilidades iniciales plausibles (distribución de probabilidad inicial).
6. Usar las distribuciones Beta y Normal como distribuciones iniciales y finales.
7. Comprender el concepto de distribución inicial uniforme y distribución informativa.
8. Identificar los valores en la muestra.
9. Aprender a introducir los datos anteriores en la hoja Excel para obtener los valores de verosimilitud y la distribución final.
10. Extracción de inferencias: Calcular probabilidades asociadas a determinados valores del parámetro en la distribución final, obtener un mejor estimador, calcular intervalos de credibilidad y realizar un contraste de hipótesis.
11. Entender cómo las distribuciones finales se transforman en iniciales en experimentos sucesivos.
12. Analizar las diferencias entre la metodología clásica y bayesiana.

Los objetivos específicos para cada tema se detallan en el Anexo 7.

3.2.3. CONTENIDOS Y SECUENCIACIÓN

Una vez fijados los objetivos, se seleccionaron los contenidos necesarios para cubrir cada uno de ellos. Se tomaron nuevas decisiones, referentes a la extensión (menos contenidos, pero estudiados con mayor profundidad), variedad (uno o unos pocos casos en cada uno de los capítulos fijados, en lugar de detallar todas las diferentes situaciones que se estudian en inferencia clásica) y orientación (práctica, en lugar de teórica). Todas estas decisiones se basaron en los criterios metodológicos. Puesto que se trataba sólo de un estudio inicial, se podría completar y extender el material y la experiencia en años sucesivos. Se dividió el material en cuatro temas: 1) conceptos básicos; 2) estimación de una proporción en caso discreto; 3) estimación de una proporción en caso continuo y 4) estimación de una media en caso continuo.

Con estos criterios se especificaron los siguientes contenidos que están explicados con más detalle en el Anexo 7.

- *Contenidos conceptuales:*
  - Probabilidades iniciales, finales y verosimilitudes.
  - Teorema de Bayes.
  - Población y muestra. Parámetro y estadístico. Parámetro como variable aleatoria.
  - Probabilidad subjetiva. Diferencia con una probabilidad clásica o frecuencial.
  - Distribución de probabilidad inicial y final. Distribución inicial informativa y no informativa.
  - Intervalo de credibilidad. Diferencia con el intervalo de confianza.
  - Aplicación sucesiva del procedimiento.
  - Paso de una distribución discreta a una continua.
  - Distribución Beta: sus parámetros y características. Forma, posición central y dispersión, en función de los parámetros.
  - Distribución normal: sus parámetros y características.
  - Procedimiento bayesiano de contraste.

- *Contenidos Procedimentales:*
  - Calcular las probabilidades finales a partir de las probabilidades iniciales y de las verosimilitudes.
  - Establecer una distribución inicial para la proporción p en la población

informativa y no informativa.

- Establecer una distribución inicial para la media, informativa y no informativa.
- Interpretación de tablas y cálculo de probabilidades finales mediante las distribuciones Beta y Normal.
- Cálculo de probabilidades relativas al parámetro en una distribución final.
- Utilizar las diferentes aplicaciones Excel e interpretar sus resultados.
- Extraer inferencias con la distribución a posteriori: dar un mejor estimador, calcular intervalos de credibilidad, contrastar una hipótesis.

- *Contenidos Actitudinales:*
  - Reconocer el interés del teorema de Bayes en situaciones de diagnóstico.
  - Reconocer la utilidad del uso del conocimiento previo en la extracción de inferencias sobre una población.
  - Reconocer la utilidad de la aplicación sucesiva del teorema de Bayes como procedimiento que permite aprender de la experiencia.

3.2.4. ELABORACIÓN Y REVISIÓN DEL MATERIAL MEDIANTE JUICIO DE EXPERTOS

Fijados los objetivos y contenidos se preparó un material para el estudiante, que fue sometido a juicio de expertos con objeto de confirmar su adecuación para los fines del estudio, detectar posibles errores u omisiones y mejorar su redacción, presentación e interés.

3.2.4.1. MÉTODO

3.2.4.1.1. SUJETOS

Para este juicio de expertos se pidió la colaboración de un grupo de profesores universitarios, todos ellos familiarizados con la estadística y con experiencia docente de esta materia a nivel universitario, con la particularidad de que ninguno hubiese estudiando anteriormente la inferencia bayesiana.

Se pidió colaboración a un total de quince profesores, de los cuáles respondieron doce y diez de ellos completaron el estudio del material. Este grupo de profesores estuvo compuesto por cuatro matemáticos, un educador y cinco estadísticos, de los cuales seis poseían un grado de Master o equivalente y cuatro de doctorado. Cuatro de ellos realizan su investigación en el campo de la estadística aplicada o métodos, uno en psicología educativa, uno en educación y cuatro en didáctica de las matemáticas. Tres participantes dictan cursos metodológicos en programas de doctorado. El número de años de docencia en estadística aplicada de estos profesores oscilaban entre 7 y 30 años, con una media de 15,7 años. La muestra incluía profesores tanto españoles como latinoamericanos.

3.2.4.1.2. MATERIAL

El material sometido a evaluación fue el preparado para los estudiantes y consta de material escrito, pruebas de autoevaluación y programas de cálculo. La experiencia de enseñanza se planificó en cuatro sesiones de tres horas cada una. La primera hora se dedicaría a la presentación del tema (objetivos, parte teórica y ejemplos) por parte de la profesora. La segunda a la presentación de ejercicios resueltos y discusión de posibles dudas y la tercera al manejo de las herramientas informáticas diseñadas para la experiencia y resolución de problemas.

Se planificaron también 5 horas de trabajo individual del alumno, cuatro correspondientes a actividades de autoevaluación, que se realizarían al final de cada sesión y una última hora que corresponde a una prueba de evaluación final realizada unos días más tarde.

Material escrito

El material escrito se dividió en 4 temas, correspondientes a cada una de las sesiones que se proyectaron para la enseñanza. Cada tema incluye objetivos, contenidos teóricos, ejemplos y ejercicios resueltos (ver anexo 7).

Programas auxiliares de cálculo

Se prepararon cuatro programas de cálculo que los alumnos podían usar para resolver las actividades, en la parte práctica de cada una de las sesiones. Los programas, que se describen con detalle en el Anexo 8 junto con un ejemplo de uso en la resolución de un problema, se describen a continuación.

El Programa *Bayes* se introduce en la primera sesión y calcula las probabilidades finales *P(Ai/D)* de un conjunto de sucesos *Ai*, dadas sus probabilidades iniciales *P(Ai)* y las verosimilitudes *P(D/Ai)* de unos ciertos datos *D*, dador los sucesos *Ai* para un máximo de ocho sucesos. Proporciona el producto de las probabilidades iniciales por las verosimilitudes y las probabilidades finales (figura 6.3) mediante la fórmula de Bayes:

$$P(A_i / B) = \frac{P(A_i) \times P(B / A_i)}{P(A_1) \times P(B / A_1) + P(A_2) \times P(B / A_2) + \ldots + P(A_n) \times P(B / A_n)}$$

Figura 6.3. Programa Bayes

CALCULO DE PROBABILIDADES FINALESI MEDIANTE TEOREMA DE BAYES				
	-----DATOS----			
Sucesos	Probabilidad inicial	Verosimilitud	Producto	Probabilidad final
A1	0,1	0,5	0,05	0,1220
A2	0,9	0,4	0,36	0,8780
A3			0	0,0000
A4			0	0,0000
A5			0	0,0000
A6			0	0,0000
A7			0	0,0000
A8			0	0,0000
	1		0,41	1,0000

El programa *Prodist* transforma una distribución inicial $P(p=p_0)$ de la proporción *p* en una población en su distribución final $P(p=p_0/datos)$ a partir de la aplicación del teorema de Bayes. Se requiere especificar el número de éxitos y fracasos en una muestra. Calcula las verosimilitudes $\mathrm{P(datos / p{=}p_0)}$ de cada posible valor de la proporción $\mathrm{p_0}$ mediante la expresión:

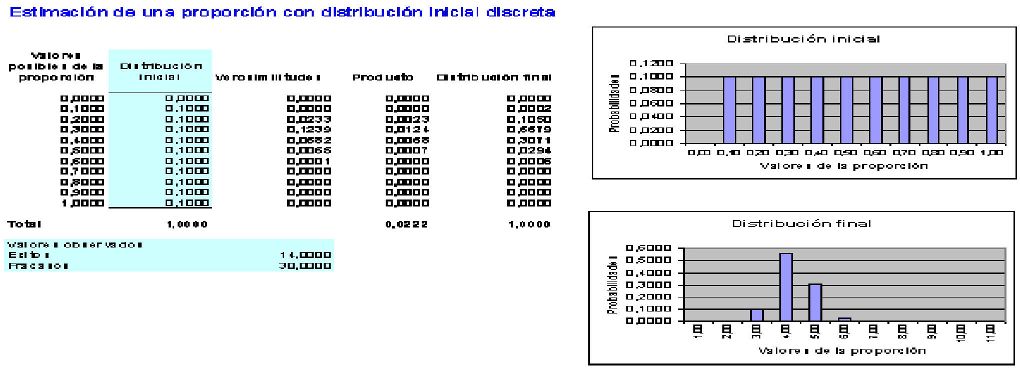
$$P(datos / p = p_0) = \binom{e+f}{e} p_0^{\,e} (1 - p_0)^f$$

y a partir de ellas proporciona las probabilidades en la distribución final:

$$P(p = p_0 / datos) = \frac{\mathrm{P(datos / p = p_0) \times P(p = p_0)}}{\sum \mathrm{P(datos / p) \times P(p)}}$$

También representa gráficamente las distribuciones iniciales y finales de la proporción, con objeto de hacer ver al alumno como cambia nuestro conocimiento sobre la proporción después de tomar los datos (figura 6.4).

Figura 6.4. Cálculo de Distribución final de la proporción (caso discreto)



Estimación de una proporción con distribución inicial discreta

Valores posibles de la proporción	Distribución inicial	Verosimilitudes	Producto	Distribución final
0,0000	0,0000	0,0000	0,0000	0,0000
0,1000	0,1000	0,0000	0,0000	0,0002
0,2000	0,1000	0,0233	0,0023	0,1050
0,3000	0,1000	0,1239	0,0124	0,5579
0,4000	0,1000	0,0682	0,0068	0,3071
0,5000	0,1000	0,0065	0,0007	0,0294
0,6000	0,1000	0,0001	0,0000	0,0006
0,7000	0,1000	0,0000	0,0000	0,0000
0,8000	0,1000	0,0000	0,0000	0,0000
0,9000	0,1000	0,0000	0,0000	0,0000
1,0000	0,1000	0,0000	0,0000	0,0000
Total	1,0000		0,0222	1,0000

Valores observados	
Exitos	14,0000
Fracasos	30,0000

El programa se presenta en dos versiones en diferentes hojas del libro Excel: 1) Valores de p variando en intervalos de 0,1 y 2) valores de p variando en intervalos de 0,05. Se introduce en la segunda sesión para que el alumno pueda calcular la mejor estimación de la proporción en la población, calcular probabilidades referidas al verdadero valor de la proporción en la población y usarlas para contrastar hipótesis o calcular intervalos de credibilidad de la proporción.

El programa *Beta* calcula probabilidades y valores críticos en una distribución *Be(e,f),* donde e y f se interpretan como el número de éxitos y fracaso en una muestra de $n= e+b$ ensayos y para el caso de distribución inicial no informativa[37]. Para valores de p_0 variando en intervalos de 0,05, calcula y representa gráficamente las probabilidades $P(0<p< p_0)$ y $P(p_0<p<1)$:

$$P(0<p< p_0) = \beta_{0,5+e,\, 0,5+f}(p_0)$$

$$P(p_0<p<1) = 1-\beta_{0,5+e,\, 0,5+f}(p_0)$$

donde $\beta^{-1}_{a,b}$ es la función de distribución *Be(a,b)*. Calcula también los valores críticos x de la proporción tal que la probabilidad $P(0<p< x)$ está fijada desde 0 a 0,5 variando en intervalos de 0,05 y desde 0,95 a 1 variando en intervalos de 0,05 dados por.

$$x = \beta^{-1}_{0,5+e,\, 0,5+f}(p)$$

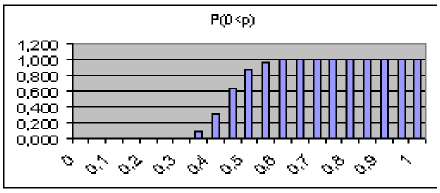
[37] En el caso de distribuciones iniciales informativas, basta especificar el total de éxitos y fracasos totales (sumando los parámetros de la distribución inicial y los de los datos), puesto que la familia de distribuciones Beta es reproductiva respecto a sus parámetros (Serrano, 2003).

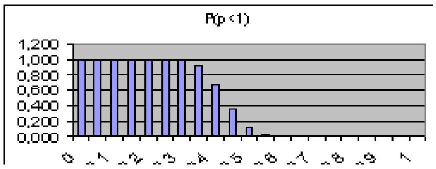
Figura 6.5. Cálculo de probabilidades y valores críticos en la distribución Beta

Cálculo de probabilidades y valores críticos con la distribución Beta

	Datos
Parámetro a	30
Parámetro b	40

Probabilidades			Valores críticos	
pu	P(0<p<p0)	P (p0 <p<1)	Probabilidad	Valor crítico
0	0,000	1,000	0,000	0,000
0,05	0,000	1,000	0,005	0,296
0,1	0,000	1,000	0,010	0,304
0,15	0,000	1,000	0,015	0,311
0,2	0,000	1,000	0,020	0,316
0,25	0,001	0,999	0,025	0,320
0,3	0,012	0,988	0,030	0,324
0,35	0,090	0,910	0,035	0,327
0,4	0,318	0,682	0,040	0,330
0,45	0,645	0,355	0,045	0,330
0,5	0,886	0,114	0,050	0,333
0,55	0,979	0,021	0,950	0,526
0,6	0,998	0,002	0,955	0,529
0,65	1,000	0,000	0,960	0,533
0,7	1,000	0,000	0,965	0,536
0,75	1,000	0,000	0,970	0,541
0,8	1,000	0,000	0,975	0,546
0,85	1,000	0,000	0,980	0,551
0,9	1,000	0,000	0,985	0,558
0,95	1,000	0,000	0,990	0,567
1	1,000	0,000	1,000	1,000





El programa se introduce en la tercera sesión y permite calcular probabilidades referidas al verdadero valor de la proporción en la población; calcular valores críticos, y usar estos resultados para contrastar hipótesis o calcular intervalos de credibilidad de la proporción.

Finalmente el programa *Media* calcula la media μ_f y la desviación típica σ_f de la distribución final de la media de una población normal $N(\mu_f, \sigma_f)$, cuando se conocen la media μ_0 y desviación típica σ_0 de la distribución inicial así como la media $\bar{x}$ y desviación típica S de los datos de la muestra.

Figura 6.6. Cálculo de media y desviación típica de la distribución final de la media en una población normal

Estimación de la media de una muestra. Distribución normal a

Tamaño de la muestra		Media	Desviación Tipica	Precisión
30	En la distribución inicial	84	5	0,04
	En los datos	88,8	2,43	0,17
	En la distribución final	87,88	2,19	0,21

Una vez conocidas la media y desviación típica de la distribución final podemos usar las tablas de la distribución $N(0,1)$ para realizar todo tipo de cálculos de probabilidades, intervalos de credibilidad o contrastes.

Finalizada la preparación del material escrito se elaboró una página de Internet (http://www.ugr.es/local/mcdiaz/bayes) para que los alumnos tuvieran el material disponible y para poder realizar una experiencia de enseñanza a distancia. La página principal contiene un menú desde donde se podía acceder a cada una de las cuatro lecciones, a la sección de descarga de los programas Excel y a un foro de discusión donde los participantes en el curso en modalidad virtual podían discutir entre ellos o con el profesor posibles dudas (ver figura 6.7). El contenido de esta página web se presenta en CD ROM como Anexo a esta Memoria.

Una vez que el alumno accede a uno de los temas se encuentra con los objetivos de ese tema y un menú en la parte superior para ir a cada una de las partes: teoría, ejemplos, ejercicios resueltos, actividades y autoevaluación. La página está diseñada para que sea de fácil navegación. El menú siempre está en la parte superior y el alumno puede navegar a través de los temas siguiendo el índice.

Figura 6.7. Pantalla de Inicio de la Página Web de Acceso al Material Didáctico

Autoevaluaciones y pruebas abiertas de resolución de problemas

Se preparó un cuestionario de autoevaluación para cada uno de los temas, cada uno de ellos formado por un conjunto pequeño de ítems, en formato de opciones múltiples. Asimismo se prepararon unas pruebas de resolución de problemas (1-2 por tema) en forma que en su conjunto se evaluase el aprendizaje de los estudiantes sobre los objetivos y contenidos del tema tratado.

Nos situamos en un dominio curricular, que se refiere a un conjunto de habilidades y conocimientos que se desarrollan como resultado de la instrucción deliberada sobre un cierto contenido curricular (Millman y Greene, 1989). El conjunto de pruebas en su conjunto es un instrumento de *medida centrada en el sujeto* y más concretamente un *test referido a criterio* (Sax, 1989). Además podemos clasificarlo como *un cuestionario de potencia*, ya que las diferencias en la puntuación entre los sujetos que respondan al cuestionario serán debidas a la calidad de su ejecución y conocimiento y un *test psicométrico* puesto que evalúa las respuestas en forma cuantitativa y se refiere a un rasgo diferenciado del sujeto (Martínez Arias, 1995).

La definición *semántica* de la variable objeto de medición se realizó a partir del estudio de los objetivos y contenidos fijados para la enseñanza (descritos en el estudio 8.1) y del análisis de contenido del material entregado a los alumnos, que se presenta en el anexo 7. Con ello se determinaron las unidades de contenido objeto de la evaluación, que se presentan en la tabla 6.30, junto con los ítems que evalúan cada uno de estos contenidos. Los ítems se especifican mediante la notación Ítem x_y y los ejercicios como Ej. x_y, donde el primer dígito indica el tema y el segundo el ítem dentro del tema.

Una vez especificados los contenidos se elaboraron ítems de nivel de aplicación y análisis en la taxonomía Bloom (1956)[38] El nivel de aplicación implica el uso de abstracciones (reglas, procedimientos, métodos generales, principios técnicos o teorías) que deben recordarse y aplicarse en situaciones particulares y concretas. El nivel de análisis supone descomponer la comunicación en sus elementos o partes constituyentes de modo que la jerarquía de ideas y relaciones entre ideas se expresen y hagan explícitas.

[38] Diversos autores (Sax, 1980; Martínez Arias, 1995) señalan esta taxonomía como la más conocida, y por otro lado es muy usada en investigaciones sobre enseñanza de las matemáticas. Esta taxonomía tiene seis niveles, conocimiento, comprensión, aplicación, análisis, síntesis y evaluación.

Tabla 6.30. Unidades de contenido de la variable objeto de medición e ítems que lo evalúan

	Contenidos evaluados en las autoevaluaciones y pruebas de resolución de problemas	Ítem
Relacionados con la inferencia bayesiana	1. Diferenciar entre probabilidades iniciales y finales y verosimilitudes.	1_1,1_2
	2. Identificar los sucesos de interés, sus probabilidades iniciales y verosimilitudes a partir del enunciado de un problema.	1_5
	3. Analizar el teorema de Bayes como herramienta para transformar probabilidades iniciales en finales.	1_3, 1_8
	4. Organizar los datos en una tabla de forma que se facilite el cálculo de probabilidades finales a partir de las iniciales y verosimilitudes.	1_6
	5. Visualizar los parámetros como variables aleatorias con una distribución de probabilidad asociada.	2_1, 2_2
	6. Identificar valores de la proporción poblacional y asignarles probabilidades iniciales plausibles (distribución de probabilidad inicial).	2_3 ,Ej.2_1
	7. Comprender el concepto de distribución inicial uniforme y distribución informativa.	2_5
	8. Asignar correctamente una distribución inicial no informativa a la proporción.	2_6
	9. Asignar correctamente una distribución inicial informativa a la proporción.	2_7
	10. Identificar los valores de éxitos y fracasos en la muestra.	Ej.2_1
	11. Aprender a introducir los datos anteriores en la hoja Excel para obtener los valores de verosimilitud y la distribución final.	2_8 ,Ej.2_1
	12. Reconocer la fórmula correcta del teorema de Bayes en formulación no estándar.	2_4
	13. Obtener la mejor estimación de la proporción en la población.	Ej.2_1, 3_2, 3_4
	14. Calcular probabilidades asociadas a determinados valores de proporción en la distribución final.	Ej.2_2,3_5
	15. Entender cómo las distribuciones finales se transforman en iniciales en experimentos sucesivos.	Ej.2_2
	16. Analizar las diferencias entre la metodología clásica y bayesiana.	2_2
	17. Usar la distribución Beta para definir una distribución inicial para la proporción en caso no informativo.	3_1, Ej.3_3
	18. Usar la distribución Beta para definir una distribución inicial para la proporción en caso informativo.	3_3
	19. Identificar los éxitos y fracasos en la muestra.	3_3, 3_4, Ej.3_3
	20. Calcular la distribución final con la fórmula $Be(a + e, b + f)$.	3_4, Ej.3_3
	21. Usar la moda de la distribución Beta $a/(a+b)$ como mejor estimador.	3_2, Ej.3_3
	22. Utilizar la herramienta Excel para obtener una tabla con la distribución de probabilidad final.	Ej.3_3
	23. Calcular intervalos de credibilidad.	3_5
	24. Interpretar el intervalo de credibilidad de una proporción.	3_6,3_7 ,Ej.3_3
	25. Contraste de hipótesis sencillas sobre una proporción siguiendo la metodología bayesiana.	3_6, Ej.3_3
	26. Diferenciar entre estadístico y parámetro en el contexto de un problema.	2_1
	27. Especificar valores plausibles de la media poblacional y asignarles probabilidades iniciales.	4_1, Ej.4_1, Ej.4_2
	28. Con los datos obtenidos en una muestra, obtener una distribución final.	4_2, 4_4, Ej.4_1, Ej.4_2
	29. Extraer inferencias sobre la población a partir de la distribución final.	4_7
	30. Cálculo de intervalos de credibilidad para la media de una población.	4_5, Ej.4_1
	31. Interpretar un intervalo de credibilidad de la media.	4_6, Ej.4_2
	32. Realizar correctamente un contraste de hipótesis sobre la media, desde una perspectiva bayesiana.	4_3, Ej.4_3
Otras	33. Diferenciar una probabilidad condicional y su transpuesta.	1_4
	34. Diferenciar probabilidades condicionales y conjuntas en el contexto de un problema.	1_7

Los ítems son de elaboración propia (puesto que no hemos encontrado cuestionarios adaptados a nuestro trabajo) y en su construcción se siguieron los criterios de calidad habituales (Osterlind, 1989):

- Alto grado de congruencia entre el ítem particular y el contenido que se trata de evaluar.
- Pertinencia del formato del ítem a los fines del cuestionario. Dado que nos encontramos ante un test de ejecución óptima (cuestionario de rendimiento) hemos considerado adecuado realizar una parte de los ítems en un formato de elección múltiple, por tanto de respuesta construida, donde se proporciona al examinado la respuesta correcta, junto con otras alternativas y se pide al examinado elegir una de las opciones (Osterlind, 1989). Se eligió un número de alternativas de 4, todas ellas plausibles (Martínez Arias, 1995).
- Además, nuestros ítems cumplen los criterios de independencia local (la repuesta a uno no depende de la respuesta a otro), unidimensionalidad (pues todos se refieren al constructo aprendizaje de inferencia bayesiana), y satisfacen las cuestiones éticas y legales.

En la figura 6.8 presentamos un ejemplo de autoevaluación. La página muestra al principio un pequeño formulario con los datos del alumno (nombre, apellidos, grupo y e-mail). El alumno debe marcar la alternativa correcta y enviar los resultados al profesor. En las tablas 6.31 a 6.38 se incluyen los cuestionarios de autoevaluación y ejercicios.

Figura 6.8. Pantalla de la Pagina Web con una autoevaluación

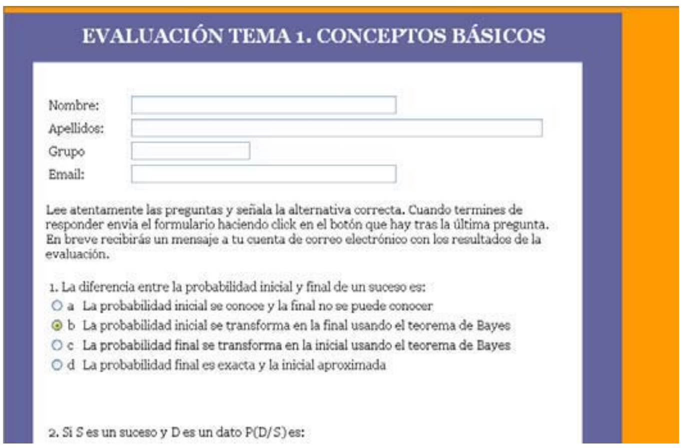


Tabla 6.31 Autoevaluación del tema 1

1. La diferencia entre la probabilidad inicial y final de un suceso es:
a. La probabilidad inicial se conoce y la final no se puede conocer
b. La probabilidad inicial se transforma en la final usando el teorema de Bayes
c. La probabilidad final se transforma en la inicial usando el teorema de Bayes
d. La probabilidad final exacta y la inicial aproximada

2. Si E es un suceso y *D* es un dato *P(D/E)* es:
a. La verosimilitud de obtener el dato si ocurre el suceso
b. La probabilidad final del suceso
c. La probabilidad inicial del suceso
d. La probabilidad conjunta de que ocurran a la vez el suceso y el dato

3. El teorema de Bayes permite calcular:
a. Las verosimilitudes cuando se conocen las probabilidades iniciales y finales
b. Las probabilidades iniciales si se conocen las versomilitudes y las probabilidades finales
c. Las probabilidades finales cuando se conocen las iniciales
d. Las probabilidades finales cuando se conocen las iniciales y verosimilitudes

4. El valor de la probabilidad *P(D/E)* es aproximadamente igual al de la probabilidad *P(E/D):*
a. En todos los casos
b. Depende de valor de *P(E)* y del valor de *P/D/E)*
c. Nunca puede ser igual

5. En una ciudad 1 de cada 100 personas estudia inglés. 90 de cada 100 personas que estudia inglés ha viajado al extranjero y también 10 de cada 100 personas que no estudian inglés. La probabilidad de que al tomar una persona al azar, de entre todas las que han viajado al extranjero, haya estudiado inglés es:
a. Una probabilidad inicial *P (I)*
b. Una probabilidad final *P(I/E)*
c. Una verosimilitud *P(E/I)*
d. Una probabilidad conjunta $P\ (E \cap I)$

6. Para calcular la probabilidad final mediante el teorema de Bayes:
a. Multiplicamos las probabilidades iniciales por las verosimilitudes y el resultado lo dividimos por la suma de todos estos productos
b. Multiplicamos las probabilidades iniciales por las verosimilitudes y el resultado lo dividimos por la suma de las verosimilitudes
c. Multiplicamos las probabilidades iniciales por las verosimilitudes y el resultado lo dividimos por la suma de las probabilidades iniciales
d. Multiplicamos las probabilidades iniciales por las verosimilitudes

7. Imagina que tienes una muestra de 100.000 personas elegidas al azar. Imagina que 5 de cada 1000 está deprimida. Supón que una prueba de depresión da positivo en 99 de cada 100 personas enfermas y también en 3 de cada 100 personas sanas. Si *D* significa depresión y + significa prueba positiva, entonces:
a. *P(D)*=0,00005; *P(D/+)*=0,99; *P(S/+)*=0,03
b. *P(D)*=0,00005; *P(+/D)*=0,99; *P(-/D)*=0,01
c. *P(D)*=0,00005; *P(D/+)*=0,99; *P(-/D)*=0,01
d. *P(D)*=0,0005; *P(+/D)*=0,99; *P(-/D)*=0,01

8. Una de las siguientes fórmulas del Teorema de Bayes es falsa, ¿Cuál de ellas es falsa?

a. Es falsa $$P(A_i / B) = \frac{P(A_i) \times P(B / A_i)}{P(A_1) \times P(B / A_1) + P(A_2) \times P(B / A_2) + \ldots + P(A_n) \times P(B / A_n)}$$

b. Es falsa $P(B) = P(B/A_1)\ P(A_1) + \ldots + P(B/A_n)\ P(A_n)$
c. Es falsa *Probabilidad final = Kx Probabilidad inicial x Verosimilitud*
d. Es falsa $P(A_i / B) = K \times P(A_i) \times P(B / A_i)$

Tabla 6.32 Autoevaluación del tema 2

1. En un estudio sobre costumbres de los jóvenes españoles se hizo una encuesta a una muestra representativa de 1000 jóvenes. De ellos 700 tenían carnet de conducir.

a. La proporción $p=0,7$ de jóvenes con carnet de conducir es un parámetro.
b. La proporción $p=0,7$ de jóvenes con carnet de conducir es una estimación del verdadero valor del parámetro en la población.
c. El 70% de los jóvenes españoles tienen carnet de conducir.
d. Si tomamos otros 1000 jóvenes más, obtendremos otros 700 con carnet de conducir.

2. La proporción p de elementos que en una población cumplen un criterio dado:

a. Se considera constante en inferencia bayesiana
b. Se considera variable en inferencia clásica
c. Se considera variable en inferencia bayesiana
d. En unas poblaciones es constante y en otras es variable

3. La distribución inicial de la proporción p en una población viene dada por:

a. El valor exacto de la proporción en la población
b. El valor exacto de la proporción en la población y el tamaño de la muestra
c. El número de éxitos y fracasos en la muestra
d. Los valores posibles de la proporción en la población y la probabilidad de cada uno de ellos.

4. Conocidas las probabilidades iniciales de la proporción p en la población las probabilidades finales se calculan con la siguiente fórmula:

a. $P\ (p = p_0\ /\ datos) = K\ x\ P(p = p_0)\ x\ P(datos/\ p = p_0)$
b. $P\ (p = p_0) = K\ x\ P(p = p_0/\ datos)\ x\ P(datos/\ p = p_0)$
c. $P(datos/\ p = p_0) = K\ x\ P(p = p_0)\ x\ P\ (p = p_0\ /\ datos)$
d. $P\ (p = p_0\ /\ datos) = K\ x\ P(p = p_0)\ /P(datos/\ p = p_0)$

5. La distribución inicial de probabilidades de un parámetro:

a. Recoge toda la información de la persona sobre la población antes de recoger los datos
b. Recoge toda la información de la persona sobre la población después de recoger los datos
c. Sirve para calcular el intervalo de credibilidad del parámetro
d. Es una distribución fija; no podemos cambiarlas en nuevos experimentos

6. Cuando no tenemos ninguna información sobre los valores posibles de una proporción en la población, y queremos asignar una distribución inicial a la proporción en la población:

a. Usaremos sólo el valor p=1/2, dando a este valor probabilidad 1 y al resto probabilidad cero
b. Usaremos varios valores de la proporción entre 0 y 1 dando a todos la misma probabilidad
c. No podemos asignar una distribución inicial, porque no sabemos nada de la proporción
d. Usaremos sólo los valores 0 y 1 dando a cada uno probabilidad ½

7. En un edificio de 10 plantas sabemos que entre 6 y 8 están equipados de alarma contra incendios. También sabemos que la probabilidad de que haya exactamente 6 equipados de alarma es doble que la de que el número sea 6 o de que sea 7¿Cuál de las columnas A, B, C o D describe mejor la distribución inicial de la proporción de edificios con alarma?

Posibles valores de p	(A)	(B)	(B)	(B)
0,00	0,00	0,00	0,00	0,00
0,10	0,1	0,00	0,00	0,00
0,20	0,1	0,00	0,00	0,00
0,30	0,1	0,00	0,00	0,00
0,40	0,1	0,00	0,00	0,00
0,50	0,1	0,00	0,00	0,00
0,60	0,2	0,50	0,40	0,30
0,70	0,1	0,25	0,20	0,30
0,80	0,1	0,25	0,20	0,30
0,90	0,1	0,00	0,00	0,00
1,00	0,00	0,00	0,00	0,00

8. En una Facultad se quiere estimar la proporción de estudiantes que estudia inglés. A partir de una muestra de 10 estudiantes 7 estudiaban inglés. ¿Cuál es la verosimilitud de obtener estos datos, si exactamente el 50 % de los alumnos estudia inglés?

a.P (datos / p=0,5)= $P(X=7)=\binom{10}{7}0,5^7(0,5)$

b.P (datos / p=0,5)= $P(X=7)=\binom{10}{7}0,5^7(0,5)^3$

c.P (datos / p=0,5)= $P(X=7)=0,5^7(0,5)^3$

$d.P(X=7)=\binom{10}{7}0,5^7$

9. Al comenzar a resolver un problema de estimación de una proporción un alumno ha completado las tres primeras columnas de la tabla Bayes, obteniendo los datos siguientes:

(1) Posibles valores de *p*	(2) Probabilidad inicial	(3) Verosimilitud	--------	-----------
0,0000	0,0000	0,0000		
0,1000	0,1000	0,0000		
0,2000	0,1000	0,0233		
0,3000	0,1000	0,1239		
0,4000	0,1000	0,0682		
0,5000	0,1000	0,0065		
0,6000	0,1000	0,0001		
0,7000	0,1000	0,0000		
0,8000	0,1000	0,0000		
0,9000	0,1000	0,0000		
1,0000	0,1000	0,0000		
Suma			0,0222	

La probabilidad final de que el verdadero valor de la proporción sea igual a ½ es:
a. 0,065 b. 0,100 c. 0,0294 d. 0,0007

Tabla 6.33. Autoevaluación del tema 3

1. Cuando no tenemos ninguna información sobre los valores posibles de una proporción en la población, usaremos como distribución inicial de la proporción para el caso continuo:
a. La distribución normal *N*(0,1) de media cero y desviación típica 1.
b. La distribución uniforme discreta en el intervalo [0,1].
c. La distribución *B* (0,1).
d. La distribución *B* (1,1).

2. Supongamos que describo la proporción de emigrantes en una ciudad española mediante una distribución inicial *B* (3, 97). Esto quiere decir que mi mejor estimación inicial de la proporción de emigrantes en la ciudad es:
a. El 3 %
b. El 97%
c. Algo más del 3%
d. Exactamente igual a 3 /97 %

3. Un estudio médico indica una incidencia de 10% de depresión en las mujeres en alguna época de su vida. Una posible distribución aproximada inicial para definir esta población sería:
a. *B* (10, 100)
b. *B* (10, 90)
c. *B* (10,10)
d. *B* (90, 10)

4. En un estudio sobre conducta social en el juego de los niños de preescolar se acepta como distribución inicial de la proporción de niños que acepta sus equivocaciones una distribución *B* (40, 60). En una muestra de 100 nuevos casos, 55 niños aceptaron sus equivocaciones durante el juego. La mejor estimación para la proporción de niños que acepta sus equivocaciones es

a. El 55%
b. El 40%
c. El 60%
d. El 47,5%

5. La siguiente tabla presenta las probabilidades y valores críticos de la distribución *B*(40,60).

p	*P(0<p)*	*P (p<1)*	*P(0<x)*	Valor crítico *(x)*
0	0,000	1,000	0,000	0,000
0,05	0,000	1,000	0,005	0,290
0,1	0,000	1,000	0,010	0,297
0,15	0,000	1,000	0,015	0,302
0,2	0,000	1,000	0,020	0,307
0,25	0,001	0,999	0,025	0,310
0,3	0,018	0,982	0,030	0,313
0,35	0,153	0,847	0,035	0,316
0,4	0,505	0,495	0,040	0,319
0,45	0,846	0,154	0,045	0,319
0,5	0,978	0,022	0,050	0,321
0,55	0,999	0,001	0,950	0,481
0,6	1,000	0,000	0,955	0,484
0,65	1,000	0,000	0,960	0,487
0,7	1,000	0,000	0,965	0,490
0,75	1,000	0,000	0,970	0,493
0,8	1,000	0,000	0,975	0,497
0,85	1,000	0,000	0,980	0,502
0,9	1,000	0,000	0,985	0,508
0,95	1,000	0,000	0,990	0,516
1	1,000	0,000	1,000	1,000

El intervalo de credibilidad del 95% para la proporción de una población descrita por una distribución final *B* (40, 60) es aproximadamente:

a. $(0,35 < p < 0,5)$
b. $(0 < p < 0,516)$
c. $(0,310 < p < 0,497)$
d. $(0,25 < p < 0,8)$

6. A partir de los datos de la tabla anterior lo más razonable es aceptar la siguiente hipótesis sobre la proporción de niños de preescolar que acepta sus equivocaciones:

a. $H: p<0,3$
b. $H: p > 0,55$
c. $H: p> 0,35$
d. $H: p >0,45$

7. Para un mismo valor de la proporción en una muestra y una misma distribución inicial el intervalo de credibilidad del *r*% para la proporción en la población es:

a. Más ancho si aumento el tamaño de muestra
b. Más ancho si aumento el valor de *r*
c. Más estrecho si aumento el valor de *r*
d. Depende de la distribución final.

Tabla 6.34. Autoevaluación del tema 4

1. Estoy realizando inferencias sobre la media de una población normal de desviación típica σ=2, pero no tengo ninguna información sobre su posible valor. En una muestra de 25 elementos he obtenido una media $\bar{x}$ =10. Entonces tomaré como distribución <u>inicial</u>:
a. La distribución normal N (0,1)
b. La distribución normal N (10, 0'4)
c. La distribución normal N (10, 5)
d. Una distribución uniforme.

2. Si en una muestra de 100 elementos de una población normal de desviación típica σ=5 donde no tengo información inicial sobre la distribución de la media en la población he obtenido una media muestra $\bar{x}$ =10, la distribución <u>final</u> de la media en la población es:
a. Una distribución uniforme.
b. Una distribución normal N (0,1)
c. Una distribución normal N (10,5)
d. Una distribución normal N (10, 0'5)

3. Para contrastar la hipótesis de que la media μ en una población normal con desviación típica σ=1 es menor que 5, mediante los datos obtenidos en una muestra de 16 elementos, siguiendo el procedimiento bayesiano:
a. Calculo la media de la muestra $\bar{x}$; calculo luego P ($\frac{\bar{x}-5}{0,25}$ <5); si esta probabilidad es muy pequeña, acepto la hipótesis.
b. Calculo la media de la muestra $\bar{x}$; calculo luego P ($\frac{\bar{x}-5}{1}$ <Z); siendo Z la distribución normal N (0,1); si esta probabilidad es muy pequeña acepto la hipótesis.
c. Calculo la media de la muestra $\bar{x}$; calculo luego P ($\frac{\bar{x}-5}{0,25}$ > Z); siendo Z la distribución normal N (0,1); si esta probabilidad es muy pequeña, acepto la hipótesis.
d. Calculo la media de la muestra $\bar{x}$; calculo luego P ($\frac{\bar{x}-5}{0,25}$ <5); si esta probabilidad es muy pequeña, acepto la hipótesis.

4. En una muestra de 100 elementos de una población normal he obtenido una media $\bar{x}$. Suponiendo una distribución inicial uniforme de la media en la población, la distribución final de la media es:
a. Aproximadamente N ($\bar{x}$, 1)
b. Aproximadamente N ($\bar{x}$, s), siendo s la desviación típica de la muestra
c. No puedo decir nada, porque no conozco la desviación típica de la población
d. Aproximadamente N (0,1)

5. En una muestra de 25 elementos de una distribución normal con desviación típica he obtenido una media $\bar{x}$ =10 σ=1, el intervalo de credibilidad del 95% de la media en la población es:
a. [10-1'96 x 0'2; 10 + 1'96 x 0'2]
b. [10-1'96; 10 +1'96]
c. [10 x 0'2 –1'96; 10 x 0'2 + 1'96]
d. [10-1'96 x 5; 10 +1'96 x 5]

6. Si en una misma población calculo en el intervalo de credibilidad para la media de la muestra, el intervalo es:
a. Más ancho si aumento el tamaño de la muestra
b. Más ancho si aumenta la desviación típica de la muestra

c. Más ancho si aumento la credibilidad
d. Más ancho si disminuyo la credibilidad

7. La tabla adjunta muestra los resultados de estimar la distribución final de la media mediante inferencia bayesiana.

Estimación de la media de una muestra. Distribución normal a prio

Tamaño de la muestra		Media	Desviación Tipica	Precisión
24	En la distribución inicial	100	5	0,04
	En los datos	88,8	2,43	0,16935088
	En la distribución final	90,939948	2,18555936	0,20935088

Para calcular un intervalo de credibilidad de la media en la población a partir de esta tabla usaríamos:
a. Una distribución normal *N* (100, 5)
b. Una distribución normal *N* (90, 2'18)
c. Una distribución normal *N* (88, 2'43)
d. Una distribución normal *N* (88, 2'43 /24)

Tabla 6.35. Ejercicios de respuesta abierta tema 1

1_1. Una fábrica de enlatados produce 5000 envases diarios. La máquina A produce 3000 de estos envases, de los que el 2% son defectuosos y la máquina B produce los 2000 restantes de los que se sabe que el 4% son defectuosos.
a. Organiza los cálculos utilizando una tabla como la 1.1.
b. Si seleccionamos al azar un envase defectuoso, calcula la probabilidad de haber sido fabricado por A. Calcula la probabilidad de que haya sido fabricado por B.

1_2. Un edificio está equipado de un sistema de alarma contra incendios. De existir peligro la alarma se activa en el 99 por ciento de las veces. También puede producirse una falsa alarma con probabilidad 0,005 en el caso de no haber incendio. Si la probabilidad de incendio es 0,002, responde a las siguientes preguntas:
a. Organiza en una tabla los datos identificando las probabilidades iniciales, verosimilitudes y calcula las probabilidades finales en este problema.
b. Si hay una alarma, ¿Cuál es la probabilidad de que sea infundada? ¿Y de que sea cierta?

Tabla 6.36. Ejercicios de respuesta abierta tema 2

2_1. Una marca de bombillas se vende en cajas de 10.
1. Escribe una distribución inicial para la proporción de bombillas fundidas en la caja;
2. En una muestra de 60 bombillas, 5 fueron defectuosas. Calcula la distribución final de la proporción de bombillas fundidas en las cajas de 10.
3. ¿Cuál es la proporción más probable de bombillas fundidas en las cajas?

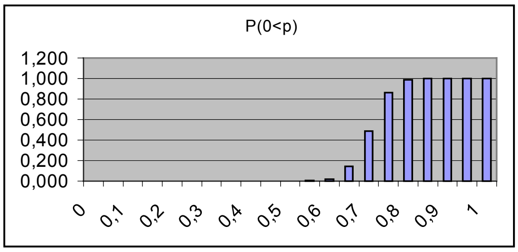
2_2. En el ejemplo de las votaciones, ¿Cuál es la probabilidad de que el candidato A gane el 60% o más de los votos? Revisa la distribución final si en una nueva muestra de 10 votantes obtenemos 6 partidarios a A.

Tabla 6.37. Ejercicios de respuesta abierta tema 3

3.1. En un estudio sobre satisfacción en el consumo se encontraron 69 consumidores satisfechos y 29 insatisfechos. No tenemos información previa sobre la satisfacción del consumidor. Contesta las siguientes preguntas:

1. ¿Qué distribución Beta usarías para reflejar tu conocimiento previo sobre la población antes de coger los datos?
2. ¿Cuál sería la distribución Beta final que refleja tu creencia modificada?
3. ¿Cuál es el mejor estimador para la proporción en la población antes de coger los datos?
4. ¿Cuál es el mejor estimador de la proporción en la población después de coger los datos?
5. Calcula el intervalo de credibilidad del 95% del valor de la proporción en la población
6. ¿Cuál es tu mejor inferencia para la probabilidad de que más de la mitad de los consumidores en la población están satisfechos?
7. ¿Aceptarías la hipótesis de que más del 60% de los consumidores están satisfechos?

p	P(0<p)	P (p<1)	P(0<x)	Valor crítico (x)
0	0,000	1,000	0,000	0,000
0,05	0,000	1,000	0,010	0,589
0,1	0,000	1,000	0,015	0,597
0,15	0,000	1,000	0,020	0,602
0,2	0,000	1,000	0,025	0,607
0,25	0,000	1,000	0,030	0,611
0,3	0,000	1,000	0,035	0,614
0,35	0,000	1,000	0,040	0,617
0,4	0,000	1,000	0,045	0,620
0,45	0,000	1,000	0,005	0,576
0,5	0,000	1,000	0,500	0,701
0,55	0,001	0,999	0,950	0,773
0,6	0,018	0,982	0,055	0,625
0,65	0,138	0,862	0,960	0,777
0,7	0,488	0,512	0,965	0,779
0,75	0,864	0,136	0,970	0,782
0,8	0,990	0,010	0,975	0,785
0,85	1,000	0,000	0,980	0,789
0,9	1,000	0,000	0,985	0,794
0,95	1,000	0,000	0,990	0,800
1	1,000	0,000	1,000	1,000



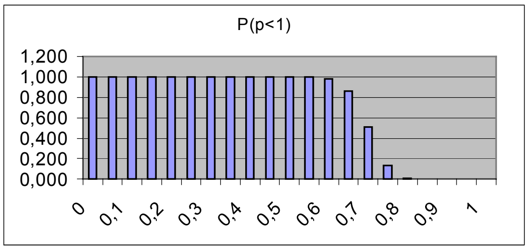


Tabla 6.38. Ejercicios de respuesta abierta tema 4

4_1. Una muestra de 25 alumnas de una facultad se pesa, dando un peso medio de 52 kg. No sabemos nada del peso medio en la población, aunque sabemos que su desviación típica es igual a 5 kg. Calcula el intervalo de credibilidad del 95% para el peso medio de la población.

4_2. En una comunidad autónoma se estudia el número medio de hijos por mujer a partir de los datos disponibles en cada municipio. Se supone que este número sigue una distribución normal con desviación típica igual a 0,08. El valor medio de estos datos para 36 municipios resulta ser 1,17. ¿Podemos suponer que el número medio es mayor que 1?

3.2.4.1.3. PROCEDIMIENTO Y ANÁLISIS

En primer lugar se contactó con los expertos mediante correo electrónico solicitando su colaboración en el estudio y explicando cual sería su tarea.

A los participantes que ofrecieron su ayuda se les pidió, en primer lugar que estudiasen ellos mismos los temas (utilizando el material para enseñanza a distancia disponible en Internet), siguiendo el orden establecido, leyendo los ejemplos y ejercicios resueltos y resolviendo ellos mismos los propuestos. Se les pidió también completar y enviar la prueba de autoevaluación de cada tema. Cada uno de ellos podía trabajar a su propio ritmo, aprovechando la facilidad que para ello ofrece la enseñanza a distancia. Se solicitó también de los participantes sus comentarios sobre legibilidad del material, posibles mejoras, detección de pequeñas erratas y cualquier otro punto que sirviera para mejorar el material.

Obtenidas todas las respuestas, se realizó un estudio descriptivo del número de respuestas correctas en cada uno de los ítems que componen cada una de las cuatro pruebas de autoevaluación, así como número total de errores cometidos por cada participante en el total de pruebas, con el fin de identificar posibles ítems cuyo enunciado fuese confuso o demasiado difíciles.

Los comentarios adicionales sobre el material se utilizaron para corregir algunas erratas, cambiar algunos ejemplos o ejercicios y mejorar la redacción en general.

3.2.4.2. RESULTADOS

Las pruebas de autoevaluación resultaron sencillas para los participantes, que en su mayoría, completaron correctamente los ítems propuestas, como se puede muestra en la distribución del el número de errores cometidos por los expertos en cada una de las autoevaluaciones (figura 6.9). El tema 2 resultó el mas sencillo, con una mayoría de 0 errores.

Tabla 6.39. Frecuencias de respuestas correctas en las pruebas de autoevaluación (*n*=10)

Evaluación 1		Evaluación 2		Evaluación 3		Evaluación 4	
Ítem	Respuestas correctas	Item	Respuestas correctas	Ítem	Respuestas correctas	Ítem	Respuestas correctas
1	10	1	10	1	10	1	8
2	10	2	10	2	10	2	10
3	10	3	10	3	9	3	10
4	6	4	9	4	9	4	10
5	7	5	10	5	10	5	10
6	10	6	10	6	7	6	10
7	9	7	10	7	10	7	9
8	9	8	10				
		9	10				

El promedio de errores por persona fue 2,5 en el total de las cuatro evaluaciones (31 ítems), lo que supone un promedio de 0,6 errores por tema y 0,08 por ítem. Aunque estos datos indicaban que las pruebas de autoevaluación eran sencillas, realizamos un análisis adicional de las preguntas donde se concentraron la mayor parte de los errores, que se presentan en la tabla 6.40 y se revisó la redacción de estas preguntas.

Figura 6.9. Frecuencia de errores en el total de la muestra por tema

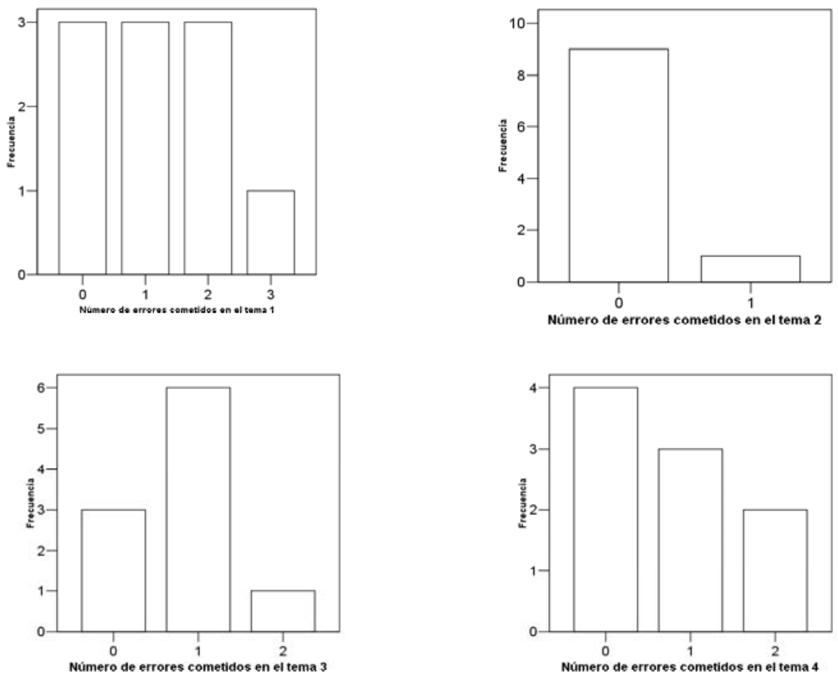


Tabla 6.40. Ítems de la autoevaluación que produjeron 3 o más errores en la muestra de expertos (n=10)

Ítem	Nº errores
El valor de la probabilidad *P(D/S)* es aproximadamente igual al de la probabilidad P(S/D): a) En todos los casos b) Sólo si *P(S)* es igual a *P(D)* c) Nunca puede ser igual	4
Imagina que tienes una muestra de 100.000 personas elegidas al azar. Imagina que 5 de cada 1000 está deprimida. Supón que una prueba de depresión da positivo en 99 de cada 100 personas enfermas y también en 3 de cada 100 personas sanas. Si D significa depresión y + significa prueba positiva, entonces: a) *P(D)*=0,005; *P(D/+)*=0,99; *P(S/+)*=0,03 b) *P(D)*=0,0005; *P(+/D)*=0,99; *P(-/D)*=0,01 c) *P(D)*=0,005; *P(D/+)*=0,99; *P(-/D)*=0,01 d) *P(D)*=0,005; *P(+/D)*=0,99; *P(-/D)*=0,01	3
A partir de los datos de la tabla anterior lo más razonable es aceptar la siguiente hipótesis sobre la proporción de niños de preescolar que acepta sus equivocaciones: H: p<0,3 H: p > 0,55 H: p> 0,35 H: p >0,45	3

Se analizaron también los comentarios y sugerencias recibida sobre interés del material, completitud, redacción y posibles errores, en forma cualitativa.

3.2.5. DISCUSIÓN DEL ESTUDIO 8.1

En este estudio se ha diseñado una experiencia realista de enseñanza de conceptos básicos de inferencia bayesiana para estudiantes de primer año de psicología. En concreto se preparó un material escrito y en versión virtual organizado en cuatro temas, que abarca el teorema de Bayes, la inferencia para una proporción (caso discreto y continuo) e inferencia para la media (caso continuo). El material se complementa con cuatro pruebas de autoevaluación y programas auxiliares de cálculo en Excel.

Todo ello fue sometido a revisión por un grupo de 10 profesores de estadística, con variada experiencia y provenientes de diversas áreas de conocimiento, que, por un lado, actuaron como muestra piloto en la prueba del material, en una experiencia de enseñanza a distancia.

Puesto que era la primera vez que estos profesores estudiaban el tema, el estudio sirvió para comprobar la legibilidad del material y su interés, así como su dificultad razonable para el tipo de alumnos a que va dirigido. Se constató que el material es comprensible, mediante los resultados muy satisfactorios de las pruebas de autoevaluación, así como por los comentarios al respecto:

- *Considero que el material es claro en las explicaciones y orienta al estudiante. Está bien escrito y el orden es adecuado.*
- *Quiero felicitarte por la elaboración de un sitio web tan atinado para este tema. He aprendido mucho sobre inferencia bayesiana.*
- *Me han gustado mucho los problemas y la manera de utilizar las tablas para resolverlo*
- *Lo primero es que debo decirles que me han dado trabajo... pero interesante, la felicito en esta propuesta de la inferencia bayesiana. En algunos puntos me he sentido un verdadero estudiante.*

La coincidencia de errores en algunos ítems sirvió para identificar aquellos que se precisaba revisar, los cuáles fueron mejorados en su redacción. Las sugerencias y comentarios enviados y supervisión del material por parte de este grupo, fue importante, al permitir detectar algunas erratas, que fueron corregidas, por ejemplo:

- *Errata en organización de datos 1.3, en el cálculo de verosimilitudes de maquina B aparece 0,052 y en la tabla 0,0052.*

- *En sección ejercicios resueltos, errata del valor 0,88 en vez de 0,78 del primer ejercicio.*
- *En el Tema 2 creo que hay un error en la formula P(0,5|datos) = K x P(0,5|datos) = ... en el punto 2.2 de la teoría.*
- *En las actividades hay una nota donde se señala que se puede descargar el programa Excel Bayes en material complementario, pero no hay ningún link con ese nombre.*
- *En el penúltimo párrafo de la sección 4.4, al final se dice, ... la distribución final de la media μ de la población es también una normal $N(\mu_0, \sigma_0)$. Debería ser: ... la distribución final de la media μ de la población es también una normal $N(\mu_f, \sigma_f)$.*

También nos ha permitido identificar partes del contenido que no estaban suficientemente claras: *Al leer el ejemplo 1 sobre la Narcolepsia se me presenta una duda respecto a la tabla donde se vacían los datos del problema. No logro ver de donde se obtienen los datos sobre "No sufren Narcolepsia", allí me indican un total general de 100.000, que no logro comprender de donde se obtiene.*

Por último también consideramos valiosos los comentarios y opiniones de estos profesores que nos sirven para tratar de hacer más o menos énfasis en determinadas partes del material: *En la teoría, creo que el tema de contraste bayesiano será el de más dificultad de enseñar a los alumnos, cuesta seguir el procedimiento y sobre todo al plantear la probabilidad final y sus consecuencias. Uno tarda algo en acostumbrarse a la terminología "prob iniciales", "prob finales", "verosimilitudes" y las primeras preguntas del examen son interesantes para repensar esa terminología.*

En resumen en este estudio hemos preparado un material original, contextualizado en el campo de la psicología y razonablemente satisfactorio para abordar la enseñanza de conceptos de inferencia bayesiana a los estudiantes de psicología, que ha sido revisado con ayuda de un grupo de profesores de estadística. En el estudio siguiente se probará que la enseñanza diseñada y el material preparado pueden producir un aprendizaje significativo de un número de conceptos y procedimientos bayesianos en nuestros estudiantes y permite alcanzar la mayoría de los objetivos fijados en la enseñanza.

3.3. ESTUDIO 8.2. EVALUACIÓN DEL APRENDIZAJE DE CONCEPTOS BAYESIANOS EN UNA EXPERIENCIA DIDÁCTICA

3.3.1. INTRODUCCIÓN

Una vez finalizado el diseño y revisión de la propuesta didáctica (estudio 8.1), se aborda en este nuevo estudio la evaluación de su adecuación para los estudiantes a los que va dirigida.

El objetivo general es, por tanto, avanzar en el análisis de la idoneidad del material y proceso didáctico planificado y descrito en el estudio 8.1 para conseguir los objetivos fijados en la investigación. Para ello se organiza una experiencia de enseñanza, que es llevada a cabo con 78 estudiantes de primer año de psicología en un total de 20 horas. Mediante unas pruebas de evaluación durante y al finalizar la enseñanza se hace un diagnóstico sobre el aprendizaje logrado. Todo ello con la finalidad de tomar una decisión sobre la viabilidad de introducir algunas ideas básicas de inferencia bayesiana a estos alumnos.

En las secciones que siguen se describen las características de la muestra participante y el método seguido en la observación de la enseñanza. Los conocimientos adquiridos por los estudiantes se evalúan mediante un cuestionario de opciones múltiples preparado como parte del material docente y que se analiza, asimismo en este capítulo. Finalizamos exponiendo nuestras conclusiones sobre las dificultades percibidas y los logros de los estudiantes, en cuanto al aprendizaje de métodos bayesianos elementales.

3.3.2. MÉTODO

El estudio se llevó a cabo durante el mes de Mayo del curso académico 2004-2005. Para ello se organizó un seminario de carácter voluntario y se invitó a participar a los alumnos de Primer Curso de Psicología que cursaban la asignatura de análisis de datos. Se motivó la participación ofreciendo a los alumnos participantes que asistiesen con constancia y mostrasen un aprendizaje significativo.

Se trata, en consecuencia de un experimento, puesto que el experimentador produce un evento, puede describir sus condiciones para que otros experimentadores puedan reproducirla (Losada y López Feal, 2003); más concretamente un cuasi-experimento, pues no hay asignación aleatoria (Cook y Campbell, 1979).

Los participantes asistieron a cuatro sesiones de enseñanza, de tres horas de duración (dividida cada una en dos partes, con un descanso de media hora entre ellas). La primera parte de cada sesión (dos horas) se dedicó a la presentación del tema, ejemplos y actividades resueltas por parte de la profesora (una hora y media), apoyada por el material escrito, que se dio a los alumnos por anticipado y que debían traer leído a la sesión, así como en diapositivas que resumían las principales ideas. Los alumnos podían, en cualquier momento, plantear preguntas o solicitar aclaraciones.

Seguidamente, se pidió a cada alumno que completase la prueba de auto evaluación (media hora), para lo cual podía consultar el material escrito, aunque no podían consultar a sus compañeros o a la profesora. El objetivo es hacer reflexionar a los estudiantes sobre lo aprendido, comprobar que eran capaces de encontrar la información necesaria en el material escrito y evaluar el grado en que se estaban logrando los objetivos de aprendizaje.

La segunda parte de cada sesión (una hora) se trabajaba en el laboratorio de informática, donde los alumnos resolvían las actividades prácticas en forma individual, haciendo uso de los programas Excel preparados al efecto. Las respuestas escritas de los alumnos a las actividades teóricas y prácticas fueron recogidas, para ser analizadas. Dichas actividades eran corregidas en la sesión siguiente, haciendo especial énfasis en las dificultades más comunes.

La autora de la Memoria actuó como profesora en dicho seminario, contando con la colaboración de una observadora, quien registraba los posibles incidentes, utilizando un diario de observación. El método de recogida de datos se trataría de una observación (Fox, 1981) puesto que los datos pueden tomarse directamente porque son accesibles al investigador.

3.3.2.1. SUJETOS

Participaron un total de 78 alumnos, todos ellos matriculados en la asignatura de análisis de datos y que colaboraron voluntariamente a la investigación. El total de la muestra se dividió en cuatro grupos, con objeto de poder trabajar en grupos pequeños y atender mejor las dudas durante las sesiones.

La muestra está compuesta por un 17,9% de chicos y un 82,1% de chicas, todos alumnos de primer curso de Psicología. 57 (73%) alumnos de los que participaron en la experiencia de aprendizaje, realizaron también el cuestionario RPC, obteniendo una puntuación media de 22,7 puntos (sobre un máximo de 34) y desviación típica 3,7. Los alumnos participantes obtuvieron una media de 4,83 en el examen final de la asignatura (convocatoria de Junio), con desviación típica de 2,07. De estos alumnos 62 completaron la evaluación final del aprendizaje.

3.3.2.2. MATERIAL

Para el desarrollo de la experiencia se utilizaron los materiales descritos en el estudio 8.1 y presentados en los anexos 7 a 9. Consta de:

- Material escrito para el alumno.
- Autoevaluaciones y ejercicios de cada tema, los cuales fueron resueltos individualmente por los alumnos, finalizada la exposición teórica.
- Programas auxiliares de cálculo en Excel.

3.3.2.3. INSTRUMENTOS

Como instrumentos de recogida de datos, se utilizaron los siguientes:

Diario de observación

Fue elaborado durante el desarrollo de la secuencia de enseñanza. El objetivo principal fue realizar una observación sistemática de la enseñanza, tal como fue llevada a cabo, lo que nos permitió luego, comparar entre la planificación de la enseñanza y la forma en que efectivamente se llevó a cabo, además de tomar nota de las dificultades planteadas por los alumnos, para poder interpretar mejor los datos recogidos por escrito de las actividades planteadas durante las sesiones. Se trató de la observación participante y estructurada de un experimento de campo (León y Montero, 2002).

No se usó guión estructurado de observación, pero al realizar las observaciones, se tuvieron en cuenta algunas de las recomendaciones dadas por Bisquerra (1989). Por ejemplo, se realizaron los registros en el campo e inmediatamente después de cada sesión, se reunieron la investigadora con la observadora, para revisar las notas de observación y agregar cualquier hecho que hubiera pasado inadvertido. Las notas de campo se tomaron lo más rápidamente que fue posible, con el fin de cubrir todo lo que sucedía en la clase.

La observación se centró en recolectar todas las preguntas y dudas planteadas por los alumnos con el fin de detectar los contenidos que presentaban mayores dificultades. También, se detalló la manera en que se fue desarrollando cada clase, para poder comparar luego con lo que se tenía previsto y de esta manera poder determinar qué elementos de significado se habían trabajado, cuáles se habían dejado de lado o cuáles presentaron más dificultades.

Cuestionario AIB

Se preparó un cuestionario sobre Aprendizaje de Inferencia Bayesiana (cuestionario AIB) que se presenta en la tabla 6.42 que sería utilizado, una vez terminada la enseñanza y completada individualmente por escrito, con la finalidad de informar del aprendizaje de los estudiantes y la adecuación del material y la enseñanza realizada para los fines de la investigación[39].

Nos situamos en un dominio curricular, que se refiere a un conjunto de habilidades y conocimientos que se desarrollan como resultado de la instrucción deliberada sobre un cierto contenido curricular (Millman y Greene, 1989). Es un instrumento de *medida centrada en el sujeto* y más concretamente un *test referido a criterio* (Sax, 1989). Además podemos clasificarlo como *un cuestionario de potencia*, ya que las diferencias en la puntuación entre los sujetos que respondan al cuestionario serán debidas a la calidad de su ejecución y conocimiento y un *test psicométrico* puesto que evalúa las respuestas en forma cuantitativa y se refiere a un rasgo diferenciado del sujeto (Martínez Arias, 1995).

Se comenzó el proceso de construcción de la prueba AIB mediante la definición *semántica* de la variable objeto de medición. A partir de los objetivos y contenidos fijados para la enseñanza (descritos en el estudio 8.1) y del análisis de contenido del material entregado a los alumnos, que se presenta en el anexo 7, de determinaron las

[39] El primer paso en el desarrollo de un instrumento es determinar su propósito y la naturaleza de las inferencias que vamos a extraer de las puntuaciones obtenidas a partir de su aplicación a una muestra de sujetos (Sax, 1989).

unidades de contenido objeto de la evaluación, que se presentan en la tabla 6.41, junto con los ítems del cuestionario AIB que evalúan cada uno de estos contenidos.

Para cada uno de estos contenidos se elaboraron ítems de nivel de aplicación y análisis en la taxonomía Bloom (1956). Los ítems son de elaboración propia (puesto que no hemos encontrado cuestionarios adaptados a nuestro trabajo) y en su construcción se siguieron los criterios de calidad habituales (Osterlind, 1989):

- Alto grado de congruencia entre el ítem particular y el contenido que se trata de evaluar.
- Pertinencia del formato del ítem a los fines del cuestionario. Dado que nos encontramos ante un test de ejecución óptima (cuestionario de rendimiento) hemos considerado adecuado realizar una parte de los ítems en un formato de elección múltiple, por tanto de respuesta construida, donde se proporciona al examinado la respuesta correcta, junto con otras alternativas y se pide al examinado elegir una de las opciones (Osterlind; 1989). Se eligió un número de alternativas de 4, todas ellas plausibles (Martínez Arias, 1995).
- Además, nuestros ítems cumplen los criterios de independencia local (la repuesta a uno no depende de la respuesta a otro), unidimensionalidad (pues todos se refieren al constructo aprendizaje de inferencia bayesiana), y satisfacen las cuestiones éticas y legales.

Al considerar el número total de ítems, se trató de cubrir adecuadamente el contenido, asegurar una fiabilidad satisfactoria y restricción del tiempo disponible para completarlo (Millan y Greene, 1989). El cuestionario consta de 20 ítems (algunas de ellas con varios subítems) y se planificó para que pudiera completarse en una hora[40]. Se llevó a cabo varias revisiones del cuestionario, con el fin de mejorar la legibilidad.

[40] Por razones de eficiencia el cuestionario fue administrado de forma grupal; por ello hubo de incluirse instrucciones claras y precisas sobre su cumplimentación.

Tabla 6.41. Contenidos evaluados en el cuestionario AIB e ítems que los evalúan

	Contenidos evaluados en el cuestionario AIB	Ítem
Relacionados con la inferencia bayesiana	Diferenciar entre probabilidades iniciales y finales y verosimilitudes.	1
	Identificar los sucesos de interés, sus probabilidades iniciales y verosimilitudes a partir del enunciado de un problema.	1
	Analizar el teorema de Bayes como herramienta para transformar probabilidades iniciales en finales.	7
	Organizar los datos en una tabla de forma que se facilite el cálculo de probabilidades finales a partir de las iniciales y verosimilitudes.	18
	Visualizar los parámetros como variables aleatorias con una distribución de probabilidad asociada.	3, 5
	Identificar valores de la proporción poblacional y asignarles probabilidades iniciales plausibles (distribución de probabilidad inicial).	6
	Comprender el concepto de distribución inicial uniforme y distribución informativa.	4
	Asignar correctamente una distribución inicial no informativa a la proporción.	6
	Asignar correctamente una distribución inicial informativa a la proporción.	8
	Identificar los valores de éxitos y fracasos en la muestra.	8
	Aprender a introducir los datos anteriores en la hoja Excel para obtener los valores de verosimilitud y la distribución final.	18
	Calcular probabilidades asociadas a determinados valores de proporción en la distribución final.	7,10,11
	Usar la distribución Beta para definir una distribución inicial para la proporción en caso informativo.	20
	Identificar los éxitos y fracasos en la muestra.	9,18
	Utilizar la herramienta Excel para obtener una tabla con la distribución de probabilidad final.	18
	Calcular intervalos de credibilidad..	10
	Interpretar el intervalo de credibilidad de una proporción	10, 12
	Contraste de hipótesis sencillas sobre una proporción siguiendo la metodología bayesiana.	11
	Diferenciar entre estadístico y parámetro en el contexto de un problema.	19
	Con los datos obtenidos en una muestra, obtener una distribución final.	13,14
	Extraer inferencias sobre la población a partir de la distribución final.	43
	Cálculo de intervalos de credibilidad para la media de una población.	16
	Interpretar un intervalo de credibilidad de la media.	16
	Realizar correctamente un contraste de hipótesis sobre la media, desde una perspectiva bayesiana.	14
Otras	Diferenciar una probabilidad condicional y su transpuesta.	2
	Diferenciar probabilidades condicionales y conjuntas en el contexto de un problema.	2

Tabla 6.42. Prueba AIB

1. En una Facultad 10 de cada 100 personas estudia el último año. 60 de cada 100 estudiantes del último año comparte un piso con otros compañeros y también 30 de cada cien estudiantes de los cursos anteriores. *C* es el suceso compartir piso y *U* el suceso estar en el último año. Cogemos al azar un estudiante y resulta ser de último año, la probabilidad de que comparta un piso es:

a. Una probabilidad inicial *P (C)*
b. Una probabilidad final *P(C/U)*
c. Una verosimilitud *P(U/C)*
d. Una probabilidad conjunta $P\ (U \cap C)$

2. Imagina que tienes una muestra de 100.000 personas elegidas al azar. Imagina que 10 de cada 1000 está deprimida. Supón que una prueba de depresión da positivo en 99 de cada 100 personas enfermas y también en 2 de cada 100 personas sanas. Si *D* significa depresión y + significa prueba positiva, entonces, calcula las siguientes probabilidades:

a) *P(D)=* b) *P(+/D)=* c) *P(-/D)=* d) $P(D \cap +)=$

3. El valor medio μ de una variable (por ejemplo la altura) en una población:
a. Se considera constante en inferencia bayesiana
b. Se considera variable en inferencia clásica
c. Se considera variable en inferencia bayesiana
d. En unas poblaciones es constante y en otras es variable

4. La distribución inicial de probabilidades de un parámetro:
a. Contiene toda la información de la persona sobre la población antes de recoger los datos
b. Se calcula mediante el teorema de Bayes, a partir de la distribución final
c. Sirve para calcular el intervalo de credibilidad del parámetro
d. Es una distribución uniforme

5. En un estudio sobre costumbres de los jóvenes españoles se hizo una encuesta a una muestra representativa de 1000 jóvenes. El tiempo medio que dedicaban al deporte fue 3 horas semanales.
a. La media de 3 horas es un parámetro de la población de jóvenes españoles
b. La media en la población es una variable aleatoria, los valores más probables de la misma serían alrededor de 3 horas.
c. La media en la población es constante, pero no la conocemos.
d. Cada joven español emplea 3 horas semanales en deporte.

6. Una fábrica de bombillas las vende en cajas de cuatro en cuatro. No sabemos nada sobre la proporción de bombillas defectuosas entre todas las fabricadas. ¿Cuál de las distribuciones A, B. C o D describe mejor la distribución inicial de la proporción de bombillas defectuosas en una caja?

(A)		(B)		C)		(D)	
Valores prop.	Probabilidad	Valores Prop.	Probabilidad	Valores Prop.	Probabilidad	Valores prop	Probabilidad
0,00	0,1	0,00	0,2	0,00	0,00	0,00	¼
0,25	0,1	0,25	0,2	0,01	0,25	0,25	¼
0,50	0,1	0,50	0,2	0,02	0,50	0,50	¼
0,75	0,1	075	0,2	0.03	0,75	075	¼
1	0,1	1	0,2	0.04	1	1	¼

7.Al comenzar a resolver un problema de estimación de una proporción un alumno ha completado las tres primeras columnas de la tabla Bayes, obteniendo estos datos

Posibles valores de p	Probabilidad inicial	Verosimilitud	--------	---------
0,0000	0,0000	0,0000		
0,1000	0,1000	0,0000		
0,2000	0,1000	0,0233		
0,3000	0,1000	0,1239		
0,4000	0,1000	0,0682		
0,5000	0,1000	0,0065		
0,6000	0,1000	0,0001		
0,7000	0,1000	0,0000		
0,8000	0,1000	0,0000		
0,9000	0,1000	0,0000		
1,0000	0,1000	0,0000		
Suma			0,0222	

La probabilidad final de que el verdadero valor de la proporción sea igual a 0,4 es:
a) 0,00682 b) 0,1000 c) 0,3072 d) 0,00015

8. Un estudio médico indica una incidencia de 15% de adicción al tabaco en las mujeres jóvenes. Una posible distribución aproximada inicial para definir esta población sería:
a) *Be* (15, 100) b) *Be* (15, 85) c) *Be* (85,15) d) *Be* (100, 15)

9. La media de la distribución *Be(a,b)* es igual a:
a) a/b b) $(a+1)/(a+b)$ c) $(a+1)/(b+1)$ d) $a/(a+b)$

10. La siguiente tabla presenta las probabilidades y valores críticos de la distribución *Be(30,40)*.

Probabilidades			Valores críticos	
p0	$P(0<p<p0)$	$P(p0<p<1)$	Probabilidad	Valor crítico
0	0,000	1,000	0,000	0,000
0,05	0,000	1,000	0,005	0,296
0,1	0,000	1,000	0,010	0,304
0,15	0,000	1,000	0,015	0,311
0,2	0,000	1,000	0,020	0,316
0,25	0,001	0,999	0,025	0,320
0,3	0,012	0,988	0,030	0,324
0,35	0,090	0,910	0,035	0,327
0,4	0,318	0,682	0,040	0,330
0,45	0,645	0,355	0,045	0,330
0,5	0,886	0,114	0,050	0,333
0,55	0,979	0,021	0,950	0,526
0,6	0,998	0,002	0,955	0,529
0,65	1,000	0,000	0,960	0,533
0,7	1,000	0,000	0,965	0,536
0,75	1,000	0,000	0,970	0,541
0,8	1,000	0,000	0,975	0,545
0,85	1,000	0,000	0,980	0,551
0,9	1,000	0,000	0,985	0,558
0,95	1,000	0,000	0,990	0,567
1	1,000	0,000	1,000	1,000

El intervalo de credibilidad del 98% para la proporción de una población descrita por una distribución final *Be* (30, 40) es aproximadamente:

a. ($0,316 < p < 0,551$)
b. ($0,304 < p < 0,567$)
c. ($0,3 < p < 0,6$)
d. ($0,1 < p < 0,9$)

11, La distribución final de la proporción de votantes favorables a un partido viene dado por la distribución *Be* (30,40). A partir de los datos de la tabla anterior lo más razonable es aceptar la siguiente hipótesis sobre la proporción en la población:

a) H: $p<0,25$ b) H: $p > 0,55$ c) H: $p> 0,25$ d) H: $p >0,45$

12, Para una misma distribución final del parámetro en la población, el intervalo de credibilidad del **r%** para el parámetro es:

a. Más ancho si aumento el valor de *r*
b. Más ancho si aumento el tamaño de muestra
c. Más estrecho si aumento el valor de *r*
d. Depende de la distribución inicial

13. En una población normal con desviación típica σ=5 donde no tengo información inicial de la media poblacional, sacamos una muestra de 25 elementos y obtenemos una media $\bar{x}$ =100. La distribución final de la media en la población es:

a. Una distribución normal *N* (100, 0,5)
b. Una distribución normal *N* (0,1)
c. Una distribución normal *N* (100,5)
d. Una distribución normal *N* (100, 1)

14.Para contrastar la hipótesis de que la media μ en una población normal con desviación típica σ=1 es mayor que 5, mediante los datos obtenidos en una muestra de 100 elementos, siguiendo el procedimiento bayesiano:

a) Calculo la media de la muestra $\bar{x}$; calculo $P(\frac{\bar{x}-5}{0,1} < 5)$; si esta probabilidad es muy pequeña, acepto la hipótesis

b) Calculo la media de la muestra $\bar{x}$; calculo $P(\frac{\bar{x}-5}{0,1} < Z)$; siendo *Z* la distribución normal *N* (0,1); si esta probabilidad es muy pequeña acepto la hipótesis

c) Calculo la media de la muestra $\bar{x}$; calculo $P(\frac{\bar{x}-5}{0,1} > Z)$ siendo *Z* la distribución normal *N* (0,1); si esta probabilidad es muy pequeña, acepto la hipótesis

d) Calculo la media de la muestra $\bar{x}$; calculo $P(\frac{\bar{x}-5}{0,1} > 5)$ si esta probabilidad es muy pequeña, acepto la hipótesis

15. En una muestra de 100 elementos de una población normal he obtenido una media igual a 50. Suponiendo una distribución inicial uniforme de la media en la población la distribución final de la media es:

a. Aproximadamente *N (50, s),* siendo s la desviación típica de la muestra
b. Aproximadamente *N (50, s/10),* siendo s la desviación típica de la muestra
c. No puedo decir nada, porque no conozco la desviación típica de la población
d. Aproximadamente *N (0,1)*

16. La distribución final de la media en una población es *N* (100, 15). También sabemos que el valor *P* (-1,96 < *Z* < 1,96) =0,95, siendo *Z* la distribución normal *N* (0,1). El intervalo de credibilidad del 95% para la media en la población es:

a. (100-1,96 *x* 1,5; 100 + 1,96 *x*1,5)
b. (100-1,96; 100+1,96)
c. (100 *x* 1,5 –1,96; 100 *x* 1,5 + 1,96)
d. (100-1,96 x 15; 100 +1,96 *x* 15)

17,Tenemos los siguientes datos de un estudio sobre la altura media de las chicas españolas obtenido de datos de una muestra de 400 chicas

	Media	Desviación típica
En la muestra	160	10
En la distribución inicial	156	13
En la distribución final	158,5	7,9

Para hallar el intervalo de credibilidad de la media en la población utilizaríamos:

a. La distribución normal N(160,10)
b. La distribución normal N (156,13)
c. La distribución normal N (158,5; 7,9)

La distribución normal N(160, 0,5)

18. En un centro de preescolar el 20 % de los niños son hijos de inmigrantes y el 10% de las niñas. En el centro hay un 50% de niños y otro 50% de niñas. Utiliza la tabla siguiente para calcular la probabilidad de que si cogemos al azar un alumno inmigrante, sea varón.

Sucesos	Probabilidades iniciales	Verosimilitudes	Producto	Probabilidades finales
Suma	1			1

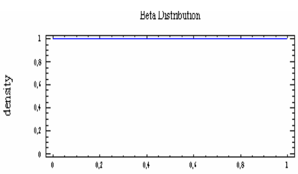
19. En un centro geriátrico se quiere estimar la proporción de residentes con deterioro cognitivo, a partir de una muestra de 10 residentes 2 mostraron deterioro cognitivo. La versomilitud para el valor del parámetro p=0,1 es de 0,1937. ¿Qué significa este valor?

a. P (datos), es decir, probabilidad de obtener esta muestra.
b. P (datos $\cap$ p=0,1), es decir, probabilidad de obtener esta muestra y además que la proporción en la población sea 0,1.
c. P (p=0,1/datos), es decir, probabilidad de que conocida la muestra, el valor del parámetro sea 0,1.
d. P (datos / p=0,1), es decir, supuesto que el valor del parámetro sea 0,1, probabilidad de obtener esa muestra.

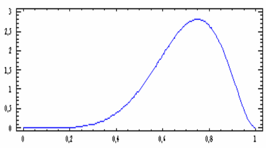
20. Observa las siguientes curvas Beta e indica cuál de ellas elegirías como distribución inicial de la proporción en las siguientes situaciones:

Estimación de la proporción de chinos que saben hablar otro idioma ____________
Estimación de la proporción de niños varones que nace en un cierto país _________
Estimación de la proporción de alumnos de la Facultad con sobrepreso___________

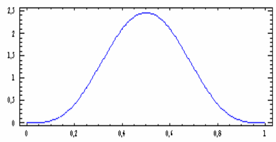
a= 1 , b=1



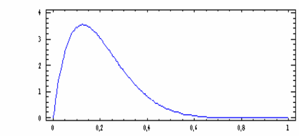
a=7 , b=3



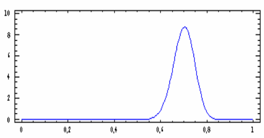
a=5 , b=5



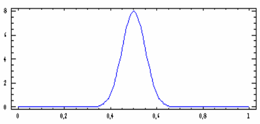
a=2 , b=8



a=70 , b=30



a=50 , b=50



3.3.3. ANÁLISIS

Se realizó un estudio descriptivo de los resultados de cada una de las pruebas de autoevaluación y del cuestionario AIB obteniendo la proporción de respuestas correctas, proporcionando estimaciones clásica y bayesiana de sus índices de dificultad.

Se realizó un análisis de contenido[41] (Weber, 1985; Ghiglione y Matalón, 1989) de los protocolos escritos producidos por los alumnos participantes en las tareas de resolución de problemas, dividiendo cada problema en fases necesarias en su resolución e identificando para cada estudiante la resolución correcta o incorrecta de cada una de

[41] En el análisis de contenido se recurren a la lógica y el conocimiento previo sobre el tema, para resumir el contenido del texto, definir categorías y verificar la validez de las mismas. (Ghiglione y Matalón, 1989).

las fases. A partir de ello se definieron una serie de variables estadísticas sobre las cuales se llevó a cabo un estudio descriptivo similar al realizado para las pruebas de autoevaluación y AIB. Para la estimación clásica en todos estos casos, se calcularon los intervalos de confianza para la proporción, utilizando la aproximación normal, puesto que el tamaño de la muestra era suficientemente elevado.

La estimación bayesiana se realizó con distribución inicial no informativa, tomando, en consecuencia una distribución inicial *Be*(0,5, 0,5) como recomiendan entre otros Lecoutre (1996) y Serrano (2003). La distribución final viene dada por la distribución *Be*(0,5+*p,* 0,5+*q*), siendo *p* y *q* el número de respuestas correctas e incorrectas en el ítem[42]. El intervalo se calcula mediante la siguiente expresión,

$$\left[\beta^{-1}_{0{,}5+\mathrm{p},\,0{,}5+\mathrm{q}}(\alpha/2)-\beta^{-1}_{0{,}5+\mathrm{p},\,0{,}5+\mathrm{q}}(1-\alpha/2)\right]$$

donde $\beta^{-1}_{\mathrm{a,b}}$ es la función de distribución inversa de la distribución *Be(a,b)* y α el coeficiente de credibilidad. Los cálculos fueron programados en Excel.

Tanto para el conjunto de pruebas de evaluación (consideradas globalmente como un único instrumento), como para el cuestionario AIB se analizó la fiabilidad desde la perspectiva de consistencia interna fundamentada en la teoría clásica de tests (Carmines y Zeller, 1979; López Feal, 1986; Morales, 1988; Múñiz, 1992; Martínez Arias, 1995). Los cálculos del coeficiente Alfa e llevaron a cabo mediante el programa SPSS.

Para la estimación bayesiana de la media de la puntuación en la evaluación final, se tomó una distribución inicial no informativa (uniforme), puesto que era la primera vez que utilizamos la prueba AIB y no teníamos información al respecto. Puesto que no se conoce la desviación típica de la población σ, se estimó mediante la desviación típica de la muestra *s* (raíz cuadrada de la cuasi-varianza). Puesto que el tamaño de la muestra era mayor de 30 elementos la distribución final de la media muestral es $\mathrm{N}(\bar{x}, s/\sqrt{n})$, siendo *n* el tamaño de la muestra.

[42] La ventaja de utilizar una distribución inicial Beta para la estimación de la proporción, es que esta distribución es reproductiva respecto a sus parámetros.

3.3.4. RESULTADOS

3.3.4.1. OBSERVACIÓN DE LAS SESIONES

Sesiones teóricas

La observación permitió detectar las principales dificultades observadas, que en las sesiones teóricas fueron las siguientes:

1. En unos pocos casos algunos alumnos faltaron a una de las sesiones. Aunque se les pidió que leyesen el material de la sesión anterior, en estos casos hubo que dedicar un tiempo complementario, fuera de las sesiones para atender a las dudas planteadas.
2. Al estar finalizando el curso, los alumnos se encontraban próximos a los exámenes. No obstante, el seguimiento de las sesiones fue muy constante y el interés por realizar el trabajo pedido se mantuvo.
3. Se presentaron algunas dificultades al diferenciar una probabilidad condicional y su inversa en los enunciados verbales de los problemas (no cuando se expresaba mediante una fórmula). Ello confirma la opinión de Falk (1986) y Pollatsek, Well, Konold y Hardiman (1987) de que la ambigüedad en el lenguaje es uno de los obstáculos de los estudiantes para diferenciar estas dos probabilidades.
4. Hubo que recordar algunos conceptos previos, como los de parámetro y estimador, que no eran objetos específicos de enseñanza en la experiencia y se suponían aprendidos por los alumnos.
5. Algunos alumnos plantearon la duda de si era válido el uso de probabilidades subjetivas, dentro de un contexto científico. Hubo que plantear una discusión sobre la objetividad /subjetividad en la ciencia, que siempre es relativa. Asimismo se resaltó el hecho de que al aumentar el tamaño de la muestra progresivamente los datos iniciales subjetivos van perdiendo su importancia.
6. Surgieron algunos problemas en diferenciar el intervalo de confianza y el de credibilidad, debido a que ellos habían dado siempre una interpretación bayesiana (incorrecta) al intervalo de confianza, por lo que no veían la diferencia entre los dos. Se insistió bastante en este tema con objeto de que se hicieran consciente que daban una interpretación incorrecta a los resultados de inferencia clásica.

Sesiones prácticas

Las principales dificultades observadas en las sesiones prácticas fueron las siguientes:

1. El diferente ritmo de trabajo de los alumnos hace que, en ocasiones sea complicado mantener el ritmo de la clase, puesto que algunos tienen más facilidad que otros para el manejo de la herramienta informática.
2. En algunos ordenadores se produjo un error, consistente en indicar que la aplicación Excel no estaba instalada (cuando si lo estaba). Esto causó confusión a los estudiantes que usaban estos ordenadores y requirió un tiempo adicional para solventar el problema.
3. En el programa PRODIST se produjo un desplazamiento de la escala del eje de abcisas, que pasó desapercibida para la profesora. Una de las alumnas notó este desplazamiento y avisó a la profesora, quien corrigió el problema y alertó al resto de los estudiantes, con objeto de que no cometiesen un error en la interpretación de la gráfica.
4. Los alumnos encontraron sorprendentes las soluciones de algunos de los problemas, por ejemplo, el hecho de que la probabilidad de que una alarma fuese infundada sea alta. Pero, ello sirvió para interesarlos más en resolver los otros ejercicios del tema.

3.3.4.2. AUTOEVALUACIONES DE CONOCIMIENTOS TEÓRICOS

Como complemento de la observación de clases, se completó una autoevaluación para cada una de las sesiones teóricas, las cuales fueron resueltas individualmente por los alumnos, finalizada la exposición teórica. Como hemos indicado el objetivo principal era comprobar si el alumno había comprendido la exposición teórica y era capaz de identificar las respuestas a las cuestiones planteadas por si mismo, con ayuda de dicho material.

En la tabla 6.43 presentamos los resultados del análisis de fiabilidad para el total de las autoevaluaciones, donde se obtuvo un valor para el coeficiente *Alfa* = 0,776, con n=78 sujetos y un total de 30 ítems. Consideramos que el valor es razonablemente alto, dado el número de sujetos participantes. La discriminación de los ítems (medida por la correlación ítem-elemento total corregida) fue en general aceptable, superando el valor

0,20 considerado por García Cueto (1992) y Barbero (1993) los mínimos aceptables. Algunos ítems (por ejemplo i1_1) resultaron muy sencillos para los estudiantes y, aunque esto supone que se consiguieron los objetivos de la enseñanza, al mismo tiempo producen una discriminación baja.

Tabla 6.43. Resultados del análisis de fiabilidad del conjunto de autoevaluaciones

	Media si se elimina el elemento	Varianza si se elimina el elemento	Correlación elemento-total corregida	Alfa si se elimina el elemento
i1_1	19,88	19,298	0,059	0,780
i1_2	20,05	19,634	0,254	0,788
i1_3	19,78	18,952	0,289	0,771
i1_4	20,36	20,181	0,280	0,796
i1_5	19,90	18,821	0,198	0,774
i1_6	19,86	18,980	0,174	0,774
i1_7	20,49	19,344	0,229	0,782
i1_8	19,78	19,393	0,083	0,777
i2_1	20,19	18,391	0,230	0,773
i2_2	19,82	17,889	0,639	0,757
i2_3	19,77	18,673	0,475	0,766
i2_4	20,09	17,875	0,370	0,765
i2_5	19,86	18,201	0,438	0,763
i2_6	19,86	17,837	0,566	0,758
i2_7	19,81	18,651	0,362	0,767
i2_8	19,78	18,692	0,413	0,767
i3_1	19,78	18,277	0,615	0,761
i3_2	19,83	17,881	0,607	0,757
i3_3	19,86	18,253	0,420	0,764
i3_4	20,19	17,768	0,381	0,764
i3_5	19,81	18,157	0,568	0,760
i3_6	20,17	18,089	0,304	0,769
i3_7	20,32	18,325	0,253	0,772
i4_1	20,18	18,435	0,220	0,774
i4_2	19,99	17,623	0,483	0,759
i4_3	20,18	18,097	0,301	0,769
i4_4	20,54	19,265	0,265	0,780
i4_5	20,26	18,037	0,316	0,768
i4_6	20,27	18,251	0,265	0,771
i4_7	20,17	17,387	0,478	0,758

En la tabla 6.44 presentamos la estimación del índice de dificultad, medido como porcentaje de respuestas correctas en cada uno de los ítems del **primer tema**, junto con los intervalos de confianza y credibilidad del 95%, calculados estos últimos con distribución inicial no informativa.

Tabla 6.44. Porcentaje de respuestas correctas a la primera autoevaluación, intervalos de confianza y de credibilidad (n=78)

Ítem	% Respuestas correctas	Intervalo confianza 95%		Intervalo credibilidad 95%	
		Lim inf	Lim inf	Lim sup	Lim sup
1	83,3	0,750	0,916	0,739	0,903
2	66,7	0,562	0,772	0,583	0,788
3	93,6	0,882	0,990	0,865	0,975
4	35,9	0,253	0,465	0,259	0,468
5	82,1	0,736	0,906	0,687	0,876
6	85,9	0,782	0,936	0,823	0,956
7	23,1	0,137	0,325	0,148	0,332
8	93,6	0,882	0,990	0,845	0,945

Los resultados nos indican que la mayoría de los alumnos (83,3% en ítem 1 y 93’6% en ítem 3) han comprendido la finalidad principal del Teorema de Bayes, así como la finalidad y diferencias entre distribución inicial y final. Un 66,7% (ítem 2) comprende el concepto de verosimilitud y un 82,1%, (ítem 5) identifica la probabilidad final, y la diferencia respecto de la verosimilitud, probabilidad inicial y probabilidad conjunta, dentro de un contexto práctico.

El 85,9% (ítem 6) ha aprendido el uso de la tabla de Bayes para realizar los cálculos. Resaltamos la mayor facilidad de comprensión del teorema de Bayes dado mediante esta formulación, en comparación con los resultados obtenidos en el cuestionario de evaluación. El 93,6% (ítem 3) han sido capaces de identificar diferentes formulaciones alternativas de la fórmula de Bayes.

En cuanto a las dificultades encontradas, el ítem 4 y el 7 resultaron difíciles. En el ítem 4, pretendíamos analizar si los alumnos creen que el valor de una probabilidad final y una verosimilitud pueden ser similares o no, pero también interfiere el concepto de una probabilidad condicionada y su transpuesta.

El ítem 7 muestra que solo un 23,1% son capaces de identificar correctamente los datos (probabilidad inicial y verosimilitudes) en un enunciado. Esto se puede deber a que los alumnos buscaran en la alternativa datos que estuvieran en el enunciado. La respuesta correcta incluye un dato que ellos tienen que elaborar (la probabilidad de un suceso complementario). En realidad, la respuesta más elegida (la alternativa a) con un 60,3%) es la única respuesta con los números del enunciado. Por otro lado, esta respuesta indica, de nuevo, la confusión entre una probabilidad condicional y su transpuesta, cuando los datos se dan en forma verbal.

En la tabla 6.45 presentamos la estimación del índice de dificultad, medido como porcentaje de respuestas correctas en cada uno de los ítems del tema 2, junto con los intervalos de confianza y credibilidad del 95%, calculados estos últimos con distribución inicial no informativa.

Tabla 6.45. Porcentaje de respuestas correctas a la segunda autoevaluación, intervalos de confianza y de credibilidad (*n*=78)

Ítem	% Respuestas Correctas	Intervalo confianza 95%		Intervalo credibilidad 95%	
		Lim inf	Lim inf	Lim sup	Lim sup
1	52,6	0,415	0,637	0,416	0,632
2	89,7	0,830	0,964	0,810	0,946
3	94,9	0,900	0,998	0,875	0,979
4	62,8	0,521	0,735	0,517	0,726
5	85,9	0,782	0,936	0,764	0,918
6	85,9	0,782	0,936	0,764	0,918
7	91,0	0,846	0,974	0,826	0,955
8	93,6	0,882	0,990	0,858	0,971

Las diferencias entre los dos tipos de intervalo se deben, por un lado, a que para el cálculo de intervalos de confianza usamos la aproximación normal a la distribución muestral de la proporción, puesto que el tamaño de muestra (*n*=78) es suficientemente elevado, mientras que para los intervalos de credibilidad utilizamos la distribución final exacta, que es una distribución Beta. Además, puesto que tomamos como distribución inicial no informativa la *Be*(0,5, 0,5), el número total de éxitos y fracasos en la muestra viene incrementado por media unidad, respecto a la estimación clásica, lo que produce una ligera diferencia en la estimación del intervalo. Por otro lado, mientras que la distribución normal es siempre simétrica, la distribución Beta tiene más o menos asimetría, dependiendo de sus parámetros. Finalmente, hacemos notar que el intervalo de credibilidad nos da un rango de valores posibles para la proporción de éxitos en la población no así el intervalo de confianza.

La **segunda unidad didáctica** ha resultado sencilla a los alumnos, quienes en su mayoría diferencian una distribución inicial de la final (ítem 3, 94,9%; ítem 5, 85,9%), son capaces de asignar distribuciones iniciales no informativas (ítem 6, 85,9%) e informativas (ítem 7, 91%), reconocen y saben aplicar la fórmula de Bayes para estimación de la proporción en una población (ítem 4, 62,8%; ítem 8, 93,6%) y comprenden la diferente consideración del parámetro en inferencia clásica y bayesiana (ítem 2, 89,7%). El único problema presentado es que parte de los alumnos no recordaba el significado del término parámetro (ítem 1, 52,6%).

En la tabla 6.46 mostramos los resultados obtenidos en la **tercera sesión**, donde pocos alumnos mostraron dificultad al pasar del caso discreto al continuo en el estudio de la estimación de la proporción.

Tabla 6.46. Porcentaje de respuestas correctas a la tercera autoevaluación, intervalos de confianza y de credibilidad (n=78)

Ítem	% Respuestas correctas	Intervalo confianza 95%		Intervalo credibilidad 95%	
		Lim inf	Lim inf	Lim sup	Lim sup
1	93,6	0,882	0,990	0,858	0,971
2	88,5	0,814	0,956	0,795	0,937
3	85,9	0,782	0,936	0,764	0,918
4	52,6	0,415	0,637	0,416	0,632
5	91,0	0,846	0,974	0,826	0,955
6	55,1	0,441	0,661	0,441	0,656
7	39,7	0,288	0,506	0,296	0,508

La mayoría de los alumnos son capaces de asignar una distribución inicial razonable, tanto informativa (ítem 3, 85,9%) como no informativa (ítem 1, 93,6%) y manejan con facilidad las tablas de probabilidad y valores críticos de la distribución Beta (ítem 5, 91%; ítem 6, 38,5%).

Igualmente, (88,5% en ítem 2; 52,6% en ítem 4) comprenden el concepto de mejor estimador y su cálculo en el caso de estimación de una proporción a partir de los parámetros de la distribución Beta. Ello presupone la correcta interpretación de los parámetros de dicha distribución como número de éxitos y fracasos en una muestra y su aplicación en un contexto problemático.

Los principales errores que se producen guardan relación con el razonamiento por contradicción, la lógica global del contraste de hipótesis y el significado del coeficiente de credibilidad. Estos errores son comunes a los descritos en inferencia bayesiana, aunque la incidencia en nuestra muestra no fue tan grande como la descrita para ella (15,4%, ítem 6). Hay que tener en cuenta el poco tiempo de enseñanza, así como la desaparición de otros errores, como son los de interpretación de los riesgos de error.

Finalmente el 34,6% (ítem 4) no tiene en cuenta la información previa, dando un estimador, basándose sólo en los datos de la muestra, es decir, usando una distribución a priori no informativa.

En la **sesión cuarta** (tabla 6.47) los alumnos han adquirido un conocimiento sobre la estimación de la media en una distribución normal, sabiendo elegir una distribución inicial no informativa adecuada (ítem 1, 53,8%) y calcular, en este caso, la distribución final resultante (ítem 2, 73,1%). La mayoría aplica correctamente, y usan la distribución final adecuada, el procedimiento de contraste y cálculo de intervalos de credibilidad a partir de situaciones prácticas (ítem 5, 46,2 %; ítem 6, 28,2%) y contraste de hipótesis (ítem 3, 53,8%). Conocen el programa de cálculo y lo usan correctamente para sacar conclusiones sobre problemas (ítem 7, 55,1%).

Tabla 6.47 Porcentaje de respuestas correctas a la cuarta autoevaluación, intervalos de confianza y de credibilidad (n=78)

Ítem	% Respuestas correctas	Intervalo confianza 95%		Intervalo credibilidad 95%	
		Lim inf	Lim inf	Lim sup	Lim sup
1	53,8	0,427	0,649	0,499	0,726
2	73,1	0,633	0,829	0,623	0,816
3	53,8	0,427	0,649	0,428	0,644
4	64,1	0,535	0,747	0,530	0,738
5	46,2	0,351	0,573	0,355	0,571
6	28,2	0,182	0,382	0,194	0,390
7	55,1	0,441	0,661	0,441	0,656

Se han observado en una proporción de los alumnos dificultades comunes con la inferencia clásica, como son la comprensión de la lógica del contraste, recordar la fórmula de la desviación típica del estimador. Hay algunos errores de aplicación de fórmulas, por ejemplo, al despejar incógnitas en desigualdades. Ninguno de estos problemas está directamente relacionado con la enseñanza de inferencia bayesiana sino que son fallos en otros conceptos que el alumno debiera tener aprendidos previamente.

3.3.4.3. RESOLUCIÓN DE PROBLEMAS

Durante cada una de las sesiones los alumnos resolvieron, adicionalmente, los problemas que se describieron en el estudio 8.1, y que se trataba de pruebas de ensayo (Fox, 1981; Osterlind, 1989) utilizando para ello los programas Excel que proporcionamos como apoyo del material escrito y que fueron descritos en dicho estudio. Los alumnos trabajaron en el laboratorio de informática y entregaron sus soluciones, junto con sus explicaciones por escrito a la profesora. A continuación analizamos las soluciones aportadas por estos alumnos a las tareas planteadas.

En la **primera sesión** se propusieron dos ejercicios de aplicación del Teorema de Bayes, que el alumno debía resolver utilizando el Programa Bayes. El primer apartado del primer ítem pide organizar los datos en una tabla Bayes, para lo cual el alumno ha de deducir correctamente las probabilidades iniciales y verosimilitudes del enunciado verbal del problema, e introducirlas en el programa. En el segundo apartado del ejercicio se pide al estudiante que, usando el programa deduzca las probabilidades a posteriori de los sucesos.

Tabla 6.48. Porcentaje de respuestas correctas en los ejercicios de la primera sesión, intervalos de confianza y de credibilidad (n=78)

			Intervalo confianza 95%		Intervalo credibilidad 95%	
		% Respuestas correctas	Lim inf	Lim inf	Lim sup	Lim sup
Ej. 1	Calcula correctamente probabilidades iniciales y verosimilitudes	85,9	0,782	0,936	0,765	0,919
	Calcula correctamente las dos probabilidades finales	97,4	0,939	1,009	0,912	0,992
Ej. 2	Calcula correctamente tanto probabilidades iniciales como verosimilitudes	91,3	0,850	0,976	0,796	0,964
	Calcula correctamente las dos probabilidades finales	93,4	0,879	0,989	0,825	0,976

En ambos ejercicios, la mayor parte de los alumnos (ejercicio 1, 85,9%; ejercicio 2, 91,3%) ha organizado correctamente los datos, según la tabla de Bayes. También la mayoría de los alumnos realiza correctamente los cálculos de las probabilidades finales.

Un de 13,04% los alumnos que resolvieron el problema tuvieron problemas en la interpretación de los resultados. Al preguntarles por la probabilidad de que la alarma sea infundada *P(I/A)*, estos alumnos dieron la probabilidad contraria *P(NI/A)*. Esto es debido a que la probabilidad de que sea infundada es mucho mayor a la probabilidad de que la alarma sea cierta. Este hecho, contraintuitivo, hizo que este grupo de alumnos cometiera el fallo de dar la probabilidad inversa como respuesta correcta.

Los ejercicios de la **sesión segunda** (tabla 6.49) piden al alumno que asigne una distribución inicial de la proporción, calcular la distribución final con la ayuda del programa PRODIST y hacer diferentes inferencias (dar un mejor estimador y dar la probabilidad de que la proporción tome un determinado valor). Un 96,1% (ejercicio 1) de los alumnos escribe la distribución inicial y la mayoría de los alumnos calcula la distribución final (ejercicio 1, 84,7%; ejercicio 2, 70,5%).

Tabla 6.49. Porcentaje de respuestas correctas en los ejercicios de la segunda sesión, intervalos de confianza y de credibilidad (*n*=78)

			Intervalo confianza 95%		Intervalo credibilidad 95%	
		% Respuestas correctas	Lim inf	Lim inf	Lim sup	Lim sup
Ej. 1	Escribe la distribución inicial	96,1	0,918	1,004	0,893	0,986
	Escribe la distribución final	84,7	0,767	0,927	0,750	0,909
	Obtiene inferencias (mejor estimador o probabilidad de un valor)	93,6	0,882	0,990	0,858	0,971
Ej. 2	Escribe la distribución inicial					
	Escribe la distribución final	70,5	0,604	0,806	0,596	0,794
	Obtiene inferencias (mejor estimador o probabilidad de un valor)	85,9	0,782	0,936	0,810	0,946

El 93,6% (ejercicio 1) de los alumnos encuentra la moda de la distribución final a partir de los resultados y el 85,9% de los alumnos calculan correctamente la probabilidad de que el parámetro tome un valor (o mayor). Tenemos que comentar que 29 alumnos (37,2%) confundieron éxitos con fracasos y obtuvieron la distribución complementaria. Con lo que al final la respuesta que daban es que la proporción más probable de bombillas fundidas era 0,9.

El ejercicio de la **tercera sesión** (tabla 6.50) plantea varias preguntas con relación a la inferencia sobre la proporción en el caso continuo. En primer lugar se le pide que de una distribución inicial, una distribución final y un estimador en cada una de las situaciones (antes y después de recoger los datos).

Tabla 6.50. Porcentaje de respuestas correctas en los ejercicios de la tercera sesión, intervalos de confianza y de credibilidad (*n*=78)

		Intervalo confianza 95%		Intervalo credibilidad 95%	
	% Respuestas correctas	Lim inf	Lim inf	Lim sup	Lim sup
Distribución inicial	89,7	0,830	0,964	0,810	0,946
Distribución final	89,7	0,830	0,964	0,810	0,946
Mejor estimación antes de tomar los datos	28,2	0,182	0,382	0,184	0,372
Mejor estimador después de tomar los datos	85,9	0,782	0,936	0,765	0,919
Intervalo de credibilidad	92,3	0,864	0,982	0,842	0,963
Probabilidad de un valor	76,9	0,675	0,863	0,664	0,848
Contraste de hipótesis	85,9	0,782	0,936	0,765	0,919

El 89,7% asigna la distribución inicial en situación de ausencia de información y determina los parámetros de la distribución beta final a partir de los datos del problema. Sólo un 28,2% de los alumnos da la respuesta correcta al mejor estimador antes de tener información. La mayoría de los alumnos calcula correctamente el intervalo de credibilidad (92,3%). El 76,9% de los estudiantes obtiene el valor de probabilidad correctamente y el 85,9% realizan correctamente el contraste de hipótesis.

Tabla 6.51. Porcentaje de respuestas correctas en los ejercicios de la cuarta sesión, intervalos de confianza y de credibilidad (*n*=78)

			Intervalo confianza 95%		Intervalo credibilidad 95%	
		% Respuestas correctas	Lim inf	Lim inf	Lim sup	Lim sup
Ej. 1	Llegan a la solución final	78,2	0,690	0,874	0,678	0,858
	Tipifican / Destipifican	83,3	0,750	0,916	0,724	0,891
	Identifican el intervalo Z / Definen hipótesis	84,6	0,766	0,926	0,750	0,909
	Calculan la distribución final	85,9	0,782	0,936	0,765	0,919
	Identifican los datos	88,5	0,814	0,956	0,795	0,937
Ej. 2	Llegan a la solución final	67,9	0,575	0,783	0,569	0,772
	Tipifican / Destipifican	87,1	0,797	0,945	0,780	0,928
	Identifican el intervalo Z / Definen hipótesis	88,5	0,814	0,956	0,795	0,937
	Calculan la distribución final	82,0	0,735	0,905	0,721	0,889
	Identifican los datos	78,2	0,690	0,874	0,678	0,858

Finalmente en la **cuarta sesión** (tabla 6.51) se plantearon dos ejercicios de estimación de una media con distribuciones continuas y le pedimos al alumno que calcule el intervalo de credibilidad y que realizara un contraste de hipótesis, respectivamente. Para que el alumno llegue a la respuesta correcta, tiene que realizar varios pasos. En primer lugar tendría que identificar los datos (88,5% de respuestas correctas); a continuación calcular la distribución final (88,5%); identificar qué intervalo en la tabla de la normal tipificada corresponde al 95% de probabilidad (84,6%) y por último transformar ese intervalo, destipificando (83,3%). Si el alumno no se equivoca en alguno de los pasos ni comete algún error de cálculo, obtendría el intervalo de credibilidad correcto (78,20%).

3.3.4.4. EVALUACIÓN FINAL DEL APRENDIZAJE

En primer lugar se estimó la fiabilidad de la prueba final de autoevaluación desde la perspectiva del estudio de la consistencia interna, fundamentada por la teoría clásica de los tests (Carmines y Zéller, 1979; López Feal, 1992; Muñiz, 1992; Martínez Arias, 1995). Mediante el paquete estadístico SPSS se obtuvo un valor del coeficiente Alfa = 0,661, con n = 62 sujetos que se consideró suficiente.

Tabla 6.52. Correlación de la puntuación total en AIB con la calificación en la asignatura

		examen	TOTAL
examen	Correlación de Pearson	1	,410(**)
	Sig. (bilateral)		,001
	n	62	62

** La correlación es significativa al nivel 0,01 (bilateral).

La correlación entre los resultados de la evaluación final de la enseñanza y el examen final de la asignatura resulta positiva y significativa, superior al valor mínimo aceptable en estudios de validez discriminante (García Cueto, 1992; Barbero, 1993), lo que permite justificar la validez respecto a criterio de la prueba AIB siendo el criterio la calificación final en la asignatura (Martínez Arias, 1995).

Los resultados de la prueba de evaluación son bastantes satisfactorios, en cuanto que indican altos porcentajes de respuestas en la mayor parte de los ítems, incluso en aquellos que plantean preguntas complejas y tareas problemáticas (tabla 6.53).

Señalamos también que los principales problemas detectados no se relacionan directamente con la enseñanza de la inferencia bayesiana. Por un lado, la dificultad en diferenciar una probabilidad condicional de su inversa y la confusión con la conjunta ha dificultado alguna de las tareas, lo que refuerza nuestra hipótesis de la importancia del razonamiento condicional para un correcto aprendizaje de la inferencia y dota de mayor valor nuestro estudio de evaluación de dicho razonamiento, así como el instrumento preparado para ello. En segundo lugar hemos observado olvido de algunos conceptos y técnicas previas, como manejo de las tablas de la distribución, confusión de estadístico y parámetro o manejo de desigualdades, aunque su incidencia no ha sido tan fuerte como la de la probabilidad condicional. En unos pocos casos los alumnos han olvidado fórmulas, como la desviación típica de la media de la muestra, lo cual es comprensible, dado que no contaban con material de consulta al realizar la prueba. Estos errores no son conceptuales sino de memoria y pensamos que se pueden superar fácilmente cuando el alumno pueda consultar el material para aplicar la inferencia bayesiana en el futuro.

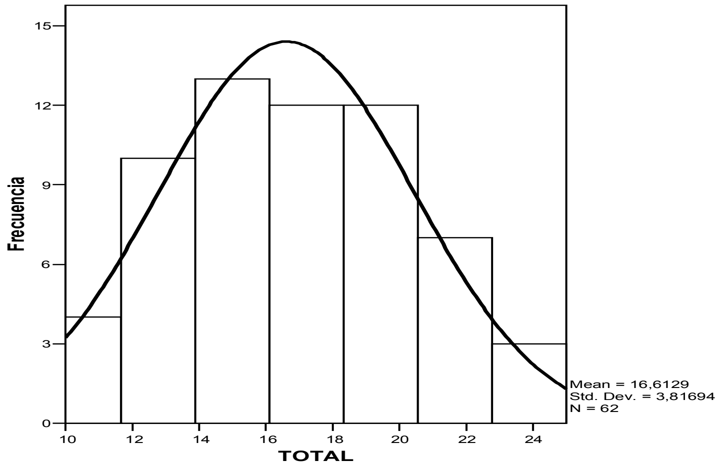
Tabla 6.53. Porcentaje de respuestas correctas en la evaluación final aprendizaje, intervalos de confianza y de credibilidad (n= 62)

	% Respuestas correctas	Intervalo confianza 95%		Intervalo credibilidad 95%	
		Lim inf	Lim inf	Lim sup	Lim sup
1	88,7	0,808	0,966	0,784	0,943
2a	79,0	0,689	0,891	0,673	0,872
2b	38,7	0,266	0,508	0,276	0,511
2c	29,0	0,177	0,508	0,192	0,412
2d	51,6	0,392	0,639	0,394	0,635
3	66,1	0,543	0,779	0,537	0,766
4	58,1	0,458	0,779	0,456	0,695
5	61,3	0,492	0,734	0,488	0,723
6	50,0	0,376	0,624	0,366	0,604
7	93,5	0,874	0,996	0,845	0,973
8	53,2	0,408	0,656	0,409	0,650
9	85,5	0,767	0,943	0,746	0,921
10	64,5	0,526	0,764	0,520	0,752
11	58,1	0,458	0,704	0,456	0,695
12	53,2	0,408	0,656	0,409	0,650
13	69,4	0,579	0,809	0,570	0,793
14	30,6	0,191	0,421	0,206	0,429
15	40,3	0,281	0,525	0,290	0,527
16	69,4	0,579	0,809	0,570	0,793
17	69,4	0,579	0,809	0,570	0,793
18	79,0	0,689	0,891	0,673	0,872
19	58,1	0,458	0,704	0,456	0,695
20a	82,3	0,728	0,918	0,709	0,897
20b	72,6	0,615	0,837	0,582	0,800

Puntuación total en la prueba de evaluación

Para resumir los resultados de la prueba final de evaluación, presentamos en la figura 6.10 el histograma de frecuencias de la puntuación total obtenida por los estudiantes.

Figura 6.10. Distribución de puntuaciones en la prueba de evaluación final



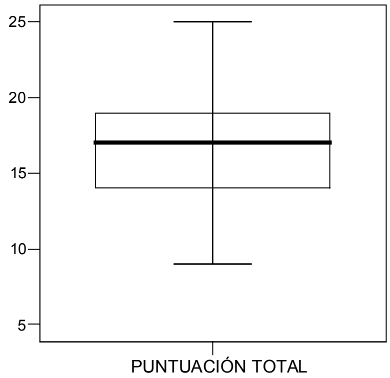
Teniendo en cuenta que la prueba tenía 26 subapartados los resultados son satisfactorios, teniendo en cuenta que la puntuación media teórica es 13 y la media obtenida en la muestra la supera. Además solo 13 alumnos (%) puntuaron por debajo de la media, es decir obtuvieron menos de la mitad de las preguntas respondidas correctamente.

Tabla 6.54. Estadísticos descriptivos para la puntuación total en el cuestionario AIB

		TOTAL
Mínimo		9
Máximo		25
Media		16,6129
	Error típico	,48475
Desv. típ.		3,81694
Varianza		14,569
Asimetría		,050

Esta misma conclusión se obtiene examinando el gráfico de caja, donde observamos que la mediana se sitúa en 18 preguntas y el 50% central de alumnos responde entre 14 y 19. Los estadísticos calculados indican una distribución simétrica (aproximadamente normal) y permiten calcular los intervalos de confianza y credibilidad. En este caso, al tratarse de una distribución normal, con distribución inicial no informativa coinciden, aunque no así su interpretación, como hemos sugerido anteriormente.

Figura 6.11. Diagrama de caja para la puntuación total en el cuestionario AIB



La media de las puntuaciones de la evaluación final sigue una distribución *N*(16’61, 0’48). El intervalo de credibilidad al 95% sería [15’67, 17’55].

3.3.4.5. OBJETIVOS CONCEPTUALES ALCANZADOS

La experiencia de enseñanza que hemos llevado a cabo en un contexto de aula real (con 75 alumnos voluntarios que cursaban primer año de la licenciatura de Psicología) y durante un periodo de tiempo limitado (12 horas) ha producido unos resultados de aprendizaje razonable, teniendo en cuenta que estos alumnos, al encontrarse al final del curso, no podían dedicar demasiado tiempo al estudio personal en sus casas.

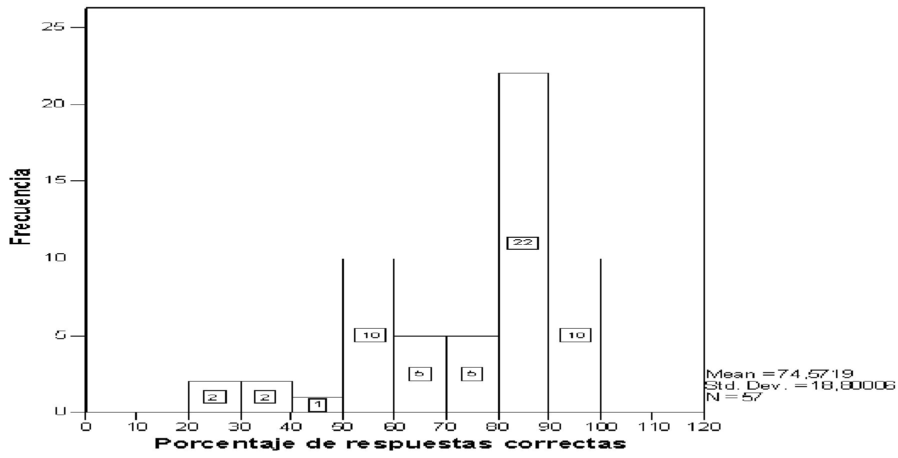
Para sintetizar los logros de aprendizaje en la tabla 6.55 presentamos el porcentaje de estudiantes que ha logrado cada uno de los objetivos de aprendizaje en las tareas propuestas durante el desarrollo de la enseñanza.

El porcentaje de estudiantes que ha sido capaz de completar con éxito las tareas teóricas y prácticas en el aula durante la experiencia ha sido notablemente alto en todas las tareas. Es verdad que durante el desarrollo de estas tareas el alumno podía consultar el material, pero esta es la situación habitual de trabajo de un profesional que no sólo cuenta con su memoria, sino puede consultar documentos que le faciliten el trabajo. Además cada alumno resolvía las tareas sin consultar al compañero o al profesor y el éxito también es indicativo de que los documentos preparados para los estudiantes, así como la forma en que se ha diseñado y organizado la enseñanza han cumplido el objetivo propuesto.

Tabla 6.55. Porcentaje de alumnos que han conseguido cada objetivo

	Objetivos alcanzados en la evaluación	Continua		Final	
		Ítem	%	Ítem	%
Relacionados con la inferencia bayesiana	Diferenciar entre probabilidades iniciales y finales y verosimilitudes.	1_1	83,3	1	88,7
		1_2	66,7		
	Identificar sucesos, sus probabilidades iniciales y verosimilitudes	1_5	82,1	1	88,7
	Analizar el teorema de Bayes como herramienta para transformar probabilidades iniciales en finales.	1_3	93,6	7	93,5
		1_8	93,6		
	Organizar los datos en una tabla Bayes.	1_6	85,9	7	93,5
	Visualizar los parámetros como variables aleatorias con una distribución de probabilidad asociada.	2_1	52,6	3	66,1
		2_2	89,7	5	61,3
	Identificar valores de la proporción poblacional y asignarles probabilidades iniciales plausibles (distribución de probabilidad inicial).	2_3	94,9	6	50
		Ej.2_1	69,2		
	Comprender la distribución inicial uniforme y distribución informativa.	2_5	85,9	4	58,1
	Asignar una distribución inicial no informativa a la proporción.	2_6	85,9	6	50
	Asignar correctamente una distribución inicial informativa a la proporción	2_7	91,0	8	53,2
	Identificar los valores de éxitos y fracasos en la muestra.	Ej.2_1	84,7	8	53,2
	Aprender a introducir los datos anteriores en la hoja Excel para obtener los valores de verosimilitud y la distribución final.	2_8	93,6	7	93,5
		Ej.2_1	84,7	18	85.5
	Reconocer la fórmula del teorema de Bayes en formulación no estándar.	2_4	62,8		
	Obtener la mejor estimación de la proporción en la población.	Ej.2_1	93,6		
		3_2	88,5		
		3_4	52,6		
	Calcular probabilidades asociadas a determinados valores de proporción en la distribución final.	Ej.2_2	85,9		
		3_5	91,0		
	Entender la transformación de dist. iniciales en experimentos sucesivos.	Ej.2_2	70,5		
	Analizar las diferencias entre la metodología clásica y bayesiana.	2_2	89,7		
	Usar la distribución Beta para definir una distribución inicial para la proporción en caso no informativo.	3_1	93,6		
		Ej.3_3	89,7		
	Usar la distribución Beta para definir una distribución inicial para la proporción en caso informativo.	3_3	85,9	20	72,6
	Identificar los éxitos y fracasos en la muestra.	3_3	85,9	9	85,5
		3_4	52,6	18	85.5
		Ej.3_3	89,7		
	Calcular la distribución final con la fórmula *Be(a + e, b + f).*	3_4	52,6		
		Ej.3_3	89,7		
	Usar la moda de la distribución Beta *a/(a+b)* como mejor estimador.	3_2	88,5		
		Ej.3_3	85,9		
	Utilizar xcel para obtener la distribución de probabilidad final.	Ej.3_3	92.3	18	79
	Calcular intervalos de credibilidad.	3_5	91,0	10	54,5
	Interpretar el intervalo de credibilidad de una proporción.	3_6	55,1	10	54,5
		3_7	39,7	12	53,2
		Ej.3_3	92,3		
	Contraste de hipótesis sencillas sobre una proporción siguiendo la metodología bayesiana.	3_6	55,1	11	58,1
		Ej.3_3	85,9		
	Diferenciar entre estadístico y parámetro en el contexto de un problema.	2_1	52,6	19	58,1
	Especificar valores plausibles de la media poblacional y asignarles probabilidades iniciales.	4_1	53,8		
		Ej.4_1	88,5		
		Ej.4_2	78,2		
	Con los datos obtenidos en una muestra, obtener una distribución final.	4_2	73,7	13	69,4
		4_4	64,1	14	40,6
		Ej.4_1	85,9		
		Ej.4_2	82,0		
	Extraer inferencias sobre la población a partir de la distribución final.	4_7	55,1	43	69,4
	Cálculo de intervalos de credibilidad para la media de una población.	4_5	46,2	16	69,4
		Ej.4_1	78,2		
	Interpretar un intervalo de credibilidad de la media.	4_6	28,2	16	69,4
		Ej.4_2	78,2		
	Realizar correctamente un contraste de hipótesis sobre la media, desde una perspectiva bayesiana.	4_3	53,8	14	30,6
		Ej.4_3	67,9		
	Diferenciar una probabilidad condicional y su transpuesta.	1_4	35,9	2	38,3
	Diferenciar probabilidades condicionales y conjuntas en un problema.	1_7	23,1	2	61,6

Figura 6.12. Porcentajes de objetivos alcanzados en la evaluación continua



Asimismo, el porcentaje de estudiantes que ha dado respuestas correctas en las tareas propuestas en la evaluación ha sido bastante alto, algo menos que el anterior porque el estudiante debía basarse solo en su memoria y no podía consultar sus apuntes.

A pesar de ello, los resultados son muy buenos, lo que indica un alto grado de consecución de los objetivos de nuestra propuesta didáctica.

Figura 6.13. Porcentajes de objetivos alcanzados en la evaluación final

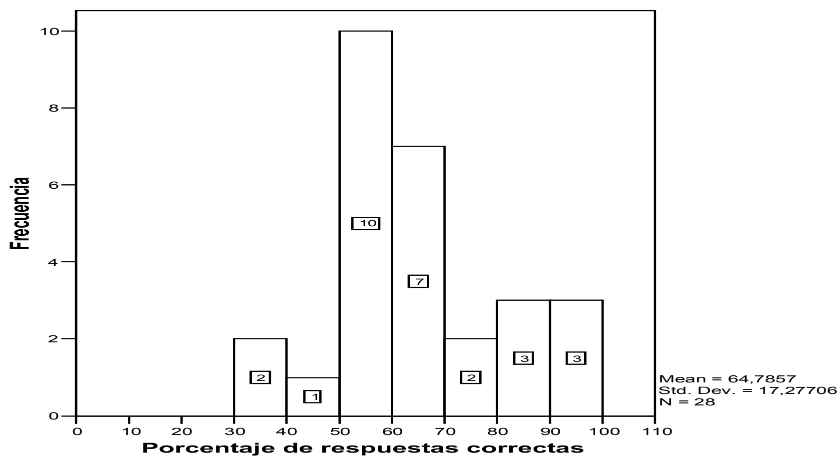


Figura 6.14. Grafico de la caja

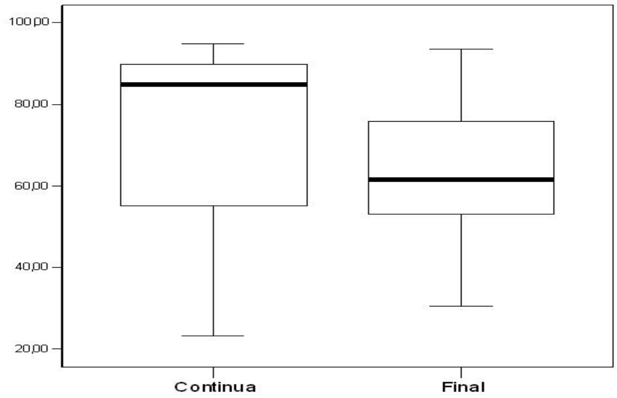


Tabla 6.56. Estadísticos descriptivos

	Continua	Final
Número de objetivos	57	28
Porcentaje de aciertos mínimo	23,10	30,60
Porcentaje de aciertos máximo	94,90	93,50
Porcentaje de aciertos medio	74,571	64,78
Desv. típ.	18,80	17,28
Varianza	353,44	298,49
Asimetría	-,981	,186
Error típico	,316	,441
Curtosis	-,031	-,611
Error típico	,623	,858

Mas allá del éxito de la experiencia, pensamos que nuestro trabajo proporciona una información detallada del grado en que se han cumplido cada uno de los objetivos específicos de la enseñanza, informando también de los principales errores y sus causas. En este sentido pensamos que nuestro trabajo es pionero en la didáctica de la inferencia bayesiana, en la que los pocos trabajos existentes tan sólo presentan resultados muy globales o bien resultados subjetivos no apoyados por datos empíricos sobre los logros de los estudiantes.

3.3.5. DISCUSIÓN DEL ESTUDIO 8.2

En este estudio se llevó a cabo una experimentación del material y la propuesta de enseñanza de conceptos elementales de inferencia bayesiana diseñado en el estudio 8.1 en una muestra de 78 estudiantes de psicología. Las condiciones de la experimentación fueron muy similares a las de los cursos habituales, salvo por el tamaño reducido de los grupos (de 15 a 20 alumnos cada uno), restricción que fue dada por el tamaño del aula de informática necesaria para la parte práctica de la enseñanza.

Los resultados del estudio apoyan la viabilidad de la enseñanza propuesta a estos alumnos, mostrados en el análisis detallado de los resultados de las pruebas de autoevaluación y ejercicios propuestos en cada una de las sesiones y los resultados de la evaluación final llevada a cabo mediante la prueba AIB.

En concreto, los objetivos de enseñanza planificados para la primera sesión se consiguieron en la mayoría de los alumnos, que demostraron una comprensión de los conceptos y su aplicación en diversos contextos prácticos. El único objetivo que parece no conseguirse es la correcta diferenciación entre una probabilidad condicional y su transpuesta, que no es específico de la inferencia bayesiana. Hacemos notar que la enseñanza comenzó por el Teorema de Bayes (y no por la probabilidad condicional propiamente dicha), aunque tratamos de incorporar a la misma las dificultades notadas en el cuestionario de evaluación de razonamiento condicional. Este punto también apoya nuestra hipótesis discutida en los capítulos anteriores de las dificultades de los estudiantes con el razonamiento en probabilidad condicional.

La segunda unidad didáctica ha resultado sencilla a los alumnos quienes, en su mayoría diferencian una distribución inicial de la final, son capaces de asignar distribuciones iniciales no informativas e informativas, reconocen y saben aplicar la fórmula de Bayes para estimación de la proporción en una población (caso discreto) y comprenden la diferente consideración del parámetro en inferencia clásica y bayesiana. El único problema presentado es que parte de los alumnos no recordaba el significado del término parámetro, por lo que hubo que recordárselos.

Pocos alumnos mostraron dificultad al pasar del caso discreto al continuo en el estudio de la estimación de la proporción. La mayoría de ellos comprendieron bien la distribución Beta y el significado de sus parámetros, siendo capaces de aplicarlos en varios contextos prácticos. Son capaces de asignar una distribución inicial razonable, tanto informativa como no informativa y manejan con facilidad las tablas de probabilidad y valores críticos de la distribución Beta.

Los principales errores que se producen guardan relación con el razonamiento por contradicción, la lógica global del contraste de hipótesis y el significado del coeficiente de credibilidad. Estos errores son comunes a los descritos en inferencia bayesiana, aunque la incidencia en nuestra muestra no fue tan grande como la descrita para ella.

Hay que tener en cuenta el poco tiempo de enseñanza, así como la desaparición de otros errores, como son los de interpretación de los riesgos de error.

Los alumnos han adquirido un conocimiento sobre la estimación de la media en una distribución normal, sabiendo elegir una distribución inicial no informativa adecuada y calcular, en este caso, la distribución final resultante. La mayoría aplica correctamente, y usando la distribución final adecuada, el procedimiento de contraste y cálculo de intervalos de credibilidad a partir de situaciones prácticas. Conocen el programa de cálculo y lo usan correctamente para sacar conclusiones sobre problemas.

Se han observado en una proporción de los alumnos dificultades comunes con la inferencia clásica, como son la comprensión de la lógica del contraste, recordar la fórmula de la desviación típica del estimador. Hay algunos errores de aplicación de fórmulas, por ejemplo, al despejar incógnitas en desigualdades. Ninguno de estos problemas está directamente relacionado con la enseñanza de inferencia bayesiana sino que son fallos en otros conceptos que el alumno debiera tener aprendidos previamente.

Respecto a los aspectos metodológicos, el estudio proporciona unas primeras evidencias de fiabilidad y validez de contenido para el conjunto de pruebas de autoevaluación y la prueba AIB, que dotan a estos instrumentos de utilidad para futuras experiencias de enseñanza o futuras investigaciones. Sería interesante continuar en el futuro el estudio de estos instrumentos, completando el estudio de su validez en nuevas muestras de estudiantes.

3.4. CONCLUSIONES SOBRE LA VIABILIDAD DE LA ENSEÑANZA DE CONCEPTOS ELEMENTALES DE INFERENCIA BAYESIANA EN PSICOLOGÍA

Los resultados del estudio 8 apoyan la viabilidad de la enseñanza propuesta a estos alumnos, mostrados en el análisis detallado de los resultados de las pruebas de autoevaluación y ejercicios propuestos en cada una de las sesiones y los resultados de la evaluación final llevada a cabo mediante la prueba AIB.

Las dificultades encontradas en el aprendizaje se pueden resumir en:

- Diferenciación entre una probabilidad condicional y su transpuesta, que no es específico de la inferencia bayesiana. Esto apoya nuestra hipótesis discutida en los capítulos anteriores de las dificultades de los estudiantes con el razonamiento en probabilidad condicional.

- Razonamiento por contradicción, la lógica global del contraste de hipótesis y el significado del coeficiente de credibilidad. Estos errores son comunes a los descritos en inferencia bayesiana, aunque la incidencia en nuestra muestra no fue tan grande como la descrita para ella.
- Dificultades comunes con la inferencia clásica, como son la comprensión de la lógica del contraste, recordar la fórmula de la desviación típica del estimador.
- Errores de aplicación de fórmulas. Ninguno de estos problemas está directamente relacionado con la enseñanza de inferencia bayesiana sino son fallos en otros conceptos que el alumno debiera tener aprendidos previamente.

4. ESTUDIO 9. INTERRELACIÓN ENTRE RAZONAMIENTO CONDICIONAL Y APRENDIZAJE DE INFERENCIA BAYESIANA

4.1. INTRODUCCIÓN

En el capítulo 1 dedicado a los fundamentos justificábamos teóricamente la importancia de la probabilidad condicional como base de la definición de conceptos elementales de inferencia bayesiana. Asimismo, al establecer los objetivos e hipótesis de la investigación en el capítulo 2, plateábamos nuestra hipótesis sobre la existencia de interrelación entre el razonamiento sobre probabilidad condicional de los estudiantes y su aprendizaje de conceptos elementales de inferencia bayesiana. Una vez analizado el material elaborado para la enseñanza de estos conceptos a estudiantes de psicología (estudio 8.1) y realizada una evaluación del aprendizaje logrado (estudio 8.2) dedicamos este nuevo estudio a aportar evidencias que apoyen la hipótesis mencionada.

En consecuencia, el objetivo que nos planteamos en este estudio es analizar si los alumnos que tienen un mejor razonamiento en probabilidad condicional, también muestran un mejor aprendizaje en todos o parte de los conceptos bayesianos introducidos en nuestra enseñanza. Asimismo pretendemos identificar la existencia de interrelaciones e implicaciones entre los diferentes objetivos de aprendizaje logrados por los estudiantes en particular de aquellos en los que interviene la probabilidad condicional.

4.2. MÉTODO

4.2.1. SUJETOS

La muestra que forma parte de este estudio es parte de la utilizada en la evaluación de la enseñanza (estudios 8.1 y 8.2), es decir, se trata de 78 estudiantes de primer año de psicología que cursaban la asignatura de análisis de datos y participaron voluntariamente en la experiencia. Como hemos justificado anteriormente, la muestra está formada por alrededor de un tercio de los estudiantes que habitualmente asisten a clase. Una submuestra (57 estudiantes) habían completado el cuestionario RPC y de ellos 50 lo hicieron en dos ocasiones (test y retest). 60 estudiantes se presentaron al examen final de la asignatura de análisis de datos, por lo cual disponíamos de su calificación en la misma.

La muestra está compuesta por un 17,9% de chicos y un 82,1% de chicas, todos alumnos de primer curso de Psicología. 57 (73%) alumnos de los que participaron en la experiencia de aprendizaje, realizaron también el cuestionario RPC, obteniendo una puntuación media de 22,7 puntos y desviación típica 3,7. Los alumnos participantes obtuvieron una media de 4,83 en el examen final de la asignatura (convocatoria de Junio), con desviación típica de 2,07.

Siendo la muestra no aleatoria, como es casi general en los estudios educativos, nos encontraríamos en un diseño cuasi-experimental, de tipo correlacional (Cook y Campbell, 1979)[43]. La finalidad del análisis será exploratoria, en cuanto a la intención de generalización a poblaciones diferentes de estudiantes.

4.2.2. MATERIAL Y PROCEDIMIENTO

Se utilizó el material y procedimiento descrito en los dos estudios anteriores que consta de: a) El cuestionario RPC, para evaluación del razonamiento en probabilidad condicional y b) la prueba AIB, utilizada para la evaluación del aprendizaje de la inferencia bayesiana (estudio 8.2). Remitimos a las secciones correspondientes para su

[43] Contemplamos diversas variables explicativas que queremos correlacionar con la principal variable dependiente del estudio que es la puntuación final en la prueba AIB

descripción.

4.2.3. ANALISIS

El estudio de la interrelación entre la comprensión de la probabilidad condicional y aprendizaje de la inferencia bayesiana se lleva a cabo mediante las correlaciones de Pearson y Spearman entre puntuaciones en el cuestionario RPC, la puntuación total en la prueba de evaluación de Aprendizaje de Inferencia Bayesiana (AIB) y la calificación final en la asignatura de análisis de datos.

Con estos análisis proporcionamos una aproximación a la *validez referida a un criterio* (Thorndike, 1991; Martínez Arias, 1995) para el cuestionario RPC, medido como la correlación con el criterio externo, en nuestro caso la puntuación en la prueba AIB. Se trataría de validez predictiva puesto que el criterio ha sido medido con posterioridad al cuestionario RPC que sirve, además para predecir el comportamiento respecto al criterio (Martínez Arias, 1995). Santisteban (1990) indica que ha de transcurrir un tiempo entre la medición del predictor y el criterio[44]; en nuestro caso transcurrieron aproximadamente tres meses desde que los estudiantes pasaron el cuestionario RPC y completaron el curso y posterior evaluación del Aprendizaje de Inferencia Bayesiana (AIB). El análisis es realizado con el programa SPSS.

Para estudiar las interrelaciones e implicaciones entre objetivos de aprendizaje hemos llevado a cabo varios análisis multivariantes entre las respuestas a los ítems de la prueba AIB, utilizando el software CHIC, Classification Hierarchical, Implicative et Cohesive (Couturier, y Gras, 2005)[45].

En primer lugar se realiza un estudio de aglomeración jerárquica (Bacelar-Nicolau, 1992; 2003), tomando como medida de similaridad entre ítems el índice de Lerman y suponiendo una distribución binomial para cada variable. Siendo a y b dos variables aleatorias dicotómicas en una población E y A y B los subconjuntos donde se verifican a y b, el índice de similaridad viene dado por la expresión siguiente (Lerman, 1981):

[44] En los estándares de APA, AERA y NCME (1999) se minimiza la importancia de la distinción entre validez predictiva y concurrente, es decir de la relación temporal.

[45] Programa desarrollado en la Universidad de Rennes, a partir de los trabajos de Lerman y Grass, entre otros.

$$\partial(a,b) = \frac{card\ (A,B) - \dfrac{card\ (A)card\ (a,b)}{n}}{\sqrt{\dfrac{card\ (A)card\ (a,b)}{n}}}$$

Las variables a y b tendrán mayor similaridad cuando el número de elementos comunes sea mayor en relación a la frecuencia esperada en caso de independencia y tiene en cuenta el tamaño muestra. La medida de similaridad induce un pre-orden parcial en el conjunto de variables. A partir de ella se define la similaridad entre clases: A y B. Sea:

$$\{U(\alpha,\beta)/(\alpha,\beta) \in AxB\}$$

La medida de proximidad entre clases viene dada por:

$$P'(A,B) = \max_{(\alpha,\beta)\in AxB} P(\alpha,\beta)$$

donde $P(\alpha,\beta)$ se calcula en la hipótesis de muestreo de dos variables independientes (Lerman, 1981).

El algoritmo de clasificación jerárquica se construye teniendo en cuenta la mayor proximidad entre elementos de una clase y la mayor distancia entre clases separadas. Proporciona también una prueba de significación de las clases obtenidas, que se calcula teniendo en cuenta no sólo la intensidad de la similaridad, sino el número total de sujetos que da una respuesta correcta al conjunto de ítems incluidos en el grupo.

Seguidamente hemos llevado a cabo un análisis implicativo entre ítems, siguiendo el modelo binomial, aunque el método se extiende a variables multinomiales (Bailleul y Gras, 1994) y variables continuas (Lagrange, 1996; Gras, Diday, Kuntz y Couturier, 2001).

Siendo a y b dos variables aleatorias dicotómicas en una población E y A y B los subconjuntos donde se verifican a y b, el índice de implicación de Gras (Gras, 1993; 1996; Gras y Ratsima-Rajohn, 1996), viene dado por la expresión siguiente:

$$q(a,\overline{b}) = \left[\frac{card(A \cap \overline{B}) - \dfrac{card(A)card(\overline{B})}{n}}{\sqrt{\dfrac{card(A)card(\overline{B})}{n}}}\right]$$

La significación del índice, que sigue una distribución *N(0,1),* da lugar a una intensidad de implicación. Se acepta que $a \Rightarrow b$ con intensidad:

$$\varphi(a,\overline{b}) = 1-\alpha \Leftrightarrow \Pr ob[car(X \cap \overline{Y}) \leq card(A \cap \overline{B})] \leq \alpha \ ,$$

siendo X e $\overline{Y}$ variables aleatorias dicotómicas e independientes, que tienen los mismos cardinales que A y $\overline{B}$ respectivamente (Lerman, Gras, y Rostam, 1981a y b).

En nuestro caso las variables aleatorias son las respuestas (correcta- incorrecta) a cada ítem, con lo que podemos calcular un total de $\binom{21}{2}$ índices de implicación entre los 21 ítems de la prueba AIB. La relación de implicación entre variables establece un preorden asimétrico dentro del conjunto de ítems, que podemos representar en un grafo ordenado (Gras, Kuntz y Briand, 2001). El programa CHIC calcula los índices de implicación entre todos los pares de variables de un conjunto de datos y proporciona un grafo mostrando todas las implicaciones que son significativas hasta el nivel pedido por el usuario (Bailleul, 2001), teniendo en cuenta tanto la intensidad de implicación, como el número de sujetos en que se cumple la relación de implicación[46].

Una vez estudiadas las implicaciones aisladas de unos ítems sobre otros, hemos llevado a cabo un estudio de clasificación implicativa. Se trata de un algoritmo que utiliza las intensidades de implicaciones entre conjuntos de variables como índice no simétrico para estudiar la cohesión interna de algunos subconjuntos de variables (Lahanier-Reuter, 2001; Couturier, Gras y Guillet, 2004). La cohesión de una clase es un concepto opuesto al de entropía[47] y se calcula por medio de una relación algebraica que tiene en cuenta la debilidad entrópica de la clase de variables considerada (David, Guillet, Philipp y Gras, 2005; Ritschard, 2005).

Dadas dos variables *a* y *b*, en nuestro caso respuestas a dos ítems, y siendo $p = \max(\varphi(a,\overline{b}), \varphi(b,\overline{a}))$ y $H = -p\log_2 p - (1-p)\log_2(1-p)$, la entropía del conjunto de variables, el grado de cohesión entre dos variables se define como $c(a,b) = \sqrt{1-H^2}$

[46] La implicación entre un ítem y otro se interpreta en el sentido de que si un estudiante es capaz de resolver correctamente un ítem mejora su probabilidad de resolver correctamente otro implicado por aquél. En este sentido el árbol implicativo nos proporciona una pauta de posible orden de introducción de los conceptos y procedimientos evaluados por los diferentes ítems.

[47] Puesto que el concepto de entropía está inversamente relacionado al de cantidad de información proporcionada por un conjunto de variables, el índice se puede interpretar como cantidad de información que una variable proporciona sobre otra.

que también varía entre 0 (mínimo) y 1 (máximo). El grado de cohesión de una clase se calcula en cada paso con la siguiente fórmula (Gras, Kuntz y Briand, 2001):

$$C(\underline{A}) = \left[\prod_{\substack{i \in \{1,\ldots r-1\} \\ j \in \{2,\ldots r\}, j>i}} c(a_i, a_j) \right]^{\frac{2}{r(r-1)}}$$

Por último, dados $\underline{A}$ y $\underline{B}$ dos conjuntos de variables, formados por variables a_i y b_j y C($\underline{A}$) y C($\underline{B}$), sus índices de cohesión, la intensidad de implicación de la clase $\underline{A}$ sobre la clase $\underline{B}$ se define (Couturier, 2001):

$$\psi(\underline{A}, \underline{B}) = \left[\sup_{i \in \{1,\ldots r\} j \in \{1,\ldots s\}} \varphi(a_i, \overline{b}_j) \right]^{rs} \left[C(\underline{A}).C(\underline{B}) \right]^{1/2}$$

El programa CHIC calcula el nivel de significación de los diferentes nodos en una jerarquía implicativa, así como las contribuciones de los sujetos El algoritmo forma las clases teniendo en cuenta los siguientes criterios:

- la cohesión máxima dentro de cada clase
- el mayor grado de implicación entre una clase y otra que es implicada por ella

4.3 RESULTADOS

En primer lugar, presentamos en la tabla 6.57 los coeficientes de correlación de Pearson y Spearman entre las puntuaciones en el cuestionario RPC y la prueba AIB para los alumnos que, habiendo completado el cuestionario de probabilidad condicional, también participaron en la experiencia de enseñanza. En total 57 alumnos participaron en ambos estudios, 50 de los cuales completaron el cuestionario en dos ocasiones (test y retest) y 60 alumnos se presentaron al examen final de análisis de datos.

Observamos que todas las correlaciones son estadísticamente significativas a nivel p= 0,01 y aunque la intensidad de la correlación es baja, superan el valor 0,20 considerado por García Cueto (1992) y Barbero (1993) los mínimos aceptables para el estudio de la validez discriminante. Por otro lado, algunos de los ítems en el cuestionario RPC se refieren a sesgos muy arraigados, incluso en alumnos con alto razonamiento en probabilidad condicional y del mismo modo algunos ítems de la prueba AIB se refieren a conocimientos que no requieren el uso de la probabalidad condicional, lo que justifica que las correlaciones no sean mayores.

Tabla 6.57. Correlación de Pearson y Spearman entre puntuaciones totales en RPC y AIB

	Puntuación total en cuerstionario RPC		
Puntuación total en prueba AIB	Total en Retest (*n*=50)	Ttotal en Test (*n*=57)	A. Datos (*n*=60)
Correlación de Pearson	0,294(*)	0,306(*)	0,410(**)
Sig. (unilateral)	0,019	0,010	0,001
Coeficiente de correlación Rho de Spearman	0,349(**)	0,315(**)	****
Sig. (unilateral)	0,006	0,008	****

** La correlación es significativa al nivel 0,01 (unilateral).

En la tabla 6.58 presentamos las correlaciones de Spearman y Pearson para aquellos ítems del cuestionario RPC relativos a elementos de la probabilidad condicional incluidos en la definición de otros conceptos bayesianos, observando que se obtienen valores significativos, aunque moderados o bajos por las mismas razones anteriormente expuestas. Estas correlaciones significativas y superiores al mínimo aceptable en el estudio de la validez discriminante indican que los ítems citados tienen validez discriminante predictiva respecto al aprendizaje de métodos elementales de inferencia bayesiana, medido mediante la puntuación total en la prueba AIB.

Tabla 6.58. Correlación de Pearson y Spearman entre puntuación en ítems de RPC y total en AIB

	Correlación Pearson	Rho de Spearman
i5 Probabilidad condicional e Independencia	0,223(*)	0,244(*)
Sig. (unilateral)	0,048	0,034
i9 Probabilidad compuesta y falacia conjunción	0,489(**)	0,462(**)
Sig. (unilateral)	0,000	0,000
i12 Independencia en experimentos repetidos	0,226(*)	0,263(*)
Sig. (unilateral)	0,046	0,024
i13 Resolución de problemas mediante regla del producto	0,313(**)	0,322(**)
Sig. (unilateral)	0,009	0,007
i16. Resolución de problemas mediante Teorema de Bayes	0,318(**)	0,356(**)
Sig. (unilateral)	0,008	0,003

* La correlación es significativa al nivel 0,05 (unilateral).
** La correlación es significante al nivel 0,01 (unilateral).

La correlación entre los resultados de la evaluación final de la enseñanza y el examen final de la asignatura resulta positiva y significativa.

Seguidamente realizamos el estudio de interrelaciones entre respuestas a ítems en la prueba AIB. Se comprobaron las hipótesis para aplicación del método: unidad experimental de las variables e independencia de las respuestas a los diferentes ítems.

Se eligió el algoritmo para ítems dicotómicos, el método clásico de análisis que es más conservador y produce menos agrupamientos. Asimismo se aceptó el supuesto de que las variables se ajustan al modelo bimomial, es decir que cada alumno tiene la misma probabilidad de acierto en cada ítem y las respuestas de los alumnos son ensayos independientes (Lerman, 1981), De hecho esto coincide con nuestro supuesto al seguir la perspectiva del Modelo Lineal de la Teoría Clásica de Test (TCT).

Comenzamos mostrando en la tabla 6.59 los estadísticos descriptivos de los ítems de la prueba AIB y en la tabla 6.60 los coeficientes de Lerman en cada paso de proceso, alguno de los cuáles resultaron estadísticamente significativos. En la figura 6.15 mostramos el árbol de similaridad, construido por el software CHIC.

Tabla 6.59. Estadísticos descriptivos de los ítems de la prueba AIB

Ítem	Frecuencia de respuestas correctas	Proporción de aciertos	Desv. Tip.
f1	49,00	0,79	0,41
f2_1	24’00	0,39	0,49
f2_2	18,00	0,29	0,45
f2_3	32,00	0,52	0,50
f2_4	41,00	0,66	0,47
f3	36,00	0,58	0,49
f4	38,00	0,61	0,49
f5	31,00	0,50	0,50
f6_1	36,00	0,58	0,49
f6_2	58,00	0,94	0,25
f7	33,00	0,53	0,50
f8	53,00	0,85	0,35
f9	40,00	0,65	0,48
f10	36,00	0,58	0,49
f11	33,00	0,53	0,50
f12	43,00	0,69	0,46
f13	19,00	0,31	0,46
f14	25,00	0,40	0,49
f15	43,00	0,69	0,46
f16	43,00	0,69	0,46
f17	53,00	0,85	0,35
f18_1	46,00	0,74	0,44
f18_2	49,00	0,79	0,41
f18_3	51,00	0,82	0,38
f20_1;f20_2	45,00	0,73	0,45

Observamos que, entre los grupos formados uno corresponde a los ítems f2_1 y f2_2 (coeficiente de similaridad del 0,999) que se refieren al cálculo de probabilidades condicionales, que a su vez se agrupan con el ítem f6_1 (similaridad 0,746) donde el alumno tiene que asignar una distribución inicial para una proporción. Interpretamos esta agrupación desde la perspectiva que los tres ítems requieren que el alumno identifique casos favorables (o valores posibles de la distribución) y que tenga que asignar una probabilidad a cada caso y que el mejor desempeño en probabilidad condicional facilita la asignación de una distribución inicial correcta en situaciones no informativas.

Este grupo de variables se relacionan a su vez con los siguientes ítems: f3, f5, f7, f9, f10 y f18_3. Los dos primeros se refieren al concepto de distribución inicial y de parámetro como variable aleatoria, el f7 a la determinación del mejor estimador a partir de la distribución final de la variable y el f10 a la determinación de un intervalo de credibilidad; el f18.3 al cálculo de probabilidades finales aplicando la regla de Bayes. Todas estas capacidades aparecen ligadas a las anteriores referidas a la probabilidad condicional.

Tabla 6.60. Coeficientes de similaridad en análisis jerárquico según pasos en la clasificación

Nivel clasificación	Ítems	Coef. Similaridad
1	(f2_1 f2_2)	0,9999**
2	(f2_3 f14)	0,9220
3	(f18_1 f18_2)	0,8647
4	(f11 f12)	0,8574
5	(f3 f9)	0,8390
6	(f5 f7)	0,8055
7	(f10 f18_3)	0,7899
8	((f2_3 f14) f2_4)	0,7466
9	((f2_1 f2_2) f6_1)	0,7461**
10	(f4 f20_1;f20_2)	0,7425
11	(f17 (f18_1 f18_2))	0,7336**
12	(f13 f16)	0,6921
13	(f6_2 f15)	0,6106
14	(1 f8)	0,5682
15	(((f2_3 f14) f2_4) (f11 f12))	0,4292**
16	((f3 f9) (f10 f18_3))	0,3179
17	(((f2_1 f2_2) f6_1) (f5 f7))	0,2965
18	((f4 f20_1;f20_2) (f13 f16))	0,1532
20	((f6_2 f15) (f17 (f18_1 f18_2)))	0,1262
21	((((f2_1 f2_2) f6_1) (f5 f7)) ((f3 f9) (f10 f18_3)))	0,0164***

*** Estadísticamente significativos

Figura 6.15. Árbol de similaridad con todas las variables, método clásico, ley binomial

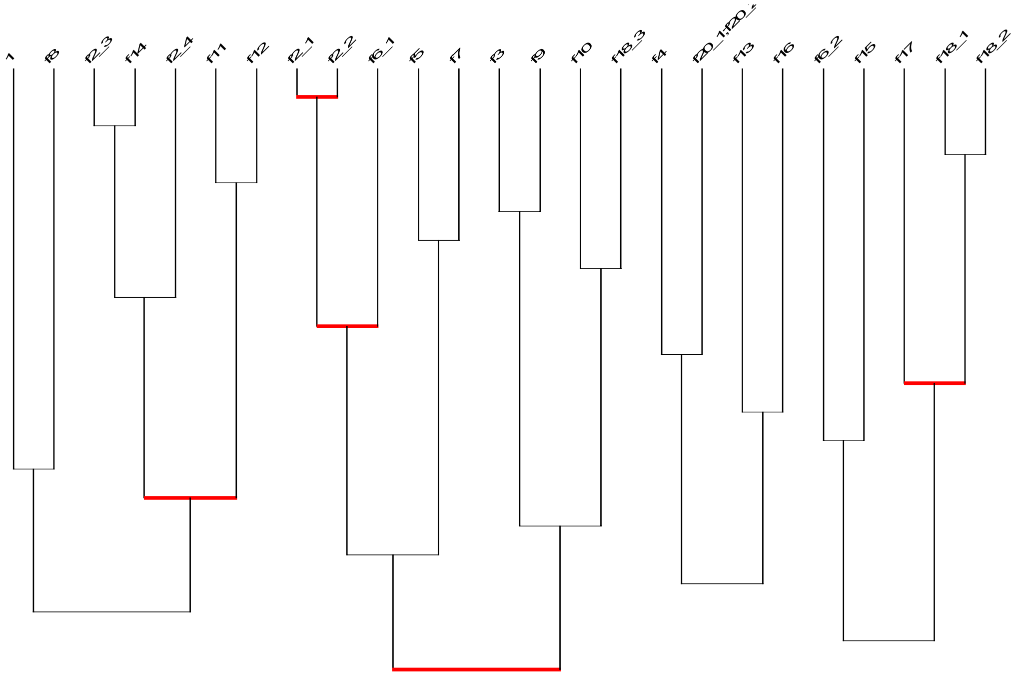


Tabla 6.61. Índices de implicación (teoría clásica)

	f1	f2_1	f2_2	f2_3	f2_4	f3	f4	f5	f6_1	f6_2	f7	f8	f9	f10
1	0	67	64	80	40	49	27	49	40	21	44	57	57	66
f2_1	88	0	100	72	91	67	33	42	79	7	79	46	85	67
f2_2	89	100	0	77	86	63	40	54	87	32	60	27	76	63
f2_3	96	67	67	0	40	47	18	34	37	34	43	32	46	58
f2_4	36	76	68	43	0	45	42	25	55	49	64	71	78	36
f3	48	62	57	48	45	0	53	27	65	41	71	89	89	74
f4	18	39	45	23	41	53	0	53	43	23	70	32	28	72
f5	47	44	53	34	17	24	54	0	54	32	78	47	42	65
f6_1	34	70	72	38	56	65	42	53	0	21	42	42	82	74
f6_2	34	36	47	45	50	47	40	45	39	0	38	34	41	47
f7	39	72	56	43	68	73	73	76	41	17	0	70	50	63
f8	55	49	43	42	62	71	40	49	46	26	59	0	51	71
f9	60	72	63	47	79	85	29	44	78	26	50	52	0	61
f10	76	62	57	57	34	74	73	62	74	41	61	89	62	0
f11	26	55	64	53	56	31	62	58	52	36	68	86	39	73
f12	41	40	49	76	78	44	50	32	63	76	54	59	46	63
f13	37	38	38	22	46	40	45	25	54	13	40	52	36	40
f14	89	67	69	91	92	72	63	37	72	22	73	97	53	82
f15	55	48	49	68	59	35	24	40	63	76	54	43	36	53
f16	30	48	49	21	49	53	60	40	35	52	45	59	46	44
f17	55	35	43	50	34	55	49	27	38	45	51	36	42	46
f18_1	50	26	36	55	30	56	42	29	47	80	40	35	37	56
f18_2	58	38	44	67	49	58	53	33	40	82	52	84	38	74
f18_3	50	61	53	50	65	39	51	64	48	42	52	61	63	80
f20_1;f20_2	60	49	45	59	46	43	75	57	43	17	53	48	43	70

Por otro lado tenemos un grupo con los ítems f2_3, f2_4, f11, f12 y f14. Este grupo pone en relación las tareas de cálculo de probabilidades condicionales y conjuntas con las tareas de realizar un contraste de hipótesis y cálculo de intervalos de confianza. Por último las variables f18_1 y f18_2 están también relacionadas, al tratarse de la identificación de las probabilidades iniciales y las verosimilitudes.

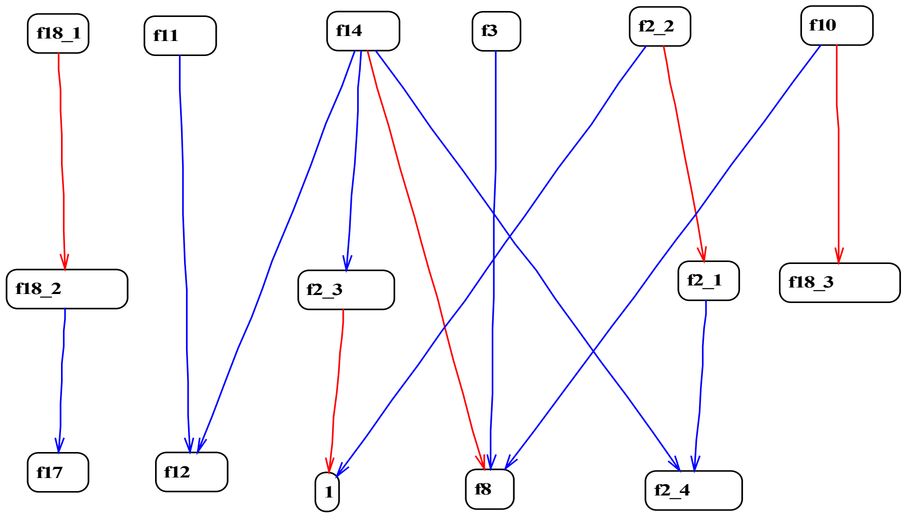
Para analizar la interrelación entre objetivos de aprendizaje mostramos los índices de implicación entre ítems de la prueba AIB (tablas 6.61 y 6.62), así como el grafo implicativo (figura 6.16), donde se presentan en rojo las relaciones significativas al 99% y en azul las significativas al 95%.

Tabla 6.62. Índices de implicación (teoría clásica)

	f11	f12	f13	f14	f15	f16	f17	f18_1	f18_2	f18_3	f20_1;f20_2
1	36	43	45	69	54	34	57	50	58	50	58
f2_1	57	32	40	67	45	45	14	10	24	80	49
f2_2	74	47	37	74	47	47	27	19	33	62	37
f2_3	53	86	30	86	76	12	50	58	80	50	65
f2_4	55	80	48	79	60	49	25	27	49	73	45
f3	33	42	44	65	31	54	60	58	63	31	40
f4	60	50	48	58	20	61	47	39	54	51	82
f5	59	25	32	39	35	35	9	18	21	80	61
f6_1	52	66	52	65	66	31	27	45	34	46	40
f6_2	46	60	41	43	60	51	47	63	67	45	33
f7	68	56	44	68	56	43	52	35	54	53	55
f8	67	56	51	70	46	56	36	40	78	60	49
f9	41	45	43	52	35	45	36	34	33	71	42
f10	71	66	44	73	54	42	42	58	87	95	77
f11	0	94	28	68	43	32	35	35	54	37	55
f12	85	0	37	79	85	25	75	86	89	49	51
f13	19	23	0	35	23	69	30	37	21	25	16
f14	73	95	38	0	88	64	49	62	77	65	81
f15	45	85	37	73	0	45	59	78	55	49	51
f16	37	25	58	58	45	0	59	43	41	49	63
f17	43	65	44	50	56	56	0	96	86	34	49
f18_1	40	83	45	56	75	44	100	0	99	30	29
f18_2	52	82	39	62	54	43	92	97	0	37	47
f18_3	43	50	41	56	50	50	32	34	38	0	64
f20_1;f20_2	53	51	33	67	51	62	48	28	47	68	0

Hacemos notar que la relación de implicación es asimétrica, indicándose el sentido de la implicación por la dirección de la flecha en el grafo. Por tanto, si estudiamos las relaciones significativas al 99%, que aparecen en rojo en la figura 6.16, la interpretación es que los alumnos que realizan correctamente el ítem f18_1 tienen más probabilidad de realizar el f18_2, es decir el estudiante que identifica correctamente las probabilidades iniciales, tiene mayor probabilidad de identificar correctamente las verosimilitudes. También la correcta realización de f10 (identificar probabilidades y valores críticos en una tabla y cálculo de intervalo de credibilidad) facilita el cálculo de probabilidades finales usando el teorema de Bayes (f18_3). Finalmente el cálculo correcto de probabilidades condicionales facilita del de las simples y conjuntas (f2_1, f2_2 y f2_4).

Figura 6.16. Grafo implicativo con nodos significativos al 99 y 95%



Al analizar las implicaciones significativas el 95%, que aparecen en azul en el grafo (figura 6.16) observamos que el alumno que realiza correctamente el proceso de contraste de hipótesis bayesiano (f14), tiene mayor probabilidad de realizar correctamente el cálculo del intervalo de credibilidad (f12), y éste a su vez calculará mejor las probabilidades condicionales (f2_3 y f2_4) lo que a su vez facilita la asignación correcta de distribuciones iniciales informativas (f8).

Todos estos resultados aportan evidencias a favor de nuestra hipótesis previa de la importancia de comprensión de la probabilidad condicional para llegar a dominar los métodos bayesianos.

Finalmente completamos el estudio con la determinación de una jerarquía implicativa en el conjunto de variables, presentando en la tabla 6.63 los coeficientes de cohesión en los distintos pasos del procedimiento y en la figura 6.17 el árbol de cohesión implicativa. En el grafo de cohesión implicativa obtenemos cuatro grupos significativos

- Grupo 1. Ítems (f2_2), y (f2_1) que a su vez unen con (f2_4), todos ellos relacionados con la probabilidad. El alumno que realiza el cálculo de probabilidades condicionales (f2_2), realiza el cálculo de probabilidades simples (f2_1) y compuestas (f2_4). Se confirma así la dificultad relativa de la probabilidad condicional respecto a las simples y compuestas.

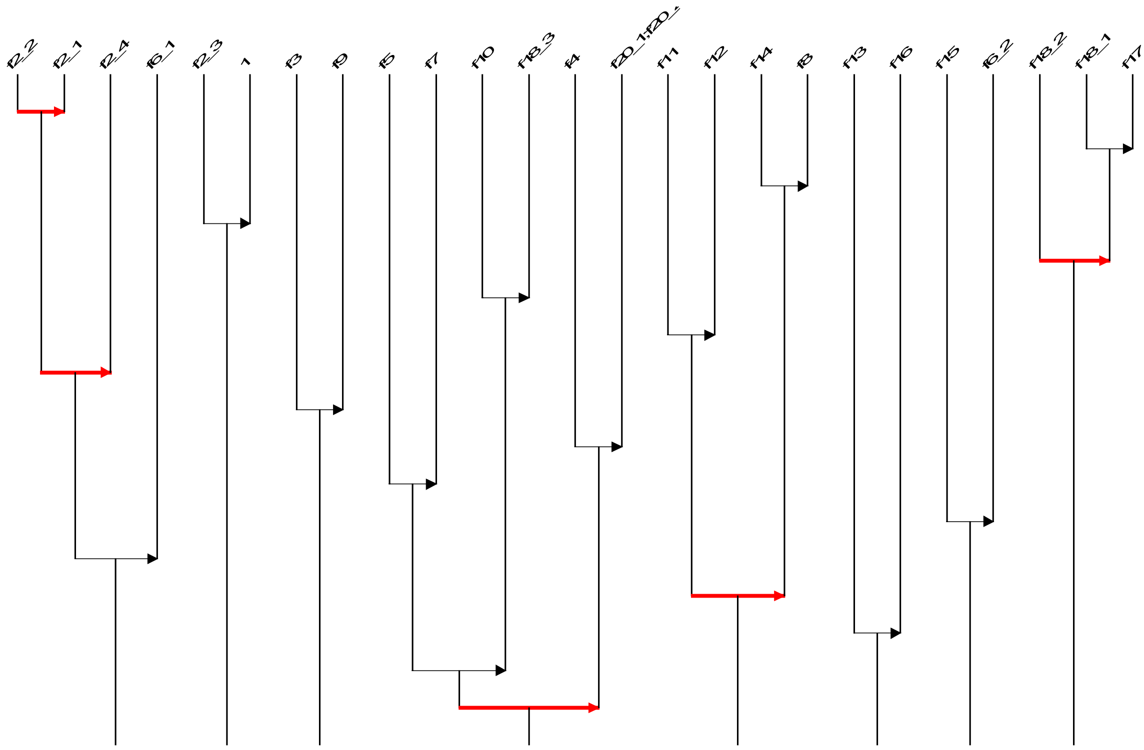
- Grupo 2. Actividades relacionadas con la lectura de datos en tablas de doble entrada, que fue otro tema en que algunos estudiantes mostraron dificultades en el cuestionario de probabilidad condicional. Dentro de estas actividades de lectura de tablas, el que identifica las verosimilitudes (f18_1), que es una probabilidad condicional a partir del enunciado de un problema ha identificado correctamente las probabilidades iniciales (f18_2), que son probabilidades simples así como lee correctamente las tablas de datos sobre medias y varianzas, actividad de no requiere calcular probabilidades (f17).

Tabla 6.63 Niveles en la construcción de la clasificación implicativa

Nivel de clasificación	Ítems que agrupa	Coef. cohesión
1**	(f2_2 f2_1)	1
2	(f18_1 f17)	1
3	(f14 f8)	0,984
4	(f2_3 1)	0,974
5**	(f18_2 (f18_1 f17))	0,966
6	(f10 f18_3)	0,962
7	(f11 f12)	0,941
8**	((f2_2 f2_1) f2_4)	0,897
9	(f3 f9)	0,865
10	(f4 f20_1;f20_2)	0,723
11	(f5 f7)	0,649
12	(f15 f6_2)	0,616
13	(((f2_2 f2_1) f2_4) f6_1)	0,604
14**	((f11 f12) (f14 f8))	0,601
15	(f13 f16)	0,451
16	((f5 f7) (f10 f18_3))	0,386
17**	(((f5 f7) (f10 f18_3)) (f4 f20_1;f20_2))	0,287

** Significativos al 99%

Figura 6.17. Árbol de cohesión implicativa con grupos estadísticamente significativos a nivel 95%



- Grupo 3. Resolución de problemas bayesianos típicos. El grafo también asocia las variables de cálculo de intervalos de credibilidad (f12) con el contraste de hipótesis (f11 y f14) y con la asignación correcta de una distribución inicial (f8). Todos los ítems de este grupo corresponden a conocimientos específicos del método bayesiano y en todos ellos intervienen probabilidades condicionales.
- Finalmente aparecen unidos algunos otros ítems conceptuales referidos a conceptos bayesianos, que no se relacionan tan directamente con la probabilidad condicional (parámetro, f5 y algoritmos de cálculo de probabilidades finales (f7, f10 y f18_3)) que a su vez implican a los más sencillos, como el concepto de distribución inicial o su correcta asignación (f4, f20_1 y f20_2).

4.4. DISCUSIÓN DEL ESTUDIO 9

Los diferentes análisis llevados a cabo nos aportan evidencias empíricas a favor de nuestra hipótesis de partida sobre relación entre comprensión de la probabilidad condicional y el aprendizaje de conceptos elementales de inferencia bayesiana.

Las correlaciones de Pearson y Spearman obtenidas entre las puntuaciones totales en el cuestionario RPC y la prueba AIB fueron estadísticamente significativas ($p<0,1$) y aunque su magnitud es moderada supera el criterio mínimo admitido de validez predictiva, por lo que, además, proporciona nuevas evidencias de validez para el cuestionario RPC en cuya definición sintáctica hipotetizamos sobre la correlación con la prueba AIB como criterio externo.

Por otro lado, las correlaciones de Pearson y Spearman entre algunos ítems del cuestionario RPC y la prueba AIB fueron estadísticamente significativas ($p<0,1$) y aunque su magnitud es moderada supera el criterio mínimo admitido de validez discriminante, lo que sugiere que estos ítems tienen propiedades discriminantes respecto al aprendizaje de conceptos elementales de inferencia bayesiana. Hacemos notar que tales ítems incluyen precisamente elementos de la probabilidad condicional, tales como independencia, teorema de Bayes, regla del producto, etc., que son particularmente relevantes en la resolución de problemas de inferencia bayesiana.

Por otro lado, el hecho de que otros ítems referidos a sesgos como la falacia de la condicional transpuesta u otros sesgos ampliamente extendidos entre el alumnado, incluidos aquellos que muestran comprensión de la probabilidad condicional, hacen que la correlación de la puntuación total en RPC no sea mayor que la observada. Otro factor que influye en la débil correlación entre estas dos pruebas es el éxito en la enseñanza de inferencia bayesiana, incluso en alumnos con razonamiento condicional insuficiente. Es decir, la homogeneidad en la alta puntuación de la prueba AIB resta intensidad a su correlación con el cuestionario RPC, a pesar de que los alumnos con mayor puntuación en probabilidad condicional también la tienen en AIB.

El estudio de clasificación cohesiva, el análisis implicativo y la clasificación implicativa realizadas confluyen en mostrar la interrelación entre ítems de la prueba AIB en los cuáles los alumnos han de identificar o calcular probabilidades condicionales y compuestas o bien aplicar el teorema de Bayes y otros relacionados con la construcción de intervalos de credibilidad, realización de contrastes de hipótesis bayesianos, cálculo de probabilidades finales y asignación correcta de distribuciones iniciales informativas.

En el caso del estudio implicativo la dirección de la implicación es precisamente la esperada por nosotros, puesto que la resolución correcta de problemas de inferencia bayesiana o el dominio de conceptos bayesiano supone otros previos de probabilidad condicional, conjunta y teorema de Bayes.

Por otro lado, las clases obtenidas, tanto en el análisis jerárquico como en el implicativo nos proporcionan información sobre conceptos cuya comprensión está relacionada y la dificultad relativa de los mismos dentro de cada grupo. Esto supone una ayuda potencial para la mejora del material didáctico preparado para los estudios 8.1 y 8.2 y en general, para la organización de la enseñanza de estos temas.

Finalmente resaltamos que los resultados de este estudio justifican el interés de la evaluación del razonamiento sobre probabilidad condicional de los estudiantes y por tanto de la construcción del cuestionario RPC que hemos llevado a cabo en la primera parte de nuestro trabajo, y que puede usarse como instrumento diagnóstico por los profesores para detectar las dificultades de sus alumnos sobre este razonamiento y prepararlos para superarlos y abordar con éxito el aprendizaje de la inferencia bayesiana.

Asimismo en este estudio proporcionamos razones para no suprimir el estudio de la probabilidad condicional en los cursos de análisis de datos en psicología, no sólo por su interés en si mismo- toma de decisiones, diagnóstico- sino como base para el estudio futuro de la inferencia bayesiana.

CAPITULO 7.
CONCLUSIONES

1. INTRODUCCIÓN

Finalizada la descripción y análisis de los diferentes estudios que configuran la memoria se dedica este capítulo a la exposición de las principales conclusiones.

En primer lugar, se discute el grado de consecución de los objetivos planteados para el trabajo y se analizan las conclusiones alcanzadas respecto a cada una de las hipótesis iniciales expuestas ambos en el Capítulo 2.

A continuación se describen las principales aportaciones de la investigación a la metodología de las ciencias del comportamiento, tanto desde el punto de vista del análisis y aplicación de los métodos bayesianos en el proceso de construcción y análisis de datos de cuestionarios, como de la didáctica de dichos métodos.

Dedicamos también una sección al análisis de las limitaciones del estudio, que permitan valorar mejor al lector el alcance de nuestros resultados.

Finalmente se sugieren algunas posibles líneas para continuar la investigación descrita en esta Memoria.

2. CONCLUSIONES SOBRE LOS OBJETIVOS

En lo que sigue se analizan los resultados y conclusiones alcanzados respecto a los objetivos expuestos en el Capítulo 2.

- *Objetivo 1. Reformular la Teoría Clásica de los Tests (TCT) desde un punto de vista bayesiano, analizando las implicaciones de este cambio de perspectiva sobre los métodos de estimación de algunas características psicométricas de los ítems y del cuestionario en su conjunto.*

En el Capítulo 3, partiendo de un análisis de los objetivos y métodos en inferencia bayesiana se analizaron las implicaciones de la perspectiva bayesiana sobre los supuestos de la TCT siguiendo, entre otros, los textos de Box y Tiao (1992), Bernard (1994), Berry (1995), Bernardo (2003) y Bolstad (2004). Una formulación consistente de una perspectiva bayesiana, llevó a mantener los principales supuestos de la TCT, siendo el cambio principal, el considerar variables aleatorias la puntuación verdadera en el test, así como los diferentes parámetros que definen su distribución.

En consecuencia, el objetivo principal de la inferencia se convierte ahora en la actualización de las distribuciones iniciales de los diferentes parámetros que definen tanto la puntuación verdadera en el test como las características de los ítems que la componen.

Abordar este estudio de una forma completa sería una tarea excesivamente amplia. En la memoria se comenzó por el estudio de la estimación de puntuaciones medias y diferencia de puntuaciones medias en el test, índices de dificultad y discriminación en los ítems. Para cada uno de ellos se estudiaron los casos de distribución inicial informativa y no informativa, comparando los resultados con la estimación frecuencial y preparando programas que permitan realizar los cálculos. En el caso de la media y diferencia de medias, se diferenciaron los casos de varianzas conocidas y desconocidas, en este último caso, varianzas iguales y diferentes.

- *Objetivo 2. Aplicar el análisis anterior en el proceso de construcción de un cuestionario, comparando los resultados de la estimación clásica y bayesiana de algunas de las características psicométricas citadas.*

La construcción del cuestionario RPC, comenzó con la definición semántica de la variable objeto de medición mediante análisis de contenido de textos de estadística dirigidos a estudiantes de psicología[1]. El proceso de construcción se recoge en el Capítulo 4 y siguió las normas psicométricas (APA, AERA y NCME, 1985; 1999). Se realizaron pruebas piloto de ítems, juicio de experto para fijar el contenido y seleccionar los ítems, pruebas piloto y nuevo juicio de experto para mejorar la redacción de los ítems. Las muestras de estudiantes oscilaron entre 49 y 117 casos, según ítem. En la prueba piloto del instrumento participaron 57 alumnos de psicología y 39 de Matemáticas en el estudio de discriminación.

Asimismo se realizó un proceso de validación (contenido, criterio y constructo) y un estudio de fiabilidad (consistencia interna, test- retest, generalizabilidad) con una

[1] El análisis de los textos se ha presentado en un grupo de trabajo sobre didáctica de la probabilidad en el Congreso Iberoamericano de Investigación Matemática (Díaz y Batanero, 2005).

muestra de 591 estudiantes de psicología de diversas universidades españolas (Capítulo 6). Los resultados indicaron una alta fiabilidad y proporcionaron evidencias suficientes de validez del instrumento[2]. El estudio de validación de constructo sugirió la existencia de diferentes factores en el razonamiento sobre probabilidad condicional.

Todo este proceso se complementó con la aplicación de métodos bayesianos en la estimación de las puntuaciones medias del cuestionario, diferencias de puntuaciones medias en diferentes grupos, índices de dificultad y discriminación.

Siguiendo el método bayesiano, se comenzó con distribuciones iniciales no informativas en las primeras fases de elaboración del instrumento (análisis de ítems para las características psicométricas de éstos; prueba piloto para el estudio de la puntuación global en el cuestionario). En las etapas sucesivas se usaron como distribuciones iniciales las distribuciones finales obtenidas en las etapas anteriores, consiguiendo con ello una mayor precisión, respecto a la inferencia clásica.

- *Objetivo 3. Realizar un estudio de evaluación del razonamiento sobre probabilidad condicional de los estudiantes de psicología con el cuestionario RPC para decidir sobre la viabilidad de la enseñanza de inferencia bayesiana a estos estudiantes.*

Construido el cuestionario, se aplicó a una muestra de 414 estudiantes de Psicología de primer curso de diferentes universidades[3]. En el Capitulo 7 se presentan los resultados del estudio de evaluación, organizándose el análisis de los ítems, respecto a los factores identificados en el estudio previo de validación de constructo.

Dichos factores son agrupados en dos categorías:

a) Conocimientos lógico –matemáticos, que se agrupan en tres factores: 1) Resolución de problemas complejos de probabilidad condicional; b) lectura de tablas y discriminación entre probabilidades simples, compuestas y condicionales; c) comprensión conceptual de la probabilidad condicional.
b) Sesgos en el razonamiento sobre probabilidad condicional, no relacionados con el conocimiento lógico matemático, que fueron los siguientes: 4) Falacia del eje

[2] Presentados parcialmente en el Simposio Nacional de Estadística en Colombia y Séptimo Congreso International de Enseñanza de la Estadística (Díaz y de la Fuente, 2005d; 2006).

[3] Resultados parciales fueron publicados en Díaz, de la Fuente y Batanero (2004b), Díaz (2005); Díaz y de la Fuente (2005c).

de tiempo y causación; 5) falacia de la conjunción; 6) falacia de la condicional transpuesta y 7) independencia y falacia de las tasas base.

Los resultados del estudio indicaron una buena capacidad lógico- matemática y un conocimiento suficiente de los estudiantes para abordar la enseñanza de la inferencia bayesiana. Por otro lado se observaron que los sesgos de razonamiento descritos era generalizados y no mejoran con el mayor conocimiento lógico matemático, lo que hacía necesario tenerlos en cuenta en el diseño del material de enseñanza sobre inferencia bayesiana.

- *Objetivo 4. Preparar materiales didácticos para la enseñanza de inferencia bayesiana a estudiantes de psicología que se apoyen en el estudio de evaluación previo y evaluar su efectividad*

Se elaboró un material didáctico, tanto en versión escrita, como para ser utilizada en enseñanza no presencial (véase www.ugr.es/local/mcdiaz/bayes) que fue revisado un grupo de expertos. El material tuvo en cuenta el estudio previo de evaluación, el análisis del tema en libros destinados a estudiantes de primeros cursos de universidad, así como una serie de principios didácticos expuestos en el Capítulo 6. Cubre el teorema de Bayes, estimación de la proporción (caso discreto y continuo) estimación de la media (caso continuo). Para cada tema incluye objetivos, teoría, ejercicios resueltos, ejemplos, actividades y pruebas de autoevaluación. El material fue complementado con un cuestionario AIB de evaluación de los conocimientos adquiridos que cubre los objetivos de aprendizaje.

Todo este material fue sometido a prueba en un experimento de enseñanza llevado a cabo con 78 estudiantes de psicología que combina las sesiones en aula tradicional y el trabajo en aula de informática con programas Excel preparados para la enseñanza. Los datos recogidos de las pruebas de autoevaluación, actividades y cuestionario AIB son analizados y muestran la viabilidad de la enseñanza para los estudiantes a los que va dirigida.

3. CONCLUSIONES SOBRE LAS HIPÓTESIS

A continuación se discuten las conclusiones obtenidas respecto las hipótesis iniciales de la investigación

3.1. HIPÓTESIS DEL ESTUDIO DE EVALUACIÓN

Hipótesis 1. *Los alumnos en Psicología adquieren suficientes conocimientos lógico – matemáticos sobre probabilidad condicional, condición necesaria para abordar el estudio de la inferencia bayesiana*[4].

El análisis factorial exploratorio llevado a cabo sobre las respuestas al cuestionario RPC de una muestra de 591 estudiantes de psicología como parte del estudio de validación del cuestionario sugirió la existencia de dos grupos de factores independientes. El primero de estos grupos, formado por los tres primeros factores, que explicaron en su conjunto el 34% de la inercia está definido por ítems (que contribuyen con pesos fuertes) relacionados con la resolución de problemas de probabilidad condicional (incluido la probabilidad total y Bayes), la lectura de tablas y cálculo de probabilidades simples, compuestas y condicionales y la comprensión conceptual (definición, restricción del espacio muestral, discriminación entre situaciones de muestreo con y sin reemplazamiento).

Los índices de dificultad en los ítems del cuestionario RPC que contribuyen a estos factores fueron altos. El porcentaje de estudiantes que resolvieron correctamente los problemas osciló entre en 54% y 70%; los que leyeron correctamente las tablas y calcularon las distintas probabilidades entre el 58% y 91%; 38% da una definición correcta de la probabilidad condicional y 65% restringen correctamente el espacio muestral; 88% diferencian situaciones con y sin reemplazamiento. Todo ello aporta evidencias a favor de nuestra hipótesis de que los conocimientos sobre probabilidad

[4] En el capítulo 1, sección 4.5 se justifica suficientemente la importancia de la comprensión de la probabilidad condicional para abordar el aprendizaje de la inferencia bayesiana.

condicional de los estudiantes de psicología que siguen el curso de análisis de datos son suficientes para abordar el éxito en el estudio de la inferencia bayesiana[5].

Posteriormente la hipótesis se ve reforzada de nuevo por los resultados obtenidos en el experimento de enseñanza, así como por la correlación obtenida entre las puntuaciones en los cuestionarios RPC y AIB en los estudiantes que participaron en el experimento de enseñanza.

Hipótesis 2. *Determinados sesgos de tipo psicológicos son resistentes a la enseñanza tradicional que se realiza en las universidades, por tanto, los alumnos, incluso después de estudiar el tema de probabilidad condicional, continúan presentando dichos sesgos.*

La hipótesis es que los sesgos se mantienen después de la enseñanza, incluso en alumnos con buena comprensión lógico – matemática. Se comprobó en el hecho de que los ítems relacionados con dichos sesgos aparecieron definiendo factores diferenciados – por tanto independientes de los ítems lógico matemáticos- en el estudio de validación. Asimismo, los resultados del estudio de evaluación sobre estos ítems mostraron la prevalencia de la falacia del eje de tiempos, sesgo de equiprobabilidad y concepciones incorrectas sobre la independencia Las falacias de la conjunción y tasas base no obstante fueron menos frecuentes de lo esperado[6].

3.2. HIPÓTESIS DEL ESTUDIO DIDÁCTICO

Hipótesis 3. *En un tiempo limitado de enseñanza es posible conseguir un aprendizaje de conocimientos elementales sobre inferencia bayesiana, dentro de una dinámica habitual de clase, con estudiantes de primer año de Psicología.*

---

[5] Precisamente uno de los argumentos para no introducir la inferencia bayesiana en los primeros cursos universitarios es la supuesta dificultad de la probabilidad condicional, en la que se basa dicha inferencia (Moore, 1997). Nuestro estudio contradice en cierto modo esta postura al mostrar que los estudiantes tienen conocimientos suficientes de probabilidad condicional para abordar dicho estudio.

[6] Resultados parciales se han publicado en Díaz, de la Fuente y Batanero (2004 b), Díaz (2005), Díaz y de la Fuente (2005 c).

Esta hipótesis se vio apoyada por los resultados del estudio 8.2 en el cual se evaluó un experimento de enseñanza llevado a cabo en una muestra de 78 estudiantes de psicología. A lo largo de 12 horas y apoyándose en el material didáctico preparado al efecto, que se describe en el estudio 8.1 se introdujeron el teorema de Bayes y la inferencia para la proporción (casos discreto y continuo) y media (caso continuo).

Los resultados obtenidos con el cuestionario AIB (que evalúa el aprendizaje de los objetivos planteados en el estudio 8.1 fueron altamente satisfactorios, con un 64% de alumnos que alcanzan en promedio cada uno de los objetivos instruccionales (rango de variación entre 31 y 91% con valores de cuartiles 55 y 75% respectivamente). Los resultados en las pruebas de evaluación intermedias y actividades en el aula de informática, en las que el alumno podría consultar el material didáctico fueron aún mejores (promedio 75% de alumnos alcanzan cada objetivo /resuelven cada tarea; rango entre 23 y 93% cuartiles 58 y 90%).

Todo ello refuerza, por un lado la hipótesis de que es posible introducir ideas elementales de inferencia bayesiana a estos estudiantes, al mismo tiempo que la hipótesis anterior, de que los conocimientos sobre probabilidad condicional de estos estudiantes eran suficientes para llevar a cabo este aprendizaje.

Hipótesis 4. *Las principales dificultades que se esperan encontrar en los estudiantes, tras la instrucción, no serán específicas de la inferencia bayesiana, sino que serán de tipo lógico o relacionadas con otros temas.*

De hecho, los objetivos alcanzados en menor grado, tanto en las pruebas intermedias como en la evaluación final del aprendizaje llevada a cabo en el estudio 8.2, se refieren a los siguientes temas, no específicos de la inferencia bayesiana:

- Diferenciar una probabilidad condicional y su transpuesta (36% en la evaluación final y 39% en la continua). Este es uno de los sesgos sobre razonamiento condicional encontrados en el estudio de evaluación y los resultados sugieren que se debe insistir más en este punto.
- Diferenciar entre probabilidad condicional y conjunta en el contexto de un problema (22% en la evaluación intermedia, aunque esta dificultad desaparece en la

evaluación final donde el 62% de los estudiantes diferencian estas probabilidades correctamente).

- Dificultad en comprender la lógica del razonamiento por contradicción durante la evaluación intermedia (respuestas correctas sólo en 40% de alumnos). A pesar de ello la mayor parte de los alumnos (55 y 58%) lleva a cabo correctamente el contraste de hipótesis sobre una proporción en las evaluaciones intermedia y final y sobre la media (64 y 60% en la evaluación intermedia).
- Olvido de la fórmula de la desviación típica del estimador de la media (35%) y algunos errores al despejar variables en inecuaciones durante las evaluaciones intermedias (43%). Estas dificultades persisten en la evaluación final y hace que solo en 30% lleven a cabo correctamente el contraste de hipótesis sobre la media.
- El único problema que aparece en la resolución de problemas a lo largo de la enseñanza fue dar el valor de la mejor estimación de la proporción (solo lo dan el 28%) pues los alumnos confundieron la distribución inicial y final, aunque este problema se supera en la evaluación final. El resto de los problemas a resolver resultó muy sencillo para los alumnos.

Hipótesis 5. *Se espera una correlación entre los resultados de los estudiantes en el cuestionario RPC y la evaluación del aprendizaje de inferencia bayesiana.*

Esta hipótesis se vio apoyada en el estudio 9. Por un lado se encontraron correlaciones ordinales y de Pearson estadísticamente significativas entre las puntuaciones en los cuestionarios AIB y RPC en los estudiantes participantes en el experimento de enseñanza. Asimismo hubo correlaciones estadísticamente significativas entre la puntuación en AIB y la puntuación en los problemas de probabilidad condicional, incluidos los relacionados con el Teorema de Bayes, independencia, probabilidad compuesta y regla del producto.

Las correlaciones fueron todas positivas, aunque de pequeña magnitud, seguramente porque, además de la probabilidad condicional, otros muchos conceptos previos intervienen en la comprensión de la inferencia bayesiana.

Por otro lado, la comprobación de esta hipótesis aporta nuevas evidencias de validez predictiva al cuestionario RPC, al utilizar la prueba de Aprendizaje de Inferencia Bayesiana (AIB) como criterio externo.

4. APORTACIONES DEL TRABAJO

La investigación presentada realiza diversas aportaciones al área de Metodología de las Ciencias del Comportamiento, que se describieron brevemente en la introducción de la Memoria.

El primer bloque de aportaciones lo constituye la síntesis realizada en el Capítulo 1 que justificación de la investigación y consta de tres partes esenciales:

- Problemática del uso actual de la estadística (Sección 2 del Capítulo 1).
- Razones que justifican la inferencia bayesiana como una alternativa interesante y razonable, tanto desde el punto de vista metodológico como didáctico (Sección 3 del Capítulo 1).
- Investigaciones sobre razonamiento en probabilidad condicional (Sección 3 del Capítulo 1).

Un segundo bloque de aportaciones consiste en una reinterpretación de la Teoría Clásica de los Tests desde un punto de vista Bayesiano (Capítulo 3), conservando los supuestos esenciales de esta teoría, pero considerando los parámetros del tests e ítems como variables aleatorias.

Como consecuencia se aplican algunos métodos bayesianos de estimación a la determinación de las distribuciones finales de los índices de dificultad y discriminación, puntuación media del test y diferencia de puntuaciones (diferenciando los casos de varianza conocida y desconocida y dentro de ella de varianzas iguales y diferentes), tanto en el caso de distribución inicial no informativa como informativa.

Se comparan para cada uno de estos parámetros y casos los resultados de la metodología clásica y bayesiana y se ponen a punto subrutinas de cálculo para aplicar cómodamente los métodos bayesianos a la estimación de estos parámetros. Dichos resultados y programas se han aplicado a lo largo del trabajo, en el proceso de construcción y validación del cuestionario RPC, así como en los procesos de evaluación del aprendizaje de la inferencia bayesiana descritos en el Capítulo 6.

El tercer bloque de aportaciones lo constituye la construcción del cuestionario RPC, que se organiza mediante 4 estudios consecutivos:

- Estudio 1. Definición semántica de la variable objeto de medición para lo cual se realizó un análisis de contenido de los libros de texto usados en la asignaturas "Análisis de datos" en las facultades de Psicología españolas.
- Estudio 2. Se compone de:
    - Estudio 2.1. Elaboración del banco inicial de ítems y ensayos piloto de ítems con alumnos de Psicología.
    - Estudio 2.2. Juicio de expertos (investigadores en didáctica de la probabilidad) para validar la tabla de especificaciones de la variable y seleccionar los ítems del cuestionario. Con ello se justifica la validez de contenido del cuestionario.
- Estudio 3. Prueba piloto del cuestionario, estimando características psicométricas de los ítems y aportando una primera estimación de la fiabilidad de consistencia interna.
- Estudio 4. Revisión del instrumento mediante juicio de expertos (metodólogos). Con ello se completa la construcción del cuestionario.

El proceso de validación del cuestionario RPC se compone de los estudios 5 y 6

- El estudio 5 aborda la estimación de la fiabilidad mediante los métodos de consistencia interna, prueba repetida y generalizabilidad.
- Estudio 6. Aborda la validación del cuestionario con un estudio de validez referida a criterio y validez de constructo.

Con todo este proceso se finaliza con un instrumento de evaluación válido y fiable que tiene utilidad tanto para la enseñanza como para la investigación sobre la probabilidad condicional.

El último bloque de aportaciones lo constituye el material de enseñanza fundamentado en todos los estudios anteriores y su evaluación. Consta de los siguientes estudios:

- Estudio 7. Evaluación del razonamiento sobre probabilidad condicional en una muestra de estudiantes de psicología de diversas universidades, una vez estudiado el tema y categorización de las principales dificultades encontradas, que se tuvieron en cuenta en la elaboración del material.
- Estudio 8. Se compone a su vez de dos partes:
  - Estudio 8.1. Elaboración del material didáctico que incluye apuntes para los estudiantes (versión presencia y a distancia), ejemplos y ejercicios, actividades de autoevaluación, problemas abiertos para resolver y programas auxiliares de cálculo.
  - Estudio 8.2. Experimentación de este material en un grupo de estudiantes con un total de 12 horas de enseñanza, análisis de las tareas realizadas por los alumnos durante la experiencia y de los resultados finales de una prueba de evaluación. Se muestra la consecución de la mayor parte de los objetivos educativos.
- Estudio 9. Se correlaciona las puntuaciones de los alumnos participantes en la prueba de aprendizaje de inferencia bayesiana (AIB) y el cuestionario RPC y se aportan nuevas evidencias de validez para el cuestionario RPC.

Todas estas aportaciones se acompañan de una serie de publicaciones referidas en revistas incluidas en el catálogo Latindex y congresos internacionales, incluidos algunos trabajos invitados que se incluyen en la lista de referencias.

5. LIMITACIONES DEL TRABAJO Y FUTURAS LINEAS DE INVESTIGACIÓN

La investigación llevada a cabo no está finalizada, puesto que la necesidad de acotar en un tiempo dado el problema de investigación ha llevado, necesariamente a limitar algunas de sus características. En este sentido se está iniciando la traducción al portugués con el propósito de llevar a cabo estudios comparativos de evaluación con estudiantes de psicología en otros países; en particular en Portugal y Brasil, en colaboración con investigadores que se han interesado por la temática.

Los participantes son alumnos de primer curso de psicología de cuatro universidades españolas y han sido elegidos intencionalmente. Sería interesante repetir los estudios de evaluación y de enseñanza con otros estudiantes de psicología o incluso con estudiantes de otras especialidades. Asimismo una parte restringida del cuestionario se está utilizando para evaluar el razonamiento en probabilidad condicional en maestros de enseñanza primaria en formación, este último trabajo en colaboración con la Facultad de Educación de Lleida[7].

Sería también necesario repetir la experiencia de enseñanza y ampliarla al estudio de otros temas, como la estimación de diferencias de medias y proporciones. En este sentido pensamos traducir el material docente al portugués, como parte de un proyecto de innovación docente realizado en colaboración con la Facultad de Psicología de la Universidad de Lisboa en el marco del Espacio Europeo de Educación Superior.

La última línea de investigación en la que pensamos seguir trabajando es la aplicación de la inferencia bayesiana en psicometría. Más concretamente, pensamos continuar aplicando este tipo de inferencia a la estimación de otras características de los cuestionarios, dentro de la TCT, así como en la teoría de generalizabilidad y teoría de respuesta al ítem.

---

[7] Algunos resultados han sido publicados en Estrada y Díaz (2006)

CAPÍTULO 8.

ENGLISH SUMMARY: SUITABILITY OF TEACHING BAYESIAN INFERENCE IN DATA ANALYSIS COURSES DIRECTED TO PSYCHOLOGISTS

1. INTRODUCTION

In this Thesis we focus on the use of Bayesian inference in the field of psychology from different perspectives:

1. The reflection on the current statistical practices in psychology, the reported errors and possible contribution of Bayesian inference to solve these problems. This analysis is carried out from the philosophical and psychological points of views (Chapter 1).
2. The study of some applications of Bayesian methods in psychometrics to estimate different indicators used in the Classical Tests Theory. These possibilities are analysed from the theoretical (Chapter 3) and practical (Chapters 4, 5) points of view and are applied in the process of building a questionnaire to assess conditional probability reasoning (CPR), which is also justified in the thesis.
3. The feasibility of teaching basic Bayesian elements in undergraduate psychology courses. We develop a teaching material that takes into account the previous analyses, as well as previous research in statistics education and the type of students. This material was tested with a sample of 78 students, and data on the students' learning at the end of the experience are provided (Chapter 6).

Below we describe the aims and structure of the thesis and summarize the different studies included in the same.

2. RESEARCH AIMS AND STRUCTURE

There are four main goals in this research. For each of them we carry out one or more studies, which are related one to another as shown in Figure 8.1.

- *Objective 1. Rethinking the Classical Tests Theory CTT from the Bayesian point of view and analyzing the implications of this change of perspective on the estimation of some psychometric features in the tests and items.*

In Chapter 3 we analyse the implications of a Bayesian perspective on the estimation of tests mean scores, differences in mean scores, difficulty and discrimination indexes. For each of these parameters we consider both informative and non informative priors and prepare some Excel programs to carry out the computations of posterior distributions and credibility intervals. Results are useful to build and adapt other questionnaires, in particular when prior information for the psychometric features is available.

- *Objective 2. Applying the above analysis in the process of building a questionnaire and comparing results from classical and Bayesian estimates in some of the test features.*

The building of the CPR questionnaire starts from the semantic definition of the variable through content analysis of 18 statistics textbooks directed to psychology students (Studies 1-4; Chapter 4). The process follows the recommendations by APA, AERA and NCME (1999) and includes items trials, expert judgment to fix the content and select the items, pilot trial of the questionnaire and a second expert judgment to improve the items wording. Reliability and validity studies (Studies 5 and 6; Chapter 5) are carried out on different sample of students. All this process is complemented with application of Bayesian methods. The CPR questionnaire is useful to assess students' understanding of conditional probability in statistics courses and future research.

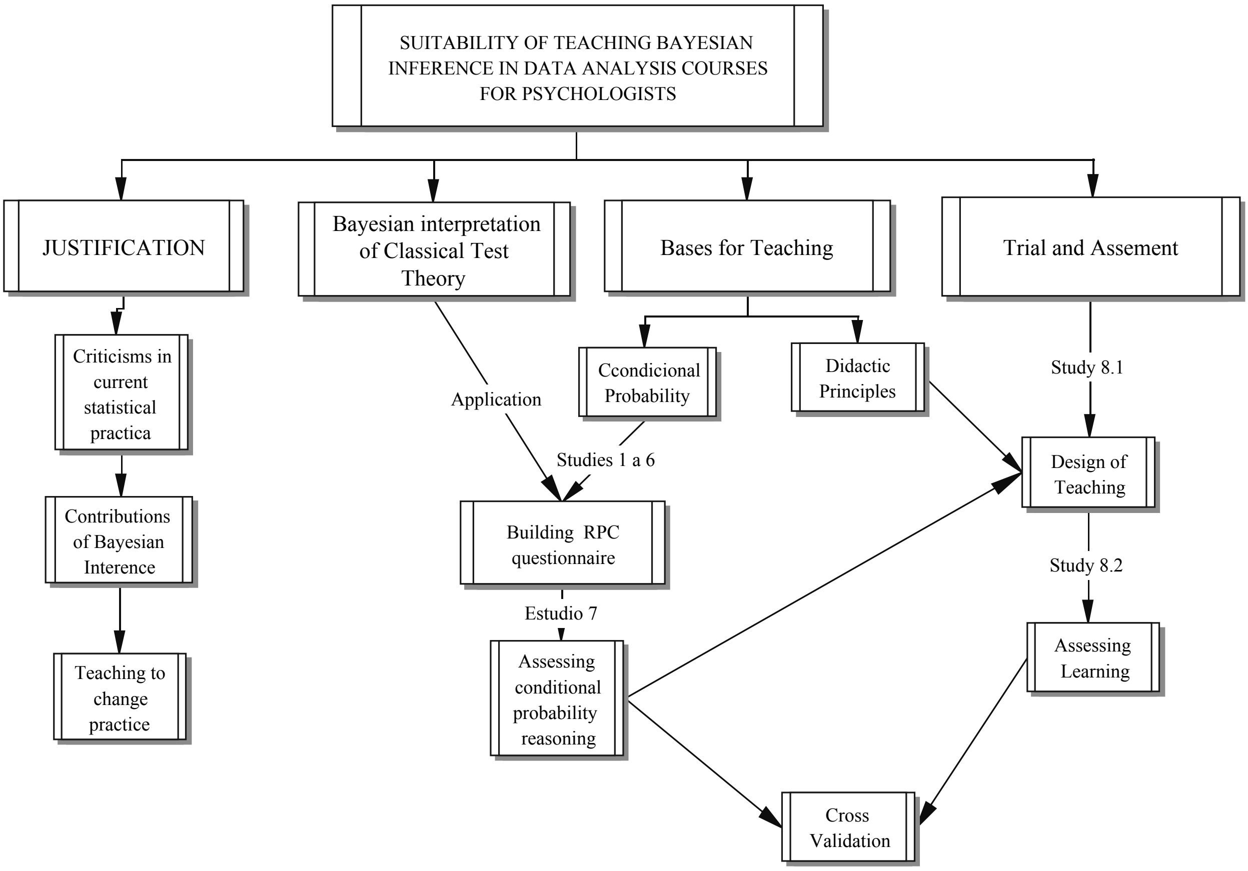
- *Objective 3. Assessing conditional probability reasoning in psychology students to decide the suitability of teaching Bayesian methods to these students.*

The RPC questionnaire is applied to a sample of 413 psychology students (Study 7) and their responses are analysed from different points of views. Students showed enough understanding of conditional probability to start the learning of Bayesian inference, but, at the same time, we found some widespread misconceptions that were taken into account in the next stage (designing a curricular proposal).

- *Objective 4. Preparing and assessing didactic materials to introduce elementary Bayesian inference to Psychology students that takes into account the previous assessment.*

The teaching materials are based on several textbooks of Bayesian inference and include activities, assessment questionnaires and Excel programs. It is available from the web page http://www.ugr.es/~mcdiaz/bayes/. An experiment is organized with a sample of 78 students (working in small groups) to try these materials (Studies 8 and 9). The posterior learning, structure of responses to assessment items and relationship with understanding conditional probability are analysed.

Figure 8.1. Research Structure



3. JUSTIFICATION

In Chapter 1 we present the foundations of the Thesis that can be classified in three main parts: a) Current situation in the practice of statistics inference and the need for a change; b) Possible contributions of Bayesian inference to improve the situation and need to include these methods in undergraduate courses; c) Relevance of assuring correct reasoning on conditional probability in the students before trying to teach them Bayesian inference and need for a comprehensive questionnaire to assess this reasoning (CPR questionnaire). In the following we summarize the main points in this justification.

3.1. CRITICISMS IN THE CURRENT PRACTICE OF STATISTICS IN EMPIRICAL RESEARCH

Empirical sciences heavily rely on establishing the existence of effects using the statistical analysis of data. Statistical inference dates back almost 300 years. However, since the logic of statistical inference is difficult to grasp, its use and interpretation are not always adequate and have been criticized for nearly 50 years (for example, in Morrison & Henkel, 1970; Harlow, Mulaik & Steiger, 1997). This controversy has increased in the past ten years within professional organizations (Menon, 1993; Ares, 1999; Wilkinson, 1999; Batanero, 2000; Frías, Pascual y García, 2000, Fidler, 2002; de la Fuente y Díaz, 2003, 2004), which are suggesting important shifts in their editorial policies regarding the use of statistical significance testing.

Despite the arguments that statistical tests are not adequate to justify scientific knowledge, researchers persist in relying on statistical significance (Hager, 2000; Borges, San Luis, Sánchez & Cañadas, 2001; Finch, Cumming & Thomason, 2001). Some explanations for this persistence include inertia, conceptual confusion, lack of better alternative tools, and psychological mechanisms such as invalid generalization from deductive logic to inference under uncertainty (Falk & Greenbaum, 1995). Below we summarize some of the problems that were analyzed in Batanero (2000) and Díaz and De la Fuente (2004).

Common Errors in Interpreting Statistical Tests

Misconceptions related to statistical tests mainly refer to the level of significance α, which is defined as the probability of rejecting a null hypothesis, given that it is true. The most common misinterpretation of this concept consists of switching the two terms in the conditional probability. For example, Birnbaum (1982) reported that his students found the following definition reasonable: *"A level of significance of 5% means that, on average, 5 out of every 100 times we reject the null hypothesis, we will be wrong"*. Falk (1986) found that most of her students believed that α was the probability of being wrong when rejecting the null hypothesis at a significance level α. Similar results were described in Pollard and Richardson (1987), Lecoutre, Lecoutre and Poitevineau (2001) and Haller and Krauss (2002) in their studies using researchers.

Another common error is the belief in the conservation of the significance level value when successive tests are carried out on the same data set, which produces the problem of multiple comparisons (Moses, 1992). Some people believe that the p-value is the probability that the result is due to chance. The p-value however is the probability of obtaining the particular result or one more extreme when the null hypothesis is true and there are no other possible factors influencing the result. What is rejected in a statistical test is the null hypothesis, and therefore we cannot infer the existence of a particular cause in an experiment from a significant result.

Another erroneous belief is that the .05 and .01 levels of significance are justified by mathematical theory. In his book "Design of Experiments", Fisher (1935) suggested selecting a significance level of 5% as a convention to recognize significant results in experiments. In later writings, however, Fisher considered that *"in fact, no scientific worker has a fixed level of significance at which from year to year and in all circumstances, he rejects hypotheses"* (Fisher, 1956, p. 42). Instead, Fisher suggested publishing the exact p-value obtained in each particular experiment which, in fact, implies establishing the significance level after the experiment. In spite of these recommendations, research literature shows that the common arbitrary levels of .05, .01 and .001 are almost universally selected for all types of research problems and are sometimes used as criteria for publication.

Misinterpretations of the significance level are linked to misinterpreting significant results; we should distinguish between statistical and practical significance, since we might have obtained a higher level of significance with a smaller experimental effect and a larger sample size. Practical significance involves statistical significance plus a sufficiently large experimental effect.

Philosophical and Psychological Issues

Several reasons explain the difficulties in understanding statistical tests. On one hand, statistical tests involve a series of concepts such as null and alternative hypotheses, Type I and Type II errors, probability of errors, significant and non significant results, population and sample, parameter and statistics, sampling distribution. Some of these concepts are misunderstood or confused by students and experimental researchers.

Moreover, the formal structure of statistical tests is superficially similar to that of proof by contradiction. However, there are fundamental differences between these two types of reasoning that are not always well understood. In proof by contradiction we reason in the following way: If A implies B cannot happen, then, if B happens, we deduce A is false. In statistical testing, it is tempting to apply similar reasoning as follows: If A implies B is very unlikely to happen. However, this does not imply that if B happens, A is very unlikely and here lays the confusion.

The controversy surrounding statistical inference involves the philosophy of inference and the logical relations between theories and facts. We expect from statistical testing more than it can provide us, and underlying this expectation is the philosophical problem of finding scientific criteria to justify inductive reasoning, as stated by Hume. The contribution made by statistical inference in this direction is important but it does not give a complete solution to this problem (Hacking, 1975; Cabria, 1994).

On the other hand, there are two different views about statistical tests that sometimes are confused or mixed. Fisher saw the aims of significance testing as confronting a null hypothesis with observations and for him a p-value indicated the strength of the evidence against the hypothesis (Fisher, 1956). However, Fisher did not believe that statistical tests provided inductive inferences from samples to population, but rather, a deductive inference from the population of possible samples to the particular sample obtained in each case.

For Neyman, the problem of testing a statistical hypothesis occurs when circumstances force us to make a choice between two courses of action. To accept a hypothesis means only to decide to take one action rather than another. This does not mean that one necessarily believes that the hypothesis is true. For Neyman and Pearson, a statistical test is a rule of inductive behaviour, a criterion for decision-making, which allows us to accept or reject a hypothesis by assuming some risks.

The dispute between these authors has been hidden in applications of statistical inference in psychology and other experimental sciences, where it has been assumed that there is only one statistical solution to inference (Gingerenzer et al, 1989). Today, many researchers apply the statistical tools, methods, and concepts of the Neyman-Pearson theory with a different aim, namely, to measure the evidence in favour of a given hypothesis. Therefore, the current practice of statistical tests contains elements from Neyman-Pearson (it is a decision procedure) and from Fisher (it is an inferential procedure, whereby data are used to provide evidence in favour of the hypothesis), which apply at different stages of the process. We should also add that some researchers often give a Bayesian interpretation to the result of (classical) hypothesis tests, in spite of the fact that the view from Bayesian statistics is very different from the theories of either Fisher or Neyman and Pearson.

Moreover, biases in inferential reasoning can be seen simply as examples of adults' poor reasoning in probabilistic problems (Nisbett & Ross, 1980; Kahneman, Slovic & Tversky, 1982). In the specific case of misinterpreting statistical inference results, Falk and Greenbaum (1995) describe the *illusion of probabilistic proof by contradiction*, which consists of the erroneous belief that one has rendered the null hypothesis improbable by obtaining a significant result. Misconceptions around the significance level are also related to difficulties in discriminating between the two directions of conditional probabilities, otherwise known as *the fallacy of the transposed conditional*. Although α is a well defined conditional probability, the expression "Type I error" is not conditionally phrased, and does not spell out to which combination of the two events it refers. This leads us to interpret the significance level as the conjunction of the two events "the null hypothesis is true" and "the null hypothesis is rejected" (Menon, 1993).

The Statistical Tests Controversy

For many years, criticisms have been raised against statistical testing, and many suggestions have been made to eliminate this procedure from academic research. However, significant results continue to be published in research journals, and errors around statistical tests continue to be spread throughout statistics courses and books, as well as in published research. An additional problem is that other statistical procedures suggested to replace or complement statistical tests (such as confidence intervals, measuring the magnitude of experimental effects, power analysis, and Bayesian inference) do not solve the philosophical and psychological problems we have described (see Fidler, 2002; Cumming, Williams & Fidler, 2004). Below we revisit some frequent criticisms that either are not justified or refer to researchers' use of statistical tests more than to the procedure itself.

Criticism 1. *The null hypothesis is never true and therefore statistical tests are invalid, as they are based on a false premise (that the null hypothesis is true).* This criticism is not pertinent because what is asserted in a test is that a significant result is improbable, given that the null hypothesis is true. This is a mathematical property of the sampling distribution that has nothing to do with the truth or falsity of the null hypothesis.

Criticism 2. *Statistical significance is not informative about the practical significance of the data, since the alternative hypothesis says nothing about the exact magnitude of the effect.* In significance testing (Fisher's approach) the aim of experimental research is directed towards theory confirmation in providing support for a substantive hypothesis and the magnitude of effect is not so important. In the context of taking a decision (Neyman- Pearson), however, the magnitude of the effect could be relevant to the decision. In these cases, the criticism applies and statistical tests should be complemented with power analysis and/ or estimates of the magnitude of the effects (Levin, 1998; Frías, Pascual & García, 2000; Vacha-Haase, 2001).

Criticism 3. *The choice of the level of significance is arbitrary; therefore some data could be significant at a given level and not significant at another different level.* It is true that the researcher chooses the level of significance. This arbitrariness does not, however, mean that the procedure is invalid. Moreover, it is also possible, following the approach of Fisher, to use the exact p-value to reject the null hypothesis at different levels, though in the current practice of statistical testing it is advisable to choose the significance level before taking the data to give more objectivity to the decision.

Criticism 4. *Statistical significance does not provide the probability of the hypothesis being true. Nor is statistical significance informative of the true value of the parameter.* The posterior probability of the null hypothesis, given a significant result, depends on the prior probability of the null hypothesis, as well as on the probabilities of having a significant result given the null and the alternative hypotheses. These probabilities cannot be determined in classical inference. It is only within Bayesian inference that posterior probability of the hypotheses can be computed, although these are subjective probabilities (Cabria, 1994; Lecoutre, 1999; 2006).

Criticism 4. *Type I error and Type II errors are inversely related. Researchers seem to ignore Type II errors while paying undue attention to Type I error.* Though the probabilities of the two types of errors are inversely related, there is a fundamental difference between them. While the probability of Type I error α is a constant that can be chosen before the experiment is done, the probability of Type II error is a function of the true value of the parameter, which is unknown. To solve this problem, power analysis assumes different possible values for the parameter and computes the probability of Type II error for these different values.

3.2. POSIBLE CONTRIBUTIONS OF BAYESIAN INFERENCE TO IMPROVE METHODOLOGICAL PRACTICE

In this section we begin summarizing the characteristics of Bayesian inference. We then present some arguments in favour of the Bayesian methodology: a) Bayesian inference does not contain greater subjectivity than other statistical methods; b) it provides the information that researchers need and c) there is statistical software available that facilitates the application of this methodology. We then suggest that the basic Bayesian concepts are understandable by psychology students, if a necessary didactic effort is made.

Bayesian inference

Bayesian inference is based on the systematic application of the Bayes Theorem, whose publication in 1763 disturbed the contemporary mathematicians. While in the previous conceptions of probability[1] it was assumed an objective value of probability, the possibility of revising the prior probabilities based on the new information opened by this theorem, lead to a

[1] Classical (quotient between favourable and possible cases) and frequentist (limit of relative frequency) conceptions.

new subjective view (Hacking, 1975; Cabriá, 1994). This new point of view also enlarges the applications of probability, since the repetition of an experience in exactly the same conditions was no more a requirement. Gradually, a distinction between frequentist probability, empirically accessible through frequencies, and epistemic probability or degree of belief in the occurrence of an event in a unique experiment (Rouanet, 1998a) and two schools of inference were developed.

In Bayesian inference a parameter θ is a random variable and we associate to it a prior epistemic distribution of probabilities $p(\theta)$, which represents the knowledge (or lack of knowledge) about θ before collecting the data. Let $y = (y_1,..., y_n)$ be a data set, whose likelihood function $p(y/\theta)$ depends on the parameter, then the conditional distribution of θ given the observed data y is given by the Bayes theorem:

$$p(\theta / y) = \frac{p(y/\theta)p(\theta)}{p(y)} \tag{1}$$

In (1) $p(y) = \sum p(y/\theta)p(\theta)$, where the sum extends through the admissible range of θ (Box and Tiao, 1992; Lee, 2004). The posterior distribution $p(\theta/y)$ contains all the information about θ once the data are observed. The Bayes theorem can be successively applied in new experiments, taking as prior probabilities of the second experiment the posterior probabilities obtained in a first experiment and so on. We speak of "learning process" (Box and Tiao, 1992).

The main method in Bayesian inference is the systematic application of the Bayes theorem, and the basic aim is updating the parameters prior distributions. The posterior distribution is the essence of Bayesian estimation. The answer to the question: once we see the data, what do we know about the parameter? is the posterior distribution, since this distribution synthesizes all the information about the parameter, once the data have been gathered and contains all the inferences that can be done from it (O'Hagan & Forster, 2004). The point estimate for the parameter is the mean of the posterior distribution, since it minimizes the expected quadratic error (O'Hagan & Forster, 2004). The posterior distribution will also allow us to compute the probability that the parameter is included in a given interval (credible interval) and the probability that the hypothesis is either true or false. The aim of Bayesian inference is to compute the hypothesis' posterior probability, contrary to classical inference, where the hypothesis is either accepted or rejected, which is not an inference, but a decision (O'Hagan & Forster, 2004).

The predictive or marginal distribution

$$p(y) = \propto \int p(y/\theta)p(\theta)d\theta$$

is used to predict future values of y. It takes into account the uncertainty about the parameter value θ, as well as the residual uncertainty about y when θ is known (Lee, 2004). This kind of probability cannot be computed in classical inference (Bolstad, 2004).

Subjectivity in Bayesian methods

A fundamental difference between Bayesian and classical inference is the subjective character (not frequentist) of probabilities, since neither the problem of repeated sampling is considered nor the sample distribution is required. Subjective probabilities can be defined for any situation, whereas frequentist probabilities are only defined for events in a space sample (O'Hagan & Forster, 2004). Moreover, Bayesian methods use all the previous information available, whereas in classical inference previous information is not considered.

Since the researcher specifies the prior distribution, the Bayesian approach takes into account the researcher's perspective, his/her knowledge of the problem. There is not just one way to choose the prior distribution, which conditions the results of inference. This fact has originated strong criticisms towards Bayesian methods since they can lead researchers to obtain different results from the same data set, depending on their previous knowledge or experience. The use of non informative priors at the beginning of the application of these methods, and updating these prior distributions in new applications, with the results of the previous steps has been suggested in order to confront these criticisms.

There is also the possibility of changing the models and interpretations throughout the analysis, whereas in classical inference both hypotheses and models are settled down before gathering the data and cannot be changed. This is not reasonable, since "allowing data to speak by themselves" is a basic idea in the mathematical modelling, where models are assumed to be useful to describe data but not to be exactly equal to data and it is therefore possible to change the model throughout the analysis (Pruzek, 1997; McLean, 2001).

The influence of prior distributions also depends on the sample size and the possible initial biases are corrected in successive experiments, since the weight lays on the likelihood as the sample size is progressively increased (Lindley, 1993). It is also advisable to repeat the analysis with different priors and inform about the differences obtained in the posterior distributions (Zhu and Lu, 2004). Procedures are standardised, using conjugated distributions, so that both the prior and posterior distribution belongs to the same functions family (Cabriá, 1994).

On the other hand, frequentist methods are not free of subjectivity: the significance level is arbitrarily defined, so that the same data is statistically significant or not depending on the chosen significance level (Skipper, Guenter & Nash, 1970). Statistical significance has no sense when the sample size is so big that any detected difference led to rejecting the null hypothesis. The variable definition, scale of measurement, significance tests used, are other subjective choices and even more, subjectivity is unavoidable in the interpretation of the results (Ayçaguera & Benavides, 2003). Of course, subjectivity does not imply arbitrairity; it is inevitable in social sciences due to the inherent randomness in its variables and has an important paper in the scientific research. The scientific community accepts the different findings, by establishing methodological or plausibility criteria (Matthews, 1998).

What are the Bayesian answers to researcher's needs?

Several works suggest that Bayesian inference provides a better answer to the researcher's needs as compared with frequentists inference (Lindley, 1993; Lecoutre, 1999; 2006).

Firstly, the meaning of probability in Bayesian statistics is identical to that of ordinary language: *conditional measurement of uncertainty* associated to the occurrence of an event, when some assumptions are assumed (Bernardo, 2003). This is the intuitive - although incorrect - interpretation that many scientists give to the frequentist probabilities associated to hypotheses tests, whose results are unconsciously interpreted in Bayesians terms (Falk, 1986; Gingerenzer, 1993; Rouanet, 1998; Lecoutre, 2006; Lecoutre, Lecoutre & Poitevineau, 2001; Haller & Krauss, 2002).

Consequently, the Bayesian interpretation of inference is simpler and more natural than that of frequentists inference (Pruzek, 1997), besides providing a base for coherent decision making in uncertainty situations (Western, 1999). In addition, Bayesian inference provides a totally general method, because its application does not require a particular kind of distribution and sampling distributions do not need to be deduced (Bernardo, 2003). Next we analyze the Bayesian answer to several questions of interest for researchers.

Effect size

A recommendation to complement hypothesis tests is to study the effect size, but a point estimation is insufficient, since it does not consider the sample error (Poitevineau, 1998). A power study would be recommendable to avoid erroneous conclusions about the absence of an effect when the result is nonsignificant (Cohen, 1990), but power computations does not depend on the statistical value observed in the sample and is therefore not pertinent to interpret a particular result, once the data are gathered (Falk & Greenbaum, 1995). Confidence intervals have the same frequentist interpretation than hypotheses tests, since they only indicate the proportion of intervals with a given sample size computed from the same population that would cover the parameter value, but they do not give information about whether the calculated interval covers the parameter or not (Cumming, Williams & Fidler, 2004).

Effect sizes and their magnitude appear in natural way in the Bayesians methods, which consider the parameter as a random variable. The probability that this parameter takes a certain value can be computed via the posterior distribution; for example it is possible to use sentences such as "the probability that the effect is larger than a is equal to 0.25". The credibility interval also provides the limits in which the parameter is included with a certain probability (Poitevineau, 1998; Lecoutre, 2006).

Hypothesis tests

The p-value provides a probability that is not useful for researchers: the probability of collecting data more extreme than the obtained if we repeated many times the experiment and the hypothesis were true (Matthews, 1998). But no researcher is interested in repeating the same experiment indefinitely and the aim of the scientific research is not to make a decision about the certainty of the hypothesis but adjusting our degree of belief in the hypothesis that is being tested (Rozeboom, 1970).

Interpreting the rejection of the null hypothesis as direct support to the research hypothesis (alternative) is incorrect, since a significant result does not indicate the magnitude of the effect, so that the statistical hypothesis does not inform on the practical meaning of the data (Hager, 2000; Finch, Cumming & Thomason, 2001). This can produce situations in which rejecting a null hypothesis does not provide any new information, since the only thing we can deduce when we reject a hypothesis is that there is an effect, but not its direction or magnitude (Falk & Greenbaum, 1995; Lecoutre, 1999).

On the contrary, in Bayesian inference we can compute the hypothesis posterior probabilities and the probabilities that the effect has a given size (Lindley, 1993). Moreover, the Bayesian method is comparative. It compares the probability of the observed event under the null hypothesis and under different alternative hypotheses (Lindley, 1993). Besides, in some situations, as bioequivalence tests, the interest is centred in verifying the null hypothesis, that is, we hope the treatments are equivalent (Molinero, 2002). In these cases the Bayesian approach is much more natural than the frequentists one, since we try to accept (not to reject) the null hypothesis.

Predictive probabilities and replication

Interpreting statistical significance as support to data replicability does not have a statistical base (Falk, 1986; Gingerenzer, 1994; Cohen, 1994; Falk & Greenbaum, 1995; Pascual, García & Frías, 2000). Statistical significance neither can be taken as an evidence that the research hypothesis is true; nor it provides the probability of the hypothesis; there is therefore no base to study replication and it does not provide verifiable evidence to replication either (Sohn, 1998).

In the Bayesian approach we can compute the probability of a future event, using the predictive distribution, which is given by the denominator in the Bayes formula, that is, the weighted average of the probability function, weighted by the prior probabilities (Berry, 1995). This distribution serves to study the possibility of replication of our results or to compute the sample size needed for a future study to be conclusive (Lecoutre, 1996). Of course, in case the requirements of data precision and sound procedures are fulfilled (Sohn, 1998). Correctly understood, replicability is related to the data reliability and consistency, and the only way to achieve it is sucessive empirical trials (Pascual, García & Frías, 2000).

Use of previous information

Whereas frequentist methods consider each sample as completely new and do not incorporate the information of previous studies, in the Bayesian framework we conceive a sequence of articulated experiments, where the information of each of them is used in the following step (Pruzek, 1997); the possibility of different opinions or knowledge is also accepted (Lindley, 1993). Although is possible to use Bayesian inference when there is no previous information about the parameter, the most interesting characteristic is the use of "informative" priors whenever this is possible, or even to investigate the effect of different priors. The central idea of Bayesian approach is updating the probabilistic knowledge about the phenomenon, based on the information available.

Computational viability of Bayesian methodology

A requirement to introduce new data analysis methods is the availability of calculation programs that facilitate their application. In the last years several researchers are developing diverse Bayesian programs, so that this approach is being introduced gradually in Social Sciences. For example, Albert (1996) published some Minitab subroutines for elementary Bayesian analysis that can be downloaded from the author's website (http://bayes.bgsu.edu/).

First Bayes (http://www.tonyohagan.co.uk/1b/) was prepared at Sheffield University to teach elementary Bayesian concepts. It admits different families of distributions and calculates posterior and predictive probabilities in uniparametrics models, analysis of variance and regression (Lawrence, 2003).

PAC (Lecoutre, 1996) also allows the analysis of data from general experimental designs, incorporating univariate and multivariate variance analysis, including repeated and covariable measures. The program includes frequentist and Bayesian analysis, with prior informative and non informative. It was developed by a research group that tries to incorporate Bayesian analysis in the statistical methods more frequently used in psychology. A reduced version is freely distributed from the group website (http://www.univ-rouen.fr/LMRS/Persopage/ Lecoutre/Eris.html).

For more complex analyses Buggs (Bayesian inference Using Gibbs Sampling) is an interactive and flexible software Windows compatible, that allows complex Bayesians calculations, based on simulation (see in http://www.mrc-bsu.cam.ac.uk/bugs/). There are on line facilities, such as tutorial, user groups and examples.

BACC (Bayesian Analysis Computation and Communication) was developed from a project funded by the National Science Foundation in the United States, and offers resources for Bayesian calculations, freely available. The emphasis is put on the combination of models and the development of predictive distributions. There are versions available for Matlab, S-PLUS and R, for Windows, UNIX and Linux systems (http://www2.cirano.qc.ca/~bacc/).

Other Bayesian computation programs are listed in http://www.mas.ncl.ac.uk/~ndjw1/bookmarks/Stats/Software-Statistical_computing/Bayesian _software/index.html/.

Didactic viability of elementary Bayesians methods

Introducing a new methodology in psychology will require its understanding by the possible users, that is to say, will depend on the degree to which we are able to transmit its main ideas in applied statistic courses. Iglesias et al. (2000) suggest the following content to introduce Bayesian inference, along with classical inference in undergraduates' courses following the approach by De Groot (1988):

- Basic concepts: population, parameter, sample, statistics, likelihood function, prior and posterior distributions.

- Point estimation: Classical and Bayesian methods.
- Interval estimation: Confidence and credibility intervals.
- Hypothesis tests: Classical and Bayesian tests, multiple decision problems.

In this sense, we found a increasing number of textbooks whose understanding does not require much mathematical knowledge and where basic Bayesian inference elements are contextualized in examples interesting and familiar for the students (for example Berry, 1995 or Albert & Rossman, 2001). These materials can be complemented with many references that explain in a simple way the basics of Bayesian inference (e.g. Ayçaguera & Benavides, 2003; Ayçaguera & Suárez, 1995). We can also find Internet didactic resources that facilitate the learning of these concepts, such as applets that visualize the Bayes theorem or the probability distributions, or compute posterior distributions, inference for means and proportions with discrete or continuous prior distributions (see, for example Jim Albert site, http://bayes.bgsu.edu/).

Most of the authors mentioned in this section have incorporated Bayesian methods to their teaching and have reported that students seem to understand better Bayesian inference than classical inference. We also found descriptions of concrete teaching experiments and suggestions about the way to carry them out (Bolstad, 2002). We are conscious, nevertheless, that this position is still controversial (e.g. Moore, 1997) due to the scarce empirical research on the students learning within statistics courses. Moreover, biases in conditional probability reasoning, as described below, may affect students' learning of Bayesian inference.

3.3. CONDITIONAL REASONING AND ITS RELEVANCE FOR UNDERSTANDING BAYESIAN INFERENCE

Research on understanding conditional probability has been carried out with both secondary school and University students. Fischbein and Gazit (1984) organized teaching experiments with 10-12 year-olds and found that conditional probability problems were harder in without replacement situation as compared to with replacement problems. Following that research Tarr and Jones (1997) identified the following four levels of thinking about conditional probability and independence in middle school students (9-13 year-olds):

- Level 1 (subjective): students ignore given numerical information in making predictions.
- Level 2 (transitional): students demonstrate some recognition of whether consecutive

events are related or not; however, their use of numbers to determine conditional probability is inappropriate.

- Level 3 (informal quantitative): students' differentiation of "with and without replacement situations" is imprecise as is the quantification of the corresponding probabilities; they are also unable to produce the complete composition of the sample space in judging independence.
- Level 4 (numerical): students state the necessary conditions for two events to be related, they assign the correct numerical probabilities and they distinguish between dependent and independent events in "with (e.g. item 15 in appendix) and without (items 4, 9) replacement situations".

Even when students progress towards the upper level in this classification (see also Tarr & Lannin, 2005), difficulties still remain at high school and University. This is shown in the various studies we summarize below, from which we have taken some of the items in our questionnaire. The full questionnaire is included in Appendix 1.

Conditioning and causation

It is well known that if an event *B* is the cause of another event *A* whenever *B* is present *A* is also present and therefore $P(A/B)=1$. On the contrary $P(A/B)=1$ does not imply that *B* is a cause for *A*, though the existence of a conditional relationship indicates a possible causal relationship. From a psychological point of view, the person who assesses the conditional probability $P(A/B)$ may perceive different type of relationships between *A* and *B* depending on the context (Tversky & Kahneman, 1982a). If *B* is perceived as a cause of *A*, $P(A/B)$ is viewed as a causal relation, if *A* is perceived as a possible cause of *B*, $P(A/B)$ is viewed as a diagnostic relation. At other times people confuse the two probabilities $P(A/B)$ and $P(B/A)$; this confusion was termed the *fallacy of the transposed conditional* (Falk, 1986). Item 10 in Appendix 1 was included to assess these difficulties.

Causal reasoning and the fallacy of the time axis

Falk (1989) gave item 17 in the Appendix 1 to 88 university students and found that while students easily answered part (a), in part (b) they typically argued that the result of the second draw could not influence the first, and claimed that the probability in Part B is 1/2. Falk suggested that these students confused conditional and causal reasoning and termed *fallacy of the time axis* their belief that an event could not condition another event that occurs before it.

This is a false reasoning, because even though there is no causal relation from the second event to the first one, the information in the problem that the second ball is red has reduced the sample space for the first drawing. Hence, *P (B1 is red/ B2 is red) =1/3*. Similar results were found by Gras and Totohasina (1995) who identified two different misconceptions about conditional probability in a survey of seventy-five 17 to 18 year-old secondary school students:

- The *chronological conception* where students interpret the conditional probability *P(A/B)* as a temporal relationship; that is, the conditioning event *B* should always precede event *A*.
- The *causal conception* where students interpret the conditional probability *P(A/B)* as an implicit causal relationship; that is, the conditioning event *B* is the cause and *A* is the consequence.

Synchronical and diachronical situations

Another issue involving time and conditional probability has been identified in the literature. In *diachronical* situations (e.g. items 5 and 17 in the Appendix) the problem is formulated as a series of sequential experiments, which are carried out over time. *Synchronical* situations (e.g. items 4, 8 and 10 in the Appendix) are static and do not incorporate an underlying sequence of experiments. Formally the two situations are equivalent, however Sánchez and Hernández (2003) found that students did not always perceive the situations as equivalent and produce additive solutions to synchronical conditional problems.

Solving Bayes problems

As regards Bayesian reasoning (see a summary in Koehler, 1996), early research by Tversky and Kahneman (1982a) suggests that people do not employ this reasoning intuitively and establish the robustness and spread of the base-rate fallacy in students and professionals (Bar-Hillel, 1983). Totohasina (1992) suggested that part of the difficulty in solving Bayes' problems is due to the representation chosen by the student to solve the problems and that using a two way table is an obstacle to perceive the sequential nature of some problems, and therefore can lead students to confuse conditional and joint probability.

Recent research suggests that Bayesian computations are simpler when information is given in natural frequencies, instead of using probabilities, percentages or relative frequencies (Cosmides & Tooby, 1996; Gigerenzer, 1994; Gigerenzer & Hoffrage, 1995). The reason is that natural frequencies (absolute frequencies) correspond to the format of information humans have encountered throughout their evolutionary development. In particular, Bayes problems transform to simple probability problems if the data are given in an adequate format of absolute frequencies. Sedlmeier (1999) analyzes and summarizes recent teaching experiments carried out by psychologists that follow this approach and involve the use of computers. The results of these experiments suggest that statistical training is effective if students are taught to translate statistical tasks to an adequate format, including tree diagrams and absolute frequencies (Martignon & Wassner, 2002).

Other difficulties and need for a comprehensive assessment questionnaire

Other difficulties include problems in defining the conditioning event (Bar-Hillel & Falk, 1982) and misunderstanding of independence (Sánchez, 1996; Truran & Truran, 1997). People also have problems with compound probabilities. Kahneman and Tversky (1982a) termed conjunction fallacy people's unawareness that a compound probability cannot be higher than the probability of each single event.

The previous study of literature showed us that there is a large amount of research on this topic but we found no comprehensive questionnaires to globally assess students' understanding and misconceptions on these topics and relate one to another. As a result, one of the goals in this research was constructing a questionnaire, which takes into account the content of conditional probability taught in the Spanish universities to psychology students, as well as the biases and misconceptions described in the literature. Studies 1-6 were oriented to construct and validate the questionnaire; Study 7 was directed to assess conditional reasoning with this questionnaire in a sample of 414 psychology students after teaching of the topic. We also analyse possible relationships between formal knowledge of the topics and psychological biases (Study 6) and relationship between understanding conditional probability and learning Bayesian inference (Study 9). Even when we focus on psychology students, the questionnaire is useful in assessing conditional probability reasoning for other undergraduate or high school students.

4. A BAYESIAN APPROACH TO CLASSICAL TESTS THEORY

In the Classical Tests Theory (Muñiz, 1994; Martinez Arias, 1995), formulated by Spearman (1904), the empirical score X obtained by a subject in a test is a random variable and it is made of two components: the subject's true score (V) in that test, that it is assumed to be constant and the error measurement (e). The model makes the following hypotheses

(Muñiz, 1994):

$$X=V+e$$

$$E(X)=V$$

- $E\ (e_i)=0,$ for the population of subjects being measured, as well as for the infinite repetitions of the test in a subject. It is supposed that errors follow a normal distribution.
- $\rho(V,e)=0; \rho(e_i,e_j)=0$. It is assumed that the measurement error is not correlated with the true score and the measurement errors of different subjects are also independent.

In a consistent Bayesian formulation of the Classical Tests Theory, the basic assumptions should be respected and the main difference is considering the model parameters as random variables, with prior and posterior distributions. Accepting this assumption, the estimation of these parameters should be carried out with a Bayesian methodology, following its procedures and objectives. Consequently, the true score is now a random variable with a normal prior distribution[2]. From these assumptions we derive the following equalities, similar to those in CTT since they are still applicable when V is a random variable:

$$E\ (X)=E\ (V)$$

$$\sigma_X^2=\sigma_V^2+\sigma_e^2$$

$$\rho_{XV}^2=\frac{\sigma_V^2}{\sigma_X^2}=1-\frac{\sigma_e^2}{\sigma_X^2}=1-\rho_{Xe}^2$$

Mean score

We can use Bayesian inference to estimate the population mean, or the difference of two different means, with both informative and non informative priors. For non informative prior distribution two cases appear:

- The standard deviation σ_i of the average prior distribution is known. In this case, for a uniform prior distribution, the average posterior distribution is normal $N\ (\bar{x}, \sigma/\sqrt{n})$ where $\bar{x}$ is the sample mean. The equation $Z=\frac{\mu_f-\bar{x}}{\sigma/\sqrt{n}}$ follows a distribution N (0,1) (Berry, 1995). The point estimator of the mean on the posterior distribution μ_f is the sample mean $\bar{x}$ of the data. The credibility interval for a credibility coefficient α is given by: $(\bar{x}-Z_{1-(1-\alpha)/2}\sigma/\sqrt{n};\bar{x}+Z_{1-(1-\alpha)/2}\sigma/\sqrt{n})$, being Z a percentile of the standard normal

[2] Since the true score is sum of scores in the different items, approximated normality is reasonable.

distribution.

- If σ (population standard deviation) is not known, we can use *s*, the unbiased estimation of the standard deviation (sample cuasivariance square root) and the *T* distribution with *n-1* degrees of freedom, being *n* the sample size of data (Bolstad, 2004).

- When the prior distribution for the population mean follows a normal distribution $N(\mu_i, \sigma_i)$ and the standard deviation σ_i on the prior distribution of the mean is known, the posterior distribution also follows a normal distribution $N(\mu_f, \sigma_f)$. The values of the mean and standard deviation of the posterior distribution are given by the following formulas:

$$\mu_f = \frac{\dfrac{n\bar{x}}{s^2} + \dfrac{\mu_i}{\sigma_i^2}}{\dfrac{n}{s^2} + \dfrac{1}{\sigma_0^2}} \qquad \sigma_f = \frac{1}{\sqrt{n/s^2 + 1/\sigma_i^2}}$$

In previous expressions *n* is the sample size, $\bar{x}$ and *s* the mean and standard deviation of the sample. For the case that the standard deviation σ_i in the prior distribution of the mean is not known, this one is estimated from the square root of sample cuasivariance *s*. The previous formulas of the mean and standard deviation of the posterior distribution are the same, but now the distribution will be *T* with *n-1* degrees of freedom (being *n* the sample size), that can be approximated to the normal distribution with a sufficient sample size (Bolstad, 2004).

Difference of two mean scores

The commonest situation is the comparison of two independent simples, where different cases can be found. We will only deal with the case of prior informative distributions, since the non informative case can be included in this one.

Case 1. Identical known variances. The mean and variance of the score difference in the posterior distribution are given by:

$$\mu_d^f = \mu_1^f - \mu_2^f$$
$$\sigma_d^{2f} = \sigma_1^{2f} + \sigma_2^{2f}$$

which would coincide with the mean and variance of the sample distribution in the case of non informative prior distribution. The credibility interval of the means difference for a α credibility coefficient would be:

$$(\mu_1^f - \mu_2^f \pm Z_{1-(1-\alpha)/2}\sqrt{\sigma_1^{2f} + \sigma_2^{2f}})$$

Case 2. Different known variances. When and prior distributions are independent in both samples, posterior distributions will be also independent. The mean and variance of the posterior distribution will be again[3]:

$$\mu_d^f = \mu_1^f - \mu_2^f$$

$$\sigma_d^{2f} = \sigma_1^{2f} + \sigma_2^{2f}$$

and the credibility is given by:

$$(\mu_1^f - \mu_2^f \pm Z_{1-(1-\alpha)/2}\sqrt{\sigma_1^{2f} + \sigma_2^{2f}})$$

Case 3. Variances are not known. In this case, each of the variances should be estimated from the sample data (using the sample cuasivariances $s_1^2; s_2^2$). This increases the uncertainty of the estimation, and therefore a *T* distribution will be used, instead of the normal distribution (Box & Tiao, 1992). The degrees of freedom are given by the Satterhwaite formula: (Bolstad, 2004):

$$\upsilon = \frac{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}{\frac{(s_1^2/n_1)^2}{n_1+1} + \frac{(s_2^2/n_2)^2}{n_2+1}}$$

The approximated credibility interval is given by

$$(\mu_1^f - \mu_2^f \pm T_{1-(1-\alpha)/2}\sqrt{\sigma_1^{2f} + \sigma_2^{2f}})$$

where mean and variance on the posterior distributions are given by (1), the prior variances are estimated by the sample cuasivariances and the Satterhwaite formula is used for calculating the degrees of freedom. For the non informative prior distribution case, this expression is:

$$(\mu_1^f - \mu_2^f \pm T_{1-(1-\alpha)/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}})$$

which coincides with the frecuentist confidence interval, but with a different interpretation.

Estimation of difficulty indexes

The difficulty index is defined as the proportion *p* of subjects that will get right the item,

[3] In this case the initial variances are different.

between all those that try to solve it in a certain population (Thorndike, 1991). Whereas in classical inference, the proportion p is considered constant, in Bayesian inference the difficulty index p is a random variable. Given a prior probability function *Be(a,b)* for a proportion, if in a new sample we observe e successes and f failures, the posterior probability function is *Be(a+e, b+f)* (Serrano, 2003).

Any Beta with $a=b$ can be used as non informative prior, that is to say, a uniform distribution of the parameter p (Lecoutre, 1996). In our study we use *Be(0.5,0.5)* as recommended by Lecoutre (1996) or Serrano (2003). The credibility interval is given by:

$$\left[\beta^{-1}_{0.5+p,\,0.5+q}(\alpha/2) - \beta^{-1}_{0.5+p,\,0.5+q}(1-\alpha/2)\right]$$

where $\beta^{-1}_{a,b}$ is the *Beta (a,b)* distribution inverse function and α the credibility coefficient.

Estimation of discrimination indexes

A first approach to study the discrimination indexes is analysing the difference in the proportion of the item success in two groups of students with different competence[4]. Let p_s and p_i be the difficulty indexes in the higher and lower groups. In the classical theory these parameters are unknown constants in their respective populations and the point estimation of the discrimination index is:

$$d = \hat{p}_s - \hat{p}_i$$

where $\hat{p}_s\ \hat{p}_i$ are the point estimators of p_s and p_i respectively. In the Bayesian interpretation, the previous proportions and their difference would be random variables. If the prior distribution for $p_s\,p_i$ are taken from the Beta family, we will obtain posterior Beta distributions for each of these proportions. Since the populations are independent, the posterior joint distribution of the bidimensional variable ($p_s\,p_i$) is the product of two posterior distributions for each proportion.

In case of non informative prior (for example, *Be(1,1)*), let e_s be the successes and f_s the failures in the higher group and e_i the successes and f_i the failures in the lower group. The respective estimators for the proportions are:

$$\hat{p}_s = \frac{e_s + 1_s}{2 + e_s + f_s} \qquad \hat{p}_i = \frac{e_i + 1}{e_i + f + 2_i} \quad \text{(Albert, 1995; 1996)}$$

Let the prior distribution for p_s be *Be(*a_s*,* b_s*)* and the prior distribution for p_i *Be(*a_i*,* b_i*)*. If

[4] For example, students with and without instruction.

we achieve e_s successes and f_s failures in the higher group and e_i successes and f_i failures in the lower group, the respective estimators of the proportions are:

$$\hat{p}_s = \frac{a_s + e_s}{a_s + b_s + e_s + f_s} \qquad \hat{p}_i = \frac{a_i + e_i}{a_i + b_i + e_i + f_i} \quad \text{(Albert, 1996)}$$

In both cases the posterior distributions of the populations p_s is $Be(a'_s, b'_s)$ and that of p_i is $Be(a'_i, b'_i)$, that will be given by the previous formulas and are independent (Bolstad, 2004). Following Berry (1995) the estimators for the means in the posterior distribution are:

$$\hat{p}_s = \frac{a'_s}{a'_s + b'_s} \qquad \hat{p}_i = \frac{a'_i}{a'_i + b_i'}$$

The estimators for the standard deviations in the posterior distributions will be (Bolstad, 2004):

$$\partial_s = \sqrt{\frac{\hat{p}_s \hat{q}_s}{n_s + 1}} \quad \partial_i = \sqrt{\frac{\hat{p}_i \hat{q}_i}{n_i + 1}}$$

Consequently, the difference of proportions is approximately a normal distribution $N(\hat{p}_s - \hat{p}_i, \sqrt{\partial_s^2 + \partial_i^2})$, so that the approximated credibility interval is given by:

$$\hat{p}_s - \hat{p}_i \pm Z_{1-(1-\alpha)/2}^{\;-1} \sqrt{\partial_s^2 + \partial_i^2}$$

where Z is the normal *N(0,1)* distribution.

Estimating correlations and reliability coefficients

There are diverse procedures to estimate the reliability coefficient, some of which are based on estimating the correlation coefficient between scores in two administrations of the questionnaire or between scores in two equivalent forms of the questionnaire: test-retest; parallel forms and split-half reliability. In estimating these coefficients and other psychometric features[5] the correlation coefficient is used, which is a random variable in the Bayesian interpretation. Given a set of observed pairs $(x_1, y_1), (x_1, y_1), \ldots (x_1, y_1)$ for a bidimensional random variable *(X, Y)* with bivariate normal distribution, let's assume that the mean, variances and correlation of the scores are given by:

[5] E.g. the discrimination index can also be assessed as correlation between the item score and total score in the test.

$$E(X)=\mu;Var(X)=\sigma^2$$
$$E(Y)=\eta;Var(X)=\varphi^2$$
$$\rho(X,Y)=\rho$$

Assume we have computed the means $\bar{x}$ and $\bar{y}$ and correlation r in the data. In case of non informative priors for the means and variances of X and Y, and given a prior distribution for the correlation coefficient $p(\rho)$, a reasonable estimation for the correlation coefficient posterior distribution is given by (Lee, 2004):

$$p(\rho/(x,y)) \propto p(\rho)\frac{\left(1-\rho^2\right)^{(n-1)/2}}{\left(1-\rho r\right)^{n-3/2}}$$

Replacing $\rho=\tanh\xi;r=\tanh z$, a new estimation is obtained, this time through the normal distribution:

$$\xi \sim N(z,1/n)$$

This approximation can be used to find credibility intervals for the hyperbolic tangent of the correlation coefficient and from these intervals, inverting the change of variable; we find the interval for the correlation coefficient.

For informative priors, let's assume that in the first occasion we observe a correlation coefficient r_1 in a sample size n_1, which lead to a posterior distribution N($tanh^{-1}$ $r_{1,}$ $1/n_1$). In a second occasion we observe a correlation coefficient r_2 in a sample size of n_2. When taking the posterior distribution in the first observation as a prior distribution in the second experiment, we can apply the formulas for estimating the mean of the normal distribution. Therefore, to estimate $tanh^{-1}$ r we have a normal posterior distribution, whose mean and variance are given by the following expressions:

$$Variance=\frac{1}{n_1+n_2}$$

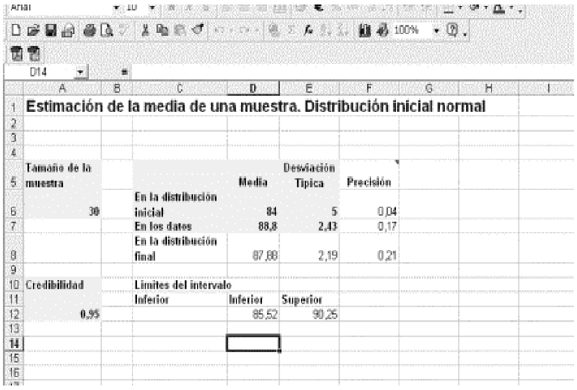
$$Mean=Variance(n_1\tanh^{-1}r_1+n_1\tanh^{-1}r_1)$$

Again this transformation is applied to obtain a credibility interval of the hyperbolic tangent arc for the correlation coefficient, and inverting the transformation we obtain the credibility interval for the correlation coefficient.

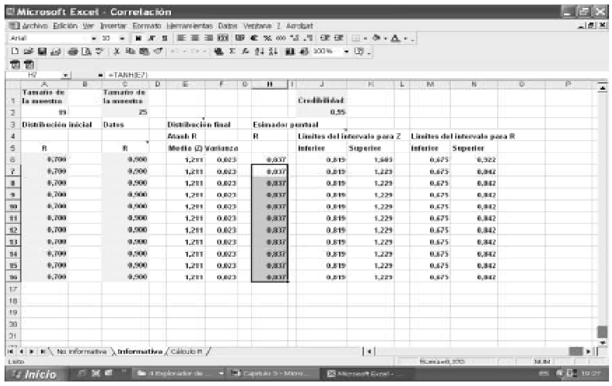
Computation software

In order to make the above calculations we prepare a set of Excel programs (See examples in Figure 8.2), using the formulas given in the previous sections, for each of the cases described. We also have distinguished (in different sheets of Excel files) the informative and non informative prior cases. The programs permit the variation of credibility and confidence coefficients, sample sizes, prior distributions parameters, sample statistics etc. The data statistics required can be computed with SPSS or another statistical program.

Figure 8.2. Some Excel programmes developed

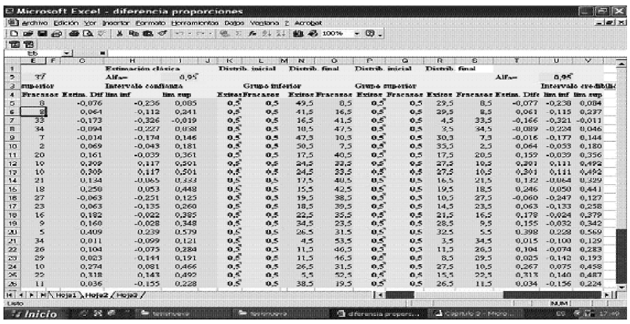


Mean



Correlation

Difference of means



Difference of proportions

In summary the above analysis was carried out to follow the Research Objective 1: *Rethinking the Classical Tests Theory (TCT) from the Bayesian point of view and analyzing the implications of this change of perspective on the estimation of some psychometric features in the tests and items.*

5. BUILDING AND VALIDATING THE CPR QUESTIONNAIRE

The objective 2 in this research *was to apply the above analysis in the process of building a questionnaire and compare results from classical and Bayesian estimates in some of the test features.* At the same time Objective 3 was *assessing conditional probability reasoning in psychology students to decide the suitability of teaching Bayesian methods to these students*. In Chapters 4 and 5 of the thesis we describe the process of building and validating the CPR questionnaire with the purpose of fulfilling these two aims.

The instrument should be useful to assess in just one application the biases and misunderstanding related to conditional probability described in previous research and summarized in section 3.3 in addition to the conceptual and procedural knowledge included in the teaching of the topic in the training of psychologists in Spain. Below we briefly describe the process of building the questionnaire which is explained in detail in Chapters 4 and 5 of the thesis. This procedure includes the use of Bayesian methods to estimate difficulty and discrimination indexes (both as difference in averages and as item- total correlation), test-retest and split-half reliability coefficients at different stages in the process. We use non informative priors in the first application of each estimation procedure; in next steps the previous final distributions are used as new informative priors.

Steps in building the questionnaire

The building of the CPR questionnaire was based on a rigorous methodological process, which included the following steps:

1. *Semantic definition of the variable* (Study 1). In educational measurement (e.g. Millman & Greene, 1989) a distinction is made between constructs (unobservable psychological traits, such as *understanding of conditional probability*) and the variables (e.g. score in a questionnaire) we use to make inferences regarding the construct. In order to achieve objectivity in defining our variable, we decompose the construct "understanding conditional probability" in semantic units. These semantic units were defined after a content analysis of 19 text books used in the teaching of statistics to psychologists. The conditional probability content in the textbooks was analysed and the definitions, properties, relationships with other concepts and procedures were classified in a reduced number of categories by means of a systematic and objective identification (Ghiglione & Matalón, 1991). To select the books, the list of references recommended in statistics courses was requested to the 31 Faculties of Psychology in Spain. All the textbooks recommended by at least 4 different Universities were analysed, after discarding some books in which conditional probability was not included.
2. *Constructing an item bank*. The aforementioned analysis was complemented with our revision of previous research on conditional probability reasoning, that also served to compile a sample of n=49 different items used in this research, some of which had been used by different authors. These items were translated into Spanish and reworded to make their format homogeneous and improve their understanding.

3. *Selection of items* (Study 2). The item difficulty (percentage of correct answers) and discrimination (correlation with test total score) were estimated from the answers by different samples of psychology students (between 49 and 117 students answered each pilot item) by classical and Bayesian procedures. Final selection of items took into account these two parameters as well as results from expert judgment. Ten statistics education researchers from five different countries (Brazil, Colombia, Mexico, Spain and Venezuela) who had themselves carried out research related to conditional probability or independence were asked to collaborate. They were asked to value (in a 5-point scale) the adequacy of the content units to understanding conditional probability as well as the suitability of each item to assess understanding for each specific content unit. The final items in the questionnaire were selected in such a way that a) the intended content of the questionnaire was covered (see Table 8.1); b) there was an agreement from the experts about the item adequacy; and c) item difficulty and discrimination were suitable.
4. *Formatting and revising the items.* We included two different formats: a) Multiple choice items with 3-4 possible responses were used to allow quick evaluation in the sample of some of the most pervasive biases described in the previous literature (e.g. item 3 taken from Tversky and Kahneman (1982a) which evaluates the base-rate fallacy, item 5 taken from Sánchez (1996) assesses the confusion between independent and mutually exclusive events and item 9 taken from Tversky and Kahneman (1982b) assesses the conjunction fallacy); b) Open- ended items were also used to better understand students' strategies in problem solving (e.g. item 16) and their understanding of definitions and properties (e.g., items 1, 2).
5. *The pilot trial of the instrument* (Study 3) took place in the academic year 2003-2004 with a small sample of $n=57$ Psychology students in order to make a preliminary estimation of the questionnaire reliability and validity. A second sample of $n=37$ students majoring in Mathematics was used to compare the performances in the two groups and to identify items with and without discriminative properties. Classical and Bayesian estimates of items difficulties and discrimination (both as item-total correlation and as difference of averages) were provided. A first estimation of internal consistency reliability provided a value Alfa= 0.787. Content validity was assessed through content analysis of items in the pilot questionnaire and through expert judgment of both content units and fitting of items to assess each content unit.
6. *Revising the pilot questionnaire* (Study 4). After discarding those items with bad psychometric features, a new expert judgment served to improve the wording of the items.

Thirteen expert methodology instructors were given three alternative wordings for each item and were asked to order the three versions, as regards methodology standards, as well as give the reasons for their choice. Rank statistics were used to summarise the data. Non parametric tests (Kendall & Friedman) showed clear agreement in the option selected by the experts for each item. This version was included in the final questionnaire and additional suggestions by the methodology instructors were used to still improve readability.

Table 8.1. Primary content assessed by each item

Content	Item
1. Defining conditional probability; giving appropriate examples	1
2. Recognising that a conditional probability involves a restriction in the sample space	2
3. Base rates fallacy	3
4. Distinguishing conditional, simple and joint probabilities	6
5. Distinguishing a conditional probability and its inverse (transposed conditional fallacy	6
6. Conjunction fallacy	9
7. Distinguishing independent and mutually exclusive events	4
8. Computing conditional probabilities in a single experiment	8
9. Solving conditional probability problems in a sampling with replacement setting	12
10. Solving conditional probability problems in a sampling without replacement setting	5
11. Computing conditional probabilities from joint and compound probabilities	7
12. Solving conditional probability problems when the time axis is reverted	17
13. Distinguishing conditional, causal and diagnosis situations	10
14. Solving conditional probability problems in a diachronic setting	14
15. Solving conditional probability problems in a synchronic setting	15
16. Solving compound probability problems by applying the product rule to independent events	13
17. Solving compound probability problems by applying the product rule to dependent events	18
18. Solving total probability problems	11
19. Solving Bayes problems	3, 16

The final questionnaire (see Appendix 1) is composed by 18 items, with some sub-items, which score independently and some open-ended items. In Table 1 we present the items primary contents that cover the content in the books analysed as well as main biases described in the literature. There is one item covering each content (item primary content); additionally each item also assesses some other secondary contents (described in detailed in Study 3).

CPR reliability

Once the questionnaire was finished we performed reliability and validity analyses (Studies 5 and 6).

A first approach to the reliability of the instrument was carried out by computing the Alpha coefficient in a sample of n=591 students from 4 different Universities, that gave a moderate value (Alpha=0.797). This value is reasonable, given that the questionnaire tries to assess a wide range of knowledge (see Table 1), so that a particular student can understand some of these concepts and do not understand others (Thorndike, 1991; Melia, 2001). We also computed two reliability coefficient based on factor analysis (Barbero, 2003):

1. $\theta = \frac{n}{n-1}\left(1 - \frac{1}{\lambda_1}\right)$=0.82; was high, since the first eigenvalue explained a relatively high percentage of variance and most items contributed to that factor before rotation, which is an indication of an underlying construct being measured by the questionnaire.
2. $\Omega = 1 - \frac{n - \sum h_j^2}{n + 2\sum r_{jh}} = 0.896$; was still higher, according what is theoretically expected; this coefficient measures the commonalities (common factors) in the items.

In the same sample (*n=591*) we also carried out a generalizability analysis (López Feal, 1987; Feldt & Brennan, 1991; Martínez Arias, 1995), an approach that considers the different sources of error in measurement, analyses the component of these errors and provides different coefficients. In this method it is possible to fix some sources of errors and use the analysis of variance to estimate the different components in the total variance, including the variance of errors. We took into account two different sources of variations in the tests scores:

1. Generalizability of results to other items (fixing the students and considering the items as the only source of variation). We obtained a coefficient G_i=0.799; very close to the Crombach's Alpha value, as, in this case the generalizability coefficient coincides with Alpha; the small difference is due to round-off in the computations.
2. Generalizability of results to other students (fixing the items and considering the students as the only source of variation). We obtained a coefficient G_i=0.987, which indicates a very high possibility of extending the results to other students similar to those taking part in the sample, when the items are fixed.

Another estimation of reliability using test- retest was carried out in a sample of 106 students, each of which completed the questionnaire in two different occasions with about a month between the two applications. We obtained a test- retest reliability coefficient of 0.871 (Pearson correlation) and 0.861 (Spearman Rho), which are quite high. The Pearson's correlations coefficients between responses to same items in the two applications were all statistically significant and positive, ranging between 0.29 and 0.79. Split-half reliability coefficient (when considering each application as half the total questionnaire) gave very high values (0.91); the means, variances, inter-element covariances and correlations were very similar in the two occasions; all of which assures a high test-retest reliability. The computation of test-restest reliability was complemented with the estimation of confidence and credibility intervals for all the correlations coefficients.

CPR validity

We carried out different studies to provide evidences for the validity of the questionnaire that was considered a unitary construct according Messick (1989; 1995; 1998) and AERA/ APA/ NCME (1999):

1. The theoretical analysis of the questionnaire content as well as the results from experts' judgment served to justify *content validity*, by comparing the content evaluated by each item to the semantic units included in the semantic definition (Study 3).
2. Studying the questionnaire capacity to discriminate between two groups of psychology students before and after studying conditional probability served to justify *criteria validity* (Study 6.1). We used discriminant analysis (Cuadras, 1981; Afifi y Clark, 1990) to compare results from 208 students without instructions and 177 students with instructions. Most items discriminated between the two groups (significant difference); the scarce exceptions were items measuring psychological biases. The canonical correlation was equal to 0.697 and the probability of correct classification was 82.34%, all of which suggest good criteria validity for the questionnaire. This study was complemented with statistical summaries, difference tests, confidence and credibility intervals for the mean of the total scores in the two groups that again favoured the group with instruction.
3. We analysed the structure of responses to the questionnaire in a sample of n=591 students and compared with the assumed structure of the construct (Study 6.2) to study the *construct validity* (Muñiz, 1994; Martínez Arias, 1995). We performed an exploratory factor analysis (Tabachnick & Fidell, 2001). We expected the analysis would confirm a main underlying construct, but, at the same time we also expected to find other factors that included the biases described in the literature and that would not correlate with the mathematical problem solving competence of students. All of this was confirmed in the factor analysis (main components extraction; varimax rotation), which lead to two different groups of interrelated factors, as described in Section 6.1.

Details and statistical results of all the different steps in the process of building the questionnaire are included in Chapters 4 and 5 of the thesis. We applied Bayesian methods along all these steps, in order to fulfil the research objective 2. The result is the CPR questionnaire with reasonable reliability and validity that will be used in the next stage of the research and is also useful to other teachers and researchers.

6. DESING AND VALIDATION OF DIDACTIC RESOURCES TO INTRODUCE ELEMENTARY BAYESIAN INFERENCE IN PSICHOLOGY

Objective 3 in this research was *assessing conditional probability reasoning in psychology students to decide the suitability of teaching Bayesian methods to these students.* In Study 7 we applied the CPR questionnaire to a sample of 413 psychology students and analysed their responses from different points of views. Students showed enough understanding of conditional probability to start the learning of Bayesian inference, but, at the same time, we found some widespread misconceptions that were taken into account in the next stage (designing a curricular proposal).

Objective 4 in this research was *preparing and assessing didactic resources to introduce elementary Bayesian inference to Psychology students that takes into account the previous assessment.* To attain this aim we designed some teaching materials that were based on results of Study 7, some didactic principles and the literature on teaching Bayesian inference. These materials were tried in Studies 8 and 9. Below we summarise these three studies, which are described in detail in Chapter 6 of the thesis.

6.1. ASSESSING CONDITIONAL REASONING IN PSYCHOLOGY STUDENTS

Once the CPR questionnaire was finished, we carried out an assessment study (Study 7). Students from the Universities of Granada (4 different groups of students; *n*=308 students) and Murcia (two different group of students; *n*=106 students) took part in the sample (*n*=414). The students were enrolled in an introductory statistics course in the first year of University studies (typically, 18-19 year-olds). They had studied conditional probability at secondary school level and were taught conditional probability and the Bayes theorem with the help of tree diagrams, two-way tables and examples in the field of psychology, for about 2 weeks before they completed the questionnaire. The questionnaire was given to the students as an activity in the course of data analysis. Participation was optional and all the students were collaborative with the research.

Once the data were collected, we analysed the response of each student in each item. The scoring for open-ended items took into account the completeness of response. In items 2, 8, 11, 12, 13, 15 the students were given a point in case they identified correctly the problem data; correct built a tree diagram and identified the conditional probability, and 2 points for a totally correct solution. In item 1 and 16 the scoring ranged from 1 to 4 (see Table 4). The maximum possible scoring in the questionnaire was 34 points. The empiric distribution of scoring ranged between 3 and 30 with an average value of 19.12, a little higher than half the maximum possible score and the standard deviation was 5.91

In computing several probabilities from a two-way table (item 1) 90% of the students correctly computed the simple probability, 61%, the joint probability and 59% and 56%, respectively the two conditional probabilities. This confirms Falk´s (1989) opinion that verbal ambiguity in linguistic expression of conditional probability still makes it difficult for the student to distinguish conditional and joint probabilities after instruction.

Results in Table 8.2 suggest the existence of the following reasoning conflicts among the students in the sample:

Table 8.2. Percentage of responses in multiple-choice items (n=414)

	a	b	c	d	Blank
I3	8	7	29	50(+)	5
I4	28	15	29	20(+)	8
I5	1	89(+)	10	(*)	0
I7	35(+)	31	34	(*)	0
I9	25(+)	9	62	(*)	4
I10	6	32(+)	59	(*)	9
I14	77	9	10 (+)	2	2
I17a	6	17	69 (+)	7	1
I17b	24(+)	25	9	36	6
I18	9	13	76(+)	(*)	2

(*) Does not apply (+) Correct

1. *As regards independence*: we found confusion of independence with mutual exclusiveness in 28 % of the responses to distractor a) in item 4; a bias also noticed by Sánchez (1996). The chronological conception of independence described by Gras and Totohasina (1995) was also shown in 29% of the responses to distractor b) in item 4.
2. *Concerning conditional probability:* 31% of the students confused it with a joint probability (response b in item 7) or with a simple probability (34% responses c in item 7). The conjunction fallacy was observed in 62% of the responses to item 9 and the confusion of the transposed conditional in 59% of the responses in item 10. Difficulties in computing probabilities when the time axis is inverted are suggested by the responses to items 14 and 17b, although the chronological conception of conditional probability described by Gras and Totohasina (1995) was not so clearly shown in these two items.
3. The base rate fallacy was not as pervasive as suggested in previous research (Bar-Hillel, 1983) as shown in the responses to distractors (a) and (b) in item 3; since the majority of

students gave the correct response (d) in this item, then showing improvement of base rate with instruction. Item 18 was also very easy.

Table 8.3. Completeness of solutions in open-ended items

	I1	I2	I8	I11	I12	I13	I15
Blank or totally wrong	29	15	47	18	21	30	24
Partly correct	30	21	18	21	9	18	16
Correct solution	41	64	35	61	70	52	60

As regards responses in open-ended items, results in Table 8.3 suggest that students had difficulties in giving a sound definition and an example of conditional probability (item 1) but were conscious of the restriction of sample space (item 2). They had difficulties in solving a conditional probability problem in a single experiment (item 8) due to a lack of distinction of dependent and independent experiments in the context (synchronic situation), so that many of them did not appear to have completely reached Level 4 in the conditional probability reasoning scheme by Tarr and Jones (1997).

Solving total probability (item 15) and solving conditional probability problems with replacement problems (item 12) and computing compound probability in the case of independent (item 11) events were easier than computing compound probability in dependent (item 13) events.

Table 8.4. Completeness of solutions in solving a Bayes problems (Item 16)

	Percentage
Blank or totally wrong	16
Correct identification of data	15
Identifies the inverse conditional probability,	16
Correct computation of denominator (total probability)	7
Correct solution	46

As regards solving an open Bayes problem (item 16), more than half the students were able to compute the total probability and a little less gave the complete solution; the majority was at least capable of correctly identifying the data and even identifying the probability to be computed although 16% failed in developing the total probability formula. We remark that data were given in the percentage format, which is considered harder than absolute frequency formats in Gigerenzer (1994) and Gigerenzer and Hoffrage´s (1995) research. We can conclude that, in general, the instruction was successful as regards problem solving capabilities, whenever there were no psychological biases involved in the situation. However, part of the biases described in the literature seemed not to be overcome with instruction.

To explore our conjecture that biases on conditional probability reasoning are unrelated to mathematical performance in the tasks, we carried out a factor analysis of the set of responses to the items (correct-incorrect responses to each item by the different students) using the SPSS software. The factor extraction method was principal components, which is the most conservative method, as it does not distort the data structure. In Table 8.5 we present the factor loadings (correlations) of items with the different factors after Varimax rotation (orthogonal rotation; maximizing variance of the original variable space).

Table 8.5. Factor Loadings for Rotated Components in Exploratory Factor Analysis of Responses to Items

Item	Component						
	1	2	3	4	5	6	7
Item 16. Bayes rule	.76						
Item 11. Total probability	.76						
Item 15. Product rule in dependent, synchronic events	.75						
Item 13. Product rule in independent events	.67						
Item 12. Conditional probability with replacement	.43		.42				
Item 6b. Conditional probability. Table		.79					
Item 6c. Joint probability. Table		.77					
Item 6a. Simple probability. Table	.32	.61					
Item 6d. Conditional probability. Table		.61					
Item 8. Conditional probability in single experiment			.67				
Item 1. Definition			.59				
Item 2. Sample space	.40		.45				
Item 17b. Time axis fallacy, diachronic experiment				.71			
Item 14. Time axis fallacy, diachronic experiment				.70			
Item 7. Cond prob. from joint and compound probability, synchronic					.66		
Item 9. Conjunction fallacy					.62		
Item 5. Conditional probability, without replacement, diachronic			.39		.44		
Item 17a. Conditional probability, without replacement						.66	
Item 10. Transposed conditional /causal-diagnostic						-.65	
Item 3. Independence /mutually exclusiveness							.68
Item 3. Base rates/ Bayes rule	.34						.48
Item 18. Product rule dependence, diachronic				.35			-.46

We found 7 factors with eigenvalue higher than 1 that explained the following percentages of the total variance: 21% (first factor), 7 % (second factor), and about 6% in the remaining factors; that is, a total of 59% of the variance was explained by the set of factors, which suggests the specificity of each item, and the multidimensional character of the construct, even when there is a common part shared by all of the items.

These percentages of variance also revealed the greater importance of the first factor, to which most of the open- ended problems contribute, in particular solving Bayes' problems had the higher contribution, followed by solving total probability and compound probability problems. All of these problems require a solving process with at least two stages, in the first of which a conditional probability is computed, which is used in subsequent steps (e.g.

product rule). We could interpret this factor as *solving complex conditional probability problems ability.*

Computing simple, joint and conditional probability from a two-way table (item 6) appeared as a separate component, probably because the task format affected performance, a fact which has also been noticed by Ojeda (1996) and Gigerenzer (1994), among other researchers. A third factor showed the relationships between *definition, sample space and computation of conditional probabilities in, with and without replacement situations*; that is, we interpreted this factor as *Level 4 reasoning* in Tarr and Jones (1997) classification.

The remaining factors suggested that the different biases affecting conditional probability reasoning that are described in the justification, appeared unrelated to mathematical performance in problem solving, understanding, building the sample space and computing conditional probability, and to Tarr and Jones's (1997) level 4 reasoning (as related items were not included in the three first factors). Each of the biases (transposed conditional, time axis fallacy, conjunction fallacy, independence/mutually exclusiveness/synchronic setting) also appeared unrelated to one another; in some cases some of them were opposed or related to some semantic units in the mathematical component of understanding conditional probability. For example, independence was linked to the base rate fallacy (where people have to judge whether if the events are independent or not) and opposed to the idea of dependence.

In summary, these results supported our previous hypothesis that biases in reasoning about conditional probability are unrelated to mathematical performance in problem solving and, at the same time, support construct validity evidence for the questionnaire. At the same time it provides information about potential biases students might hold that were used in the design of the teaching experience in the next step of this research.

6.2. EVALUATION OF A TEACHING EXPERIENCE

There is nowadays a tendency to recommend that teaching of Bayesian inference might be included in undergraduate statistics courses as an adequate and desirable complement to classical inference (Lecoutre, 1999; 2006; Lecoutre, Lecoutre & Poitevineau, 2001; Iglesias, Leiter, Mendoza, Salinas & Varela, 2005). Situations where available a priori information can help making an accurate decision and software that facilitates the application of these methods are becoming increasingly available,

Some excellent textbooks whose understanding does not involve advance mathematical knowledge and where basic elements of Bayesian inference are contextualized in interesting examples (e.g., Berry, 1995 or Albert & Rossman, 2001) can help follow these recommendations. There are also a great number of Internet didactic resources that might facilitate the teaching of these concepts (e.g. those available from Jim Albert's web page at, http://bayes.bgsu.edu/). These and other authors (Bolstad, 2002) have incorporated Bayesian methods to their teaching and are suggesting that Bayesian inference is easier to understand than classical inference. This is however a controversial question (see Moore, 1997) and moreover empirical research that analyze the learning of students in natural teaching contexts is still very scarce.

The aim of Study 8 in this thesis was *to explore the possibility of introduce basic ideas of Bayesian inference to undergraduate psychology students and report the extent to which the learning goals were achieved.* The goal of Study 9 *was identifying groups of related concepts, as well as implications between learning objectives with the aim of providing some recommendations about how best organised the teaching of the topics*. In both studies we took into account the results of the previous assessment Study 7.

The sample taking part in this research included 78 students (18-20 year-olds) in the first year of the Psychology Major at the University of Granada, Spain. These students were in the introductory statistics course and volunteered to take part in the experiment. The sample was composed by 17.9% boys and 82.1% girls, which is the normal proportion of boys and girls in the Faculty. These students scored an average of 4.83 (in a scale 0-10) in the statistics course final examination with standard deviation of 2.07.

The students were organized into four groups of about 15-20 students each and attended a short 12 hours long course given by the same lecturer with the same material. The 12 hours were organized into 4 days. Each day there were two teaching sessions with a half-an hour break in between. The first session (2 hours) was devoted to presentation of the materials and examples, followed by a short series of multiple choice items that each student should complete, in order to reinforce their understanding of the theoretical content of the lesson.

In the second session, students in pairs worked in the computer lab with some Excel programs provided by the lecturer to solve a set of inference problems. The Excel programs were as follows:

1. Program Bayes: This program computes posterior probabilities from prior probabilities and likelihood (that should be identified by the students from the problem statement).
2. The program Prodist transforms a prior distribution $P(p=p_0)$ for a population proportion p in the posterior distribution $P(p=p_0|data)$, once the number of successes and failures in the sample are given. Prior and posterior distribution is represented graphically.
3. The program Beta computes probabilities and critical values in the Beta distribution $Be(s,f)$,

where s and f are the successes and failures in the sample.

4. The program Mean computes the mean and standard deviation in the posterior distribution for the mean of a normal population, when the mean and standard deviation are given in the sample and prior population.

In table 8.6 we present a summary of the teaching content. Students were given a printed version of the didactic material that covered this content. Each lesson was organized in the following sections: a) Introduction, describing the lesson goals and introducing a real life situation; b) theory development, using the situation previously presented; c) additional examples of other situations where the same procedures and concepts could be applied, d) some solved exercises, with description of main steps in solving the exercises; e) new problems for students to solve in the computer lab; and f) self assessment. All this material together with the Excel programs was also made available to the students on the web site (http://www.ugr.es/~mcdiaz/bayes) and is also included as Appendix 7 to 9 in the thesis. We added a forum, so that students could consult the teacher or discuss themselves their difficulties, if needed.

Table 8.6. Teaching content and its organization

Lesson	Content	In classroom Session 1	Computer lab Session 2
1	Bayes theorem in the context of clinical diagnose	Prior and posterior probabilities; likelihood; Bayes theorem Subjective probability. Comparison with classical and frequentist probability. Revision of beliefs; sequential application of Bayesian procedures	Solving Bayes Problems: (Program Bayes)
2	Inference for proportion. Discrete case in the context of voting	Sample and population; parameters and statistics; Parameter as random variable; Prior and posterior distribution; Informative and non informative prior distribution. Credible intervals	Computing credible intervals for proportion; assigning non informative and informative prior distributions (Program Prodist)
3	Inference for proportion. Continuous case in the context of production	Generalization to continuous case. Beta distribution, its parameters and shape. Credible intervals; Bayesian tests	Assigning non informative and informative prior distributions Computing credible intervals for proportion; testing simple hypotheses (Program Beta)
4	Inference for the mean of a normal population in the context of psychological assessment	Normal distribution and its parameters; credible intervals and tests for the mean of a normal distribution with known variance; non informative and informative prior distributions	Assigning non informative and informative prior distributions Computing credible intervals for means; testing simple hypotheses (Program Mean)

Two weeks after the end of the teaching, the students were given a questionnaire to assess their understanding of the topic. They were warned to study the topic and prepare for the assessment and were motivated to get a good result in the test.

Questionnaire. A-priori analysis

The BIL (Bayesian Inference Learning) questionnaire (which is presented in Appendix) was made of multiple choice and some open ended items that were developed by the author with the specific aim to cover the most important contents in the teaching. In table 8.7 we describe the contents assessed by the different items in the BIL questionnaire. (In item I18 we considered three different scores).

Table 8.7. Contents assessed in the BLI Questionnaire

Item	Content assessed
I1	Likelihood, conditional probability
I2a	Simple probability
I2b	Conditional probability
I2c	Conditional probability of contrary event
I2d	Joint probability
I3	Parameter as random variable
I4	Prior distribution
I5	Parameter as random variable; difference with statistics
I6	Correct assignment of a non informative prior distribution for proportion
I7	Using the Bayes' theorem as a tool to transform prior into posterior probabilities; table given
I8	Parameters in Beta distribution, defining prior informative distribution for proportion
I9	Parameters in Beta distribution,
I10	Computing credible intervals for proportion; reading Beta tables
I11	Testing simple hypotheses for proportion; reading Beta tables
I12	Properties of credible intervals
I13	Posterior distribution of mean; non informative prior. Known variance
I14	Testing simple hypotheses for means
I15	Posterior distribution of mean; non informative prior. unknown variance
I16	Credible intervals for means
I17	Posterior distribution for mean, informative prior
I18.1	Identifying prior probabilities from a problem statement
I18.2	Identifying likelihood from a problem statement
I18.3	Using the Bayes' theorem as a tool to transform prior into posterior probabilities;
I19	Meaning of likelihood
I20a	Parameters in Beta curve. Spread
I20b	Parameters in Beta curve. Centre

The aim was to assess learning in the following groups of concepts, which in our a-priori analysis were assumed to be the core content of basic Bayesian inference and might cause different types of difficulties to students. We also assumed learning of one of these groups of concepts would not automatically assure the learning of the other groups:

1. *Conditional probability and the Bayes' theorem.* As was argued before, different authors pointed to students' difficulties in understanding conditional probability: *fallacy of the transposed conditional*; causal and chronological conception of conditional probability; confusion between simple, joint and conditional probability. All these errors might cause

difficulties in computing different types of probabilities (item2), understanding of the differences between prior and posterior probability and likelihood (items 1 and 18), and using the Bayes' theorem as a tool to transform prior into posterior probabilities (item 7 and 18).

2. *Parameters as random variables, their distribution, distinction between prior and posterior distribution.* In Bayesian inference, parameters are considered to be random variables with a prior distribution, while in frequentist inference they are assumed to be unknown constants (items 3, 5), a distinction which is not too clear for some students (Bolstad, 2002). Moreover, the aim of Bayesian inference is to transform the prior into a posterior distribution via the Bayes' theorem (item 18). A prior distribution provides all the information for the parameter before collecting the data (item 4), non informative priors are given by uniform distributions and are used when no previous information is available for the parameter (item 6).

   There are different models to represent prior distributions. The Beta distribution was introduced in the teaching, and students had to learn the meaning of its parameters (item 8, 20) and how to select a specific Beta distribution in a particular inference problem (item 9). Students knew the normal distribution from previous lessons. However, they had to learn the rule to compute the posterior distribution for a mean when the prior distribution is normal (item 13; 14, 15, 16). In managing all these distributions, Bayesian statistics uses the rules of probability to make inferences, and that requires dealing with formulae, but actual calculus used is minimal as students only have to understand that probability is given by different types of areas under a density function (Bosltad, 2002). However, the extent to which all of this is grasped by psychology students has still to be assessed.

3. *Logic of Bayesian inference.* The aim of Bayesian inference is updating the prior distribution via the likelihood to get the posterior distribution, which provides all the information for the parameter, once the data have been collected (Bolstad, 2004). However, it is also possible to carry out procedures similar to those used in frequentist statistics, although the interpretation and logic is a little different (Berry, 1995; Lecoutre, 2006). Credible intervals provide the epistemic probability that the parameter is included in a specific interval of values, for the particular sample, while confidence intervals provide the frequentist probability that in a percentage of samples from the same population the parameter will be included in intervals of values computed in those samples. Credible intervals are computed from the posterior distribution (item 17) and students should be able to compute them by using the tables of different distributions

(items 10, 16); they should understand that the interval width increases with the credibility coefficient and decreases with the sample size (item 12).

In Bayesian inference we can compare at the same time different hypotheses; in this case we compute the probabilities for those hypotheses given the data by using the posterior distribution and select the hypothesis with higher probability (item 11). In testing only one hypothesis we either compute the probability for the hypothesis or for the contrary event (item 14); acceptance or rejection will depend on the value of that probability. So, there are some conceptual and interpretative differences between classical and frequentist approaches, but, since both approaches often lead to approximately the same numerical results, students might not understand these differences and confuse both approaches (Iversen 1998).

Results

There were only 4 difficult tasks (percentage of correct responses under 50%). These tasks were (See table 8.8) the following: In item 14 (testing a hypothesis about the mean) the students either made an error in the reasoning by contradiction (choosing distractor c) or did not understand the standardization operation and choose distractor a).

Table 8.8. Results in BIL questionnaire

	% Correct responses	Confidence interval 95%		Credible interval 95%	
		Lim inf	Lim inf	Lim sup	Lim sup
1	88.7	0.808	0.966	0.784	0.943
2a	79.0	0.689	0.891	0.673	0.872
2b	38.7	0.266	0.508	0.276	0.511
2c	29.0	0.177	0.508	0.192	0.412
2d	51.6	0.392	0.639	0.394	0.635
3	66.1	0.543	0.779	0.537	0.766
4	58.1	0.458	0.779	0.456	0.695
5	61.3	0.492	0.734	0.488	0.723
6	50.0	0.376	0.624	0.366	0.604
7	93.5	0.874	0.996	0.845	0.973
8	53.2	0.408	0.656	0.409	0.650
9	85.5	0.767	0.943	0.746	0.921
10	64.5	0.526	0.764	0.520	0.752
11	58.1	0.458	0.704	0.456	0.695
12	53.2	0.408	0.656	0.409	0.650
13	69.4	0.579	0.809	0.570	0.793
14	30.6	0.191	0.421	0.206	0.429
15	40.3	0.281	0.525	0.290	0.527
16	69.4	0.579	0.809	0.570	0.793
17	69.4	0.579	0.809	0.570	0.793
18	79.0	0.689	0.891	0.673	0.872
19	58.1	0.458	0.704	0.456	0.695
20a	82.3	0.728	0.918	0.709	0.897
20b	72.6	0.615	0.837	0.582	0.800

Of course this is a highly complex item, where the logic of testing hypotheses is mixed with knowledge of probability calculus and the standard Normal distribution. Students also found much difficulty in items 2b, and 2c where they confused a conditional probability and its inverse, a problem that have been repeatedly denounced (Bar-Hillel & Falk, 1982; Falk, 1986). We remark that distractors in this item are given only by formulas (instead of using a verbal description such as in item 1) while we found a high percentage of correct responses in item 1 and 7, in spite of the many difficulties and misconceptions described for conditional probability (see Batanero & Sánchez, 2005 for a survey). We conclude that the expressions prior and posterior probabilities and likelihood helped students to better distinguish a conditional probability and its inverse in these items. Finding a posterior distribution for the mean (item 15) was also difficult because students forgot to divide by the square root of the sample size to find the standard deviation in the posterior distribution. All the other tasks had a medium difficulty (between 50-60% correct responses).

Table 8.9. Results in problem solving in lesson 4 (Inference about a mean) (*n*=78)

		% Correct responses	95% Conf, interval		95% Credible interval	
			Lim inf	Lim inf	Lim sup	Lim sup
Ej.1	Correct solution	78.2	0.690	0.874	0.678	0.858
	Typify	83.3	0.750	0.916	0.724	0.891
	Identify the Z interval / Define hypothesis	84.6	0.766	0.926	0.750	0.909
	Compute final distribution	85.9	0.782	0.936	0.765	0.919
	Identify data	88.5	0.814	0.956	0.795	0.937
Ej.2	Correct solution	67.9	0.575	0.783	0.569	0.772
	Typify	87.1	0.797	0.945	0.780	0.928
	Identify the Z interval / Define hypothesis	88.5	0.814	0.956	0.795	0.937
	Compute final distribution	82.0	0.735	0.905	0.721	0.889
	Identify data	78.2	0.690	0.874	0.678	0.858

We also gave the students some problem solving activities and short self-assessment questionnaires in each lesson. In Table 8.9 we show the results of solving the problems related to inference in a mean (normal population). Details of results in the other intermediate assessment are included in Chapter 8 of the thesis and again show that students were capable of solve simple activities of Bayesian inference for proportions and means, including computing credible intervals and carrying out hypotheses tests.

6.3. INTERRELATIONSHIP BETWEEN CONDICIONAL PROBABILITY REASONING AND LEARNING OF BAYESIAN INFERENCE

To study the interrelations and implications between learning objectives we carried out several multivariate analyses, using the CHIC software, Classification Hierarchical, Implicative et Cohesive (Couturier and Gras, 2005). The implication index between two dichotomous variables *a* and *b* in a population is defined by

$$q(a,\overline{b}) = \left[\frac{card(A \cap \overline{B}) - \frac{card(A)card(\overline{B})}{n}}{\sqrt{\frac{card(A)card(\overline{B})}{n}}} \right]$$

where *A* and *B* are the population subgroups where *a* and *b* take the value 1 (Gras, 1993; 1996; Gras & Ratsima-Rajohn, 1996). This index follows the normal distribution *N(0,1)*, and from there an intensity for the implication $a \Rightarrow b$ is defined by

$$\varphi(a,\overline{b}) = \Pr ob[car(X \cap \overline{Y}) \leq card(A \cap \overline{B})],$$

where *X* and $\overline{Y}$ are dichotomous independent random variables having the same cardinal than *A* and $\overline{B}$ respectively (Lerman, Gras & Rostam, 1981a & b). In our study we have a total of $C_{21,\,2}$ implication indexes among the 21 subítems in the LBI questionnaire. The software CHIC computes these indexes and provides a graph with all the implications which are significant to a given significance level.

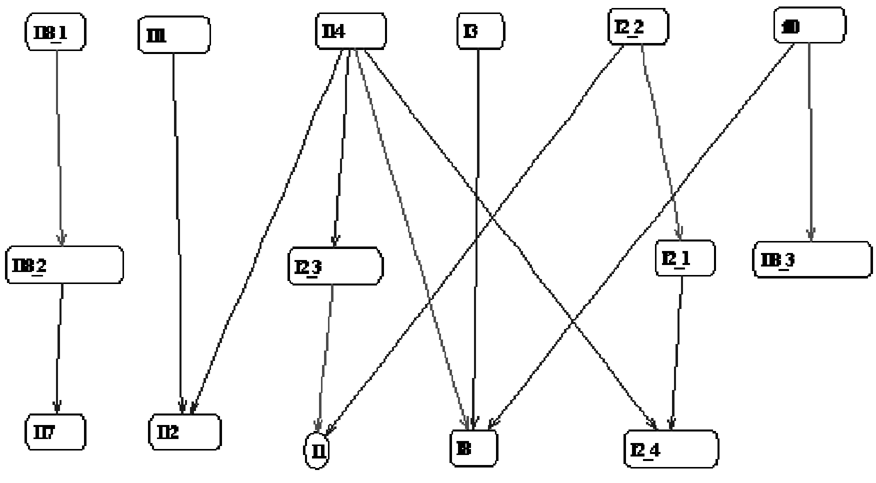
The implication $a \Rightarrow b$ in our study is interpreted in the sense that when a students correctly solves item *a* there is higher probability for him /her to solve item *b*. In this sense the implicative graph provides a possible order to introduce different concepts and procedures whose understanding is assessed in those items in the teaching of the topic. Before carrying out the implicative analysis we checked the assumptions of the method; experimental units of variables, and independence of responses by different students. We assumed a binomial model for the responses; that is, we assumed each student having same likelihood to correctly solve the items (Lerman, 1991), as in fact these are the hypotheses assumed in classical theory of tests.

In Figure 8.3 we present the implicative graph with all the relationship that were significant at 99% level (red) or 95% level (blue), We observe that the implication relationship is asymmetrical and the sense of implication is showed by the arrows in the graph.

If we study the relationships higher than 99% in the graph, we observe that students who correctly answer item I18_2 (correct identification of likelihood, which is given by a conditional probability) have better likelihood to answer I18_1 (correct identification of prior probabilities, which are given by simple probabilities). Correct performance in I10 (identifying probabilities and critical values from the Beta distribution table and computing credible intervals for a proportion) facilitate correct computation of posterior probabilities with Bayes theorem (I18_3). Both tasks involve computing probabilities but the first one is more complex. Then correct computation of conditional probabilities implies correct computation of join and single probabilities (I2_1, I2_2, I2_4).

As regards implications higher than 95% (blue in the diagram) we observe that students who correctly perform a Bayesian hypothesis test (I14 or I11) increase their likelihood to correctly interpret credible intervals (I12), possibly because all the ideas in understanding the second task are involved in the first one, which adds the need to understand the logic of proof by contradiction. I14 implies I2_3, the computation of conditional probability for a contrary event, but, again mastering the idea of proof by contradiction involves correct reasoning on both conditional reasoning and complementation. Students who visualize parameters as radon variables (I3) or compute probabilities for the Beta function and credible intervals for proportions (I10) perform better in correctly assigning a Beta informative prior distribution (I8), a task that is also facilitated by I14.

Figure 8.3. Implicative graph with significant implications at 99 and 95%



I2_3 (computing the conditional probability for the contrary event) or I2_2 (computing conditional probability) facilitate I1, distinguishing prior and posterior probabilities and likelihood (all these ideas are supported on correct conditional reasoning); I2_2 facilities computing simple probability (I1) and both of them together facilitate the computation of join probabilities (I2_4), another task which is easier for those who succeeded in I14 (testing hypotheses).

Implicative hierarchy of learning outcomes

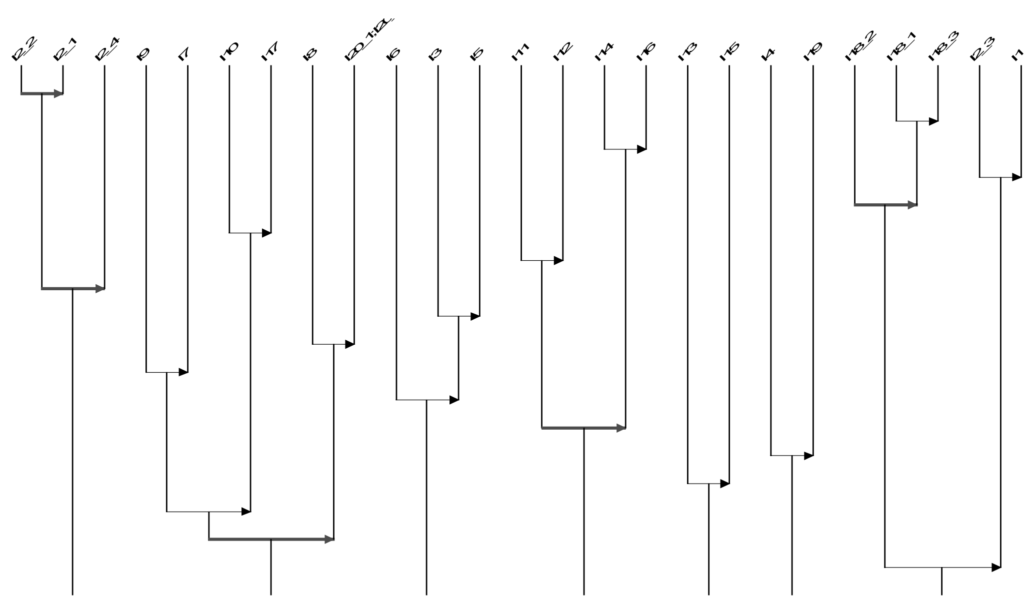
Once the isolate implications between items were studied we carried out an implicative classification analysis. This is an algorithm, which uses the implicative indexes in a set of variables to study the internal cohesion of some variables subsets (Lahanier-Reuter, 2001; Couturier, Gras & Guillet, 2004). The cohesion between two variables *a* and *b* is defined by $c(a,b) = \sqrt{1-H^2}$ where H is the entropy for the two variables, and varies between 0 and 1. The cohesion for a class of variables is defined by (Gras, Kuntz & Briand, 2001):

$$C(\underline{A}) = \left[\prod_{\substack{i \in \{1,\dots r-1\} \\ j \in \{2,\dots r\}, j>i}} c(a_i, a_j) \right]^{\frac{2}{r(r-1)}}$$

Then, given two sets of variables $\underline{A}$ and $\underline{B}$ the strength of implication from $\underline{A}$ to $\underline{B}$ is defined by (Couturier, 2001):

$$\psi(\underline{A}, \underline{B}) = \left[\sup_{i \in \{1,\dots r\}\, j \in \{1,\dots s\}} \varphi(a_i, \overline{b}_j) \right]^{rs} \left[C(\underline{A}).C(\underline{B}) \right]^{1/2}$$

Figure 8.4. Implicative hierarchy with 95% node



The software CHIC builds an implicative hierarchy in the set of variables, taking into account both the maximal cohesion into each class and the higher implication from a class to another. In Figure 8.4 we present the hierarchy produced. There are four significant clusters:

- Group 1. Items (I2_2) and (I2_1) which join to (I2_4), all of them related to probability. The student who correctly computes conditional probabilities (I2_2), correctly perform simple (I2_1) and compound probability (I2_4). The higher difficulty of conditional probability as regards simple and compound probabilities is then confirmed.
- Group 2: Prior and posterior distributions and Beta curves. Item I9, I7, I10, I17, I8 and the two parts of I20. Students who are able to interpret the parameters in the Beta curve (I9) and understand how posterior distributions are get from prior distributions and likelihood through the Bayes theorem (I7) succeeded better in getting a credible interval for proportions in the continuous case; a task that requires interpreting probabilities of Beta curves, and understanding the concept of posterior probability, as well as the concept of credible interval. They also performed better in discriminating prior and posterior distribution of the mean (I17). All of this lead to better choosing a non informative prior distribution for proportions in the continuous case through the Beta Curve (I8) and graphically interpreting the parameters in Beta curves (I20).
- Group 3 (Items I11, I12, I14 and I 16) group a set of Bayesian inference tasks. Being able of correctly test a hypothesis for proportions (I11) increases the likelihood of correctly interpret credible intervals (I12); and these two task are associated with correctly testing a hypothesis about the mean (I14), and correctly computing a credible interval for the mean (I16). All this knowledge is specifically related to the Bayesian methods which are based on conditional probability and also in the logic of scientific inference.
- Group 4: Moreover there is a second group of tasks related to conditional probability (the different parts of Item 18, I2-3 and I1). Correct identification of likelihood from a problem statement (I18_2) facilitates correct identification of prior probability (I18-1) and this lead to correct computation of posterior probabilities (I18_3). These three abilities lead to better identification of conditional probabilities for the contrary event (I2-3) and discrimination between prior probability, likelihood and posterior probabilities in the context of a problem (I1).

Other groupings of items that are non significant were as follows:

- Group 5: Items I6 (assigning adequate prior distribution for the non informative case to proportions in the discrete case), I3 (understanding parameters as random variables) and I5 (discrimination between parameters and statistics); all these tasks are related to understanding parameters from a Bayesian point of view.
- Group 6: Items I13 (Posterior distribution of mean when variance is known) and I15 (posterior distribution of mean when the variance is unknown: they are related to specific knowledge the students should remember.
- Group 7: I4 (concept of prior distribution) and I19 (concept of likelihood).

In summary these implications point to three groups of concepts relevant for students' introduction to the elementary ideas of Bayesian inference that should be taken into account in planning the teaching and support our previous a-priori analysis of the BIL questionnaire:

1. *Conditional probabilistic reasoning* (as shown in groups 1 and 4), a theme where many biases have been described in the literature, but which is basic in defining posterior probabilities and distributions and likelihood, as well as in understanding the logic of credible intervals and hypothesis testing. Results also suggested that formulas for different types of probability were harder than verbal expressions for students to understand. Perhaps we should take into account Feller's suggestion (1973, p. 114) that "*conditional probability is a basic tool of probability theory, and it is unfortunate that its great simplicity is somewhat obscured by a singularly clumsy terminology".*
2. *Probability distributions,* its parameters (visualized as random variables), the distinction of prior and posterior distribution of parameters and assignment of prior distributions for informative and non informative cases (Groups 2, 5, 6 and 7). In our teaching we limited to Beta and Normal distributions, since the time available for teaching was restricted, but still so, the understanding of Beta curves appeared as a separated subgroup, as well as remembering the rules for known and unknown variance in inference about normal distributions. The difficulties to understand the different conception of parameters in Bayesian and frequentists statistics also appeared as a separated subgroup.
3. *Logic of Bayesian inference* (Group 3), that is, understanding the logic for computing and interpreting credible intervals and testing simple hypothesis. Performance in these tasks is in fact supported in understanding the previous two groups of concepts, most of which are not specific to Bayesian reasoning. However, limitation of teaching time leads some lecturers to reduce the teaching of the same and to try to pass directly from data analysis to

inference. Teaching of Bayesian inference therefore should only be started when the previous groups of concepts are well understood by the students.

7. SUMMARY AND MAIN CONTRIBUTIONS

In this Thesis we focus on the use of Bayesian inference in the field of Psychology from different perspectives. Below we summarise these perspectives and the main conclusions /contributions achieved for each of them.

Current practice of statistics

We produced a synthesis of main criticisms of current statistical practices in psychology, the reported errors and the possible contribution of Bayesian inference to solve part of the denounced errors. As a consequence we suggested the need to introduce the teaching of elementary Bayesian methods in psychology and to carry out empirical research to assess the suitability of this teaching.

Application of Bayesian methods in psychometrics

We analysed the implications of a Bayesian approach to Classical Tests Theory and deduced estimation procedures for some of the psychometrics features of items and questionnaires. These procedures were applied in the process of building a questionnaire to assess conditional probability reasoning (CPR), which is also justified in the thesis. We also developed some Excel programmes to carry out the main computations.

Assessing conditional probability reasoning in undergraduates

We used the CPR questionnaire to carry out a detailed evaluation in a sample of 414 students after teaching of the topics. The complex relationship between probabilistic concepts and intuition was shown in our study, where probabilistic biases were widespread in students, even in those with good problem solving probability. Consequently, our research suggest the need of reinforcing the study of conditional probability in the teaching of data analysis at University level, although it also provides arguments for a change of approach in this teaching. Following Nisbett and Ross' recommendations (1980, p. 280) students should be "*given greater motivation to attend closely to the nature of the inferential tasks that they perform and the quality of their performance"* and consequently "*statistics should be taught in conjunction with material on intuitive strategies and inferential errors*" (p. 281) of the kind presented in their book. In this sense we support Rossman and Short (1995), who suggest conditional probability can be taught in line with new statistics education ideas, in presenting a variety of applications to realistic problems, proposing interactive activities and using technology to facilitate learning.

Studying the suitability of teaching elementary Bayesian inference to undergraduates

We developed a teaching material that takes into account the previous analyses, as well as previous research in statistics education and the type of students. This material was trailed with a sample of 78 students, and data on the students' learning at the end of the experience showed that most instructional objectives were achieved by the students.

The implicative and cohesive classification analyses also supported the interrelationship between learning Bayesian inference and understanding conditional probability as it was previously assumed. On the other hand, the obtained classes in the implicative hierarchy provided us with information about the concepts whose understanding is related and their relative difficulty. This is a potential help to prepare didactic materials and to organize the teaching of the topic.

In summary, we think that this thesis opens a new perspective for research in the Behavioural Sciences Research Methods, both from the strictly methodological point of view (implementing and applying Research Methods) and from the didactic point of view. Partial results of each of the mentioned contributions have been published in diverse journals and international conferences (See references).

In the present convergence process to the European Space of Higher Education, it is not only possible, but required that lecturers in this area carry out research on the didactics of research methods, including non-traditional topics. Only by means of systematic research we can enrich our educational practice and contribute to improve the application of research methods. It is therefore expected that new studies continue the research started in this Thesis.

REFERENCIAS

Abelson, R. P. (1997). On the surprising longevity of flogged horses: Why there is a case for the significance test? *Psychological Science, 8 (1)*, 12 – 14.

Afifi, A, A. y Clark, V. (1990). *Computer-aided multivariate analysis.* New York: Van Nostrand.

Agnoli, F. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. *Cognitive Psychology*, *21(4),* 515.

Albert, J. (1995). Teaching inference about proportion. Using Bayes and discrete models. *Journal of Statistics Education, 3(3).* On line: http://www.amstat. org/publications/jse/v3n3/albert.html.

Albert, J. (1996). *Bayesian computation using Minitab*. Belmont, CA: Duxbury Press.

Albert, J. (2000). Using a sample survey project to assess the teaching of statistical inference. *Journal of Statistics Education, 8(1).* On line: http://www.amstat. org/publications/jse/secure/v8n1/albert.html.

Albert, J. (2002). Teaching introductory statistics from a bayesian perspective. En B. Philips (Ed), *Proceedings of the 6th International Conference on Teaching Statistics.* Cape Town, South Africa: International Statistical Institute.

Albert, J. H. y Rossman, A. (2001). *Workshop statistics. Discovery with data. A bayesian approach.* Bowling Green. OH: Key College Publishing.

Alvarado, H. (2004). *Significado del teorema central del límite en textos universitarios para ingenieros.* Trabajo de Investigación Tutelada. Universidad de Granada.

American Psychological Association, American Educational Research Association y National Council on Measurement in Education (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

American Psychological Association, American Educational Research Association y National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Anastasi, A. y Urbina, S. (1998). *Test psicológicos.* Prentice Hall.

Ares, V. M. (1999). La prueba de significación de la «hipótesis cero» en las investigaciones por encuesta. *Metodología de Encuestas,* 1, 47-68.

Ayçaguera, L. y Benavides, A. (2003). Apuntes sobre subjetividad y estadística en la investigación en salud. *Revista Cubana de Salud Pública, 29(2),* 170-173. On line:

http://scielo.sld.cu/scielo.php?pid=S0864-34662003000200012&script=sci_arttext&tlng=es.

Ayçaguera, L. y Suárez, P. (1995). ¿Qué es la inferencia bayesiana? *JANO,* 1132, 1542. On line: http://www.atheneum.doyma.es/Socios/sala_l/lec06est.htm.

Aydýnlý, G., Härdle, W. y Rön, B. (2003). E-learning/e-teaching of statistics: A new challenge. En J. Engel (Ed.), *Proceedings of the IASE Satellite Conference Statistics Education and the Internet*. Berlin: International Association for Statistical Education. On line: http://www.stat.auckland.ac. nz/~iase/publications/6.

Azorín, F. y Sánchez-Crespo, J. L. (1986). *Métodos y aplicaciones del muestreo.* Madrid: Alianza Editorial.

Bacelar-Nicolau, H.(1992). Probabilistic similarity coefficients for variables in hierarchical agglomerative clustering models, En S. Jolly y G. Le Calve (Eds.), *Abstracts of the International Meeting on Distance Analysis, DISTANCIA'92* (pp. 93-94). Universidad de Rennes.

Bacelar-Nicolau, H. (2003). *Introdução à Análise Classificatória Hierárquica Ascendente: Modelos de ACHA*. Notas e Comunicações do LEAD

Bailleul, M. y Gras, R. (1994). L'implication statistique entre variables modales. *Mathématiques, Informatique et Sciences Humaines*, *128*, 41-57.

Bailleul, M. (2001). Des réseaux implicatifs pour mettre en évidence des representations. *Mathématiques, Informatique et Sciences Humaines*, *135,* 154-155.

Bakan, D. (1997). The test of significance in psychological research. En D. E. Morrison y R. E. Henkel, (Eds.). *The significance tests controversy: A reader* (pp. 231 – 251). Chicago: Aldine.

Bar – Hillel, M. (1987). The base rate fallacy controversy. En R. W. Scholz (Ed.), *Decision making under uncertainty.* (pp 39 – 61) Amsterdam: North Holland.

Barbero, M. (2003). *Psicometría II. Métodos de elaboración de escalas*. Madrid: UNED.

Batanero, C. (2000). Controversies around significance tests. *Mathematical Thinking and Learning, 2(1-2),* 75 – 98.

Batanero, C. (2001). (Ed.). *Training researchers in the use of statistics.* Granada: International Association for Statistical Education e International Statistical Institute.

Batanero, C. y Díaz, C. (2005). Análisis del proceso de construcción de un cuestionario sobre probabilidad condicional. Reflexiones desde el marco de la TFS. En A.

Contreras (Ed.). *Investigación en Didáctica de las Matemáticas. I Congreso Internacional sobre Aplicaciones y Desarrollos de la Teoría de las Funciones Semióticas* (pp. 13 – 36). Jaén: Universidad de Jaén.

Batanero, C. y Díaz, C. (2006). Methodological and didactical controversies around statistical inference. *Actes du 36iémes Journées de la Societé Française de Statistique.* CD ROM. Paris: Societé Française de Statistique.

Batanero, C. y Díaz, C. (En prensa). Meaning and understanding of mathematics. The case of probability. En J. P Van Bendegen y K. Fraçois (Eds), *Philosophical Dimensions in Mathematics Education.* New York: Springer.

Batanero, C., Díaz, C. y de la Fuente, I. (En prensa). Alcune consideración sull'insegnemento della probabilitá condizionata. *Nuova Secondaria.*

Batanero, C., Estepa, A., Godino, J. D. y Green D. R. (1996). Judgments of association in contingency tables: An empirical study of students' intuitive strategies and preconceptions. *Journal for Research in Mathematics Education,* 27 (2), 151-169.

Batanero, C., Henry, M. y Parzysz, B. (2005). The nature of chance and probability. En G. A. Jones (Ed.), *Exploring probability in schools: Challenges for teaching and learning* (pp. 15-37). New York: Springer.

Bauersfeld, H. (1995).The structuring of the structures: Development and function of mathemathizing as a social practice. En L. Steffe y J. Gale (Eds.), *Constructivism in Education.* (pp. 137-158). Hillsdale, NJ: Lawrence Erlbaum.

Beltrán, M. (Ed.) (2001). *Actas de las Jornadas Europeas de Enseñanza y Difusión de la Estadística.* Mallorca: Instituto Balear de Estadística.

Ben-Zvi, J. y Garfield, J. (2004) (Eds.), *The challenge of developing statistical literacy, reasoning and thinking.* Dordrecht: Kluwer.

Benzecri (1983). J. P. *Histoire et préhistoire de l'analyse des données*. Paris: Dunod.

Bernard, J. M. (1998). Bayesian inference for categorised data. En H. Rouanet et al. (Eds.), *New ways in statistical methodology* (pp. 159 – 226). Berna: Peter Lang.

Bernardo, J. M. (1981). *Bioestadística. Una perspectiva bayesiana.* Barcelona: Vicens-Vives.

Bernardo, J. M. (2003). Bayesian Statistics. En R. Viertl (Ed.), *Encyclopaedia of life support systems (EOLSS). Probability and statistics*. Oxford, UK: UNESCO, On line: http://www.uv.es/~bernardo/BayesStat.pdf.

Bernardo, J. M. (2006). A Bayesian mathematical statistics primer. En A. Rossman y B. Chance, (Eds.), *Proceedings of the Seventh International Conference on Teaching*

Statistics. CD ROM. Salvador de Bahia: International Association for Statistical Education.

Bernardo, J. M. y Smith, A. F. M. (1994). *Bayesian Theory*. New Your. Wiley.

Berry, D. A. (1995). *Basic statistics: A Bayesian perspective.* Belmont, CA: Wadsworth.

Biehler, R. (1997a). Software for learning and for doing statistics. *International Statistical Review*, *65(2),* 167-190.

Biehler, R. (1997b). Students' difficulties in practicing computer-supported data analysis: Some hypothetical generalizations from results of two exploratory studies. En J. B. Garfield y G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics* (pp. 176-197). Voorburg, The Netherlands: International Statistical Institute.

Biehler, R. (2001). Developing and assessing students' reasoning in comparing statistical distributions in computer supported statistics courses. Trabajo presentado en el *Statistics Literacy and Reasoning Research Forum 3.* Armidale. Australia.

Birnbaum, I. (1982). Interpreting statistical significance. *Teaching statistics. 4,* 24 – 27.

Bisquerra, R. (1989). *Métodos de investigación educativa.* Barcelona: P.P.U.

Black, M. (1979). *Inducción y probabilidad*. Madrid: Cátedra.

Bloom, B. S. (1956). *Taxonomy of educational objectives, Handbook 1: The cognitive domain.* New York: McGraw Hill.

Bolívar, A. (1998). Tiempo y contenido del discurso curricular en España. *Revista de currículum y formación del profesorado, 2(2),* On line: http://www.ugr.es /~recfpro/rev22ART4.pdf.

Botella, J., León, O. G. y San Martín, R. (1993). *Análisis de datos en psicología I.* Madrid: Pirámide.

Boldstad, W. M. (2002). Teaching Bayesian statistics to undergraduates: Who, what, where, when, why, and how. En B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics.* Ciudad del Cabo: International Association for Statistical Education. CD ROM.

Bolstad, W. (2004). *Introduction to Bayesian statistics.* New York: Wiley.

Borges, A., San Luis, C., Sánchez, J. A. y Cañadas, I. (2001). El juicio contra la hipótesis nula: muchos testigos y una sentencia virtuosa. *Psicothema, 13 (1),* 174-178.

Borges, A. y Sánchez Bruno, A. (2004). Algunas consideraciones metodológicas relevantes para la investigación aplicada. *Revista Electrónica de Metodología Aplicada, 9 (1),* pp. 1-11.

Box, G. P. y Tiao, G. C. (1992). *Bayesian inference in statistical analysis.* Nueva York: Wiley.

Brennan, R. L. (1983). *Elements of generalizability theory.* Iowa: Blackwell.

Brousseau, G. (1997). *Theory of didactical situations in mathematics.* Dordrecht: Kluwer.

Burrill, G. y Camden (Eds.) (2006). *Curricular development in statistics education: International Association for Statistical Education 2004 Roundtable.* Voorburg, the Netherlands: International Statistical Institute. On line: http://www.stat.auckland.ac.nz/~iase/publications.php

Cabriá, S. (1994). *Filosofía de la estadística.* Valencia: Servicio de Publicaciones de la Universidad.

Cañadas, I., Borges, A., Sánchez, A. y San Luis, C. (2000). Estudio de la potencia de los contrastes de medias con dos y tres grupos con tamaño de efecto pequeño y en condiciones de no normalidad y homo-heterocedasticidad. *Psicothema,* 12 (2) , 114-116.

Carmines, E. G. y Zeller, R. A. (1979). *Reliability and validity assesment*. London: Sage.

Carmona, J. (2004). Una revisión de las evidencias de fiabilidad y validez de los cuestionarios de actitudes y ansiedad hacia la estadística. *Statistics Education Research Journal*. *3(1).* On line. Disponible en www.stat.auckland.ac.nz/ ~iase/serj/SERJ3(1) _marquez.pdf.

Carvalho, C. (2005). Literacia estatística. En CIEFCUL (Ed.), *Itinerários Investigar em Educação (*pp. 35-44). Lisboa: Centro de Investigação em Educação da FCUL. CD-ROM.

Castro-Posada, J. (2001). *Metodología de la investigación*. Fundamentos. Salamanca: Amaru.

Chow, L. S. (1996). *Statistical significance: Rationale, validity and utility.* London: Sage.

Cohen, J. (1990). Things I have learnt so far. *American Psychologist, 45,* 1304 - 1312.

Cohen, J. (1994). The earth is round (p<.05). *American Psychologist, 49(12),* 997 – 1003.

Congdon, P. (2003). *Applied Bayesian modelling.* New York: Wiley.

Cook, T. y Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field set tings.* Chicago: Rand. Mc. Nally Publishing Company.

Couturier, R. (2001). Subjects categories contribution in the implicative and the similarity analysis. En A. Gagatsis (Ed.), *Learning in mathematics and science and educational technology* (Vol. 2), (pp. 369-376). University of Cyprus.

Couturier, R. y Gras, R. (2005). CHIC: Traitement de données avec l'analyse implicative. En G. Ritschard y C. Djeraba (Eds.), *Journées d'extraction et gestion des connaissances (EGC'2005)* (Vol. 2), (pp. 679.684). Universidad de Lille.

Couturier, R., Gras, R. y Guillet, F. (2004). Reducing the number of variables using implicative analysis. En D. Banks, L. House, F. R. McMorris, P. Arabie y W. Gaul (Eds.), *Classification, clustering, and data mining application. Proceedings of the Meeting of the International Federation of Classification Societies Conference* (pp.277—285), Springer-Verlag.

Cronbach, L. J. (1988). Five perspectives on the validity argument. En H. Wainer y H. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Erlbaum.

Cuadras, C. M. (1981). *Análisis multivariante*: Barcelona: Eudeba.

Cuadras, C. M., Echevarría B., Mateo, J. y Sánchez, P. (1984). *Fundamentos de estadística. Aplicación a las ciencias humanas.* Madrid: Promociones Publicaciones Universitarias.

Cumming, G., Williams, J. y Fidler, F. (2004). Replication, and researchers' understanding of confidence intervals and standard error bars. *Understanding Statistics, 3,* 299-311.

Curtis, C. (Ed.) (2002). *Actas de las Jornadas Interamericanas de Enseñanza de la Estadística.* Buenos Aires: Universidad Nacional Tres de Febrero. CD ROM.

Dane, F. C. (1990). *Research methods*. Thompson. Pacific Grow. CA.

David, J., Guillet, F., Philipp, V. y Gras, R. (2005). Implicative statistical analysis applied to clustering of terms taken from a psychological text corpus *Applied Stochastic Models and Data Analysis (ASMDA 2005).* On Line http://asmda2005.enst-bretagne.fr/IMG/pdf/proceedings/201.pdf.

Davidson, R. y Swift, J. (Eds.) (1988). *Proceedings, 2nd International Conference on Teaching Statistics.* British Columbia, Canada: International Association for Statistics Education.

De la Fuente, E. I., De la Fuente, L., García, J., Iglesias, S., Martín I., Ramírez, I, y Cañadas, G. (2002). El análisis bayesiano de datos en psicología. Distribuciones conjugadas en un problema de inferencia estadística. *Metodología de las Ciencias del Comportamiento.* Volumen Monográfico, 189-191.

De la Fuente, E. I. y Díaz, C. (2003). Reflexiones sobre los métodos inferenciales en psicología. *Libro de resúmenes del VIII Congreso de Metodología de las Ciencias Sociales y de la Salud* (pp. 326 – 327). Valencia: Departamento de Metodología de las Ciencias del Comportamiento y Asociación Española de Metodología de las Ciencias del Comportamiento.

De la Fuente, E. I. y Díaz, C. (2004). Controversias en el uso de la inferencia en la investigación experimental. *Metodología de las Ciencias del Comportamiento*, Volumen especial 2004, 161-167

De la Fuente, E. I., Díaz, C. y Cañadas, G. (2005). Algunas razones para introducir la inferencia bayesiana en la formación metodológica en el campo de la psicología. En I. de la Fuente et al. (Eds.), *IX Congreso de Metodología de las Ciencias Sociales y de la Salud. Libro de resúmenes* (p. 104). Universidad de Granada.

De la Fuente, E. I., García, J. y De la Fuente, L. (2002). Estadística bayesiana en la investigación psicológica. *Metodología de las Ciencias del Comportamiento*, *4*, 185-200.

De Finetti, B. (1974). *Theory of probability.* London: John Wiley.

De Groot, M. H. (1988). *Probabilidad y estadística*. Delaware: Addisson Wesley.

Delgado, J. M. y Gutiérrez, J. (1994) *Métodos y técnicas cualitativas de investigación en ciencias sociales*. Madrid: Síntesis Psicología.

DelMas, R. C., Garfield, J. B. y Chance, B. L. (1998). Exploring the role of computer simulations in developing understand of sampling distributions. Trabajo presentado en el *American Educational Research Association*. *Annual Meeting*. Montreal.

Díaz, C. (2004). *Elaboración de un instrumento de evaluación del razonamiento condicional. Un estudio preliminar*. Trabajo de Investigación Tutelada. Universidad de Granada.

Díaz, C. (2005). Evaluación de la falacia de la conjunción en alumnos universitarios. *Suma, 48,* 45-50.

Díaz, C. y Batanero, C. (2005). La probabilidad condicional en los textos de estadística para psicología *Actas del V CIBEM, Congreso Iberoamericano de Educación Matemática.* Oporto: Sociedad Portuguesa de Profesores de Matemáticas. CD ROM.

Díaz, C., Batanero, C. y Cobo, B. (2003). Fiabilidad y generalizabilidad. Aplicaciones en evaluación educativa. *Números, 54*, 3 – 21.

Díaz, C. y de la Fuente, E. I. (2004). Controversias en el uso de la inferencia en la investigación experimental. *Metodología de las Ciencias del Comportamiento,* Volumen especial 2004, 161-167.

Díaz, C. y de la Fuente, E. I. (2005a). Recursos para la enseñaza del razonamiento condicional en Internet. Trabajo presentado en el *Congreso Internacional* el *Profesorado ante el Reto de las Nuevas Tecnologías en la Sociedad del Conocimiento.* Universidad de Granada.

Díaz, C., y de la Fuente, I. (2005b). Razonamiento sobre probabilidad condicional e implicaciones para la enseñanza de la estadística. *Epsilon, 59*, 245-260.

Díaz, C. y de la Fuente, E. I. (2005c). Conflictes semiòtics en el càlcul de probabilitats a partir de taules de doble entrada1. *Biaix*, *24,* 85-91.

Díaz, C., y de la Fuente, E. I. (2005d). Construcción de un cuestionario sobre comprensión de la probabilidad condicional. En J. Ortiz y A. Montenegro (Eds.), *Actas del XV Simposio de Estadística.* Bogotá: Universidad Nacional de Colombia. CD ROM.

Díaz, C. y de la Fuente, E. I. (2006). Assessing psychology students' difficulties with conditional probability and bayesian reasoning. En A. Rossman y B. Chance (Eds.), *Proceedings of Seventh International Conference on Teaching of Statistics.* Salvador de Bahia: International Association for Statistical Education. CD ROM.

Díaz, C. y de la Fuente, E. I. (En prensa). Dificultades en la resolución de problemas bayesianos: un estudio exploratorio en estudiantes de psicología. *Educación Matemática.*

Díaz, C., de la Fuente, I y Bacelar-Nicolau, H. (En prensa). Cluster and factor analysis of students' responses to the CPR questionnaire. En L. Bacelar y H. Bacelar (Eds.), *Actas de las XIII Jornadas de Classificaçao e Analise de Dados*. Lisboa: Associaçao Portuguesa de Analise de Dados.

Díaz, C., de la Fuente, E. I. y Batanero, C. (2004a). Statistical inference and experimental research. should we revise our educational practices. *Libro de resúmenes de ICME-10* (p. 15). Copenhague: International Comission on Mathematical Instruction.

Díaz, C., de la Fuente, E. I. y Batanero, C. (2004b). Competencia de estudiantes de psicología en la resolución de problemas bayesianos. *Libro de resúmenes de la*

XVIII Reunión Latinoamericana de Matemáticas Educativa (p. 244). Tuxtla Gutiérrez, México: Comité Latinoamericano de Matemática Educativa.

Dunn, O, J. y Clarck,V. A. (1987). *Applied statistics: Analysis of variance and regressio*n. New York: John Wiley.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. En D. Kahneman, P. Slovic y Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Edwards, W., Lindman, H. y Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70*, 193-242.

Einhorn, H. J. y Hogart, R. M. (1986). Judging probable cause. *Psychological Bulletin, 99,* 3 – 19.

Engel, J. (2002). Activity-based statistics, computer simulation and formal mathematics. En Phillips, B. (Ed.). *Proceedings of the Sixth International Conference on Teaching of Statistics.* Ciudad del Cabo: International Association for Statistical Education. CD ROM.

Engel, J. (2003). (Ed.). *Statistics education and the Internet.* Berlín: International Association for Statistical Education.

Engel, J. y Vogel, M. (2004). Mathematical problem solving as modeling process. En H. Henn y W. Blum (Ed.). *Proceedings of ICMI-Study 14: Applications and Modelling in Mathematics Education* (pp. 83-88). Dormunt: International Commission on Mathematical Instruction.

Ernest, P. (1994). Varieties of constructivism: Their metaphors, epistemologies and pedagogical implications. *Hiroshima Journal of Mathematics Education 2,* 1-14.

Ernest, P. (1998). *Social constructivism as a philosophy of mathematics*. New York: SUNY.

Estepa, A. (1993). *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores*. Tesis doctoral. Universidad de Granada.

Estes, W. K. (1997). Significance testing in psychological research: Some persisting issues. *Psychological Science*, 8(1), 18 – 20.

Estrada, A. y Díaz, C. (2006). Computing probabilities from two way tables. An exploratory study with future teachers. En A. Rossman y B. Chance (Eds.), *Proceedings of Seventh International Conference on Teaching of Statistics.* Salvador (Bahia): International Association for Statistical Education. CD ROM.

Estrada, A., Díaz, C. y de la Fuente, E. I. (2006). Un estudio inicial de sesgos en el razonamiento sobre probabilidad condicional en alumnos universitarios En P. Bolea, M. J. Gonzáles y M. Moreno (Eds.), *Actas del IX Simposio de la SEIEM* (pp. 277-284). Huesca: Sociedad Española de Investigación en Educación Matemática.

Falk, R. (1986a). Misconceptions of statistical significance. *Journal of Structural Learning, 9,* 83 – 96.

Falk, R. (1986b). Conditional probabilities: insights and difficulties. En R. Davidson y J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics*. (pp. 292 – 297). University of Victoria.

Falk, R. (1989). Inference under uncertainty via conditional probability. En R. Morris (Ed.), *Studies in mathematics education,* vol. 7 (pp. 175 – 184). Paris: UNESCO.

Falk, R. y Greenbaum, C. W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory and Psychology, 5 (1),* 75 – 98.

Feller, W. (1973). *Introducción a la teoría de las probabilidades y sus aplicaciones.* México: Limusa.

Feldt, L. S. y Brennan, R. L. (1991). Reliability. En R. Linn (Ed.), *Educational measurement* (pp. 105-146). Nueva York: McMillan.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research, 50,* 123-129.

Fidler, F. (2002). The fifth edition of the APA publication manual: Why statistics recommendations are so controversial. *Educational and Psychological Measurement*, *62 (5),* 749-770.

Finch, S., Cumming, G. y Thomason, N. (2001). Reporting of statistical inference in the Journal of Applied Psychology: Little evidence of reform. *Educational and Psychological Measurement, 61,* 181-210.

Fisher, R. A. (1956). Mathematics of a lady testing tea, En J. Newman (Ed.), *The world of mathematics* Vol., III. Simon and Schuster. Traducido como Las matemáticas de la catadora de té. En J. R. Newman (Ed.), *El mundo de las matemáticas* (Vol. 3, pp. 194 – 203). Barcelona: Grijalbo, 1979.

Frías, M. D., Pascual, J. y García, J. F. (2000). Tamaño del efecto del tratamiento y significación estadística. *Psicothema, 12 (2)*, 236-240.

Frías, M. D., Pascual, J. y García, J. F. (2002). La hipótesis nula y la significación práctica. *Metodología de las Ciencias del Comportamiento, 4 (especial),* 181-185.

Fox, D. J. (1981). *El proceso de investigación en la educación*. Pamplona: Eunsa.

Gal, I. y Garfield, J. (Eds.) (1997). *The assessment challenge in statistics education*. The Netherland: IOS Press.

Garfield, J. B. y Burrill, G. (Eds.) (1997). *Research on the role of technology in teaching and learning statistics.* Voorburg: International Association for Statistical Education e International Statistical Institute.

Gelman, A., Carlin, J. B. Stern, H. S. y Rubin, D. B. (2003). *Bayesian data analysis*. Londres: Chapman and Hall.

Ghiglione, R. y Matalón, B. (1991). *Les enquêtes sociologiques. Théorie et practique.* París: Armand Colin.

Gil Flores, J. (1994). *Análisis de datos cualitativos*. Barcelona. P.P.U.

Gingerenzer, G. (1993). The superego, the ego and the id in statistical reasoning. En G. Keren y C. Lewis (Eds.), *A handbook for data analysis in the behavioural sciences: Methodological issues* (pp. 311 – 339). Hillsdale, NJ: Erlbaum.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice-versa). En G. Wright y P. Ayton (Eds.), *Subjective probability* (pp. 129 – 161). Chichester: Wiley.

Gigerenzer, G. y Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats (pp. 129-161). *Psychological Review,* 102, 684 – 704.

Gingerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J. y Kruger, L. (1989). *The empire of chance. How probability changed science and everyday life*. Cambridge: Cambridge University Press.

García Cueto, E. (1993). *Introducción a la psicometría*. Madrid: Siglo XXI.

Goetz, J. P. y Lecompte, M. D. (1988). *Etnografía y diseño cualitativo en investigación educativa*. Madrid: Morata.

Granaas, M. (2002). Hypothesis testing in psychology: throwing the baby out with the bathwater? En B. Phillips (Ed.). *Proceedings of the Sixth International Conference on Teaching of Statistics.* Ciudad del Cabo: International Association for Statistical Education. CD ROM.

Gras, R. (1993). Une méthode de classification non symétrique: L'implication statistique. *Bulletin de la Société. Française de Classification*, 1.

Gras, R. (1996). *L'implication statistique: nouvelle méthode exploratoire de données applications a la didactique*. Grenoble: La Pensée Sauvage.

Gras, R., Diday, E., Kuntz, P. y Couturier, R. (2001). Variables sur intervalles et variables intervalles en analyse statistiquee implicative. *SFC'2001* (pp 166—173). Guadeloupe, France: *Société Francophone de Classification,*

Gras, R., Kuntz P. y Briand, H. (2001). Les fondements de l'analyse statistique implicative et quelques prolongements pour la fouille de données. *Mathématiques et Sciences Humaines*, *154-155*, 9-29.

Gras, R. y Ratsima-Rajohn, H. (1996). L'implication statistique, une nouvelle méthode d'analyse de données. RAIRO *Recherche Opérationnelle 30 (3),* 217-232.

Gras, R. y Totohasina, A. (1995). Chronologie et causalité, conceptions sources d'obstacles épistémologiques à la notion de probabilité conditionnelle. *Recherches en Didactique des Mathématiques*. *15(1),* 49 – 95.

Grey, D. R. (Ed.) (1982). *Proceedings of the First International Conference on Teaching Statistics.* Sheffield: Centre for Statistical Education, University of Sheffield.

Godino, J. D. (2005). *Marcos teóricos de referencia sobre la cognición matemática*. Documento de trabajo del curso de doctorado "Teoría de la educación Matemática". On line: http://www.ugr.es/local/jgodino.

Hacking, I. (1975). *The emergence of probability.* Cambridge, MA: Cambridge University Press.

Hager, W. (2000). About some misconceptions and the discontent with statistical tests in psychology. *Methods on Psychological Research, 5(1).* On line: http://www.mpr-online.de.

Hagod, M. J. (1970). The notion of hypothetical universe. En D. E. Morrison y R. E. Henkel, (Eds.), *The significance tests controversy: A reader* (pp. 65 – 79). Chicago: Aldine.

Haller, H. y Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research, 7(1).* On line: http://www.mpr-online.de/issue16/art1/article.html.

Hamerton, M. (1973). A case of radical probability estimation. *Journal of Experimental Psychology, 101,* 252 – 254.

Harlow, L. L. (1997). Significance testing: Introduction and overview. En L. L. Harlow, S. A. Mulaik, y J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 1 – 20). Mahwah, NJ: Erlbaum.

Harlow, L. L., Mulaik, S. A. y Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Harris, R. J. (1997). Significance tests have their place. *Psychological Science. 8(1),* 8 – 11.

Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics, 6,* 187 – 205.

Hernández, R., Fernández, C. y Baptista, C. (1998). *Metodología de investigación.* México: McGraw-Hill.

Hertwig, R. y Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: how intelligent inferences look like reasoning errors. *Journal of Behavioural Decision Making, 12(4),* 275.

Huberman, A. M. y Miles, M. (1994). Data management and analysis methods. En N. K. Denzin y Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 428 – 444). London: Sage Publications.

Hunter, J. E. (1997). Needed: A ban on the significance test. *Psychological Science, 8(1),* 3 – 7.

Iglesias, P., Leiter, J., Mendoza, M., Salinas, V. y Varela, H. (2000). Mesa redonda sobre enseñanza de la estadística bayesiana. *Revista de la Sociedad Chilena de Estadística, 16-17*, 105-120.

Ito, P. K. (1999). Reaction to invited papers on statistical education and the significance tests controversy. *Proceedings of the Fifty-Second Session of the International Statistical Institute* (Tome 58, Book 3) (pp. 101-103). Helsinki, Finland: International Statistical Institute.

Iversen, G. R. (1998), Student perceptions of Bayesian statistics. En Pereira- Mendoza (Ed.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 234-240). Singapore: International Statistical Institute.

Jones, G. A. (2005) (Ed.), *Exploring probability in school: Challenges for teaching and learning.* New York: Springer.

Kahneman, D., Slovic, P., y Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Kaput, J. (1991). Notations and representations as mediators of constructive processes. En E. von Glasersfeld (Ed.), *Constructivism and mathematics education* (pp. 53– 74). Boston: Reidel.

Kelly, I. W. y Zwiers, F. W. (1986). Mutually exclusive and independence: Unravelling basic misconceptions in probability theory. *Teaching Statistics, 8,* 96 – 100.

Kish, L. (1970). Some statistical problems in research design. En D. E. Morrison y R. E. Henkel, (Eds.), *The significance tests controversy: A reader* (pp. 127 – 141). Chicago: Aldine.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behaviour and Brain Sciences*, *19,* 1-54.

Kurzenhauser, S. y Hoffrage, U. (2002). Teaching Bayesian reasoning: An evaluation of a classroom tutorial for medical students. *Medical Teacher*, *24 (5),* 516 – 521.

Labovitz, S. (1970). Criteria for selecting a significance level: A note on the sacredness of .05. En D. E. Morrison y R. E. Henkel, (Eds.), *The significance tests controversy: A reader* (pp. 166-170). Chicago: Aldine.

Lagrange J. B. (1996). Analyse implicative d'un ensemble de variables numériques, *Revue de Statistique Appliquée,* 71-93.

Lahanier-Reuter, D. (2001). Grouping together variables values: an algorithm in implicative analysis. *Mathématique et Sciences Humaines*, *154,* 47-59.

Lawrence, J. (2003). *A quick introduction to First Bayes.* Montreal: Mc Gill University: On line*:* http://www.medicine.mcgill.ca/epidemiology/Joseph/pdf/First.Bayes.pdf.

Lecoutre, B. (1996). *Traitement statistique des données experimentales.* París: CISIA.

Lecoutre, B. (1998). From signicance tests to fiducial Bayesian inference. En H. Rouanet, J. M. Bernard, M.C. Bert, B. Lecoutre, M .P. Lecoutre y B. Le Roux (Eds.), *New ways in statistical methodology. From significance tests to Bayesian inference*. (2nd edition) (123-157). Berna: Peter Lang.

Lecoutre, B. (1999). Beyond the significance test controversy: Prime time for Bayes? *Bulletin of the International Statistical Institute: Proceedings of the Fifty-second Session of the International Statistical Institute* (Tome 58, Book 2) (pp. 205 – 208). Helsinki: International Statistical Institute.

Lecoutre, B. (2006). Training students and researchers in Bayesian methods for experimental data analysis. *Journal of Data Science*, *4(2),* 207-232.

Lecoutre, B., Lecoutre M. P., y Poitevineau J. (2001). Uses, abuses and misuses of significance tests in the scientific community: Won't the Bayesian choice be unavoidable? *International Statistical Review*, *69,* 399-418.

Lee, P. M. (2004). *Bayesian statistics. An introduction.* York, UK : Arnold.

León, O. G. y Montero, I. (2002). *Métodos de investigación en psicología y educación.* Madrid: McGraw-Hill.

Leonard, T. y Hsu, J. S. (2001). *Bayesian methods*. Cambridge: Cambridge Universisty

Press.

Lerman, I. C. (1981). *Classification et analyse ordinale des données*, Dunod.

Lerman, I. C; Gras, R. y Rostam, H. (1981a). Elaboration d'un indice d'implication pour données binaires I. *Mathématiques et Sciences Humaines, 74,* 5-35.

Lerman, I. C; Gras, R. y Rostam, R. (1981b). Elaboration d'un indice d'implication pour données binaires II. *Mathématiques et Sciences Humaines, 75,* 5-47.

Lindley, D. L. (1993). The analysis of experimental data. The appreciation of tea and wine. *Teaching Statistics, 15 (1),* 22-25.

Lipset, S. M., Trow, M. A. y Coleman, J. S. (1970). Statistical problems. En D. E. Morrison y R. E. Henkel, (Eds.), *The significance tests controversy: A reader* (pp. 81 – 86). Chicago: Aldine.

Lonjedo, M. A. y Huerta, P. (2004). Una clasificación de los problemas escolares de probabilidad condicional. Su uso para la investigación y el análisis de textos. En Castro, E., y De la Torre, E. (Eds.), *Investigación en Educación Matemática. Octavo Simposio de la Sociedad Española de Investigación en Educación Matemática*, (pp 229-238). Universidade da Coruña.

Lonjedo, M, A. y Huerta, P. (2005). The nature of the quantities in a conditional probability problem. Its influence in the problem resolution. *Proceedings of CERME IV.* On line: http://cerme4.crm.es/Papers%20definitius/5/wg5litofpapers.

Losada, J. L. y López- Feal, R. (2003). *Métodos de investigación en ciencias humanas y sociales*. Madrid: Thompson.

López Feal, R. (1986). *Construcción de instrumentos de medida en ciencias conductuales y sociales.* Barcelona: Alamex.

Martínez Arias, R. (1995). *Psicometría: teoría de los tests psicológicos y educativos.* Madrid: Síntesis.

Martínez Bonafé, J. (1995). Interrogando al material curricular (Guión para el análisis y la elaboración de materiales para el desarrollo del curriculum). En J. García y M. Beas (Eds.), *Libro de Texto y construcción de materiales curriculares* (pp.221 – 240). Granada: Proyecto Sur de Ediciones.

Martignon, L. y Wassner, C. (2002). Teaching decision making and statistical thinking with natural frequencies. En B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics.* Ciudad del Cabo: International Association for Statistical Education. CD ROM.

Matthews, R. A. (1998). Facts versus factions: the use and abuse of subjectivity in scientific research. En J. Morris (Ed.), *Rethinking risk and the precautionary principle* (pp. 247-282). Oxford: Butterworth.

Maury, S. (1985). Influence de la question dans une épreuve relative á la notion d'independance. *Educational Studies in Mathematics, 16,* 283 – 301.

Maury, S. (1986). *Contribution a l'étude didactique de quelques notions de probabilité et de combinatoire à travers la résolution de problémes.* Tesis doctoral. Universidad de Montpéllier II.

McLean, A. (2001). Statistics in the catwalk. The importance of models in training researchers in statistics. En C. Batanero (Ed), *Training researchers in the use of statistics.* Granada, Spain: International Association for Statistics Education and International Statistical Institute.

Meliá, J. L. (2001). *Teoría de la fiabilidad y la validez*. Valencia: Cristóbal Serrano.

Menon, R. (1993). Statistical significance testing should be discontinued in mathematics education research. *Mathematics Education Research Journal, 5(1),* 4 – 18.

Messick, S. (1989). Validity. En R. L. Linn (Ed.), *Educational Measurement*. 3ª. Edición (pp. 13-103). Nueva York: Collier Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into scoring meaning. *American Psychologist, 9*, 741-749.

Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research, 45,* 35-44.

Millman, J. y Greene, J. (1989). The specification and development of test of achievement and ability. En R. L. Linn (Ed.), *Educational Measurement* (pp. 335 – 366). London: Macmillan.

Molinero, A. (2002). *El método bayesiano en la investigación médica.* Asociación española contra la hipertensión arterial. On line: http://www.seh-lelha.org/bayes1.htm.

Monterde-Bort, H., Pascual, J. y Frías, M. D. (2005). Incomprensión de los conceptos metodológicos y estadísticos: La encuesta "USABE". Presentado en el *IX Congreso de Metodología de las Ciencias Sociales y de la Salud.* Granada.

Monterde-Bort, H., Pascual, J. y Frías, M. D. (En prensa). Errores de interpretación de los métodos estadísticos: importancia y recomendaciones, *Psicothema.*

Moore, D. S. (1991). Teaching statistics as a respectable subject. En F. Gordon y S. Gordon (Eds.), *Statistics for the Twenty-First Century* (pp. 14-25). The Mathematical Association of America.

Moore, D. S. (1995). *Estadística aplicada básica*. Barcelona: Antoni Bosch.

Moore, D. S. (1997a). New pedagogy and new content: The case of statistics. *International Statistical Review, 65(2),* 123 – 155.

Moore, D. S. (1997b). Bayes for beginners? Some pedagogical questions. En S. Panchapakesan (Ed.), *Advances in statistical decision theory* (pp. 3-17). Boston: Birkhäuser.

Moore, D. S. (1997c). Bayes for beginners? Some reason to hesitate. *The American Statistician*, *51(3),* 254-261.

Morales, P. (1988). *Medición de actitudes en psicología y educación*. San Sebastián: Universidad de Comillas.

Moses, L. E. (1992). The reasoning of statistical inference. En D. C. Hoaglin y D. S. Moore (Eds.), *Perspectives on contemporary statistics* (pp. 107 – 122). Washington, DC: Mathematical Association of America.

Morrison, D. E., y Henkel, R. E. (Eds.) (1970). *The significance tests controversy. A reader*. Chicago: Aldine.

Muñiz, J. (1994). *Teoría clásica de los tests.* Madrid: Pirámide.

Nisbett, R., y Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgments*. Englewood Cliffs, NJ: Prentice Hall.

Nortes Checa, A. (1993). *Estadística teórica y aplicada*. Barcelona: PPU.

National Organising Committee ICOTS-4 (1994). *Proceedings of the 4th International Conference on Teaching Statistics.* Voorburg, The Netherlands: International Statistical Institute.

Navarro-Pelayo, V. (1994). *Estructura de los problemas combinatorios simples y del razonamiento combinatorio en alumnos de secundaria.* Tesis doctoral. Universidad de Granada.

Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioural sciences*. Chichester, England: Wiley.

O'Hagan, A. y Forster, J. (2004). *Bayesian inference.* Vol. 2B en Kendall's Advanced Theory of Statistics. London: Arnold.

Ojeda, A. M. (1995). Dificultades del alumnado respecto a la probabilidad condicional. *UNO, 5,* 37 – 55.

Ortega, A. R. (1991). *Contingencia y juicios de covariación en humanos.* Granada: Servicio de Publicaciones de la Universidad de Granada.

Ortega, A. I., Martos, R. y Ortega, A. R. (1992). Juicios de contingencia. *Revista de la Facultad de Humanidades de Jaén, 1 (3),* 115-140.

Osterlind, S. J. (1989). *Constructing test items.* Boston: Kluwer.

Pascual, J., García, J. F. y Frías, M. D. (2000). Significación estadística, importancia del efecto y replicabilidad de los datos. *Psicothema, 12 (2)*, 408-412.

Peña, D. (1986). *Estadística. Modelos y Métodos 1. Fundamentos.* Madrid: Alianza Editorial.

Peña, D. (2002). *Análisis de datos multivariables.* Madrid: McGraw-Hill.

Peña. D. y Romo, J. (1997). *Introducción a la estadística para las ciencias sociales.* Madrid: McGraw-Hill.

Pereira-Mendoza, L., Seu Kea, L., Wee Kee, T. y Wong, W.K. (Eds.) (1998). *Statistical Education-Expanding the Network. Proceedings of the Fifth International Conference on Teaching Statistics.* Voorburg, The Netherlands: International Statistical Institute.

Pérez Echeverría, M. P. (1990). *Psicología del razonamiento probabilístico.* Madrid: Ediciones de la Universidad Autónoma de Madrid.

Phillips, B. (Ed.) (1996). *Papers on statistical education, presented at ICME-8.* Swinburne, Australia: University of Technology.

Phillips, B. (Ed.) (2002). *Proceedings of the Sixth International Conference on Teaching of Statistics.* Ciudad del Cabo: International Association for Statistical Education. CD ROM.

Piaget, J. (1979). *Introducción a la epistemología genética*. Buenos Aires: Paidos.

Poitevineau, J. (1998). *Méthodologie de l'analyse des données expérimentales: Ëtude de la pratique des tests statistiques chez les chercheurs en psychologies: approches normative, prescriptive et descriptive*. Tesis doctoral., Universidad de Rouen.

Pollard, P., y Richardson, J. T. E. (1987). On the probability of making type I errors. *Psychological Bulletin, 10,* 159 – 163.

Pollatsek, A., Well, A. D., Konold, C. y Hardiman, P. (1987). Understanding conditional probabilities. *Organization, Behavior and Human Decision Processes, 40,* 255 – 269.

Popper, K. R. (1967). *La lógica de la investigación científica.* Madrid: Tecnos.

Pruzek. R. M. (1997). An introduction to bayesian inference and its applications. En L. L. Harlow, S. A. Mulaik y J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 287-318). Mahwah, NJ: Lawrence Erlbaum.

Rios, S. (1967). *Métodos estadísticos.* Madrid: Ediciones del Castillo.

Rindskopf, D. M. (1997). Classical and bayesian approaches. En L. L. Harlow, S. A. Mulaik y J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 319-334). Mahwah, NJ: Lawrence Erlbaum Associates.

Ritschard, G. (2005). De l'usage de la statistique implicative dans les arbres de classification, Trabajo presentado en las *Jounrnées ASI'05.* Palermo.

Rivadulla, A. (1991). *Probabilidad e inferencial científica.* Barcelona: Anthropos.

Robert, C. P. (2001). *The Bayesian choice*. New York: Springer.

Rossman, A. y Chance, B. (2006). *Proceedings of the Seventh International Conference on Teaching Statistics.* Salvador (Bahia), Brasil: International Association for Statistical Education. CD ROM.

Rossman, A. y Short, T. (1995). Conditional probability and education reform: Are they compatible? *Journal of Statistics Education, 3(2).* On line: http://www.amstat. org/publications/jse/v3n2/rossman.html.

Rouanet, H. (1998a). Statistics for researchers. En H. Rouanet et al. (Eds.), *New ways in statistical methodology* (pp. 1 – 28). Berna: Peter Lang.

Rouanet, H. (1998b). Statistical practice revisited. En H. Rouanet et al. (Eds.), *New ways in statistical methodology* (pp. 29 – 64). Berna: Peter Lang.

Rozeboon, W. W. (1970). The fallacy of the null hypothesis significance test. En. D. E. Morrison y R. E. Henkel, (Eds.), *The significance tests controversy: A reader* (pp. 216 – 230). Chicago: Aldine.

Sacerdote, A. y Balima, G. (En preparación). *Estadística bayesiana.* Buenos Aires: Universidad de Buenos Aires. On line: http://www.fi.uba.ar/materias/6109 /libro.html.

Sánchez, E. (1996). Dificultades en la comprensión del concepto de eventos independientes. En F. Hitt (Ed.), *Investigaciones en Educación Matemática* (pp. 389 – 404). México.

Sánchez, E. y Hernández, R. (2003). Variables de tarea en problemas asociados a la regla del producto en probabilidad. En E. Filloy (Coord.), *Matemática educativa, aspectos de la investigación actual* (pp. 295 –313). México: Fondo de Cultura Económica.

Santisteban, C. (1990). *Psicometría aplicada.* Madrid: Norma.

Sax, G. (1989). *Principles of educational and psychological measurement and evaluation.* Belmont, CA: Wadsworth.

Scholz, R. W. (1991). Psychological research in probabilistic understanding. En R. Kapadia y M. Borovcnik (Eds.)*, Chance encounters: Probability in education* (pp. 213 – 249). Dordrecht: Kluwer.

Schuyten, G. (1991). Statistical thinking in psychology and education. En D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (pp. 486 – 490). Voorburg, The Netherlands: International Statistical Institute.

Sedlmeier, P. (1999). *Improving statistical reasoning. Theoretical models and practical implications.* Mahwah, NJ: Erlbaum.

Selander, S. (1990). Análisis del texto pedagógico. En J. García y M. Beas (Comp.), *Libro de texto y construcción de materiales curriculares,* (pp. 131 – 161). Granada: Proyecto Sur de Ediciones.

Selvin, H. C. (1970). A critique of tests of significance in survey research. En. D. E. Morrison y R. E. Henkel, (Eds.), *The significance tests controversy: A reader* (pp. 94 – 106). Chicago: Aldine.

Serrano Angulo, J. (2003). *Iniciación a la estadística bayesiana.* Madrid: La Muralla.

Sfard, A. (2000). Symbolizing mathematical reality into being -or how mathematical discourse and mathematical objects create each other. En, P. Coob, E. Yackel y K. McClain (Eds.), *Symbolizing and communicating in mathematics classrooms* (pp.38-75). London: Lawrence Erlbaum.

Shaughnessy, J. M. (1977). Misconceptions of probability: An experiment with a small – group, activity – based, model building approach to introductory probability at the college level. *Educational Studies in Mathematics*, *8,* 295 – 316.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. En D. A. Grows (Eds.), *Handbook of research on mathematics teaching and learning* (pp. 465 – 494). New York: MacMillan.

Shaughnessy, J. M. (2006). Research on students' understanding of some big concepts in statistics. En G. Burrill (Ed.), *NCTM 2006 Yearbook: Thinking and reasoning with data and chance* (pp. 77-95). Reston, VA: NCTM.

Shaughnessy, J. M. (En prensa). Research on statistics learning and reasoning. En F. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning.* Greenwich, CT: Information Age Publishing, Inc., and NCTM.

Shaughnessy, J. M., Garfield, J., y Greer, B. (1996). Data handling. En A. Bishop et al. (Eds.), *International handbook of mathematics education* (v.1, pp. 205-237). Dordrecht, Netherlands: Kluwer.

Sohn, D. (1998). Statistical significance and replicability: Why the former does not presage the latter. *Theory & Psychology*, *8(3),* 291-311.

Skipper, J. K., Guenter, A. L. y Nass, G. (1970). The sacredness of .05: A note concerning the uses of statistical levels of significance in social sciences. En D. E. Morrison y R. E. Henkel, (Eds.), *The significance tests controversy: A reader* (pp. 155 – 160). Chicago: Aldine.

Spatz, C. (1993). *Basic statistics. Tales of distributions.* Pacific Grove, CA: Brooks /Cole.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology 15*, 201-293. Disponible en Internet: http://psychclassics.yorku.ca/Spearman.

Spiegel, M. R. (1991). *Estadística.* Mc Graw Hill (2ª edición).

Stangl D. K. (1998). Classical and Bayesian paradigms: can we teach both? En L. Pereira Mendoza (Ed.), *Proceedings of the 5th International Conference on Teaching Statistics.* Singapore: International Statistical Institute.

Starkings, S. (2000). *Stochastics papers presented at ICME-9.* Tokio: International Association for Statistics Education.

Stolarz-Fantino, S., Fantino, E., Zizzo, D.J. y Wen, J. (2003). The conjunction effect: New evidence for robustness. *The American Journal of Psychology, 116 (1),* 15 – 34.

Tabachnick, B. G. y Fidell, E. S. (2001). *Using multivariate statistics.* Needham Heights, MA: Allyn & Bacon.

Tarr, J. E. y Jones, G. A. (1997). A framework for assessing middle school students' thinking in conditional probability and independence. *Mathematics Education Research Journal, 9*, 39-59.

Tarr, J. E. y Lannin, J. K. (2005). How can teachers build notions of conditional probability and independence? En G. Jones (Ed,), *Exploring probability in school. Challenges for teaching and learning.* New York: Springer.

Tauber, L. (2001). *La construcción del significado de la distribución normal en un curso de análisis de datos.* Tesis doctoral. Universidad de Sevilla.

Taylor, S. J. y Bogdan, R. (1986). *Introducción a los métodos cualitativos de investigación.* Buenos Aires: Paidós.

Teigen, K. H., Brun, W. y Frydenlund, R. (1999). Judgments of risk and probability: the role of frequentist information. *Journal of Behavioral Decision Making*, *12(2),* 123.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher, 25(2),* 26 – 30.

Thorndike, R. L. (1989). *Psicometría aplicada.* Mexico: Limusa.

Totohasina, A. (1992). *Méthode implicative en analyse de données et application á l'analyse de concetpions d'étudiants sur la notion de probabilité conditionnelle.* Tesis Doctoral. Universidad Rennes I.

Truran, J. M. y Truran, K. M. (1997). Statistical independence: One concept or two? En B. Phillips (Ed.), P*apers from Statistical Education Presented at ICME 8* (pp. 87 – 100). Swinburne University of Technology.

Tversky, A. y Kahneman, D. (1982a). Causal schemas in judgment under uncertainty. En D. Kahneman, P. Slovic y A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 117 – 128). Cambridge, MA: Cambridge University Press.

Tversky, A. y Kahneman, D. (1982b). On the psychology of prediction. En D. Kahneman, P. Slovic y A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 69 – 83). Cambridge, MA: Cambridge University Press.

Tversky, A. y Kahneman, D. (1982c). Evidential impact of base rates. En D. Kahneman, P. Slovic y A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 153 – 160). Cambridge, MA: Cambridge University Press.

Tversky, A. y Kahneman, D. (1982d). Judgement of and by representativeness. En D. Kahneman, P. Slovic y A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 84 – 98). Cambridge, MA: Cambridge University Press.

Vacha-Haase, T. (2001). Statistical significance should not be considered one of life's guarantees: Effect sizes are needed. *Educational and Psychological Measurement*, 61, 219-224.

Valera, A. y Sánchez, J. (1997). Pruebas de significación y magnitud del efecto: reflexiones y propuestas. *Anales de Psicología*, *13*, 85-90.

Valera, S., Sánchez, J. y Marín, F. (2000). Contraste de hipótesis e investigación psicológica española: Análisis y propuestas. *Psicothema, 12(2),* 549-582.

Valera, A., Sánchez, J., Marín, F. y Velandrino, A.P. (1998). Potencia estadística en la Revista de Psicología General y Aplicada (1990-1992). *Revista de Psicología General y Aplicada. 51 (2).*

Vallecillos, A (1994). *Estudio teórico experimental de errores y concepciones sobre el contraste de hipótesis en estudiantes universitarios.* Tesis doctoral. Universidad de Granada.

Vallecillos, A. (1996). Comprensión de la lógica del contraste de hipótesis en estudiantes universitarios. *Recherches en Didactique des Mathématiques, 15(3),* 53 – 81.

Vallecillos, A. (1999). Some empirical evidence on learning difficulties about testing hypotheses. *Bulletin of the International Statistical Institute: Proceedings of the Fifty-second Session of the International Statistical Institute* (Tome 58, Book 2) (pp. 201 – 204). Helsinki, Finland: International Statistical Institute.

Vallecillos, A. y Batanero, C. (1996). Conditional probability and the level of significance in tests of hypotheses. En L. Puig y A. Gutiérrez (Eds.), *Proceedings of the Twentieth Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 271 – 378). Valencia: Universidad de Valencia.

Visauta, R. (1989). *Técnicas de investigación social I. Recogida de datos*. Barcelona: P.P.U.

Weber, R. P. (1985). *Basic content analysis.* Londres: Sage.

Western, B. (1999). Bayesian analysis for sociologists: An introduction. *Sociological Methods & Research*, *28 (1),* 7-34.

White, A. L. (1980). Avoiding errors in educational research. En R. J. Shumway (Ed.), *Research in mathematics education* (pp. 47 – 65). Reston, Va: National Council of Teachers of Mathematics.

Wild, C. y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry (with discussion). *International Statistical Review*, 67(3), 223-265.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594 – 604.

Zhu, M. y Lu, A.Y. (2004). The counter-intuitive non-informative prior for the Bernoulli family. *Journal of Statistics Education, 12 (2),* On line: http://www.amstat.org/publications/jse/v12n2/zhu.pdf.

ANEXOS

ANEXO 1.

PRIMER CUESTIONARIO PARA RECOGIDA DE DATOS DE EXPERTOS

A continuación presentamos una lista de contenidos que consideramos relevantes para evaluar la comprensión del tema de la probabilidad condicional. Junto a cada contenido, presentamos una lista de ítems de evaluación.

Requerimos su colaboración para evaluar:

- El grado en que el contenido propuesto es relevante para la comprensión de la probabilidad condicional.
- El grado en que cada ítem es adecuado para evaluar la comprensión del contenido específico propuesto.

Le agradeceríamos marque para cada uno de estos dos aspectos su opinión en la escala 1 a 5, donde 1 indica: nada relevante y 5 muy relevante.

Contenido 1: Definición de la probabilidad condicional.

Ítem 1. Explica con tus propias palabras la diferencia entre una probabilidad simple y una probabilidad condicional.

Ítem 2. ¿Qué quiere decir la expresión "la probabilidad condicional de A dado B es 1/4"?
a) En la cuarta parte de los experimentos obtenemos A y B simultáneamente
b) A ocurre la cuarta parte de las veces en que ocurre B
c) B ocurre la cuarta parte de las veces en que ocurre A
d) A o B ocurren la cuarta parte de las veces

	1: Nada	2	3	4	5: Mucho
El contenido "Definición de la probabilidad condicional" es relevante					
El ítem 1 es adecuado para este contenido					
El ítem 2 es adecuado para este contenido					

Contenido 2: Reconocer que P(B) >0 para poder definir P(B/A)

Ítem 3. ¿Cuál de las siguientes condiciones es necesaria para poder definir la probabilidad condicional P(A/B)?
a) $P(A) > 0$
b) $P(B) > 0$
c) A y B sean independientes
d) No hace falta ninguna de estas condiciones

Ítem 4. Para poder definir la probabilidad condicional $P(A/B)$ es necesario que $P(B) > 0$
Verdadero___ Falso ___

	1: Nada	2	3	4	5: Mucho
El contenido "Reconocer que la probabilidad de *P(B)>0* para poder definir *P(B/A)"* es relevante					
El ítem 3 es adecuado para este contenido					
El ítem 4 es adecuado para este contenido					

Contenido 3: Reconocer que una probabilidad condicional cumple los axiomas de probabilidad

Ítem 5. Sea *A* el suceso "Pedro instala una alarma antirrobo en su casa". Sea *B* el suceso "La casa de Pedro es robada antes de fin de año". Sean $\overline{A}$ y $\overline{B}$ las negaciones de *A* y *B* respectivamente. ¿Cuál de las siguientes afirmaciones es correcta?

a) P (A/B) + P (A/ $\overline{B}$ ***) = P (A)***

b) $P(A \cap B) >$ P (A)

c) P (A/B) $= P(A) \times P(B)$

d) P (A/B) $= P(A \cap B) / P(A)$

Ítem 6. Sea *A* el suceso "Pedro instala una alarma antirrobo en su casa". Sea *B* el suceso "La casa de Pedro es robada antes de fin de año". Sean $\overline{A}$ y $\overline{B}$ las negaciones de *A* y *B* respectivamente. ¿Cuál de las siguientes afirmaciones es incorrecta?

a) $P\,(A/B) \leq 1$

b) P (A/B) + P (A/$\overline{B}$*) = P (A)*

c) P (A/A) = 1

d) P (A/B) = $P(A \cap B) / P(A)$

	1: Nada	2	3	4	5: Mucho
El contenido "Reconocer que la probabilidad condicional cumple los axiomas" es relevante					
El ítem 5 es adecuado para este contenido					
El ítem 6 es adecuado para este contenido					

Contenido 4: Reconocer que la probabilidad condicional supone una restricción del espacio muestral

Ítem 7. ¿Cuál es el espacio muestral de una probabilidad condicionada P(A/B)?
a) El mismo que el de la probabilidad simple
b) El suceso condicionante
c) El suceso condicionado
d) La intersección de los dos sucesos

Ítem 8. El espacio muestral de una probabilidad condicionada *P(A/B)* es el suceso condicionante *B*

Verdadero___ Falso _____

	1: Nada	2	3	4	5: Mucho
El contenido "Reconocer que la probabilidad condicional supone una restricción del espacio muestral" es relevante					
El ítem 7 es adecuado para este contenido					
El ítem 8 es adecuado para este contenido					

Contenido 5: Distinguir una probabilidad condicional y su inversa: Diferenciar *P(A/B)* y *P(B/A)*

Ítem 9. En un centro médico se han observado a 250 personas para estudiar si el hábito de fumar tiene relación con los trastornos bronquiales, obteniéndose los siguientes resultados:

	Fuma	**No Fuma**	**Total**
Padece trastornos	90	60	150
No padece trastornos	60	40	100
Total	150	100	250

Si escogemos al azar una de estas personas:

I. ¿Cuál es la probabilidad de que fume si sabemos que tiene trastornos? ___________

II. ¿Cuál es la probabilidad de que tenga trastornos, si sabemos que fuma? ___________

Ítem 10. En un barrio de Granada se ha realizado una entrevista a un grupo de hombres, obteniendo los siguientes resultados:

	menos de 55 años	**más de 55 años**	**Total**
Ha sufrido un ataque al corazón	29	75	104
No ha sufrido ataque	401	275	676
Total	430	350	780

Si elegimos al azar una de estas personas:

a) Dado que la persona escogida tiene más de 55 años, ¿cuál es la probabilidad de que haya tenido un ataque al corazón?

b) Dado que la persona escogida ha tenido un ataque al corazón, ¿cuál es la probabilidad de que tenga más de 55 años?

Ítem 11. ¿En cuál de las siguientes predicciones tendrás mayor seguridad o confianza?

a) Una persona adicta a la heroína también fuma marihuana

b) Una persona que fuma marihuana es también adicta a la heroína

c) Los dos sucesos son iguales de probables

	1: Nada	2	3	4	5: Mucho
El contenido “Distinguir una probabilidad condicional de su inversa” es relevante					
El ítem 9 es adecuado para este contenido					
El ítem 10 es adecuado para este contenido					
El ítem 11 es adecuado para este contenido					

Contenido 6: Distinguir probabilidad conjunta, condicional y simple

Ítem 12. En un centro médico se han observado a 250 personas para estudiar si el hábito de fumar tiene relación con los trastornos bronquiales, obteniéndose los siguientes resultados:

	Fuma	**No Fuma**	**Total**
Padece trastornos	90	60	150
No padece trastornos	60	40	100
Total	150	100	250

Si escogemos al azar una de estas personas:

I. ¿Cuál es la probabilidad de tener trastornos bronquiales? ___________

II. ¿Cuál es la probabilidad de que tenga trastornos y fume? ___________

III. ¿Cuál es la probabilidad de que fume si sabemos que tiene trastornos? ___________

Ítem 13. En un barrio de Granada se ha realizado una entrevista a un grupo de hombres, obteniendo los siguientes resultados:

	menos de 55 años	**más de 55 años**	**Total**
Ha sufrido un ataque al corazón	29	75	104
No ha sufrido ataque	401	275	676
Total	430	350	780

Si elegimos al azar una de estas personas:

a) ¿Cuál es la probabilidad de que tenga más de 55 años y además haya tenido uno o más ataques al corazón?

b) ¿Cuál es la probabilidad de que haya tenido un ataque al corazón?

c) Dado que la persona escogida tiene más de 55 años, ¿cuál es la probabilidad de que haya tenido un ataque al corazón?

	1: Nada	2	3	4	5: Mucho
El contenido "Distinguir probabilidad condicional, conjunta y simple" es relevante					
El ítem 12 es adecuado para este contenido					
El ítem 13 es adecuado para este contenido					

Contenido 7: Reconocer que la probabilidad conjunta *P(A y B)* ha de ser menor o igual que la probabilidad simple *P(A)*

Ítem 14. Supón que Carlos Ferrero alcanza la final de Roland Garros en 2004. ¿Cuál de los siguientes sucesos consideras más probable?

a) Carlos Ferrero pierde el primer set

b) Carlos Ferrero pierde el primer set pero gana el partido

c) Los dos sucesos son iguales de probables

Ítem 15. ¿Cuál de los siguientes sucesos consideras más probable?

a) Aznar mandará más tropas a Irak y aumentará el presupuesto de becas

b) Aznar aumentará el presupuesto de becas.

c) Los dos sucesos son igual de probables

	1: Nada	2	3	4	5: Mucho
El contenido "Reconocer que la probabilidad conjunta *P(A y B)* ha de ser menor o igual que la probabilidad simple *P(A)*" es relevante					
El ítem 14 es adecuado para este contenido					
El ítem 15 es adecuado para este contenido					

Contenido 8: Distinguir sucesos independientes, dependientes y mutuamente excluyentes

Ítem 16. Se extrae una carta al azar de una baraja americana: sea A el suceso "se extrae un trébol" y B el suceso "se extrae una reina" ¿Los sucesos A y B son independientes?

a) Sí, en todos los casos.

b) No son independientes porque en la baraja hay una reina de tréboles.

c) Sólo si sacamos primero una carta para ver si es reina y se vuelve a colocar en la baraja y luego sacamos una segunda carta y para mirar si es trébol.

d) No, porque P(reina de trébol)= P(reina) x P(trébol)

Ítem 17. ¿Qué quiere decir el término “mutuamente excluyentes” cuando se usa en referencia a la probabilidad?

a) La ocurrencia de un suceso no afecta la ocurrencia del otro suceso
b) La probabilidad de la intersección es el producto de las probabilidades simples
c) Si un suceso ocurre, el otro no puede ocurrir
d) A y B son correctas

Ítem 18. En un grupo de mil sujetos se pasó un test de inteligencia y se midió su rendimiento, clasificando a cada sujeto según fuese superior o inferior a la media en inteligencia, y según fuese apto o no apto en la prueba de rendimiento académico. Se encontraron 500 aptos, de los que 300 tenían inteligencia superior. De los 400 con inteligencia superior, 300 resultaron aptos. Extraemos un sujeto al azar y definimos los sucesos, A: “Ser superior a la media en inteligencia”, y B: “Ser apto en rendimiento”. ¿Son los sucesos A y B independientes?

	1: Nada	2	3	4	5: Mucho
El contenido “Distinguir sucesos independientes, dependientes y mutuamente excluyentes” es relevante					
El ítem 16 es adecuado para este contenido					
El ítem 17 es adecuado para este contenido					
El ítem 18 es adecuado para este contenido					

Contenido 9: Calcular una probabilidad condicional dentro de un único experimento

Ítem 19. Cogemos un número al azar menor que 1000. Sabiendo que es múltiplo de cinco. ¿Cuál es la probabilidad de que sea múltiplo de cuatro?

a) 1/20 **b) 1/4** c) 1/5 d) 4/5

Ítem 20. Hemos lanzado dos dados y sabemos que el producto de los dos números obtenidos ha sido 12 ¿Cual es la probabilidad de que ninguno de los dos números sea un 6?

	1: Nada	2	3	4	5: Mucho
El contenido “Calcular una probabilidad condicional dentro de un único experimento” es relevante					
El ítem 19 es adecuado para este contenido					
El ítem 20 es adecuado para este contenido					

Contenido 10: Calcular la probabilidad condicional en un contexto de muestreo con reposición

Ítem 21. Una persona lanza un dado y anota si saca un número par o impar. Estos son los resultados:
par, impar, impar, par, par, impar, par, par, par, par, impar, impar, par, par, par
Lanza el dado de nuevo
¿Cuál es la probabilidad de sacar un número par en la siguiente tirada?

Ítem 22. Una persona lanza 3 veces la misma moneda, obteniendo en este orden los resultados siguientes: cara, cruz, cruz. Lanza la moneda una cuarta vez. ¿Qué resultado piensas que es más probable que salga a continuación?

a) Cara b) Cruz c) **Los dos sucesos son iguales de probables**.

	1: Nada	2	3	4	5: Mucho
El contenido "Calcular una probabilidad condicional en un contexto de muestreo con reposición" es relevante					
El ítem 21 es adecuado para este contenido					
El ítem 22 es adecuado para este contenido					

Contenido 11: Calcular la probabilidad condicional en un contexto de muestreo sin reposición

Ítem 23. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin reemplazamiento.

- ¿Cuál es la probabilidad de extraer una bola negra en segundo lugar, habiendo extraído una bola negra en primer lugar? $P\ (N_2/\ N_1)$

  a) 1/ 3 b) 1/ 2 c) 1/ 6 d) 1/ 4

Ítem 24. Una caja tiene cuatro focos, de los cuales dos son defectuosos. Se sacan dos al azar, uno tras otro sin reemplazamiento.

- Si el primer foco fue defectuoso, ¿qué es más probable?
  a) que el segundo sea defectuoso
  b) que el segundo no sea defectuoso
  c) los dos sucesos son iguales de probables

	1: Nada	2	3	4	5: Mucho
El contenido "Calcular la probabilidad condicional en un contexto de muestreo con reposición" es relevante					
El ítem 23 es adecuado para este contenido					
El ítem 24 es adecuado para este contenido					

Contenido 12: Calcular una probabilidad condicional a partir de probabilidades conjuntas y simples: Aplicando directamente la definición $P(A/B) = P(A \cap B)/P(B)$

Ítem 25. La probabilidad de que una mujer de 40 años tenga una mamografía positiva es el 10.3 %. La probabilidad de que una mujer de 40 años tenga cáncer de pecho y una mamografía positiva es de 0.8%. Una mujer de 40 años se realiza una mamografía y resulta positiva. ¿Cuál es la probabilidad de que realmente tenga cáncer?

a) $\frac{0,8}{10,3} = 0,0776$**, probabilidad de 7,76%**

b) $10,3 \times 0,8 = 8,24$, probabilidad del 8'24%

c) 0,8 %

Ítem 26. 103 de cada 1000 mujeres de 40 años tienen una mamografía positiva. 8 de cada 1000 mujeres de 40 años tienen cáncer de pecho y obtienen una mamografía positiva. En una muestra representativa de mujeres de 40 años que se realizan una mamografía y resulta positiva. ¿Cuántas de ellas esperas que realmente tengan cáncer?

a) 7,7 de cada 100
b) 8,24 de cada 100
c) 0,8 de cada 100

Ítem 27. En un país el 30 % de las personas tiene los ojos claros y el 15 % tiene los ojos claros y el pelo rubio. Si en este país cogemos una persona rubia al azar, ¿Cuál es la probabilidad de que tenga los ojos oscuros?

	1: Nada	2	3	4	5: Mucho
El contenido "Calcular una probabilidad condicional a partir de probabilidades conjuntas y simples" es relevante					
El ítem 25 es adecuado para este contenido					
El ítem 26 es adecuado para este contenido					
El ítem 27 es adecuado para este contenido					

Contenido 13: Resolver correctamente problemas condicionales cuando se invierte el eje de tiempo

Ítem 28. ¿Cuál de los siguientes sucesos es más probable?

a) Que una niña tenga los ojos azules si su madre tiene los ojos azules
b) Que una madre tenga los ojos azules si su hija tiene los ojos azules
c) Los dos sucesos son iguales de probables
d) La probabilidad de que la madre tenga los ojos azules dado que su hija tiene los ojos azules no puede calcularse ya que es anterior en el tiempo.

Ítem 29. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin reemplazamiento. ¿Cuál es la probabilidad de extraer una bola negra en primer lugar, sabiendo que hemos extraído una bola negra en segundo lugar?
a) 1/ 3 b) No se puede calcular c) 1/ 6 d) 1/ 2

	1: Nada	2	3	4	5: Mucho
El contenido "Reconocer que un suceso puede condicionarse por otro posterior en el tiempo" es relevante					
El ítem 28 es adecuado para este contenido					
El ítem 29 es adecuado para este contenido					

Contenido 14: Distinguir situación condicional, causal y diagnóstica

Ítem 30. ¿Cuál de los siguientes sucesos es más probable?

a) Un divorcio es resultado de un matrimonio infeliz
b) Un matrimonio infeliz lleva a un divorcio
c) Los dos enunciados son iguales de probables.

Ítem 31. ¿En cuál de las siguientes predicciones tendrás mayor seguridad?

a) Predecir que una persona que está enferma, tiene fiebre.
b) **Predecir que una persona que tiene fiebre está enferma.**
c) Tendría igual seguridad en ambas predicciones.

Ítem 32. Un test diagnóstico de cáncer fue administrado a todos los residentes en una gran ciudad. Un resultado positivo en el test es indicativo de cáncer y un resultado negativo es indicativo de ausencia de cáncer. ¿En cuál de las siguientes predicciones tienes más confianza?

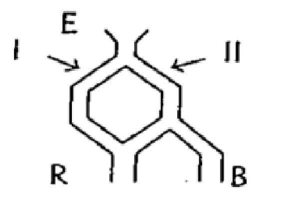
a) Predecir que una persona tiene cáncer si ha dado positivo en el test de diagnóstico.
b) Predecir un resultado positivo en el test de diagnóstico si la persona tiene cáncer.
c) Tengo la misma confianza en ambas predicciones.

	1: Nada	2	3	4	5: Mucho
El contenido "Resolver correctamente problemas condicionales en situaciones diagnósticas" es relevante					
El ítem 30 es adecuado para este contenido					
El ítem 31 es adecuado para este contenido					
El ítem 32 es adecuado para este contenido					

Contenido 15: Resolver correctamente problemas en situaciones diacrónicas (experimentos simples consecutivos)

Ítem 33. Una bola se suelta en E. Si sale por R, ¿Cuál es la probabilidad de que haya pasado por el canal 1?

a) 0,50
b) 0,66
c) 0,33
d) No se puede calcular



Ítem 34. Una caja tiene cuatro focos, de los cuales dos son defectuosos. Se sacan dos al azar, uno tras otro sin reemplazamiento.

- Si el segundo foco es defectuoso, ¿qué es lo más probable?

a) que el primero sea defectuoso
b) que el primero no sea defectuoso
c) los dos sucesos son iguales de probables
d) No se puede saber

	1: Nada	2	3	4	5: Mucho
El contenido "Resolver correctamente problemas en situaciones diacrónicas es relevante					
El ítem 33 es adecuado para este contenido					
El ítem 34 es adecuado para este contenido					

Contenido 16: Resolver correctamente problemas en situaciones sincrónicas (experimentos simples simultáneos)

Ítem 35. En una urna hay tres tarjetas: una es azul por ambos lados, otra es roja por ambos lados y la última es por un lado azul y por otro roja. Seleccionamos una tarjeta al azar y la ponemos sobre la mesa tal y como la sacamos de la urna. La cara que vemos es roja, ¿cuál es la probabilidad de que la otra cara sea roja también?

a) **2/ 3** **b)** 1/ 2 c) 3/4 d) 1/3 x 1/2

Ítem 36. En un grupo de mil sujetos se pasó un test de inteligencia y se midió su rendimiento, clasificando a cada sujeto según fuese superior o inferior a la media en inteligencia, y según fuese o no reflexivo. Se encontraron 500 sujetos reflexivos, de los que 300 tenían inteligencia superior. ¿Cuál es la probabilidad de que un sujeto reflexivo sea de inteligencia superior?

Ítem 37. En una encuesta publicada en un periódico se informa que el 91% de los habitantes de una ciudad mienten usualmente y de ellos el 36% mienten sobre cosas importantes. Si nos encontramos con una persona al azar de esta ciudad, ¿cuál es la probabilidad de que mienta sobre cosas importantes?

	1: Nada	2	3	4	5: Mucho
El contenido “Resolver correctamente problemas en situaciones sincrónicas” es relevante					
El ítem 35 es adecuado para este contenido					
El item 36 es adecuado para este contenido					
El ítem 37 es adecuado para este contenido					

Contenido 17: Calcular una probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos independientes

Ítem 38. Un grupo de alumnos de un colegio se examina de Matemáticas e Inglés. La proporción de alumnos que aprueban Matemáticas es del 80% y la proporción de alumnos que aprueba Inglés es del 70%. Suponiendo que las notas en cada una de las asignaturas es independiente, ¿Cuál es la probabilidad de que un alumno escogido al azar haya aprobado ambas asignaturas?

Ítem 39 Se lanza cuatro veces un dado. ¿Cuál es la probabilidad de no obtener ningún uno?

Ítem 40. En una secuencia de tres dígitos, cada uno de los dígitos es seleccionado al azar del conjunto {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}. ¿Cuál es la probabilidad de que de este proceso resulten tres números diferentes?

	1: Nada	2	3	4	5: Mucho
El contenido “Calcular una probabilidad compuesta haciendo uso de la regla del producto en sucesos independientes ” es relevante					
El ítem 38 es adecuado para este contenido					
El ítem 39 es adecuado para este contenido					
El ítem 40 es adecuado para este contenido					

Contenido 18: Calcular una probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos dependientes

Ítem 41. Se tiene una bolsa con 1 bola azul y 2 bolas rojas. Se saca una bola al azar. Sin reponerla otra vez en la bolsa se saca una segunda bola. ¿Cuál suceso es más probable?

a) Sacar dos bolas rojas
b) Sacar primero una bola roja y luego una azul.
c) Los dos sucesos son iguales de probables
d) Sacar primero una bola azul y luego una roja

Ítem 42. Un hombre, por accidente, mezcla dos baterías gastadas con tres baterías nuevas del mismo modelo. Con ayuda de un voltímetro va probando una batería tras otra hasta encontrar las dos baterías gastadas. ¿Cuál es la probabilidad de que tenga que probar sólo dos veces para encontrar las baterías gastadas?

Ítem 43. De cada 1000 niños varones nacidos vivos 746 llegan a cumplir 65 años. La probabilidad de que una persona que acaba de cumplir 65 muera en los cinco años siguientes es 0,160. ¿Cuál es la probabilidad de que un niño varón alcance su 70 cumpleaños?

	1: Nada	2	3	4	5: Mucho
El contenido “Calcular una probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos dependientes” es relevante					
El ítem 41 es adecuado para este contenido					
El ítem 42 es adecuado para este contenido					
El ítem 43 es adecuado para este contenido					

Contenido 19: Aplicar correctamente la regla de la probabilidad total en un problema

Ítem 44. En una escuela hay en marcha dos programas educativos, A y B. El 60% de los estudiantes están en el programa A y el 40% en el programa B. En el programa A, el 50% son chicos y en el programa B, el 25% son chicos. Encontramos al azar un estudiante ¿cuál es la probabilidad de que sea chico?

Ítem 45. En una ciudad hay 60 hombres y 40 mujeres por cada 100 habitantes. 50 de cada 100 hombres y 25 de cada 100 mujeres fuman. Si cogemos 200 personas al azar, ¿cuántas de ellas fumarán?

Ítem 46. Al azar se seleccionan dos de cuatro cartas numeradas 1, 2, 3 y 4, una después de otra. Los números de las cartas se suman y si la suma es par, se gana un premio. Qué es más probable, ¿obtener el premio o no?

	1: Nada	2	3	4	5: Mucho
El contenido “Aplicar correctamente la regla de la probabilidad total en un problema” es relevante					
El ítem 44 es adecuado para este contenido					
El ítem 45 es adecuado para este contenido					
El ítem 46 es adecuado para este contenido					

Contenido 20: Aplicar correctamente el teorema de Bayes en un problema

Ítem 47. Un taxi se vio implicado en un accidente nocturno con choque y huida posterior. Hay dos compañías de taxis en la ciudad, la Verde y la Azul. El 85% de los taxis de la ciudad son Verde y el 15% Azul. Un testigo identificó el taxi como Azul. El tribunal comprobó la fiabilidad del testigo bajo las mismas circunstancias que había la noche del accidente y llegó a la conclusión de que el testigo identificaba correctamente cada uno de los colores en el 80% de las ocasiones y fallaba en el 20%. ¿Cuál es la probabilidad de que el taxi implicado en el accidente fuera en efecto Azul?

a) 80 % **b) 48 %** c) 15% d) (15/100) x (80/100)

Ítem 48. Una fábrica dispone de dos máquinas M1 y M2 que fabrican bolas. La máquina M1 fabrica el 40 % de las bolas y la M2 el 60%. El 5% de las bolas fabricadas por M1 y el 1% de las fabricadas por M2 son defectuosas. Tomamos una bola al azar que resulta ser defectuosa. ¿Cuál es la probabilidad de que haya sido fabricada por M1?

Ítem 49. Se sabe que 100 de cada 10000 personas está enferma. De cada 100 personas enfermas, el test será positivo en 80 de ellas. De cada 9900 personas sanas el test saldrá positivo en 950 de ellas. El test resulta positivo en 103 personas. ¿Cuantas de estas personas están realmente enfermas?

	1: Nada	2	3	4	5: Mucho
El contenido “Recordar el teorema de Bayes” es relevante					
El contenido “Aplicar correctamente la regla de Bayes en un problema” es relevante					
El ítem 47 es adecuado para este contenido					
El ítem 48 es adecuado para este contenido					
El ítem 49 es adecuado para este contenido					

Como resumen le presentamos la lista de contenidos previstos en el cuestionario. Si piensa que hemos omitido algún contenido importante, le agradeceríamos lo añada en la tabla

	Principales áreas del contenido
Conoci. conceptual	1. Definición de la probabilidad condicional
	2. Reconocer que la probabilidad de *P(B) >0* para poder definir *P(B/A)*
	3. Reconocer que una probabilidad condicional cumple los axiomas.
	4. Reconocer que la probabilidad condicional supone una restricción del espacio muestral
	5. Distinguir probabilidad condicional con inversa
	6. Distinguir probabilidad conjunta, condicional y simple
	7. Probabilidad conjunta menor que probabilidad simple
	8. Distinguir situación condicional, causal y diagnóstica
	9. Distinguir sucesos independientes, dependientes y mutuamente excluyentes
Conocim. procedimental	10. Resolver correctamente problemas de probabilidad condicional dentro de un único experimento
	11. Resolver correctamente problemas de probabilidad condicional en un contexto de muestreo con reposición
	12. Resolver correctamente problemas de probabilidad condicional en un contexto de muestreo sin reposición
	13. Resolver correctamente problemas de probabilidad condicional a partir de probabilidades conjuntas y simples
	14. Resolver correctamente problemas condicionales cuando se invierte el eje de tiempo
	15. Resolver correctamente problemas en situaciones sincrónicas
	16. Resolver correctamente problemas en situaciones diacrónicas
	17. Resolver correctamente problemas de probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos independientes
	18. Resolver correctamente problemas de probabilidad compuesta haciendo uso de la regla del producto en caso de sucesos dependientes
	19. Aplicar correctamente el cálculo de la probabilidad condicional en situaciones de sucesos múltiples (regla de la probabilidad total)
	20. Aplicar correctamente el cálculo de la probabilidad condicional en situaciones de probabilidad inversa (regla de Bayes)

ANEXO 2.

RESULTADOS DE LAS PRUEBAS INICIALES DE ÍTEMS

Ítem 1. Explica con tus propias palabras la diferencia entre una probabilidad simple y una probabilidad condicional.

Este ítem se ha puntuado siguiendo el siguiente criterio:
0: No se responde o se responde incorrectamente
1: Define correcta, pero imprecisamente una de las probabilidades pedidas
2: Define correcta, y precisamente una de las probabilidades pedidas
3. Define correcta, pero imprecisamente las dos probabilidades pedidas
4. Define correcta y precisamente las dos probabilidades pedidas

Tabla A2.1. Resultados en el ítem 1

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
1	2	4,1	4,7	4,7
2	4	8,2	9,3	14,0
3	20	40,8	46,5	60,5
4	17	34,7	39,5	100,0
Total	43	87,8	100,0	
Blanco	6	12,2		
Total	49	100,0		

Ítem 2. ¿Qué quiere decir la expresión "la probabilidad condicional de A dado B es 1/4"?

Tabla A2.2. Resultados en el ítem 2

	Frecuencia	Porcentaje	Porcentaje válido
a) En la cuarta parte de los experimentos obtenemos A y B simultáneamente	6	11,5	11,8
b) **A ocurre la cuarta parte de las veces en que ocurre B**	37	71,2	72,5
c) B ocurre la cuarta parte de las veces en que ocurre A	8	15,4	15,7
d) A o B ocurren la cuarta parte de las veces	0	0	0
Total	51	98,1	100,0
Blanco	1	1,9	
Total	52	100,0	

Ítem 3. ¿Cuál de las siguientes condiciones es necesaria para poder definir la probabilidad condicional *P(A/B)*?

Tabla A2.3. Resultados en el ítem 3

	Frecuencia	Porcentaje	Porcentaje válido
a) $P(A) > 0$	3	6,1	6,7
b) $\boldsymbol{P(B) > 0}$	12	24,5	26,7
c) *A* y *B* sean independientes	15	30,6	33,3
d) No hace falta ninguna de estas condiciones	15	30,6	33,3
Total	45	91,8	100,0
Blanco	4	8,2	
Total	49	100,0	

Ítem 4. Para poder definir la probabilidad condicional *P(A/B)* es necesario que *P(B) > 0*
Verdadero___ Falso ___

Tabla A2.4. Resultados en el ítem 4

	Frecuencia	Porcentaje	Porcentaje válido
Verdadero	38	73,1	73,1
Falso	14	26,9	26,9
Total	52	100,0	100,0

Ítem 5. Sea *A* el suceso "Pedro instala una alarma antirrobo en su casa". Sea *B* el suceso "La casa de Pedro es robada antes de fin de año". Sean $\overline{A}$ y $\overline{B}$ las negaciones de *A* y *B* respectivamente. ¿Cuál de las siguientes afirmaciones es correcta?

Tabla A2.5. Resultados en el ítem 5

	Frecuencia	Porcentaje	Porcentaje válido
a. $P(A \cap B) > P(A)$	16	10,2	16,2
b. $\boldsymbol{P(A/B) + P(A/\overline{B}) = P(A)}$	**39**	**24,8**	**39,4**
c. $P(A/B) = P(A) \times P(B)$	44	28,0	44,4
Total	99	63,1	100,0
En blanco	58	36,9	
Total	157	100,0	

Ítem 6. Sean $\overline{A}$ y $\overline{B}$ las negaciones de *A* y *B* respectivamente. ¿Cuál de las siguientes afirmaciones es incorrecta?

Tabla A2.6. Resultados en el ítem 6

	Frecuencia	Porcentaje	Porcentaje válido
a) $P(A/B) \leq 1$	6	12,2	14,3
b) $P(A/B) + P(A/\overline{B}) = P(A)$	7	14,3	16,7
c) $P(A/A) = 1$	11	22,4	26,2
d) $\boldsymbol{P(A/B) = P(A \cap B) / P(A)}$	18	36,7	42,9
Total	42	85,7	100,0
Blanco	7	14,3	
Total	49	100,0	

Ítem 7. ¿Cuál es el espacio muestral de una probabilidad condicionada *P(A/B)*?

Tabla A2.7. Resultados en el ítem 7

	Frecuencia	Porcentaje	Porcentaje válido
a) El mismo que el de la probabilidad simple	4	7,7	8,0
b) **El suceso condicionante**	8	15,4	16,0
c) El suceso condicionado	14	26,9	28,0
d) La intersección de los dos sucesos	24	46,2	48,0
Total	50	96,2	100,0
Blanco	2	3,8	
Total	52	100,0	

Ítem 8. El espacio muestral de una probabilidad condicionada *P(A/B)* es el suceso condicionante *B*

Tabla A2.8. Resultados en el ítem 8

	Frecuencia	Porcentaje	Porcentaje válido
Verdadero	18	36,7	38,3
Falso	29	59,2	61,7
Total	47	95,9	100,0
Blanco	2	4,1	
Total	49	100,0	

Ítem 9. En un centro médico se han observado a 250 personas para estudiar si el hábito de fumar tiene relación con los trastornos bronquiales, obteniéndose los siguientes resultados:

	Fuma	**No Fuma**	**Total**
Padece trastornos	90	60	150
No padece trastornos	60	40	100
Total	150	100	250

a) ¿Cuál es la probabilidad de que fume si sabemos que tiene trastornos?

Tabla A2.9. Resultados en el ítem 9a

	Frecuencia	Porcentaje	Porcentaje válido
Incorrecto	12	23,1	26,1
Correcto	**34**	**65,4**	**73,9**
Total	46	88,5	100,0
Blanco	6	11,5	
Total	52	100,0	

b) ¿Cuál es la probabilidad de que tenga trastornos si sabemos que fuma?

Tabla A2.10. Resultados en el ítem 9b

	Frecuencia	Porcentaje	Porcentaje válido
Incorrecto	13	25,0	35,1
Correcto	**24**	**46,2**	**64,9**
Total	37	71,2	100,0
Blanco	15	28,8	
Total	52	100,0	

Ítem 10. En un barrio de Granada se ha realizado una entrevista a un grupo de hombres, obteniendo los siguientes resultados:

	menos de 55 años	**más de 55 años**	**Total**
Ha sufrido un ataque al corazón	29	75	104
No ha sufrido ataque	401	275	676
Total	430	350	780

Si elegimos al azar una de estas personas:

a) Dado que la persona escogida tiene más de 55 años, ¿cuál es la probabilidad de que haya tenido un ataque al corazón?

Tabla A2.11. Resultados en el ítem 10a

	Frecuencia	Porcentaje	Porcentaje válido
Incorrecto	12	24,5	34,3
Correcto	23	46,9	65,7
Total	35	71,4	100,0
Blanco	14	28,6	
Total	49	100,0	

b) Dado que la persona escogida ha tenido un ataque al corazón, ¿cuál es la probabilidad de que tenga más de 55 años?

Tabla A2.12. Resultados en el ítem 10b

	Frecuencia	Porcentaje	Porcentaje válido
Incorrecto	12	24,5	38,7
Correcto	19	38,8	61,3
Total	31	63,3	100,0
Blanco	18	36,7	
Total	49	100,0	

Ítem 11. ¿En cuál de las siguientes predicciones tendrás mayor seguridad o confianza?

Tabla A2.13. Resultados en el ítem 11

	Frecuencia	Porcentaje	Porcentaje válido
a. Una persona adicta a la heroína también fuma marihuana	**74**	**47,1**	**47,4**
b. Una persona que fuma marihuana es también adicta a la heroína	3	1,9	1,9
c. Los dos sucesos son igual de probables.	79	50,3	50,6
Total	156	99,4	100,0
En blanco	1	,6	
Total	157	100,0	

Item 12. En un centro médico se han observado a 250 personas para estudiar si el hábito de fumar tiene relación con los trastornos bronquiales, obteníendose los siguientes resultados:

	Fuma	**No Fuma**	**Total**
Padece trastornos	90	60	150
No padece trastornos	60	40	100
Total	150	100	250

a) ¿Cuál es la probabilidad de tener trastornos bronquiales?

Tabla A2.14. Resultados en el ítem 12a

	Frecuencia	Porcentaje	Porcentaje válido
Incorrecto	11	21,2	21,6
Correcto	40	76,9	78,4
Total	51	98,1	100,0
Blanco	1	1,9	
Total	52	100,0	

b) ¿Cuál es la probabilidad de que tenga trastornos y fume?

Tabla A2.15. Resultados en el ítem 12b

	Frecuencia	Porcentaje	Porcentaje válido
Incorrecto	28	53,8	54,9
Correcto	23	44,2	45,1
Total	51	98,1	100,0
Blanco	1	1,9	
Total	52	100,0	

c) ¿Cuál es la probabilidad de que fume si sabemos que tiene trastornos?

Tabla A2.16. Resultados en el ítem 12c

	Frecuencia	Porcentaje	Porcentaje válido
Incorrecto	12	23,1	26,1
Correcto	34	65,4	73,9
Total	46	88,5	100,0
Blanco	6	11,5	
Total	52	100,0	

Ítem 13. En un barrio de Granada se ha realizado una entrevista a un grupo de hombres, obteniendo los siguientes resultados:

	menos de 55 años	**más de 55 años**	**Total**
Ha sufrido un ataque al corazón	29	75	104
No ha sufrido ataque	401	275	676
Total	430	350	780

Si elegimos al azar una de estas personas:

a) ¿Cuál es la probabilidad de que tenga más de 55 años y además haya tenido uno o más ataques al corazón?

Tabla A2.17. Resultados en el ítem 13a

	Frecuencia	Porcentaje	Porcentaje válido
Incorrecto	26	53,1	68,4
Correcto	12	24,5	31,6
Total	38	77,6	100,0
Blanco	11	22,4	
Total	49	100,0	

b) ¿Cuál es la probabilidad de que haya tenido un ataque al corazón?

Tabla A2.18. Resultados en el ítem 13b

	Frecuencia	Porcentaje	Porcentaje válido
Incorrecto	5	10,2	12,5
Correcto	35	71,4	87,5
Total	40	81,6	100,0
Blanco	9	18,4	
Total	49	100,0	

c) Dado que la persona escogida tiene más de 55 años, ¿cuál es la probabilidad de que haya tenido un ataque al corazón?

Tabla A2.19. Resultados en el ítem 13c

	Frecuencia	Porcentaje	Porcentaje válido
Incorrecto	12	24,5	34,3
Correcto	23	46,9	65,7
Total	35	71,4	100,0
Blanco	14	28,6	
Total	49	100,0	

Ítem 14. Supón que Anna Kournikova alcanza la final de Roland Garros en 2004. ¿Cuál de las siguientes afirmaciones consideras más probable?

Tabla A2.20. Resultados en el ítem 14

	Frecuencia	Porcentaje	Porcentaje válido
a. Anna Kournikova gana el primer set pero pierde el partido	6	7,4	8,0
b. Anna Kournikova gana el primer set	**30**	**37,0**	**40,0**
c. Los dos sucesos son igual de probables	39	48,1	52,0
Total	75	92,6	100,0
En blanco	6	7,4	
Total	81	100,0	

Ítem 15. ¿Cuál de los siguientes sucesos consideras más probable?

Tabla A2.21. Resultados en el ítem 15

	Frecuencia	Porcentaje	Porcentaje válido
a. Aznar mandará más tropas a Irak y aumentará el presupuesto de becas	14	18,4	20,3
b. Aznar aumentará el presupuesto de becas	**21**	**27,6**	**30,4**
c. Los dos sucesos son igual de probables	34	44,7	49,3
Total	69	90,8	100,0
En blanco	7	9,2	
Total	76	100,0	

Ítem 16. Se extrae una carta al azar de una baraja americana: sea A el suceso "se extrae un trébol" y B el suceso "se extrae una reina" ¿Los sucesos A y B son independientes?

Tabla A2.22. Resultados en el ítem 16

	Frecuencia	Porcentaje	Porcentaje válido
a. Sí, en todos los casos.	**26**	**16,6**	**18,6**
b. No son independientes porque en la baraja hay una reina de tréboles.	53	33,8	37,9
c. Solo si sacamos primero una carta para ver si es reina y se vuelve a colocar en la baraja y luego sacamos una segunda carta y para mirar si es trébol.	61	38,9	43,6
Total	140	89,2	100,0
En blanco	17	10,8	
Total	157	100,0	

Ítem 17. ¿Qué quiere decir el término "mutuamente excluyentes" cuando se usa en referencia a la probabilidad?

Tabla A2.23. Resultados en el ítem 17

	Frecuencia	Porcentaje	Porcentaje válido
a) La ocurrencia de un suceso no afecta la ocurrencia del otro suceso	1	2,0	2,1
b) La probabilidad de la intersección es el producto de las probabilidades simples	5	10,2	10,4
c) **Si un suceso ocurre, el otro no puede ocurrir**	37	75,5	77,1
d) A y B son correctas	5	10,2	10,4
Total	48	98,0	100,0
Blanco	1	2,0	
Total	49	100,0	

Ítem 18. En un grupo de mil sujetos se pasó un test de inteligencia y se midió su rendimiento, clasificando a cada sujeto según fuese superior o inferior a la media en inteligencia, y según fuese apto o no apto en la prueba de rendimiento académico. Se encontraron 500 aptos, de los que 300 tenían inteligencia superior. De los 400 con inteligencia superior, 300 resultaron aptos. Extraemos un sujeto al azar y definimos los sucesos, A: "Ser superior a la media en inteligencia", y B: "Ser apto en rendimiento". ¿Son los sucesos A y B independientes?

Este ítems se ha puntuado siguiendo el siguiente criterio:
0: No se responde o se responde incorrectamente
1: Construye correctamente la tabla
2: Define correcta, y precisamente la fórmula de independencia
3. Calculo correcto
4. Identifica la independencia

Tabla A2.24. Resultados en el ítem 18

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
0	22	42,3	59,5	59,5
1	10	19,2	27,0	86,5
2	2	3,8	5,4	91,9
3	1	1,9	2,7	94,6
4	2	3,8	5,4	100,0
Total	37	71,2	100,0	
Blanco	15	28,8		
Total	52	100,0		

Ítem 19. Cogemos un número al azar menor que 1000. Sabiendo que es múltiplo de cinco. ¿Cuál es la probabilidad de que sea múltiplo de cuatro?

a) 1/20 **b) 1/4** c) 1/5 d) 4/5

Tabla A2.25. Resultados en el ítem 19

	Frecuencia	Porcentaje	Porcentaje válido
1	5	9,6	13,2
2	25	48,1	65,8
3	7	13,5	18,4
4	1	1,9	2,6
Total	38	73,1	100,0
Blanco	14	26,9	
Total	52	100,0	

Ítem 20. Hemos lanzado dos dados y sabemos que el producto de los dos números obtenidos ha sido 12 ¿Cual es la probabilidad de que ninguno de los dos números sea un 6?

Este ítem se ha puntuado siguiendo el siguiente criterio:
0: No se responde o se responde incorrectamente
1: Identifica los casos favorables o posibles
2: Identifica los casos favorables y posibles
3: Resuelve el problema correctamente bien por la regla de Laplace, bien por la fórmula del producto.

Tabla A2.26. Resultados en el ítem 20

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
0	15	30,6	40,5	40,5
1	8	16,3	21,6	62,2
2	4	8,2	10,8	73,0
3	10	20,4	27,0	100,0
Total	37	75,5	100,0	
Blanco	12	24,5		
Total	49	100,0		

Ítem 21. Una persona lanza un dado y anota si saca un número par o impar. Estos son los resultados:
par, impar, impar, par, par, impar, par, par, par, par, impar, impar, par, par, par
Se lanza el dado de nuevo
¿Cuál es la probabilidad de sacar un número par en la siguiente tirada?

Este ítem se ha puntuado siguiendo el siguiente criterio:
0: No se responde o se responde incorrectamente
1: Se da una estimación frecuencial de la probabilidad correctamente
2: Se calcula correctamente la probabilidad

Tabla A2.27. Resultados en el ítem 21

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
0	3	6,1	6,7	6,7
1	6	12,2	13,3	20,0
2	36	73,5	80,0	100,0
Total	45	91,8	100,0	
Blanco	4	8,2		
Total	49	100,0		

Ítem 22. Una persona lanza 3 veces la misma moneda, obteniendo en este orden los resultados siguientes: cara, cruz, cruz. Lanza la moneda una cuarta vez. ¿Qué resultado piensas que es más probable que salga a continuación?

Tabla A2.28. Resultados en el ítem 22

	Frecuencia	Porcentaje	Porcentaje válido
a) Cara	9	5,7	5,8
b) Cruz	3	1,9	1,9
c) Los dos sucesos son igual de probables	**144**	**91,7**	**92,3**
Total	156	99,4	100,0
En blanco	1	,6	
Total	157	100,0	

Ítem 23. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin reemplazamiento.

- ¿Cuál es la probabilidad de extraer una bola negra en segundo lugar, habiendo extraído una bola negra en primer lugar? *P* (N_2/ N_1)

Tabla A2.29. Resultados en el ítem 23

	Frecuencia	Porcentaje	Porcentaje válido
a. 1/ 3	**46**	**56,8**	**61,3**
b. 1/ 2	20	24,7	26,7
c. 1/ 6	9	11,1	12,0
Total	75	92,6	100,0
En blanco	6	7,4	
Total	81	100,0	

Ítem 24. Una caja tiene cuatro focos, de los cuales dos son defectuosos. Se sacan dos al azar, uno tras otro sin reemplazamiento.

Si el primer foco fue defectuoso, ¿qué es más probable?

Tabla A2.30. Resultados en el ítem 24

	Frecuencia	Porcentaje	Porcentaje válido
a. Que el segundo no sea defectuoso	**45**	**59,2**	**59,2**
b. Que el segundo sea defectuoso	6	7,9	7,9
c. Los dos sucesos son igual de probables	25	32,9	32,9
Total	76	100,0	100,0

Ítem 25. La probabilidad de que una mujer de 40 años tenga una mamografía positiva es el 10,3 %. La probabilidad de que una mujer de 40 años tenga cáncer de pecho y una mamografía positiva es de 0,8%. Una mujer de 40 años se realiza una mamografía y resulta positiva. ¿Cuál es la probabilidad de que realmente tenga cáncer?

Tabla A2.31. Resultados en el ítem 25

	Frecuencia	Porcentaje
a) $\frac{0,8}{10,3} = 0,0776$, **probabilidad de 7’76%**	**27**	**33,8**
b) $10,3 \times 0,8 = 8,24$, probabilidad del 8’24%	30	37,0
c) 0,8 %	24	29,2
Total	81	100,0

Ítem 26. 103 de cada 1000 mujeres de 40 años tienen una mamografía positiva. 8 de cada 1000 mujeres de 40 años tienen cáncer de pecho y obtienen una mamografía positiva. En una muestra representativa de mujeres de 40 años que se realizan una mamografía y resulta positiva. ¿Cuántas de ellas esperas que realmente tengan cáncer?

Tabla A2.32. Resultados en el ítem 26

	Frecuencia	Porcentaje
a) 7,7 de cada 100	**13**	**25,3**
b) 8,24 de cada 100	19	36,5
c) 0,8 de cada 100	20	38,2
Total	52	100,0

Ítem 27. En un país el 30 % de las personas tiene los ojos claros y el 15 % tiene los ojos claros y el pelo rubio. Si en este país cogemos una persona rubia al azar, ¿Cuál es la probabilidad de que tenga los ojos oscuros?

Este ítem se ha puntuado siguiendo el siguiente criterio:
0: No se responde o se responde incorrectamente
1: Construye correctamente el árbol
2: Da una fórmula correcta y/o resuelve

Tabla A2.33. Resultados en el ítem 27

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
0	5	9,6	17,2	17,2
1	16	30,8	55,2	72,4
2	8	15,4	27,6	100,0
Total	29	55,8	100,0	
Blanco	23	44,2		
Total	52	100,0		

Ítem 28. ¿Cuál de los siguientes sucesos es más probable?

Tabla A2.34. Resultados en el ítem 28

	Frecuencia	Porcentaje	Porcentaje válido
a. Que una madre tenga los ojos azules si su hija tiene los ojos azules	8	5,1	5,2
b. Que una niña tenga los ojos azules si su madre tiene los ojos azules	94	59,9	61,0
c. Los dos sucesos son igual de probables.	**52**	**33,1**	**33,8**
Total	154	98,1	100,0
En blanco	3	1,9	
Total	157	100,0	

Ítem 29. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin reemplazamiento.

- ¿Cuál es la probabilidad de extraer una bola negra en primer lugar, sabiendo que hemos extraído una bola negra en segundo lugar? $P\ (N_1/\ N_2)$

Tabla A2.35. Resultados en el ítem 29

	Frecuencia	Porcentaje	Porcentaje válido
a. 1/ 6	6	7,4	8,5
b. No se puede calcular	40	49,4	56,3
c. 1/ 3	**25**	**30,9**	**35,2**
Total	71	87,7	100,0
En blanco	10	12,3	
Total	81	100,0	

Ítem 30. ¿Cuál de los siguientes sucesos es más probable?

Tabla A2.36. Resultados en el ítem 30

	Frecuencia	Porcentaje	Porcentaje válido
a. Un divorcio es resultado de un matrimonio infeliz	**42**	**26,8**	**26,9**
b. Un matrimonio infeliz lleva a un divorcio	18	11,5	11,5
c. Los dos enunciados son igual de probables.	96	61,1	61,5
Total	156	99,4	100,0
En blanco	1	,6	
Total	157	100,0	

Ítem 31. ¿En cuál de las siguientes predicciones tendrás mayor seguridad?

Tabla A2.37. Resultados en el ítem 31

	Frecuencia	Porcentaje	Porcentaje válido
a. Predecir que una persona que está enferma, tiene fiebre.	2	1,3	1,3
b. Predecir que una persona que tiene fiebre está enferma.	**143**	**91,1**	**91,1**
c. Tendría igual seguridad en ambas predicciones.	12	7,6	7,6
Total	157	100,0	100,0

Ítem 32. Un test diagnóstico de cáncer fue administrado a todos los residentes en una gran ciudad. Un resultado positivo en el test es indicativo de cáncer y un resultado negativo es indicativo de ausencia de cáncer. ¿En cuál de las siguientes predicciones tienes más confianza?

Tabla A2.38. Resultados en el ítem 32

	Frecuencia	Porcentaje	Porcentaje válido
a. Predecir que una persona tiene cáncer si ha dado positivo en el test de diagnóstico.	38	24,2	25,2
b. Predecir un resultado positivo en el test de diagnóstico si la persona tiene cáncer.	**50**	**31,8**	**33,1**
c. Tengo la misma confianza en ambas predicciones.	63	40,1	41,7
Total	151	96,2	100,0
En blanco	6	3,8	
Total	157	100,0	

Ítem 33. Una bola se suelta en E. Si sale por R, ¿cuál es la probabilidad de que haya pasado por el canal 1?

a) 0,50

b) 0,66

c) 0,33

d) No se puede calcular

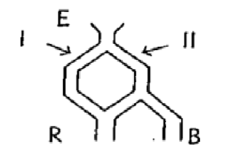


Tabla A2.39. Resultados en el ítem 33

	Frecuencia	Porcentaje	Porcentaje válido
1	38	73,1	76,0
2	7	13,5	14,0
3	2	3,8	4,0
4	3	5,8	6,0
Total	50	96,2	100,0
Blanco	2	3,8	
Total	52	100,0	

Ítem 34. Una caja tiene cuatro focos, de los cuales dos son defectuosos. Se sacan dos al azar, uno tras otro sin reemplazamiento.

- Si el segundo foco es defectuoso, ¿qué es lo más probable?

Tabla A2.40. Resultados en el ítem 34

	Frecuencia	Porcentaje	Porcentaje válido
a. Que el primero sea defectuoso	0	0	0
b. Que el primero no sea defectuoso	**31**	**40,8**	**40,8**
c. Los dos sucesos son igual de probables	45	59,2	59,2
Total	76	100,0	100,0

Ítem 35. En una urna hay tres tarjetas: una es azul por ambos lados, otra es roja por ambos lados y la última es por un lado azul y por otro roja. Seleccionamos una tarjeta al azar y la ponemos sobre la mesa tal y como la sacamos de la urna. La cara que vemos es roja, ¿cuál es la probabilidad de que la otra cara sea roja también?

Tabla A2.41. Resultados en el ítem 35

	Frecuencia	Porcentaje	Porcentaje válido
a.1/ 2	60	38,2	41,4
b. 2/ 3	**76**	**48,4**	**52,4**
c. 3/ 4	9	5,7	6,2
Total	145	92,4	100,0
En blanco	12	7,6	
Total	157	100,0	

Ítem 36. En un grupo de mil sujetos se pasó un test de inteligencia y se midió su rendimiento, clasificando a cada sujeto según fuese superior o inferior a la media en inteligencia, y según fuese apto o no apto en la prueba de rendimiento académico. Se encontraron 500 reflexivos, de los que 300 tenían inteligencia superior. ¿Cuál es la probabilidad de que un sujeto con inteligencia superior sea reflexivo?

Tabla A2.42. Resultados en el ítem 36

	Frecuencia	Porcentaje	Porcentaje válido
Incorrecto	84	53,8	86,6
Correcto	**13**	**8,0**	**13,4**
Total	97	61,8	100,0
En blanco	60	38,2	
Total	157	100,0	

Ítem 37. En una encuesta publicada en un periódico se informa que el 91% de los habitantes de una ciudad mienten usualmente y de ellos el 36% mienten sobre cosas importantes. Si nos encontramos con una persona al azar de esta ciudad, ¿cuál es la probabilidad de que mienta sobre cosas importantes?

Este ítem se ha puntuado siguiendo el siguiente criterio:
0: No se responde o se responde incorrectamente
1: Identifican los datos, realizan el diagrama en árbol correctamente
2: Resuelven el problema correctamente.

Tabla A2.43. Resultados en el ítem 37

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
0	10	20,4	22,7	22,7
1	7	14,3	15,9	38,6
2	27	55,1	61,3	100,0
Total	44	89,8	100,0	
Blanco	5	10,2		
Total	49	100,0		

Ítem 38. Un grupo de alumnos de un colegio se examina de Matemáticas e Inglés. La proporción de alumnos que aprueban Matemáticas es del 80% y la proporción de alumnos que aprueba Inglés es del 70%. Suponiendo que las notas en cada una de las asignaturas es independiente, ¿Cuál es la probabilidad de que un alumno escogido al azar haya aprobado ambas asignaturas?

Este ítem se ha puntuado siguiendo el siguiente criterio:
0: No se responde o se responde incorrectamente
1: Construye correctamente el árbol o dejar indicada la fórmula
2: Resuelve correctamente

Tabla A2.44. Resultados en el ítem 38

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
0	12	23,1	30,0	30,0
1	17	32,7	42,5	72,5
2	11	21,2	27,5	100,0
Total	40	76,9	100,0	
Blanco	12	23,1		
Total	52	100,0		

Ítem 39. Se lanza cuatro veces un dado. ¿Cuál es la probabilidad de no obtener ningún uno?

Este ítem se ha puntuado siguiendo el siguiente criterio:
0: No se responde o se responde incorrectamente
1: Calcula la probabilidad en un sólo lanzamiento
2: Resuelve correctamente

Tabla A2.45. Resultados en el ítem 39

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
0	12	23,1	40,0	40,0
1	13	25,0	43,3	83,3
2	5	9,6	16,7	100,0
Total	30	57,7	100,0	
Blanco	22	42,3		
Total	52	100,0		

Ítem 40. En una secuencia de tres dígitos, cada uno de los dígitos es seleccionado al azar del conjunto {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}. ¿Cuál es la probabilidad de que de este proceso resulten tres números diferentes?

Este ítem se ha puntuado siguiendo el siguiente criterio:
0: No se responde o se responde incorrectamente
1: Identifican el problema como de probabilidad compuesta
2: Llegan a la probabilidad pedida

Tabla A2.46. Resultados en el ítem 40

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
0	11	22,4	61,1	61,1
1	7	14,3	38,9	100,0
2	0	0	0	100,0
Total	18	36,7	100,0	
Blanco	31	63,3		
Total	49	100,0		

Ítem 41. Se tiene una bolsa con 1 bola azul y 2 bolas rojas. Se saca una bola al azar. Sin reponerla otra vez en la bolsa se saca una segunda bola. ¿Cuál suceso es más probable?

Tabla A2.47. Resultados en el ítem 41

	Frecuencia	Porcentaje	Porcentaje válido
a. Sacar primero una bola roja y luego una azul.	26	16,6	17,2
b. Sacar dos bolas rojas	15	9,6	9,9
c. Los dos sucesos son igual de probables	**110**	**70,1**	**72,8**
Total	151	96,2	100,0
En blanco	6	3,8	
Total	157	100,0	

Ítem 42. Un hombre, por accidente, mezcla dos baterías gastadas con tres baterías nuevas del mismo modelo. Con ayuda de un voltímetro va probando una batería tras otra hasta encontrar las dos baterías gastadas. ¿Cuál es la probabilidad de que tenga que probar sólo dos veces para encontrar las baterías gastadas?

Este ítem se ha puntuado siguiendo el siguiente criterio:
0: No se responde o se responde incorrectamente
1: Identifican los casos favorables y posibles
2: Identifica el problema como de probabilidad compuesta y que el segundo suceso depende del primero
3: Resuelven el problema correctamente.

Tabla A2.48. Resultados en el ítem 42

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
0	1	2,0	2,9	2,9
1	14	28,6	40,0	42,9
2	6	12,2	17,1	60,0
3	14	28,6	40,0	100,0
Total	35	71,4	100,0	
Blanco	14	28,6		
Total	49	100,0		

Ítem 43. De cada 1000 niños varones nacidos vivos 746 llegan a cumplir 65 años. La probabilidad de que una persona que acaba de cumplir 65 muera en los cinco años siguientes es 0,160. ¿Cuál es la probabilidad de que un niño varón alcance su 70 cumpleaños?

Este ítem se ha puntuado siguiendo el siguiente criterio:
0: No se responde o se responde incorrectamente
1: Construye correctamente el árbol
2: Dejar indicada la fórmula correctamente
3: Resuelve correctamente

Tabla A2.49. Resultados en el ítem 43

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
0	14	26,9	50,0	50,0
1	6	11,5	21,4	71,4
2	1	1,9	3,6	75,0
3	7	13,5	25,0	100,0
Total	28	53,8	100,0	
Blanco	24	46,2		
Total	52	100,0		

Ítem 44. En una escuela hay en marcha dos programas educativos, A y B. El 60% de los estudiantes están en el programa A y el 40% en el programa B. En el programa A, el 50% son chicos y en el programa B, el 25% son chicos. Encontramos al azar un estudiante ¿cuál es la probabilidad de que sea chico?

Tabla A2.50. Resultados en el ítem 44

	Frecuencia	Porcentaje	Porcentaje válido
a. 0,75	30	39,5	46,2
b. 0,715	5	6,6	7,7
c. 0,40	**30**	**39,5**	**46,2**
Total	65	85,5	100,0
En blanco	11	14,5	
Total	76	100,0	

Ítem 45. En una ciudad hay 60 hombres y 40 mujeres por cada 100 habitantes. 50 de cada 100 hombres y 25 de cada 100 mujeres fuman. Si cogemos 200 personas al azar, ¿cuántas de ellas fumarán?

Tabla A2.51. Resultados en el ítem 45

	Frecuencia	Porcentaje	Porcentaje válido
a. 150 personas	26	32,1	37,7
b. 80 personas	**35**	**43,2**	**50,7**
c. 40 personas	8	9,9	11,6
Total	69	85,2	100,0
En blanco	12	14,8	
Total	81	100,0	

Ítem 46. Al azar se seleccionan dos de cuatro cartas numeradas 1, 2, 3 y 4, una después de otra. Los números de las cartas se suman y si la suma es par, se gana un premio. Qué es más probable, ¿obtener el premio o no?

Este ítem se ha puntuado siguiendo el siguiente criterio:
0: No se responde o se responde incorrectamente
1: punto más por cada uno de los siguientes conocimientos aplicados correctamente:
2: Identifican los casos favorables y posibles
3: Identifica el problema como de probabilidad compuesta
4: Identifica que el orden es relevante
5: Identifica la regla de la suma
6: Llegan a la probabilidad pedida

Tabla A2.52. Resultados en el ítem 46

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
0	4	8,2	10,3	10,3
1	13	26,5	33,3	43,6
3	2	4,1	5,1	48,7
4	13	26,5	33,3	82,1
5	7	14,3	17,9	100,0
Total	39	79,6	100,0	
Blanco	10	20,4		
Total	49	100,0		

Ítem 47. Un taxi se vio implicado en un accidente nocturno con choque y huida posterior. Hay dos compañías de taxis en la ciudad, la Verde y la Azul. El 85% de los taxis de la ciudad son Verde y el 15% Azul. Un testigo identificó el taxi como Azul. El tribunal comprobó la fiabilidad del testigo bajo las mismas circunstancias que había la noche del accidente y llegó a la conclusión de que el testigo identificaba correctamente cada uno de los colores en el 80% de las ocasiones y fallaba en el 20%. ¿Cuál es la probabilidad de que el taxi implicado en el accidente fuera en efecto Azul?

Tabla A2.53. Resultados en el ítem 47

	Frecuencia	Porcentaje	Porcentaje válido
a. 80 %	91	58,0	74,6
b. 85 %	8	5,1	6,6
c. 48 %	**23**	**14,6**	**18,9**
Total	122	77,7	100,0
En blanco	35	22,3	
Total	157	100,0	

Ítem 48. Una fábrica dispone de dos máquinas M1 y M2 que fabrican bolas. La máquina M1 fabrica el 40 % de las bolas y la M2 el 60%. El 5% de las bolas fabricadas por M1 y el 1% de las fabricadas por M2 son defectuosas. Tomamos una bola al azar que resulta ser defectuosa. ¿Cuál es la probabilidad de que haya sido fabricada por M1?

Este ítem se ha puntuado siguiendo el siguiente criterio:
0: No se responde o se responde incorrectamente
1: Construye un diagrama en árbol adecuado
2: Calcula la probabilidad total
3. Da la fórmula de Bayes
4: Resuelve correctamente

Tabla A2.54. Resultados en el ítem 48

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
0	1	1,9	2,3	2,3
1	12	23,1	27,9	30,2
2	9	17,3	20,9	51,2
3	4	7,7	9,3	60,5
4	17	32,7	39,5	100,0
Total	43	82,7	100,0	
Blanco	9	17,3		
Total	52	100,0		

Ítem 49. Se sabe que 100 de cada 10000 personas está enferma. De cada 100 personas enfermas, el test será positivo en 80 de ellas. De cada 9900 personas sanas el test saldrá positivo en 950 de ellas. El test resulta positivo en 103 personas. ¿Cuantas de estas personas están realmente enfermas?

Este ítem se ha puntuado siguiendo el siguiente criterio:
0: No se responde o se responde incorrectamente
1: Construye un diagrama en árbol adecuado
2: Calcula la probabilidad total
3. Llega a la probabilidad pedida por Bayes o mediante una proporción o la plantea
4. Calcula tanto la probabilidad como el número esperado de enfermos

Tabla A2.55. Resultados en el ítem 49

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
0	7	14,3	17,9	17,9
1	20	40,8	51,3	69,2
2	8	16,3	20,5	89,7
3	3	6,1	7,7	97,4
4	1	2,0	2,6	100,0
Total	39	79,6	100,0	
Blanco	10	20,4		
Total	49	100,0		

ANEXO 3.

CUESTIONARIO PILOTO

Lee atentamente los enunciados y elabora o selecciona la respuesta que consideres correcta.

Ítem 1. Explica con tus propias palabras la diferencia entre una probabilidad simple y una probabilidad condicional.

Ítem 2. Define el espacio muestral de los siguientes experimentos aleatorios:
a) Observar el género (masculino/ femenino) de los hijos de las familias con tres descendientes.
b) Observar el género (masculino/ femenino) de los hijos de las familias con tres descendientes en los que dos son varones.

Ítem 3. Un taxi se vio implicado en un accidente nocturno con choque y huida posterior. Hay dos compañías de taxis en la ciudad, la Verde y la Azul. El 85% de los taxis de la ciudad son Verde y el 15% Azul. Un testigo identificó el taxi como Azul. El tribunal comprobó la fiabilidad del testigo bajo las mismas circunstancias que había la noche del accidente y llegó a la conclusión de que el testigo identificaba correctamente cada uno de los colores en el 80% de las ocasiones y fallaba en el 20%. ¿Cuál es la probabilidad de que el taxi implicado en el accidente fuera en efecto Azul?

☐ a. 80 %
☐ b. 15%
☐ c. (15/100) x (80/100)
☐ d. 41 %

Ítem 4. Se extrae una carta al azar de una baraja española (40 cartas con números del 1 al 7, sota caballo y rey): sea A el suceso "se extrae una carta de oros " y B el suceso "se extrae un rey" ¿Los sucesos A y B son independientes?

☐ a. No son independientes porque en la baraja hay un rey de oros.
☐ b. Sólo si sacamos primero una carta para ver si es rey y se vuelve a colocar
c. en la baraja. Luego sacamos una segunda carta para mirar si es oros.
☐ d. No, porque P(rey de oros)= P(rey) x P(oros)
☐ e. Sí, en todos los casos

Ítem 5. Una caja tiene cuatro focos, de los cuales dos son defectuosos. Se sacan dos al azar, uno tras otro sin reemplazamiento. Si el primer foco fue defectuoso, ¿qué es más probable?

☐ a. Que el segundo sea defectuoso
☐ b. Que el segundo no sea defectuoso
☐ c. Los dos sucesos son iguales de probables

Ítem 6. En un barrio de Granada se ha realizado una entrevista a un grupo de hombres, obteniendo los siguientes resultados:

	menos de 55 años	**más de 55 años**	**Total**
Ha sufrido un ataque al corazón	29	75	104
No ha sufrido ataque	401	275	676
Total	430	350	780

Si elegimos al azar una de estas personas:

a) ¿Cuál es la probabilidad de que haya tenido un ataque al corazón? _______________
b) ¿Cuál es la probabilidad de que tenga más de 55 años y además haya tenido uno o más ataques al corazón? ____
c) Dado que la persona escogida tiene más de 55 años, ¿cuál es la probabilidad de que haya tenido un ataque al corazón? ____
d) Dado que la persona escogida ha tenido un ataque al corazón, ¿cuál es la probabilidad de que tenga más de 55 años? ____

Ítem 7. La probabilidad de que una mujer de 40 años tenga una mamografía positiva es el 10,3%. La probabilidad de que una mujer de 40 años tenga cáncer de pecho y una mamografía positiva es de 0,8%. Una mujer de 40 años se realiza una mamografía y resulta positiva. ¿Cuál es la probabilidad de que realmente tenga cáncer?

- [ ] a. $\frac{0,8}{10,3} = 0,0776$, probabilidad de 7’76%
- [ ] b. $10,3 \times 0,8 = 8,24$, probabilidad del 8’24%
- [ ] c. 0,8 %

Ítem 8. Hemos lanzado dos dados y sabemos que el producto de los dos números obtenidos ha sido 12 ¿Cual es la probabilidad de que ninguno de los dos números sea un 6?

Ítem 9. Supón que Carlos Ferrero alcanza la final de Roland Garros en 2004. Para ganar el partido hay que ganar tres sets. ¿Cuál de los siguientes sucesos consideras más probable?

- [ ] a. Carlos Ferrero pierde el primer set.
- [ ] b. Carlos Ferrero pierde el primer set pero gana el partido.
- [ ] c. Los dos sucesos son iguales de probables.

Ítem 10. Un test diagnóstico de cáncer fue administrado a todos los residentes de una gran ciudad. Un resultado positivo en el test es indicativo de cáncer y un resultado negativo es indicativo de ausencia de cáncer. ¿En cuál de las siguientes predicciones tienes más confianza?

- [ ] a. Predecir que una persona tiene cáncer si ha dado positivo en el test de diagnóstico.
- [ ] b. Predecir un resultado positivo en el test de diagnóstico si la persona tiene cáncer.
- [ ] c. Tengo la misma confianza en ambas predicciones.

Ítem 11. En una ciudad hay 60 hombres y 40 mujeres por cada 100 habitantes. 50 de cada 100 hombres y 25 de cada 100 mujeres fuman. Si cogemos 200 personas al azar, ¿cuántas de estas 200 personas fumarán?

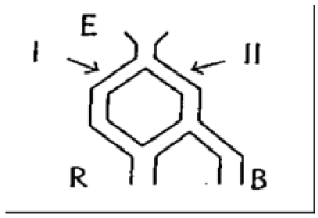
Ítem 12. Una persona lanza un dado y anota si saca un número par o impar. Estos son los resultados al lanzarlo 15 veces:

par, impar, impar, par, par, impar, par, par, par, par, impar, impar, par, par, par

Lanza el dado de nuevo ¿Cuál es la probabilidad de sacar un número par en esta nueva tirada?

Ítem 13. Un grupo de alumnos de un colegio se examina de Matemáticas e Inglés. La proporción de alumnos que aprueban Matemáticas es del 80% y la proporción de alumnos que aprueba Inglés es del 70%. Suponiendo que las notas en cada una de las asignaturas es independiente, ¿Cuál es la probabilidad de que un alumno escogido al azar haya aprobado ambas asignaturas?

Ítem 14. Una bola se suelta por la entrada E. Si sale por R, ¿Cuál es la probabilidad de que haya pasado por el canal 1?



☐ a. 0,50
☐ b. 0,33
☐ c. 0,66
☐ d. No se puede calcular

Ítem 15. En una encuesta publicada en un periódico se informa que el 91% de los habitantes de una ciudad mienten usualmente y de ellos el 36% mienten sobre cosas importantes. Si elegimos al azar a una persona de esta ciudad, ¿cuál es la probabilidad de que mienta sobre cosas importantes?

Ítem 16. Una fábrica dispone de dos máquinas M1 y M2 que fabrican bolas. La máquina M1 fabrica el 40 % de las bolas y la M2 el 60%. El 5% de las bolas fabricadas por M1 y el 1% de las fabricadas por M2 son defectuosas. Tomamos una bola al azar que resulta ser defectuosa. ¿Cuál es la probabilidad de que haya sido fabricada por M1?

Ítem 17. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin reemplazamiento. ¿Cuál es la probabilidad de extraer una bola negra en segundo lugar, habiendo extraído una bola negra en primer lugar? P (N_2/ N_1)

☐ a. 1/ 2
☐ b. 1/ 6
☐ c. 1/ 3
☐ d. 1/ 4

¿Cuál es la probabilidad de extraer una bola negra en primer lugar, sabiendo que hemos extraído una bola negra en segundo lugar? P (N_1/ N_2)

☐ a. 1/ 3
☐ b. No se puede calcular
☐ c. 1/ 6
☐ d. 1/ 2

Ítem 18. Se tiene una bolsa con 1 bola azul y 2 bolas rojas. Se saca una bola al azar. Sin reponerla otra vez en la bolsa se saca una segunda bola. ¿Cuál suceso es más probable?

☐ a. Sacar dos bolas rojas
☐ b. Sacar primero una bola roja y luego una azul.
☐ c. Los dos sucesos son iguales de probables

ANEXO 4.

RESULTADOS DE LA PRUEBA DEL CUESTIONARIO PILOTO

Ítem 1. Explica con tus propias palabras la diferencia entre una probabilidad simple y una probabilidad condicional.

Este ítem se ha puntuado siguiendo el siguiente criterio:

0. Respuesta totalmente incorrecta o en blanco
1. Define correcta, pero imprecisamente una de las probabilidades pedidas.
2. Define correcta, y precisamente una de las probabilidades pedidas.
3. Define correcta, pero imprecisamente las dos probabilidades pedidas.
4. Define correcta y de manera precisa las dos probabilidades pedidas.

Tabla A4.1. Resultados en el ítem 1

	Matemáticas			Psicología		
	Frecuencia	Porcentaje	Porcentaje acumulado	Frecuencia	Porcentaje	Porcentaje acumulado
Blanco	0	0	0	5	8,8	8,8
1	1	2,7	2,7	2	3,5	12,3
2	7	18,9	21,6	1	1,8	14,1
3	6	16,2	37,8	20	35,1	49,2
4	23	62,2	100,0	29	50,9	100,0
Total	37	100,0		57	100,0	

Ítem 2. Define el espacio muestral de los siguientes experimentos aleatorios:

a. Observar el género (masculino/ femenino) de los hijos de las familias con tres descendientes.
b. Observar el género (masculino/ femenino) de los hijos de las familias con tres descendientes en los que dos son varones.

0: No responde o responde incorrectamente.
1: Falta más de un suceso del espacio muestral o no tiene en cuenta el orden.
2: Responde correctamente.

Tabla A4.2. Resultados en el ítem 2

	Matemáticas			Psicología		
	Frecuencia	Porcentaje	Porcentaje acumulado	Frecuencia	Porcentaje	Porcentaje acumulado
Blanco	2	5,4	5,4	9	15,8	15,8
0	6	16,2	21,6	7	12,3	28,1
1	21	56,8	78,4	18	31,6	59,7
2	8	21,6	100	23	40,4	100,0
Total	37	100,0		57	100,0	

Ítem 3. Un taxi se vio implicado en un accidente nocturno con choque y huida posterior. Hay dos compañías de taxis en la ciudad, la Verde y la Azul. El 85% de los taxis de la ciudad son Verde y el 15% Azul. Un testigo identificó el taxi como Azul. El tribunal comprobó la fiabilidad del testigo bajo las mismas circunstancias que había la noche del accidente y llegó a la conclusión de que el testigo identificaba correctamente cada uno de los colores en el 80% de las ocasiones y fallaba en el 20%. ¿Cuál es la probabilidad de que el taxi implicado en el accidente fuera en efecto Azul?
a) 80 %
b) 15%
c) (15/100) x (80/100)
d) 41 %

Tabla A4.3. Resultados en el ítem 3

	Matemáticas		Psicología	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
a) 80 %	7	18,9	10	17,5
b) 15%	6	16,2	2	3,5
c) (15/100) x (80/100)	18	48,6	17	29,8
d) 41 %	4	10,8	16	28,1
Blanco	2	5,4	12	21,1
Total	37	100,0	57	100,0

Ítem 4. Se extrae una carta al azar de una baraja americana: sea A el suceso "se extrae un trébol" y B el suceso "se extrae una reina" ¿Los sucesos A y B son independientes?
a. No son independientes porque en la baraja hay una reina de tréboles.
b. Sólo si sacamos primero una carta para ver si es reina y se vuelve a colocar en la baraja y luego sacamos una segunda carta y para mirar si es trébol.
c. No, porque P(reina de trébol)= P(reina) x P(trébol) .
d. Sí, en todos los casos.

Tabla A4.4. Resultados en el ítem 4

	Matemáticas		Psicología	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
a. No son independientes porque en la baraja hay una reina de tréboles.	12	32,4	15	26,3
b. Sólo si sacamos primero una carta para ver si es reina y se vuelve a colocar en la baraja. Luego sacamos una segunda carta para mirar si es trébol.	18	48,6	9	15,8
c. No, porque P(reina de trébol)= P(reina) x P(trébol)	4	10,8	18	31,6
d. Sí, en todos los casos	3	8,1	10	17,5
Blanco	0	0	5	8,8
Total	37	100,0	57	100,0

Ítem 5. Una caja tiene cuatro focos, de los cuales dos son defectuosos. Se sacan dos al azar, uno tras otro sin reemplazamiento. Si el primer foco fue defectuoso, ¿qué es más probable?

a) que el segundo sea defectuoso
b) **que el segundo no sea defectuoso**
c) los dos sucesos son iguales de probables

Tabla A4.5. Resultados en el ítem 5

	Matemáticas		Psicología	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
a) Que el segundo sea defectuoso	0	0	0	0
b) Que el segundo no sea defectuoso	30	81,1	47	82,5
c) Los dos sucesos son iguales de probables	7	18,9	8	14,0
Blanco	0	0	2	3,5
Total	37	100,0	57	100,0

Ítem 6. En un barrio de Granada se ha realizado una entrevista a un grupo de hombres, obteniendo los siguientes resultados:

	menos de 55 años	**más de 55 años**	**Total**
Ha sufrido un ataque al corazón	29	75	104
No ha sufrido ataque	401	275	676
Total	430	350	780

Si elegimos al azar una de estas personas:
a) ¿Cuál es la probabilidad de que haya tenido un ataque al corazón?
b) ¿Cuál es la probabilidad de que tenga más de 55 años y además haya tenido uno o más ataques al corazón?
c) Dado que la persona escogida tiene más de 55 años, ¿cuál es la probabilidad de que haya tenido un ataque al corazón?
d) Dado que la persona escogida ha tenido un ataque al corazón, ¿Cuál es la probabilidad de que tenga más de 55 años?

Tabla A5.6. Resultados en el ítem 6a

	Matemáticas		Psicología	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Incorrecto	1	2,7	2	3,5
Correcto	35	94,6	50	87,7
Blanco	1	2,7	5	8,8
Total	37	100,0	57	100,0

Tabla A4.7. Resultados en el ítem 6b

	Matemáticas		Psicología	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Incorrecto	16	43,2	31	54,4
Correcto	17	45,9	17	29,8
Blanco	4	10,8	9	15,8
Total	37	100,0	57	100,0

Tabla A4.8. Resultados en el ítem 6c

	Matemáticas		Psicología	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Incorrecto	9	24,3	22	38,6
Correcto	27	73,0	24	42,1
Blanco	1	2,7	11	19,3
Total	37	100,0	57	100,0

Tabla A4.9. Resultados en el ítem 6d

	Matemáticas		Psicología	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Incorrecto	9	24,3	17	29,8
Correcto	27	73,0	24	42,1
Blanco	1	2,7	16	28,1
Total	37	100,0	57	100,0

Ítem 7. La probabilidad de que una mujer de 40 años tenga una mamografía positiva es el 10,3 %. La probabilidad de que una mujer de 40 años tenga cáncer de pecho y una mamografía positiva es de 0,8%. Una mujer de 40 años se realiza una mamografía y resulta positiva. ¿Cuál es la probabilidad de que realmente tenga cáncer?

a) $\frac{0,8}{10,3}=0,0776$ **, probabilidad de 7'76%**	
b) $10,3\times 0,8=8,24$, probabilidad del 8'24%	
c) 0,8 %	

Tabla A4.10. Resultados en el ítem 7

	Matemáticas		Psicología	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
a) $\frac{0,8}{10,3}=0,0776$, probabilidad de 7,76%	16	43,2	17	29,8
b) $10,3\times 0,8=8,24$, probabilidad del 8,24%	8	21,6	16	28,1
c) 0,8 %	12	32,4	18	31,6
Blanco	1	2,7	6	10,5
Total	37	100,0	57	100,0

Ítem 8. Hemos lanzado dos dados y sabemos que el producto de los dos números obtenidos ha sido 12 ¿Cuál es la probabilidad de que ninguno de los dos números sea un 6?

0. No se responde, o se responde incorrectamente.
1. Identifica los casos favorables o posibles, pero no resuelve el problema.
2. Identifica los casos favorables y posibles, plantea el problema pero comete algún error.
3. Resuelve el problema correctamente bien por la regla de Laplace, bien por la fórmula del producto.

.

Tabla A4.11. Resultados en el ítem 8

	Matemáticas			Psicología		
	Frecuencia	Porcentaje	Porcentaje acumulado	Frecuencia	Porcentaje	Porcentaje acumulado
Blanco	1	2,7	2,7	7	12,3	12,3
0	13	35,1	37,8	23	40,4	52,7
1	4	10,8	48,6	12	21,1	73,8
2	7	18,9	67,5	9	15,8	89,6
3	12	32,4	100,0	6	10,5	100,0
Total	37	100,0		57	100,0	

Ítem 9. Supón que Carlos Ferrero alcanza la final de Roland Garros en 2004. ¿Cuál de los siguientes sucesos consideras más probable?

a) Carlos Ferrero pierde el primer set
b) Carlos Ferrero pierde el primer set pero gana el partido
c) Los dos sucesos son iguales de probables

Tabla A4.12. Resultados en el ítem 9

	Matemáticas		Psicología	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
a. Carlos Ferrero pierde el primer set.	10	27,0	19	33,3
b. Carlos Ferrero pierde el primer set pero gana el partido.	6	16,2	6	10,5
c. Los dos sucesos son iguales de probables.	19	51,4	31	54,4
Blanco	2	5,4	1	1,8
Total	37	100,0	57	100,0

Ítem 10. Un test diagnóstico de cáncer fue administrado a todos los residentes en una gran ciudad. Un resultado positivo en el test es indicativo de cáncer y un resultado negativo es indicativo de ausencia de cáncer. ¿En cuál de las siguientes predicciones tienes más confianza?

a) Predecir que una persona tiene cáncer si ha dado positivo en el test de diagnóstico.
b) Predecir un resultado positivo en el test de diagnóstico si la persona tiene cáncer.
c) Tengo la misma confianza en ambas predicciones.

Tabla A4.13. Resultados en el ítem 10

	Matemáticas		Psicología	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
a. Predecir que una persona tiene cáncer si ha dado positivo en el test de diagnóstico.	4	10,8	9	15,8
b. Predecir un resultado positivo en el test de diagnóstico si la persona tiene cáncer.	14	37,8	18	31,6
c. Tengo la misma confianza en ambas predicciones.	17	45,9	28	49,1
Blanco	2	5,4	2	3,5
Total	37	100,0	57	100,0

Ítem 11. En una ciudad hay 60 hombres y 40 mujeres por cada 100 habitantes. 50 de cada 100 hombres y 25 de cada 100 mujeres fuman. Si cogemos 200 personas al azar, ¿cuántas de ellas fumarán?

0. No se responde, o se responde incorrectamente
1. Identifica los datos o construye un diagrama de árbol.
2. Calcula la probabilidad total de personas que fuman.
3. Llega a dar el número esperado de personas que fuman.

Tabla A4.14. Resultados en el ítem 11

	Matemáticas			Psicología		
	Frecuencia	Porcentaje	Porcentaje acumulado	Frecuencia	Porcentaje	Porcentaje acumulado
Blanco	3	8,1	8,1	10	17,5	17,5
0	6	16,2	24,3	15	26,3	43,8
1	7	18,9	43,2	10	17,5	61,3
2	4	10,8	54,1	6	10,5	71,8
3	17	45,9	100,0	16	28,1	100,0
Total	37	100,0		57	100,0	

Ítem 12. Una persona lanza un dado y anota si saca un número par o impar. Estos son los resultados:
par, impar, impar, par, par, impar, par, par, par, par, impar, impar, par, par, par.
Lanza el dado de nuevo.
¿Cuál es la probabilidad de sacar un número par en la siguiente tirada?

0. No se responde, o se responde incorrectamente.
1. Se da una estimación frecuencial de la probabilidad correctamente.
2. Se calcula correctamente la probabilidad, dando el valor ½.

Tabla A4.15. Resultados en el ítem 12

	Matemáticas			Psicología		
	Frecuencia	Porcentaje	Porcentaje acumulado	Frecuencia	Porcentaje	Porcentaje acumulado
Blanco	5	13,5	13,5	13	22,8	22,8
0	1	2,7	16,2	3	5,3	28,1
1	3	8,1	24,3	7	12,3	40,4
2	28	75,7	100	34	59,6	100
Total	37	100,0		57	100,0	

Ítem 13. Un grupo de alumnos de un colegio se examina de Matemáticas e Inglés. La proporción de alumnos que aprueban Matemáticas es del 80% y la proporción de alumnos que aprueba Inglés es del 70%. Suponiendo que las notas en cada una de las asignaturas es independiente, ¿Cuál es la probabilidad de que un alumno escogido al azar haya aprobado ambas asignaturas?

0. No se responde o se responde incorrectamente.
1. Construye correctamente el árbol o dejar indicada la fórmula, pero no resuelve el problema.
2. Resuelve correctamente.

Tabla A4.16. Resultados en el ítem 13

	Matemáticas			Psicología		
	Frecuencia	Porcentaje	Porcentaje acumulado	Frecuencia	Porcentaje	Porcentaje acumulado
Blanco	3	8,1	8,1	18	31,6	31,6
0	0	0	8,1	7	12,3	43,9
1	2	5,4	13,5	6	10,5	54,4
2	32	86,5	100	26	45,6	100
Total	37	100,0		57	100,0	

Ítem 14. Una bola se suelta en E. Si sale por R, ¿Cuál es la probabilidad de que haya pasado por el canal 1?

a) 0,50
b) 0,66
c) 0,33
d) No se puede calcular

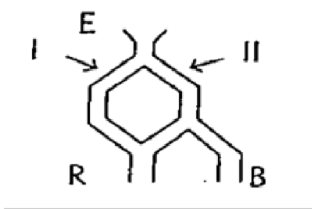


Tabla A4.17. Resultados en el ítem 14

	Matemáticas		Psicología	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
a) 0,50	27	73,0	43	75,4
b) **0,66**	3	8,1	4	7,0
c) 0,33	5	13,5	3	5,3
d) No se puede calcular	2	5,4	5	8,8
Blanco	0	0	2	3,5
Total	37	100,0	57	100,0

Ítem 15. En una encuesta publicada en un periódico se informa que el 91% de los habitantes de una ciudad mienten usualmente y de ellos el 36% mienten sobre cosas importantes. Si nos encontramos con una persona al azar de esta ciudad, ¿cuál es la probabilidad de que mienta sobre cosas importantes?

0. No se responde o se responde incorrectamente.
1. Identifican los datos o realizan el diagrama en árbol correctamente.
2. Resuelven el problema correctamente empleando la notación adecuada

Tabla A4.18. Resultados en el ítem 15

	Matemáticas			Psicología		
	Frecuencia	Porcentaje	Porcentaje acumulado	Frecuencia	Porcentaje	Porcentaje acumulado
Blanco	4	10,8	10,8	16	28,1	28,1
0	7	18,9	29,7	23	40,4	68,5
1	15	40,5	70,2	7	12,3	80,8
2	11	29,7	100,0	11	19,3	100,0
Total	37	100,0		57	100,0	

Ítem 16. Una fábrica dispone de dos máquinas M1 y M2 que fabrican bolas. La máquina M1 fabrica el 40 % de las bolas y la M2 el 60%. El 5% de las bolas fabricadas por M1 y el 1% de las fabricadas por M2 son defectuosas. Tomamos una bola al azar que resulta ser defectuosa. ¿Cuál es la probabilidad de que haya sido fabricada por M1?

0. No se responde o se responde incorrectamente
1. Construye un diagrama en árbol adecuado.
2: Diagrama correcto e identifica el problema como de probabilidad condicional,
3. Calcula correctamente la probabilidad total.
4: Resuelve correctamente, como en el caso siguiente

Tabla A4.19. Resultados en el ítem 16

	Matemáticas			Psicología		
	Frecuencia	Porcentaje	Porcentaje acumulado	Frecuencia	Porcentaje	Porcentaje acumulado
Blanco	3	8,1	8,1	15	26,3	26,3
0	5	13,5	21,6	8	14,0	40,3
1	13	35,1	56,7	10	17,5	57,8
2	8	21,6	78,3	13	22,8	80,6
3	5	13,5	91,8	2	3,5	84,1
4	3	8,1	99,9	9	15,8	99,9
Total	37	100,0		57	100,0	

Ítem 17a. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin reemplazamiento. ¿Cuál es la probabilidad de extraer una bola negra en segundo lugar, habiendo extraído una bola negra en primer lugar? P (N_2/ N_1)

a) 1/ 2
b) 1/ 6
c) **1/ 3**
d) 1/ 4

Tabla A4.20. Resultados en el ítem 17a

	Matemáticas		Psicología	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
a) 1/2	0	0	5	8,8
b) 1/ 6	8	21,6	14	24,6
c) 1/ 3	26	70,3	27	47,4
d) 1/ 4	2	5,4	7	12,3
Blanco	1	2,7	4	7,0
Total	37	100,0	57	100,0

Ítem 17b. ¿Cuál es la probabilidad de extraer una bola negra en primer lugar, sabiendo que hemos extraído una bola negra en segundo lugar? *P (N1/ N2)*

a) **1/ 3**
b) No se puede calcular
c) 1/6
d) 1/ 2

Tabla A4.21. Resultados en el ítem 17b

	Matemáticas		Psicología	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
a) 1/ 3	5	13,5	15	26,3
b) No se puede calcular	10	27,0	19	33,3
c) 1/ 6	2	5,4	6	10,5
d) 1/ 2	15	40,5	11	19,3
Blanco	5	13,5	6	10,5
Total	37	100,0	57	100,0

Ítem 18. Se tiene una bolsa con 1 bola azul y 2 bolas rojas. Se saca una bola al azar. Sin reponerla otra vez en la bolsa se saca una segunda bola. ¿Cuál suceso es más probable?

a) Sacar dos bolas rojas
b) Sacar primero una bola roja y luego una azul.
c) Los dos sucesos son iguales de probables

Tabla A4.22. Resultados en el ítem 18

	Matemáticas		Psicología	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
a. Sacar dos bolas rojas	2	5,4	10	17,5
b. Sacar primero una bola roja y luego una azul.	5	13,5	3	5,3
c. Los dos sucesos son iguales de probables	28	75,7	41	71,9
Blanco	2	5,4	3	5,3
Total	37	100,0	57	100,0

ANEXO 5.

SEGUNDO CUESTIONARIO PARA RECOGIDA DE DATOS DE EXPERTOS

A continuación presentamos, para cada uno de los 18 ítems que componen el actual instrumento piloto, tres enunciados alternativos. Todos ellos evalúan el mismo contenido relacionado con la probabilidad condicional.

Requerimos su colaboración para:

- Ordenar los tres enunciados propuestos para cada uno de los ítems, según considere que son más o menos adecuados, metodológicamente. En esta ordenación 1 indica el ítem más adecuado y 3 el menos adecuado.
- Sugerir mejoras en la redacción del ítem (sin variar el contenido).

Contenido 1: Definición de la probabilidad condicional

Ítem 1a. Explica con tus propias palabras la diferencia entre una probabilidad simple y una probabilidad condicional.

Ítem 1b. Define probabilidad simple y probabilidad condicional, dando un ejemplo de cada una de ellas.

Ítem 1c. Define probabilidad simple y probabilidad condicional.

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

Contenido 2: Reconocer que la probabilidad condicional supone una restricción del espacio muestral

Ítem 2a. Define el espacio muestral de los siguientes experimentos aleatorios:
a. Observar el género (varón/ mujer) de los hijos de las familias con tres descendientes, teniendo en cuenta el orden de nacimiento.
b. Observar el género (varón/ mujer) de los hijos de las familias con tres descendientes en los que dos o tres son varones, teniendo en cuenta el orden de nacimiento.

Ítem 2b. Define el espacio muestral de los siguientes experimentos aleatorios:
a. Observar el género (varón/ mujer) de los hijos de las familias con tres descendientes.
b. Observar el género (varón/ mujer) de los hijos de las familias con tres descendientes en los que dos o tres son varones.

Ítem 2c. Complete el espacio muestral de los siguientes experimentos aleatorios:
a. Observar el género (varón/ mujer) de los hijos de las familias con tres descendientes (ejemplo MVM,...).

b. Observar el género (varón/ mujer) de los hijos de las familias con tres descendientes en los que dos o tres son varones (ejemplo VVM,...).

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

Contenido 3: Teorema de Bayes y falacia de tasas base (no tener en cuenta las probabilidades a priori)

Ítem 3a. Un taxi se vio implicado en un accidente nocturno con choque y huida posterior. Hay dos compañías de taxis en la ciudad, la Verde y la Azul. El 85% de los taxis de la ciudad son Verdes y el 15% Azules. Un testigo identificó el taxi como Azul. El tribunal comprobó la fiabilidad del testigo bajo las mismas circunstancias que había la noche del accidente y llegó a la conclusión de que el testigo identificaba correctamente cada uno de los colores en el 80% de las ocasiones y fallaba en el 20%. ¿Cuál es la probabilidad de que el taxi implicado en el accidente fuera en efecto Azul?

	a. 80 %
	b. 15 %
	c. 12 %
	d. 41 %

Ítem 3b. Un taxi se vio implicado en un accidente nocturno con choque y huida posterior. Hay dos compañías de taxis en la ciudad, la Verde y la Azul. El 85% de los taxis de la ciudad son Verdes y el 15% Azules. Un testigo identificó el taxi como Azul. El tribunal comprobó la fiabilidad del testigo bajo las mismas circunstancias que había la noche del accidente y llegó a la conclusión de que el testigo identificaba correctamente cada uno de los colores en el 80% de las ocasiones y fallaba en el 20%. ¿Cuál es la probabilidad de que el taxi implicado en el accidente fuera en efecto Azul?

	a. $\frac{80}{100}$
	b. $\frac{15}{100}$
	c. $\frac{15}{100}\times\frac{80}{100}$
	d. $\frac{15\times 80}{85\times 20+15\times 80}$

Item 3c. Un taxi se vio implicado en un accidente nocturno con choque y huida posterior. Hay dos compañías de taxis en la ciudad, la Verde y la Azul. 85 de cada 100 taxis de la ciudad son Verdes y 15 de cada 100 Azules. Un testigo identificó el taxi como Azul. El tribunal comprobó la fiabilidad del testigo bajo las mismas circunstancias que había la noche del accidente y llegó a la conclusión de que el testigo identificaba correctamente cada uno de los colores 4 de cada 5 ocasiones y fallaba en 1 de cada 5. ¿Cuál es la probabilidad de que el taxi implicado en el accidente fuera en efecto Azul?

- a. 4/5
- b. 15/100
- c. (15/100) X (4/5)
- d. 12 / (17 + 12)

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

Contenido 4: Distinguir sucesos dependientes, independientes y mutuamente excluyentes

Ítem 4a. Se extrae una carta al azar de una baraja española (40 cartas con cuatro palos: oros, bastos, espadas y copas. Cada palo tiene los números del 1 al 7, sota caballo y rey). Sea A el suceso "se extrae una carta de oros" y B el suceso "se extrae un rey". ¿Los sucesos A y B son independientes?

- a. No son independientes porque en la baraja hay un rey de oros.
- b. Sólo si sacamos primero una carta para ver si es rey y se vuelve a colocar en la baraja y luego sacamos una segunda carta para mirar si es oros.
- c. No, porque P(rey de oros)= P(rey) x P(oros)
- d. Sí, en todos los casos

Ítem 4b. Se extrae una carta al azar de una baraja española (40 cartas con cuatro palos: oros, bastos, espadas y copas. Cada palo tiene los números del 1 al 7, sota caballo y rey). Sea A el suceso "se extrae una carta de oros" y B el suceso "se extrae un rey". ¿Los sucesos A y B son independientes?

- a. No son independientes porque en la baraja hay un rey de oros.
- b. Sólo si sacamos primero una carta para ver si es rey y se vuelve a colocar en la baraja y luego sacamos una segunda carta para mirar si es oros.
- c. Si, porque P(rey de oros)= P(rey) x P(oros)
- d. No, porque P(rey / oros) $\neq$ P(rey)

Ítem 4c. Una baraja española tiene 40 cartas con cuatro palos: oros, bastos, espadas y copas. Cada palo tiene los números del 1 al 7, sota caballo y rey. Se extrae una carta al azar. Sea A el suceso "se extrae una carta de oros " y B el suceso "se extrae un rey". ¿Son los sucesos A y B independientes?

- a. No son independientes porque en la baraja hay un rey de oros.
- b. Sólo si sacamos primero una carta para ver si es rey y se vuelve a colocar en la baraja. Luego sacamos una segunda carta para mirar si es oros.
- c. Si porque la probabilidad de obtener el rey de oros se obtiene multiplicando la probabilidad de rey por la probabilidad de oros
- d. No, porque la probabilidad de obtener el rey si sabemos que la carta es oros es diferente de la probabilidad de ser rey.

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

Contenido 5: Resolver problemas en contexto de muestreo sin reposición

Ítem 5a. Una caja tiene cuatro bombillas, de las cuales dos son defectuosas. Se sacan dos al azar, una tras otra sin reemplazamiento. Si la primera bombilla fue defectuosa, ¿Cuál de las siguientes afirmaciones es cierta?

- [ ] a. Hay más probabilidad de que la segunda bombilla sea también defectuosa
- [ ] b. Hay más probabilidad de que la segunda bombilla no sea defectuosa
- [ ] c. La probabilidad que la segunda bombilla sea defectuosa es la misma que la probabilidad de que no lo sea.

Ítem 5b. Una caja tiene cuatro bombillas, de las cuales dos son defectuosas. Se sacan dos al azar, una tras otra sin reemplazamiento. Si la primera bombilla fue defectuosa, ¿Qué es más probable?

- [ ] a. Que la segunda sea defectuosa
- [ ] b. Que la segunda no sea defectuosa
- [ ] c. Los dos sucesos son iguales de probables

Ítem 5c. Una caja tiene cuatro bombillas, de las cuales dos son defectuosas. Se sacan dos al azar, una tras otra sin reemplazamiento. Si la primera bombilla fue defectuosa, ¿Cuál de las siguientes afirmaciones es cierta?

- [ ] a. La probabilidad de que la segunda bombilla sea defectuosa es mayor de 1/2
- [ ] b. La probabilidad de que la segunda bombilla sea defectuosa es menor de 1/2
- [ ] c. La probabilidad de que la segunda bombilla sea defectuosa es igual a 1/2

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

Contenido 6: Distinguir probabilidad simple, condicional y compuesta. Distinguir una probabilidad simple y su inversa

Ítem 6a. En una población se ha realizado una entrevista a un grupo de hombres, obteniendo los siguientes resultados:

	menos de 55 años	**más de 55 años**	**Total**
Ha sufrido un ataque al corazón	29	75	104
No ha sufrido ataque	401	275	676
Total	430	350	780

Si elegimos al azar una de estas personas:
a. ¿Cuál es la probabilidad de que haya tenido un ataque al corazón?
b. ¿Cuál es la probabilidad de que tenga más de 55 años y además haya tenido un ataque al corazón?
c. Dado que la persona escogida tiene más de 55 años, ¿cuál es la probabilidad de que haya tenido un ataque al corazón?
d. Dado que la persona escogida ha tenido un ataque al corazón, ¿cuál es la probabilidad de que tenga más de 55 años?

Ítem 6b. En una población se ha realizado una entrevista a un grupo de hombres, obteniendo los siguientes resultados:

	menos de 55 años	**más de 55 años**	**Total**
Ha sufrido uno o más ataques al corazón	29	75	104
No ha sufrido ataques al corazón	401	275	676
Total	430	350	780

Si elegimos al azar una de estas personas:
a. ¿Cuál es la probabilidad de que haya tenido un ataque al corazón?
b. ¿Cuál es la probabilidad de que tenga más de 55 años y además haya tenido uno o más ataques al corazón?
c. Dado que la persona escogida tiene más de 55 años, ¿cuál es la probabilidad de que haya tenido un ataque al corazón?
d. Dado que la persona escogida ha tenido uno o más ataque al corazón, ¿cuál es la probabilidad de que tenga más de 55 años?

Ítem 6c. En una población se ha realizado una entrevista a un grupo de hombres, obteniendo los siguientes resultados:

	menos de 55 años	**más de 55 años**	**Total**
Ha sufrido un ataque al corazón	29	75	104
No ha sufrido ataque	401	275	676
Total	430	350	780

Si elegimos al azar una de estas personas:
a. ¿Cuál es la probabilidad de que haya tenido un ataque al corazón?
b. ¿Cuál es la probabilidad de que tenga más de 55 años y además haya tenido uno o más ataques al corazón?
c. Sabiendo que la persona escogida tiene más de 55 años, ¿cuál es la probabilidad de que haya tenido un ataque al corazón?
d. Sabiendo que la persona escogida ha tenido uno o más ataque al corazón, ¿cuál es la probabilidad de que tenga más de 55 años?

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

Contenido 7: Resolver problemas de probabilidad condicional a partir de probabilidades simples y compuestas

Ítem 7a. Al realizarse una mamografía, el 10'3 % de las mujeres de 40 años tienen un resultado positivo. El 0'8 % de las mujeres de 40 años tienen un resultado positivo en la mamografía y además tienen cáncer de pecho. Una mujer de 40 años se realiza una mamografía y resulta positiva. ¿Cuál es la probabilidad de que realmente tenga cáncer?

☐ a. $\frac{0,8}{10,3} = 0,0776$, probabilidad de 7,76%

☐ b. $10,3 \times 0,8 = 8,24$, probabilidad del 8,24%

☐ c. 8/1000, probabilidad del 0,8 %

Ítem 7b. La probabilidad de que una mujer tenga una mamografía positiva es el 10.3%. La probabilidad de que una mujer tenga cáncer de pecho y una mamografía positiva es de 0.8%. Una mujer se realiza una mamografía y resulta positiva. ¿Cuál es la probabilidad de que realmente tenga cáncer?

☐ a. $\frac{0,8}{10,3} = 0,0776$, probabilidad de 7,76%

☐ b. $10,3 \times 0,8 = 8,24$, probabilidad del 8,24%

☐ c. 0,8 %

Ítem 7c. La probabilidad de que una mujer de 40 años tenga una mamografía positiva es el 10.3 %. La probabilidad de que una mujer de 40 años tenga cáncer de pecho y una mamografía positiva es de 0.8%. Una mujer de 40 años se realiza una mamografía y resulta positiva. ¿Cuál es la probabilidad de que realmente tenga cáncer?

☐ d. $\frac{0,8}{10,3} = 0,0776$, probabilidad de 7,76%

☐ e. $10,3 \times 0,8 = 8,24$, probabilidad del 8,24%

☐ f. 0,8 %

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

Contenido 8: Calcular la probabilidad condicional en un experimento simultáneo. Restricción del espacio muestral en la probabilidad condicional

Ítem 8a. Hemos lanzado dos dados y sabemos que el producto de los dos números obtenidos ha sido 12. ¿Cuál es la probabilidad de que ninguno de los dos números sea un 6?

Ítem 8b. Hemos lanzado dos dados y sabemos que el producto de los dos números obtenidos ha sido 12. ¿Cuál es la probabilidad de que alguno de los dos números sea un 6?

Ítem 8c. Hemos lanzado dos dados y sabemos que el producto de los dos números obtenidos ha sido 12. ¿Cuál es la probabilidad de que ninguno de los dos números sea un 6? (El orden de los resultados se tiene en cuenta).

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

Contenido 9: Falacia de la conjunción (no darse cuenta que la probabilidad conjunta de dos sucesos es menor que la probabilidad simple de uno de ellos)

Ítem 9a. Supón que una estrella del tenis alcanza la final de Roland Garrós en 2005. Para ganar el partido hay que ganar tres sets. ¿Cuál de los siguientes sucesos consideras más probable?

- [ ] a. El jugador pierde el primer set.
- [ ] b. El jugador pierde el primer set pero gana el partido.
- [ ] c. Los dos sucesos son iguales de probables.

Ítem 9b. Supón que una estrella del tenis alcanza la final de Roland Garros en 2005. Para ganar el partido hay que ganar tres sets. ¿Cuál de los siguientes sucesos consideras más probable?

- [ ] a. El jugador pierde el primer set.
- [ ] b. El jugador pierde el primer set y luego gana los otros dos, ganando el partido.
- [ ] c. Los dos sucesos son iguales de probables.

Ítem 9c. Supón que una estrella del tenis alcanza la final de Roland Garros en 2005. Para ganar el partido hay que ganar tres sets. ¿Cuál de los siguientes sucesos consideras más probable?

- [ ] a. El jugador pierde el primer set
- [ ] b. El jugador pierde el primer set y luego gana los otros dos
- [ ] c. El jugador pierde los tres sets

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

Contenido 10: Distinguir situación condicional causal y diagnóstica

Ítem 10a. Un test diagnóstico de cáncer fue administrado a todos los residentes de una gran ciudad. Un resultado positivo en el test es indicativo de cáncer y un resultado negativo es indicativo de ausencia de cáncer. ¿En cuál de las siguientes predicciones tienes más confianza?

- [ ] a. Predecir que una persona tiene cáncer si ha dado positivo en el test de diagnóstico.
- [ ] b. Predecir un resultado positivo en el test de diagnóstico si la persona tiene cáncer.
- [ ] c. Tengo la misma confianza en ambas predicciones.

Ítem 10b. Un test diagnóstico de cáncer fue administrado a todos los residentes de una gran ciudad. Un resultado positivo en el test es indicativo de cáncer y un resultado negativo es indicativo de ausencia de cáncer. ¿Cuál de las siguientes predicciones es más fiable?

☐ a. Predecir que una persona tiene cáncer si ha dado positivo en el test de diagnóstico.
☐ b. Predecir un resultado positivo en el test de diagnóstico si la persona tiene cáncer.
☐ c. Las dos predicciones son igual de fiables.

Ítem 10c. Un test diagnóstico de cáncer fue administrado a todos los residentes de una gran ciudad. Un resultado positivo en el test es indicativo de cáncer y un resultado negativo es indicativo de ausencia de cáncer. ¿Qué te parece más probable?

☐ a. Que una persona tenga cáncer si ha dado positivo en el test de diagnóstico.
☐ b. Que un test de diagnóstico de positivo si la persona tiene cáncer.
☐ c. Las dos situaciones tienen la misma probabilidad

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

Contenido 11: Teorema de la probabilidad total

Ítem 11a. En una ciudad hay 60 hombres y 40 mujeres por cada 100 habitantes. 50 de cada 100 hombres y 25 de cada 100 mujeres fuman. Si cogemos 200 personas al azar, ¿Cuántas de estas 200 personas fumarán?

Ítem 11b. En una ciudad el 60 % son hombres y el 40 % mujeres. El 50 % de los hombres y el 25 % de las mujeres fuman. Si cogemos una persona al azar, ¿Cuál es la probabilidad de que sea fumadora?

Ítem 11c. En una ciudad hay 60 hombres y 40 mujeres por cada 100 habitantes. 50 de cada 100 hombres y 25 de cada 100 mujeres fuman. Si cogemos 400 personas al azar, ¿Cuántas de estas 400 personas fumarán?

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

Contenido 12: Resolver problemas en situaciones diacrónicas (experimentos consecutivos)

Ítem 12a. Una persona lanza un dado y anota si saca un número par o impar. Estos son los resultados al lanzarlo 15 veces:

par, impar, impar, par, par, impar, par, par, par, par, impar, impar, par, par, par

Lanza el dado de nuevo. ¿Cuál es la probabilidad de sacar un número par en esta nueva tirada?

Ítem 12b. Una persona lanza un dado y anota si saca un número par o impar. Se trata de un dado no sesgado (es decir, todos los números tienen la misma probabilidad). Estos son los resultados al lanzarlo 15 veces:

par, impar, impar, par, par, impar, par, par, par, par, impar, impar, par, par, par

Lanza el dado de nuevo. ¿Cuál es la probabilidad de sacar un número par en esta nueva tirada?

Ítem 12c. Una persona lanza una moneda y anota el resultado. Se trata de una moneda no sesgada (es decir, todos los números tienen la misma probabilidad). Estos son los resultados al lanzarla 15 veces:

cara, cruz, cruz, cara, cara, cruz, cara, cara, cara, cara, cruz, cruz, cara, cara, cara

Lanza la moneda de nuevo. ¿Cuál es la probabilidad de sacar cara en esta nueva tirada?

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

Contenido 13: Calcular probabilidades condicionales haciendo uso de la regla del producto en caso de experimentos independientes

Ítem 13a. Un grupo de alumnos de un colegio se examina de matemáticas e inglés. La proporción de alumnos que aprueban matemáticas es del 80% y la proporción de alumnos que aprueban inglés es del 70%. Suponiendo que las notas en cada una de las asignaturas es independiente, ¿Cuál es la probabilidad de que un alumno escogido al azar haya aprobado ambas asignaturas?

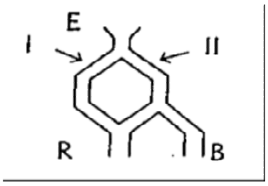
Ítem 13b. La proporción de alumnos que cursan la asignatura Matemáticas en una Facultad es del 80% y la proporción de alumnos que cursan Inglés es del 70%. Los alumnos eligen independientemente las dos asignaturas, ¿Cuál es la probabilidad de que un alumno elegido al azar curse ambas asignaturas?

Ítem 13c. 80 de cada 100 alumnos de una Facultad cursan la asignatura Matemáticas y 70 de cada 100 alumnos cursan Inglés. Los alumnos pueden cursar independientemente una, las dos o ninguna de estas asignaturas. Si tomamos 100 alumnos al azar en la Facultad, ¿Cuántos, aproximadamente, cursaran las dos asignaturas?

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

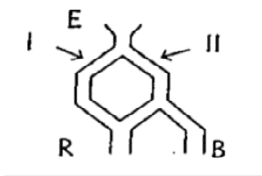
Contenido 14: Resolver problemas de probabilidad condicional en situaciones diacrónicas (experimentos sucesivos**) y falacia del eje de tiempos** (pensar que un suceso no se puede condicionar por otro posterior).

Ítem 14a. Una bola se suelta por la entrada E. Si sale por R, ¿Cuál es la probabilidad de que haya pasado por el canal I?



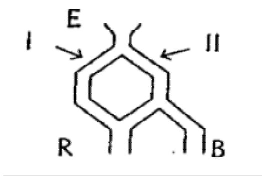
- ☐ a. 0'50
- ☐ b. 0'33
- ☐ c. 0'66
- ☐ d. No se puede calcular

Ítem 14b. Soltamos un montón de bolas por la entrada E. 100 bolas han caído en R. ¿Cuántas de ellas (aproximadamente) han pasado por el canal I?



- ☐ a. 50
- ☐ b. 33
- ☐ c. 66
- ☐ d. No se puede calcular

Ítem 14c. Una bola se suelta por la entrada E. Si sale por R, ¿Cuál es la probabilidad de que haya pasado por el canal I?



- ☐ a. 1/2
- ☐ b. 1/3
- ☐ c. 2/3
- ☐ d. No se puede calcular

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

Contenido 15: Resolver problemas de probabilidad condicional en situaciones sincrónicas (experimentos simultáneos).

Ítem 15a. En una encuesta publicada en un periódico se informa que el 91% de los habitantes de una ciudad mienten usualmente y de ellos el 36% mienten sobre cosas importantes. Si elegimos al azar a una persona de esta ciudad, ¿Cuál es la probabilidad de que mienta sobre cosas importantes?

Ítem 15b. En una encuesta publicada en un periódico se informa que el 91 de cada 100 habitantes de una ciudad mienten usualmente. También se informa que 36 de cada 100 personas que mienten, lo hacen sobre cosas importantes. Si elegimos al azar 10000 personas de esta ciudad, ¿Cuántos, aproximadamente, mentirán sobre cosas importantes?

Ítem 15c. En una encuesta publicada en un periódico se informa que el 91% de los habitantes de una ciudad mienten usualmente. También se informa que el 36% de los mentirosos, lo hacen sobre cosas importantes. Si elegimos al azar a una persona de esta ciudad, ¿Cuál es la probabilidad de que mienta sobre cosas importantes?

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

Contenido 16: Aplicar correctamente el teorema de Bayes

Ítem 16a. Una fábrica dispone de dos máquinas M1 y M2 que fabrican bolas. La máquina M1 fabrica el 40 % de las bolas y la M2 el 60%. El 10% de las bolas fabricadas por M1 y el 5% de las fabricadas por M2 son defectuosas. Tomamos una bola al azar que resulta ser defectuosa. ¿Cuál es la probabilidad de que haya sido fabricada por M1?

Ítem 16b. Una fábrica dispone de dos máquinas M1 y M2 que fabrican bolas. La máquina M1 fabrica 40 de cada 100 bolas y la M2 60 de cada 100. 10 de cada 100 bolas fabricadas por M1 y 5 de cada 100 bolas fabricadas por M2 son defectuosas. Tomamos 1 bola azar que resulta ser defectuosa. ¿Cuál es la probabilidad de que haya sido fabricada por M1?

Ítem 16c. Una fábrica de bolas dispone de dos máquinas M1 y M2. De cada 100 bolas fabricadas, 40 han sido producidas por la máquina M1 y 60 por la M2. 4 de cada 40 bolas fabricadas por M1 y 3 de cada 60 bolas fabricadas por M2 son defectuosas. Tomamos 1000 bolas al azar, de las que 70 resultan ser defectuosas. ¿Cuántas de estas bolas defectuosas habrán sido fabricadas por M1?

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

Contenido 17: Resolver problemas de probabilidad condicional en contexto de muestreo sin reposición. Falacia del eje de tiempos (parte 2)

Ítem 17a. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin devolver la primera a la urna. ¿Cuál es la probabilidad de extraer una bola negra en segundo lugar, habiendo extraído una bola negra en primer lugar? $P(N_2/N_1)$

☐ a. 1/ 2
☐ b. 1/ 6
☐ c. 1/ 3
☐ d. 1/ 4

¿Cuál es la probabilidad de extraer una bola negra en primer lugar, sabiendo que hemos extraído una bola negra en segundo lugar? $P(N_1/N_2)$

☐ a. 1/ 3
☐ b. No se puede calcular
☐ c. 1/ 6
☐ d. 1/ 2

Ítem 17b. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin devolver la primera a la urna. ¿Cuál es la probabilidad de extraer una bola negra en segundo lugar, habiendo extraído una bola negra en primer lugar? $P(N_2/N_1)$

☐ a. 1/ 2
☐ b. 1/ 6
☐ c. 1/ 3
☐ d. 1/ 4

¿Cuál es la probabilidad de extraer una bola negra en primer lugar, sabiendo que hemos extraído una bola negra en segundo lugar? P (N_1/ N_2)

- [ ] a. 1/ 3
- [ ] b. 1/ 4
- [ ] c. 1/ 6
- [ ] d. 1/ 2

Ítem 17c. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin devolver la primera a la urna. ¿Cuál es la probabilidad de extraer una bola negra en segundo lugar, habiendo extraído una bola negra en primer lugar? P (N_2/ N_1)

¿Cuál es la probabilidad de extraer una bola negra en primer lugar, sabiendo que hemos extraído una bola negra en segundo lugar? P (N_1/ N_2)

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

Contenido 18: Usar la regla del producto en el caso de sucesos dependientes

Ítem 18a. Se tiene una bolsa con 1 bola azul y 2 bolas rojas. Se saca una bola al azar. Sin reponerla otra vez en la bolsa se saca una segunda bola. ¿Cuál suceso es más probable?

- [ ] a. Sacar dos bolas rojas
- [ ] b. Sacar primero una bola roja y luego una azul.
- [ ] c. Los dos sucesos son iguales de probables

Ítem 18b. Se tiene una bolsa con 1 bola azul y 2 bolas rojas. Se saca una bola al azar. Sin reponerla otra vez en la bolsa se saca una segunda bola. ¿Cuál de las siguientes afirmaciones es cierta?

- [ ] a. Lo más probable es sacar dos bolas rojas
- [ ] b. Lo más probable es sacar primero una bola roja y luego una azul.
- [ ] c. Es igual de probable sacar azul - roja o roja - roja

Ítem 18c. Se tiene una bolsa con 1 bola azul y 2 bolas rojas. Se saca una bola al azar. Sin reponerla otra vez en la bolsa se saca una segunda bola. ¿Cuál de las siguientes afirmaciones es cierta?
¿Cual es la probabilidad de sacar dos bolas rojas?
¿Cual es la probabilidad de sacar una roja en primer lugar y una azul en segundo?

	Número del ítem
Orden de adecuación	
Primero (el más adecuado)	
Segundo	
Tercero (el menos adecuado)	

ANEXO 6.

CUESTIONARIO DE EVALUACIÓN DEL RAZONAMIENTO CONDICIONAL RPC

Ítem 1. Define probabilidad simple y probabilidad condicional, dando un ejemplo de cada una de ellas.

Ítem 2. Completa el espacio muestral de los siguientes experimentos aleatorios:

a. Observar el género (hombre/mujer) de los hijos de las familias con tres descendientes (ejemplo MHM,...).
b. Observar el género (hombre/mujer) de los hijos de las familias con tres descendientes en los que dos o más son hombres.

Ítem 3. Un taxi se vio implicado en un accidente nocturno con choque y huida posterior. Hay dos compañías de taxis en la ciudad, la Verde y la Azul. El 85% de los taxis de la ciudad son Verdes y el 15% Azules. Hubo un testigo del accidente. El tribunal comprobó la fiabilidad del testigo bajo las mismas circunstancias que había la noche del accidente y llegó a la conclusión de que el testigo identificaba correctamente cada uno de los colores en el 80% de las ocasiones y fallaba en el 20%. Sabiendo que el testigo identificó el taxi como azul, ¿Cuál es la probabilidad de que el taxi implicado en el accidente fuera en efecto Azul?

☐ a. 80/100
☐ b. 15 /100
☐ c. (15/100) x(80/100)
☐ d. $\dfrac{15\times 80}{85\times 20+15\times 80}$

Ítem 4. Se extrae una carta al azar de una baraja española (40 cartas con cuatro palos: oros, bastos, espadas y copas. Cada palo tiene los números del 1 al 7, sota caballo y rey). Sea A el suceso "se extrae una carta de oros" y B el suceso "se extrae un rey". ¿Los sucesos A y B son independientes?

☐ a. No son independientes, porque en la baraja hay un rey de oros.
☐ b. Sólo si sacamos primero una carta para ver si es rey y se vuelve a colocar
c. en la baraja y luego sacamos una segunda carta para mirar si es oros.
☐ d. Si, porque P(rey de oros)= P(rey) x P(oros),
☐ e. No, porque P(rey / oros) $\neq$ P(rey).

Ítem 5. Una caja tiene cuatro focos, de los cuales dos son defectuosos. Se sacan dos al azar, uno tras otro sin reemplazamiento. Si el primer foco fue defectuoso, entonces:

☐ a. Es mas probable que el segundo sea defectuoso;
☐ b. Es más probable que el segundo no sea defectuoso;
☐ c. La probabilidad de que el segundo sea defectuoso es igual a la probabilidad de que no lo sea.

Ítem 6..En una población se ha realizado una entrevista a un grupo de hombres, obteniendo los siguientes resultados:

	menos de 55 años	más de 55 años	Total
Ha sufrido un ataque al corazón	29	75	104
No ha sufrido ataque	401	275	676
Total	430	350	780

Si elegimos al azar una de estas personas:

a. ¿Cuál es la probabilidad de que haya tenido un ataque al corazón?
b. ¿Cuál es la probabilidad de que tenga más de 55 años y además haya tenido un ataque al corazón?
c. Sabiendo que la persona escogida tiene más de 55 años, ¿Cuál es la probabilidad de que haya tenido un ataque al corazón?
d. Sabiendo que la persona escogida ha tenido un ataque al corazón, ¿Cuál es la probabilidad de que tenga más de 55 años?

Ítem 7. La probabilidad de que una mujer tenga una mamografía positiva es el 10.3%. La probabilidad de que una mujer tenga cáncer de pecho y una mamografía positiva es de 0.8%. Una mujer se realiza una mamografía y resulta positiva. ¿Cuál es la probabilidad de que realmente tenga cáncer?

☐ **a.** $\frac{0,8}{10,3} = 0,0776$, probabilidad de 7,76%

☐ **b.** $10,3 \times 0,8 = 8,24$, probabilidad del 8,4%

☐ **c.** 0,8 %

Ítem 8. Hemos lanzado dos dados y sabemos que el producto de los dos números obtenidos ha sido 12 ¿Cual es la probabilidad de que ninguno de los dos números sea un 6? (Se diferencia si un número ha aparecido en un dado o en otro).

Ítem 9. Supón que una estrella del tenis alcanza la final de Roland Garros en 2005. Para ganar el partido hay que ganar tres sets de cinco. ¿Cuál de los siguientes sucesos consideras más probable?

☐ a. El jugador pierde el primer set.
☐ b. El jugador pierde el primer set pero gana el partido.
☐ c. Los dos sucesos son iguales de probables.

Ítem 10. Un test diagnóstico de cáncer fue administrado a todos los residentes de una gran ciudad en la que hay pocos casos de cáncer. Un resultado positivo en el test es indicativo de cáncer y un resultado negativo es indicativo de ausencia de cáncer. ¿Qué te parece más probable?

☐ a. Que una persona tenga cáncer si ha dado positivo en el test de diagnóstico.
☐ b. Que un test de diagnóstico resulte positivo si la persona tiene cáncer.
☐ c. Los dos sucesos tienen la misma probabilidad.

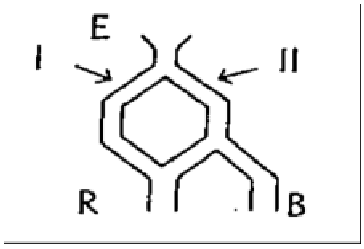
Ítem 11. En una ciudad el 60 % son hombres y el 40 % mujeres. El 50 % de los hombres y el 25 % de las mujeres fuman. Si se escoge una persona al azar, ¿Cuál es la probabilidad de que sea fumadora?

Ítem 12. Una persona lanza un dado y anota si saca un número par o impar. Se trata de un dado no sesgado (es decir todos los números tienen la misma probabilidad). Estos son los resultados al lanzarlo 15 veces: par, impar, impar, par, par, impar, par, par, par, par, impar, impar, par, par,

par. Lanza el dado de nuevo ¿Cuál es la probabilidad de sacar un número par en esta nueva tirada?

Ítem 13. Un grupo de alumnos de un colegio se examina de matemáticas e inglés. La proporción de alumnos que aprueban matemáticas es del 80% y la proporción de alumnos que aprueba inglés es del 70%. Suponiendo que las notas en cada una de las asignaturas son independientes, ¿Cuál es la probabilidad de que un alumno escogido al azar haya aprobado ambas asignaturas?

Ítem 14. Una bola se suelta por la entrada E. Si sale por R, ¿Cuál es la probabilidad de que haya pasado por el canal 1?



- ☐ a. 1/2
- ☐ b. 1/3
- ☐ c. 2/3
- ☐ d. No se puede calcular

Ítem 15. En una encuesta publicada en un periódico se informa que el 91% de los habitantes de una ciudad mienten usualmente y, de ellos, el 36% mienten sobre cosas importantes. Si elegimos al azar a una persona de esta ciudad, ¿Cuál es la probabilidad de que mienta sobre cosas importantes?

Ítem 16. Una fábrica dispone de dos máquinas M1 y M2 que fabrican bolas. La máquina M1 fabrica el 40 % de las bolas y la M2 el 60%. El 5% de las bolas fabricadas por M1 y el 1% de las fabricadas por M2 son defectuosas. Tomamos una bola al azar que resulta ser defectuosa. ¿Cuál es la probabilidad de que haya sido fabricada por M1?

Ítem 17. Una urna contiene dos bolas blancas y dos bolas negras. Extraemos a ciegas dos bolas de la urna, una detrás de otra, sin devolver la primera a la urna. ¿Cuál es la probabilidad de extraer una bola negra en segundo lugar, habiendo extraído una bola negra en primer lugar? $P(N_2/ N_1)$

- ☐ a. 1/ 2
- ☐ b. 1/ 6
- ☐ c. 1/ 3
- ☐ d. 1/ 4

¿Cuál es la probabilidad de haber extraído una bola negra en primer lugar, sabiendo que hemos extraído una bola negra en segundo lugar? P (N_1/ N_2)

- ☐ a. 1/ 3
- ☐ b. No se puede calcular
- ☐ c. 1/ 6
- ☐ d. 1/ 2

Ítem 18 Se tiene una bolsa con 1 bola azul y 2 bolas rojas. Se saca una bola al azar. Sin reponerla otra vez en la bolsa se saca una segunda bola. ¿Qué suceso es más probable?

- ☐ a. Sacar dos bolas rojas.
- ☐ b. Sacar primero una bola roja y luego una azul.
- ☐ c. Los dos sucesos son iguales de probables.

ANEXO 7.

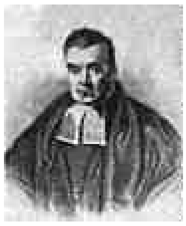
MATERIAL DIDÁCTICO ENTREGADO A LOS ALUMNOS

Tema 1. Conceptos Básicos

Objetivos del tema 1

En este tema aprenderás a:

- Diferenciar entre probabilidades iniciales y finales y verosimilitudes
- Identificar los sucesos de interés, sus probabilidades iniciales y verosimilitudes a partir del enunciado de un problema
- Analizar el teorema de Bayes como herramienta para transformar probabilidades iniciales en finales
- Organizar los datos en una tabla de forma que se facilite el cálculo, bien manual o con recursos informáticos (hoja Excel, applet) para calcular las probabilidades finales a partir de las probabilidades iniciales y de las verosimilitudes.



Thomas Bayes

1.1 Probabilidad inicial, final y verosimilitud

El primer concepto importante en inferencia bayesiana es el de probabilidad inicial, final y verosimilitud. En lo que sigue, analizamos una situación común en que es importante diferenciar entre probabilidades iniciales, finales y verosimilitudes. Son las *pruebas médicas*, donde tenemos:

- *Probabilidad inicial* de sufrir una cierta enfermedad (es decir, cuando no tenemos ningún otro dato del enfermo). En este caso, la mejor predicción de la probabilidad de sufrir la enfermedad *P(E)* viene dada por la proporción de sujetos que la padece en la población (prevalencia de la enfermedad).
- *Probabilidad final* de tener la enfermedad, una vez que se le ha hecho una prueba al enfermo y ha sido positiva; es decir, a la luz de un nuevo dato *P(E/+)*.
- *Verosimilitud* de que la prueba sea positiva si el sujeto está enfermo *P(+/Enfermo)* (sensibilidad del test).

El Teorema de Bayes nos permite pasar de la probabilidad inicial *P(E)* a la probabilidad final *P(E/+)*, cuando se conocen las verosimilitudes *P(+/Enfermo)*.

1.2. Teorema de Bayes

En el año 1763, dos años después de la muerte de *Thomas Bayes* (1702-1761), se publicó una memoria en la que aparece, por vez primera, la determinación de la probabilidad de las causas a partir de los efectos que han podido ser observados. El cálculo de dichas probabilidades recibe el nombre de teorema de Bayes.

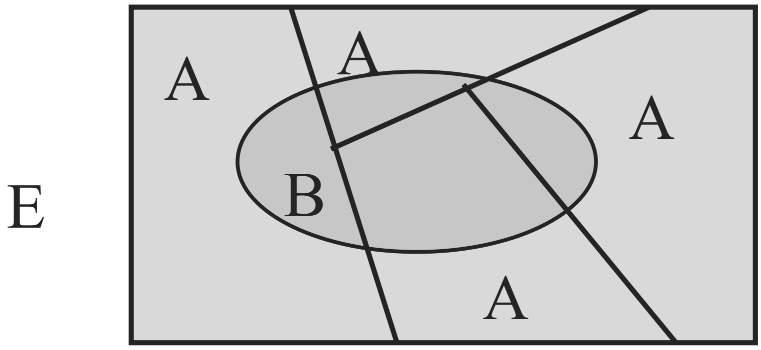
El Teorema de Bayes es uno de los más importantes teoremas en estadística, porque nos permite *aprender de la experiencia*. Más concretamente, la regla de Bayes nos permite calcular cómo se modifican las probabilidades de determinados sucesos, cuando se conoce alguna información adicional.

Este teorema es de gran utilidad para un psicólogo, que tiene que diagnosticar diferentes trastornos a partir de pruebas y también en general, para cualquier persona. Nosotros podemos pensar que la probabilidad de que tengamos un trastorno, inicial es *P(T),* pero una vez que la prueba ha salido positiva, esa probabilidad cambia *P(T/+)*.

El teorema de Bayes nos permite calcular esta última, conocida la primera, pero también puede aplicarse en muchas otras ocasiones. Veámoslo de una forma más general.

Teorema de Bayes:

Consideremos un experimento aleatorio y supongamos que su espacio muestral asociado es *E*. Sean los sucesos A_1, A_2, A_n una partición de *E*, cuyas probabilidades se conocen. Sea *B* un suceso cualquiera del espacio muestral, del que conocemos las probabilidades *P(B/Ai)*. El siguiente esquema representa esta situación.



El teorema de Bayes permite calcular las probabilidades $P(A_i/B)$, mediante la siguiente fórmula:

$$(1.1) \qquad P(A_i/B)=\frac{P(A_i)\times P(B/A_i)}{P(A_1)\times P(B/A_1)+P(A_2)\times P(B/A_2)+...+P(A_n)\times P(B/A_n)}$$

Las probabilidades $P(A_i)$ se denominan usualmente *probabilidades iniciales*, y $P(A_i/B)$ *probabilidades finales*, que se calcularán conocida la información $P(B/A_i)$, también llamadas *verosimilitudes*.

El denominador de esta fórmula si lo recordáis es el teorema de la probabilidad total P(B). Es constante, por lo que si definimos K de la forma siguiente:

$$(1.2) \qquad K=\frac{1}{P(A_1)\times P(B/A_1)+P(A_2)\times P(B/A_2)+...+P(A_n)\times P(B/A_n)}$$

La fórmula de Bayes quedaría así: (1.3) $P(A_i/B)=K\times P(A_i)\times P(B/A_i)$

Recuerda: El Teorema de Bayes (fórmula 1.3) nos indica que la Probabilidad final de A_i dado B es proporcional al producto de la probabilidad inicial de A_i por la verosimilitud de los datos dado A_i. La constante de proporcionalidad K es la inversa de la probabilidad de B que se calcula en (1.2).

1.3. Organización de los cálculos

El teorema de Bayes puede también generalizarse a mayor número de sucesos, como en el ejemplo que mostramos a continuación:

Tres máquinas denominadas A, B y C, producen un 43%, 26% y 31% de la producción total de una empresa respectivamente. Se ha detectado que un 8%, 2% y 1,6% del producto manufacturado por estas máquinas es defectuoso. Se selecciona un producto al azar y se encuentra que es defectuoso, ¿cuál es la probabilidad de que el producto haya sido fabricado en la máquina B?

La probabilidad pedida se calcula usando el Teorema de Bayes:

$$P(B/D) = \frac{P(B \cap D)}{P(D)} = \frac{P(B) \times P(D/B)}{P(A) \times P(D/A) + P(B) \times P(D/B) + P(C) \times P(D/C)} =$$

$$P(B/D) = \frac{0,26 \times 0,02}{(0,43 \times 0,08) + (0,26 \times 0,02) + (0,31 \times 0,016)} = \frac{0,0052}{0,04456} = 0,116697$$

Una forma sencilla de organizar los cálculos para pasar de la probabilidades iniciales *P(A), P(B), P(C)* a las probabilidades finales *P(A/D), P(B/D), P(C/D),* es utilizando una tabla similar a la 1.1.

Tabla 1.1. Organización de cálculos de probabilidad final

(1) Sucesos de interés	(2) Probabilidad inicial	(3) Verosimilitud	(4) Producto	(5) Probabilidad final
A	0,43	0,08	0,0344	0,7720
B	0,26	0,02	0,0052	0,1167
C	0,31	0,016	0,00496	0,1113
Suma			0,04456	1

- En la columna (1) ponemos los sucesos de interés, en este caso las máquinas A, B y C.
- En la columna (2) ponemos las probabilidades inicial y en la (3) las verosimilitudes, que son los datos del problema.
- Calculamos ahora los numeradores de la fórmula de Bayes (producto (2)x(3) en la columna (4).
- Sumamos la columna (4) para obtener el denominador de la fórmula de Bayes.
- En la columna (5) obtenemos las probabilidades final, dividiendo cada celda en la columna (4) por la suma anterior.

Ejemplo 1. 1. Pruebas clínicas: Narcolepsia

La **narcolepsia** es un trastorno primario del sueño, cuya sintomatología principal es la aparición recurrente e irresistible de ataques de sueño reparador. Las personas que tienen este trastorno no descansan bien; por tanto se vuelven irritables y no rinden lo que quisieran.

En una gran ciudad una de cada 1000 personas sufre narcolepsia. Seleccionamos al azar una persona en una gran ciudad ¿Cuál es la probabilidad de que tenga narcolepsia?

- Probabilidad inicial: probabilidad inicial de sufrir narcolepsia en una persona tomada al azar de la población *P(N)= 1/1000*

Supón que el test es positivo en 99 de cada 100 personas enfermas y también en 2 de cada 100 *personas sanas.*

- *Verosimilitud:* probabilidad de que el test sea positivo si se tiene narcolepsia (sensibilidad del test) *P(+/N)= 99/100*
- Probabilidad *final:* probabilidad final de sufrir narcolepsia si el test ha dado positivo *P(N/+).* Para calcular esto, hay que aplicar el teorema de Bayes, que vemos a continuación

Vamos a completar en la tabla siguiente los datos que conocemos de la situación:

	Test +	Test -	Total
Sufren narcolepsia	99	1	100
No sufren narcolepsia	1998	97902	99.900
Total	2097	97903	100.000

Para calcular la probabilidad final, habrá que calcular:

$$P(N/+) = \frac{99}{99+1998} = 0{,}0472$$

El resultado no deja de sorprendernos a primera vista: ¡Puede parecer que las pruebas médicas son poco fiables! Lo que pasa es que el número de personas que tienen la enfermedad, en el total de la población es muy pequeño.

- La probabilidad de que el test de positivo si la persona está enferma es muy alta. Casi todas ellas son detectadas en el test (el test tiene mucha *sensibilidad).*
- Por otro lado, la probabilidad de un resultado positivo si se está sano (*falso positivo)* es muy pequeña.
- Pero un suceso con probabilidad pequeña no es un suceso *imposible* y puede ocurrir. Más aún, si el número de personas que pasan la prueba es muy grande, pueden aparecer más falsos positivos que positivos reales (como en el ejemplo).

Ejemplo 1. 2. Pruebas clínicas. Insomnio

Las pruebas médicas son bastante fiables, pero el resultado anterior se explica porque el número de personas sanas a las que la prueba da positiva es mucho mayor (aunque sólo una pequeña proporción de sanos tiene un test positivo) que el número de personas enfermas a las que la prueba da positiva (aunque casi todos los enfermos tienen el test positivo).

¡Pero es que el número de enfermos en la población es muy pequeño!

¿Qué ocurriría si, en lugar de narcolepsia, el test da positivo en insomnio?

La prevalencia del insomnio en la población es bastante mayor que la de la narcolepsia (un 15%), es decir, la probabilidad inicial de insomnio es P (I) = 15/100.

La probabilidad de que el test de positivo en una persona con insomnio es de 0,99 y la probabilidad de que de positivo en una persona sana es de 0,02.

Calculemos ahora la probabilidad final de sufrir insomnio si la prueba es positiva

$$P(I/+) = \frac{P(I \cap +)}{P(+)} = 0{,}8973$$

Vemos que el resultado de la prueba de insomnio es mucho más fiable, porque la prevalencia de insomnio en la población es mucho mayor que la narcolepsia.

Ejercicio resuelto 1.1. Seguros de accidente

Enunciado

El 22% de los clientes de una compañía de seguros de accidentes de automóvil es menor de 30 años. Las estadísticas indican que el 11% de los conductores menores de 30 años tiene un accidente y que sólo el 5% de los mayores o iguales de 30 años tienen un accidente. ¿Qué porcentaje de accidentes pagados por la compañía son de clientes menores de 30 años?

Datos del ejercicio

- P(menor 30 años)=0,22; es una probabilidad inicial.
- P (mayor de 30 años)=0,78; es una probabilidad inicial.
- P(accidente /menor 30 años)=0,11; es una verosimilitud.
- P(accidente /mayor o igual de 30 años)=0,05; es una verosimilitud.

¿Qué pide el ejercicio?

- P(menor 30 años / accidente) es una probabilidad final.

Solución

Puesto que tenemos que pasar de probabilidades iniciales a finales, hay que usar el teorema de Bayes. Una forma sencilla es usar la tabla para organizar los datos y cálculos. En primer lugar colocamos en ella los datos conocidos (sucesos de interés, sus probabilidades iniciales y verosimilitudes

(1) Sucesos de interés	(2) Probabilidad inicial	(3) Verosimilitud de accidente	(4) Producto	(5) Probabilidad final
Menor 30 años	0,22	0,11		
Mayor 30 años	0,78	0,05		
Suma				

A continuación realizamos los productos de las columnas (2) y (3), obtenemos la suma de los productos (columna 4)

(1) Sucesos de interés	(2) Probabilidad inicial	(3) Verosimilitud de accidente	(4) Producto	(5) Probabilidad final
Menor 30 años	0,22	0,11	0,0242	
Mayor 30 años	0,78	0,05	0,039	
Suma			0,0632	

Finalmente dividimos cada uno de los productos (columna 4) por la suma de todos los productos (0,0632) para obtener las probabilidades finales. Si no te equivocas en los cálculos la suma de las probabilidades finales es igual a 1

(1) Sucesos de interés	(2) Probabilidad inicial	(3) Verosimilitud de accidente	(4) Producto	(5) Probabilidad final
Menor 30 años	0,22	0,11	0,0242	0,3829
Mayor 30 años	0,88	0,05	0,044	0,6171
Suma			0,0682	1

El 38% de los accidentes que tiene que pagar la compañía son de conductores de menos de 30 años.

Ejercicio resuelto 1.2. Depresión post-parto

Enunciado

Alrededor del 15% de las madres primerizas sienten ansiedad en los días posteriores al parto. En el 90% de los casos esta ansiedad disminuye y en pocos días desaparece, pero en los casos restantes puede aparecer una depresión que necesite tratamiento. Un 2% de madres primeriza que no sufrieron ansiedad en los primeros días desarrolla una depresión tras las primeras semanas del nacimiento. Si una madre primeriza tuvo una depresión post-parto, ¿Qué porcentaje de depresiones post- parto no fueron previstas los primeros días después del parto?

Datos del ejercicio

- P(ansiedad en las primeras semanas)=0,15; es una probabilidad inicial.
- P (no sufrir ansiedad en las primeras semanas)=0,85; es una probabilidad inicial.
- P(depresión /si hubo ansiedad inicial)=0,10; es una verosimilitud.
- P(depresión /si no hubo ansiedad inicial)=0,02; es una verosimilitud.

¿Qué pide el ejercicio?

- P(ansiedad inicial/ depresión) es una probabilidad final.

Solución

Puesto que tenemos que pasar de probabilidades iniciales a finales, hay que usar el teorema de Bayes. Usaremos la tabla, como hemos hecho en el ejercicio 1.

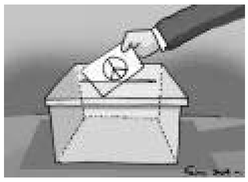
(1) Sucesos de interés	(2) Probabilidad inicial	(3) Verosimilitud de depresión	(4) Producto	(5) Probabilidad final
Ansiedad inicial	0,15	0,1	0,015	0,4688
No ansiedad	0,85	0,02	0,017	0,5313
Suma			0,032	1

Casi la mitad de las depresiones post-parto no se previeron en las primeras semanas.

Tema 2.Inferencia para una proporción. Caso discreto

Objetivos del tema 2

En este tema aprenderás a:



Votaciones

- Visualizar los parámetros como variables aleatorias con una distribución de probabilidad asociada
- Identificar valores de la proporción poblacional y asignarles probabilidades iniciales plausibles (distribución de probabilidad inicial)
- Comprender el concepto de distribución inicial uniforme y distribución informativa
- Identificar los valores de éxitos y fracasos en la muestra
- Aprender a introducir los datos anteriores en la hoja Excel para obtener la distribución final
- Obtener la mejor estimación de la proporción en la población
- Calcular probabilidades asociadas a determinados valores de proporción en la distribución final
- Entender cómo las distribuciones finales se transforman en iniciales en experimentos sucesivos
- Aplicar los conceptos y procedimientos anteriores al contraste de hipótesis sencillas sobre una proporción siguiendo la metodología bayesiana
- Analizar las diferencias entre el contraste de hipótesis clásico y bayesiano.

2.1. El parámetro como variable aleatoria.

En el sistema democrático los ciudadanos eligen individualmente al candidato favorito. El resultado de unas elecciones es siempre aleatorio, ya que el número final de personas que acuden a votar es desconocido. Por diversas razones se producen abstenciones y, además, algunas personas cambian su voto a última hora.

En unas votaciones dos candidatos A y B tienen que enfrentarse. ¿Qué candidato piensas que ganará?

Para resolver el problema, se precisa estimar la proporción de votantes que votará a cada candidato, lo que requiere discriminar entre los conceptos de población y muestra, aplicándolos al contexto del problema:

- *Población*: es el conjunto de todas las posibles observaciones que nos gustaría estudiar. En este caso el conjunto de todos los votantes que votarán a A o a B.
- *Muestra*: Es la parte de la población que podemos estudiar. El estudio completo de la población es, con frecuencia, costoso, lento o imposible. Es difícil que conozcamos la población hasta que ocurran las elecciones, pero podríamos hacer un sondeo con una muestra para intentar inferir algo acerca de la población.

La población queda caracterizada por un *parámetro*, que es la “proporción p de casos que cumplen A”, que desconocemos. Como hemos dicho, tomaremos una muestra para *estimar* dicha proporción teórica en la población, es decir, averiguar cuál de los dos candidatos es preferido.

En la **inferencia clásica**, suponemos que *p* es un valor constante. Para estimar su valor, calculamos la proporción en la muestra y a partir de él calculamos un intervalo de confianza para los valores de *p*.

En **inferencia bayesiana** admitimos las *probabilidades subjetivas*, es decir, la asignación de probabilidades puede ser diferente según los conocimientos de la persona que la asigna.

La proporción de votantes favorables a *A* puede variar de una ciudad a otra o incluso de un día a otro, dependiendo de muchos factores, como el precio del petróleo, noticias internacionales, terrorismo, etc. Por tanto, sería razonable considerar que el parámetro *p* es una *variable aleatoria*. Sería también razonable admitir que diferentes personas pueden asignar diferentes valores iniciales a *p*, en función de sus conocimientos.

Una distinción importante en inferencia bayesiana, es, por tanto:

- *Distribución de probabilidad inicial*: distribución asumida por la persona antes de llevar a cabo el experimento (o recoger los datos). Una vez observados los datos, la persona puede cambiar su opinión sobre el valor de la proporción.
- *Distribución de probabilidad final*: distribución de probabilidad que resulta cuando actualizamos la información usando los datos recogidos. Recoge todo el conocimiento de la persona, una vez analizados los datos y sirve para hacer inferencias.

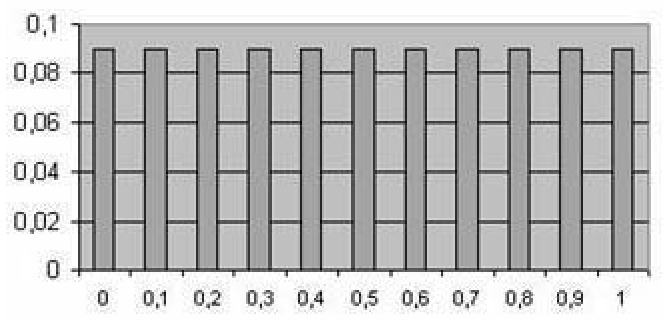
A continuación analizamos los pasos en el proceso bayesiano de estimación. Esto es lo que hacemos en *inferencia bayesiana* donde se siguen los pasos siguientes

- Especificar un conjunto de valores posibles para el parámetro, asignar a cada valor su probabilidad; es decir, establecer una *distribución inicial para la proporción p* en la población.
- Recoger datos y calcular la *verosimilitud* de cada posible valor de *p* a partir de ellos.
- Usar el teorema de Bayes para calcular *distribución final de p* a partir de la distribución a priori y las verosimilitudes.
- Extraer inferencias con la distribución a posteriori.

2.2. Distribuciones iniciales en ausencia de información

Supongamos que no sabemos nada del valor *p*. Una distribución inicial plausible para reflejar esta opinión en las siguiente: asignar una probabilidad igual a *1/11* a cada uno de los valores de *p =0, 0,1, 0,2, ..., 0,9, 1.*

La forma de esta distribución sería así:



En el eje horizontal se muestran los valores de proporción posibles y en el eje vertical se muestra la probabilidad. Como se ve, en una distribución inicial no informativa, todos los valores de proporción tienen la misma probabilidad de ocurrencia, es por tanto una distribución uniforme.

Si en una muestra de 10 votantes obtenemos 7 favorables al candidato *A*, podremos revisar nuestra opinión sobre la probabilidad de que la proporción poblacional sea igual a *p= 0,5,* aplicando la regla de Bayes:

$$P(p = 0{,}5 \,/\, datos) = K \text{ x } P(p = 0{,}5) \text{ x } P(datos / p = 0{,}5),$$

donde:

- *P(p=0,5)* es nuestra probabilidad a priori de *1/11*
- *P(datos / p = 0,5)* es la verosimilitud de obtener nuestros datos para el valor obtenido de la proporción muestral $\hat{p}$ *=0,7* (7 votos favorables de 10) si la proporción real de votantes partidarios A fuese igual a *0,5*. Aplicando la fórmula binomial esta probabilidad es:

$$\binom{10}{7} 0{,}5^7 (1 - 0{,}5)^{10-7} = 0{,}1172$$

- *K* es la inversa del denominador de la fórmula de Bayes, es decir, de la suma de los productos de las verosimilitudes por las probabilidades a priori

Aplicando el teorema de Bayes, podríamos obtener la probabilidad a posteriori de que la proporción en la población sea igual a 0,5:

$$\text{P(p = 0,5 / datos) = K x P(p = 0,5 / datos) = } K \times \frac{1}{10} x \binom{10}{7} 0{,}5^7 (1 - 0{,}5)^{10-7}$$

En general, la probabilidad a posteriori de que la proporción poblacional *p* sea exactamente igual a p_0 es:

$$\text{P (p = / datos) = K x } \frac{1}{10} x \binom{10}{7} p_0^{\,7} (1 - p_0)^{10-7}$$

Es decir, la probabilidad final es igual a la inicial multiplicada por la verosimilitud y por K (una constante que es el denominador de la fórmula de Bayes).

Si queremos calcular las nuevas probabilidades finales para el resto de los valores de la proporción poblacional *p = 0, 0,1,... 1,* es útil organizar los datos en una tabla, como la tabla 2.1.

Las columnas (1) y (2) son los posibles valores de la proporción poblacional y sus probabilidades iniciales, definidas por el investigador.

Tabla 2.1. Transformación de distribución inicial en distribución final

(1) Posibles valores de p	(2) Probabilidad inicial	(3) Verosimilitud	(4) Producto	(5) Probabilidad final
0,00	0,09			
0,10	0,09			
0,20	0,09			
0,30	0,09			
0,40	0,09			
0,50	0,09			
0,60	0,09			
0,70	0,09			
0,80	0,09			
0,90	0,09			
1,00	0,09			
Suma				

La columna (3) son las verosimilitudes de los datos, para cada valor posible de p= p0 y se calculan aplicando la fórmula:

$$P(datos / p = p_0) = \binom{e+f}{e} p_0^{\,e} (1 - p_0)^f$$

Siendo *p* el valor de la proporción, y *e* y *f* el número de éxitos y fracasos en los datos*.

Finalmente, las probabilidades finales (columna 5) se calculan con la fórmula de Bayes:

$$P(p = p_0 / datos) = \frac{\mathrm{P(datos / p = p_0) \times P(p = p_0)}}{\sum \mathrm{P(datos / p) \times P(p)}}$$

y llamando K al inverso del denominador de la fórmula de Bayes, quedaría así:

$$P(p = p_0 / datos) = k \times \mathrm{P(datos / p = p_0) \times P(p = p_0)}$$

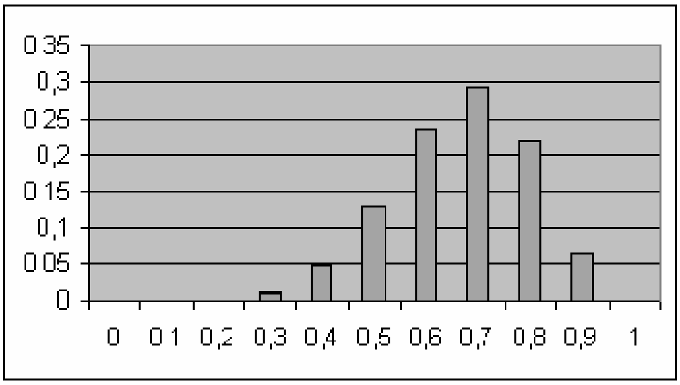
Es decir, multiplicamos la probabilidad inicial (columna 2) por la verosimilitud (columna 3) y por k (es como dividir por la suma de los productos, columna 4).

* No te preocupes si no entiendes cómo se calcula la verosimilitud. Hemos diseñado un programa de Excel que hace la función de "calculadora". Está disponible en la siguiente página web (sección descargas): http://www.ugr.es/local/mcdiaz/bayes

La tabla quedaría de la siguiente forma:

(1) Posibles valores de p	(2) Probabilidad inicial	(3) Verosimilitud	(4) Producto	(5) Probabilidad final
0,00	0,09	0,0000	0,0000	0,0000
0,10	0,09	0,0000	0,0000	0,0000
0,20	0,09	0,0008	0,0001	0,0009
0,30	0,09	0,0090	0,0008	0,0099
0,40	0,09	0,0425	0,0039	0,0467
0,50	0,09	0,1172	0,0107	0,1288
0,60	0,09	0,2150	0,0195	0,2363
0,70	0,09	0,2668	0,0243	0,2932
0,80	0,09	0,2013	0,0183	0,2212
0,90	0,09	0,0574	0,0052	0,0631
1,00	0,09	0,0000	0,0000	0,0000
Suma			0,0827	

La gráfica de la distribución final que resulta es la siguiente. Compárala con la que teníamos en un principio.



2.3. Obtención de inferencias a partir de la distribución final

El objetivo del cálculo de la distribución final es usarla para realizar inferencias sobre los valores del parámetro, es decir, para predecir la proporción de votantes que en la población votará al candidato *A*.

Mejor estimador de una proporción

A partir de la distribución final podemos estimar la proporción poblacional desconocida. Nuestra apuesta más razonable sería el valor de algún promedio de la distribución final, como la moda, que es el valor más probable.

En el ejemplo, la proporción más probable de votantes en la población es *0'7*, que tiene una probabilidad aproximada de *0,29.*

Contraste bayesiano

Podemos también usar las distribuciones finales para *contrastar una hipótesis* sobre un posible valor de la proporción en la población. Por ejemplo supongamos que quiero contrastar la hipótesis de que el candidato *A* no gana las elecciones, es decir $H: p \leq 0,5$

De la distribución final puedo ver que la probabilidad $P(p \leq 0,5/datos)=0,1863$ es pequeña, pero no lo suficiente para que la consideramos como una buena evidencia de que, de hecho, el candidato *A* ganará, porque es un suceso que puede ocurrir 18 de cada 100 veces (por tanto no es muy raro). Nuestra conclusión es que no podemos rechazar la hipótesis, aunque tampoco podemos aceptarla, porque la probabilidad de perder o empatar tampoco es demasiado grande.

De nuevo, los resultados clásico y bayesiano coinciden cuando usamos distribuciones iniciales no informativas. En la inferencia clásica el valor α sería también igual a 0,1863 en este ejemplo, por lo tanto, no rechazaríamos la hipótesis. Pero la interpretación es diferente:

- El valor α nos da la probabilidad de encontrar los datos, si fuese cierta la hipótesis nula, esto es, en este ejemplo $P\ (Datos/\ p \leq 0,5)$.
- En inferencia bayesiana calculamos a partir de la distribución final la probabilidad de que la hipótesis sea cierta, a partir de los datos, $P\ (p \leq 0,5/\ datos)$.

Contrastar una hipótesis significa (desde el punto de vista bayesiano) encontrar su probabilidad final.

2.4. Aplicación sucesiva del Teorema de Bayes en nuevos experimentos

La distribución de probabilidad final representa nuestra nueva creencia sobre la proporción.

Si hacemos un nuevo experimento y obtenemos en una nueva muestra de *e'* éxitos y *f'* fracasos, podemos tomar como nueva distribución inicial, las distribución final obtenida en el primer experimento, porque no sería ahora razonable suponer que todos los valores de la proporción son equiprobables. Aplicando de nuevo el teorema de Bayes obtenemos una nueva distribución final.

Es decir, al comenzar una serie de experimentos no tenemos ninguna información y usamos una distribución no informativa, asignando igual probabilidad a todos los valores.

Después de obtener *e* resultados favorables a *A*, nuestra creencia cambia y la distribución final obtenida en el paso 1 es la que refleja mejor nuestra información. Luego la usaremos como distribución inicial en un segundo experimento.

En futuros experimentos los nuevos datos nos servirán para mejorar todavía nuestro conocimiento.

En general, en inferencia Bayesiana podemos usar cualquier distribución inicial que refleje nuestra experiencia, ya que sería absurdo suponer que no sabemos nada sobre la proporción si tenemos datos anteriores.

Es decir, aplicamos el principio de *aprender de la experiencia* y de asignar probabilidades inicial, *teniendo en cuenta la información disponible*.

Ejemplo 2.1. Revisión de las inferencias en el ejemplo de las votaciones.

Si en el ejemplo anterior de las votaciones hacemos un nuevo experimento y obtenemos en una nueva muestra de 10 casos 6 éxitos y 4 fracasos, podemos tomar como nueva distribución inicial, las distribución final obtenida en el primer experimento (Tabla 2.3), porque no sería ahora razonable suponer que todos los valores de la proporción son equiprobables. Aplicando de nuevo el método, obtenemos la tabla 2.4.

En la figura 2.1 vemos como ha cambiado nuestra creencia sobre la proporción de votantes favorables a *A*.

Al comenzar no teníamos ninguna información y usamos una distribución no informativa, asignando igual probabilidad a todos los valores.

Después de obtener 7 resultados favorables a *A*, nuestra creencia cambia y ahora los valores próximos a 0 o a 1 son muy poco probables.

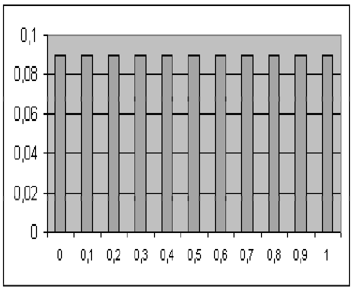
Un nuevo experimento (6 resultados favorables de 10) nos restringe aún más las posibilidades. En futuros experimentos los nuevos datos nos servirán para mejorar todavía nuestro conocimiento sobre las posibilidades del candidato *A*.

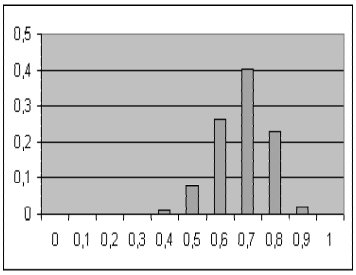
Tabla 2.2. Transformación de distribución inicial en distribución final con 6 nuevos éxitos y 4 nuevos fracasos (segundo experimento)

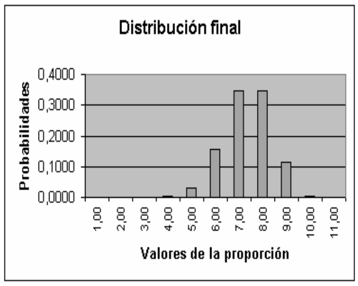
(1) Posibles valores de p	(2) Probabilidad inicial	(3) Verosimilitud	(4) Producto	(5) Probabilidad final
0,00	0,0000	0,0000	0,0000	0,0000
0,10	0,0000	0,0001	0,0000	0,0000
0,20	0,0009	0,0055	0,0000	0,0000
0,30	0,0099	0,0368	0,0004	0,0021
0,40	0,0467	0,1115	0,0052	0,0306
0,50	0,1288	0,2051	0,0264	0,1553
0,60	0,2363	0,2508	0,0593	**0,3484**
0,70	0,2932	0,2001	0,0587	0,3449
0,80	0,2212	0,0881	0,0195	0,1145
0,90	0,0631	0,0112	0,0007	0,0041
1,00	0,0000	0,0000	0,0000	0,0000
Suma			0,1701	

Figura 2.1. Revisión de nuestra creencia sobre la distribución de la proporción

Al comienzo de la experiencia	Después del primer experimento	Después del segundo experimento







Podemos ahora usar la nueva distribución final para hacer inferencias, como:

- La mejor estimación de la proporción de votantes a *A* es ahora *0,6*, que tiene una probabilidad aproximada de *0,348.*

La probabilidad de que la proporción de votantes favorables a A tome un valor comprendido entre *0,5* y *0,7* (*intervalo de credibilidad* del 84,86%) es:

$$P\ (0{,}5<p<0{,}7/\ datos)=0{,}1553+0{,}3484+0{,}3449=0{,}8486$$

Supongamos que quiero contrastar la hipótesis de que A va a perder las elecciones, es decir *H:* $p \leq 0{,}5$.

De la distribución final puedo ver que la probabilidad $P(p<0{,}5/datos)=0{,}0327$ es pequeña, y consideramos esto es una buena evidencia de que, de hecho, el candidato A ganará (y por tanto rechazamos la hipótesis).

Ejercicio resuelto 2.1. Encuesta de satisfacción

Enunciado

Una empresa de productos alimenticios quiere evaluar el grado de satisfacción de sus usuarios mediante entrevistas personales. Como hipótesis inicial, supone que la proporción de clientes satisfechos puede ser el 50%, 60%, 70% u 80%, aunque no está seguro cuál de estas cifras es más acertadas.

a) Asigna una distribución discreta de probabilidad inicial para la proporción de clientes satisfechos con el producto

b) La empresa entrevista a 4 clientes, de los cuales 3 estuvieron satisfechos con el producto. Calcula las verosimilitudes y distribución final de la proporción de clientes satisfechos, de acuerdo a los datos.

Datos del ejercicio

- No nos da información inicial sobre la probabilidad de cada uno de los cuatro posibles valores de la proporción de clientes satisfechos. Por tanto asumimos que los cuatro valores son equiprobables.
- Sabemos que de un total de 4 entrevistados, 3 están satisfechos, es decir en una muestra de 4 elementos encontramos 3 aciertos y un fallo.

¿Qué pide el ejercicio?

- En la parte a) nos pide la distribución inicial de la proporción.
- En la b) se nos pide la distribución final. Previamente hay que calcular las verosimilitudes P (datos / p=50%); P (datos / p=60%); etc.

Solución parte a

Basta asignar igual probabilidad a los cuatro valores. Para mayor comodidad en los cálculos que siguen en la parte 2 usamos una tabla Bayes.

(1) Posibles valores de p	(2) Probabilidad inicial	(3) Verosimilitud	(4) Producto	(5) Probabilidad final
0,00	0,00			
0,10	0,00			
0,20	0,00			
0,30	0,00			
0,40	0,00			
0,50	0,25			
0,60	0,25			
0,70	0,25			
0,80	0,25			
0,90	0,0			
1,00	0,00			
Suma				

Solución parte b

La distribución final la calculamos en dos pasos. Primeramente calculamos la columna de verosimilitudes. Si no quieres hacer estos cálculos puedes usar el programa de Excel.

En primer lugar vemos que la verosimilitud de que los valores de p sean igual a 0 o a 1 es cero, puesto que no es posible ninguno de los casos (ya que la muestra tenia clientes satisfechos e insatisfechos).

Para el resto de los casos en esta columna la calculamos usando la fórmula binomial, teniendo en cuenta que estamos en el caso de una binomial B (4, p) donde hemos obtenido 3 éxitos. Estamos interesados en el caso de 3 éxitos en 4 ensayos para cada valor de p en la columna 1. Esta probabilidad viene dada por:

Verosimilitud de 3 éxitos en 4 ensayos = p= P (datos / p)=$P(X=3)=\binom{4}{3}p^3(1-p)$

Por ejemplo, P (datos / p=0,1)= $P(X=3)=\binom{4}{3}0,1^3(0,9)$

e igual para el resto de los casos. En la tabla que sigue damos la solución a los cálculos

(1) Posibles valores de p	(2) Probabilidad inicial	(3) Verosimilitud	(4) Producto	(5) Probabilidad final
0,0000	0,0000	0,0000	0,0000	0,0000
0,1000	0,0000	0,0036	0,0000	0,0000
0,2000	0,0000	0,0256	0,0000	0,0000
0,3000	0,0000	0,0756	0,0000	0,0000
0,4000	0,0000	0,1536	0,0000	0,0000
0,5000	0,2500	0,2500	0,0625	0,1765
0,6000	0,2500	0,3456	0,0864	0,2439
0,7000	0,2500	0,4116	0,1029	0,2905
0,8000	0,2500	0,4096	0,1024	0,2891
0,9000	0,0000	0,2916	0,0000	0,0000
1,0000	0,0000	0,0000	0,0000	0,0000
Suma	1,0000		0,3542	1,0000

Nota: Puedes ayudarte en los cálculos usando el programa Excel "Estimación proporción con distribución inicial discreta", que puedes descargar en:
http://www.ugr.es/local/cmetodo/bayes

Ejercicio resuelto 2.2. ¿Cuantos mayores viven en tu barrio?

Enunciado

En un barrio se desconoce la proporción de viviendas donde viven personas mayores.

1. *Asigna una distribución de probabilidad a priori que indique la proporción de viviendas con personas mayores en la ciudad, usando valores p= 0'1, 0'2, 0'3...1.*

2. *Preguntamos en 3 viviendas, y en dos de ellas viven ancianos. Calcula la verosimilitud de este resultado, en el supuesto de que la proporción de viviendas con ancianos sea p= 0'1, 0'2, 0'3...1.*

3. *¿Cuál es la proporción más probable de viviendas con personas mayores en el barrio, dados estos datos?*

4. *¿Cuál es la probabilidad de que más de la mitad de las viviendas tengan ancianos?*

Datos del ejercicio

- No nos da información inicial sobre el número de viviendas con ancianos. Estamos en el caso de ausencia de información inicial (antes de recoger los datos).
- Sabemos que de un total de 3 viviendas, se encontraron ancianos en 2, es decir, los datos de una muestra de 3 ensayos son 2 aciertos y un fallo.

¿Qué pide el ejercicio?

- En la parte 1 nos pide la distribución inicial de la proporción.
- En la 2 las verosimilitudes P (datos / p=0); P (datos / p=0,1); etc.
- En la parte 3 nos pide el mejor estimador de la proporción de viviendas con ancianos.
- En la parte 4 se nos pide una probabilidad en la distribución a posteriori.

Solución parte 1

Puesto que no tenemos información sobre la verdadera proporción de viviendas con ancianos, la distribución inicial de la proporción será uniforme, es decir, podemos suponer igual probabilidad a la proporción de viviendas con ancianos p=0, p=0,1, p=0,2, p=0,3,... p=0,9, p=1. Para ayudarnos a resolver el problema, usaremos una Tabla Bayes, incluyendo en ella nuestros datos

(1) Proporción posible de viviendas con ancianos	(2) Probabilidad inicial	(3) Verosimilitud de 2 éxitos en 3 ensayos	(4) Producto	(5) Probabilidad final
P=0	1/11			
P=0,1	1/11			
P=0,2	1/11			
P=0,3	1/11			
P=0,4	1/11			
P=0,5	1/11			
P=0,6	1/11			
P=0,7	1/11			
P=0,8	1/11			
P=0,9	1/11			
P=1	1/11			
Suma	1			

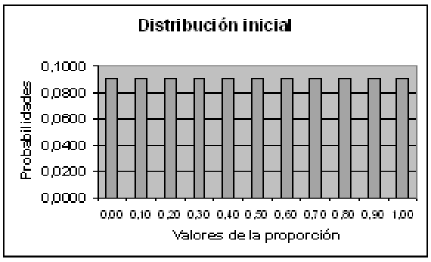
Solución parte 2

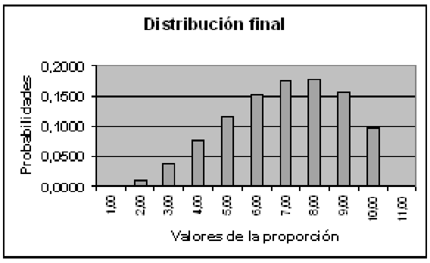
Para calcular la verosimilitud usaremos el programa Excel "Estimación proporción con distribución inicial discreta", introduciendo los datos de la distribución inicial así como el número de éxitos y fracasos: 2 éxitos y 1 fracaso en 3 ensayos

Estimación de una proporción con distribución inicial discreta

Valores posibles de la proporción	Distribución inicial	Verosimilitudes	Producto	Distribución final
0,0000	0,0909	0,0000	0,0000	0,0000
0,1000	0,0909	0,0270	0,0025	0,0109
0,2000	0,0909	0,0960	0,0087	0,0388
0,3000	0,0909	0,1890	0,0172	0,0764
0,4000	0,0909	0,2880	0,0262	0,1164
0,5000	0,0909	0,3750	0,0341	0,1515
0,6000	0,0909	0,4320	0,0393	0,1745
0,7000	0,0909	0,4410	0,0401	0,1782
0,8000	0,0909	0,3840	0,0349	0,1552
0,9000	0,0909	0,2430	0,0221	0,0982
1,0000	0,0909	0,0000	0,0000	0,0000
Total	1,0000		0,2250	1,0000

Valores observados	
Exitos	2,0000
Fracasos	1,0000





Solución parte 3

Para obtener el mejor estimador de la proporción de viviendas con ancianos, usamos la distribución final de la proporción, que hemos calculado en el paso anterior

Vemos que la moda (valor más probable en esta distribución final) es 0,7. Por tanto lo más probable es que la proporción de viviendas con ancianos sea el 70%.

Solución parte 4

Finalmente para calcular la probabilidad de que más de la mitad de las viviendas tengan ancianos usamos la distribución final que es la que mejor refleja nuestro conocimiento actual de la población.

Sumando las probabilidades correspondientes a los valores de p=0,5 o mayores en esta distribución obtenemos:

$$P\ (p \geq 0{,}5\ /\text{datos}) = 0{,}7576$$

Tema 3. Estimación de la proporción con distribución inicial continua. Distribución Beta

Objetivos del tema 3

En este tema aprenderás a:

- Asignar una distribución inicial, cuando el valor supuesto de la proporción en la población es continuo
- Usar la distribución Beta para asignar probabilidades iniciales
- Podemos interpretar los parámetros en la distribución Beta *Be(a,b): a* como el número de éxitos y *b* como el número de fracasos al realizar *a+b* experimentos
- Reconocer las formas diferentes dependiendo de *a* y *b*
- Reconocer que la moda de la distribución *Be(a, b)* se sitúa en el punto *a/(a+b) y que a* igual valor promedio la dispersión es menor cuanto mayor es *a+b*

3.1. Distribución inicial continua para la proporción

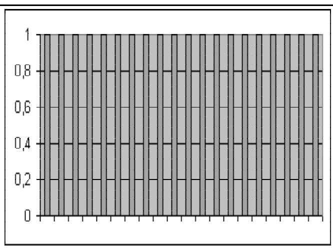
En muchas situaciones es razonable asumir un conjunto continuo de valores para proporción. El principal cambio es el uso de *distribuciones continuas de probabilidad* tanto en las probabilidades iniciales del parámetro como en las finales.

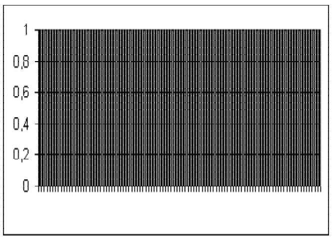
Por ejemplo, en vez de usar 11 valores diferentes de *p* (0, 0,1, 0,2, 0,3, 0,4...) podríamos tomar sucesivamente 21 valores (0, 0,05, 0,10, 0,15, 0,20, 0,25...), 100 valores y seguir así aumentando dicho número.

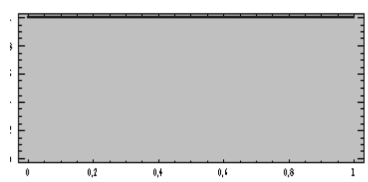
Si para cada uno de estos pasos representáramos las distribuciones iniciales, el gráfico de barras sería poco apropiado y debiéramos pasar poco a poco a un histograma que, progresivamente tendería al gráfico de una *función densidad uniforme* (ver Figura 3.1).

Figura 3.1. Distintas distribuciones inicial para la proporción

21 valores para p: 0, 05, 0,1, etc.	100 valores para p=0,01, 0,02,etc.	Cualquier valor de p entre 0 y 1







3.2. La distribución Beta

La distribución uniforme podría ser razonable para representar la situación en que no tenemos ninguna información previa sobre la proporción. Pero en muchos casos, este supuesto es poco razonable, por ejemplo:

- Proporción de varones entre todos los recién nacidos en un mes en un cierto hospital. Sabemos que lo más probable es que esta proporción esté alrededor de 0,51, aunque podría variar más o menos alrededor de este número, sobre todo si el número de niños que nace es pequeño.
- Proporción de ancianos con enfermedad de Alzheimer entre los residentes en un centro de mayores.

En los casos en que no es razonable pensar que todos los valores son equiprobables, una buena solución es usar la distribución Beta.

Las funciones de densidad beta vienen definidas por dos números, *a* y *b* y la llamamos *Be(a,b)*. La función de densidad en la distribución beta toma la forma siguiente:

$$f(p)= K\, p^{a-1}(1-p)^{b-1}$$

En esta fórmula *K* es una constante que usamos para convertir estas alturas de la curva en una función de densidad. Puesto que la suma de todas las probabilidades debe ser uno, habría que dividir cada altura por la suma (integral) para todos los valores de *p*. En la práctica no tenemos que preocuparnos de la fórmula de cálculo para usar la distribución Beta, ya que podemos usar las tablas de la distribución o bien programas estadísticos para calcularlas.

Una ventaja de la distribución Beta es que puede tomar formas muy diferentes, dependiendo de los valores a y b, por lo que es sencillo encontrar una distribución beta que exprese nuestras creencias sobre las probabilidades iniciales de la proporción.

3.3. Elección de la distribución inicial y cálculo de la distribución final

Para encontrar una función que se ajuste a la distribución de probabilidad inicial hay que tener en cuenta que:

- Para *a=b=1* obtenemos la distribución uniforme (distribución no informativa) donde todos los valores de p tienen la misma probabilidad

- Podemos interpretar *a* como el número de éxitos y *b* el número de fracasos al hacer un experimento *a+b* veces e interpretar *Be(a,b)* como una generalización de la distribución binomial. Pero mientras en la distribución binomial p era fijo y *a* y *b* eran variables ya que nos interesábamos por el número de éxitos, ahora la variable es *p*, siendo *a* y *b* fijos.

- El valor medio es *a/(a+b)*. Así para *a= 2; b=8* el valor medio es 0,2; para *a=b=5,* el valor medio es 0,5; para a=*7, b=3*, el valor medio es 0,7. Esto nos da el valor más probable de la proporción en la distribución inicial.

- Para el mismo valor medio, la mayor dispersión depende de la suma total de *a +b*. Así al comparar *Be(7, 3)* con *Be(70, 30)* o *Be(5,5)* con *Be(50, 50)* la dispersión disminuye cuando aumenta la suma de *a+ b*. Esto es lógico si pensamos que la información que proporcionan 50 éxitos en un experimento de 100 ensayos es mayor de la que proporcionan 5 éxitos en un experimento de 10 ensayos. Por lo tanto, al estar más seguros del valor más probable para la proporción, elegimos una distribución con menor dispersión para representar nuestra creencia.

Las funciones Beta también facilitan el cálculo probabilidades final (actualización de las probabilidades, a la luz de nuevos datos) sin tener que calcular la verosimilitud.

Dada una distribución de probabilidad inicial *Be(a,b),* si al hacer un nuevo experimento observamos *e* éxitos y *f* fracasos, la distribución de probabilidad final sería una nueva distribución Beta *Be(a+ e, b+ f)*.

En resumen, una vez seleccionada la función de densidad inicial, es fácil actualizarla con los datos. Bastaría con sumar los éxitos a *a* y sumar los fracasos a *b* al realizar el experimento.

3.4. Cálculo de probabilidades y obtención de inferencias con la distribución Beta

- Para calcular probabilidades en las distribuciones discretas bastaba con sumar las probabilidades asociadas a cada valor.

- Con las distribuciones continuas no tiene sentido hablar de probabilidades de que p tome un valor puntual. Ahora operamos con probabilidades asociadas a intervalos y tenemos que calcular el *área que encierra la curva* entre dos puntos.

Podemos ayudarnos en los cálculos con una hoja de cálculo Excel (ver ejemplo 3.2) y de este modo calcular intervalos de credibilidad para la proporción a partir de la distribución final.

Mejor estimador de una proporción

A partir de la distribución final podemos estimar la proporción poblacional desconocida. Nuestra apuesta más razonable sería el valor de algún promedio de la distribución final, como la moda, que es el valor más probable. En este caso la moda de la distribución final Beta.

Intervalos de credibilidad

El *intervalo de credibilidad* del 95% para la proporción poblacional se calcula tomando los valores centrales (alrededor de la moda) que den una probabilidad del 95% en la distribución final. Igualmente podemos calcular el intervalo de confianza del 99% o de cualquier otro valor de credibilidad.

De hecho, los límites del intervalo de credibilidad bayesiano coinciden con el intervalo de confianza clásico, cuando no hay información previa (distribución inicial uniforme) pero la interpretación de uno y otro es muy diferente:

- El intervalo de credibilidad del *r%* para la proporción nos indica que hay una probabilidad igual al *r%* de que la proporción de la población se encuentre en el intervalo.

- El intervalo de confianza del *r %* no nos da la probabilidad de que la proporción esté en el intervalo. Lo que nos dice es la proporción de intervalos que, con el mismo tamaño de muestra, contienen la proporción de la población. Es decir, si tomamos 100 muestras del mismo tamaño y calculamos para cada una de ellas el intervalo de confianza, a partir de los datos obtenidos, en *r* intervalos estará incluida la proporción verdadera de la población. Pero, en concreto, no sabemos si la proporción está o no incluida en el nuestro.

- Los límites de los intervalos de confianza y credibilidad no coinciden si hubiésemos usado una distribución inicial informativa, es decir cuando hay información previa.

Observa en el ejemplo 3.3 que:

- A mayor credibilidad, más anchura del intervalo.
- Para la misma credibilidad y distribución inicial, a mayor tamaño de muestra menor anchura de intervalo.

Contraste Bayesiano

Podemos también usar las distribuciones finales para *contrastar una hipótesis* sobre un posible valor de la proporción en la población.

Contrastar una hipótesis significa (desde el punto de vista bayesiano) encontrar su probabilidad final.

Por ejemplo supongamos que quiero contrastar la hipótesis de que una moneda está trucada y produce pocas caras, es decir H: $p \leq 0,5$, siendo p la proporción de caras que produce la moneda. Parto de una distribución inicial *Be(1,1)* (ausencia de información); si al lanzar la moneda 10 veces obtengo 3 caras, la distribución final es *Be(4, 8),* porque he obtenido 3 éxitos y 7 fracasos.

De esta distribución final puedo ver en las tablas de la distribución que la probabilidad *P(p ≤ 0,5/ datos)=0,887;* es grande, pero no lo suficiente para que la consideramos como una buena evidencia de nuestra hipótesis (ya que la complementaria 0,113 indica un suceso que ocurre 113 de cada 1000 veces (por tanto no es muy raro). Nuestra conclusión es que no podemos rechazar la hipótesis, aunque tampoco podemos aceptarla.

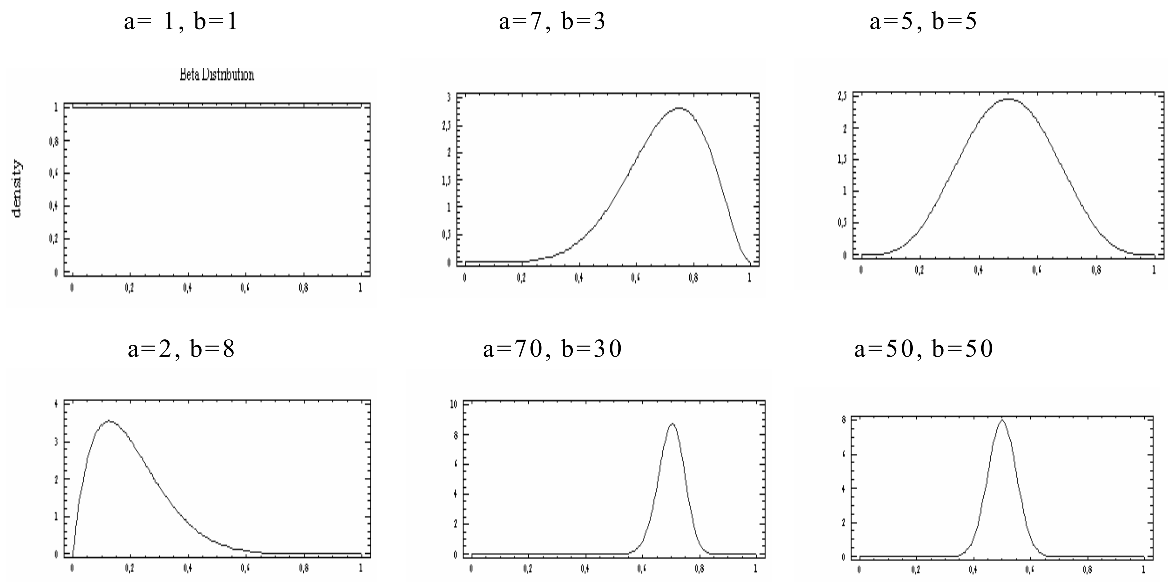
De nuevo, los resultados clásico y bayesiano coinciden cuando usamos distribuciones inicial no informativas. En la inferencia clásica el valor α para este contraste de hipótesis sería también igual a 0,113 en este ejemplo, por lo tanto, no rechazaríamos la hipótesis. Pero la interpretación es diferente:

- El valor α nos da la probabilidad de encontrar los datos, si fuese cierta la hipótesis nula, esto es, en este ejemplo *P(Datos/ p ≤ 0,5).*
- En inferencia bayesiana calculamos a partir de la distribución final la probabilidad de que la hipótesis sea cierta, a partir de los datos, *P(p ≤ 0,5/ datos).*

Ejemplo 3.1. Formas de la distribución Beta para valores de a y b

Algunos ejemplos de gráficas de la distribución Beta se presentan en la figura 3.2.

Figura 3.2. Forma de la distribución Beta para distintos valores de a y b



Ejemplo 3.2. Cálculo de probabilidades en la distribución beta

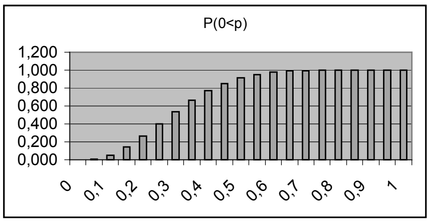
En la figura 3.3 representamos la distribución *Be (3,7*) junto con una tabla obtenida mediante uno de los programas Excel en que se proporcionan diferentes probabilidades.

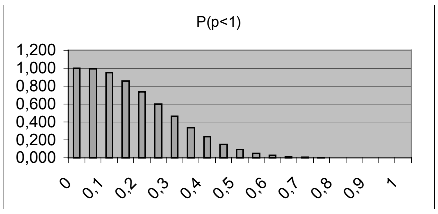
Figura 3.3. Cálculo de probabilidades y valores críticos para B(3,7)

Cálculo de probabilidades con la distribución Beta

Parámetro a	**3**
Parámetro b	**7**

p	P(0<p)	P (p<1)	P(0<x)	Valor crítico (x)
0	0,000	1,000	0,000	0,000
0,05	0,008	0,992	0,005	0,053
0,1	0,053	0,947	0,010	0,062
0,15	0,141	0,859	0,015	0,069
0,2	0,262	0,738	0,020	0,075
0,25	0,399	0,601	0,025	0,080
0,3	0,537	0,463	0,030	0,085
0,35	0,663	0,337	0,035	0,090
0,4	0,768	0,232	0,040	0,094
0,45	0,850	0,150	0,045	0,094
0,5	0,910	0,090	0,050	0,098
0,55	0,950	0,050	0,950	0,550
0,6	0,975	0,025	0,955	0,558
0,65	0,989	0,011	0,960	0,567
0,7	0,996	0,004	0,965	0,577
0,75	0,999	0,001	0,970	0,588
0,8	1,000	0,000	0,975	0,600
0,85	1,000	0,000	0,980	0,615
0,9	1,000	0,000	0,985	0,633
0,95	1,000	0,000	0,990	0,656
1	1,000	0,000	1,000	1,000





- Calculemos *P(0,3 < p <0,6)* en una probabilidad *B(3, 7)*. Esta proporción viene dada por el área de la curva entre 0,3 y 0,6

P(0,3 < p < 0,6)= P(p < 0,6)- P(0,3 < p)=0,975-0,537=0,438

- *P (p<0,3)=0,537*
- *Valor crítico x, tal que P (p<x)=0,45; x=0,094*
- *Percentil.* Los *percentiles* de la proporción son los valores de la proporción que dejan por debajo un tanto por ciento dado de valores. Por ejemplo, el percentil del *5 %* deja por debajo un *5 %* de casos, y es igual a 0,098, o lo que es lo mismo

P(p <0,098) = 0,05

Del mismo modo, *P(p<0,55)= 0,95*, es decir el percentil del 95% corresponde a una proporción igual a 0,55.

Ejemplo 3.3. Cálculo de intervalos de credibilidad

En una distribución *Be(3,7)* la probabilidad de que la proporción sea menor a 0,098 es el 0,05% y la probabilidad de que sea mayor a 0,55 también otro 5%. (Véase el ejemplo 3.2). En consecuencia, el 90% de los valores centrales están comprendidos entre (0, 098 <p<055).

El intervalo de credibilidad del 90 % para la proporción en una distribución final *Be(3,7*) es: *P(0, 098 <p<055)=0,90*

Del mismo modo podemos calcular intervalos de credibilidad para otro valor cualquiera de credibilidad, por ejemplo, para la misma distribución final *Be(3,7):*

- Intervalo de credibilidad del 95% (dejando arriba 0,025% de probabilidad y otro 0,025% abajo): *P(0, 008 <p<0,6)=0,95*
- Intervalo de credibilidad del 99 %: *P(0, 063 <p<0,656)=0,99*

Observa en estos ejemplos que a mayor credibilidad, más anchura del intervalo

Ejercicio resuelto 3.1. Hipertensión

Enunciado

En un estudio de salud se encontraron 72 mujeres hipertensas en una muestra de 412 mujeres.

a. ¿Cuál es el mejor estimador de la verdadera proporción de mujeres hipertensas en la población?

b. ¿Cuál es el límite en que variará la proporción de mujeres hipertensas en la población, con una credibilidad del 90%?

c. ¿Podemos admitir que el riesgo de hipertensión es menor en las mujeres que en los hombres, supuesto que la incidencia de hipertensión en los varones es 0,22?

Datos del ejercicio

El ejercicio nos da como datos el número de éxitos $e=72$ en una muestra de tamaño $n=412$. No nos da información sobre la distribución inicial en la población

¿Qué pide el ejercicio?

- La parte a pide la estimación más probable de la proporción en la población, dada por la moda de la distribución final de la proporción.
- La parte b pide el intervalo de credibilidad del 90% para la proporción en la población
- La parte c pide contrastar la hipótesis de que la proporción en la población es mayor que 0,22, que es la proporción en la población de varones.

Solución parte a

Puesto que no tenemos información sobre la distribución inicial de la proporción tomaremos como distribución inicial la *Be(1,1)*. La distribución final de la proporción será igual a *Be(73, 341),* ya que sumaremos los éxitos y fracasos a los parámetros de la distribución beta inicial.

La moda en esta distribución es igual a su media:

$$P=73/(73+341)=0,176$$

Solución parte b

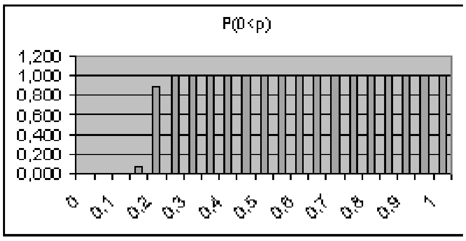
Para calcular el intervalo de credibilidad necesitamos las tablas de la distribución Beta o bien, el programa de cálculo de probabilidades de la distribución Beta, que puedes descargar de la pagina web. En la figura 3.4 aparece este programa en el que hemos cambiado los parámetros a y b por los datos del problema: $a=73; b=341$.

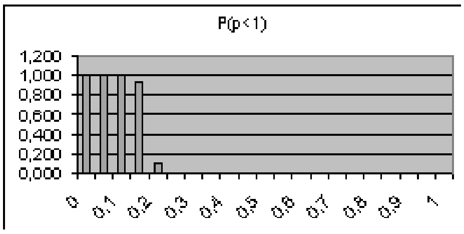
Figura 3.4. Distribución Be(73,341)

Cálculo de probabilidades con la distribución Beta

Parámetro a	73
Parámetro b	341

p	P(0<p)	P(p<1)	P(0<x)	Valor crítico (x)
0	0,000	1,000	0,000	0,000
0,05	0,000	1,000	0,005	0,135
0,1	0,000	1,000	0,010	0,138
0,15	0,075	0,925	0,015	0,140
0,2	0,894	0,106	0,020	0,141
0,25	1,000	0,000	0,025	0,143
0,3	1,000	0,000	0,030	0,144
0,35	1,000	0,000	0,035	0,145
0,4	1,000	0,000	0,040	0,146
0,45	1,000	0,000	0,045	0,146
0,5	1,000	0,000	0,050	0,146
0,55	1,000	0,000	0,950	0,208
0,6	1,000	0,000	0,955	0,209
0,65	1,000	0,000	0,960	0,210
0,7	1,000	0,000	0,965	0,211
0,75	1,000	0,000	0,970	0,213
0,8	1,000	0,000	0,975	0,214
0,85	1,000	0,000	0,980	0,216
0,9	1,000	0,000	0,985	0,219
0,95	1,000	0,000	0,990	0,222
1	1,000	0,000	1,000	1,000





Puesto que nos piden el intervalo de credibilidad del 90%, debemos buscar el intervalo central de valores que deja a cada lado un 5% de probabilidad.

Buscando en la tabla observamos que el valor crítico x, tal que $P(p<x_1)=0,05$ es igual a 0,146 y el valor crítico x, tal que $P(p<x_2)=0,95$ es igual a 0,208. El intervalo de credibilidad estará comprendido entre estos dos valores, ya que:

$$P(x_1 < p<x_2)=0,95-0,05=0,90$$

Solución parte c

Para contrastar esta hipótesis tendremos que calcular su probabilidad. De la tabla anterior obtenemos: $P\ (p<0,22)=0,99$

Por tanto, puesto que esta probabilidad es grande, podemos aceptar que la proporción de hipertensos es menor en las mujeres que en los hombres.

Ejercicio resuelto 3.2. Fallos en producción

Enunciado

En el proceso de control de calidad de una fábrica de reproductores de discos compactos se encontraron fallos en 20 reproductores, en una muestra aleatoria de 500.

a) Encuentre un intervalo de credibilidad de 90% para la proporción de los reproductores de discos compactos defectuosos en la población.

b) Revisadas las máquinas que realizan la producción se encontró un total de 5 discos defectuosos en un total de 200 reproductores. ¿Cuál será ahora el nuevo intervalo de credibilidad?

c) ¿Podemos suponer que el proceso de revisión ha mejorado la producción?

Datos del ejercicio

El ejercicio nos da como datos el número de éxitos $e=20$ en una muestra de tamaño $n=500$. No nos da información sobre la distribución inicial en la población

¿Qué pide el ejercicio?

- La parte a pide el intervalo de credibilidad del 95% para la proporción en la población.
- La parte b pide el nuevo intervalo de credibilidad del 95% para la proporción en la población, una vez tomamos una segunda muestra de datos. Ahora tendremos como distribución inicial la distribución final calculada en la fase anterior.
- La parte c pide contrastar la hipótesis de que la proporción en la población es menor que 0,04, que es la proporción en la población de partida.

Solución de la parte a

Puesto que no tenemos información sobre la distribución inicial de defectos en la población, tomaremos como distribución inicial la *Be(1,1)*. Una vez recogidos los datos, hemos encontrado 20 éxitos (número de defectos) y 480 fallos (número de piezas correctas). Luego la distribución final de defectos en la población viene descrita por la distribución *Be(21, 481)*.

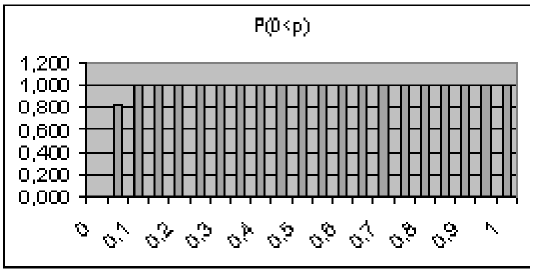
Utilizando el programa de cálculo de probabilidades Beta en Excel obtenemos los datos de la figura 3.5.

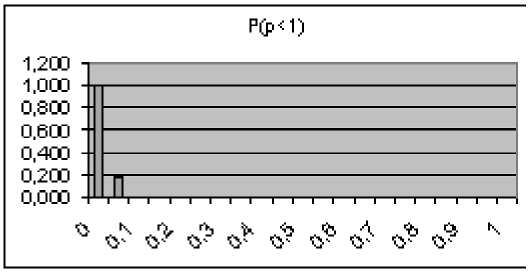
Figura 3.5. Distribución Be(21,481)

Cálculo de probabilidades con la distribución Beta

Parámetro a	21
Parámetro b	481

p	P(0<p)	P (p<1)	P (0<x)	Valor crítico (x)
0	0,000	1,000	0,000	0,000
0,05	0,824	0,176	0,005	0,024
0,1	1,000	0,000	0,010	0,025
0,15	1,000	0,000	0,015	0,026
0,2	1,000	0,000	0,020	0,026
0,25	1,000	0,000	0,025	0,027
0,3	1,000	0,000	0,030	0,027
0,35	1,000	0,000	0,035	0,028
0,4	1,000	0,000	0,040	0,028
0,45	1,000	0,000	0,045	0,028
0,5	1,000	0,000	0,050	0,028
0,55	1,000	0,000	0,950	0,057
0,6	1,000	0,000	0,955	0,058
0,65	1,000	0,000	0,960	0,059
0,7	1,000	0,000	0,965	0,059
0,75	1,000	0,000	0,970	0,060
0,8	1,000	0,000	0,975	0,061
0,85	1,000	0,000	0,980	0,062
0,9	1,000	0,000	0,985	0,063
0,95	1,000	0,000	0,990	0,065
1	1,000	0,000	1,000	1,000





Puesto que nos piden el intervalo de credibilidad del 95%, debemos buscar el intervalo central de valores que deja a cada lado un 2,5% de probabilidad.

Buscando en la tabla observamos que el valor crítico *x*, tal que $P(p<x_1)=0,025$ es igual a 0,027 y el valor crítico *x*, tal que $P(p<x_2)=0,975$ es igual a 0,061. El intervalo de credibilidad estará comprendido entre estos dos valores, ya que: $P(0,027 < p<0,061)=0,975-0,025=0,95$

Solución de la parte b

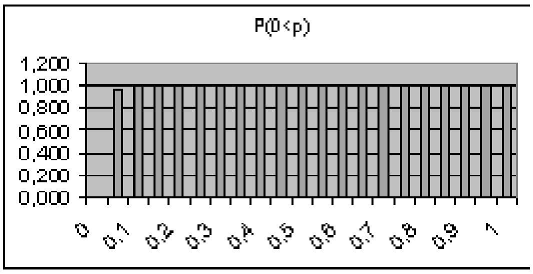
En este caso ya tenemos una información sobre la distribución inicial de defectos en la población, que es la distribución final obtenida en el paso anterior *Be(21, 481)*.

Una vez recogidos los nuevos datos, hemos encontrado 5 nuevos éxitos (número de defectos) y 195 nuevos fallos (número de piezas correctas). Luego la distribución final de defectos en la población viene descrita por la distribución *Be(26, 676)*. Utilizando el programa de cálculo de probabilidades Beta en Excel obtenemos los datos de la figura 3.6.

Cálculo de probabilidades con la distribución Beta

Parámetro a	26
Parámetro b	676

p	P(0<p)	P (p<1)	P(0<x)	Valor crítico (x)
0	0,000	1,000	0,000	0,000
0,05	0,956	0,044	0,005	0,022
0,1	1,000	0,000	0,010	0,023
0,15	1,000	0,000	0,015	0,024
0,2	1,000	0,000	0,020	0,024
0,25	1,000	0,000	0,025	0,025
0,3	1,000	0,000	0,030	0,025
0,35	1,000	0,000	0,035	0,026
0,4	1,000	0,000	0,040	0,026
0,45	1,000	0,000	0,045	0,026
0,5	1,000	0,000	0,050	0,026
0,55	1,000	0,000	0,950	0,049
0,6	1,000	0,000	0,955	0,050
0,65	1,000	0,000	0,960	0,050
0,7	1,000	0,000	0,965	0,051
0,75	1,000	0,000	0,970	0,052
0,8	1,000	0,000	0,975	0,052
0,85	1,000	0,000	0,980	0,053
0,9	1,000	0,000	0,985	0,054
0,95	1,000	0,000	0,990	0,056
1	1,000	0,000	1,000	1,000



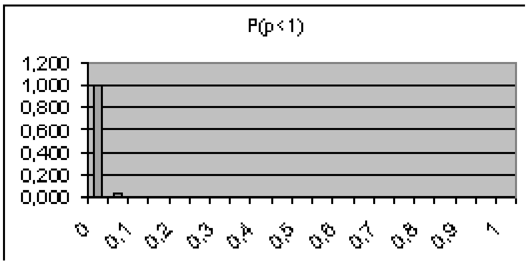


Figura 3.6. Distribución Be(26,676)

Puesto que nos piden el intervalo de credibilidad del 95%, debemos buscar el intervalo central de valores que deja a cada lado un 2,5% de probabilidad.

Buscando en la tabla observamos que el valor crítico *x*, tal que $P(p<x_1)=0,025$ es igual a 0,025 y el valor crítico *x*, tal que $P(p<x_2)=0,975$ es igual a 0,052. El intervalo de credibilidad estará comprendido entre estos dos valores, ya que:

$$P(0,025 < p<0,052)=0,975-0,025=0,95$$

Solución de la parte c

Al comparar los dos intervalos de credibilidad obtenidos:

$$P(0,027 < p<0,061)=0,975-0,025=0,95$$

$$P(0,025 < p<0,052)=0,975-0,025=0,95$$

Observamos que se solapan. Aunque el segundo intervalo (después de la revisión) tiene los límites más bajos para esta muestra particular, no se llega a observar una clara mejoría de la producción.

Tema 4. Inferencias sobre la media

Objetivos del tema 4

En este tema aprenderás a:

- Diferenciar entre media poblacional y media muestral
- Especificar valores plausibles de la media poblacional y asignarles probabilidades iniciales.
- Calcular verosimilitudes para diferentes valores posibles de la media poblacional a partir del valor obtenido en una muestra.
- Revisar las probabilidades iniciales de los valores plausibles de la proporción poblacional transformándolas en probabilidades finales utilizando el teorema de Bayes.
- Extraer inferencias sobre la población a partir de los datos de la muestra.

4.1. Distribución inicial no informativa o uniforme, conocida la desviación típica

Hasta ahora el parámetro de interés ha sido una proporción. La inferencia bayesiana también permite hacer estimaciones de la media poblacional, que es el parámetro que interesa en muchas situaciones, por ejemplo cuando queremos conocer el cociente de inteligencia medio (C.I.) de un grupo de sujetos.

Igual que en el caso de la proporción, usaremos una distribución inicial para el parámetro (la media) para representar muestro conocimiento previo. Siguiendo la regla de Bayes sabemos que:

Distribución final = K x Verosimilitud x Distribución inicial

Siendo *K* una constante y la *verosimilitud* la probabilidad de obtener un valor dado $\bar{x}$ de la media muestral para un valor dado μ de la media en la población, esto es:

$$\text{Verosimilitud} = P(\bar{x}/\mu)$$

Si conocemos la desviación típica de la población σ, entonces, si tomamos una muestra aleatoria de n elementos de esta población, la media de la muestra $\bar{x}$ tiene una desviación típica $\sigma/\sqrt{n}$, por tanto:

$$\text{Verosimilitud} = \mathrm{N}(\bar{x}, \sigma/\sqrt{n})$$

En consecuencia, obtenemos como verosimilitud la curva normal de media $\bar{x}$ y desviación típica $\sigma/\sqrt{n}$.

Se presentan dos casos: que tengamos un conocimiento previo sobre el valor probable de la media o que no lo tengamos.

- Si la distribución inicial es uniforme la probabilidad de todos los valores posibles de la media en la población es constante, y, en consecuencia

Distribución final = K x Verosimilitud x Distribución inicial =

K x Verosimilitud x K′ = K″ x Verosimilitud = $\mathrm{K'''} \times \mathrm{N}(\bar{x}, \sigma/\sqrt{n})$

Es decir, si la distribución inicial es uniforme, la distribución final coincide con la verosimilitud, que a su vez coincide con la curva normal de media $\bar{x}$ y desviación típica $\sigma/\sqrt{n}$.

Resumiendo:
Si en una muestra de *n* observaciones de una población normal N(μ, σ) hemos obtenido una media muestral $\bar{x}$ y *distribución inicial de la media* μ de la población *es uniforme,* la distribución final de la media μ de la población es también normal $N(\bar{x}, \sigma/\sqrt{n}.)$.

Por ejemplo, si en una población normal de desviación típica σ=10 obtenemos una media $\bar{x}=5$ en una muestra de 100 elementos y no tenemos información previa de la distribución inicial para la media de la población, la distribución final será $N(\bar{x}, \sigma/\sqrt{n}) = N(5, 10/\sqrt{100}) = N(5, 1)$.

¿Qué hacemos cuando no conocemos la desviación típica?

Si σ (desviación típica poblacional) se desconoce, podemos usar *s*, la desviación típica muestral, siempre que la muestra sea lo suficientemente grande (al menos n > 30).

4.2. Uso de la distribución normal tipificada *N (0,1)* para el cálculo de probabilidades

Puesto que la distribución final para la media de una población normal N(μ, σ) y distribución inicial uniforme es también $N(\bar{x}, \sigma/\sqrt{n}.)$ podemos usar las tablas de la distribución normal tipificada (Tabla 4.1) para calcular probabilidades en esta distribución final.

En el ejemplo anterior, si queremos calcular el intervalo de credibilidad de la media en la población y puesto que la distribución final es N(5,1) tenemos:

$$0{,}95 = P(-1{,}96 < Z < 1{,}96) = P(-1{,}96 < \frac{1{,}96 < \mu - 5}{1} < 1{,}96) = P(-1{,}96 < \mu - 5 < 1{,}96) =$$

$$P(5-1{,}96 < \mu < 5+1{,}96) = P(3{,}94 < \mu < 6{,}96)$$

El intervalo de credibilidad del 95% de la media en la población es [3,35-6,65]

Tabla 4.1. Valores de la distribución normal N (0,1), P (Z ≤ Z_x)

Z_x	0,0000	0,0100	0,0200	0,0300	0,0400	0,0500	0,0600	0,0700	0,0800	0,0900
-3,6	0,00016	0,00015	0,00015	0,00014	0,00014	0,00013	0,00013	0,00012	0,00012	0,00011
-3,5	0,00023	0,00022	0,00022	0,00021	0,0002	0,00019	0,00019	0,00018	0,00017	0,00017
-3,4	0,00034	0,00032	0,00031	0,0003	0,00029	0,00028	0,00027	0,00026	0,00025	0,00024
-3,3	0,00048	0,00047	0,00045	0,00043	0,00042	0,0004	0,00039	0,00038	0,00036	0,00035
-3,2	0,00069	0,00066	0,00064	0,00062	0,0006	0,00058	0,00056	0,00054	0,00052	0,0005
-3,1	0,00097	0,00094	0,0009	0,00087	0,00084	0,00082	0,00079	0,00076	0,00074	0,00071
-3	0,00135	0,00131	0,00126	0,00122	0,00118	0,00114	0,00111	0,00107	0,00104	0,001
-2,9	0,00187	0,00181	0,00175	0,00169	0,00164	0,00159	0,00154	0,00149	0,00144	0,00139
-2,8	0,00256	0,00248	0,0024	0,00233	0,00226	0,00219	0,00212	0,00205	0,00199	0,00193
-2,7	0,00347	0,00336	0,00326	0,00317	0,00307	0,00298	0,00289	0,0028	0,00272	0,00264
-2,6	0,00466	0,00453	0,0044	0,00427	0,00415	0,00402	0,00391	0,00379	0,00368	0,00357
-2,5	0,00621	0,00604	0,00587	0,0057	0,00554	0,00539	0,00523	0,00508	0,00494	0,0048
-2,4	0,0082	0,00798	0,00776	0,00755	0,00734	0,00714	0,00695	0,00676	0,00657	0,00639
-2,3	0,01072	0,01044	0,01017	0,0099	0,00964	0,00939	0,00914	0,00889	0,00866	0,00842
-2,2	0,0139	0,01355	0,01321	0,01287	0,01255	0,01222	0,01191	0,0116	0,0113	0,01101
-2,1	0,01786	0,01743	0,017	0,01659	0,01618	0,01578	0,01539	0,015	0,01463	0,01426
-2	0,02275	0,02222	0,02169	0,02118	0,02068	0,02018	0,0197	0,01923	0,01876	0,01831
-1,9	0,02872	0,02807	0,02743	0,0268	0,02619	0,02559	0,025	0,02442	0,02385	0,0233
-1,8	0,03593	0,03515	0,03438	0,03362	0,03288	0,03216	0,03144	0,03074	0,03005	0,02938
-1,7	0,04457	0,04363	0,04272	0,04182	0,04093	0,04006	0,0392	0,03836	0,03754	0,03673
-1,6	0,0548	0,0537	0,05262	0,05155	0,0505	0,04947	0,04846	0,04746	0,04648	0,04551
-1,5	0,06681	0,06552	0,06426	0,06301	0,06178	0,06057	0,05938	0,05821	0,05705	0,05592
-1,4	0,08076	0,07927	0,0778	0,07636	0,07493	0,07353	0,07215	0,07078	0,06944	0,06811
-1,3	0,0968	0,0951	0,09342	0,09176	0,09012	0,08851	0,08692	0,08534	0,08379	0,08226
-1,2	0,11507	0,11314	0,11123	0,10935	0,10749	0,10565	0,10383	0,10204	0,10027	0,09853
-1,1	0,13567	0,1335	0,13136	0,12924	0,12714	0,12507	0,12302	0,121	0,119	0,11702
-1	0,15866	0,15625	0,15386	0,15151	0,14917	0,14686	0,14457	0,14231	0,14007	0,13786
-0,9	0,18406	0,18141	0,17879	0,17619	0,17361	0,17106	0,16853	0,16602	0,16354	0,16109
-0,8	0,21186	0,20897	0,20611	0,20327	0,20045	0,19766	0,19489	0,19215	0,18943	0,18673
-0,7	0,24196	0,23885	0,23576	0,2327	0,22965	0,22663	0,22363	0,22065	0,2177	0,21476
-0,6	0,27425	0,27093	0,26763	0,26435	0,26109	0,25785	0,25463	0,25143	0,24825	0,2451
-0,5	0,30854	0,30503	0,30153	0,29806	0,2946	0,29116	0,28774	0,28434	0,28096	0,2776
-0,4	0,34458	0,3409	0,33724	0,3336	0,32997	0,32636	0,32276	0,31918	0,31561	0,31207
-0,3	0,38209	0,37828	0,37448	0,3707	0,36693	0,36317	0,35942	0,35569	0,35197	0,34827
-0,2	0,42074	0,41683	0,41294	0,40905	0,40517	0,40129	0,39743	0,39358	0,38974	0,38591
-0,1	0,46017	0,4562	0,45224	0,44828	0,44433	0,44038	0,43644	0,43251	0,42858	0,42465
0	0,5000	0,49601	0,49202	0,48803	0,48405	0,48006	0,47608	0,4721	0,46812	0,46414
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5754
0,	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7258	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7518	0,7549

0,7	0,7580	0,7612	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7996	0,8023	0,8051	0,8079	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9430	0,9441
1,6	0,9452	0,9463	0,9474	0,9485	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9700	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9762	0,9767
2,0	0,9773	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9865	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9980	0,9980	0,9981
2,9	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

4.3. Predicción de una nueva observación

Supongamos que, en lugar de estar interesados en el valor de la media, queremos predecir qué valor tomará una nueva observación. Para ello tenemos dos datos:

- Una nueva observación x tendría la misma distribución de las anteriores, esto es, la distribución final $N(\mu, \sigma)$.
- Lo único que sabemos de μ es su distribución $N(\bar{x}, \sigma/\sqrt{n})$.
- En consecuencia tenemos dos fuentes de variación para predecir x:
    - La variación de μ alrededor de $\bar{x}$ que viene dada por $\sigma/\sqrt{n}$
    - La variación de x alrededor de μ que viene dada por σ

También sabemos que cuando se suman dos fuentes de variación sus varianzas se suman, pero no su desviación típica. Por tanto la varianza de una nueva observación viene dada por:

$$Var\ (x) = \sigma^2 + (\sigma/\sqrt{n})^2 = \sigma^2 + (\sigma^2/n) = \sigma^2\ (n+1)/n$$

De aquí deducimos que la desviación típica de la siguiente observación x es $\sigma\sqrt{(n+1)/n}$.

Resumen: Para predecir la siguiente observación, bastaría con usar la normal $N(\bar{x}, \sigma\sqrt{(n+1)/n})$. Dada la mayor variabilidad de la densidad predictiva de la siguiente observación, la distribución es mucho más dispersa que la densidad posterior de la media poblacional.

Ejemplo: En el ejemplo anterior para predecir la siguiente observación usaríamos la distribución normal $N(\bar{x}, \sigma\sqrt{(n+1)/n}) = N(5, 10\sqrt{(100+1)/100}) = N(5,\ 5{,}24)$.

4.4. Distribución inicial normal

En los apartados anteriores hemos supuesto que no tenemos información sobre el valor medio μ en la población. Pero hay ocasiones en que podemos tener cierta idea de su valor, o bien datos de un experimento anterior. En estos casos sería poco razonable suponer una distribución inicial uniforme.

Un modelo más razonable para la distribución inicial de la media es la distribución normal, ya que tiene las siguientes ventajas:

- Al aplicar la regla de Bayes a una distribución inicial normal obtendremos una distribución final normal
- La curva normal queda determinada simplemente por dos datos: su media y valor típico.

Supongamos ahora que nuestra distribución inicial para la media de la muestra sigue una curva normal $N(\mu_i,\ \sigma_i)$. Como recuerdas, esta distribución tiene forma de campana invertida, centrada alrededor de su valor medio μ_0, que también coincide con su moda (el valor más alto en la función de densidad).

Elegiremos como media μ_0 el valor que nos parezca más probable en la población. Para elegir la desviación típica σ_0 en la distribución inicial tenemos una regla sencilla que consiste en pensar cuál es el intervalo central de valores donde pensamos puede variar la media de esta población, teniendo en cuenta que si a la media de una distribución normal sumamos y restamos dos veces la desviación típica, aproximadamente en este intervalo están el 95% de los casos

Podemos aplicar de nuevo la fórmula de Bayes para deducir la distribución final:

$$Distribución\ final = K \times Verosimilitud \times Distribución\ inicial$$

Puede demostrarse que, en este caso la distribución final también sigue la distribución normal $N(\mu_f, \sigma_f)$. Los valores de la media y desviación típica de la distribución final vienen dados por las fórmulas siguientes:

$$\mu_f = \frac{\frac{n\bar{x}}{S^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{n}{S^2} + \frac{1}{\sigma_0^2}} \qquad \sigma_f = \frac{1}{\sqrt{n/S^2 + 1/\sigma_0^2}}$$

Donde n es el tamaño de la muestra, $\bar{x}$ y S la media y desviación típica de la muestra.

Resumiendo:

Si en una muestra de *n* observaciones de una población normal N(μ, σ) hemos obtenido una media muestral $\bar{x}$ y una desviación típica S y *distribución inicial de la media μ* de la población *es Normal N* (μ_0, σ_0) *,* la distribución final de la media μ de la población es también normal $N(\mu_0, \sigma_0)$.

En este caso, la media y desviación típica de la distribución final vienen dadas por las fórmulas siguientes:

$$\mu_f = \frac{\frac{n\bar{x}}{S^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{n}{S^2} + \frac{1}{\sigma_0^2}} \quad \sigma_f = \frac{1}{\sqrt{n/S^2 + 1/\sigma_0^2}}$$

Ejemplo 4.1. Evaluación de un programa de intervención

Sabemos que la distribución del C.I. toma una forma aproximadamente normal y esta distribución queda caracterizada cuando se conoce su media y desviación típica.

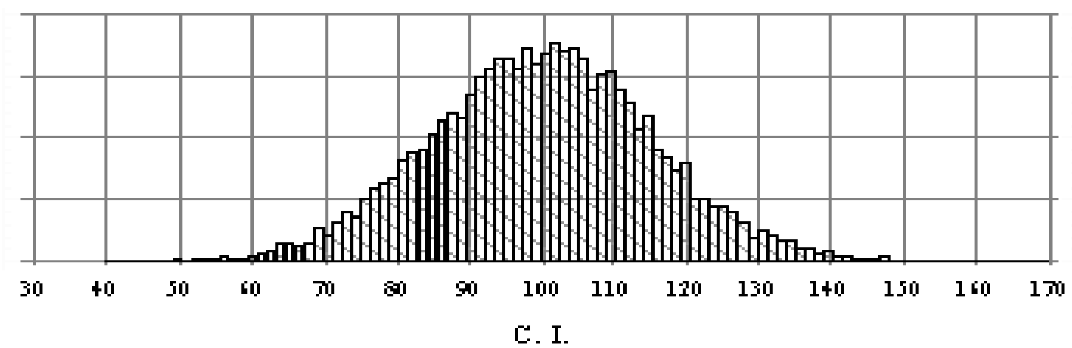


Figura 4.1. Distribución del C.I. en una muestra de 10000 sujetos

En la Figura 4.1 representamos la distribución del C.I. de 10.000 sujetos, que sigue una distribución *N (100, 15)* es decir con media 100 y desviación típica 15.

Supongamos que un grupo de educadores ha diseñado una nueva metodología para la enseñanza que se supone aumenta el C.I. y la prueba en un colegio. Para evaluar los resultados realizan una prueba consideran que la metodología ha tenido éxito si los alumnos tienen una mayor C.I. que el habitual. Si los investigadores sólo pueden evaluar los resultados con una muestra de 100 alumnos, en la que obtienen un valor *(*$\bar{x}$ *= 110).* ¿Podemos deducir que el método ha sido efectivo?

Como no sabemos nada sobre el valor medio posible del C.I. antes del experimento, usaremos una distribución uniforme, que, como recuerdas, asigna igual probabilidad a todos los valores posibles.

Por ejemplo, podríamos suponer que el C.I. puede ser igual que antes o aumentar como mucho 50 puntos. Cualquier valor comprendido entre 100 y 150 podría ser igualmente probable, entonces la distribución inicial sería la que se presenta en la Figura 4.2.

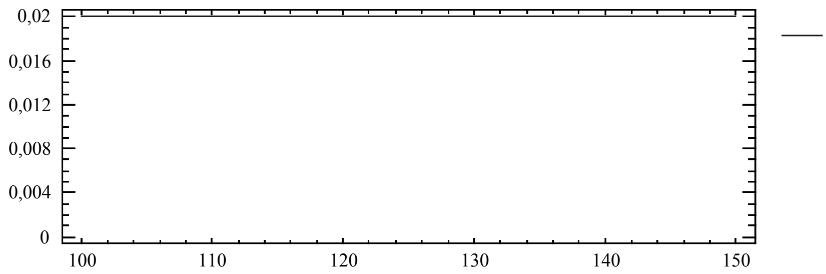


Figura 4.2. Distribución inicial uniforme para la media

Siguiendo la regla de Bayes sabemos que:

Distribución final = K x Verosimilitud x Distribución inicial

Siendo *K* una constante y la *verosimilitud* la probabilidad de obtener el valor dado de la media muestral *(*$\bar{x}$ *=110)* para un valor dado μ de la media en la población, esto es:

$$\text{Verosimilitud} = P(\bar{x} = 110 / \mu)$$

Si en una población, la desviación típica es σ, entonces, si tomamos una muestra aleatoria de n elementos de esta población, la media de la muestra $\bar{x}$ tiene una desviación típica $\sigma/\sqrt{n}$, por tanto:

$$\text{Verosimilitud} = P(\bar{x} = 110 / \mu) = P(Z = \frac{110 - \mu}{\sigma/\sqrt{n}} / \mu) = P(Z / \mu)$$

Siendo *Z* la distribución normal tipificada *N (0,1)*.

Pero la distribución inicial es en el caso que nos ocupa constante (uniforme), y, en consecuencia:

Distribución final = K x Verosimilitud x Distribución inicial = K' x P (Z /μ*)*

Es decir, si la distribución inicial es uniforme, la distribución final coincide con la verosimilitud, que a su vez coincide con la distribución de la media de la muestra (normal), considerándola como función de μ..

En consecuencia, obtenemos como verosimilitud la curva normal de media $\overline{x}$ y desviación típica $\sigma/\sqrt{n}$.

En nuestro ejemplo, la distribución final de la media μ en la población sería normal *N(110, 15/*$\sqrt{100}$*)*, es decir *N (110, 1,5).*

La variable $Z = (\mu - 110)/1{,}5$ sería *N (0,1)* es decir seguiría la distribución normal tipificada. Podemos ahora usar la tabla de la distribución *N(0,1)* para realizar inferencias.

¿Es efectivo el programa de intervención?

Para responder a esta pregunta debemos contrastar la hipótesis de que la nueva media μ es mayor de *100*.

Como de costumbre en el contraste de hipótesis, supondremos que se cumple la hipótesis contraria a la que queremos contrastar, es decir, que el programa es inefectivo.

Calculemos P ($\mu \leq 100$) con la distribución final:

$$\mathrm{P}\,(\mu \leq 100) = \mathrm{P}\left(\frac{\mu - 110}{1{,}5} \leq \frac{100-110}{1{,}5}\right)$$

$$= P\,(Z \leq -6{,}66) = P\,(Z > 6{,}66) = 0{,}000000$$

Pero esta probabilidad es muy pequeña.

La conclusión es que los datos *nos proporcionan una evidencia* de que el tratamiento es efectivo.

Ejemplo 4.2. Actitudes hacia la estadística

La escala SATS es un cuestionario formado por afirmaciones que tratan de reflejar los sentimientos y creencias que el alumno siente hacia la estadística. En total hay 28 afirmaciones, cada una de las cuales puntúa de 1 a 5 (a mayor puntuación las actitudes son más positivas).

Un investigador ha pasado la prueba SATS a una muestra de 248 estudiantes y ha obtenido un valor medio de la media muestral 88,8 con una desviación típica muestra igual a 13,33. El

Un investigador ha pasado la prueba SATS a una muestra de 248 estudiantes y ha obtenido un valor medio de la media muestral 88,8 con una desviación típica muestra igual a 13,33. El investigador desea predecir el valor de la puntuación media en el cuestionario en la población de estudiantes donde se tomó la muestra.

¿Cómo especificar una distribución inicial normal?

En este ejemplo es algo absurdo suponer que todos los valores posibles de la escala tienen la misma probabilidad.

Puesto que hay 28 ítems, los valores posibles pueden ir de 28 (si un estudiante puntúa todos los ítems con un 1) a 140 (si un estudiante puntúa todos los ítems con un 1). Un estudiante que puntuase todos los ítems con 3 tendría una actitud neutra, ni positiva ni negativa. En este caso el estudiante obtendría 84 puntos.

Supondremos entonces que nuestra distribución inicial para la media de la muestra sigue una curva normal $N(\mu_0, \sigma_0)$ centrada alrededor del valor más probable (en este caso 84 puntos).

En la figura 4.3 hemos representado tres posibles distribuciones inicial normales de la puntuación media en la escala SATS, todas con la misma media $\mu=84$.

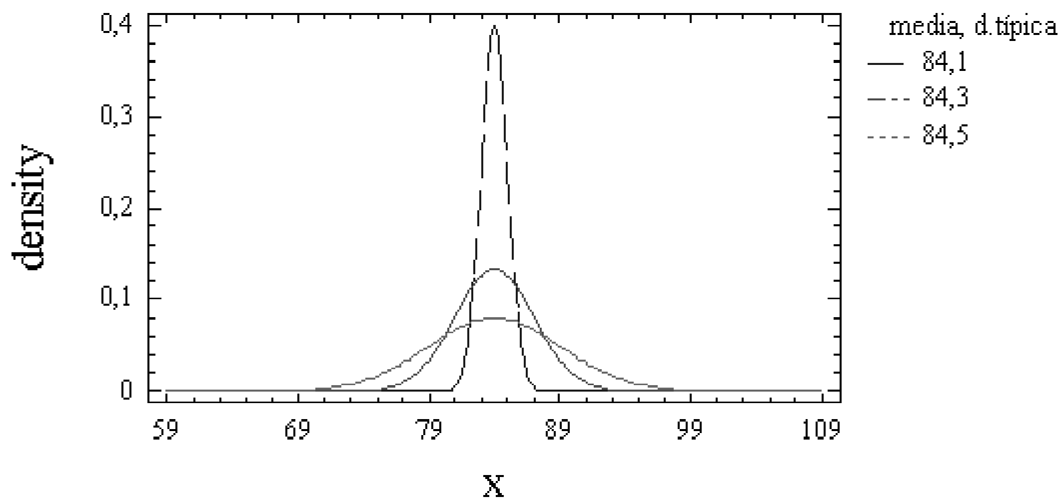
Para elegir la desviación típica σ_0 en la distribución inicial tenemos una regla sencilla que consiste en pensar cuál es el intervalo central de valores donde pensamos puede variar la media de esta población:

- Si estamos muy seguros de que la media en la población debería ser 84, podríamos tomar una desviación típica muy pequeña para esta distribución inicial, por ejemplo $\sigma_0=1$. Puesto que sabemos que en una distribución normal el 95% de los casos está aproximadamente a una distancia 2σ de la media, esto quiere decir que si tomamos $\sigma_0=1$ estamos pensamos que la media de la distribución inicial debería estar comprendida entre 82 y 86 con probabilidad 0,95.

- Del mismo modo, un valor $\sigma_0=3$ implica que esperamos que con probabilidad 0,95 el valor inicial de la media varíe entre 78 y 90

- El valor $\sigma_0=5$ supone que esperamos una variación de la media entre 74 y 94 con probabilidad 0,95.

Figura 4.3. Distintas posibles distribuciones inicial para la media en la escala SATS



Ejercicio resuelto 4.1. Efectividad de un tratamiento

Enunciado

En el ejemplo 4.1 sobre efectividad de un tratamiento para mejorar el cociente intelectual habíamos obtenido una distribución final de la media de la población N (110, 1,5) mientras que el valor teórico de la media de cocientes intelectuales es μ=100.

a) Calcular el intervalo de credibilidad para la media para medir la efectividad del tratamiento

b) Dar un valor esperado de la puntuación de un nuevo alumno que siga este programa

Datos del ejercicio

En este ejercicio se nos da la distribución final de la media de la población en una distribución normal. También se nos da el valor teórico de la media para contrastar una hipótesis.

¿Qué pide el ejercicio?

En la parte a se nos pide el intervalo de credibilidad del 95% para la media y contrastar la hipótesis de que la media en la población es mayor que 100.

En la parte b se nos pide predecir el CI de un nuevo sujeto cuando siga el tratamiento.

Solución de la parte a

Sabemos que esta distribución final es normal, con centro en el valor de la media muestral 100. El intervalo estará, entonces centrado en 100 y será el intervalo centrado en este valor que contenga a una probabilidad de 0,95. Puesto que la curva normal es simétrica, a cada lado del valor central habrá que dejar una probabilidad igual a 0,425. Mirando la tabla de la distribución normal tipificada Z vemos que:

$$P(-1{,}44 \leq Z \leq 1{,}44) = 0{,}95$$

$$P(-1{,}44 \leq \mu\text{-}110/1{,}5 \leq 1{,}44) = 0{,}95$$

$$P(-1{,}44 \times 1{,}5 \leq \mu\text{-}110 \leq 1{,}44 \times 1{,}5) = 0{,}95$$

$$P(-2{,}94 + 100 \leq \mu \leq 2{,}94 + 110) = 0{,}95$$

$$P(107{,}06 \leq \mu \leq 112{,}94) = 0{,}95$$

Luego el intervalo de credibilidad del 95 % pedido para la media después del tratamiento sería (107,06-112,94). Ha habido claramente una mejora y podemos aceptar la hipótesis de que el tratamiento sube el CI puesto que valor 100 no está incluido en el intervalo.

Solución de la parte b

Supongamos que un alumno ingresa al nuevo programa de intervención descrito en el apartado 3.1. Calculemos los valores más probables (95%) del valor de su C.I. cuando termine el tratamiento

La distribución predictiva de la población *es N* $N(\bar{x}, \sigma\sqrt{(n+1)/n})$ *(110, 15 √(101/100))* *≈N (110, 15).* Por tanto:

$$P\ (110\text{-}a < x < 110+a) = P\ (-a/15 < Z < a/15) =$$
$$P\ (-1{,}44 < Z < 1{,}44)$$
$$1{,}44 = a/15;\ a = 15 \times 1{,}44 = 16{,}44$$
$$P\ (93{,}56 < x < 126{,}44)$$

Los valores más probables de la siguiente observación son (93,56- 126,44). Aunque este intervalo contiene valores por debajo del 100, podemos ver que de hecho los valores más probables han subido respecto a la distribución de sujetos sin tratamiento.

También podemos ver que, conforme aumenta *n* la desviación típica de la distribución predictiva de la siguiente observación tiende al valor de la desviación típica en la población.

Ejercicio resuelto 4.2. Actitudes hacia la estadística

Enunciado

Supongamos que en el ejemplo del test de actitudes somos conservadores y suponemos que la distribución inicial de la puntuación media en la escala SATS es N (84, 5). En una muestra de 248 estudiantes y ha obtenido un valor medio de la media muestral 88,8 con una desviación típica muestra igual a 13,33.

a. Calcular el intervalo de credibilidad del 95% para la media

b. Contrastar la hipótesis de que la actitud de estos alumnos es superior a la teórica.

Datos del problema

El problema nos da la distribución a priori de la media *N (84, 5)* y los datos de una muestra; n=248; media muestral $\bar{x}$=88,8; desviación típica muestral s=13,33.

¿Qué pide el problema?

En la parte a) pide el intervalo de credibilidad del 95%

En la parte b) pide contrastar la hipótesis de que $\mu > 84$.

Solución de la parte a

Para calcular el intervalo de confianza, en primer lugar hemos de determinar la distribución final de la media, que es también normal. Su media y desviación típica viene dada por las fórmulas siguientes:

$$\mu_f = \frac{\dfrac{n\bar{x}}{S^2} + \dfrac{\mu_0}{\sigma_0^2}}{\dfrac{n}{S^2} + \dfrac{1}{\sigma_0^2}} \qquad \sigma_f = \frac{1}{\sqrt{n/S^2 + 1/\sigma_0^2}}$$

Podemos realizar estos cálculos directamente, o bien utilizar el programa de cálculo preparado para esta finalidad que puedes descargar de la página web (ver Figura 4.4).

Figura 4.4.Estimación de la media con distribución inicial normal

Estimación de la media de una muestra. Distribución inicial normal

Tamaño de la muestra		Media	Desviación Tipica	Precisión
30	En la distribución inicial	84	5	0,04
	En los datos	88,8	2,43	0,16935088
	En la distribución final	87,8828794	2,18555936	0,20935088

En este caso la media obtenida N (84,5) se acerca más a la de los datos, debido a que el tamaño de la muestra es muy grande y a que el valor obtenido en la media de la muestra era diferente del esperado. Pero para el caso de muestras pequeñas el valor de la media podría acercarse más a la distribución inicial.

Sabemos que en la distribución normal tipificada N (0,1) se verifica que:

$$P\ (-1{,}96 < Z < 1{,}96)=0{,}95$$

Y que la media μ en la distribución final tiene una distribución *N(87'88, 2'18).* Por tanto:

$$0{,}95 = P\ (-1{,}96<\frac{\mu-87{,}88}{2{,}13}<1{,}96)= P(-1{,}96\ x\ 2{,}13 < \mu-87{,}88 <1{,}96\ x\ 2{,}13) =$$

$$P(87{,}88- 4{,}175 < \mu < 87{,}88+ 4{,}175) = P\ (83{,}7 < \mu < 92{,}05)$$

Por tanto, el intervalo de credibilidad del 95% es $(83{,}7 < \mu < 92{,}05)$.

Solución parte b

Para contrastar la hipótesis de que $\mu >84$ calcularemos la probabilidad de la hipótesis contraria $\mu <84$ usando la distribución final de μ que es *N(87,88, 2,13).*

$$P(\mu < 84)= P(\frac{\mu-87{,}88}{2{,}13}<\frac{84-87{,}88}{2{,}13}) = P\ (Z <-1{,}82)$$

Siendo Z la distribución *N(0,1).*

Buscando en la tabla de la Normal *N (0,1):*

$$P\ (Z <-1{,}82)=1- P\ (Z >1{,}82)=0{,}0344$$

Luego podemos aceptar la hipótesis al ser pequeña la probabilidad de la hipótesis contraria.

Referencias

Albert, J. H. y Rossman, A. J. (2001). *Workshop statistics. Discovery with data. A Bayesian approach.* Emeryville, CA: Key College Publishing.

Berry, D. A. *Statistics. A Bayesian perspective*. Belmont, CA: Duxbury Press.

Corroyer, D. y Wolff, M. (2003). *L'analyse statistique des donneés en psychologie. Concepts et méthodes de base*. París: Armand Colin.

Lecoutre, B. (1996). *Traitement statistique des données experimentales.* París: CISIA.

Serrano Angulo. J. (2003). *Iniciación a la estadística bayesiana.* Madrid: La Muralla.

Página web:

http://www.ugr.es/local/mcdiaz/bayes

ANEXO 8.

DESCRIPCIÓN DE PROGRAMAS DE CÁLCULO UTILIZADOS EN LA ENSEÑANZA

1. DESCRIPCIÓN DEL PROGRAMA BAYES

1. **Objetivos:** Calcular las probabilidades finales $P(A_i/D)$ de un conjunto de sucesos A_1, A_2, etc. cuando se conocen sus probabilidades iniciales $P(A_i)$ y las verosimilitudes $P(D/A_i)$ de unos ciertos datos *D*, dados los sucesos A_i

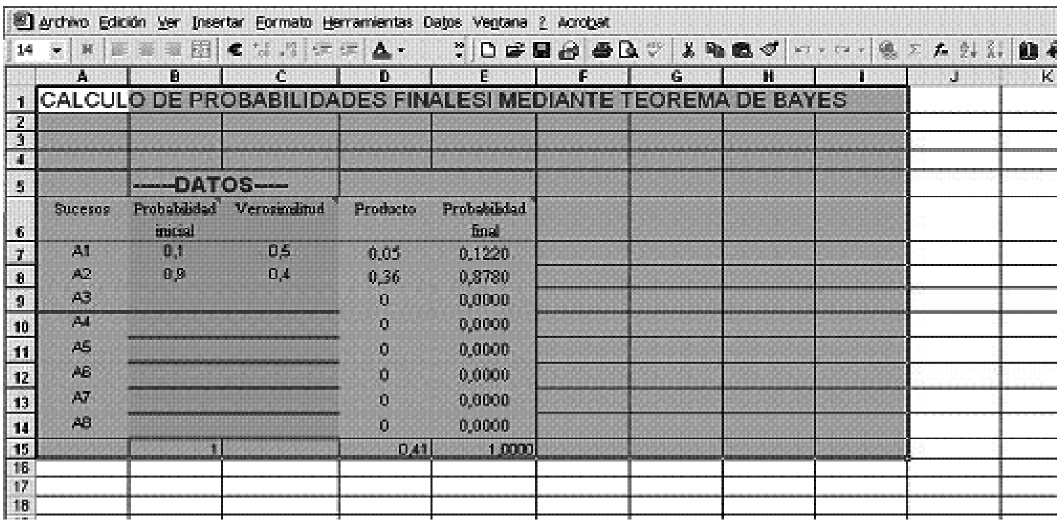
Este cálculo se hace mediante el teorema de Bayes

$$P(A_i/B) = K \times P(A_i) \times P(B/A_i)\text{ , siendo}$$

$$K = \frac{1}{P(A_1)\times P(B/A_1) + P(A_2)\times P(B/A_2) + ... + P(A_n)\times P(B/A_n)}$$

2. Datos requeridos: Los datos que pide el programa son las probabilidades iniciales de los sucesos y sus verosimilitudes (celdas sombreadas en azul en la Figura A8.1). En el ejemplo hemos considerado solo dos sucesos *A1* y *A2* con probabilidades 0,1 y 0,9. Las verosimilitudes de obtener los datos son respectivamente 0,5 si se verifica A1 y 0,4 si se verifica *A2*. Nota que la suma de las probabilidades iniciales tiene que ser igual a 1.

Figura A8.1. Programa de Cálculo de Probabilidades finales



CALCULO DE PROBABILIDADES FINALESI MEDIANTE TEOREMA DE BAYES

-----DATOS-----

Sucesos	Probabilidad inicial	Verosimilitud	Producto	Probabilidad final
A1	0,1	0,5	0,05	0,1220
A2	0,9	0,4	0,36	0,8780
A3			0	0,0000
A4			0	0,0000
A5			0	0,0000
A6			0	0,0000
A7			0	0,0000
A8			0	0,0000
	1		0,41	1,0000

3. Funcionamiento del programa: El programa calcula el producto de las probabilidades iniciales por las verosimilitudes y las probabilidades finales. Nota que la suma de probabilidades finales tiene que ser igual a 1.

2. DESCRIPCIÓN DEL PROGRAMA PRODIS

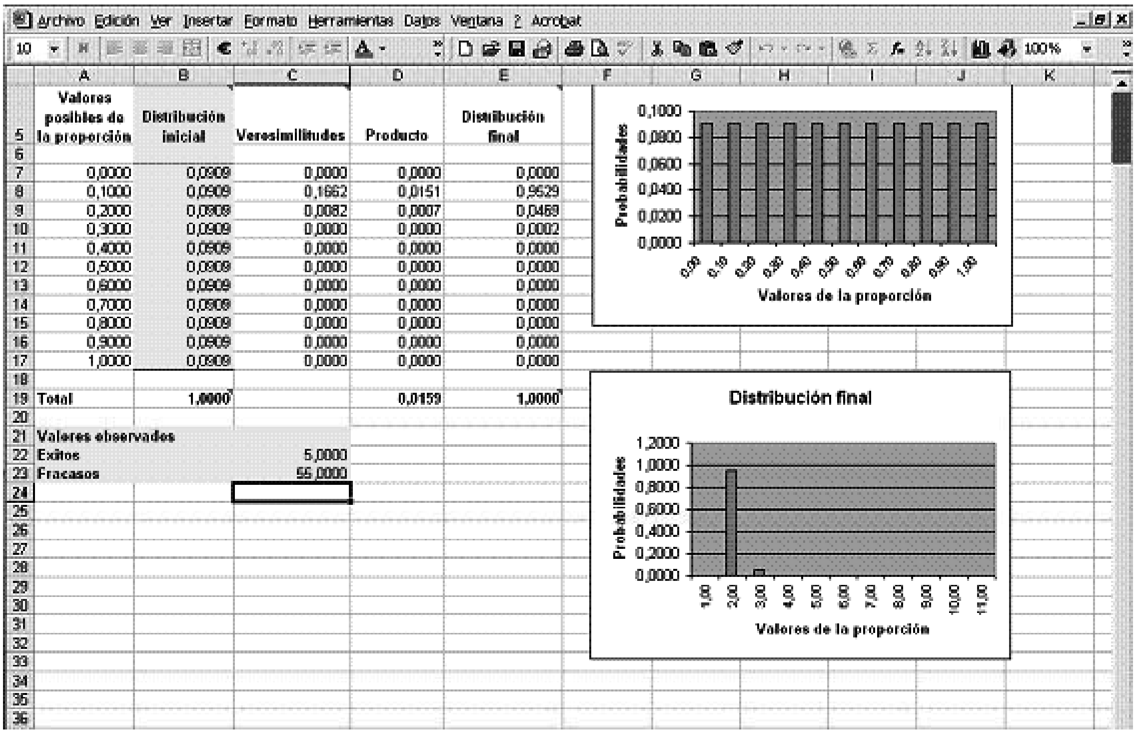
1. **Objetivos:** Transformar una distribución inicial $\mathrm{P}\,(\mathrm{p} = \mathrm{p}_0\,)$ de la proporción p en una población en su distribución final $P(p = p_0 \,/\, datos)$, para el caso de distribución inicial discreta. Como hemos estudiado en la teoría, este cálculo se hace mediante el teorema de Bayes:

$$P(p = p_0 \,/\, datos) = kx\mathrm{P}(\mathrm{datos}\ /\ \mathrm{p{=}p}_0\ )\ \mathrm{x}\ \mathrm{P}\ (\mathrm{p} = \mathrm{p}_0)$$

2. **Datos requeridos:** El programa pide dos tipos de datos (celdas sombreadas en azul en la Figura A8.2):

- Distribución inicial de la proporción (probabilidades iniciales $\mathrm{P}(\mathrm{p} = \mathrm{p}_0)$ para cada valor posible de la proporción). En el ejemplo hemos considerado que la proporción puede tomar los valores 0,1; 0,2...... 1 todos con igual probabilidad (0.1). Nota que la suma de las probabilidades en la distribución inicial tiene que ser igual a 1.
- Número de éxitos y fracasos en los datos (también sombreado de azul en la Figura A8.2). En el ejemplo hemos tenido 14 éxitos y 30 fracasos.

Figura A8.2. Cálculo de Distribución final de la proporción (caso discreto)



	A	B	C	D	E
5	Valores posibles de la proporción	Distribución inicial	Verosimilitudes	Producto	Distribución final
6					
7	0,0000	0,0909	0,0000	0,0000	0,0000
8	0,1000	0,0909	0,1662	0,0151	0,9529
9	0,2000	0,0909	0,0082	0,0007	0,0469
10	0,3000	0,0909	0,0000	0,0000	0,0002
11	0,4000	0,0909	0,0000	0,0000	0,0000
12	0,5000	0,0909	0,0000	0,0000	0,0000
13	0,6000	0,0909	0,0000	0,0000	0,0000
14	0,7000	0,0909	0,0000	0,0000	0,0000
15	0,8000	0,0909	0,0000	0,0000	0,0000
16	0,9000	0,0909	0,0000	0,0000	0,0000
17	1,0000	0,0909	0,0000	0,0000	0,0000
18					
19	Total	1,0000		0,0159	1,0000
20					
21	Valores observados				
22	Exitos		5,0000		
23	Fracasos		55,0000		

3. **Funcionamiento del programa:** El programa calcula las verosimilitudes $\mathrm{P}(\mathrm{datos}\ /\ \mathrm{p{=}p}_0)$ de cada posible valor de la proporción p_0 y sus probabilidades en la distribución final $P(p = p_0 \,/\, datos)$. Nota que la suma de probabilidades finales tiene que ser igual a 1.

El programa también representa gráficamente las distribuciones iniciales y finales de la proporción: Mientras que la distribución inicial indica nuestro conocimiento sobre la proporción ANTES de tomar los datos, la distribución final indica nuestro conocimiento sobre la proporción DESPUÉS de tomar los datos.

4. Posible utilización de los resultados: Podemos usar los resultados para:

- Calcular la mejor estimación de la proporción en la población (moda en la distribución final; en el ejemplo $p= 0,3$
- Calcular probabilidades referidas al verdadero valor de la proporción en la población; en el ejemplo $P\ (p< 0,3)=0,0002+0,1050= 0,1052$
- Usar estas probabilidades para contrastar hipótesis o calcular intervalos de credibilidad de la proporción. En el ejemplo $P(0,2<p<0,4)=0,96$ es el intervalo de credibilidad del 96%.

Nota. El programa se presenta en dos versiones: 1) Valores de p variando en intervalos de 0,1 (en la hoja 1) y 2) valores de p variando en intervalos de 0,05 (en la hoja 2).

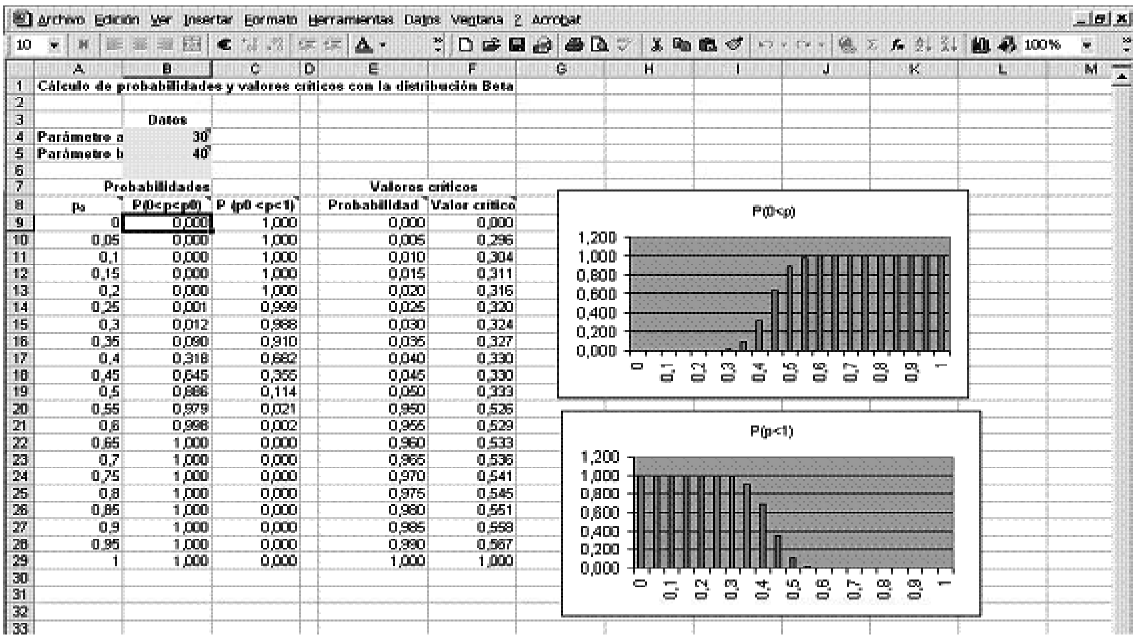
3. DESCRIPCIÓN DEL PROGRAMA BETA

1. **Objetivos:** Calcular probabilidades y valores críticos en una distribución Beta *B(a,b)*, donde a y b se interpretan como el número de éxitos y fracaso en una muestra de $n= a+b$ ensayos.

2. **Datos requeridos:** El programa pide lo parámetros a y b en una distribución *B(a,b),* esto es, el número de éxitos y fracasos en una muestra de $n= a+b$ ensayos (celdas sombreadas en azul en la Figura 3).

3. **Funcionamiento del programa:** El programa calcula dos tipos de probabilidades directas para la proporción para valores posibles p_0 que se han fijado en la primera columna, y varían en intervalos de 0,05:
   - Probabilidades de que la proporción p sea menor que un valor dado p_0, esto es, $P(0<p< p_0)$. Estas probabilidades se calculan en la columna 2 de la tabla.
   - Probabilidades de que la proporción p sea mayor que un valor dado p_0, esto es, $P(p_0<p<1)$. Estas probabilidades se calculan en la columna 3 de la tabla.
   - El programa calcula también los valores críticos x de la proporción tal que la probabilidad $P(0<p< x)$ está fijada. Para ello, en la columna 4 de la tabla se han fijado probabilidades desde 0 a 0,5 variando en intervalos de 0,05 y desde 0,95 a 1 variando en intervalos de 0,05. En la columna 5 se dan los valores críticos x correspondientes a cada una de estas probabilidades.

Figura A8.3. Cálculo de probabilidades y valores críticos en la distribución Beta



Cálculo de probabilidades y valores críticos con la distribución Beta

	Datos
Parámetro a	30
Parámetro b	40

Probabilidades			Valores críticos	
p0	P(0<p<p0)	P (p0 <p<1)	Probabilidad	Valor crítico
0	0,000	1,000	0,000	0,000
0,05	0,000	1,000	0,005	0,296
0,1	0,000	1,000	0,010	0,304
0,15	0,000	1,000	0,015	0,311
0,2	0,000	1,000	0,020	0,316
0,25	0,001	0,999	0,025	0,320
0,3	0,012	0,988	0,030	0,324
0,35	0,090	0,910	0,035	0,327
0,4	0,318	0,682	0,040	0,330
0,45	0,645	0,355	0,045	0,330
0,5	0,886	0,114	0,050	0,333
0,55	0,979	0,021	0,950	0,526
0,6	0,998	0,002	0,955	0,529
0,65	1,000	0,000	0,960	0,533
0,7	1,000	0,000	0,965	0,536
0,75	1,000	0,000	0,970	0,541
0,8	1,000	0,000	0,975	0,545
0,85	1,000	0,000	0,980	0,551
0,9	1,000	0,000	0,985	0,558
0,95	1,000	0,000	0,990	0,567
1	1,000	0,000	1,000	1,000

4. **Posible utilización de los resultados:** Podemos usar los resultados para:

- Calcular probabilidades referidas al verdadero valor de la proporción en la población; en el ejemplo P (p< 0,3)=0, 012; P (p> 0,3) =0,988; P (0,3<p< 0,5)=0,886-0,012=874
- Calcular valores críticos, por ejemplo, el valor x tal que P (0<p<x) sea fija. Por ejemplo, para P (0<p<x)=0,95; x=0,526; Esto quiere decir que la probabilidad de que la proporción sea menor que 0,526 es igual a 0,95.
- Usar estas probabilidades y valores críticos para contrastar hipótesis o calcular intervalos de credibilidad de la proporción. En el ejemplo el intervalo de credibilidad del 95% es (0,329-0,545) ya que P (0<p< 0,320)=0,025; P (0,545<p< 1)=0,0975 y P(0,329<p< 0,545)=0,95.

4. DESCRIPCIÓN DEL PROGRAMA MEDIA

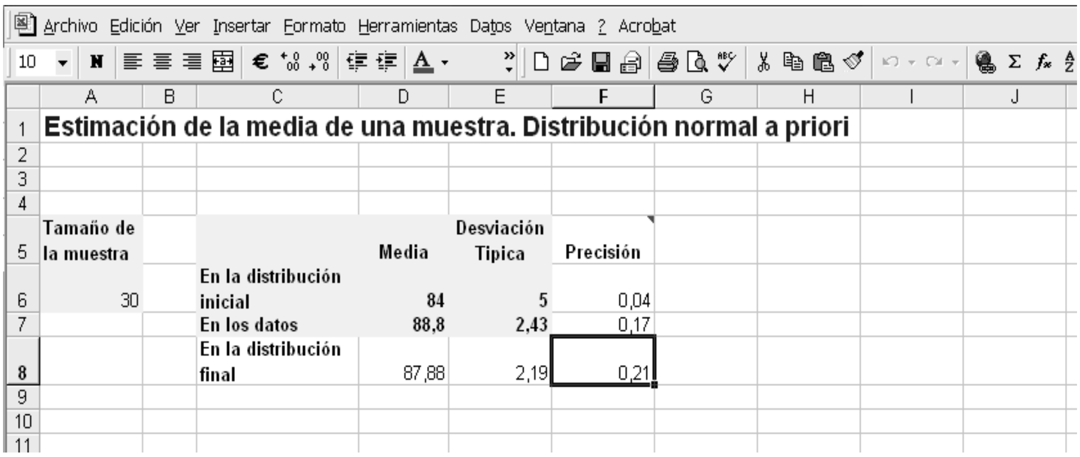
1. Objetivos: Calcular la media μ_f y la desviación típica σ_f de la distribución final de la media de una población normal $N(\mu_f, \sigma_f)$, cuando se conocen la media μ_0 y desviación típica σ_0 de la distribución inicial así como la media $\bar{x}$ y desviación típica S de los datos de la muestra.

2. Datos requeridos: El programa pide la media μ_0 y desviación típica σ_0 de la distribución inicial así como la media $\bar{x}$ y desviación típica S de los datos de la muestra.

3**. Funcionamiento del programa:** El programa calcula la media μ_f y la desviación típica σ_f de la distribución final de la media de una población normal $N(\mu_f, \sigma_f)$, mediante las fórmulas siguientes:

$$\mu_f = \frac{\dfrac{n\bar{x}}{S^2} + \dfrac{\mu_0}{\sigma_0^2}}{\dfrac{n}{S^2} + \dfrac{1}{\sigma_0^2}} \qquad \sigma_f = \frac{1}{\sqrt{n / S^2 + 1 / \sigma_0^2}}$$

Figura A8.4. Cálculo de media y desviación típica de la distribución final de la media en una población normal



Estimación de la media de una muestra. Distribución normal a priori				
Tamaño de la muestra		**Media**	**Desviación Tipica**	**Precisión**
30	**En la distribución inicial**	**84**	**5**	0,04
	En los datos	**88,8**	**2,43**	0,17
	En la distribución final	87,88	2,19	0,21

4. Posible utilización de los resultados. Una vez conocidas la media y desviación típica de la distribución final podemos usar las tablas de la distribución $N(0,1)$ para realizar todo tipo de cálculos, como en los siguientes ejemplos:

- $P(\mu_f < 90) = P\left(\dfrac{\mu_f - 87{,}88}{2{,}19} < \dfrac{90 - 87{,}88}{2{,}19} = 0{,}96\right) = P(Z < 0{,}96) = 0{,}9315$
- Intervalo de credibilidad del 95%; este intervalo deja por debajo y por encima una probabilidad del 2,5%; como $P(Z<-1{,}96)=0{,}025$ y $P(Z<1{,}96)0\ 0.0975$ el intervalo es $(-1{,}96 < Z < 1{,}96)$; es decir $\left(-1{,}96 < \dfrac{\mu_f - 87{,}88}{2{,}19} < 1{,}96\right)$ y despejando tenemos $(87{,}88-4{,}29 < \mu_f < 87{,}88-4{,}29)$ o lo que es lo mismo $(83{,}59 < \mu_f < 92{,}17)$.

ANEXO 9.

EVALUACIONES PARCIALES Y FINAL

Autoevaluación, tema 1

1. La diferencia entre la probabilidad inicial y final de un suceso es:
   a. La probabilidad inicial se conoce y la final no se puede conocer
   b. La probabilidad inicial se transforma en la final usando el teorema de Bayes
   c. La probabilidad final se transforma en la inicial usando el teorema de Bayes
   d. La probabilidad final exacta y la inicial aproximada

2. Si *E* es un suceso y *D* es un dato *P(D/E)* es:
   a. La verosimilitud de obtener el dato si ocurre el suceso
   b. La probabilidad final del suceso
   c. La probabilidad inicial del suceso
   d. La probabilidad conjunta de que ocurran a la vez el suceso y el dato

3. El teorema de Bayes permite calcular:
   a. Las verosimilitudes cuando se conocen las probabilidades inciales y finales
   b. Las probabilidades iniciales si se conocen las versomilitudes y las probabilidades finales
   c. Las probabilidades finales cuando se conocen las iniciales
   d. Las probabilidades finales cuando se conocen las iniciales y verosimilitudes

4. El valor de la probabilidad *P(D/E)* es aproximadamente igual al de la probabilidad *P(E/D)*:
   a. En todos los casos
   b. Depende de valor de P(E) y del valor de P/D/E)
   c. Nunca puede ser igual

5. En una ciudad 1 de cada 100 personas estudia inglés. 90 de cada 100 personas que estudia inglés ha viajado al extranjero y también 10 de cada 100 personas que no estudian inglés. La probabilidad de que al tomar una persona al azar, de entre todas las que han viajado al extranjero, haya estudiado inglés es:
   a. Una probabilidad inicial P (I)
   b. Una probabilidad final P(I/E)
   c. Una verosimilitud P(E/I)
   d. Una probabilidad conjunta P ($E \cap I$)

6. Para calcular la probabilidad final mediante el teorema de Bayes:
   a. Multiplicamos las probabilidades iniciales por las verosimilitudes y el resultado lo dividimos por la suma de todos estos productos
   b. Multiplicamos las probabilidades iniciales por las verosimilitudes y el resultado lo dividimos por la suma de las verosimilitudes
   c. Multiplicamos las probabilidades iniciales por las verosimilitudes y el resultado lo dividimos por la suma de las probabilidades iniciales
   d. Multiplicamos las probabilidades iniciales por las verosimilitudes

7. Imagina que tienes una muestra de 100.000 personas elegidas al azar. Imagina que 5 de cada 1000 está deprimida. Supón que una prueba de depresión da positivo en 99 de cada 100 personas enfermas y también en 3 de cada 100 personas sanas. Si D significa depresión y + significa prueba positiva, entonces:

a. *P(D)*=0,00005; *P(D/+)*=0,99; *P(S/+)*=0,03
b. *P(D)*=0,00005; *P(+/D)*=0,99; *P(-/D)*=0,01
c. *P(D)*=0,00005; *P(D/+)*=0,99; *P(-/D)*=0,01
d. *P(D)*=0,0005; *P(+/D)*=0,99; *P(-/D)*=0,01

8. Una de las siguientes fórmulas del Teorema de Bayes es falsa, ¿Cuál de ellas es falsa?

a. Es falsa $P(A_i / B) = \frac{P(A_i) \times P(B / A_i)}{P(A_1) \times P(B / A_1) + P(A_2) \times P(B / A_2) + \ldots + P(A_n) \times P(B / A_n)}$

b. Es falsa $P(B) = P(B/A_1)\ P(A_1) + \ldots + P(B/A_n)\ P(A_n)$
c. Es falsa *Probabilidad final= KxProbabilidad inicialx Verosimilitud*

d. Es falsa $P(A_i / B) = K \times P(A_i) \times P(B / A_i)$

Autoevaluación, tema 2

1. En un estudio sobre costumbres de los jóvenes españoles se hizo una encuesta a una muestra representativa de 1000 jóvenes. De ellos 700 tenían carnet de conducir.
   a. La proporción $p=0,7$ de jóvenes con carnet de conducir es un parámetro.
   b. La proporción $p=0,7$ de jóvenes con carnet de conducir es una estimación del verdadero valor del parámetro en la población.
   c. El 70% de los jóvenes españoles tienen carnet de conducir.
   d. Si tomamos otros 1000 jóvenes más, obtendremos otros 700 con carnet de conducir.

2. La proporción p de elementos que en una población cumplen un criterio dado:
   a. Se considera constante en inferencia bayesiana
   b. Se considera variable en inferencia clásica
   c. Se considera variable en inferencia bayesiana
   d. En unas poblaciones es constante y en otras es variable

3. La distribución inicial de la proporción p en una población viene dada por:
   a. El valor exacto de la proporción en la población
   b. El valor exacto de la proporción en la población y el tamaño de la muestra
   c. El número de éxitos y fracasos en la muestra
   d. Los valores posibles de la proporción en la población y la probabilidad de cada uno de ellos.

4. Conocidas las probabilidades iniciales de la proporción p en la población las probabilidades finales se calculan con la siguiente fórmula:
   a. $P\ (p = p_0\ /\ datos) = K\ x\ P(p = p_0)\ x\ P(datos/\ p = p_0)$
   b. $P\ (p = p_0) = K\ x\ P(p = p_0/\ datos)\ x\ P(datos/\ p = p_0)$
   c. $P(datos/\ p = p_0) = K\ x\ P(p = p_0)\ x\ P\ (p = p_0\ /\ datos)$
   d. $P\ (p = p_0\ /\ datos) = K\ x\ P(p = p_0)\ /P(datos/\ p = p_0)$

5. La distribución inicial de probabilidades de un parámetro:
   a. Recoge toda la información de la persona sobre la población antes de recoger los datos
   b. Recoge toda la información de la persona sobre la población después de recoger los datos
   c. Sirve para calcular el intervalo de credibilidad del parámetro
   d. Es una distribución fija; no podemos cambiarlas en nuevos experimentos

6. Cuando no tenemos ninguna información sobre los valores posibles de una proporción en la población, y queremos asignar una distribución inicial a la proporción en la población:
   a. Usaremos sólo el valor $p=1/2$, dando a este valor probabilidad 1 y al resto probabilidad cero
   b. Usaremos varios valores de la proporción entre 0 y 1 dando a todos la misma probabilidad
   c. No podemos asignar una distribución inicial, porque no sabemos nada de la proporción
   d. Usaremos sólo los valores 0 y 1 dando a cada uno probabilidad ½

7. En un edificio de 10 plantas sabemos que entre 6 y 8 están equipados de alarma contra incendios. También sabemos que la probabilidad de que haya exactamente 6 equipados de alarma es doble que la de que el número sea 6 o de que sea 7.

¿Cuál de las columnas A, B. C o D describe mejor la distribución inicial de la proporción de edificios con alarma?

Posibles valores de p	(A)	(B)	(C)	(D)
0,00	0,00	0,00	0,00	0,00
0,10	0,1	0,00	0,00	0,00
0,20	0,1	0,00	0,00	0,00
0,30	0,1	0,00	0,00	0,00
0,40	0,1	0,00	0,00	0,00
0,50	0,1	0,00	0,00	0,00
0,60	0,2	0,50	0,40	0,30
0,70	0,1	0,25	0,20	0,30
0,80	0,1	0,25	0,20	0,30
0,90	0,1	0,00	0,00	0,00
1,00	0,00	0,00	0,00	0,00

8. En una Facultad se quiere estimar la proporción de estudiantes que estudia inglés. A partir de una muestra de 10 estudiantes 7 estudiaban inglés. ¿Cuál es la verosimilitud de obtener estos datos, si exactamente el 50 % de los alumnos estudia inglés?

a. P (datos / p=0,5)= $P(X=7)=\binom{10}{7}0,5^7(0,5)$

b. P (datos / p=0,5)= $P(X=7)=\binom{10}{7}0,5^7(0,5)^3$

c. P (datos / p=0,5)= $P(X=7)=0,5^7(0,5)^3$

d. $P(X=7)=\binom{10}{7}0,5^7$

9. Al comenzar a resolver un problema de estimación de una proporción un alumno ha completado las tres primeras columnas de la tabla Bayes, obteniendo los datos siguientes:

(1) Posibles valores de p	(2) Probabilidad inicial	(3) Verosimilitud	--------	-----------
0,0000	0,0000	0,0000		
0,1000	0,1000	0,0000		
0,2000	0,1000	0,0233		
0,3000	0,1000	0,1239		
0,4000	0,1000	0,0682		
0,5000	0,1000	0,0065		
0,6000	0,1000	0,0001		
0,7000	0,1000	0,0000		
0,8000	0,1000	0,0000		
0,9000	0,1000	0,0000		
1,0000	0,1000	0,0000		
Suma			**0,0222**	

La probabilidad final de que el verdadero valor de la proporción sea igual a ½ es:

a. 0,065 b. 0,100 c. 0,0294 d. 0,0007

Autoevaluación, tema 3

1. Cuando no tenemos ninguna información sobre los valores posibles de una proporción en la población, usaremos como distribución inicial de la proporción para el caso continuo:

   a. La distribución normal $N(0,1)$ de media cero y desviación típica 1.
   b. La distribución uniforme discreta en el intervalo [0,1].
   c. La distribución $Be(0,1)$.
   d. La distribución $Be(1,1)$.

2. Supongamos que describo la proporción de emigrantes en una ciudad española mediante una distribución inicial *Be(3, 97).* Esto quiere decir que mi mejor estimación inicial de la proporción de emigrantes en la ciudad es:

   a. El 3 %
   b. El 97%
   c. Algo más del 3%
   d. Exactamente igual a 3 /97 %

3. Un estudio médico indica una incidencia de 10,1% de depresión en las mujeres en alguna época de su vida. Una posible distribución aproximada inicial para definir esta población sería:

   a. *Be(10, 100)* b) *Be(10, 90)* c) *Be(10,10)* d) *Be(90, 10)*

4. En un estudio sobre conducta social en el juego de los niños de preescolar se acepta como distribución inicial de la proporción de niños que acepta sus equivocaciones una distribución B (40, 60). En una muestra de 100 nuevos casos, 55 niños aceptaron sus equivocaciones durante el juego. La mejor estimación para la proporción de niños que acepta sus equivocaciones es

   a. El 55% b. El 40% c. El 60% d. El 47,5%

5. La siguiente tabla presenta las probabilidades y valores críticos de la distribución B (40,60).

p	P(0<p)	P (p<1)	P(0<x)	Valor crítico (x)
0	0,000	1,000	0,000	0,000
0,05	0,000	1,000	0,005	0,290
0,1	0,000	1,000	0,010	0,297
0,15	0,000	1,000	0,015	0,302
0,2	0,000	1,000	0,020	0,307
0,25	0,001	0,999	0,025	0,310
0,3	0,018	0,982	0,030	0,313
0,35	0,153	0,847	0,035	0,316
0,4	0,505	0,495	0,040	0,319
0,45	0,846	0,154	0,045	0,319
0,5	0,978	0,022	0,050	0,321
0,55	0,999	0,001	0,950	0,481
0,6	1,000	0,000	0,955	0,484
0,65	1,000	0,000	0,960	0,487
0,7	1,000	0,000	0,965	0,490
0,75	1,000	0,000	0,970	0,493
0,8	1,000	0,000	0,975	0,497
0,85	1,000	0,000	0,980	0,502
0,9	1,000	0,000	0,985	0,508
0,95	1,000	0,000	0,990	0,516
1	1,000	0,000	1,000	1,000

El intervalo de credibilidad del 95% para la proporción de una población descrita por una distribución final *Be*(40, 60) es aproximadamente:

a. $(0,35 < p < 0,5)$
b. $(0 < p < 0,516)$
c. $(0,310 < p < 0,497)$
d. $(0,25 < p < 0,8)$

6. A partir de los datos de la tabla anterior lo más razonable es aceptar la siguiente hipótesis sobre la proporción de niños de preescolar que acepta sus equivocaciones:

a. H: $p<0,3$
b. H: $p > 0,55$
c. H: $p> 0,35$
d. H: $p >0,45$

7. Para un mismo valor de la proporción en una muestra y una misma distribución inicial el intervalo de credibilidad del *r*% para la proporción en la población es:

a. Más ancho si aumento el tamaño de muestra
b. Más ancho si aumento el valor de r
c. Más estrecho si aumento el valor de r
d. Depende de la distribución final.

Autoevaluación, tema 4

1. Estoy realizando inferencias sobre la media de una población normal de desviación típica *σ=2*, pero no tengo ninguna información sobre su posible valor. En una muestra de 25 elementos he obtenido una media $\bar{x}$=10. Entonces tomaré como distribución <u>inicial</u>:

a. La distribución normal *N* (0,1)
b. La distribución normal *N* (10, 0’4)
c. La distribución normal *N* (10, 5)
d. Una distribución uniforme.

2. Si en una muestra de 100 elementos de una población normal de desviación típica σ=5 donde no tengo información inicial sobre la distribución de la media en la población he obtenido una media muestra $\bar{x}$=10, la distribución <u>final</u> de la media en la población es:

a. Una distribución uniforme.
b. Una distribución normal *N* (0,1)
c. Una distribución normal *N* (10,5)
d. Una distribución normal *N* (10, 0’5)

3. Para contrastar la hipótesis de que la media μ en una población normal con desviación típica σ=1 es menor que 5, mediante los datos obtenidos en una muestra de 16 elementos, siguiendo el procedimiento bayesiano:

a. Calculo la media de la muestra $\bar{x}$; calculo luego P ($\frac{\bar{x}-5}{0,25}$<5); si esta probabilidad es muy pequeña, acepto la hipótesis.
b. Calculo la media de la muestra $\bar{x}$; calculo luego P ($\frac{\bar{x}-5}{1}$<Z); siendo Z la distribución normal N (0,1); si esta probabilidad es muy pequeña acepto la hipótesis.
c. Calculo la media de la muestra $\bar{x}$; calculo luego P ($\frac{\bar{x}-5}{0,25}$ > Z); siendo Z la distribución normal N (0,1); si esta probabilidad es muy pequeña, acepto la hipótesis.
d. Calculo la media de la muestra $\bar{x}$; calculo luego P ($\frac{\bar{x}-5}{0,25}$<5); si esta probabilidad es muy pequeña, acepto la hipótesis.

4. En una muestra de 100 elementos de una población normal he obtenido una media $\bar{x}$. Suponiendo una distribución inicial uniforme de la media en la población, la distribución final de la media es:

a. Aproximadamente *N* ($\bar{x}$, 1)
b. Aproximadamente *N* ($\bar{x}$, s), siendo s la desviación típica de la muestra
c. No puedo decir nada, porque no conozco la desviación típica de la población
d. Aproximadamente *N* (0,1)

5. En una muestra de 25 elementos de una distribución normal con desviación típica he obtenido una media $\bar{x}$=10 σ=1, el intervalo de credibilidad del 95% de la media en la población es:

a. [10-1’96 x 0’2; 10 + 1’96 x 0’2]
b. [10-1’96; 10 +1’96]
c. [10 x 0’2 –1’96; 10 x 0’2 + 1’96]
d. [10-1’96 x 5; 10 +1’96 x 5]

6. Si en una misma población calculo en el intervalo de credibilidad para la media de la muestra, el intervalo es:

a. Más ancho si aumento el tamaño de la muestra
b. Más ancho si aumenta la desviación típica de la muestra
c. Más ancho si aumento la credibilidad
d. Más ancho si disminuyo la credibilidad

7. La tabla adjunta muestra los resultados de estimar la distribución final de la media mediante inferencia bayesiana.

Estimación de la media de una muestra. Distribución normal a prio

Tamaño de la muestra		Media	Desviación Tipica	Precisión
24	En la distribución inicial	100	5	0,04
	En los datos	88,8	2,43	0,16935088
	En la distribución final	90,939948	2,18555936	0,20935088

Para calcular un intervalo de credibilidad de la media en la población a partir de esta tabla usaríamos:

a. Una distribución normal *N* (100, 5)
b. Una distribución normal *N* (90, 2’18)
c. Una distribución normal *N* (88, 2’43)
d. Una distribución normal *N* (88, 2’43 /24)

Evaluación final

1. En una Facultad 10 de cada 100 personas estudia el último año. 60 de cada 100 estudiantes del último año comparte un piso con otros compañeros y también 30 de cada cien estudiantes de los cursos anteriores. C es el suceso compartir piso y U el suceso estar en el último año. Cogemos al azar un estudiante y resulta ser de último año, la probabilidad de que comparta un piso es:

a. Una probabilidad inicial *P (C)*
b. Una probabilidad final *P(C/U)*
c. Una verosimilitud *P(U/C)*
d. Una probabilidad conjunta *P* $(U\cap C)$

2. Imagina que tienes una muestra de 100.000 personas elegidas al azar. Imagina que 10 de cada 1000 está deprimida. Supón que una prueba de depresión da positivo en 99 de cada 100 personas enfermas y también en 2 de cada 100 personas sanas. Si *D* significa depresión y + significa prueba positiva, entonces, calcula las siguientes probabilidades:

$P(D)=$
$P(+/D)=$
$P(-/D)=$
$P(D\cap +)=$

3. El valor medio μ de una variable (por ejemplo la altura) en una población:

a. Se considera constante en inferencia bayesiana
b. Se considera variable en inferencia clásica
c. Se considera variable en inferencia bayesiana
d. En unas poblaciones es constante y en otras es variable

4. La distribución inicial de probabilidades de un parámetro:

a. Contiene toda la información de la persona sobre la población antes de recoger los datos
b. Se calcula mediante el teorema de Bayes, a partir de la distribución final
c. Sirve para calcular el intervalo de credibilidad del parámetro
d. Es una distribución uniforme

5. En un estudio sobre costumbres de los jóvenes españoles se hizo una encuesta a una muestra representativa de 1000 jóvenes. El tiempo medio que dedicaban al deporte fue 3 horas semanales.

a. La media de 3 horas es un parámetro de la población de jóvenes españoles
b. La media en la población es una variable aleatoria, los valores más probables de la misma serían alrededor de 3 horas.
c. La media en la población es constante, pero no la conocemos.
d. Cada joven español emplea 3 horas semanales en deporte.

6. Una fábrica de bombillas las vende en cajas de cuatro en cuatro. No sabemos nada sobre la proporción de bombillas defectuosas entre todas las fabricadas.
¿Cuál de las distribuciones A, B. C o D describe mejor la distribución inicial de la proporción de bombillas defectuosas en una caja?

(A)		(B)		(C)		(D)	
Valores proporción	Probabilidad	Valores proporción	Probabilidad	Valores proporción	Probabilidad	Valores proporción	Probabilidad
0,00	0,1	0,00	0,2	0,00	0,00	0,00	¼
0,25	0,1	0,25	0,2	0,01	0,25	0,25	¼
0,50	0,1	0,50	0,2	0,02	0,50	0,50	¼
0,75	0,1	075	0,2	0.03	0,75	075	¼
1	0,1	1	0,2	0.04	1	1	¼

6. En un centro geriátrico se quiere estimar la proporción de residentes con deterioro cognitivo, a partir de una muestra de 10 residentes 2 mostraron deterioro cognitivo. La versomilitud para el valor del parámetro p=0,1 es de 0,1937. ¿Qué significa este valor?

a. P (datos), es decir, probabilidad de obtener esta muestra.
b. P (datos ∩ p=0,1), es decir, probabilidad de obtener esta muestra y además que la proporción en la población sea 0,1.
c. P (p=0,1/datos), es decir, probabilidad de que conocida la muestra, el valor del parámetro sea 0,1.
d. P (datos / p=0,1), es decir, supuesto que el valor del parámetro sea 0,1, probabilidad de obtener esa muestra.

7. Al comenzar a resolver un problema de estimación de una proporción un alumno ha completado las tres primeras columnas de la tabla Bayes, obteniendo estos datos:

(1) Posibles valores de p	(2) Probabilidad inicial	(3) Verosimilitud	--------	-----------
0,0000	0,0000	0,0000		
0,1000	0,1000	0,0000		
0,2000	0,1000	0,0233		
0,3000	0,1000	0,1239		
0,4000	0,1000	0,0682		
0,5000	0,1000	0,0065		
0,6000	0,1000	0,0001		
0,7000	0,1000	0,0000		
0,8000	0,1000	0,0000		
0,9000	0,1000	0,0000		
1,0000	0,1000	0,0000		
Suma			0,0222	

La probabilidad final de que el verdadero valor de la proporción sea igual a 0,4 es:

a. 0,00682
b. 0,1000
c. 0,3072
d. 0,00015

8. Un estudio médico indica una incidencia de 15% de adicción al tabaco en las mujeres jóvenes. Una posible distribución aproximada inicial para definir esta población sería:

a. *Be* (15, 100)
b. *B*e (15, 85)
c. *Be* (85,15)
d. *Be* (100, 15)

9. La media de la distribución B(a,b) es igual a:

a. *a/b*
b. *(a+1)/(a+b)*
c. *(a+1)/(b+1)*
d. *a/(a+b)*

10. La siguiente tabla presenta las probabilidades y valores críticos de la distribución *Be*(30,40).

Probabilidades			**Valores críticos**	
p_0	**$P(0<p<p_0)$**	**$P(p_0<p<1)$**	**Probabilidad**	**Valor crítico**
0	0,000	1,000	0,000	0,000
0,05	0,000	1,000	0,005	0,296
0,1	0,000	1,000	0,010	0,304
0,15	0,000	1,000	0,015	0,311
0,2	0,000	1,000	0,020	0,316
0,25	0,001	0,999	0,025	0,320
0,3	0,012	0,988	0,030	0,324
0,35	0,090	0,910	0,035	0,327
0,4	0,318	0,682	0,040	0,330
0,45	0,645	0,355	0,045	0,330
0,5	0,886	0,114	0,050	0,333
0,55	0,979	0,021	0,950	0,526
0,6	0,998	0,002	0,955	0,529
0,65	1,000	0,000	0,960	0,533
0,7	1,000	0,000	0,965	0,536
0,75	1,000	0,000	0,970	0,541
0,8	1,000	0,000	0,975	0,545
0,85	1,000	0,000	0,980	0,551
0,9	1,000	0,000	0,985	0,558
0,95	1,000	0,000	0,990	0,567
1	1,000	0,000	1,000	1,000

El intervalo de credibilidad del 98% para la proporción de una población descrita por una distribución final *Be* (30, 40) es aproximadamente:

a. $(0{,}316 < p < 0{,}551)$
b. $(0{,}304 < p < 0{,}567)$
c. $(0{,}3 < p < 0{,}6)$
d. $(0{,}1 < p < 0{,}9)$

11. La distribución final de la proporción de votantes favorables a un partido viene dado por la distribución *Be*(30,40). A partir de los datos de la tabla anterior lo más razonable es aceptar la siguiente hipótesis sobre la proporción en la población:

a) H: p<0,25 b) H: p > 0,55 c) H: p> 0,25 d) H: p >0,45

12. Para una misma distribución final del parámetro en la población, el intervalo de credibilidad del **r%** para el parámetro es:

a. Más ancho si aumento el valor de r
b. Más ancho si aumento el tamaño de muestra
c. Más estrecho si aumento el valor de r
d. Depende de la distribución inicial

13. En una población normal con desviación típica σ=5 donde no tengo información inicial de la media poblacional, sacamos una muestra de 25 elementos y obtenemos una media $\bar{x}=100$. La distribución final de la media en la población es:

a. Una distribución normal N (100, 0,5)
b. Una distribución normal N (0,1)
c. Una distribución normal N (100,5)
d. Una distribución normal N (100, 1)

14. Para contrastar la hipótesis de que la media μ en una población normal con desviación típica *σ=1* es mayor que 5, mediante los datos obtenidos en una muestra de 100 elementos, siguiendo el procedimiento bayesiano:

a. Calculo la media de la muestra $\bar{x}$; calculo luego $P(\frac{\bar{x}-5}{0,1} < 5)$; si esta probabilidad es muy pequeña, acepto la hipótesis.
b. Calculo la media de la muestra $\bar{x}$; calculo luego $P(\frac{\bar{x}-5}{0,1} < Z)$; siendo *Z* la distribución normal *N (0,1);* si esta probabilidad es muy pequeña acepto la hipótesis.
c. Calculo la media de la muestra $\bar{x}$; calculo luego $P(\frac{\bar{x}-5}{0,1} > Z)$ siendo *Z* la distribución normal *N (0,1);* si esta probabilidad es muy pequeña, acepto la hipótesis.
d. Calculo la media de la muestra $\bar{x}$; calculo luego $P(\frac{\bar{x}-5}{0,1} > 5)$; si esta probabilidad es muy pequeña, acepto la hipótesis.

15. En una muestra de 100 elementos de una población normal he obtenido una media igual a 50. Suponiendo una distribución inicial uniforme de la media en la población la distribución final de la media es:

a. Aproximadamente *N (50, s),* siendo s la desviación típica de la muestra
b. Aproximadamente *N (50, s/10),* siendo s la desviación típica de la muestra
c. No puedo decir nada, porque no conozco la desviación típica de la población
d. Aproximadamente *N (0,1).*

16. La distribución final de la media en una población es *N* (100, 15). También sabemos que el valor *P* (-1,96 < *Z* < 1,96) =0,95, siendo *Z* la distribución normal *N* (0,1). El intervalo de credibilidad del 95% para la media en la población es:

a. (100-1,96 *x* 1,5; 100 + 1,96 *x*1,5)
b. (100-1,96; 100+1,96)
c. (100 *x* 1,5 –1,96; 100 *x* 1,5 + 1,96)
d. (100-1,96 *x* 15; 100 +1,96 *x* 15)

17. Tenemos los siguientes datos de un estudio sobre la altura media de las chicas españolas obtenido de datos de una muestra de 400 chicas

	Media	Desviación típica
En la muestra	160	10
En la distribución inicial	156	13
En la distribución final	158,5	7,9

Para hallar el intervalo de credibilidad de la media en la población utilizaríamos:

a. La distribución normal $N(160,10)$
b. La distribución normal $N\ (156,13)$
c. La distribución normal $N\ (158{,}5;\ 7{,}9)$
d. La distribución normal $N(160,\ 0{,}5)$

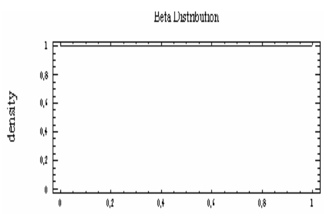
18. En un centro de preescolar el 20 % de los niños son hijos de inmigrantes y el 10% de las niñas. En el centro hay un 50% de niños y otro 50% de niñas. Utiliza la tabla siguiente para calcular la probabilidad de que si cogemos al azar un alumno inmigrante, sea varón.

Sucesos	Probabilidades iniciales	Verosimilitudes	Producto	Probabilidades finales
Suma	1			1

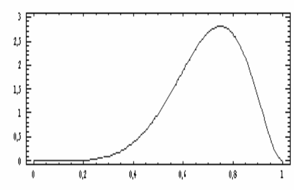
20. Observa las siguientes curvas Beta e indica cuál de ellas elegirías como distribución inicial de la proporción en las siguientes situaciones:

1. Estimación de la proporción de chinos que saben hablar otro idioma ____________
2. Estimación de la proporción de niños varones que nace en un cierto país _________
3. Estimación de la proporción de alumnos de la Facultad con sobrepreso___________

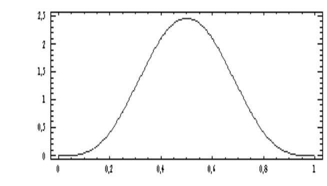
a= 1, b=1



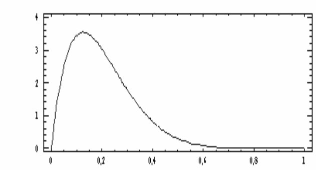
a=7, b=3



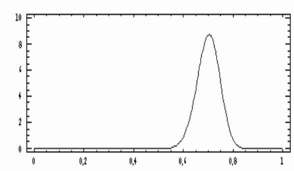
a=5, b=5



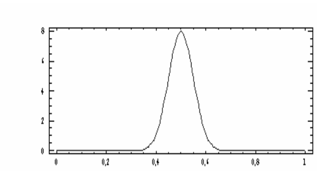
a=2, b=8



a=70, b=30



a=50, b=50



ANEXO 10.

TRADUCCIÓN DE CUESTIONARIOS RPC Y AIB

CPR QUESTIONNAIRE

Item 1. Explain in your own words what a simple and a conditional probability is and provide an example for each.

Item 2. Complete the sample space in the following random experiments:
a. Observing gender (male/female) of the children in a three children family (e.g. MFM,...)
b. Observing gender (male/female) of the children in a three children family when two or more children are male.

Item 3. A witness sees a crime involving a taxi in a city. The witness says that the taxi is blue. It is known from previous research that witnesses are correct 80% of the time when making such statements. The police also know than 15% of the taxis in the city are blue, the other 85% being green. What is the probability that a blue taxi was involved in the crime?
a. 80/100
b. b) 15 /100
c. (15/100) X (80/100)
d. $\frac{15 \times 80}{85 \times 20 + 15 \times 80}$

Item 4. A standard deck of playing cards has 52 cards. There are four suits (clubs, diamonds, hearts, and spades), each of which has thirteen numbered cards (2,..., 9, 10, Jack, Queen, King, Ace). We pick a card up at random. Let A be the event "getting diamonds" and B the event ·getting a Queen". Are events A and B independent?
a. They are not independent, since there is the Queen of diamonds
b. Only when we first get a card to see if it is a diamond, return the card to the pack and then get a second card to see if it is a Queen.
c. They are independent, since P(Queen of diamonds)= P(Queen)) x P(diamonds),
d. They are not independent, since P(Queen /diamonds) $\neq$ P(Queen).

Item 5. There are four lamps in a box, two of which are defective. We pick up two lamps at random from the box, one after another, without replacement. Given that the first lamp was defective:
a. The second lamp is more likely to be defective
b. The second lamp is most likely to be correct.
c. The probabilities for the second lamp being either correct or defective are the same.

Item 6. In a medical centre a group of people were interviewed with the following results:

	55 years-old or younger	Older than 55	Total
Previous heart stroke	29	75	104
No previous heart stroke	401	275	676
Total	430	350	780

Suppose we select at random a person from this group:

a. What is the probability that the person had a heart stroke?
b. What is the probability that the person had a heart stroke and, at the same time is older than 55?
c. When the person is older than 55, what is the probability of having had a heart stroke?
d. When the person had a heart stroke, what is the probability of being older than 55?

Item 7. 10.3 % of women in a given city have a positive mammogram. The probability that a woman in this city has both positive mammogram and a breast cancer is 0.8%- A mammogram given to a woman taken at random in this population was positive. What is the probability that she actually has breast cancer?

a. $\frac{0.8}{10.3} = 0.0776$, 7.76%   b. $10.3 \times 0.8 = 8.24$, 8.24%   c. 0.8 %

Item 8. In throwing two dice the product of the two numbers was 12. What is the probability that none of the two numbers is a six (we differentiate the order of numbers in the two dice).

Item 9. Suppose a tennis player goes to the Roland Garros posterior in 2005. He has to win 3 out of 5 sets to win. Which of the following events are more likely?

a. The player wins the first set
b. He wins the first set but looses the match
c. Both events a) and b) are equally likely

Item 10. A cancer test was given to all the residents in a large city. A positive result was indicative of cancer and a negative result of no cancer. Which of the following results is more likely?

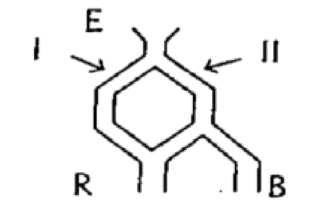
a. That a person had cancer if they got a positive result
b. Having a positive test if the person had cancer.
c. The two events are equally likely.

Item 11. 60% of the population in a city are men and 40% women. 50% of the men and 35% of the women smoke. If we pick a person from the city at random, what is the probability that the person is a smoker?

Item 12. A person throws a die and writes down the result (odd or even). It is a fair die (that is all the numbers are equally likely). These are the results after 15 throws:
Odd, even, even, odd, odd, even, odd, odd, odd, odd, even, even, odd, odd, odd
The person throws once more. What is the probability to get an odd number this time?

Item 13. A group of students in a school take a mathematics test and an English test.80% of the students pass the mathematics test and 70% of the students pass the English test. Assuming the two subjects score independently, what is the probability that a student passes both tests (mathematics and English)?

Item 14. Ojeda We throw a ball in the entrance E of a machine (see the figure). If the ball goes out through R, what is the probability of having passed by channel I?



a. 1/2
b. 1/3
c. 2/3
d. Cannot be computed

Item 15. According to a recent survey, 91% of the population in a city usually lie and 36% of them usually lie about important matters. If we pick a person at random from this city, what is the probability that the person usually lies about important matters?

Item 16. Two machines M1 and M2 produce balls. Machine M1 produces 40 % and M2 60% of balls. 5% of the balls produced by M1 and 1% of those produced by M2 are defective. We take a ball at random and it is defective. What is the probability that that ball was produced by machine M1?

Item 17. Two black marbles and two white marbles are put in an urn. We pick a white marble from the urn. Then, without putting the white marble in the urn again, we pick a second marble at random from the urn.

1. If the first marble is white, what is the probability that this second marble is white? $P\ (N_2/\ N_1)$

a. 1/ 2
b. 1/ 6
c. 1/ 3
d. 1/ 4

2. If the second marble is white, what is the probability that the first marble is white? $P\ (N_1/\ N_2)$

a. 1/ 3
b. Cannot be computed
c. 1/ 6;
d. 1/ 2

Item 18. An urn contains one blue marble and two red marbles. We pick up two marbles at random, one after the other without replacement. Which of the events below is more likely?

a. Getting two red marbles.
b. The first marble is red and the second is blue
c. The two events a) and b) are equally likely.

LBI QUESTIONNAIRE

Item1. 10 out of every 100 students in a Faculty study mathematics; 30 out of every 100 students doing mathematics share an apartment with other students. Let *S* be the event "sharing the apartment" and M the event the student is doing mathematics course. If we pick a student at random and the student is doing mathematics, the probability that he shares the apartment is:
a. A prior probability *P (S)*
b. A posterior probability *P(S|M)*
c. A likelihood *P(M|S)*
d. A joint probability *P* $(M \cap S)$

Item 2. Imagine you pick 1000 people at random. You know that 10 out of every 1000 people get depression. A depression test is positive for 99 out of every 100 depressed people as well as for 2 out of every 100 non depressed people. Given that D means depression and + means a positive test, compute the following probabilities:
a. P(D)= *b. P(+|D)=* *c. P(-|D)=* *d.* $P(D \cap +)=$

3. The mean value μ for a variable (for example height) in a population:
a. Is a constant in Bayesian inference
b. Is a random variable in classical inference
c. Is a random variable in Bayesian inference
d. Could be constant or variable, depending on the population

Item 4. The prior probability distribution for a parameter:
a. Provides all the information about the population before collecting the data
b. Is computed from the posterior distribution by using the Bayes theorem.
c. It can be used to compute the credible interval for the parameter
d. Is an uniform distribution

Item 5. 1000 young Spanish people were interviewed in a survey. On average they spent 3 hours a week in practicing some sports. In Bayesian inference:
a. 3 hours is a parameter in the population of young Spanish people
b. The average in this population is a random variable; the most likely value is about 3 hours.
c. The average in this population is an unknown constant.
d. Each young Spanish person spends 3 hours a week in doing some sport.

Item 6. In a factory lamps are sold in boxes of four lamps. We have no information about the proportion of defective lamps. Which of the distributions A, B. C or D better describes the prior distribution for the proportion of defective lamps in a box?

(A)		**(B)**		(C)		(D)	
Values of Proportion	Probability	**Values of proportion**	**Probability**	Values of proportion	Probability	Values of Proportion	Probability
0.00	0.1	**0.00**	**0.2**	0.00	0.00	0.00	1/4
0.25	0.1	**0.25**	**0.2**	0.01	0.25	0.25	1/4
0.50	0.1	**0.50**	**0.2**	0.02	0.50	0.50	1/4
0.75	0.1	**0,75**	**0.2**	0.03	0.75	0,75	1/4
1	0.1	**1**	**0.2**	0.04	1	1	1/4

Item 7. In trying to estimate a proportion a student filled three columns in the Bayes table. He got these data:

Values of proportion	Prior Probability	Likelihood	--------	---------
0.0000	0.0000	0.0000		
0.1000	0.1000	0.0000		
0.2000	0.1000	0.0233		
0.3000	0.1000	0.1239		
0.4000	0.1000	0.0682		
0.5000	0.1000	0.0065		
0.6000	0.1000	0.0001		
0.7000	0.1000	0.0000		
0.8000	0.1000	0.0000		
0.9000	0.1000	0.0000		
1.0000	0.1000	0.0000		
Sum			0.0222	

The posterior probability that the true value of proportion in the population is 0.4 would be:
a. 0.00682 b. 0.1000 c. **0.3072** **d.** 0.00015

Item 8. A clinical survey showed a 15% incidence of tobacco addition in young women. A possible prior distribution to approximately describe this proportion is:
a. *Be(15, 100)* *b.* ***Be(15, 85)*** ***c.*** *Be(85,15)* *d. Be(100, 15)*

Item 9. The mean for a *Beta Be(a,b)* distribution is:
a. a/b b. (a+1)/(a+b) c. (a+1)/(b+1) d. **a/(a+b)**

Item 10. In the following table probabilities and critical values for the B(30,40) distribution are given

Probabilities			Critical values	
p_0	$P(0<p<p_0)$	$P(p_0<p<1)$	$P(0<p<p_0)$	p_0
0	0.000	1.000	0.000	0.000
0.05	0.000	1.000	0.005	0.296
0.1	0.000	1.000	0.010	0.304
0.15	0.000	1.000	0.015	0.311
0.2	0.000	1.000	0.020	0.316
0.25	0.001	0.999	0.025	0.320
0.3	0.012	0.988	0.030	0.324
0.35	0.090	0.910	0.035	0.327
0.4	0.318	0.682	0.040	0.330
0.45	0.645	0.355	0.045	0.330
0.5	0.886	0.114	0.050	0.333
0.55	0.979	0.021	0.950	0.526
0.6	0.998	0.002	0.955	0.529
0.65	1.000	0.000	0.960	0.533
0.7	1.000	0.000	0.965	0.536
0.75	1.000	0.000	0.970	0.541
0.8	1.000	0.000	0.975	0.545
0.85	1.000	0.000	0.980	0.551
0.9	1.000	0.000	0.985	0.558
0.95	1.000	0.000	0.990	0.567
1	1.000	0.000	1.000	1.000

The 98 % credible interval for the proportion in a population described by a posterior distribution *Be(30, 40)* is about:

a. ($0.316 < p < 0.551$) b. **($0.304 < p < 0.567$)** **c.** ($0.3 < p < 0.6$) d. ($0.1 < p < 0.9$)

Item 11. The posterior distribution for the proportion of voters favourable to a political party is given by the *Be(30, 40)* distribution. From the above data table, the most reasonable decision is accepting the following hypothesis for the population proportion

a. *H:* $p<0.25$
b. *H:* $p > 0.55$
c. ***H:*** $\boldsymbol{p > 0.25}$
d. *H:* $p > 0.45$

Item 12. For the same posterior distribution of the parameter in a population the r% credible interval for the parameter is:

a. Wider if r increases
b. Wider if the sample size increases
c. Narrower if r increases
d. It depends on the prior distribution

Item 13. In a normal population with standard deviation $\sigma=5$ and with no prior information about the population mean, we pick a random sample of 25 elements and get a sample mean $\bar{x} = 100$. The posterior distribution of the population mean is:

a. A normal distribution *N (100, 0,5)*
b. A normal distribution *N (0,1)*
c. A normal distribution *N (100,5)*
d. **A normal distribution *N (100, 1)***

Item 14. To test the hypothesis that the mean μ in a normal population with standard deviation $\sigma=1$ is larger than 5, we take a random sample of 100 elements. To follow the Bayesian method:

a. We compute the sample mean $\bar{x}$ and then compute $P(\frac{\bar{x}-5}{0,1} < 5)$; when this probability is very small, we accept the hypothesis.

b. We compute the sample mean $\bar{x}$ and then compute $P(\frac{\bar{x}-5}{0,1} < Z)$; when Z is the normal distribution N (0,1); when this probability is very small, we accept the hypothesis.

c. We compute the sample mean $\bar{x}$ and then compute $P(\frac{\bar{x}-5}{0,1} > Z)$ when Z is the normal distribution N (0,1); when this probability is very small, we accept the hypothesis.

d. We compute the sample mean $\bar{x}$ and then compute $P(\frac{\bar{x}-5}{0,1} > 5)$ when this probability is very small, we accept the hypothesis.

Item 15. In a sample of 100 elements from a normal population we got a mean equal to 50. If we assume a prior uniform distribution for the population mean, the posterior distribution for the population mean is:

a. About *N (50, s),* where s is the sample standard estimation.
b. About *N (50, s/10),* where s is the sample standard estimation.
c. We do not know, since we do not know the standard deviation in the population
d. About *N (0,1)*

Item 16. The posterior distribution for a population mean is *N (100, 15)*. We also know that *P (-1.96 < Z < 1.96) =0.95*, where *Z* is the normal distribution *N (0,1)*. The 95% credible interval for the population mean is:

a. (100-1.96 x 1.5; 100 + 1.96 x1.5)

b. (100-1.96; 100+1.96)

c. (100 x 1.5 –1.96; 100 x 1.5 + 1.96)

d. (100-1.96 x 15; 100 +1.96 x 15)

Item 17. In a survey to 100 Spanish girls the following data were obtained:

	Mean	Standard dev.
Sample	160	10
Prior distribution	156	13
Posterior distribution	158.5	7.9

To get the credible interval for the population mean we use:

a. The normal distribution *N(160,10)*

b. The normal distribution *N (156,13)*

c. The normal distribution *N (158.5; 7.9)*

d. The normal distribution *N(160, 0.5)*

Item 18. 20 % of boys and 10% of girls in a kindergarten are immigrant. There are about 60% boys and 40% girls in the center. Use the following table to compute the probability that an immigrant child taken at random is a boy.

Events	Prior probabilities	Likelihoods	Product	Posterior probabilities
Sum	1			1

Item 19. In a geriatric center we want to estimate the proportion of residents with cognitive impairment. 2 out of 10 residents taken at random in the residence showed cognitive impairment. The likelihood for the parameter p=0.1 is 0.1937. What is the meaning of this value?

a. *P (data),* that is, probability of getting this sample.

b. *P (data ∩ p=0.1),* that is, probability of getting the sample and that, in addition, the population proportion is 0,1.

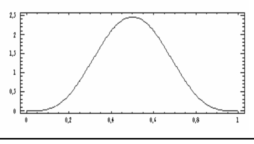
c. *P (p=0.1|data),* that is, probability of a population proportion is 0.1. given the sample

d. *P (data| p=0.1),* that is, given than p= 0.1, probability of getting this sample.

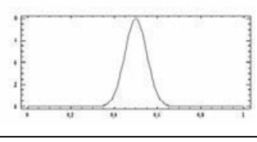
Item 20. Observe the following Beta curves

a) Which of them has a greater spread?

a=5, b=5

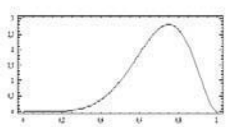


a=50, b=50



b) Which of them predict a greater value of proportion in the population?

a=7, b=3



a=2, b=8