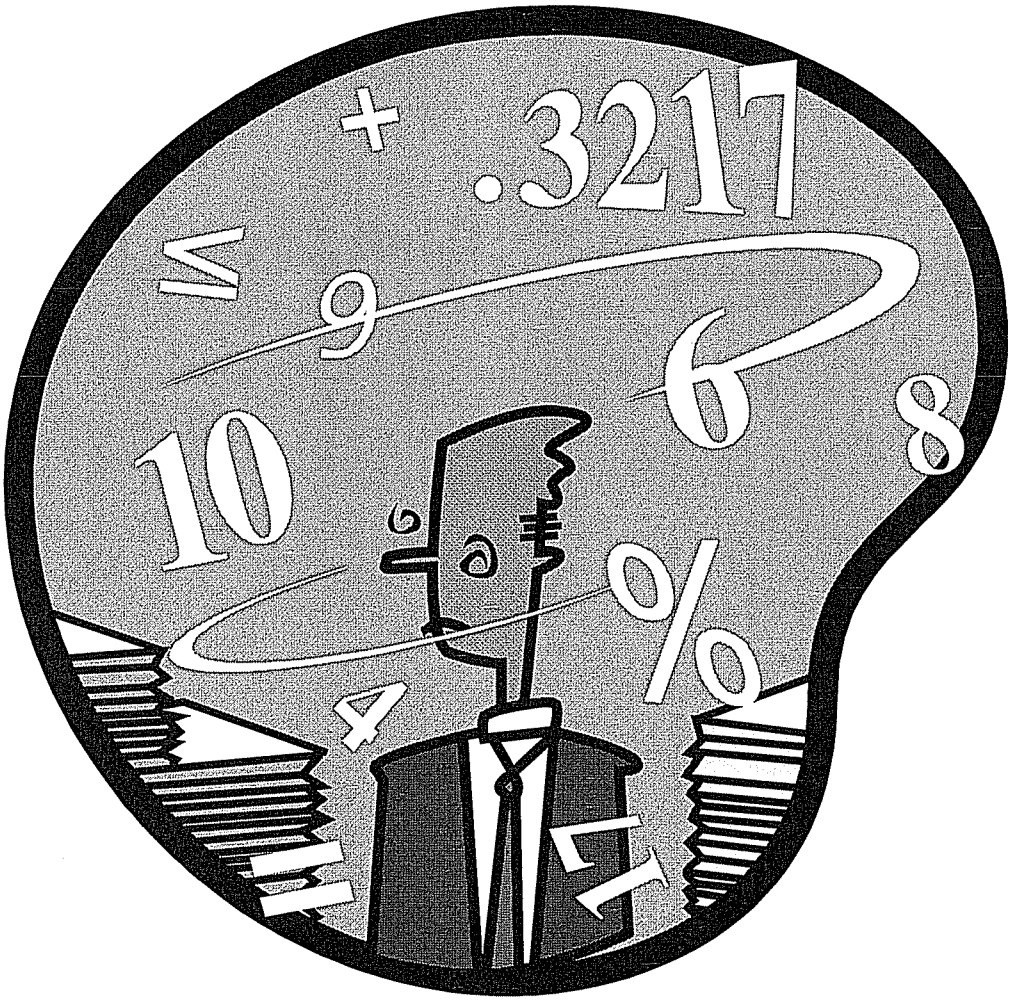
STATS 101/108 WORKSHOP

8-10Qs intest

TEST PREP 1: CHAPTERS 1, 2 AND 3

MONDAY 6 SEPTEMBER, 2010



Students **MUST REGISTER** for all workshops with
The Student Learning Centre, 3rd Floor, Information Commons

Student Learning Centre

Topics we teach and can provide advice on include:

- ✓ Essay writing
- ✓ Computer skills
- ✓ Reading and notetaking
- ✓ Memory and concentration
- ✓ Report writing
- ✓ Test and examination skills
- ✓ Thesis and dissertation writing
- ✓ Tutorial skills
- ✓ Research skills
- ✓ Time and stress management
- ✓ Mathematics
- ✓ **Statistics**
- ✓ Oral presentation and seminar skills
- ✓ Language learning
- ✓ Specific learning disabilities
- ✓ Motivation and goal setting
- ✓ Survival skills (in the University system)

Programmes within SLC include:

- Te Puni Wananga
  Maori university tutors committed to enhancing Maori students' success

- Fale Pasifika
  Pacific Island tutors committed to enhancing success for Pacific Island students

- Students with impairments
  Learning assessments are available for students with specific learning disabilities; academic assistance is available for these students and those with mental health impairments.
  If you have any special learning requirements, please feel free to discuss this with Leila in person or via email.

- Academic English Conversation Groups
  Improve your academic English; develop communication skills including critical/creative thinking and clear expression of ideas and opinions. Weekly class held at the SLC on Thursdays, 3-5pm (during semester)

Statistical help available at the SLC

The Student Learning Centre (SLC) offers help for STATS 101/108 by offering:

- one-on-one tutoring help, and
- a number of workshops

One-on-one help

The SLC employs tutors specifically to help students with one-on-one assistance for STATS 101/108. One-on-one tutoring must be booked at SLC reception on the third floor of the Information Commons in person or by calling 373-7599 X 88850. Enquire at the SLC reception for available times.

Note: SLC tutors are not allowed to help students complete their assignments.

SLC STATS 101/108 Workshops

Any questions regarding STATS 101/108 workshops should be forwarded to:

Leila Boyle
SLC Statistics Co-ordinator
l.boyle@auckland.ac.nz

Workshops are run in a relaxed environment, typically set at a pace for those students that find the Statistics Department's tutorials too fast. All workshops allow plenty of time for questions. In fact, this is encouraged ☺

1) Saturday Workshops

These five 3-hour workshops are held on Saturdays throughout the semester to help students with different sections of the course.

2) Computer Workshops: Excel / PASW (SPSS) x 3

These three computer-based workshops introduce students to the skills needed for Excel and PASW (SPSS) use in STATS 101/108 assignments.

3) Pre-test Workshops

These three workshops will cover the basics that you need for the test.

4) Pre-exam Workshops

These six workshops will cover the basics that you need for the exam.

Note: All workshops concentrate on questions reviewing the basic concepts, rather than questions on finer details. They are designed to assist students to achieve a pass; they are not designed to cover all material.

The timetable for these workshops is available with this handout. Currently the SLC website is still partly down so online enrolments are not available until further notice. Please enrol in each of your preferred classes at the Student Learning Centre by:

- **Going to the SLC in person**
- **Emailing slc@auckland.ac.nz with your name, ID number and the workshop/s you wish to attend.**
- **Calling the SLC reception on 373-7599 ext. 88850 and enrol over the phone. Make sure you know which workshop/s you want to enrol in and have your ID number handy.**

Useful Websites

- SLC webpage: **www.slc.auckland.ac.nz** (The SLC website currently has all functionality except online enrolment! Download an undergraduate brochure and enrol in workshops in person or by emailing/phoning the SLC Reception as per above instructions).
- Cecil: **https://cecil.auckland.ac.nz/**
- **Leila's website for STATS 101/108 SLC workshop handouts & information: www.stat.auckland.ac.nz/~leila**

Revision Notes

Chapter 1 – What is Statistics?

3 or 4 Qs in test

Look at blue pages for good notes and test/exam questions for practice

- **Polls and Surveys**
  - **Target population/Population of interest**
    Complete set of individuals, objects, or units that we want information about.
  - **Study population**
    Complete set of units that might possibly be included in the study. Ideally the same as the target population but often different.
  - **Sampling frame**
    List of units in the study population from which the sample will be drawn.
  - **Sampling design**
    The way the sample is to be chosen from the sampling frame.
  - **Sample**
    Subset of units in the study population which information is collected on.
  - **Census**
    Attempt to sample the whole population.
  - **Variable**
    A characteristic of each unit that we measure.
  - **Parameter** usually unknown eg. μ, p
    Numerical characteristic of the population or distribution.
  - **Statistic** / Estimate, always known
    A number calculated from the data, usually used to estimate an unknown parameter. Sample eg. $\bar{x}, \hat{p}$

- **Randomisation - Obtain representative samples**
  - A representative sample reflects the characteristics in the population.
  - Random sampling
    Technique where each unit is selected entirely by chance.
  - Simple Random Sample (SRS) } usually used for people
    Sampling without replacement (ie no repeats)

- **Errors in Surveys:**

true error = sampling error + non sampling error(s)

1. Sampling errors

- Arise from taking a sample rather than a census, unavoidable.
- Also known as chance or random errors.
- Are bigger in smaller samples than larger ones.
- Size may be estimated by statistical methods. → Ch 8 std. error calc.s

2. Non-sampling errors

- Errors that occur during the data collection process → try to minimise in design of survey by using a pilot survey.
- Can be much larger than sampling errors – always present
- Can be virtually impossible to correct for afterwards
- Virtually impossible to determine how large they can be

  - **Selection bias**
    Population sampled is not exactly the population of interest.
  - **Non-response bias**
    Not everyone in the sample who had been specifically chosen responded. Non-respondents often behave or think differently from respondents.
  - **Self selection**
    People themselves decide whether or not to participate.
  - **Question effects**
    Wording and sentence structure of questions. Even slight differences in question wording can produce measurable differences in how people respond.
  - **Survey format effects**
    Factors such as type of survey (mail/phone/face-to-face interview), question order, layout of written survey, self-administered questionnaire or interviewer, … etc, can affect the results.
  - **Interviewer effects**
    Gender, ethnicity, age of the interviewer, facial expression…etc. Different interviewers asking the same questions may obtain different results.
  - **Behavioural considerations**
    Social desirability of answers.
  - **Transferring findings**
    Using data gathered from one population and using the results to comment on another.

Can be one or other, never both

- **Experiments**
  - Experimenter decides who or what receives which treatment (ideally using some form of random allocation)
  - Randomisation used for treatment allocation.
  - Can prove cause and effect
  - Types of experiments include:
    - ***Completely Randomised Design***
      Allocate treatments to units entirely by chance to try to make the treatment groups as similar as possible.
    - ***Randomised Block Design***
      Group (block) units by some known factor, then randomly allocate treatments to units within each block to try to balance out the unknown factors.
  - ***Control group***
    Group of experimental units given no treatment, a placebo or an existing treatment.
  - ***Placebo*** – inert (inactive) "dummy" treatment
  - ***Placebo effect*** – people show signs of "improvement" when they believe they have taken the real treatment.
  - ***Blinding***
    Prevent people involved in experiment from knowing which experimental subjects have received which treatment.
    - ***Single Blind*** – subjects themselves
    - ***Double Blind*** – subjects and people administering the treatments
- **Observational Studies**
  - CANNOT prove cause and effect – often useful for identifying possible causes of effects, but cannot reliably establish causation.
  - Should use some form of random sampling → representative samples.
  - Unit/person/thing "decides" what treatment they want/get.
    - ***Cross-sectional***:
      A study which observes a group of individuals or units at a point in time. It is a descriptive study, providing a "snapshot" at a particular point in time.
    - ***Longitudinal***:
      A study which observes the same group of individuals or units over a long period of time. Comprised of a series of cross-sectional studies.

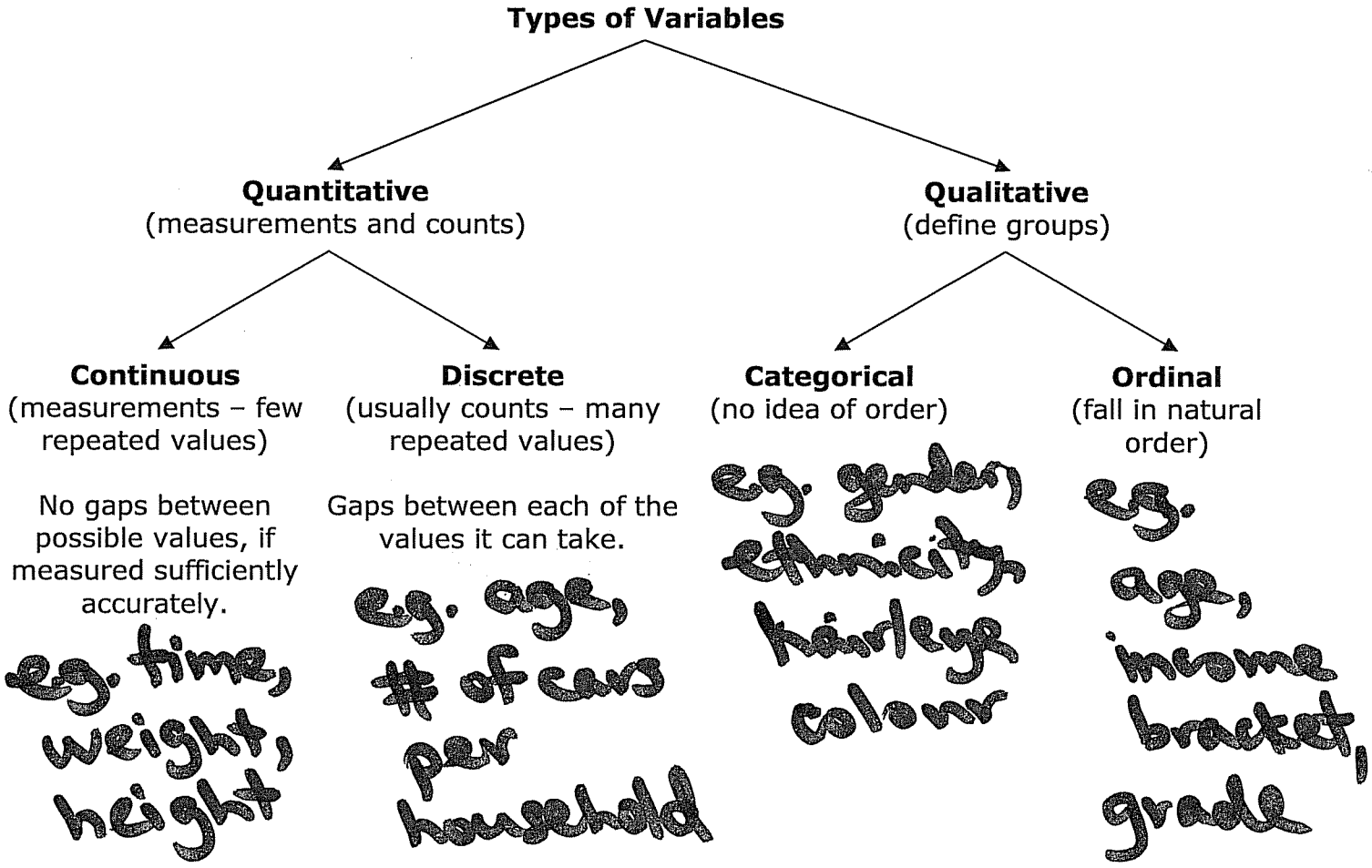
3-5Qs in test

Chapter 2 – Tools for Exploring Univariate Data

Look at blue pages for good notes and test/exam questions for practice

- **Presentation of Data in Tables**
  Guidelines for conveying information quickly and easily:
  - Round drastically
  - Arrange the numbers you want compared in columns, not rows
  - Sort by appropriately chosen column(s)
  - Use row and/or column averages if appropriate

- **Types of Variables**



- **Numerical summaries**

  - **Centre**
    - Sample mean, $\bar{x}$ (also known as the average or expected value) $\sum \frac{x_i}{n}$ – affected by outliers
    - Median (= Med – also known as the 50th percentile) = middle number of the ordered data – not affected by outliers
    - Mode, most frequently occurring number/most common value – not affected by outliers, useful for qualitative data
  - **Spread**
    - Inter-quartile range: IQR = Upper quartile – Lower quartile (middle 50% of data) – not affected by outliers
      - Lower quartile (Q_1) – upper boundary of the lower quarter of the data – not affected by outliers
      - Upper quartile (Q3) – lower boundary of the upper quarter of the data – not affected by outliers
    - Range (maximum – minimum) – affected by outliers
    - Sample standard deviation, σ_{n-1} / s / s_x – affected by outliers
  - **Five number summary** (Min, Q1, Med, Q3, Max)

- **Using your scientific/graphics calculator**

  Do you know how to use your calculator to find the **mean** ($\bar{x}$) and **standard deviation** (σ_{n-1} / s / s_x) of a **sample**? There are three possible types of **sample mean/ standard deviation** problem, can you do all of them?

  - A list of numbers (usually separated by commas, tabs or spaces) ☐ *(tick)*
  - A frequency table with single numbers in the left hand column ☐
  - A frequency table with ranges in the left hand column ☐

- **Outliers / Outside values**
- **Shape / Distributions of data**
    - How many modes/peaks does the data have?
        - Unimodal **OR**
        - Bimodal
    - Is the data symmetric or skewed?
        - Positively/right skewed **OR**
        - Negatively/left skewed **OR**
        - Roughly/approximately symmetric

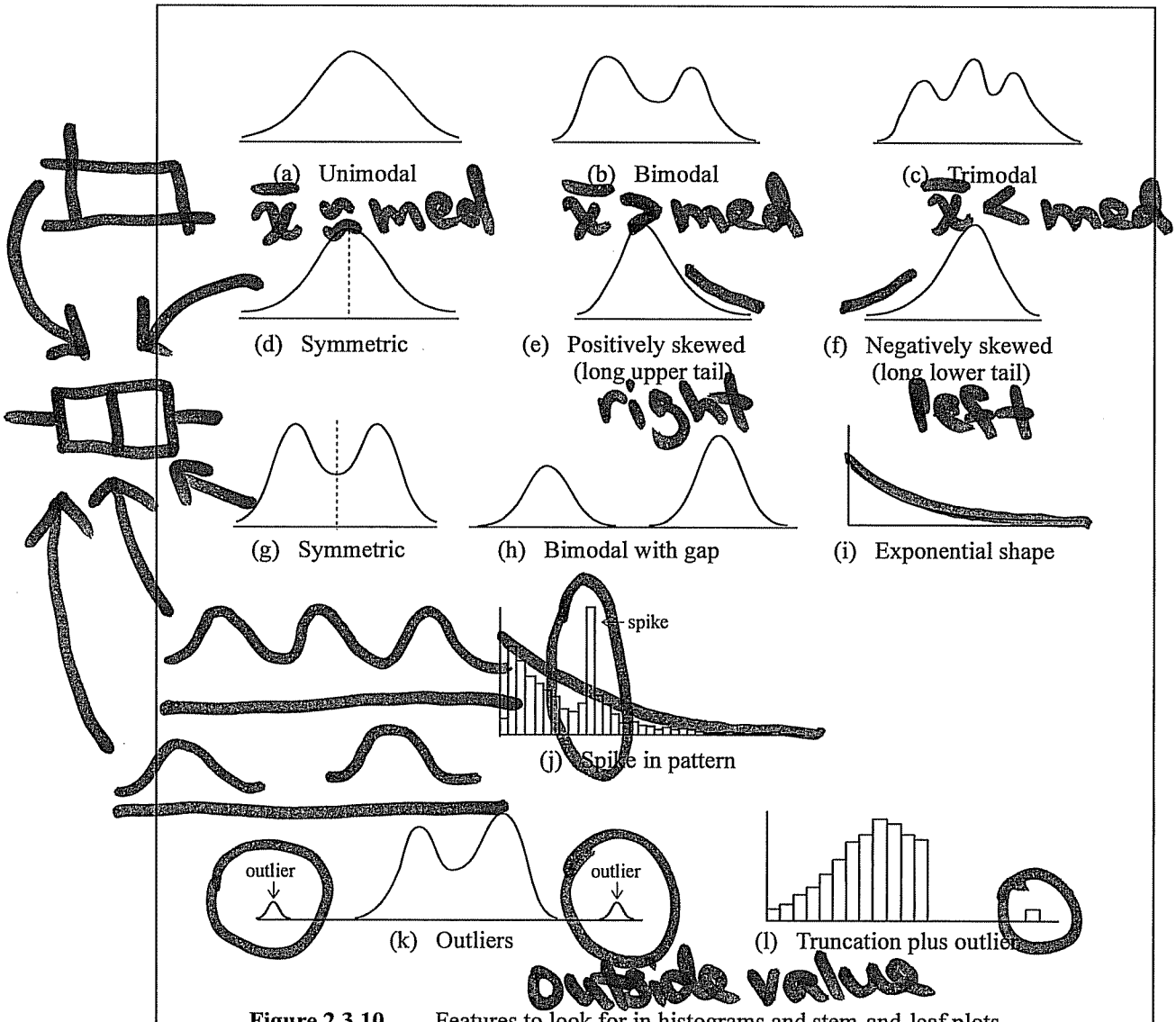


Figure 2.3.10 Features to look for in histograms and stem-and-leaf plots.

From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.

Types of graphs

- 2D vs 3D – Always use 2D!
- Avoid pie graph if you can!
- Order items sensibly.
- Discrete variables:
  - Bar – order categories by size
- Continuous variables:
  - Dot plot – small data sets, $n \leq 20$
  - Stem-and-leaf plot– moderate data sets, $15 \leq n \leq 150$
  - Box plot – moderate to large data sets, $n \geq 30$
  - Histogram – large data sets, $n \geq 50$

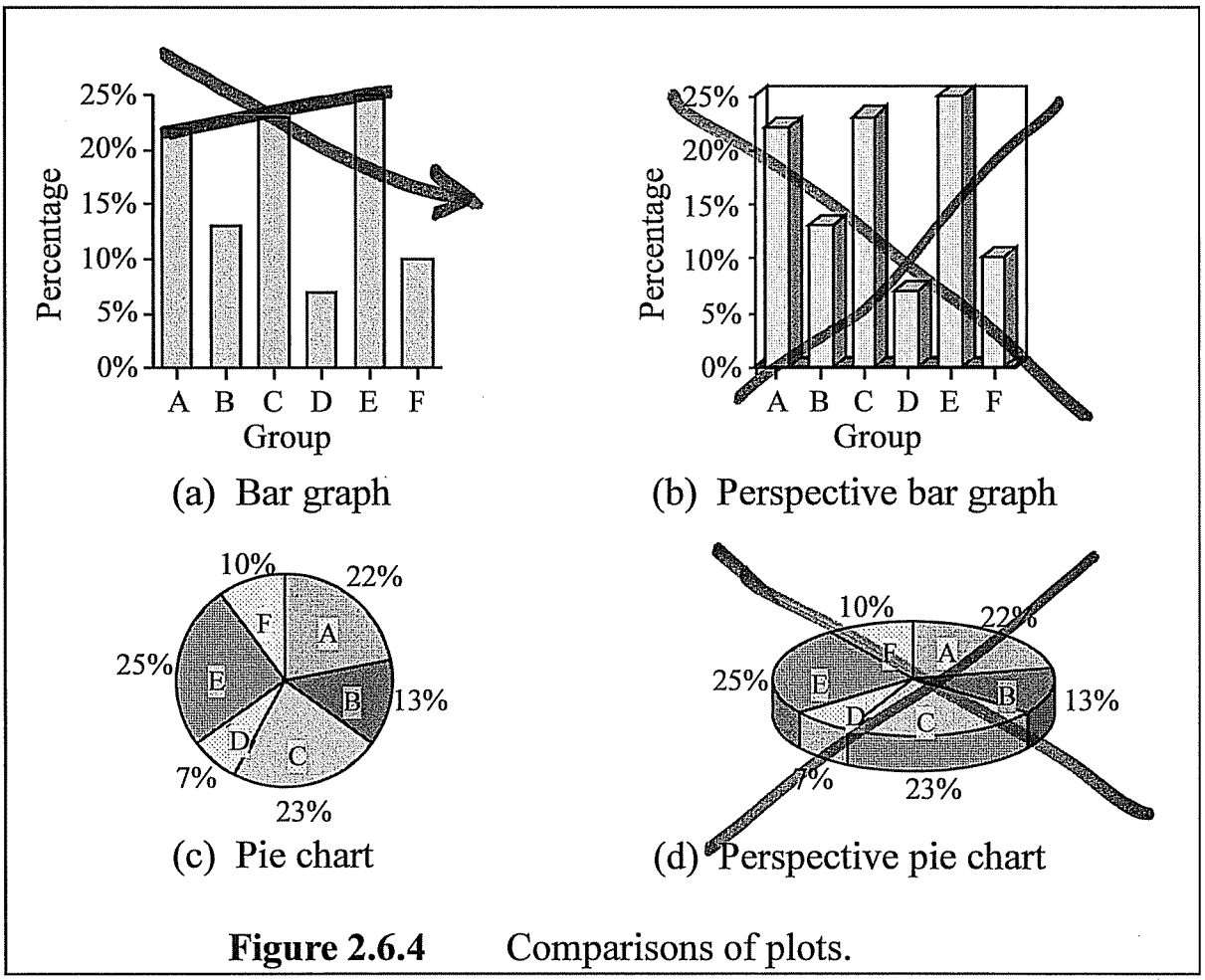


Figure 2.6.4 Comparisons of plots.

From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.

1-40s in test

Chapter 3 – Exploratory Tools for Relationships

Look at blue pages for good notes and test/exam questions for practice

Exploring Relationships between Two Variables

Ch 12 | Ch 10 | Ch 11 (Ch 4)

two quantitative	quantitative vs qualitative	two qualitative
exploratory tool to use to explore relationship		
Scatter plot	Side-by-side plots on the same scale	Two-way table of counts or Bar graphs of proportions
	dot stem & leaf box histogram	
Features to look for:		
▪ Trend (none) - linear - non-linear ▪ Scatter - constant? - little or lots? ▪ Strength of relationship strong or weak? ▪ Association +ve or -ve ▪ Outliers ▪ Groupings	▪ Any group differences: o averages/centres o variability/spread o shapes • symmetric/skewed • modes → not box Details of individual groups: o outliers, gaps, clusters, groupings Think about reasons *why* these differences, similarities and features are seen	▪ Most common and least common combinations ▪ Differences in distributions (e.g. row or column bar graphs)

Chapters 1, 2 & 3 – Questions

Questions 1 to **3** refer to the following information.

Last year an Auckland Statistics student was required to survey Statistics students on transport as part of a Diploma project. One aim of the survey was to test whether changing the wording of a question produced a different level of response. One such question was posed in the following two versions:

Q1 (1) *Taking into account the problems and cost associated with parking, do you think that the cost of catching a bus to university is too high? (Answer: Yes/No/Don't know.)*

Q1 (2) *Do you think that the cost of catching a bus to university is too high? (Answer: Yes/No/Don't know.)*

Two questionnaires were prepared, but each questionnaire only included one of the above versions. The two questionnaires were then distributed to two groups of students - one questionnaire to each group.

1. A pilot survey should have been used:

   (1) Before the full survey to try to determine the precision and bias of the sample estimates.

   (2) Before the full survey to try to get good estimates of the proportions and thereby avoiding the cost of having to do a full survey.

   (3) After the full survey to make sure that it is an experiment rather than an observational study.

   (4) After the full survey to try to determine the precision and bias of the sample estimates.

   (5) Before the full survey to try to identify any problems with the questionnaires.

2. Suppose we decide to conduct a pilot survey involving 10 students; 5 students receiving version (i) and 5 receiving version (ii). The first 5 students are to be selected by taking a random sample without replacement from a class of 75 students. To do this, the names of the 75 students are numbered from 1 to 75. The following numbers from our table of random digits are used to select the students:

   38905 93159 25252 29004 54972 73607 49557 98263

   By starting sampling at the beginning of the above line of random digits and taking consecutive pairs of digits, the first 5 students selected would be numbered:

(1)	38	59	31	59	25
(2)	38	59	31	25	22
(3)	38	15	25	29	00
(4)	38	05	31	59	25
(5)	38	15	25	29	54

3. The two questionnaires are to be distributed to all first and second year Statistics classes. Which one of the following designs will result in the fairest allocation of the two versions?

   (1) Take copies of each version to every class and distribute version (i) to the front half of the lecture room and version (ii) to the back half of the lecture room.

   (2) Make sure that version (i) is distributed to all the first year classes and that version (ii) is distributed to all second year classes.

   (3) Take copies of each version to every class and distribute version (i) to all females in the class and version (ii) to all the males in the class.

   (4) Take copies of each version to every class and distribute version (i) to those students who travel to university by bus and version (ii) to the rest of the students.

   (5) Have the pile of questionnaires alternating (i), (ii), (i), (ii) etc., and distribute the pile one row at a time in each class so that about every second student in each class receives version (ii).

4. The New Zealand Medical Journal (December 1995) has a report on a heart and health investigation that is being carried out by the School of Medicine in conjunction with Fletcher Challenge. The study involves 10,529 New Zealanders. To determine what proportions of the participants were exposed to various risk factors such as smoking, high blood pressure, high cholesterol, etc., each participant was given a medically oriented questionnaire and a physical examination. The aim of the investigation is to follow up what happens to these people in the future. Such an investigation is called a:

   (1) Longitudinal observational study.

   (2) Cross-sectional sample survey.

   (3) Randomised experiment.

   (4) Double-blind study.

   (5) Controlled experiment.

5. Below is a frequency table of the ages of passengers who arrived in Port Chalmers in Otago on the John Whitcliffe on the 23rd of March 1848 (Source: John MacGibbon).

Ages of Passengers

Age	Frequency
0–9	17
10–19	16
20–29	26
30–39	11
40–49	4
50–59	3
60–69	2
Total	79

The sample mean, $\bar{x}$, and the sample standard deviation, s, of this data set are:

(1) $\bar{x} = 22.7, \quad s = 14.56$

(2) $\bar{x} = 22.2, \quad s = 14.65$

(3) $\bar{x} = 22.7, \quad s = 14.65$

(4) $\bar{x} = 23.2, \quad s = 14.56$

(5) $\bar{x} = 22.2, \quad s = 14.56$

6. Which one of the following statements about experiments is false?

(1) If blocking is used in the design of an experiment, then nothing is gained from using randomisation in that design.

(2) Only a well-designed and well-executed experiment can reliably establish causation.

(3) In experiments on people, we should be aware of the possibility of a placebo effect.

(4) An observational study can be useful for identifying possible causes of effects.

(5) Randomisation does not guarantee that the treatment groups are exactly balanced with respect to unknown factors.

7. In March 1998 it was suggested that the city of Hamilton change its name to Waikato City. On the evening of 12 March TV One ran an opinion poll on its Holmes show. Viewers were asked the question "Should Hamilton retain its name or should it be changed to Waikato City?" and were invited to dial an 0900 number at a cost of $1.49 per minute to record their choice. A total of 4103 votes were received. 77% of these indicated that Hamilton should retain its name and 23% indicated the name should be changed to Waikato City.

   The types of non-sampling error likely to cause the most bias in the results of this poll are:

   (1) Behavioural effects and non-response.

   (2) Transferring findings and behavioural effects.

   (3) Selection bias and transferring findings.

   (4) Non-response bias and question effects.

   (5) Self-selection bias and selection bias.

8. A study was carried out to investigate the association between bank interest rates and mortgage interest rates in Canada for the period 1987 to 1989. To explore the relationship between these two types of interest rates we should first look at (choose the most appropriate) a:

   (1) Box plot of bank interest rates and a box plot of mortgage interest rates.

   (2) Histogram of bank interest rates and a histogram of mortgage interest rates, on the same scale.

   (3) Dot plot of bank interest rates and a dot plot of mortgage interest rates.

   (4) Scatter plot of bank interest rates and mortgage interest rates.

   (5) Dot plot of the differences between bank interest rates and mortgage interest rates.

9. In presenting a table to communicate some general features of a set of data, which one of the following statements is true?

   (1) Always order the categories in a table alphabetically.

   (2) Use as little white space as possible; compact tables convey information more easily.

   (3) Averages can be helpful for indicating overall patterns in a table.

   (4) A verbal summary, which highlights the important features of the table, will confuse the reader.

   (5) Don't round the original numbers when they are presented in the table, as the rounded numbers will be misleading.

10. The five number summary for a set of data is (21, 43, 77, 85, 92). Which one of the following statements is true?

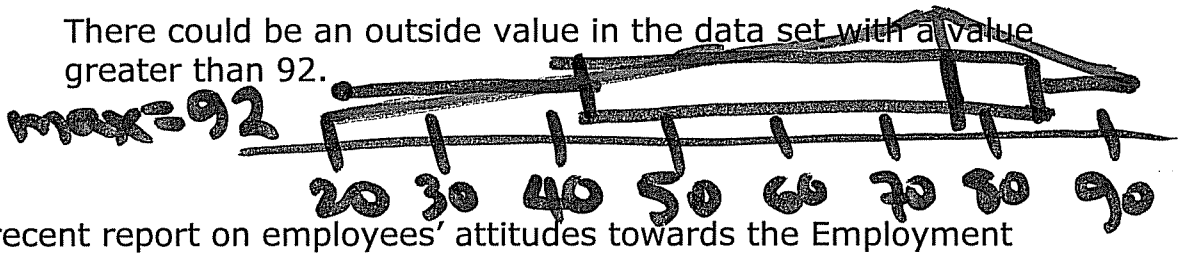
(1) The interquartile range is 42.

(2) There are no observations greater than 85.

(3) The range is 61.

(4) The distribution is symmetric.

(5) There could be an outside value in the data set with a value greater than 92.



11. In a recent report on employees' attitudes towards the Employment Contracts Act it was stated that the mean hourly rate for salary and wage earners is $15.79 per hour. The median hourly rate for salary and wage earners is believed to be less than $10 per hour. Assuming this belief to be true, the best explanation of why the difference between these two rates is so large is (choose one only):

(1) A mistake must have been made in calculating the mean hourly rate.

(2) A relatively large number of wage and salary earners have an extremely high hourly rate.

(3) The sample of wage and salary earners used to determine the mean hourly rate must have been nonrandom.

(4) A relatively small number of wage and salary earners have an extremely high hourly rate.

(5) The distribution of the hourly rates of wage and salary earners is symmetric.

12. Which one of the following statements is false?

(1) The interquartile range is much less sensitive to the presence of outliers than the range.

(2) The range, interquartile range, and sample standard deviation measure the spread of the data.

(3) If a data set is positively (right) skewed, then the sample median is ~~larger~~ smaller than the sample mean.

(4) To help characterize a distribution of data, both a measure of 'centre', and a measure of 'spread' are useful.

(5) A plot is useful for identifying special features of a data set.

Questions 13 and **14** are about the following information.

A class of first year Statistics students was asked to take part in an experiment, which involved each person measuring their pulse. The numbers of beats per minute for the 29 male students are given in the following stem-and-leaf plot.

Pulse rates of Male 1st-year Statistics students

Units: 6|4 = 64 beats per minute

5	4
5	8
6	0 2 2 3 3 4
6	5 6 6 6 7 7 8 8
7	0 0 1 2 2 4 4
7	5 6 6
8	4
8	
9	
9	8 9

13. Which one of the following statements is not a feature of the above data?

(1) Most of the observations lie between 60 and 76 beats per minute.

(2) Ignoring the two largest observations, the data appears to be distributed approximately symmetrically.

(3) There appear to be two outliers.

(4) The distribution appears to be bimodal.

(5) The range is 45 beats per minute.

14. The smallest 8 pulse rates are 54, 58, 60, 62, 62, 63, 63 and 64. The sample mean, $\bar{x}$, and the sample standard deviation, s, of these observations are:

(1) $\bar{x} = 60.75,\ s = 3.33$

(2) $\bar{x} = 60.75,\ s = 3.11$

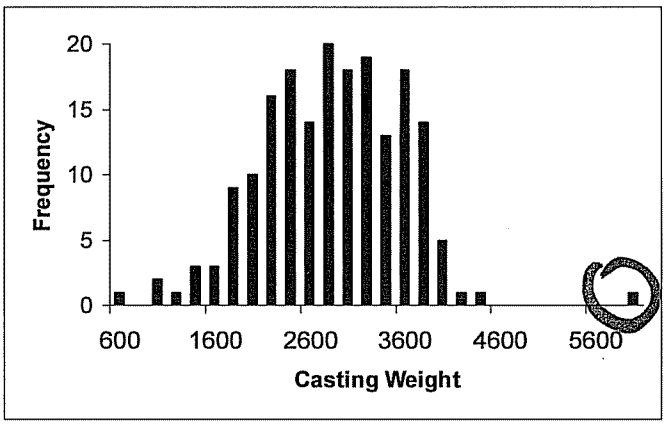
(3) $\bar{x} = 60.75,\ s = 2.33$

(4) $\bar{x} = 62,\ s = 10$

(5) $\bar{x} = 62,\ s = 3.33$

Questions 15 to **17** are about the following information.

15. Recently, investigators became interested in the weights of metal castings for transmission housing. A histogram of the weight of metal castings is given below. The investigators looked at the histogram and decided that the high value of 5990g should be removed from the data set. Which one of the following statements is the best response to their decision?

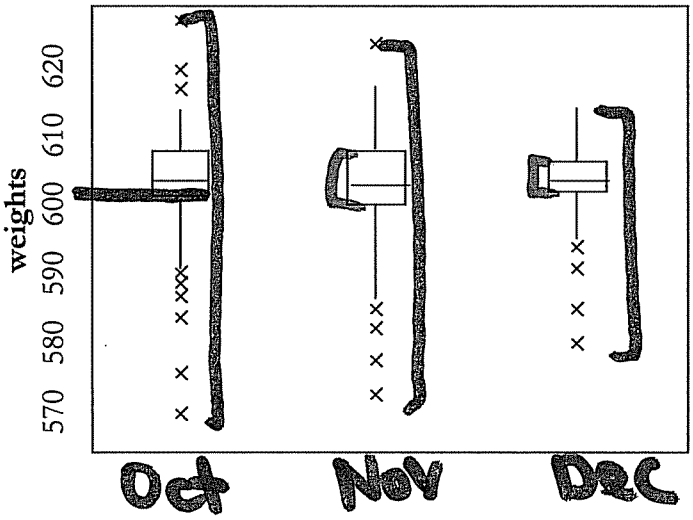


Weight of metal castings in transmission housing

(1) I disagree, you can't remove a data value unless you know that it is an error.
(2) I agree, it is probably an error.
(3) I disagree, data values should never be removed even if they are errors.
(4) I agree, it will cause problems in the data analysis.
(5) I agree, it makes the sample data more Normally distributed.

Box plots for the weights of stacks of wafers sampled in October, November and December 1997 are shown.

Boxplots of Wafer Stack Weights by Month



16. Which one of the following statements is false?

(1) The median weight for November is slightly lower than the median weights for October and December.

(2) November has the largest range.

(3) Approximately 75% of the weights for October are greater than 600g.

(4) December contains the fewest outside values.

(5) December has a smaller interquartile range than November.

17. A batch of 7 stacks of wafers yielded the following values (in grams):

605, 601, 604, 601, 597, 607, 604

The sample mean, $\bar{x}$, and the sample standard deviation, s, for this data set are given by:

(1) $\bar{x} = 602.7\text{g}, \quad s = 3.30\text{g}$

(2) $\bar{x} = 604.0\text{g}, \quad s = 3.06\text{g}$

(3) $\bar{x} = 604.0\text{g}, \quad s = 3.30\text{g}$

(4) $\bar{x} = 602.7\text{g}, \quad s = 10\text{g}$

(5) $\bar{x} = 602.7\text{g}, \quad s = 3.06\text{g}$

Questions 18 and **19** are about the following information.

Information on 5387 school children in Caithness, Scotland, was collected by sending out questionnaires to randomly selected schools. In each school, every child was asked to take the questionnaire home and fill it out. Information gathered on the children included their eye and hair colour, and their favourite pastimes.

18. This study is **best** described as:

(1) a longitudinal study.
(2) a cross-sectional study.
(3) a census.
(4) a case-control study.
(5) an experiment.

19. Which **one** of the following is **not** a potential source of error in this study?

(1) Interviewer effects.
(2) Behavioural considerations.
(3) Nonresponse bias.
(4) Question effects.
(5) Sampling error.

Questions 20 and **21** are about the following information.

Data on 56 hospital births from a single week at the Wellington hospital were collected. Researchers wanted to examine the relationship between various characteristics of the mother and the Apgar Score for the baby. This is a score given to the baby in the first minute after birth and measures the overall physical appearance of the baby.

The characteristics that were measured include:

Age	Age of mother in years.
Mass	Mass of baby in grams.
Gravida	Number of pregnancies including this.
Para	Number of births including this.
Term	Time to delivery in weeks.
Apgar	Apgar score out of 10.

Age	Mass	Gravida	Para	Term	Sex	Apgar
18	3850	1	1	40	M	9
22	2590	1	1	34	M	7
23	3500	1	1	41	M	8
29	2850	2	2	36	F	9
26	3480	3	3	41	M	9
28	3210	2	1	40	F	6
30	3310	8	4	39	F	9
30	4220	3	3	42	M	9
27	4400	2	2	41	M	9
27	2900	1	1	40	F	9

Table: Ten observations from the Apgar Study

20. The main problem with using this data to draw conclusions about Apgar scores in New Zealand is that:

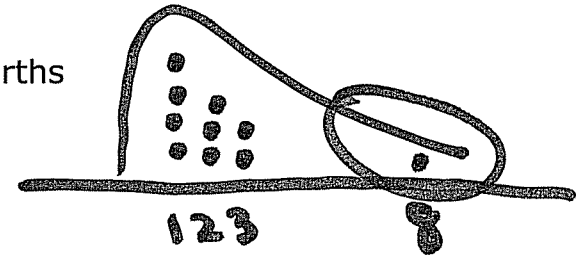
(1) not all pregnancies resulted in births

(2) there was an interviewer effect

(3) there was not control group

(4) there was selection bias

(5) the data contain sensitive personal information



21. Which one of the following numerical summaries is the best summary for the ten observations of Gravida reported in the Table above?

(1) The mean and the range.

(2) The median and the standard deviation.

(3) The mode and the median.

(4) The median, the lower quartile and the upper quartile.

(5) The mean and the standard deviation.

Questions 22 to **25** refer to the following information.

Sports Foundation grants for sports which won the right to represent New Zealand at the Sydney Olympics are shown in the table below.

No.	Sport	1998–1999	1999–2000	2000–2001
01	Archery	$16,500	$15,000	$25,000
02	Athletics	$515,340	$485,400	$299,000
03	Basketball	$133,100	$90,000	$40,000
04	Boxing	$166,950	$55,000	$44,350
05	Cycling	$678,500	$747,182	$688,140
06	Equestrian	$691,000	$717,000	$558,620
07	Gymnastics	$94,500	$34,500	$22,400
08	Hockey	$498,500	$478,460	$554,000
09	Judo	$153,650	$93,179	$124,500
10	Rowing	$533,100	$466,700	$707,265
11	Shooting	$327,000	$106,000	$405,616
12	Softball	$251,913	$425,259	$254,542
13	Swimming	$431,470	$205,000	$280,594
14	Table Tennis	$26,250	$3,000	$29,000
15	Triathlon	$343,110	$548,255	$86,300
16	Weightlifting	$98,900	$48,125	$79,500
17	Wrestling	$13,520	$8,000	$15,000
18	Yachting	$947,000	$1,131,000	$622,356

Table: Sports Foundation Grants

22. Suppose the purpose of this table was to convey the information so that the reader could make visual comparisons between different sports with respect to the size of the grant awarded. One change in the presentation of the data which would **not** be an improvement would be to:

(1) interchange the rows and columns in the table.
(2) round all grants to the nearest thousand dollars.
(3) list the sports in order of the amount of the grant received in the year 2000-2001.
(4) add a column on the right of the table for the 'Average Amount Awarded per Year (1998–2001)'.
(5) add a row at the bottom of the table for the 'Average Amount Awarded per Sport'.

23. The figure below is a dot plot of the grants for each of the 18 sports in the year 2000–2001.

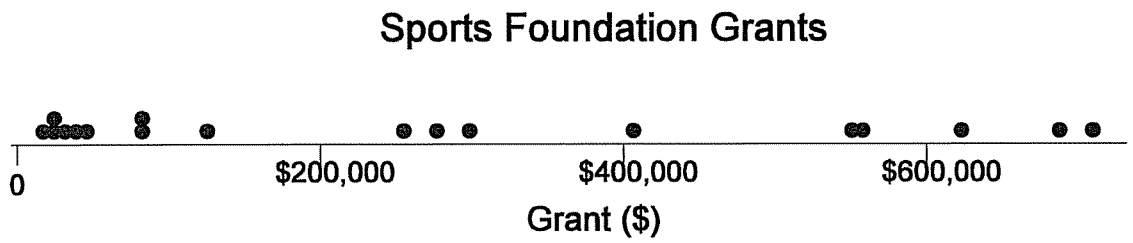


Figure: Sports Foundation Grants, 2000–2001

A **better** graph to highlight the difference between the grants obtained by different sports would be:

(1) side-by-side box plots with the same scaled x-axes.

(2) a labelled bar graph ordered by the size of the grant.

(3) a histogram with equal width class intervals for **Grants** on the x-axis.

(4) a scatter plot with **Grants** as the response variable and **Sport** as the explanatory variable.

(5) a pie chart with the sectors labelled and ordered by the size of the grant.

24. The five sports that received the highest grants in the year 2000–2001 are given in the table below.

Sport	Grant
Rowing	$707,265
Cycling	$688,140
Yachting	$622,356
Equestrian	$558,620
Hockey	$554,000

Table: Highest Five Grants Awarded, 2000–2001

The sample mean, $\bar{x}$, and sample standard deviation, s, for these five sports are:

(1) $\bar{x} = \$626{,}076$, $s = \$71{,}068$

(2) $\bar{x} = \$622{,}356$, $s = \$71{,}068$

(3) $\bar{x} = \$622{,}356$, $s = \$63{,}565$

(4) $\bar{x} = \$622{,}000$, $s = \$71{,}068$

(5) $\bar{x} = \$626{,}076$, $s = \$63{,}565$

25. Suppose we wish to randomly select five of the sports listed in the table on page 22. The method for randomly selecting the five sports uses the number (in the **No.** column of the table on page 22) associated with each sport and random number digits. Use the row of random digits below to select a simple random sample of five sports. You must start at the beginning of the row and use consecutive pairs of digits.

09874 11018 39090 54804 17130

The five sports selected are:

(1) Judo, Rowing, Yachting, Judo, Cycling

(2) Judo, Shooting, Archery, Wrestling, Swimming

(3) Judo, Shooting, Archery, Cycling, Wrestling

(4) Judo, Rowing, Yachting, Cycling, Boxing

(5) Judo, Hockey, Gymnastics, Boxing, Archery

26. Which **one** of the following statements is **false**?

(1) Blinding and double blinding are techniques often used by researchers when people are used as experimental units.

(2) Blocking is used in experiments to ensure fair comparisons with respect to factors the experimenter believes are important.

(3) In an experiment, the control group always receives no treatment.

(4) The placebo effect is the response caused in human subjects by the idea that they are being treated.

(5) Randomisation in experiments allows the calculation of the likely size of sampling errors.

27. Which **one** of the following statements is **false**?

(1) A well designed, and well executed experiment can reliably establish causation.

(2) Random sampling errors always have an identifiable cause.

(3) Experiments on humans should be double blind, if possible.

(4) Non-response in surveys can cause bias because non-respondents often tend to behave differently from people who do respond.

(5) Observational studies are not reliable for proving causation.

ANSWERS

1. **(5)**	2. **(2)**	3. **(5)**	4. **(1)**	5. **(3)**	6. **(1)**	7. **(5)**
8. **(4)**	9. **(3)**	10. **(1)**	11. **(4)**	12. **(3)**	13. **(4)**	14. **(1)**
15. **(1)**	16. **(2)**	17. **(1)**	18. **(2)**	19. **(1)**	20. **(4)**	21. **(4)**
22. **(1)**	23. **(2)**	24. **(1)**	25. **(4)**	26. **(3)**	27. **(2)**	