Ch 8-10: Ch 8: 3-6 Qs, Ch 9: 2-9 Qs, Ch 10: 7-10 Qs

17-21 Qs

STATS 10x WORKSHOP

EXAM PREP ~~3~~4: CHAPTERS 8 ~ 10 MEANS

SAT 16 & TUE 26 OCT 2010

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

Statistical help available at the SLC

The Student Learning Centre (SLC) offers help for STATS 10x by offering:

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

One-on-one help over S2 2010 including exam period

One-on-one assistance for STATS 10x is available at the SLC. Check appointment availability and book at SLC reception in person (third floor, Information Commons building) or by calling 373-7599 ext. 88850.

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

SLC STATS 10x Exam Prep Workshops

Any questions regarding STATS 10x workshops should be forwarded to:

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

These twelve workshops (six different sessions, each repeated twice) are held prior to the exam, from Saturday 2 October until Monday 1 November 2010 (inclusive).

These workshops concentrate on questions reviewing the **basic concepts**, rather than questions on finer details. They are designed to assist students to achieve a pass and **don't cover all material**.

The timetable for these workshops is available at this workshop, at SLC Reception and on Leila's website. Please enrol in each of your preferred workshops by EITHER:

- ***Dropping by*** **the SLC Reception to enrol in person (Room 320, Level 3, Information Commons Building, 11 Symonds Street) OR**
- ***Emailing*** **slc@auckland.ac.nz with your name, ID number, and the name, date and time of the workshop/s you wish to attend OR**
- ***Calling*** **the SLC Reception on 373-7599 ext. 88850 and book over the phone.**

Useful Websites

- SLC webpage: **www.slc.auckland.ac.nz**
- Cecil: **https://cecil.auckland.ac.nz**
- **Leila's website for STATS 10x SLC workshop handouts & information: www.stat.auckland.ac.nz/~leila**

Recall: **Chapter 8 – Confidence Intervals**

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

- A **confidence interval** gives a range of plausible values for the parameter of interest that is consistent with the data (at the specified level of confidence). It determines the **size** of the effect or difference.
- You can do all kind of CI's, 90%, 95%, 99%...
- Increasing the confidence level will **increase** the width of the interval.
- Increasing the sample size will make the confidence interval more precise. narrower

  halve the width
- To double the accuracy of the confidence interval we **need 4 times** as many observations.

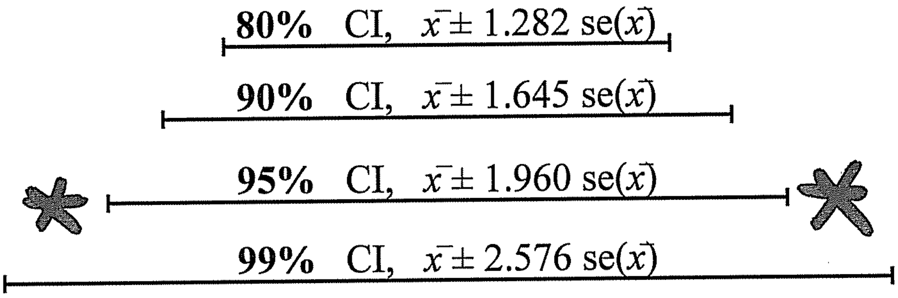


Figure 8.1.3 The greater the confidence level, the wider the interval

- To triple the accuracy of the confidence interval we **need 9 times** as many observations. third the width
- 95% confidence interval
  - ✓ Range of plausible values for the parameter of interest that contains the **true value** of our parameter of interest for 95% of samples taken.
  - ✓ 5% of samples taken will not have the parameter within the calculated confidence interval.
  - ✓ We do not know if the sample we have taken is one of the 95% that contains the true unknown parameter. All we can say is that 95% of the time it will.
  - ✓ If you take 1000 samples, based on the same sampling protocol, then you can expect approximately 950 of these samples will contain the true value (e.g. true mean, true proportion) of the population.

Step-by-Step Guide to Producing a Confidence Interval by Hand

1. State the **parameter** to be estimated (symbol and words).
   Is it μ, ~~p~~, $\mu_1 - \mu_2$ ~~or $p_1 - p_2$~~?
   1 3
2. State the **estimate** and its value.
3. Write down the **formula** for a CI, from the Formula Sheet — $\boxed{\textit{estimate} \pm \underbrace{t \times \text{se}(\textit{estimate})}_{\text{margin of error}}}$,
4. Find the appropriate **standard error** from the Formula Sheet.
5. Find the appropriate ***df***.
6. Find the appropriate ***t*-value.** This is called the '*t*-multiplier'.
   Need to know the confidence level and *df*.
7. Calculate the **confidence limits**, the end points of the confidence interval.
8. Interpret the interval using plain English.
   Use the confidence limits to construct an answer to the original question in plain English.

- There are four different types of problem covered in Chapter 8:
  1. Single mean ~~2. Single proportion~~. 3. Difference between two means
  ~~4. Difference between two proportions~~

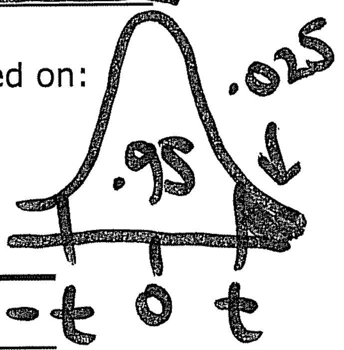
first bit

- The **estimate** is based on the **parameter** of interest we are investigating:

Parameter	Estimate
1. Single mean $\boldsymbol{\mu}$:	*estimate* = $\bar{x}$
3. Difference between two means $\boldsymbol{\mu_1 - \mu_2}$: (independent samples)	*estimate* = $\bar{x}_1 - \bar{x}_2$

second bit

- The ***t*-multiplier** (means) / ~~**z-multiplier** (proportions)~~ is based on:
  ✓ Whether we are investigating means ~~or proportions~~
  ✓ The desired level of confidence 95%
  ✓ The degrees of freedom



Estimate	Degrees of Freedom
1. *estimate* = $\bar{x}$	$df = n - 1$
3. *estimate* = $\bar{x}_1 - \bar{x}_2$	$df = minimum(n_1 - 1, n_2 - 1)$

df	.025
$n-1$	(t)
∞	1.960

- The **standard error** is based on the estimate, the number of samples and sample size(s).

Estimate	**se(estimate)**
1. *estimate* $= \overline{x}$	$se(\overline{x}) = \dfrac{s}{\sqrt{n}}$
3. *estimate* $= \overline{x}_1 - \overline{x}_2$	$se(\overline{x}_1 - \overline{x}_2) = \sqrt{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)}$

- CI's for the difference between two means/[redacted]:
  - $(-7, 13)$ If the CI contains 0 (i.e. one negative and one positive number), there may be no difference between the two means/[redacted].
  - $(7, 13)$ If CI is positive, then μ_1/[redacted] is higher/larger than μ_2/[redacted]. $\mu_1 > \mu_2$
  - $(-13, -7)$ If CI is negative, then μ_1/[redacted] is lower/smaller than μ_2/[redacted] $\mu_1 < \mu_2$

Chapter Nine

In hypothesis testing and confidence intervals we use **sample** data to make inferences (draw conclusions) about **population** parameters.

- A **confidence interval** gives a range of plausible values for the parameter of interest that is consistent with the data (at the specified level of confidence).
- A **significance test,** tests one possible value for the parameter, called the **hypothesised** value. We determine the strength of evidence provided by the data against the null hypothesis, H_0.

A significance test determines the **strength** of the evidence **against** the **hypothesised** value, while a confidence interval determines the **size** of the effect or difference.

Significance testing is a method used to deal with the **uncertainty** about the true value of a parameter caused by the sampling variation in estimates.

$\boldsymbol{H_0}$ denotes the **null hypothesis** whereas $\boldsymbol{H_1}$, denotes the **alternative hypothesis**.

The null hypothesis, H_0

✓ It is our best guess as to what we think the parameter of interest is – a single plausible value.

✓ General form: H_0: *parameter = hypothesised value (some number)*

✓ The hypothesised value is **not** the parameter of interest. Remember that the parameter of interest is an unknown quantity.

The alternative hypothesis, H_1

✓ Specifies the type of departure from H_0 that we expect to detect.

✓ Corresponds to the research hypothesis.

✓ There are three different types:

- H_1: *parameter ≠ hypothesised value (some number)* } 2-sided
- H_1: *parameter < hypothesised value (some number)* } 1-sided
- H_1: *parameter > hypothesised value (some number)*

✓ ONLY use a one-sided alternative when you have prior information or a theory.

The t-test statistic

- ✓ Is the number of standard errors our estimate is from the hypothesised value.
- ✓ We calculate it using: $\boxed{t_0 = \dfrac{estimate - hypothesised\ value}{std\ error}}$

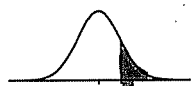
on formulae sheet

The P-value

- ✓ Is the probability that sampling variation would produce an estimate that is further away from the hypothesised value than the estimate we obtained from our data, assuming that the null hypothesis is true.
- ✓ It is a measure of evidence **against** H_0. memorise!
- ✓ We calculate the P-value using the t-test statistics and the appropriate Student's t distribution.
- ✓ We interpret the size of the P-value using the following table:

	P-value	Evidence against H_0
>12%	> 0.12	None
≈10%	≈ 0.10	Weak
≈5%	≈ 0.05	Some
≈1%	≈ 0.01	Strong
<.1%	< 0.001	Very Strong

memorise

Alternative hypothesis	P-value = area of shaded region $T \sim Student(df)$
H_1: parameter ≠ hypothesised value (2-sided)	 2-tailed test
H_1: parameter **>** hypothesised value (1-sided)	 1-tailed test
H_1: parameter **<** hypothesised value (1-sided)	 1-tailed test

Step-by-Step Guide to Performing a *t*-test by Hand

some # from story

1. State the **parameter**, being discussed (symbol/s and words).
   Is it μ, [redacted] $\mu_1 - \mu_2$, [redacted]
   1. 3.

   1. $H_0: \mu = \mu_0$

2. State the **null hypothesis**, $\boldsymbol{H_0}$.

   i.e. H_0: parameter = *hyp. val.*

   3. $H_0: \mu_1 - \mu_2 = 0$

3. State the **alternative hypothesis**, $\boldsymbol{H_1}$.

   i.e. H_1: parameter ≠ *hyp. val.*
   OR H_1: parameter < *hyp. val.*
   OR H_1: parameter > *hyp. val.*

4. State the **estimate**, and its value.

   1. $\bar{x} = \ldots$
   3. $\bar{x}_1 - \bar{x}_2 = \ldots$

   get from story

5. Calculate the ***t*-test statistic**:

   - Use: $\boxed{t_0 = \dfrac{\textit{estimate} - \textit{hypothesised value}}{\textit{std error}}}$ (from the Formula Sheet)

     Steps 2&3

     Write down the estimate and the hypothesised value.

     Find the appropriate standard error and the *df* using the Formula Sheet.

     1. $n-1$ 3. $\min(n_1-1, n_2-1)$

   - Calculate t_0.

6. ~~Calculate the ***P*-value**.~~ See *Finding P-values*, pages 9-12, *Excel* Tutorial, Section F – Lecture Workbook

7. **Interpret** the *P-value*. See *Interpreting the P-value*, page 11, Chapter 9 Notes, Section B – Lecture Workbook

8. Calculate the **confidence interval**. $\boxed{\text{est} \pm t \times se(\text{est})}$

9. **Interpret** the confidence interval using plain English.

step 4 "look up" step 5

Statistical significance

- ✓ Relates to having evidence of the **existence** of an effect or difference.
- ✓ Determined by examining the ***P*-value** of your significance test.

stat sig @ 1% level if p-val ≤ .01 (1%)

Practical significance

- ✓ Depends on the **size** of the effect or difference.
- ✓ Determined by examining the **confidence interval** in relation to the research in question.

Chapter Ten

In this chapter we dealt with:

special case

- 1. One sample
- Paired data comparisons
- 3. Two independent samples
- More than 2 samples
- Parametric vs Nonparametric issues
- Experimental design

Non-parametric tests

✓ Used when there are strong concerns about the "normality" of the data. You should make sure data don't show separation into clusters or have a multi-modal nature and then apply the 15-40 rule as follows:

Sample Size Guidelines – "15 – 40 Rule"

Small (total $n \leq 15$ or so)	Medium ($15 <$ total $n < 40$)	Large (total $n \geq 40$ or so)
no outliers	no outliers	no gross outliers
at most, slight skewness	not strongly skewed	data may be strongly skewed

✓ They test the **median(s)**, $\tilde{\mu}$, NOT the mean(s).

1. 1-sample *t*-test

✓ 1 sample (group) of data

✓ Parameter = μ

✓ **Hypotheses**: $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$

Some # from story

✓ **Assumptions** for 1-sample *t*-tests

1. **Random sample → Independence** of observations – *CRITICAL!*
2. **Normality** – quite robust, but sensitive to outliers in small-medium samples. No clusters or multi-modes allowed.

✓ Data **not** Normal? Use **Sign Test** *OR* **Wilcoxon Signed-Rank Test**.

- The **Sign Test** is a nonparametric equivalent of the one sample *t*-test.
- Observations above the hypothesised value are given a positive sign (**+**).
- Observations below the hypothesised value are given a negative sign (**–**).
- Observations with the same value as the hypothesised value are ignored.
- Evidence against the null hypothesis is provided by a large imbalance in the number of ' **+** ' and ' **–** ' signs.

$H_0: \tilde{\mu} = \mu_0$

$H_1: \tilde{\mu} \neq \mu_0$

1a Paired data (special case of single sample)

✓ 1 sample (group) of data

✓ Two measurements taken on each experimental unit

✓ With related or paired data we analyse the <u>differences</u> and use a 1-sample *t*-test, i.e. we treat the differences as a single sample.

✓ Parameter = μ_{diff}

✓ **Hypotheses**: H_0: $\mu_{\text{diff}} = 0$

vs H_1: $\mu_{\text{diff}} \neq 0$

✓ **Assumptions** for paired-*t*-tests

1. **Random sample → Independence** of <u>pairs of observations</u> – *CRITICAL!*
2. **Normality** – quite robust, but sensitive to outliers in small-medium samples. No clusters or multi-modes allowed.

✓ Data **not** Normal? Use **Sign Test** *OR* **Wilcoxon Signed-Rank Test.**

✓ **Note:** You don't need to know how to do **Sign** or **Wilcoxon Signed-Rank** tests by hand – just remember the names and be able to interpret the **Sign** test's *P-value* from SPSS computer output.

3. 2-sample *t*-tests

✓ Two independent samples

✓ Parameter = $\mu_1 - \mu_2$

✓ **Hypotheses**: H_0: $\mu_1 - \mu_2 = 0$

vs H_1: $\mu_1 - \mu_2 \neq 0$

✓ **Assumptions** for 2-sample *t*-tests

1. **Independence** of <u>samples</u> – *CRITICAL!*
   <u>AND</u>
   Random samples → Independence of <u>observations</u> – *CRITICAL!*
2. **Normality** – even more robust against non-Normality than one sample *t*-procedures. No clusters or multi-modes allowed.

✓ Data **not** Normal? Use **Mann Whitney Test** *OR* **Wilcoxon Rank-Sum Test**.

✓ **Note:** You don't need to know how to do **Mann Whitney** or **Wilcoxon Rank-Sum** tests by hand – just remember the names.

3 or more samples ⇒ *F*-test for 1-way ANOVA

ANalysis Of VAriance

✓ Three or more independent samples

✓ **Hypotheses:** H_0 : all the underlying population means are the same

i.e. $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$ ← # of groups

vs H_1 : at least 2 of the underlying population means are not the same

Note: SHOULD NOT be written using symbols

✓ *F*-statistic $\boxed{f_0 = \frac{s_B^2}{s_W^2}}$ where s_B^2 = the **between** group variation and

s_W^2 = the **within** group variation

✓ **Assumptions** for 1-way ANOVA

1. **Independence** of samples – *CRITICAL!*
   AND
   Random samples → Independence of observations – *CRITICAL!*
   Check by thinking about the way the data was obtained.
2. **Normality**
   Check by looking at plots of each sample and consider sample sizes. Plots should be unimodal and not too skewed for the n_{tot} you have.
3. **Equality of Standard Deviations**/variance/spread/variability
   Check by using the ratio (fraction) $\frac{\text{largest sd}}{\text{smallest sd}} < 2$ as a guide

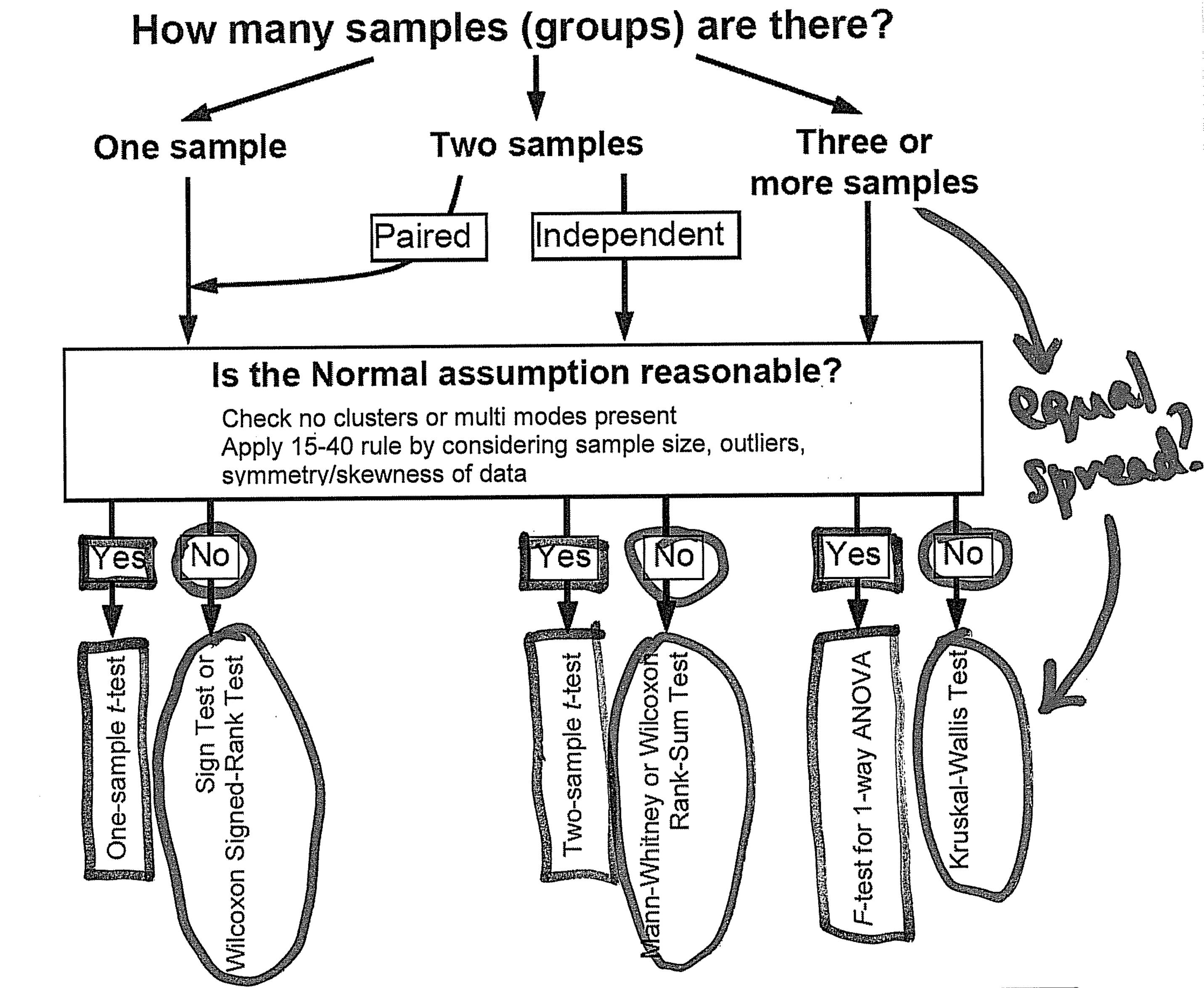
✓ **One-way ANOVA Table** k = # of samples

	Sum of Squares (SS)	*df*	Mean Sum of Squares (MSS)	*F*-statistic	*P*-value
Between Groups (BG)	BGSS	$\boxed{df_1 = k - 1}$	$MSS_B = s_B^2$	$\boxed{f_0 = \frac{s_B^2}{s_W^2}}$	$\Pr(F \geq f_0)$
Within Groups (WG)	WGSS	$\boxed{df_2 = n_{\text{tot}} - k}$	$MSS_W = s_W^2$		
Total (T)	TSS	$df_{\text{tot}} = n_{\text{tot}} - 1$			

n_{tot} = total # of obs

✓ Data **not** Normal or standard deviations **not** equal? Use **Kruskal-Wallis Test**.

✓ **Note:** You don't need to know how to do **Kruskal-Wallis** test by hand – just remember the name.



Assumptions	Checks
1. Independence – All tests - Single sample assumes indep. between observations. - Paired data assumes indep. between pairs of observations. - 2 or more samples assume independence between observations and between samples.	- The design of the experiment
2. Normality – Parametric tests only - one-sample *t*-test - two-sample *t*-test - *F*-test	- Plot the data and apply 15-40 rule
3. Equal Spread – *F*-test only	- Plot the data - Check the standard deviation ratio: $\frac{\text{largest sd}}{\text{smallest sd}} < 2$

Chapters 8~10 – Questions

1. Suppose the hypothesis test H_0: μ = 100 versus H_1: $\mu \neq$ 100 obtained a *P-value* = 0.001. Which of the following statements is **true**?

   (1) The *P-value* is very small, therefore H_0 is probably false.

   (2) We would reject H_0 at the 1% level of significance.

   (3) A 95% confidence interval for μ does not ~~contains~~ contain the value 100.

   (4) A 99% confidence interval for μ contains the hypothesised value.

   (5) We will accept that H_0 is true.

95% CI : hypval → in CI ⟷ p-val > 0.05
→ out of CI ⟷ p-val < .05

2. In order to study the harmful effects of DDT poisoning, the pesticide was fed to 6 randomly chosen rats out of a group of 12 rats. The other 6 unpoisoned rats comprised of the control group. The following data gives measurements of the amount of tremor detected in the bodies of each rat after the experiment. The more tremor, the more harmful.

   Poisoned group: 12.207, 16.869, 25.050, 22.429, 8.456, 20.589
   Control group: 11.074, 12.064, 9.351, 6.642, 9.686, 8.182

   We wish to test 2 indep samples

(1)	$H_0 : \mu_{diff} = 0$	versus	$H_1 : \mu_{diff} \neq 0$
(2)	$H_0 : \mu_1 - \mu_2 = 0$	versus	$H_1 : \mu_1 - \mu_2 \neq 0$
(3)	$H_0 : \bar{x}_1 - \bar{x}_2 = 0$	versus	$H_1 : \bar{x}_1 - \bar{x}_2 \neq 0$
(4)	$H_0 : p_1 - p_2 = 0$	versus	$H_1 : p_1 - p_2 \neq 0$
(5)	$H_0 : \bar{x}_{Diff} = 0$	versus	$H_1 : \bar{x}_{Diff} \neq 0$

3. Which one of the following statements about the one-way analysis of variance *F*-test is **false**?

   (1) T The evidence of differences between the true group means comes from comparing the variability between group means with the variability within the groups. $f_0 = s_B^2 / s_W^2$

   (2) T It should only be used when comparing independent samples.

   (3) T It provides partial protection against multiple comparisons.

   (4) T The null hypothesis is that all the true group means are the same.

   (5) It is ~~not~~ maybe badly affected by the presence of only one or two outliers. (equal sd assumption)

4. Does too much sleep impair intellectual performance? Taub *et al.* (1971) examined this commonly held belief by comparing the performance of 12 subjects on the mornings following (1) two normal nights' sleep and (2) two nights of "extended sleep". In the morning they were given a number of tests of ability to think quickly and clearly. One test was for vigilance where the lower the score, the more vigilant the subject. The following data was collected:

Subject	1	2	3	4	5	6	7	8	9	10	11	12
Normal Sleep	8	9	14	4	12	11	3	26	3	11	10	1
Extended Sleep	8	9	15	2	21	16	9	38	10	11	16	41

To see if the data supports the view that too much sleep can be bad for you, we would test which of the following hypotheses?

paired data

(1) $H_0 : \bar{x}_1 - \bar{x}_2 = 0$ versus $H_1 : \bar{x}_1 - \bar{x}_2 < 0$

(2) $H_0 : \bar{x}_{Diff} = 0$ versus $H_1 : \bar{x}_{Diff} \neq 0$

(3) $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 - \mu_2 \neq 0$

(4) $H_0 : p_1 - p_2 = 0$ versus $H_1 : p_1 - p_2 \neq 0$

(5) $H_0 : \mu_{diff} = 0$ versus $H_1 : \mu_{diff} \neq 0$

5. Suppose that a 95% confidence interval for the difference in true mean HOSP.RATE level between the *small cars* and *medium cars*, $\mu_{Small} - \mu_{Medium}$, is given by (-0.05, 0.9). Which one of the following statements is **true**?

$H_0: \mu_1 - \mu_2 = 0$

(1) F There is not a significant difference between the true means at the 5% level.

(2) F There is not a significant difference between the ~~sample~~ true means at the 5% level.

(3) F It is likely that mean HOSP.RATE for *small cars* is much smaller than the mean HOSP.RATE for *medium cars*.

(4) With 95% confidence the true mean HOSP.RATE for *small cars* is somewhere between 0.05 units smaller and 0.9 units bigger than the mean HOSP.RATE for *medium cars*.

(5) F The difference between the ~~sample~~ true means will be outside ~~this~~ a 95% interval 5% of the time.

6. Analysis of variance (ANOVA) is (select **one** only):

(1) an overall test of no difference between sample variances.

(2) an *F*-test of no difference between population means.

(3) an overall test of no difference between population variances.

(4) an *F*-test for the equality of population variances.

(5) an *F*-test of no difference between sample means.

7. Which one of the following statements about significance tests is false?

(1) Formal tests can help determine whether effects we see in our data may just be due to sampling error.

(2) A test statistic is a measure of discrepancy between what we see in our data and what we would expect to see if H_0 was true.

(3) The *P-value* says nothing about the size of an effect.

(4) The data should be carefully examined in order to determine whether the alternative hypothesis needs to be one-sided or two-sided.

(5) The *P-value* describes the strength of evidence against the null hypothesis.

8. Which one of the following statements is **true**?

F (1) A two-sided test of H_0: *parameter = hypothesised value* has *P-value* ~~less~~ greater than 0.05 if the *hypothesised value* lies within a 95% confidence interval for the *parameter*.

F (2) The larger the *P-value,* the ~~stronger~~ weaker the evidence against H_0.

F (3) The larger the test statistic, $|t_0|$, for a two-sided test, the ~~larger~~ smaller the *P-value* will be.

(4) Tests of hypotheses can only deal with random errors and sampling variation. They are ineffective when confronted with data that has systematic bias.

F (5) An extremely small *P-value* means that the actual effect differs markedly from that claimed in the null hypothesis.

9. Which one of the following statements about hypothesis testing is **false**?

(1) The larger the *P-value,* the stronger the evidence against the null hypothesis.

(2) The *P-value* is the probability that, if the null hypothesis were true, sampling variation would produce an estimate that is further away from the hypothesised value than our data estimate.

(3) We cannot establish an hypothesised value for a parameter, we can only determine whether there is evidence to reject a hypothesised value.

(4) H_0 is typically a sceptical reaction to a research hypothesis.

(5) The *P-value* measures the strength of evidence against the null hypothesis.

10. Before conducting formal tests, one should look at plots of the data. Which **one** of the following statements is **false**?

(1) Plots may highlight strange or interesting features of the data which cannot be seen in a formal test.

(2) Summaries of the important features of the data can often be obtained from looking at plots.

(3) Plots are used to check the validity of the assumptions for the formal tests.

(4) Inferences, i.e. conclusions about the population, drawn from plots do not need to be verified by formal tests.

(5) Additional points of interest are often suggested by plots.

n

11. Thirty observations of the relative return of over-the-counter stocks bought in the week of the 9th to the 13th of May, 1994 are given below.

-0.2940	-0.1092	-0.1053	-0.0707	-0.0563
-0.0541	-0.0423	-0.0398	-0.0396	-0.0390
-0.0381	-0.0323	-0.0221	-0.0169	-0.0139
-0.0081	0.0038	0.0057	0.0156	0.0172
0.0182	0.0192	0.0423	0.0459	0.0476
0.0667	0.0714	0.0780	0.1176	0.1224

Note: $\bar{x} = -0.01030$ and $s = 0.0786$

Investors would like to know how the market performed. One measure of market performance is the mean relative return for the week.

A 95% confidence interval for the mean relative return is [-0.040, 0.019]. Which **one** of the following statements is **false**?

T (1) The *P*-value for testing $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$ is larger than 0.05.

(2) There is NO evidence, at the 5% level, to believe that the mean return is different from zero.

T (3) A 99% confidence interval for the mean return would be wider than the 95% confidence interval. $n=30$

T (4) Because the sample is large the assumption that the sample mean, $\bar{X}$, is approximately Normal is reasonable.

T (5) An estimate of the mean relative return is –0.01030. $\bar{x}$

Questions 12 to **17 refer to the following situation:**

The weight gain of women during pregnancy has an important effect on the birth weight of their children. In a study conducted in three countries, weight gains (in kg) of women during the last three months of pregnancy were measured. The results are summarised in the following table.

Country		n	$\overline{x}$	s_x
Egypt	(E)	11	4.55	1.83
Kenya	(K)	11	3.29	0.851
Mexico	(M)	11	2.9	1.8

mean

12. We wish to determine the size of the difference in average weight gain between Egyptian and Kenyan women. Given that the data shows no non-Normal features when plotted, the most appropriate procedure to use here is (select **one** only):

size!

(1) a confidence interval based on the paired t-test.

(2) a two-independent sample proportion test.

(3) a confidence interval based on the two-independent sample t-test.

(4) a confidence interval based on the paired sign test.

(5) a confidence interval based on the Mann-Whitney test.

13. Suppose a two-independent sample t-test is appropriate here for testing H_0: $\mu_E - \mu_K = 0$, against H_1: $\mu_E - \mu_K \neq 0$. Then the value of the standard error, $se(\overline{x}_E - \overline{x}_K)$, is (select **one** only):

(1) 0.5990

(2) 0.6085

(3) 0.3703

(4) 0.7739

(5) 0.9288

$$\sqrt{\frac{s_E^2}{n_E} + \frac{s_K^2}{n_K}}$$

$$= \sqrt{\left(\frac{1.83^2}{11} + \frac{.851^2}{11}\right)}$$

$$= .6085$$

14. Suppose the *P-value* for the test in Question 9 is 0.1654 (it is not). Which **one** of the following statements is **correct**?

(1) The data provides no evidence against H_0.

(2) The data provides strong evidence that H_0 is true.

(3) The data provides evidence that H_0 is true.

(4) The data provides evidence in favour of H_1.

(5) The data provides strong evidence that H_1 is true.

15. When calculating a 90% confidence interval for $\mu_E - \mu_K$, the value of the *t*-multiplier obtained from a *Student's* *t*-table is 2.228. Assuming $se(\bar{x}_E - \bar{x}_K) = 0.7$ (it is not), then the **margin of error** for the 90% confidence interval is:

(1) $(4.55 \cancel{-} 3.29) \pm 2.228 \times \frac{0.7}{\cancel{\sqrt{11}}}$

(2) $\pm 2.228 \times \frac{0.7}{\sqrt{11}}$

(3) $(3.29 \cancel{-} 4.55) \pm 2.228 \times 0.7$

(4) $(3.29 \cancel{-} 4.55) \pm 2.228 \times \frac{0.7}{\cancel{\sqrt{11}}}$

(5) $\pm 2.228 \times 0.7$

$\pm t \times se(est)$

single mean

$\Rightarrow \pm 2.228 \times .7$

16. We wish to determine if there are differences in average weight gain in any of the three countries. The most appropriate procedure to use here is (select **one** only):

(1) a confidence interval for the highest-lowest average weight gain: $\mu_{high} - \mu_{low}$.

(2) an *F*-test for H_0: $\mu_E = \mu_K = \mu_M$.

(3) a paired t-test for H_0: $\mu_{diff} = 0$, where $\mu_{diff} = \mu_{high} - \mu_{low}$.

(4) three paired t-tests.

(5) a Tukey interval for the highest-lowest average weight gain: $\mu_{high} - \mu_{low}$.

17. We wish to perform an *F*-test for the weight gain data. The appropriate degrees of freedom are (select **one** only):

(1) $df_1 = 2$ and $df_2 = 30$

~~(2) $df_1 = 3$ and $df_2 = 33$~~

~~(3) $df_1 = 33$ and $df_2 = 3$~~

~~(4) $df_1 = 30$ and $df_2 = 2$~~

(5) $df_1 = 2$ and ~~$df_2 = 33$~~

$df_1 = k - 1 = 3 - 1 = 2$

$df_2 = n_{tot} - k = 33 - 3$
$= 30$

18. Which one of the following statements is **true**?

(1) A small *P-value* provides evidence ~~of the size~~ of an effect.
(2) Statistical significance is the same as practical significance.
(3) Practical significance depends on the size of the effect.
(4) A small *P-value* provides ~~no~~ evidence against H_0.
(5) A confidence interval estimates the ~~strength~~ of an effect.

size

19. Which one of the following statements about *t*-tests is **false**?

(1) *t*-tests may not be valid if there are outliers present and the sample is not large.
(2) *t*-tests may not be valid when the data show clustering.
(3) In general, *t*-tests are ~~not~~ robust against the Normality assumption.
(4) *t*-tests will generally work well for any large sample.
(5) *t*-tests may not be valid if the data are clearly skewed and the sample is not large.

Question 20 refers to the following information.

n

The heights (in cm) of the carapaces (shells) of a sample of 48 painted turtles were recorded. Shown below is a stem-and-leaf plot of this data set.

Units: 3|5 = 35cm

```
3 | 55778888999
4 | 0000111222344
4 | 55566789
5 | 0111113
5 | 567
6 | 01233
6 | 7
```

Figure: Stem-and-leaf plot of carapace height of painted turtles.

20. Based on this sample of size 48, a 95% confidence interval for the underlying mean carapace height of all painted turtles is 44.0cm to 48.8cm. The number of painted turtle carapace heights we would need to sample in order to halve the width of this interval is, approximately:

(1) 24
(2) 192
(3) 12
(4) 96
(5) 72

$4 \times 48 = 192$

21. Which one of the following statements about the validity of confidence intervals of the form sample mean ± t standard errors is false?

T (1) It is critical that the sample is random.
(2) It is not critical that the distribution being sampled is Normal.
T (3) It is critical that the observations come from the same distribution.
T (4) Outliers and clusters of data can invalidate confidence intervals.
T (5) It is critical that observations are independent.

22. Which one of the following statements is false?

T (1) In a t-test for no difference between two proportions, being able to demonstrate that the difference was of practical significance (importance) would almost always imply statistical significance.
T (2) In hypothesis testing, large samples can lead to small *P-values* without the results having any practical significance (importance).
T (3) In hypothesis testing, statistical significance does not imply practical significance (importance).
(4) In a hypothesis test for no difference between two proportions, a very small *P-value* always indicates a very large difference in the proportions.
T (5) In hypothesis testing, a nonsignificant test result does not imply that the null hypothesis is true.

23. Which one of the following is **false**?

(1) T A *P-value* calculated for a hypothesis formulated after looking at the data provides less convincing evidence than if the study had been designed to investigate the hypothesis.

(2) T Formulae for the standard errors of data estimates do not take into account systematic biases in the experiment or survey.

(3) The fact that multiple comparisons have been made from a single set of data ~~can~~ cannot be ignored when reporting the results.

(4) T If 100 people independently collect data and calculate a 95% confidence interval for a population proportion we expect approximately 95 people to capture the true proportion in their interval and 5 to miss it.

(5) T If 100 people independently collect data and test a true hypothesis, then just by chance, we expect about 5 to obtain results, which were significant at the 5% level.

Questions 24 to **27** refer to the following information.

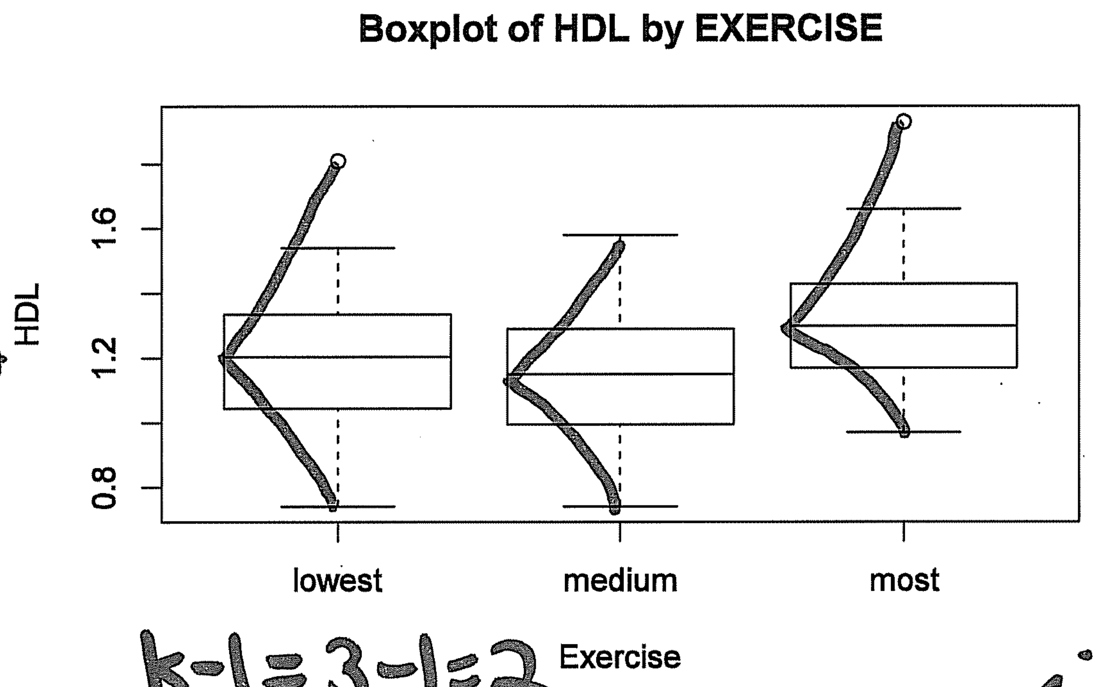
HDL cholesterol is known as the "good cholesterol" as it is associated with lower risks of problems like heart disease. The following data were collected on men working in New Zealand companies.

The men were divided into exercise groups by the amount of exercise they reported.

3 groups

Levels of exercise were classified as lowest, medium, and most. Sixty men were randomly sampled from each exercise group, and their HDL cholesterol level was measured.

Box plots and some PASW (SPSS) output analysing the data follow.



$k-1=3-1=2$

Figure 3: Box plot of HDL by exercise level.

$\frac{.3507}{.0382} = 9.181$ (3dp)

Analysis of Variance for HDL

Source	DF	SS	MS	F	P
EXERCISE	df1	0.7013	0.3507	f0	0.000
Error	df2	6.7632	0.0382		
Total	179	7.4645			

$180-3\ (n_{tot}-k) = 177$

	n	sample mean	sample std dev
lowest	60	1.1998	0.1926
medium	60	1.1497	0.2301
most	60	1.2998	0.1906

$\frac{.2301}{.1906} = 1.2 < 2$

$n_{tot} = 180$

Table 3: PASW (SPSS) output for the cholesterol data.

24. Which one of the following statements about the *F*–test shown in Table 3 is false?

- (1) The outside values in the lowest and most groups show that the Normality assumption of the *F*–test is violated.
- T (2) The alternative hypothesis states that at least one of the exercise groups has a different underlying mean HDL level from another.
- T (3) The differences in the sample standard deviations of the lowest, medium and most exercise groups do not affect the validity of the *F*–test in practice. .2301/.190622
- T (4) The box plots in Figure 3 give us no information on the independence of the exercise groups.
- T (5) The null hypothesis states that the underlying mean HDL level is the same for each exercise group.

25. The values for the degrees of freedom, df_1 and df_2, at the top of Table 3 are:

- (1) $df_1=2$, $df_2=177$
- (2) $df_1=2$, $\cancel{df_2=178}$
- (3) $df_1=2$, $\cancel{df_2=59}$
- ~~(4)~~ $\cancel{df_1=3}$, $df_2=177$
- ~~(5)~~ $\cancel{df_1=3}$, $\cancel{df_2=176}$

26. The value of the *F*–test statistic, f_0, at the top of Table 3 is nearest to:

- (1) 9.643
- (2) 10.643
- (3) 9.181
- (4) 0.104
- (5) 0.109

27. Which one of the following statements is the best interpretation of the *P-value* in the analysis of variance for HDL shown in Table 3? 0.000

(1) There is ~~no~~ evidence that all of the exercise groups have different underlying mean HDL cholesterol levels.

(2) There is extremely strong evidence that at least one of the exercise groups has a different underlying mean HDL cholesterol level from another.

(3) There is ~~no~~ evidence that any of the exercise groups have different underlying mean HDL cholesterol levels.

(4) There is extremely strong evidence that ~~all of~~ the exercise groups have different underlying mean HDL cholesterol levels.

(5) There is ~~some~~ evidence that at least one of the exercise groups has a different underlying mean HDL cholesterol level from another.

28. Which one of the following statements about paired data is false?

(1) For paired data, we analyse the differences.

(2) Pairing is beneficial when the variability within pairs is small compared with the variability between pairs.

(3) The Wilcoxon signed-rank test can be used to analyse the differences in paired data.

(4) Pairing can~~not~~ be used in observational studies.

(5) The carryover effect occurs when the first treatment alters the effect of the second treatment.

29. Which one of the following statements is **true**?

(1) A small *P-value* provides evidence of ~~the size of~~ an effect.

(2) Statistical significance is not the same as practical significance.

(3) Practical significance depends on the size of the effect.

(4) A small *P-value* provides ~~no~~ evidence against H_0.

(5) A confidence interval estimates the ~~strength~~ size of an effect.

typo in answers for Q8

Answers

1.	**(2)**	2.	**(2)**	3.	**(5)**	4.	**(5)**	5.	**(4)**	6.	**(2)**
7.	**(4)**	8.	~~**(2)**~~ (4)	9.	**(1)**	10.	**(4)**	11.	**(2)**	12.	**(3)**
13.	**(2)**	14.	**(1)**	15.	**(5)**	16.	**(2)**	17.	**(1)**	18.	**(3)**
19.	**(3)**	20.	**(2)**	21.	**(2)**	22.	**(4)**	23.	**(3)**	24.	**(1)**
25.	**(1)**	26.	**(3)**	27.	**(2)**	28.	**(4)**	29.	**(3)**		