Stats 101/101G/108 Workshop

Hypothesis Tests:
Proportions [HTP]

2020

by Leila Boyle

Stats 101/101G/108 Workshops

The Statistics Department offers workshops and one-to-one/small group assistance for Stats 101/101G/108 students wanting to improve their statistics skills and understanding of core concepts and topics.

Leila's website for Stats 101/101G/108 workshop hand-outs and information is here: **www.tinyURL.com/stats-10x**

Resources for this workshop, including pdfs of this hand-out and Leila's scanned slides showing her working for each problem are available here: **www.tinyURL.com/stats-HTP**

Leila Boyle
Undergraduate Statistics Assistance, Department of Statistics
Room 303S.288 (second floor of the Science Centre, Building 303S)
l.boyle@auckland.ac.nz; (09) 923-9045; 021 447-018

Want help with Stats?

Stats 101/101G/108 appointments

Book your preferred time with Leila here: **www.tinyURL.com/appt-stats**, or contact her directly (see above for her contact details).

Stats 101/101G/108 Workshops

One computing workshop, four exam prep workshops and four drop-in sessions are held during the second half of the semester.

Workshops are run in a relaxed environment and allow plenty of time for questions. In fact, this is encouraged! ☺

Please make sure you bring your calculator with you to all of these workshops!

No booking is required – just turn up to any workshop! You are also welcome to come along virtually on Zoom if you prefer. Search your emails for "Leila" to find the link – email Leila at l.boyle@auckland.ac.nz if you can't find it.

- **Computer workshop: *Hypothesis Tests in SPSS***
  www.tinyURL.com/stats-HTS
  Computing for Assignment 3 **–** covers the **computing** you need to do for **Questions 3 and 4** (iNZight plots & SPSS output). There are **six identical sessions**:
  - Friday 16 October, 3-4pm
  - Monday 19 October, 10-11am
  - Monday 19 October, 2-3pm
  - Tuesday 20 October, 4-5pm
  - Wednesday 21 October, 11am-midday
  - Wednesday 21 October, 3-4pm

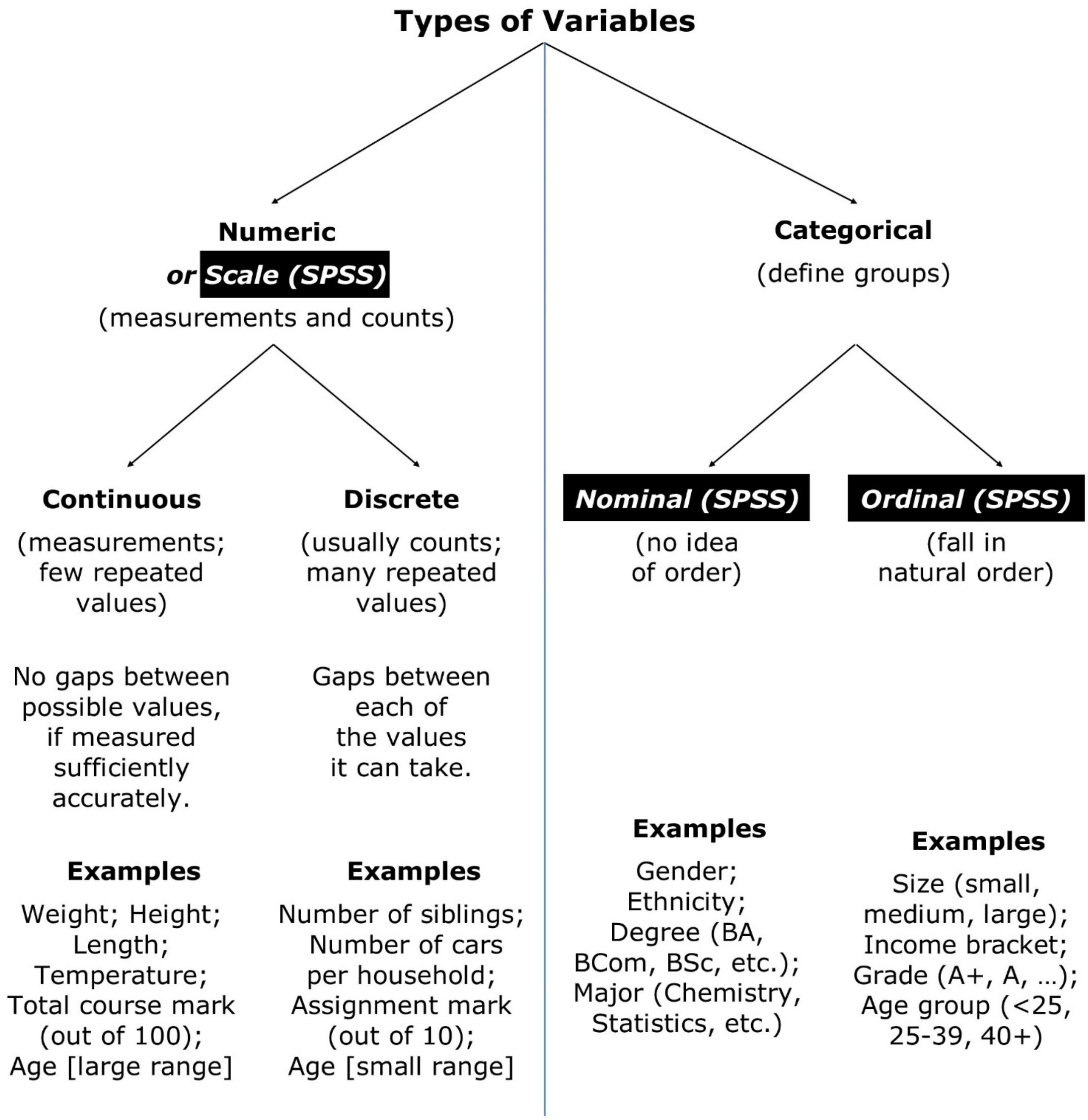
- **Exam prep workshops**
  - ***Chi-Square Tests*** **www.tinyURL.com/stats-CST**
    Exam revision for Chapter 9 **–** Saturday 24 October, 1-4pm, LibB15 (useful exam prep and also useful for the **Chapter 9 Quiz** due at 11pm on Wednesday 28 October!)
  - ***Regression and Correlation*** **www.tinyURL.com/stats-RC**
    Exam revision for Chapter 10 **–** Saturday 31 October, 9.30am-12.30pm, LibB10 (useful exam prep and also useful for the **Chapter 10 Quiz** due at 11pm on Wednesday 4 November!)
  - ***Hypothesis Tests: Proportions*** **www.tinyURL.com/stats-HTP**
    Exam revision for Chapters 6 & 7 *(with a focus on proportions)* **–** Tuesday 3 November, 9.30am-12.30pm, LibB10 (useful exam prep)
  - ***Hypothesis Tests: Means*** **www.tinyURL.com/stats-HTM**
    Exam revision for Chapter 6, 7 & 8 *(with a focus on means)* **–** Tuesday 3 November, 1-4pm, LibB10 (useful exam prep)

- **Drop-in sessions**
  - Saturday 17 October, 9.30am-4pm, LibB10
  - Saturday 24 October, 9.30am-12.30pm, LibB15
  - Monday 26 October, 9.30am-4pm, LibB10
  - Saturday 31 October, 1-4pm, LibB10

Hypothesis Tests: *Proportions* [HTP]

This material builds on a number of workshops already held in the first half of this semester, which you may or may not have attended.

If you want to learn more about how to **extract a proportion/probability from a two-way table of counts**, see the ***Proportions and Proportional Reasoning [PPR]*** workshop material. For more practice on how to **quantify the size of a single proportion or difference between two proportions**, see the ***Confidence Intervals: Proportions [CIP]*** workshop material.



📖**Useful reference:** Chance Encounters, pages 40 – 42

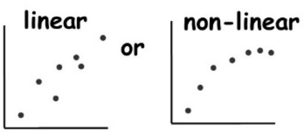
Exploring relationships between two variables

Two Numeric	**Numeric versus Categorical**	**Two Categorical**
Exploratory tools to use to explore relationship/s		
Scatter plot	Side-by-side plots on the same scale: any n { • Dot plots $n \geq 20$ { • Stem-and-leaf plots, • Box plots $n \geq 50$ { • Histograms	Two-way table of counts and/or Bar graphs of proportions (on rows/columns)

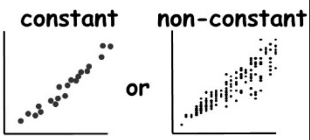
Features to look for:

Two Numeric

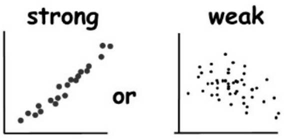
- Trend



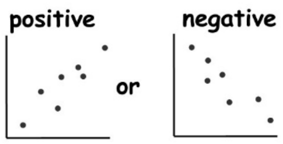
- Scatter



- Strength of relationship

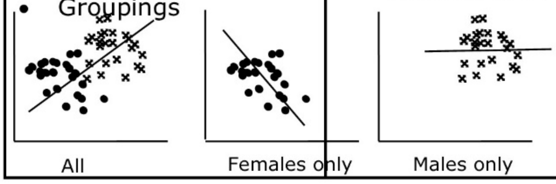


- Association



- Outliers

- Groupings



Numeric versus Categorical

Compare the groups by looking at:

- Any group differences:
  - averages (centres) → **medians**, **means**
  - variability (spread) → **IQRs**, **ranges**
  - shapes
    - ♦ symmetric/skewed → **left/negative**, **right/positive**
    - ♦ modes → **unimodal**, **bimodal**, **trimodal** } **not box!**
- Details of individual groups:
  - outliers, gaps, clusters, groupings

Think about reasons *why* these differences, similarities and features are seen

Two Categorical

- Most common and least common combinations
- Differences in distributions (e.g. row and/or column bar graphs)

Recall that:

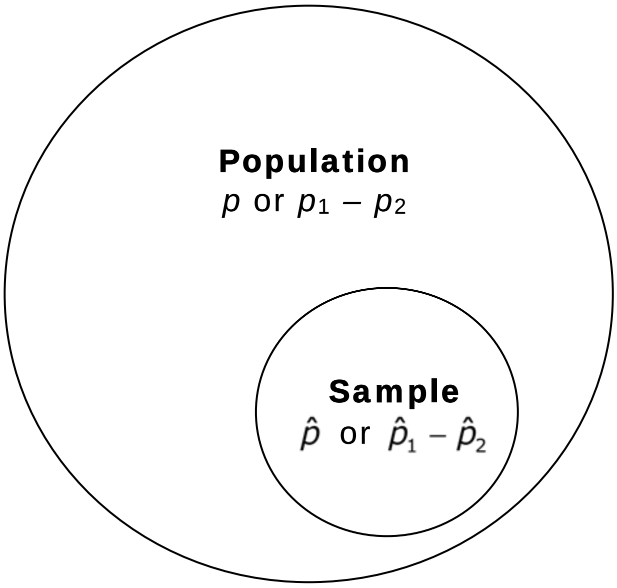
- A **proportion** is a number between 0 and 1 that estimates the likelihood of an event occurring.
- Our main source of proportions is from data which will usually be presented in a table of counts.

t-tests by Hand – One and Two Proportion/s

We use statistics to find out about the real world and aspects of it specific to our area of interest. Statistical tools allow us to deal with the **uncertainty** present in all samples due to **sampling variation** which occurs because we are unable to survey the entire population of interest.

We are usually unable to survey the entire population (take a census) as it is too large and/or there are:

- budget constraints
- time limits
- logistical barriers



This means we are unable to establish the **parameters** of interest within our population, such as:

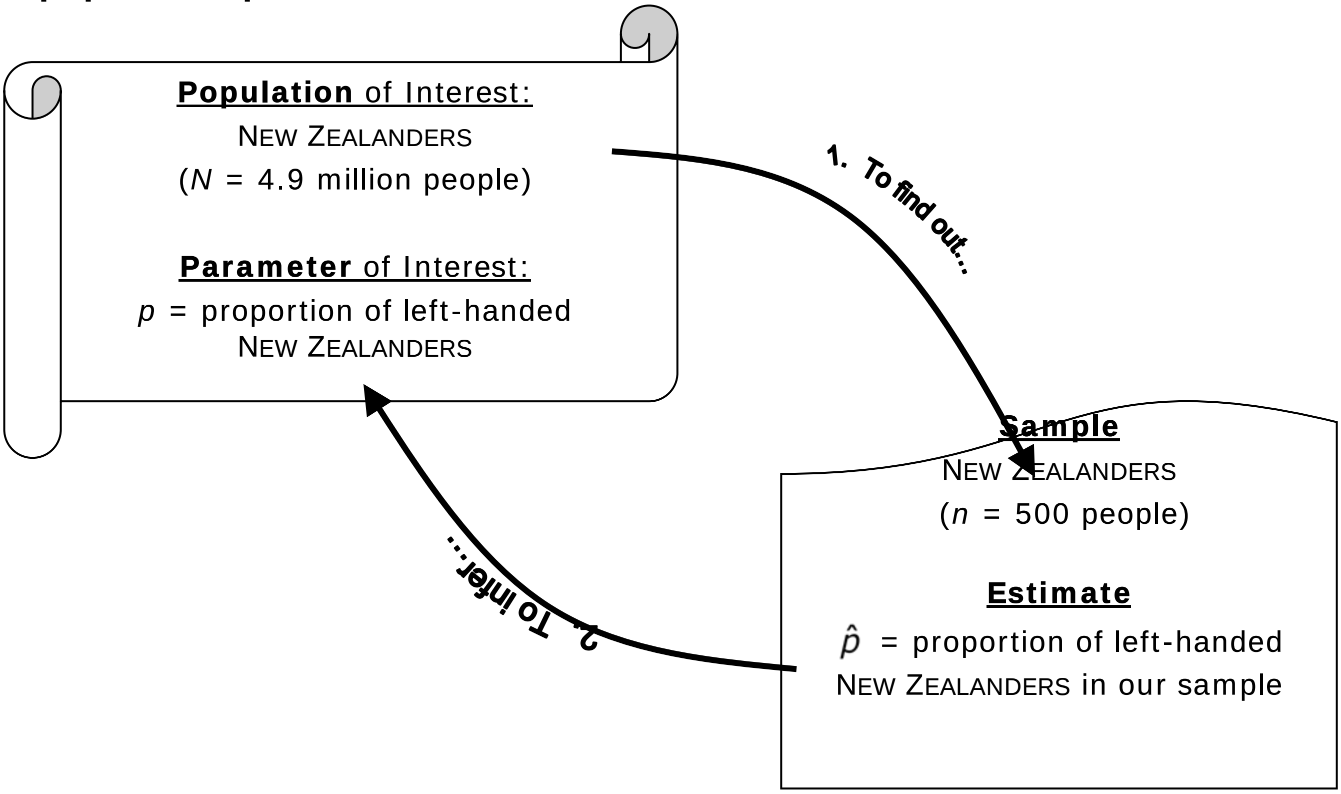
2. Population proportion, p or
4. Difference in population proportions, $p_1 - p_2$

This means that the **parameter** of interest
is an ***<u>unknown</u> numerical characteristic***
for that particular population.

To estimate an ***<u>unknown</u> numerical characteristic* (parameter)** for our population of interest, we take a sample and find a sample **estimate** from it (that is, we make a **statistical inference**). The **sample estimates** of the above **population parameters** are:

2. Sample proportion, $\hat{p}$
4. Difference in sample proportions, $\hat{p}_1 - \hat{p}_2$

Usually ^HATS or ¯BARS are used to distinguish between **sample estimates** and **population parameters**.



We use **sample** data to make inferences (draw conclusions) about **population** parameters by carrying out hypothesis tests and constructing confidence intervals.

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

Significance testing and confidence intervals are methods used to deal with the **uncertainty** about the true value of a parameter caused by the **sampling variation** in estimates.

Step-by-Step Guide to Performing a Hypothesis Test by Hand

1. State the **parameter** of interest (symbol and words).

For example, is it μ, p, $\mu_1 - \mu_2$, or $p_1 - p_2$?

2. State the **null hypothesis**, H_0. **e.g.** H_0: parameter = *hyp. val.*

3. State the **alternative hypothesis**, H_1. **e.g.** H_1: parameter ≠ *hyp. val.*

or H_1: parameter > *hyp. val.*

or H_1: parameter < *hyp. val.*

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

5. Calculate the **test statistic**:

see back page for Formulae Sheet

For example, for a ***t*-test statistic**:

- Use: $t_0 = \dfrac{\text{estimate} - \text{hypothesised value}}{\text{std error}}$

- Use the estimate from Step 4 and the hypothesised value from Steps 2&3.
- Use the appropriate standard error. (Will be provided)
- Calculate t_0.

6. Estimate the ***P*-value**. (Will be provided)

7. **Interpret** the *P-value*. (see page 13)

8. Calculate the **confidence interval**.

For example, for a **Normality-based confidence interval**:

- Use: *estimate ± t × se(estimate)*

- Use the estimate from Step 4 and the standard error from Step 5.
- Use the appropriate *t*-multiplier. (Will be provided)

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

10. Give an overall **conclusion**.

- There are four different types of problem:

~~1. Single mean~~ 2. Single proportion ~~3. Difference between two means~~

4. Difference between two proportions:

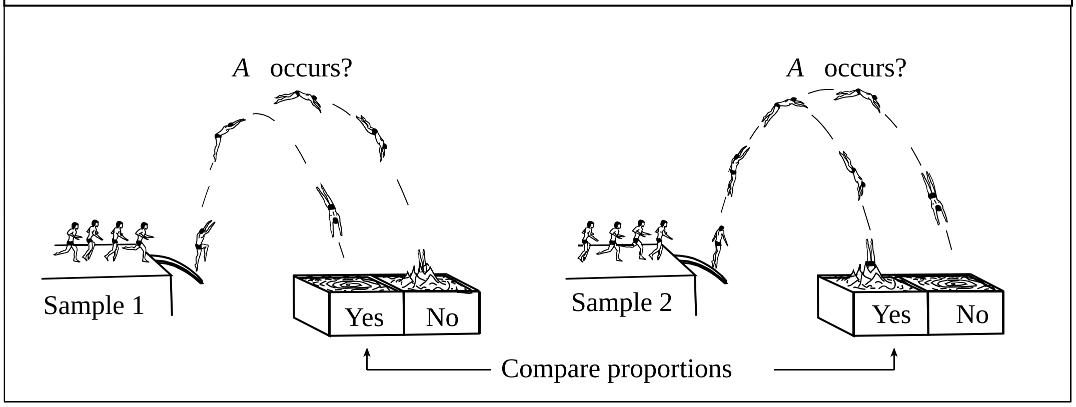
Situation (a) ***Proportions from two independent samples***

Situation (b) ***One sample of size n, several response categories***

Situation (c) ***One sample of size n, many yes/no items***

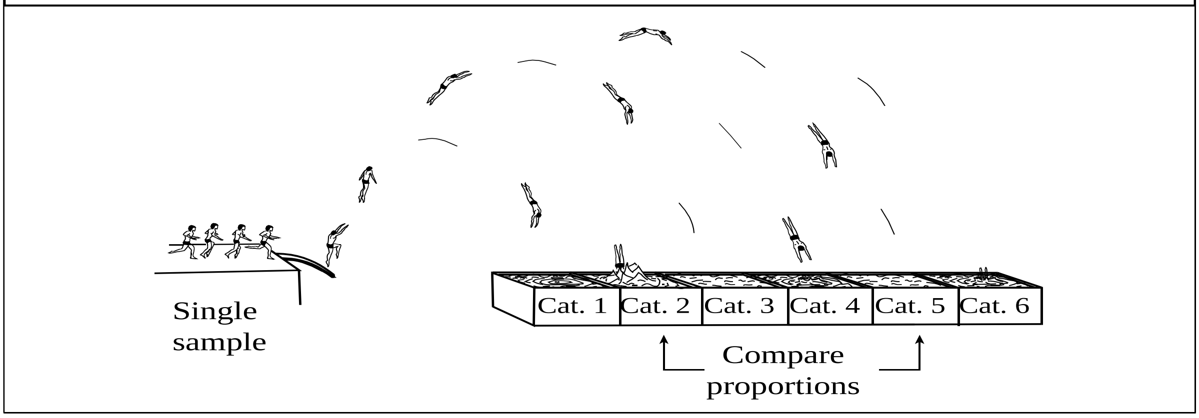
- **3 sampling situations for the difference between two proportions**

Situation (a) *Proportions from two independent samples*



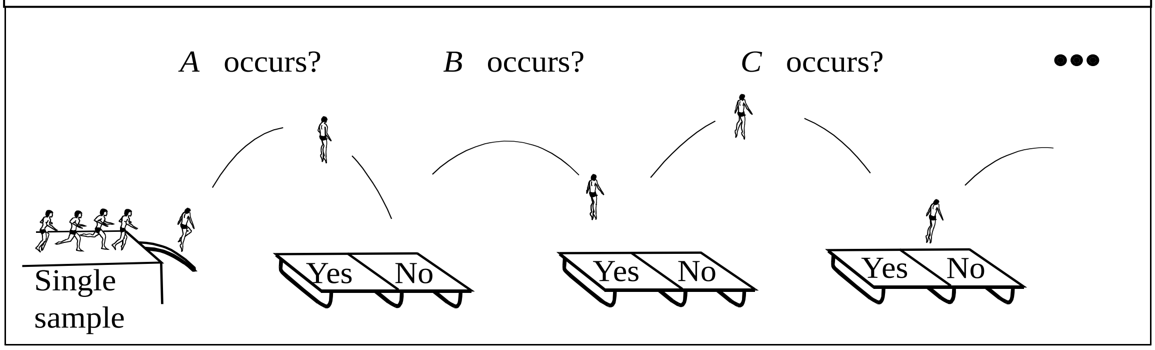
From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 1999.

Situation (b) *One sample of size n, several response categories*



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

Situation (c) *One sample of size n, many yes/no items*



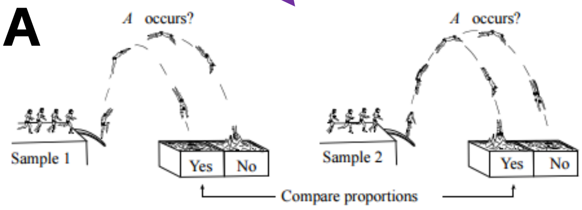
From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.

Identifying the Sampling Situations:

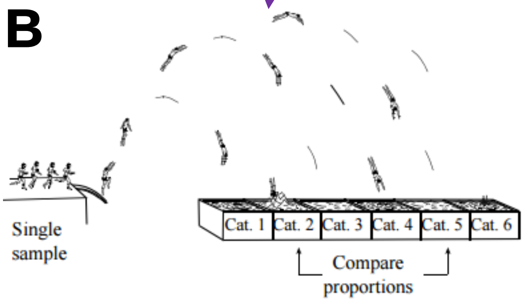
How many independent samples / independent subgroups?

- **One** → Can you answer "yes" more than once / pick more than one option?
  - **No** → **B**
  - **Yes** → **C**
- **Two** → **A**

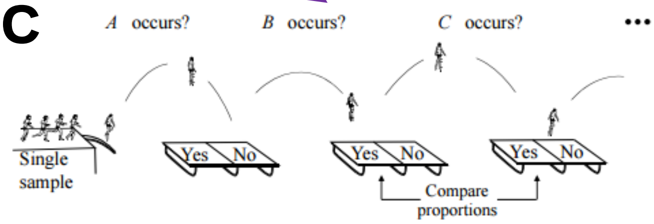
A



B



C



Step 1

The **parameter** of interest we are investigating depends on the problem type:

Parameter
~~1. Single mean μ:~~
2. Single proportion p:
~~3. Difference between two means $\mu_1 - \mu_2$: (independent samples)~~
4. Difference between two proportions $p_1 - p_2$:

Steps 2 & 3

The null hypothesis, H_0

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

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

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

2. H_0: p =

4. H_0: $p_1 - p_2$ =

✓ It's the **boring** thing – **there is no** effect or difference.

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)*
- H_1: *parameter > hypothesised value (some number)*
- H_1: *parameter < hypothesised value (some number)*

2. H_1: p

4. H_1: $p_1 - p_2$

✓ When do we use a 1-sided alternative hypothesis?
* if in doubt * data * research

✓ It's the **interesting** thing – **there is an** effect or difference.

Step 4 (and Step 8)

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

Parameter	**Estimate**
~~1. Single mean μ:~~	*estimate* = $\bar{x}$
2. Single proportion ***p***:	*estimate* = $\hat{p}$
~~3. Difference between two means $\mu_1 - \mu_2$: (independent samples)~~	*estimate* = $\bar{x}_1 - \bar{x}_2$
4. Difference between two proportions $p_1 - p_2$:	*estimate* = $\hat{p}_1 - \hat{p}_2$

Step 5 (and Step 8)

- The **standard error** can be found from the *t*-procedures tool.

In the exam situation, the standard error will be provided.

- The **degrees of freedom** are based on the problem type:

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

- **The *t*-test statistic, t_0:**
  - ✓ tells us how many standard errors the estimate is away from the hypothesised value.
  - ✓ is calculated using: $t_0 = \dfrac{estimate - hypothesised\ value}{std\ error}$

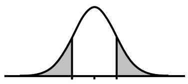
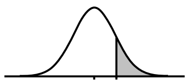
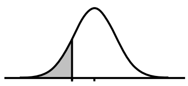
see back page for Formulae Sheet

  - ✓ is **positive**, if the estimate is **above** the hypothesised value.
  - ✓ is **negative**, if the estimate is **below** the hypothesised value.
  - ✓ is a **measure** of **difference/distance/discrepancy** between the estimate and the hypothesised value in terms of standard errors.

Step 6

- The ***P*-value:**
  - ✓ is the **conditional** probability of observing a test statistic as extreme as that observed or more so, given that the null hypothesis, H_0, is true.
  - ✓ is the probability that sampling variation would produce an estimate that is at least as far from the hypothesised value than the estimate we obtained from our data, assuming that the null hypothesis is true.
  - ✓ measures the strength of evidence **against** H_0.
  - ✓ is calculated using the t-test statistic and the appropriate Student's t-distribution for the t-test.

In the exam situation, the *P*-value will be provided.

Alternative hypothesis	***P*-value ≈ area of shaded region**
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

Student(*df*) or Re-randomisation distribution

Typical generic exam question about *P-values*:

Q. Which **one** of the following statements about a *P-value* is **false**?

(1) A *P-value* measures the strength of evidence against the null hypothesis.

(2) A relatively large test statistic results in a relatively small *P-value*.

(3) A *P-value* is the conditional probability of observing a test statistic as extreme as that observed or even more so, if the null hypothesis were true.

(4) A *P-value* says nothing about the size of an effect or difference.

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

Step 7

- The **P-value** measures the strength of evidence against the null hypothesis, H_0. We interpret the *P-value* as a description of the **strength of evidence against the null hypothesis, H_0**. The **smaller** the *P-value*, the **stronger** the evidence against H_0:

P-value	Evidence against H_0
> 0.10	None
≈ 0.07	Weak
≈ 0.05	Some
≈ 0.01	Strong
≤ 0.001	Very Strong

- An alternative approach often found in research articles and news items is to describe the test result as (statistically) significant or not significant. A test result is said to be significant when the *P-value* is "small enough"; usually people say a *P-value* is "small enough" if it is less than 0.05 (5%):

Testing at a 5% level of significance:

P-value	Test result	Action
< 0.05	Significant	Reject H_0 in favour of H_1
> 0.05	Nonsignificant	Do not reject H_0

Testing can be done at any level of significance; 1% is common but 5% is what most researchers use.

The level of significance can be thought of as a false alarm error rate, i.e. it is the proportion of times that the null hypothesis will be rejected when it is actually true (which can result in action being taken when really no action should be taken).

Thus, a statistically significant result means that a study has produced a "small" *P-value* (usually < 5%).

Normality-based (Chapter 6) Confidence Interval:

________ *of* _________

Step 8

$estimate \pm t \times se(estimate)$

The ***t*-multiplier** is based on:

- Whether we are investigating means or proportions
- The desired level of confidence
- The degrees of freedom:

Estimate	**Degrees of freedom**
1. $estimate = \bar{x}$	$df = n - 1$
2. $estimate = \hat{p}$	$df = \infty$
3. $estimate = \bar{x}_1 - \bar{x}_2$	$df = minimum(n_1 - 1,\ n_2 - 1)$
4. $estimate = \hat{p}_1 - \hat{p}_2$	$df = \infty$

In the exam situation, you will be given the *t*-multiplier for a 95% confidence interval for a single proportion or a difference between proportions (it's 1.96!)

Typical generic exam question about confidence intervals:

Q. Which **one** of the following statements about a Normality-based confidence interval for a parameter *p* is **false**?

(1) Large samples tend to yield narrower 95% confidence intervals than small samples.

(2) In the long run, if we repeatedly take samples and calculate a 95% confidence interval from each sample, we expect that 95% of the intervals will contain the true value of *p*.

(3) If I plan to do a study in the future in which I will take a random sample and calculate a 90% confidence interval, there is a 90% chance that I will catch the true value of *p* in my interval.

(4) If a large number of researchers independently perform studies to estimate *p*, about 95% of them will catch the true value of *p* in their 95% confidence intervals.

(5) The process of using a population parameter to construct an interval for the data estimate is an example of statistical inference.

Step 9

- 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.

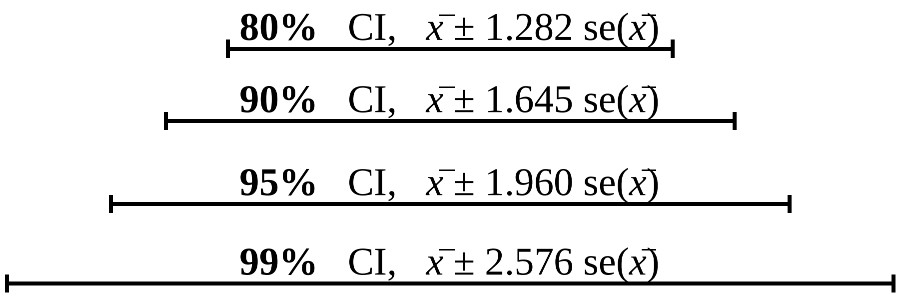


Figure 8.1.3 The greater the confidence level, the wider the interval

- Increasing the sample size will make the confidence interval more precise.

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

- To triple the precision of the confidence interval we **need 9 times** as many observations.

- 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 difference between means) of the population.

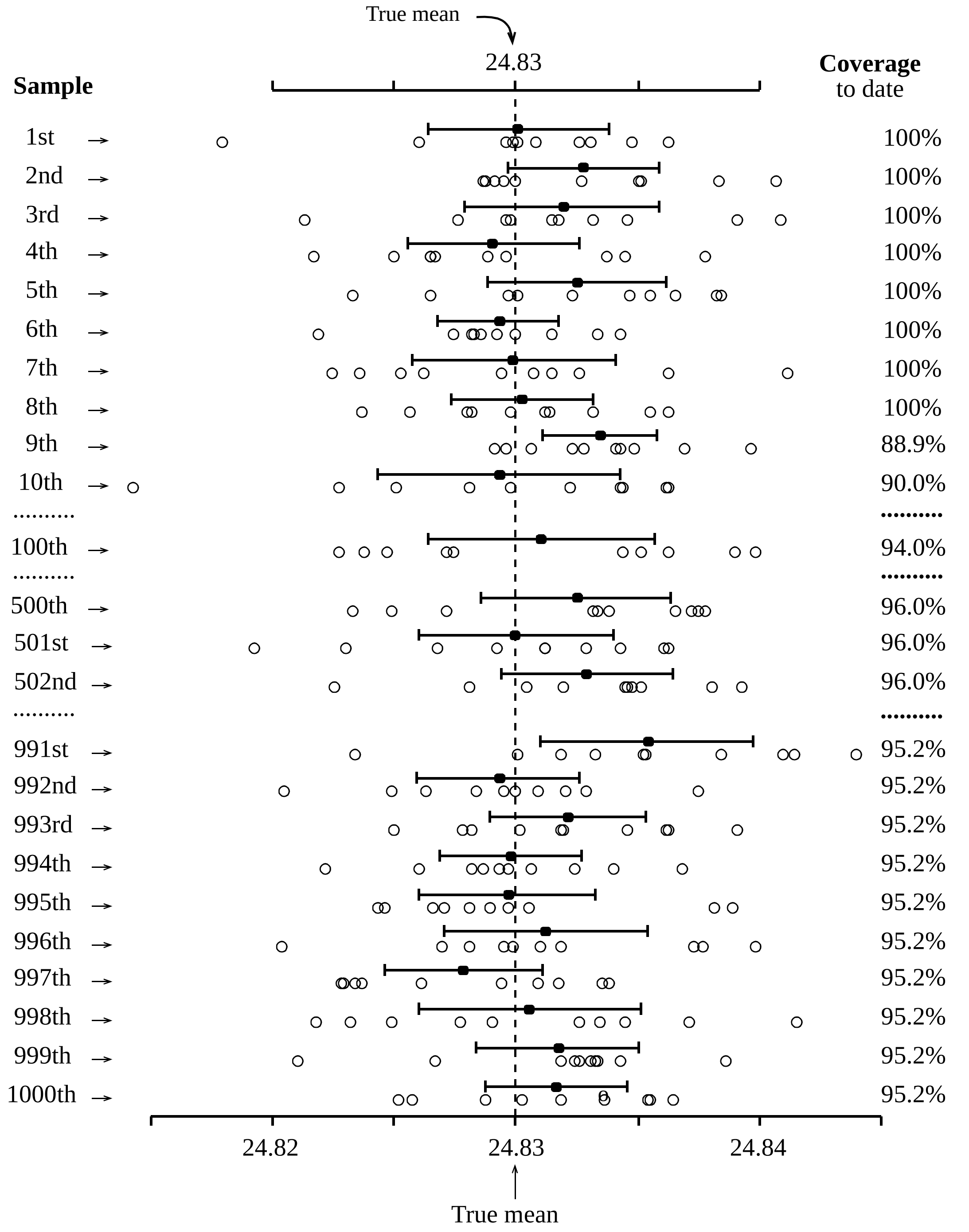


Figure 8.1.2 Samples of size 10 from a Normal(μ=24.83, s=.005) distribution and their 95% confidence intervals for μ..

Interpreting the CI limits → Step 9 for story type 4:

- CIs for the difference between two proportions:

Examples:

  - ✓ If the CI contains 0 (i.e. one negative and one positive number), there may be no difference between the two proportions. (–.05, .03)
  - ✓ If CI is positive, then p_1 is higher/larger than p_2. (0.03, 0.05)
  - ✓ If CI is negative, then p_1 is lower/smaller than p_2. (–0.05, –0.03)

Practical significance versus ***Statistical significance***

You may find it useful to use Anna Fergusson's online tool to visualise both statistical and practical significance: www.tinyURL.com/stats-sig

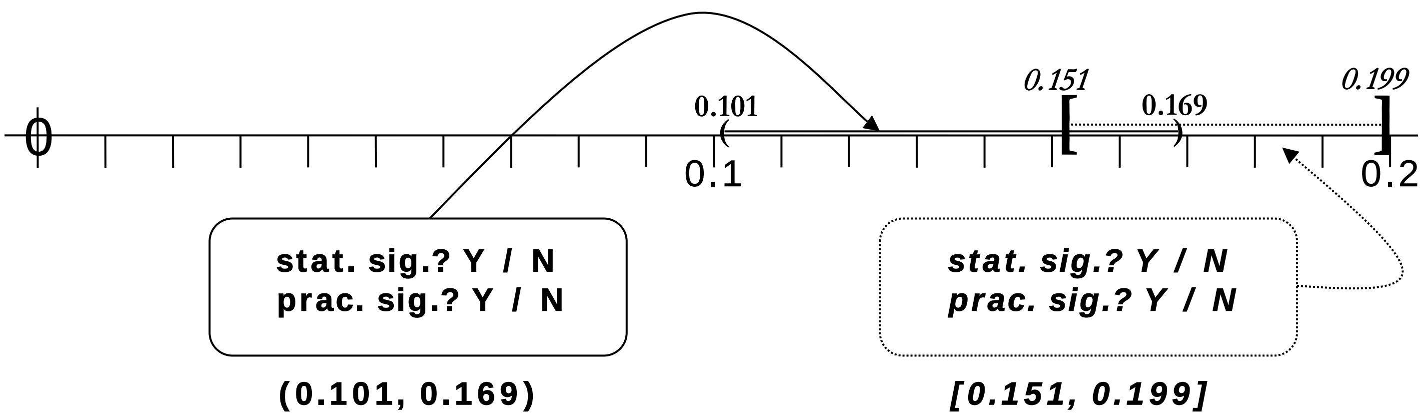
Statistical significance

- Relates to the ***P-value***.
- A small *P-value* provides evidence of the **existence** of an effect or difference.
- To be statistically significant at the 5% level, the *P-value* must be less than / greater than 0.05 (5%).

Practical significance

- Relates to the **size** of an effect or difference.
- Determined by examining the **confidence interval** in relation to the context of the question/s (i.e. the story).

For example: Left-handed New Zealanders: H_0: p = 0.1
vs H_1: p > 0.1



The link between the P-value *and the confidence interval*

Recall that a confidence interval for a parameter gives a range of plausible (believable) values for the unknown true parameter value.

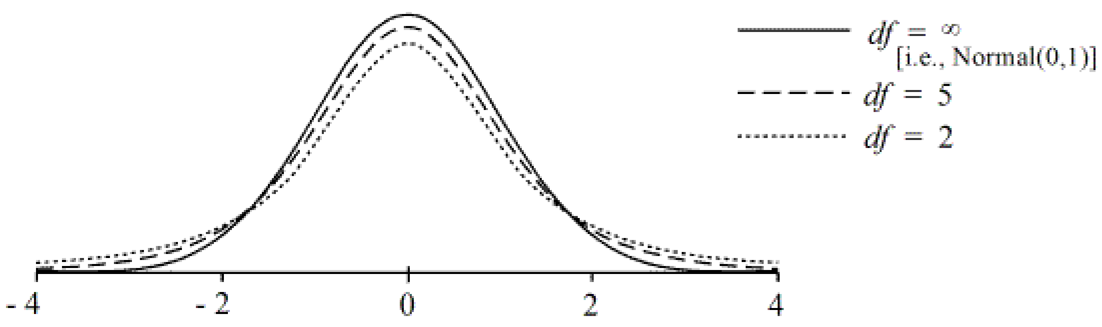
If a 2-tailed test has a *P-value* less than 5% then the test is significant at the 5% level of significance and the hypothesised value is not plausible (not believable) and it will / will not be in the 95% confidence interval. Conversely, if the hypothesised value is not in the 95% confidence interval it is not a plausible value and so the test is significant at the 5% level of significance and the *P-value* will be less than / greater than 5%.

If a 2-tailed test has a *P-value* greater than 5% then the test is not significant at the 5% level of significance and the hypothesised value is plausible (is believable) and so it will / will not be in the 95% confidence interval. Conversely, if the hypothesised value is in the 95% confidence interval it is a plausible value and so H_0 will be not rejected at the 5% level and the *P-value* will be less than / greater than 5%.

Note: The same relationship applies to 90% confidence intervals and *P-values* less than 10% (tests at the 10% level of significance), or 99% confidence intervals and *P-values* less than 1% (tests at the 1% level).

Student's *t*-distribution (background understanding)

- ✓ The parameter is the degrees of freedom, *df*.
- ✓ Smooth symmetric, bell-shaped curve centred at 0 like the Standard Normal distribution [Z ~ Normal (μ = 0, σ = 1)] but it's more variable (it's more spread out).



- ✓ As *df* becomes larger, the Student (*df*) distribution becomes more and more like the Standard Normal distribution.
- ✓ Student's *t*-distribution (*df* = ∞) and Normal (0,1) are the same distribution.

Check you understand! Practice Questions...

Questions 1 to **6** refer to the following information.

The Marketing Department of the University of Auckland carried out a survey of New Zealand bank customers. A random sample of 761 customers was selected and these customers were asked a wide range of questions about banks and the services the banks provide. From the responses, measurements were made on many variables. Some of these variables were:

Bank: The main bank used by the customer

— ANZ, BNZ, Westpac (for WestpacTrust), Other (all other banks)

Closeness: The customer's opinion of the closeness of their relationship with their main bank

— Not Close, Quite Close, Very Close

Performance: The customer's opinion of the overall performance of their main bank

— Poor/Fair, Good, Excellent

Two of the questions in the bank survey, each together with a table showing some of the percentage results, are given below.

Closeness:
How close is the relationship you have with your main bank?

	Response (Closeness)			**Sample size**
	Not Close %	Quite Close %	Very Close %	
Main Bank				
ANZ	57.3%	33.1%	9.6%	157
BNZ	48.3%	40.8%	10.8%	120
Westpac	48.7%	39.1%	12.2%	230
Other	39.4%	44.9%	15.7%	254
Total Sample	47.3%	40.0%	12.6%	761

Table 1: Responses to closeness of relationship with main bank

Performance:
How would you describe the overall level of performance of your main bank to date?

	Response (Performance)			Sample size
	Poor/Fair %	Good %	Excellent %	
Main Bank				
ANZ	29.9%	52.9%	17.2%	157
BNZ	32.5%	53.3%	14.2%	120
Westpac	24.3%	59.6%	16.1%	230
Other	13.8%	57.1%	29.1%	254
Total Sample	23.3%	56.4%	20.4%	761

Table 2: Responses to main bank's performance

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

Let:

p_{ANZ} be the proportion of bank customers, with ANZ as their main bank, who would describe their relationship with ANZ as 'Not Close'

and

$p_{Westpac}$ be the proportion of bank customers, with WestpacTrust (Westpac) as their main bank, who would describe their relationship with Westpac as 'Not Close'.

1. From the information in Table 1, page 19, an estimate of the difference $p_{ANZ} - p_{Westpac}$ is:

   (1) 0.86 (4) 0.054
   (2) 0.086 (5) 0.004
   (3) 0.54

2. A 95% confidence interval is constructed for the difference between p_{ANZ} and $p_{Westpac}$. For the purpose of calculating $se(\hat{p}_{ANZ} - \hat{p}_{Westpac})$, the sampling situation can be described as:

   (1) one sample of size 761, several response categories.
   (2) one sample of size 387, several response categories.
   (3) one sample of size 761, many yes/no items.
   (4) two independent samples of sizes 157 and 230.
   (5) one sample of size 387, many yes/no items.

3. A 95% confidence interval for the difference $p_{ANZ} - p_{Westpac}$ is (−0.014, 0.187).

The **best** interpretation of this interval is:

With 95% confidence, the percentage of bank customers, with ANZ as their main bank, who would describe their relationship with ANZ as 'Not Close' is somewhere between:

(1) 1.4 percentage points and 18.7 percentage points higher than the percentage of bank customers, with Westpac as their main bank, who would describe their relationship with Westpac as 'Not Close'.

(2) 1.4 percentage points lower and 18.7 percentage points higher than the percentage of bank customers, with Westpac as their main bank, who would describe their relationship with Westpac as 'Not Close'.

(3) 1.4 percentage points and 18.7 percentage points lower than the percentage of bank customers, with Westpac as their main bank, who would describe their relationship with Westpac as 'Not Close'.

(4) 1.4 percentage points higher and 18.7 percentage points lower than the percentage of bank customers, with Westpac as their main bank, who would describe their relationship with Westpac as 'Not Close'.

(5) −1.4% and 18.7%.

Questions 4 and **5** refer to the following additional information.

Consider only customers with **ANZ** as their main bank.

Let:

p_{Close} be the proportion who would describe their relationship with ANZ as either 'Quite Close' or 'Very Close'

and

$p_{Perform}$ be the proportion who would describe the ANZ performance as 'Good' or 'Excellent'.

Information from Tables 1 and 2, pages 19 and 20, is used to conduct a two-tailed *t*-test for no difference between p_{Close} and $p_{Perform}$.

4. The formula for the standard error of the estimate, $se(\hat{p}_{Close} - \hat{p}_{Perform})$, is:

   (1) two independent samples of sizes 157 and 174.

   (2) one sample of size 157, several response categories.

   (3) one sample of size 761, many yes/no items.

   (4) one sample of size 157, many yes/no items.

   (5) one sample of size 761, several response categories.

5. The expression for evaluating the test statistic for the null hypothesis, H_0: $p_{Close} - p_{Perform} = 0$, is:

   (1) $\dfrac{p_{Close} - p_{Perform}}{se(\hat{p}_{Close}) + se(\hat{p}_{Perform})}$

   (2) $\dfrac{\hat{p}_{Close} - \hat{p}_{Perform}}{se(\hat{p}_{Close} - \hat{p}_{Perform})}$

   (3) $\dfrac{\hat{p}_{Close} - \hat{p}_{Perform}}{\sqrt{se(\hat{p}_{Close})^2 - se(\hat{p}_{Perform})^2}}$

   (4) $\dfrac{\hat{p}_{Close} - \hat{p}_{Perform}}{se(\hat{p}_{Close}) + se(\hat{p}_{Perform})}$

   (5) $\dfrac{p_{Close} - p_{Perform}}{se(\hat{p}_{Close} - \hat{p}_{Perform})}$

Question 6 refers to the following additional information.

Consider only customers with **BNZ** as their main bank.

Let:

p_{Good} be the proportion who would describe the BNZ performance as 'Good'.

and

$p_{Excellent}$ be the proportion who would describe the BNZ performance as 'Excellent'.

6. Using information from Table 2, page 20, the formula for the standard error of the estimate, $se(\hat{p}_{Good} - \hat{p}_{Excellent})$, is:

   (1) one sample of size 120, several response categories.

   (2) one sample of size 761, many yes/no items.

   (3) two independent samples of sizes 157 and 174.

   (4) one sample of size 761, several response categories.

   (5) one sample of size 120, many yes/no items.

Questions 7 to **14** refer to the following information.

In 2015 Research New Zealand published the report 'Gender Equality in New Zealand' which was based on a public opinion survey completed in June 2015.

The survey was conducted by telephone with a nationally-representative sample of 500 New Zealanders aged 18 or over. You may consider the sample as a random sample of 500 adult New Zealanders.

Two of the questions in the survey were:

> *In your personal opinion, are males and females in New Zealand treated the same way in the workplace?*

and

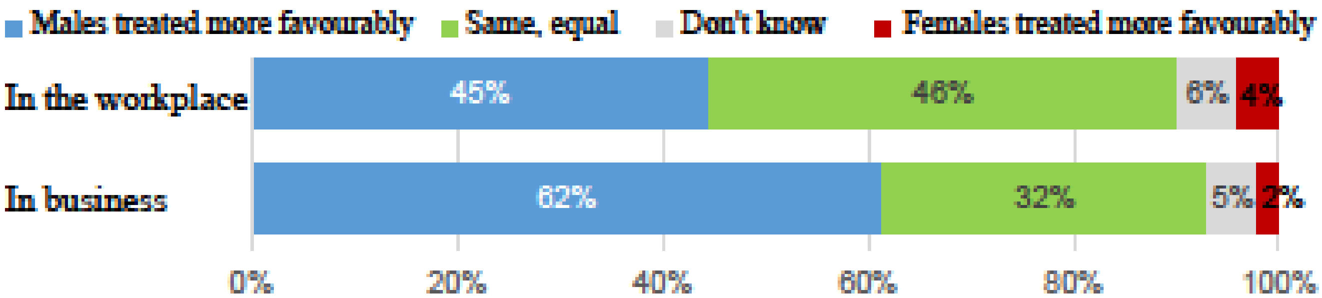
> *In your personal opinion, are males and females in New Zealand treated the same way in business?*

Responses to the above questions are summarised in Table 3 and Figure 1.

	Males treated more favourably	Treated the same	Don't know	Females treated more favourably
In the workplace	45%	46%	6%	4%
In business	62%	32%	5%	2%

Note: Totals may not sum to 100% due to rounding

Table 3: Gender equality questions' responses



Note: Totals may not sum to 100% due to rounding

Figure 1: Responses to gender equality questions

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

   (1) While 46% of the respondents believe that, in the **workplace**, males and females are treated the same; 45% believe males are treated more favourably compared to only 4% who believe females are treated more favourably.

(2) A respondent in this survey was nearly twice as likely to believe that, in **business**, males are treated more favourably than to believe that males and females are treated the same.

(3) The percentage of respondents who believe that males and females are treated the same in **business** is higher than the percentage of respondents who believe they are treated the same in the **workplace**.

(4) A respondent in this survey was almost as likely to believe that, in the **workplace**, males are treated more favourably as to believe that males and females are treated the same.

(5) While 32% of the respondents believe males and females are treated the same in **business**; 62% believe males are treated more favourably compared to only 2% who believe females are treated more favourably.

8. From the survey responses, we may report that 46% of adult New Zealanders hold the opinion that males and females are treated **the same way** in the **workplace**, with a margin of error of 4.37% (calculated for a 95% confidence level).

Which **one** of the following statements is **correct**?

With 95% confidence, we may estimate that the percentage of adult New Zealanders who hold this opinion is somewhere between:

(1) 43.82% and 48.19%

(2) 41.63% and 50.37%

(3) 2.3% and 89.7%

(4) 32.89% and 59.11%

(5) 37.43% and 54.56%

9. From the survey responses, we may estimate that 62% of adult New Zealanders hold the opinion that males are treated **more favourably** than females in **business**. The standard error associated with this estimate is 0.0217.

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

With 95% confidence, we estimate that the proportion of adult New Zealanders who hold this opinion is somewhere between:

(1) $0.62 - (2.5 \times 0.0217)$ and $0.62 + (2.5 \times 0.0217)$

(2) $0.62 - 0.0217$ and $0.62 + 0.0217$

(3) $0.62 - (1.5 \times 0.0217)$ and $0.62 + (1.5 \times 0.0217)$

(4) $0.62 - (3 \times 0.0217)$ and $0.62 + (3 \times 0.0217)$

(5) $0.62 - (1.96 \times 0.0217)$ and $0.62 + (1.96 \times 0.0217)$

Questions 10 to **14** refer to the following additional information.

Let:

$p_{\text{Workplace}}$ be the proportion of adult New Zealanders who hold the opinion that males and females are treated the same way in the workplace

and

p_{Business} be the proportion of adult New Zealanders who hold the opinion that males and females are treated the same way in business.

A two-tailed *t*-test is carried out for no difference between $p_{\text{Workplace}}$ and p_{Business}. (Assume that it is appropriate to use a t-test.)

Use Table 3 and/or Figure 1, page 23, to answer **Questions 10** to **14**.

10. The sampling situation associated with $\text{se}(\hat{p}_{Workplace} - \hat{p}_{Business})$ is best described as:
    (1) two independent samples both of size 250.
    (2) one sample of size 500, many yes/no items.
    (3) two independent samples both of size 500.
    (4) one sample of size 250, several response categories.
    (5) one sample of size 500, several response categories.

11. Given that $\text{se}(\hat{p}_{Workplace} - \hat{p}_{Business}) = 0.0390$, the value of the test statistic, t_0, is approximately:
    (1) −4.359
    (2) 0.513
    (3) 3.590
    (4) 0.256
    (5) −7.692

12. In this *t*-test, the *P-value* is 0.0003. Which **one** of the following is an **incorrect** interpretation of this *t*-test?
    (1) At the 1% level of significance, there is no evidence of a difference between $p_{\text{Workplace}}$ and p_{Business}.
    (2) It is highly unlikely that sampling variation alone would produce a difference at least as big as the observed difference, $\hat{p}_{Workplace} - \hat{p}_{Business}$.
    (3) At the 5% level of significance, it can be claimed that there is a difference between $p_{\text{Workplace}}$ and p_{Business}.
    (4) The observed difference, $\hat{p}_{Workplace} - \hat{p}_{Business}$, is statistically significant at the 5% level of significance.
    (5) The observed difference, $\hat{p}_{Workplace} - \hat{p}_{Business}$, is so large that we can not reasonably attribute it to chance alone.

13. A 95% confidence interval for the difference $p_{\text{Workplace}} - p_{\text{Business}}$ is (0.0636, 0.2164). Which **one** of the following statements is **true**?

With 95% confidence, the proportion of adult New Zealanders who hold the opinion that males and females are treated the same way in the workplace is:

(1) somewhere between 6 and 22 percentage points higher than the proportion who hold the opinion that males and females are treated the same way in business.

(2) 14 percentage points different from the proportion of adult New Zealanders who hold the opinion that males and females are treated the same way in business.

(3) −14 percentage points different from the proportion of adult New Zealanders who hold the opinion that males and females are treated the same way in business.

(4) somewhere between 6 percentage points lower than and 22 percentage points higher than the proportion of adult New Zealanders who hold the opinion that males and females are treated the same way in business.

(5) 6% and the proportion of adult New Zealanders who hold the opinion that males and females are treated the same way in business is 22%.

14. Suppose that a 90% confidence interval for the difference $p_{\text{Workplace}} - p_{\text{Business}}$ is also to be constructed using the information in Table 3 and/or Figure 1, page 23.

When comparing the 90% confidence interval with the 95% confidence interval given in Question 13, page 26, which **one** of the following statements is **false**?

(1) The 90% confidence interval will be narrower than the 95% confidence interval.

(2) The value of the estimate, $\hat{p}_{Workplace} - \hat{p}_{Business}$, for the 90% confidence interval will be the same as that for the 95% confidence interval.

(3) The margin of error for the 90% confidence interval will be smaller than that for the 95% confidence interval.

(4) The value of the standard error, se($\hat{p}_{Workplace} - \hat{p}_{Business}$), for the 90% confidence interval will be smaller than that for the 95% confidence interval.

(5) The *t*-multiplier for the 90% confidence interval will be smaller than that for the 95% confidence interval.

Questions 15 to **19** refer to the following information.

Four single-sex and two co-educational schools in Melbourne, Australia, were asked to participate in a recent study designed to examine adolescents' attitudes towards confidentiality in the school counselling situation. All six schools were private schools. Three of the single-sex schools agreed to take part; one of the single-sex schools and both of the co-educational schools declined to take part in the study.

The students were advised that participation was voluntary and anonymous, and that they were free to withdraw from the study at any time.

Questionnaires were completed in school. Some results from the study are given in Table 4 below. It shows the percentage of students (aged 14–18 years) agreeing, disagreeing, or unsure as to whether the school counsellor should tell parents in situations of contraceptive use, and/or pregnancy.

There were 174 female respondents and 221 male respondents.

	Response			Sample
Situation	**Agree %**	**Disagree %**	**Unsure %**	**size**
Contraception				
females	13	79	8	174
males	33	52	15	221
Pregnancy				
females	15	74	11	174
males	41	43	16	221

Table 4: Adolescents' Attitudes Towards Confidentiality

Let p_{contra} be the proportion of Australian female students (aged 14–18 years) who disagree that a counsellor should tell parents in situations of contraceptive use

and p_{preg} be the proportion of Australian female students (aged 14–18 years) who disagree that a counsellor should tell parents in situations of pregnancy

15. Information from Table 4 is used to construct a 95% confidence interval for the difference $p_{contra} - p_{preg}$. The formula for the standard error of the estimate, $se(\hat{p}_{contra} - \hat{p}_{preg})$, would be calculated using the sampling situation described as:
    (1) two independent samples of sizes 221 and 174.
    (2) one sample of size 221, several response categories.
    (3) one sample of size 174, many yes/no items.
    (4) one sample of size 221, many yes/no items.
    (5) one sample of size 174, several response categories.

Let p_{female} be the proportion of all Australian female secondary school students (aged 14–18 years) who are unsure whether a counsellor should tell parents in situations of contraceptive use

and p_{male} be the proportion of all Australian male secondary school students (aged 14–18 years) who unsure whether a counsellor should tell parents in situations of contraceptive use

The results from the study are used to conduct a 2-tailed test for no difference between p_{female} and p_{male}.

16. For the purpose of calculating $se(\hat{p}_{female} - \hat{p}_{male})$, the sampling situation can be described as:

    (1) one sample of size 174, several response categories.

    (2) one sample of size 174, many yes/no items.

    (3) two independent samples of sizes 221 and 174.

    (4) one sample of size 221, several response categories.

    (5) one sample of size 221, many yes/no items.

Let p_{agree} be the proportion of all Australian male secondary school students (aged 14–18 years) who agree that a counsellor should tell parents in situations of pregnancy

and $p_{disagree}$ be the proportion of all Australian male secondary school students (aged 14–18 years) who disagree that a counsellor should tell parents in situations of pregnancy

The results from the study are used to conduct a 2-tailed test for no difference between p_{agree} and $p_{disagree}$.

17. An estimate of the difference between p_{agree} and $p_{disagree}$ is:

    (1) −1.9

    (2) −0.02

    (3) −0.2

    (4) −0.59

    (5) −0.19

18. For the purpose of calculating $se(\hat{p}_{agree} - \hat{p}_{disagree})$, the sampling situation can be described as:

    (1) one sample of size 395, several response categories.

    (2) one sample of size 395, many yes/no items.

    (3) two independent samples of sizes 221 and 174.

    (4) one sample of size 221, several response categories.

    (5) one sample of size 221, many yes/no items.

19. The expression for evaluating the *t*-test statistic for the null hypothesis, H_0: $p_{agree} - p_{disagree} = 0$, is:

(1) $\dfrac{\hat{p}_{agree} - \hat{p}_{disagree}}{se(\hat{p}_{agree} - \hat{p}_{disagree})}$

(2) $\dfrac{\hat{p}_{agree} - \hat{p}_{disagree}}{\sqrt{se(\hat{p}_{agree})^2 - se(\hat{p}_{disagree})^2}}$

(3) $\dfrac{p_{agree} - p_{disagree}}{se(\hat{p}_{agree}) + se(\hat{p}_{disagree})}$

(4) $\dfrac{p_{agree} - p_{disagree}}{se(\hat{p}_{agree} - \hat{p}_{disagree})}$

(5) $\dfrac{\hat{p}_{agree} - \hat{p}_{disagree}}{se(\hat{p}_{agree}) + se(\hat{p}_{disagree})}$

Questions 20 to **22** refer to the following information.

Research New Zealand conducts monthly surveys in which they survey a sample of 500 adult New Zealanders, aged 18 years or over. (Assume the surveys use simple random sampling.) In June 2015, the survey included questions on gender equality in such areas as the health system, the education system and in the workplace (ResearchNZ, 2015). One of the questions about a variety of situations was:

> *"In which of the following areas do you feel male and females have equal or different levels of opportunity in New Zealand?"*

With respect to opportunity in Senior Management Levels in the Public Sector, the response to the question has been cross-classified by **Opportunity** and **Sex** in Table 5 below. (Note: The exact numbers of each sex in the survey were not provided. For the purpose of this analysis we will assume that 267 females and 233 males answered this question.)

	Opportunity				
Sex	Men have more	Same/Equal	Women have more	Don't know	**Total**
Female	184	70	5	8	267
Male	93	105	23	12	233
Total	277	175	28	20	500

Table 5: Opportunity in senior management levels in the public sector

We are interested in comparing the proportion of adult New Zealand males who in June 2015 thought men had more opportunity in senior management (p_{MM}) and the proportion of adult New Zealand males who in June 2015 thought women had more opportunity in senior management (p_{MW}).

20. Use Table 5, page 29, to find the estimate for $p_{\text{MM}} - p_{\text{MW}}$.

(1) 0.300
(2) 0.486
(3) 0.329
(4) 0.290
(5) 0.140

21. The sampling situation for calculating the standard error of the estimate, se($\hat{p}_{MM} - \hat{p}_{MW}$), is:

(1) two independent samples, one of size 93 and one of size 23.
(2) one sample of size 233, several response categories.
(3) two independent samples, one of size 277 and one of size 28.
(4) one sample of size 500, several response categories.
(5) one sample of size 233, many yes/no items.

22. A 95% confidence interval for $p_{\text{MM}} - p_{\text{MW}}$ is (0.22, 0.38). Which **one** of the following statements is **false**?

(1) At the 5% level of significance we can not claim that the observed difference between $\hat{p}_{MM}$ and $\hat{p}_{MW}$ could be due to sampling variability alone.
(2) At the 5% level of significance we can claim that p_{MM} is higher than p_{MW}.
(3) It would be surprising to see a different random sample of the same size at the same time produce a result with $\hat{p}_{MW}$ larger than $\hat{p}_{MM}$.
(4) The proportion of adult New Zealand males who thought that men have more opportunities is estimated to be between 22 and 38 percentage points higher than the proportion of adult New Zealand males who thought that women have more opportunities.
(5) The margin of error for this survey is 0.16.

23. Which **one** of the following statements about a *P-value* is **false**?

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

    (2) A *P-value* measures the strength of evidence against the null hypothesis.

    (3) A relatively large test statistic results in a relatively small *P-value*.

    (4) A *P-value* is the conditional probability of observing a test statistic as extreme as that observed or even more so, if the null hypothesis were true.

    (5) A *P-value* says nothing about the size of an effect or difference.

Questions 24 to **29** refer to the following information.

The Washington Post, The Henry J Kaiser Family Foundation and *Harvard University* conducted a poll (8 March – 22 April, 2001) 'to gauge the racial attitudes of American adults'. The telephone poll surveyed 1709 adults including 779 whites, 323 African Americans, 315 Hispanics and 254 Asian Americans. Assume this sample of 1709 adults is a random sample of American adults. Two of the questions in the survey were:

> **Question 1:**
> Do you feel that African Americans have more, less or about the same opportunities in life as whites have?

and

> **Question 2:**
> Do you feel that Asian Americans have more, less or about the same opportunities in life as whites have?

The percentage results for these two questions are shown in Table 6 below.

	Response				Sample
	More %	Less %	Same %	Unsure %	size
Question 1					
White	13	27	58	2	779
African American	1	74	23	2	323
Hispanic	8	46	44	2	315
Asian American	10	44	39	7	254
Total Sample	11	35	51	2	1709
Question 2					
White	13	14	70	4	779
African American	15	38	39	8	323
Hispanic	18	24	55	3	315
Asian American	7	34	53	5	254
Total Sample	14	18	63	4	1709

Table 6: Americans' responses to racial attitudes survey

Let p_{more} be the proportion of whites who feel that African Americans have more opportunities in life than whites have

and p_{less} be the proportion of whites who feel that African Americans have less opportunities in life than whites have.

Note that these two proportions describe two of the whites' responses to **Question 1**.

24. An estimate of the difference between p_{more} and p_{less} is:

(1) 0.018
(2) -0.01
(3) -0.14
(4) 0.01
(5) -0.10

25. Information from Table 6, page 31, is used to construct a 95% confidence interval for the difference p_{more} - p_{less}. For the purpose of calculating $se(\hat{p}_{more} - \hat{p}_{less})$, the sampling situation can be described as:

(1) two independent samples of sizes 779 and 254.
(2) one sample of size 779, several response categories.
(3) one sample of size 1709, many yes/no items.
(4) one sample of size 1709, several response categories.
(5) one sample of size 779, many yes/no items.

26. A 95% confidence interval for the difference p_{more} - p_{less} is (-0.1833, -0.09668). The **best** interpretation of this interval is:

With 95% confidence, the percentage of whites who feel that African Americans have more opportunities in life than whites have is somewhere between:

(1) 10 percentage points higher than and 18 percentage points lower than the percentage who feel that African Americans have less opportunities in life than whites have.
(2) 10 and 18 percentage points.
(3) 10 and 18 percentage points higher than the percentage who feel that African Americans have less opportunities in life than whites have.
(4) 10 percentage points lower than and 18 percentage points higher than the percentage who feel that African Americans have less opportunities in life than whites have.
(5) 10 percentage points and 18 percentage points lower than the percentage who feel that African Americans have less opportunities in life than whites have.

Questions 27 and **28** refer to the following additional information.

Let $p_{question1}$ be the proportion of Asian Americans who feel that African Americans have more opportunities in life than whites have

and $p_{question2}$ be the proportion of Asian Americans who feel that Asian Americans have more opportunities in life than whites have.

Information from Table 2, page 21,, is used to conduct a 2-tailed test for no difference between $p_{question1}$ and $p_{question2}$.

27. The formula for the standard error of the estimate, $se(\hat{p}_{question1} - \hat{p}_{question2})$, is:

    (1) two independent samples of sizes 323 and 254.
    (2) one sample of size 254, several response categories.
    (3) one sample of size 323, many yes/no items.
    (4) one sample of size 323, several response categories.
    (5) one sample of size 254, many yes/no items.

28. The expression for evaluating the test statistic for the null hypothesis, H_0: $p_{question1} - p_{question2} = 0$, is:

    (1) $\dfrac{\hat{p}_{question1} - \hat{p}_{question2}}{se(\hat{p}_{question1}) + se(\hat{p}_{question2})}$

    (2) $\dfrac{\hat{p}_{question1} - \hat{p}_{question2}}{\sqrt{se(\hat{p}_{question1})^2 - se(\hat{p}_{question2})^2}}$

    (3) $\dfrac{p_{question1} - p_{question2}}{se(\hat{p}_{question1}) + se(\hat{p}_{question2})}$

    (4) $\dfrac{\hat{p}_{question1} - \hat{p}_{question2}}{se(\hat{p}_{question1} - \hat{p}_{question2})}$

    (5) $\dfrac{p_{question1} - p_{question2}}{se(\hat{p}_{question1} - \hat{p}_{question2})}$

Question 29 refers to the following additional information.

Let $p_{WhiteQ2}$ be the proportion of Whites who feel that Asian Americans have more opportunities in life than whites have

and $p_{AsianQ2}$ be the proportion of Asian Americans who feel that Asian Americans have more opportunities in life than whites have.

29. If we use information from Table 6, page 31, for the purpose of calculating $se(\hat{p}_{WhiteQ2} - \hat{p}_{AsianQ2})$, then the sampling situation can be described as:

    (1) two independent samples of sizes 779 and 254.
    (2) one sample of size 254, several response categories.
    (3) one sample of size 779, many yes/no items.
    (4) one sample of size 254, many yes/no items.
    (5) one sample of size 779, several response categories.

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

    (1) In hypothesis testing, statistical significance does not imply practical significance.

    (2) In a hypothesis test for no difference between two proportions, a very small *P-value* indicates a very large difference in the proportions.

    (3) In hypothesis testing, a non-significant test result does not imply that H_0 is true.

    (4) In hypothesis testing, large samples can lead to small *P-values* without the results having any practical significance.

    (5) In a hypothesis test for no difference between two proportions, a two-sided test should be used when the idea of doing the test has been triggered as a result of looking at the data.

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

    (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.

    (2) In hypothesis testing, large samples can lead to small *P-values* without the results having any practical significance (importance).

    (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.

    (5) In hypothesis testing, a nonsignificant test result does not imply that the null hypothesis is true.

32. 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) The *P-value* associated with a two-sided alternative hypothesis is obtained by doubling the *P-value* associated with a one-sided alternative hypothesis.

    (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) A large *P-value* says the null hypothesis is believable based on the evidence (the data) presented.

Questions 33 to **37** refer to the following information.

The *New Zealand Herald* (3 November 2016) reported a study that explored what people wanted in a relationship where 2000 British adults who were in a heterosexual relationship were surveyed. The actual numbers of females and males were not provided in the article so we will suppose that 850 males and 1150 females responded to the survey. You may assume that these 2000 respondents are a random sample of all British adults in a relationship.

Let:

p_M be the proportion of British **men** in a heterosexual relationship who would describe their partner as their best friend

and

p_W be the proportion of British **women** in a heterosexual relationship who would describe their partner as their best friend.

Using a *t*-procedure, a 95% confidence interval for $p_M - p_W$ is (0.06, 0.14).

33. Which **one** of the following statements is **false**?
    (1) Since we are dealing with proportions, the degrees of freedom for obtaining the t-multiplier used for this confidence interval are ∞.
    (2) The estimate of $p_M - p_W$ is 10% and the margin of error is 4%.
    (3) The standard error of $\hat{p}_M - \hat{p}_W$ is smaller than the margin of error of this confidence interval.
    (4) The corresponding 99% confidence interval would have a larger standard error than this confidence interval.
    (5) The corresponding 90% confidence interval would not contain 0.

34. When considering this confidence interval and a two-tailed *t*-test for no difference between p_M and p_W, which **one** of the following statements is **false**?
    (1) With 95% confidence we estimate that the proportion of British men who would describe their partner as their best friend is somewhere between 6 and 14 percentage points higher than the corresponding proportion of British women.
    (2) The observed difference between the two proportions is significant at the 5% level of significance.
    (3) There is evidence against the proportion of British men who would describe their partner as their best friend being the same as the corresponding proportion of British women.
    (4) It is plausible that there is no difference between the proportion of British men who would describe their partner as their best friend and the corresponding proportion of British women.
    (5) It is not believable that a higher proportion of British women would describe their partner as their best friend than the corresponding proportion of British men.

Questions 35 to **37** refer to the following additional information.

The *New Zealand Herald* article also included a list of “The 30 things women really want from a man in a relationship”. Some of these things are listed in Table 7. Recall that 2000 British adults (850 males and 1150 females) responded to the survey.

1	Makes you feel safe	66%
2	Completely trusts you	62%
3	Truly appreciates everything you do	59%
4	Is a good laugh	51%
⋮	⋮	⋮
30	Never leaves the car without petrol	12%

Table 7: What women want from a man in a relationship

Let p_{Laugh} be the proportion of British women in a heterosexual relationship who really want their male partner to be a good laugh (option 4 in Table 7).

The required information to find a 95% confidence interval for p_{Laugh} is shown in Figure 2.

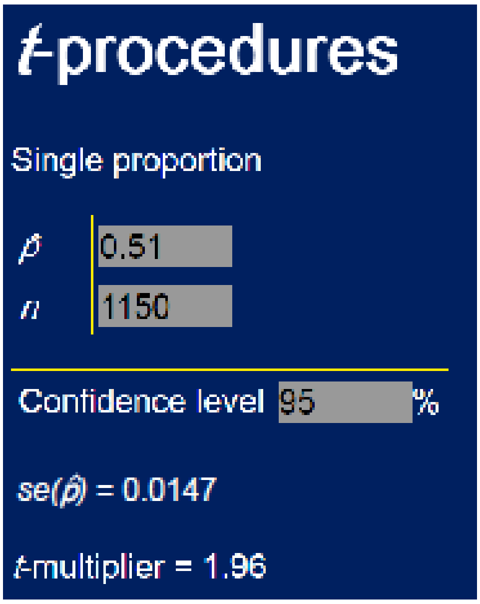


Figure 2: Screen-shot of the *t*-procedures tool

35. The 95% confidence interval for p_{Laugh} is:

(1) (−0.539, 0.539)

(2) (−0.481, 0.539)

(3) (−1.450, 2.470)

(4) (0.495, 0.563)

(5) (0.481, 0.539)

36. Which **one** of the following statements about the value se($\hat{p}_{Laugh}$) = 0.0147 is **true**?

(1) 0.0147 approximately measures the average distance between the proportion of women who want their male partner to be a good laugh in a sample of 1150 women over all possible samples of 1150 women, and the corresponding population proportion.

(2) 0.0147 approximately measures the average distance between the responses of the 1150 women in the survey who want their male partner to be a good laugh, and the corresponding population proportion.

(3) 0.0147 approximately measures the average of all sample proportions (size 1150) of women in population who want their male partner to be a good laugh.

(4) 0.0147 estimates the difference between the men and women in the survey who want their male partner to be a good laugh.

(5) 0.0147 estimates the difference between the men and women in the population who want their male partner to be a good laugh.

37. Refer to options 1 and 2 in Table 7, page 36, and let:

p_{Safe} be the proportion of British women in a heterosexual relationship who want their male partner to make them feel safe

and

p_{Trust} be the proportion of British women in a heterosexual relationship who want their male partner to completely trust them.

For the purposes of calculating se($\hat{p}_{Safe} - \hat{p}_{Trust}$), the sampling situation can be described as:

(1) one sample of size 2000, many yes/no items.

(2) one sample of size 1150, many yes/no items.

(3) one sample of size 2000, several response categories.

(4) two independent samples of sizes 1150 and 850.

(5) one sample of size 1150, several response categories.

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

(1) We make hypotheses about sample estimates.

(2) In the *t*-test, the null hypothesis, H_0, always involves an "=" sign.

(3) We investigate whether a hypothesised value is plausible in light of our sample data.

(4) If we get a *t*-test statistic with a value of 3, we know the sample estimate is 3 standard errors above the hypothesised value.

(5) A large *P-value* does not imply that H_0 is true.

Questions 39 to **41** refer to the following information.

Death Penalty Survey Results

"Should convicted murderers be put to death?"

	Australia	N.Z.
Yes	46%	42%
No	39%	41%
Can't Say	15%	17%

[Polls of 1307 Australians & 1010 New Zealanders]

39. Based on previous studies, a researcher believes that the proportion of New Zealanders who agree that convicted murderers should be put to death would be more than forty percent. The hypotheses for this test would be:

(1) H_0: p = 0.40; H_1: $p \neq 0.40$

(2) H_0: $p < 0.40$; H_1: p = 0.40

(3) H_0: p = 0.40; H_1: $p > 0.40$

(4) H_0: $p < 0.40$; H_1: $p \neq 0.40$

(5) H_0: $p < 0.40$; H_1: $p > 0.40$

40. Let's assume instead that the researcher tested H_0: p = 0.40 versus H_1: $p \neq 0.40$ where p = the proportion of New Zealanders who agree that convicted murderers should be put to death. If the standard error, se($\hat{p}$) = 0.0155, then the value of the t-test statistic, t_0, and the degrees of freedom, *df*, to be used to calculate the *P-value* are given by:

(1) t_0 = 1.290, *df* = ∞

(2) t_0 = 1.290, *df* = 1009

(3) t_0 = -1.290, *df* = ∞

(4) t_0 = 1.290, *df* = 1.96

(5) t_0 = -1.290, *df* = 1009

A difference considered by the researcher was between the proportion of Australians supporting the *death penalty for convicted murderers* and the proportion of New Zealanders supporting the *death penalty for convicted murderers*.

41. To test for a difference in the two proportions given above the hypotheses would be:

(1) $H_0: \mu_1 - \mu_2 = 0$ vs $H_1: \mu_1 - \mu_2 > 0$

(2) $H_0: \hat{p}_1 - \hat{p}_2 = 0$ vs $H_1: \hat{p}_1 - \hat{p}_2 \neq 0$

(3) $H_0: p_1 - p_2 = 0$ vs $H_1: p_1 - p_2 \neq 0$

(4) $H_0: \hat{p}_1 - \hat{p}_2 < 0$ vs $H_1: \hat{p}_1 - \hat{p}_2 > 0$

(5) $H_0: p_1 - p_2 > 0$ vs $H_1: p_1 - p_2 = 0$

42. The general formula for a Normality-based confidence interval for the difference between two proportions is:

$$\hat{p}_1 - \hat{p}_2 \pm t \times se(\hat{p}_1 - \hat{p}_2)$$

Which **one** of the following statements about Normality-based confidence intervals for the difference between two proportions is **false**?

(1) The value of the t-multiplier depends on the confidence level.
(2) In the long run, if we repeatedly take samples and calculate a 95% confidence interval from each sample, we expect that 95% of the intervals will contain the true value of $p_1 - p_2$.
(3) The confidence interval is centred on $\hat{p}_1 - \hat{p}_2$.
(4) The value of the t-multiplier depends on the sample size.
(5) The size of the standard error depends on the sampling situation.

Questions 43 to **47** refer to the following information.

A survey of 2171 men and 2412 women in Auckland in the early 1990s found that 10% of men abstained from drinking alcohol compared with 16% of women.

We wish to compare the proportion of female abstainers, p_{female}, with the proportion of male abstainers, p_{male}.

43. The sampling situation is **best** described as:

(1) two independent samples.
(2) one sample, several response categories.
(3) one sample, many yes/no items.
(4) two samples, several response categories.
(5) two samples, many yes/no items.

44. Based on the data, a 95% confidence interval for $p_{\text{female}} - p_{\text{male}}$ is (0.041, 0.079). Which **one** of the following statements is **false**?

(1) Based on the data, a 99% confidence interval would be wider than 0.038.
(2) The point estimate of $p_{\text{female}} - p_{\text{male}}$ is 0.06.
(3) We are confident that the proportion of female abstainers is larger than the proportion of male abstainers.
(4) Zero is a plausible value for $p_{\text{female}} - p_{\text{male}}$.
(5) Based on the data, a 95% confidence interval for $p_{\text{male}} - p_{\text{female}}$ is (−0.079, −0.041).

45. Consider the *P-value* associated with a two-tailed *t*-test for no difference between p_{female} and p_{male}. Based on the confidence interval in Question **13**, which **one** of the following statements is **true**?

(1) The *P-value* is much less than 5%.
(2) The *P-value* is around 10%.
(3) We do not have enough information to determine the approximate *P-value*.
(4) The *P-value* is greater than 5%.
(5) The *P-value* is just below 5%.

Questions 46 and **47** refer to the following **additional** information.

Overall, 13% of the 4583 people surveyed abstained from alcohol. We are interested in p_{abstain}, the proportion of people who abstain from alcohol.

A *t*-test of the hypotheses:

$$H_0 : p_{\text{abstain}} = 0.1$$

$$H_1 : p_{\text{abstain}} \neq 0.1$$

gives a *t*-test statistic of 6.04 and a *P-value* of 0.000.

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

(1) The test is significant at the 1% level of significance.
(2) If the null hypothesis is true, it is extremely unlikely that sampling variability would give values further away from the hypothesised value, 0.1, than our sample estimate.
(3) The sample estimate, $\hat{p}_{abstain}$, is approximately 6 standard errors above the hypothesised value, 0.1.
(4) If the null hypothesis is true, sampling variability could never give values further away from the hypothesised value, 0.1, than our sample estimate.
(5) The hypothesised value, 0.1, would be outside a 99% confidence interval for p_{abstain}.

47. Which **one** of the following statements gives the **best** interpretation of the hypothesis test result?

(1) There is some evidence that the true population proportion, p_{abstain}, is **not** 0.1.
(2) There is very strong evidence that the true population proportion, p_{abstain}, is **not** 0.1.
(3) There is no evidence that the true population proportion, p_{abstain}, is **not** 0.1.
(4) There is very strong evidence that the true sample proportion, $\hat{p}_{abstain}$, is **not** 0.1.
(5) There is very strong evidence that the true population proportion, p_{abstain}, is 0.1.

Questions 48 to **50** refer to the following information.

In 2008, as part of the first World Internet Project New Zealand survey, the Institute of Culture, Discourse and Communication published *The Internet in New Zealand 2007 Final Report*.

A random sample of 1430 people aged 12 and over were surveyed via telephone and were asked many questions in order to ascertain New Zealanders' usage of, and attitudes towards, the Internet in 2007.

One of the questions asked:

> *'How important is the Internet in your daily life — important, neutral, not important?'*

The respondents were also categorised by ethnicity. Table 2 shows the response to this question.

	Importance of the Internet			
Ethnicity	Important	Neutral	Not important	**Total**
Pakeha	476	137	302	915
Maori	46	25	44	115
Pasifika	50	26	10	86
Asian	130	16	11	157
Other	77	33	47	157
Total	779	237	414	1430

Table 8: Importance of the Internet in daily life

Assume these 1430 respondents form a random sample from the population of all New Zealanders aged 12 and over.

Let:

> $p_{Pasifika}$ be the true proportion of Pasifika New Zealanders aged 12 and over who think that the Internet is important in their daily life

and

> p_{Pakeha} be the true proportion of Pakeha New Zealanders aged 12 and over who think that the Internet is important in their daily life.

48. An estimate of $p_{Pasifika}$ - p_{Pakeha}, rounded to two decimal places, is:

(1) 0.30
(2) 0.06
(3) −0.30
(4) −0.58
(5) −0.06

49. The sampling situation associated with $se(\hat{p}_{Pasifika} - \hat{p}_{Pakeha})$ is best described as:

(1) one single sample of size 779, several response categories.
(2) two independent samples, of sizes 86 and 915.
(3) one single sample of size 779, many yes/no items.
(4) one single sample of size 1001, several response categories.
(5) two independent samples, of sizes 50 and 476.

50. A 95% confidence interval for $p_{Pasifika}$ - p_{Pakeha} is (−0.0480, 0.1704). Which **one** of the following statements is **true**?

(1) It would be surprising to see a different sample of 1430 New Zealanders aged 12 and over produce a poll result with $\hat{p}_{Pasifika}$ smaller than $\hat{p}_{Pakeha}$.
(2) The difference between $\hat{p}_{Pasifika}$ - $\hat{p}_{Pakeha}$ is significant at the 5% level of significance.
(3) The estimated difference between $p_{Pasifika}$ - p_{Pakeha} could just be sampling error.
(4) The margin of error for this 95% confidence interval for $p_{Pasifika}$ - p_{Pakeha} is approximately 22%.
(5) Since $\hat{p}_{Pasifika}$ is bigger than $\hat{p}_{Pakeha}$, we can claim that $p_{Pasifika}$ is bigger than p_{Pakeha}.

51. Which **one** of the following statements about a *P-value* is **false**?

(1) The larger a *P-value*, the stronger the evidence against the null hypothesis.
(2) A *P-value* measures the strength of evidence against the null hypothesis.
(3) A relatively large test statistic results in a relatively small *P-value*.
(4) A *P-value* is the conditional probability of observing a test statistic as extreme as that observed or even more so, if the null hypothesis were true.
(5) A *P-value* says nothing about the size of an effect or difference.

Questions 52 to **55** refer to the following information in this appendix.

Researchers (Ridker et al., 2008) conducted a study to see whether taking a particular drug reduces the rate of a first major cardiovascular event (FMCE).

17 802 apparently healthy men and women were randomly assigned to receive 20 mg of this drug daily or to receive a placebo. The researchers recorded whether or not an FMCE occurred within five years for each participant and found that 142 of the 8901 subjects in the drug group had an FMCE within five years compared to 251 of the 8901 subjects in the placebo group.

Let:

p_{drug} be the proportion who would have had an FMCE within five years if all 17 802 subjects received 20mg of the drug daily

and

p_{placebo} be the proportion who would have had an FMCE within five years if all 17 802 subjects received a placebo daily.

Using the *t*-procedure tool, a 95% confidence interval for $p_{\text{drug}} - p_{\text{placebo}}$ is calculated to be $(-0.0166, -0.0079)$.

52. The sampling situation associated with se($\hat{p}_{drug} - \hat{p}_{placebo}$) can be described as:
    (1) one sample of size 17 802, several response categories.
    (2) two independent samples, of sizes 142 and 251.
    (3) two independent samples, both of size 8901.
    (4) one sample of size 8901, many yes/no items.
    (5) one sample of size 8901, several response categories.

53. When considering a two-tailed *t*-test for no difference between p_{drug} and p_{placebo}, which **one** of the following statements is **false**?
    (1) At the 5% level of significance, it may be claimed that p_{drug} is less than p_{placebo}.
    (2) If a similar (second) study was conducted using the same number of participants, it would be surprising if the second study produced a result with $\hat{p}_{drug}$ greater than $\hat{p}_{placebo}$.
    (3) The observed difference between $\hat{p}_{drug}$ and $\hat{p}_{placebo}$ is significant at the 5% level of significance.
    (4) At the 10% level of significance, it is not plausible that the observed difference between $\hat{p}_{drug}$ and $\hat{p}_{placebo}$ could be due to chance alone.
    (5) The observed difference between $\hat{p}_{drug}$ and $\hat{p}_{placebo}$ is not significant at the 10% level of significance.

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

(1) This study is an experiment and we can conclude that when an apparently healthy adult is treated with the drug it will lower the rate of having a first major cardiovascular event within five years.

(2) This study is an experiment and, for the 17 802 participants, the observed reduction in the proportion in the drug (treatment) group who had a first major cardiovascular event within five years compared to the proportion in the placebo group can be attributed to the drug.

(3) This study is an observational study but, for the 17 802 participants, the observed reduction in the proportion in the drug (treatment) group who had a first major cardiovascular event within five years compared to the proportion in the placebo group cannot be attributed to the drug.

(4) This study is an experiment but, for the 17 802 participants, the observed reduction in the proportion in the drug (treatment) group who had a first major cardiovascular event within five years compared to the proportion in the placebo group cannot be attributed to the drug.

(5) This study is an observational study and, for the 17 802 participants, the observed reduction in the proportion in the drug (treatment) group who had a first major cardiovascular event within five years compared to the proportion in the placebo group can be attributed to the drug.

55. Suppose the study had used 5000 participants (i.e. 2500 in each group), instead of the 17 802 participants it actually used, and that $\hat{p}_{drug}$ and $\hat{p}_{placebo}$ (i.e. the observed group proportions) remained unchanged.

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

If 5000 participants took part in the study instead of 17 802 participants, then the resulting 95% confidence interval for $p_{\text{drug}} - p_{\text{placebo}}$ would be:

(1) wider and the estimation statement would be less precise.

(2) narrower and the estimation statement would be less precise.

(3) narrower and the estimation statement would be more precise.

(4) wider and the estimation statement would have the same precision.

(5) unchanged and the estimation statement would have the same precision.

Question 56 refers to the following information.

The Auckland Marathon is an annual running event which involves thousands of runners. Within the event, there are five different races: the full marathon (42 km), the half marathon (21 km), the traverse (12 km), the challenge (5 km), and the kids marathon (2.2 km). Information about each runner who enters the Auckland Marathon is made available to the public each year on the website www.aucklandmarathon.co.nz. Data about a random sample of 1000 runners who entered the Auckland Marathon in 2015 was scraped (with permission) from this website to create a data set.

An excerpt of this sample data set is shown in Figure 3.

bib number	name	gender	division	age division	distance in km	completed event	time in hours	place	mean pace km per hr
11055	KATIE CARROLL	F	F0034	Up to 34	21	No			
755	JULIE CAHILL	F	F4549	45 to 49	21	Yes	3.4	4,793	6.2
10610	VICKY FOSTER	F	F0034	Up to 34	21	No			
9998	GEORJIARNA MACCORMACK	F	F0034	Up to 34	21	No			
4528	DEON STOLTZ	M	M6064	60 to 64	42	Yes	7	1,507	6
4648	GARRY DONOGHUE	M	M7074	70 to 74	42	Yes	4.7	1,069	8.9
25459	JOHANN RLYNON	F	F0034	Up to 34	12	No			
21940	FERNANDA STEWART	F	F0034	Up to 34	12	No			
21872	GABOR PERJESSY	M	M0034	Up to 34	12	Yes	1.1	235	10.9
3804	CHRISTOPHER ABESAMIS	M	M4044	40 to 44	42	Yes	4.2	694	10
20910	LIZA CLARK	F	F5054	50 to 54	12	Yes	1.3	566	9.2
25236	HELEN MARINOVICH	F	F5054	50 to 54	12	Yes	2.1	1,771	5.7
14887	ZHENG DONG	M	M0034	Up to 34	21	Yes	1.9	937	11.1
11737	ELLA STENSNESS	F	F0034	Up to 34	21	No			
20282	HEATHER IRVINE	F	F4044	40 to 44	12	Yes	2	1,703	6
11773	ALEXANDRA BARNETT	F	F0034	Up to 34	21	Yes	2.2	2,365	9.5
4364	SAMUEL EASTON	M	M0034	Up to 34	42	Yes	4.7	1,044	8.9
388	SHARON RANDELL	F	F4549	45 to 49	42	Yes	4	594	10.5
2660	TREVOR THEYS	M	M4549	45 to 49	21	Yes	2.3	2,806	9.1

Figure 3: Excerpt of the sample data set created

56. Assuming the conditions for the underlying assumptions are satisfied, which **one** of the following types of analysis would be **most appropriate** to investigate the relationship between **gender** and **completed event**?

(1) Correlation.

(2) Simple linear regression.

(3) *t*-test on a difference between two proportions.

(4) *F*-test for one-way analysis of variance.

(5) One-sample *t*-test on a proportion.

For "Spot the Analysis" practice across Chapters 7 to 10, use Anna Fergusson's app here: www.tinyURL.com/stats-spot

Questions 57 to **59** refer to the following information.

Students enrolled in stage one statistics courses at the University of Auckland were surveyed regarding their access to, and experience with, computers. The survey was included as a question in an assignment, and students were given marks for completing it (irrespective of the answers they gave). Staff administering the courses wished to use the results of this survey to draw conclusions about future stage one statistics students.

One question asked: 'At the start of the course, how would you describe your Excel experience?'. A total of 918 students answered this question. Each of the 918 answers were classified according to the response given by the student, and the stream the student attended. The results are given in the table below, where 101G, 108 and 101 refer to the various streams.

	Stream			
Response	101G	101	108	Total
None	15	36	102	153
Very Little	44	89	119	252
Some	74	150	200	424
Lots	9	29	51	89
Total	142	304	472	918

57. A hypothesis test is performed on the data for no difference between the proportion of **101** students who responded **None** and the proportion of **108** students who responded **None**. Which **one** of the following statements about this hypothesis test is **true**?

    (1) The degrees of freedom used depends on the number of 101 and 108 students in the sample.

    (2) The one sample t-test should be used.

    (3) The test should be two-tailed.

    (4) The test could only be used to show a difference existed in the sample proportions.

    (5) An appropriate null hypothesis is that the difference between the proportion of 101 students who responded None and the proportion of 108 students who responded None, is not zero.

58. The standard error for the difference in the proportions tested in Question 57 would be calculated using the sampling situation described as:

    (1) one sample of size 304, several response categories.

(2) one sample of size 472, many yes/no items.

(3) one sample of size 304, many yes/no items.

(4) two independent samples of sizes 304 and 472.

(5) one sample of size 472, several response categories.

59. The *P-value* for the statistical test mentioned in Question 57 is 0.004. Which **one** of the following statements gives the **best** interpretation of this *P–value*?

(1) There is some evidence that the underlying proportions of 101 students and 108 students that would respond None are different.

(2) There is strong evidence that the sample proportions of 101 students and 108 students that responded None are different.

(3) There is strong evidence that the underlying proportions of 101 students and 108 students that would respond None are different.

(4) There is no evidence that the underlying proportions of 101 students and 108 students that would respond None are different.

(5) There is weak evidence that the underlying proportions of 101 students and 108 students that would respond None are different.

ANSWERS

1. **(2)**	2. **(4)**	3. **(2)**	4. **(4)**	5. **(2)**	6. **(1)**
7. **(3)**	8. **(2)**	9. **(5)**	10. **(2)**	11. **(3)**	12. **(1)**
13. **(1)**	14. **(4)**	15. **(3)**	16. **(3)**	17. **(2)**	18. **(4)**
19. **(1)**	20. **(1)**	21. **(2)**	22. **(5)**	23. **(1)**	24. **(3)**
25. **(2)**	26. **(5)**	27. **(5)**	28. **(4)**	29. **(1)**	30. **(2)**
31. **(4)**	32. **(4)**	33. **(4)**	34. **(4)**	35. **(5)**	36. **(1)**
37. **(2)**	38. **(1)**	39. **(3)**	40. **(1)**	41. **(3)**	42. **(4)**
43. **(1)**	44. **(4)**	45. **(1)**	46. **(4)**	47. **(2)**	48. **(2)**
49. **(2)**	50. **(3)**	51. **(1)**	52. **(3)**	53. **(5)**	54. **(2)**
55. **(1)**	56. **(3)**	57. **(3)**	58. **(4)**	59. **(3)**	

WHAT SHOULD I DO NEXT?

- **Do Question 2 of Assignment 3!**
- **Go through the proportions material in Chapter 6 and 7 of your Lecture Workbook. (Chapter 6: pages 7 & 8, 12-18, 20, 23, 25 & 26; Chapter 7: Pages 5 & 6, 11-16, 18-20, 22).**
- **Try Chapter 6 & 7 questions from three of the past five exams that are relevant to this workshop (i.e. on proportions not means!)**

FORMULAE

Confidence intervals and *t*-tests

Confidence interval: $estimate \pm t \times \text{se}(estimate)$

t-test statistic: $t_0 = \dfrac{estimate - hypothesised\ value}{standard\ error}$

Applications:

1. Single mean $\boldsymbol{\mu}$: $estimate = \overline{x}$; $df = n - 1$

2. Single proportion $\boldsymbol{p}$: $estimate = \widehat{p}$; $df = \infty$

3. Difference between two means $\boldsymbol{\mu_1 - \mu_2}$: (independent samples)

   $estimate = \overline{x}_1 - \overline{x}_2$; $df = \min(n_1 - 1, n_2 - 1)$

4. Difference between two proportions $\boldsymbol{p_1 - p_2}$:

   $estimate = \widehat{p}_1 - \widehat{p}_2$; $df = \infty$

   Situation (a): *Proportions from two independent samples*

   Situation (b): *One sample of size n, several response categories*

   Situation (c): *One sample of size n, many yes/no items*

The *F*-test (ANOVA)

F-test statistic: $f_0 = \dfrac{s_B^2}{s_W^2}$; $df_1 = k - 1,\ df_2 = n_{\text{tot}} - k$

The Chi-square test

Chi-square test statistic: $\chi_0^2 = \displaystyle\sum_{\text{all cells in the table}} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$

Expected count in cell $(i, j) = \dfrac{R_i C_j}{n}$

$df = (I - 1)(J - 1)$

Regression

Fitted least-squares regression line: $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x$

Inference about the intercept, β_0, and the slope, β_1: $df = n - 2$