please get a handout from the back!

Stats 101/101G/108 Workshop

Hypothesis Tests: *Proportions* [HTP]

we'll start @ 9.35am...

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?

in person, on the phone or via zoom!

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.

Types of Variables

Numeric *or Scale (SPSS)*

(measurements and counts)

Continuous

(measurements; few repeated values)

No gaps between possible values, if measured sufficiently accurately.

Examples

Weight; Height; Length; Temperature; Total course mark (out of 100); Age [large range]

Discrete

(usually counts; many repeated values)

Gaps between each of the values it can take.

Examples

Number of siblings; Number of cars per household; Assignment mark (out of 10); Age [small range]

Categorical

(define groups)

Nominal (SPSS)

(no idea of order)

Examples

Gender; Ethnicity; Degree (BA, BCom, BSc, etc.); Major (Chemistry, Statistics, etc.)

Ordinal (SPSS)

(fall in natural order)

Examples

Size (small, medium, large); Income bracket; Grade (A+, A, ...); Age group (<25, 25-39, 40+)

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

Exploring relationships between two variables

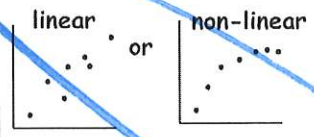
Two Numeric	Numeric versus Categorical	Two Categorical

Exploratory tools to use to explore relationship/s

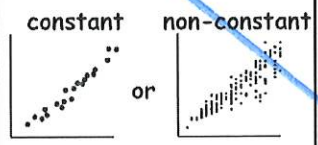
Scatter plot	Side-by-side plots on the same scale: any n { • Dot plots $n \geq 20$ { • Stem-and-leaf plots • Box plots $n \geq 50$ { • Histograms	Two-way table of counts and/or Bar graphs of proportions (on rows/columns)

Features to look for:

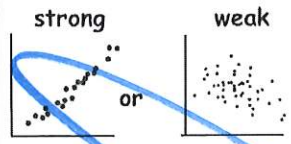
- Trend



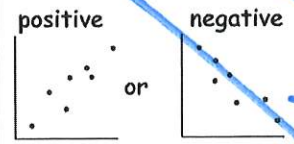
- Scatter



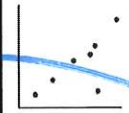
- Strength of relationship



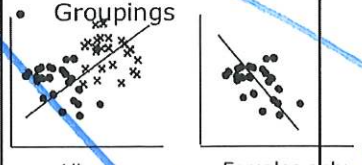
- Association



- Outliers



- Groupings



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

- 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

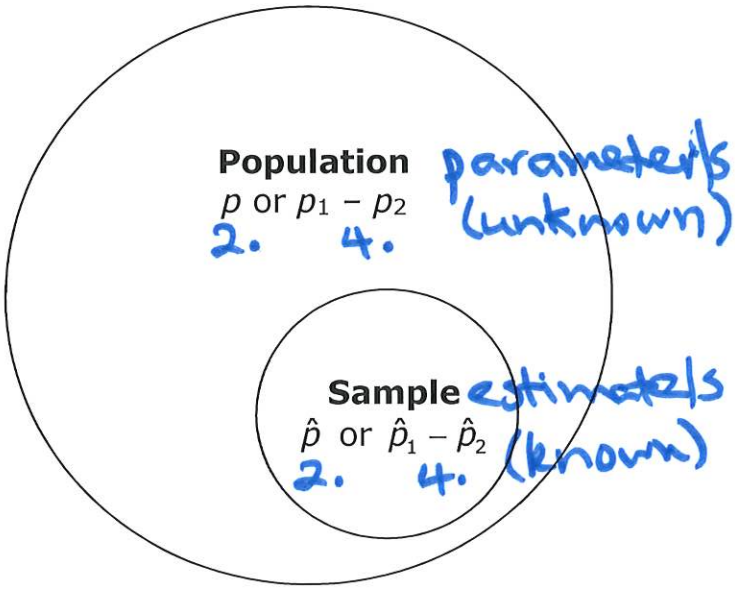
We use statistics to find out about the real world and aspects of it specific to our area of interest. Statistical tools allow us to deal with the **uncertainty** present in all samples due to **sampling variation** which occurs because we are unable to survey the entire population of interest.

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, $p_1 - p_2$

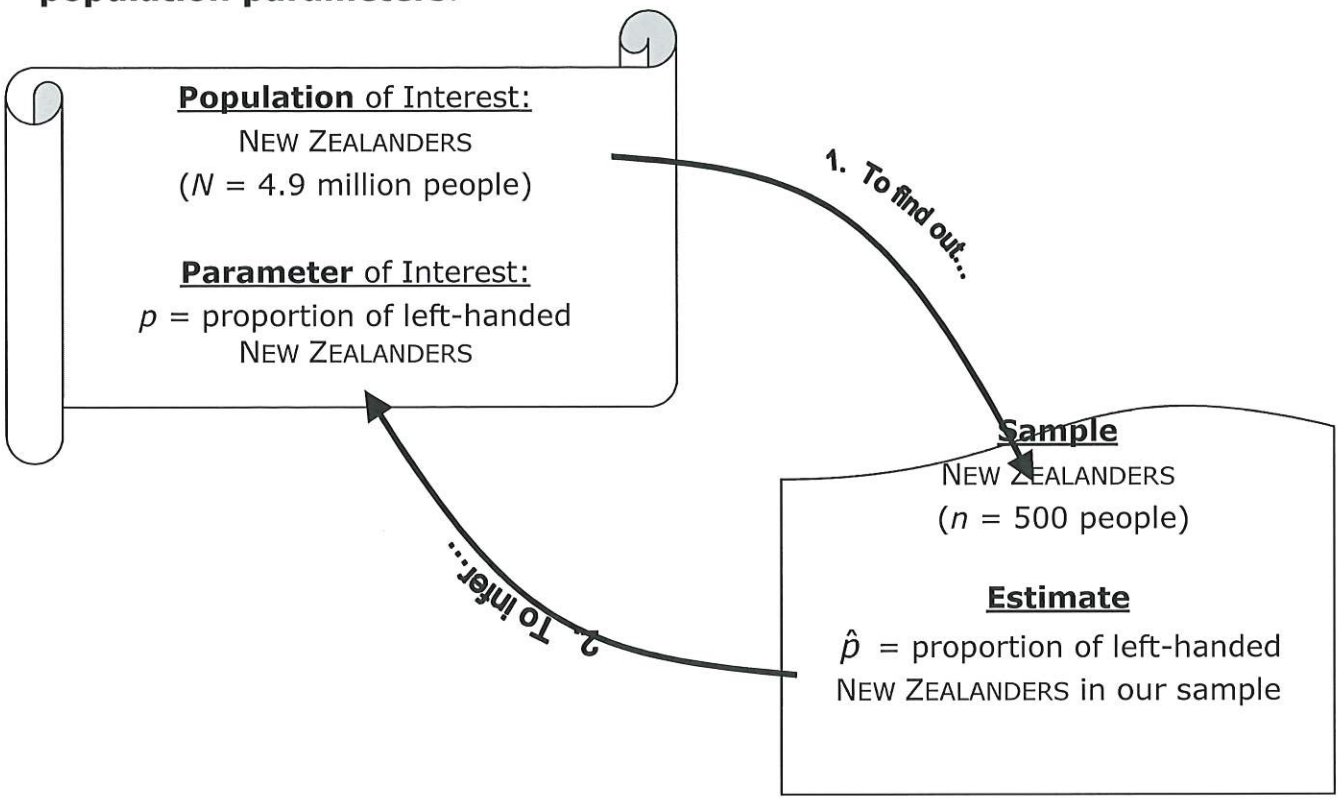


This means that the **parameter** of interest
is an ***unknown* numerical characteristic**
for that particular population.

To estimate an ***unknown* 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 $\bar{}$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.

hypothesis test

- 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**:

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

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

see back page for Formulae Sheet

- 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**:

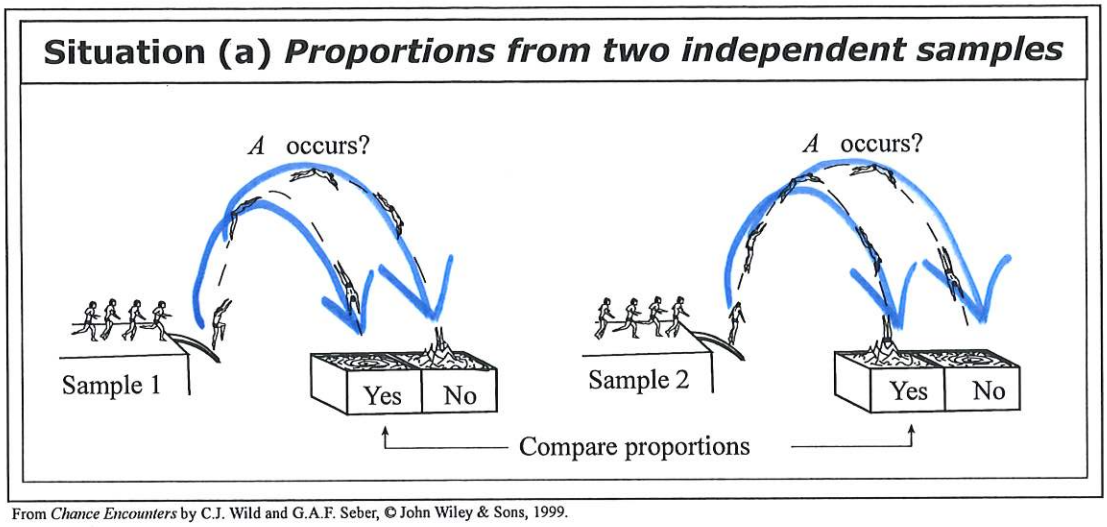
- Use: $\textit{estimate} \pm t \times \textit{se(estimate)}$
- Use the estimate from Step 4 and the standard error from Step 5.
- Use the appropriate t-multiplier. (Will be provided)

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

10. Give an overall **conclusion**.

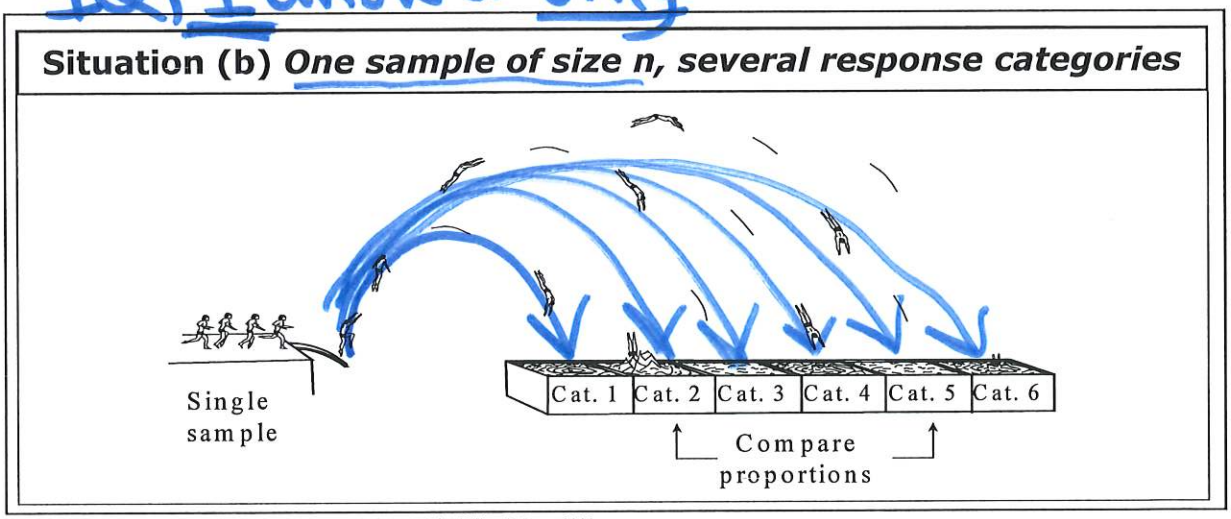
- There are four different types of problem:
  ~~1. Single mean~~ 2. Single proportion ~~3. Difference between two means~~
  4. Difference between two proportions:
  Situation (a) ***Proportions from two independent samples***
  Situation (b) ***One sample of size n, several response categories***
  Situation (c) ***One sample of size n, many yes/no items***

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



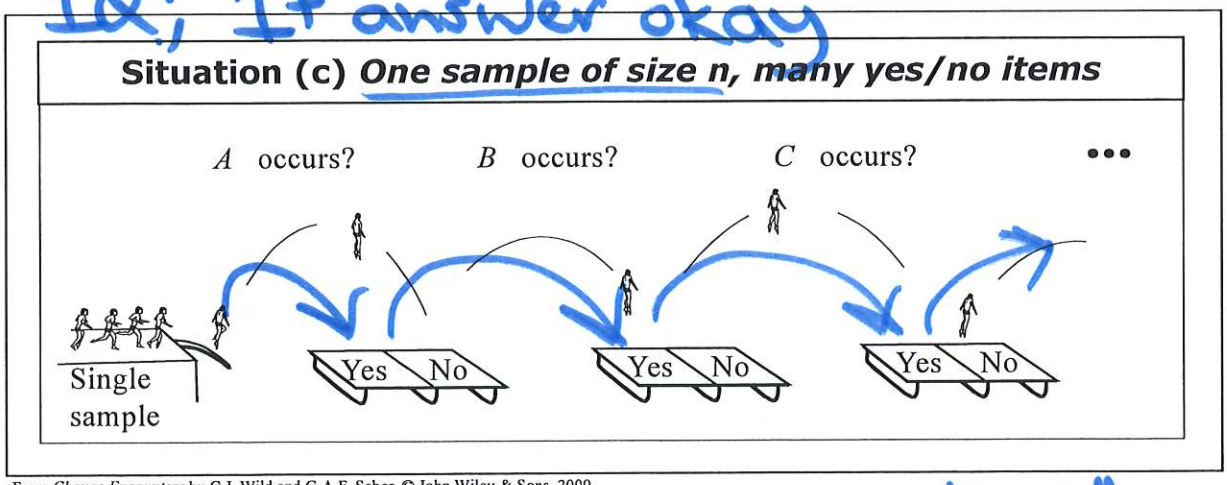
From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 1999.

1Q; 1 answer only



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

1Q; 1+ answer okay



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

2 separate Qs, can say "yes" to each, if so wish.

Identifying the Sampling Situations: