Please get a handout from the back!

Stats 101/101G/108 Workshop

Hypothesis Tests: *Means* [HTM]

We'll start @1.05pm...

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-HTM** (*)

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).

(*) in person, on the phone or via Zoom!

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** 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** 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!)
  - this morning!
  - ***Hypothesis Tests: Proportions*** **www.tinyURL.com/stats-HTP** 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** 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: *Means* [HTM]

> This material builds on a couple of workshops already held in the first half of this semester, which you may or may not have attended.
>
> For more practice on how to **quantify the size of a single mean or difference between two means**, see the ***Confidence Intervals: Means*** workshop materials. If you want to learn more about how to **explore our sample data**, see the ***Exploratory Data Analysis*** workshop.

Types of Variables

(A) (D) 1. & 3. means

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]

(B) (E) 2. & 4. proportions

Categorical

(define groups)

Nominal (SPSS)

(no idea of order)

Examples

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

Ordinal (SPSS)

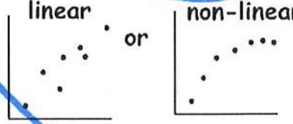
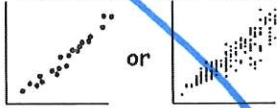
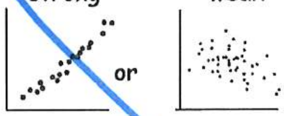
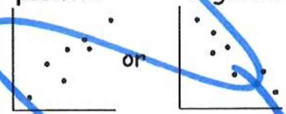
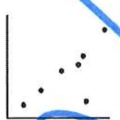
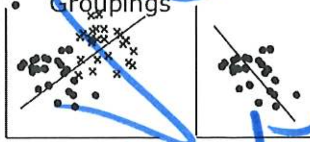
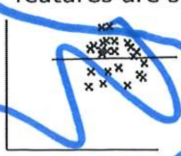
(fall in natural order)

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 **linear** or **non-linear**  • Scatter **constant** or **non-constant**  • Strength of relationship **strong** or **weak**  • Association **positive** or **negative**  • Outliers  • Groupings  All Females only	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  Males only	• Most common and least common combinations • Differences in distributions (e.g. row and/or column bar graphs)

NTP (B)&(E); Chi-sq test for indep CST; regression! RC; t-test! → cate: 2 levels/groups 3.(D); or F-test! → cate: 3+ levels or groups (E); NTM

Recall that two **numerical summaries** of **centre** are:

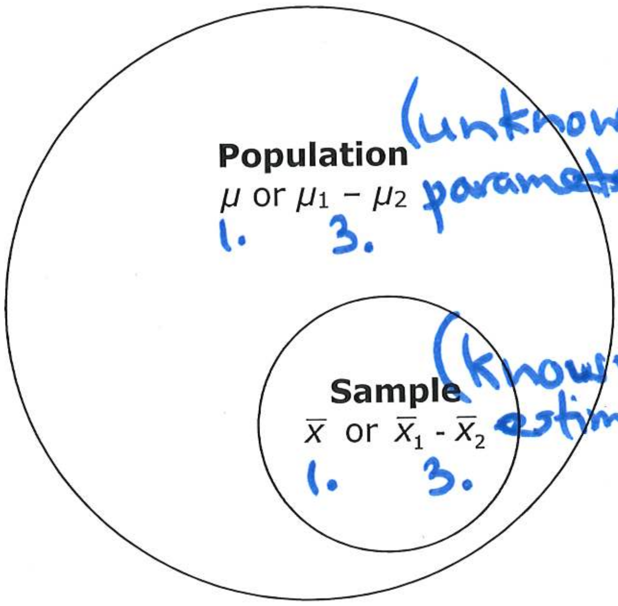
- Sample mean, $\bar{x}$ (also known as the average or expected value)
  $\sum \frac{x_i}{n}$ – affected by outliers
- Median (= Med – also known as the 50th percentile) = middle number of the ordered data – not affected by outliers

t-tests by Hand – One and Two Mean/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:

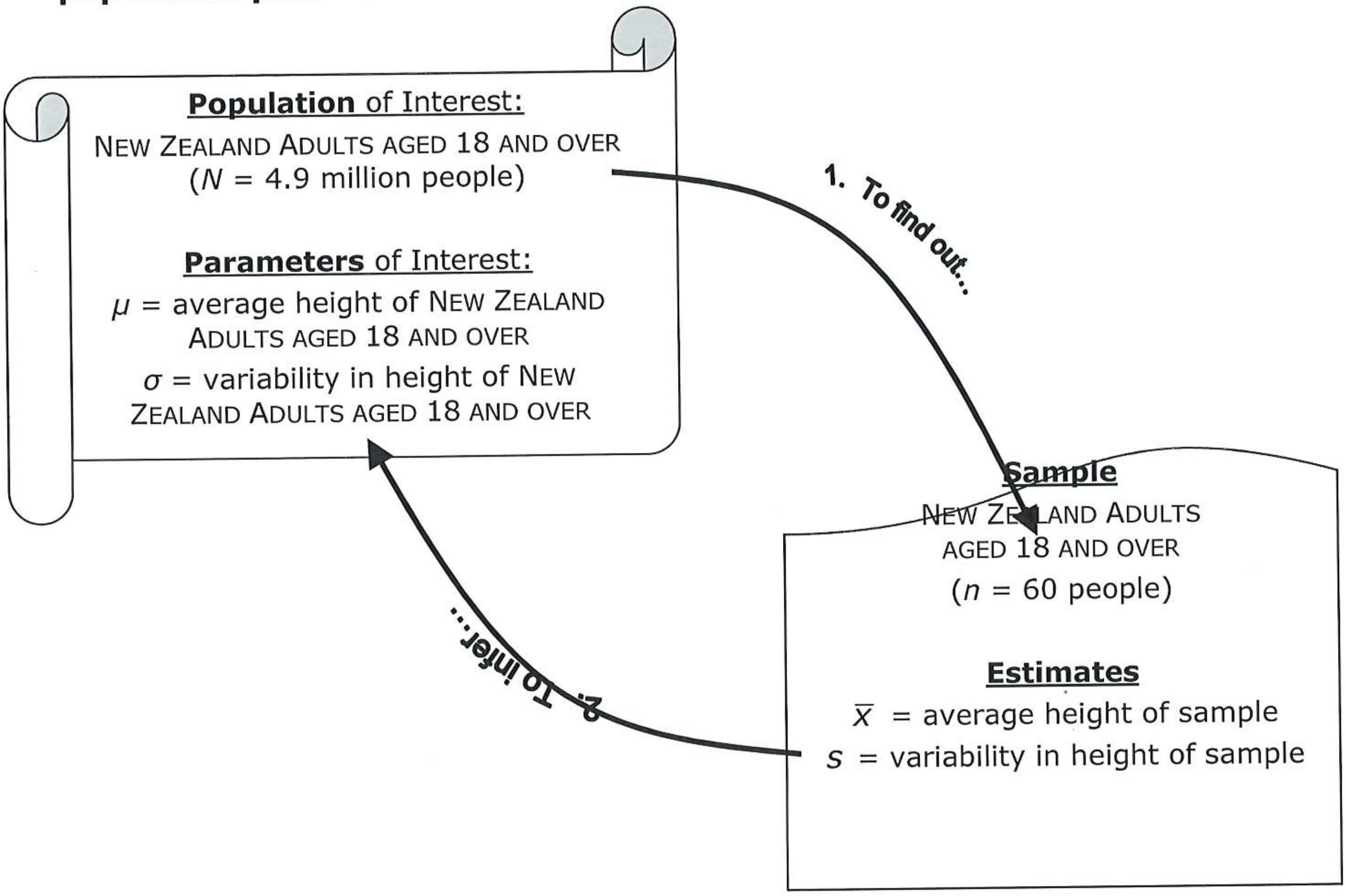
1. Population mean, μ
3. Difference in population means, $\mu_1 - \mu_2$

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:

1. Sample mean, $\bar{x}$
3. Difference in sample means, $\bar{x}_1 - \bar{x}_2$

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**:
   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.

   see back page for Formulae Sheet
6. Estimate the ***P*-value**. (Will be provided)
7. **Interpret** the *P-value*. (see page 10)
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***

Step 1

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

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

Steps 2 & 3

The null hypothesis, H_0 Step 2

- ✓ 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)*

c.f. blue pg 17, Ch 7

1. $H_0: \mu =$ 500g ← some # from story!   3. $H_0: \mu_1 - \mu_2 =$ 0

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

The alternative hypothesis, H_1 Step 3

- ✓ Specifies the type of departure from H_0 that we expect to detect.
- ✓ Corresponds to the research hypothesis.
- ✓ There are three different types:
  - H_1: *parameter ≠ hypothesised value (some number)* } 2-sided
  - H_1: *parameter > hypothesised value (some number)* } 1-sided
  - H_1: *parameter < hypothesised value (some number)* } 1-sided

1. $H_1: \mu$ ≠ 500g   3. $H_1: \mu_1 - \mu_2$ > 0

- ✓ When do we use a 1-sided alternative hypothesis?
  * if in doubt ≠   * data ≠   * research >< ← exam will have a sentence
- ✓ It's the **interesting** thing – **there is an** effect or difference.

c.f. blue pg 13, RNJ, Ch 7

Step 4 (and Step 8)

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

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

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 sample size(s):

Degrees of Freedom	
1.	$df = n - 1$
3.	$df = minimum(n_1 - 1, n_2 - 1)$

***e.g.* $\boldsymbol{n_1 = 50}$ and $\boldsymbol{n_2 = 30}$:**

$$df = minimum(n_1 - 1, n_2 - 1)$$
$$= min(50 - 1, 30 - 1)$$
$$= min(49, 29)$$
$$= 29$$

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

see back page for Formulae Sheet

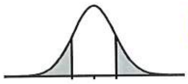
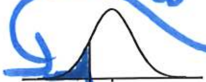
Step 6

- 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

A. 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.

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

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

(2) Statistical significance is the same as practical significance.

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

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

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