Step 1

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

Parameter
~~1. Single mean μ:~~
2. Single proportion $\boldsymbol{p}$:
~~3. Difference between two means $\mu_1 - \mu_2$: (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)*

2. H_0: p = 0.1 ← some # from story

4. H_0: $p_1 - p_2$ = 0

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

2. H_1: p > 0.1

4. H_1: $p_1 - p_2$ ≠ 0

✓ When do we use a 1-sided alternative hypothesis?
* if in doubt ≠ * data ≠ * research → context → prev. studies → knowledge

✓ 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 $\boldsymbol{p}$:	*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 $\boldsymbol{p_1 - p_2}$:	*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* = $\bar{x}_1 - \bar{x}_2$~~	~~$df = \text{minimum}(n_1 - 1, n_2 - 1)$~~
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: $$t_0 = \frac{\overset{\text{step 4}}{estimate} - \overset{\text{steps 2 \& 3}}{hypothesised\ value}}{std\ error\ \text{step 5}}$$
  - ✓ 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:**

memorise!

✓ 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.

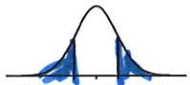
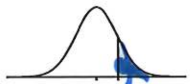
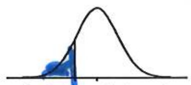
similar

✓ 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, given 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**?

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

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

T (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.

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

F (5) The ~~larger~~ smaller 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

memorise!

0.03

- 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	.049 ✓
> 0.05	Nonsignificant	Do not reject H_0	.051 ✗

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:

margin of error

Step 8

$\boxed{estimate \pm \underbrace{t \times se(estimate)}}$

step 4 step 5

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

- Whether we are investigating means or proportions
- The desired level of confidence 95%
- The degrees of freedom:

Estimate	Degrees of Freedom
~~1. $estimate = \bar{x}$~~	~~$df = n - 1$~~
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**? 2.

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

T (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.

T (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.

T (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.

F (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.

2. 4.

- You can do all kind of CI's, 90%, 95%, 99%...
- Increasing the confidence level will **increase** the width of the interval.
- Increasing the sample (n) size will make the confidence interval more precise. narrower

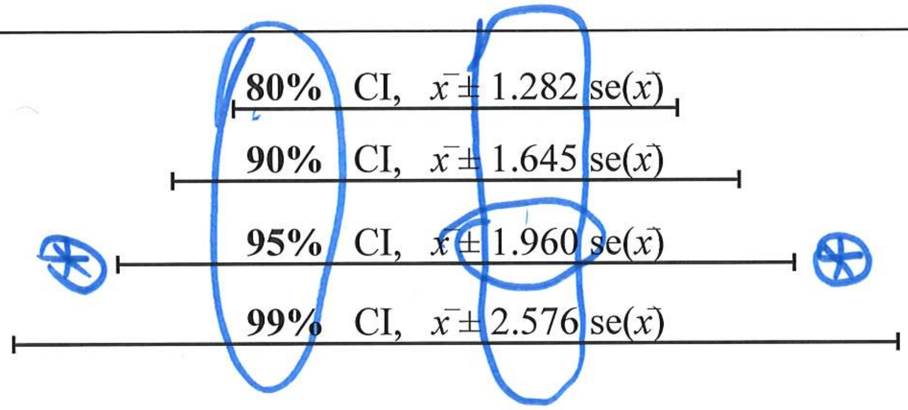


Figure 8.1.3 The greater the confidence level, the wider the interval

- To double the precision of the confidence interval we **need 4 times** as many observations. halve the width
- To triple the precision of the confidence interval we **need 9 times** as many observations. third the width
- 95% confidence interval
  - ✓ Range of plausible values for the parameter of interest that contains the **true value** of our parameter of interest for 95% of samples taken.
  - ✓ 5% of samples taken will not have the parameter within the calculated confidence interval.
  - ✓ We do not know if the sample we have taken is one of the 95% that contains the true unknown parameter. All we can say is that 95% of the time it will.

✓ If you take 1000 samples, based on the same sampling protocol, then you can expect approximately 950 of these samples will contain the true value (e.g. true mean, true difference between means) of the population.

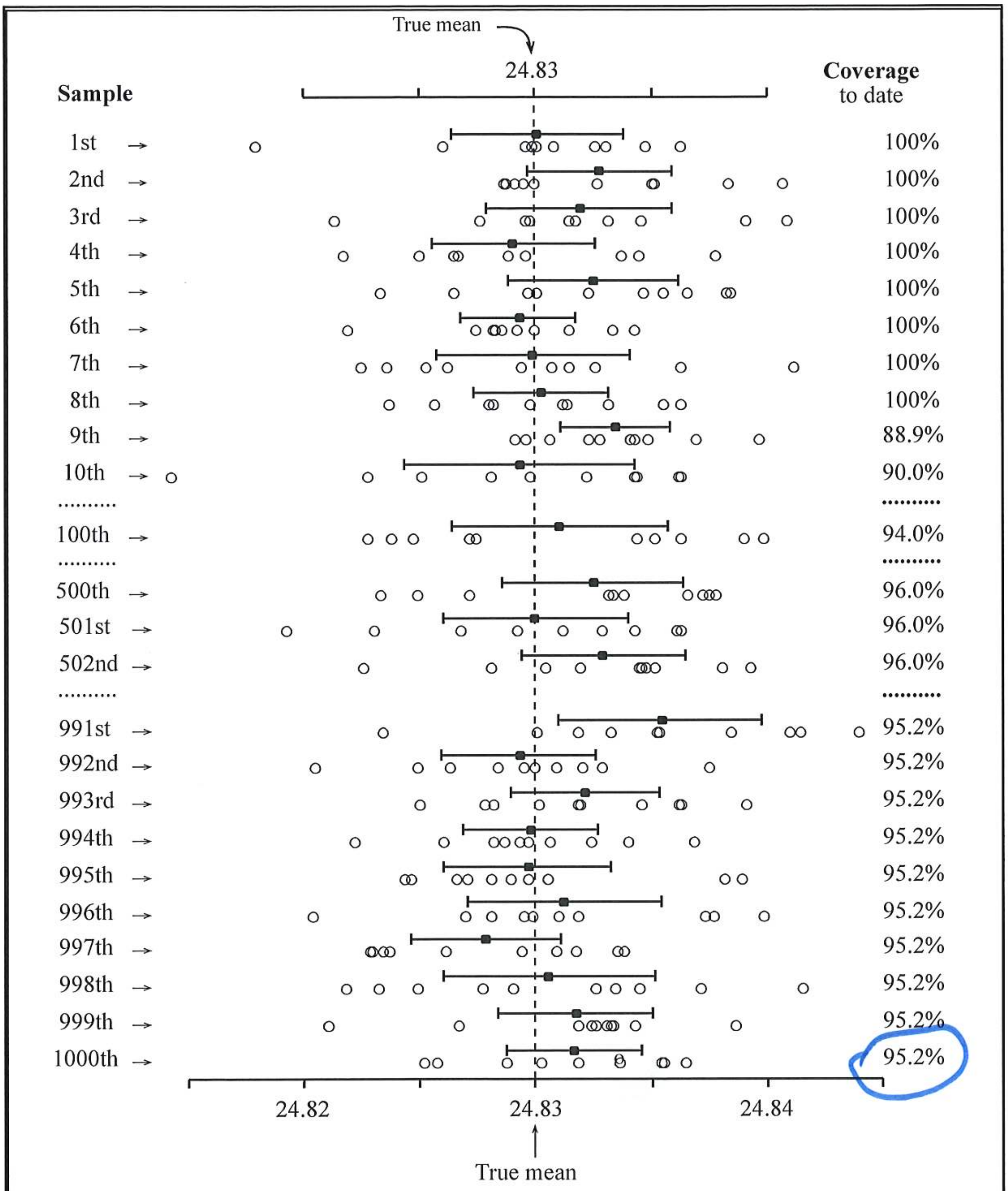


Figure 8.1.2 Samples of size 10 from a Normal(μ=24.83, s=.005) distribution and their 95% confidence intervals for μ..

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. $p_1 \approx p_2$ (-.05, .03)

  ✓ If CI is positive, then p_1 is higher/larger than p_2. $p_1 > p_2$ (.03, .05)

  ✓ If CI is negative, then p_1 is lower/smaller than p_2. $p_1 < p_2$ (-.05, -.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

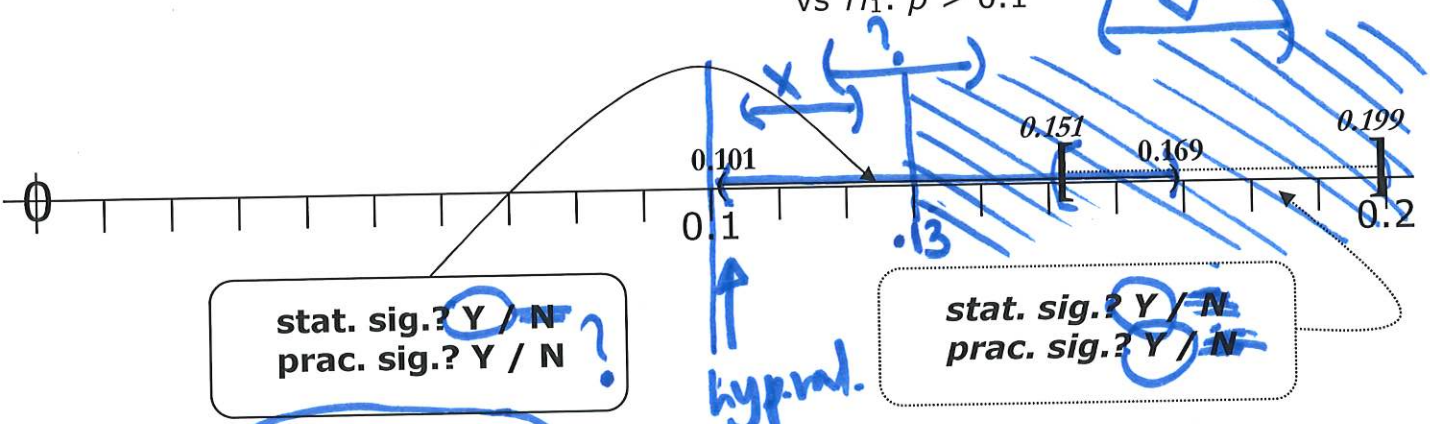
Statistical significance

- 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: $p > 0.1$



(0.101,0.169)

[0.151,0.199]

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.

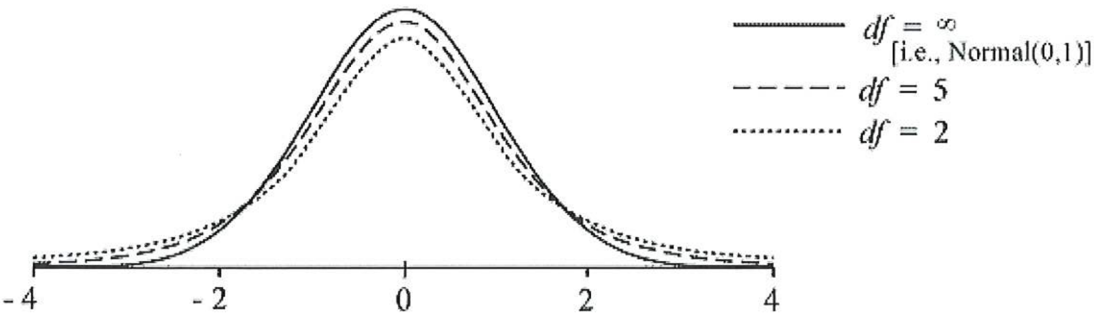
If a 2-tailed test has a *P-value* less than 5% then the test is significant at the 5% level of significance and the hypothesised value is not plausible (not believable) and it ~~will~~ / will not be in the 95% confidence interval. Conversely, if the hypothesised value is not in the 95% confidence interval 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 H0 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 [$Z \sim$ Normal ($\mu = 0$, $\sigma = 1$)] 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 ($df = \infty$) and Normal (0,1) are the same distribution.