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

.035
0.03

memorise

- 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%).

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

F (1) A small *P-value* provides evidence of ~~the size of~~ an effect.
F (2) Statistical significance is not the same as practical significance.
T (3) Practical significance depends on the size of the effect.
F (4) A small *P-value* provides ~~no~~ evidence against H_0.
F (5) A confidence interval estimates the ~~strength~~ size of an effect.

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

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

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

(3) T H_0 is typically a sceptical reaction to a research H_1 hypothesis.

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

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

less than 0·01 ∴ st. ev. against H_0, stat. sig. @ 1% level

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

(∴ also stat sig @ 5% level)

F (1) The *P-value* is very small, therefore H_0 is probably false.
T (2) We would reject H_0 at the 1% level of significance.
F (3) A 95% confidence interval for μ would not contain the value 100.
F (4) A 99% confidence interval for μ would not contain the hypothesised value. (100)
F (5) ~~We will accept that~~ H_0 is ~~true.~~ probably false.

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

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

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

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

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

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

– zero is not a plausible val.
– ∴ hyp val will be out of a 95% CI / 99% CI !

Normality-based (Chapter 6) Confidence Interval:

margin of error

<u>**Step 8**</u>

$$\boxed{\textbf{estimate} \pm \underbrace{t \times \textbf{se(estimate)}}_{\text{margin of error}}}$$

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 appropriate *t*-multiplier for a 95% confidence interval.

G. Which **one** of the following statements about confidence intervals is **false**?

T (1) For a given level of confidence, increasing the sample size generally decreases the width of a confidence interval.

T (2) For a given level of confidence, decreasing the standard error decreases the width of a confidence interval.

T (3) For a given sample size, increasing the confidence level increases the width of a confidence interval.

F (4) For a given sample size, increasing the confidence level ~~in~~creases (de) the precision of a confidence interval.

T (5) For a given level of confidence, decreasing the sample size generally decreases the precision (width) of a confidence interval.

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.

  1. μ 3. $\mu_1 - \mu_2$

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

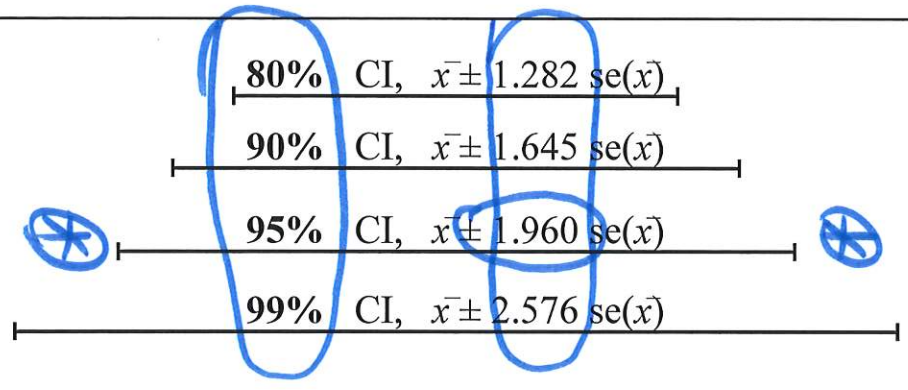


Figure 8.1.3 The greater the confidence level, the wider the interval

- 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.

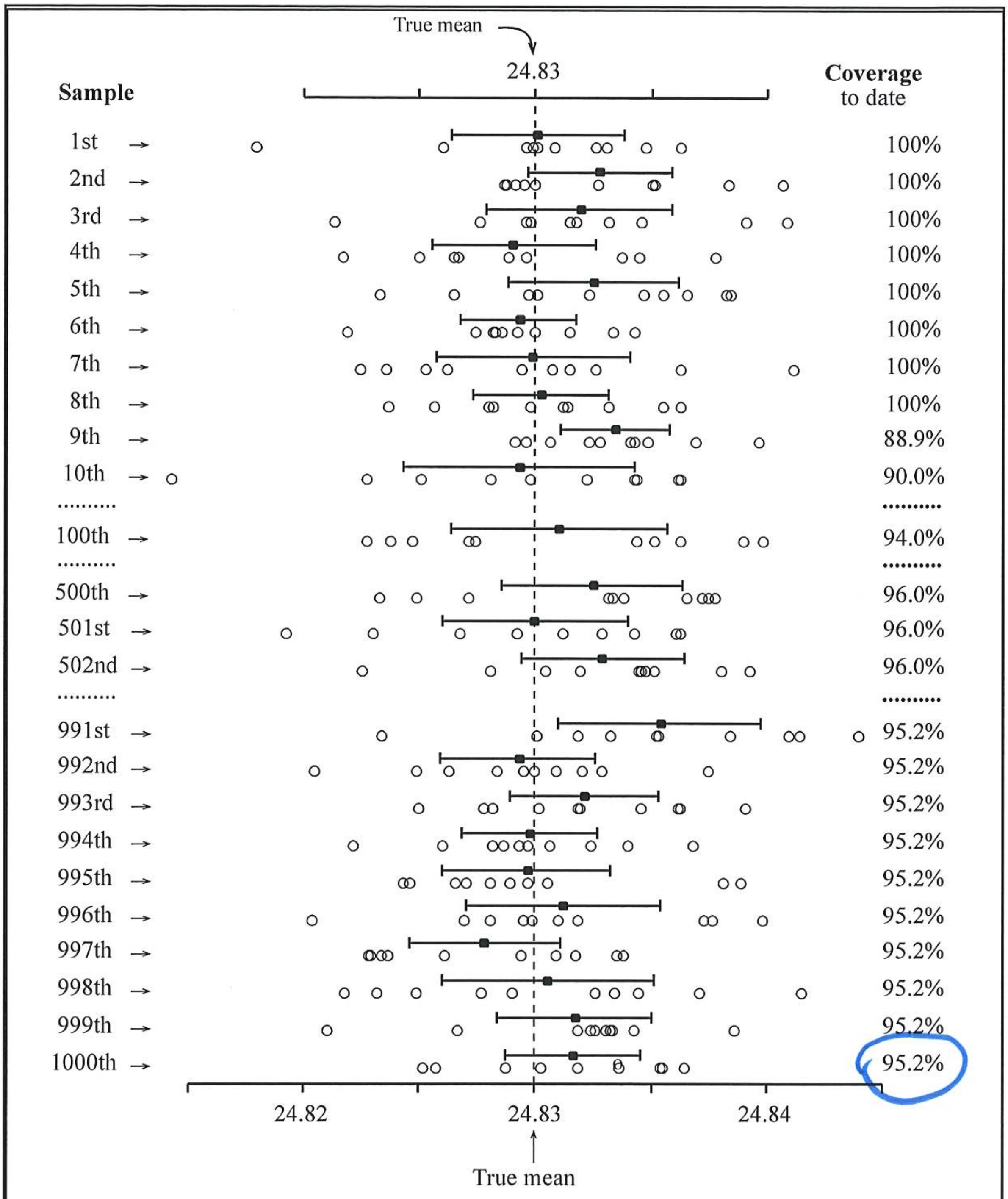


Figure 8.1.2 Samples of size 10 from a Normal(μ=24.83, s=.005) distribution and their 95% confidence intervals for μ..

Interpreting the CI limits → Step 9 for story type 3:

- CIs for the difference between two means:

Examples:

✓ If the CI contains 0 (i.e. one negative and one positive number), there may be no difference between the two means. $\mu_1 \approx \mu_2$ — (-7, 19)

✓ If CI is positive, then μ_1 is higher/larger than μ_2. $\mu_1 > \mu_2$ — (7, 19)

✓ If CI is negative, then μ_1 is lower/smaller than μ_2. $\mu_1 < \mu_2$ — (-19, -7)

zero in CI! $\mu_S \approx \mu_m$ $H_0: \mu_S - \mu_m = 0$

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

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

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

(not stat. sig. at 5% level)

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

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

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

est ± t × se(est)

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

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

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

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

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

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

size comes from CI!

$H_0: \mu_1 - \mu_2 = 0$
vs $H_1: \mu_1 - \mu_2 \neq 0$

Questions J to **N** refer to the following information.

Lai *et al.* (2017) measured illegal drugs in water processed by wastewater treatment plants to estimate drug use in Auckland, New Zealand. Random sampling and testing of wastewater occurred between 2 May and 18 July, 2014, at two treatment plants.

A plant's 'catchment area' refers to where the water it treats comes from:
- Plant 1 catchment area: Auckland City, Papakura, Waitakere and Manukau
- Plant 2 catchment area: North Shore

Methamphetamine consumption in milligrams per day per 1000 people was estimated for each plant's catchment area from the samples of wastewater collected at each plant.

Let:

μ_1 be the underlying mean methamphetamine consumption in milligrams per day per 1000 people in the wastewater treatment plant 1 catchment area between 2 May and 18 July, 2014,

and

μ_2 be the underlying mean methamphetamine consumption in milligrams per day per 1000 people in the wastewater treatment plant 2 catchment area between 2 May and 18 July, 2014.

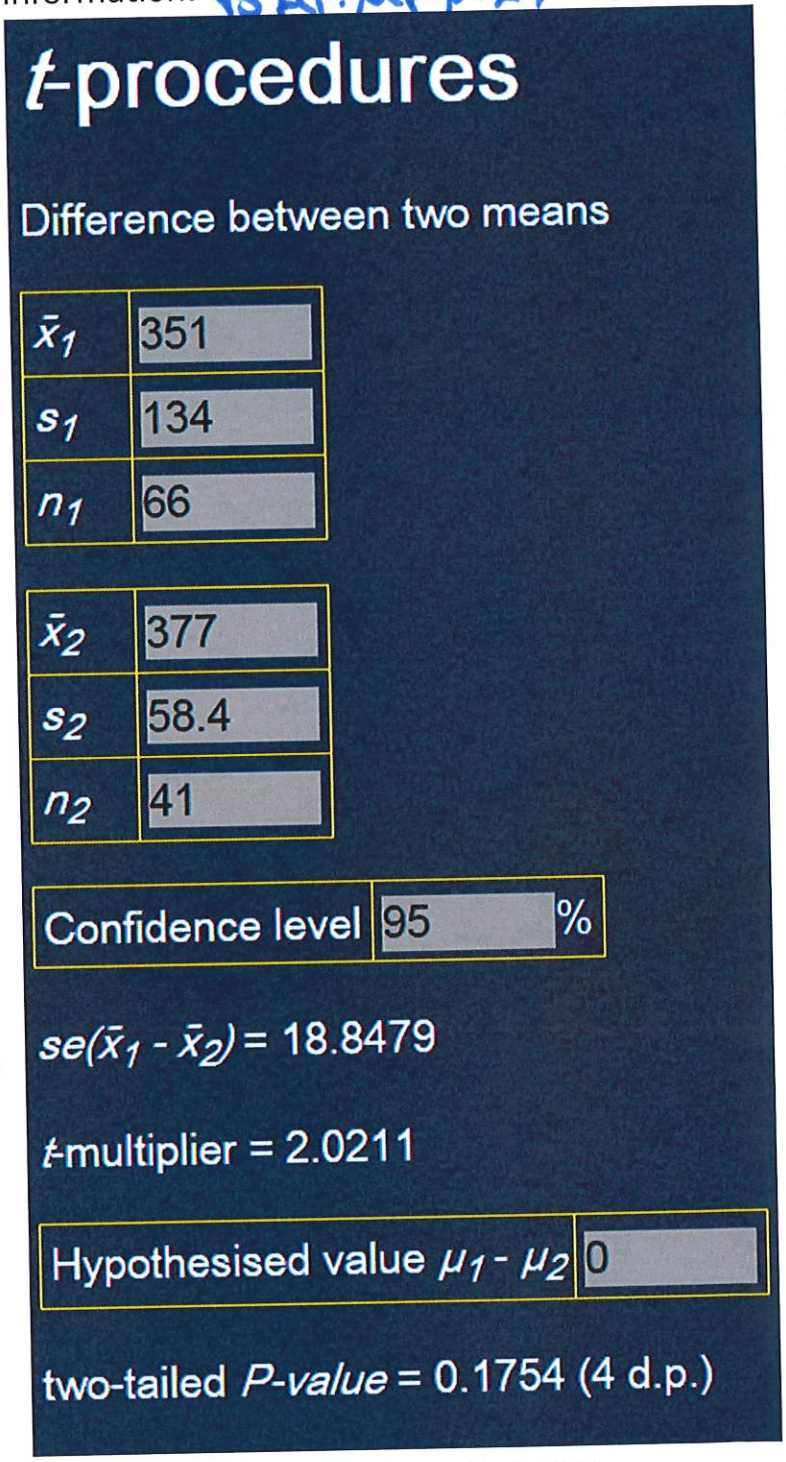


Figure 1: Screen-shot of the *t*-procedures tool

A two-tailed *t*-test for no difference between μ_1 and μ_2 was conducted. The *t*-procedures tool with the means, standard deviations and sample sizes for the estimated methamphetamine consumption for the two wastewater catchment areas is shown to the right where $\bar{x}_1$ is the sample mean for the plant 1 catchment area and $\bar{x}_2$ is the sample mean for the plant 2 catchment area.

J. Which **one** of the following is a **correct** pair of hypotheses for this *t*-test?

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

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

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

~~(4)~~ $H_0 : \cancel{\bar{x}_1 - \bar{x}_2} = 0$ versus $H_1 : \cancel{\bar{x}_1 - \bar{x}_2} \neq 0$

(5) $H_0 : p_D - p_S > 0$ versus $H_1 : p_D - p_S = 0$

K. The value of the *t*-test statistic, t_0, is approximately: Step 5

(1) −1.38

(2) −26

(3) 2.02

(4) 26

(5) 1.38

$$t_0 = \frac{\text{est} - \text{hyp.val.}}{\text{std err}} = \frac{\bar{x}_1 - \bar{x}_2 - 0}{se(\bar{x}_1 - \bar{x}_2)} = \frac{(351 - 377)}{18.8479} = -1.38$$

L. To the nearest whole number, which **one** of the following is the **correct** margin of error for the 95% confidence interval for $\mu_1 - \mu_2$?

(1) 74

(2) 19

(3) 37

(4) 38

(5) 76

$$\text{est} \boxed{\pm\ t \times se(\text{est})}$$

$$t \times se(\bar{x}_1 - \bar{x}_2) = 2.0211 \times 18.8479 = 38$$

M. If the confidence level in the *t*-procedures tool was changed from 95% to 99%, which **one** of the following statements would be **true**?

F (1) The standard error would ~~increase~~. remain the same.

T (2) The *t*-multiplier would increase.

F (3) The *P-value* would ~~change~~. remain the same

F (4) The confidence interval would be ~~narrower~~. wider

F (5) The margin of error would in~~de~~crease by 4%.

N. Which **one** of the following is a **correct** interpretation of the *P-value* shown in Figure 1, page 17? p-val = .1754 no ev against H_0

F (1) There is a 0.1754 chance that the null hypothesis is true.

F (2) If the alternative hypothesis was true, the probability of getting a test statistic at least as extreme as the observed test statistic is 0.1754.

F (3) There is a 0.1754 chance that the alternative hypothesis is true.

T (4) If the null hypothesis was true, the probability of getting a test statistic at least as extreme as the observed test statistic is 0.1754.

F (5) The probability of getting an estimate of at least −26 is 0.1754.

assuming H_0 is true

Practical significance versus ***Statistical significance***

Statistical significance

- Relates to having evidence of the **existence** of an effect or difference.
- Determined by examining the ***P-value*** of your significance test.
- To be statistically significant at the 5% level, the *P-value* must be ~~greater than~~ / less than 0.05 (5%).

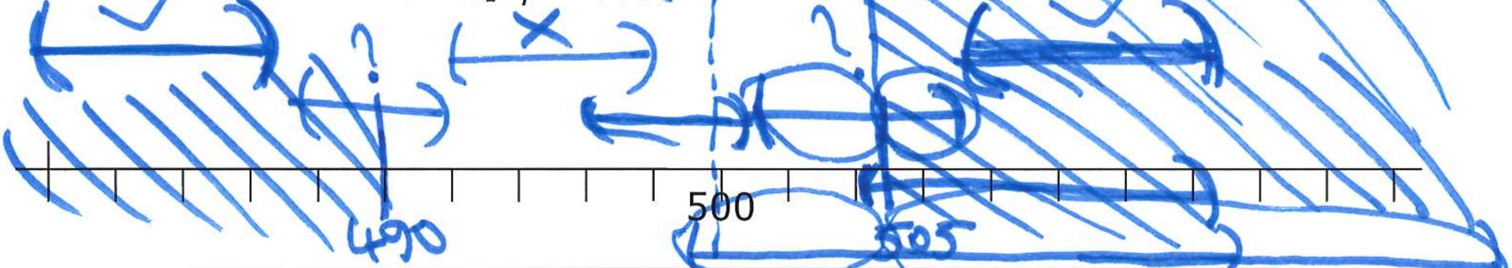
Practical significance

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

C.f. S219 Exam: Q23

For example: Jam jar example (please see Lecture Workbook, Chapter 7, page 19):

H_0: $\mu = 500$
vs H_1: $\mu \neq 500$



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%.

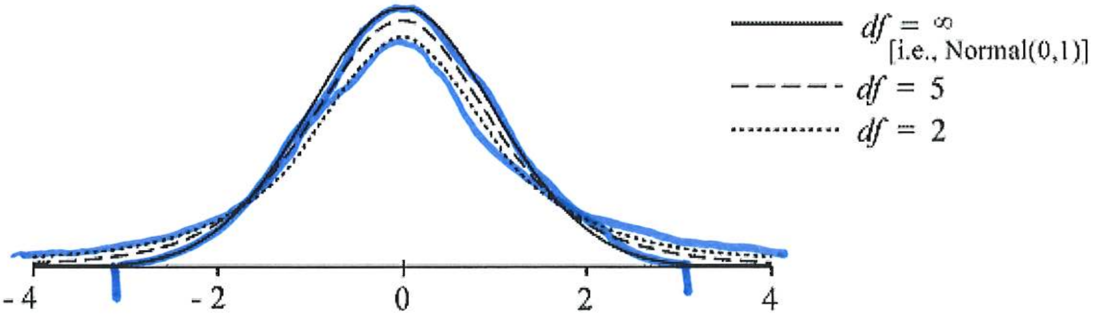
If a 2-tailed test has a *P-value* greater than 5% then the test is not significant at the 5% level of significance and the hypothesised value is plausible (is believable) and so it will / ~~will not~~ be in the 95% confidence interval. Conversely, if the hypothesised value is in the 95% confidence interval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

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

T (1) In a *t*-test for no difference between two means, being able to demonstrate that the difference was of practical significance (importance) would almost always imply statistical significance.

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

T (3) In hypothesis testing, statistical significance does not imply practical significance (importance).

F (4) In a hypothesis test for no difference between two means, a very small *P-value* always indicates a very large difference in the means. CI

T (5) In hypothesis testing, a non-significant test result does not imply that the null hypothesis is true.

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

F (1) The Student's *t*-distribution has tails which become fatter as the degrees of freedom ~~increase~~. de

F (2) An estimate is more precise if it has ~~more~~ less variability.

T (3) Student($df = \infty$) and Normal($\mu = 0, \sigma = 1$) are identical distributions.

F (4) If greater confidence in a confidence interval calculated from our data is desired, then a ~~narrower~~ wider interval needs to be used.

(5) A ~~parameter~~ estimate is a numerical characteristic which can be calculated from a sample.

the parameter is a fixed value from the pop/distb.