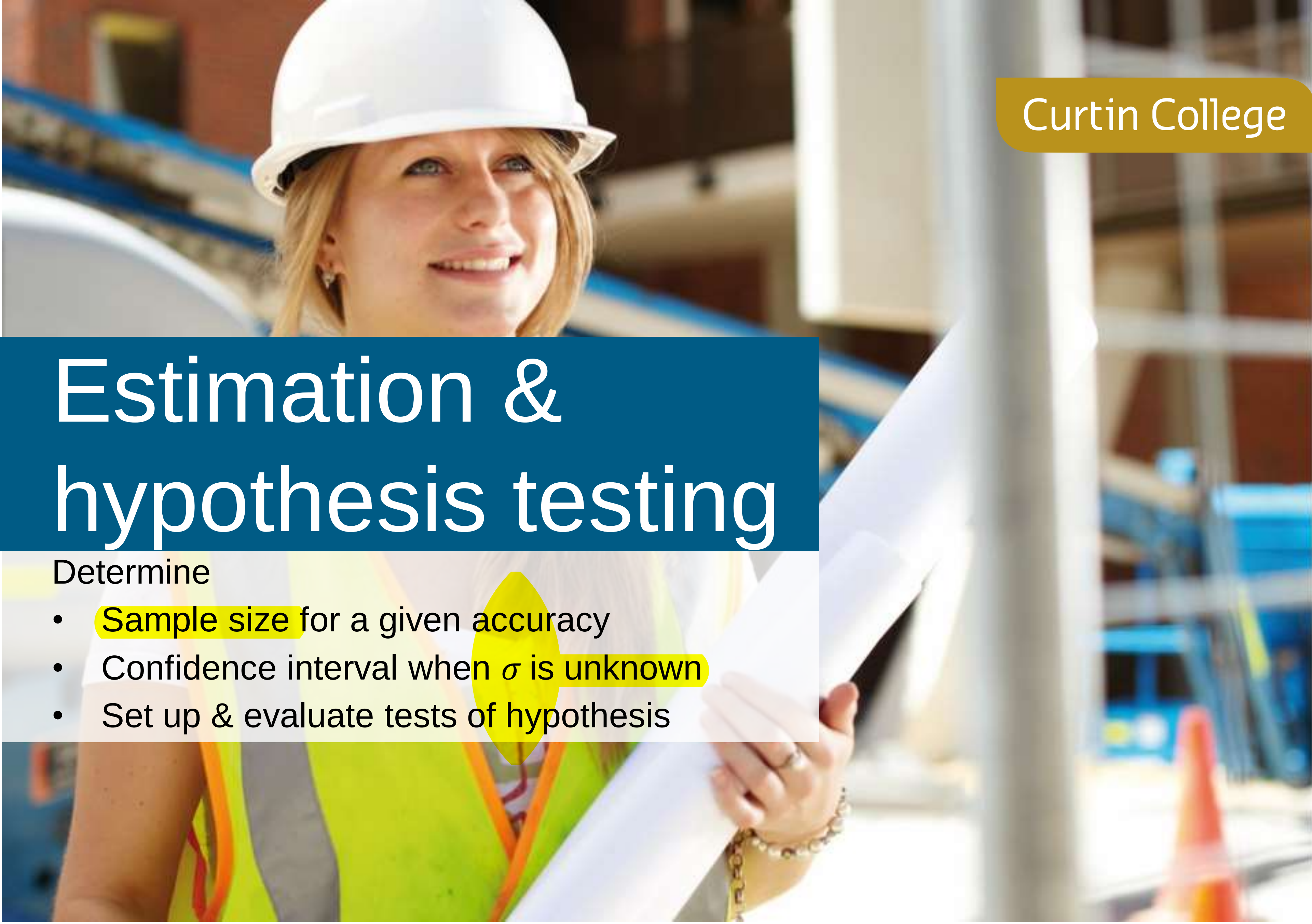
Curtin College

DIPLOMA OF ENGINEERING

EMTH1019 W4 ESTIMATION & HYPOTHESIS TESTING



Curtin College

Estimation & hypothesis testing

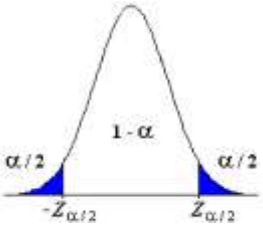
Determine

- Sample size for a given accuracy
- Confidence interval when σ is unknown
- Set up & evaluate tests of hypothesis

Curtin College

HYPOTHESIS & T – SUMMARY SHEET

- t distribution $t = \frac{\bar{x}-\mu}{(\frac{s}{\sqrt{n}})}$ use when you only have sample data but it is reasonable to assume population is Normally distributed & sample is SRS.
- Significance (α or sometimes p) is the value of the tails.
- DOF = n -1 – you must be able to read t tables.



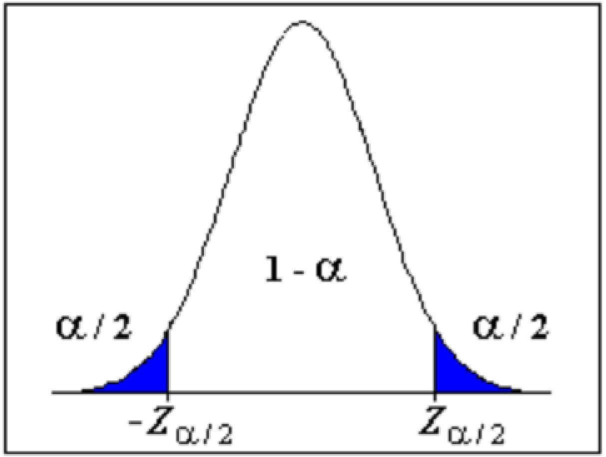
- Null Hypothesis is usually historical data or a claim. H_0 : $\mu_0 = 500g$
- Alternate Hypothesis is H_A : $\mu \neq 500g$
    - You can have < or > in H_A but ≠ is easiest
- If significance = 0.05 for a 2-tailed distribution the remaining 0.95 is the 95% confidence interval. I call 0.95 the “happy accept Null Hypothesis region”
- Test statistic: $t = \frac{\bar{x}-\mu}{(\frac{s}{\sqrt{n}})}$ (only sample data & $\mu = H_0$) or $z = \frac{\bar{x}-\mu}{(\frac{\sigma}{\sqrt{n}})}$ (σ is known & $\mu = H_0$)
- If test statistic lies in “happy accept null hypothesis region” you have reason to believe that you can accept H_0 at stated significance & assumptions. Your sample is not statistically different from H_0 .
- You need a conclusion that refers back to the original question.

Curtin College

MARGIN OF ERROR

When we looked at confidence intervals last week…

$$-z_{\alpha/2} < \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}$$



If we rearrange the equation by $*\frac{\sigma}{\sqrt{n}}$

$$\left(-z_{\alpha/2}\right) * \frac{\sigma}{\sqrt{n}} < \bar{x}-\mu < \boxed{\left(z_{\alpha/2}\right) * \frac{\sigma}{\sqrt{n}}}$$

The margin of error is $|\bar{x}-\mu|$

Sample mean

population mean

USING MARGIN OF ERROR TO CALCULATE SAMPLE SIZE

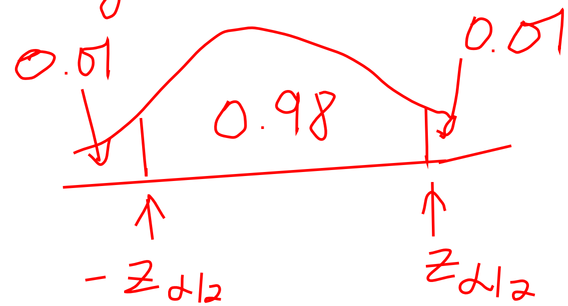
If I let the margin of error $|\bar{x} - \mu| = E$

from confidence interval

$\Rightarrow$ get $z_{\alpha/2}$

Then

$$\boxed{E = (z_{\alpha/2}) * \frac{\sigma}{\sqrt{n}}}$$



Re-arrange

$$\sqrt{n} = (z_{\alpha/2}) * \frac{\sigma}{E}$$

Square both sides

$$n = ((z_{\alpha/2}) * \frac{\sigma}{E})^2$$

Whatever n you calculate, always round up

EXAMPLE: HOW MANY MEASUREMENTS?

To assess the accuracy of a laboratory scale a standard weight is weighed repeatedly.

The scale readings are normally distributed with unknown mean

The standard deviation is 0.0002 grams

population SD = 0.0002
⇒ use z table

Question

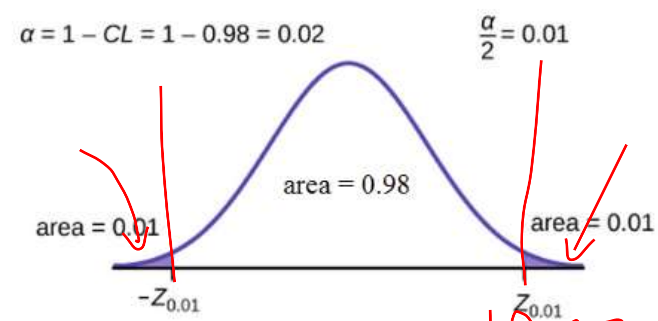
How many measurements must be made to get a margin or error of $\pm\ 0.0001$ with 98% confidence?

Hint: always work out your confidence interval first. What are the $\pm z$ for 98% confidence interval?

SOLUTION: HOW MANY MEASUREMENTS?

SD of population = 0.0002

- 1 Draw a 98% confidence
- What values of z are the critical z values for 98% confidence (from tables).
  - $1 - 0.98 = 0.02 = \alpha$ then 0.02/2 = 0.01
  - In z tables look for a probability of 0.01 this gives a $z = -2.33$ for the LHS.
  - By symmetry $z = +2.33$ for the RHS.
  - $\boldsymbol{z_{\alpha/2} = \pm 2.33}$ is another way of writing it.



-2.33 $+2.33$

$$\frac{1 - 0.98}{2} = 0.01$$

You know $\sigma = 0.0002$ and error $E = \pm 0.0001$

- $\boldsymbol{n = ((z_{\alpha/2}) * \frac{\sigma}{E})^2} = ((\mathbf{2.33}) * \frac{\mathbf{0.0002}}{\mathbf{0.0001}})^2$
- $n = 21.72$ Always round up $n = 22$

$$E = z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$$

$$\sqrt{n} = \frac{z_{\alpha/2} * \sigma}{E}$$

$$n = \left(\frac{z_{\alpha/2} * \sigma}{E}\right)^2$$

ANSWER

- You need at least $n = 22$ measurements to get a margin or error of ± 0.0001 with 98% confidence

CU PAST EXAM QUESTION

Part (b) is a margin of error question

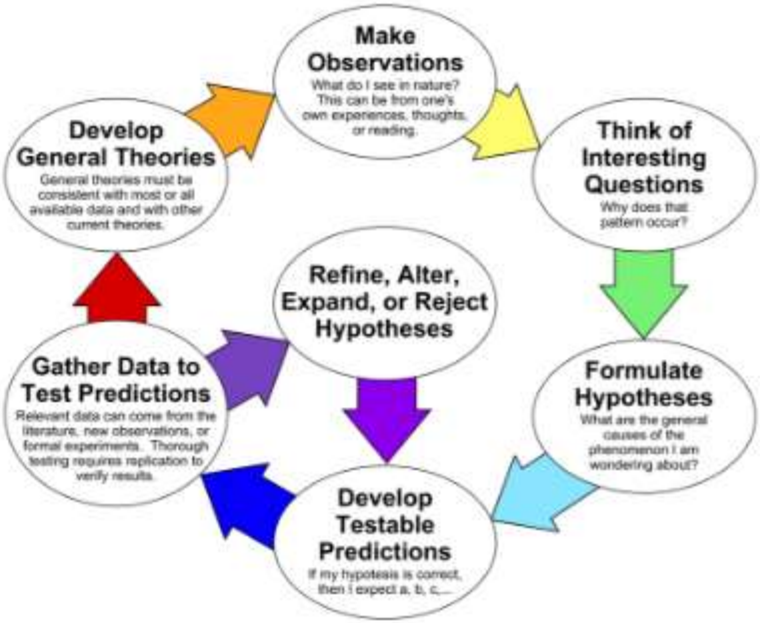
Question 4. This is a continuation of Question 3 above. ACCC wants to assess independently whether there is any significant evidence to support Kellogg's claim that each cereal box contains 500g of cereal. Using the randomly chosen 20 boxes of the product, ACCC finds that the average weight per box is $\bar{x} = 495$g with the sample standard deviation $s = 6$g.

(a) Perform a test of hypothesis at the 5% significance level with the intent to show that Kellogg over-estimates the average weight of cereal in a box. (6 marks)

(b) If we accept Kellogg's claim of the average cereal weight per box is $\mu = 500$g with the standard deviation $\sigma = 10$g, how large a sample is required if we want a 98% confidence interval for the mean μ to have a margin of error of ± 5g? (4 marks)

Curtin College

SCIENTIFIC METHOD



https://raeonscience.weebly.com/the-scientific-method.html

HYPOTHESIS

Is a proposed explanation for a phenomenon

Hypothesis testing is at the heart of the scientific method.

This week we will be dealing with:

- Null Hypothesis H_0
  - Is the commonly accepted fact
  - States a claim that we assume to be true unless we can find sufficient evidence to indicate it is not true.
- Alternate Hypothesis H_A
  - Opposite of the Null Hypothesis.
  - Is a statement that we hope or suspect is true but we need sufficient evidence to support it before we are willing to accept it.

EXAMPLE OF NULL & ALTERNATE HYPOTHESES

A researcher is studying the effects of an exercise program on knee surgery patients.

There is a good chance the therapy will improve recovery time, but there's also the possibility it will make it worse. The researcher does not know the answer

The historical average recovery times for knee surgery patients is 8.2 weeks.

1. The null hypothesis (no significant change) is that the exercise program will have no impact on recovery time
   $H_0 = 8.2$
2. Th alternate hypothesis (there is a significant change) is the opposite of the null hypothesis
   $H_A \neq 8.2$
   I chose a 2-tailed alternate hypothesis as this includes the exercise program having improved and WORSE recovery times.

PROBLEMS WITH HYPOTHESIS TESTS

The out come of a statistical test is to either accept or not accept the Null Hypothesis.

- If you cannot accept H_0 then you are inclined to accept hypothesis H_A
- Your data sample may provide a good representation of the population then your conclusion might be true

Or

- Your sample data leads you to not accept H_O even though it is actually true for the population.

Or

- You accept the H_O as your sample indicates there is no significant change, but your sample is not representative of the population

THE BOY WHO CRIED WOLF

This story is attributed to Aesop (620 – 564 BCE)

A naughty shepherd boy repeatedly fools villagers into believing a wolf is attacking his flock of sheep calling out " wolf, wolf " when there is no wolf.

A real wolf appears and the boy calls for help " wolf, wolf ". The villagers hear the cry but believe it is another false alarm and do not help.

The sheep are eaten by the wolf and in some versions of the fable so is the boy!

Moral of the story

Liars are not believed – even when they speak the truth.

https://www.pinterest.com.au/pin/507288345498566353/

YOU ARE GUARDING YOUR SHEEP

- ***The null hypothesis is H_O = no wolf is present***
- ***The alternate hypothesis H_A= there is a wolf present***

Type 1 error or False Positive or False Alarm

- Assert something is true when it is not true
- A false positive is crying "***wolf, wolf!!***" when there is no wolf present
- You have rejected H_O, and accepted H_A , but it is H_O that is true

Type 2 error or False Negative

- Assert something is false when it is true.
- There is a ***dangerous wolf present***, but you think there is no wolf.
- You have accepted H_O, when it is H_A that is true

https://infocus.dellemc.com/william_schmarzo/understanding-type-i-and-type-ii-errors/

TYPE 1 AND 2 ERRORS CONTINUED

		The Truth (Based on Entire Population)	
		Nothing Is There (H_0 Is True)	Something Is There (H_0 Is False)
Your Conclusion (Based on Your Sample)	I Don't See Anything (Nonsignificant)	Right!	Wrong (Type II Error)
	I See Something (Significant)	Wrong (Type I Error)	Right!

https://www.dummies.com/education/science/biology/type-i-and-type-ii-errors-in-hypothesis-testing/

1 & 2 SIDED ALTERNATE HYPOTHESES

The null hypothesis is usually written as $H_0 : \mu_0 = \dots$

The alternate hypothesis is usually written as $H_A : \mu_A \neq or \mu_A < or\ \mu_A >$

You can have either a 1 or 2 sided alternate hypothesis

- **1 sided**
  - $\boldsymbol{\mu_A < \mu_0}$ **or** $\boldsymbol{\mu_A > \mu_0}$
- **2 sided**
  - $\boldsymbol{\mu \neq \mu_0}$
  - **I usually do a 2 sided alternate hypothesis.**

- The alternate hypothesis chosen depends on what we are trying to prove or suspect to be true.
- You must form the hypothesis before collecting the data
- You cannot fit an hypothesis after you have collected the data – this would be unethical (*and very dodgy!)*

Curtin College

SETTING UP HYPOTHESES

Apartments

The mean area of new homes in Perth is claimed to be 520 m^2. You think the average area is smaller than advertised.

- The Null hypothesis (claim or no change) is that the mean area is 520 m^2. $H_0 = 520$ or $H_0 : \mu = 520$
- The Alternate hypothesis could be either:
  - The mean area is less than 520 m^2. $H_A < 520$ one-tail: $H_A < 520$ or $H_A : \mu < 520$
  - The mean area is not equal to 520 m^2. $H_A \neq 520$ two-tail: $H_A \neq 520$ or $H_A : \mu \neq 520$
  - For me the 2 tail alternate hypothesis $H_A \neq 520$ is a good alternate hypothesis as you are not making judgement on what you will find as the area could be lower or higher.
  - The important thing is that you have to decide on an alternate hypothesis and justify your use of it.

Cars

two-tail

Your car averages 10 km/litre. A new fuel says it will improve fuel consumption.

- The new fuel might cause: no change, an increase or a decrease in fuel consumption.
- Identify an appropriate hypothesis. $H_0 = 10$ $H_A \neq 10$

$H_0 : \mu = 10$

Bolts

The diameter of a bolt is supposed to be 6mm. If it is too big or small the nut will not fit properly.

two-tail

- $H_0 = 6$ $H_A \neq 6$

Curtin College

HYPOTHESES FOR BREAKFAST

A bored statistician suspects that the 500 gram packets of breakfast cereal he buys actually contain less than 500 grams.

He cannot accuse the company of short changing consumers without some evidence.

- He needs to be ***confident*** that he is correct!

- The **Null Hypothesis** is that the packets do contain 500 grams of delicious cereal.

$$H_0: \mu_0 = 500\ grams$$

- The **Alternate Hypothesis** is that the packets contain< 500 grams

$$H_A: \mu < 500\ grams$$

This is a ***one sided test*** as the statistician is only worried that he is not getting enough cereal. He does not care if they are giving him >500 grams.

Curtin College

TO TEST THE CEREAL HYPOTHESIS…

1. State the hypothesis ✓ Ho and HA
2. Take a simple random sample and calculate the sample mean

If the sample mean is $490\ grams, <\ 500\ grams$ is this enough proof that $\boldsymbol{H_A: \mu < 500\ grams}$ is true?

Think back to last weeks lecture…

- You took a SRS of n=5, from a set of numbers and calculated the mean, and then calculated the mean of the means.
- Even though the mean of the population was actually ~56, your sample means were all over the place and then the mean of the means was ~48
- Couldn't the same be true here?
- How much less than 500 does my sample mean have to be before I am confident that $\boldsymbol{H_A: \mu < 500\ grams}$ is true?

What to do?

Curtin College

CALCULATE THE TEST STATISTIC

490 < 500

What to do?

A packet of cereal may say it contains 500 grams but we know that there will be some variation in the actual weight.

We calculate the test statistic (**z**) of the sample

$$z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

- μ is the value under the Null Hypothesis
- σ is the know population value of the variance of random variable

P-VALUE

A SRS of 30 boxes of cereal yields a sample mean of 498 grams.

You are given that the population has a standard deviation of 5.

Question

Is there enough evidence to suggest that the packet population means are underweight?

Method

- State the null hypothesis and alternate hypotheses
- Calculate the test statistic and work out $P(\bar{x} < 498)$

This probability is called the ***p-value***

- If the ***p-value*** is sufficiently small we might be able to suggest that there is some evidence to support the alternate hypothesis.

CAN I JUST USE P-VALUES?

NO. By itself, a *p*-value does not provide a good measure of evidence regarding a model or hypothesis. It appears in lots of scholarly papers but that does not make it good practice.

In 2016 the American Statistical Association stated that:

•P-values can indicate how incompatible the data are with a specified statistical model.

•*P*-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.

•Scientific conclusions and business or policy decisions should not be based only on whether a *p*-value passes a specific threshold.

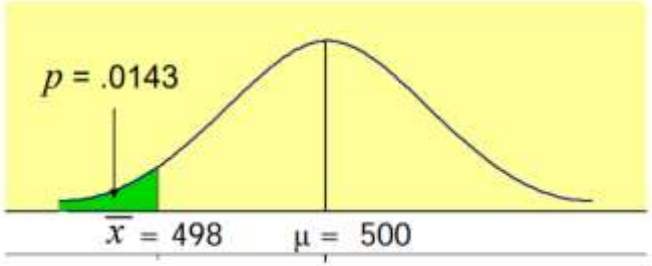
•Proper inference requires full reporting and transparency.

•A *p*-value, or statistical significance, does not measure the size of an effect or the importance of a result.

https://www.tandfonline.com/doi/full/10.1080/00031305.2016.1154108

IS THE P-VALUE SMALL ENOUGH?

- H_0: $\mu_0 = 500\ grams$
- H_A: $\mu < 500\ grams$
- We calculate $P(\bar{x} < 498)$ assuming that H_0 that is $\bar{x} \sim N\left(500, \frac{5}{\sqrt{30}}\right)$
- $P(\bar{x} < 498) = 0.0143 = 1.43\%$



0.0143 is a small number.

- It means that just by chance this $\bar{x}$ would occur 1.43% of the time
- Is it small enough to say that the Null Hypothesis is probably wrong?

Curtin College

LEVEL OF SIGNIFICANCE

You must *decide in advance* how small a p-value is needed to reject H_0

- The *size of p-value chosen* is called the level of significance α.
- Common values of α are 0.05 or 0.01
- If the data provides evidence against H_0 that is so strong that it would only happen no more than 5% of the time (or 1%)

For our cereal question if we had chosen a:

- ***1% level of significance*** our p-value of 0.0143 is greater than 0.01 so at the 1% significance level we have insufficient evidence to support the claim that the average weight of the cereal is <500 grams
- ***5% level of significance*** our p-value of 0.0143 is less than 0.05 so at the 5% significance level we have sufficient evidence to support the claim that the average weight of the cereal is <500 grams

Curtin College

NO ABSOLUTES IN STATISTICS

- **At the 1% significance level we are *not* saying** the boxes do contain 500 grams, just that we have insufficient evidence to support the alternative hypothesis
- **At the 5% significance level we are *not* saying** that the boxes do contain less than 500 grams, just that we have evidence to support the alternative hypothesis at this level of significance.

Curtin College

P VALUES ARE NOT ENOUGH

You will see many papers that only have p-values it is not enough!

The American Statistical Association advises not to use p-values as the sole analysis.

In many papers you will see p-value claims but no access to the raw data. Be suspicious. The raw data should be accessible.

You may also have people say “but that is the way we do it in our specialty”. That is not a convincing argument.

So why are p-values here?

> You are going to see them so this is a way of teaching you to be cautious of what people claim their analysis means.

The more complicated the statistical analysis the more cautious you should be about believing the conclusion.

Curtin College

EFFECT OF DANCE ON GAIT

What does this mean?

Data were analyzed using Sigmastat software (Systat, Richmond, VA).

Participant baseline demographics and the exit questionnaire responses were compared for differences with:

- One-way ANOVAs or Kruskall-Wallis one-way ANOVAs on ranks for non-parametric data.
- Two way repeated measures ANOVAs (group [Partner, Non-partner] × time [pre, post, follow-up]), with Holms-Sidak post-hoc tests,
- Determined statistical significance of changes from pre to post to follow-up.
- Level of significance was set at p = 0.05. ← Why?

Abreu, Mauro PT, DPT, GCS[1]; Hartley, Greg PT, DPT, GCS[1,2] The Effects of Salsa Dance on Balance, Gait, and Fall Risk in a Sedentary Patient With Alzheimer's Dementia, Multiple Comorbidities, and Recurrent Falls, Journal of Geriatric Physical Therapy: April/June 2013 - Volume 36 - Issue 2 - p 100-108 doi: 10.1519/JPT.0b013e318267aa54

Curtin College

METHOD: HYPOTHESIS TESTING

1. Choose an appropriate hypothesis.
   - You need to think about what you are testing for.
2. Decide on the level of significance
   - This is a chosen number.
3. Discuss all necessary assumptions and limitations.
   - It is not enough to say “assume srs”
4. Decide on a distribution.
   - For us it is either t or z.
5. Draw a bell curve & mark critical values of t/z that result from significance.
6. Calculate the test statistic and locate it on the curve.
7. State your conclusion but remember to include:
   - The word question
   - Assumptions and limitations.
   - A numerical statement.
   - Are you accepting the null or alternate hypothesis and why; refer calculations.
   - Make a recommendation for further action if necessary.

page 29

Curtin College

EXAMPLE: TELEPHONE CALLS

population mean

population SD

```
If Population SD is given , then use Z table
If Population SD is not known, then use T table
```

According to a telephone company's records from 1999 the average length of a long distance call was 12.44 minutes with a standard deviation of 2.65 minutes

$H_0 = 12.44$

Management wants to check if the mean length of current long distance calls has changed from 12.44 minutes

$H_A \neq 12.44$

A sample of 150 calls places through the company has a mean duration of 13.71 minutes.

sample $n = 150$

$\bar{x} = 13.71$

Using a 5% significance level can you conclude that the mean length of all current long distance calls is different from 12.44 minutes?

Curtin College

DISCUSSION OF ASSUMPTIONS

- The question doesn't say but for the millions of phone calls it is probably ok to assume the population is normally distributed.
- The historical data is the μ and σ of a normal distribution.
- The sample is random, large enough and hopefully representative of the current year.
- The big assumption in this analysis is that σ of the current year is the same as the historical data. That might not be true...

5%

- P=0.05 is given but why choose p=0.05?
- I am choosing a 2 tailed hypothesis as I am unsure if call duration has gone up, down or no change.
- I am using z based on earlier discussion and assumptions.

CALCULATIONS

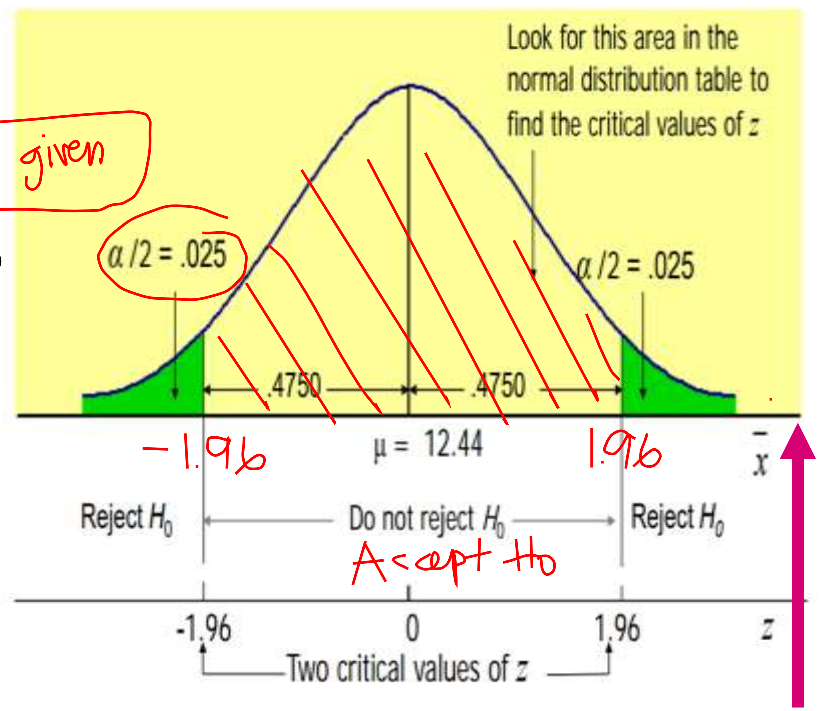
5 %

95%

- $H_0 \colon \mu_0 = 12.44\, minutes$
- $H_A \colon \mu_A \neq 12.44\, minutes$
- $significance = 0.05 = \alpha,$ 5%
    - $therefore\ \frac{\alpha}{2} = 0.025$

population S.D is given

- $Critical\ values\ from\ z\ tables\ z = \pm 1.96$
- $\mu = 12.44, \sigma = 2.65, n = 150\ , \bar{x} = 13.71$



H_A mean

H0 mean

- Test statistic
    - $\boxed{z_{test} = \frac{\bar{x}-\mu}{\left(\sigma/\sqrt{n}\right)}} = \frac{13.71-12.44}{2.65/\sqrt{150}} = 5.87$

pop s.D

- $z_{test} = 5.87 > \mathrm{z} = +1.96$

Z=5.87

Curtin College

CONCLUSION

$$z_{test} = 5.87 > z = +1.96$$

At the 5% significance level I have reason to be confident that I can reject the null hypothesis H_0: $\mu_0 = 12.44\, minutes$ (average call duration is still 12.44 minutes) and in fact the current duration of 13.77 minutes (n=150) is statistically significantly higher.

This analysis assumes:

- The population is normally distributed.
- It is still valid to use the population standard deviation from 1999.
- The sample of n=150 is random and representative of the population.

Recommendation:

Whilst the $z_{test} = 5.87 > z = +1.96$ provides strong evidence to reject the null hypothesis, I recommend redoing the analysis with a t-distribution which is a more conservative analysis.

BIG PROBLEM: YOU ONLY HAVE THE SAMPLE STANDARD DEVIATION…

Unfortunately we often do not know the population standard deviation

If we only know the sample standard deviation (s) we can use it as an estimate of the population standard deviation (σ)

BUT…

Instead of calculating z we calculate the t statistic.

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$$

The formula looks *very, very* similar to the z statistic but:

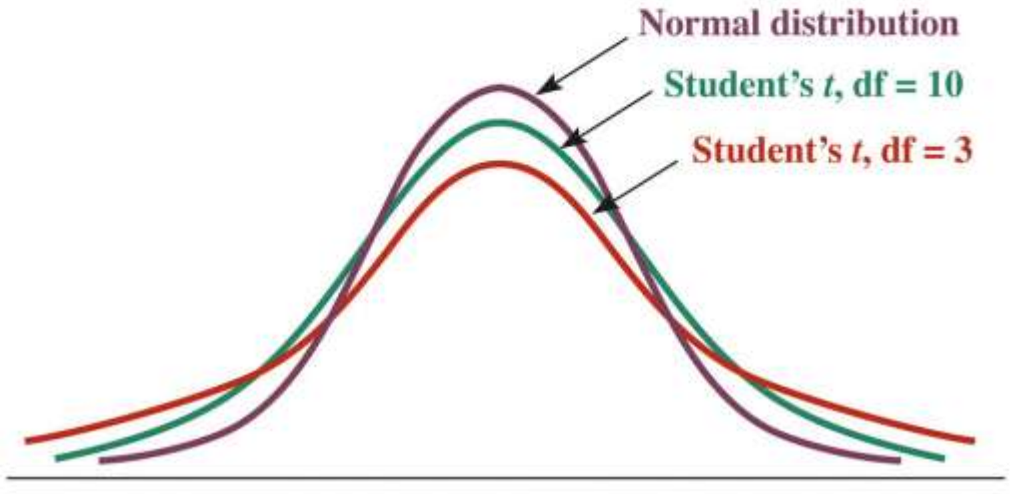
- A t-distribution is **not** a normal distribution.
- A t-distribution has more variation than a normal distribution.
- As $n \to \infty$ the t statistic starts approaching the z statistic (normal distribution)

Curtin College

T-DISTRIBUTION

- A t-distribution look similar to a normal distribution but you can see that it is more spread out. It has more variation
- There are different t-distributions for each sample size
- A t-distribution is defined by its degrees of freedom (DOF)
- DOF = n-1
- The larger the DOF the more "normal" the t-distribution looks.
- You need to use tables!

SD of population is unknown



IF I KNOW…

- Population μ and σ then use $z = \frac{x-\mu}{\sigma}$ and Normal Distribution tables
- Population σ and the sample is SRS of size n then use $z = \frac{\bar{x}-\mu}{(\frac{\sigma}{\sqrt{n}})}$ and then use Normal Distribution tables
- Sample mean and sample standard deviation (s) but do not know any population parameters you ***cannot*** use the Normal Distribution tables

KNOWN VERSUS UNKNOWN POPULATION STANDARD DEVIATION - σ

Known σ population SD

When you have a sample but you know the population standard deviation you can calculate:

$$\boxed{z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}}$$

Then you can use the normal distribution tables to calculate the probabilities

Unknown σ

You have to use your sample standard deviation, ***s***

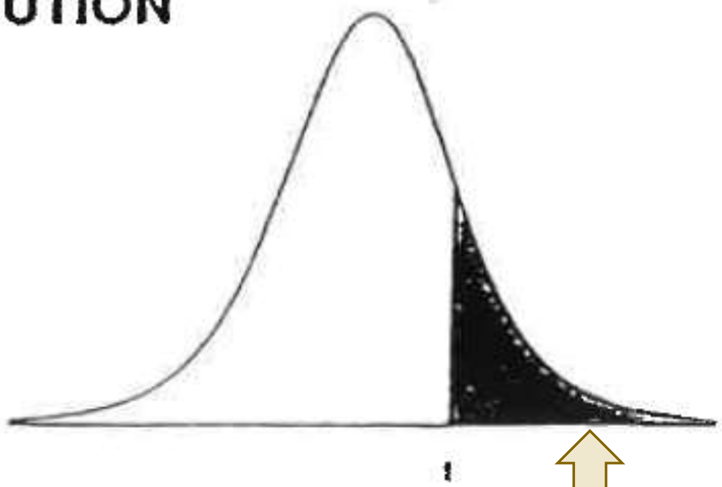
You can not use the normal distribution tables

You have to use the t-statistic and therefore t-distribution tables.

$$\boxed{t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}}$$

- **Degrees of freedom = n-1**
- **A p-value**

CRITICAL POINTS OF THE t-DISTRIBUTION



Degrees of freedom = n-1

Entry is t where P(T ≥ t) = p for t-distribution with v degrees of freedom

v	p 0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.657	318.317	636.607
2	1.886	2.920	4.303	6.965	9.925	22.327	31.598
3	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437

Curtin College

EXAM QUESTION

Population Standard Deviation is unknown
so use t-table and t-statistic

$\bar{x} = 2959$

$s = 39.4$

$n = 8$

Question 4

(a) Muzzle velocities of eight shells tested with a new gunpowder yield a sample mean of $\bar{x} = 2959$ feet per second and a standard deviation of $s = 39.4$. The manufacturer claims that the new gunpowder produces an average velocity of more than 3000 feet per second.

Does the sample provide enough evidence to support the manufacturer's claim? Set up and test an appropriate hypothesis at the 5% level of significance.

Use the following headings as a guide:

$H_0 = 3000$

$H_A > 3000$

1 must state the assumptions

- Assumptions
- Test Statistic
- Critical Region(s)
- Conclusion.

(10 marks)

SOLUTION – 2 TAIL STEP BY STEP

Assumptions

simple random sample

- The eight shells constitute an SRS.
- The muzzle velocities can be reasonably modelled by a normal probability distribution.
- σ is unknown. population standard deviation
- State hypothesis
  - H_0: $\mu_0 = 3000\ feet/sec$ (manufacturers claim)
  - H_A: $\mu \neq 3000\ feet/sec$ (even though the question talks about <, it is actually more logical to ask is the data significantly different to 3000)
    - When you look at question $\bar{x} = 2959$ so the sample data mean appears to be less than the manufacturers claim. Is it significantly lower?

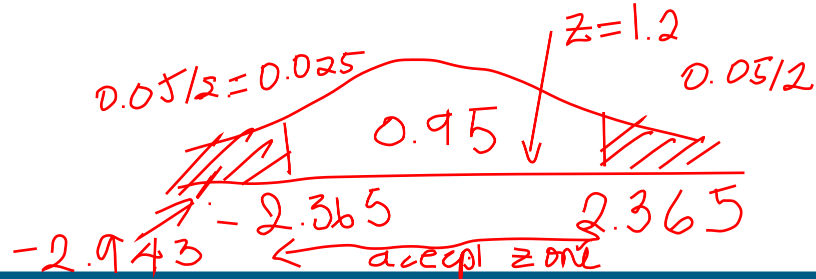
Test Statistic

from sample

$H_A - H_0$

$$\boxed{t = \frac{\bar{x} - \mu}{(\frac{s}{\sqrt{n}})}} = \frac{2959 - 3000}{(\frac{39.4}{\sqrt{8}})} = -2.943$$

ex $t = 1.2$



Curtin College

SOLUTION – 2 TAIL CRITICAL REGION

Use T-table · dof = n-1 = 8-1 = 7

Critical region

- To get a critical value for a t-distributions you have to define a p-value. The question says 5% level of significance
- If you assume 2 tail distribution (which I always would) then the p=0.05 is shared with each side of the distribution so new p=0.05/2=0.025.
- From tables: A dof=8-1=7 and p=0.025 for yields $t = \pm 2.365$
  - Remember the t statistic for the sample data was -2.943

Conclusion

$-2.943 < -2.365$

Since t (sample data)< t(p=0:025, dof=8-1=7)), we reject the null hypothesis at the 2.5% level of significance with a 2 tailed distribution. There appears to be good reason to doubt the manufacturer’s claims.

Curtin College

SOLUTION – 1 TAIL STEP BY STEP

You could also treat this problem as a 1 tailed distribution

Assumptions

- The eight shells constitute an SRS.
- The muzzle velocities can be reasonably modelled by a normal probability distribution.
- σ is unknown.
- State hypothesis
  - H_0: $\mu_0 = 3000\ feet/sec$ (manufacturers claim)
  - H_A: $\mu < 3000\ feet/sec$ (not worried about it being >3000)
    - When you look at question $\bar{x} = 2959$ so the sample data mean appears to be less than the manufacturers claim. Is it significantly lower?

Test Statistic

HA Ho

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{2959 - 3000}{\left(\frac{39.4}{\sqrt{8}}\right)} = -2.943$$

from sample

SOLUTION – 1 TAIL CRITICAL REGION

Critical region

- To get a critical value for a t-distributions you have to define a p-value. The question says 5% level of significance and this is a 1 tailed distribution so p=0.05.
- From tables: A dof=8-1=7 and p=0.05 for a 1 tailed yields t=1.895
  - Remember the t statistic for the sample data was -2.943

t=-2.943

Conclusion

Since t (sample data)< t(p=0:05, dof=8-1=7)), we are inclined to reject the null hypothesis at the 5% level of significance and accept the alternate. There appears to be good reason to doubt the manufacturer’s claims about the velocity and it facts it appears to be significantly less than claimed..

1 TAIL OR 2 TAIL?

There are no definite answers in statistics

I have spoken to statisticians and doing a 2 tail distribution is usually valid and I think easier.

- You just need to establish the confidence interval for the data and state a valid hypothesis
- Then determine the t value for the significance and degrees of freedom
  - Remember you only use t distributions for certain conditions
- If the t value for the significance and degrees of freedom, lies outside the confidence interval you can *probably* not accept the null hypothesis
- If you chose a 1 tail distribution and your assumption was valid/reasonable, your work can still be correct even though your calculation and conclusion might be different to mine.

Curtin College

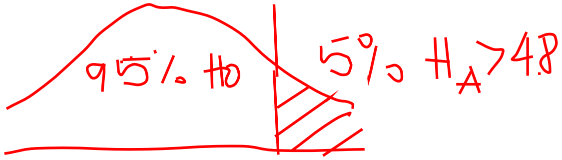
BLOOD QUESTION

The level of phosphate in blood varies normally over time. A dialysis patient has his level monitored on 6 visits to the clinic. $n = 6$

$H_0 = 4.8$

Set up and test an hypothesis at a 5% level of significance to decide if there is evidence that his level is above 4.8 which is the top of the range for healthy people?

5.6, 5.1, 4.6, 4.8, 5.7, 6.4



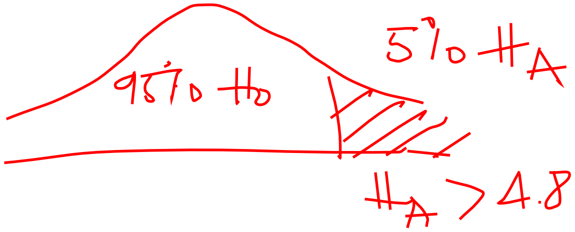
Use the following headings as a guide: H_0 H_A

1. Assumptions (this includes statement of Null & Alternate Hypothesis)
2. Test statistic
3. Critical region(s)
4. Conclusion

Curtin College

Assumptions:

- σ is unknown but it is reasonable to model phosphate level in blood as a t-distribution. (in reality you need evidence from research or historical records to know if this is valid)
- The 6 readings constitute a valid SRS, you are confident that the data is credible and hopefully a good representation of the actual patient levels.
- Hypothesis
  - $\boldsymbol{H_0 = 4.8}$
  - Alternate hypothesis: 1-sided or 2-sided?
    - Does it matter if potassium levels are low or is it only high levels that are an issue for dialysis patients?
    - High potassium levels (hyperkalemia) can be life threatening as can low potassium level (hypokalemia).
    - What are you particularly interested in checking for the patient?
  - For this question the concern is that the levels are high so I am going to choose… $\boldsymbol{H_A > 4.8}$

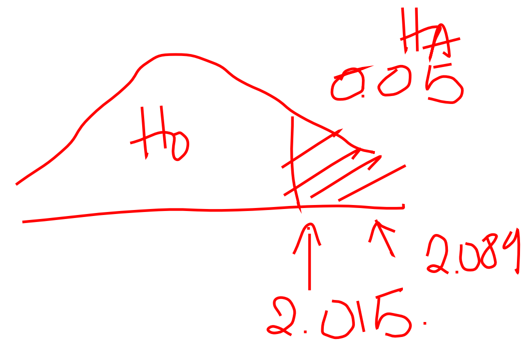


TEST STATISTICS & CRITICAL REGION

- Using calculator $\bar{x} = 5.367$ and s=0.665
- Test statistic for sample data

HA − Ho

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{5.367 - 4.8}{\left(\frac{0.665}{\sqrt{6}}\right)} = \boxed{2.089}$$



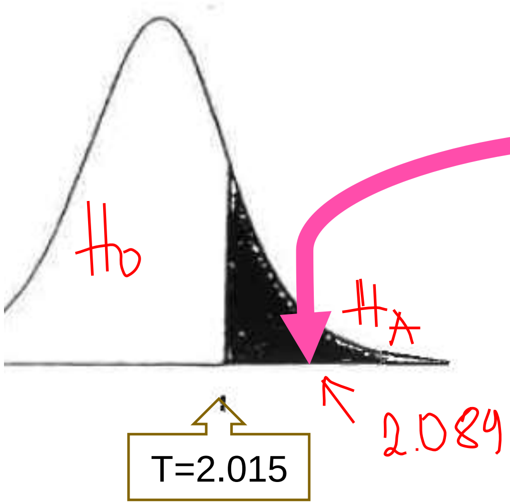
To obtain the critical values for a t-distributions you must define a significance at the beginning and identify sample size. The question has 5% level of significance and we chose a 1 tail hypothesis so p=0.05 is the right hand tail.

- From tables: A $dof = 6 - 1 = 5$, $\alpha = 0.05$ for a 1 tailed $t = \pm 2.015$

BLOOD SOLUTION CONCLUSION

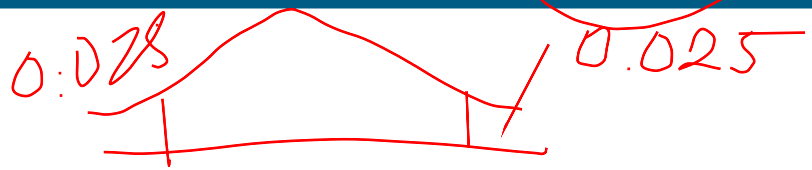
t (sample=2.089)> t=2.015 (p=0:05, dof=6-1=5)),

At the 5% significance level we have reason to be confident that we can reject the null hypothesis that the patients level are in the safe range and that the sample data indicates the average phosphate level of 5.367 is statistically significantly higher than 4.8 and the patient is at risk of hyperkalemia .



Curtin College

COULD I HAVE DONE THIS QUESTION AS A 2-TAIL?



Yes or Maybe

- I hope that you are seeing that the calculations are only a small part of the analysis.
- You need to understand the topic you are analyzing.
  - Is the test for low phosphate the same as for high phosphate?
- Why 5% significance? Why not 1%? (good question)

COULD PHONE CALLS BE A T QUESTION?

Yes… just let’s change the information in the question a little bit.

According to a telephone company’s records from 1999 the average length of a long distance call was 12.44 minutes with a standard deviation of 2.65 minutes

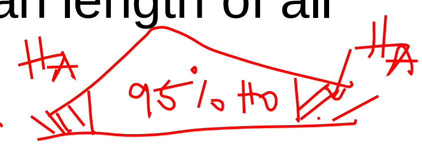
population standard deviation = 2.65 -> use Z table

Management wants to check if the mean length of current long distance calls has changed from 12.44 minutes

A sample of 150 calls places through the company has a mean duration of 13.71 minutes *and a sample standard deviation of 2.65 minutes. It is not known if the current σ is the same as all those years back.*

population standard deviation is unknown -> t table

Using a 5% significance level can you conclude that the mean length of all current long distance calls is different from 12.44 minutes?



$H_0 = 12.44$

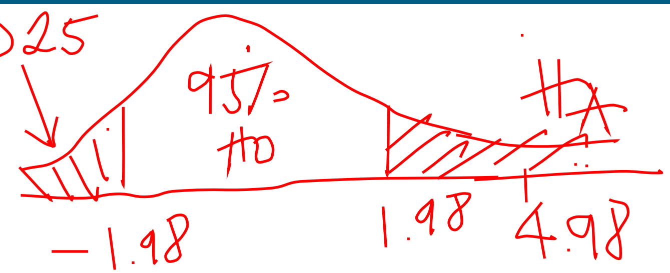
$H_A \neq 12.44 \rightarrow$ 2 tail test

Curtin College

DISCUSSION OF ASSUMPTIONS

- The question doesn't say but for the millions of phone calls it is probably ok to assume the population is normally distributed.
- The historical data is the μ and σ of a normal distribution.
- The sample is random, large enough and hopefully representative of the current year. SRS
- ~~The big assumption in this analysis is that σ of the current year is the same as the historical data. That might not be true...~~
- This time we only have sample standard deviation. Sample standard deviations are usually > population so let's use $s = 3.2$
- P=0.05 is given. (why choose p=0.05?)
- I am choosing a 2 tailed hypothesis as I am unsure if call duration has gone up, down or no change.
- I am using t based on earlier discussion and assumptions.

CALCULATIONS



- H_0: $\mu_0 = 12.44\, minutes$
- H_A: $\mu_A \neq 12.44\, minutes$
- $significance = 0.05 = \alpha,$
  - $therefore\ \frac{\alpha}{2} = 0.025$
- $Critical\ values\ from\ t\ tables, DOF = 150 - 1 = 149,\ t = \pm 1.98$
- $\mu = 12.44, s = 3.20, n = 150\ , \bar{x} = 13.71$
- $Test\ statistic$

  HA − Ho
  - $\boxed{t_{test} = \frac{\bar{x}-\mu}{({}^{s}/_{\sqrt{n}})}} = \frac{13.71-12.44}{3.20/_{\sqrt{150}}} = 4.86$
- $t_{test} = +4.86 > \mathrm{t} = +1.98$

$$\frac{13.71 - 12.44}{3.2} \times \sqrt{n}$$

$\uparrow n \quad \uparrow t_{test}$

Curtin College

CONCLUSION WITH A T

$$t_{test} = 4.86 > \text{t} = +1.98$$

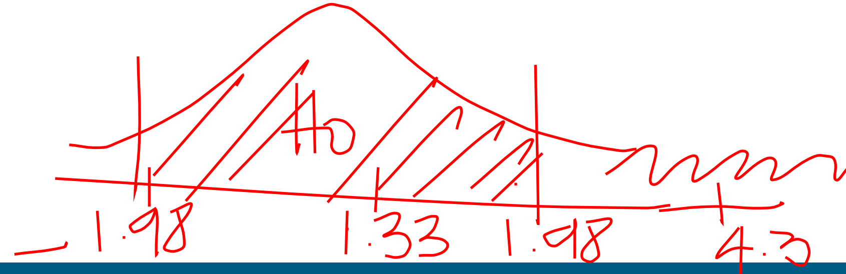
At the 5% significance level I have reason to be confident that I can reject the null hypothesis $H_0\colon \mu_0 = 12.44\, minutes$ (call duration is still 12.44 minutes) and in fact the current duration of 13.77 minutes (n=150) is statistically significantly higher.

This analysis assumes:

- The population is normally distributed.
- The sample of n=150 is random and representative of the population.
- σ is unknown or not valid

Recommendation:

Whilst this analysis appears to providing convincing evidence to reject the null hypothesis, I recommend repeating the the analysis with further samples of the same size or larger to ensure this result is not a false negative.



HOW SAMPLE SIZE IMPACTS T

Phone calls

In t calculations n and sample standard deviation s impact critical values.

- If n=15 (instead of 150) then the critical values of $t = \pm 2.145$ (from tables).
- The t_{test} gets smaller if n decreases.

$$t_{test} = \frac{\bar{x} - \mu}{\left({}^{s}/_{\sqrt{n}}\right)} = (\bar{x} - \mu) * \frac{\sqrt{n}}{s}$$

- Leave $s = \mathit{3.2\ minutes\ then}\ \ t_{test} = \frac{\bar{x}-\mu}{\left({}^{s}/_{\sqrt{n}}\right)} = \frac{13.71-12.44}{{}^{3.2}/_{\sqrt{15}}} = +1.33$

Then $t = -2.145 < t_{test} = +1.33 < \mathrm{t} = +2.145$

These numbers suggests that you cannot reject the Null Hypothesis and average call duration has not changed significantly since 1999 (based on the assumptions and limitations of your analysis)

Curtin College

CONFUSED?

What does this mean?

- If I want to mislead you I can CHOOSE a significance and a sample size that suits me.
- You need to understand the assumptions and limitations of an analysis.
- Do not just accept the final answer or conclusion of any analysis.
- An analysis can be any calculations, not just statistics!

This does not mean all statistics are lies but they must be fit for purpose.

Page 55

if population standard deviation is unknown --> use t table

Curtin College

KITCHEN AND CONFIDENCE

A manufacturer of kitchen mixers employs a market research firm to estimate retail sales of its product by gathering information from a *sample* of retail stores.

simple random sample $n = 50$

In the last month a SRS of 50 stores found the average number of their mixers sold is 24 with standard deviation of 11.

sample / not population $\Rightarrow$ use t table

Question

- Determine the 95% confidence interval for the mean number of mixers sold by all stores in the region
- The distribution of sales is strongly right skewed because there are many smaller stores and only a few large stores. The use of t in the first question is reasonable safe despite the store sizes not being normally distributed. **WHY?**
  - **Central Limit Theorem (CLT) of course!**
  - As the sample size is large the sample means will be approximately normally distributed even though the parent population is not.

sample mean = $\bar{x}$
sample s.d = s

Curtin College

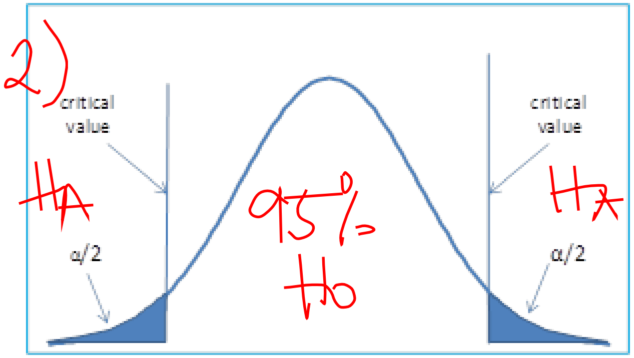
KITCHEN - DO NOT KNOW σ USE T

Sample (must use correct notation)

- $\bar{x} = 24, \mathrm{n} = 50, s = 11$
- 95% confidence, but we are looking at $\pm$ away from the sample mean.
- This is a 2 tailed distribution so the p= 0.05/2 = 0.025
- Looking up t tables p=0.025 and dof=50-1=49 get t=2.010
- Remember that $t = \frac{\bar{x}-\mu}{(\frac{s}{\sqrt{n}})}$ and for a 2 tail there is an upper & lower t
- Rearrange to get $\bar{x} \pm t_{(n-1,\frac{\alpha}{2})} * \frac{s}{\sqrt{n}}$ (page 62)
- $= 24 \pm 2.010 * \frac{11}{\sqrt{50}}$
- $= 24 \pm 3.127$

Confidence Interval = (20.873,27.127)

https://i2.wp.com/www.real-statistics.com/wp-content/uploads/2012/11/two-tailed-significance-testing.png?resize=303%2C181



t=-2.010 $\bar{x} = 20.873$

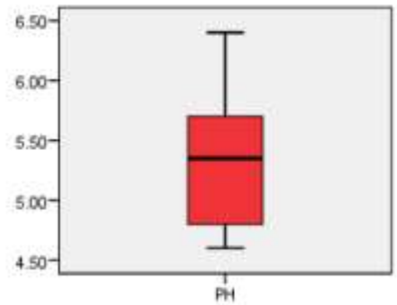
t=+2.010 $\bar{x} = 27.127$

t=0 or $\bar{x}$= 24

Curtin College

USING STATISTICAL PACKAGES FOR BIGGER DATA SETS

Check underlying assumption where ever possible



Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
PH	.156	6	.200*	.955	6	.783

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

- No extreme skewness, no outliers
- No departure from Normality

⇒Test is valid

Curtin College

FOOD POISONING

This table provides data on a group of people who contracted botulism, a potentially fatal food poisoning.

The variables record: age, incubation period (the time in hours between eating the infected food and the first sign of illness),and whether the person survived (S) or died(D).

Test if the mean incubation period for the population is more than 30 days with a confidence of 95%.

$H_0 = 30$

$H_A > 30$

Person	1	2	3	4	5	6	7	8	9
Age	29	39	44	37	42	17	38	43	51
Incubation	13	46	43	34	20	20	18	72	19
Outcome	D	S	S	D	D	S	D	S	D
Person	10	11	12	13	14	15	16	17	18
Age	30	32	59	33	31	32	32	36	50
Incubation	36	48	44	21	32	86	48	28	16
Outcome	D	D	S	D	D	S	D	S	D

$n = 18$

HYPOTHESIS & TEST ASSUMPTIONS

Null Hypothesis – mean incubation period = 30 days

Alternate Hypothesis – mean incubation period is more than 30 days

First of all check Normality assumptions

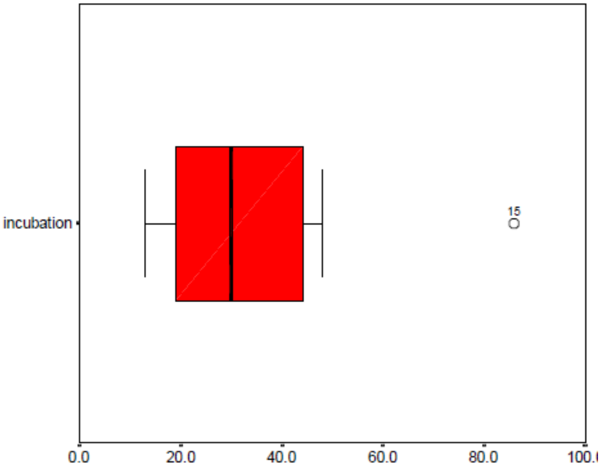
- Box plot looks "ok" – fairly symmetrical
- Shapiro-Wilk Sig is >0.05
- Test is valid

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
incubation	.187	18	.098	.847	18	.010**

**. This is an upper bound of the true significance.

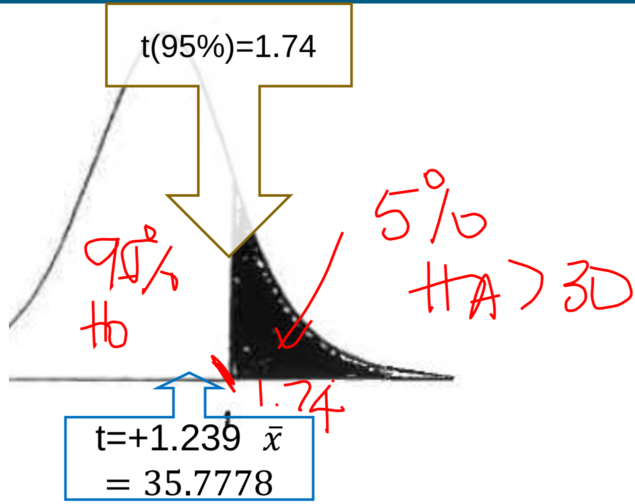
a. Lilliefors Significance Correction



TEST STATISTIC, CRITICAL REGION & CONCLUSION

The outputs for the sample are.

- $\bar{x} = 35.7778$, s= 19.7917   HA - Ho
- **Test statistic**: $t = \frac{\bar{x}-\mu}{(\frac{s}{\sqrt{n}})} = \frac{35.7778-30}{(\frac{19.7917}{\sqrt{18}})} = 1.239$
- **Critical region** :
  - 95% confidence 1 tail with dof = 18-1=17
  - t=1.74



Test t=1.239 < t(95% confidence)=1.74

With a 95% confidence we can accept the Null Hypothesis that a sample mean incubation period as high as 35.78 would occur if the μ= 30 days.

Conclusion

HA

The data does not support the claim that the mean incubation period is more than 30 days.(t=1.239, dof=17, one tailed)

As Test t=1.239 < t(95% confidence)=1.74

.

DO NOT "GUESSTIMATE" SIGNIFICANCE

I find this result amazing. Remember the question was:

Test if the mean incubation period for the population is more than 30 days with a confidence of 95%.

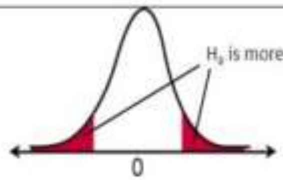
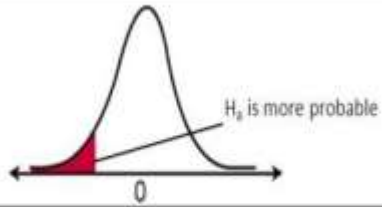
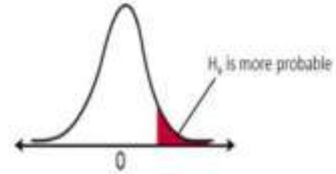
Because we did not know the population standard deviation we had to use a t distribution which is more spread out.

Even though the sample mean was 35.7778, it is not sufficiently statistically different from 30 for us to be **confident** that this is more than natural variation in sample means based on the information we had.

page 62

Curtin College

FORMULAE SUMMARY

Week 2 Basic Normal distribution	Week 3 Sampling	Week 4 Sampling
Known: population	Population is ND σ is known sample	Population is ND σ is unknown t - table
$z = \dfrac{\bar{x} - \mu}{\sigma}$	$z = \dfrac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$	$t = \dfrac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)}$
	Confidence interval $\mu = \bar{x} \pm \dfrac{z\sigma}{\sqrt{n}}$	Confidence interval $\boxed{\mu = \bar{x} \pm \dfrac{ts}{\sqrt{n}}}$
	Margin of error $\pm \dfrac{z\sigma}{\sqrt{n}}$	Margin of error $\pm \dfrac{ts}{\sqrt{n}}$
	Hypothesis testing	
$H_0 = \mu_0 = 3$ $H_A = \mu_A \neq 3$	$H_0 = \mu_0 = 10$ $H_A = \mu_A < 10$	$H_0 = \mu_0 = 5$ $H_A = \mu_A > 5$
		

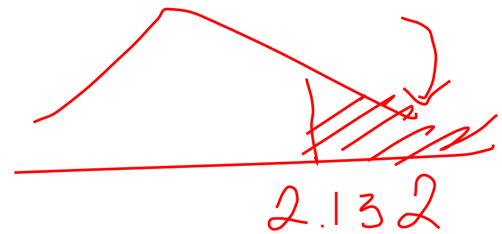
Curtin College

PAST CU EXAM QUESTIONS

Don't Know pop S.D
⇒ use t- table

sample
n=5 $\bar{x} = 55$, s=3

t table



Question 4. Curtin University claims that the mean salary of its graduates is \$67K (\$67,000). You think Curtin over-estimates the average salary with the aim of attracting more students. You then conduct a survey by randomly sampling 5 Curtin's graduates' salaries and find that the mean salary is $\bar{x}$ = \$55K with the standard deviation s = \$3K.

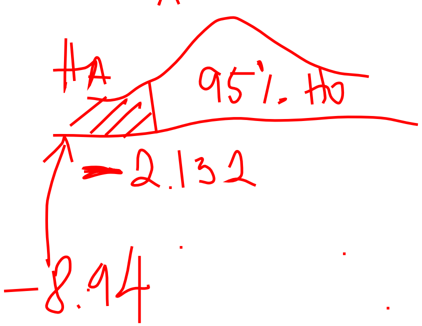
p=0.05

(a) Perform a test of hypothesis at the 5% significance level with the intent to prove that Curtin University over-estimates the average salary. (6 marks)

(b) If we accept Curtin University's claim of the graduate's mean salary μ = \$67K with the standard deviation σ = \$1K, how large a sample is required if we want a 95% confidence interval for the mean μ to have a margin of error of ± 0.05? (4 marks)

(a) $H_0 = 67$
$H_A < 67$

Assumptions:
1/normal distributed population
2/ sample is a simple random sample



dof = n-1 = 4
⇒ t = 2.132

$$t_{test} = \frac{H_A - H_0}{s/\sqrt{n}}$$

$$= \frac{55-67}{3/\sqrt{5}} = -8.94 < -2.132 \Rightarrow \text{reject } H_0$$

t - table

sample n=20
$\bar{x} = 495$
s = 6

Question 4. This is a continuation of Question 3 above. ACCC wants to assess independently whether there is any significant evidence to support Kellogg's claim that each cereal box contains 500g of cereal. Using the randomly chosen 20 boxes of the product, ACCC finds that the average weight per box is $\bar{x}$ = 495g with the sample standard deviation s = 6g.

P=0.05

(a) Perform a test of hypothesis at the 5% significance level with the intent to show that Kellogg over-estimates the average weight of cereal in a box. (6 marks)

(b) If we accept Kellogg's claim of the average cereal weight per box is μ = 500g with the standard deviation σ = 10g, how large a sample is required if we want a 98% confidence interval for the mean μ to have a margin of error of $\pm$5g? (4 marks)