MM104 / MM107
Statistics and Data Presentation Semester 2

Project 3: Hypothesis Testing: 2 Sample t-Tests

Ainsley Miller
ainsley.miller@strath.ac.uk

Lecture Overview

In this lecture we will

- discuss hypothesis testing.
- introduce the concept of one and two tailed hypothesis testing.
- introduce test statistics.
- learn the assumptions of the z and t test.
- learn the formal definition of a p-value and how to interpret them.
- learn the steps of a hypothesis test
- go through an example of how to carry out a t test in Minitab. We will interpret the results and check the assumptions.
- introduce effect size and how to calculate and interpret the obtained results.

Hypothesis

Generally when you collect data you have a research question you are trying to address. This is called your **hypothesis**.

Examples

1. Are the average marks of mathematics students higher than those for chemistry students ?
2. Does exercise increase heart rate?
3. Does one drug for cancer perform better than another?
4. Do men earn more than women ?

Hypothesis cont...

As statisticians we want to formalise things a bit. We separate our hypothesis into the:

- H_0 : Null Hypothesis
- H_1 : Alternative Hypothesis

The alternative hypothesis is the claim we are trying to make.

Relating that to the example

Are the average marks for mathematics students higher than those for chemistry students ?

- H_0 : Average marks for mathematics students are the **same or lower** than chemistry students.
- H_1 : Average marks for mathematics are **higher** than chemistry students.

The null hypothesis is the opposite of the alternative hypothesis and should contain all other possibilities (this is the reason we have the same or lower in the null hypothesis).

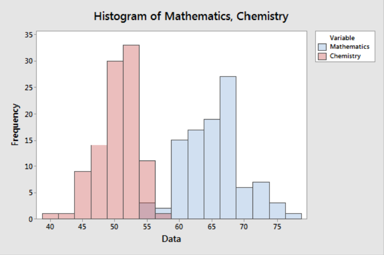
How do we go about this ?

Firstly we should plot the data to get a feel for it and find **summary statistics**.

Summary statistics are simply a collection of descriptive statistics e.g. the mean, median, standard deviations and the inter-quartile range.

Simple Plot and Summary Statistics

Subject	Mean	Median	St Dev	IQR
Chemistry	50.3409	50.44	2.95	3.38
Maths	65.2459	65.50	4.35	6.13



Conclusion

Some people may simply conclude that since the mean mathematics mark is greater than the mean chemistry mark.

There is a problem with this !

This is very subjective, some people have have different ideas about how far apart the means need to be for there to be a difference.Also is the sample big enough to make this conclusion ?

We formalise this with **hypothesis testing**.

Hypothesis Testing

- In hypothesis testing we are testing a specific claim about a population.
- We take a sample from the population and calculate summary statistics e.g. the sample mean.
- Then we would look at the relative position of the population mean (what we are interested in comparing it to) and sample mean, but any difference here could be due to randomness. This is the sample if you are comparing two samples, any difference in the samples can be due to randomness.
- Hypothesis testing allows us to decide whether our data support the claim.

Hypothesis Testing cont...

Recall that the claim we want to make provides our alternative hypothesis, and our null hypothesis is the opposite of the claim.

We want to know if our sample gives evidence in favour of H_1 over H_0.

There are two conclusions you can draw from a hypothesis test

1. Reject H_0 in favour of H_1.
2. Do not reject H_0.

One or Two Tailed ?

The alternative hypothesis can take the following forms:

- is there a difference between the means ($\neq$): **two tailed**. (Note you are not saying what the difference is, you are simply acknowledging there is one).
- one mean is greater than the other ($>$): **one tailed test**
- one mean is smaller than the other ($<$): **one tailed test**.

This week you will need to decide if you want to carry out a one or two tailed test. This will depend on the research question you are posing.

Test Statistic

The decision of whether to reject (or not) H_0 is based on information contained with the sample.

We use a **test statistic** which is a single number calculated from the sample data.

This week we are looking at comparing means so there are two situations we may be in

1. We know the population standard deviation so therefore we use a z-statistic which is based on the standard normal distribution i.e. mean 0 and standard deviation 1.
2. We only know the sample standard deviation so therefore use a t-statistic which is based on the t-distribution.

Test Statistics: Z test

One sample

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

where $\bar{x}$ is the sample mean, μ is what we are interested in comparing that mean too (population mean), σ is the known population standard deviation and n is the sample size.

Test Statistics: Z test cont...

Two sample

$$z = \frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $\bar{x_1}$ and $\bar{x_2}$ are means of sample 1 and 2 respectively.
$(\mu_1 - \mu_2)$ is called the **hypothesised difference**.
The hypothesised difference is the difference we are interested in seeing. If we were interested in for example if $\bar{x_1} = \bar{x_2}$ then $(\mu_1 - \mu_2)$ would disappear (equal zero). σ^2 is the population variance, and n_1 and n_2 are sample sizes of sample 1 and 2 respectively.
Remember that sample 1 and sample 2 may not be the same size.

Test Statistics: t test

One sample

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}, \qquad s = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

where $\bar{x}$ is the sample mean, μ is what we are interested in comparing that mean too (population mean), s is the sample standard deviation, n is the sample size and x_i where $i = 1 \ldots n$ is each observation.

Test Statistics: t test cont...

Two sample

$$t = \frac{(\bar{x_1} - \bar{x_2}) - (\mu_1 - \mu_2)}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \qquad s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

where $\bar{x_1}$ and $\bar{x_2}$ are means of sample 1 and 2 respectively. $(\mu_1 - \mu_2)$ is called the **hypothesised difference**. s_1^2 and s_2^2 are the sample variance of sample 1 and 2 respectively, and n_1 and n_2 are sample sizes of sample 1 and 2 respectively.

You do not need to know the formula for the t-test by heart.

Degrees of Freedom for a t-test

The degrees of freedom in a two sample t test is $n_1 + n_2 - 2$ where n_1 is size of sample 1 and n_2 is the size of sample 2.

We will use the degrees of freedom for a two sample t-test when we learn about written statistics in weeks 8-10.

Assumptions

Both the t-test and z-test have underlying assumptions which have to hold for the test to be valid:

- The sample (or samples) must be normally distributed – or the sample size must be large (above 20).
- The test statistics described also assume that both samples have equal variances (i.e. the spread of values should be similar).
- The Z-test also assumes that population variances are known but in the real world this is unlikely to be the case and so a t-test is usually adopted.
- If these don't hold there are other options, but these are not covered in this course.

p-Value

From a hypothesis test you get a p-value.

> The p-value is the probability of observing a value at least as extreme as the test statistic assuming that the null hypothesis is true.

We compare the p-value to our significance level α (usually 0.05)

- If your p-value $< \alpha$ then we reject the null hypothesis.
- If your p-value $> \alpha$ then we fail to reject the null hypothesis.

Steps of a Hypothesis Test

1. State the null and alternative hypotheses. (You should not have looked at your data first).
2. Calculate your descriptive statistics to get a feel for the data (think about skewness etc).
3. Calculate the test statistic and p-value.
4. Given the significance level α determine whether the result is significant.
5. Decide whether to reject or fail to reject the null hypothesis.
6. Write your results in the context of the question.
7. Check your assumptions.

Let's go back to our example.

We are going to be doing a t-test since we do not know the population standard deviation.
Using our steps we will firstly write our null and alternative hypothesis.

- H_0 : Average marks for mathematics students are $\leq$ chemistry students.
- H_1 : Average marks for mathematics are $>$ chemistry students.

This is a one tailed t-test.

We can also say that $\mu_{maths} - \mu_{chemistry} > 0$.

Calculating Summary Statistics in Minitab

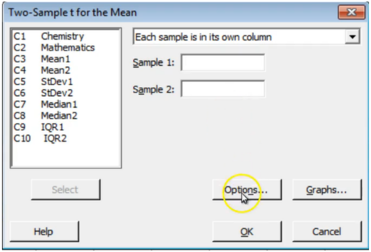
To calculate these in Minitab we do

> Stat > Basic Statistics > Store Descriptive Statistics choose the appropriate columns of your data set. Press Statistics and select the descriptive statistics you are interested in.

An informal test for skewness is if the mean and median are close, from the descriptive statistics above the mean and median are similar for both the maths (65.2459 & 65.50) and chemistry (50.3409 & 50.44) so we can say these are approximately symmetric. Note the standard deviations are quite different, we will come back to this...

t-Tests in Minitab

Stat > Basic Statistics > 2 Sample t.



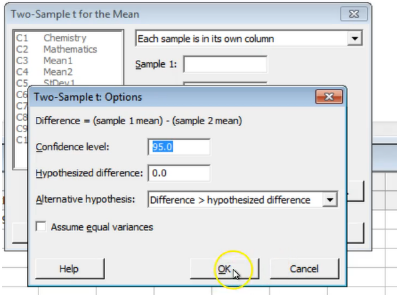
Note that "Each sample is in its own column" is not the default you need to change this. Do not select your Sample 1 and 2 until you have pressed *Options*.

t-Tests in Minitab cont...

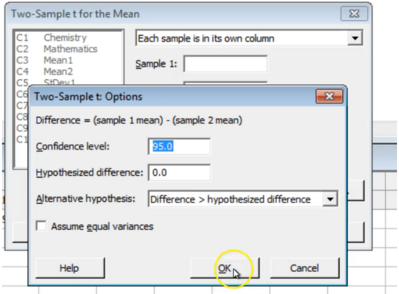
After clicking on *Options* type in the appropriate confidence level i.e. 95.0 since we have a 5 % significance level.

Select the appropriate *Alternative hypothesis*. From our earlier conjecture we want difference > hypothesised difference.

Minitab states the difference as (sample 1 mean)-(sample 2 mean). Therefore sample 1 is Maths and sample 2 is Chemistry.

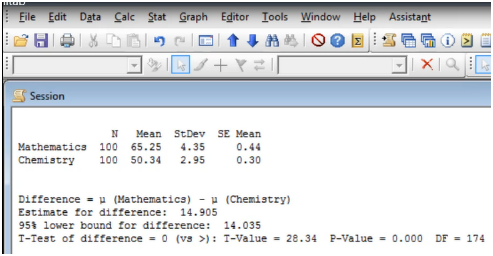


Assume Equal Variances ?



In the above we have left *Assume equal variances* unticked as the variances (or standard deviations) are not equal. Recall, the standard deviation for Chemistry is 2.95, and the standard deviation for Mathematics is 4.35.

Results



The estimate for the difference is 14.905. This comes from the difference of the means i.e. $65.2459 - 50.3409$.
Note the means have been rounded to 2 dp in the output, so make sure you use *Store Descriptive Stats* for more accurate results.

Results cont...

The t-test statistic is 28.34, and we are looking at the probability of a t distributed variable being greater than 28.34.

Generally if your t-statistic is > 2 (or has an absolute value > 2) then your test is likely to be significant. Since our t-test statistic is much greater than 2 it is very significant hence why our p-value is so small.

The p-value is stated as 0.000 but **this does not mean your p-value is 0**. It is extremely unlikely that your p-value is exactly 0. If you get a result like this you report it as > 0.001.

Never say that your p-value is 0 !

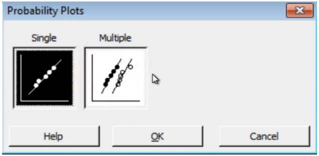
Interpreting the Results

Since our p-value is < 0.001 and our t-test statistic is > 2, we reject our null hypothesis and conclude that average marks for mathematics are statistically significantly higher than the average marks for chemistry.

Checking Assumptions

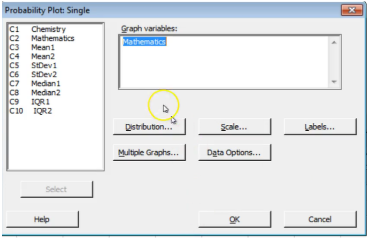
For our result to be valid we need to check that the marks for Chemistry and Mathematics and approximately normal. To do that we do a test for Normality.

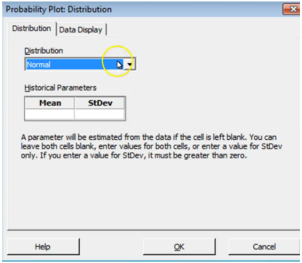
Graph > Probability Plot > Single.
It is best practise to do them one at a time (i.e. single) as if you press multiple the results can be hard to interpret.



Checking Assumptions cont...

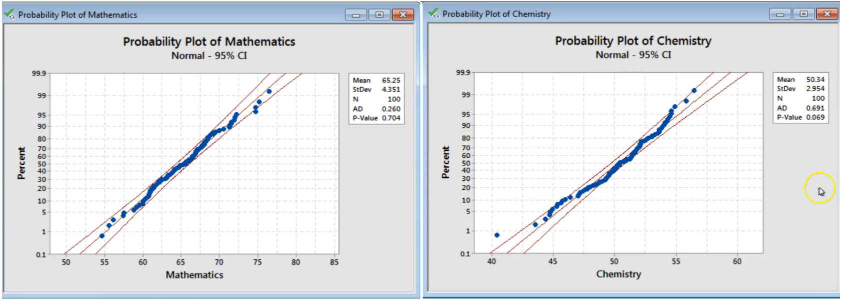
We will do Mathematics first, and clicking on *Distribution* brings up the menu on the right. Just check the distribution is Normal. We do not know the historical means or standard deviations so leave that blank and Minitab will estimate it.





Checking Assumptions cont...

We are looking for the points to lie roughly on the line.



Anderson Darling Test for Normality

These results also carry out the **Anderson Darling test for Normality** this is where the p-value and the AD are coming from. The Anderson Darling has the following hypotheses

1. H_0: Data are normally distributed.
2. H_1: Data are not normally distributed.

Since the p-values for both of these results are greater than 0.05 we fail to reject our null hypothesis and conclude that chemistry and mathematics marks are normally distributed.

What does this mean?

Since we can say that our data for mathematics and chemistry follow a normal distribution, the data meets the assumptions for the t-test and the results hold.

What about constant variance ?

Although our data-sets do not have constant variance this has been taken into account in our test by leaving the *Assume equal variances* check box unticked.

Effect Size

Effect size is a measure of how different two groups are from one another.
You should only calculate an effect size if your result is significant i.e. p-value < 0.05.

- A small effect size: $0 \leq ES \leq 0.2$.
- A medium effect size: $0.2 < ES \leq 0.5$
- A large effect size: $ES > 0.5$

An Effect Size of 1 indicates the two groups differ by 1 standard deviation, an Effect Size of 2 indicates they differ by 2 standard deviations, and so on.

Calculating Effect Size

We calculate the effect size for a two sample t-test via

$$ES = \frac{|\bar{x_1} - \bar{x_2}|}{\sqrt{\left(\frac{s_1^2 + s_2^2}{2}\right)}},$$

where $\bar{x_1}$ is the sample mean of sample 1, $\bar{x_2}$ is the sample mean of sample 2, and s_1^2 and s_2^2 are the sample variances of sample 1 and sample 2 respectively.

Effect Size - Interpreting the Result

If the effect size is zero then this means that both groups tend to be very similar and overlap entirely.

As the effect size gets larger it reflects an decreasing overlap between the two groups.

Effect size and p-value

A lower p-value is sometimes interpreted as meaning there is a stronger relationship between two variables. However, statistical significance means that it is unlikely that the null hypothesis is true (less than 5 %).

> Therefore, a significant p-value tells us that an intervention works, whereas an effect size tells us how much it works.