

## Title: Conducting a Manual t-test

**Target:** On completion of this worksheet you should be able to identify key assumptions and conduct a manual t-test for normal distribution

### Key Information to start:

A t-test is a method in statistical hypothesis testing to compare two means, and assess whether the means of these two groups are statistically different from each other. There are 3 different types: independent samples, paired samples (ST3 and ST4) and one sample t-tests.

### One sample t-test

Compares the means of your sample data to a known value. You should run a one sample t-test when you don't know the Population Standard Deviation or if you have a small sample size. Upper or lower one sample t-test will have its alternative hypothesis as either greater than or less than, while a two tailed one sample t-test has its alternative hypothesis as not equal to. An example would be: The manager believes the production rate has a mean above the original mean at 20 per hour. Our hypothesis would be

$$H_0 : \mu = 20$$

$$H_1 : \mu > 20$$

Another example would be: A professor believes the amount of students turning up to a particular class throughout the year is not the mean of last year at 59. Our hypothesis would be

$$H_0 : \mu = 59$$

$$H_1 : \mu \neq 59$$

### Assumptions of one sample t-test

- 1) the data has to be independent and continuous
- 2) data has to be collected at randomly
- 3) the data is approximately Normally Distributed

### t-score

the t-score is given by

$$\text{t-score} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

where

$\bar{X}$  is the sample mean

$\mu$  is the population mean

$s$  is the sample standard deviation

$n$  is the number of observations

$H_0$  is the null hypothesis saying the means are the same or a certain value

$H_1$  is the alternative hypothesis where the means are less than, greater than or unequal to each other or a certain value.

### Accept or Reject Null Hypothesis

We would look in the t-table, using our Alpha level, usually 5%, and use degrees of freedom,  $n - 1$ , to find our t-crit value. We compare our t-crit value and t-score and evaluate their values. If  $\text{t-score} > \text{t-crit}$ , then we reject null hypothesis. If  $\text{t-score} < \text{t-crit}$  we accept null hypothesis.

### Confidence Interval

We can calculate our 95% confidence level as long as we know our  $\bar{X}$ . If we rejected the null hypothesis, then we would need a new  $\bar{X}$  based upon the data. Our confidence interval at 95% tells us that we are 95% confident that "true" value lies within this region. We calculate this by the following formula

$$\bar{X} \pm \text{t-crit} \times \frac{s}{\sqrt{n}}$$

**Example 1**

Your company wants to improve sales. Past sales data indicate that the average sale was £100 per transaction. After training your sales force, recent sales data from 25 salesmen and women indicates an average sale of £130, with a standard deviation of £15. Using 5% alpha level, determine if the training was successful.

**Answer**

Null hypothesis is  $H_0 : \mu = £100$

Alternative hypothesis is  $H_1 : \mu > £100$

Therefore this is a one-tailed t-test

$$\bar{X} = £130$$

$$\mu = £100$$

$$s = £15$$

$n = 25$  So by using our formula

$$t\text{-score} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

We get

$$t\text{-score} = \frac{130 - 100}{\frac{15}{\sqrt{25}}} = 10$$

Our degrees of freedom is 24, so using 5%, our t-crit value is 1.711.

Since  $10 > 1.711$ ,  $t\text{-score} > t\text{-crit}$ , we reject our null hypothesis, meaning the training did work in improving sales.

**Example 2**

Calculate the 95% confidence interval for this group of data

5.4 6.1 5.7 4.9 5.8 5.6 5.1

**Answer**

Firstly we find the mean and standard deviation

$$\bar{X} = \frac{5.4 + 6.1 + 5.7 + 4.9 + 5.8 + 5.6 + 5.1}{7}$$

$$= 5.514285714$$

sample standard deviation is given by using the formula

$$\sigma = \sqrt{\frac{\sum(X - \bar{X})^2}{(n - 1)}}$$

Putting our values in the equation gives us a standard deviation of 0.414039

Our t-crit is 2.447 since we use two tail as there's no direction so alpha becomes half of 0.05 to be 0.025

Our confidence interval is

$$5.514285714 \pm 2.447 \times \frac{0.414039}{\sqrt{7}}$$

$$= 5.514285714 \pm 0.383325939$$

(5.13, 5.90) to 2d.p

this means we are 95% confident that our value lies in this range

**Exercise using 2 independent samples t-test**

A teacher wanted to see if students performed better in a classroom or online, both within timed conditions. There was 9 students for classroom and 9 students for online, and both were marked out of 45. Here is the table

Online	Classroom
32	35
37	31
35	29
28	25
41	34
44	40
35	27
31	32
34	31

Do these data present sufficient evidence to indicate that the average grade for students who take the exam online is significantly higher than those who performed in a classroom?

**Answer**

We let  $\mu_1$  be the mean scores for the online group and  $\mu_2$  for the classroom group. Our null and alternative hypothesis would be

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 > \mu_2$$

The standard deviation using the formula

$$\sigma = \sqrt{\frac{\sum(X - \bar{X})^2}{(n - 1)}}$$

We get  $\sigma_1 = 4.9441$  and  $\sigma_2 = 4.4752$ . Since we have two independent samples, we would need to calculate the pooled estimate which gives us a combined variance that is reasonable between the two sets of data. The formula is

$$S^2 = \frac{(n_1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}$$

Ours will become

$$S^2 = \frac{(9 - 1)4.9441^2 + (9 - 1)4.4752^2}{9 + 9 - 2} = 22.2361$$

### Answer Continued

Our t-value is calculated by using a pooled sample is the following fomula

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

Using our data we get

$$t = \frac{35.22 - 31.56}{\sqrt{22.2361(\frac{1}{9} + \frac{1}{9})}} = 1.65$$

Since it is a one tailed test shown by the greater than sign on the alternative hypothesis, we only look at the upper tail and have  $\alpha = 0.05$  will remain as 0.05. Our degrees of freedom for pooled is

$$n_1 + n_2 - 2 = DF$$

In our case our degrees of freedom are

$$9 + 9 - 2 = 16$$

Looking in our t-tables, our t-crit is 1.746

Comparing our t-crit to our t-value, we have

$1.65 < 1.746$ , so therefore since crit is higher than

our value we accept  $H_0$  at a 5% level of significance.

### t-test in Regression

To use a t-test regression, we must look at the regression line in our graph. A linear regression line has the equation  $y = mx + c$  where  $x$  and  $y$  are the axis or values plotted and  $m$  is the gradient and  $c$  as the intercept. We must find our  $m$  and  $c$ . We can do this by plotting two points on the line to find our equation. When we got our coordinates we do

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

We can of course have more than 1  $x$  and more than 1  $m$ , e.g.  $y = m_1x_1 + m_2x_2 + c$

We can then put one of the points of coordinates as  $y$  and  $x$  and use the value for  $m$  into our equation to find the remainder which will be  $c$ . In an ANOVA table, our  $m$  will be our (constant)B value and  $c$  will be our (variable)B. Our standard error should be given to us. Our Beta is what the regression coefficient would be if the model were fitted to standardised data and should already have been told. Our t-values are calculated by doing  $B \div std.error$ . We compare these values to our t-crit in our statistical tables, and if our  $t\text{-value} > t\text{-crit}$  we put in our significance  $< 0.05$ , if it is less than then we put  $> 0.05$  or what value alpha is when it becomes less than.

An example in the ANOVA tables will look like

Model	B	Std. Error	Beta	t	Sig.
(Constant)	373.743	55.067		6.787	< 0.05
Score (50)	13.897	1.748	0.882	7.951	< 0.05

### Example on regression

We are given this table from a car dealer in the price of cars from old to new depending on their milage.

They have looked at 2500 cars

Model	B	Std. Error	Beta	t	Sig.
(Constant)	6000	35.267		?	?
Age	-20	7.834	0.462	?	?
Miles	-65	4.862	0.225	?	?

Find the regression equation mathematically. Then work out the missing values using the table figures and our t-test table.

### Answer

Our regression equation is given by having variables at each section except the constant, as the constant is the intercept. So our equation will look like (Car price) $y = 6000 - 20x_1(\text{Age}) - 65x_2(\text{Miles})$

To work out the t-value we use  $B \div \text{Std. Error}$

$$6000 \div 35.267 = 170.131$$

$$-20 \div 7.834 = -2.553$$

$$-65 \div 4.862 = -13.379$$

We would now compare to the t-crit value. Since our data size is large (2500), we would use the  $\infty$  row. We would compare till we either run out of numbers or our t-value is smaller than the t-crit.

For 170.131 our Sig. would be  $< 0.05$

For  $-2.553$ , we ignore the minus sign, our Sig. would be  $> 0.05$

For  $-13.379$  we get our Sig. to be  $< 0.05$

### Exercises

- Find the 95% confidence interval for this group of data  
6, 7, 9, 3, 6, 10, 7, 9, 5, 8, 10
- A company creates Jam and put these into Jars of 500ml, however because of the production of the Jam, each jar weighs differently and wants to put an average weight figure on the jar. the intended weight for the Jam in the jar is 210g but in a sample of 20 jars, the average weight is 215g with a standard deviation of 3.75g. Using the t-test, check whether the company can use 210g as weight or needs to be changed.
- A large chain supermarket sells 5000 packets of cereal in each of its stores each month. It decides to test-market a different brand of cereal in 15 of its stores. A month later, the 15 stores sold an average of 5200 packets each, with a standard deviation of 500 packets. Should all supermarkets switch to a selling the new brand? Find the 95% confidence interval of the accepted hypothesis.

**Answers**

1. t-crit is 2.228,  
 $7.2727 \pm 1.474553$   
(5.798, 8.747)

2. t-score > t-crit  
 $5.96284794 > 1.729$   
reject null hypothesis, company needs a new  
weight on the jar.

3. the null is accepted as 1.55 is less than the critical  
value of 2.62. the confidence interval is  
(4861.76, 5538.24)