

# community project

encouraging academics to share statistics support resources

All stcp resources are released under a Creative Commons licence

stcp-marshall-indepE

The following resources are associated:  
 Checking normality in Excel and the Excel dataset 'Birthweight reduced'

## Independent t-test in Excel

**Dependent (outcome) variable:** Continuous (scale)

**Independent (predictor) variable:** Binary (2 independent groups)

**Common Applications:** Comparing means of data from two unrelated groups on the same continuous, dependent variable; for example, you could use an independent t-test to test whether first year graduate salaries differed based on gender or whether there is a difference in test anxiety based on educational level which has two groups: "undergraduates" and "postgraduates".

**Data:** The data set '*Birthweight\_reduced*' contains details of 42 babies and their parents at birth. The research question is whether the mother smoking has an effect on the birthweight of a baby so the dependant variable is Birth weight (lbs) and the independent variable is whether or not the mother smokes (smoker).

Birthweight	Gestation	smoker	motherage	mmocig	rheight
5.80	33	0	24	0	58
4.20	33	1	20	7	63
6.40	34	0	26	0	65

In order to use the t-test command, the data for the two groups (Smokers and Non-smokers) has to be in separate columns so it is best to take the two variables of interest (Birthweight and smoker) to a new sheet. Sort the data by the variable 'smoker' and move the birthweights for the smokers to column B.

Before carrying any analysis, summarise birthweight (Birthweight) by smoking (smoker) using some summary statistics. The command `=AVERAGE(A:A)` will calculate the mean for everything in column A (which is the non-smoker data here) and `=ST.DEV.S(A:A)` will calculate the sample standard deviation.

	A	B	C	D	E	F
1	Non-smoker	smoker			Non-smoker	smoker
2	5.8	4.2		Means	7.69	6.88
3	6.4	4.5		Standard Deviation	1.148	1.389
4	6.8	5.8				

Do the group means and standard deviations look similar or very different?

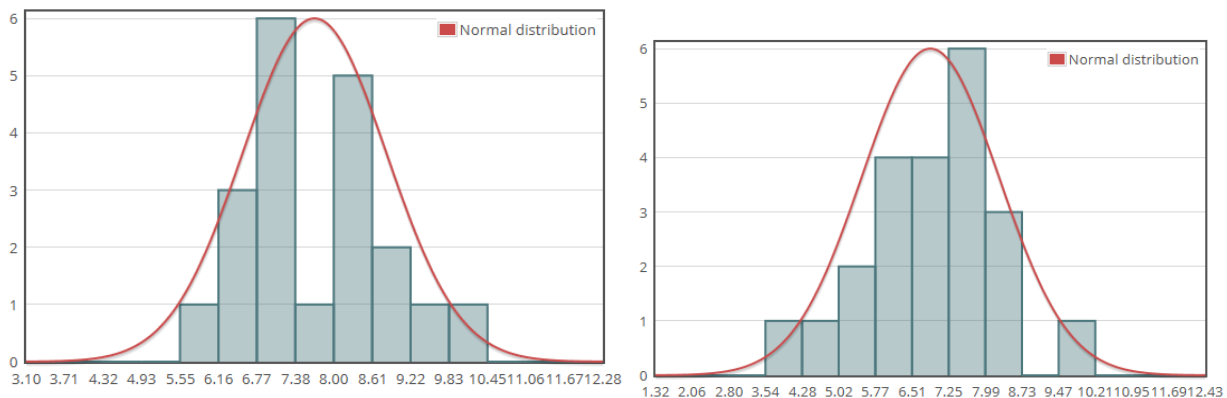
## Checking the assumptions

Assumptions	How to check	What to do if the assumptions is not met
The dependent variable should be approximately normally distributed for each group	Use histogram, QQ plots, normality tests by group For more details see the <i>Checking normality in Excel</i> resource.	If the data for either group are very skewed, use the Mann-Whitney test
Homogeneity (equality) of variance: The variances (SD squared) should be similar for all the groups	If one SD is more than twice the other, the assumption has not been met.  Alternatively, use the F-test command <code>=F.TEST(group1,group2)</code> where $p < 0.05$ indicates unequal variances.	If one SD is more than twice the other, the results of the t-test are less reliable. An adjustment can be made to the t-test command to account for this.

## Checking the assumptions for this data

To check the assumption of normality, produce histograms of the dependent variable birthweight for each group. Excel can produce histograms using the 'Data analysis toolpak' add in but using a website such as the one suggested here is often simpler and produces more options:

<https://www.ai-therapy.com/psychology-statistics/distributions/normal>



Both histograms are approximately normally distributed so the assumption has been met. QQplots or normality tests could also be used to assess normality (see *checking normality in Excel* resource for more details).

For checking the equality of variances, the ratio of the standard deviations is calculated by dividing the largest by the smallest standard deviations so here

`=MAX(E3:F3)/MIN(E3:F3)` gives the value 1.21. As this is smaller than 2, the assumption has been met. The assumption can also be checked using the F-test command and the raw data, which is in columns A and B here. The command `=F.TEST(A:A,B:B)` gives a p-value of 0.4069. A p-value of less than 0.05 indicates the variances may not be equal. Again the assumption has been met.

Checking equality of variance		
Test	Value	Unequal variances if:
SD ratio	1.21	SD ratio > 2
F-test	0.4069	$p < 0.05$

Warning: If you have a large sample size the F-test may give a significant p-value even when the standard deviations are similar so look at the standard deviations as well as the test results.

## Steps for the independent t-test in Excel and output

The independent t-test tests the null hypothesis 'There is no difference in the mean birthweights of babies whose mothers smoke and don't smoke'. The null is rejected if the p-value for the t-test is less than 0.05 and statistically significant evidence of a difference concluded. There are several methods for carrying out a t-test in Excel and two will be covered here; using Excel commands and using the standard Excel 'add-in' the data analysis toolpak.

### Method 1: Using the commands

Excel has a general t-test command which calculates a p-value for the data.

`=T.TEST(group1,group2,tails,test)`

Group1 = the cells containing the group 1 values e.g. non-smokers

Group 2 = the cells containing the group 2 values e.g. smokers

Tails: Are you doing a one or a two tailed test? 1=one, 2=two-tailed. If you are not sure, put two as this is standard practice.

Test: there are three options

1 = paired t-test

2=Independent t-test assuming equal variances

3= Independent- t-test not assuming equal variances

For this example, two independent groups are being compared and the assumption of equal variances has been met so choose option 2. It is a two-tailed test so tails = 2 as well. Therefore, the command used is `=t.test(A:A,B:B,2,2)` which gives a p-value of 0.0465. As this is less than 0.05, there is significant evidence of a difference between the two groups. If there is a significant difference, it is common to report the difference between the means.

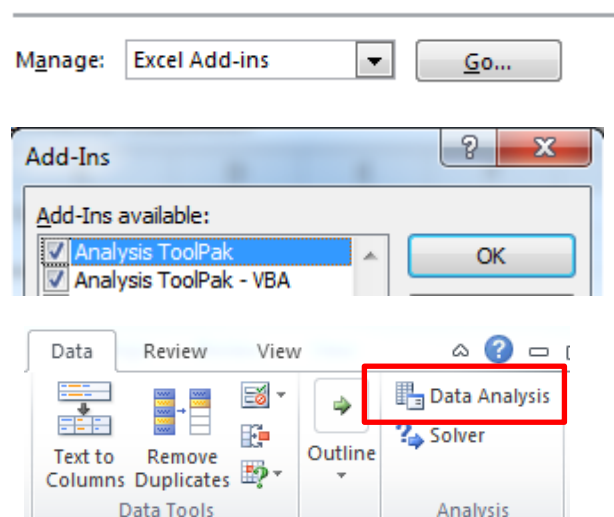
*There is significant evidence ( $p = 0.0465$ ) that there is a difference in the birthweight of babies whose mothers smoke and don't smoke. In this dataset, smokers have babies who weigh 0.81 lbs less on average.*

### Method 2: Using the data analysis toolpak

Excel has an add-in package which needs to be activated through the options. Go to **File > Options > Add-Ins** and select **Go**.

In the box that appears, select 'Analysis Toolpak' and then click **OK**.

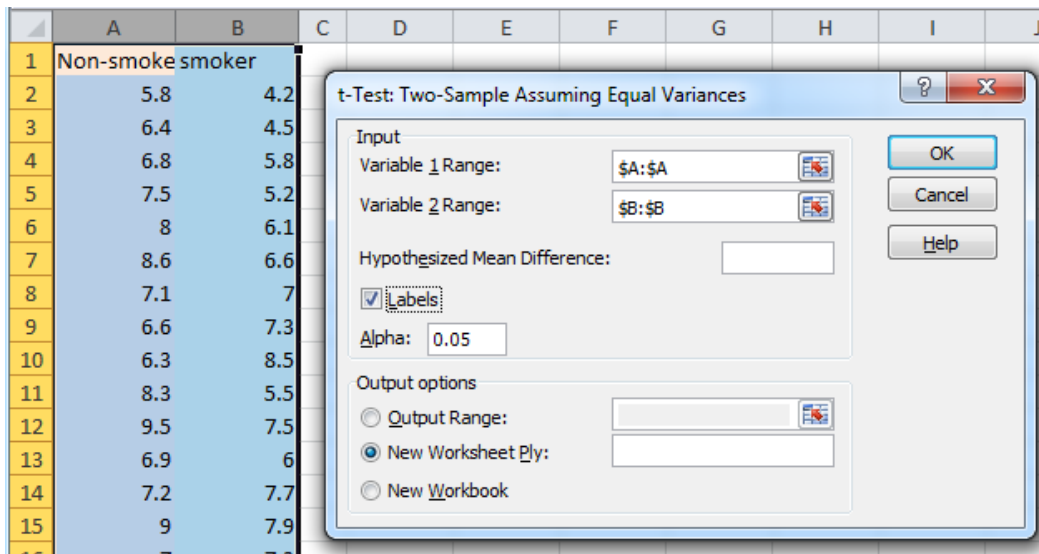
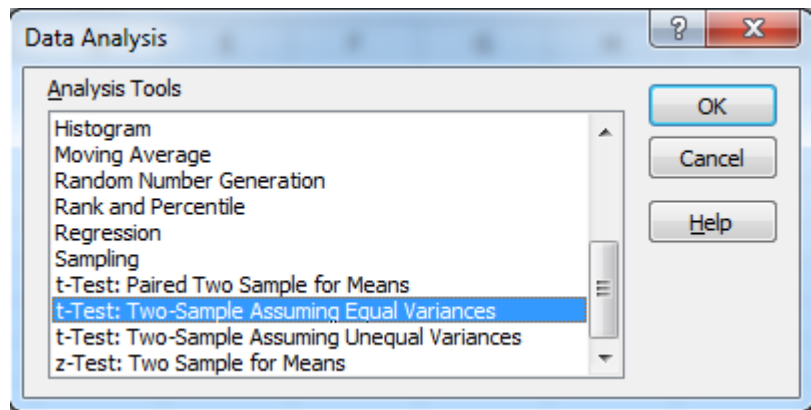
Once the data analysis toolpak is loaded, it will appear in the **Data** tab.



# Independent t-test in Excel

Click on the **Data Analysis** button to open the first menu and choose '*t-test: Two-Sample Assuming Equal Variances*' and then **OK**.

If one standard deviation is more than twice the other, select '*t-test: Two-Sample Assuming Unequal Variances*' instead.



Enter the column (range of data) for the first group in the '*Variable 1 range*' box and the second group in the '*Variable 2 Range*' box. Tick the '*Labels*' box if variable names are included in the selected range. The default '*Output*'

option is for the results to appear on a new sheet. Click **OK** to run the analysis.

Excel gives a large amount of output. As this is a two-tailed test, the main value of interest is the two-tailed p-value of 0.0465. As the p-value is less than 0.05, there is significant evidence of a difference between the two groups.

t-Test: Two-Sample Assuming Equal Variances		
	Non-smoker	smoker
Mean	7.69	6.877272727
Variance	1.317789474	1.930411255
Observations	20	22
Pooled Variance	1.639415909	
Hypothesized Mean	0	
df	40	
t Stat	2.054480171	
P(T<=t) one-tail	0.023251361	
t Critical one-tail	1.683851013	
P(T<=t) two-tail	0.046502722	
t Critical two-tail	2.02107539	

## Reporting t-tests

It is common practice to report the test statistic (2.05), the degrees of freedom (40) and the p-value (0.0465). If the result is significant, then report the difference between the means. Excel gives the output to many decimal places but 3 decimal places for the p-value is enough. Look to the raw data for guidance on how many decimals for the mean difference. Here, the numbers are small so a mean difference of 0.81 should be reported.

*There is evidence ( $t(40)=2.05$ ,  $p = 0.047$ ) that there is a difference in the birthweight of babies whose mothers smoke and don't smoke. In this dataset, smokers have babies who weigh 0.81 lbs less on average.*