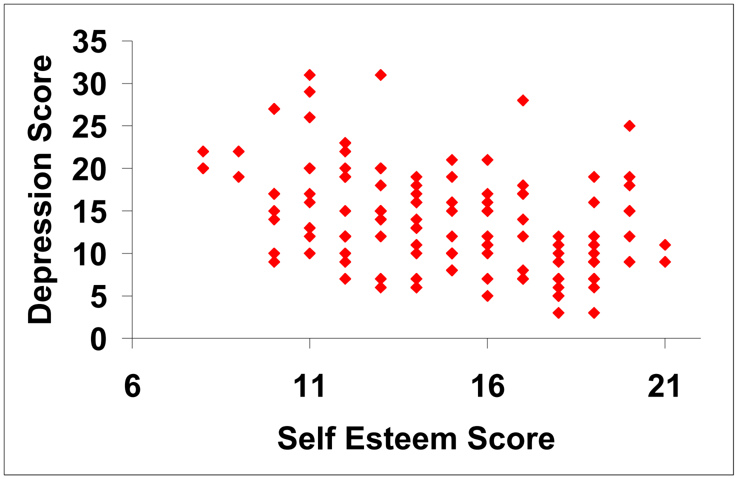
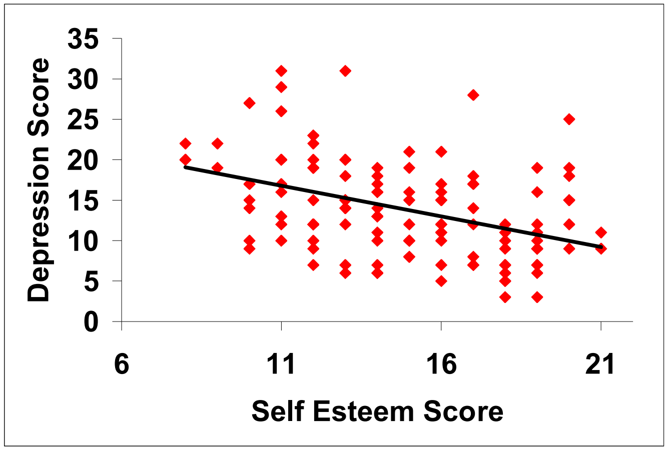
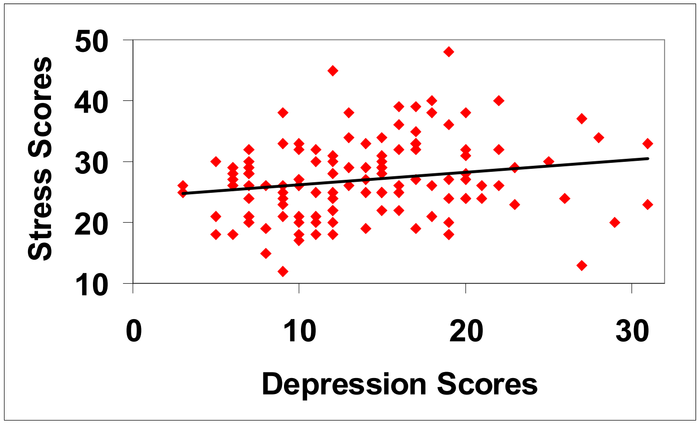
Sample Post-processed Data

Sex	Age	Ethnicity	MomEthnic	DadEthnic	SES	SelfEsteem	Stress
1	18	2	2	2	4	14.00	28.00
1	18	3	3	3	5	14.00	26.00
2	18	3	3	3	4	18.00	28.00
1	18	6	6	5	3	15.00	24.00
2	18	3	3	3	4	16.00	17.00
1	18	5	4	2	4	8.00	32.00
2	18	2	2	2	5	19.00	21.00
2	17	2	4	2	4	16.00	18.00
2	18	5	3	4	4	13.00	27.00
2	18	5	6	4	4	14.00	29.00
2	18	4	4	4	2	18.00	21.00
1	18	1	1	1	4	12.00	25.00
1	18	1	1	1	5	20.00	36.00

To easily do tests in Excel, you will need to copy portions of the data into new worksheets. Right-click on a current tab, select **Insert…**, select **Worksheet**, and then choose **OK**.




Depression Score
35
30
25
20
15
10
5
0
6
11
16
21
Self Esteem Score


Stress Scores
50
40
30
20
10
0
10
20
30
Depression Scores

SelfEsteem (SE)	Depression (D)	Stress (S)			
14.00	6.00	28.00	**cor(SE,D)**	**=pearson(a2:a54,d2:d54)**	**-0.413**
14.00	10.00	26.00	**cor(D,S)**	**=pearson(b2:b54,d2:d54)**	**0.193**
18.00	7.00	28.00			
15.00	21.00	24.00			

Correlation provides a measure of the strength of the linear association between two quantitative variables

Correlation is a unitless quantity that falls between -1 and +1
Perfect correlation is -1 or +1
Correlation close to zero ➔ no linear association

High correlation does not imply causation

Hypothesis Tests--Correlation

H_0: There is no relationship between self-esteem scores and depression scores.

versus

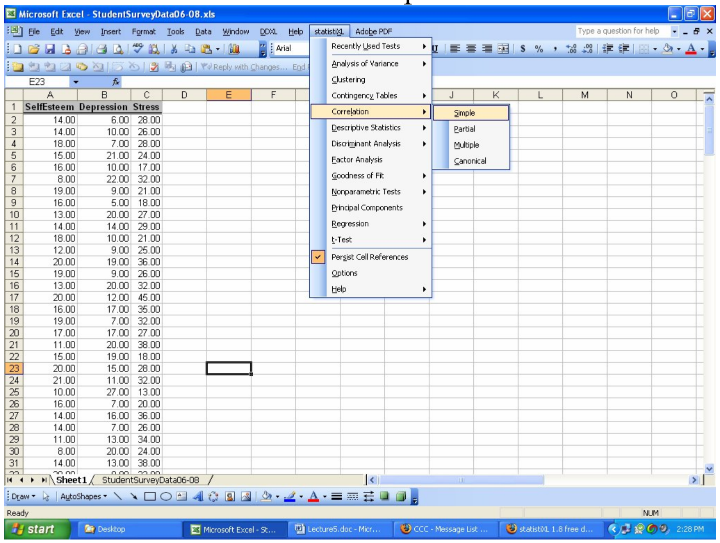
H_a: There is a relationship (linear) between self-esteem scores and depression scores.

$$H_0 : \rho_{SE,D} = 0 \text{ versus } H_a : \rho_{SE,D} \neq 0$$

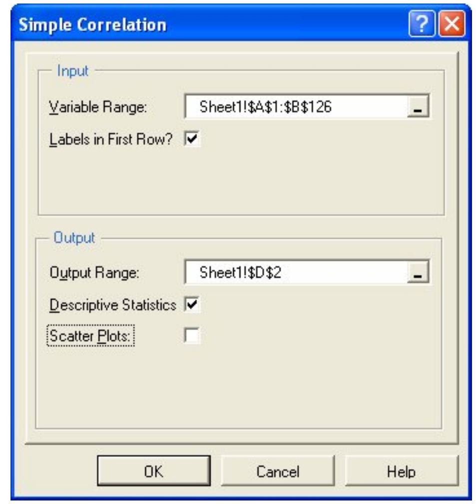
Go to
http://3d2f.com/download/10-791-statistixl-free-download.shtml
and download this Excel add-in. You can install with a 30 day free trial.

To do tests of correlation for the relationship between self-esteem/depression and depression/stress, copy the data for these variables onto a new worksheet.

Choose statistiXL > Correlation > Simple



This will bring up the following dialog:



Enter the variable range either by typing it in or by highlighting the data. If you included the data labels as part of the variable range, make sure to check the “**Labels in First Row?**” box. Select a cell for the upper left-hand corner of the output. Optionally, you can choose to display “**Descriptive Statistics**” and “**Scatter Plots:**” by checking the appropriate boxes. Click **OK** to get the following output:

Pearson Correlation Results for:
Variable Range = Sheet1!$A$1:$B$126

Descriptive Statistics

Variable	Mean	Std Dev.	Std Err	N
SelfEsteem	**14.696**	**3.278**	**0.293**	**125**
Depression	**13.992**	**6.034**	**0.540**	**125**

Correlation Matrix (R)

	SelfEsteem	Depression
SelfEsteem	1.000	**-0.413**
Depression	-0.413	1.000

t Statistic

	SelfEsteem	Depression
SelfEsteem	-	**5.025**
Depression	5.025	-

Correlation Significance (P)

	SelfEsteem	Depression
SelfEsteem	-	**0.000**
Depression	**0.000002**	-

The test for correlation gave $t = 5.025$ with 123 df.

The sample data provides sufficient evidence to reject the null hypothesis at $\alpha = 0.05$ level of significance ($P < 0.00005$).

There is a statistically significant relationship between self-esteem scores and depression scores. The sample data indicate that as self-esteem tends to increase, as depression decreases.

What if we had tested:

H_0: There is no relationship between self-esteem scores and depression scores.

versus

H_a: There is an inverse relationship between self esteem scores and depression scores; as self-esteem scores increase, depression scores tend to decrease.

$$H_0 : \rho_{SE,D} = 0 \text{ versus } H_a : \rho_{SE,D} < 0$$

The test would be done in an identical fashion, but we would need to adjust the reported P-value since Excel gives a 2-tailed value. If you are conducting a 1-tailed test, then you must divide the reported P-value in half.

For this example, however, since the P-value is so small, you can leave it at $P < 0.00005$.

H_0: There is no relationship between self-esteem scores and stress scores.

versus

H_a: There is a relationship between self-esteem scores and stress scores.

$H_0 : \rho_{D,S} = 0$ versus $H_a : \rho_{D,S} \neq 0$

Pearson Correlation Results for:
Variable Range = Sheet1!$B$1:$C$126

Descriptive Statistics

Variable	Mean	Std Dev.	Std Err	N
Depression	**13.992**	**6.034**	**0.540**	**125**
Stress	**27.024**	**6.421**	**0.574**	**125**

Correlation Matrix (R)

	Depression	Stress
Depression	1.000	**0.193**
Stress	0.193	1.000

t Statistic

	Depression	Stress
Depression	-	**2.181**
Stress	2.181	-

Correlation Significance (P)

	Depression	Stress
Depression	-	**0.0311**
Stress	0.0311	-

The test for correlation gave $t = 2.181$ with 123 df.

The sample data provides sufficient evidence to reject the null hypothesis at $\alpha = 0.05$ level of significance ($P = 0.0311$).

NOTE: If you had conducted a 1-tailed test, you would report $P = 0.0156$.

There is a statistically significant relationship between depression scores and stress scores. The sample data indicate that as depression increases, as stress increases.

Hypothesis Tests--Between two means for independent samples

Use this when you want to compare the mean values between two groups.

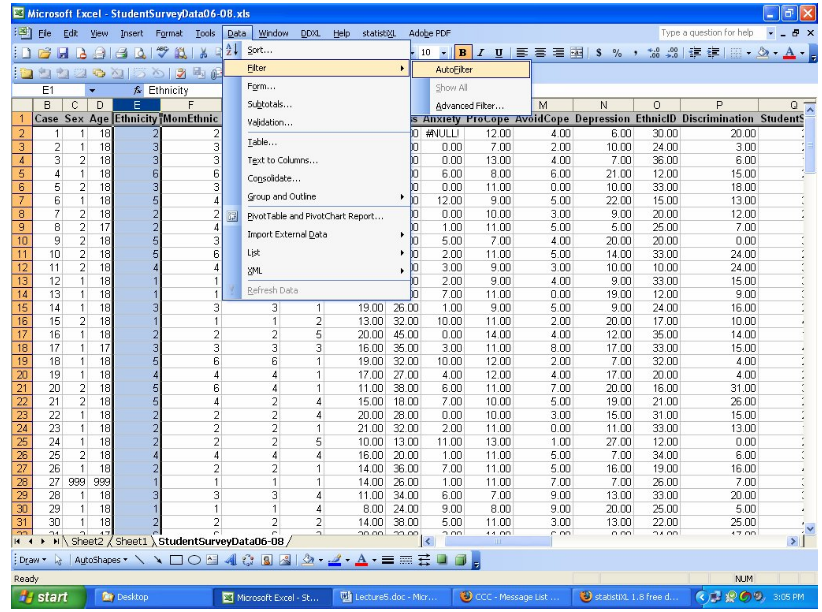
For the psychology data set this test would be used when comparing “scores” between genders, ages, races, ethnicity, or SES.

Let’s consider ethnicity and the stress score. Since ethnicity has six categories, you can’t directly apply the two-sample *t*-test. One option, however, is to break ethnicity into two categories by grouping responses.

For this example, consider ethniciy = 4 (White, Caucasian, Anglo, European American, not Hispanic… versus IDs 1, 2, 3, 5, 6 (everything else).

In a new worksheet (or use two new columns in an existing worksheet), put stress scores for ethnicity = 4 in one column and the stress scores for all other ethnicities in another column. Using “**Filters**” can help with this.

Highlight the Ethnicity column and then Select Data > Filters > AutoFilter as shown below:



This will give you a dropdown arrow in the Ethnicity column heading, which can be used to select data by the value of Ethnicity.

Click the dropdown arrow and select 4—this will give you all observations with ethnicity of 4. Now highlight all the Stress scores and use Copy > Paste special…. Select Values to copy the stress data into a new column.

Repeat the process for ethnicity values 1, 2, 3, 5, 6—but be sure to put these stress scores into a second column. You should now have two columns with stress scores—one with scores for ethnicity = 4 and one with stress scores for all other ethnicities.

When using the *t*-test, we are comparing the means of the two groups.

H_0: The mean racism score for Caucasian students is the same as the mean racism score for students from other ethnicities.

versus

H_a: The mean racism scores between Caucasian students and all others are different.

$$H_0 : \mu_C = \mu_O \text{ or } \mu_C - \mu_O = 0$$

versus

$$H_a : \mu_C \neq \mu_O \text{ or } \mu_C - \mu_O \neq 0$$

In Excel, choose **Tools** > **Data Analysis** > **t-Test: Two Samples Assuming Unequal (equal) Variances** (Note: you might have to load the Data Analysis toolpak).

t-Test: Two-Sample Assuming Unequal Variances

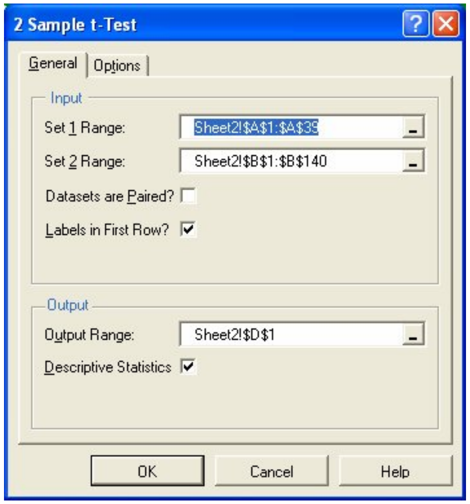
	Stress-4	*Stress-other*
Mean	25.684	27.410
Variance	44.654	48.678
Observations	38	139
Hypothesized Mean Difference	**0**	
df	**61**	
t Stat	**-1.3974**	
P(T<=t) one-tail	0.0837	
t Critical one-tail	1.6702	
P(T<=t) two-tail	**0.1674**	
t Critical two-tail	1.9996	

What do we conclude?

Fail to reject H_0. There is insufficient sample evidence to reject the null hypothesis and conclude there is a statistically significant difference between the mean stress scores. The mean stress score for Caucasian students is not different from the mean stress score for other students at the $\alpha = 0.05$ significance level (P = 0.1674).

You can also conduct this test using statistiXL. Choose statistiXL > t-Test > Two Sample.

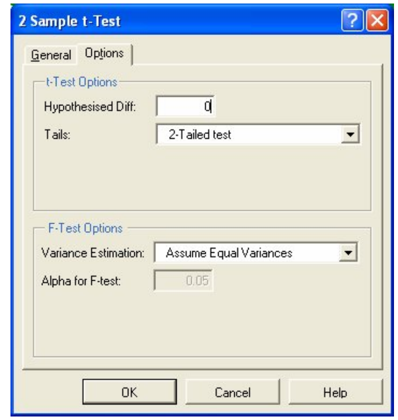
This gives the following dialog:



Enter the range for the stress data for ethnicity = 4 and the range for the stress data for ethnicity = other (not 4). If you include Labels in the data ranges, check the “Labels in First Row?” box. Select a starting point for the Output. Check “Descriptive Statistics” to get means, and confidence intervals.

Now choose the Options tab and select the options as indicated below:

Choose the 2-Tailed test or choose an appropriate 1-Tailed test. Then select OK.



t-Test Result for Datasets:
Set 1 Range = Sheet2!$A$1:$A$39
Set 2 Range = Sheet2!$B$1:$B$140
Unequal variances assumed

Descriptive Statistics

Variable	Mean	Std Dev.	Std Err	Lower 95% CL	Upper 95% CL	N
Stress-4	25.684	6.682	1.084	**23.488**	**27.881**	38
Stress-other	27.410	6.977	0.592	**26.240**	**28.580**	139

1-tailed t-Test (Stress-4 < Stress-other)

Ho. Diff	Mean Diff.	SE Diff.	T	DF	P
0.000	-1.726	1.235	-1.397	60.889	0.084