Sample Post-processed Data

SexA	lge E	Ethnicity MomE	thnic DadE	thnic SE	S	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**.

1











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  $\rightarrow$  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.

$$H_0: \rho_{SE,D} = 0$$
 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

⊠ N	licrosoft Ex	cel - Student	SurveyD	ata06-0	8.xls		_								JÞ
:0)	Ele Edit	View Insert	Format	Took	Data <u>W</u> indow	DOXI Help	st	tistig. Adobe PDF		-		Typ	be a question for h	40 • _ 6	9
10	💕 🖬 🔒	A 3 4	1 2 12	18 00	🚬 + 🛄	Arial		Recently Used Tests	•		S 9	% , <u>.0</u> .0	8 律律 图	- 🖄 - <u>A</u>	•
1	2 2 2	0015	813	时间	Po Reply with	changes Egd		Analysis of Variance	•						
_	E23	• fx		-				⊆lustering							
	A	В	C	D	E	F		Contingency Tables		JK		L M	N	0	
1	SelfEsteen	Depression	Stress					Correlation	•	Simple					
2	14.0	0 6.00	28.00					Description Chableties			_				
3	14.0	0 10.00	26.00					Describuye scarpors	'	Barbal					
4	18.0	3 7.00	28.00					Discriminant Analysis	•	Multiple					
5	15.0	21.00	24.00					Factor Analysis		Caponical					
6	16.0	0 10.00	17.00							Saucurea					
7	8.0	22.00	32.00					Goodness of Hit	'						
8	19.0	9.00	21.00					Nonparametric Tests	•						
9	16.0	5.00	18.00					Principal Components							
10	13.0	20.00	27.00						2						
11	14.0	3 14.00	29.00					Regression	,						
12	18.0	0 10.00	21.00					t-Test	•	· · · · · · · · · · · · · · · · · · ·					
13	12.0	9.00	25.00					Durative Call Disferences							
14	20.0	) 19.00	36.00				Ľ	Pergor Centrelence							
15	19.0	9.00	26.00					Options							
16	13.0	20.00	32.00					Help							
17	20.0	12.00	45.00							-					
18	16.0	0 17.00	35.00												
19	19.0	3 7.00	32.00												
20	17.0	17.00	27.00												
21	11.0	0 20.00	38.00												
22	15.0	0 19.00	18.00												
	20.0	0 15.00	28.00												
24	21.0	11.00	32.00												
25	10.0	27.00	13.00												
26	16.0	7.00	20.00												
27	14.0	16.00	36.00												
28	14.0	7.00	26.00												
29	11.0	13.00	34.00												
30	8.0	20.00	24.00												
31	14.0	3 13.00	38.00												
- +	► H\She	et1 / Stude	ntSurvey	Data06-0	в /			<						2	1
Dra	w - 🗟   Aut	oShapes * 🔪	10		4 0 3 3	101-2-	A	· = = = = •							
Read	y								-				N.	м	
4	start	Ca Desitop			Microsoft Exce	I-9 🗐		ure5.doc - Micr	D ccc		😺 statis			99, 2:281	

This will bring up the following dialog:

imple Correlation		? 🗙
Input Variable Range: Labels in First Row?	Sheet1!\$A\$1:\$B\$126	_
Output Output Range: Descriptive Statistics Scatter Plots:	Sheet1!\$D\$2 I⊽	_
ОК	Cancel	Help

7

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	Ν
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.

9

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.

# $H_0: \rho_{SE,D} = 0$ versus $H_a: \rho_{SE,D} \leq 0$

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	Ν
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:

<b>1</b>	🛿 Microsoft Excel - StudentSurveyData06-08.xls																	
:8)	Eile	Edit	View	Insert	Format Tools	Data	Window	DDXL	Help statis	tõ <u>s</u> L A	dobe P	PDF				Туре	a question for help	8×
in	12		a	AR	** 61 X E	21	Sort				10	- B	<i>I</i> Π ≡		5 %			- A -
	-		100				Filter			•		AutoFilte				.000	1	
		9 4	-02		DI 🛛 🖬 🕪	-	Form					Theme All		-				
_	E1		•	∱∗ E	thnicity							210H AL				-	-	
4	B	C	0	E	Page - Calcula		Sugcocais					Advance	d Filter	M	N	U	P	Q <mark>_</mark>
-	Case	Sex	Age	Eunicity	IMOMETINIC		Validation				s An	xiety i	ToCope AV	/oracope	Depression	Ethnicib	Discrimination S	tudents
2	1	1	18	4	2		Table				U #AN	IULLI	12.00	4.00	6.00	30.00	20.00	
5	4	2	18	-	3		- Tauk ka Calu				0	0.00	7.00	2.00	7.00	24.00	3.00	
4	-	2	10		3		Text to Colu	nns			0	0.00	13.00	4.00	7.00	42.00	5.00	
5	4	2	10	6	6		Consolidate.				0	6.00	8.00	0.00	21.00	12.00	15.00	
7	C (		10		3		Group and O	utline			0	10.00	0.00	5.00	10.00	35.00	12.00	
0	7	2	10			and				- (	0	0.00	10.00	2.00	22.00	20.00	13.00	
0	6	2	17	4	4	12	Prvot l'able al	AG PINOCO	.nart Report		n	1.00	11.00	5.00	5.00	20.00	7.00	
10	6	2	10	-	2		Import Exter	nal <u>D</u> ata		•	0	6.00	7.00	4.00	20.00	20.00	0.00	
11	10	2	18	-	6		List				n	2.00	11.00	5.00	14.00	33.00	24.00	
12	11	2	18	7	Å						n	3.00	9.00	3.00	10.00	10.00	24.00	
13	17	1	18	1	1		Shur			•	n	2.00	9.00	4.00	9.00	33.00	15.00	
14	12	1	18	1	1	Y	Refresh Data	3			n	7.00	11.00	0.00	19.00	12.00	9.00	$\rightarrow$
15	14	1	18		3	_	3	1	19.00	26.0	n	1.00	9.00	5.00	9.00	24.00	16.00	
16	16	2	18	1	1		1	2	13.00	32.0	n	10.00	11.00	2.00	20.00	17.00	10.00	
17	16	1	18		2		2	5	20.00	45.0	n	0.00	14.00	4.00	12.00	35.00	14.00	
18	17	1	17		3		3	3	16.00	35.0	n	3.00	11.00	8.00	17.00	33.00	15.00	
19	16	1	18	F	6		6	1	19.00	32.0	n	10.00	12.00	2.00	7.00	32.00	4.00	-
20	19	1	18	4	4		4	1	17.00	27.0	0	4.00	12.00	4.00	17.00	20.00	4.00	
21	20	2	18	5	6		4	1	11.00	38.0	0	6.00	11.00	7.00	20.00	16.00	31.00	
22	21	2	18	5	4		2	4	15.00	18.0	0	7.00	10.00	5.00	19.00	21.00	26.00	:
23	22	1	18	2	2		2	4	20.00	28.0	0	0.00	10.00	3.00	15.00	31.00	15.00	
24	23	1	18	2	2		2	1	21.00	32.0	0	2.00	11.00	0.00	11.00	33.00	13.00	
25	24	1	18	2	2		2	5	10.00	13.0	0	11.00	13.00	1.00	27.00	12.00	0.00	:
26	25	2	18	4	4		4	4	16.00	20.0	0	1.00	11.00	5.00	7.00	34.00	6.00	
27	26	1	18	2	2		2	1	14.00	36.0	0	7.00	11.00	5.00	16.00	19.00	16.00	
28	27	999	999	1	1		1	1	14.00	26.0	0	1.00	11.00	7.00	7.00	26.00	7.00	(
29	28	1	18	3	3		3	4	11.00	34.0	0	6.00	7.00	9.00	13.00	33.00	20.00	
30	29	1	18	1	1		1	4	8.00	24.0	0	9.00	8.00	9.00	20.00	25.00	5.00	
31	30	1	18	2	2		2	2	14.00	38.0	0	5.00	11.00	3.00	13.00	22.00	25.00	
11	F H	\ She	et2 /	Sheet1	StudentSurv	eyDa	ta06-08/	2	20.00	22.0	1	2.00	11.00	C 00	0.00	24.00	47.00	> i
Dra	w = D	Au	oShap	es • 🔪		-AI 1	3 🔒 🛃	₫.	<u>2 • A</u> • :	-	Ħ		-					
Read	ły												_				NUM	
1	sta	rt 🛛	6	Desktop		<b>X</b> M	crosoft Excel	- st	🔮 Lecture	5.doc - 1	Micr	6	CCC - Messag	ge List	🕲 statistiXL (	1.8 free d	C # 2 @ 9	3:05 PM

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-
	Stress-4	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:

2 Sample t-Test		? 🔀							
General Options									
- Input									
Set <u>1</u> Range:	Sheet2!\$A\$1:\$A\$39								
Set <u>2</u> Range:	Sheet2!\$B\$1:\$B\$140	-							
Datasets are Paired?	Г								
Labels in First Row?	Labels in First Bow?								
Output									
Output Range:	Sheet2l\$D\$1	_							
Descriptive Statistics	<b>v</b>								
ОК	Cancel H	lelp							

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.

2 Sample t-Test		? 🛛
General Options		1
Hypothesised Diff:	a l	
Tails:	2-Tailed test	•
F-Test Options Variance Estimation: Alpha for F-test:	Assume Equal Variances	•
OK	Cancel	Help

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% CLUpper	r 95% CL	Ν
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.	Т	DF	Р
0.000	-1.726	1.235	-1.397	60.889	0.084



