

Correlation and Regression Lab

An educational researcher wanted to predict student's anxiety levels using information about teacher rating, student age, number of math classes completed, and gender.

Statistics Anxiety (1 = low, 10 = high)	Teacher Rating (1 = awful, 10 = great)	Student Age	Math Classes	Gender (0 = male, 1 = female)
3	8	21	2	0
5	5	19	1	0
2	9	22	4	1
6	5	22	1	1
7	1	23	2	1
1	6	18	3	0
3	3	19	2	1
5	3	19	1	1

Calculate the correlation between statistics anxiety and teacher rating by hand.

Calculate the linear regression model for predicting statistics anxiety using teacher rating.

Calculate the standard error of the estimate and R^2 for the regression model by hand.

Test whether the correlation is significant using an alpha of .05 by hand.

Correlation By Hand

Anx.	(Anx – Mean)	(Anx – Mean) ²	Teach	(Teach – Mean)	(Teach – Mean) ²
3	3 – 4 = -1	1	8	8 – 5 = 3	9
5	5 – 4 = 1	1	5	5 – 5 = 0	0
2	2 – 4 = -2	4	9	9 – 5 = 4	16
6	6 – 4 = 2	4	5	5 – 5 = 0	0
7	7 – 4 = 3	9	1	1 – 5 = -4	16
1	1 – 4 = -3	9	6	6 – 5 = 1	1
3	3 – 4 = -1	1	3	3 – 5 = -2	4
5	5 – 4 = 1	1	3	3 – 5 = -2	4
Mean 4	0 always	Total Variability 30	Mean 5	0 always	Total Variability 50

Statistics Anxiety	Teacher rating
Mean =	Mean =
Variance =	Variance =
SD =	SD =

(Anx – Mean)	(Teach – Mean)	(Anx – Mean)(Teach – Mean)
3 – 4 = -1	8 – 5 = 3	
5 – 4 = 1	5 – 5 = 0	
2 – 4 = -2	9 – 5 = 4	
6 – 4 = 2	5 – 5 = 0	
7 – 4 = 3	1 – 5 = -4	
1 – 4 = -3	6 – 5 = 1	
3 – 4 = -1	3 – 5 = -2	
5 – 4 = 1	3 – 5 = -2	
0 always	0 always	Sum of Products

Correlation

$$r = \frac{\text{Sum of Products}}{\sqrt{\text{Total Variability}_x \text{Total Variability}_y}}$$

Linear Regression by Hand

Y = *dependent variable* =

X = *predictor* =

Slope: $b = r \frac{SD_Y}{SD_X}$

Intercept: $a = \bar{Y} - b(\bar{X})$

Linear regression model

Using the linear regression model to predict scores for others

R^2

$$\text{Standard Error of the Estimate} = \sqrt{\frac{(1 - R^2)\text{Total Variability}_Y}{n - 2}}$$

Significance Testing for Correlation and Linear Regression

We could calculate a t test to test the significance of the predictor.

We could calculate an ANOVA F test to evaluate the entire regression model.

We could compare the r to critical value(s) to test the correlation.

When there is only one predictor all three tests are equivalent.

The correlation, the predictor, and the entire regression model are significant.

OR

The correlation, the predictor, and the entire regression model are nonsignificant.

Therefore, we will use the easiest method of comparing r to critical values.

Null Hypothesis:

Alternative Hypothesis:

Alpha level =

Degrees of freedom = $n - 2$

Critical Values =

Determine the significance by comparing the calculated r to the critical values.

Write a sentence to interpret the results. For example,

Use SPSS to predict statistics anxiety using teacher rating.

Interpret the intercept.

Interpret the slope.

Interpret the standard error of the estimate.

Interpret the R^2 .

Is the predictor significant?

Untitled - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

19 :

	statsanxiety	teacherrating	gender
1	3.00	8	
2	5.00	5	
3	2.00	9	
4	6.00	5	
5	7.00	1	1.00 Female
6	1.00	6	2.00 Female
7	3.00	3	3.00 Male
8	5.00	3	2.00 Female
			1.00 Female

Analyze > Correlate > Bivariate...

Bivariate Correlations

Variables:

- statsanxiety
- teacherrating
- studentage
- mathclasses
- gender

Correlation Coefficients

Pearson Kendall's tau-b Spearman

Test of Significance

Two-tailed One-tailed

Flag significant correlations

Options...

Bivariate Correlations: Options

Statistics

Means and standard deviations

Cross-product deviations and covariances

Missing Values

Exclude cases pairwise

Exclude cases listwise

Continue

Cancel

Help

Correlations

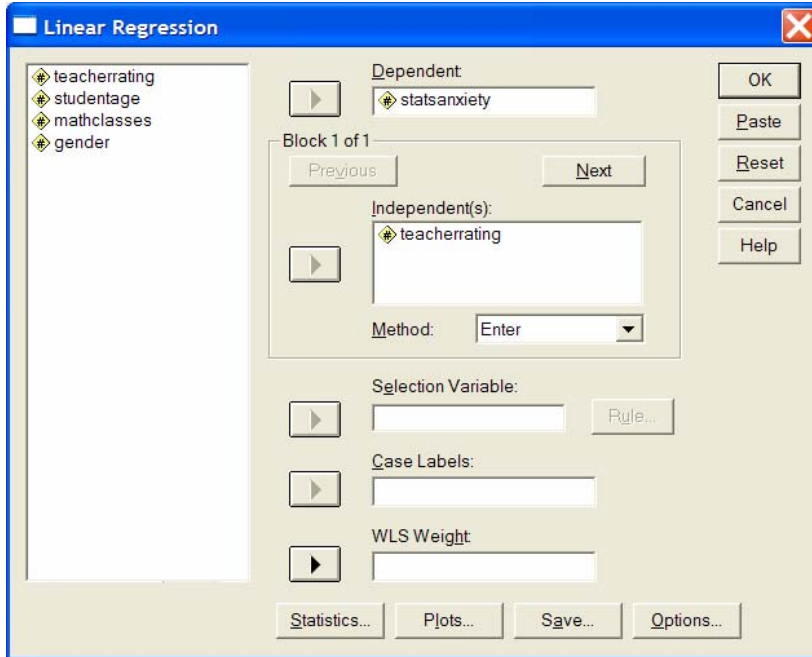
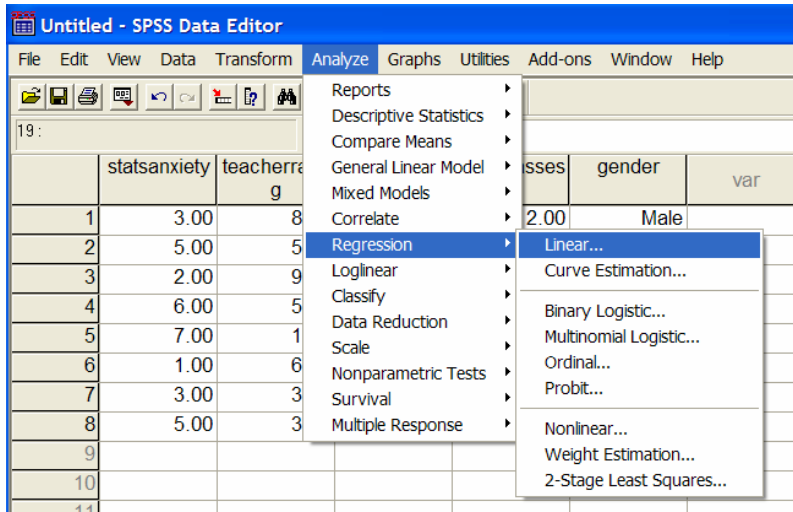
Descriptive Statistics

	Mean	Std. Deviation	N
statsanxiety	4.0000	2.07020	8
teacherrating	5.0000	2.67261	8
studentage	20.3750	1.84681	8
mathclasses	2.0000	1.06904	8
gender	.6250	.51755	8

Correlations

		statsanxiety	teacherrating	studentage	mathclasses	gender
statsanxiety	Pearson Correlation	1	-.671	.486	-.710*	.400
	Sig. (2-tailed)	.	.068	.222	.048	.326
	N	8	8	8	8	8
teacherrating	Pearson Correlation	-.671	1	.029	.550	-.413
	Sig. (2-tailed)	.068	.	.946	.158	.309
	N	8	8	8	8	8
studentage	Pearson Correlation	.486	.029	1	.145	.467
	Sig. (2-tailed)	.222	.946	.	.732	.243
	N	8	8	8	8	8
mathclasses	Pearson Correlation	-.710*	.550	.145	1	.000
	Sig. (2-tailed)	.048	.158	.732	.	1.000
	N	8	8	8	8	8
gender	Pearson Correlation	.400	-.413	.467	.000	1
	Sig. (2-tailed)	.326	.309	.243	1.000	.
	N	8	8	8	8	8

*. Correlation is significant at the 0.05 level (2-tailed).



Regression

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	teacherrating ^a	.	Enter

a. All requested variables entered.

b. Dependent Variable: statsanxiety

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.671 ^a	.451	.359	1.65731

a. Predictors: (Constant), teacherrating

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13.520	1	13.520	4.922	.068 ^a
	Residual	16.480	6	2.747		
	Total	30.000	7			

a. Predictors: (Constant), teacherrating

b. Dependent Variable: statsanxiety

Coefficients^a

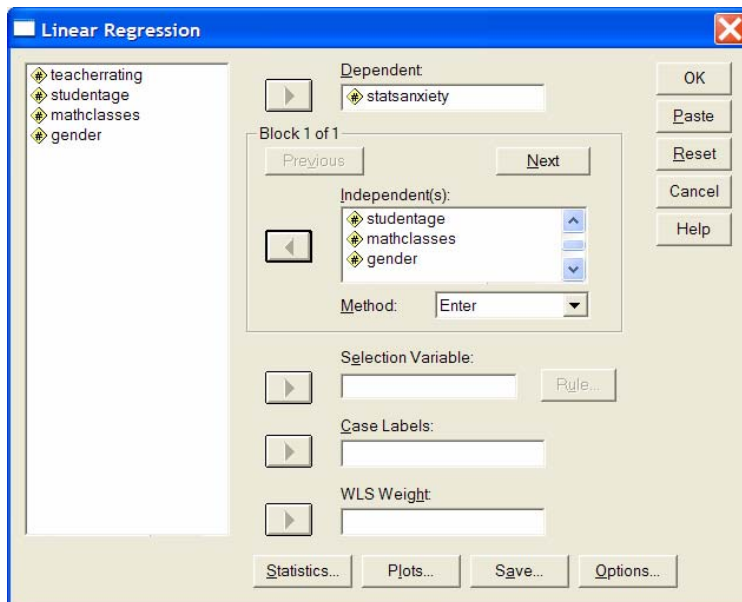
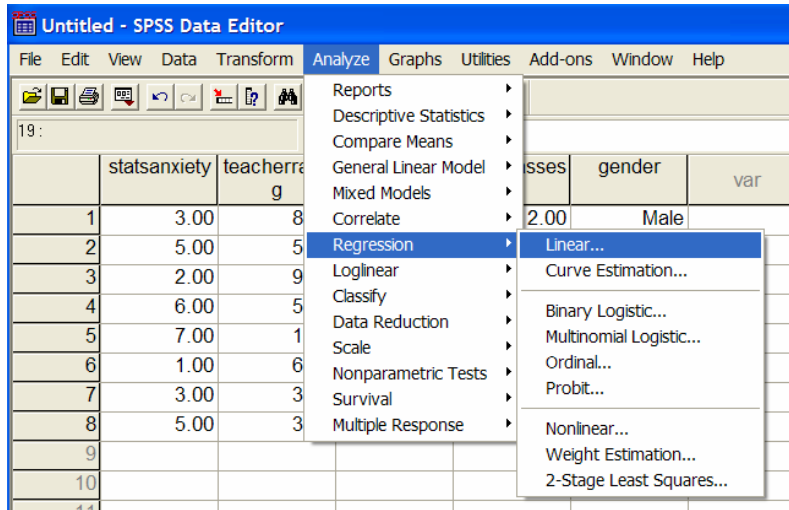
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.600	1.310		5.037	.002
	teacherrating	-.520	.234	-.671	-2.219	.068

a. Dependent Variable: statsanxiety

Use SPSS to predict statistics anxiety using teacher rating, student age, math classes, and gender.

Is the entire regression model significant?

Which of the predictors are significant?



Regression

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	gender, mathclasses, studentage, teacherrating ^a	.	Enter

a. All requested variables entered.

b. Dependent Variable: statsanxiety

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.974 ^a	.949	.881	.71497

a. Predictors: (Constant), gender, mathclasses, studentage, teacherrating

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	28.466	4	7.117	13.922	.028 ^a
	Residual	1.534	3	.511		
	Total	30.000	7			

a. Predictors: (Constant), gender, mathclasses, studentage, teacherrating

b. Dependent Variable: statsanxiety

Coefficients^a

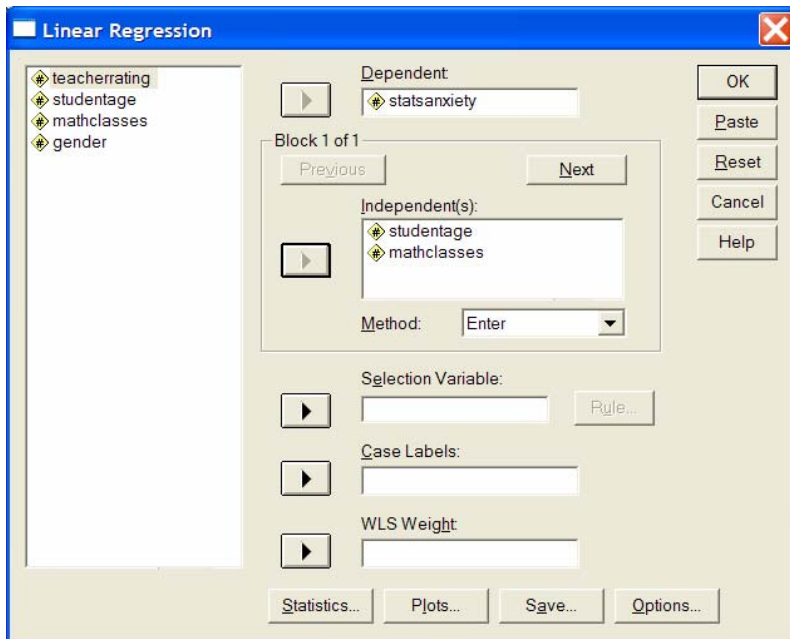
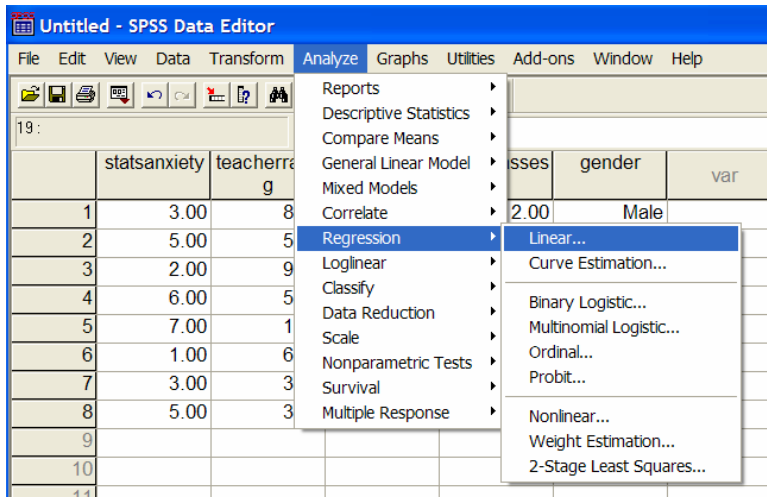
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-5.841	3.233		-1.807	.169
	teacherrating	-.295	.143	-.381	-2.065	.131
	studentage	.672	.172	.599	3.902	.030
	mathclasses	-1.137	.317	-.587	-3.585	.037
	gender	-.149	.698	-.037	-.214	.844

a. Dependent Variable: statsanxiety

Remove the nonsignificant predictors and re-estimate the regression model.

Is the revised regression model significant?

Which of the predictors are significant for the revised model?



Regression

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	mathclasses, studentage ^a	.	Enter

a. All requested variables entered.

b. Dependent Variable: statsanxiety

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.926 ^a	.858	.801	.92333

a. Predictors: (Constant), mathclasses, studentage

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	25.737	2	12.869	15.094	.008 ^a
	Residual	4.263	5	.853		
	Total	30.000	7			

a. Predictors: (Constant), mathclasses, studentage

b. Dependent Variable: statsanxiety

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-6.642	3.865		-1.718	.146
	studentage	.674	.191	.601	3.528	.017
	mathclasses	-1.543	.330	-.797	-4.678	.005

a. Dependent Variable: statsanxiety