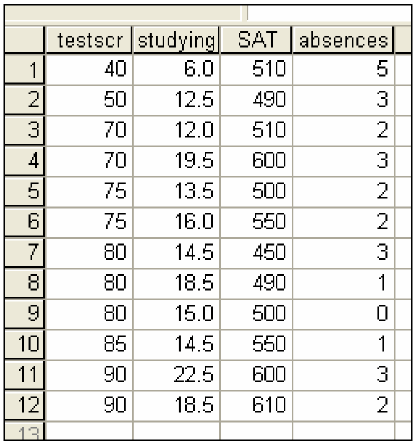
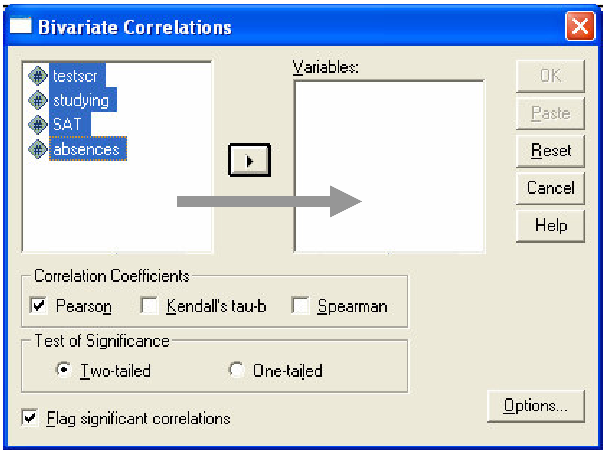
SPSS Guide: Correlation & Regression



	testscr	studying	SAT	absences
1	40	6.0	510	5
2	50	12.5	490	3
3	70	12.0	510	2
4	70	19.5	600	3
5	75	13.5	500	2
6	75	16.0	550	2
7	80	14.5	450	3
8	80	18.5	490	1
9	80	15.0	500	0
10	85	14.5	550	1
11	90	22.5	600	3
12	90	18.5	610	2
13				



Correlation Coefficients

☑ Pearson ☐ Kendall's tau-b ☐ Spearman

Test of Significance

● Two-tailed ○ One-tailed

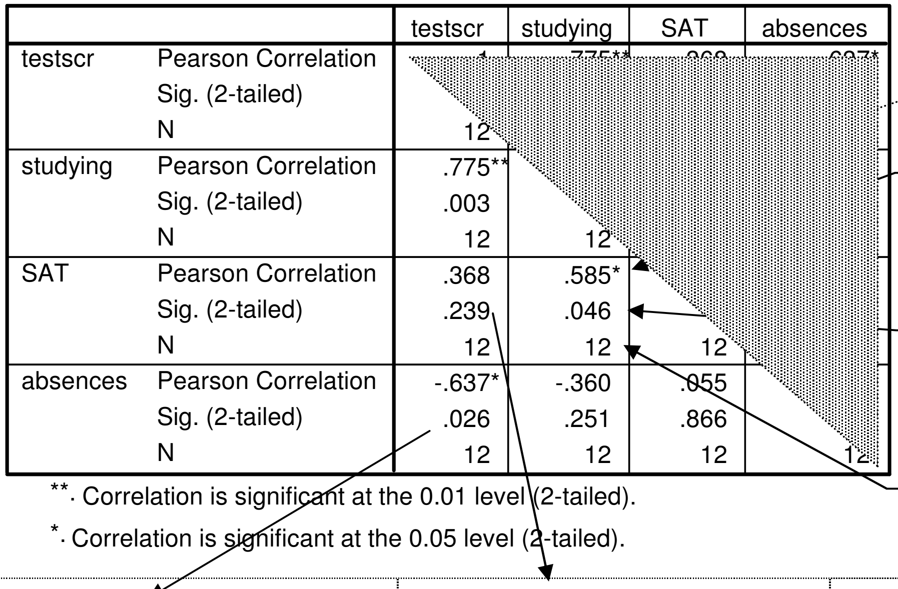
☑ Flag significant correlations

Once the data are entered, go to *Analyze, Correlation, Bivariate* to get this dialogue box.

Move the variables (quantitative only) that you wish to correlate into the variables box and hit *OK*.

This is a correlation matrix. It gives results for ***six*** correlations.

Correlations



		testscr	studying	SAT	absences
testscr	Pearson Correlation				
	Sig. (2-tailed)				
	N	12			
studying	Pearson Correlation	.775**			
	Sig. (2-tailed)	.003			
	N	12	12		
SAT	Pearson Correlation	.368	.585*		
	Sig. (2-tailed)	.239	.046		
	N	12	12	12	
absences	Pearson Correlation	-.637*	-.360	.055	
	Sig. (2-tailed)	.026	.251	.866	
	N	12	12	12	12

**. Correlation is significant at the 0.01 level (2-tailed).

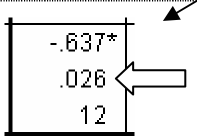
*. Correlation is significant at the 0.05 level (2-tailed).

You can ignore info above the diagonal. It's redundant.

The r-value. Indicates strength and direction (±) of the correlation. Bigger is better. The "*" means we can reject the null hypothesis (Ho).

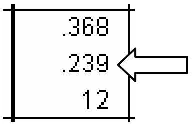
The p-value. Probability that you'd see an r-value of this size just by chance. Smaller is better. Reject Ho if p≤.05 [e.g., .046 is ≤ .05, so Reject.]

Number of pairs in sample. Degrees of freedom (df) equals n-2.



In this case, the p-value is below the magic .05 so we **REJECT** the Ho. [We think absences really do correlate negatively with test score].

We're HAPPY!! ☺



In this case, the p-value is NOT below the magic .05 so we **RETAIN** the Ho. [We are NOT confident that there is a correlation between SAT and test score].

We're SAD!! ☹

Statistical Hypotheses

Every r value (a sample statistic) strives to represent ρ (The actual correlation value in the population). When r gets bigger, we get more confident that there really is a correlation. We know one of two things must be true.

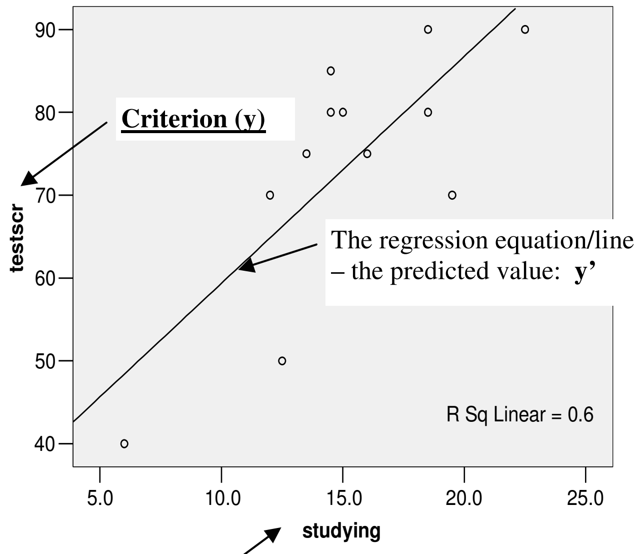
H_0: $\rho = 0$ *[There is NO actual correlation]*

H_A: $\rho \neq 0$ *[This is a correlation]*

KEY POINT: If p (the middle number) drops below .05, we REJECT the Ho. This makes us happy. We want to reject the null hypothesis because it means we have evidence that we found a true relationship.

We explain a finding as follows: The [research] hypothesis was supported. Absences correlate significantly with Test Score, r (10) = -.637, p ≤ .05. *Note: More on this later.* ***Degrees of freedom (df) = n-2.***

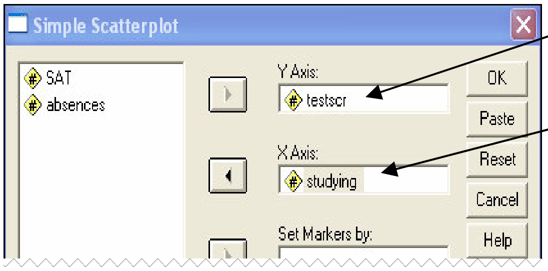
Scatterplot & Regression (using the same data)



Criterion (y)

The regression equation/line – the predicted value: **y'**

Predictor (x)



Y Axis: testscr

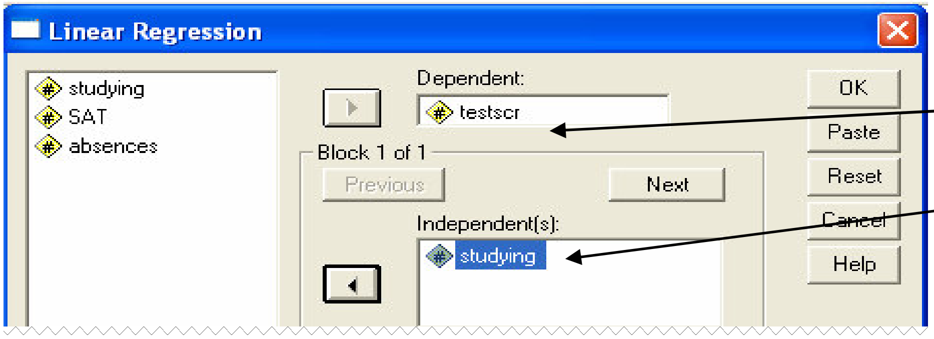
X Axis: studying

Criterion (y)

Predictor (x)

Scatterplot: Once the data are entered, go to *Graphs, Scatter, [leave on Simple]* to get to this box. Put the **criterion** (the variable you will predict) on the y-axis and the **predictor** (the variable to predict with) on the x-axis. To add a regression line…

- Double click the graph to open the Chart Editor
- Go to *Elements, Fit Line at Total* [change nothing]
- Close the Fit Line box & close the Chart Editor



Dependent: testscr

Independent(s): studying

TO DO REGRESSION: go to *Analyze, Regression, Linear* to get to this box.

Criterion (y): *What you predict.*

Predictor (x): What you predict with, what you already know.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.775[a]	.600	.560	10.012

a. Predictors: (Constant), studying

r^2

Coefficients[a]

p

a

b

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	31.960	11.171		2.861	.017
	studying	2.740	.708	.775	3.873	.003

a. Dependent Variable: testscr

You will get four tables, but you need only these two.

Model Summary: gives you the r-value, the r^2 value.

Coefficients: gives you the a & b values, and the p-value to check for significance. We reject Ho if p≤ .05. This means the relationship is reliable and can be used to make predictions. [Note: It's the same p value you see on the correlation matrix for these two variables.]

- In this case, our regression equation [$y' = bx + a$] becomes……… **$y' = 2.740(x) + 31.960$.**
- We can now predict **test score** (y) given any value of **studying** (x).