

Correlation and Regression Example in SPSS

Dean Wormer of Madeup University wants to limit the number of TV channels available for on-campus cable TV. He believes that students who watch more hours of TV tend to have lower GPAs. So, he hopes that limiting TV access will cause their GPAs to increase.

Is Dean Wormer correct that students who watch more TV tend to have lower GPAs? Does knowing how much TV a student watches help us to *predict* their GPA?

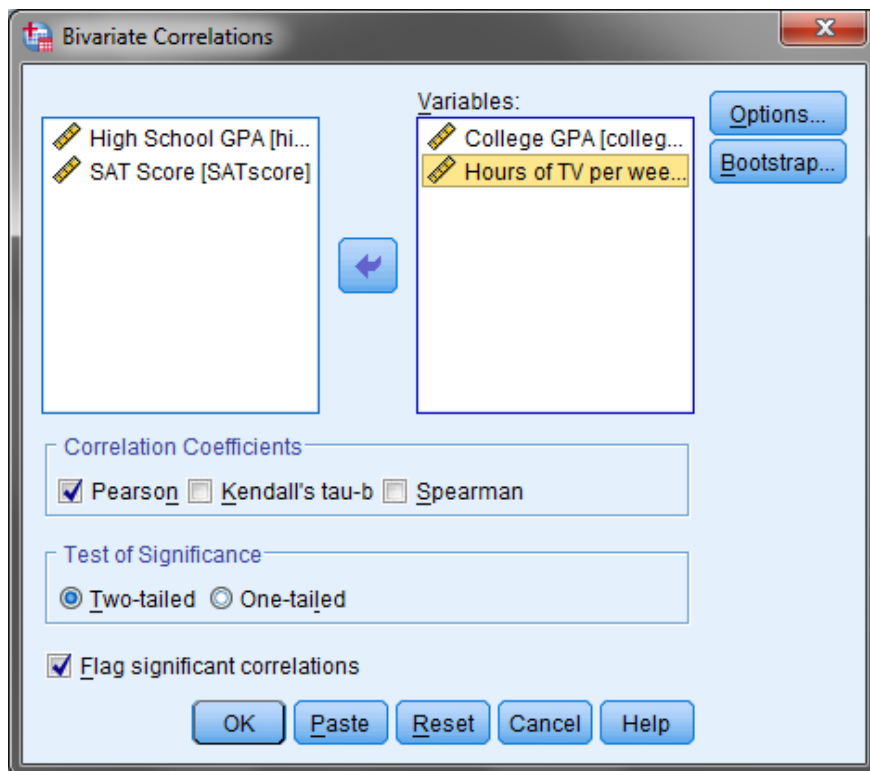
The datafile **CollegeGPA.sav** contains the data for 20 Madeup students.

Step 1: What test should we use to test this hypothesis?

Because we are looking at a relationship between TWO *continuous* variables (i.e., hours of TV and GPA), we can use a Pearson correlation and a simple regression.

Step 2a: How do we run this test in SPSS?

- First, to run the **correlation**, go to “Analyze” → “Correlate” → “Bivariate”



- The variables that we want to correlate with each other go in the “Variables” box.

Note: If we put more than two variables into the “Variables” box, SPSS will correlate every variable in the box with every other variable in the box. This will produce a much larger table than the one below, containing many Pearson r values instead of just one. This can save us a lot of time when we are working with many variables.

Step 3a: How do we read the output from SPSS?

Correlations			
		College GPA	Hours of TV per week
College GPA	Pearson Correlation	1	-.451*
	Sig. (2-tailed)		.046
	N	20	20
Hours of TV per week	Pearson Correlation	-.451*	1
	Sig. (2-tailed)	.046	
	N	20	20

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

1: This is our Pearson correlation ($r = -0.45$, which is a moderate to strong correlation).

2: Here is our p -value ($p = .046$). Because it is less than .05, we know that this correlation is significant. As hours of TV watching increases, GPA tends to decrease.

3: Annoyingly, SPSS doesn’t give us the df value; instead, it gives us N , and makes us calculate df ourselves. Fortunately this is easy: $df = N - 2$. So in this case, $df = 18$.

Step 4a: How do we write up the results in APA format?

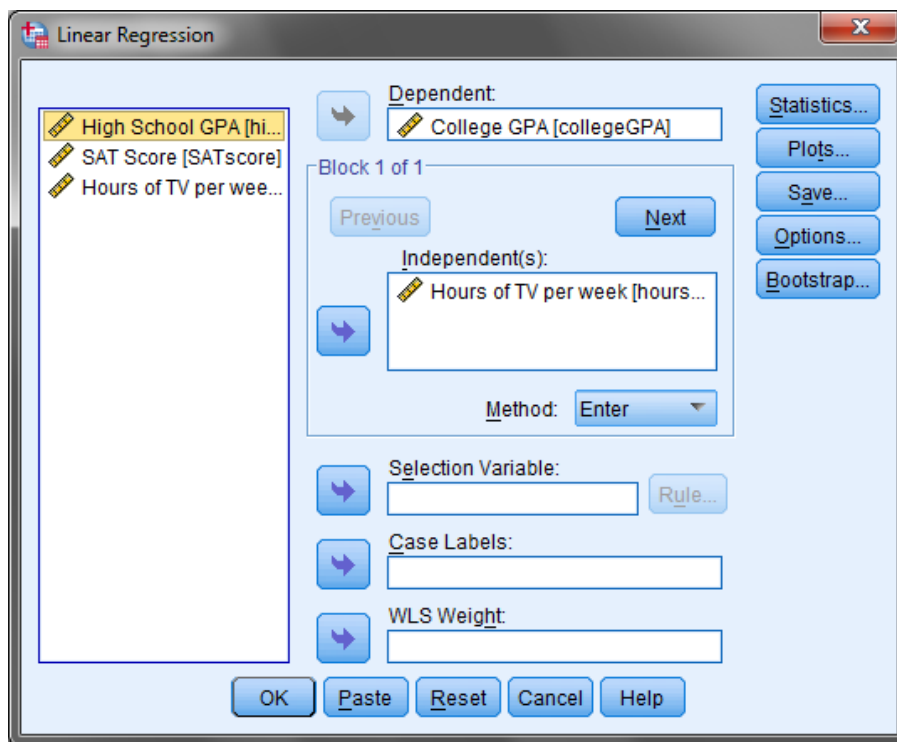
We hypothesized that students who watch more TV would tend to have lower GPAs. A Pearson correlation showed a significant relationship between TV watching and GPA, $r(18) = -0.45$, $p = .046$. Supporting our hypothesis, this correlation was negative and moderately strong, indicating that as students watched more hours of TV per week, their GPAs tended to decrease in a linear fashion.

A few things to notice about our write-up...

- 1) We always begin our write-up by explicitly stating our hypothesis.
- 2) Then, we name which test we did, and explain what the test showed. Did we, or didn't we, find a significant correlation? This is where we will provide our test statistic (in this case, our r value), degrees of freedom ($N - 2$), and p -value.
- 3) Finally, we note whether the correlation was positive or negative, describe its strength, and explain in plain language what the correlation is (or isn't) telling us.

Step 2b: How do we run this test in SPSS?

- Next, to run a **simple regression**, go to “Analyze” → “Regression” → “Linear”



- Our *predictor* variable (which acts as our IV) goes in the “Independent(s)” box. (If we had multiple predictor variables – i.e., a “multiple regression” – they would all go here.)
- Our *outcome* variable (which acts as our DV) goes in the “Dependent” box.

Step 3b: How do we read the output from SPSS?

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.451 ^a	.204	.160	.42280

a. Predictors: (Constant), Hours of TV per week

4: R^2 is a measure of “model fit.” It tells us how much of the variability in our outcome variable is explained by our predictor(s). So, in this example, 20.4% of the variability in students’ GPAs is explained by how many hours of TV they watch.

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.823	1	.823	4.606	.046 ^b
	Residual	3.218	18	.179		
	Total	4.041	19			

a. Dependent Variable: College GPA

b. Predictors: (Constant), Hours of TV per week

^ This table tests whether or not our model (i.e., all of our predictors) helps us to predict our outcome variable. When we only have one predictor, this table isn’t very important.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5 3.505	.254		13.812	.000
	Hours of TV per week	6 -.076	.035	7 -.451	8 -2.146	9 .046

a. Dependent Variable: College GPA

5: This is our *intercept* ($\alpha = 3.51$), i.e., the predicted value of y when x is zero. So, if a student watches zero hours of TV per week, our model predicts their GPA to be 3.51.

6: This is our *unstandardized* beta (i.e., slope; $\beta = -0.08$). For every additional hour of TV that a student watches, our model would predict their GPA to decrease by .08 points.

^ Using these two numbers, we can predict any student's GPA if we know how much TV they watch per week, by using the regression equation $\hat{y} = \alpha + \beta x (+ e)$. However, these predictions will not be perfect (hence the 'e' error term in the equation).

7: This is our *standardized* beta. If we had multiple predictors, these would allow us to compare them and see which ones are most useful to our prediction.

8: This is the *t* value for our predictor variable, which we will need for our write-up. (If we had multiple predictors, each would have its own *t* value.)

9: Here is the *p*-value for our predictor variable, which tells us whether it is a significant predictor of our outcome variable. In this case, because *p* is less than .05 (*p* = .046),

Step 4b: How do we write up the results in APA format?

We hypothesized that the number of hours of TV that a student watches per week would predict their GPA. A simple linear regression found that TV watching was a significant predictor of GPA, $t(18) = -2.15, p = .046$. Specifically, the model predicted that a student who watches zero hours of TV would have a GPA of 3.51, and that for each additional hour of TV that a student watches per week, their GPA will decrease by 0.08.

A few things to notice about our write-up...

- 1)** We always begin our write-up by explicitly stating our hypothesis.
- 2)** Then, we name which test we did, and explain what the test showed. Which of our predictor variables were (and/or were not) significant? This is where we will provide beta(s), and *t* value(s), *df* (*N* – 2), and *p*-value(s) for our predictor(s).
- 3)** It is often useful to then also mention (and interpret) unstandardized beta.
- 4)** Notice that all non-Greek letters are italicized, every equals sign has a space both before and after it, and every number is rounded to two decimal places (except for *p* values, which are rounded to three decimal places).

Epilogue: Dean Wormer is a chump who doesn't understand that correlation doesn't imply causation. Even though TV watching and GPA are negatively correlated, it is unlikely that taking away cable channels on campus would have any effect on students' GPAs.