

Correlation and Bivariate Regression Program Transcript

MATT JONES: This week, we're performing a Pearson Correlation Test. To do this, we can go to SPSS to perform this rather simple procedure. Like many of our tests, go ahead and activate the Analyze button to get the drop down menu. Because we're performing a correlation, we can move down to Correlate and across to Bivariate. The Pearson Correlation Test is a bivariate test.

If you click on that, you'll see a box come up, Bivariate Correlations. Let's go ahead and perform a bivariate correlation for respondent's socioeconomic status index and the respondent's highest level of education.

Now, it's important to remember, in this GSS data set, that respondent's highest level of education is measured in two different ways, one, as a categorical variable, and one as an interval ratio level variable. The categorical variable is the respondent's highest degree obtained. The respondent's highest level of education is measured in number of years of education.

We want to use respondent's highest level of education as measured in years, the interval ratio level variable, because a Pearson correlation test is easier to understand when we use two metric level variables. We're going to want to use the respondent's highest level of education as measured in number of years. That is the interval ratio level measurement for this test.

So again, I see my variable listings off to the left. And I can scroll down to find the appropriate variables that I want to test for a possible correlation.

Here, I can see the highest year of school completed. I place my cursor over it. It's highlighted. Again, I know this is the interval ratio level of measurement because I can see the scale ruler next to it. I highlight that. Move it over.

If I scroll down to find socioeconomic status index, again, placing my cursor over it, activating it, and moving it over, you'll see that SPSS automatically, by default, clicks on this Pearson correlation coefficient. Note that there are two other correlation coefficients that we will talk about later in the class.

The output for the Pearson correlation coefficient is rather simplistic. Since it's a bivariate test, you'll see the bivariate combinations here. We can see that there is a correlation coefficient of 0.610 between the highest year of school completed and the respondent's socioeconomic index.

If we move below, we can see the test of significance and see that the p value for this test is 0.000, which is well below the conventional 0.05 threshold. Therefore, we can reject the null hypothesis that there is no relationship between the respondent's highest year of school completed and their socioeconomic index.

Looking at the Pearson correlation coefficient, we know that this is a positive relationship and that the relationship is somewhat moderate.

Again, remember that a Pearson correlation coefficient is a standardized index that has a range of values from negative 1.0 to positive 1.0 with a 0 indicating no relationship whatsoever. The closer you move to 1.0 on either side, the stronger the relationship becomes.

You can see, by default, SPSS flags significant correlations. If we move down to the bottom here, we can see that this correlation is significant at the 0.01 level.

Bivariate regression in many ways similar to a Pearson correlation coefficient. Whereas a Pearson correlation coefficient provides us with the strength of a relationship between two variables, bivariate regression provides us with just a little bit more information. Let's go to SPSS to see how we can perform this test.

To perform this bivariate regression in SPSS we click on Analyze. And we move our cursor down to Regression. Right away, you will see a number of options for regression. For bivariate regression we're using a method called ordinary least squares, which in SPSS is referred to as Linear Regression. Bivariate regression often goes by the term simple linear regression as well.

If we click on that, we'll see that we have a number of options available to us. A dependent variable and an independent variable box are the first things that we want to pay attention to. Let's go ahead and predict a respondent's socioeconomic status index from their highest level of education.

Again, we want to pay attention to levels of measurement. For our independent variable, we want to use the respondent's highest level of education measured as number of years in school. That is at the interval or ratio level of measurement. Let's go ahead and enter our dependent variable first, Socioeconomic Status Index.

So again, I can hover my cursor over this variable to make sure this is the proper variable that I want to select. Highlight it. And just use the arrow key to move it over.

We'll scroll up to my independent variable, which is, again, respondent's highest level of education measured as number of years. Move that over. And then I can click OK.

Let's go ahead and walk through some of the output that SPSS provides us for the bivariate regression model. Let's first focus on our model summary. The large R, or multiple R, in a bivariate regression model is equal to the Pearson correlation coefficient. In this case, we have a statistic of 0.610 If we ran a Pearson correlation coefficient between a respondent's socioeconomic status

and their highest level of education, we would receive a Pearson correlation coefficient statistic of 0.610

The R Square, here a statistic of 0.372 provides us with more information about the overall model. From the 0.372, we can infer that 37% of the respondent's socioeconomic status is accounted for, or explained, by their highest year of school completed.

The Adjusted R Square is similar in this case, because we only have one predictor. As we increase the number of predictors in a multiple regression model, that Adjusted R Square will change from the R Square.

Next, we go to our ANOVA box. Here, we're testing for the overall significance of the regression model. You'll see a significance level of 0.000, which is well below the conventional 0.05 threshold. Therefore, we can conclude that our model has statistical significance and the R Square can be interpreted.

Next, let's go ahead and interpret the coefficients output. You'll see here that we're provided with several statistics. The first statistic is the constant. This is where the slope of our regression line intercepts with the y-axis.

Our next coefficient to interpret is our independent variable, here, highest year of school completed. This is the unstandardized coefficient, so we can interpret this as for every one unit increase in our independent variable our dependent variable will change by this value.

So we'll say it in plain English. For every additional year of school completed, socioeconomic status will change by 3.765 units, on average.

We'll also note here that SPSS provides us with a standardized coefficient, or a beta, for our independent variable. You might notice right away that this statistic, this value, is the same as the Pearson R, 0.610. That's because the standardized coefficient standardizes the units of measure.

We, of course, also want to pay close attention to our significance. Here, we have a significance level of 0.000, which is well below the 0.05 threshold. Therefore, we can reject the null hypothesis that there is no relationship between our two variables of highest year school of completed and respondent's socioeconomic index. It appears that the more school one completes, on average, the higher their socioeconomic index will be.

This was just a basic introduction to bivariate regression in SPSS. Although the procedures are rather simple, there still is a lot more to know about bivariate regression. As you'll probably note, some of the output we didn't go over. If you have additional questions, be sure and use your textbook and also utilize your

faculty instructor. We want you to understand linear regression. And we're here to see you succeed.