

TutorTube: SPSS – Multiple Regression

Spring 2021

Introduction

Hello, welcome to another edition of TutorTube, where the Learning Center's Lead Tutors help you understand challenging course concepts with easy to understand videos. My name is Kelly Schmidt, Lead Tutor for statistics at the Learning Center. In today's video, we will run through the process of conducting a multiple regression in SPSS. Let's get started!

Research Question

First, let's define our research question. In this example, we are interested in examining predicting a student's GPA based on their scores on three other tests: a test measuring self-confidence, a test measuring adaptability, and a test measuring IQ. Here we can see the data from 14 students who were given these tests, along with their GPAs.

Subject	GPA	Self Confidence	Adaptability	IQ
1	3.8	45	60	105
2	4.0	50	10	109
3	3.2	45	50	102
4	3.5	51	25	95
5	2.5	60	15	92
6	3.0	39	80	101
7	2.1	42	41	99
8	2.8	41	14	95
9	3.6	46	57	94
10	4.0	50	68	110
11	3.6	53	24	104
12	3.4	47	95	105
13	3.2	48	25	98
14	2.0	40	36	75

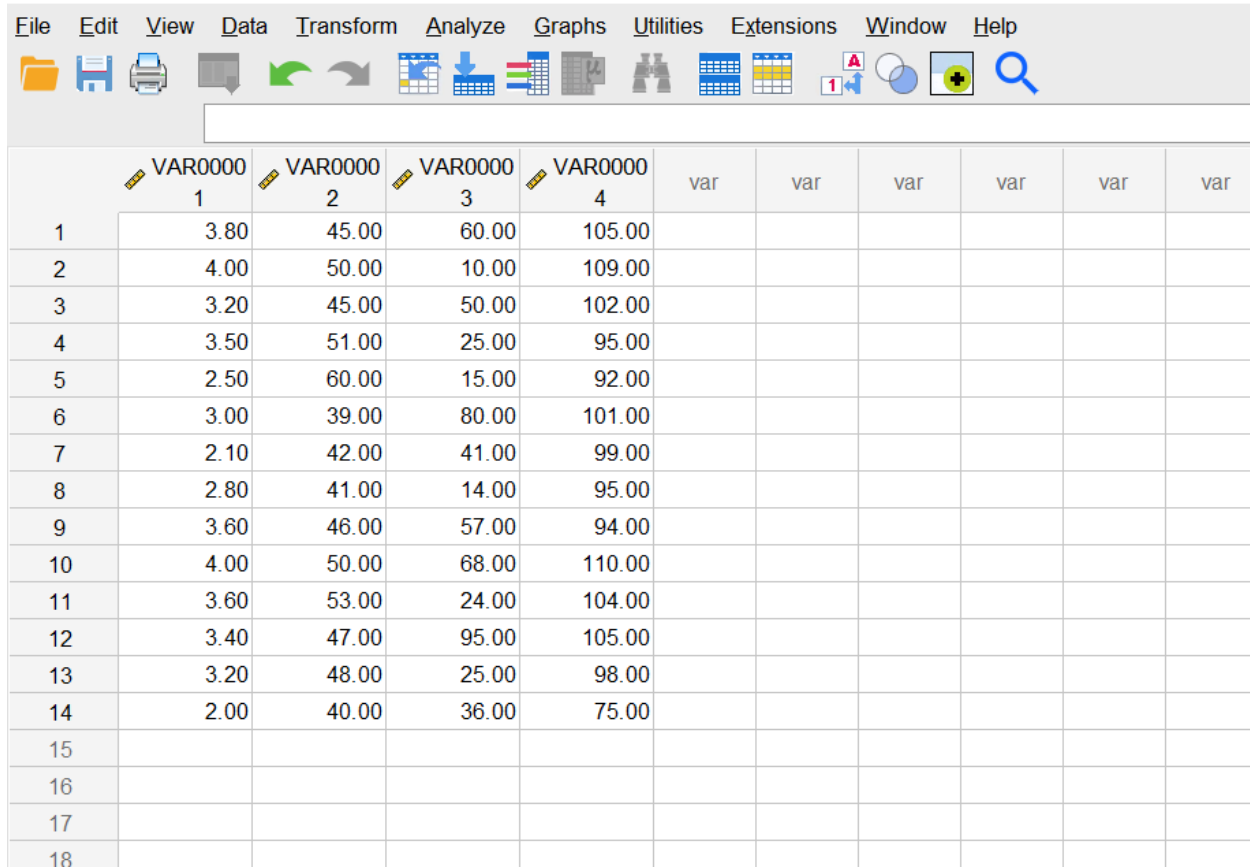
Notice that we have four numeric (or scale) variables here, not counting the subject identifier row. When doing regression problems, the first step is to identify our dependent and independent variables. In this scenario, our goal is to predict GPA, so GPA will be our dependent variable. All the others will serve as our independent variables.

Data Entry

Next, we need to enter our data into SPSS. Since each row in our dataset represents the scores from a single student, we can enter the data in columns as is into SPSS.

*Untitled1 [DataSet0] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help



	VAR0000 1	VAR0000 2	VAR0000 3	VAR0000 4	var	var	var	var	var	var
1	3.80	45.00	60.00	105.00						
2	4.00	50.00	10.00	109.00						
3	3.20	45.00	50.00	102.00						
4	3.50	51.00	25.00	95.00						
5	2.50	60.00	15.00	92.00						
6	3.00	39.00	80.00	101.00						
7	2.10	42.00	41.00	99.00						
8	2.80	41.00	14.00	95.00						
9	3.60	46.00	57.00	94.00						
10	4.00	50.00	68.00	110.00						
11	3.60	53.00	24.00	104.00						
12	3.40	47.00	95.00	105.00						
13	3.20	48.00	25.00	98.00						
14	2.00	40.00	36.00	75.00						
15										
16										
17										
18										

Figure 1. Data View

Notice that when we first copy the data in, SPSS automatically labels the columns with VAR0000 by default instead of our variable names. To change this, we can click on "Variable View" at the bottom of the screen to switch to a new window.

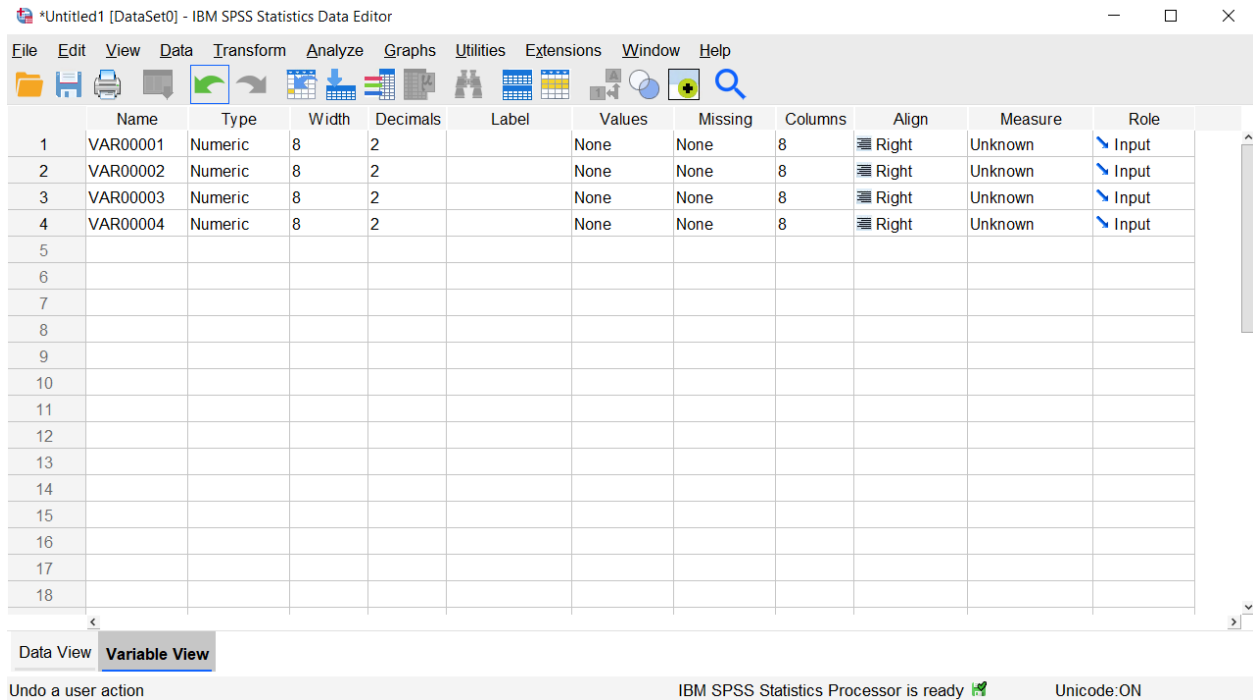


Figure 2. Initial Variable View

From here, we can click on the VAR0001 label (which corresponds to our first column with the GPA scores) and rename it "GPA." We can do the same for the next three cells with Self Confidence, Adaptability, and IQ as well.

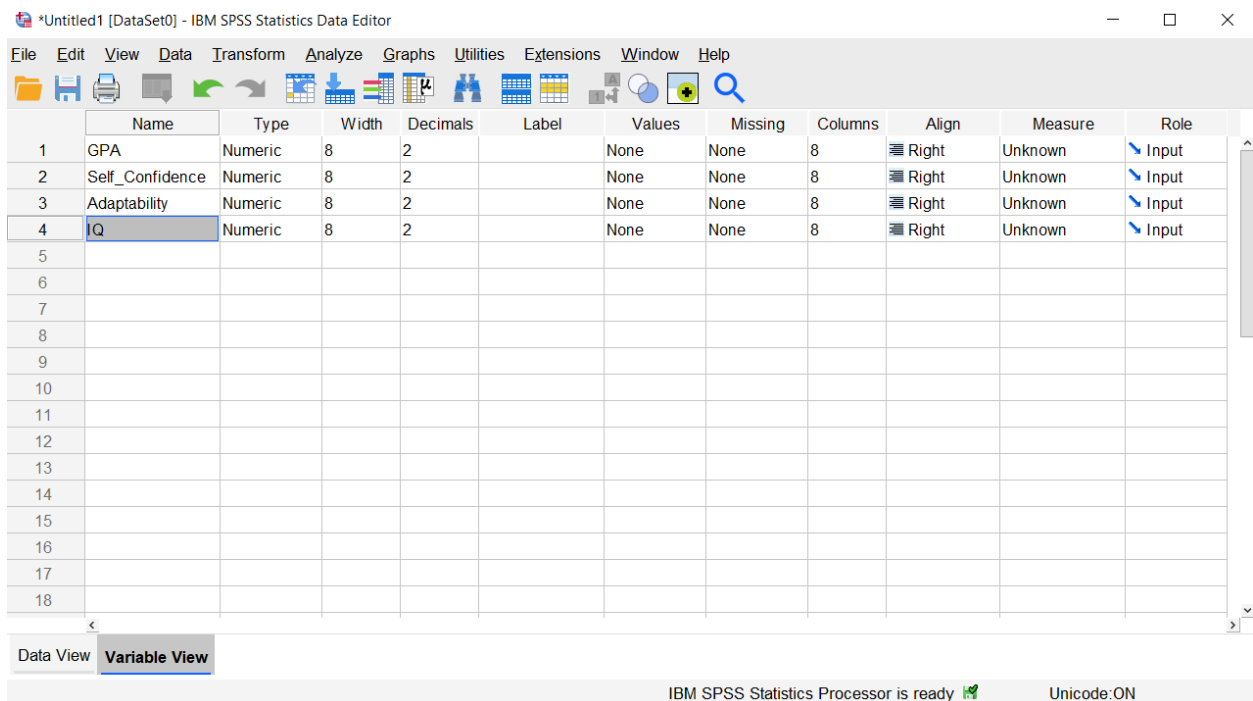


Figure 3. Variable View with Column Names

Next, we can specify the measure for each variable. Since all of our variables were numbers, not names or categories, these will all be coded as Scale.

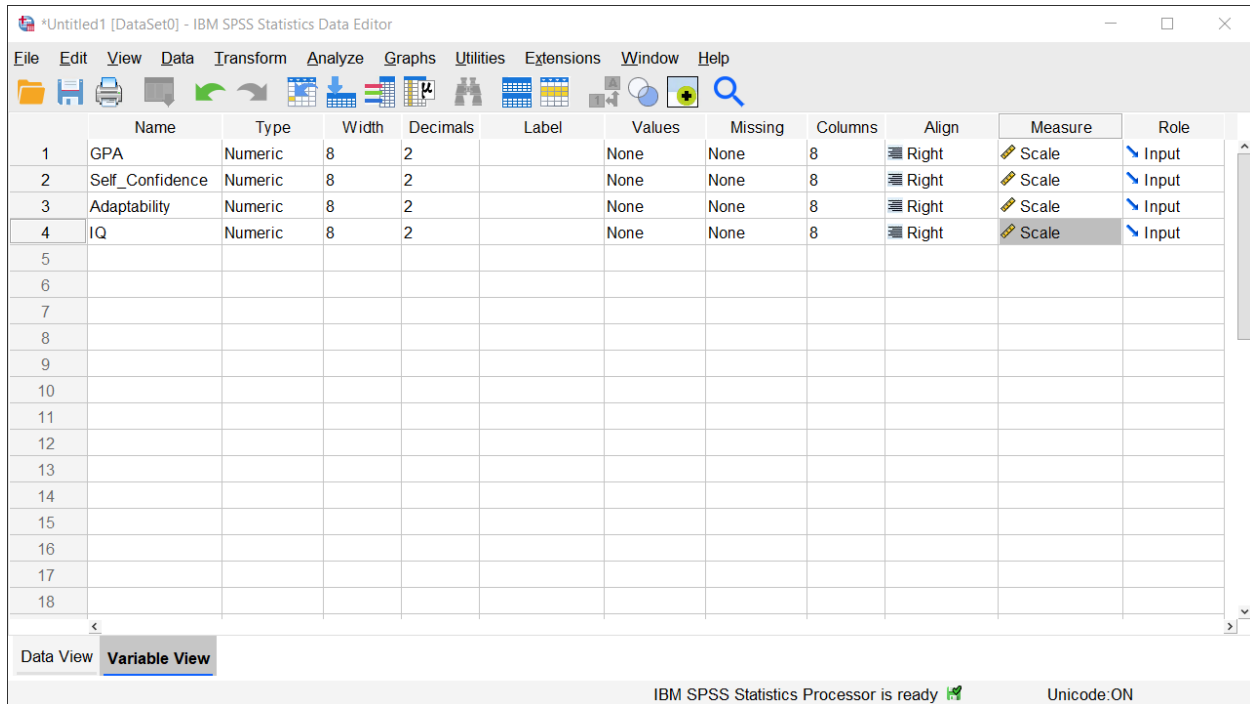


Figure 4. Variable View with Measures

Alright, with that we are ready to run the regression.

Analysis

Click on “Analyze,” then “Regression,” then “Linear.” This will pull up a new window. From here, we need to first choose our dependent variable. Remember that since we are trying to *predict* GPA, it will be our dependent or response variable.

Our independent variables will go into the Independent variable box. You can move them one at a time or use Control+Shift to highlight multiple values at once.

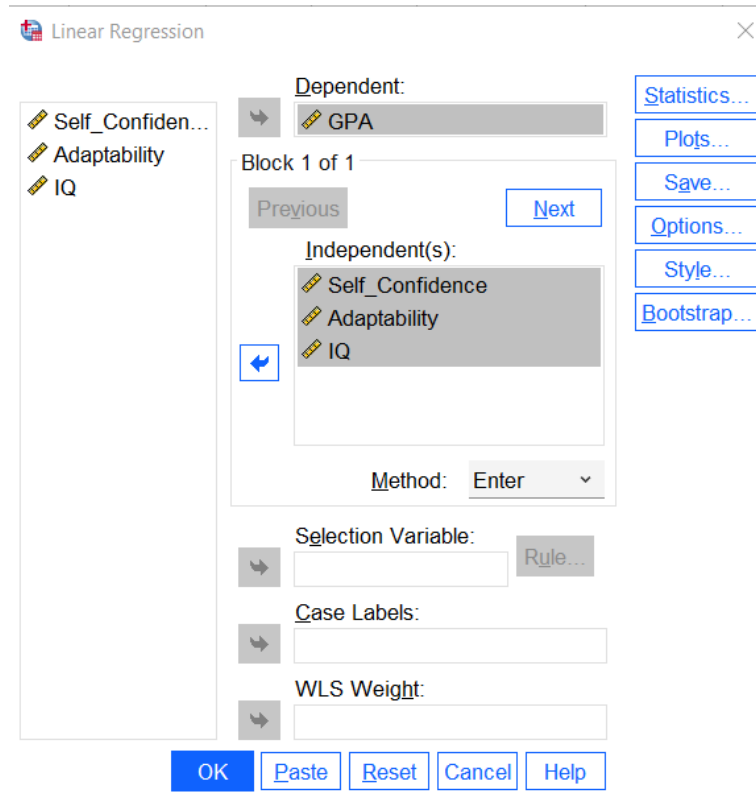


Figure 5. Analyze > Regression > Linear

Next, click on “Statistic.” Here, we will want to select Descriptives, Part and partial correlations, and Collinearity diagnostics. Descriptives will give you a table showing your summary statistics for each variable (the mean, standard deviation, and count). Part and partial correlations is useful for determining how influential each individual independent variable is, and collinearity diagnostics allow us to check and make sure that our assumption about multicollinearity hasn’t been violated.

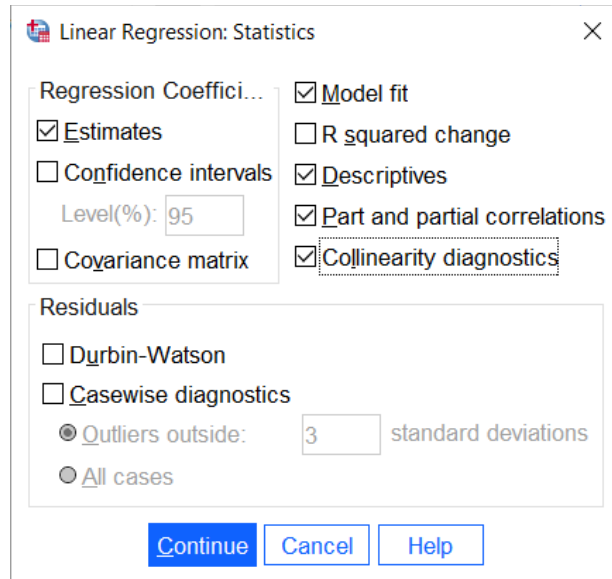


Figure 6. Linear Regression - Statistics

Click "Continue," and then from here we can click on "Plots." This window will let us make a couple of scatter plots that will help to check our assumptions on homoscedasticity and normality of residuals. Use the arrows to move *ZPRED into the X axis box and *ZRESID into the Y axis box. This will create a scatter plotting the standardized residuals against the standardized predicted values.

Next, click on the boxes next to histogram and Normal probability plot, and then click "Continue" again.

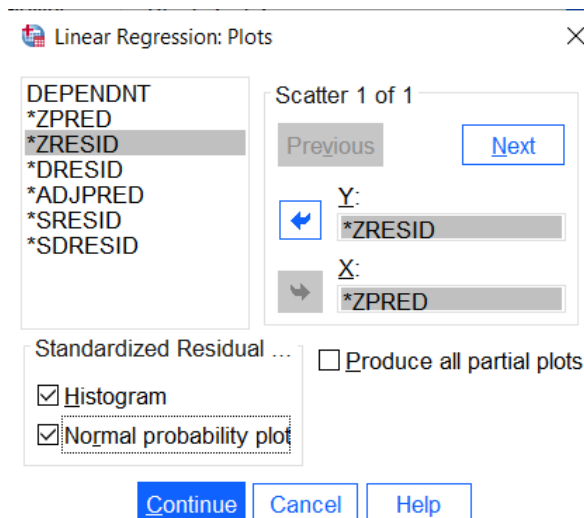


Figure 7. Linear Regression - Plots

From here, we are good to go, and we can click on OK.

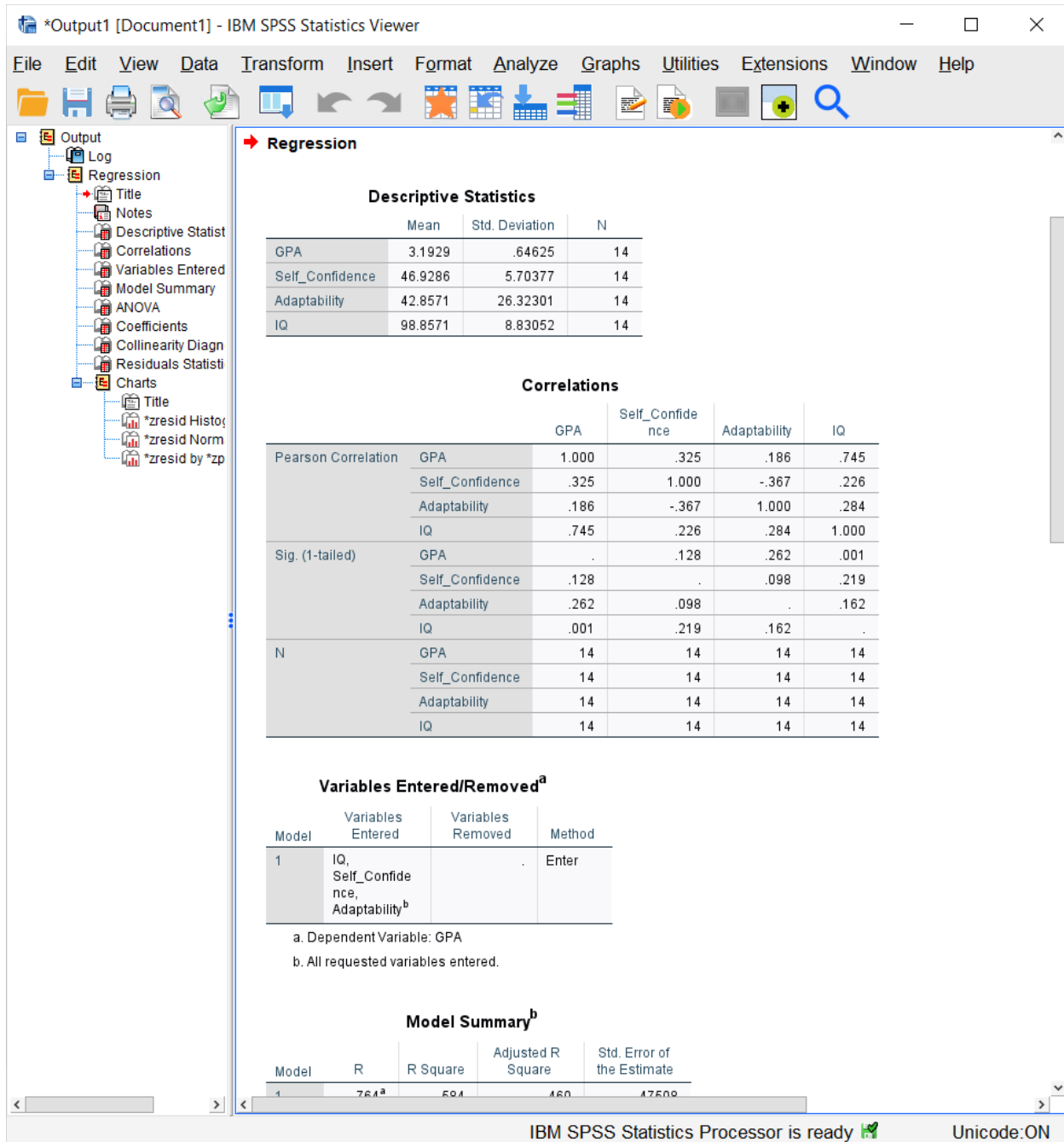


Figure 8. Output

This will open up our output window with the results of our regression. At the top, we can see our descriptive statistics table with the summary statistics for each variable. Below that, we have a correlation matrix showing the correlations between each pair of variables.

If we keep scrolling down, we can find the model summary and the ANOVA table for the regression. If we want to go about interpreting this model, we

should first look at the sig. value for the overall test. Because this value of 0.027 is less than 0.05, we can conclude that our overall regression model using these three independent variables is significant, which means that the three variables together are able to predict GPA scores better than chance.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.764 ^a	.584	.460	.47508

a. Predictors: (Constant), IQ, Self_Confidence, Adaptability

b. Dependent Variable: GPA

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.172	3	1.057	4.685	.027 ^b
	Residual	2.257	10	.226		
	Total	5.429	13			

a. Dependent Variable: GPA

b. Predictors: (Constant), IQ, Self_Confidence, Adaptability

Figure 9. Model Summary and ANOVA Tables

Now that we know this, we now want to know how good the model actually is. The R-squared value gives us the percentage of variation in the dependent variable scores that is explained by the independent variables. Here, we would say that 58.4% of the variation in GPA scores can be predicted or explained by a student's Self Confidence, Adaptability, and IQ.

Coefficients Table

We can get a lot of information from the coefficients table, including the coefficients needed to construct the multiple regression equation. Be sure to pay attention to whether or not you are asked for the unstandardized equation or the standardized equation because the coefficients used for each will change.

Coefficients ^a											
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics		
	B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF	
1	(Constant)	-2.851	1.677		-1.700	.120					
	Self_Confidence	.022	.027	.194	.824	.429	.325	.252	.168	.746	1.340
	Adaptability	.002	.006	.064	.265	.796	.186	.083	.054	.723	1.383
	IQ	.050	.017	.683	2.984	.014	.745	.686	.609	.793	1.261

a. Dependent Variable: GPA

Figure 10. Coefficients Table

For example, if we wanted to build the unstandardized (or raw score) equation, we would look at the values in the first column. Just like with a simple linear regression, we want to find both a “y-intercept” and a “slope” value for each of our variables. We start with our dependent variable (GPA) with a hat indicating that the equation will give us predicted values. Next, we take the constant value, -2.851, and leave it as is without multiplying it. Next, we add a term which has the coefficient in front of Self Confidence (0.022), multiplied by the variable. We do the same for each variable in the column to get our full equation.

$$\widehat{GPA} = -2.851 + .022(\text{Self Confidence}) + .002(\text{Adaptability}) + .050(\text{IQ})$$

We can use this equation, this model, to predict scores on GPA if we have a student's Self Confidence, Adaptability, and IQ scores.

Multicollinearity

Looking at the final column of the table, we can find the VIF values. We use these to check our assumptions on multicollinearity between the independent variables. Here all our values are pretty close to one, which is good because it tells us that multicollinearity is low and our assumption is met. If these values are bigger than five, then you might want to reconsider your choice of independent variables (Akinwande et al., 2015).

Structure Coefficients

Finally, we can calculate our structure coefficients to determine which of our independent variables contributes the most to the model. For example, if we wanted to calculate the structure coefficient for the IQ variable, we would use the formula here:

$$r_s = \frac{r_{IQ}}{R}$$

r_{IQ} is the zero-order correlation for IQ and R is the correlation coefficient for the full model. We can find the zero order correlations in the coefficient table and see that the value for IQ is 0.745. The R for the full model can be found in the model summary table at 0.764.

Dividing these, we get a value of .975 for the structure coefficient. In order to interpret this value in context, we square it first. This gives us a value of .951 or about 95% for the squared structure coefficient for IQ. Now, we can use this to say that about 95% of the variation in dependent variable scores that was explained by the model was accounted for by the IQ variable. This is a pretty high value, which tells us that IQ is probably a significant predictor of GPA based on this data.

$$r_s = \frac{r_{IQ}}{R}$$

$$r_s = \frac{.745}{.764} = .975$$

$$r_s^2 = (.975)^2 = .951 = \mathbf{95.1\%}$$

Of course, there is more to interpret with an analysis like this, but we have gone through some of the main concepts that'll be helpful when you are going through your coursework.

Outro

Thank you for watching this TutorTube presentation! I hope you enjoyed this video. Please subscribe to our channel for more exciting videos. Check out the links in the description below for more information about The Learning Center and follow us on social media. See you next time!

References

*All analyses performed with IBM SPSS Statistics 27.

Akinwande, M., Dikko, H., & Samson, A. (2015). Variance inflation factor: As a condition for the inclusion of suppressor variable(s) in regression analysis.

Open Journal of Statistics, 5, p. 754-767.
<http://dx.doi.org/10.4236/ojs.2015.57075>