SPSS Guide: Regression Analysis

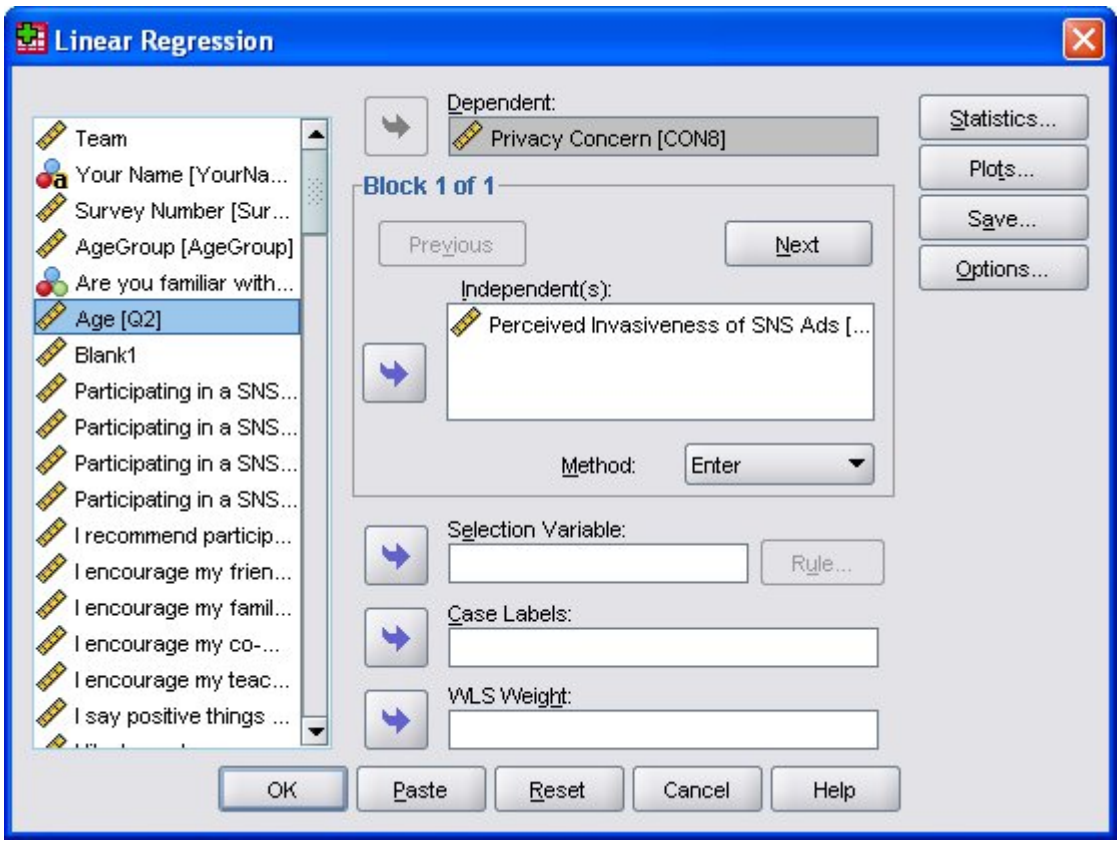
I put this together to give you a step-by-step guide for replicating what we did in the computer lab. It should help you run the tests we covered. The best way to get familiar with these techniques is just to play around with the data and run tests. As you do it, though, think of the research questions from your project and how these tests can answer them.

For the final exam, be sure you are able to:

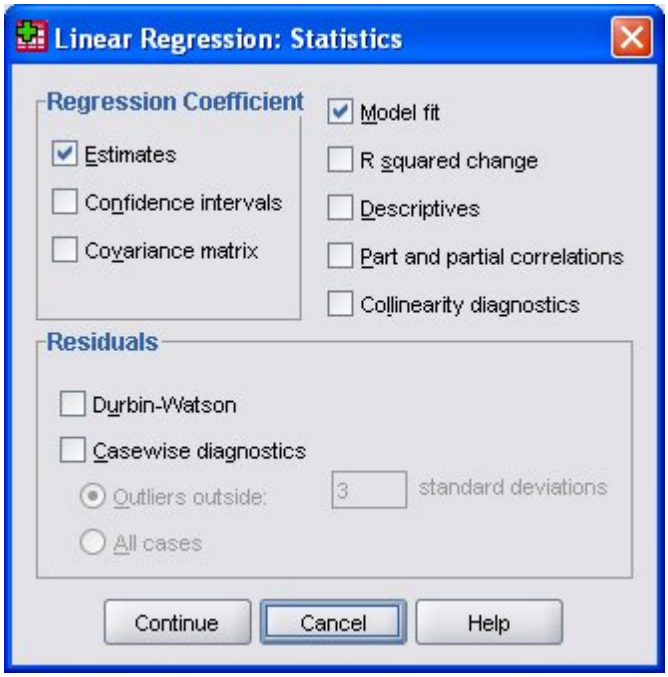
- Peform a regression analysis (both bivariate and multivariate)
- Interpret the output
    - Which variables are significant?
    - How is an unstandardized coefficient used? Standardized?
    - Which is the b0 coefficient?
    - Which are the b coefficients?
- Create a regression equation from the output
- Plug in coefficients to predict a Y variable from the X variable(s)

In SPSS, go to Analyze>Regression>Linear Regression

Enter your dependent variable (Y) in the dependent box, and your independent variable(s) in the "Independent(s)" box.



Under the "Statistics ..." button, make sure that "Estimates" is checked under Regression Coefficient, and "Model fit" is checked.



Click continue, and then "OK" on the main menu.

Your output should look like this:

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	CON3 Perceived Invasiveness of SNS Ads[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: CON8 Privacy Concern

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.259[a]	.067	.066	1.02546

a. Predictors: (Constant), CON3 Perceived Invasiveness of SNS Ads

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	96.802	1	96.802	92.055	.000[a]
	Residual	1344.953	1279	1.052		
	Total	1441.755	1280			

a. Predictors: (Constant), CON3 Perceived Invasiveness of SNS Ads

b. Dependent Variable: CON8 Privacy Concern

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.297	.096		23.968	.000
	CON3 Perceived Invasiveness of SNS Ads	.268	.028	.259	9.595	.000

a. Dependent Variable: CON8 Privacy Concern

First, look at your model summary. Your R-square tells you the "goodness of fit" of the model. You can think of it as a percentage. Our R-square for this model is .066, which means that the X variable can explain about 6.6% of the change in Y.

The next table, ANOVA, just tells us that the model can predict Y using X. The significance is .000, so we can reject the null hypothesis that "The model has no predictive value."

Finally, the most important table is the coefficients. The significance level of .000 indicates that we can reject the null hypothesis that X does not predict Y.

The first coefficient, "(Constant)", is your intercept term. That is, before you account for the dependent variable(s) – or, putting it another way, when X is zero – this is the value of Y. In this case, the intercept is 2.297, so when X=0, Y will equal 2.297. Remember, that regression equations are in this format:

$$Y = b_0 + b_1x_1$$

Each of the other coefficients are b variables, or the slope of the line. For each 1-unit change in X, Y will change by b units. Since we only have one variable in this case, we just have a b1 (the slope) and a x1 (the value of X). In this case, X is CON3 (Perceived invasiveness). For prediction, use the unstandardized coefficient.

We can plug the coefficients into the formula to predict Y using a value of X.

$$Y = b_0 + b_1x_1$$

$$Y = 2.297 + .028* \text{CON3}$$

If X=1, then Y= 2.297 + (.028*1) = 2.325
If X=2, then Y=2.297 + (.028*2) = 2.353 etc.

So when do we use the standardized coefficient? It tells us the relative size of the influence of a variable. The standardized coefficient can be interpreted like the Pearson coefficient in bivariate associative analysis (a 0-1 scale with 1 being perfectly correlated).

In multivariate regression, this allows us to compare the relative influence of multiple variables, as shown in the next example where we add some more variables to this analysis. Let's use the following variables:

Dependent: CON8 (Privacy concern)

Independent: CON6 (Perceived information security)
Q2 (Age)
(F4) How many months participating in SNS
CON3 (Perceived invasiveness)

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	F4 About how many months have you been participating in SNS sites, CON6 Perceived Information Security of SNS, CON3 Perceived Invasiveness of SNS Ads, Q2 Age[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: CON8 Privacy Concern

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.293[a]	.086	.083	1.01456

a. Predictors: (Constant), F4 About how many months have you been participating in SNS sites, CON6 Perceived Information Security of SNS, CON3 Perceived Invasiveness of SNS Ads, Q2 Age

Notice that our R-square has increased to .083, so we have a better model with the additional variables.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	118.566	4	29.641	28.797	.000[a]
	Residual	1267.112	1231	1.029		
	Total	1385.678	1235			

a. Predictors: (Constant), F4 About how many months have you been participating in SNS sites, CON6 Perceived Information Security of SNS, CON3 Perceived Invasiveness of SNS Ads, Q2 Age

b. Dependent Variable: CON8 Privacy Concern

The ANOVA shows that the model has predictive value, since it is significant. (Null hypothesis of no predictive value is rejected)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.251	.164		13.686	.000
	CON3 Perceived Invasiveness of SNS Ads	.253	.028	.244	8.905	.000
	Q2 Age	.010	.003	.081	2.850	.004
	CON6 Perceived Information Security of SNS	-.095	.029	-.093	-3.317	.001
	F4 About how many months have you been participating in SNS sites	.004	.002	.067	2.378	.018

a. Dependent Variable: CON8 Privacy Concern

Notice that all of the significance levels are < .05, so they are all significant. (Reject null hypothesis that they are not associated with the dependent variable).

The unstandardized coefficients can be used to create an equation for Y.

b_0 = (Constant)

b_1 = CON3

b_2 = Age

etc.

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$$

$$Y = (2.251) + (.253)*x_1 + .010*x_2 + (-.095)*x_3 + (.004)*x_4$$

For this example, we'll stop here, but to calculate Y you would just plug in the data from an individual observation for each of these variables.