

Introduction to SPSS
Module 4: Correlation and Regression

SLIDE 1, 2, 3

Modules 1, 2 and 3 covered preparing data for analysis, working with variables, and summaries and descriptive statistics. Module 4 will cover correlation and regression. Attachments include instructions, sample data, and sample survey.

Note: Remember to save your file now and then.

SLIDE 4

Quick Review of the 5 steps (actions) of analysis (remember this from statistics class?):

Action 1. What is/are your hypothesis(es):

List your variables:

Their level of measurement:

The number of levels:

<hr/>	<hr/>	<hr/>
<hr/>	<hr/>	<hr/>
<hr/>	<hr/>	<hr/>
<hr/>	<hr/>	<hr/>
<hr/>	<hr/>	<hr/>
<hr/>	<hr/>	<hr/>
<hr/>	<hr/>	<hr/>
<hr/>	<hr/>	<hr/>

SLIDE 5

Action 2. Select a significance level: typically .05 or .01 (.05 means that, if the findings are less than .05, chances are less than 5 in 100 that the findings are not true (very unlikely that the null is true), so we reject the null. If you need to be more sure of your findings (say in the case of a drug that has harmful side effects), you may want to choose a lower significance level to reduce the chance of making a Type I error (rejecting the null when in fact the null was true). If you need to be more sure that you don't miss a finding, you may want to choose a higher significance level to reduce the chance of making a Type II error (not rejecting the null when the null is false).

Select: One-tailed or Two-tailed (Does your hypothesis predict that the difference or relationship between the variables will be in a specific direction?)

SLIDE 6

Action 3. Determine which analysis and perform the analysis. Group differences, strength of relationship, or ? Number of DVs? Number of IVs? Number of levels? Decision trees are helpful.

SLIDE 7

Action 4.

After calculating in SPSS, review the results to determine if you have any significant findings (any with levels of probability that are smaller than the p-value you chose).

Action 5.

Decide whether you will reject or fail to reject the null hypothesis.

SLIDE 8

Run a Correlation

A Pearson correlation analyzes relationships between parametric, linear (interval, ratio, “scale” in SPSS) variables. If ordinal, use Spearman Rho even if not from a normal distribution. You can enter several variables and get a matrix of the relationships (direction and strength, -1 to 1).

SLIDE 9

Open SPSS and open you data file.

From Analyze menu, choose Correlate, and then Bivariate. In the list of variables in the left box, click on one of your variables and then click on the arrow button to move it to the Variable box. Then click on your other variable(s) and then click on the arrow button to move to the Variable box. Check Pearson or Spearman, Two or One-tailed, and Flag significant correlations.

In the Option section, check “Means and standard deviations” under Statistics section; click Continue. Then click OK.

Review Output

Sample results: There was a positive correlation ($r = .52, p < .05$) between high school GPA and college GPA.

SLIDE 10, 11

A scatter plot is a good graph for correlations. One variable on each axis and dots represent intersection of participants’ scores on the two variables. Use can display a third variable by entering it at *set markers by*.

SLIDE 12

Regression

Regression investigates whether _____ can be predicted by _____. Simple linear regression is used for two interval/ratio variables with a normal distribution (some exceptions apply). Multiple regression is used to predict a level of a variable using two or more predictor variables.

SLIDE 13

From **Analyze**, choose **Regression**, then **Linear**. Move the variable that you are trying to predict to Dependent and move the variable(s) that you are trying to predict from to Independent. Click OK. (Example: We want to know if we can predict College GPA by knowing a person’s High School GPA and RSR math score)

Review output – Model Summary (R is the absolute value of correlation coefficient, R Square is the proportion of the variation of the dependent variable that can be explained by the independent variable and Std. Error of the Estimate is the standard deviation of the confidence intervals). ANOVA Summary (check for significance). Coefficients (under “B”, constant is the intercept and the variable is the slope).

Regression results are often reported in a table that shows the regression coefficient for each predictor variable and the overall R or R².

SLIDE 14, 15

Recap and Modules