Lab Exercise #5 Correlation and Linear Regression

Tutorial Objectives

1) Learn how examine data for measures of correlation, using both visual examination (plots) and statistical measures.
2) Be able to plot data in scatterplots using SPSS.
3) Understand the Pearson Product Moment (interval/ratio) and Spearman (ordinal) Rank correlation measures and how to interpret SPSS results.
4) Be able to generate and interpret a simple linear regression model.

1. Use the Anchorage Parks & Open Space survey data set to answer the questions below.

a. Generate a simple scatterplot for age (q30) and how long respondents have lived in Alaska (q22).

For each pair, state the strength and direction of the relationship, the amount of variance explained, and whether the relationship is statistically significant (see page 128 in textbook for explanation of variance explained).

b. Respondent's age (q30) and how long they have lived in Alaska (q22).
c. Respondent's age (q30) and how long they have lived in Anchorage (q23).
d. Respondent's age (q30) and their attitudes toward parks and opens (all items q21_1 through q21_8).

Are any of the correlations statistically significant? If so, which one(s)? What would you guess might be going on with this relationship?

2. Look through the list of items (facilities and activities) in Q-2. Before running the stats, which items would you guess to have the highest correlation coefficients on Part A, "importance"?

a. In SPSS, run bivariate correlations on all the importance items in Part A for "Activities" only (q2_27a through q2_36a—10 items). Which 5 pairs of items have the highest correlation (r-values)?

3. Get the SPSS data file "examanx.sav" from the course website. This data file has hypothetical measures of student exam scores, gender, measured anxiety levels, and time spent revising (studying). What the British call "revising" is called "studying" in the U.S. Run bivariate correlations on the variables "exam", "anxiety", and "revise."

a. For each pair, state the strength and direction of the relationship, the amount of variance explained, and whether the relationship is statistically significant.

b. Now run a partial correlation between "exam" and "anxiety" while controlling for "revise". What is the partial correlation coefficient? Is the partial correlation greater or less than the bivariate correlation between "exam" and "anxiety"? (See page 98 in textbook for explanation of partial correlation).

4. Using the same "examanx" file, test whether there was any difference on exam scores, anxiety levels, or time spent revising based on gender. In other words, are there differences between males and females on these variables?

a. Any difference between on exam scores?
b. Any difference in anxiety levels?
c. Any difference in revision time?

5. Download the file "record1.sav" from the course website. The marketing department has approached you (the researcher) indicating their belief that the amount of marketing funds expended by the company is directly related to the dollar amount of sales of the product (records). Build a simple regression model to test the relationship.

a. What is the strength and direction of the relationship?
b. Is your model statistically significant? At what alpha level? (Note: an alpha level is the level of statistical significance—i.e., .1 or .05 or .001)
c. What is the regression equation (careful: do you use unstandardized or standardized coefficients)?
d. If the model is used for prediction, what would the level of sales be for the company with a marketing budget of £200,000 (careful—since units are in thousands, in the regression equation you would substitute 200—not 200,000).
e. See if you can generate an SPSS interactive scatterplot with the regression line included.