Homework 1
Due Thursday, January 30 (10:00 AM)

For all questions, please show your work or include a copy of the output, whichever is relevant. Please write your answers in report form including relevant statistics, as if you were describing results in a published study. Be sure to describe each finding in terms of the research problem. Most responses to a particular question should be approximately one paragraph in length. Text should be typed, but please do not waste your time typing hand computations. All answers should be in your own words.

Data from the next several problems come from a survey of adolescent political opinions, which can be downloaded from the data page, http://web.pdx.edu/~newsomj/data.htm.¹ The data set includes a composite measure that assesses the extent to which students used critical thinking when encountering news stories (critical), based on the average of three items ("When I see or read a news story about an issue, I try to figure out if it is biased," "When I hear news about politics, I try to figure out what is REALLY going on," and "News about people running for office makes me wonder how they might change things."), each rated on a three-point scale of agreement (thus the index has a possible range from 1 to 3). The data set also contains a variable about interest in politics (political) based on an average of four items such as "How much attention do you pay to news about politics?" and "In general, how much interest do you have in politics?", each rated on a 5-point scale of agreement (with possible range for the index from 1 to 5). A single item was used to assess intentions to attend college, "I definitely plan to attend college," rated on a 5-point scale of agreement from 1 "strongly disagree" to 5 "strongly agree". An additional variable is for the respondent's race/ethnicity (race) with 1= "White/Anglo" 2= "Hispanic" 3= "African-American."

1. a. Use R and SPSS to obtain a correlation matrix of the correlations among the variables political, critical, and colbound. (Skip any scatterplots for now—we will use some later.) Describe and explain your results. Include both outputs with your response (but only one write-up of the results is needed).

b. Use R and SPSS to conduct a simple regression analysis in which colbound predicts political. Be sure to obtain standardized and unstandardized coefficients, R-squared, and confidence intervals. Report the relevant statistics and interpret your findings in terms of the research problem. Include an interpretation of the standardized and unstandardized coefficients, significance test, and R-squared.

c. Using R and SPSS, run a simultaneous regression analysis with colbound and critical predicting political. Report your results, including standardized and unstandardized coefficients, significance tests, confidence intervals, R-squared, and F-test, and provide an interpretation of the coefficients and R-squared.

d. Using R or SPSS, conduct a multiple regression analysis using colbound and critical as predictors of political. This time use a hierarchical approach, entering colbound in the first block and then critical in the second block. Report your results, including standardized and unstandardized coefficients, significance tests, confidence intervals, R-squared, F-test, R-squared change, and F change, and provide an interpretation of the coefficients and R-squared. Include a comment that compares your results to those obtained in the simultaneous regression above. How are they similar or different and why? (be specific, give values)

e. Based on your results from the hierarchical multiple regression in the previous problem, sketch a Venn diagram and label it with the approximate percentage of variance in the dependent variable accounted

¹ These data consist of a random sample from the study conducted by McDevitt, M., & Kiousis, S. (2007). The red and blue of adolescence: Origins of the compliant voter and the defiant activist. American Behavioral Scientist, 50(9), 1214-1230.
for each of the predictors and the percentage variance overlap in the two predictors (be sure to clearly indicate which quantity from the output goes with which percentage).

f. Using R or SPSS, test a regression model which allows you to compare race/ethnicity groups in predicting political interest (political). Hint: you will not be able to use race in its current form. Report your results, including standardized and unstandardized coefficients, significance tests, confidence intervals, $R$-squared, and $F$-test, and provide an interpretation of the coefficients and $R$-squared. What are the special interpretations of the intercept and slopes in this model (be specific, give values)?

g. Using R or SPSS, run a simple regression analysis with critical predicting political. Obtain a histogram of residuals and residual plot using $X$ values (x-axis) with standardized residual values (y-axis), and make an assessment about nonlinearity, homoscedasticity, and outliers. Request the regression diagnostics that were discussed in class, obtaining at least one $X$ outlier diagnostic index, one $Y$ outlier diagnostic index, and one influence statistic. Use syntax/code or save the diagnostics to a data file to obtain the diagnostics for each case. Interpret the regression output results, diagnostics, and discuss any outliers (the type of outlier and justifying why they meet the criteria for an outlier) or other potential problems with the data. [Please include only relevant excerpts of the individual case diagnostics to save paper].

h. Remove any outliers that might have an important influence on the results and rerun the regression analysis (if none of the diagnostics cutoffs are exceeded or there are many, choose one or two, and no more than a hand full of, cases with the highest values on any of the indices and remove it/them). Report your results, including standardized and unstandardized coefficients, significance tests, confidence intervals, $R$-squared, and $F$-test, and provide an interpretation of the coefficients and $R$-squared. Discuss how your results changed and whether this remedy fixed any problems you noted in the prior analysis including specific statements about assumptions and diagnostic values that were examined in the prior analysis.

2. Use your own data for the following problems. Make sure your dependent variable is a continuous variable (ordinal with at least 5 ordinal values, interval, or ratio scale). You will need to have at least two continuous predictors.

a. Use SPSS or R to test a simple regression with one predictor of your outcome. Provide a very brief statement about the purpose of the analysis. In your analyses, be sure to obtain standardized and unstandardized coefficients, $R$-square, confidence intervals, and a scatterplot. Report the relevant statistics and interpret your findings in terms of the research problem. Include an interpretation of the standardized and unstandardized coefficients, significance test, and R-squared. Response should be no longer than two paragraphs.

b. Use SPSS or R to test a simultaneous multiple regression with the predictor used in the previous problem plus one other predictor. Include an interpretation of the standardized and unstandardized coefficients, significance test, and $R$-squared.

3. Read one of the following articles (copies available from the class website [http://web.pdx.edu/~newsomj/mvclass/](http://web.pdx.edu/~newsomj/mvclass/)) and write two paragraphs summarizing the article. First, describe the study design (e.g., randomized experiment, non-equivalent control group design, cross-sectional survey; for a quick refresher, see [http://sphweb.bumc.bu.edu/otlt/mph-modules/programevaluation/ProgramEvaluation7.html](http://sphweb.bumc.bu.edu/otlt/mph-modules/programevaluation/ProgramEvaluation7.html)) and purpose of the study in your own words. Be sure to include who/what was studied (e.g., who were the participants?) and the number of cases. Then, choose one regression analysis used in the article that you have learned about in the course so far (i.e., simple or multiple regression), and, in your own words, describe the hypothesis that is being tested, the results obtained, and what the findings mean. Be sure to include the relevant statistical
values and whether the results were significant. Write your paragraphs as if you were describing results in a published article or reporting someone else’s results as in a review article.


