Homework 3
Due 12/5/16 (noon)

For all questions, please show your work or include a copy of the output, whichever is relevant. Please type your answers in report form, as if you were describing results in a published study. Include the relevant descriptive and statistical values in your write-up (e.g., percents, regression coefficients). **Your answers should be in your own words** and most answers should be approximately one paragraph. Data sets are available at http://web.pdx.edu/~newsomj/data.htm.

1. Use the second Early Head Start data set (child2.sav) from HW 2 (Problem 2d) for the following problems. The variable NEGLECT, which is a count of the number of child neglect reports, has been added to this data set.

   a. Use SPSS or SAS (the R approach does not appear to rescale the variables appropriately) to test a hypothesized mediational model in which the race/ethnicity of the mother (WHITE) leads to having a boyfriend that is not the biological father child neglect reports (BOYFRIEND NEGLECT) which leads a higher risk of reported abuse (ABUSE). Report and interpret your findings. Be sure to include the direct effect regression coefficients, indirect effect coefficient, and bootstrap confidence limits.

   b. Use SPSS, R, or SAS to conduct a multiple logistic regression predicting ABUSE with PROGRAM, BOYFRIEND, WHITE, and WELFARE as predictors (same as HW 2, problem 2d). Obtain a residual plot (with standardized residuals on the y-axis and predicted probabilities on the x-axis) and casewise values for $\Delta\chi^2$ or ΔD and $\Delta\beta$ to check for outliers and influential data points. State what each diagnostic test tells you. Describe your findings, reporting the diagnostic values for any cases that you judge to be possible outliers or influential data points. (Because reports of diagnostic tests are rare or nonexistent in research papers, you do not need to attempt to emulate a published report format).

2. For the problem below, a new data set (widow.sav) was taken from a study conducted by David Morgan and Margaret Neal that examined social relationships among recent widows. A yes/no question about whether the respondent felt lonely was asked at two different time points (LONELY1, LONELY2), each six months apart. AGE1 is the age of the respondent at Time 1 and EDUC1 is the years of education. A question about income adequacy (INCADQ1), a 4-point rating of the extent to which she felt she had enough money each month to get by, and self-rated health (HEALTH1) were both assessed at the first time point.

   a. Use SPSS, R, or SAS to test a lagged regression model to investigate whether income adequacy (INCADQ1) and health (HEALTH1) predicted loneliness at Time 2 (LONELY2) after controlling for loneliness at Time 1 (LONELY1). Report and interpret your findings, with special attention to the interpretation of the longitudinal model. Be sure to include the regression coefficients, the odds ratios, confidence limits, model fit information, and a pseudo-R^2 measure.

   b. Use SPSS, R, or SAS to test a cumulative logistic model predicting the ordinal variable income adequacy (INCADQ1) with age (AGE1), education (EDUC1), and health (HEALTH1) as predictors. Report and interpret your findings. Be sure to include the regression coefficients, the odds ratios, confidence limits, model fit information, and a pseudo-R^2 measure.

   c. Use SPSS, R, or SAS to test an ordinal probit model with the same outcome and predictors as you used in the cumulative logistic model. Report and interpret your findings. Be sure to include the regression coefficients, confidence limits, model fit information, and a pseudo-R^2 measure. Write one or two sentences about how your results and conclusions compare to the logistic model above.

3. You used your own data set for Problem 3 in HW 2. Do the following analyses with the same variables and the same data set.

   a. Using SPSS, R, or SAS, test the same logistic regression model you tested in Problem 3 of HW 2, but this time obtain a residual plot (with standardized residuals on the y-axis and predicted probabilities on the x-axis) and casewise values for $\Delta\chi^2$ or ΔD and $\Delta\beta$ to check for outliers and influential data points. Describe your findings, reporting the diagnostic values for any cases that you judge to be possible outliers or influential data points. (Because reports of diagnostic tests are rare or nonexistent in research papers, you do not need to attempt to emulate a published report format).

   b. Replicate the model in 3a., but this time test a probit model. Report and interpret your findings. Be sure to include the regression coefficients, confidence limits, model fit information, and a pseudo-R^2 measure. Write one or two sentences about how your results and conclusions compare to the logistic model above.