Homework 2  
Due 5/13/21 (pdf format please)

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. With the police stops data set for Homework 1 (race2.sav) from the 2014 Portland Police racial profiling report,\(^1\) use SPSS, R, or SAS to test a logistic regression model using race to predict pedestrian stops (major vs. minor offense). Report and interpret your findings. Be sure to include the regression coefficient, the odds ratio, confidence limits, model fit information, and a pseudo-\(R^2\) measure. Look back at your findings from the analysis of the contingency table chi-square (HW 1 Problem 2) and the loglinear analysis (HW 1 Problem 4). Briefly compare your results from HW 1 to your results from the logistic regression model (no more than 2-3 sentences are needed). Provide specific values from each of the analyses when making your comparisons.

2. Use the Early Head Start data set (child.sav) from HW 1 Problem 3 with SPSS, R, or SAS to test a logistic regression model predicting abuse with an interaction between the boyfriend and program. Do not bother centering the variables for this analysis for this model. Just report the test of the interaction and do not conduct any follow-up analysis. Look back at your findings from the analysis of the contingency table in HW 1 and briefly compare your results from the logistic model to the Breslow-Day and loglinear analyses (no more than 2-3 sentences is needed). Provide specific values from each of the analyses when making your comparisons.

3. A second version of the Early Head Start data set has been created (child2.sav) which adds new variables, including the race/ethnicity of the mother (white: 1 = white, 0 = other race/ethnicity), the number of encounters with the child welfare system (welfare), and the number of reports of neglect (neglect).

   a. Use SPSS, R, or SAS to conduct a multiple logistic regression predicting abuse with program, boyfriend, white, and welfare 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. If you like, you may construct a table and report only significant coefficients in the text of the write-up.

   b. Use SPSS, R, or SAS and the Hayes PROCESS macro to test whether the number welfare encounters (\(w\)) moderates the association between the number of reports of neglect (\(x\)) and the probability of abuse (omit the other covariates from the model). Report and interpret the results, including the regression coefficients, the odds ratios, confidence limits, model fit information, and a pseudo-\(R^2\) measure as well as the simple effects coefficients and the plot of the interaction.

   c. Using SPSS, R, or SAS, retest the model from 3a predicting abuse with program, boyfriend, white, and welfare as predictors. This time, obtain casewise diagnostics and (at minimum) a histogram of residuals, a plot of residuals by predicted probabilities, and a plot of leverage by predicted probabilities. Examine the casewise diagnostic values and report whether, in your expert opinion, there are any outliers for \(Y\), \(X\), or influential data points (include only the relevant excerpts from the casewise table illustrating any large values in order to save space). Give the values and the case (or covariate pattern) numbers. Examine the plots and discuss any assumption likely assumption violations, including poorer prediction at lower or higher values of \(X\) (predicted probabilities), and outliers.

---

\(^1\) Not needed for the assignment, of course, but you can download the report here: https://www.portlandoregon.gov/police/article/481668.
4. For these problems, use your own data set(s), making sure the variable types match the analysis requested for the problem. For any data set you use, please provide a one paragraph description of the study and the variables you will be using.

   a. Test a multiple logistic regression model with at least two predictors in SPSS, R, or SAS. 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.

   b. Use two continuous predictors of a binary outcome test for an interaction in SPSS, R, or SAS and the Hayes PROCESS macro. 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 as well as the simple effects coefficients and the plot of the interaction if the interaction is significant (or marginally significant at $p < .10$).

   c. Select three variables from your data set, with one that can serve as a binary outcome variable, for which you can formulate a hypothesized mediational model. In SPSS, R, or SAS, use the Hayes PROCESS macro to conduct a bootstrap test of the indirect effect. Report and interpret your findings, including the regression coefficients for each of the direct effects, their odds ratios and confidence limits, and the test of the indirect effect.

5. Using the Early Head Start data set (child2.sav), test a mediational model to investigate whether welfare contacts mediate the relationship between neglect and abuse: neglect $\rightarrow$ welfare $\rightarrow$ abuse. Use SPSS, R, or SAS and the Hayes PROCESS macro to conduct a bootstrap test of the indirect effect. Report and interpret your findings, including the regression coefficients for each of the direct effects, their odds ratios and confidence limits, and the test of the indirect effect.