

Homework 2 Due 11/13/25 (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., proportions, chi-squared). Your answers should be **in your own words** and most answers should be approximately one paragraph. You may use **SPSS, R, or SAS** for any of the analyses.

For each of the problems below, I request a specific analysis for the data sets I have supplied, which are available at <http://web.pdx.edu/~newsomj/data.htm>. However, you also are **free to use your own data set** for any of the problems. If you use your data set, make sure the variable types match the analysis requested for the problem. **If using a different data set from the one I provide, please supply a one paragraph description** of the study and the variables you will be using.

1. Using the data from the Pew American Trends Panel survey from HW 1 (`pewtrend.sav`)¹, test a logistic regression model to investigate whether report of being unfairly stopped by police (`police`) predicts satisfaction in how the country is going (`country`). 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 Problems 2a and 2b) and briefly compare your results from HW 1 to your results from the logistic regression model (no more than 2-3 sentences are needed for this part of the answer). Provide specific values from each of the analyses when making your comparisons.

2. Using the same data set, test a logistic regression model predicting satisfaction (`country`) with an interaction between the gender (`gender`) and being unfairly stopped by police (`police`). 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 three-way contingency table in HW 1 (Problem 4a) and briefly compare your results from the logistic model to the Breslow-Day (no more than 2-3 sentences are needed for this part of the answer). Provide specific values from each of the analyses when making your comparisons.

3. The data for this problem come from a randomized experiment investigating the effects of Early Head Start (`child2.sav`), a preschool program for ages 0 to 3, on child maltreatment (conducted by Professor Beth Green at Portland State). There are several variables in this data set: the treatment vs. control variable (`program`), whether there was any report of abuse (`abuse`), whether or not the mother had a boyfriend who was not the biological father (`boyfriend`), 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. 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 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.

¹ <https://www.pewresearch.org/dataset/american-trends-panel-wave-86/>

- c. Using SPSS, R, or SAS, retest the model from 3a predicting `abuse` with `program`, `boyfriend`, `white`, and `welfare` as predictors (no need to repeat the write-up from the previous question). 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 large residuals, outliers for 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.
- d. Test a mediational model to investigate whether welfare contacts mediate the relationship between neglect and abuse: $\text{neglect} \rightarrow \text{welfare} \rightarrow \text{abuse}$. 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 confidence intervals for the indirect effect (including your conclusion about its significance).
4. For this problem, there is a new data set (`widow.sav`), taken from a study conducted by David Morgan and Margaret Neal that examined social relationships among recent widows, including a yes/no question about whether the respondent felt lonely was asked at two different time points (`lonely1`, `lonely2`), each six months apart, age of the respondent at Time 1 (`age1`) and years of education (`educ1`). 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. Test a lagged regression model to investigate whether income adequacy and health predict loneliness at Time 2 after controlling for loneliness at Time 1. 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.