

Assignment 3
Due April 30, 2009

For this assignment, you will utilize the General Social Survey dataset for 2002. The dataset is located at <http://www.sarkisian.net/sc704/gss2002.dta> and the codebook can be found at http://www.cpanda.org/data/a00079/ddi_cbk.html

- 1) Select a count variable and devise a theoretical argument explaining this variable using 4-6 independent variables. Some of the independent variables should be continuous, others should be dichotomous – this means that you might need to create a series of dummy variables if you want to use ordinal or multi-category nominal variables as your independent variables. Provide a brief description of this argument and your hypotheses.
- 2) Start a running log that will contain the commands and output for all of your models and diagnostics, with brief comments.
- 3) Conduct univariate and bivariate examination of your independent variables, consider potential transformations. Examine your count variable by tabulating it and graphically compare its distribution to Poisson distribution.
- 4) Fit four multivariate models: Poisson, negative binomial, zero-inflated Poisson and zero-inflated negative binomial. If applicable, adjust your models for exposure (if that is not possible or not necessary, make sure you explain why not). Compare their fit using prcounts and countfit commands and identify the best model for your data.
- 5) Do diagnostics for your model and attempt to remedy problems if you find some. Check for possible multicollinearity and error term distribution problems; examine nonlinearity and additivity problems, outliers and influential observations.
- 6) Respecify your model to include only statistically significant variables using test or lrtest command.
- 7) Interpret the results of the final model using incidence rate ratios, predicted rates (using prvalue, prtab) and predicted probabilities of specific counts (prvalue, prgen and graph twoway). Also, examine discrete and marginal changes in predicted rates (prchange and mfx compute). Write up your brief interpretation of the results.
- 8) When submitting the assignment to me, make sure that you include all the steps specified in items 1-7. There is no page limit for your assignment but please edit it to contain only the relevant commands and output, and include the relevant graphs as well (you can copy and paste them into your word processor).

Journal write-up component (optional):

Write up the results like you would for a journal publication. First, include an Introduction that will provide a short substantive description of your theoretical argument, your research questions and hypotheses (1 page max.). Second, include a brief Data and Methods section (1-2 pages) describing the variables and the analysis methodology. Include any discussion of diagnostics and modifications in this section, either in the text or in the footnotes. Also, include a table with summary statistics for the variables you use (means, standard deviations, number of observations). Third, provide a 1-3 page description of the results including a table (in journal format) and any graphs assisting in the interpretation of results (graphs can be useful if your model includes nonlinear relationships or interactions; do not include any diagnostic graphs). If it is absolutely necessary that you discuss some diagnostics or model modifications here rather than in the methods section, do so using footnotes. Finally, include a brief conclusion summarizing your findings and discussing contributions and limitations of your research (1 page max.). The page limit for this write-up is 7 pages double-spaced.

Assignment 3 Grading Sheet
Total Preliminary Grade: out of 100

- A. Model Construction (30 points):
 - 1. Variable construction and fitting the models (5 points):
 - 2. Comparative assessment of model fit (10 points):
 - 3. Selection among count models, including exposure issue (10 points):
 - 4. Model respecification (5 points):
- B. Diagnostics and remedies (15 points):
 - 1. Linearity and additivity (5 points):
 - 2. Outliers and influential data (5 points):
 - 3. Multicollinearity and error terms (5 points):
- C. Interpreting the Model (50 points)
 - 1. Coefficients and exponentiated coefficients (10 points):
 - 2. Predicted counts (10 points):
 - 3. Predicted probabilities (10 points):
 - 4. Graphical examination of predicted probabilities (10 points):
 - 5. Discrete and marginal changes (10 points):
- D. Log organization (5 points):

Journal-style Write-up Grading Sheet
Total Preliminary Grade: out of 100

- 1. Introduction (15 points):
- 2. Data and methods (30 points):
- 3. Tables and graphs (15 points):
- 4. Description of results (30 points):
- 5. Conclusion (10 points):