

Assignment 2
Due April 9, 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 categorical variable (dichotomous, ordinal, or multi-category nominal) and devise a theoretical argument explaining this categorical 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 variables, consider potential transformations. Then fit a logit model appropriate for your dependent variable – binary, ordered, or multinomial logit.
- 4) Assess goodness of fit for your model. If using ordered or multinomial logit model, test the appropriate assumptions: For ordered logit, investigate whether the parallel slopes assumption is violated, and if it is, fit a generalized ordered logit model. For multinomial logit, test whether IIA assumption holds, and if it does not, try to respecify the model to fix that problem.
- 5) Respecify your model to include only statistically significant variables--use test, fitstat, or mlogtest (for multinomial logit) to conduct hypothesis testing. Also, in multinomial logit, check if some outcomes can be combined using mlogtest, and if necessary, combine them.
- 6) Do diagnostics for your model and attempt to remedy problems if you find some. Examine linearity and additivity assumptions. Examine residuals to identify outliers and influential observations. Examine it for possible multicollinearity. Check the error term distribution by obtaining robust standard errors. (Note that for ordered and multinomial logit models, you have to estimate appropriate binary models to do nonlinearity diagnostics and check for outliers and influential observations). Attempt to remedy any problems that you find. For your final model (and the binary sub-models), assess model fit.
- 7) Interpret the results of your final model using odds ratios and percent changes in odds ratios. Examine predicted probabilities using prvalue command, as well as using prtab for categorical independent variables and prgen and graph twoway for continuous independent variables. Also, examine discrete and marginal changes in predicted probabilities. 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 2 Grading Sheet
Total Preliminary Grade: out of 100

- A. Model Construction (30 points):
 - 1. Theoretical argument (5 points)
 - 2. Variable construction and fitting the model (5 points):
 - 3. Model respecification (10 points):
 - 4. Assessment of model fit, incl. parallel slopes or IIA assumption (10 points):
- B. Diagnostics and remedies (25 points):
 - 1. Univariate and bivariate examination (5 points)
 - 2. Linearity and additivity (5 points):
 - 3. Outliers and influential data (5 points):
 - 4. Multicollinearity (5 points):
 - 5. Error term distribution and overdispersion assessment (5 points):
- C. Interpreting the Model (40 points):
 - 1. Coefficients and odds ratios (10 points)
 - 2. Predicted probabilities (10 points)
 - 3. Graphical examination of predicted probabilities and odds (10 points)
 - 4. 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)