

Assignment 2**Due Monday, December 1 for peer feedback**

For this assignment, you will utilize the General Social Survey dataset for 2002 (or a dataset of your choice; just make sure that the data are cross-sectional and not nested). The GSS 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 research question you can answer using this categorical variable as your dependent variable, with 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. Clearly state your research question as well as your hypotheses; include a brief justification for each hypothesis.

Start a running log that will contain the commands and output for all of your models and diagnostics, with brief comments. (In order to be able to do this, make sure you open a log file each time you open Stata to work on this assignment!)

2) Conduct univariate examination of your variables, consider potential transformations and deal with univariate outliers.

3) Investigate the bivariate relationships between the dependent and each of the continuous independent variables; evaluate linearity and consider potential transformations; identify bivariate outliers.

Fit a logit model appropriate for your dependent variable – binary, ordered, or multinomial logit (using the independent variables with those transformations and modifications you find appropriate). Then conduct the following diagnostics and attempt to remedy any problems that you find. This process can sometimes be circular as you may need to recheck things after applying new transformations etc.

- 4) Examine the model for potential nonlinearities (note that for ordered and multinomial logit models, you have to estimate appropriate binary models to do nonlinearity).
- 5) Consider potential 2-way interactions. If any of these interactions are statistically significant (or if they are important based on your theory, but are not significant – given that significance for interactions is tricky in logit), include them in your model.
- 6) Use hypothesis testing to evaluate whether you can omit non-significant variables, combine some categories of dummy variables (if relevant), and whether groups of dummies are jointly significant (also if relevant); try to make the model more parsimonious -- 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.
- 7) Check for possible multicollinearity problems using variance inflation factors.
- 8) Identify possible outlying and influential data (note that for ordered and multinomial logit models, you have to estimate appropriate binary models to check for outliers and influential observations).
- 9) Check the error term distribution by obtaining robust standard errors.

- 10) For your final model, assess goodness of fit. If using ordered logit model, also test whether the parallel slopes assumption is violated, and if it is, use a generalized ordered logit model. If using multinomial logit, discuss whether you think IIA assumption should hold for your model and why.
- 11) Write up your interpretation of the results of your final model using logit coefficients (for ordered logit, both unstandardized and standardized), odds ratios (both unstandardized and standardized), predicted probabilities, and discrete and marginal changes in probability. Whenever possible, use graphs to assist the interpretation (graphs can be particularly useful for multinomial and ordered logit model as well as whenever your model includes nonlinear relationships or interactions).
- 12) When submitting the assignment, make sure to include all the steps specified in items 1-11. 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).

Once your assignment is completed, please fill out and include the evaluation table that will contain a summary of what you did and will be used to evaluate your work, first by a peer and then by the instructor.