

Assignment 1**First draft is due March 12, 2009**

For this assignment, you will utilize a World Development Indicators (WDI) dataset for 2000, located at <http://www.sarkisian.net/sc704/datasets.html>. You can find more information about this dataset at <http://www.worldbank.org/data>.

- 1) Devise a theoretical argument that you can test using OLS regression on these data, with approximately 4-6 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. (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!)
- 3) Investigate the univariate distributions for each of the variables involved; consider normalizing transformations and deal with univariate outliers.
- 4) Investigate the bivariate relationships between the dependent and each of the independent variables; evaluate linearity and consider potential transformations.
- 5) Fit the OLS regression model (using the dependent and independent variables with those transformations and modifications you find appropriate).
- 6) Consider potential 2-way interactions. If any of these interactions are statistically significant, include them in your model.
- 7) Check for possible multicollinearity problems using correlations and variance inflation factors.
- 8) Examine the residuals for departures from normality.
- 9) Examine the model for potential nonlinearities.
- 10) Identify possible outlying and influential data.
- 11) Examine residuals for possible heteroscedasticity. If you detect problems, determine whether any of the independent variables might be responsible.
- 12) Attempt to remedy any problems that you find.
- 13) Write up your interpretation of the results of your final model. If necessary, use graphs to assist the interpretation. (Graphs can be useful if your model includes nonlinear relationships or interactions.)
- 14) When submitting the assignment to me, make sure that you include all the steps specified in items 1-13. 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 1 Grading Sheet
Total Preliminary Grade: out of 100

1. Theoretical argument (5 points)
2. Univariate examination of the data (10 points)
3. Bivariate examination of the data (10 points)
4. Non-normality (10 points)
5. Nonlinearity (10 points)
6. Interactions (10 points)
7. Multicollinearity (10 points)
8. Outliers and influential data (10 points)
9. Heteroscedasticity (10 points)
10. Interpretation of the results (10 points)
11. 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)