Stata Reproducibility Assignment

Clare Clingain

November 14, 2018

The goal of this assignment is to create an entirely reproducible research process. Your results should be replicable by anyone who receives your do-file(s) and documentation. For questions about this assignment, please reach out to Clare at clc586@nyu.edu

1 Scenario

Imagine that you have been approached by the Center for Disease Control (CDC) to run statistical analyses on the National Health and Nutrition Examination Survey II (NHANES II), and to compile a report of your findings in table format. The CDC requests that the research process be reproducible so that other researchers can replicate your results and any critics about table results can be more easily shut down.

2 Data

You will use data from the Center for Disease Control's National Health and Nutrition Examination Survey II (NHANES II) for this assignment. The dataset can be loaded into Stata using the following code: *use "http://www.stata-press.com/data/r14/nhanes2.dta", clear*. For the purposes of this assignment, **you are not required to use the survey weights**. Note that if you do choose to use the survey weights, you will need to use Stata's *svy, subpop():* commands to run statistical analyses.

3 Tasks

You must complete the following tasks:

1. Subset your data based on one demographic feature (eg. gender, race, region, etc.)
2. Create and/or recode 3 variables
3. Run one descriptive analysis of at least 4 variables
4. Export a descriptive table (via estout, putexcel, texdoc – your choice)
5. Run two predictive analyses: either a nested regression or two models with same predictors, different outcomes

6. Create reproducible code that will run your regressions and export the results into a table shell using putexcel

4 Submission

Please submit the following documents via NYU Classes:

1. A **clean and readable** do-file containing all your code
2. A 1-2 page PDF write-up of your research process containing your final tables
3. An Excel file with your table shell