SOCY7706: Longitudinal Data Analysis
Instructor: Natasha Sarkisian
Assignment 1: Data Management and Description
Due: September 28, 2022 at 11:59pm

For this assignment, you will conduct data management, obtain descriptive statistics, and write a brief interpretation of your findings. You will submit your do file and your annotated log that will contain the output (with your brief comments) for all of the tasks that you will perform for this assignment. Make sure to provide brief comments throughout your do-file and log file that will clarify your steps and decisions. There is no page limit for your annotated log but please edit it to contain only the relevant syntax and output (i.e., omit any unproductive steps).

1) Read the following article and use it as a guide to prepare a dataset for a study that will explore the potentially bidirectional relationship between one personality trait and physical health.

Turiano, Nicholas A., Lindsay Pitzer, Cherie Armour, Arun Karlamangla, Carol D. Ryff, Daniel K. Mroczek. 2012. Personality Trait Level and Change as Predictors of Health Outcomes: Findings From a National Study of Americans (MIDUS). Journals of Gerontology Series B: Psychological Sciences and Social Sciences 67B (1): 4-12. DOI: https://doi.org/10.1093/geronb/gbr072

2) Download the data and the codebooks from ICPSR website and select the variables you need (make sure to keep all ID variables in the dataset). That will include the measure of the personality trait you selected as well as two or more items measuring physical health in order to create a scale so that the resulting variable can be treated as continuous (make sure to standardize items using std function of egen command prior to adding them together). Also, select some control variables for your models (you can use the ones used in the article or select your own).

3) Merge datasets across waves using two approaches: first, using the append command, and second, using the merge command (make sure to adjust the variable names accordingly). Reshape one of the two datasets so that they are both in long format, drop the empty rows, and compare basic summary statistics of the two versions of the datasets.

4) Using one of the two datasets, recode the variables for your analysis.

5) Using either wide or long format, generate descriptive statistics separately for each wave, and evaluate univariate normality and outliers for your continuous variables.

6) Using long format (and making sure you dropped the empty rows), xtset the data, and examine the amount of change over time in your key variables by using xtsum, xttab, and xttrans.

7) Write up an interpretation of both sets of results. Discuss which of these statistics you would choose for presentation in a journal article and why.

Please submit your do-file and your annotated log file (including all graphs and the table) electronically (by email or using a file transfer website of your choice).