Assignment 10
LPO 9951 | Fall 2015

Data validation, like data cleaning, is a rather unclear term that refers to a set of techniques to ensure that the characteristics of your dataset conform with known characteristics of the same sample or a similar sample.

The reason for validating your data is simple: you don't want to be wrong. This is a fundamental and worthwhile step to take in putting together an analysis.

For this assignment, choose five variables from your dataset. You should definitely include your dependent variable or some variant as part of this set. Validate the estimates that you have of the mean and the standard error of the mean from another source. Your best bet is one of the published NCES reports. Failing that, you will need to find another published source that includes estimates of your particular variables.

Create a table showing your means (or frequencies) and s.e.s for the five variables, along with the published results. If there is a discrepancy, speculate as to the cause.