

**Exercise 1: Importing Data**

Let's practice importing different types of datafiles. The first thing you will want to do after you open Stata is open a Do-File. You can do this by clicking on the Do-File Icon or by clicking on "Window" and then "Do-File Editor." As a general rule, always open STATA first – don't try to access STATA by opening your saved Do-File because once you attempt to open your Do-File, STATA will automatically rerun the file, which could create problems. Remember clear and stop currently running Stata processes and to tell Stata how much memory to use when you begin the session.

1. Try opening the following files:
  - a. A comma separated value file: gss.csv
  - b. A SAS transport file: gss.xpt

**Exercise 2: Variable Labels and Value Labels**

Open the data set, gss.csv

1. Take a look at your data using one of the data review commands we discussed.
2. Rename your variables and add variable names using the following codebook:
  - v1, marital, marital status
  - v2, age, age of respondent
  - v3, educ, education
  - v4, sex, respondent's sex
  - v5, inc, household income
  - v6, happy, general happiness
  - v7, region, region of interview
3. Add value labels to your "marital" variable using the following codebook:
  - 1 "married"
  - 2 "widowed"
  - 3 "divorced"
  - 4 "separated"
  - 5 "never married"

**Exercise 3: Manipulating Variables**

Use the dataset, gss.dta

1. Generate a new variable, age<sup>2</sup>
2. Generate a new "high income" variable that will take on a value of "1" if a person has an income value greater than "15" and "0" otherwise
3. Generate a new divorced/separated dummy variable that will take on a value of "1" if a person is either divorced or separated and "0" otherwise

**Exercise 4: Descriptive statistics**

Use the dataset, gss.dta

1. Examine a few selected variables using the describe, sum and codebook commands
2. Tabulate the variable, "marital," with and without labels
3. Cross-tabulate marital with region and show gender percent by region
4. Summarize the variable, "income" separately participants based on marital status
5. Summarize the variable, "happy" for married individuals only
6. Generate a histogram of income
7. Generate a second histogram of income, but this time, split income based on participants' sex and ask Stata to print the normal curve on your histograms