

**Exercise 1: Gen, Egen, Recode & Replace**

Open the datafile, gssDM.dta.

1. Generate a new variable that represents the squared value of an individual's age.
2. Recode values "99" and "98" on the variable, "hrs1" so that they are identified as "missing."
3. Generate a new dichotomous variable that has a value of "1" if an individual's income is greater than "19."
4. Recode the marital variable into a "string" variable and then code it back into a numeric variable.
5. Create a new variable that counts the number of "yes" responses that individuals provided in regard to their compute use (usecomp) email use (usemail) and internet use (usenet).
6. Create a new variable that counts the number of missing responses for each respondent.
7. Create a new variable that associates each individual with the average number of hours worked among individuals with matching educational degrees.
8. Use Stata's help command to look up the "egen" command. Look over the various egen options and try one that would be useful to you in your own work.

**Exercise 2: Merge, Append, and Joinby**

Open the dataset, gss2.dta

1. The gss2 dataset contains only half of the variables that are in the complete gss dataset. To create a single dataset containing all the variables, merge dataset gss1 with dataset gss2. The identification variable is "id."

Open the dataset, gssDM.dta

1. Merge in data from the "marital.dta" dataset, which includes income information grouped by individuals' marital status. The marital dataset contains collapsed data regarding average statistics of individuals based on their marital status. Tell Stata that you'd like data from your marital dataset to replace data in your gss dataset when there are conflicts.
2. Additional observations for the gssDM.dta dataset can be found in "gssAddObserve.dta." Create a new dataset that combines the observations in gssDM.dta with the observations in gssAddObserve.dta.
3. Create a new dataset that summarizes mean, median, and standard deviation of income based on individuals' degree status ("degree"). In the process of creating this new dataset, rename your three new variables. Don't forget to consider how to handle missing data on the "degree" variable before you begin.

**Exercise 3: Comparing Datasets**

Open the dataset, gssDM.dta

1. Compare the gss1.dta dataset to the gssDM.dta dataset. What are the differences?
2. Compare marital.dta to gssAddObserve.dta. What are the differences?
3. Open the dataset, gsscompare1.dta. Compare the values in this dataset to those in gssDM.dta.
4. Ask Stata to identify exactly where the differences lie between gsscompare1.dta and gssDM.dta.