Exercise: Introduction for beginners

Task 1
Prepare your dofile and load the built-in dataset nlsw88.

Task 2
Familiarize yourself more with the following basic commands and their options using the helpfiles: describe, summarize, tabulate, count, graph, histogram, generate, egen, keep, drop, isid, label, display

Task 3
Install the user-written command “fre” and check its helpfile

Task 4
Get an overview of the data by checking:

- Whether there is a variable that uniquely identifies each observation
- How many college graduates are in the data, what is their share in the data
- Whether there are missing values for union membership
- What mean, minimum, maximum and median age are
- The age histogram
- How many married college graduates live in the south

Task 5
Modify the dataset by:

- Excluding individuals aged below 35 and above 45, check using a histogram whether this was successful
- Creating a new indicator variable for individuals working in sales who are in a union (make sure that the new variable is missing if either occupation or union membership are missing), label variable and values accordingly
- Creating a new variable capturing the weekly wage of each individual
- Creating a new variable that stores the mean wage for each occupation group
- Creating a new variable that captures the difference between the own and the industry’s mean wage
- Creating a new variable that captures the difference between the own and the overall mean wage (using the results stored in the return list)

Task 6
Analyze

- Display in a horizontal bar graph the mean hourly wage of each occupation group, sort them according to wage (highest on top), give the graph a title and change the color of the bars to green (use the help file to find the correct options)
- Use a t-test to see whether the mean hourly wage of unionized and non-unionized workers differs and then display again the mean of each group drawing the result from the return list