Dis 10: Stata Q&A

Check out the solution from Dis 4 (Stata Review) when doing your Stata problem set.
You can use that discussion's solution and this as a guide when completing your Stata problem set.

1 Stata Exercise

Before we start the Q&A part of this section, let's look at one Stata exercise together. This exercise uses the same dataset given in your Stata Assignment (that's due Apr 30 @ 11pm on Canvas).

You are given a dataset that is created by a collaboration between the Robert Wood Johnson Foundation and the University of Wisconsin Population Health Institute.

The dataset contains variables from the 2019 Wisconsin County Health Rankings.

1. First thing first, let's make sure we can keep track of your work and your results.
   (a) To keep track of your results, clear out the Results panel.
   (b) To keep track of your work / code, create a Do-file with the first line as a comment, and put your name down there.
   (c) Set up the Do-file environment before proceeding.

2. The dataset has been given in `.csv` and `.dta`. The files are on Canvas under the following names:

   Econ 310 Stata Assignment Data.csv
   Econ 310 Stata Assignment Data.dta

   Load the data into Stata environment.

3. Describe your data using the **describe** command.
4. Browse the variable related to Percent Unemployed only.
5. Create a histogram of Percent Unemployed, and save it as a PNG file called "q5_histogram.png". Is the distribution of Percent Unemployed skewed in any way?

6. Calculate the 93% confidence interval for mean of Percent Unemployed.

7. What's the mean, variance, and the 10th percentile of Percent Unemployed?

8. Create a scatter plot between Percent Unemployed and Percent Some College for observations where Percent Some College is above 50.
9. What is the correlation between Percent Unemployed and Percent Some College for observations where Percent Some College is above 50? Interpret the correlation measure.

10. Export your Stata output result as a PDF file.

2 A Type-II Error Question from Exam 3

1. The temperature (measured in Fahrenheit) in May in Madison is distributed with standard deviation of 10 F. Suppose that a sample of 31 random days in May over the years has been collected through simple random sampling, and the average May temperature in Madison in your sample is found to be 50.72 F.

   In the first two parts of this question, we tested the following hypothesis:

$$H_0 : \mu = 50$$
$$H_1 : \mu > 50$$

   (c) Suppose that the true population average of May temperature in Madison is 52.52 F. What's the probability of committing type-II error for the test you just conducted under 10% significance level?