Dis 4: Stata Review

Stata handout offered by Dr. Pac can be accessed here: Handout
This handout reorganizes the official handout's information, and adds more Stata resources.

1 A Probability Question from Exam 1

Since our section today is mainly introducing you to Stata, which shouldn't take up the entire duration of the section, let's first go through one fairly challenging question on your first midterm together.

- A UW Madison professor is interested in learning about and detecting the effects of long-haul Covid on individuals' health. The professor examines 10,000 long-haul Covid cases (patients) and learns that 1500 of the cases had serious illness due to long-haul Covid and the other 8500 had no health impact from long-haul Covid. The UW professor develops a new test which is accurate 80% of time in detecting whether long-haul Covid will lead to a serious illness or whether it will have no impact on the patient's health.

  The professor administers the test on a new patient with long-haul Covid. What is the probability that the patient has serious illness due to long-haul Covid if she tests negative using the professor's test?

2 Stata Overview

- **What is Stata?**
  - Stata is a software that allows you to analyze data statistically. You can think about it as the advanced version of Microsoft Excel.
  - Some comparable software / programming languages out there that can do what Stata does (and maybe some more) include SAS, SPSS, R, Python, MATLAB, Julia.
    * Because Stata is currently the most popular statistical package amongst economists, it is what we'll be learning and using for this class.
    * If you go on to take Econ 400 or 410, you will continue to use Stata in that class as well.
    * The Social Science Computing Cooperative (SSCC) here at UW-Madison offers some training classes in Stata, R, and Python. If you're interested in these classes, you can find more information on this website: https://sscc.wisc.edu/sscc_jsp/training/
- **How to access Stata?**

  You can access Stata using either one of the two following methods:

  - **Installing it onto your personal laptop (recommended)**:

    Visit UW Software Library (software.wisc.edu) for installation guide and license & activation key. The version of Stata to install is **Stata/SE**.
  - **Logging into Winstat (i.e. a remote server; great alternative for people with Chromebook or using unsupported OS)**:

    Check out the following link for information on logging into Winstat: https://kb.wisc.edu/sscc/using-winstat

3 Get Started Using Stata

Let's go through the following steps to get you started on using Stata.

1. Launch Stata, and let's go through how the Stata program looks like. Specifically, identify
   - where results show up
   - where can you find the list of variables
   - where can you find more information about the variables
   - where to run your commands (via either the Command panel or the Do-file editor)
2. Before running any commands, let's set up the working directory to tell Stata which folder on your laptop should Stata read data and save graphs or results to. The easiest way to do so is to go to the menu bar, and select

   ```
   File > Change working directory...
   ```

3. Just to make sure your name is somewhere in your results, use the display command to write your name in the log. For example, assuming you happen to be Lindsey Lohan, you would type the following command:

```
display "Lindsey Lohan"
```

4. Let's now load a set of data to do some simple statistical analysis. On your Stata problem set, we tell you exactly what command you should use to load the appropriate dataset. But for today, let's just load a sample dataset known as auto:

   ```
   sysuse auto
   ```

5. Use the describe command to determine which variable in this dataset contains "Price" and which contains "Trunk space in cubic feet":

   ```
   describe
   ```

6. Use the histogram command to graph a histogram of price and assess whether the distribution is symmetric or skewed:

   ```
   histogram price
   ```

7. Save the histogram created by clicking on the "Save" button in the graph window. Make sure you save the graph as a `.png` file.
8. Use the summarize command to calculate the mean, median, and standard deviation of trunk space:

   ```
   summarize trunk
   ```

9. Did the previous command have all the information you needed? If not, let's now try the same command with the detail option:

   ```
   summarize trunk, detail
   ```

10. Use the correlate command to calculate the correlation coefficient between price and trunk space:

    ```
    correlate price trunk
    ```

11. Use the scatter command to graph a scatterplot of price (on the y-axis) and trunk space (on the x-axis) and assess the relationship between the two variables (note that the order of variables in the following command matters):

    ```
    scatter price trunk
    ```

12. We've now finished running all practice commands. To save the output printed in the Results panel, go to

    ```
    File > Print > Results
    ```

    and then save your Stata output as a PDF document using your OS's printing dialogue.

4 Some More Commands For Future Use

The previous section covered all the Stata commands for the statistical operations that we have learned so far in class. There are some other Stata commands that could be helpful for your Stata problem set due around the end of the semester (April 30 at 11pm). They are listed in here for you to reference to when doing your Stata problem set.

1. Use the ci command to calculate a 95% confidence interval for price:

   ```
   ci means price, level(95)
   ```

2. Use the ttest command to test whether the population mean of trunk space is equal to 13 using a 10% size of test:

   ```
   ttest trunk=13, level(90)
   ```

3. Use the ttest command to test whether the population mean of trunk space is greater than 13 using a 10% size of test. Note: Will this command differ from the one above? Why or why not?

4. Use the regress command to run a regression of price on trunk space (this language means price should be the dependent variable while trunk should be the explanatory variable, so the order of variables in the following command matters):

   ```
   regress price trunk
   ```