

Economics 472

Lab Outline

1. Go to the course website <http://www.econ.iastate.edu/classes/econ472/Tobias/> and click on the link for the California Test Score data in text format. You might also want to take a look at the README file which explains this data in greater detail.

When you load the data, it will appear in a new window. Click on **File** and then **Save Page As**. In the window that opens, it should then give the filename *caschool* and then save as a text document. *Be sure to direct this file to your disk in the a: drive!*

2. From the Start menu, load the STATA 7 program (intercooled Stata 7).

3. In the command window, type in

log using a:lab, replace t

What this does is create a log file, titled lab.log, in text format, which will contain all of the results of your subsequent commands.

4. Now, we are going to load in our data. To load data in text format into STATA, we need to use the **infile** command. After the infile command, we need to assign names to our variables. **Be sure that the names are assigned to the variables in the correct order!**

We are going to load the California test score data into STATA. This is a reasonably large data set containing 420 observations on 11 variables: (NOTE: I have cleaned the raw data and omitted the first 6 variables from *DISTCODE* to *GRSPAN*, as they are not crucial to our analysis).

To load these 11 variables into data, simply type:

```
infile enr1tot teachers calwpct mealpct computer testscr compstr ex-  
penstu str avginc elpct using a:caschool.txt
```

5. To get some basic summary statistics regarding this data, type:

summarize

This will produce a table of means, standard deviations, etc.

6. We will now try and replicate the results of textbook where the test score variable, *testscr*, is regressed on student teacher ratio, *str*. That is, we consider the model:

$$TestScore_i = \beta_1 + \beta_2 STRatio + u_i.$$

Stock and Watson, page 99, claim the following estimates are obtained:

$$\widehat{TestScore} = 698.9 - 2.28STRatio.$$

To perform this regression in STATA, simply type

regress testscr str

When invoking this command, the first variable appearing after the “regress” statement must *always* be the *dependent* variable. The variables that appear after the first variable are assumed to be the set of *independent* or *explanatory* variables.

You will then see a bunch of output appear on your screen. **Do you get the same regression results? What is your interpretation of these results?**

7. To obtain a *graph* of our estimated regression line against the scatterplot of data, we must take a few steps. First, type:

predict yhat

This will generate a set of fitted (or predicted) values for every observation in the sample based on the regression estimates. Then type

graph testscr yhat str, connect(.s) symbol(oi)

The dependent variable should always appear first following the graph command. The option `connect(.s)` tells STATA not to connect the first series (*testscr*), but to connect the second series (*yhat*) smoothly. The option `symbol(oi)` tells STATA that *testscr* should be plotted with large circles (*o*) while the second option (*i*) says to plot the *yhat* values invisibly.

8. When you are done, be sure to close your log file using the **log close**

command. Finally, to clear your variables (and thus potentially to do more analysis using a different data set) simply type in **clear**.

Creating Do Files: The problem with the interactive approach outlined above is that it is difficult to fix errors when they occur or to add additional steps to the program. To this end, it is useful to create “do” files. Simply locate the do-file editor window, type in your commands, and then save the file. Let’s say you save the file as ps1.do on your a: drive. Then, to execute the file, simply go to the command window in STATA and type in **do a:ps1.do** The program will then be executed. If errors occur, the error messages will usually appear in red. Clear STATA of the variables, fix the error and re-execute the file.