

INTRODUCTION TO STATA FOR STATISTICS 110/201

Let's look at a sample session for doing Problem 1.19 on page 35. The data are in a file on the CD accompanying the book. The file is called CH01PR19.txt and has two columns of data, one row per person, with Y_i = freshman GPA listed first, then X_i = ACT score.

STEP 1: The first thing you need to do after you open Stata is create a log file, so you can later cut and paste the results you want into a Word (or other) file. For most actions in Stata you can either give a command, or use the pull down menus or icons on the tool bar. To create a log file it's easiest to use the pull down menu **File -> log -> Begin**. You will be given two choices for type of file. Choose the .log extension, or you won't be able to edit it in your word processor! Save the file using a name of your choice. Everything you type and Stata's response will be logged in this file, EXCEPT graphs. You will need to save and/or print those separately.

STEP 2: The next thing you need to do is read in the data. The easiest method again is the file menu: **File -> Import -> Unformatted ASCII Data**. A dialog box will pop up. Browse for the data set (you will need to have inserted the CD, or saved the data on a flash drive) and enter that name in the first box. Make sure you are looking at files of type ".txt" (not ".raw") or the file won't show up. In the "New variable names:" box, type the name you want for GPA first, then for ACT (that's their order in the file.) Let's use GPA and ACT. Just separate them with spaces.

Comment: Notice that there are 4 windows showing. The most prominent is the big black **Results window**. All results from your commands, except graphs, appear in this window. Below that, on the bottom right, is a small white space called the **Command window**. This is where you enter commands, if you prefer to do that rather than use the pull down menus. On the top left is the **Review window**. Here you see the commands that you typed, or that you could have typed to perform the same actions you performed using the menus. Watching these commands is a great way to learn Stata commands! For instance, notice that we could have read the data file using the command *infile GPA ACT using "CH01PR19.txt"* If you double-click on a command in the review window, Stata will repeat that command. The final window showing is the **Variables window**. Here you can see all the variables in the data set you are using, whether you read them in or created them. (More on creating variables later.)

STEP 3: Part a of Problem 1.19 asks for the least squares estimates (b_0 and b_1) of the coefficients (β_0 and β_1), and for the estimated regression function. Using the menus: **Statistics -> Linear Models and Related -> Linear Regression**. In the dialog box, specify GPA as the dependent (Y) variable and ACT as the independent (X) variable. For now, you can ignore the rest of the dialog box. Notice that the corresponding command, which now appears in the "Review window" is easy: *regress GPA ACT* and we could have entered that in the "Command window" in the bottom right instead of using the menu.

STEP 4: Find what you need in the output in the Results window. Here is where to find the coefficients:

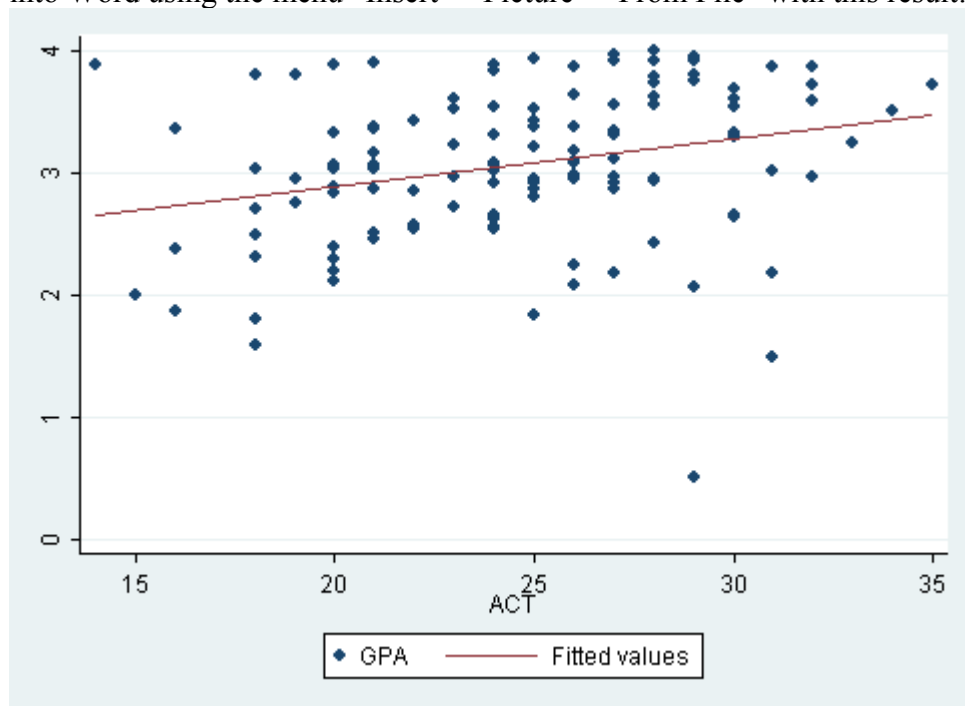
GPA	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
ACT	.0388271	.012773	3.04	0.003	.0135331 .0641212
_cons	2.114049	.3208948	6.59	0.000	1.47859 2.749508

where I have put them in bold for you to see. The "**_cons**" is b_0 and the "Coef" in the "ACT" row is b_1 . So the estimated regression function (with some rounding) is: $\hat{Y} = 2.114 + 0.0388X$. You will learn what all of the rest of the output means soon! Right now, the only other thing you have learned is $s = \text{Root MSE} = .62313$. That's the estimate for σ , the population standard deviation of GPA scores at *each* ACT value.

STEP 5: Part b of Problem 1.19 asks for a plot of the regression function AND the data, on one plot. While we're still learning commands, let's use the menu again. Later, you will find it easier to use a command to do this kind of plot. Basically, we need to ask Stata for two plots on top of each other. Here's how to do it using the menus:

Graphics -> Two-way graph (Scatter, line, etc) When the dialog box appears, left-click **Create**, then for the first plot (the actual data points), leave the default radio button **Basic plots** and the default type of plot, **Scatter**. Choose GPA for the **Y variable** and ACT for the **X variable**. Click on the **Accept** button (not the "Submit" button). We're not ready to submit the request yet, because we need to add the line! You should now be back to the original Two way graphs dialog box. Click on **Create** again. This time, when "Plot 2" dialog box pops up, click the radio button labeled **Fit plots**. Then highlight **Linear prediction** and again chose GPA for the **Y variable** and ACT for the **X variable**. Click either **Accept** or **Submit**. If you click "Accept" you will go back to the Two way graph dialog box. You will see Plot 1 and Plot 2 listed. Click **Submit** and the graph should pop up in a new window. Before you "Submit" you can also add labels to the axes, and so on. Notice that we could have used the command *tway (scatter GPA ACT) (lfit GPA ACT)* to produce this graph.

STEP 6: You need to save or print this graph. You can print it directly from the graph window itself, using **File -> Print** or clicking on the printer icon. Or, you can save it using **File -> Save** to print later. You can also modify it using **File -> Start Graph Editor**. I saved the graph as .png, and then inserted it into Word using the menu "Insert -> Picture -> From File" with this result:



STEP 7: Part d asks for a point estimate for mean GPA when ACT = 30. Here is one inelegant way:

```
generate pred30 = _b[_cons] + _b[ACT]*30
list pred30 in 1
```

[The result is given as 3.278863.]

STEP 8: Exit Stata ("Exit" on the file menu)

NOTE: When you edit your log file to insert results into your homework, to format Stata output correctly you need to put it into Courier font, size 10 or smaller.

This is a brief introduction. There is a link to a very helpful UCLA website on the class webpage, that can answer many questions you might have: <http://www.ats.ucla.edu/stat/stata/>