

Lab 1, Getting Started in Stata

In today's lab we will use PM2.5 data from Beijing collected hourly by the U.S. Embassy on July 24-25, 2012. These data were made available on Twitter by the U.S. Embassy, and China has protested this practice (see <http://www.foxnews.com/world/2012/06/05/china-tells-us-to-stop-reporting-beijing-bad-air/>).

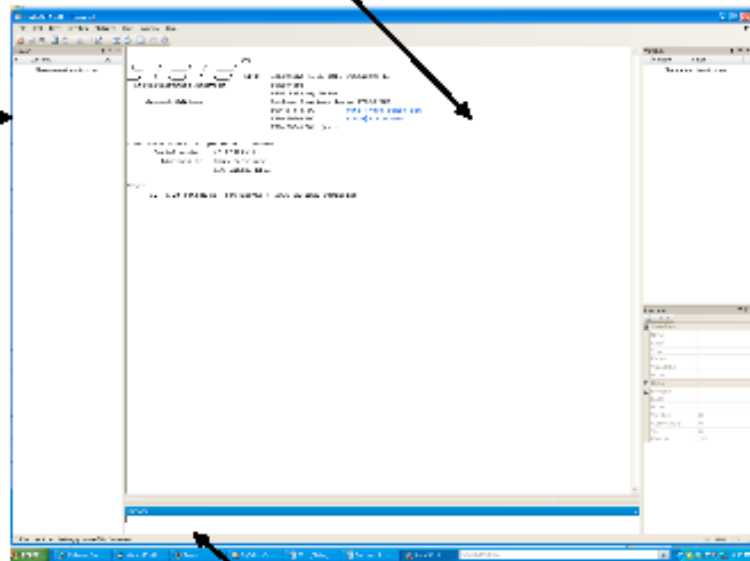
If you are not using Stata on your own machine, first you need to log in to the Virtual Computing Lab (VCL) at <https://vcl.unc.edu>, choose Shibboleth (UNC-Chapel Hill) and proceed to login, click on New Reservation, choose 'Stata 12, WinXP (vmware)' from the pull-down list of environments, and click Create Reservation, click connect (and then get RDP file if needed). Note that your next password will be the one provided on the screen and NOT your usual onyen password.

1. Open an internet browser on the VCL interface and download the Stata file chinapollution.dta from the Resources-Labs section of the Sakai website for BIOS 600.
2. Open Stata by double clicking the Stata icon
3. Familiarize yourself with the Stata windows, particularly the Variables, Command, and Results windows

There are usually five windows to use – adjust them with your mouse.

Results window: results are displayed here.

Review Window: past commands appear here – click on command in the review window, and it will appear in the command window – double click on command here and it will execute.

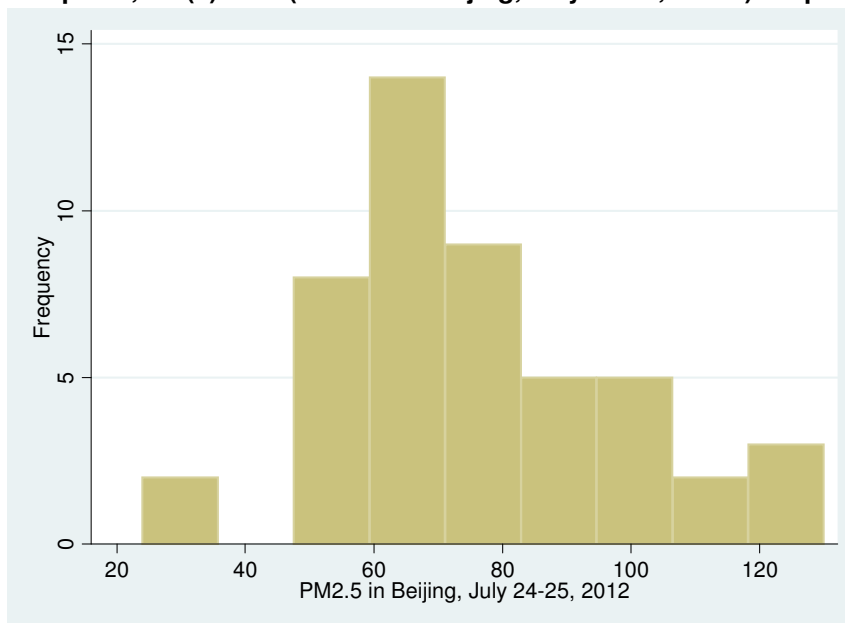


Variables window: variable list appears here. Double click on variable and it will appear in the command window.

Properties window: variable properties and dataset properties appear here.

Command window: commands are typed here – the general format is `<command>`, `<option(s)>`. To execute a command, press enter.

4. Open a Stata log file by clicking File, then Log, then Begin (and name it somewhere you can find it) . If you forget to open your log file, you lose any record of what you did!
5. Open the Stata data file
 - a. Open or click the open file button
 - b. Find the file chinapollution.dta and double click
 - c. This opens the data file in Stata
6. Click Data, then Data Browser to take a look at the data (safer than the Data Editor, which allows you to make changes to the data)
7. Create histogram of the hourly PM2.5 measures in Beijing on July 24-25, 2012 by typing
hist pm25, bin(9) xtitle("PM2.5 in Beijing, July 24-25, 2012") freq



8. Look at some descriptive statistics for PM2.5 by typing
summarize pm25, detail
9. Check on your log file to be sure it is working by going to
Go to File, then Log, then View
10. Create a boxplot by typing
graph box pm25
11. Look at the health indices by typing
tabulate healthindex

12. Generate a new variable that indicates whether the PM2.5 levels are above the US short-term standard of $35\mu\text{g}/\text{m}^3$ by typing

```
generate badpm=0  
replace badpm=1 if pm25>35
```

13. Look at this variable to determine the proportion of measures above the short-term US standard

```
tabulate badpm
```

14. Label this variable (badpm is not too sophisticated!)

```
label variable badpm "Exceeds US short-term standard"
```

15. Label the values of badpm so you do not forget whether 0 or 1 indicates the standard has been exceeded

```
label define yesno 1 "yes" 0 "no"  
label value badpm yesno
```

16. Look at this variable again to see how the labels have improved the output

```
tabulate badpm
```

17. Close your Stata log file by typing

```
log close
```

18. Open your log file to be sure everything was saved

```
Go to File, then Log, then View
```