Econ 174, Section 101/103
Week 4

Joshua Blumenstock

jblumenstock@berkeley.edu

Please Note: I am aware

If you have a laptop with STATA installed:

- Turn it on and load stata
- Type “sysuse auto, replace” to load some data

If you don’t have a laptop:

- Find someone who does, and kindly ask him/her to share

Today

- Quick review of last class
- Questions?
- Problem set logistics
- Stata exercises

Problem set

- Turn in printed copy @ beginning of lecture
- Submit electronic copy via bSpace
- Three files
  - Your solutions – name, section, group members
  - Your .do file – does everyone know this?
  - Your .log file

A model .do file

```
clear
set mem 100M
set matsize 800
set more off

cd "C:\Users\josh\Documents\My Dropbox\Classes\Econ 174\Assignments\practice"
log using practice.log, replace

****************************************
*
*   Problem Set: STATA practice
*   Freddie Freeloader
*   GSI: Joshua Blumenstock
*   Group members: Harry Redknapp, Gareth Bale
*
****************************************

****************************************
*  Question 1
****************************************
use "practice.dta"


****************************************
*  Question 2
* there are 200 observations and 8 variables
****************************************
count
```

How to find help with STATA?

- type `help command`, e.g. `help sort`
- Use google, e.g. “stata sort random”

STATA Exercise

- `sysuse auto, replace`
- Suppose the cost of manufacturing a car is the sum of the following:
  - $1.50 per pound of weight
  - $0.25 per pound to ship if it is foreign
  - $100 if its rep78 is 5 (presumably to hire better engineers)
  - $50 if its mpg is greater than 25 (better engineers again)
- Calculate the profit (price minus cost) from selling each car.
  - *What is the average profit for the cars in the dataset?*
- Generate a dummy variable called “badidea" that equals 1 if the car is not profitable to produce
  - What is wrong with this command?
  - `gen badidea if profit<0`

STATA Exercise

- Our dataset is too big, we want a random subsample

- Type `preserve` – to create a restore point
- Drop 10 observations from the dataset, ***at random***
  - Hint: create a new variable with a random number (runiform()), then sort by it
- Type `restore` – to restore to the point at which you typed preserve

STATA Exercise

1. Which cars have the lowest and highest values of mpg?
2. Do foreign cars have the same mpg as non-foreign cars? What is the p-value associated with this difference? Use `regress` and/or `ttest`
3. Are foreign cars more expensive than domestic cars? More profitable?
4. Rerun the regression in q2, but include a number of control variables (you choose). How do the controls affect the point estimate? The standard error? The R-squared? What does all this mean?
5. Test to see if there is an interaction between being foreign cars and the gear ratio, on mpg. Is there a differential effect of the gear ratio on mpg for foreign and domestic cars?

STATA Exercise

- Create a new variable rank that gives the relative price of the car: Use `sort, gen rank = _n`
- Create the variable `cost_quartile:`

  =1 if cost in first quartile (0-25%), 2 if (26-50%), etc…

  Hint: `xtile`
- Compute average mpg for each cost quartile, and save this in a new variable `quartile_mpg`
  - Hint: `bysort, egen`
- Plot average mpg by cost quartiles
  - Hint: `scatter`

STATA Exercises

- Plot the relationship between mpg and price
  - Draw the regression line and add a title to the plot. Use + signs for point markers, use a dashed yellow regression line.
  - Is the relationship statistically significant?
  - Bonus: label the points with the car model
  - Bonus: add confidence intervals

Some links

- http://data.princeton.edu/stata/graphics.html
- http://www.ssc.wisc.edu/sscc/pubs/sfr-data.htm