Handout #1: Introduction to STATA

17.846 Multivariate Political Analysis (S98)
Micky Tripathi
E40-380
x32736
mtripath@mit.edu

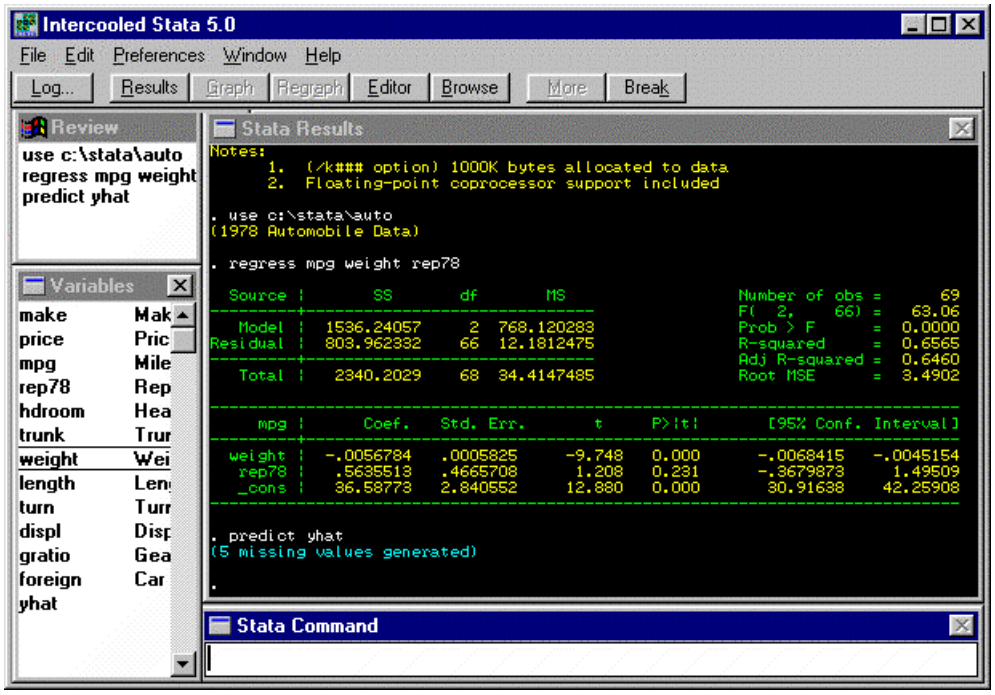
Getting in and out of STATA

- windows

```
run:  stata [\k10000]
```

- unix

```
athena%  add stata_v5.0
athena%  stata [-k10000]
```



- Command syntax -- *command* [*varlist*] [***if*** exp] [***in*** *range*] [, *options*]

```
. use \stata\auto, clear
```

```
. help command
```

I've got the data: now what?

- flat files

```
----begin file------

"AMC Concord" 22 1977
"AMC Concord" 22 1978
"AMC Pacer" 17 1977
"AMC Pacer" 18 1978
"AMC Spirit" 22 1977
"AMC Spirit" 23 1978

-----end file-----
```

- *input*

```
. input str20 make mpg year

       make  mpg  year
1.

. "AMC Concord" 22 1977
. blah blah blah
. end
```

- *infile* with dictionary and data in same file

```
----begin file------

dictionary {
      str20       make
      int         mpg
      int         year
}
"AMC Concord" 22 1977
"AMC Concord" 22 1978
blah blah blah

-----end file-----

. infile using dict.raw
```

- *infile* with dictionary (dedicated) and data in different files

```
----begin file------

dictionary using auto.raw {
      str20       make
      int         mpg
      int         year
}

-----end file-----


. infile using dict.raw
```

- *infile* with dictionary (generic) and data in different files

```
----begin file------

dictionary {
      str20       make
      int         mpg
      int         year
}

-----end file-----

. infile using dict.raw, using(auto.raw)
```

<u>The data's in: now what?</u>

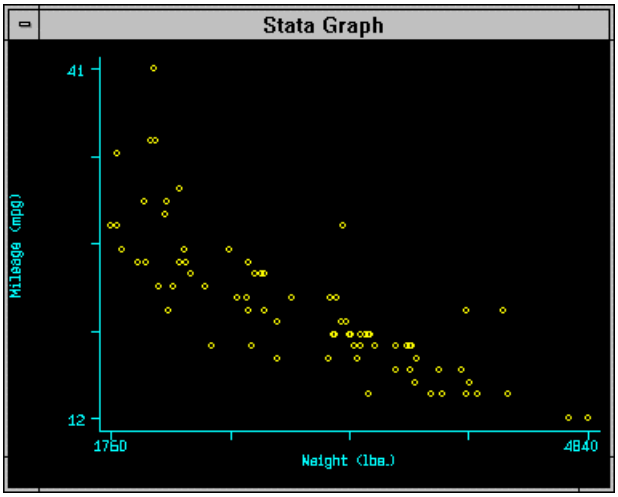
- Describing the data (*list, describe, summarize, tabulate, graph*)

```
. list make mpg in 1/10

                  make       mpg
  1.        AMC Concord        22
  2.          AMC Pacer        17
  3.         AMC Spirit        22
  4.      Buick Century        20
  5.      Buick Electra        15
  6.      Buick LeSabre        18
  7.         Buick Opel        26
  8.        Buick Regal        20
  9.      Buick Riviera        16
 10.      Buick Skylark        19
```

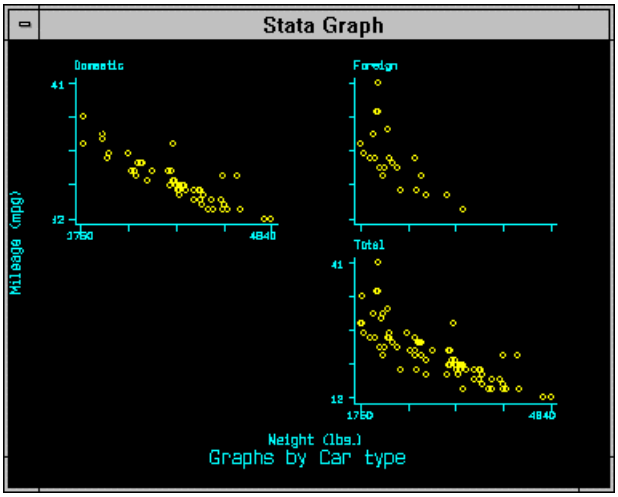
```
. describe

Contains data from \stata\auto.dta
  obs:            74                          1978 Automobile Data
 vars:            12
 size:         3,552 (99.6% of memory free)
-------------------------------------------------------------------------------
   1. make        str18  %18s                  Make and Model
   2. price       int    %8.0g                 Price
   3. mpg         int    %8.0g                 Mileage (mpg)
   4. rep78       int    %8.0g                 Repair Record 1978
   5. hdroom      float  %6.1f                 Headroom (in.)
   6. trunk       int    %8.0g                 Trunk space (cu. ft.)
   7. weight      int    %8.0g                 Weight (lbs.)
   8. length      int    %8.0g                 Length (in.)
   9. turn        int    %8.0g                 Turn Circle (ft.)
  10. displ       int    %8.0g                 Displacement (cu. in.)
  11. gratio      float  %6.2f                 Gear Ratio
  12. foreign     int    %8.0g       foreign   Car type
-------------------------------------------------------------------------------
Sorted by:  foreign


. summarize

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+--------------------------------------------------------
    make |       0
   price |      74    6165.257   2949.496       3291      15906
     mpg |      74     21.2973   5.785503         12         41
   rep78 |      69    3.405797   .9899323          1          5
  hdroom |      74         3.0        0.8        1.5        5.0
   trunk |      74    13.75676   4.277404          5         23
  weight |      74    3019.459   777.1936       1760       4840
  length |      74    187.9324   22.26634        142        233
    turn |      74    39.64865   4.399354         31         51
   displ |      74    197.2973   91.83722         79        425
  gratio |      74        3.01       0.46       2.19       3.89
 foreign |      74    .2972973   .4601885          0          1


. summarize price if mpg<21.3

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+--------------------------------------------------------
   price |      43     7091.86   3425.019       3291      15906
```

```
. sort foreign

. by foreign: summarize price mpg

-> foreign=Domestic
Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+--------------------------------------------------------
   price |      52    6072.423   3097.104       3291      15906
     mpg |      52    19.82692   4.743297         12         34

-> foreign= Foreign
Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+--------------------------------------------------------
   price |      22    6384.682   2621.915       3748      12990
     mpg |      22    24.77273   6.611187         14         41



. tabulate foreign

    Car type|      Freq.     Percent        Cum.
------------+-----------------------------------
   Domestic |         52       70.27       70.27
    Foreign |         22       29.73      100.00
------------+-----------------------------------
      Total |         74      100.00



. correlate mpg weight price weight length displ
(obs=74)

        |      mpg   weight    price   weight   length    displ
--------+------------------------------------------------------
     mpg|   1.0000
  weight|  -0.8072   1.0000
   price|  -0.4686   0.5386   1.0000
  weight|  -0.8072   1.0000   0.5386   1.0000
  length|  -0.7958   0.9460   0.4318   0.9460   1.0000
   displ|  -0.7056   0.8949   0.4949   0.8949   0.8351   1.0000
```

```
. graph mpg weight
```




```
. sort foreign
. graph mpg weight, by(foreign) total
```



Data management

- Changing the data set (*replace, drop, rename, generate, egen*)
    - replacing values

```
. replace mpg = 23 in 1
```

- replaces the value of mpg in the 1st observation

```
. replace mpg = 23
```

- replaces the value of mpg in all observations

```
. drop mpg
```

- drops the variable mpg

```
. replace mpg = mpg*(8/5)
. rename mpg kpg
```

- converts miles to kilometers
- renames variable to reflect change of units

    - generating new variables

```
. gen kpg = mpg*(8/5)
```

- creates a new variable

```
. egen kpgdiff = mean(kpg), by(foreign)
```

- creates new variable having 2 values; for foreign cars, average kpg of all foreign cars, and for domestic cars, average kpg of all domestic cars
- egen has many other functions (including *sum, count, rsum, diff, ma, max/min, median*)

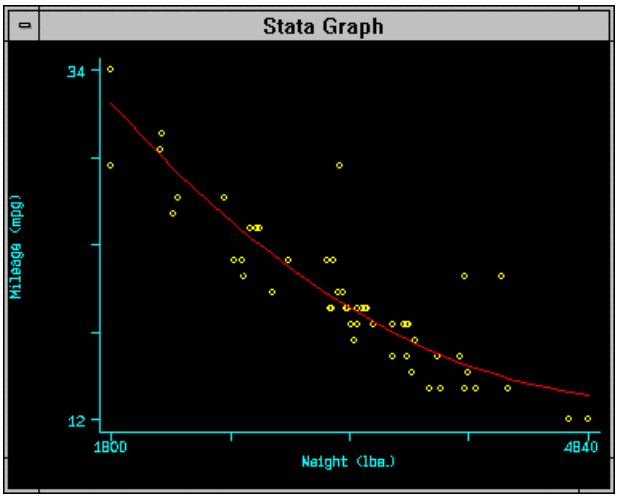
Model estimation

```
. gen wtsq = weight^2

. regress mpg weight wtsq foreign

  Source |       SS       df       MS                  Number of obs =      74
---------+------------------------------               F(  3,    70) =   52.25
   Model |  1689.15372     3   563.05124               Prob > F      =  0.0000
Residual |   754.30574    70  10.7757963               R-squared     =  0.6913
---------+------------------------------               Adj R-squared =  0.6781
   Total |  2443.45946    73  33.4720474               Root MSE      =  3.2827

------------------------------------------------------------------------------
     mpg |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  weight |  -.0165729   .0039692     -4.175   0.000      -.0244892   -.0086567
    wtsq |   1.59e-06   6.25e-07      2.546   0.013       3.45e-07    2.84e-06
 foreign |    -2.2035   1.059246     -2.080   0.041        -4.3161   -.0909003
   _cons |   56.53884   6.197383      9.123   0.000       44.17855    68.89913
------------------------------------------------------------------------------
```

```
. correlate, _coef
. correlate, _coef covariance
```

- gives correlation matrix of the coefficients and variance-covariance matrix from the last regression estimated, respectively

```
. graph mpg mpghat weight if foreign==0, connect(.l) symbol(Oi)
```




```
. predict mpghat
```

- calculates this: -.0165729*weight + 1.59*10^-6*wtsq - 2.2035*foreign + 56.53884

```
. predict mhatse, stdf
```

- calculates standard error of forecast; remember, this varies with mpghat

Programming in STATA

- batch files (or "do files") are just STATA commands in ascii files

```
. do batch.do
```

- STATA has resident programming language to create user-generated commands

```
. program define mycommand
```

- can use any STATA commands as well as special program commands