

Useful websites

- ◎ <http://data.princeton.edu/stata/>
- ◎ <http://www.ats.ucla.edu/stat/stata/notes/>
(Videos)
- ◎ <http://www.stata.com/support/faqs/>
(Stata FAQs)

Commands
being entered

Overview

Results

Enter new commands

List of variables

The screenshot displays the Stata 11.1 software interface. The main window is titled "Stata/IC 11.1 - [Results]". The menu bar includes "Edit", "Data", "Graphics", "Statistics", "User", "Window", and "Help". The toolbar contains various icons for file operations and data manipulation. The main display area shows the Stata logo and version information: "STATA (R) 11.1 Copyright 2009 StataCorp. LP, StataCorp, 4905 Lakeway Drive, College Station, Texas 77845 USA, 800-STATA-PC, http://www.stata.com, 979-696-4600, stata@stata.com, 979-696-4600 (fax)". Below this, it displays the license information: "80-student stata lab perpetual license: Serial number: 30110530930 Licensed to: Arts IT Univ of British Columbia". The "Notes" section contains two items: "1. (/m# option or -set memory-) 10.00 MB allocated to data" and "2. New update available type -update all-". The "Command" window at the bottom is empty. The "Variables" window on the left is empty, with columns for "Name", "Label", "Type", and "Format". The status bar at the bottom shows "E:\Stata11" and "CAP NUM OVR".

Name	Label	Type	Format
------	-------	------	--------

Downloading Data

- ◎ <http://faculty.arts.ubc.ca/vmarmer/econ326/401k.dta> (401k.dta)
- ◎ <http://faculty.arts.ubc.ca/vmarmer/econ326/401k.xls> (401k.xls)
- ◎ <http://faculty.arts.ubc.ca/vmarmer/econ326/401k.des> (401k.des)

Entering Data

⦿ “Use”

- To load a Stata-format dataset (.dta)
- e.g. use D:\401k.dta

⦿ “Insheet using”

- To load data from a spreadsheet (.csv)
- e.g. insheet using D:\401k.csv
- N.B. Save .xls files as .csv files before importing

Clearing Data

⦿ “Drop”

- To drop a certain variable
- e.g. drop v1

⦿ “Clear”

- To clear all the data

Modifying Data

⦿ **“Generate”**

- To generate a new variable
- e.g. generate $v9=v2^2$

⦿ **“Rename”**

- To rename a variable
- e.g. rename v9 v10

⦿ **“Replace”**

- To replace values of a variable
- e.g. replace $v10=v8$

Modifying Data

- ⦿ Exercise:
- ⦿ Using “rename” command to rename variables from 401k.csv in accordance with 401k.des
 - rename v1 prate
 - rename v2 mrate
 - rename v3 totpart
 - ...

Modifying Data

⦿ Examples:

- generate young=0

(generate a new variable “young” and the initial values are “0”s)

- replace young=1 if age<19

(if the value of “age” is below 19, then change the value of “young” from “0” to “1”)

- replace young=2 if age==19

(if the value of “age” is 19, then change the value of “young” from “0” to “2”)

Exploring Data

- ◎ **“Summarize”** (Descriptive statistics)
 - e.g. summarize prate mrate
- ◎ **“Table”** (Create a table of statistics)
 - e.g. table prate

Analyzing Data

- ⦿ **“Regress”**

- Regression (OLS)

- ⦿ **Exercise**

- Jeffrey M. Wooldridge (4e), Page 64: C2.1

C2.1

- ⦿ (i) Find the average participation rate and the average match rate in the sample of plans.
- ⦿ Solution:

```
. use D:\401k.dta
```

```
. sum prate mrate
```

Variable	Obs	Mean	Std. Dev.	Min	Max
prate	1534	<u>87.36291</u>	16.71654	3	100
mrate	1534	<u>.7315124</u>	.7795393	.01	4.91

C2.1

- ⦿ (ii) Now, estimate the simple regression equation $prate = \hat{\beta}_0 + \hat{\beta}_1 mrate$, and report the results along with the sample size and R-squared.

C2.1

. regress prate mrate

Source	SS	df	MS	Number of obs = 1534		
Model	32001.7271	1	32001.7271	F(1, 1532) =	123.68	
Residual	396383.812	1532	258.73617	Prob > F =	0.0000	
Total	428385.539	1533	279.442622	R-squared =	0.0747	
				Adj R-squared =	0.0741	
				Root MSE =	16.085	

prate	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
mrate	5.861079	.5270107	11.12	0.000	4.82734	6.894818
_cons	83.07546	.5632844	147.48	0.000	81.97057	84.18035

◎ Solution: $\hat{\beta}_0 = 83.07546$, $\hat{\beta}_1 = 5.861079$

Sample space=1534; R-squared=0.0747

C2.1

- ⦿ (iii) Interpret the intercept in your equation. Interpret the coefficient on *mrte*.

- ⦿ Solution: Recall $prate = \hat{\beta}_0 + \hat{\beta}_1 mrte$

$\hat{\beta}_0$ is the participation percentage in the pension plan when the firm contributes nothing (i.e. when $mrte=0$);

$\hat{\beta}_1$ measures the marginal effect of firm's contribution on workers' participation percentage in the pension plan.

(i.e. $\partial prate / \partial mrte$)

C2.1

- ⦿ (iv) Find the predicted prate when $mrate=3.5$. Is this a reasonable prediction?
- ⦿ Solution: $\because prate = 83.07546 + 5.861079 \cdot mrate$
when $mrate=3.5$, the predicted value
 $prate = 83.07546 + 5.861079 \cdot 3.5 = 103.589236$
which is larger than 100%, hence not reasonable.

C2.1

- ⦿ (v) How much of the variation in prate is explained by mrate ? Is this a lot in your opinion?

C2.1

◎ Solution:

```
. regress prate mrate
```

Source	SS	df	MS
<u>Model</u>	<u>32001.7271</u>	<u>1</u>	<u>32001.7271</u>
Residual	396383.812	1532	258.73617
<u>Total</u>	<u>428385.539</u>	<u>1533</u>	<u>279.442622</u>

```
Number of obs = 1534  
F( 1, 1532) = 123.68  
Prob > F = 0.0000  
R-squared = 0.0747  
Adj R-squared = 0.0741  
Root MSE = 16.085
```

prate	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
mrate	5.861079	.5270107	11.12	0.000	4.82734	6.894818
_cons	83.07546	.5632844	147.48	0.000	81.97057	84.18035

◎ $ESS=32001.7271$; $TSS=428385.539$;

◎ $ESS/TSS=32001.7271/428385.539=7.4703\%$

Other Useful Commands

- ⦿ **“Help”** (if you know the command)
 - e.g. help graph

- ⦿ **“Search”** (if you don't know the command)
 - e.g. search graph