

Stata on HPC

2023 KSU Spring HPC Workshop Series

Tom Boyle, April 07, 2023

Agenda

What is Stata?

Local: Interface example

Local: Do-file and Data example

HPC: Interact session

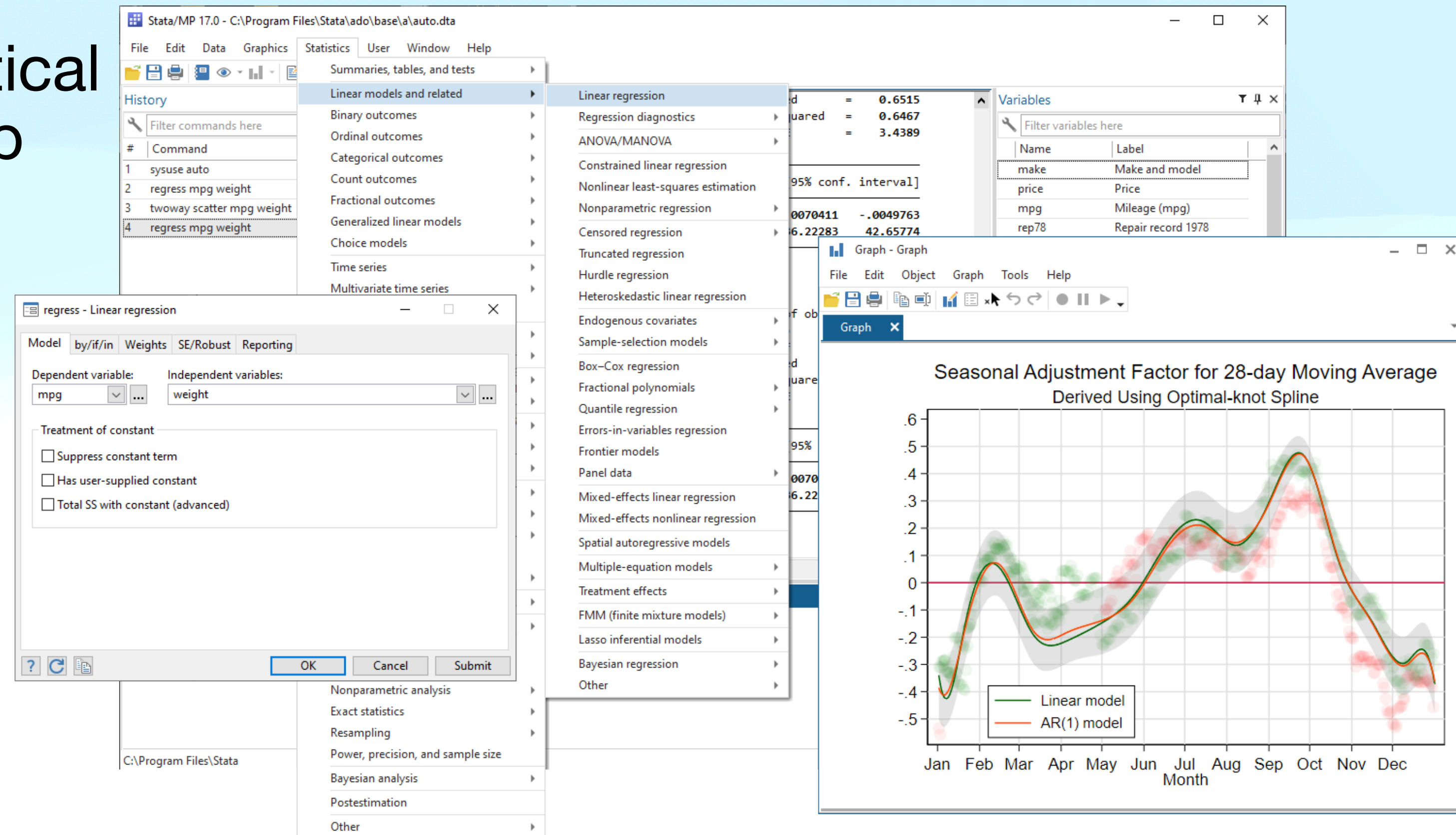
HPC: Job submission

Getting Help

What is Stata?

Statistical software for data science

- Stata is a general-purpose statistical software package from StataCorp
 - Data manipulation
 - Visualization
 - Statistics
 - Automated Reporting



Stata

local install

- Available on apps.kennesaw.edu
- Version 17
- Mac, Win & Linux versions
- Note: “update all” after install
- Tip: use “man <topic>” for info
- KSU license does not include Stata/MP



StataCorp LLC
stata.com for
Training courses
Webinars
Conferences
Happy hours
Support

Stata Interface

econometric_interface.do

- Command window
- Do-file editor
- Data editor
- Menu system & dialog boxes
 - Graph
- From “Econometric Analysis” by William Greene

The screenshot displays the Stata/SE 17.0 interface with several windows open:

- Do-file editor (econometric_interface.do):** Contains the following code:

```
1 use "https://stats.idre.ucla.edu/stat/stata/examples/greene/TBL5-1.DTA", clear
2 rename x1 age
3 rename x2 income
4 rename x3 exp
5 rename x4 ownrent
6 rename x5 selfemp
7
8 generate incomesq = income^2
9 drop if exp==0
10 save chapter12
11
12 regress exp age ownrent income incomesq
13
14 rvppplot income, xlabel(0(2)12) xline(2 4 6 8 10) ylabel(-500(500)2000) yline(0 500
15 1000 1500)
```
- Data Editor (Data Editor (Edit) — chapter...):** Shows a table with columns y1, y2, age, and ir. The first row is highlighted with a yellow box labeled "Data Editor".
- Results window (Stata/SE 17.0 — chapter12.dta):** Displays regression statistics:

Model	1749357.01	4	437339.252	Prob > F	=	0.0008
Residual	5432562.03	67	81083.0153	R-squared	=	0.2436
Total	7181919.03	71	101153.789	Adj R-squared	=	0.1984
				Root MSE	=	

exp	Coefficient	Std. err.	t	P> t	[95% conf.]
age	-3.081814	5.514717	-0.56	0.578	-14.08923
ownrent	27.94091	82.92232	0.34	0.737	-137.5727
income	234.347	80.36595	2.92	0.005	73.93593
incomesq	-14.99684	7.469337	-2.01	0.049	-29.9057
_cons	-237.46	190.3517	-1.19	0.238	-635.0541
- Graph window (Graph — Graph):** Shows a scatter plot of Residuals vs. income. The plot has a grid and a yellow box labeled "Graph" over the data points.

Local Do-file

Example: wage_educ.do

```
use http://fmwww.bc.edu/ec-p/data/wooldridge/wage1
summ wage educ
reg wage educ
display _b[educ]*12.56+_b[_cons]
```

- From Experimental Design by Roger Kirk
Chapter 5: Completely Randomized Design

Local Do-file

Command Line or via Do-file Editor

- use `http://fmwww.bc.edu/ec-p/data/wooldridge/wage1`
- use `- Load Stata dataset`
 - use `loads` into memory a Stata-format dataset previously saved by `save`.
- Note: your local machine can point to data on the internet, HPC compute nodes do not have access to the internet

Do-File

Local results from Command window

- summ wage educ

Variable	Obs	Mean	Std. dev.	Min	Max
-----+-----					
wage	526	5.896103	3.693086	.53	24.98
educ	526	12.56274	2.769022	0	18

- Note: Summ - Summarize
- Summarize calculates and displays a variety of univariate summary statistics.
 - Observations, Mean, Standard Deviation, Minimum, Maximum

Results

Local results from Command window

- reg wage educ

- Source | SS df MS Number of obs = 526
- -----+----- F(1, 524) = 103.36
- Model | 1179.73204 1 1179.73204 Prob > F = 0.0000
- Residual | 5980.68225 524 11.4135158 R-squared = 0.1648
- -----+----- Adj R-squared = 0.1632
- Total | 7160.41429 525 13.6388844 Root MSE = 3.3784

- Note: reg = regress
- Linear Regression - regress performs ordinary least-squares linear regression.
 - Source, Sum of Squares, degrees of freedom (variance), Mean Squares, Observations

Results

Local results from Command window

- `display _b[educ]*12.56+_b[_cons]`
 - 5.8946206
-
- `display` - Display strings and values of scalar expressions
 - `_` is used with underscore variables (built-in)
 - `_b` means all elements in the [vector]

Exporting data and Do-file

From local to HPC

- See your data with Data Editor
- Export your data, File -> Export ->
 - to Excel, to SAS
 - to delimited, to csv
 - to ODBC, to dBase
- See your Do-file (filename.do) with Do-file editor
 - Save Button in editor or File -> Save

HPC

Testing & Running Stata Jobs



Stata Environment Module

module avail stata

Stata/17.0.1 is the default version available on the HPC

\$ module load Stata

Note: the capital s is required.

wage_educ.do

Stata Do-file example

```
use /data/home/tboyle/workshop/2023/wage_educ.dta  
summ wage educ  
reg wage educ  
display _b[educ]*12.56+_b[_cons]
```

- **Very low** intensity tedfdsting on login node:
 - \$ module load stata
 - \$ stata-se -b do wage_educ.do
 - \$ less wage_educ.log

Testing with Interact

\$ interact

- \$ cd workshop/2023
- \$ module load Stata
- \$ stata-se -b do wage_educ.do
- \$ less wage_educ.log

run_stata.pbs

job submission script (simple)

```
#!/bin/bash
#PBS -l nodes=1:ppn=2
#PBS -j oe
#PBS -N Stata_Job
#PBS -q batch
#PBS -l walltime=00:10:00
#PBS -m abe
#PBS -M tboyle@kennesaw.edu
```

```
module load Stata
cd /data/tboyle/workshop/2023
stata-se -b do wage_educ.do
```

```
exit 0
```

to submit: \$ qsub run_stata.pbs
to view results: \$ more wage_educ.log

run_stata_pass.pbs

job submission script (pass Do-file)

```
#!/bin/bash
#PBS -l nodes=1:ppn=2
#PBS -j oe
#PBS -N Stata_Job
#PBS -q batch
#PBS -l walltime=00:10:00
#PBS -m abe
#PBS -M tboyle@kennesaw.edu
```

```
module load Stata
cd /data/home/tboyle/workshop/2023
echo stata-se -b do ${FILE}
stata-se -b do ${FILE}
```

```
exit 0
```

to submit: \$ qsub run_stata_pass.pbs -vFILE=wage_educ.do
to view results: \$ more wage_educ.log

Get More Help

type 'man stata' in the command window

- More resources - Resources for learning more about Stata
 - Links to Video tutorials, FAQs, Blog, Forum, Books, Training, Conferences, Webinars, News, Tips & Tricks and Social Media.
- Other resources (official and unofficial)
 - <https://www.stata.com/links/>

Questions?

Thank you