

A Stata Course for Undergraduates

Beyond the Numbers Conference
Federal Reserve Bank of St. Louis
September 29, 2014

James Ng, PhD
Hesburgh Libraries
University of Notre Dame

Outline

- Background
- Logistics
- Course content in detail
- Lesson learned
- Challenges ahead

History

- Pitched idea to dept. chair in Fall 2013
 - perceived need
 - target: rising juniors/seniors doing research
- Spring 2014 – first time offered
- Fall 2014 – ongoing

ECON 30300

Empirical Research in Economics using Stata

This course will equip students with substantial knowledge of Stata so that they can use it for econometric analysis in their research projects, such as senior honors thesis. Using **real datasets**, students will learn how to use Stata to **read and manipulate** data, and to produce reliable **results**. The course will therefore focus on the **practical** aspect of data analysis rather than econometric theory. The course will also highlight many **common pitfalls** that beginning researchers often fall into. By the end of the course, students should feel sufficiently competent with Stata to **independently work on any dataset**.

Logistics

	Spring 2014 -completed	Fall 2014 -ongoing
Credit hours	1 (letter grade)	1 (letter grade)
Length	Entire semester	Half semester
# meetings per week	1	1
Meeting duration	1 hour	2 hours
Capacity	19	12
Enrollees	14	6
Location	Library	Library

Library Classrooms Used

Spring 2014 -completed

Fall 2014 -ongoing



Computer Lab



Center for Digital Scholarship

Course Content

	Spring 2014 -completed	Fall 2014 -ongoing
Loading data	Week 1	Week 1
Combining data	Week 2	Week 1
Manipulating variables	Weeks 3-6	Weeks 2-3
Producing descriptive stats	Week 7	Week 2
Graphing	Week 6	Week 3
Loops and macros	Week 3	Week 4
Estimation and inference	Weeks 8-14	Week 5
Replication exercise	None in class	Weeks 6-7 in class
Writing programs	Week 5	-

Fall 2014 in detail

- Data used
 - National Health Interview Survey
 - General Social Survey
 - Penn World Table
 - Indonesian Family Life Survey
 - Census 1990 (from IPUMS)

 - Other sources: Stata built-in , UCLA tutorials, made-up

Fall 2014 in detail

- Best practices
 - setting working directory
 - using do-files, keeping logs

Fall 2014 in detail

- Loading data
 - focus on fixed-format ASCII files
 - 2012 NHIS household, family, person files

Fall 2014 in detail

- Combining data
 - merge vs. append
 - how to merge correctly
 - 1:1, 1:m, m:1
 - how to append
 - 2 waves of IFLS (longitudinal)

Fall 2014 in detail

- Manipulating data
 - operations on variables
 - generating, modifying, labeling
 - beware missing values, special codes
 - simple questions to check understanding:

After merging, some households did not have any members interviewed in 2012.
How would you drop these households from our dataset?
How many variables does the combined data contain?
How many observations? What is the unit of observation?

Fall 2014 in detail

- Manipulating data
 - within-group patterns

How many households have multiple families?
What is the age of the oldest person in each family?

- collapsing observations across groups
- reshaping data

Fall 2014 in detail

- Manipulating data
 - dealing with dates
 - handling string variables
 - automating repetitive tasks: loops and macros

Fall 2014 in detail

- In-class exercise to check understanding:
 - convert this:

	id	country	year	duration	startdate	enddate	region	location
1	201990001	Canada	1990	1	10972	10972	North America	national
2	201990002	Canada	1990	1	11133	11133	North America	Montreal, Quebec
3	201990003	Canada	1990	1	11139	11139	North America	Montreal, Quebec
4	201990004	Canada	1990	17	11190	11206	North America	Montreal, Quebec
5	201990005	Canada	1990	2	11183	11184	North America	Montreal, Quebec
6	201990006	Canada	1990	1	11219	11219	North America	Montreal, Quebec

- to this:

	id	daynum	date	duration	startdate	enddate	country	year	region	location
1	201990001	1	15jan1990	1	10972	10972	Canada	1990	North America	national
2	201990002	1	25jun1990	1	11133	11133	Canada	1990	North America	Montreal, Quebec
3	201990003	1	01jul1990	1	11139	11139	Canada	1990	North America	Montreal, Quebec
4	201990004	1	21aug1990	17	11190	11206	Canada	1990	North America	Montreal, Quebec
5	201990004	2	22aug1990	17	11190	11206	Canada	1990	North America	Montreal, Quebec
6	201990004	3	23aug1990	17	11190	11206	Canada	1990	North America	Montreal, Quebec
7	201990004	4	24aug1990	17	11190	11206	Canada	1990	North America	Montreal, Quebec
8	201990004	5	25aug1990	17	11190	11206	Canada	1990	North America	Montreal, Quebec
9	201990004	6	26aug1990	17	11190	11206	Canada	1990	North America	Montreal, Quebec
10	201990004	7	27aug1990	17	11190	11206	Canada	1990	North America	Montreal, Quebec
11	201990004	8	28aug1990	17	11190	11206	Canada	1990	North America	Montreal, Quebec
12	201990004	9	29aug1990	17	11190	11206	Canada	1990	North America	Montreal, Quebec

Using the two datasets stored in Sakai under Class #3 > GDP & Happiness, produce the following table of statistics. Then, replicate the graph.

Table 1: Annual real GDP per capita and mean happiness, 1975-1997

year	rgdppc	meanhappy
1975	22075.38	2.19798
1976	23055.56	2.215477
1977	23896.55	2.229208
1978	24999.11	2.247858
1979	25538.14	
1980	25220.93	2.205882
1981	25615.52	
1982	24869.67	2.160647
1983	25745.56	2.183725
1984	27333.99	2.217993
1985	28183.51	2.172549
1986	28864.97	2.20911
1987	29485.49	2.156742
1988	30390.34	2.24693
1989	31163.08	2.229358
1990	31430.6	2.244673
1991	31046.02	2.200798
1992	31796.23	
1993	32392.62	2.204872
1994	33384.2	2.165603
1995	33868.18	
1996	34749.89	2.183016
1997	35882.78	

Figure 1

Mean Happiness and Real GDP Per Capita between 1975 and 1997 for Repeated Cross-Sections of (Different) Americans



Notes: Right-hand scale is the average of the answers to the question from the United States General Social Survey: "Taken all together, how would you say things are these days—would you say that you are (3) very happy, (2) pretty happy, or (1) not too happy?" Real GDP per capita is measured in 1990 U.S. dollars.

Note: In Figure 1, the base year used to compute real GDP was 1990. In the data I provided, the base year used was 2005. Consequently, your numbers for real GDP per capita don't match up with the figure above, as is evident from Table 1. However, the trend will be exactly the same. Source:

<http://www.jstor.org/stable/30033632?origin=JSTOR-pdf>

Fall 2014 in detail

- Estimation and inference
 - T-test, F-test, OLS, logit/probit, IV-2SLS
 - running, interpreting, outputting in formatted table
 - Want students to realize: heavy lifting was in data manipulation

Fall 2014 in detail

- Replication exercise
 - Challenge:
 - doable in 4 hours of class time
 - not too easy or difficult
 - sufficiently interesting
 - Bleakley and Chin, 2004. Language Skills and Earnings: Evidence from Childhood Immigrants, *The Review of Economics and Statistics*, 86(2): 481–496
 - Data from IPUMS-USA (1990 Census)

Lessons Learned

- Missteps in inaugural course:
 - Too much time on complicated features e.g. programs, matrices
 - Too much time lecturing on econometrics

	Spring 2014	Fall 2014 -ongoing
Estimation and inference	Weeks 8-14	Week 5

Lessons Learned

- Missteps in inaugural course:
 - Too little time letting students DIY

	Spring 2014	Fall 2014 -ongoing
Replication exercise	None in class	Weeks 6-7 in class

- Homework? Not for a 1-credit course
- Classroom setup was not ideal:

Spring 2014 vs. Fall 2014

NOT GREAT

GREAT



Library Computer Lab



Library CDS

Challenges

- Need to increase enrollment
 - Econometrics as co-requisite instead of pre-?
- Turn it into a lab for Econometrics?
 - instructors may prefer other software
- Still need more hands-on components