Introduction to Handling Data

ECON20222 - Lecture 1

Ralf Becker and Martyn Andrews

30 October 2018

What is this course unit about?

- Help you implement and interpret the main inference techniques used in Economics
- Focus on:
  - causal inference
  - the main pitfalls of time-series analysis

At the end of this unit ...

You will be able to:

- Do intermediate data work in R
- Confidently apply regression analysis in R
- Apply more advanced causal inference techniques in R
- Find coding help for any new challenges in R
- Identify inference appropriate for the occasion
- Discuss strengths and weaknesses of particular empirical applications
- Interpret empirical results (with due caution!)

What you need to do

To learn in this unit you need to:

coding, cleaning data, struggling, amazement at what you can do

answering real questions, that there is not always a clear answer

Aim for today

Statistics/Econometrics

- Summary Statistics
- Difference between population and sample
- Hypothesis testing
- Graphical Data Representations
- Simple regression analysis

R Coding

- Introduce you to R and RStudio
- How do I learn R
- Import data into R
- Perform some basic data manipulation
- Perform hypothesis tests
- Estimate a regression

Why Data Matter

Let's assume the following

The statistical techniques we want students to be able to implement are:

- Importing data
- Cleaning data
- Merging data
- Summary stats
- Plotting data
- Regressions analysis
- Hypothesis testing

The traditional way

Here is how we tend to deliver the R-component

- Make code and data for analysis used in lectures available
- Set extra weekly worksheets (see "Week3practice.pdf")
- Have website with material to learn R (Google "ECLR R" to find it)
- Have "smallish" assessment items which assume that students have done some data work (see "CW3 201516.pdf")
- Drop-in help sessions
- Perhaps one introductory lab
- Link to Datacamp

But self-learning has limitations

The things that trip students up

- RStudio architecture
- Having to write a code file (and not click)
- Using packages
- What the heck is a function???

The things that trip students up

- RStudio architecture
- Having to write a code file (and not click)
- Using packages
- What the heck is a function???
- Error messages

The things that trip students up

- RStudio architecture
- Having to write a code file (and not click)
- Using packages
- What the heck is a function???
- Error messages
- They think they need to know everything - they Google everything else, but not how to solve their problem

The things that trip students up

- RStudio architecture
- Having to write a code file (and not click)
- Using packages
- What the heck is a function???
- Error messages
- They think they need to know everything - they Google everything else, but not how to solve their problem
- The fear of plagiarism

The things that trip students up

- RStudio architecture
- Having to write a code file (and not click)
- Using packages
- What the heck is a function???
- Error messages
- They think they need to know everything - they Google everything else, but not how to solve their problem
- The fear of plagiarism
- They think they will break the computer if they make a mistake

Plan for today

Think about how we can help students to develop these vital coding skills

Our testbed

The projects

- Measuring climate change
- Collecting and analysing data from experiments
- Measuring the effect of a sugar tax
- Measuring wellbeing
- Measuring inequality: Lorenz curves and Gini coefficients
- Measuring management practices
- Supply and demand
- Measuring the non-monetary cost of unemployment
- Credit-excluded households in a developing country
- Characteristics of banking systems around the world
- Measuring willingness to pay for climate change mitigation
- Government policies and popularity: Hong Kong cash handout

Measuring the effect of a sugar tax

Basic data structure

`dat_c` contains observations for products (`product_id`) for which prices are observed in the same store (`store_id`) at three points in time (`time` - DEC2014, JUN2015, MAR2016).

A sugar tax was introduced on some products (`dat_c$taxed == "taxed"`), sometime between DEC2014 and MAR2016.

Things we do in the project

- Find out how many products and stores there are
- Frequency Tables
- Calculating the
- Column/Bar chats
- Testing for statistical significance in Price changes

Your task

Think how you could use the available data to introduce students to one of the vital, generic programming skills

- Understand the RStudio architecture
- The need to write a code file (.r or .Rmd)
- Understand and learn from error messages
- How to google effectively for help
- How to pick someone else's code and adapt it
- Trial and error

Email finished products to ralffbecker@gmail.com (yes, two fs)