Class 1, Introduction

Adam Okulicz-Kozaryn

adam.okulicz.kozaryn@gmail.com

this version: Wednesday 4th February, 2015 18:40

outline

misc

why this class ?

descriptive statistics

paper

bonus–data sources

outline

a note

- note: i always put lots of things at the beginning in the "misc section" that are not necessarily that important (e.g. you may not need to print them)

let's introduce ourselves

- this is a small class, which is great for you
- we can customize it better
- we can have more class discussions
- and make it less formal than a usual lecture

about myself...

- public economics e.g. preferences for redistribution
- cultural economics e.g. religion, trust
- economics of happiness

  [*] `http://articles.cnn.com/2011-05-23/travel/vacation.in.america_1_vacation-germans-long-holiday?_s=PM:TRAVEL`
- regions, cities, nature, sustainability
- software: stata, Python
- data: mostly surveys: gss, wvs, eurobarometers

about yourself...

- what are your research interests?
- what data are you using?
- any ideas for civic engagement yet?
- what would you like to see in this class, any specific examples?

my teaching approach

- feel free to interrupt me during the class
- my approach to teaching methods is applied
- i believe that class is not useful if you do not use the material covered
  - yes, we still have to do the math
  - but i am really interested in teaching you how to apply your skills
  - hence, we will use stata a lot
  - and we will have lots of examples
  - and lots of extra materials for self-study

outline

why this class is useful

- there are more and more jobs that require statistical skills
- there will be more such jobs because we have more and more data
- data analysis will become more important in every field [*]
  `http://www.amazon.com/Super-Crunchers-Thinking---Numbers-Smart/dp/0553805401/ref=sr_1_1?s=books&ie=UTF8&qid=1306434931&sr=1-1`
- data analysis is usually done (in social science) with regression that you will learn in this class

data revolution

- we are now witnessing the data revolution
- what that means ?
- we had industrial revolution...
  people moved from farms to factories
- now people move towards data-oriented economy
- you will find data everywhere and you can analyze it using regression

data, regressions and randomized trials

- data, regressions and randomized trials will be more important in all fields [*] http://www.amazon.com/Super-Crunchers-Thinking---Numbers-Smart/dp/0553805401/ref=sr_1_1?s=books&ie=UTF8&qid=1306434931&sr=1-1

Example: decline of physicians

- data analyst job will be more prestigious than physician job in 10, maybe 25 years
- this is due to the rise of evidence-based medicine (resisted by physicians)
- physicians are not any more oracles; they are just consumer of data
- databases and models are starting to tell them what to do

- [*] `http://www.amazon.com/Super-Crunchers-Thinking---Numbers-Smart/dp/0553805401/ref=sr_1_1?s=books&ie=UTF8&qid=1306434931&sr=1-1`

regressions becoming more important

- "I keep saying that the sexy job in the next 10 years will be statisticians." [*] `http://www.nytimes.com/2009/08/06/technology/06stats.html`
- more and more data, e.g. surveys, blogs, twitter
- academia more quantitative, e.g. pol sci
- industry more quantitative, e.g. google, amazon, facebook, netflix

regressions becoming more important

- qualitative data (pictures, text, etc.) are just rich quantitative data and can be analyzed like quantitative – everything can be quantified – any examples of non-quantifiable things ?
- the bottom line is that no matter what you study you'll find data for it and you can analyze it with a regression

how it differs from qm1?

- qm1 just prepares for qm2
- qm1 is not very useful
- qm2 is much more useful than qm1
- qm2 reuires much more work than qm1
- qm2 is more fun than qm1
- qm2: you have to read Gujarati book (not necessary to read book in qm1)

outline

rule

- never run regressions without doing descriptive statistics first
- you need to understand data before making inferences from it

descriptive statistics first !

- what is the level of measurement ?
- interval/ratio
- ordinal
- nominal

  examples ?
- in ols regression:
  - dv (outcome var) has to be interval/ratio
  - all rhs (predictors) var have to be interval/ratio or dummies (1/0); shouldn't use ordinal variables

outline

publish or perish

- start paper early, discuss with me, other faculty, students
- some guidelines:

  `http://gking.harvard.edu/files/paperspub.pdf` [but write original paper, do not replicate existing research]

paper guidelines

- graphs are great and underutilized in social science [*]
  www.stat.columbia.edu/~gelman/presentations/cuips_presentation.pdf
- results: statistical vs substantive significance; robustness (e.g. model specification)
- results: is there a potential bias, what direction?
- you have to have replication files (data, code, text)
- concise, to the point; avoid data/text padding
- how researchers were wrong? why? why you are right? so what?
- paper should be written as if for a journal submission
  it should read like a story, not like stata manual

outline

data.gov

- http://www.data.gov/

data sources

- http://www.worldvaluessurvey.org/
- http://www.norc.uchicago.edu/GSS+Website/
- http://www.icpsr.umich.edu/icpsrweb/ICPSR/
- http://www.thearda.com/
- http://ksghome.harvard.edu/~pnorris/Data/Data.htm

more data sources

- http://www.measureofamerica.org/
- http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/0,,contentMDK:20388241~menuPK:665266~pagePK:64165401~piPK:64165026~theSitePK:469382,00.html
- http://usa.ipums.org/usa/
- https://international.ipums.org/international/

“non-traditional” data

- http://dvn.iq.harvard.edu/dvn/dv/patent
- http://www.trustlet.org/wiki/Trust_network_datasets

happiness data

- http://www.bmj.com/content/337/bmj.a2338.full
- http://apps.facebook.com/usa_gnh/
- http://www.facebook.com/notes/facebook-data-team/relationships-and-happiness/304457453858
- http://www.springerlink.com/content/757723154j4w726k/fulltext.pdf
- http://www.wefeelfine.org/

facebook data

- http://apps.facebook.com/usa_gnh/
- http://www.facebook.com/notes/facebook-data-team/relationships-and-happiness/304457453858
- http://www.facebook.com/notes/facebook-engineering/visualizing-friendships/469716398919
- http://cyber.law.harvard.edu/node/4682
- http://www.thefacebookproject.com/resource/datasets.html