final project

(if little time just discuss briefly each TOC item)

adam okulicz-kozaryn

adam.okulicz.kozaryn@gmail.com

this version: Monday 23rd March, 2020 17:26

outline

litRev:https://theaok.github.io/generic/howToGoogSch.html

opinion/activism v science

how do i produce a final project for this class?

presentations

final paper/project in general [NOT resMet]

a dillema: publishabe project or student project [NOT resMet]

inline response to comments on ps [qm*,dev,swb,dirStu]

links: good research in words of others [NOT resMet]

the end of theory, data is enough, airplane model [datMan, dirStu]

regression [qm2, dirStu]

data managemnt [datMan, dirStu]

GIS (todo)

outline

litRev:https://theaok.github.io/generic/howToGoogSch.html

opinion/activism v science

how do i produce a final project for this class?

presentations

final paper/project in general [NOT resMet]

a dillema: publishabe project or student project [NOT resMet]

inline response to comments on ps [qm*,dev,swb,dirStu]

links: good research in words of others [NOT resMet]

the end of theory, data is enough, airplane model [datMan, dirStu]

regression [qm2, dirStu]

data managemnt [datMan, dirStu]

GIS (todo)

outline

litRev:https://theaok.github.io/generic/howToGoogSch.html

opinion/activism v science

how do i produce a final project for this class?

presentations

final paper/project in general [NOT resMet]

a dillema: publishabe project or student project [NOT resMet]

inline response to comments on ps [qm*,dev,swb,dirStu]

links: good research in words of others [NOT resMet]

the end of theory, data is enough, airplane model [datMan, dirStu]

regression [qm2, dirStu]

data managemnt [datMan, dirStu]

GIS (todo)

intuition

- many beginning scholars conflate opinion/activism with science
- youthful pasion and excitement are great
- but we need to force ourselves to be objective because humans are not
- opinion/op-ed usually surfaces where objective science should appear

opinion/activism v science: a dichotomy?

- not necessarily a dichotomy
- and indeed, much of the time ideally both
- although in practice it seems to be either or

Haidit's book 'Righteous Mind'

- being right and superior: righteous: (of a person or conduct) morally right or justifiable; virtuous
- the problem may be that "science" (fake science) is the result of opinion/activism, not the other way round
- does activism binds and blinds (like morality)?
- argument building follows passion and intuition, not reason!
  - guns, abortion, etc
- eg, why is it ok to make fun of Melania Trump and not Michelle Obama?
- why is antifa better than proud boys?

a serious problem with academics! we're righteous!

- think we know better, even things we don't study!!
- like if you're an expert in one area, you think you are one in all areas!
- how many of you study guns? and yet you sound like experts on the issue
- why? because you have some expertise in some other area
- it's one thing to have an opinion
  - another thing to absolutely sure you are right
- and yet doubt is necessary for science and progress
- do you even allow a possibility that you may be wrong?
- if not, then is it belief or science?

anecdotal evidence, stereotyping

- similar statements, and you think only one is true!
-
- an undocumented immigrant killed someone: a significant number of undocumented immigrants are criminals
- a gun was used to commit a crime: a significant number of guns are used to commit a crime
- same fallacies!! but you just see one depending on which side you're on
- what about this: residential swimming pools kill more babies than guns (Freakonomics)

easier to be positive than normative

- are guns good or bad? tough to say!
  - see a philosopher or a priest
- but if we want to increase public safety, we should increase gun control

  `https://www.hsph.harvard.edu/hicrc/firearms-research/gun-threats-and-self-defense-gun-use-2`

(LUKIANOFF and HAIDT, 2015)

- microaggressions/ political reasoning
- emotional reasoning
- Common Cognitive Distortions
- I'm offended is a trump card!; if sth is unwelocomed, that's it!
- trigger warnings
- catastrophize/zero tolerance

outline

litRev:https://theaok.github.io/generic/howToGoogSch.html

opinion/activism v science

how do i produce a final project for this class?

presentations

final paper/project in general [NOT resMet]

a dillema: publishabe project or student project [NOT resMet]

inline response to comments on ps [qm*,dev,swb,dirStu]

links: good research in words of others [NOT resMet]

the end of theory, data is enough, airplane model [datMan, dirStu]

regression [qm2, dirStu]

data managemnt [datMan, dirStu]

GIS (todo)

start early

- it's high time now you know what you'll do for the final project
- if you are not sure, email me
- if you cannot find data, email me
- I'd like to meet with each of you **at least twice per your project**

kill 2 birds with one stone

- analyze something that you study for another class
- use data from your work
  - no matter where you work–they always have some data

start with good data

- representative
- easy to use
- novel/innovative (eg twitter)
- local/familiar (so that you can compare to your experience)
- long term investment (use same data for years)

treat it seriously, dont't waste your time

- not only a big chunk of the final grade
- use it or lose it!
- if you don't use tools, you will lose this skill soon
- be efficient, use this class for something beyond this class
  - do something useful for your work (civic engagement)
  - it could be analysis chapter for your capstone/thesis/dissertation/journal paper
- **important!:** email me drafts and see me few times in the second half of this class

the good news

- the good news is that you already have much of it
- just reuse your problem sets
- yes, you can reuse past (future) assignments for final project
- or you can, of course, come up with something new
- you can also reuse your work from other classes/projects (eg your job)
  - but in that case you need to clearly state what you are reusing
  - state that in the text of the final project, eg at the beginning of it

the bad news

- there is always bad news accompanying good news...
- if you are building on your past ps
- you need to extend them very substantially
  - cannot just glue them all together
- and they need to form a logical project
- it needs to be interesting/innovative
- and discuss your findings–why they are important?
- what is new there?

consensus creation or consensus shift

- perhaps your study creates consensus or shifts it
- great if it does
- [*] Hollenbeck (2008)

outline

litRev:https://theaok.github.io/generic/howToGoogSch.html

opinion/activism v science

how do i produce a final project for this class?

presentations

final paper/project in general [NOT resMet]

a dillema: publishabe project or student project [NOT resMet]

inline response to comments on ps [qm*,dev,swb,dirStu]

links: good research in words of others [NOT resMet]

the end of theory, data is enough, airplane model [datMan, dirStu]

regression [qm2, dirStu]

data managemnt [datMan, dirStu]

GIS (todo)

rules

- not too much background! cut to the chase asap
- max 1 slide per 1 minute, so 10min presentation max 10slides
- max 15 lines of text per slide (ideally $<=12$)
- use graphs/pictures/maps as much as possible/practicable

outline

litRev:https://theaok.github.io/generic/howToGoogSch.html

opinion/activism v science

how do i produce a final project for this class?

presentations

final paper/project in general [NOT resMet]

a dillema: publishabe project or student project [NOT resMet]

inline response to comments on ps [qm*,dev,swb,dirStu]

links: good research in words of others [NOT resMet]

the end of theory, data is enough, airplane model [datMan, dirStu]

regression [qm2, dirStu]

data managemnt [datMan, dirStu]

GIS (todo)

interesting to you− >fun for you

- study something that is of interest to you
- say crime if you live in high-crime area
- or agriculture if you live in high-agriculture area
- eg I study income inequality, because my family is unequal
- fun to work on something of great interest to you

be curious

- curiosity is arguably the most important reason for research
- do research about something that you are curious about
- it will be fun and you will be good at it

interesting to others

- (if interesting to you, more likely also interesting to others)
  - (if you hate your work, others won't love it)
- research must be interesting
- i am very much against typical dry research only demonstarting technical proficiency or mastery of material
- research should read like a story
  - its language should be simple
  - do not write words that you do not use when talking
- be simple and clear:
  "person", not "individual"
  "explain", not "elucidate"

the “so what” question?

- go through your final project and ask yourself “so what?”
- if what you have just read is not relevant, drop it
- this rule, as all rules here, pertain not only to text
- but also to tables, graphs, maps, etc

quality vs quantity

- do not just dump everything that you know on the topic
- in fact, the opposite is good:
  - be as brief as possible: i will **decrease grade** for padding:
    (putting irrelevant/wordy stuff into your paper)
- sure, do a lot of stats, reading, mapping
  - but give me only the best of it
  - (have to do a lot to find the best)
- be thorough–a typical paper is twenty something pages doublespaced incl rferences and everything
  - again, don't do boilerplate! better shorter than longer
- need to beef it up!
- if literature review only, need to cite really a lot: eg

make it bulletproof/various robustness checks

- triangulate (different measurement of the same concept)
  - eg: educ: years, degree, ranking on US News, actual standardized skills (eg PISA scores)
- exclude alternative hypotheses: be devil's advocate, try to throw out your research out of the window
- use different datasets: time (eg now v 70s, space (US v UK), aggregation: state v county
- etc

be smart and ethical

- ◇ do build on your earlier/concurrent work
  - · connect it somehow; dont start someting completely new
- ◇ have to say if and which parts were written outside of this class,
  - · otherwhise it is scholastic dishonesty–you can indeed plagiarize yourself!
- ◇ also per code: reuse yours (and others!): eg Glaeser's unhappy cities in JOLE; dataverse, authors websites
  - · ie: not just lit rev, but also build on others code: this is the fastest way
  - · and do acknowledge that code taken from others, too

outline

litRev:https://theaok.github.io/generic/howToGoogSch.html

opinion/activism v science

how do i produce a final project for this class?

presentations

final paper/project in general [NOT resMet]

a dillema: publishabe project or student project [NOT resMet]

inline response to comments on ps [qm*,dev,swb,dirStu]

links: good research in words of others [NOT resMet]

the end of theory, data is enough, airplane model [datMan, dirStu]

regression [qm2, dirStu]

data managemnt [datMan, dirStu]

GIS (todo)

outline

litRev:https://theaok.github.io/generic/howToGoogSch.html

opinion/activism v science

how do i produce a final project for this class?

presentations

final paper/project in general [NOT resMet]

a dillema: publishabe project or student project [NOT resMet]

inline response to comments on ps [qm*,dev,swb,dirStu]

links: good research in words of others [NOT resMet]

the end of theory, data is enough, airplane model [datMan, dirStu]

regression [qm2, dirStu]

data managemnt [datMan, dirStu]

GIS (todo)

inline response: like a peer-review process

- inline response: quote *all* comments and reply to each below it
- i will give you comments on your draft(s) (eg each ps)
  - so each subsequent ps will start with a section where you reply
  - don't forget about verbal comments from me and others in class (if you present or we discuss your research–so make notes!)
  - and i typically email you if you present–include text from that email, too
- you need to respond to *all* comments
- you may disagree, but you have to respond (eg say why you disagree)

inline response: like a peer-review process

- you need to reply inline, ie quote the comment and then respond to it
- for example see my `https://sites.google.com/site/adamokuliczkozaryn/gis_int/rev_ariq.pdf`
- if no tracked changes be specific where the change was made–page and paragraph
- if you need clarification on my comments (i tend to vbe overly parsimonious), email me!

outline

litRev:https://theaok.github.io/generic/howToGoogSch.html

opinion/activism v science

how do i produce a final project for this class?

presentations

final paper/project in general [NOT resMet]

a dillema: publishabe project or student project [NOT resMet]

inline response to comments on ps [qm*,dev,swb,dirStu]

links: good research in words of others [NOT resMet]

the end of theory, data is enough, airplane model [datMan, dirStu]

regression [qm2, dirStu]

data managemnt [datMan, dirStu]

GIS (todo)

i wish i knew it when i was a student

- instead of rephrasing what i have learned by reading other people description of good academic work
- i am just linking their writings
- following their advice should help you producing a good final project for this class
- we'll quickly scan through them
- i also list some points in slides
- read them after the class–they are very useful

Greg Mankiw

- "My rules of thumb"
- `http://scholar.harvard.edu/files/mankiw/files/my_rules_of_thumb.pdf`
- have productive mentor(s)
  - Scott Long's research shows that a student's productivity depends on mentor's productivity
- have broad interests, be interdisciplinary
- your research should be T-shaped: broad, but also deep in one area

Greg Mankiw

- http://scholar.harvard.edu/files/mankiw/files/my_rules_of_thumb.pdf
- time management is key! extremely easy to mismanage time in research:
  - ask yourself how what you are doing now gets you to your goal
  - have strategy
- write well–see other slides; essp: simple, clean

Andrew Gelman

- "Advice on writing research articles"
- `http://andrewgelman.com/2009/07/30/advice_on_writi`
- be clear about your story
- give your paper to other people to read
- ask for comments

.

- start with the conclusions and work back to abstract

Gary King [do it at home]

- "Publication Publication" and some notes under:
- `http://gking.harvard.edu/papers`
- if needed, criticize others, but step on their shoulders, not their face
- [note: this is about replication; still some good ideas]

great references on academic writing

- clarity, simplicity, conciseness
- `http://amzn.com/0060891548`
- `http://amzn.com/1577660633`

outline

litRev:https://theaok.github.io/generic/howToGoogSch.html

opinion/activism v science

how do i produce a final project for this class?

presentations

final paper/project in general [NOT resMet]

a dillema: publishabe project or student project [NOT resMet]

inline response to comments on ps [qm*,dev,swb,dirStu]

links: good research in words of others [NOT resMet]

the end of theory, data is enough, airplane model [datMan, dirStu]

regression [qm2, dirStu]

data managemnt [datMan, dirStu]

GIS (todo)

wired article

- http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory
- again, we have data revolution
- unprecedented amounts of data about pretty much anything
- with so much data, we can just look at basic correlations
- without being too serious about theory !
- note: this is computer science approach to data analysis
- such view is not mainstream in social science

theory

- there is no reason to be very serious about existing theory
- theories are only valid untill proved wrong
- we need new theories
- remember "all models are false, some are useful"
  - our model and theory is *never* right
  - world is too complicated
  - we just want to show some useful pattern
  - that's all we can do
  - still, we want to be as close to the truth as possible

airplane model

- models replicate some of the useful features of real objects
- think of an airplane model
- there are airplanes models without windows
- and models that are too heavy to ever fly
- yet they are useful eg to test airflow in a wind tunnel
- but these models are not the same as airplanes
- (and nobody claims they are "true")
- but social scientists behave as if they have "true" models
- your regression model is always false, but hopefully useful

build new theories and models...

- because all theories and models are wrong, be creative
- come up with new theories in models
- don't take well established theories and models for granted just because they are out there for a long time and everybody uses them

...but do your homework

- cannot produce new theories if don't know the old ones
  - your new theory/model may already be old
  - (reinventing the wheel)
  - rather invent the new given the old–build on other's work
- you have to defend your theory/model
  - why is it important ? "so what ?"
  - how come millions of other soc sci did not get?
  - why they got it wrong ?
- again, all models/theories are wrong, some are useful
- also, some are better than others in terms of /creativeness/logic/argument/robustness

conclusion: theory and modeling

- think out of the box
- be creative
- do not use models only because everybody else uses them
- defened your approach

and remember that no model works all the time

- eg famous now professor couldn't get into PhD
- because his GPA was low,
- and model prodicted that people with low GPA cannot do well in PhD
- model works probably well most of the time, but as any model
- it sometimes fails

outline

litRev:https://theaok.github.io/generic/howToGoogSch.html

opinion/activism v science

how do i produce a final project for this class?

presentations

final paper/project in general [NOT resMet]

a dillema: publishabe project or student project [NOT resMet]

inline response to comments on ps [qm*,dev,swb,dirStu]

links: good research in words of others [NOT resMet]

the end of theory, data is enough, airplane model [datMan, dirStu]

regression [qm2, dirStu]

data managemnt [datMan, dirStu]

GIS (todo)

what next?

- now you know the basic and powerful tool of multiple OLS
- what next ?
- use it !
- turn your ideas into new theories and hypotheses
- and test those hypothses by regressing the outcome (Y) on your main X, controling for other X's
- do data support your hunch ? find out ...
- be creative ! being social scientist you don't have to study economic development or income inequality
- you can study happiness, culture, religion, terrorism, facebook relationships, and so forth

theory, logic, explanation

- again, you need to have some theory that makes sense and that is interesting for public policy/business/philosophers, etc...
  - and be as clear and simple as possible
  - eg "Wage is a function of education and experience; it is based more on merit than on privilege, such as race and gender." [see also Alesina's paper in few slides]
- do not say that you expect that "gender affect wage" etc...
  - why ? how ? so what ?

regressions

- again, do not overemphasize Rsq
- do *not* pick the models based on the Rsq !
- use beta coeff to compare magnitude!
- see code in 1.4 Multiple Regression

  `https://stats.idre.ucla.edu/stata/webbooks/reg/chapter1/regressionwith-statachapter-1-simple-and-multiple-regr`

regressions

- e.g.: "When controlling for union membership, experience is not statistically significant; and even if it were statistically significant, it's practical significance is negligible."

- This is great ! The coeff on exp is $< .1$ depending on specification; with .1 it means that 10 more years of experience (a lot !) would produce only 1 more $ per hour

regressions

- produce alternative models , eg merit v privilige
- but then always have a combined model with both to see which one is more important
  - is privilege affecting wages controlling for merit ?
  - is merit affecting wages controlling for privilege ?
- if both merit and privilege affect wages
  - (they do–we know it from theory and models)

  then if you run separate models, you have LOVB !

general coding practices

- clean data and save it as something else
  (never overwrite the original files)
- merge/append
- cleanup, save, and then for analysis start with clean:
  - have a final clean combined data file that you will use from now on
- then descriptive statistics
- and inferential statistics
- NOTE: in the course of coding code chunks will be all over the place – rearrange them

file formats again

- again, no Microsoft files
- stata code–can append at the end of paper
  - can post online
  - can have a separate .do file
  - but never have a dofile with a non .do extension
  - (unless it is an appendix in your paper)

dropping outliers

- if dropping outliers, always say why
- and may have an analysis including them in the appendix
  - if not sure... (unless it is obvious that outliers must be dropped)
- however, rarely anything is obvious in research
  - best try different options/do robustness checks...
- yet, there is obviously a time constraint

make it meaningful

- eg better have freq or perc for histogram
- avoid ugly graphs and tables: follow published examples!
- don't forget about the practical significance!

elaboration of the model

- start with a basic model
- possibly bivariate
- and have more columns adding more covariates as per theory
- the idea is that you test competing hypotheses/alternative explanations
- and in doing so show the robustness of your results

do the whole thing

- why study only counties in South Jersey
- or only libraries in Philly
- when you can study all of them!
- and at very least compare with your small n results

speculation/opinion

- this is not an op-ed
- there cannot be any speculation/opinion
- all statements must be supported by evidence
- evidence: literature or your own results

this is soc sci, not data sci

- in social science all models must be theory driven
  - (this is not true in statistics or data science)
- choice of variables, functional form (e.g. log) must be theory-driven
- you need to be explicit why you run a model that you run !

satisfy assumptions

- you *always* have to take care of assumptions
- e.g. heteroskedascity etc
- don't have to discuss in great detail in paper
- but have to have code–you need to show that you have done it!
- if you have many obs, no need to worry about normality
- also do not spend too much time on violations; just need to take care of them in dofile!

yet, another note on collinearity

- again collinearity is just a correlation between independent vars
- you can see it with `corr`
- some people say that you have collinarity if say correlation $>.9$
- you really have collinearity most of the time
- you can also use `vif`
- `www.nd.edu/~rwilliam/stats2/l11.pdf`

yet another note on BLUE

- what BLUE really means ?
- how estimators compare ?
- lets compare efficient/inefficient and unbiased/biased estimators
  - draw a picture (based on Kennedy)

organize

- descriptive stats goes before the regressions, not after (unless in the appendix)
- if descriptive stats is not very interesting (e.g. table of means and sd) just put it into the appendix
- instead of having alternative models, elaborate models
- figures and tables need captions and numbering
  - captions need to be very detailed so that you can understand table/figure from the caption only
  - axes need to be labeled in the figure
- have to refer tables/figures in text

contribute

- don't be modest !
- your paper needs to contribute to the literature
- it should be clear how it contributes
- again, explain:
  - how come nobody else did this before
  - or/and how come they got it wrong

get intuition; make it meaningful

- use beta coefficients
- use more descriptive statistics

cite data; replication replication

- data – you should clearly cite data
  - best give URL and authors and description
  - describe sample, time, sampling, etc
- your dofile should produce final results from the raw data
  - do not just send me the dofile with few regress
  - it should have all the commands you executed after loading the fresh data

interpret!

- beginning researchers usually do not spend enough time on interpreting the results
- there should be at least 1 page (12pt, double-spaced) of discussion
  - what have you found
  - substantive meaning
  - why does it matter
  - "so what ?"
  - limitations/future research

ols almost always useful; sometimes not best

- what data you have ?
- ols is good for cross sectional data only
- if you have panel or time series or dyadic/network data you need different models !
  - in this class it is fine, again ols will often give you reasonable results
  - but you should at least acknowledge the problems

paper

- let's have a look at Alesina's "Public Goods and Ethnic Divisions"

  `http://www.google.com/search?sourceid=chrome&ie=UTF-8&q=public+goods+and+ethnic+divisions`

- note:
  - nice elaboration/sequential models, eg TABLE III
  - well-developed theory–alternative explanations
  - multiple models
  - sensitivity analysis

another example

- `http://theaok.github.io/qm2/CassPortfolioPaper-FinancialLiteracy.pdf`
- skip nonlinear logit models!
- by a former student in this class
- note that it tells a story, it is interesting, engaging
- it contributes–we learn something new
- theory first, descriptive statistics second
- then regressions, interpretation and discussion
- last but not least, this paper looks polished and "publishable"

more examples

- https://link.springer.com/article/10.1007/s11205-011-9812-y
- https://link.springer.com/article/10.1007/s12232-015-0223-2
- http://journals.sagepub.com/doi/abs/10.1177/0042098016645470

- go through at least some of them and do ask questions if anything unclear
- also do read literature with OLS in your field, practice practice
- MQE is mostly about interpreting regressions!

practice interpretation

- http://link.springer.com/article/10.1007/s11482-014-9319-1
- what is worse for wellbeing: inequality or poverty?
- Tab1: note precise definitions of vars
- Tab2: some examples: be meaningful!
- Fig1, 2: des sta
- Tab3,4: coef, and std coef
- Discussion: gini ranges 32 to 60, if goes up by 6∗.5=.3*100k (in avg county): 30k unhealthy days
- causality: alternative explanations, reverse causality

practice interpretation

- `http://link.springer.com/article/10.1007/s11205-016-1327-0`
- 70s v 00s: 50% wider happiness gap: middle class v rich
- Fig1, Table 1: des sta
- Tab2: interactions
- Fig2: $\hat{Y}$
- robustness checks: eg Fig6, Fig10

outline

litRev:https://theaok.github.io/generic/howToGoogSch.html

opinion/activism v science

how do i produce a final project for this class?

presentations

final paper/project in general [NOT resMet]

a dillema: publishabe project or student project [NOT resMet]

inline response to comments on ps [qm*,dev,swb,dirStu]

links: good research in words of others [NOT resMet]

the end of theory, data is enough, airplane model [datMan, dirStu]

regression [qm2, dirStu]

data managemnt [datMan, dirStu]

GIS (todo)

do something useful

- do not just merge, loop, reshape, etc
  - for the sake of doing it
  - eg first split dataset, and then merge it back again
- playing is fine for learning and exploration
- but the final project must do something useful!

one-on-one

- again, let's work more one-on-one in second part of the class
- the idea is that by the end of the semester you will
  - develop a great dataset
  - understand your data really well (des stats, graphics)
  - and be able to change/expand your data easily
  - also be able to manage output (tables, coeff, graphs) easily

how do i cite data

- the most proper way
  - `http://www.bu.edu/datamanagement/background/cite/`
  - `http://libguides.lib.msu.edu/citedata`
  - `https://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/citations.jsp`
- the quick way way: just give url
  - you can also then load it directly into stata
  - but keep it on hardrive as well!
  - data on websites change and disappear

outline

litRev:https://theaok.github.io/generic/howToGoogSch.html

opinion/activism v science

how do i produce a final project for this class?

presentations

final paper/project in general [NOT resMet]

a dillema: publishabe project or student project [NOT resMet]

inline response to comments on ps [qm*,dev,swb,dirStu]

links: good research in words of others [NOT resMet]

the end of theory, data is enough, airplane model [datMan, dirStu]

regression [qm2, dirStu]

data managemnt [datMan, dirStu]

GIS (todo)

HOLLENBECK, J. R. (2008): "The role of editing in knowledge development: Consensus shifting and consensus creation," in Opening the black box of editorship, ed. by Y. Baruch, A. M. Konrad, H. Aguinis, and W. H. Starbuck, Palgrave Macmillan, New York NY, 1–12.

LUKIANOFF, G. AND J. HAIDT (2015): "The Coddling of the American Mind. In the name of emotional well-being, college students are increasingly demanding protection from words and ideas they don't like. Here's why that's disastrous for education'and mental health." The Atlantic.