Empirical Methods of Data Science

PROFESSOR MICHELLE LEVINE

WEEK 3: SCIENTIFIC METHOD, CONT'D

2/6/19

Today

In-depth look at a multi-modal study

Importance / unimportance (?) of external validity

Defining, and starting to understand, big data

How to conduct and write a literature review (Assignment 1)

Example of mixed-method approach

My research:
◦ Methodologies from cognitive psychology, linguistics and computer science

One area (Levine & Schober):
◦ What communication cues are necessary for different types of coordinated action?
  ◦ Face-to-face versus computer-mediated settings

◦ The following line of my research looks at *music*
  ◦ Often (metaphorically) compared to language:
    ◦ Scripted language = notated sections
    ◦ Conversational interaction = jazz improvisation

Importance of external validity

When and how does external validity truly matter?

Mook (1983)

◦ Separate out cases when we want to and when we don't need/want to generalize results from sample to population, or from lab to life.

  ◦ Ie, survey research aims to be generalizable.

  ◦ But does all research intend, or need, to be?

Mook (1983), cont'd

How researchers present their findings and what conclusions they draw is a key.

◦ "In many cases we are not using what happens in the laboratory to "predict" the real world. Prediction goes the other way: Our theory specifies what subjects should do in the laboratory. Then we go to the laboratory to ask, Do they do it? And we modify our theory, or hang onto it for the time being, as results dictate. Thus we improve our theories, and—to say it again—it is these that generalize to the real world if anything does."

Take the idea that samples should be representative of the population.

◦ Sometimes this is true, but other times this is false.

◦ Must evaluate sampling procedure based on the purpose of the study.

What is big data? Data science?

Big data: large data sets
- Text
- Audio
- Video
- Social media

Data science: interdisciplinary field that applies scientific methods to understanding big data
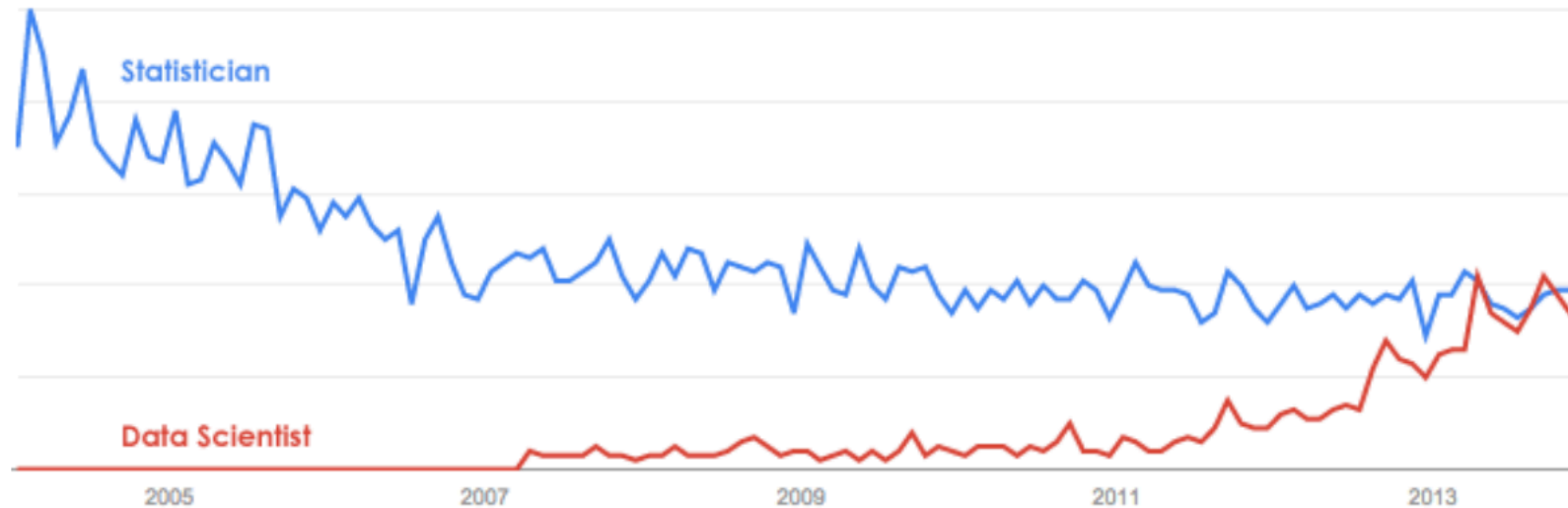
Big data

Big data

End of 2010: when "big data" started to register on Google searches

2011: explosion of interest

Data scientist surpasses statistician on Google Trends

Google Trends
https://flowingdata.com/2013/12/18/data-scientist-surpasses-statistician-on-google-trends/),
December 18, 2013

Impact of big data

What does our access to big data mean for research?
- ◦ Provides a tremendous amount of information w/o having to actively collect data from people
- ◦ BUT many times the data is still about people and can have a tremendous effect on people

E.g., transactional data (Burrows & Savage, 2007)
- ◦ Commercial and government databases that store routine transactions between citizens, consumers, business and government

E.g., social media data

Big data and the scientific methods

With big data, can we apply:

◦ Quantitative approaches?

◦ Qualitative approaches?

◦ Mixed-method approaches?

What does big data provide us with that traditional academic research does not?

Literature Reviews

- HOW TO CONDUCT A LITERATURE REVIEW
- HOW TO WRITE AN INTRODUCTION SECTION TO A PAPER

Tools to conduct literature review

Google scholar - https://scholar.google.com
◦ Focus search on articles published within last 5 years

Columbia University library electronic resources

Scientific writing (in general)

Strive for perfect clarity and simplicity of expression

Different from other types of writing assignments

Particular paper structure

Correct grammar is essential

◦ The following examples are taken from:

Goodwin, C. J. (2008) *Research in Psychology* (5th Edition). New York: Wiley.

Active vs passive voice

Active tense preferred
- The interviews were conducted by trained clinicians.
- Trained clinicians conducted the interviews.

"Significant"

Refers to description of results as being statistically significant

Avoid using it to indicate important or meaningful (use synonyms)

"Affect" vs. "Effect"

- Affect
  - As noun: means an emotional state

  Subjects demonstrated an increase in positive affect as a result of the warm weather.

  - As verb: means to influence

  The warm weather affected the students' level of participation in class.

- Effect
  - As noun: the result of some event

  The sudden change in the weather had an effect on the students' mood.

  - As verb: to bring about or accomplish some result

  The warm weather effected a change in the students' behavior.

Singular vs plural

Datum vs. data
- The data indicates that... (WRONG)
- The data indicate that... (RIGHT)

Analysis vs. analyses

Criterion vs. criteria

Phenomenon vs. phenomena

Stimulus vs. stimuli

Sections of a research paper

Title Page

Abstract

Introduction

Method

Results

Discussion

References

Tables

Figure Captions Page

Figures

Appendices

Introduction section

Think of your introduction as where you build a logical argument for conducting your experiment.

- Why perform this experiment?

- Why your hypothesis?

Structure

◦ Opening paragraph:
◦ Overall, invoke interest by explaining why this is an important topic of investigation.
◦ Include a general statement about the field of inquiry (general research question you are considering).
◦ May introduce your topic by pulling in a relevant real-world example.

Body:
◦ Cite other sources of information to explain logical basis for your experiment.
◦ Structure your paragraphs so that it leads the reader into understanding how your experiment extends, validates, disproves or replicates prior research.

Closing paragraph:
◦ State the specific goal of your study.
◦ Describe the variables you are manipulating and a statement of your hypothesis.
◦ Clearly indicate what results were expected and what they will mean
◦ (This links your introduction to the method and results sections)

Describing prior research

In some cases, you only need to mention a citation as an example of something you want to refer to.

In other cases, you will need to dive into the details to fully support what you are trying to say.

◦ In this case, you should include a brief overview:
  ◦ Overall purpose, rqs, hypotheses, method, results, conclusions
  ◦ Critiques and confounds

Citations and references

Ethical responsibility to cite where you read particular information.

Only cite work that you have actually read.

When possible, paraphrase rather than directly quote.

Two parts:
- Citation in text
- List of all sources on the References page

Week 4 readings

1) Milgram (1963) study

2) APA ethical principles (2017)

3) Galvan and Galvan – ch 3 and ch 10 (how to do lit review and write up)

4) Creswell and Creswell – ch 7 (how to form hypotheses)