Empirical Methods of Data Science

PROFESSOR MICHELLE LEVINE

WEEK 2: INTRODUCTION TO SCIENTIFIC METHOD

1/30/19

Today

Research approaches

- philosophical worldview
- Method
- Design

Key elements of scientific method

Why do we do research?

What is our end goal?

Examples?

Research approaches

Qualitative vs quantitative vs mixed-method

- Think of it as a continuum

Components in ***an approach***:

1. Philosophical worldviews
2. Distinct methods/procedures
3. Design

Research approaches

Qualitative

- words
- open-ended questions and responses
- case studies

Quantitative

- numbers
- closed-ended questions and responses
- Experiments

Mixed-methods

- incorporates elements of quantitative and qualitative research

Philosophical worldviews

Worldviews arise based on discipline orientations

- Postpositivism
- Constructivism
- Transformative
- Pragmatism

Postpositivism

Represents the traditional form of research

Lends to quantitative research

Challenges the traditional notion of the absolute truth of knowledge

- When studying human behavior and actions, we cannot prove anything in absolute terms

Deterministic philosophy

- Causes determine effects

Reductionistic

- Reduces ideas into small discrete sets to be able to be tested

Vs. constructivist

Lends to qualitative research

- Open-ended questions

Belief that we as humans develop subjective meanings that are key to understanding the world

Heavy emphasis on participants' view of the question at hand, and the interaction of participants (as we are social beings and thus understanding the world must be with this lens)

Vs. transformative Vs. pragmatic

Transformative

- Research needs to consider politics and political change
- Contains an “action agenda”

Pragmatic

- About what works − solutions and applications
- Uses all approaches

Research Designs

Once type of study decided on, then you need to decide on research design.

Research design refers to *type of inquiry or strategies of inquiry* within the research method chosen.

- Quantitative research designs:
  - These originated mainly in psychology
  - True experiments and quasi-experiments
- Qualitative research designs:
  - Narrative research
  - Phenomenological research

Experimental research

Approach used by postpositivists (quantitative research)

Rational empiricism

- Empiricism: gaining knowledge through observation
- Empirical questions: questions that can be answered through systematic observations and experiences

Scientific method: rules and techniques of observation that minimize errors allowed by simple observation

Scientific method

Steps to the scientific accumulation of knowledge

1. Begin with a question (usually prompted by theory) and form a *testable hypothesis*.
2. Gather evidence / collect data.
   - Hypothesis confirmed or disconfirmed.
3. Revise and conduct further studies.
4. Make findings public & therefore open to scrutiny.

Theory → Testable hypothesis

Theory: an idea or set of ideas that describe a particular event, process, or behavior

Testable Hypothesis: prediction that is formulated in such a way that observations are able to confirm or disconfirm

Gathering evidence

Correlational (non-experimental) research vs experimental research

What vs why

Correlations: do changes on a variable link to changes in another variable?

Correlations

If two variables are correlated, what does this tell us?

3 possible relationships:

- A causes B
- B causes A
- Some other factor causes both

Problem of directionality

Problem of third variable

Causation

Cannot make causal inferences, unless:

- A and B occur together with regularity
- A precedes B in time
- Theoretical explanation exists
- Other explanations can be ruled out

Ultimate goal of research is “causal”

Experiments clarify cause and effect by:

- Define and manipulate variable(s)
- Randomization
- Reliability measures
- Validity measures

Variables

Experimental manipulation: manipulate some variables, control other variables

- **Independent variable:** variable manipulated to measure its effect on the dependent variable
- **Dependent variable:** variable measured/recorded

Defining variables

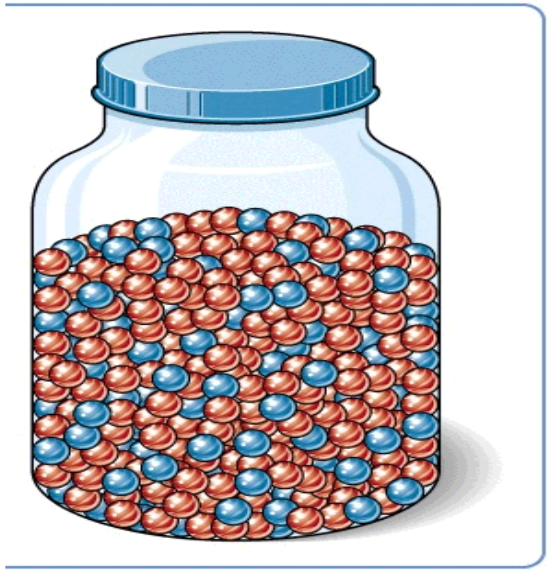
Operational definition:

- defining variable in concrete terms to make sure they are *reliable and valid*
  - Easy to measure physical properties
    - E.g., “distance”
  - Not so easy to measure psychological properties
    - E.g., “happiness”

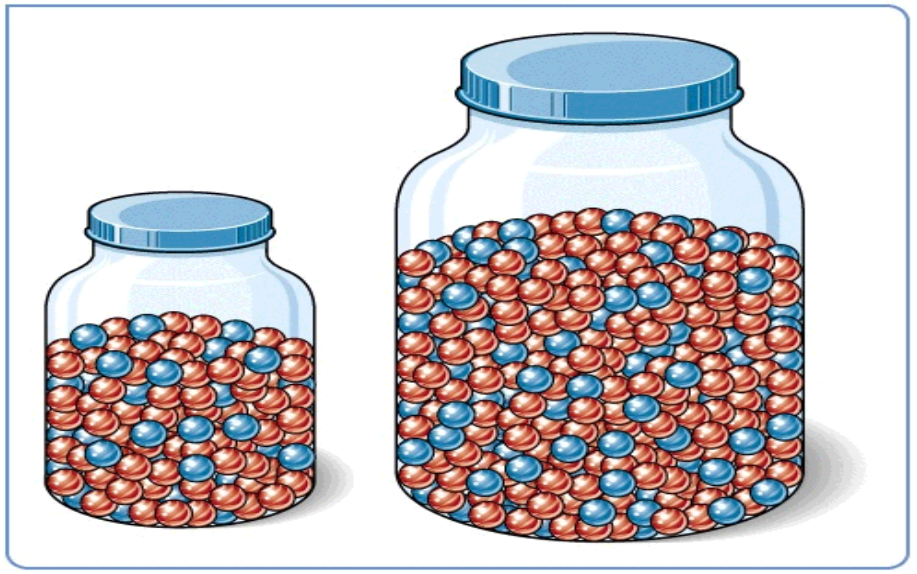
Randomization

Random sampling:

- each member of a population has an equal chance of inclusion into a sample (unbiased sample)



What is the fastest way to know about the marble color ratio in the jar?



Blindly transfer a few into a smaller jar and count them.

Randomization

Random sampling:

- each member of a population has an equal chance of inclusion into a sample (unbiased sample)

Random assignment:

- to experimental group vs control group
  - to different experimental groups

Between-subject design vs within-subject design

Reliability

Regarding research methods, refers to *reproducibility of data*

- Similar findings across trials/studies
  - E.g., The Mozart Effect

Another way to define it refers to *absence of measurement error*

- Focus is on they way researchers measure their variables
- Two forms:
  - Systematic
  - Random

Types of reliability

Test-retest

- Need to make sure observers are reliable
  - Interobserver reliability
- But what if still low test-retest? That is, what if there are participant factors coming into play.
  - Increase number of trials
  - More closely examine stimuli included – *internal consistency*

Validity

Are we researching what we think we are researching?

Four types of validity to consider:

- Construct
- Statistical conclusion
- Internal
- External

Questions to think about

Relationship between reliability and validity

- Does low in one automatically mean low in the other?
- Does high in one automatically mean high in the other?

How important is external validity?

Week 3

Read:

- Mook (1983). In defense of external invalidity.
- Schober (2006). Virtual environments for creative work in collaborative music-making.
- Gandomi & Haider (2015). Beyond the hype: Big data concepts, methods, and analytics

While reading Mook (1983), think about where you stand and what the opposing view is. We'll have an in-class discussion on this next week.

In-class assignment

Refer to page 19 in your Creswell & Creswell Chapter 1 reading.

Using the framework set in that chapter, and the concepts discussed today,

- For each the two research articles you read this week, answer the following:
  - What general theory were the researchers assessing?
  - What were the hypotheses? Independent variables? Dependent variables?
  - What approach was used? Anything in particular to note about the method?
  - What conclusions?
  - Any critiques? (take time on this…ask yourself, what does this truly answer? What would you do differently?)

- Come up with a research scenario. That is:
  - Think about a general theory you want to test.
  - Decide on a hypothesis/hypotheses (for this assignment, it does not need to be based on any theory or on prior work).
  - Go through each approach and weigh the pros/cons.
  - Decide which approach will best suit your needs.