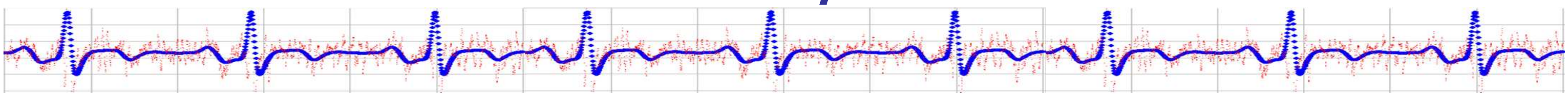
Empirical Research Methods in Information Science

IS 4800 / CS6350



Lecture 7

Finish Humans / Measures

Outline

- Homework
  - Ethnography observations
  - Lingering homework questions
- Ethics
- Next homework
- Finish research models
- Measures
  - Why worry about measures?
  - What should you consider in choosing a measure?
  - What is reliability and validity? Why care?

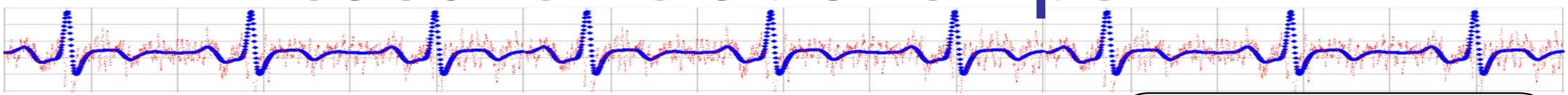
Ethnography

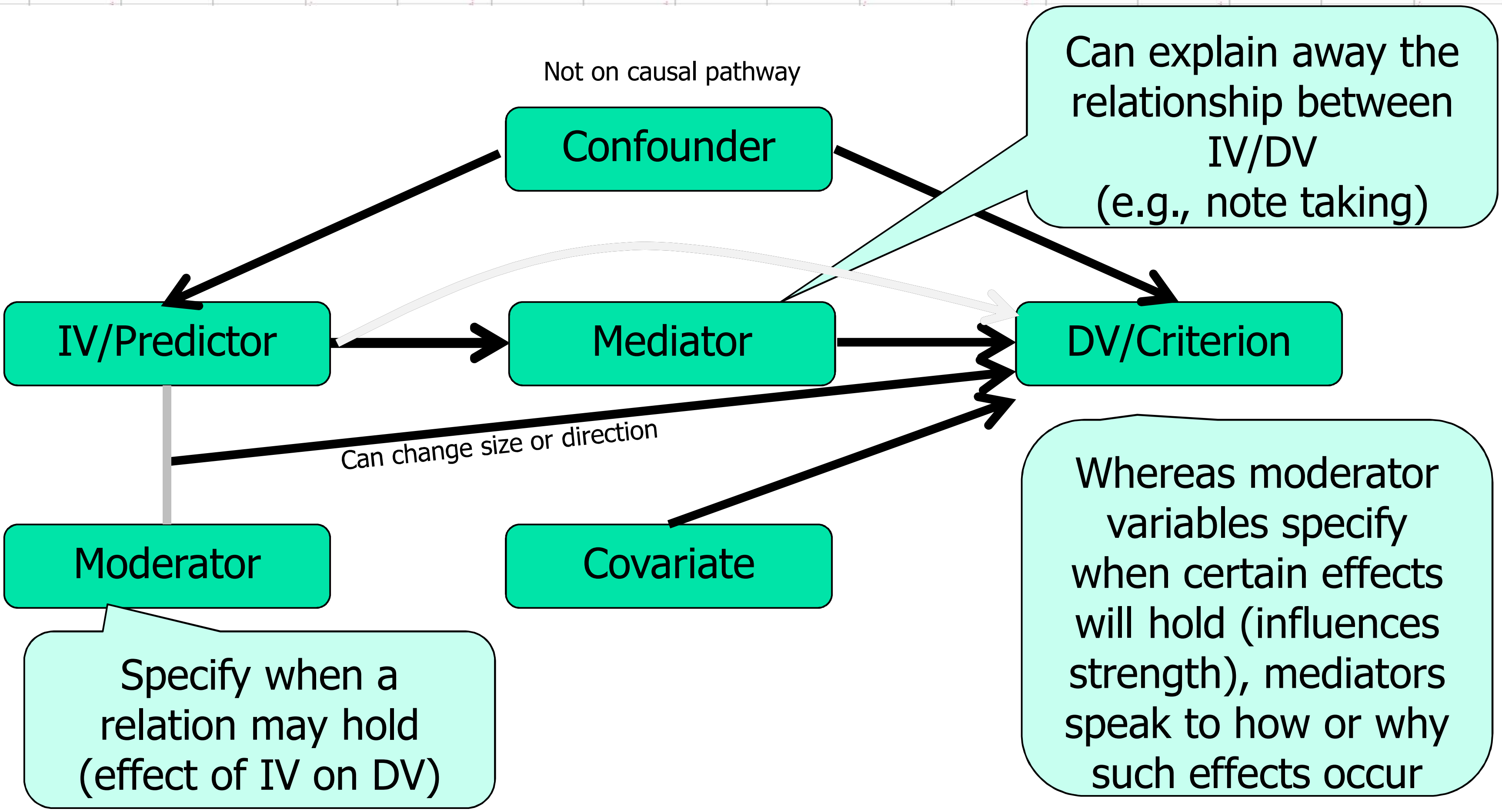
Observations?

I2b

- Identify your problem
- Identify your variables
  - Make lists
  - Think about relationships!

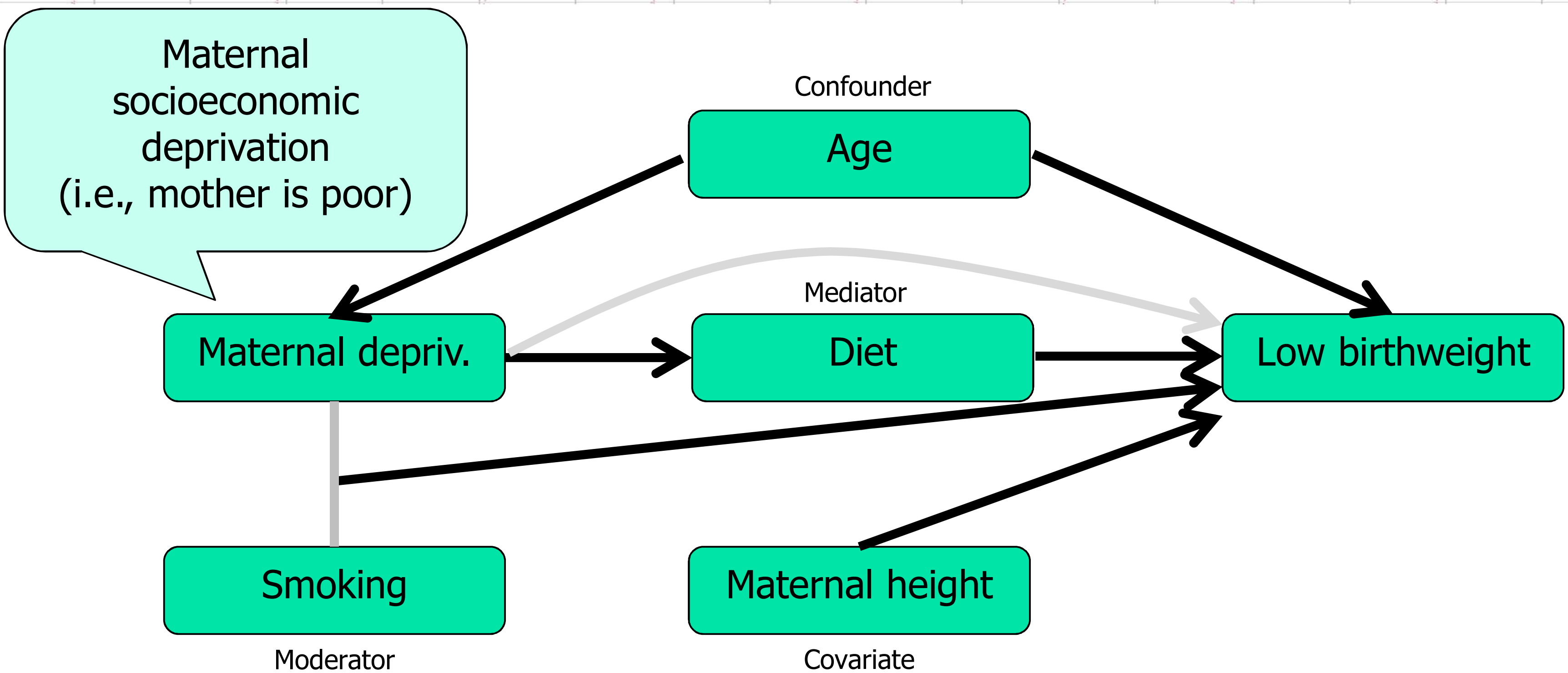
Measure relationships





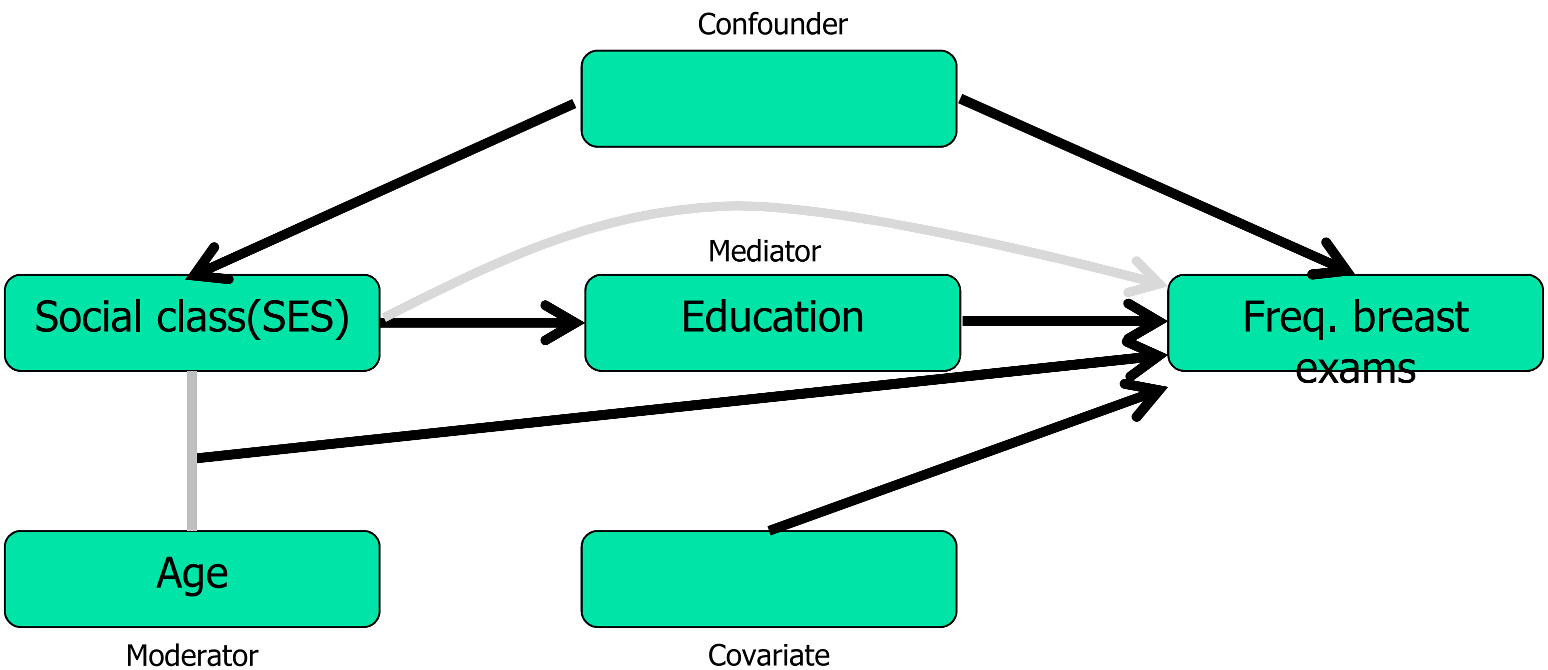
Example adapted from: https://significantlystatistical.wordpress.com/2014/12/12/confounders-mediators-moderators-and-covariates/

Measure relationships



Example adapted from: https://significantlystatistical.wordpress.com/2014/12/12/confounders-mediators-moderators-and-covariates/

Measure relationships



Example adapted from: https://significantlystatistical.wordpress.com/2014/12/12/confounders-mediators-moderators-and-covariates/

Remember, science done by people; people have flaws

- Types of research fraud
  - Outright fabrication of data
  - Altering data to make them “look better”
  - Sabotage of others’ work
  - Claiming credit for work done by others
  - Attaching your name to a study you had little to do with
- Ethical behavior is of utmost importance!

How do you avoid these problems?

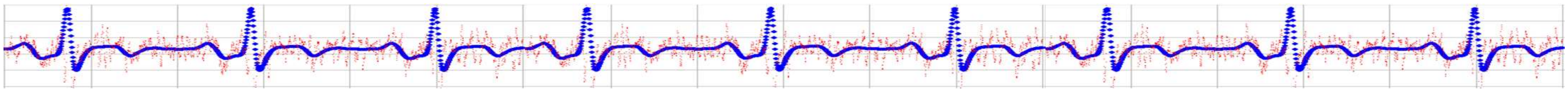
Fraud: Gray areas

- Selecting only the best data for publication
- Using the “least publishable unit” rule
- Dropping subjects (data points) without justification
- Fishing by design (using many outcome measures)
- Post-hoc fishing (methods, analyses)
- Running subjects until you get your result

Prevalence of fraud?

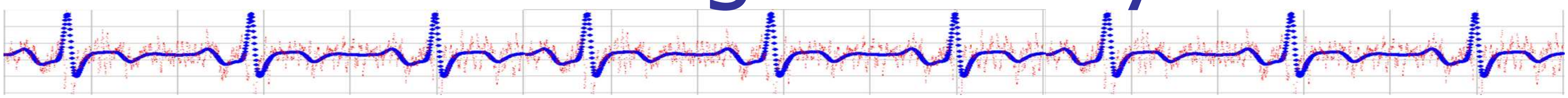
- We don’t know
  - Research fraud may be underreported
    - Many researchers who suspect fraud do not report it
    - There may be serious consequences for whistle-blowing
      - Vilified
      - Credibility called into question
      - May be fired
- But, even a few high-profile cases can be very damaging to science

What goes wrong with good scientists?



- Pursuit of funding for research
- Pressure to "publish or perish" for tenure
- Scientific elitism
- Breakdown of the review process

Integrity, especially when facing adversity



Must be at every scientist’s core!

Avoiding problems: Improving the process

- Train students early in their careers (you!)
- Note that perpetrators will eventually get caught
- Study cases of fraud: research requires honesty and the highest of standards
- Verify participant participation
- Encourage whistle-blowing

But science is not value-free

- Everyone has cultural, political, and personal biases
- Implicit bias: https://implicit.harvard.edu
- Values can influence:
  - Practices
  - Questions
  - Data interpretation (e.g., value-laden terms)
  - Specific and global assumptions

Homework I3

- Read example papers
- Pick an obscure piece of software with a user interface (ideally one you may have created for a class). Define two simple tasks using the software (something you can describe in 1-3 sentences and take less than three minutes to do) and write them down on two pieces of paper.

Homework I3 (cont.)

- Select two or more interval or ratio measures from pages 194-195 of the Nielsen reading that you think may be relevant to the software, in addition to at least one nominal or ordinal measure (could be sociodemographic)

Homework I3 (cont.)

- Ask three (or more) people to help you with a user study. Make sure they have not used the software before. Obtain consent using consent form. Provide a brief description of the software (but not how to use it). Then, give each participant each task and watch them attempt to complete it. **Do not provide any help**. Collect your measures.

Homework I3 (cont.)

- Submit a brief writeup of your test plan, descriptive statistics of your data, and any design recommendations resulting from your tests.

- I will provide you with an example assignment, an example consent form, and the template for your consent form

Example consent (start verbal)

"Hi, I'm conducting a survey to find out what people think about using [some cool software] to do [whatever people do with it.] ...ou to spend a few minutes ...are, and observe you as you ...e just [however many] minutes, in total. I will observe you and take notes and then ask you some questions. It's for a course I'm taking in Research Methods from Prof. Stephen Intille in the Khoury College of Computer Sciences. Your participation is voluntary and you can stop anytime and ask that your data not be used. Can you help me out with this?"

You will then get written consent.

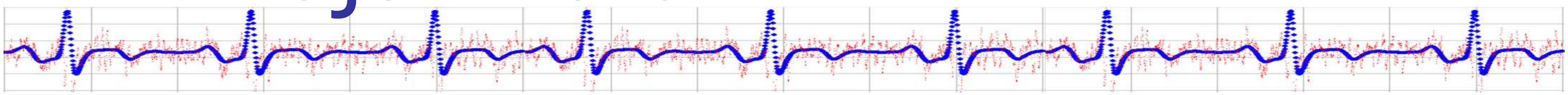
Rest of research models

Group exercise

Want to determine cleanliness of houses cleaned with Roomba

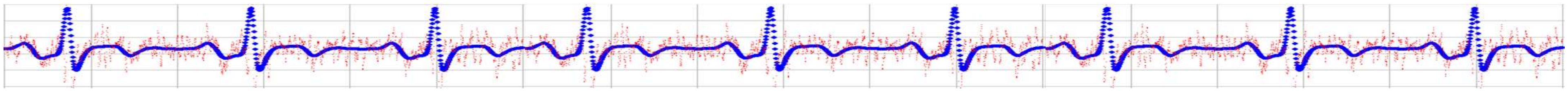
1. Design a descriptive study
2. Design a demonstration study
3. Design a correlational study
4. Design an experimental study

Correlational research: Major features



- No independent variables are manipulated
- Two or more dependent variables are measured, and a relationship is established
- Correlational relationships can be used for predictive purposes
  - A PREDICTOR VARIABLE can be used to predict the value of a CRITERION VARIABLE
- Correlational research cannot be used to establish causal relationships among variables
  - THIRD VARIABLE PROBLEM
  - DIRECTIONALITY PROBLEM

Correlational research: When is it used?



- You are gathering data in the early stages of research
- Manipulating an independent variable is impossible or unethical
- You are relating two or more naturally occurring variables
- You don’t have subjects or other resources to run an experimental study
- You are doing retrospective data analysis

Experimental research: Major features

- An independent variable is manipulated (with at least two levels)
- A dependent variable is measured
- *The most basic experiment consists of an experimental and a control group*
- Control is exercised over extraneous variables either by holding them constant or by randomizing their effects across treatments
- A causal relationship between the independent and dependent variables can be established

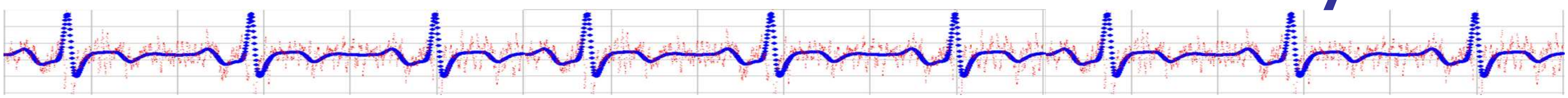
Strengths and limitations of experimental research

- Strength:
  Identification of causal relationships among variables (not possible with correlational research)
- Limitations
  - Can't use experimental method if you cannot manipulate variables
  - Tight control over extraneous variables limits generality of results (control/generality tradeoff)
  - Most expensive & difficult to do

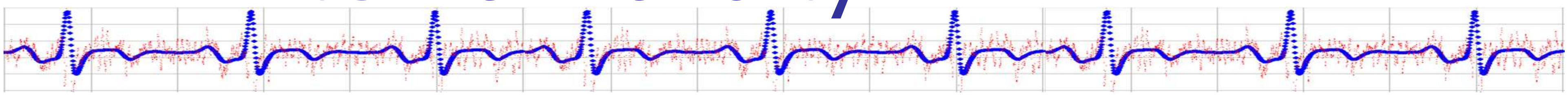
Types of study designs

	Number of Variables	Number of IV Levels	Manipulation
Descriptive	1	NA	NA
Demonstration	≥ 2	1	√
Correlational	≥ 2	NA	NA
Experimental	≥ 2	≥ 2	√

Internal vs. external validity?



Internal validity

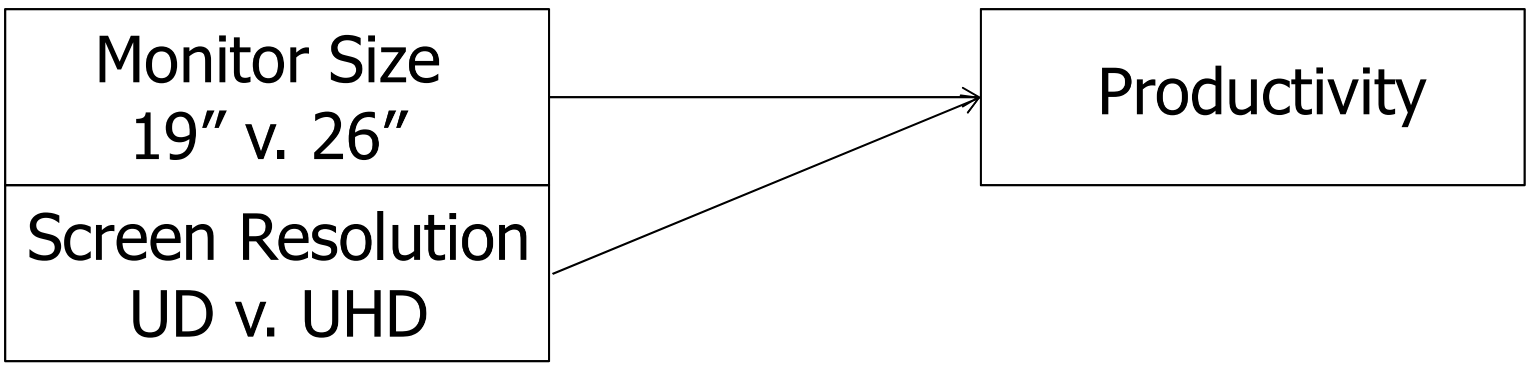


- INTERNAL VALIDITY is the degree to which your design tests what it was intended to test
  - In an experiment, internal validity means showing that variation in the dependent variable is caused *only* by variation in the independent variable
- Internal validity must be considered during the design phase of research
- Internal validity is threatened by CONFOUNDING and EXTRANEOUS VARIABLES

Internal validity

- Confounding variable – effects from IV cannot be separated
  - E.g., Larger monitors have higher-quality speakers, which could impact Performance and Satisfaction
- Extraneous variable – may impact behavior being investigated (DV), but not of interest in experiment
  - Hold constant
  - Randomize across treatments
  - E.g., Seniority impacts Performance

Example confounder



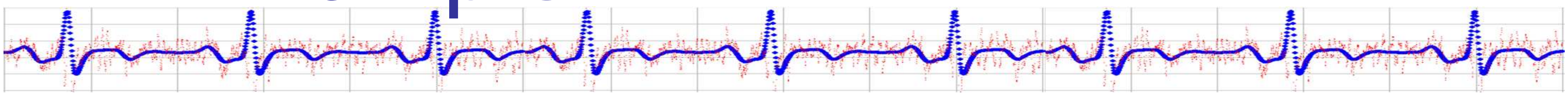
Factors affecting internal validity

History	Events may occur between multiple observations.
Maturation	Participants may become older or fatigued.
Testing	Taking a pretest can affect results of a later test.
Instrumentation	Changes in instrument calibration or observers may change results.
Statistical regression	Subjects may be selected based on extreme scores.
Biased subject selection	Subjects may be chosen in a biased fashion.
Experimental mortality	Differential loss of subjects from groups in a study may occur.

External validity

- Degree to which results generalize beyond your sample and research setting
- Threatened by the use of a highly controlled laboratory setting, restricted populations, pretests, demand characteristics, experimenter bias, and subject selection bias
- Steps taken to increase internal validity may decrease external validity and vice versa
- Internal validity may be more important in basic research; external validity, in applied research

Example:



- You want to evaluate a new sensor to detect whether people are happy or not.
- You hire actors and randomly assign them to act happy or sad, and test your sensors on them.
- What kind of validity (internal/external) might be challenged?

Example:

- You conduct the "Conversational Agents to Promote Health Literacy" study by assigning the first 30 patients who volunteer to the intervention group, and the next 30 to the control group.
- What kind of validity (internal/external) might be challenged?

Factors affecting external validity

Reactive testing	A pretest may affect reactions to an experimental variable.
Interactions between selection biases and the independent variable	Results may apply only to subjects representing a unique group.
Reactive effects of experimental arrangements	Artificial experimental manipulations or the subject's knowledge that he or she is a research subject may affect results.
Multiple treatment interference	Exposure to early treatments may affect responses to later treatments.

Research settings

- Laboratory
  - Affords greatest control over extraneous variables
  - Simulations
    - Attempt to recreate the real world in the laboratory
    - Realism is an issue
- Field
  - Study conducted in a real world environment
  - Manipulate variables in the field
  - High degree of external validity, but internal validity may be low

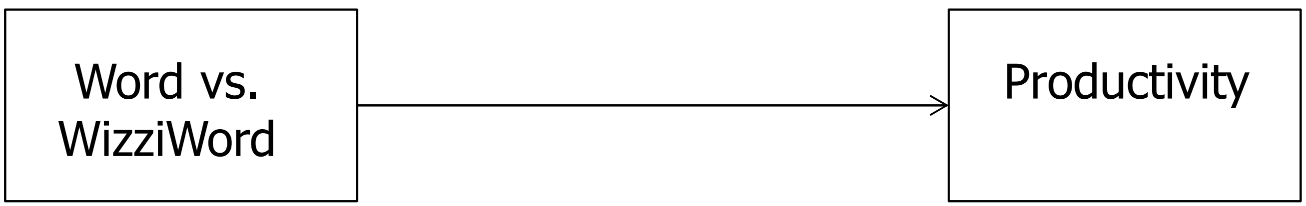
Proving causality with experiments

What's required?

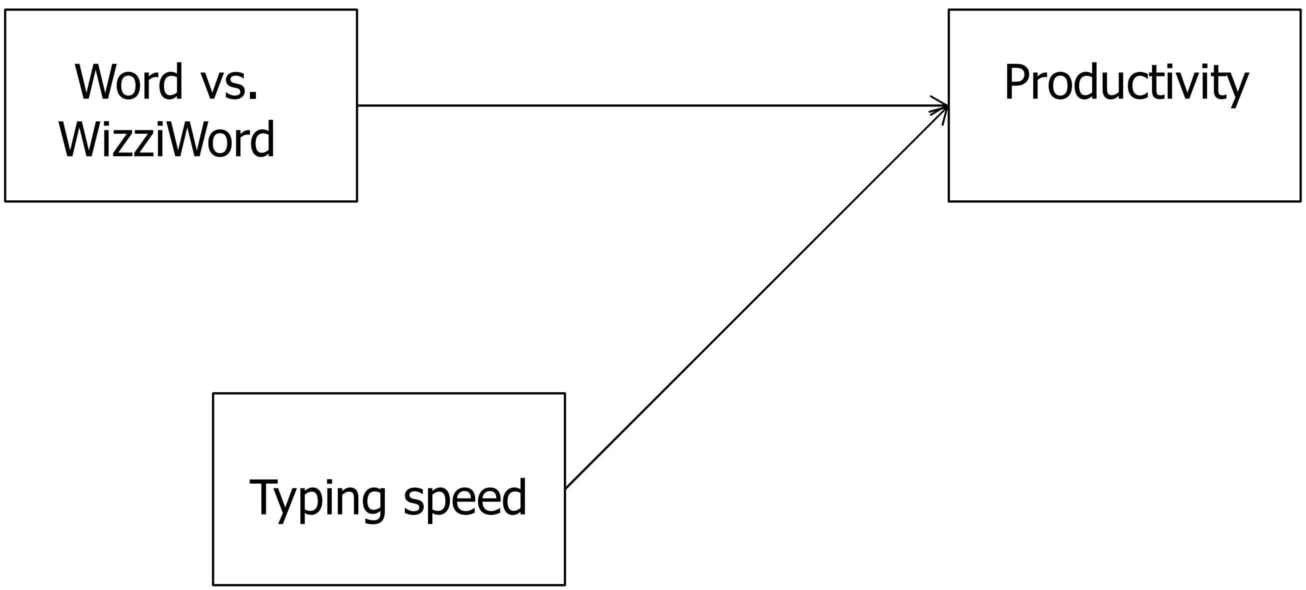
- Must manipulate the world
- Must measure an outcome/effect
- Must control extraneous variables
  - Fix
  - Randomize

What do we mean by randomize?

One example: Two treatment, between subjects design



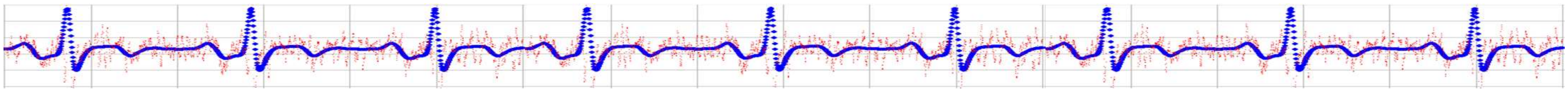
Why does randomization help?



Why does randomization help?

- On average
  - As many fast typists using WW as W
  - As many slow typists using WW as W
- The effect of typing speed “averages out” across conditions, thus is not a confound (does not systematically vary with IV)
- Same should be true for all other extraneous variables

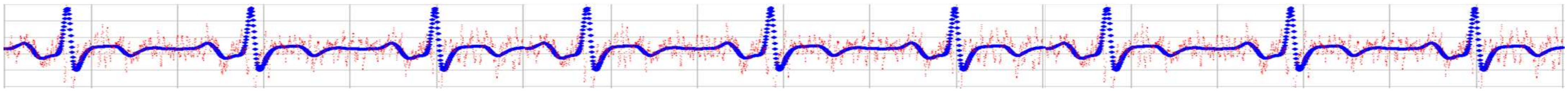
Making systematic observations



Remember your first science labs?

What to measure / how to measure it?



Given the choice, use a measure that…

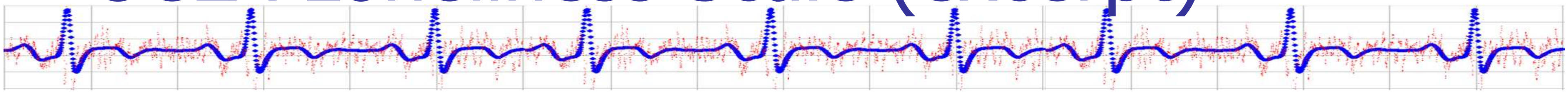
- Is "validated"
- Has been used before in your field
- Is readily accessible or inexpensive
- That takes the least time and effort

What is a validated measure?

- Has reliability
- Has validity

- For questionnaire measures, these are collectively referred to as a measure's "psychometrics"

Example 'Composite Scale Questionnaire' UCLA Loneliness Scale (excerpt)



1. I feel in tune with the people around me.

NEVER RARELY SOMETIMES ALWAYS

2. I lack companionship.

NEVER RARELY SOMETIMES ALWAYS

3. There is no one I can turn to.

NEVER RARELY SOMETIMES ALWAYS

4. I do not feel alone.

NEVER RARELY SOMETIMES ALWAYS

5. I feel part of a group of friends.

NEVER RARELY SOMETIMES ALWAYS

Measure reliability

- A reliable measure produces similar results when repeated measurements are made under identical conditions

- Reliability can be established in several ways
  - Physical measures = repeatability
  - Behavioral measures = inter-rater reliability
  - Questionnaire measures …

Questionnaire Reliability

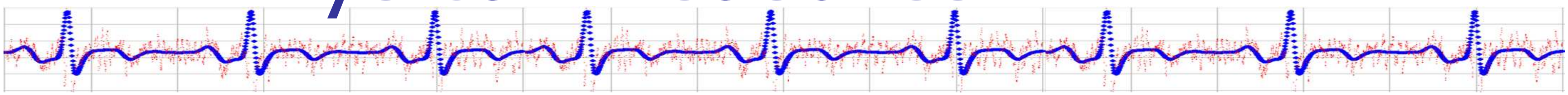
- *Test-retest reliability:* Administer the same test twice (or many times)
- *Parallel-forms reliability:* Alternate forms of the same test used
- *Split-half reliability:* Parallel forms are included on one test and later separated for comparison

Questionnaire reliability

Problem with this?

- For questionnaires using multiple questions to assess the same underlying factor, this also encompasses *internal consistency:*
  - Do all of the questions address the same underlying construct of interest?
  - That is, do scores co-vary?
  - A standard measure is Cronbach's alpha
    - 0 = no correlation
    - 1 = scores always co-vary in the same way
    - 0.7 considered "good"

Physical measures



- Length, weight, time, temperature, etc.
- Reliability = *precision*
  - Range of variation to be expected on repeated measurement
  - Reflected in amount of information in each measure (level of detail)

Measure validity

- A valid measure measures what you intend it to measure
- Carefully consider when indirectly measuring something (e.g., IQ test)
- For questionnaires, establish validity multiple ways…

Measure validity

- *Face validity:* Assessment of adequacy of content; Least powerful method
- *Content validity:* How adequately does a test sample behavior get measured?

Measure validity

- *Criterion-related validity:* How adequately does a test score match some criterion score? Takes two forms:
  - Concurrent validity: Does test score correlate highly with score from a measure with known validity?
  - Predictive validity: Does test predict behavior known to be associated with the behavior being measured?