Designs

February 17, 2010
Pedro Wolf

Today

- Sampling
- Correlational Design
- Experimental Designs
- Quasi-experimental Design
- Mixed Designs
- Multifactioral Design

Sampling Overview

- Sample- A subset of a population
  - Descriptive statistics (mean, standard deviation....)
- Population- “an exaustive group of objects having one or more common characteristics” (Klugh, 1974)
- Because a population can be so big, we have to use the following:
  - Inferential statistics (for you now correlations and t-tests)
  - Sampling
  - Design

What Makes a Good Sample?

- You want the sample to be “representative” of the population
  - Your sample is in essence a model of the population and should share the characteristics of that population
- It depends on the following:
  - What population are you trying to generalize to?
    - If I'm interested in gerontology do I want a representative sample from all humans?
    - If I am interested in fear what can I sample from?

Sampling

- Fear question
  - How about rats?
  - How about dogs?
  - Don't rats, dogs, etc share a common characteristic one may call fear?
- You don't care about people but the characteristics they may carry that you are interested in studying
  - The characteristic is what determines your population, and your population should determine your sampling

Sampling

- How do we get our sample to represent the population?
- How do we choose who to include in our study?
- Can I just call a bunch of people in my phone's contacts list and believe I have an adequate sample of humans
  - Probably not, probably a sample of my friends and family that I like enough to save in my phone, and maybe one person who I saved so I know not to answer.
- Improper sampling can deceive us.
- What do we do? How do I know who is more characteristic of the population than another?

Sampling

- We don't know
- We have some procedures to help
- Simple Random Sampling-
- You set up a procedure where:
  - Each individual carrier of the characteristic is equally likely to participate in your experiment
  - We achieve this by “randomly” sampling from our population of interest
- Problems- Difficult to do
- We don't have access to the entire population
  - I and most scientists don't have the resources to sample from all people, animals, or things who carry our characteristic of interest

Sampling

- Cluster or Multi-Stage Sampling- A two step method where you make it easier.
  - You take a given geographic area/s or place/s
  - This is in essence a sample of all possible places
  - Then you sample respondents from the first stage
    - Usually randomly sampling from there
- Although it can introduce error to your study this is a much more inexpensive and doable sampling technique when compared to simple random sampling

Matched Random Sampling

- A method of assigning participants to two groups
- You first match participants as closely as possible
  - Age, SES, Ethnicity, Identical Twins...
- Then randomly assign each individual to the two groups
- You do this to reduce the initial between group variance (standard deviation squared)

Designs

- No one design is perfect
- They all have their strengths
  - discovering potential causes
  - disconfirming hypotheses
  - cheaper to run
  -
  -
  -

The Correlational Design

- At its most basic level a correlation is testing the strength of the relationship between two variables
- How do we get our hands on those variables?

1) You try your best to obtain a representative sample
   - If it is not representative whatever inferences you make are questionable, and could be full of error
   - error in this case its non-random things you didn't measure which may influence your results

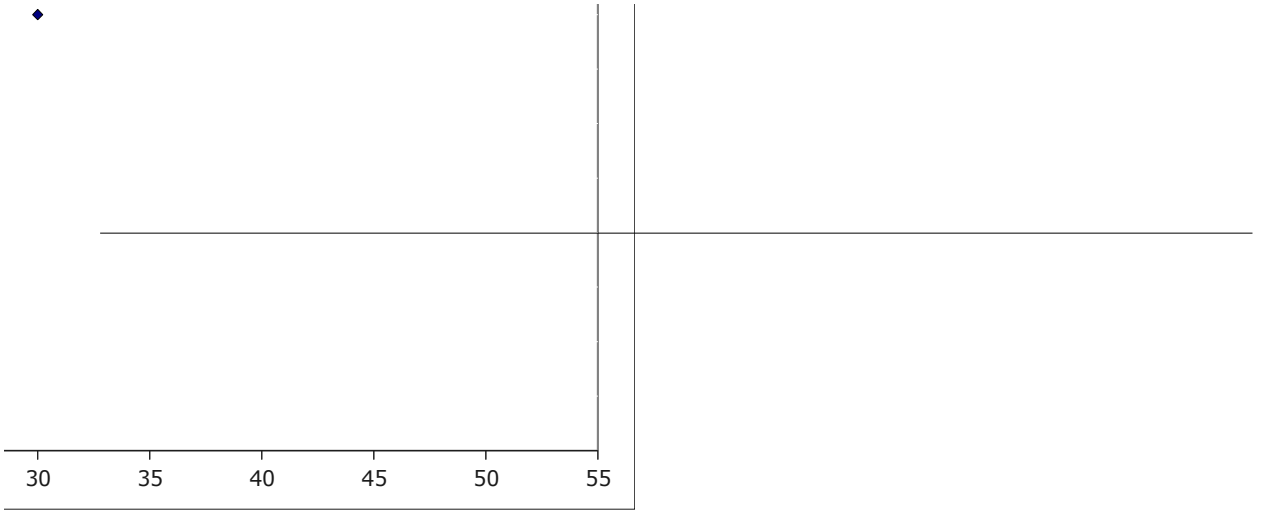
2) You try your best to measure both of the characteristics of interest for each participant
   - What happens if you only measure one characteristic in one participant and not the other?
   - For now you chuck it, you wasted both your and the participant's time
   - Your representative sample is now not so representative

3) You run the appropriate correlation on the two variables

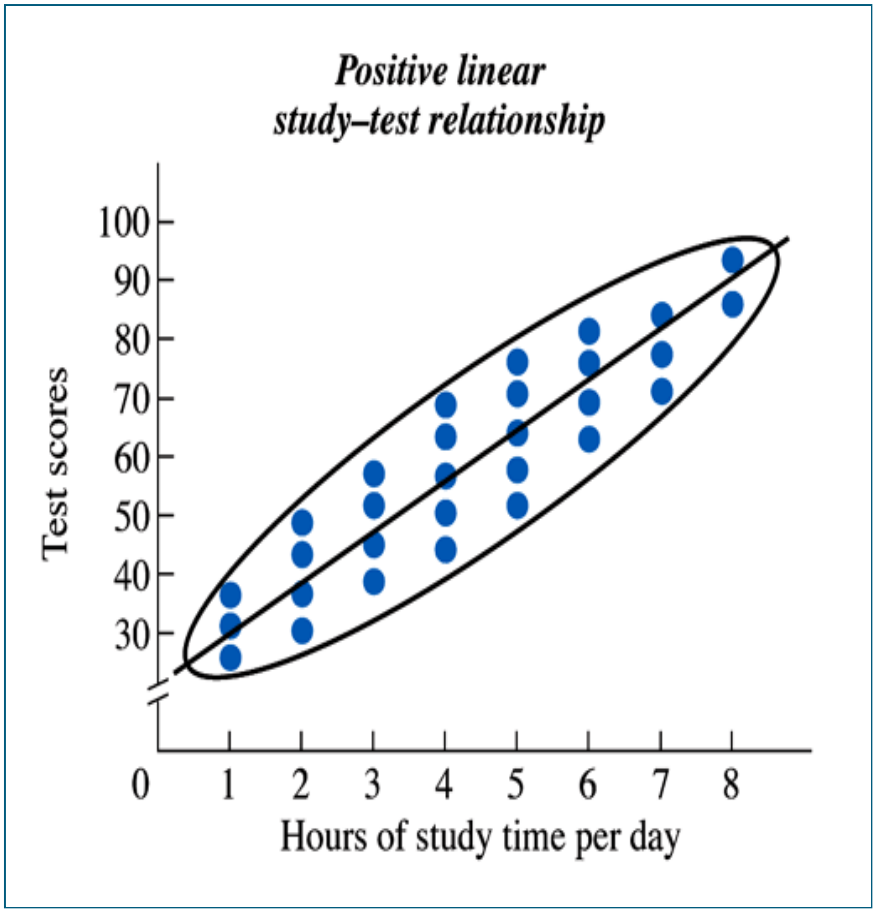
Correlations

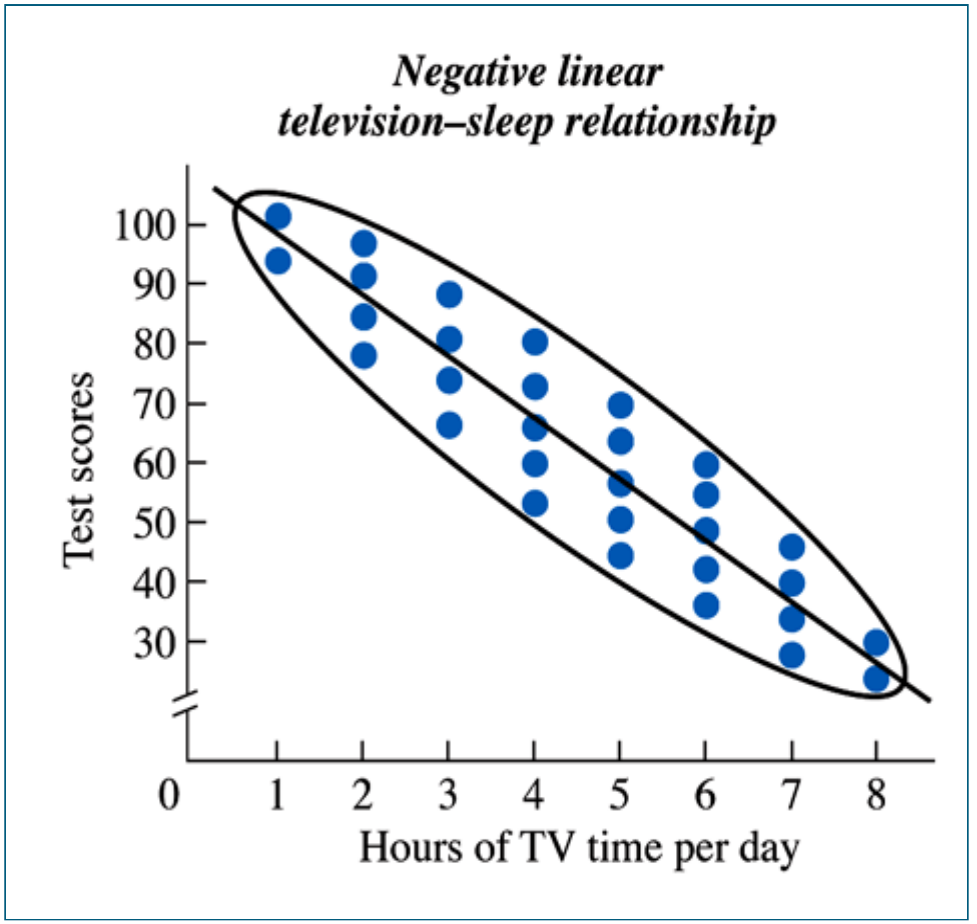
- Correlation does not imply causation
  - causation does imply correlation
- Correlational research is at its best when attempting to discover potential causes
- At a weaker level it supports discomfirmable hypotheses
- Cannot be used to imply causation
- Once you find a correlation the next step is to start running experiments

The Null Hypothesis



Alternative Hypotheses





Classic Experimental Design

- Your target is to have two groups with the following characteristics:
  - They are exactly the same in all respects except for one
  - The one difference is the characteristic you think is the cause (independent variable) of a phenomenon of interest (dependant variable)
    - e.g. you think exercise decreases depression
    - depression is the phenomenon
    - exercise causes a decrease in it
  - Finally you the experimenter controls the cause

How do we do that?

- First we sample
- We want to get the two groups both as identical and as representative as possible
  - the identical as possible ensures that are conclusion tests causality
  - the representative as possible ensures that it is generalizable
- More often than not you can't do simple random sampling
- You can do do cluster or multi-stage sampling
  - the two groups might have differences from the beginning which are due to chance alone
- You can do matched sampling
  - it may increase the likelihood that your groups are identical
  - it may make it less generalizable or introduce error
  - what if you matched on unimportant characteristics?
- In the end it is up to you, no one study is perfect.

Now What?

- We have our two samples and we are pretty sure they are as identical as possible
- We have to have a way to introduce the cause to one of the groups
  - make one of the groups exercise
- You measure each group's depression
- If depression in the group is lower in the group that exercised than we have some confidence that our hypothesis may be true

Okay how do we know which one is higher?

- We analyze the data
- In this case we have two groups (two groups run a t-test)
- We want to see if the mean of the experimental group (the one with the cause) is different from the control group (the one without the cause)
- The t-test tests the null hypothesis (the two groups are from the same population)
  - If the null hypothesis is rejected you can infer that the manipulation caused the two groups who were from the same population initially to suddenly be from two populations
  - If the null hypothesis is not rejected your hypothesis is not supported, the two groups appear to have come from the same population and your manipulation did not cause them to appear different statistically

Within Subjects Designs

- A variant of the experimental design is one where the same participants are used in both groups
- In the control group the participants are measured on the dependent variable.
- This serves as a baseline
- The independent variable (potential cause) is then introduced in the second condition
- The dependent (outcome) variable is again measured

Within Subjects Designs (repeated measures)

- The null alternative hypotheses are still the same
  - the two means are the same
  - the two means are different
- Problems with the design
  - practice effect- it could be that just being measured the first time (a questionnaire) causes the two groups to have different means
- Do a within participants t-test (ask how if you do this)

Another Within Subjects Design

- The ABA design
- You measure the same people three times on the dependent variable
  - first time no independent variable
  - second time introduce the independent variable then measure
  - third remove the independent variable then measure
- If the means across groups stay the same there is no effect
- If the means continuously change in the same direction you may have a practice effect
- If it goes up with the independent variable and goes back down when you take it away well you just may have an effect
- Statistics- more than two groups, we'll cross that bridge when we get there, use an ANOVA

The Quasi-experimental Design

- It's kind of like a mixture of an experimental design and a correlational design
- Like an experiment you have two groups
- Like a correlational design you have two variables which are naturally occurring in the world (you don't control it)
- Weakness- you analyze it like an experiment but you can't infer causation
- Strengths- when the cause is not controllable you can still study it

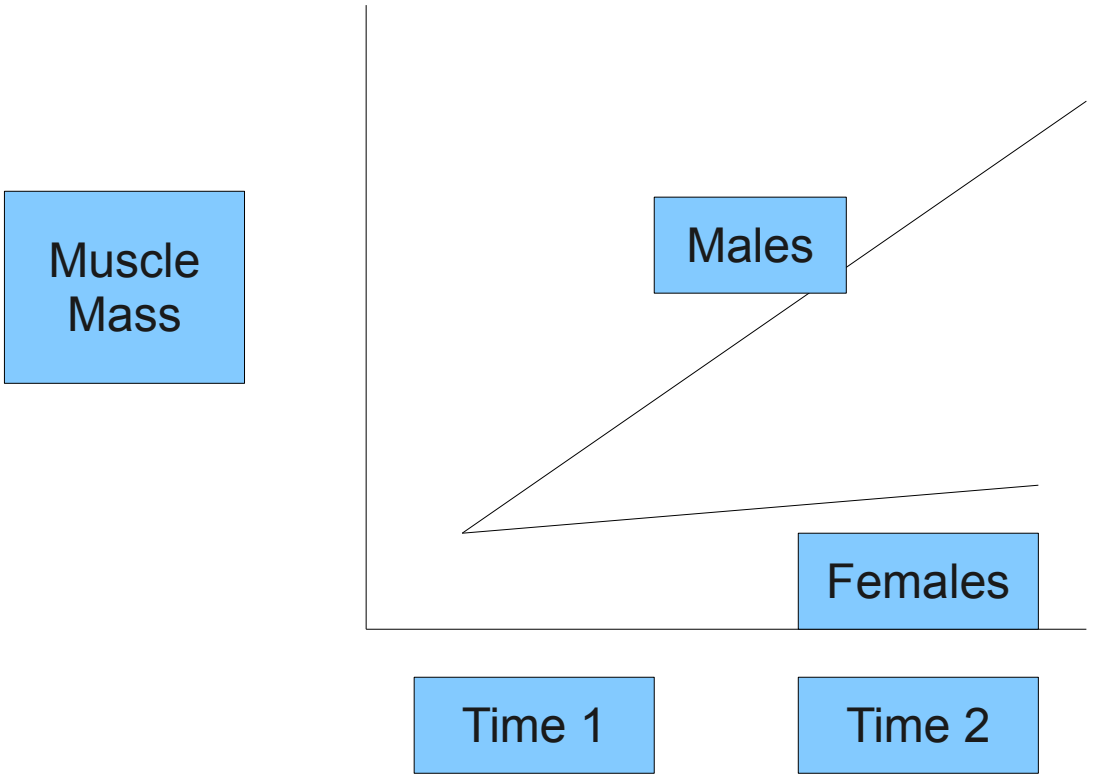
Mixed Designs

- a mixture of correlational and experimental techniques
- you set up the classic experiment
- you measure characteristics besides your dependent variable which you think may influence the dependent variable
  - e.g. sex, I can't manipulate your sex but it may be very important to my dependent variable
- I then statistically “control” for that extra variable(e.g. sex)
- I analyze this design using the General Linear Model (GLM). Don't worry, no GLM for you unless you ask for it

Multifactorial Designs

- Two independent variables
- Four groups
- Tests for interactions
  - If you think that it takes a combination of two variables to cause a change in your dependent variable
  - e.g. being a male and lifting weights for 8 hours a day for 16 weeks increases muscle mass more than if you were female
- If you can manipulate both it increases the power of the inferences

What does an interaction look like



Homework

- For next week turn in paper version of Introduction next week at the beginning of class