Experiments and Observational Studies

Grinnell College

November 2025

Motivation: Final Project

A critical part of your final will be a "self-reflection" part where you discuss...

- describe the data collection method
- the limitations of your data collection method
- what, if anything, you'd like to change if it was redone

Data Collection Methods

Broadly there are two major ways to collect data

Experiments where the variables of interest are manipulated (hopefully intelligently)

- Handing a subject a bean or letting subject choose the bean
- Assigning different fields different new soybean hybrids

Observational studies is where the researcher does not intervene in/manipulate the scientific process

- Phone screen time generally and on their favorite app (retrospectively)
- Asking someone how long they can plank for and then recording their time (prospective)

Observational Studies

Broadly two types...

- Retrospective studies
  - Looks backwards
  - Usually considered "weaker"
  - Asking about nutrition as a child for studying adult heights
- Prospective studies
  - Looks forward
  - Continues to take measurements into the future
  - Stronger than retrospective studies but weaker than experiments
  - Medical studies where they follow people around for years....

Causation: Right out

No control over variables and random assignment; it just *is*

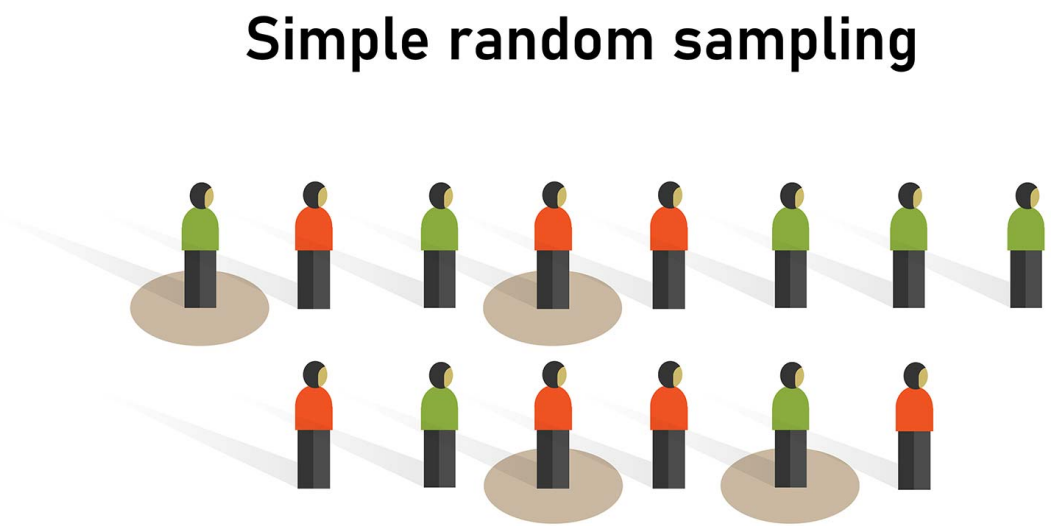
Better to talk about correlation

(Applies to experiments as well)

Survey Methods (the good ones)

Simple Random Sample: All population members have an equal chance of being chosen

- Most basic form
- Eg Grinnell sends an email survey to a random sample of students on LGBT acceptance/discrimination

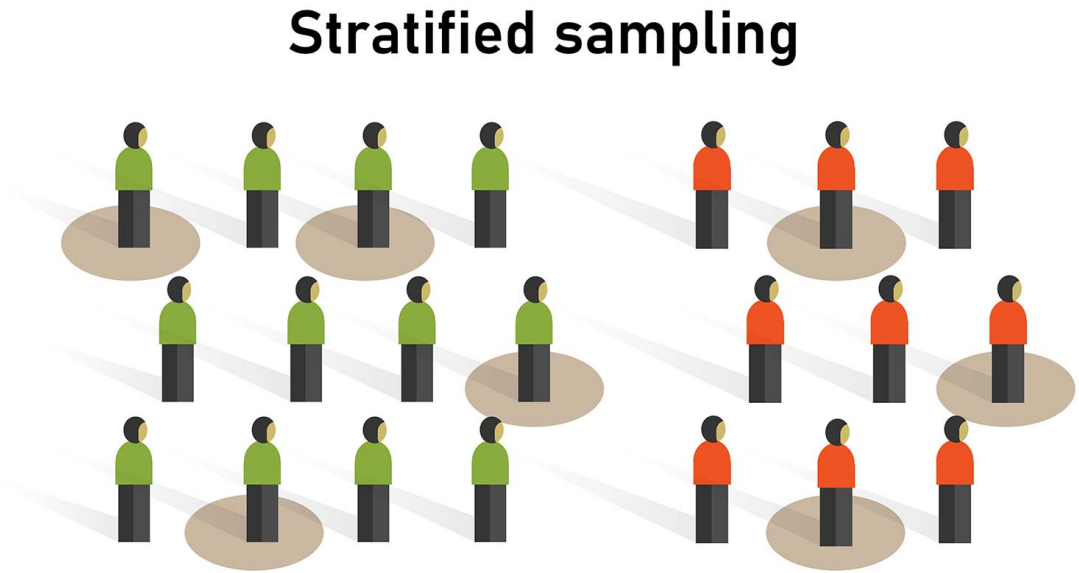


Survey Methods (the good ones)

Stratified Sampling: Simple random sampling occurs within each subpopulation

- Ie The population is broken into groups and then each group gets sampled
- Members of a subpopulation of the same probability of being picked
- Members do not have the same probability of being picked across/between subpopulations
- Eg Grinnell sends an email survey on LGBT acceptance/discrimination to..
  - LGB Students
  - Trans/NB Students
  - Het-Cis Students
- Ensures all subpops are represented

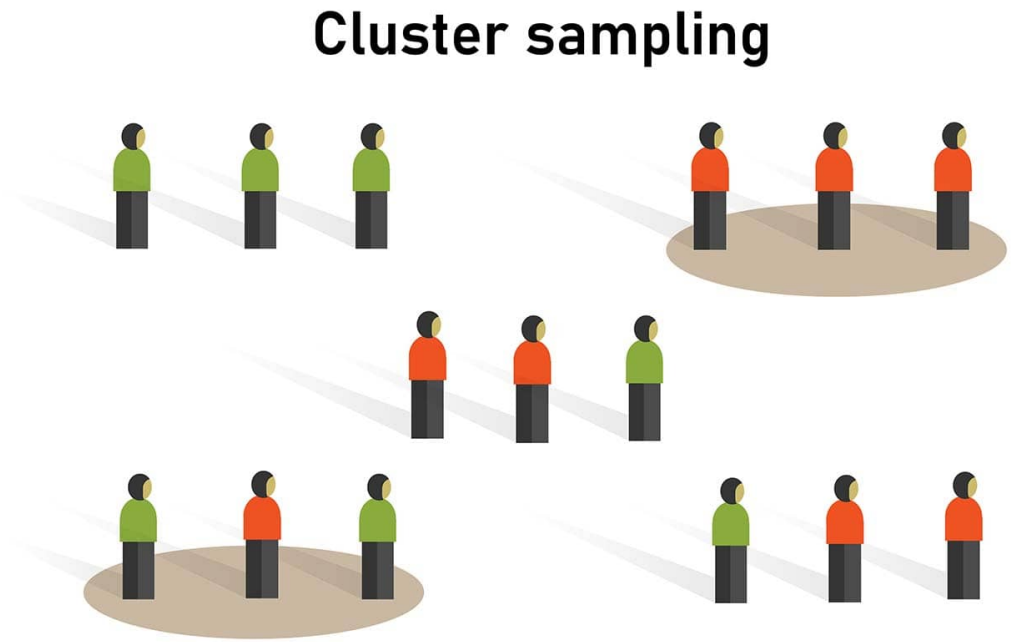
Survey Methods (the good ones)



Survey Methods (the good ones)

Cluster Sampling is where all members in a few random (usually tiny) subpopulation are studied

- Exaustive study of a few smal things
- Eg Grinnell surveys by physical paper 5 randomly selected classes, all students required to respond



Survey Methods (the good ones)

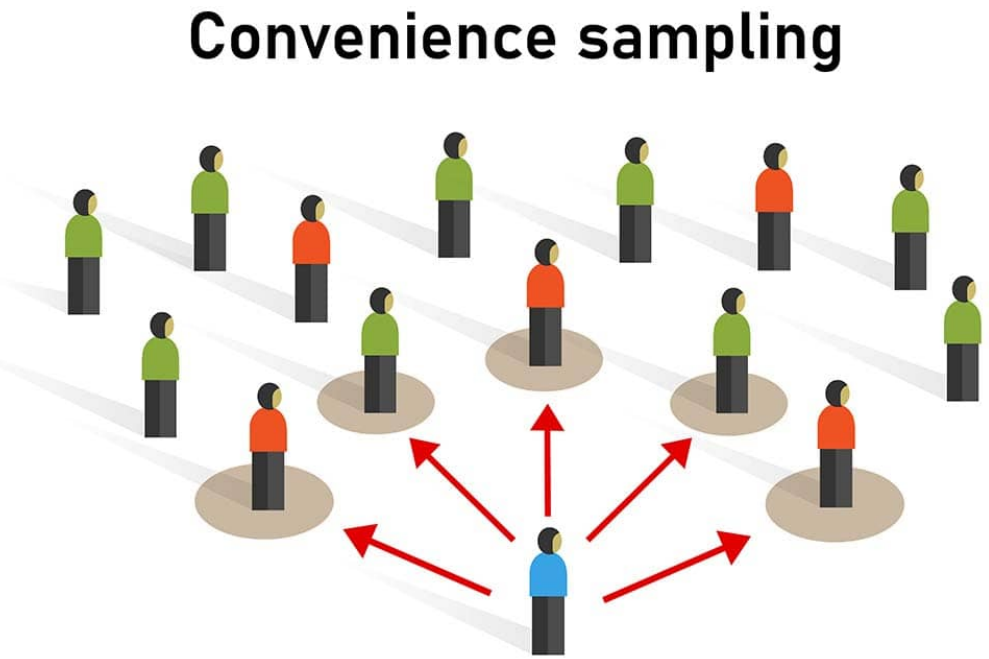
Multi-Stage Sampling uses combinations of the others to make it easier, done in stages

- Common for complicated situations
- Tedious?
- Eg Grinnell randomly samples 3 classes from each of the Humanities, STEM, Social Sciences. All students required to respond.
- Stratified (Humanities, STEM, Social Sciences)
- Clustered (entire classes)

Survey Methods (the bad ones)

Convenience Sampling is where a sample is taken because it is convenient/easy

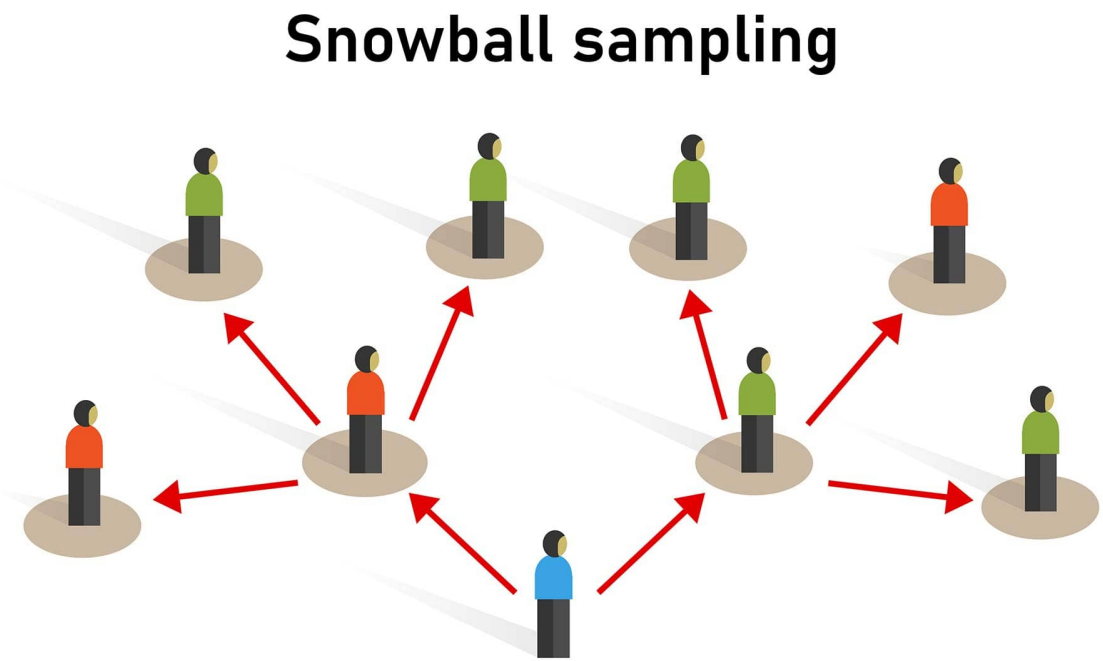
- Eg I texted my friends and we all agreed that X
- **Undercoverage** is where members of the population cannot be sampled or are sampled disproportionately low
- Eg Asking people how long they can plank *only* at the gym doesn't give you a sense for the student body as a whole (pun)



Survey Methods (the bad ones)

Snowball Sampling is where we use the first subjects to find the next few subjects

- Often useful (unfortunately) when populations are hard to reach/interact with
- Examples were early studies of homosexuals
- Problem is independence between respondents....



Other Biases

Several things can cause biases in your sample

- **Wording of the Questions**
  - wording is so charged people are "pushed" one way or the other
  - Also applies if it's so complicated/technical/run on as to be confusing
- **Response Bias**
  - Why did this weirdo take time out of their day to fill out my survey?
- **Non-Response Bias**
  - Why did this weirdo not take time out of their day to fill out my survey?

Experiments

Experiments are a scientific study where treatments are randomly *assigned* to experimental units

- Latin root: "test" or "try"
- An **Experimental Unit** is the object *a* treatment is applied to
  - Eg a farmer's field that recieves soybean hybrid A
  - Note singular: one experimental unit can have only one treatment
  - Often we can physcially touch it (safety check)
  - It's NOT the variable we are maniupulating (seed variety)
- ASSIGNED is the critical word with respect to experiments
  - Without assignment it's observational

Factors, Levels, and Treatments

A **factor** is a categorical input variable (eg fertilizer type). **Levels** of a factor are the categories the variable can take.

- Eg Varieties A, B, and C would be the three levels of a trinary variable Seed Type

A **treatment** (trt) is a particular combination of input variables

- Eg (seed variety A) + (fertilizer type 2)
- If there is only one input the treatment and the variable are the same

	Fertilizer 1	Fertilizer 2
Corn (mono crop)	Corn + Fert 1	Corn + Fert 2
Corn-Soybean rotation	Corn/Soy + Fert 1	
Corn-Soy-Alfalfa rotation		
Corn-Soy-Winter Oats rotation		

Experimental Units (EUs)

Experimental units are what we randomly (ideally) *assign* treatments to

- Without *assignment* it's an observational study
- Without *random assignemnt* it's a biased study

Observational Units is the thing we observe/measure and write down in our spreadsheets

- Common for EU's and OU's to be the same but not always
  - Eg applied seed variety A to the *entrie* field at once so the entire field is the experimetnal unit
  - Any given soybean plant in the field is an obserational unit
- The confusion is often the impetus of *pseudo – replication*

Exp. Unit Helpful Advice

THE EASIEST WAY TO IDENTIFY THE EXPERIMENTAL UNIT IS TO THINK ABOUT WHAT THE TREATMENT IS BEING ASSIGNED TO!!

Experiments: Example

There is a field split in 64 sections outside of Ames, IA owned and operated by Iowa State University. In one study, crops of different types (corn, soybean, white oats, alfalfa) were planted and, in early spring of the following year, a sample was taken of biomass from each section of the field.

- Treatment?
- Experimental Unit?
- Observational Unit?

Experiments: Example

There is a field split in 64 sections outside of Ames, IA owned and operated by Iowa State University. In one study, crops of different types (corn, soybean, white oats, alfalfa) were planted and, in early spring of the following year, a sample was taken of biomass from each section of the field.

- Treatment? The seed varieties
- Experimental Unit? The $1/64^{th}$ section of the field
- Observational Unit? The same as the experimental unit

Experiments: Example

ISU also does research in wine making (just a more interesting crop study....). 12 rows of grape vines were in the experiment. Each row was assigned one of four fertilizers and was either aggressively trimmed or was only mildly trimmed.

At the end of the season ten bottles of wine were made from each row of grapes. Then, each bottle underwent a chemical analysis and the results (nutrition content, polyphenals, etc...) were recorded.

- Variables we are controlling? What are their levels?
- Treatment?
- Experimental Unit?
- Observational Unit?

Experiments: Example

ISU also does research in wine making (just a more interesting crop study....). 12 rows of grape vines were in the experiment. Each row was assigned one of four fertilizers and was either aggressively trimmed or was only mildly trimmed.

At the end of the season ten bottles of wine were made from each row of grapes. Then, the wine underwent a chemical analysis and the results (nutrition content, polyphenals, etc...) were recorded.

- Variables we are controlling? What are their levels?
  - Trimming: Aggressively trimmed or mildly trimmed
  - Fertilizer: 1 of 4 types
- Treatment? A combination of trimming and fertilizer (eg Aggressively trimmed with fertilizer number 2)
- Experimental Unit? An entire row of grape vines
- Observational Unit? A single bottle of wine

Observation vs Experimental Units: Who Cares

The reason observational and experimental units need to be kept seperated is because an experimental unit is considered the foundational piece of information.

The total number of experimental units presents the upper limit of your information

At the end of the day we had 12 grape vines where each vine got a treatment. We didn't have 120 grape vines (120 = 10 bottles x 12 vines)

The ten bottles of wine give us a better understanding of that experimental unit and variability in the system

Experiments: Causality?

Some people claim that experiments can show causality (because the only expected difference in the final results is the different treatments....)

Can experiments show causation, that is that event A caused event B?

Experiments: Causality?

Some people claim that experiments can show causality (because the only expected difference in the final results is the different treatments....)

Can experiments show causation, that is that A caused B?

No, no, no No No NO NO no NOO no a thousand times NO!

Experiments: No Causality

Claim: Experiments cannot show causation.

Experiments: No Causality

Claim: Experiments cannot show causation.

Arguments:

1. The analysis depends on our assumptions
   - True independence between exp units?
   - "identically distributed" There is no reason any exp unit will be different? No lurking variables leading to different means?
   - Sample Distribution $\approx$ Population Distribution?
   - Ability to generalize from our *very* restrictive experimental conditions

Experiments: No Causality

Claim: Experiments cannot show causation.

Arguments:

1. The analysis depends on our assumptions
2. Our models are a simplification
   - Central limit theorem gets us *close* to normality
   - Measurement error, hand waving away lurking variables
   - Difference between a "straight" line and a $17^t h$ order polynomial that is extremely flat?
   - Did we account for relativity?

Experiments: No Causality

Claim: Experiments cannot show causation.

Arguments:

1. The analysis depends on our assumptions
2. Our models are a simplification
3. Our estimates are wrong
   - No one truly believes there sample's estimates are population's parameter (ie the true value)
   - Sample Distribution $\approx$ Population Distribution isn't good enough, we need equality (w/ probability 1)
   - Confidence intervals wouldn't exist for experiments.....

Experiments: No Causality

Claim: Experiments cannot show causation.

Arguments:

1. The analysis depends on our assumptions
2. Our models are a simplification
3. Our estimates are wrong

In conclusion: Using estimates we expect to be wrong and assumptions we don't trust we build an overly simplified model.....and then claim to have shown causality?