Study Design

Ryan Miller

Introduction

- So far we've learned that the procedure we use to obtain a sample is critically important, but this is not the only factor worthy of considering in the design of a study
- Suppose researchers develop a new COVID-19 vaccine, how would you design a study to determine if it is effective or not?

Introduction

- So far we've learned that the procedure we use to obtain a sample is critically important, but this is not the only factor worthy of considering in the design of a study
- Suppose researchers develop a new COVID-19 vaccine, how would you design a study to determine if it is effective or not?
- The only meaningful designs will compare people who receive the vaccine with another group
  - Therefore, we'll either need to use two samples, or use method for splitting a single sample into two

Two Types of Studies

This leads us to distinguish between two types of studies:

- **Observational studies**: the explanatory and response variables are *observed* by the researchers (ie: collecting two separate samples)
- **Experimental studies**: the explanatory variable is *assigned* by the researchers (ie: actively splitting a single sample)

Observational Studies

- Experimental studies are preferable for a number of reasons (which we'll soon see), but they aren't possible for every research question
- Sometimes we're interested in explanatory variables that cannot be randomly assigned, or doing so would be unethical
  - For example, researchers can't force people to expose themselves to harmful substances, nor can they randomly assign variables related to one's genetics

Example

- In late 1970s, researchers analyzed data on murders that took place during a felony in the state of Florida. (study link)
- The researchers recorded the race of the offender, as well as whether the offender was sentenced to the death penalty or not
  - Does the offender's race appear to be associated with their sentence?

	death	not
black	38	142
white	46	152

Example

- Using row proportions, we see that 21% of black offenders were sentenced to death, compared to 23% of white offenders
  - The *p*-value for such a difference is large, so does this mean there wasn't any racial bias in Florida courts in the 1970s?

	death	not
black	38	142
white	46	152

Example

- The researchers in this study collected another variable, race of the victim
  - The table below **stratifies** the data by the third variable victim's race

```
##                          DeathPenalty death not
## VictimRace OffenderRace
## black      black                         1 101
##            white                         0   8
## white      black                        37  41
##            white                        46 144
```

- Does the stratified table change your assessment?

Example

After stratifying by victim's race, racial bias is clearly evident in proportion of death penalty verdicts by offenders race (row proportions):

```
##                          DeathPenalty death  not
## VictimRace OffenderRace
## black      black                     0.01 0.99
##            white                     0.00 1.00
## white      black                     0.47 0.53
##            white                     0.24 0.76
```

Confounding Variables

- In this study, race of the victim is a **confounding variable**, or a third variable that is related to both the explanatory and response variables
  - The existence of one or more confounding variables can obscure the marginal relationship between explanatory and response variables

Confounding Variables

- In this study, race of the victim is a **confounding variable**, or a third variable that is related to both the explanatory and response variables
  - The existence of one or more confounding variables can obscure the marginal relationship between explanatory and response variables
- Can you explain why blacks are more likely to be setenced to death in each stratum of victim's race, but less likely overall?

```
##                               DeathPenalty death not
## VictimRace OffenderRace
## black      black                              1 101
##            white                              0   8
## white      black                             37  41
##            white                             46 144
```

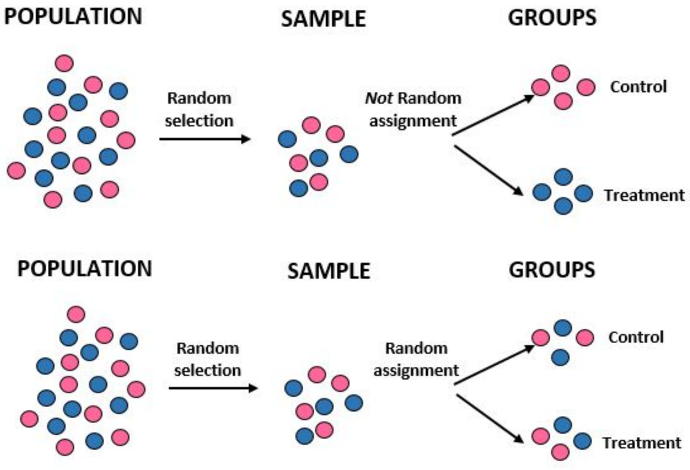
Balance

- We can view the problems caused by confounding variables as an issue of **imbalanced groups**
  - Offenders were more likely to victimize their own race, and crimes against whites tended to be punished more severely
  - The groups white offenders and black offenders were systematically different in an important way (victims race)

Balance

- We can view the problems caused by confounding variables as an issue of **imbalanced groups**
  - Offenders were more likely to victimize their own race, and crimes against whites tended to be punished more severely
  - The groups white offenders and black offenders were systematically different in an important way (victims race)
- Going back to the COVID-19 example, if study participants are allowed to choose whether they receive the vaccine we might expect the vaccine group to be older, sicker, working riskier jobs, etc.
  - However, these factors would all occur in equal proportions in the vaccinated and control groups if we **randomly assigned** which study participants received the vaccine

Balance


POPULATION
SAMPLE
GROUPS
Random selection
Not Random assignment
Control
Treatment
POPULATION
SAMPLE
GROUPS
Random selection
Random assignment
Control
Treatment

Discussion - Planning a Study

Suppose we want to know: "Is arthroscopic surgery is effective in treating arthritis of the knee?" Describe both an *observational study* and a *randomized experiment* that you could conduct to answer this question. Be sure to address the following during your discussion:

1. How costly will it be for the researchers to collect data with each design?
2. Are there any feasibility problems or ethical issues with each design?

Sham Knee Surgery

In the 1990s a study was conducted in 10 men with arthritic knees that were scheduled for surgery. They were all treated identically expect for one key distinction: only half of them actually got surgery! Once each subject was in the operating room and anesthetized, the surgeon looked at a randomly generated code indicating whether he should do the full surgery or just make three small incisions in the knee and stitch up the patient to leave a scar. All patients received the same post-operative care, rehabilitation, and were later evaluated by staff who didn't know whether they had actually received the surgery or not. The result? Both the sham knee surgery and the real knee surgery showed indistinguishable levels of improvement

Source: https://www.nytimes.com/2000/01/09/magazine/the-placebo-prescription.html

Control Groups, Placebos, and Blinding

The Sham Knee Surgery example illustrates several important aspects of a well-designed experiment that we've yet to discuss:

- **Control Group** - Some patients were randomly assigned not to receive the knee surgery, providing a comparison group that is, on average, balanced with surgery group in all baseline characteristics
- **Placebo** - Patients in the control group received a fake surgery
- **Blinding** - Using a placebo is not helpful if patients know which group they're in. Similarly, the staff interacting with the patients might treat them differently if they knew the patient's group
  - **Single-blind** - the participants don't know the treatment assignments
  - **Double-blind** - the participants *and* everyone interacting with the participants don't know the treatment assignments

Control Groups, Placebos, and Blinding

- The goal of each of these design elements is to prevent biasing the measurement of the outcome variable in a particular direction

Control Groups, Placebos, and Blinding

- The goal of each of these design elements is to prevent biasing the measurement of the outcome variable in a particular direction
- Other types of studies are susceptible to other types of biases, which we should also carefully consider

1. Social Desirability Bias - Respondents tend to answer questions in ways that portray themselves in a positive light Link
2. Habituation Bias - Respondents tend to provide similar answers for similarly worded or structured questions (the brain going on autopilot) Link
3. Leading Questions - The wording of a question impacts how people respond, great examples in the textbook
4. Cultural Bias - Questions are often to be constructed with one's own culture in mind, they might not even make sense to people from other cultures.

Study Design and Statistical Inference

- The overarching goal of a statistician is to *rule out* as many possible explanations for an observed association as possible

Study Design and Statistical Inference

- The overarching goal of a statistician is to *rule out* as many possible explanations for an observed association as possible
- So far we've considered the following design-related explanations, as well as methods for addressing for them
  - Sampling bias - Simple random sampling
  - Confounding variables - Random assignment of the explanatory variable, or stratification
  - Other biases - Using placebo, double-blinding, proper instruments, etc.
- Ideally we reduce the possible explanations to just two - random chance or a real relationship

Discussion - Can Randomization Fail?

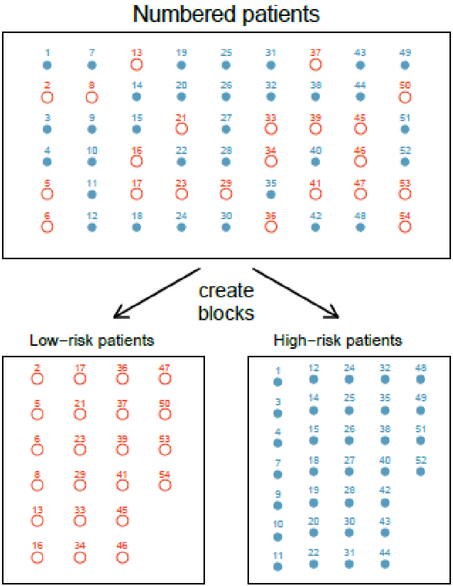
- A University of Iowa researcher was conducting an experiment on lab monkeys
- Lab monkeys are expensive, so his experiment had $n = 8$
- Having taken a statistics course, he randomly assigned treatment/control groups
- After conducting the experiment and seeing surprising results, the researcher recognizes that the 4 monkeys in the control group were also the oldest 4 monkeys
- The researcher knew that the age of the monkey had an important on the outcome variable, but he expected randomization to handle that

Should he report his results? What could he have done differently?

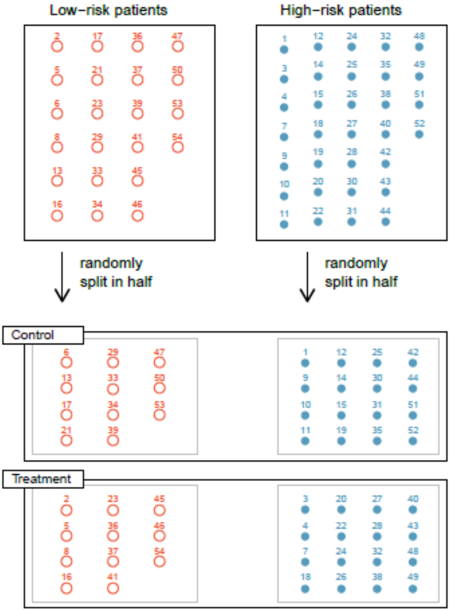
Can Randomization Fail?

- Randomization is not guaranteed to properly balance the treatment and control groups unless the sample size is relatively large
- At smaller sample sizes, strategies such as **blocking** can be used
  - In this design, cases are first split using a **blocking variable**, then random assignment is done within each block
  - This ensures the blocking variable is balanced in each group

Blocking


Numbered patients
create
blocks
Low–risk patients
High–risk patients

Blocking


Low−risk patients
High−risk patients
randomly
split in half
randomly
split in half
Control
Treatment

Minneapolis Police Case Study

- Police departments have long been uncertain about how to best respond to cases of domestic abuse
- In the 1980s, the Minneapolis Police conducted a study comparing three different response strategies:
  - Arrest
  - Advice
  - Separate
- Officers were randomly assigned a strategy to use on each case, but they were given discretion to change the strategy if necessary, but precautions were taken to ensure the officers were as faithful as possible to each assigned strategy
  - The outcome variable was whether or not violence reoccurred.

Cross-over

The columns of the table indicate the strategy assigned, the rows indicate the strategy actually used:

	Arrest	Advice	Separate
Arrest	91	18	26
Advice	0	84	5
Separate	1	6	82

Overall there was 82% adherence to the randomly assigned strategy, but do you see any problems?

Minneapolis Police Article Link

Minneapolis Police Case Study

- A common pattern was to “upgrade” to the arrest strategy
- This meant that the advice and separate groups lost their highest risk cases to the arrest group

The Intention-to-Treat Principle

- The Food and Drug Administration (FDA) mandates an **intent-to-treat principle** (ITT) as the primary design and analysis strategy for clinical trials
- This means that all subjects who are randomized be included in the final analysis, even if they, cross-over, do not adhere to any protocol, or drop out of the study
- It also means that clinical trials estimate the effect of the *treatment assignment* rather than the treatment itself

Intent-to-treat Article Link

Intention-to-Treat Example

Below are the results of the MN police case study *as randomized*:

	Recurrence	No Recurrence
Arrest	10	82
Advice	24	84
Separate	26	87

Below are the results *as treated*:

	Recurrence	No Recurrence
Arrest	18	117
Advice	16	73
Separate	26	63

Practice: compare the difference in recurrence proportions of the Advice and Separate treatments using an *ITT analysis*. Compare this with the *as treated* analysis

Intention-to-Treat Example - Solution

Using ITT, recurrence in the advice group was $24/106 = 0.22$ and for the separate group was $26/113 = 0.23$

- We can conclude that telling an officer to use the "Advice" strategy leads to essentially the same outcomes as telling the officer to use the "Separate" strategy

Analyzing the data *as treated*, the relevant proportions are $16/89 = 0.18$ and $26/89 = 0.29$

- But we have trouble concluding anything; while there is a clear difference in recurrence, we don't know if it is due to the strategies themselves, or a disproportionate movement of high/low risk cases into different strategies

Conclusions

- Randomized experiments are the *gold-standard* in yielding trustworthy results
  - Randomization can fail due to "bad luck"
  - Cross-over can be an issue in some studies
  - In some circumstances a randomized experiment is not possible

Conclusions

- Randomized experiments are the *gold-standard* in yielding trustworthy results
  - Randomization can fail due to "bad luck"
  - Cross-over can be an issue in some studies
  - In some circumstances a randomized experiment is not possible
- Observational studies have flaws, but if we carefully identify confounding variables and apply methods like stratification we can still identify meaningful relationships
  - Many of the most important public health discoveries are the results of observational studies (smoking and lung cancer as one example)