Study Design

Ryan Miller

Outline

1. Observational designs
   - Prospective vs. retrospective studies, design challenges
2. Experimental designs
   - Randomization, methods to reduce bias, causation

Introduction

- So far, we've seen how *sampling* can influence the trends seen in our data, but there are other aspects of data collection that we need to consider
- Suppose researchers want to test a new COVID-19 treatment, how would you design a study to determine if it is effective or not? Could you offer the treatment to anyone who wants it?

Introduction

- So far, we've seen how *sampling* can influence the trends seen in our data, but there are other aspects of data collection that we need to consider
- Suppose researchers want to test a new COVID-19 treatment, how would you design a study to determine if it is effective or not? Could you offer the treatment to anyone who wants it?
- Any meaningful design must compare the new treatment with something else (ie: a control group)
  - Therefore, we'll either need to *collect two samples* (people who chose this treatment and others who didn't), or proactively *split a single sample into two* (conducting an experiment)

Two different study designs

These two possibilities lead us to distinguish between two types of studies:

- **Observational studies**: the explanatory and response variables are *observed* by the researchers (separate samples)
- **Experimental studies**: the explanatory variable is *assigned* by the researchers (the researchers split up a single sample)

Observational studies

- Suppose you'd like to determine whether smoking increases the severity of the flu
  - Can you identify *two different types* of observational studies that could investigate this question?

Observational studies

- Suppose you'd like to determine whether smoking increases the severity of the flu
  - Can you identify *two different types* of observational studies that could investigate this question?
- A **retrospective study** might survey people about their smoking status and how badly (if at all) they suffered from the flu in the previous year
  - The defining feature of a retrospective study is that data is collected after both the explanatory and response variable have occurred

Observational studies

- Suppose you'd like to determine whether smoking increases the severity of the flu
  - Can you identify *two different types* of observational studies that could investigate this question?
- A **retrospective study** might survey people about their smoking status and how badly (if at all) they suffered from the flu in the previous year
  - The defining feature of a retrospective study is that data is collected after both the explanatory and response variable have occurred
- A **prospective study** (also called a cohort study) might follow the residents in an entire town for a year and track whether smokers or non-smokers are hospitalized more often
  - The defining feature of a prospective study is the researchers must wait for the outcome to be observed
  - Prospective studies generally provide stronger evidence than retrospective ones (fewer opportunities for bias)

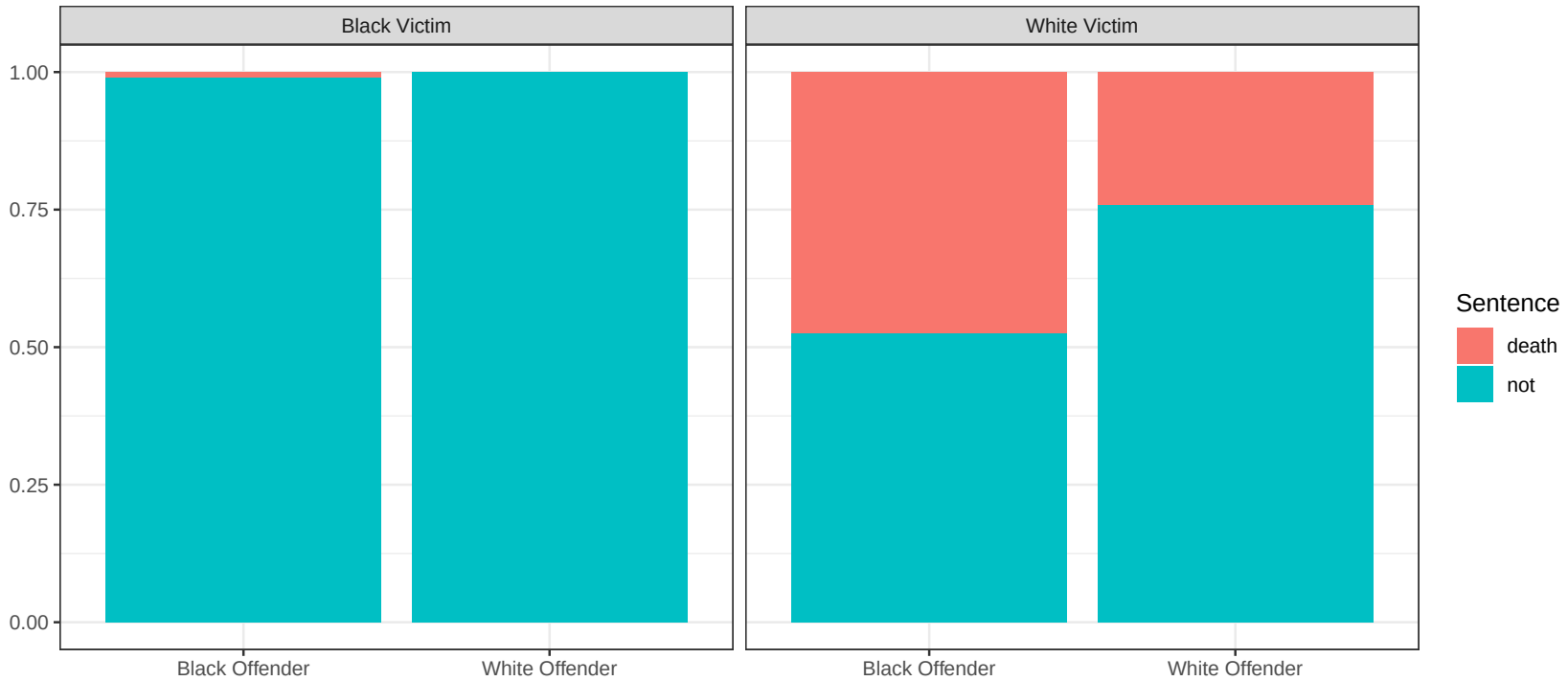
Problems with observational studies

- We've already seen an observational study in the Florida Death Penalty Sentencing case study
  - Without considering any other variables, did the offender's race appear associated with their sentence?

	death	not
black	38	142
white	46	152

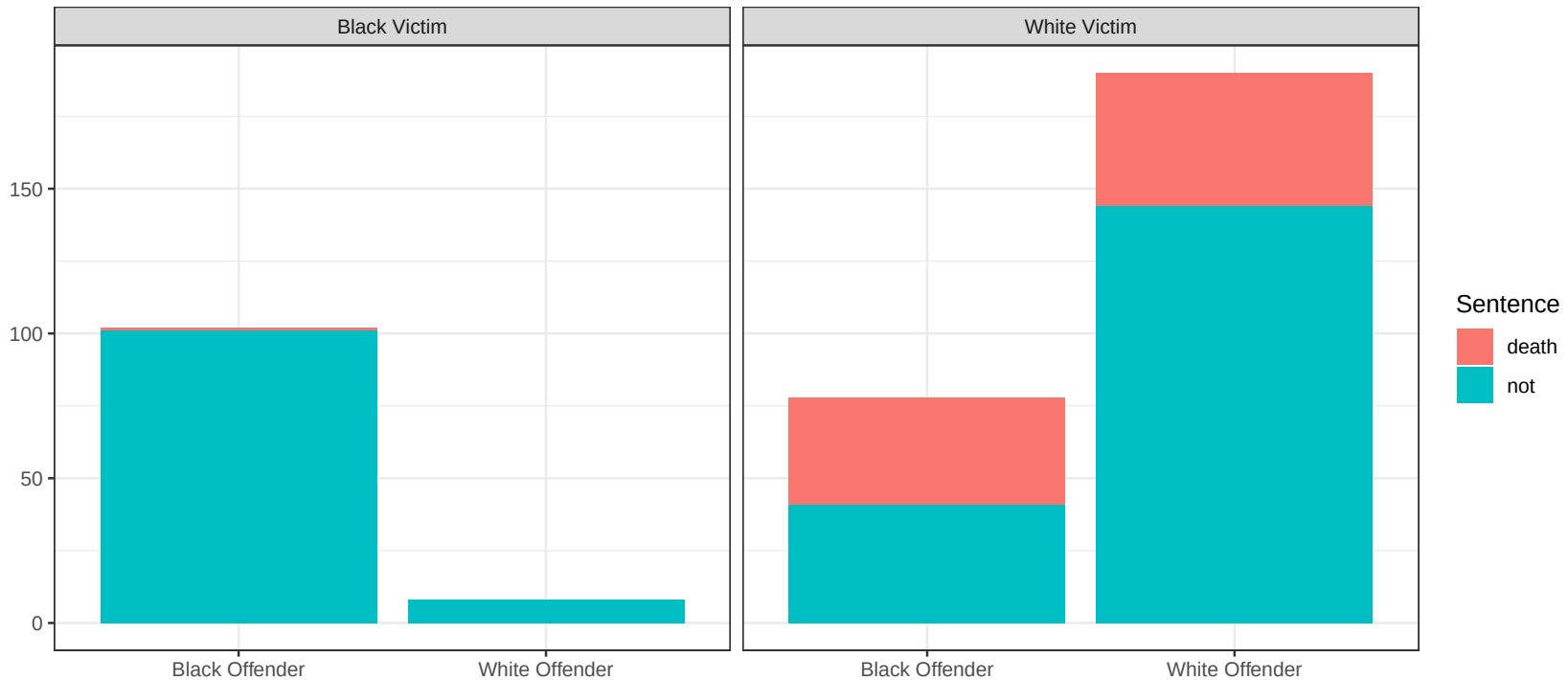
Confounding variables

Overall, white offenders received the death penalty slightly more often, but this ignored the influence of the victim's race:



Confounding variables

Because offenders *disproportionately* committed crimes against victims of their own race, the overall death penalty rates were skewed in a way that obscured the racially biased sentencing:



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)
- *Stratification* provided a method for forcing balance (in regards to identified confounding variable), which helps us make a stronger case for causation
  - Unfortunately, any observational study can contain many confounding variables. . .

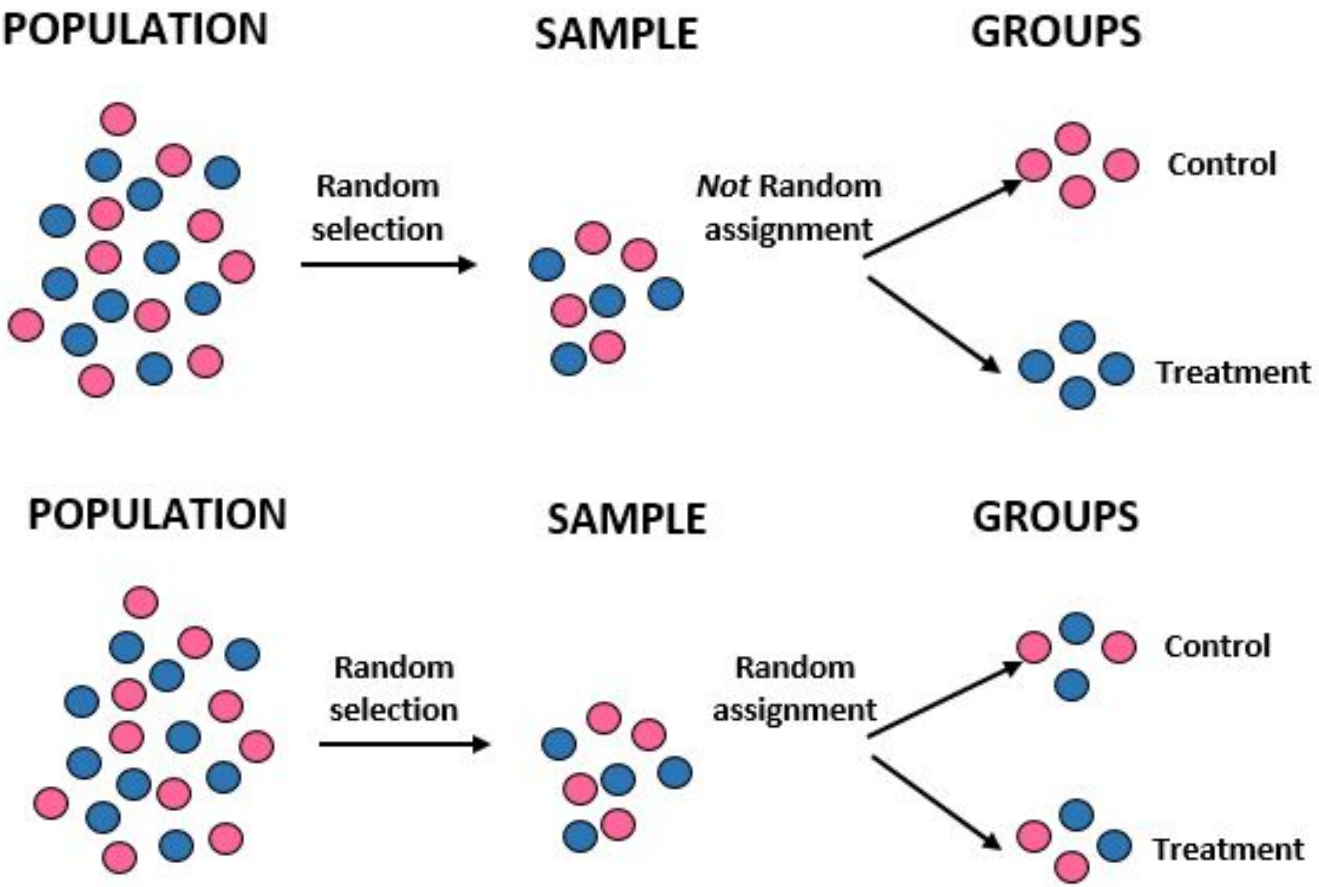
Random assignment

- In our COVID-19 treatment example, if study participants can choose their treatment it's likely that one group will be disproportionately older, sicker, working riskier jobs, etc.
  - However, these factors will be equally prevalent in both groups if we **randomly assigned** participants

Random assignment

- In our COVID-19 treatment example, if study participants can choose their treatment it's likely that one group will be disproportionately older, sicker, working riskier jobs, etc.
  - However, these factors will be equally prevalent in both groups if we **randomly assigned** participants
- Because random assignment is expected to balance *all* possible confounding variables, causation can be established in a well-designed randomized experiment

Random assignment



Practice

Suppose we want to know: "Is arthroscopic surgery is effective in treating arthritis of the knee?"

1) Describe both an *observational study* and a *randomized experiment* that could be conducted to address this research question.
2) Discuss how *ethical* each design would be.
3) Discuss the *strength of evidence* that would result from 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

Vocabulary for randomized experiments

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

Minneapolis Police Case Study

- Police departments have long been uncertain about how to best respond to cases of domestic abuse
- 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
  - The outcome of interest was whether or not violence reoccurred

Minneapolis Police Case Study

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
- The advice and separate groups likely lost their highest risk members to the arrest group
- This seemingly well-designed experiment ended up needing to be bailed out by complex statistical approaches used to jointly model recurrence of violence (the outcome variable) and adherence to the randomized strategy

The Intention-to-Treat Principle

- The Food and Drug Administration (FDA) mandates using the **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 *by treatment used*:

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

Summary

- Observational study - the explanatory and outcome variables are observed as they'd naturally occur
  - Retrospective design - outcomes have already occurred before the start of the study
  - Prospective design - cases followed forward until the outcome occurs
- Experimental study - the explanatory variable is manipulated by the researchers
  - Randomized experiment - the explanatory variable is randomly assigned, resulting in balanced groups
  - Placebo, blinding, etc. - additional steps to address possible sources of bias
  - Intent-to-treat (ITT) - a type of analysis that keeps subjects in the group they were randomized into (even if non-adherence or cross-over occur)