STAT 311: More Observational Studies and Experiments

Y. Samuel Wang

Summer 2016

Logistics

- Practice problems posted

Overview

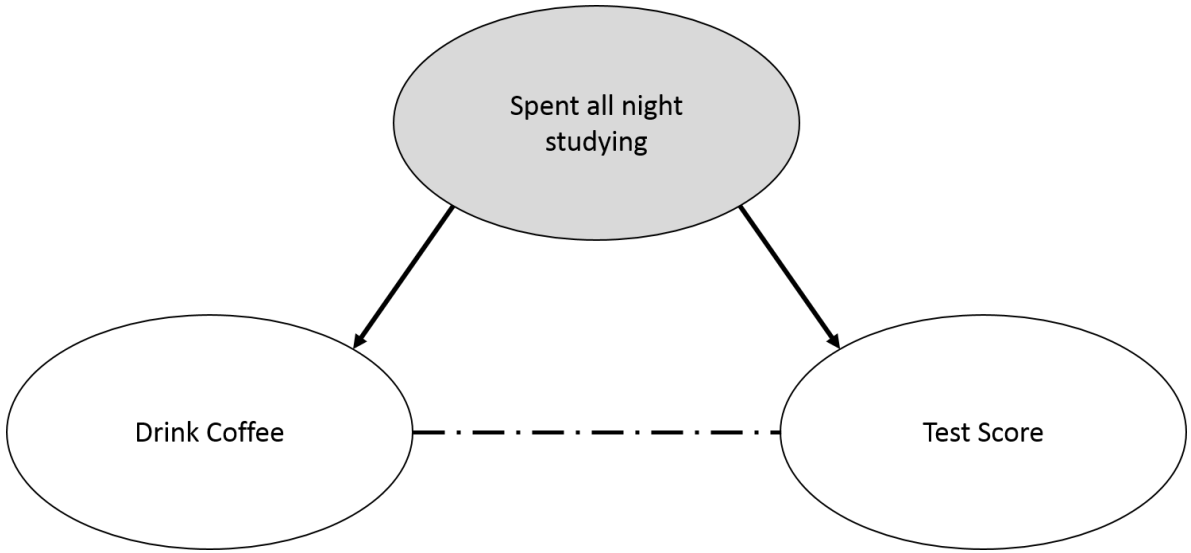
- Monday we focused primarily on surveys, a specific type of observational study where we typically are only interested in describing a population , but not necessarily drawing cause and effect relationships
- Making statements about cause and effect relationships requires a bit more care
- Today we will discuss gathering data when we have an explanatory and response variable

Causal relationships

- Discovering causal relationships is difficult because of confounding variables
- A **Confounding variable** is a variable that affects both the response and is related to the explanatory variable

Causal relationships

- Discovering causal relationships is difficult because of confounding variables
- A **Confounding variable** is a variable that affects both the response and is related to the explanatory variable



Randomized Experiment Terminology

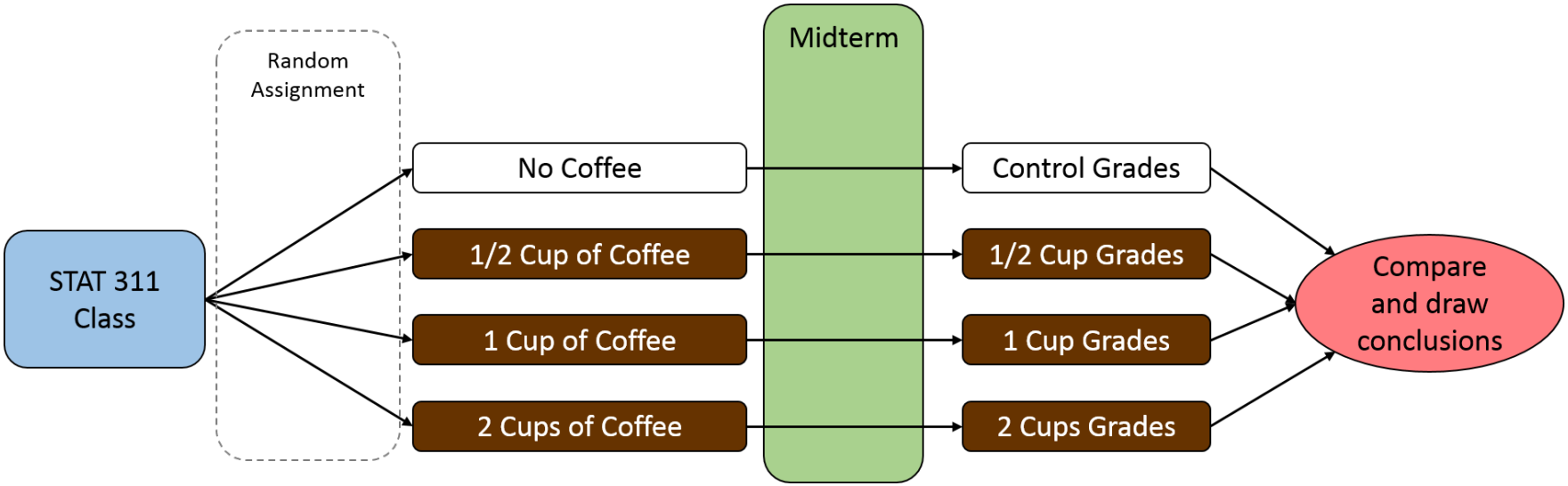
- A **treatment** is a specific condition we apply to an individual. An experiment can have multiple treatments
  - Ex- 1/2 cup of coffee, 1 cup of coffee, 2 cups of coffee
- The total sample can be divided into a **control group** which does not receive a treatment, and **treatment groups** which receive a treatment

How else can we improve experiments

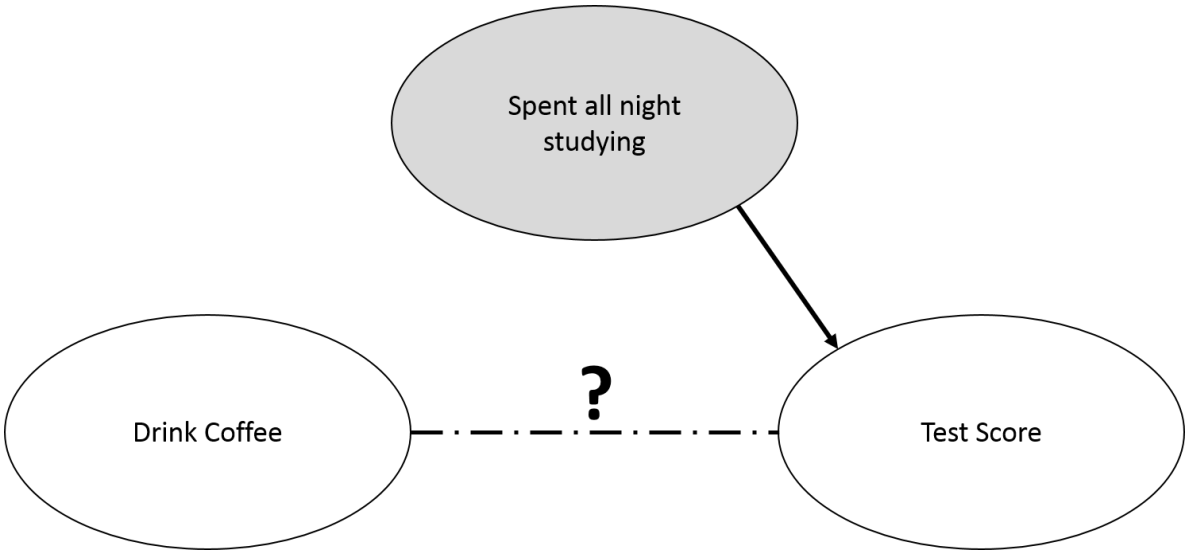
Randomization helps remove bias, but can we anything to reduce variability

- If we have observed other characteristics on our sample, we can first divide them into subsets, then randomly assign individuals within those subsets to the control/treatment groups. This is called **blocked design**.
- This ensures that the control/treatment groups will look similar. Note how this is a similar idea to stratified sampling
- When there is only 1 treatment group and 1 control group, this is sometimes called a **matched pair** experiment
- A specific type of matched pair design is when the same individual receives multiple treatment (ie they are matched with themselves)

Randomized Experiment



Randomized Experiment



Cautions

Despite the randomization, there are still several things that might cause inaccurate results

- The **placebo effect** refers to a change in the response variable due to the perception of the subject, despite an ineffectual treatment
  - Giving a **Placebo**, or ineffectual treatment, to the control group can help cancel out the placebo effect in the end data analysis
  - Also, the subjects should be **blinded**, or unaware whether they are receiving a placebo or a real treatment
  - For extra caution, some experiments are **double blind**, where even the researcher taking the measurements does not know who is getting a placebo or treatment
- The **Hawthorne Effect** refers to the fact that just being in an experiment (or study) can cause a change in the response variable

Types of observational studies

In an observational study, since we do not apply a treatment, we can't randomize. However, the data can still be useful.

There are many types of observational studies

- Longitudinal vs Cross-sectional
  - **Longitudinal**: Follow the same group of people over time
  - Individuals can serve as their own controls, but this takes time to gather and people might drop out
  - **Cross sectional**: Take data from a single point in time
  - Easier and faster to accomplish, but it can be hard to control for confounding

Types of observational studies

In an observational study, since we do not apply a treatment, we can't randomize. However, the data can still be useful.

There are many types of observational studies

- Retrospective vs Prospective
  - **Retrospective**: Gather and analyze historical data
  - Don't have to wait for data, but data which address the specific question you are interested in can be hard
  - **Prospective**: Recruit a sample, then follow them into the future
  - Takes time for the study to finish

Types of observational studies

In an observational study, since we do not apply a treatment, we can't randomize. However, the data can still be useful.

There are many types of observational studies

- **Case-Control**: Although we cannot randomize, we can match individuals who have the "treatment" with someone similar who does not have the "treatment." Similar idea to matched pair, but we don't get to choose who is the treatment and who is not.

What does our analysis actually tell us?

So we've completed a data analysis, how can we interpret the results?

- **Internal Validity**: Did we actually measure what we set out to measure?
  - Did I rule out confounding variables?
  - Did I measure the response variable accurately?
- **External Validity**: Can I generalize these results to individuals outside this study or experiment?
  - Does my sample represent the population I am interested in?
  - Do the experimental conditions reflect reality?