SECTION 4.2 PART 1

THURSDAY OCTOBER 19TH, 2017
TODAY’S OBJECTIVES

• Distinguish between an observational study and an experiment

• Explain the concept of confounding

• Identify experimental units, subjects, explanatory and response variables, and treatments in an experiment

• Explain the purpose of comparison, random assignment, control, and replication in an experiment
My ex-girlfriend thinks I'm mean and my current one thinks I'm average...

Maybe you’ve set your bar too high?
DISCUSS: WHAT ARE SOME DIFFERENCES BETWEEN AN OBSERVATIONAL STUDY AND AN EXPERIMENT?

Observational Studies:

No intervention, individuals are simply observed and variables of interest are measured.

*The goal is to describe, compare, or examine.*

Experiments:

Intervening within the population by imposing a treatment to measure individual’s response to treatment.

*The goal is to construct a credible argument for a causal purpose.*
WHAT IS A CONFOUNDING VARIABLE?

A variable that results in confounding

Confounding occurs when two variables are associated in such a way that you can’t separate their effects on the response variable from each other.
OBSERVATION VS. EXPERIMENT

IMPORTANT: An observational study may reveal correlation between two variables, but only a randomized experiment can prove cause-and-effect.

Why???

- Confounding may be present in observational studies
- Random assignment to treatment and control groups in an experiment helps equalize the groups with respect to any confounding variables so any difference in the response is attributable to the explanatory variable
**DO PARACHUTES PREVENT FATAL FALLS DUE TO GRAVITATIONAL CHALLENGE?**

Observational data have shown that their use is associated with morbidity and mortality.

There are no experiments that assign some people to use parachutes and some people to not.

Anyone who insists on randomized trials needs to come back down to earth...

The point: we can’t conduct certain studies for ethical reasons

(read Section 4.3)
CHECK YOUR UNDERSTANDING

1. Does reducing screen brightness increase battery life in laptop computers? To find out, researchers obtained 30 new laptops of the same brand. They chose 15 of the computers at random and adjusted their screens to the brightest setting. The other 15 laptop screens were left at the default setting—moderate brightness. Researchers then measured how long each machine’s battery lasted. Was this an observational study or an experiment? Justify your answer.

Questions 2 to 4 refer to the following setting. Does eating dinner with their families improve students’ academic performance? According to an ABC News article, “Teenagers who eat with their families at least five times a week are more likely to get better grades in school.”

This finding was based on a sample survey conducted by researchers at Columbia University.

2. Was this an observational study or an experiment? Justify your answer.

3. What are the explanatory and response variables?

4. Explain clearly why such a study cannot establish a cause-and-effect relationship. Suggest a variable that may be confounded with whether families eat dinner together.
THE LANGUAGE OF EXPERIMENTS
EXPERIMENT

A statistical study where we manipulate the environment or impose treatment on subjects in order to measure their response.
TREATMENT

The specific condition applied to individuals
EXPERIMENTAL UNITS

Smallest collection of individuals to which treatments are applied to the whole group

Subjects are humans
A study published in the New England Journal of Medicine (March 11, 2010) compared two medicines to treat head lice: an oral medication called ivermectin and a topical lotion containing malathion. Researchers studied 812 people in 376 households in seven areas around the world. Of the 185 households randomly assigned to ivermectin, 171 were free from head lice after 2 weeks compared with only 151 of the 191 households randomly assigned to malathion.

Problem: Identify the experimental units, explanatory and response variables, and the treatments in this experiment.
A LOUSEY SITUATION

Experimental Units: 376 households
Subjects: 812 people
Explanatory variable: Type of Medication
Response Variable: Whether the household was lice-free
Treatments: Ivermectin and Malathion
Factors are just explanatory variables while the levels are subgroups of each factor.

Example: different concentrations of Vitamin D administered to an experimental unit of subjects.
GROWING TOMATOES

Does adding fertilizer affect the productivity of tomato plants? How about the amount of water given to the plants? To answer these questions, a gardener plants 24 similar tomato plants in identical pots in his greenhouse. He will add fertilizer to the soil in half of the pots. Also, he will water 8 of the plants with 0.5 gallon of water per day, 8 of the plants with 1 gallon of water per day, and the remaining 8 plants with 1.5 gallons of water per day. At the end of 3 months, he will record the total weight of tomatoes produced on each plant.

Problem: Identify the experimental units or subjects, explanatory and response variables, and the treatments.
GROWING TOMATOES

Experimental Units:
24 tomato plants

2 Explanatory variable(factors):
Whether or not fertilizer was used and amount of water

Response Variable:
Total Weight of plants

Diagram:
6 treatments:
① Fertilizer with 0.5 gallons water
② Fertilizer with 1.0 gallon water
③ Fertilizer with 1.5 gallons water
④ No fertilizer with 0.5 gallons water
⑤ No fertilizer with 1.0 gallons water
⑥ No fertilizer with 1.5 gallons water
The Physicians’ Health Study

A well-designed experiment

Does regularly taking aspirin help protect people against heart attacks? The Physicians’ Health Study was a medical experiment that helped answer this question. In fact, the Physicians’ Health Study looked at the effects of two drugs: aspirin and beta-carotene. Researchers wondered whether beta-carotene would help prevent some forms of cancer. The subjects in this experiment were 21,996 male physicians. There were two explanatory variables (factors), each having two levels: aspirin (yes or no) and beta-carotene (yes or no). Combinations of the levels of these factors form the four treatments shown in Figure 4.3 on the next page. One-fourth of the subjects were assigned at random to each of these treatments.

Watch The Physician's Study Video
4.2 PART 2
EXPERIMENTAL DESIGN
TODAY’S OBJECTIVES

• Describe a completely randomized design for an experiment, including how to use the Hat Method, Technology, and Table D

• Describe the placebo effect and the purpose of blinding in an experiment
PRINCIPLES OF EXPERIMENTAL DESIGN

The basic principles for designing experiments are as follows:

1. **Comparison.** Use a design that compares two or more treatments.

2. **Random assignment.** Use chance to assign experimental units to treatments. Doing so helps create roughly equivalent groups of experimental units by balancing the effects of other variables among the treatment groups.

3. **Control.** Keep other variables that might affect the response the same for all groups.

4. **Replication.** Use enough experimental units in each group so that any differences in the effects of the treatments can be distinguished from chance differences between the groups.
Researchers in Canada performed an experiment with university students to examine the effects of in-class laptop use on student learning. All participants in the study were asked to attend a university-style lecture and take notes with their laptops. Half of the participants were assigned to complete other non-lecture-related online tasks during the lecture. These tasks were meant to imitate typical student Web browsing during classes. The remaining students simply took notes with their laptops. To assign the treatments, the researchers printed 40 papers with instructions (20 with multitasking and 20 without), shuffled them, and handed them out at random to students in the classroom. At the end of the lecture, all participants took a comprehension test to measure how much they learned from it. The results: students who were assigned to multitask did significantly worse (11%) than students who were not assigned to multitask. From [http://www.sciencedirect.com/science/article/pii/S0360131512002254](http://www.sciencedirect.com/science/article/pii/S0360131512002254).
Comparison: The researchers compared students who were multitasking to other students who were not multitasking.

Random assignment: It was determined at random which students (and which seat locations) received the instructions to multitask and which students (and seat locations) received the regular instructions.

Control: The experiment used undergraduate students from the same university in Canada. All participants listened to the same lecture and took the same comprehension test at the end of the lecture.

Replication: There were 20 students in each treatment group. Although this is not a very large number of subjects, the difference in performance was big enough to determine that the difference was due to multitasking and not the chance variation in random assignment.
LET’S DESIGN AN EXPERIMENT TO SEE IF CAFFEINE AFFECTS PULSE RATE.

Here is an initial plan:

Measure initial pulse rate.

Give each student some soda with caffeine.

Wait for a specified time (10 minutes).

Measure final pulse rate.

Compare initial and final pulse rates.

What are some problems with this plan? What other variables are most likely to be sources of variability in pulse rates?
Unfortunately, even if the pulse rate of every student went up, we couldn’t attribute the increase to caffeine.

Perhaps the excitement of being in an experiment made their pulse rates increase. Perhaps it was the sugar in the cola and not the caffeine. Perhaps their teacher told them a funny joke during the 10-minute waiting period and made everyone laugh!

In other words, there are many other variables that are potentially confounded with taking caffeine.
THE FIRST STEP IS TO INCLUDE A CONTROL GROUP THAT DOES NOT RECEIVE CAFFEINE.
PURPOSE OF A CONTROL GROUP:

Provides a baseline for comparison. Otherwise, any pulse-raising (or lowering) event that occurs during the experiment would be confounded with the caffeine.

For example, an amazing stats lecture during the waiting period would certainly raise pulse rates, making it hard to know how much of the pulse increase was due to the caffeine.

What should the no caffeine group get? Nothing? Sprite? Caffeine-free Coke?
PLACEBO EFFECT

1. A positive effect or cure, produced by a placebo “drug” or treatment. The effect is not caused by the placebo. The effect is solely due to the patient's belief in the treatment.
NOTE:

Not all experiments have a control group or use a placebo as long as there is comparison. If you are testing a new drug, it is usually compared to the currently used drug, not a placebo.

Also, you can do an experiment to compare four brands of paint without using a placebo.
WHAT IS A SINGLE-BLIND EXPERIMENT? A DOUBLE-BLIND EXPERIMENT? WHY IS BLINDNESS IMPORTANT?
SUPPOSE YOU HAVE A CLASS OF 30 STUDENTS WHO VOLUNTEER TO BE SUBJECTS IN THE CAFFEINE EXPERIMENT DESCRIBED EARLIER. EXPLAIN HOW YOU WOULD RANDOMLY ASSIGN 15 STUDENTS TO EACH OF THE TWO TREATMENTS:
Using 30 identical slips of paper, write A on 15 pieces of paper and B on the other 15.

Mix them thoroughly in a hat and have each student select one slip of paper.

Then ask each student who received an A to drink the cola with caffeine and each student who received a B to drink the cola without caffeine.
Number the students from 1 to 30.

Use randInt(1,30) to select 15 different numbers from 1 to 30. Ignore repeated integers.

These students will drink the cola with caffeine and the remaining 15 will drink the cola without caffeine.
Number the students from 01 to 30. Use a line from Table D and read two-digit numbers moving from left to right.

The first 15 different numbers from 01 to 30 will identify the students who will drink the cola with caffeine.

The remaining 15 students will drink the cola without caffeine.
BACK TO DESIGNING OUR EXPERIMENT..

Experimental Units?

Subjects?

Factor(s)?

Level?

Treatment?
DRAW AN OUTLINE OF THE DESIGN
2. The second step is to make sure that the two groups (caffeine and non-caffeine) are as similar as possible and are treated in exactly the same way, with the exception of the treatments. To make this happen, we use random assignment, replication, and control.
RANDOM ASSIGNMENT

To create groups that are roughly equivalent at the beginning of the experiment.

Random assignment ensures that the effects of uncontrolled variables are balanced among the treatments groups.
WHY DO WE RANDOMIZE?

We must ALWAYS randomize since there will always be other variables we cannot control or that we do not consider. Randomizing guards against what we don’t know and prevents people from asking “But what about this variable?”
WHAT IS REPLICATION?

Ensuring that there are an adequate number of units in each treatment group so that the two groups are as equivalent as possible.

Note: Replication can also refer to repeating the experiment with different subjects. This can help us feel more confident applying the results of our experiment to a larger population.
WHAT DOES IT MEAN TO CONTROL OTHER VARIABLES?

Holding other variables constant for each member of both treatment groups.
If one treatment group was given regular Coke (which has sugar) and the other treatment group was given caffeine free Diet Coke (which has no sugar), then sugar and caffeine would be confounded. To prevent sugar from becoming confounded with caffeine, we need to make sure that members of both treatment groups get the same amount of sugar.
If we let subjects in both groups drink any amount of soda they want, the changes in pulse rates will be more variable than if we made sure each subject drank the same amount of soda. In the second set of dotplots below, the additional variability in pulse rate changes makes the evidence for caffeine less convincing.
DISCUSS WITH YOUR PARTNER

1. What is the purpose of an experiment?
2. What are the principles of good experimental design?
We’ll randomly assign 16 subjects to each of the two treatments: *soda with caffeine*, and *soda without caffeine*. Our response variable is the change in pulse rate.

To implement the design, use 16 equally sized slips of paper. Label 8 of the slips “A” and 8 of the slips “B”. Mix the papers in a hat and have each student draw a paper without looking.

The letter that each student chooses will be the group to which he or she will be assigned: _____ will represent treatment with caffeine and _____ will represent treatment with no caffeine. Students will not know which treatment they are receiving.

At the end of the experiment, compare the increase in pulse rates for both treatment groups.
BLOCKING & RANDOMIZED BLOCK DESIGNS

TUESDAY OCTOBER 24TH, 2017
TODAY’S LEARNING
OBJECTIVES

• Explain the purpose of blocking in an experiment.

• Describe a randomized block design or a matched pairs design.
Answer Questions:

What is a block? What is a randomized block design?

What is the benefit of blocking? How can we determine the variables best for blocking?

*Blocking accounts for a source of variability, just like stratifying. This means that blocking is a good way to increase your chances of finding convincing evidence.*

*Use the variables that are most strongly associated with (best predict) the response variable. The units within the block should respond similarly, but differently than the units in other blocks.*
EXAMPLE: SAT PREP!

Many students enroll in prep courses to improve their SAT scores. Twenty students who have taken the SAT once have volunteered to participate in an experiment comparing online and classroom prep courses. Among the 20 volunteers, 10 are in Precalculus, 6 are in Algebra 2, and 4 are in Geometry.

1. Explain why a randomized block design is ideal.
2. Outline a randomized block design for this experiment.
3. Describe how you would carry out the random assignment required by your design.
1. EXPLAIN WHY A RANDOMIZED BLOCK DESIGN IS IDEAL.

• A randomized block design would consider each group of students with different current math classes separately. Random assignment will occur separately within each block.

• This randomized block design will allow us to account for the variation in improvement due to one’s current math course, which will help us to more accurately assess the overall improvements made in the two types of SAT prep courses.
2. **OUTLINE A RANDOMIZED BLOCK DESIGN FOR THIS EXPERIMENT.**

- 20 volunteers
  - Precalculus Student \( n=10 \)
  - Algebra 2 Student \( n=6 \)
  - Geometry Students \( n=4 \)

```
Random Assignment ➔ Online Course ➔ Compare Improvements
Random Assignment ➔ Classroom Course
Random Assignment ➔ Online Course ➔ Compare Improvements
Random Assignment ➔ Classroom Course
Random Assignment ➔ Online Course ➔ Compare Improvements
Random Assignment ➔ Classroom Course
```

---

Combine Results
3. **DESCRIBE HOW YOU WOULD CARRY OUT THE RANDOM ASSIGNMENT REQUIRED BY YOUR DESIGN.**

We will separate the 20 students into three groups (blocks) based on their current math class (10 Precalculus students, 6 Algebra 2 students, and 4 Geometry students).

We will randomly assign the members of each block into the Online course or the Classroom course. Write the names of the 10 Precalculus students on 10 identical slips of paper, place the slips in a hat, and mix them well. Pull out 5 slips to determine which students will take the Online course. The remaining 5 Precalculus students will take the classroom course.

Randomly assign the 6 Algebra 2 students into two groups of 3 and the 4 Geometry students into two groups of 2 using a similar process.

After each subject has taken the randomly assigned course, compare improvements to the two SAT prep courses within the current math class blocks. To compare the overall effectiveness of the two SAT prep courses, combine the results from the three blocks after accounting for the difference in improvement for Precalculus, Algebra 2, and Geometry classes.
NOW YOU TRY! **Microwave Popcorn**

A popcorn lover wants to know if it is better to use the “popcorn button” on her microwave oven or use the amount of time recommended on the bag of popcorn. To measure how well each method works, she will count the number of unpopped kernels remaining after popping. To obtain the experimental units, she goes to the store and buys 10 bags each of 4 different varieties of microwave popcorn (butter, cheese, natural, and kettle corn), for a total of 40 bags.

**Problem:**

(a) Explain why a randomized block design might be preferable to a completely randomized design for this experiment.

(b) Outline a randomized block design for this experiment. Describe how you would carry out the random assignment required by your design.
a) EXPLAIN WHY A RANDOMIZED BLOCK DESIGN MIGHT BE PREFERABLE TO A COMPLETELY RANDOMIZED DESIGN FOR THIS EXPERIMENT.

A completely randomized design ignores the difference between the four types of popcorn, which will probably result in a great deal of variability in the number of unpopped kernels for both treatments. For example, if there are many more unpopped kernels in bags of cheese popcorn and kettle corn than the other two varieties, it will be harder to tell if there is a difference in the methods of popping.

A randomized block design considers each variety of popcorn separately, which allows us to account for the variability in the number of unpopped kernels created by the difference in varieties.
b) OUTLINE A RANDOMIZED BLOCK DESIGN FOR THIS EXPERIMENT. DESCRIBE HOW YOU WOULD CARRY OUT THE RANDOM ASSIGNMENT REQUIRED BY YOUR DESIGN.

We will randomly assign 5 bags of each variety to each of the two treatments. To perform the random assignment, place all 10 bags of a particular variety in a large grocery bag. Shake the bag and then randomly select 5 bags to be popped using the “popcorn button.” The remaining 5 bags will be popped using the instructions on the bags. Repeat this process for the remaining 3 varieties. After popping each of the 40 bags in random order, count the number of unpopped kernels in each bag and compare the results within each variety. Then combine the results from the 4 varieties after accounting for the difference in average response for each variety.