4.2 Observational Studies and Experiments

Learning Objectives

1. Distinguish between an observational study and an experiment.
2. Explain the concept of confounding and how it limits the ability to make cause-and-effect conclusions.
3. Identify the experimental units, explanatory and response variables, and treatments in an experiment.
4. Explain the purpose of comparison, random assignment, control, and replication in an experiment.
5. Describe a completely randomized design for an experiment, including how to randomly assign treatments using slips of paper, technology, or a table of random digits.
6. Describe the placebo effect and the purpose of blinding in an experiment.
7. Interpret the meaning of statistically significant in the context of an experiment.
8. Explain the purpose of blocking in an experiment. Describe a randomized block design or a matched pairs design for an experiment.

Vocabulary: observational study, experiment, confounding, treatment, experimental units, subjects, factors, level, comparison, random assignment, control, replication, completely randomized design, control group, placebo effect, double-blind, single-blind, statistically significant, block, randomized block design, matched pairs design

ADHD Linked to Lead and Mom's Smoking, by Karen Barrow **(**February 1, 2007)

> *A mother's smoking during pregnancy and exposure to lead significantly increases her child's risk for developing attention deficit hyperactivity disorder (ADHD), say researchers. In fact, as many as one third of cases of ADHD in children are linked to exposure to tobacco smoke and lead before birth, giving moms yet another reason to quit smoking during pregnancy.*
>
> *For the study, researchers from Cincinnati Children's Hospital Medical Center surveyed over 4,700 children between the ages of 4 and 15 and their parents. Over 4 percent of the children included had ADHD. The researchers found that those children whose mother smoked during pregnancy were over twice as likely to develop ADHD than a child whose mother had not smoked.*

Based on this study, should we conclude that smoking during pregnancy *causes* an increase in the likelihood that a child develops ADHD? Explain.

_____ occurs when two variables are associated in such a way that their effects on a response variable cannot be distinguished from each other.

Explain the concept of confounding in the context of this study.

Is there any way to *prove* that smoking causes ADHD?

Read 234–236

An _____ observes individuals and measures variables of interest but does not attempt to influence the response.

An _____ deliberately imposes some treatment on individuals to measure their responses.

What are some differences between an observational study and an experiment?

Check your understanding: 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.

(a) Was this an observational study or an experiment? Justify.

(b) What are the explanatory and response variables?

(c) Explain why such a study cannot establish a cause-and-effect relationship. Suggest a variable that may be confounding with whether families eat dinner together.

The Language of Experiments

Read 237-239

Experiment Terms	Definition
	Smallest collection of individuals to which treatments are applied. When the units are human beings, they are often called **subjects**.
	A specific condition applied to the individuals in an experiment.
	The explanatory variable(s) in an experiment are sometimes called **factors**.
	In experiments that study the joint effects of several factors, each treatment is formed by combining a specific value (often called **level**) of each of the factors. *When there is only one factor (explanatory variable) in an experiment, the levels of factors are the same as the treatments.*

Alternate Example. Does adding fertilizer affect the productivity of tomato plants? How about the amount of water given to the plant? 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 gallons 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. Identify the experimental units or subjects, explanatory and response variables, and the treatments.

subjects: explanatory variable(s)/factor(s):

response variable(s): treatments:

***Watch Ted Talk video "How to Buy Happiness" for an example of a multi-factor experiment.*

HW: page 233 (37–42) page 259 (45–55 odd) **Directions for 51–56 are the same*

Designing Experiments

Suppose we wanted to design an experiment to see if caffeine affects pulse rate.

Here is an initial plan:

- measure initial pulse rate
- give each student some caffeine
- wait for a specified time
- measure final pulse rate
- compare final and initial rates

What are some problems with this plan? What other variables are most likely to be sources of variability in pulse rates?

There are several steps we should take to solve these problems:

1. The first step is to include a _____ that does not receive caffeine so we have something to compare to. 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.

2. The second step in solving the problem 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 randomization, replication, and control.

   (a) We _____ subjects to treatments 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. 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?"

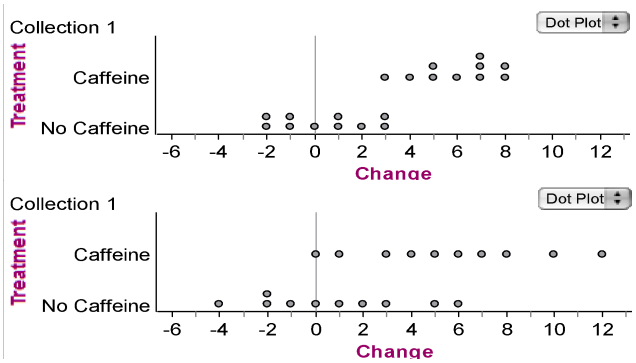
In a _____, the experimental units are assigned to the treatments completely by chance.

Explain how to set up a randomized experiment for the caffeine example. *Let's say we want 15 subjects in each treatment.*

(b) _____ means holding other variables constant for each member of both treatment groups. This prevents these other variables from becoming confounded with caffeine and from adding additional variability to the distribution of the response variable.

(c) _____ means ensuring that there are an adequate number of observations in each treatment group so that the two groups are as equivalent as possible. Then, differences in the effects of the treatments can be distinguished from chance differences between the groups.

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 _____.



Since the subjects have been randomly assigned to the treatments, can the weight of the students be considered a confounding variable in this experiment?

It is also important that all subjects in both groups are _____ so that the expectations are the same for the subjects in both groups. Otherwise, members of the caffeine group might suffer from the _____. If the people measuring the response are also blind, the experiment is _____.

Note: Not all experiments have a control group or use a placebo as long as there is comparison. For example, 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.

***Interesting Articles on the Placebo Effect: A More Expensive Placebos Bring More Relief, Hooked on a Feeling: This is Your Brain on Placebo, The Growing Power of the Sugar Pill, 60 Minutes Video: Treating Depression: Is there a placebo effect?*

IN SUMMARY: The 4 Principles of Experimental Design

1. Comparison
2. Randomization
3. Control
4. Replication

Read 243–249

Alternate Example: Explain how the following experiment demonstrates the 4 principles.

Multitasking: 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.

HW: page 260 (57, 59, 61, 63, 67, 69, 71)

Inference for Experiments

Read 249

The results of an experiment are called _____ if they are unlikely to occur by random chance. That is, if it is unlikely that the results are due to the possible imbalances created the random assignment.

For example, if caffeine really has no effect on pulse rates, then the average change in pulse rate of the two groups should be exactly the same. However, because the results will vary depending on which subjects are assigned to which group, the average change in the two groups will probably differ slightly. Thus, whenever we do an experiment and find a difference between two groups, we need to determine if this difference could be attributed to the chance variation in random assignment or because there really is a difference in effect of the treatments.

Sample Results: Increases in the heart rate of 14 students are shown below. (Note that the negative values mean their heart rate decreased).

Caffeinated Today	8	12	10	14	2	0	0
Non-Caffeinated Soda	-6	0	1	2	-3	-4	2

How can we determine if the results of our experiment are statistically significant?

HW: page 259 (48, 58, 64, 68, 73)

Blocking & Matched Pairs

Read 251–255

Alternate Example*: SAT schools*

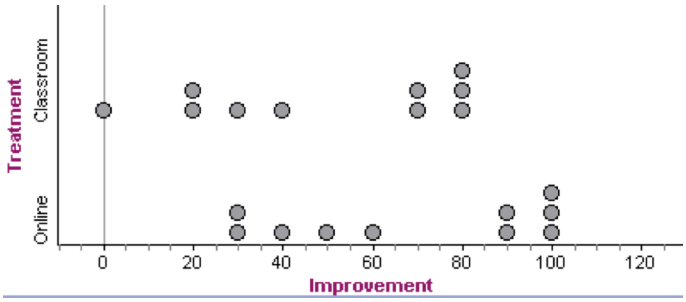
Many students enroll in prep courses to improve their SAT scores. Twenty students who have taken the SAT once volunteered to participate in an experiment comparing online and classroom prep courses.

1. Describe how we can use a completely randomized design to compare online and classroom SAT prep courses.

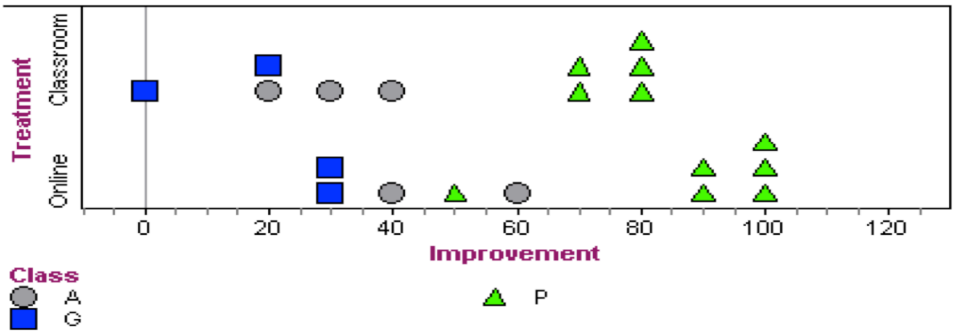
2. Among the 20 volunteers, 10 of are in Precalculus, 6 are in Algebra 2, and 4 are in Geometry. What problem does this cause? How can we address this problem?

3. Here are the results of the experiment, using math level as a blocking variable. The dotplots compare the improvements of the students in the online course and the improvements of students in the classroom course. Based on the dotplots, does there appear to be convincing evidence that the online course is better?

Class	Treatment	Improvement
P	Online	100
P	Online	100
P	Online	90
P	Online	90
P	Online	100
P	Classroom	70
P	Classroom	70
P	Classroom	80
P	Classroom	80
P	Classroom	80
A	Online	50
A	Online	60
A	Online	40
A	Classroom	30
A	Classroom	40
A	Classroom	20
G	Online	30
G	Online	30
G	Classroom	0
G	Classroom	20



4. The dotplots in #3 ignored the fact that we blocked by math level. Here is the dotplot again, using different symbols for students in each math level.



Notice that within each math level, the online students clearly did better. We couldn't see this difference when we ignored the blocks. The average improvement for students was:

Precal: Alg 2: Geo:

How can we use this information to account for the variability created by differences in class level?

Class	Treatment	Improvement	Adj. Imp.
P	Online	100	
P	Online	100	
P	Online	90	
P	Online	90	
P	Online	100	
P	Classroom	70	
P	Classroom	70	
P	Classroom	80	
P	Classroom	80	
P	Classroom	80	
A	Online	50	
A	Online	60	
A	Online	40	
A	Classroom	30	
A	Classroom	40	
A	Classroom	20	
G	Online	30	
G	Online	30	
G	Classroom	0	
G	Classroom	20	

_____ in experiments is similar to stratification in sampling.

- Blocking accounts for a source of variability, just like stratifying. This means that blocking is a good way to increase power.
- Blocks should be chosen like strata: the units within the block should be similar, but different than the units in the other blocks. You should only block when you expect that the blocking variable is associated with the response variable.
- Blocks, like strata, are not formed at random!

What are some variables that we can block for in the caffeine experiment? In general, how can we determine which variables might be best for blocking? What about a matched pairs design?

A common type of randomized block design for comparing two treatments is a _____ _____. The idea is to create blocks of matching pairs of similar experimental units.

Could we use a matched pairs design for the caffeine experiment?

Alternate Example: *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. She goes to the store and buys 10 bags each of 4 different varieties of microwave popcorn (movie butter, light butter, natural, and kettle corn), for a total of 40 bags.

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

Outline a randomized block design for this experiment.

What is a matched pairs design? Could we use a matched pairs design for the caffeine experiment?

HW: page 257 (75–85 odd)