Chapter 4: Designing Studies

Section 4.2
Experiments

Observational Study versus Experiment
In contrast to observational studies, experiments don't just observe individuals o them questions. They actively impose some treatment in order to measure the response.

An **observational study:** observes individuals and measures variables of interest but does not attempt to influence the responses.

An **experiment:** deliberately imposes some treatment on individuals to measure their responses.

Observational studies of the effect of one variable on another often fail because of **confounding** between the explanatory variable and one or more **lurking variables.**

A **lurking variable** is a variable that is not among the explanatory or response variables in a study but that may influence the response variable.

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

Example 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.

An experiment because a treatment (brightness of screen) was imposed on the laptops.

Example 2: 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." Th finding was based on a sample survey conducted by researchers at Columbia University.

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

An observational study. Students were not assigned to a particular number of meals to eat with their family per week.

b) What are the explanatory and response variables?

Explanatory: number of meals per week eaten with their family. Response: GPA.

c) Explain clearly why such a study cannot establish a cause-and-effect relationship.

There are probably other variables that are influencing the response variable. For example, students who have part-time jobs may not be able to eat many meals with their families and may not have much time to study, leading to lower grades.

The Language of Experiments

A specific condition applied to the individuals in an experiment is called a **treatment**. If an experiment has several explanatory variables, a treatment is a combination of specific values of these variables.

The **experimental units** are the smallest collection of individuals to which treatments are applied. When the units are human beings, they often are called **subjects**.

Example 3: Researchers at the University of North Carolina were concerned about the increasing dropout rate in the state's high schools, especially for low-income students. Surveys of recent dropouts revealed that many of these students had started to l interest during middle school. They said they saw little connection between what they were studying in school and their future plans. To change this perception, researchers developed a program called CareerStart. The big idea of the program is that teachers show students how the topics they learn get used in specific careers. To test the effectiveness of CareerStart, the researchers recruited 14 middle schools in Forsyth County to participate in an experiment. Seven of the schools, determined at random, used CareerStart along with the district's standard curriculum. The other seven schools just followed the standard curriculum. Researchers followed both groups of students for several years, collecting data on students' attendance, behavior, standardized test scores, level of engagement in school, and whether or not the students graduated from high school. *Results:* Students at schools that used CareerStart generally had better attendance and fewer discipline problems, earned higher test scores, reported greater engagement in their classes, and were more likely to graduate. Identify the experimental units, explanatory and response variables, and the treatments in the CareerStart experiment.

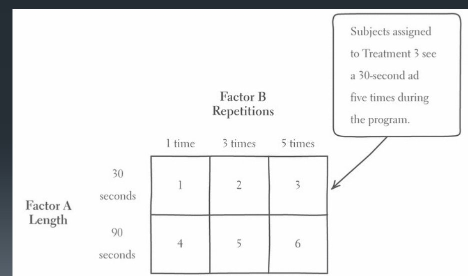
The experimental units are 14 middle schools in Forsyth County, NC. The explanatory variable is whether the school used the CareerStart program with its students. Several response variables were measured, including test scores, attendance, behavior, student engagement, and graduation rates. This experiment compares two treatments: (1) the standard middle school curriculum and (2) the standard curriculum plus CareerStart.

Sometimes, the explanatory variables in an experiment are called **factors**. Many experiments study the joint effects of several factors. In such an experiment, each treatment is formed by combining a specific value (often called a **level**) of each of the factors.

Example 4: What are the effects of repeated exposure to an advertising message? The answer may depend on both the length of the ad and on how often it is repeated. An experiment investigated this question using 120 undergraduate students who volunteered to participate. All subjects viewed a 40-minute television program that included ads for a digital camera. Some subjects saw a 30-second commercial; others, a 90-second version. The same commercial was shown either 1, 3, or 5 times during the program. After viewing, all the subjects answered questions about their recall of the ad, their attitude toward the camera, and their intention to purchase it. For the advertising study, identify the experimental units or subjects, explanatory and response variables, and the treatments.

The subjects are the 120 undergraduate students. This experiment has 2 explanatory variables (factors): length of the commercial and number of repetitions. The response variables include measures of subjects' recall of the ad, their attitudes about the digital camera, and whether they intend to purchase it.

Example 4 cont.:
There are 2 different lengths of commercial (30 and 90 seconds) and three different numbers of repetitions (1, 3, and 5). The 6 combinations consisting of one level of each factor form the 6 treatments shown in the figure below: (1) 30 seconds, 1 time; (2) 30 seconds, 3 times; (3) 30 seconds, 5 times; (4) 90 seconds, 1 time; (5) 90 seconds, 3 times; (6) 90 seconds, 5 times.



How to Experiment Badly

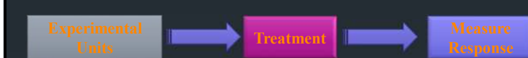
Example 5: A high school regularly offers a review course to prepare students for the SAT. This year, budget cuts will allow the school to offer only an online version of the course. Suppose the group of students who take the online course earn an average increase of 45 points in their math scores from a pretest to the actual SAT test. Can we conclude that the online course is effective?

This experiment has a very simple design. A group of subjects (the students) were exposed to a treatment (the online course), and the outcome (increase in math scores) was observed. Here is the design:

Students → Online course → increase in math scores

A closer look showed that many of the students in the online review course were taking advanced math classes in school. Maybe the students in the online course improved their math scores because of what they were learning in their school math classes, not because of the online course. This confounding prevents us from concluding that the online course is effective.

How to Experiment Badly
Many laboratory experiments use a design like the one in the online SAT course example:



In the lab environment, simple designs often work well. Field experiments and experiments with animals or people deal with more varied conditions. *Outside the lab, badly designed experiments often yield worthless results because of confounding.*

How to Experiment Well
Comparison alone isn't enough to produce results we can trust. If the treatments are given to groups that differ greatly when the experiment begins, *bias* will result. For example, if we allow students to select online or classroom instruction, more self-motivated students are likely to sign up for the online course. Allowing personal choice will bias our results in the same way that volunteers bias the results of online opinion polls. The solution to the problem of bias in sampling is random selection. In experiments, the solution is **random assignment.**

In an experiment, **random assignment** means that experimental units are assigned to treatments using a chance process.

Example 6: This year, the high school has enough budget money to compare the online SAT course with the classroom SAT course. Fifty students have agreed to participate in an experiment comparing the two instructional methods. Describe how you would randomly assign 25 students to each of the three methods:

a) Using 50 identical slips of paper

The simplest way would be to use the "hat method." Write each subject's name on one of the slips. Put all the slips in a hat and mix them thoroughly. Draw them out one at a time until you have 25 slips. These 25 students will take the online course. The remaining 25 students will take the classroom course. Alternatively, you could write "online" on 25 of the slips and "classroom" on the other 25 slips. Then put the slips in a hat and mix them well. Have students come up one by one and (without looking) pick a slip from the hat. This guarantees 25 students per group, with the treatments assigned by chance.

b) Using technology

Give numbers 1, 2, 3, …, 49, 50 to the subjects in alphabetical order by last name. Then use your calculator's randInt command or a computer's random number generator to produce numbers between 1 and 50. Ignore any repeated numbers. The first 25 different numbers chosen select the students for the online course. The remaining 25 subjects will take the classroom course.

c) Using Table D

We can use the two-step process from random sampling to do the random assignment.

Step 1: Label. Give labels 01, 02, 03,…, 49, 50 to the subjects in alphabetical order by last name.

Step 2: Table. Go to a line of Table D and read two-digit groups moving from left to right. The first 25 distinct labels between 01 and 50 identify the 25 students that are assigned to the online course. The remaining 25 students will take the classroom course. Ignore repeated labels and groups of digits from 51 to 00.

PRINCIPLES OF EXPERIMENTAL DESIGN

The basic principles for designing experiments are as follows:

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

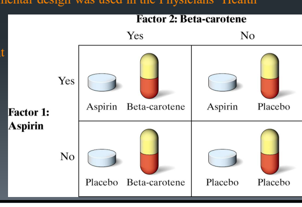
2. **Random assignment**. Use chance to assign experimental units to treatments. Doing so helps create roughly equivalent groups of experimental units by balancing the effects of other variables among the treatment groups.

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.

Example 7: 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 below. One-fourth of the subjects were assigned at random to each of these treatments. On odd-numbered days, the subjects took either a tablet that contained aspirin or a dummy pill that looked and tasted like the aspirin but had no active ingredient (a **placebo**). On even-numbered days, they took either a capsule containing beta-carotene or a placebo. There were several response variables—the study looked for heart attacks, several kinds of cancer, and other medical outcomes. After several years, 239 of the placebo group but only 139 of the aspirin group had suffered heart attacks. This difference is large enough to give good evidence that taking aspirin does reduce heart attacks. It did not appear, however, that beta-carotene had any effect on preventing cancer. Explain how each of the four principles of experimental design was used in the Physicians' Health Study.

A **placebo** is a "dummy pill" or inactive treatment that is indistinguishable from the real treatment.



Comparison: Researchers used a design that compared both of the active treatments to a placebo.

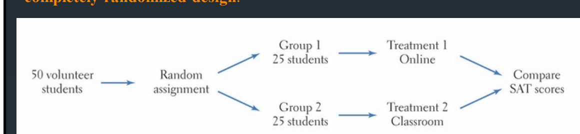
Random assignment: Was used to determine which subjects received each of the four treatment combinations. This helped ensure that the treatment groups were roughly equivalent to begin with.

Control: The experiment used subjects of the same gender and occupation. All subjects followed the same schedule of pill taking.

Replication: There were over 5000 subjects per treatment group. This large number of subjects helped ensure that the difference in heart attacks was due to the aspirin and not to chance variation in the random assignment.

Completely Randomized Designs

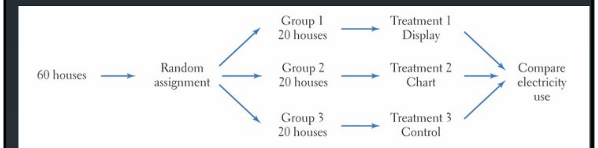
The diagram below presents the details of the SAT prep experiment: random assignment, the sizes of the groups and which treatment they receive, and the response variable. There are, as we will see later, statistical reasons for using treatment groups that are about equal in size. This type of design is called a **completely randomized design**.



In a **completely randomized design,** the experimental units are assigned to the treatments completely by chance.

Example 8: Many utility companies have introduced programs to encourage energy conservation among their customers. An electric company considers placing small digital displays in households to show current electricity use and what the cost would be if this use continued for a month. Will the displays reduce electricity use? One cheaper approach is to give customers a chart and information about monitoring their electricity use from their outside meter. Would this method work almost as well? The company decides to conduct an experiment to compare these two approaches (display, chart) with a group of customers who receive information about energy consumption but no help in monitoring electricity use. Describe a completely randomized design involving 60 single-family residences in the same city whose owners are willing to participate in such an experiment. Write a few sentences explaining how you would implement your design.

The figure below outlines the design. We'll randomly assign 20 houses to each of three treatments: digital display, chart plus information, and information only. Our response variable is the total amount of electricity used in a year.



To implement the design, start by labeling each house with a distinct number from 1 to 60. Write the labels on 60 identical slips of paper, put them in a hat, and mix them well. Draw out 20 slips. The corresponding homes will be given digital displays showing current electricity use. Now draw out 20 more slips. Those homes will use a chart. The remaining 20 houses will be given information about energy consumption but no way to monitor their usage. At the end of the year, compare how much electricity was used by the homes in the three groups.

AP EXAM TIP: If you are asked to describe the design of an experiment on the AP exam, you won't get full credit for a diagram like the figure in example 8. You are expected to describe how the treatments are assigned to the experimental units and to clearly state what will be measured or compared. Some students prefer to start with a diagram and then add a few sentences. Others choose to skip the diagram and put their entire response in narrative form.

Experiments: What can go wrong?

A response to a dummy treatment is called a **placebo effect**. The strength of the placebo effect is a strong argument for randomized comparative experiments.

Example 9: Want to help balding men keep their hair? Give them a placebo. One study found that 42% of balding men maintained or increased the amount of hair on their heads when they took a placebo. In another study, researchers zapped the wrists of 24 test subjects with a painful jolt of electricity. Then they rubbed a cream with no active medicine on subjects' wrists and told them the cream should help soothe the pain. When researchers shocked them again, 8 subjects said they experienced significantly less pain. When the ailment is vague and psychological, like depression, some experts think that the placebo effect accounts for about three-quarters of the effect of the most widely used drugs. Others disagree. In any case, "placebos work" is a good place to start when you think about planning medical experiments.

Whenever possible, experiments with human subjects should be **double-blind**.

In a **double-blind experiment,** neither the subjects nor those who interact with them and measure the response variable know which treatment a subject received.

Inference for Experiments

An observed effect so large that it would rarely occur by chance is called **statistically significant.**

A statistically significant association in data from a well-designed experiment does imply causation.

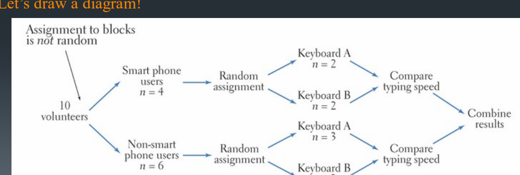
Blocking

Example 10: Suppose that a mobile phone company is considering two different keyboard designs(A and B) for its new smart phone. The company decides to perform an experiment to compare the two keyboards using a group of 10 volunteers. The response variable is typing speed, measured in words per minute.

How should the company deal with the fact that four of the volunteers already use a smart phone, whereas the remaining six volunteers do not? They could use a completely randomized design and hope that the random assignment distributes the smart-phone users and non-smart-phone users about evenly between the group using keyboard A and the group using keyboard B. Even so, there might be a lot of variability in typing speed in both groups because some members of each group are much more familiar with smart phones than others. This additional variability might make it difficult to detect a difference in the effectiveness of the two keyboards. What should the researchers do?

Because the company knows that experience with smart phones will affect typing speed, they could start by separating the volunteers into two groups—one with experienced smart-phone users and one with inexperienced smart-phone users. Each of these groups of similar subjects is known as a **block**. Within each block, the company could then randomly assign half of the subjects to use keyboard A and the other half to use keyboard B. To control other variables, each subject should be given the same passage to type while in a quiet room with no distractions. This **randomized block design** helps account for the variation in typing speed that is due to experience with smart phones.

Let's draw a diagram!



A **block** is a group of experimental units that are known before the experiment to be similar in some way that is expected to affect the response to the treatments.

In a **randomized block design**, the random assignment of experimental units to treatments is carried out separately within each block.

Form blocks based on the most important unavoidable sources of variability (lurking variables) among the experimental units.

Randomization will average out the effects of the remaining lurking variables and allow an unbiased comparison of the treatments.

Control what you can, block on what you can't control, and randomize to create comparable groups.

Example 11: Women and men respond differently to advertising. Researchers would like to design an experiment to compare the effectiveness of three advertisements for the same product.

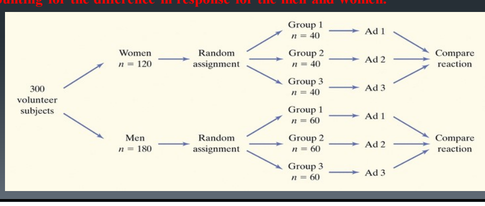
a) Explain why a randomized block design might be preferable to a completely randomized design for this experiment.

A completely randomized design considers all subjects, both men and women, as a single pool. The random assignment would send subjects to three treatment groups without regard to their gender. This ignores the differences between men and women, which would probably result in a great deal of variability in responses to the advertising in all three groups. For example, if an ad appealed much more to men, you would get a wide range of reactions to that ad from the two genders. That would make it harder to determine whether one ad was more effective.

A randomized block design would consider women and men separately. In this case, the random assignment would occur separately in each block. Blocking will account for the variability in responses to advertising due to gender. This will allow researchers to look separately at the reactions of men and women, as well as to more effectively assess the overall response to the ads.

b) Outline a randomized block design using 300 volunteers (180 men and 120 women) as subjects. Describe how you would carry out the random assignment required by your design.

We randomly assign the 120 women into three groups of 40, one for each of the advertising treatments. Write the women's names on 120 identical slips of paper, place the slips in a hat, and mix them well. Pull out 40 slips to determine which women will view Ad 1. Pull out another 40 slips to determine which women will view Ad 2. The remaining 40 women will view Ad 3. Randomly assign the 180 men into three groups of 60 using a similar process. After each subject has viewed the assigned ad, compare reactions to the three ads within the gender blocks. To compare the overall effectiveness of the three ads, combine the results from the two blocks after accounting for the difference in response for the men and women.



Matched Pair Designs

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

A **matched-pairs design** is a randomized blocked experiment in which each block consists of a matching pair of similar experimental units.

Chance is used to determine which unit in each pair gets each treatment.

Sometimes, a "pair" in a matched-pairs design consists of a single unit that receives both treatments. Since the order of the treatments can influence the response, chance is used to determine with treatment is applied first for each unit.