Teacher Notes for Section III: Experiments

Observational studies are a very useful tool for gaining information about the world around us. With observational studies, we can determine the answers to many questions: Has drivers' seat belt use in the states increased since last year? How often do people wash their hands after using the bathroom? Is there a relationship between the height and shoe print length of teenagers? Are meat hot dogs less healthy than poultry hot dogs?

But there are many other important questions that observational studies cannot help us answer. Here are a few: Does smoking cause lung cancer? Is a new medication for treating migraine headaches more effective than the current treatment that doctors most often prescribe? Which is more effective for reducing weight in obese adults, a low-fat diet or a low-carbohydrate diet? Does listening to Mozart help people memorize better than working in silence? To get answers to these questions, which suggest some sort of cause-and-effect conclusions, we must perform experiments. In this section, students will examine experiments in more detail.

There are three investigations in this section.

Investigation # 10: Do Diets Work?

This investigation presents students with the results of two well-designed experiments that compared the effectiveness of low-carb and low-fat diets in reducing weight and lowering cholesterol in obese adults. Students are led in a step-by-step fashion to identify and describe specific design elements of these two experiments. Then, students are asked to interpret the results of the two studies in context, taking into account some possible limitations of each study.

Investigation #11: Distracted Learning

Students begin by incrementally designing an experiment to test whether listening to Mozart improves performance on a memorization task. With their design established, students carry out the experiment using the members of their class as subjects. Once the data have been produced, students set about the task of analyzing and drawing conclusions from the data.

Investigation #12: Would You Drink a Blue Soda?

This investigation serves as a culminating investigation on experiments. From what they have learned in the Overview and from completing investigations #11 and #12, students should be ready to design an experiment on their own. This time, they can use random selection to choose subjects, which will extend their ability to generalize results to a larger population of interest.

Prerequisites

Students should be able to distinguish an observational study from a survey or an experiment.

Learning Objectives

As a result of completing the Overview, students should be able to:

- Identify the experimental units/subjects, factor(s)/explanatory variable(s), treatments, and response variable(s) in an experimental setting
- Explain the purpose of randomly assigning treatments to subjects in an experiment
- Determine whether an experiment was carried out in a single-blind or double-blind manner
- Explain the purpose of control in an experimental design
- Explain what is meant by "statistical significance"
- Explain why replication is an important experimental design principle
- Identify a potential confounding variable in a study and explain how the variable could result in confounding
- Explain how the way in which data were produced affects our ability to generalize results to a larger population of interest

Teaching Tips

The Overview is chock full of important terminology and issues related to experimental design. Our advice is to have students take turns reading the material aloud, pausing at appropriate spots to clarify definitions and ideas.

In the first paragraph, we distinguish an experiment from an observational study. Simply put, an experiment requires that researchers deliberately impose specific conditions and measure some response.

As our first example of an experiment, we consider a biologist who wants to compare the effects of two brands of weed killer on a particular variety of broad-leafed plant found in a university's garden. A primary purpose of this example is to convince students that the method of assigning treatments to experimental units is vitally important. More specifically, we argue that the "best" method of determining which experimental units receive which treatments is to let chance decide. This process of "random assignment" gives researchers the best hope of starting out with fairly equivalent groups of experimental units prior to administering treatments. Without random assignment, researchers risk creating groups of experimental units that differ in some important way that could systematically affect their response to the treatments. Then, any differences in response between the groups could be due to these initial differences, rather than to the effects of the treatments. This circumstance, in which the effects of the treatments are hopelessly mixed up with the effects of some other variable on individuals' responses, is known as *confounding*. Random assignment of treatments to experimental units gives researchers a powerful tool for avoiding confounding.

Random assignment also helps with the primary goal of an experiment: establishing that the difference in treatments *caused* a difference in responses. This is a key advantage of experiments over observational studies; well designed experiments allow researchers to make cause-and-effect conclusions. An observational study comparing two or more groups—even one involving random selection of individuals from the corresponding populations of interest—cannot provide convincing evidence of causation. Why not? Because we can't isolate the effects of the variable(s) we're interested in from the effects of other variables. We discuss this limitation of observational studies in detail in the final paragraph of the Overview using the well-known setting of trying to determine whether smoking cigarettes causes cancer in humans.

When a well-designed experiment reveals differences in responses between treatment groups, there are two possible explanations: (1) the difference in responses was caused by the different effects of the treatments, or (2) the treatments actually have the same effect on experimental units, so the difference in responses is not due to the effects of the treatments, but rather to the chance involved in the random assignment of treatments to experimental units. More experienced users of statistics can calculate the probability (chance) of obtaining a difference in responses as large as or larger than the one actually observed in the study just from the random assignment. Based on this probability, we can determine whether explanation (2) is a plausible explanation for the observed difference. If not, we conclude that the observed difference is **statistically significant** and that we favor explanation (1). Such decisions based on probability are the foundation of inference, which is introduced in Section IV of this module.

Once students have read about and discussed the three essential experimental design principles—random assignment, control, and replication—in the context of the weed killer example, you may want to ask them to explain how these principles apply in the subsequent example describing the Physicians' Health Study.

The Physicians' Health Study is a famous example of a well-designed experiment that showed taking aspirin regularly helps reduce the risk of heart attack—at least for middle-aged, male physicians.

For reference, here is a complete listing of the vocabulary from the Overview:

Experimental units/subjects: The individuals who take part in an experiment

Treatments: The specific conditions that researchers impose on experimental units

Confounding: When it is impossible to separate the effects of the treatments from the effects of another variable on the response variable in an experiment

Random assignment: A fundamental principle of experimental design that involves using a chance mechanism to allocate treatments to experimental units

Explanatory variable/factor: A variable that is deliberately manipulated by the researcher to measure experimental units' responses

Response variable: A variable that measures experimental units' responses to the treatments

Control: An important principle of experimental design that entails trying to ensure that variables other than the explanatory variable(s) have roughly equivalent effects on the experimental units that are assigned to the different treatment groups. Researchers can either try to hold the values of such variables constant throughout the experiment or rely on the random assignment to balance out the effects of these variables on the experimental units in different treatment groups.

Replication: A fundamental principle of experimental design that involves giving each treatment to enough experimental units so that any difference in the overall effects of the treatments can be detected

Placebo: A fake treatment

Double-blind: When neither the subjects nor the individuals measuring subjects' responses know who is receiving which treatment

Single-blind: When either the subjects or the people measuring subject's responses, but not both, are unaware of who is receiving which treatment

Statistically significant: A difference in responses that cannot be accounted for by the chance involved in the random assignment of treatments to experimental units

Possible Extensions

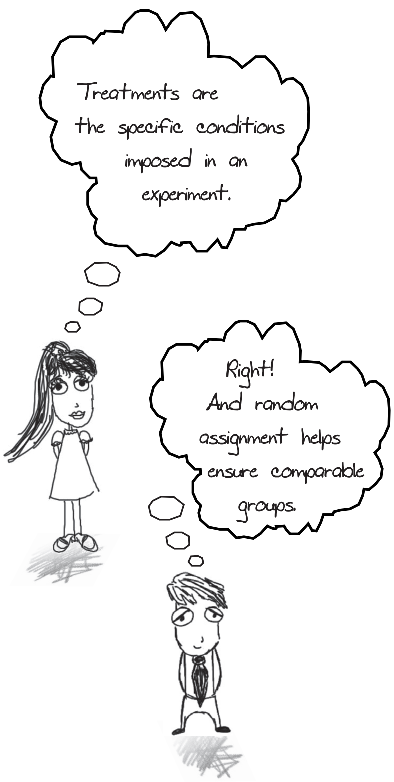
You might want to show students a video clip describing the Physicians' Health Study experiment. The Annenberg/Corporation for Public Broadcasting web site, *www.learner.org*, houses a series of instructional statistics videos called "Against All Odds: Inside Statistics." By completing a free registration process, you can play any of these videos as streaming downloads on your computer. The Physicians' Health Study clip is in Video 12: Experiments. The Physicians' Health Study web site, *phs.bwh.harvard.edu*, contains additional information about the experiment, including the results of the beta carotene treatment (no statistically significant difference from placebo beta carotene).

The more recent Women's Health Initiative (WHI), begun in 1991, included clinical trials and an observational study that examined the effects of hormone therapy, diet, and vitamin supplements in postmenopausal women. The WHI's web site is *www.nhlbi.nih.gov/whi*.

Section III: Experiments

Corresponds to pp. 62-66
in Student Module

In an observational study, researchers make observations and record data. As much as possible, the observer tries not to influence what is being observed. In an experiment, researchers deliberately do something and then measure a response. The "participants" in an experiment are called **experimental units**. Experimental units can be people, animals, or objects. When the experimental units are people, they are often referred to as **subjects**. The specific conditions researchers impose on the experimental units are called **treatments**. As experimental units may differ from one another in many important ways, the method of assigning treatments to experimental units is an important concern in the experimental design process.



Let's look at an example. A biologist would like to determine which of two leading brands of weed killer is less likely to harm the broad-leafed plants in a garden at the university. Before spraying near the plants in the garden, the biologist decides to conduct an experiment that will allow her to compare the effects of these two brands of weed killer on broad-leafed pansy plants (one of the varieties in the garden). The biologist obtains 24 individual pansy plants to use in the experiment. In this simple experiment, the *experimental units* are the individual pansy plants and the *treatments* are the two brands of weed killer.

Consider the following two plans for assigning treatments to the pansy plants:

Plan A: Choose the 12 healthiest looking pansy plants. Apply brand X weed killer to all 12 of those plants. Apply brand Y weed killer to the remaining 12 pansy plants.

Plan B: Choose 12 of the 24 individual pansy plants at random. Apply brand X weed killer to those 12 plants and brand Y weed killer to the remaining 12 plants.

Which plan seems preferable? Let's evaluate what might happen with each of these plans.

Under Plan A, suppose the pansy plants treated with brand Y weed killer have many more dead or dying leaves than the pansy plants treated with brand X. Can the biologist feel confident recommending brand X to the campus gardener as the safer weed killer? Not at all. Since the healthier plants received the brand X treatment and the less healthy plants received the brand Y treatment, it could be that more leaves were dead or dying on the pansy plants treated with brand Y because those plants were less healthy to begin with. We really can't separate the effects of the two brands of weed killer from the effect of the original healthiness of the plants in the two groups. The inability to separate the effects of the treatments from the effects of another variable in a study is known as **confounding**.

With Plan B, individual pansy plants are assigned at random to one of the two weed killer treatments. This **random assignment** helps to ensure that the group of plants treated with brand X and the group of plants treated with brand Y are fairly similar to begin with in terms of all characteristics that might affect the plants' responses to the treatments. If the biologist then observes that the pansy plants treated with brand Y weed killer have many more dead or dying leaves than the pansy plants treated with brand X, there are two plausible explanations for the observed difference.

First, it is possible that there is no difference in the effects of the two brands of weed killer on pansy plants. Some pansies are heartier than others, and, just by chance, the random assignment placed more of those healthy plants in the group that was treated with brand X. In other words, the observed difference could be simply due to chance.

The second possible explanation is that brand X weed killer actually results in greater harm to pansy plants than brand Y. In that case, we could say the difference in the number of dead or dying leaves between the two groups of pansy plants is a direct result of the brand of weed killer used. Put another way, the difference in brand of weed killer *caused* the difference in the number of dead or dying leaves.

Random assignment of treatments to subjects is an essential component of well-designed experiments. One of the big advantages of such experiments is their ability to help the researcher establish that changes in one variable (like brand of weed killer) cause changes in another variable (like number of dead or dying leaves). Since establishing causation is often a goal of experiments, we find it useful to give names to the two variables mentioned in the previous sentence. We call the variables that the experimenters directly manipulate the **explanatory variables** or **factors** and the variables that measure the subjects' responses to the treatments the **response variables**. The treatments in an experiment correspond to the different possible values of the explanatory variables. For the weed killer experiment above, there is one factor—brand of weed killer—and one response variable—number of dead or dying leaves.

In addition to randomly assigning treatments to experimental units, there are two other important considerations in designing experiments. The first is to **control** for the effects of variables that are not factors in the experiment but that might affect experimental units' responses to the treatments. Some variables can be controlled by trying to keep them at a constant value. For example, the biologist would want to ensure that the plants all receive the same amount of water and are exposed to the same amount of light. If everything is roughly equivalent for the two groups of plants except for the treatments, and we observe a difference in the response variable, then that difference is either a result of the random assignment or is caused by the difference in treatments.

Some variables can't be easily controlled by keeping them at a constant value. One such variable in the weed killer example was the current state of health of the plant. In this case, the random assignment of plants to treatments should help spread the healthy and less healthy plants out in a fairly balanced way between the two groups of pansy plants. Then, any differences in the number of dead or dying leaves that appear should not be a result of differences in initial plant health.

The other important experimental design principle is **replication**. In a nutshell, replication means giving each treatment to enough experimental units so that any difference in the effects of the treatments is likely to be detected. Imagine the biologist treating one pansy plant with brand X weed killer and one pansy plant with brand Y weed killer. If the plant treated with brand Y has more dead or dying leaves, can the biologist conclude that brand X is safer to use on the university's pansy plants? Of course not. Individual pansy plants vary widely in terms of general health and other characteristics that might affect their response to a particular brand of weed killer. With only one experimental unit available for each treatment, the random assignment can't be counted on to produce roughly "equivalent" groups prior to administering the treatments. Any difference we observe in the number of dead or dying leaves on the two pansy plants could simply be due to the difference in the initial health of the plants.

Now imagine the biologist conducting the same weed killer experiment, but with 50 pansy plants receiving each treatment. If the pansies treated with brand Y have a much higher number of dead or dying leaves than the pansies treated with brand X, the biologist should feel much more confident concluding that the difference in treatments caused the observed difference in the response variable.

Let's look at one more example. In the fall of 1982, researchers launched a now famous experiment investigating the effects of aspirin and beta carotene on heart disease and cancer. Over 22,000 healthy male physicians between the ages of 40 and 84 agreed to serve as *subjects* in the experiment. The two *factors* being manipulated by the researchers were whether a person took aspirin regularly and whether a person took beta carotene regularly. Researchers decided to use four treatments: (1) aspirin every other day and beta carotene every other day, (2) aspirin every other day and "fake" beta carotene every other day, (3) "fake" aspirin every other day and beta carotene every other day, and (4) "fake" aspirin every other day and "fake" beta carotene every other day.

The "fake" pills looked, tasted, and smelled like the pills with the active ingredient, but had no active ingredient themselves. (We call such "fake" treatments **placebos**.) Subjects were randomly assigned in roughly equal numbers to the four groups. Several *response variables* were measured in the study, including whether the individual had a heart attack and whether the individual developed cancer. Neither the subjects nor the people measuring the response variable knew who was receiving which treatment. We say this experiment was carried out in a **double-blind** manner. If either the subjects or the people measuring the response variable knows who is receiving which treatment, but the other doesn't, then the experiment is **single-blind**.

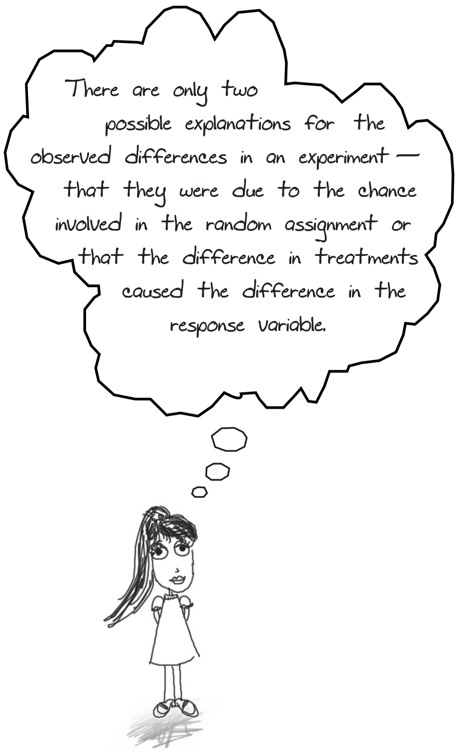
An outside group of statisticians that was monitoring the Physicians' Health Study reviewed data from the experiment on a regular basis. To everyone's surprise, the data monitoring board stopped the aspirin part of the experiment several years ahead of schedule. Why? Because there was compelling evidence that the subjects taking aspirin were having far fewer heart attacks than those who were taking placebo aspirin. It would have been unethical to continue allowing some physicians to take a placebo with clear evidence that aspirin reduced the risk of heart attack.

Even though the Physicians' Health Study was an exceptionally well-designed experiment, it does have some limitations. Researchers decided to use male physicians as subjects because they felt doctors would be more likely to understand the importance of taking the pills every other day for the duration of the study. That may be true, but because only male physicians were used in the study, we cannot generalize the findings of this study to women, or even to all male adults. We can feel pretty confident concluding that taking aspirin regularly *caused* a reduction in heart attack risk. However, the benefits of taking aspirin regularly might be offset by other effects of the drug, such as an increased risk of stroke. In spite of its limitations, the Physicians' Health Study provided a template for other researchers who wanted to design experiments to help answer important questions.

In many published reports of experimental studies, we see conclusions such as "the observed difference in heart attack rates was **statistically significant**." This tells us that the differences in the response variable between those in different treatment groups cannot reasonably be explained by the chance involved in the random assignment of treatments to subjects. Recall what we said earlier: There are only two possible explanations for the observed differences in an experiment—that they were due to the chance involved in the random assignment or that the difference in treatments caused the difference in the response variable. Saying that the results of a particular experiment are *not* statistically significant means that we can't rule out the possibility that there is no difference in the effects of the treatments, and that the differences in response are simply due to the random assignment.

You may have noticed that in both the examples presented here, the subjects were *not* randomly selected from a larger population. This is usually the case with experiments. It often isn't practical to choose subjects at random from the population of interest. Consider how you would go about randomly selecting 24 pansy plants from the population of *all* pansy plants, for example. Or how researchers might randomly select 22,000 male physicians. As you learned earlier, the lack of random selection limits our ability to generalize results to the population of interest.

However, even if experimental units are not randomly selected, well-designed experiments can give convincing evidence that changes in one variable cause changes in another variable. Establishing causation is much more difficult with observational studies, because researchers cannot hold other variables constant and cannot assign individuals at random to treatment groups. As an example, consider early observational studies that suggested people who smoked were much more likely to get lung cancer than people who didn't smoke. Cigarette company executives argued that *confounding* was at work. They claimed that the kinds of people who smoked were also much more likely to engage in other unhealthy activities—such as drinking, overeating, and failing to exercise—than people who didn't smoke. It was these other unhealthy behaviors, they said, that led to increased risk of cancer, not smoking cigarettes. After many other observational studies showed the strong connection between smoking and lung cancer, and experiments on animal subjects demonstrated that smoking caused cancerous growths, cigarette company executives finally conceded.



Teacher Notes for Investigation #10: Do Diets Work?

In this investigation, students will review and critique two experiments designed to compare the effectiveness of low-carbohydrate and low-fat diets in reducing weight and cholesterol in obese adults.

Prerequisites

Students should be able to:

- Identify the subjects, factor(s)/explanatory variable(s), treatments, and response variable(s) in an experimental setting
- Distinguish an observational study from a survey or an experiment
- Explain the purpose of randomly assigning treatments to subjects in an experiment
- Determine whether an experiment was carried out in a single-blind or double-blind manner
- Explain the purpose of control in an experimental design
- Explain what is meant by "statistical significance"
- Identify a potential confounding variable in a study and explain how the variable could result in confounding
- Explain how the way in which data were produced affects our ability to generalize results to a larger population of interest

Learning Objectives

As a result of completing this investigation, students should be able to:

- Explain how the design principle of control applies in a specific experimental setting
- Interpret experimental results in context
- Explain what it means for a result to not be statistically significant in the context of an experiment
- Describe possible limitations of an experiment, such as side effects and dropouts
- Summarize and critique an experiment based on written information about the experiment

Teaching Tips

One of the primary goals of this first investigation in the Experiments section is to increase students' familiarity with and comfort in applying the terminology of experiments. Students may want to refer to the Overview as they complete the investigation.

Be sure to discuss how data ethics apply in these experimental settings: informed consent, anonymity and confidentiality, and external review board.

We recommend having students work through the questions in pairs initially. The questions are divided into four distinct groups. Questions 1 through 5 focus on the design of the two experiments. Questions 6 through 8 ask students to draw preliminary conclusions about low-carb versus low-fat diets based on the results of these two studies.

Questions 9 through 12 address some possible limitations of these experiments. Finally, Questions 13 and 14 ask students to refine their preliminary conclusions in light of the possible limitations.

To promote effective communication, you may want to have students discuss their responses with a partner prior to sharing answers with the class. You might also ask students to provide feedback on each other's answers in a whole class setting before you evaluate the accuracy and clarity of their responses.

Suggested Answers to Questions

1. The completed table is shown below.

	Duke University Study	**Philadelphia Study**
Subjects	120 volunteers, aged 18 to 65, with high cholesterol	132 obese adult volunteers
Factor(s)/ explanatory variable(s)	Type of diet followed	Type of diet followed
Treatments	Low-carb, high-protein diet Low-fat, low cholesterol diet	Low-carbohydrate diet Low-fat diet
Response variable(s)	Change in weight Change in cholesterol	Change in weight Change in cholesterol

2. In both the Duke University study and the Philadelphia study, researchers deliberately imposed treatments—either a low-carbohydrate diet or a low-fat diet—on the subjects. When something is deliberately done to individuals in a study to measure their responses, the study is an experiment.

3. Researchers assigned subjects at random to either a low-fat or low-carbohydrate diet. By letting chance divide the available subjects into two groups, the researchers were attempting to ensure the groups were roughly equivalent in terms of variables other than the specific diets assigned that might affect subjects' responses to the treatments. The researchers were also trying to avoid any bias that might have resulted from subjectively assigning subjects to treatment groups.

4. These experiments could have been conducted in a single-blind manner if the individuals who interacted with the subjects and measured the response variables did not know who was assigned to each of the diet treatments. As the subjects would know what kinds of foods they were eating, it would not have been possible to carry out either experiment in a double-blind fashion.

5. (a) If any of the subjects had dieted recently, their bodies might have responded differently to the diet regimens assigned in the Duke experiment than if they had not been dieting. Likewise, subjects who had used weight loss medications during the previous six months might have responded differently to the diet treatments assigned in the Duke study as a result of lingering effects of those medications. By using only subjects who had not dieted or used weight loss medications in the previous six months, researchers attempted to control for the effects of other variables that might have systematically affected subjects' responses to the diet treatments.

(b) Because exercise could affect subjects' weight loss and change in cholesterol level, it was important for researchers to try to ensure that all participants in the experiment engaged in similar amounts of exercise. Otherwise, any differences in weight loss or cholesterol level between the two groups of subjects could have been the result of differing exercise habits, rather than the specific diets assigned to those groups.

6. Both experiments suggest that following a low-carbohydrate diet caused a greater decrease in weight over a six-month period than following a low-fat diet. Likewise, both experiments suggest that following a low-carb diet caused a greater increase in HDL (good) cholesterol than following a low-fat diet. The Philadelphia experiment did not show a significant difference in weight loss for subjects on a low-carb diet when compared to those on a low-fat diet over a one-year period. So it is possible that a low-carb diet is more effective at reducing weight in the short-run than a low-fat diet, but that the two diet regimens result in similar amounts of weight loss over longer periods of time. One important caveat: These conclusions only apply to individuals like those who were willing to take part in these two experiments—somewhat motivated, otherwise healthy, obese adults.

7. This difference in average weight loss (2 kg) for subjects in the two groups was not large enough to rule out the possibility that the observed difference was simply due to the luck of the random assignment, and not to the effects of the two diet treatments.

8. Although the low-carb diet showed significant benefits in terms of weight loss and decrease in cholesterol over a six-month period, it also resulted in more minor side effects, such as constipation and headaches, than did the low-fat diet.

9. With such a high dropout rate in both experiments, our conclusions would be open to challenge. Researchers don't know what would have happened to the subjects who dropped out in terms of weight loss or change in cholesterol level. It is possible that the results of the experiment would have been different if all the subjects had participated for the full duration of the study. We have no way of knowing in what way the results might have differed.

What if most of the dropouts in the Philadelphia study had been from the low-carb diet group? Maybe those people withdrew from the study because they weren't experiencing a decrease in weight loss. If that was the case, then had those subjects remained in the experiment for the entire six months, researchers might not have observed a significant difference in weight loss for the two diet treatments. The fact that a much higher percentage of subjects in the low-fat diet group than of subjects in the low-carb diet group dropped out of the Duke University experiment is concerning.

Researchers should follow up with individuals who drop out of an experiment to find out why they made that decision.

10. If subjects did not follow their assigned diet treatments, then the results of the experiment are no longer as convincing. Researchers are drawing conclusions based on the belief that subjects are following their assigned diet plans. If some subjects deviate from the assigned diet regimen, researchers can no longer attribute any significant differences in weight loss or cholesterol level to the difference in the diet treatments.

11. As the daily nutritional supplement represents another systematic difference between the two groups of subjects (in addition to the diet plan they're following), researchers would need to rule out the possibility that differences in the response variables between the two groups could be due to the daily nutritional supplement and not the low-carb or low-fat diet.

12. In the Duke University study, a potential confounding variable is whether subjects took a daily nutritional supplement. To be potentially confounding, the variable must be associated with group membership and have an effect on the response variables. Since only the subjects in the low-carb diet group took the daily nutritional supplement, there is a clear association between this variable and group membership in the experiment.

As another example, consider the variable "amount of exercise." Amount of exercise could clearly affect weight loss or change in cholesterol level. In order for this to be a potential confounding variable, however, it would also have to be the case that subjects in one group tended to exercise more than subjects in the other group. As researchers randomly assigned subjects to the two diet treatments, the groups should have started out fairly balanced in terms of exercise habits.

13. No. The subjects who participated in both these experiments were recruited to do so. That is, they were willing volunteers. Perhaps these individuals were more motivated to begin with than the general population of obese adults. Also note that the subjects in both experiments were obese adults. Consequently, the results of the experiments apply only for otherwise healthy, obese adults, not to overweight adults in general. We can only generalize the findings of these two experiments to a population of individuals like the subjects who actually participated.

14. Answers will vary. Students should include the following points in their summaries:

Both experiments suggested a low-carb diet resulted in greater weight loss over a six-month period than did a low-fat diet.

The Philadelphia experiment found no significant difference in weight loss between the low-fat diet and the low-carb diet over a one-year period. The 2 kg difference in average weight loss that researchers observed could have been due to the random assignment of subjects to groups, and not due to the difference in diet regimens.

Both experiments suggested that a low-carb diet resulted in a significantly higher increase in LDL (good) cholesterol than a low-fat diet.

The high dropout rates in both experiments are concerning. We don't know how the results would have been affected if these subjects had completed the experiment.

In the Duke experiment, subjects in the low-carb group were given a daily nutritional supplement, but those in the low-fat group weren't. This is a potential source of confounding.

Researchers can only generalize the results of these experiments to the population of otherwise healthy, obese adults like the ones who agreed to participate in these studies.

Possible Extensions

You might want to have students find an article describing the results of another experiment on dieting and weight loss, and then have them perform an analysis similar to the one outlined in this investigation.

Investigation #10: Do Diets Work?

Corresponds to pp. 67-71 in Student Module

The Atkins Diet is one of many popular weight loss diets. It is based on reducing the consumption of carbohydrates. For years, such "low-carb" diets have been touted as being effective for weight loss and other health benefits. But before 2001, no one had attempted to demonstrate the effectiveness of a low-carb diet in a well-designed comparative experiment. Then, two separate groups of researchers attempted to do just that.

At Duke University Medical Center, Dr. William Yancy and his colleagues recruited 120 people between the ages of 18 and 65. All of the participants were obese and had high cholesterol, but were otherwise in generally good health. Researchers randomly assigned half of the participants to a low-carbohydrate, high-protein diet (similar to an Atkins Diet) and the other half to a low-fat, low-cholesterol diet. At the end of six months, researchers measured the change in each participant's weight and cholesterol levels.[1]

In the second study, Dr. Linda Stern and her colleagues recruited 132 obese adults at the Philadelphia Veterans Affairs Medical Center in Pennsylvania. Half of the participants were randomly assigned to a low-carbohydrate diet and the other half were assigned to a low-fat diet. Researchers measured each participant's change in weight and cholesterol level after six months and again after one year.[2]

1. Complete the following table using the details provided above about the two studies.

	Duke University Study	**Philadelphia Study**
Subjects		
Factor(s)/Explanatory Variable(s)		
Treatments		
Response Variable(s)		

---

1 "A Low-Carbohydrate, Ketogenic Diet versus a Low-Fat Diet To Treat Obesity and Hyperlipidemia," by Yancy, William S. et al, *Annals of Internal Medicine*, May 2004, 140(10) 769-777.

2 "The Effects of Low-Carbohydrate versus Conventional Weight Loss Diets in Severely Obese Adults: One-Year Follow-up of a Randomized Trial," by Stern, Linda et al, *Annals of Internal Medicine*, May 2004, 140(10) 778-785.

2. Explain why both of these studies are experiments, and not observational studies or surveys.

3. How did the researchers in both studies determine which subjects received which treatments? Why did they use the method they did?

4. Could these experiments have been carried out in a single-blind or double-blind manner? Justify your answer.

5. Each of the following quotations describes the subjects in the Duke University experiment. Explain how each is an example of control and why it is important in terms of the design of the study.

(a) "None had dieted or used weight loss medications in the previous six months."

(b) "All subjects were encouraged to exercise 30 minutes at least three times per week and had regular group meetings at an outpatient research clinic for six months."

Let's look at some results from the two studies.

> In the Duke University experiment, over the six-month duration of the study, weight loss was 12.9% of original body weight in the low-carbohydrate diet group and 6.7% of original body weight in the low-fat diet group. The low-carb diet group showed a greater increase in HDL (good) cholesterol than the low-fat diet group.
>
> In the Philadelphia experiment, subjects in the low-carbohydrate diet group lost significantly more weight than subjects in the low-fat diet group during the first six months of the study. At the end of a year, however, the average weight loss for subjects in the two groups was not significantly different. The low-carbohydrate diet group did show greater increase in HDL (good) cholesterol level after a year than the low-fat diet group.

6. Briefly summarize what the results of these two experiments seem to suggest about the relative effectiveness of low-carbohydrate diets and low-fat diets on weight and cholesterol.

7. In the Philadelphia experiment, the subjects in the low-carbohydrate diet group lost an average of 5.1 kg in a year. The subjects in the low-fat diet group lost an average of 3.1 kg. Explain how this information could be consistent with the statement above about the average weight loss in the two groups not being significantly different.

8. Here is an excerpt from a report about the Duke University experiment: "Participants in the low-carbohydrate diet group had more minor adverse effects, such as constipation and headaches, than did patients in the low-fat diet group." How would you modify your summary in question 6 based on this additional information?

When you look at experimental results, it's important to consider possible limitations of the study. The next few questions will help you look critically at the two experiments described earlier.

9. Explain how the following excerpts from a report about the two experiments might affect your conclusions about the effectiveness of low-carb versus low-fat diets:

Duke University study: "The study was completed by 76% of participants in the low-carbohydrate diet group and by 57% of participants in the low-fat diet group."

Philadelphia study: "Study limitations include high dropout rate of 34% …"

10. In both experiments, participants were assigned at random to a low-fat or low-carbohydrate diet group. What exactly does that mean? The subjects in the low-fat diet group attended counseling sessions about how to restrict their caloric intake from fat. The subjects in the low-carbohydrate group attended counseling sessions about how to restrict their carbohydrate intake. These counseling sessions continued on a weekly or monthly basis throughout the experiment. It is possible that some people in each group did not restrict their diets as instructed. How might this affect conclusions based on the experiment?

11. In the Duke University study, subjects in the low-carbohydrate group all received daily nutritional supplements. Subjects in the low-fat group did not. How might this affect conclusions based on the experiment?

12. Give an example of a potential confounding variable in one of the two experiments. Explain carefully how the factor you choose could result in confounding.

13. Is it reasonable to generalize the results of these two experiments to the population of all overweight adults? Justify your answer.

14. Now that you have considered possible limitations of these two experiments, summarize what the results of these two experiments seem to suggest about the relative effectiveness of low-carbohydrate diets and low-fat diets on weight and cholesterol. You may want to refer to what you wrote earlier in response to question 6.

Teacher Notes for Investigation #11: Distracted Learning

In this investigation, students will design, carry out, and analyze results from an experiment to determine whether listening to Mozart while performing a memorization task helps students remember better than doing a similar task with no music playing.

Prerequisites

Students should be able to:

- Explain how the way in which data were produced affects our ability to generalize results to a larger population of interest
- Identify a potential confounding variable in a study and explain how the variable could result in confounding
- Explain the purpose of randomly assigning treatments to subjects in an experiment
- Carry out the random assignment of treatments to subjects in an experiment
- Identify the subjects, factor(s)/explanatory variable(s), treatments, and response variable(s) in an experimental setting
- Construct and interpret a comparative dotplot for a quantitative variable, describing shape, center, spread, and any unusual values
- Construct and interpret a dotplot of differences for paired data, describing shape, center, spread, and any unusual values
- Choose the most appropriate numerical measures of center and spread to use in a given setting (mean and standard deviation OR median and interquartile range [IQR])
- Determine whether an experiment was carried out in a single-blind or double-blind manner

Learning Objectives

As a result of completing this investigation, students should be able to:

- Consider alternative designs for an experiment, and then choose the best one for answering a given research question
- Explain why it is important for the order of treatments to be randomly assigned to subjects in a design that requires each subject to receive both treatments
- Draw appropriate conclusions from an experiment involving paired data from volunteer subjects
- Make at least one suggestion for improving the design of an experiment based on the actual experience of carrying out that experiment

Teaching Tips

Questions 1 through 4 of this investigation walk students through the process of designing an experiment to test whether listening to Mozart improves memorization skills for students in their class. Students are steered away from the design used in the two experiments of the previous investigation, in which subjects were randomly assigned into two roughly equal treatment groups. This type of design is known as a *completely* *randomized design*. Instead, students are nudged toward using a *matched pairs design*, in which each subject receives both treatments in a random order.

Why is a matched pairs design preferable in this case? We know that individuals vary widely in their memorization abilities. If we used a completely randomized design, with about half the students in the class assigned to the Mozart treatment and the other half assigned to work in silence, we would expect considerable variation in the individual scores on the memorization task within each group. If we observe a difference in the mean scores for the two groups, we would like to know whether that difference was caused by listening to Mozart. Of course, there is another possible explanation for any difference that emerges. Maybe subjects would perform the same whether they listened to Mozart or not, so the observed difference is simply a result of which subjects were randomly assigned to each group. With lots of variation present, it will be more difficult to rule out this second possible explanation in favor of a causal connection between listening to Mozart and memorization.

By using a matched pairs design, we isolate the variation among individuals by comparing each individual's performance on two similar memorization tasks—one while listening to Mozart and one done in silence. We perform our analysis on the difference in memorization scores for the students in the class. There should be less variation in the difference values than there would have been with data produced using a completely randomized design. As a result, it should be easier to detect a "Mozart effect" if there is one by ruling out chance variation from the random assignment as a plausible explanation.

Question 5 asks students to review the details of their design before implementing it. In Question 6, students carry out the random assignment for their design. In Question 7, students actually perform the experiment. Here are two memorization tasks that students can use:

Task A: 12 09 96 62 66 52 26 82 25 18 98 31 06 48 47 72 28 67 85 57

Task B: 38 07 18 85 73 90 31 12 37 39 87 33 06 44 43 34 08 27 24 99

Questions 8 through 16 take students through the process of analyzing data, identifying possible limitations, and drawing conclusions. If students use a matched pairs design for their experiment, it would be inappropriate for them to analyze the "with Mozart" and "in silence" data as if they came from two unrelated groups of individuals, as Question 8 seems to suggest. Make the point that **the appropriate method of data analysis is determined by the design of the study**. If data are paired by design, then students should analyze the pairs of data values. In this case, that means examining the differences in performance scores for the subjects.

Suggested Answers to Questions

1. Since the subjects are available volunteers, and are not randomly selected from a larger population of interest, we will only be able to generalize our findings to the population of students that are similar to the people in this class.

2. With this design, the two groups of subjects would be performing the experiment in two different locations. It is possible that students will perform differently on the task as a result of the conditions in the two rooms. If so, then "room conditions" would be a confounding variable. The process of relocating to another room may affect the subject's performance on the task in a systematic way. Perhaps the movement will stimulate these students' brains, resulting in better performance on the memorization task than for those students who stay put. Because individuals vary widely in their ability to memorize, it might be better to have each subject perform a similar memorization task twice—once while listening to Mozart and once in silence—so that individual differences in memorization are planned for, rather than distributed between, the two groups with random assignment. After all, the random assignment could lead to two groups with large amounts of variability in their memorization skills, which would make it more difficult to detect any effect of listening to Mozart on memorization.

3. (a) By separating the "good" and "not-so-good" memorizers in advance based on performance on the initial memory task, we would expect less variability in memorization abilities for the randomly assigned groups of subjects in each performance category than for the two randomly assigned groups in the design proposed in question 2. With less variability present, it should be easier to detect any effect of listening to Mozart on memorization for "good" memorizers and for "not-so-good" memorizers.

(b) Have each subject perform two similar memorization tasks, one while listening to Mozart and one in silence. Randomly assign the subjects into two approximately equal groups. Have one group do the first task while listening to Mozart and the second task in silence. Have the other group do the first task in silence and the second task while listening to Mozart.

4. (a) Even if the two memorization tasks are similar, subjects may still find one task more difficult than the other. Suppose the subjects find Task A easier than Task B. If subjects perform better while listening to Mozart, it might be because they are doing the easier task, and not because of the music. In other words, which task subjects perform would be a potential confounding variable.

(b) Students may learn from doing the first memorization task, and perform better on the second memorization task as a result. This is known as a *learning effect*. In this scenario, if students performed better while listening to Mozart, we wouldn't know whether this was due to a learning effect or due to the effects of the music.

(c) Students should design a method of random assignment in which about equal numbers of students will perform the experiment under each of the following four conditions:

Task A with Mozart, then Task B in silence

Task A in silence, then Task B with Mozart

Task B with Mozart, then Task A in silence

Task B in silence, then Task A with Mozart

(d) Answers will vary, depending on the random assignment plan that was agreed upon in (c). One method would be to have students write their names on roughly identical slips of paper, put the slips in a hat, and mix them thoroughly. Then, you could draw out names one at a time, with the first person assigned to the first set of experimental conditions from (c), the second person to the second set of experimental conditions from (c), and so on. Of course, students could use a variation of the hat method by assigning distinct numbers to the members of the class, and then using a random digit table or random number generator to mimic the process described in the previous sentence.

Students could opt to roll a four-sided die (or a six-sided die, ignoring two of the numbers) for each member of the class to determine which of the four experimental conditions from (c) that person would follow. Note that this method could result in somewhat unequal numbers of students following each of the four experimental conditions just by chance.

5. (a) The students in this class.

(b) The explanatory variable is what a person listens to while performing a memorization task.

(c) Treatments are connected with values of the explanatory variable. In this case, the two possible values of the explanatory variable are “listen to Mozart” and “work in silence.” The two treatment combinations for our experiment are (1) listen to Mozart during the first task; work in silence during the second task, and (2) work in silence during the first task, then listen to Mozart during the second task. We’re going to measures students’ performances on the tasks as part of the experiment. However, the tasks themselves are not treatments, because we are not deliberately imposing the tasks on the students to measure their responses to those tasks.

(d) A scoring system such as one point for each number remembered correctly, and minus one point for each incorrect number that is listed, might be a good way to measure performance and avoid haphazard guessing.

(e) The response variable is the difference in score on the memorization tasks while listening to Mozart and while working in silence.

6. Answers will vary.

7. Data will vary!

8. Comparative dotplots will vary. Note that the horizontal axis in the plot represents the score on the memorization task, which could be positive, negative, or zero based on the scoring system that was suggested in 5(d). When describing similarities and differences, students should discuss issues of shape, center, spread/variability, and any unusual values.

9. Difference values will vary.

10. Dotplots will vary. Note that the horizontal axis in the plot represents the difference in score on the two memorization tasks for each student. Since students are testing the belief that Mozart might help improve memorization, they might want to define difference = score with Mozart – score in silence. When interpreting the plot, students should discuss issues of shape, center, spread/variability, and any unusual values in the context of this experiment.

11. The dotplot in question 10 shows the difference in score for each student when listening to Mozart versus when performing the memory task in silence. The dotplot in question 8 treated the two scores for each student as unrelated values, simply showing all students' memorization scores with Mozart and all students' memorization scores without Mozart. Because the two scores for each student are related (by virtue of being produced by the same individual), it is more appropriate to focus on the difference in scores when making a graphical display of the data. The plot in question 10 makes it easier to see whether listening to Mozart helped increase memorization performance for students in the class, which was the goal of the experiment.

12. Answers will vary. Students could use the mean and standard deviation to summarize center and spread, respectively, if the distribution of differences is roughly symmetric and there are no potential outliers. If the distribution is clearly skewed, or potential outliers are present, then the median and interquartile range (IQR) would be more appropriate summaries of center and spread.

13. This experiment was neither single-blind nor double-blind. Both the subjects and the individuals measuring the response variable (memorization score) knew which treatment combination the students were receiving.

14. Answers will vary. Students should be evaluated on how well they use the evidence from their graphs and numerical summaries to support their answer.

15. No. We can only generalize our findings about listening to Mozart to memorization tasks that are similar to the ones used in this experiment.

16. Having each student listen to the same Mozart selection was a form of control. It is possible that students would respond differently to other Mozart pieces or other kinds of music when performing similar memorization tasks. Consequently, we can't generalize the results of this study to all Mozart tunes or other types of music.

Possible Extensions

There are plenty of possible variations on this experiment that students could design and carry out. For instance, the original claim of researchers who discovered the so called "Mozart effect" was that listening to Mozart helps improve performance on spatial reasoning tasks. Students could use mazes as the task, rather than lists of numbers to memorize.

Investigation #11: Distracted Learning

Corresponds to pp. 72-79
in Student Module

While you study, do you watch TV, listen to music, check your MySpace page, surf the Internet, chat on e-mail, talk or text on your cell phone? Do your parents insist that you can't possibly concentrate on studying while you're distracted by one of these activities? Maybe the conversation goes something like this:

> **Parent:** "Take off your headphones and do your homework!"
>
> **Student:** "I am doing my homework, and I work better with my music on."
>
> **Parent:** "Turn it off! You can't study with that distraction!"
>
> **Student:** "Yes I can. It helps me relax."
>
> **Parent:** "Turn off that racket and concentrate on your school work!"
>
> **Student:** "I study better with it on!"

Who is right? Some say that any distraction might interfere with your focus on the work you're doing, which may in turn affect the quality of the finished product. But others argue that listening to music actually helps them concentrate because the music "drowns out" other potential distractions. What do you think? Can previous research help us sort this out?[1]

In 1993, Frances Raucher and his colleagues designed an experiment to test whether listening to Mozart would help students improve their performance on a spatial reasoning task. They recruited 36 college students to participate in the experiment. The subjects were randomly assigned to three groups, with 12 students per group. Subjects in Group 1 listened to a 10-minute selection from a Mozart piece. Group 2 listened to a relaxation tape for 10 minutes. Subjects in Group 3 sat in silence for 10 minutes. Each subject took a pretest on spatial reasoning two days before the experiment and a post-test on spatial reasoning immediately after the 10-minute treatment. The results of the experiment seemed surprising: Students who listened to Mozart showed significantly higher gains in their scores on spatial-reasoning tasks than students in the other two groups.

After hearing the results of Rauscher's experiment, some eager parents started playing Mozart tapes for their children in hopes of increasing their spatial reasoning skills. One state even passed legislation requiring preschools to play 30 minutes of classical music a day. Other researchers tried to confirm this so-called "Mozart effect" in experiments of their own, but with little success.

So the question remains: Does listening to music help or hinder students' learning? The answer may depend on what type of "learning" we mean. In this investigation, your class will design and carry out an experiment to test whether listening to music helps or

---

1 *www.madsci.org/posts/archives/mar98/889467626.Ns.r.html* served as inspiration for part of this investigation.

hinders students as they perform a memorization task. Then, you will analyze data from the experiment and draw some preliminary conclusions from your research.

1. For simplicity, the members of your class will serve as the subjects in your experiment. How might this affect your ability to generalize the results of your study?

2. One possible design for the experiment would be to randomly assign about half of the students in your class to perform the memorization task while listening to Mozart, and the other half to perform the task in a silent room nearby. Then, you could compare the scores of students who listened to Mozart while memorizing with the scores of students who didn't. What flaw(s) do you see in using this design to conduct the experiment?

3. Some people are better at memorizing things than others. Here's another possible design for your experiment that takes this fact into account. Begin by having each student perform a memory task. Based on students' performance on this task, split the class into two roughly equal-sized groups containing the "good memorizers" and the "not-so-good memorizers." Randomly assign about half of the good memorizers to perform a second memory task while listening to Mozart, and the other half to perform the task in a silent room nearby. Use the same random assignment strategy for the not-so-good memorizers. To analyze the data from the experiment, you would compare the change in scores from the first memory task to the second for the good memorizers who listened to Mozart and those who didn't, and separately for the not-so-good memorizers who did and didn't listen to Mozart while memorizing.

(a) In what ways does this design improve on the design from question 2?

(b) How might you further improve the design of this experiment using the idea that some people are better memorizers than others? Explain.

4. Perhaps the best way to take individual differences in memorization skills into account in this experiment is to have each person perform two memory tasks—one while listening to Mozart and one in silence. Then, you can analyze data on the difference in performance for all students in your class and determine whether listening to Mozart seems to help or hurt memorization.

To carry out the experiment in this way, you will need two different but similar memory tasks. Let's call them task A and task B.

(a) Explain why you should not have all students perform task A while listening to Mozart and task B while in a silent room.

(b) Explain why you should not have all students perform their first memory task while sitting in a silent room and their second memory task while listening to Mozart, or vice versa.

(c) Discuss with your classmates how you could use random assignment to most effectively address the issues raised in parts (a) and (b). Once you have settled on a plan, propose it to your teacher.

(d) Describe carefully how you will perform the random assignment required by your approved plan from part (c).

5. Now that we have settled on a design for the experiment, let's confirm some of the details.

(a) Who are the subjects in this experiment?

(b) What factor(s)/explanatory variable(s) is this experiment investigating?

(c) What treatments are being administered? Explain why task A and task B are not treatments.

(d) Let's take a look at the tasks. Each subject will be presented with a list of 20 randomly generated two-digit numbers, such as the list shown below. The student will then have one minute to memorize as many of the numbers in the list as possible. At the end of the minute, each student will have two minutes to write down as many of the numbers as he or she can remember.

26	86	64	65	75	11	49	47	85	19
23	57	97	00	62	43	66	94	79	50

A wily student might just write down a bunch of two-digit numbers during the two minute period, hoping to match as many as possible. How might you score performance on this task to reward students for actual memorization and not for guessing?

(e) Based on your answer to (d), describe the response variable(s) this experiment will measure.

Now it's time to do the experiment! Your teacher will assist with logistics so that all students can participate.

6. Carry out the random assignment required for your experiment from question 4(d). Indicate clearly what each student will be doing first and second. You may find it helpful to make a chart like the one below that summarizes how the experiment will be carried out.

Subject	First Task	First Treatment	Second Task	Second Treatment
1	A	Music	B	Silence
2	A	Silence	B	Music
3	B	Music	A	Silence
4	B	Silence	A	Music

Subject	Which Task First? (A or B)	Music First? (Yes or No)	Score With Music	Score Without Music	Difference

7. Have students perform the two memorization tasks as specified in question 6. Record data from the experiment in the table on the previous page.

8. Construct comparative dotplots or boxplots of the scores with music and the scores without music. Describe any similarities and differences you see in a few sentences.

9. Calculate the difference in scores for each student when listening to Mozart versus sitting in a silent room. As a class, decide on which order you will subtract the values. Record these values in the right-most column of the table on the previous page.

10. Construct an appropriate graph of the difference in memorization scores. Describe what the graph tells you in a couple of sentences.

11. In what way is the graph you constructed for question 10 more informative than the comparative graph from question 8?

12. Calculate a measure of center (mean or median) and a measure of spread that you think summarize the differences well. Explain why you chose the measures you did.

13. Was this experiment single-blind, double-blind, or neither? Justify your answer.

14. Based on the results of your experiment, does it appear that listening to Mozart helps or hinders students' performance on memorization tasks? Give appropriate graphical and numerical evidence to support your answer.

15. Can we generalize the results of this experiment to any kind of task that requires memorization? Justify your answer.

16. Why did we have all students listen to the same piece of Mozart music, rather than letting each student choose music he or she liked? Explain.

Teacher Notes for Investigation #12: Would You Drink Blue Soda?

In this culminating investigation for the Experiments Section of the module, students will design, carry out, and analyze data from an experiment to test whether people have a preference for blue-colored soda. By this point, students should feel fairly comfortable with the terminology and basic concepts of experimental design. If students completed the previous investigation using a matched pairs design, then they should need little prodding to come up with a similar design for this taste test experiment.

Prerequisites

Students should be able to:

- Define a research question
- Explain why it is important for the order of treatments to be randomly assigned to subjects in a design that requires each subject to receive both treatments
- Carry out the random assignment of treatments to subjects in an experiment
- Identify the subjects, factor(s)/explanatory variable(s), treatments, and response variable(s) in an experimental setting
- Determine whether an experiment can be carried out in a single-blind or double-blind manner
- Explain how the way in which data were produced affects our ability to generalize results to a larger population of interest
- Consider alternative designs for an experiment, and then choose the best one for answering a given research question
- Use appropriate graphical and numerical techniques for describing the distribution of a categorical variable and for describing the relationship between two categorical variables

Learning Objective

As a result of completing this investigation, students should be able to carry out a complete analysis of an experiment involving one or more categorical variables using counts, percents, and bar graphs to support their narrative conclusions.

Teaching Tips

We have designed this investigation so that students can formulate a plan for their experiment with little or no prompting, using only the first page of the student investigation to get started. Questions 1 through 7 then ask students to review their proposed design in light of several important issues before finalizing their experimental design in Question 8.

Obtain permission from your administration before allowing students to conduct the experiment. You may be required to get parental consent before students can participate in the experiment.

You will need to provide clear instructions to your students about obtaining informed consent, preserving anonymity and confidentiality, and ensuring subject's health and safety.

Next, students carry out their beverage preference experiment. Using the data they have collected, students are asked to perform an analysis and draw conclusions about students' preferences in Questions 9 through 11. Note that Question 10 focuses on the issue of whether order of presentation seems to have affected student preference, while Question 11 addresses the original research question.

Finally, students are asked to write a report about teenagers' preference for blue-colored beverages based on the results of this experiment. This question gives students a final opportunity to showcase their ability to analyze results from an experiment.

Suggested Answers to Questions

1. People may have a tendency to prefer the beverage they taste first (or last), regardless of the actual qualities of the beverages themselves (color, taste, etc.). That is, it is possible that the order in which people taste the beverages might affect their stated preference. If so, then you wouldn't want to present the same beverage first (or last) to more than about half of the subjects. Randomizing the order should help ensure that about half of the subjects taste one beverage first and about half taste the other beverage first. Then, any sizable differences that emerge in terms of preference for one beverage over the other should not be due to the order in which the beverages were presented.

2. One way to determine the order would be to flip a coin for each subject. If the coin shows "heads," then the subject would drink the clear beverage followed by the blue beverage. If the coin shows "tails," then the subject would drink the blue beverage followed by the clear beverage. Note that this method of randomly assigning the order could result in unequal numbers of subjects drinking the beverages in the two possible orders. An alternative method would be to put subjects' names on roughly identical slips of paper, drop them in a hat, and mix them up. Draw one slip at a time without looking. The person whose name is drawn first will drink clear then blue; the person whose name is drawn second would drink blue then clear, and so forth. Of course, you could use a modified version of the hat method by giving each subject a distinct numeric label, and then using a random number table or random number generator to select individuals one at a time. As before, the person whose name is drawn first will drink clear then blue; the person whose name is drawn second would drink blue then clear, and so forth.

3. The specific treatments in this experiment are "clear then blue" and "blue then clear."

4. One of the aims of the experiment is to see how color affects subjects' perceived preferences for a beverage. To study this, you must allow the subjects to see whether the beverage they are tasting is clear or blue. Hence, the subjects cannot be blind.

5. Answers will vary. Students should use some form of random selection to choose subjects to participate in the experiment. A true random sample of students may not be practical, but there's no need to go to the opposite extreme and use volunteers, either.

6. Answers will vary. If students use random selection to choose the subjects for their experiment, it should be reasonable to generalize the results to the larger population from which the subjects were selected. That's the benefit of random selection!

7. Answers will vary. One possible question is: "Of the two beverages that you tasted, which did you prefer?"

8. Answers will vary. Students' plans should include:

Research question, clearly stated

Subjects: how many; how they will be selected

Explanatory variable and treatments

How subjects will be assigned to treatment combinations

Response variable: what will be measured and how

9. Answers will vary. Students should construct a well-labeled, comparative bar graph to display the categorical variable of drink preference for the two experimental groups. If the number of subjects in the two groups differs, then students should use percents rather than counts to compare subjects' drink preferences.

10. Answers will vary. Students should be evaluated based on the strength and clarity of the graphical evidence they provide about whether preference differs based on order of tasting.

11. Answers will vary. Students should be evaluated based on the strength and clarity of the graphical and numerical evidence they provide about whether students clearly prefer either blue or clear soda.

12. Answers will vary. Students should be evaluated based on the strength and clarity of the graphical and numerical evidence they provide in support of their recommendations.

Possible Extensions

Can people distinguish bottled water from tap water? Coke from Pepsi? Students could design and carry out a taste test experiment to help answer questions such as these.

After completing Section IV of the module, students could use simulation to test for a significant difference in preference to reinforce some of the ideas associated with statistical inference.

Investigation #12: Would You Drink Blue Soda?

Corresponds to pp. 80-84 in Student Module

Does what you see affect your perception of how it tastes? If color can influence how people think a food tastes, what implications does this have for companies that make and market food and beverages?[1]

PepsiCo might be interested in your answer to these questions, as they have had two marketing failures based on introducing nontraditional colored beverages. In the early 1990s, PepsiCo introduced Pepsi Clear, a cola-flavored drink that was clear instead of brown in color. Pepsi Clear was later discontinued because sales were low. In 2002, PepsiCo tried again with Pepsi Blue.[2] Pepsi Blue was a berry-flavored cola drink that was blue in color. The Pepsi web site (*www.pepsi.com*) says that Pepsi Blue was "created by and for teens. Through nine months of research and development, Pepsi asked young consumers what they want most in a new cola. Their response: Make it berry and make it blue."

Unfortunately for PepsiCo, Pepsi Blue, like Pepsi Clear, was not a successful product, and it was discontinued a few years later. So what happened? Was the mistake adding a berry flavoring to cola, making the cola blue, or a combination of both?

In this investigation, you'll investigate whether teens have a preference for or a dislike for blue-colored soda.

Getting Started

To decide whether coloring a soda blue is a good or bad strategy if the drink is going to be marketed to teenagers, you will design and conduct an experiment, collect and analyze the data, and then make a recommendation.

For this experiment, you can start with a clear-colored soda, such as 7-Up or Sprite. Experiment with adding blue food coloring to the soda to create a "recipe" for a blue version of the soda. Food coloring is tasteless, so the addition of food coloring will not change the actual taste of the soda.

Once you have developed your new product, think carefully about how you would design an experiment to determine if teens have a preference for the clear soda or the blue soda.

Note: Be sure to discuss the ethical considerations involved in performing an experiment with human subjects. Your teacher will require you to obtain informed consent from all students (and possibly their parents) before they can participate in your experiment.

Once you have a plan in mind, answer the following questions. Be as specific as possible in your answers. It is OK to modify the design of your experiment if any of these

---

1 The page titled "Does the Color of Foods and Drinks Affect the Sense of Taste?" on the Neuroscience for Kids web site, *http://faculty.washington.edu/chudler/coltaste.html*, has a list of references to studies that have examined how color affects perceived taste.

2 You can find an announcement describing the launch of Pepsi Blue at *http://money.cnn.com/2002/05/07/news/companies/pepsi*.

questions reveal a weakness in your original plan. Now is the time to revise, before you actually carry out the experiment and collect the data!

1. In taste test experiments like the one you are designing, it is usual to randomize the order in which subjects taste the two drinks. That is, some subjects should taste the clear drink first and then the blue drink, while others should taste the blue drink first and then the clear. A random mechanism would be used to determine the order for each subject. Why do you think it is important to randomize the order in which the drinks are presented in an experiment of this type?

2. What would be a good way to determine the order (clear then blue or blue then clear) for each subject?

3. What are the two treatments for this experiment? *Hint:* In an experiment, subjects are assigned at random to one of the treatments.

4. Explain why it is not possible in this experiment to "blind" the subjects with respect to which experimental group they are in.

5. How will you select the subjects for your experiment, and how many subjects will participate? Be specific!

6. To what group, if any, will you be able to generalize the results of your experiment? Explain why you think it is reasonable to generalize to this particular group.

7. What question will you ask each subject after he or she has tasted the two sodas? Make sure that you will be able to determine from the response which of the two drinks was preferred.

8. After considering your answers to questions 1 through 7 and modifying your plan as needed, write a summary of your plan for conducting the experiment on *separate paper*. Include enough detail that someone who has not been part of your design team could read the summary and be able to carry out the experiment as you intended. Be sure to address ethical issues of using human subjects.

After your teacher has approved your experimental plan, carry out the experiment and collect data. Be sure to record the order in which the two drinks were tasted and the response for each subject.

Once you have collected the data, use it to fill in the four cells of the table below.

		Order	
		Clear then Blue	Blue then Clear
Preference	Clear		
	Blue		

9. Construct a graphical display that allows you to compare the preferences for the two experimental groups (clear then blue and blue then clear).

10. Based on your display, do you think there is a difference in preference for the two experimental groups? That is, do you think the order in which the drinks were tasted makes a difference? Explain.

11. Based on the data from this experiment, do you think there is a preference for one of the drinks (clear or blue) over the other? Explain, justifying your answer using the data from the experiment.

12. Write a report that makes recommendations to a soft drink company that is considering introducing a blue soft drink that will be marketed to teens. Include appropriate data and graphs to support your recommendations.