
Chapter 8 Data Sources: Observational Studies and Surveys



8.1 Observational Studies and Found Data

8.2 Sample Surveys

8.3 Populations and Parameters

8.4 Common Sampling Designs

8.5 The Valid Survey

8.6 How to Sample Badly

Roper Polls

Public opinion polls are a relatively new phenomenon. In 1948, as a result of telephone surveys of likely voters, all of the major organizations—Gallup, Roper, and Crossley—consistently predicted, throughout the summer and into the fall, that Thomas Dewey would defeat Harry Truman in the November presidential election. By October the results seemed so clear that *Fortune* magazine declared, “Due to the overwhelming evidence, *Fortune* and Mr. Roper plan no further detailed reports on change of opinion in the forthcoming presidential campaign. . . .”

Of course, Harry Truman went on to win the 1948 election, and the picture of Truman in the early morning after the election holding up the *Chicago Tribune* (printed the night before), with its headline declaring Dewey the winner, has become legend.

The public's faith in opinion polls plummeted after the election, but Elmo Roper vigorously defended the pollsters. Roper was a principal and founder of one of the first market research firms, Cherington, Wood, and Roper, and director of the *Fortune Survey*, which was the first national poll to use scientific sampling techniques. He argued that rather than abandoning polling, business leaders should learn what had gone wrong in the 1948 polls so that market research could be improved. His frank admission of the mistakes made in those polls helped to restore confidence in polling as a business tool.

For the rest of his career, Roper split his efforts between two projects, commercial polling and public opinion. He established the Roper Center for Public Opinion Research at Williams College as a place to house public opinion archives, convincing fellow polling leaders Gallup and Crossley to participate as well. Now located at Cornell University, the Roper Center is one of the world's leading archives of social science data. Roper's market research efforts started as Roper Research Associates and later became the Roper Organization, which was acquired in 2005 by GfK. Founded in Germany in 1934 as the Gesellschaft für Konsumforschung (literally, "Society for Consumption Research"), GfK now stands for "growth from knowledge." It is the fourth largest international market research organization, with over 130 companies in 70 countries and more than 7700 employees worldwide.

GfK Roper Consulting conducts a yearly, global study to examine cultural, economic, and social information that may be crucial to companies doing business worldwide. These companies use the information provided by GfK Roper to help make marketing and advertising decisions in different markets around the world.

How do the researchers at GfK Roper know that the responses they get reflect the real attitudes of consumers? After all, they don't ask everyone, but they don't want to limit their conclusions to just the people they surveyed. Generalizing from the data at hand to the world at large is something that market researchers, investors, and pollsters do every day.

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

8.1 Observational Studies and Found Data

Data are everywhere. IBM estimates that 90% of all data in the world were generated in the past two years. We currently generate about 2.5 billion terabytes (TB) a day (a TB is 1000 gigabytes). Most of these data are simply recording an event: a text message (about 15.2 million per minute), an Uber trip (about 46,000/min), a tweet (456,000/min), an Instagram post (47,000/min), or a Google search (3.6 million/min).¹

¹www.iflscience.com/technology/how-much-data-does-the-world-generate-every-minute/

We've discussed visualizing, analyzing, and modeling data, but we haven't talked about where data come from. And that matters. The conclusions you can draw from data and the soundness of your business decisions are limited by how your data were collected and their quality. You should ask about your data's "pedigree" as part of your analysis. Start with the five W's, especially how, where, and why the data were collected. And be skeptical of the suggestion that large amounts of data substitute for knowing their pedigree. A large amount of unreliable or biased data is just as unreliable as a smaller amount.

Can you trust your data? Companies and government agencies collect data in many different ways. Recording customer transactions or "loyalty cards" enables a company to record every purchase. A company might study such transactional data to identify associations between customer behavior and demographic information about the customer. For example, they might notice that customers with pets tend to spend more on home cleaning supplies. But, because these data have no structure, it would be dangerous to conclude that owning a pet *causes* increased spending.

What's wrong with concluding that owning a pet causes purchases of cleaning supplies? That claim depends on there being no other differences between the groups that could account for the differences in purchasing behavior. But there are lots of variables that might cause differences. Perhaps more affluent families both have pets and spend more freely. Or maybe families with children both spend more on cleaning supplies and tend to have pets.

When it is practical, a somewhat better approach is to observe individuals, recording the variables of interest, and seeing how things turn out. This kind of study, in which individuals are observed and outcomes are measured, is called an **observational study** ⓘ.

For example, if a home improvement store thought pet ownership might be a way to identify profitable customers, it might start by selecting a random sample of new customers and ask whether they have a pet. Analysts could then track their spending on home improvement and compare those who own pets to those who don't.

A study like this, in which individuals are followed to observe future outcomes is called a **prospective** ⓘ study. One that identified pet owners and reviewed previous purchasing behavior would be called a **retrospective** ⓘ study.

Observational studies are used widely in public health and marketing. Those that study rare outcomes, such as specific diseases, are often retrospective. They first identify people with the disease and then look into

their history and heritage in search of things that may be related to their condition. But retrospective studies have a restricted view of the world because they are usually limited to a small part of the entire population. And because retrospective records are based on historical data and memories, they can have errors. (Do you recall *exactly* what you ate even yesterday? How about last Wednesday?)

Retrospective Studies Can Give Valuable Clues

For rare illnesses, it's not practical to draw a large enough sample to see many ill respondents, so the only option remaining is to develop retrospective data. For example, researchers can interview those who have become ill. The likely causes of both Legionnaires' disease and HIV were identified from such retrospective studies of the small populations who were first infected. But to confirm the causes, researchers needed laboratory-based experiments.

Although an observational study may identify important variables related to an outcome of interest, there is no guarantee that we have found the right or the most important related variables. Perhaps people who own a home with a yard are more likely to have a pet and spend more on home improvement. It may be this fact—whether we know it or not—that drives spending on home improvement.

Observational studies, with appropriate caution and under certain circumstances, can be used to demonstrate causal relationships. However, getting to a causal explanation from observational data is extremely difficult under the best of conditions—and impossible in many cases. There is an entire field called causal inference that attempts to infer causes from observational studies, through careful reasoning and extensive case matching. These methods are extremely technical and they lie beyond the methods of this book. When it is important to identify factors that cause an outcome, the most effective approach is a designed experiment. A **designed experiment** differs from observational studies because the factors are actively manipulated with the intention of affecting the outcomes. Experimental designs are the subject of the next chapter.

In Practice 8.1 Gaining insights from observational studies

Amtrak launched its high-speed train, the Acela, in December 2000. It is the only high-speed line in the United States and it is by far the most profitable of all Amtrak lines. The Acela line generates about one quarter of Amtrak's entire revenue.² It is often used by business professionals because of its fast travel times, business class seats, and free Wi-Fi. The manager you work for at Amtrak wants to boost Acela's ridership among Millennials. You examine a sample of last year's customers for whom you have demographic information and find that only 5% of last year's riders were 21 years old or younger, but of those, 90% used the Internet while on board as opposed to 37% of riders older than 21 years.

2reasonrail.blogspot.fr/2014/11/amtrak-routes-by-2014-cost-recovery.html

Manager That's great. So, if we increase the quality and speed of the Wi-Fi service on board, will we attract more young riders?

Analyst Not necessarily. This is only a retrospective observational study. Although I compared the rates of Internet use between those older and younger than 21 years, I can't come to any conclusions about why they chose to ride the Acela. We certainly can't conclude that changing our Wi-Fi quality will boost their ridership based on these data. On the other hand, we could possibly design an experiment to try to determine the effect of Wi-Fi quality on usership by millennials.

Big Data and Data on the Web

In recent years, it has become common to work with large amounts of routinely collected transactional data or data available on the Internet—commonly called “big data.” And such data are increasingly being “mined” to make business and policy decisions. But even when the data are of high quality and record what they claim to record, it can be hazardous to use such data to infer conclusions. “Found” data have not been collected with the aim of learning anything specific. It can be dangerous to use them as if they were collected with such a purpose. One hazard is that variables that should have been measured to obtain an accurate view of patterns and relationships, and that would therefore have been included in a designed study, might not be available in found data. Another is that when there is a large amount of data, the sheer size of the dataset can make conclusions appear to be convincing. You should be very cautious with such data.

8.2 Sample Surveys

Data collected to understand a larger population or estimate some attributes of the population are often found by a **sample survey** ⓘ. A **sample survey** ⓘ asks questions of potential respondents and records their responses in an attempt to understand the larger population.

Three Ideas of Sampling

Idea 1: Sample—Examine a Part of the Whole

We'd like to know about an entire collection of individuals, called a **population** ⓘ, but examining all of them is usually impractical, if not impossible. So we settle for examining a smaller group of individuals—a **sample** ⓘ—selected from the population. For the Roper researchers the population of interest is the entire world, but it's not practical, cost-effective, or feasible to survey everyone. So they examine a sample selected from the population.

We take samples all the time. For example, if a restaurant chef wants to be sure that the vegetable soup she's cooking is up to her standards, she'll taste a spoonful. She doesn't need to consume the whole pot. She can trust that the taste will *represent* the flavor of the population—the entire pot. The idea of tasting is that a small sample, if selected properly, can represent the larger population. Sampling is common in many aspects of business practice. For example, auditors may sample some records rather than reading through all of them. Manufacturers monitor quality by testing a small sample off the line.

The W's and Sampling

The population we are interested in is usually determined by the *why* of our study. The participants or cases in the sample we draw will be the *who*. *When* and *how* we draw the sample may depend on what is practical.

The GfK Roper Reports® Worldwide poll is an example of a **sample survey** ⓘ, designed to ask questions of a small group of people in the hope of learning something about the entire population. Most likely, you've never been selected to be part of a national opinion poll. That's true of most people. So how can the pollsters claim that a sample represents the entire population? As we'll see, a representative sample can often provide a good idea of what the entire population is like. But the sample must be selected with care.

Selecting a sample to represent the population fairly is easy in theory, but in practice, it's more difficult than it sounds. For example, a sample may fail to represent part of the population. If a retail business samples customers as they come in the door, they may be missing an important part of their potential customer population—those who choose to shop elsewhere. Samples that over- or underemphasize some characteristics of the population are said to be biased. When a sample is **biased** ⓘ, the summary characteristics of a sample differ from the corresponding characteristics of the population it is trying to represent, so they can produce misleading information. Conclusions based on biased samples are

inherently flawed. There is usually no way to fix bias after the sample is drawn and no way to salvage useful information from it.

To make the sample as representative as possible, the best strategy is to select individuals for the sample *at random*. This may seem almost careless at first, but, as we will see, it is essential.

Idea 2: Randomize

Think back to our soup example. Suppose the chef adds some salt to the pot (the population). If she samples from the top before stirring, she'll get the misleading idea that the soup is salty. If she samples from the bottom, she'll get the equally misleading idea that it's bland. But by stirring the soup, she'll randomly distribute the salt throughout the soup making each spoonful a more representative sample of the soup, so each taste is more typical of the saltiness of the whole pot. Deliberate randomization is one of the great tools of statistics.



Randomization can also protect against factors that you aren't aware of. Suppose, while the chef isn't looking, an assistant adds a handful of peas to the soup. The peas sink to the bottom of the pot, mixing with the other vegetables. Stirring in the salt *also* randomizes the peas throughout the pot, making the sample taste more typical of the overall pot *even though the chef didn't know the peas were there*. So randomizing protects us by giving us a representative sample even for effects we were unaware of.

For a survey, we select participants at random, and this helps us represent *all* the features of our population, making sure that *on average* the sample looks like the rest of the population.

The essential feature of randomness is that the selection is "fair." We have discussed many facets of randomness in [Chapter 5](#), and we can use some of those concepts here. What makes the sample fair is that each participant has an equal chance to be selected.

- **Why not match the sample to the population?** Rather than randomizing, we could try to design a sample to include every possible, relevant characteristic: income level, age, political affiliation, marital status, number of children, place of residence, etc. But we can't possibly think of all the things that might be important. Even if we could, we wouldn't be able to match our sample to the population for all these characteristics.

How well can a sample represent the population from which it was selected? Here's an example using the database of the Paralyzed Veterans of America, a philanthropic organization with a donor list of about 3.5

million people. We've taken two samples, each of 8000 individuals at random from the population. **Table 8.1**  shows how the means and proportions match up on seven variables.

Table 8.1

Means and proportions for seven variables from two samples of size 8000 from the Paralyzed Veterans of America data. We drew these samples using Microsoft Excel's RAND function (Excel 2016), but you can use almost any statistics software to draw similar random samples. The fact that the summaries of the variables from these two samples are so similar gives us confidence that either one would be representative of the entire population.

	A	B	C	D	E	F	G	H
1		Age (yr)	White (%)	Female (%)	# of children	Income Bracket (1-7)	Wealth Bracket (1-9)	Homeowner? (% Yes)
2	Sample 1	61.4	85.12	56.2	1.54	3.91	5.29	71.36
3	Sample 2	61.2	84.44	56.4	1.51	3.88	5.33	72.30

The two samples match closely in every category. You can see how well randomizing has stirred the population. We didn't preselect the samples for these variables, but randomizing has matched the results closely. The two samples don't vary much from each other, so we can assume that they don't differ much from the rest of the population either.

Idea 3: The Sample Size Is What Matters

You probably weren't surprised by the idea that a sample can represent the whole. And the idea of sampling randomly to make the sample fair makes sense too. But the third important idea of sampling often surprises people. The third idea is that the *size of the sample* determines what we can conclude from the data *regardless of the size of the population*. Many people think that to provide a good representation of the population, the sample must be a large percentage, or *fraction*, of the population, but in fact all that matters is the size of the sample. The size of the *population* doesn't matter at all.³ A random sample of 100 students in a college represents the student body just about as well as a random sample of 100 voters represents the entire electorate of the United States. This is perhaps the most surprising idea in designing surveys.

³Well, that's not exactly true. If the population is smaller than about 10 times the size of the sample it *can* matter. It doesn't matter whenever, as usual, our sample is a very small fraction of the population.

Think about the pot of soup again. The chef is probably making a large pot of soup. But she doesn't need a really big spoon to decide how the soup tastes. She'll get the same information from an ordinary spoonful no matter how large the pot—as long as the pot is sufficiently stirred. That's what randomness does for us. What *fraction* of the population you sample doesn't matter. It's the **sample size**  itself that's important. This idea is of key importance to the design of any sample survey, because it determines the balance between how well the survey can measure the population and how much the survey costs.

How big a sample do you need? That depends on what you're estimating, but too small a sample won't be representative of the population. To get an idea of what's really in the soup, you need a large enough taste to be a *representative* sample from the pot, including, say, a selection of the vegetables. For a survey that

tries to find the proportion of the population falling into a category, you'll usually need at least several hundred respondents.

- **What do the professionals do?** How do professional polling and market research companies do their work? The most common polling method today is to contact respondents by telephone. But technology is changing the way polling can be done. We learned about Dalia back in [Chapter 2](#). Dalia samples from a sampling frame of up to 500 million people across 150 countries, accessing them by way of any of about 40,000 apps on their mobile devices. Dalia partners with the publishers of these apps to offer rewards for participation in the form of access to premium content, extra lives in games, gift cards, or charitable donations. Dalia uses stratification to invite a representative sample that matches known distributions in the population of interest. For example, they will sample the number of women in a narrow age range from a specified region based on census data that specify what fraction of the population fits those criteria. By contacting sampled individuals electronically, Dalia can complete a survey in less time, less expensively, and for a larger sample than with previous methodologies.

Do these methods work? The Pew Research Center for the People and the Press reports on survey completion rates about every three years. Pew reports that by 2012 a telephone survey could contact about 62% of households whose phone numbers had been randomly generated. However, only 14% of those contacts yielded an interview, amounting to only 9% of the households originally sampled. By contrast, Dalia reports Internet penetration of more than 85% in many countries (www.internetworldstats.com/top20.htm), and smartphone penetration is especially high in developing countries compared to computer or landline use. Nevertheless, Pew concludes that “telephone surveys that include landlines and cell phones and are weighted to match the demographic composition of the population continue to provide accurate data on most political, social and economic measures.” (www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/)

A Census—Does It Make Sense?

Why bother determining the right sample size? If you plan to open a store in a new community, why draw a sample of residents to understand their interests and needs? Wouldn't it be better to just include everyone and make the “sample” be the entire population? Such a special sample is called a **census**. Although a census would appear to provide the best possible information about the population, there are a number of reasons why it might not.

First, it can be difficult to complete a census. There always seem to be some individuals who are hard to locate or hard to measure. Do you really need to contact the folks away on vacation when you collect your data? How about those with no telephone or mailing address? The cost of locating the last few cases may far exceed the budget. It can also be just plain impractical to take a census. The quality control manager for Hostess® Twinkies® doesn't want to taste *all* the Twinkies on the production line to determine their quality. Aside from the fact that nobody could eat that many Twinkies, it would defeat their purpose: There would be none left to sell.

Second, the population you're studying may change. For example, in any human population, babies are born, people travel, and folks die during the time it takes to complete the census. News events and

advertising campaigns can cause sudden shifts in opinions and preferences. A sample, surveyed in a shorter time frame, may actually generate more accurate information.

Finally, taking a census can be cumbersome. A census usually requires a team of pollsters and the cooperation of the population. Even with both, it's almost impossible to avoid errors. Because it tries to count everyone, the U.S. Census records too many college students. Many are included both by their families and in a report filed by their schools. Errors of this sort, of both under- and overcounting, can be found throughout the U.S. Census.

In Practice 8.2 Identifying the sample

A nonprofit organization has taken over the historic State Theater and hopes to preserve it with a combination of attractive shows and fundraising. The board of directors of the organization has asked a team of student consultants to help them design a survey to better understand the customer base likely to purchase tickets. Fortunately, the theater's computerized ticket system records contact and some demographic information for ticket purchasers, and that database of 7345 customers is available.

Board of Directors Why can't we just perform a census?

Consultants The population of interest is all potential ticket purchasers. A census would have to reach all potential purchasers. We don't know who they are or have any way to contact them. We do have information about those who have purchased tickets and we can hope that potential purchasers would be like them.

8.3 Populations and Parameters

GfK Roper Reports Worldwide reports that 60.5% of people over 50 worry about food safety, but only 43.7% of teens do. What does this claim mean? We can be sure the Roper researchers didn't take a census. So they can't possibly know exactly what percentage of teenagers worry about food safety. So what does "43.7%" mean?

Statistic

Any quantity that we calculate from data could be called a "statistic." But in practice, we usually obtain a statistic from a sample and use it to estimate a population parameter.

To generalize from a sample to the world at large, we need a model of reality. Such a model doesn't need to be complete or perfect. Just as a model of an airplane in a wind tunnel can tell engineers what they need to know about aerodynamics even though it doesn't include every rivet of the actual plane, models of data can give us summaries that we can learn from and use even though they don't fit each data value exactly. It's important to remember that they're only models of reality and not reality itself. But without models, what we can learn about the world at large is limited to only what we can say about the data we have at hand.

Models use mathematics to represent reality. We call the key numbers in those models **parameters**. Sometimes a parameter used in a model for a population is called (redundantly) a **population parameter**.

Parameter

Population model parameters are not just unknown—usually they are unknowable. We take a sample and use the sample statistics to estimate them.

But let's not forget about the data. We use the data to try to estimate values for the population parameters. Any summary found from the data is a **statistic**. Those statistics that estimate population parameters are particularly interesting. Sometimes—and especially when we match statistics with the parameters they estimate—we use the term **sample statistic**.

We draw samples because we can't work with the entire population. We hope that the statistics we compute from the sample will estimate the corresponding parameters accurately. A sample that does this is said to be **representative**.

Just Checking

1. Various claims are often made for surveys. Why is each of the following claims not correct?

- a. It is always better to take a census than to draw a sample.
- b. Stopping customers as they are leaving a restaurant is a good way to sample opinions about the quality of the food.
- c. We drew a sample of 100 from the 3000 students in a school. To get the same level of precision for a town of 30,000 residents, we'll need a sample of 1000.
- d. A poll taken at a popular website garnered 12,357 responses. The majority of respondents said they enjoy doing statistics. With a sample size that large, we can be sure that most Americans feel this way.
- e. The true percentage of all Americans who enjoy statistics is called a "population statistic."

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

8.4 Common Sampling Designs

We've said that every individual in the population should have an equal chance of being selected in a sample. That makes the sample fair, but it's not quite enough to ensure that the sample is representative. Consider, for example, a market analyst who samples customers by drawing at random from product registration forms, half of which arrived by mail and half by online registration. She flips a coin. If it comes up heads, she'll draw 100 mail returns; tails, she'll draw 100 electronic returns. Each customer has an equal chance of being selected, but if tech-savvy customers are different, then the samples are hardly representative.

Simple Random Sample (SRS)

To make the sample representative, we must ensure that our sampling method gives each *combination* of individuals an equal chance as well. A sample drawn in this way is called a **simple random sample** ⓘ, usually abbreviated **SRS** ⓘ. An SRS is the sampling method on which the theory of working with sampled data is based and thus the standard against which we measure other sampling methods.

We'd like to select from the population, but often we don't have a list of all the individuals in the population. The list we actually draw from is called a **sampling frame** ⓘ. A store may want to survey all its regular customers. But it can't draw a sample from the population of all regular customers, because it doesn't have such a list. The store may have a list of customers who have registered as "frequent shoppers." That list can be the sampling frame from which the store can draw its sample.

Of course, whenever the sampling frame and the population differ (as they almost always will), we must deal with the differences. Are the opinions of those who registered as frequent shoppers different from the opinions of the rest of the regular shoppers? What about customers who used to be regulars but haven't shopped there recently? The answers to questions like these about the sampling frame may depend on the purpose of the survey and may impact the conclusions that one can draw.

Once we have a sampling frame, we need to **randomize** ⓘ it so we can choose an SRS. Fortunately, random numbers are readily available these days in spreadsheets, statistics programs, and even on the Internet. Before this technology existed, people used to literally draw numbers out of a hat to randomize. But now, the easiest way to randomize your sampling frame is to match it with a parallel list of random numbers and then sort the random numbers, carrying along the cases so that they get "shuffled" into random order. Then you can just pick cases off the top of the randomized list until you have enough for your sample.

Samples drawn at random generally differ one from another. If we were to repeat the sampling process, a new draw of random numbers would select different people for our sample. These differences would lead to different values for the variables we measure. We call these sample-to-sample differences **sampling variability** ⓘ. Sometimes they are called **sampling error** ⓘ even though no error has taken place. Surprisingly, sampling variability isn't a problem; it's an opportunity. If different samples from a population vary little from each other, then most likely the underlying population harbors little variation. If the samples show much sampling variability, the underlying population probably varies a lot. In the coming chapters, we'll

spend much time and attention working with sampling variability to better understand what we are trying to measure.

Sampling Errors vs. Bias

Referring to sample-to-sample variability as sampling error, makes it sound like it's some kind of mistake. It's not. We understand that samples will vary, so "sampling errors" are to be expected. It's bias we must strive to avoid. Bias means our sampling method distorts our view of the population. Of course, bias leads to mistakes. Even more insidious, bias introduces errors that we cannot correct with subsequent analysis.

In Practice 8.3 Choosing a random sample

Continuing In Practice from page 259, the student consultants select 200 ticket buyers at random from the database. First, the State Theater database is placed in a spreadsheet. Next, to draw random numbers, the students use the Excel command `RAND()`. (They type `=RAND()` in the top cell of a column next to the data and then use *Fill Down* to populate the column down to the bottom.) They then sort the spreadsheet to put the random column in order and select ticket buyers from the top of the randomized spreadsheet until they complete 200 interviews. This makes it easy to select more respondents when (as always happens) some of the people they select can't be reached by telephone or decline to participate.

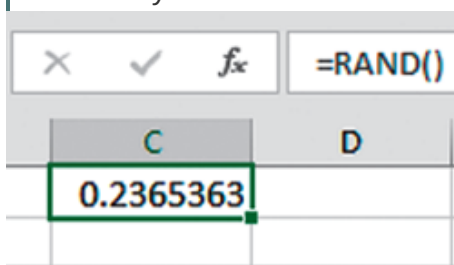
Board of Directors If we draw another sample will we get the same results? Also, next year, our database will have twice as many records. Will we need a sample that is twice as large?

Consultants With random samples, we expect results to vary from one sample to another. We call that *sampling error* or *sampling variability*. We call the list from which we sample respondents our *sampling frame*. But a larger sampling frame does not require a larger sample. The size of the sample is all that matters.

The sampling frame is the customer database.

The size of the sample is all that matters, not the size of the population. We would need a sample of 200.

The differences in the responses from one sample to another are called sampling error, or sampling variability.



A Different Answer Every Time?

The RAND() function in Excel can take you by surprise. Every time the spreadsheet reopens, you get a new column of random numbers. But don't worry. Once you've shuffled the rows, you can ignore the new numbers. The order you got by shuffling won't keep changing. (Image created in Microsoft Excel)

Simple random sampling is not the only fair way to sample. More complicated designs may save time or money or avert sampling problems. All statistical sampling designs have in common the idea that chance, rather than human choice, is used to select the sample.

Stratified Sampling

Designs that are used to sample from large populations—especially populations residing across large areas—are often more complicated than simple random samples. Sometimes we slice the population into homogeneous groups, called **strata**, and then use simple random sampling within each stratum, combining the results at the end. This is called **stratified random sampling**.

Why would we want to stratify? Suppose we want to survey how shoppers feel about a potential new anchor store at a large suburban mall. The shopper population is 60% women and 40% men, and we suspect that men and women have different views on their choice of anchor stores. If we use simple random sampling to select 100 people for the survey, we could end up with 70 men and 30 women or 35 men and 65 women. Our resulting estimates of the attractiveness of a new anchor store could vary widely. To help reduce this sampling variability, we can force a representative balance, selecting 40 men at random and 60 women at random. This would guarantee that the proportions of men and women within our sample match the proportions in the population, and that should make such samples more accurate in representing population opinion.

You can imagine that stratifying by race, income, age, and other characteristics can be helpful, depending on the purpose of the survey. When we use a sampling method that restricts by strata, additional samples are more like one another, so statistics calculated for the sampled values will vary less from one sample to another. This reduced sampling variability is the most important benefit of stratifying, but the analysis of data sampled with these designs is beyond the scope of our text.

Cluster and Multistage Sampling

Sometimes dividing the sample into homogeneous strata isn't practical, and even simple random sampling may be difficult. For example, suppose we wanted to assess the reading level of a product instruction manual based on the length of the sentences. Simple random sampling could be awkward; we'd have to number each sentence and then find, for example, the 576th sentence or the 2482nd sentence, and so on. Doesn't sound like much fun, does it?

Strata or Clusters?

We create strata by dividing the population into groups of similar individuals so that each stratum is different from the others. (For example, we often stratify by age, race, or sex.) By contrast, we create clusters that all look pretty much alike, each representing the wide variety of individuals seen in the population.



We could make our task much easier by picking a few pages at random and then counting the lengths of the sentences on those pages. That's easier than picking individual sentences and works if we believe that the pages are all reasonably similar to one another in terms of reading level. Splitting the population in this way into parts or **clusters** ⓘ that each represent the population can make sampling more practical. We select one or a few clusters at random and perform a census within each of them. This sampling design is called **cluster sampling** ⓘ. If each cluster fairly represents the population, cluster sampling will generate an unbiased sample.

What's the difference between cluster sampling and stratified sampling? We stratify to ensure that our sample represents different groups in the population, and sample randomly within each stratum. This reduces the sample-to-sample variability. Strata are homogeneous, but differ from one another. By contrast, clusters are more or less alike, each heterogeneous and resembling the overall population. We cluster to save money or even to make the study practical.

Sometimes we use a variety of sampling methods together. In trying to assess the reading level of our instruction manual, we might worry that the "quick start" instructions are easy to read, but the "troubleshooting" chapter is more difficult. If so, we'd want to avoid samples that selected heavily from any one chapter. To guarantee a fair mix of sections, we could randomly choose one section from each chapter of the manual. Then we would randomly select a few pages from each of those sections. If altogether that made too many sentences, we might select a few sentences at random from each of the chosen pages. So, what is our sampling strategy? First we stratify by the chapter of the manual and randomly choose a section to represent each stratum. Within each selected section, we choose pages as clusters. Finally, we consider an SRS of sentences within each cluster. Sampling schemes that combine several methods are called

multistage samples ⓘ. Most surveys conducted by professional polling organizations and market research firms use some combination of stratified and cluster sampling as well as simple random samples.

In Practice 8.4 More complex designs

The theater board wants to encourage people to come from out of town to attend theater events. They know that, in general, about 40% of ticket buyers are from out of town. These customers often purchase dinner at a local restaurant or stay overnight in a local inn, generating business for the town.

Board of Directors We would like to encourage local businesses to advertise in the theater program. They will want to know about out-of-town customers. Will they be properly represented in the sample?

Consultants The data includes ZIP codes, so we can identify out-of-town customers. We propose a stratified sample, consisting of a sample of 80 out-of-town customers and a sample of 120 local customers.

By stratifying, we can guarantee that 40% of the sample is from out of town, reflecting the overall proportions among ticket buyers. If out-of-town customers differ in important ways from local ticket buyers, a stratified sample will reduce the variation in the estimates for each group so that the combined estimates can be more precise.

Systematic Samples

Sometimes we draw a sample by selecting individuals systematically. For example, a **systematic sample** ⓘ might select every tenth person on an alphabetical list of employees. To make sure our sample is random, we still must start the systematic selection with a randomly selected individual—not necessarily the first person on the list. When there is no reason to believe that the order of the list could be associated in any way with the responses measured, systematic sampling can give a representative sample. Systematic sampling can be much less expensive than true random sampling. When you use a systematic sample, you should justify the assumption that the systematic method is not associated with any of the measured variables.

Think about the reading level sampling example again. Suppose we have chosen a section of the manual at random, then three pages at random from that section, and now we want to select a sample of 10 sentences from the 73 sentences found on those pages. Instead of numbering each sentence so we can pick a simple random sample, it would be easier to sample systematically. A quick calculation shows $73/10=7.3$, so we can get our sample by picking every seventh sentence on the page. But where should you start? At random, of course. We've accounted for $10 \times 7 = 70$ of the sentences, so we'll throw the extra three into the starting group and choose a sentence at random from the first 10. Then we pick every seventh sentence after that and record its length.

Just Checking

2. We need to survey a random sample of the 300 passengers on a flight from San Francisco to Tokyo. Name each sampling method described.
 - a. Pick every tenth passenger as people board the plane.

- b. From the boarding list, randomly choose five people flying first class and 25 of the other passengers.
- c. Randomly generate 30 seat numbers and survey the passengers who sit there.
- d. Randomly select a seat position (right window, right center, right aisle, etc.) and survey all the passengers sitting in those seats.

Guided Example Market Demand Survey



In a course at a business school in the United States, the students form business teams, propose a new product, and use seed money to launch a business to sell the product on campus.

Before committing funds for the business, each team must complete the following assignment: “Conduct a survey to determine the potential market demand on campus for the product you are proposing to sell.” Suppose your team’s product is a 500-piece jigsaw puzzle of the map of your college campus. Design a marketing survey and discuss the important issues to consider.

Define the problem—state the goals and objectives of the survey. Characterize the Population and parameters. Identify the population to be studied and the associated sampling frame. What are the parameters of interest? Sampling Plan Specify the sampling method and the planned sample size, n. Specify how the sample was actually drawn. What is the sampling frame? The description should, if possible, be complete enough to allow someone to replicate the procedure, drawing another sample from the same population in the same manner. A good description of the procedure is essential, even if it could never	Our team designed a study to find out how likely students at our school are to buy our proposed product—a 500-piece jigsaw puzzle of the map of our college campus. The population studied will be students at our school. We have obtained a list of all students currently enrolled to use as the sampling frame. The parameter of interest is the proportion of students likely to buy this product. We’ll also collect some demographic information about the respondents. We will select a simple random sample of 200 students. The sampling frame is the master list of students we obtained from the registrar. We decided against stratifying by sex or class because we thought that students were all more or less alike in their likely interest in our product.
---	---

practically be repeated. The question you ask is important, so state the wording of the question clearly. Be sure that the question is useful in helping you with the overall goal of the survey.	We will ask the students we contact: <i>Do you solve jigsaw puzzles for fun?</i> Then we will show them a prototype puzzle and ask: <i>If this puzzle sold for \$10, would you purchase one?</i> We will also record the respondent's sex and class.
Sampling Practice Specify <i>when, where,</i> and <i>how</i> the sampling will be performed. Specify any other details of your survey, such as how respondents were contacted, any incentives that were offered to encourage them to respond, how nonrespondents were treated, and so on.	The survey will be administered in the middle of the fall semester during October. We have a master list of registered students, which we will randomize by matching it with random numbers from www.random.org and sorting on the random numbers, carrying the names. We will contact selected students by phone or e-mail and arrange to meet with them. If a student is unwilling to participate, the next name from the randomized list will be substituted until a sample of 200 participants is found.
Communicate and Present This report should include a discussion of all the elements needed to design the study. It's good practice to discuss any special circumstances or other issues that may need attention.	MEMO Re: Survey plans Our team's plans for the puzzle market survey call for a simple random sample of students. Because subjects need to be shown the prototype puzzle, we must arrange to meet with selected participants. We have arranged an office for that purpose. We will also collect demographic information so we can determine whether there is in fact a difference in interest level among classes or between men and women.

We have been discussing sampling in a somewhat idealized setting. In the real world, things can be a bit messier. Here are some things to consider.

The population may not be as well defined as it seems. For example, if a company wants the opinions of a typical mall “shopper,” who should they sample? Should they only ask shoppers carrying a purchase? Should they include people eating at the food court? How about teenagers just hanging out in the mall? Even when the population is clear, it may not be possible to establish an appropriate sampling frame.

Usually, the practical sampling frame is not the group you *really* want to know about. For example, election polls want to sample from those who will actually vote in the next election—a group that is particularly tricky to identify before election day. The sampling frame limits what your survey can find out.

Then there’s your target sample. These are the individuals selected according to your sample design for whom you *intend* to measure responses. You’re not likely to get responses from all of them. (“I know it’s dinner time, but I’m sure you wouldn’t mind answering a few questions. It’ll only take 20 minutes or so. Oh, you’re busy?”) Nonresponse is a problem in many surveys.

What’s the Sample?

The population we want to study is determined by asking *why*. When we design a survey, we use the term “sample” to refer to the individuals selected, from whom we hope to obtain responses. Unfortunately, the real sample is just those we can reach to obtain responses—the *who* of the study. These are slightly different uses of the same term **sample** ⓘ. The context usually makes clear which we mean, but it’s important to realize that the difference between the two samples could undermine even a well-designed study.

Sample designs are usually about the target sample. But in the real world, you won’t get responses from everyone your design selects. So in reality, your sample consists of the actual respondents. These are the individuals about whom you *do* get data and can draw conclusions. Unfortunately, they might not be representative of either the sampling frame or the population.

At each step, the group we can study may be constrained further. The *who* of our study keeps changing, and each constraint can introduce biases. A careful study should address the question of how well each group matches the population of interest. The *who* in an SRS is the population of interest from which we’ve drawn a representative sample. That’s not always true for other kinds of samples.

When people (or committees!) decide on a survey, they often fail to think through the important questions about who are the *who* of the study and whether they are the individuals from whom the answers would be interesting or have meaningful business consequences. This is a key step in performing a survey and should not be overlooked.

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

duce.

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

8.5 The Valid Survey

It isn't sufficient to draw a sample and start asking questions. You want to feel confident your survey can yield the information you need about the population you are interested in. We want a *valid survey*.

To help ensure a valid survey, you need to ask four questions:

- What do I want to know?
- Who are the right respondents?
- What are the right questions?
- What will be done with the results?

These questions may seem obvious, but there are a number of specific pitfalls to avoid:

Know what you want to know

Far too often, decision makers decide to perform a survey without any clear idea of what they hope to learn. Before considering a survey, you must be clear about what you hope to learn and what population you want to learn about. If you don't know that, you can't even judge whether you have a valid survey. The survey *instrument*—the questionnaire itself—can be a source of errors. Perhaps the most common error is to ask unnecessary questions. The longer the survey, the fewer people will complete it, leading to greater nonresponse bias. For each question on your survey, you should ask yourself whether you really want to know this and know what you would do with the responses if you had them. If you don't have a good use for the answer to a question, don't ask it.

Use the right sampling frame

A valid survey obtains responses from appropriate respondents. Be sure you have a suitable sampling frame. Have you identified the population of interest and sampled from it appropriately? A company looking to expand its base might survey customers who returned warranty registration cards—after all, that's a readily available sampling frame—but if the company wants to know how to make its product more attractive, it needs to survey customers who rejected its product in favor of a competitor's product. This is the population that can tell the company what about its product needs to change to capture a larger market share. The errors in the presidential election polls of 1948 were likely due to the use of telephone samples in an era when telephones were not affordable by the less affluent—who were the folks most likely to vote for Truman.

It is equally important to be sure that your respondents actually know the information you hope to discover. Your customers may not know much about the competing products, so asking them to compare your product with others may not yield useful information.

Ask specific rather than general questions

It is better to be specific. "Do you usually recall TV commercials?" won't be as useful as "How many TV commercials can you recall from last night?" or better, yet, "Please describe for me all the TV commercials

you can recall from your viewing last night.”

Watch for biases

Even with the right sampling frame, you must beware of bias in your sample. If customers who purchase more expensive items are less likely to respond to your survey, this can lead to **nonresponse bias** ⓘ. Although you can't expect all mailed surveys to be returned, if those individuals who don't respond have common characteristics, your sample will no longer represent the population you hope to learn about. Surveys in which respondents volunteer to participate, such as online surveys, suffer from **voluntary response bias** ⓘ. Individuals with the strongest feelings on either side of an issue are more likely to respond; those who don't care may not bother.

Be careful with question phrasing

Questions must be carefully worded. A respondent may not understand the question—or may not understand the question the way the researcher intended it. For example, “Does anyone in your family own a Ford truck?” leaves the term “family” unclear. Does it include only spouses and children or parents and siblings, or do in-laws and second cousins count too? A question like “Was your Twinkie fresh?” might be interpreted quite differently by different people.

Be careful with answer phrasing

Respondents and survey-takers may also provide inaccurate responses, especially when questions are politically or sociologically sensitive. This also applies when the question does not take into account all possible answers, such as a true-false or multiple-choice question to which there may be other answers. Or the respondent may not know the correct answer to the question on the survey. In 1948, there were four major candidates for President,⁴ but some survey respondents might not have been able to name them all. A survey question that just asked “Who do you plan to vote for?” might have underrepresented the less prominent candidates. And one that just asked “What do you think of Wallace?” might yield inaccurate results from voters who simply didn't know who he was. We refer to inaccurate responses (intentional or unintentional) as **measurement errors** ⓘ. One way to cut down on measurement errors is to provide a range of possible responses. But be sure to phrase them in neutral terms.

⁴Harry Truman, Thomas Dewey, Strom Thurmond, and Henry Wallace.

The best way to protect a survey from measurement errors is to perform a pilot test. In a **pilot test** ⓘ, a small sample is drawn from the sampling frame, and a draft form of the survey instrument is administered. A pilot test can point out flaws in the instrument. For example, during a staff cutback at one of our schools, a researcher surveyed faculty members to ask how they felt about the reduction in staff support. The scale ran from “It's a good idea” to “I'm very unhappy.” Fortunately, a pilot study showed that everyone was very unhappy or worse. The scale was re-tuned to run from “unhappy” to “ready to quit.”

In Practice 8.5 Survey design

Board of Directors We are confused by some of the technical terms. Can you explain the population, sampling frame, target sample, and sample to us? Are there concerns about which customers we can contact?

Consultants The population is all potential ticket buyers.

The sampling frame consists of only those who have previously purchased tickets. Anyone who wasn't attracted to previous productions wouldn't be surveyed. That could keep the board from learning of ways to make the theater's offering more attractive to those who hadn't purchased tickets before.

The target sample is those selected from the database who can be contacted by telephone. Those with unlisted numbers or who had declined to give their phone number can't be contacted. It may be more difficult to contact those with caller ID.

The actual sample will be those previous customers selected at random from the database who can be reached by telephone and who agree to complete the survey.

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

8.6 How to Sample Badly

Bad sample designs yield worthless data. Many of the most convenient forms of sampling can be seriously biased. And there is no way to correct for the bias from a bad sample. So it's wise to pay attention to sample design—and to beware of reports based on poor samples.

Voluntary Response Sample

One of the most common dangerous sampling methods is the voluntary response sample. In a **voluntary response sample**, a large group of individuals is invited to respond, and all who do respond are counted. This method is used by call-in shows, 900 numbers, Internet polls, and letters written to members of Congress. Voluntary response samples are almost always biased, and so conclusions drawn from them are almost always wrong.

It's often hard to define the sampling frame of a voluntary response study. Practically, the frames are groups such as Internet users who frequent a particular website or viewers of a particular TV show. But those sampling frames don't correspond to the population you are likely to be interested in.

Even if the sampling frame is of interest, voluntary response samples are often biased toward those with strong opinions or those who are strongly motivated—and especially from those with strong negative opinions. A request that travelers who have used the local airport visit a survey site to report on their experiences is much more likely to hear from those who had long waits, cancelled flights, and lost luggage than from those whose flights were on time and carefree. The resulting voluntary response bias invalidates the survey.

Convenience Sampling

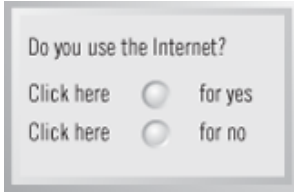
Another sampling method that doesn't work is convenience sampling. As the name suggests, in **convenience sampling** we simply include the individuals who are convenient. Unfortunately, this group may not be representative of the population. A survey of 437 potential home buyers in Orange County, California, found, among other things, that

all but 2 percent of the buyers have at least one computer at home, and 62 percent have two or more. Of those with a computer, 99 percent are connected to the Internet (Jennifer Hieger, "Portrait of Homebuyer Household: 2 Kids and a PC," Orange County Register, July 27, 2001).

Later in the article, we learn that the survey was conducted via the Internet. That was a convenient way to collect data and surely easier than drawing a simple random sample, but perhaps home builders shouldn't conclude from this study that every family has a computer and an Internet connection.

Many surveys conducted at shopping malls suffer from the same problem. People in shopping malls are not necessarily representative of the population of interest. Mall shoppers tend to be more affluent and include a

larger percentage of teenagers and retirees than the population at large. To make matters worse, survey interviewers tend to select individuals who look “safe,” or easy to interview.



Do you use the Internet?

Click here ☐ for yes

Click here ☐ for no

Internet Surveys

Internet convenience surveys are often worthless. As voluntary response surveys, they have no well-defined sampling frame (all those who use the Internet and visit their site?) and thus report no useful information. Do not use them.

Convenience sampling is not just a problem for beginners. In fact, convenience sampling is a widespread problem in the business world. When a company wants to find out what people think about its products or services, it may turn to the easiest people to sample: its own customers. But the company will never learn how those who *don't* buy its product feel about it.

Bad Sampling Frame?

An SRS from an incomplete sampling frame introduces bias because the individuals included may differ from the ones not in the frame. It may be easier to sample workers from a single site, but if a company has many sites and they differ in worker satisfaction, training, or job descriptions, the resulting sample can be biased. There is serious concern among professional pollsters that the increasing numbers of people who can be reached only by cell phone may bias telephone-based market research and polling.

Undercoverage

Many survey designs suffer from **undercoverage** ⓘ, in which some portion of the population is not sampled at all or has a smaller representation in the sample than it has in the population. Undercoverage can arise for a number of reasons, but it's always a potential source of bias. Are people who use answering machines to screen callers (and are thus less available to blind calls from market researchers) different from other customers in their purchasing preferences?

In Practice 8.6 Common mistakes in survey design

Board of Directors Why not simply post someone at the door to ask theatergoers their opinions? Another Board member asks, wouldn't it be even easier to post a questionnaire on the theater website and invite responses there? A third Board member suggests that rather than working with random numbers, we simply phone every 200th person on the list of past customers.

Consultants Questioning customers at the door would be a convenience sample. It would be cheap and fast but is likely to be biased by the nature and quality of the particular performance where the

survey takes place.

Inviting responses on the website would be a voluntary response sample. Only customers who frequented the website and decided to respond would be surveyed. This might, for example, underrepresent older customers or those without home Internet access.

Sampling every 200th name from the customer list would be a systematic sample. It is slightly easier than randomizing. If the order of names on the list is unrelated to any questions asked, then this might be an acceptable method. But if, for example, the list is kept in the order of first purchases (when a customer's name and information were added to the database), then there might be a relationship between opinions and location on the list.

What Can Go Wrong?

- **Nonrespondents.** No survey succeeds in getting responses from everyone. The problem is that those who don't respond may differ from those who do. And if they differ on just the variables we care about, the lack of response will bias the results. Rather than sending out a large number of surveys for which the response rate will be low, it is often better to design a smaller, randomized survey for which you have the resources to ensure a high response rate.
- **Long, dull surveys.** Surveys that are too long are more likely to be refused, reducing the response rate and biasing all the results. Keep it short.
- **Response bias.** Response bias includes the tendency of respondents to tailor their responses to please the interviewer and the consequences of slanted question wording.
- **Push polls.** Push polls, which masquerade as surveys, present one side of an issue before asking a question. For example, a question like
Would the fact that the new store that just opened by the mall sells mostly goods made overseas by workers in sweatshop conditions influence your decision to shop there rather than in the downtown store that features American-made products?

is designed not to gather information, but to spread ill-will toward the new store.

THE WIZARD OF ID



The Wizard of Id © 2001 John L. Hart/Distributed by Creators Syndicate. Reprinted with permission. All rights reserved.

How to Think about Biases

- **Look for biases in any survey.** If you design a survey of your own, ask someone else to help look for biases that may not be obvious to you. Do this *before* you collect your data. There's no way to recover from a biased sample or a survey that asks biased questions.

A bigger sample size for a biased study just gives you a bigger useless study. A really big sample gives you a really big useless study.

- **Spend your time and resources reducing biases.** No other use of resources is as worthwhile as reducing the biases.
- **If you possibly can, pretest or pilot your survey.** Administer the survey in the exact form that you intend to use it to a small sample drawn from the population you intend to sample. Look for misunderstandings, misinterpretation, confusion, or other possible biases. Then redesign your survey instrument.
- **Always report your sampling methods in detail.** Others may be able to detect biases where you did not expect to find them.

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Chapter 8 From Learning to Earning

Learning Objectives

Recognize observational studies.

- A retrospective study looks at an outcome in the present and looks for facts in the past that relate to it.
- A prospective study selects subjects and follows them as events unfold.

Know the three ideas of sampling.

- Examine a part of the whole: A sample can give information about the population.
- Randomize to make the sample representative.
- The sample size is what matters. It's the size of the sample—and not its fraction of the larger population—that determines the precision of the statistics it yields.

Be able to draw a simple random sample (SRS) using a table of random digits or a list of random numbers from technology or an Internet site.

- In a **simple random sample** (SRS), every possible group of n individuals has an equal chance of being our sample.

Know the definitions of other sampling methods:

- **Stratified samples** can reduce sampling variability by identifying homogeneous subgroups and then randomly sampling within each.
- **Cluster samples** randomly select among heterogeneous subgroups that each resemble the population at large, making our sampling tasks more manageable.
- **Systematic samples** can work in some situations and are often the least expensive method of sampling. But we still want to start them randomly.
- **Multistage samples** combine several random sampling methods.

Identify and avoid causes of bias.

- **Nonresponse bias** can arise when sampled individuals will not or cannot respond.
- **Response bias** arises when respondents' answers might be affected by external influences, such as question wording or interviewer behavior.
- **Voluntary response samples** are almost always biased and should be avoided and distrusted.
- **Convenience samples** are likely to be flawed for similar reasons.
- **Undercoverage** occurs when individuals from a subgroup of the population are selected less often than they should be.

Terms

Bias

Any systematic failure of a sampling method to represent its population.

Census

An attempt to collect data on the entire population of interest.

Cluster

A subset of a population aggregated into larger sampling units. These units, chosen for reasons of cost or practicality, are often natural groups thought to be representative of the population.

Cluster sampling

A sampling design in which groups, or clusters, representative of the population are chosen at random and a census is then taken of each.

Convenience sampling

A sample that consists of individuals who are conveniently available.

Measurement error

Any inaccuracy in a response, from any source, whether intentional or unintentional.

Multistage sample

A sampling scheme that combines several sampling methods.

Nonresponse bias

Bias introduced to a sample when a large fraction of those sampled fails to respond.

Observational study

A study based on data in which no manipulation of factors has been employed.

Parameter

A numerically valued attribute of a model for a population. We rarely expect to know the value of a parameter, but we do hope to estimate it from sampled data.

Pilot test

A small trial run of a study to check that the methods of the study are sound.

Population

The entire group of individuals or instances about whom we hope to learn.

Population parameter

A numerically valued attribute of a model for a population.

Prospective study

An observational study in which subjects are followed to observe future outcomes. Because no treatments are deliberately applied, a prospective study is not an experiment. Nevertheless, prospective studies typically focus on estimating differences among groups that might appear as the groups are followed during the course of the study.

Randomization

A defense against bias in the sample selection process, in which each individual is given a fair, random chance of selection.

Representative sample

A sample from which the statistics computed accurately reflect the corresponding population parameters.

Response bias

Anything in a survey design that influences responses.

Retrospective study

An observational study in which subjects are selected and then their previous conditions or behaviors are determined. Because retrospective studies are not based on random samples, they usually focus on estimating differences between groups or associations between variables.

Sample

A subset of a population, examined in hope of learning about the population.

Sample size

The number of individuals in a sample.

Sample survey

A study that asks questions of a sample drawn from some population in the hope of learning something about the entire population.

Sampling frame

A list of individuals from which the sample is drawn. Individuals in the population of interest but who are not in the sampling frame cannot be included in any sample.

Sampling variability (or sampling error)

The natural tendency of randomly drawn samples to differ, one from another.

Simple random sample (SRS)

A sample in which each set of n elements in the population has an equal chance of selection.

Statistic, sample statistic

A value calculated for sampled data, particularly one that corresponds to, and thus estimates, a population parameter. The term "sample statistic" is sometimes used, usually to parallel the corresponding term "population parameter."

Strata

Subsets of a population that are internally homogeneous but may differ one from another.

Stratified random sample

A sampling design in which the population is divided into several homogeneous subpopulations, or strata, and random samples are then drawn from each stratum.

Systematic sample

A sample drawn by selecting individuals systematically from a sampling frame.

Undercoverage

A sampling scheme that biases the sample in a way that gives a part of the population less representation than it has in the population.

Voluntary response bias

Bias introduced to a sample when individuals can choose on their own whether to participate in the sample.

Voluntary response sample

A sample in which a large group of individuals are invited to respond and decide individually whether or not to participate. Voluntary response samples are generally worthless.

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Tech Support

Computer-generated random numbers are usually quite good enough for drawing random samples. But there is little reason not to use the truly random values available on the Internet. Here's a convenient way to draw an SRS of a specified size using a computer-based sampling frame. The sampling frame can be a list of names or identification numbers arrayed, for example, as a column in a spreadsheet, statistics program, or database:

1. Generate random numbers of enough digits so that each exceeds the size of the sampling frame list by several digits. This makes duplication unlikely. (For example, in Excel, use the RAND function described in detail in Technology Help, [Chapter 5](#), to fill a column with random numbers between 0 and 1. With many digits they will almost surely be unique.)
2. Assign the random numbers arbitrarily to individuals in the sampling frame list. For example, put them in an adjacent column.
3. Sort the list of random numbers, *carrying along* the sampling frame list.
4. Now the first n values in the sorted sampling frame column are an SRS of n values from the entire sampling frame.

Most statistics packages also offer commands to sample from your data, but you should be careful to see that they do what you intend.

Chapter 8 Exercises

Section 8.1

1. For the following observational studies, indicate whether they are prospective or retrospective.
 - a. A company looked at a sample of returned registration cards to estimate the income level of households that purchased their product.
 - b. A retail outlet encouraged customers to join their “frequent buyers” program and studied whether those who joined were more likely to make use of discount coupons than those who were not members.
2. For the following observational studies, indicate whether they are prospective or retrospective studies.
 - a. An airline was concerned that new security measures might discourage air travelers. A year after the new security restrictions were put into place, the airlines compared the miles traveled by their frequent fliers before and after the change.
 - b. Does giving children a flu shot protect parents? Researchers questioned a random sample of families at the end of a flu season. They asked whether the children had been immunized, whether the parents had received flu shots, and who in the family had contracted the flu.

Section 8.2

3. Indicate whether each statement below is true or false. If false, explain why.
 - a. We can eliminate sampling error by selecting an unbiased sample.
 - b. Randomization helps to ensure that our sample is representative.
 - c. Sampling error refers to sample-to-sample differences and is also known as sampling variability.
 - d. It is better to try to match the characteristics of the sample to the population rather than relying on randomization.
4. Indicate whether each statement below is true or false. If false, explain why.
 - a. To get a representative sample, you must sample a large fraction of the population.
 - b. Using modern methods, it is best to select a representative subset of a population systematically.
 - c. A census is the only true representative sample.
 - d. A random sample of 100 students from a school with 2000 students has the same precision as a random sample of 100 from a school with 20,000 students.

Section 8.3

5. An environmental advocacy group is interested in the perceptions of farmers about global climate change. Specifically, they wish to determine the percentage of organic farmers who are concerned that climate change will affect their crop yields. They use an alphabetized list of members of the Northeast Organic Farming Association (www.nofa.org), a nonprofit organization of over 5000 members with chapters in Connecticut, Massachusetts, New Hampshire, New Jersey, New York, Rhode Island, and Vermont. They use Excel to generate a randomly shuffled list of the members. They then select members to contact from this list until they have succeeded in contacting 150 members.
- What is the population?
 - What is the sampling frame?
 - What is the population parameter of interest?
6. An airline company is interested in the opinions of their frequent flyer customers about their proposed new routes. Specifically they want to know what proportion of them plan to use one of their new hubs in the next six months. They take a random sample of 10,000 from the database of all frequent flyers and send them an e-mail message with a request to fill out a survey in exchange for 1500 miles.
- What is the population?
 - What is the sampling frame?
 - What is the population parameter of interest?

Section 8.4

7. As discussed in the chapter, GfK Roper Consulting conducts a global consumer survey to help multinational companies understand different consumer attitudes throughout the world. In India, the researchers interviewed 1000 people aged 13–65 (www.gfkamerica.com). Their sample is designed so that they get 500 males and 500 females.
- Are they using a simple random sample? How do you know?
 - What kind of design do you think they are using?
8. For their class project, a group of Business students decides to survey the student body to assess opinions about a proposed new student coffee shop to judge how successful it might be. Their sample of 200 contained 50 first-year students, 50 sophomores, 50 juniors, and 50 seniors.
- Do you think the group was using an SRS? Why?
 - What kind of sampling design do you think they used?
9. The environmental advocacy group from **Exercise 5** that was interested in gauging perceptions about climate change among organic farmers has decided to use a different method to sample. Instead of randomly selecting members from a shuffled list, they listed the members in alphabetical order and took every tenth member until they succeeded in contacting 150 members. What kind of sampling method have they used?
10. The airline company from **Exercise 6**, interested in the opinions of their frequent flyer customers about their proposed new routes, has decided that different types of customers might have different opinions. Of their customers, 50% are silver-level, 30% are blue, and 20% are red. They first compile

separate lists of silver, blue, and red members and then randomly select 5000 silver members, 3000 blue members, and 2000 red members to e-mail. What kind of sampling method have they used?

For **Exercises 11** and **12**, identify the following if possible. (If not, say why.)

- a. The population
 - b. The population parameter of interest
 - c. The sampling frame
 - d. The sample
 - e. Any potential sources of bias you can detect and any problems you see in generalizing to the population of interest
11. A business magazine mailed a questionnaire to the human resources directors of all *Fortune* 500 companies, and received responses from 23% of them. Those responding reported that they did not find that such surveys intruded significantly on their workday.
 12. A question posted on the Lycos website asked visitors to the site to say whether they thought that businesses should be required to pay for their employees' health insurance.

Section 8.5

13. An intern for the environmental group in **Exercise 5** has decided to make the survey process simpler by calling 150 of the members who attended the recent symposium on coping with climate change that was recently held in Burlington, VT. He has all the phone numbers, so it will be easy to contact them. He will start calling members from the top of the list, which was generated as the members enrolled for the symposium. He has written a script to read to them that follows, "As we learned in Burlington, climate change is a serious problem for farmers. Given the evidence of impact on crops, do you agree that the government should be doing more to fight global warming?"
 - a. What is the population of interest?
 - b. What is the sampling frame?
 - c. Point out any problems you see either with the sampling procedure and/or the survey itself. What are the potential impacts of these problems?
14. The airline company in **Exercise 6** has realized that some of its customers don't have e-mail or don't read it regularly. They decide to restrict the mailing only to customers who have recently registered for a "Win a trip to Miami" contest, figuring that those with Internet access are more likely to read and to respond to their e-mail. They send an e-mail with the following message: "Did you know that National Airlines has just spent over \$3 million refurbishing our brand new hub in Miami? By answering the following question, you may be eligible to win \$1000 worth of coupons that can be spent in any of the fabulous restaurants or shops in the Miami airport. Might you possibly think of traveling to Miami in the next six months on your way to one of your destinations?"
 - a. What is the population?
 - b. What is the sampling frame?
 - c. Point out any problems you see either with the sampling procedure and/or the survey itself. What are the potential impacts of these problems?

15. An intern is working for Pacific TV (PTV), a small cable and Internet provider, and has proposed some questions that might be used in the survey to assess whether customers are willing to pay \$50 for a new service.

Question 1: If PTV offered state-of-the-art, high-speed Internet service for \$50 per month, would you subscribe to that service?

Question 2: Would you find \$50 per month—less than the cost of a daily cappuccino—an appropriate price for high-speed Internet service?

- Do you think these are appropriately worded questions? Why or why not?
- Which one has more neutral wording? Explain.

16. Here are more proposed survey questions for the survey in **Exercise 15**:

Question 3: Do you find that the slow speed of DSL Internet access reduces your enjoyment of web services?

Question 4: Given the growing importance of high-speed Internet access for your children's education, would you subscribe to such a service if it were offered?

- Do you think these are appropriately worded questions? Why or why not?
- Suggest a question with better wording.

Section 8.6

17. Indicate whether each statement below is true or false. If false, explain why.

- A local television news program that asks viewers to call in and give their opinion on an issue typically results in a biased voluntary response sample.
- Convenience samples are generally representative of the population.
- Measurement error is the same as sampling error.
- A pilot test can be useful for identifying poorly worded questions on a survey.

18. Indicate whether each statement below is true or false. If false, explain why.

- Asking viewers to call into an 800 number is a good way to produce a representative sample.
- When writing a survey, it's a good idea to include as many questions as possible to ensure efficiency and to lower costs.
- A recent poll on a website was valid because the sample size was over 1,000,000 respondents.
- Malls are not necessarily good places to conduct surveys because people who frequent malls may not be representative of the population at large.

19. For your marketing class, you'd like to take a survey from a sample of all the Catholic Church members in your city to assess the market for a DVD about Pope Francis's first year as pope. A list of churches shows 17 Catholic churches within the city limits. Rather than try to obtain a list of all members of all these churches, you decide to pick 3 churches at random. For those churches, you'll ask to get a list of all current members and contact 100 members at random.

- What kind of design have you used?
- What could go wrong with the design that you have proposed?

20. The U.S. Fish and Wildlife Service plans to study the fishing industry around Saginaw Bay. To do that, they decide to randomly select five fishing boats at the end of a randomly chosen fishing day and count the numbers and types of all the fish on those boats.
- What kind of design have they used?
 - What could go wrong with the design that they have proposed?

Chapter Exercises

21. **Software licenses.** The website www.gamefaqs.com asked, as their question of the day to which visitors to the site were invited to respond, “Do you ever read the end-user license agreements when installing software or games?” Of the 98,574 respondents, 63.47% said they never read those agreements—a fact that software manufacturers might find important.
- What kind of sample was this?
 - How much confidence would you place in using 63.47% as an estimate of the fraction of people who don’t read software licenses?
22. **Drugs in baseball.** Major League Baseball, responding to concerns about their “brand,” tests players to see whether they are using performance-enhancing drugs. Officials select a team at random, and a drug-testing crew shows up unannounced to test all 40 players on the team. Each testing day can be considered a study of drug use in Major League Baseball.
- What kind of sample is this?
 - Is that choice appropriate?
23. **Pew.** Pew Research Center publishes polls on issues important in the news and about American life at its website, www.pewinternet.org. At the end of a report about a survey you can find a paragraph such as this one:
- These readings come from a national survey conducted between November 14 and December 9, 2012 of U.S. adults on landline and cell phones and in English and in Spanish. The results reported here come from the 1,802 respondents who are internet users and the margin of error is +/-2.6 percentage points.*
- For this survey, identify the population of interest.
 - Pew performs its surveys by phoning numbers generated at random by a computer program. What is the sampling frame? Does this seem representative of the population?
24. **Defining the survey.** At its website (www.gallup.com) the Gallup World Poll reports results of surveys conducted in various places around the world. At the end of one of these reports about the reliability of electric power in Africa, they describe their methods, including explanations such as the following:
- Results are based on face-to-face interviews with 1,000 adults, aged 15 and older, conducted in 2010 in Botswana, Burkina Faso, Cameroon, Central African Republic, Chad, Ghana, Kenya, Liberia, Mali, Niger, Nigeria, Senegal, Sierra Leone, South Africa, Tanzania, Uganda, and Zimbabwe. For results based on the total sample of national adults, one can say with 95% confidence that the maximum margin of sampling error ranges from ± 3.4 percentage points to ± 4.0 percentage points. The margin*

of error reflects the influence of data weighting. In addition to sampling error, question wording and practical difficulties in conducting surveys can introduce error or bias into the findings of public opinion polls.⁵

⁵Copyright © 2011 Gallup Inc. All rights reserved. The content is used with permission; however, Gallup retains all rights of republication.

- a. Gallup is interested in the opinions of Africans. What kind of survey design are they using?
- b. Some of the countries surveyed have large populations. (South Africa is estimated to have over 50 million people.) Some are quite small. (Zimbabwe has fewer than 13,000,000 people.) Nonetheless, Gallup sampled 1000 adults in each country. How does this affect the precision of its estimates for these countries?

25–32. Survey details. For the following reports about statistical studies, identify the following items (if possible). If you can't tell, then say so—this often happens when we read about a survey.

- a. The population
 - b. The population parameter of interest
 - c. The sampling frame
 - d. The sample
 - e. The sampling method, including whether or not randomization was employed
 - f. Any potential sources of bias you can detect and any problems you see in generalizing to the population of interest
25. **Teens and technology.** Pew Internet & American Life Project surveyed 802 pairs of parents and teens (aged 12–17). They report that 93% of teens have access to a computer. 25% of teens access the Internet primarily on their cell phone rather than on a computer.
26. **Global warming.** The Gallup Poll interviewed 1022 randomly selected U.S. adults aged 18 and older, March 7–10, 2013. Gallup reports that when asked whether respondents thought that global warming was due primarily to human activities, 57% of respondents said it was.
27. **At the bar.** Researchers waited outside a bar they had randomly selected from a list of such establishments. They stopped every tenth person who came out of the bar and asked whether he or she thought drinking and driving was a serious problem.
28. **Election poll.** Hoping to learn what issues may resonate with voters in the coming election, the campaign director for a mayoral candidate selects one block at random from each of the city's election districts. Staff members go there and interview all the residents they can find.
29. **Toxic waste.** The Environmental Protection Agency took soil samples at 16 locations near a former industrial waste dump and checked each for evidence of toxic chemicals. They found no elevated levels of any harmful substances.
30. **Housing discrimination.** Inspectors send trained “renters” of various races and ethnic backgrounds, and of both sexes to inquire about renting randomly assigned advertised apartments. They look for evidence that landlords deny access illegally based on race, sex, or ethnic background.
31. **Quality control.** A company packaging snack foods maintains quality control by randomly selecting 10 cases from each day's production and weighing the bags. Then they open one bag from each

- case and inspect the contents.
32. **Contaminated milk.** Dairy inspectors visit farms unannounced and take samples of the milk to test for contamination. If the milk is found to contain dirt, antibiotics, or other foreign matter, the milk will be destroyed and the farm is considered to be contaminated pending further testing.
33. **Instant poll.** A local TV station conducted an “Instant Poll” to predict the winner in the upcoming mayoral election. Evening news viewers were invited to phone in their votes, with the results to be announced on the late-night news. Based on the phone calls, the station predicted that Amabo would win the election with 52% of the vote. They were wrong: Amabo lost, getting only 46% of the vote. Do you think the station’s faulty prediction is more likely to be a result of bias or sampling error? Explain.
34. **Paper poll.** Prior to the mayoral election discussed in [Exercise 33](#), the newspaper also conducted a poll. The paper surveyed a random sample of registered voters stratified by political party, age, sex, and area of residence. This poll predicted that Amabo would win the election with 52% of the vote. The newspaper was wrong: Amabo lost, getting only 46% of the vote. Do you think the newspaper’s faulty prediction is more likely to be a result of bias or sampling error? Explain.
35. **Cable company market research.** A local cable TV company, Pacific TV (PTV), with customers in 15 towns is considering offering high-speed Internet service on its cable lines. Before launching the new service they want to find out whether customers would pay the \$75 per month that they plan to charge. An intern has prepared several alternative plans for assessing customer demand. For each, indicate what kind of sampling strategy is involved and what (if any) biases might result.
- Put a big ad in the newspaper asking people to log their opinions on the PTV website.
 - Randomly select one of the towns and contact every cable subscriber by phone.
 - Send a survey to each customer and ask him or her to fill it out and return it.
 - Randomly select 20 customers from each town. Send them a survey, and follow up with a phone call if they do not return the survey within a week.
36. **Cable company market research, part 2.** Four new sampling strategies have been proposed to help PTV determine whether enough cable subscribers are likely to purchase high-speed Internet service. For each, indicate what kind of sampling strategy is involved and what (if any) biases might result.
- Run a poll on the local TV news, asking people to dial one of two phone numbers to indicate whether they would be interested.
 - Hold a meeting in each of the 15 towns, and tally the opinions expressed by those who attend the meetings.
 - Randomly select one street in each town and contact each of the households on that street.
 - Go through the company’s customer records, selecting every 40th subscriber. Send employees to those homes to interview the people chosen.
37. **Amusement park riders.** An amusement park has opened a new roller coaster. It is so popular that people are waiting for up to three hours for a two-minute ride. Concerned about how patrons (who paid a large amount to enter the park and ride on the rides) feel about this, they survey every tenth person in line for the roller coaster, starting from a randomly selected individual.
- What kind of sample is this?
 - Is it likely to be representative?
 - What is the sampling frame?

38. **Playground.** Some people have been complaining that the children's playground at a municipal park is too small and is in need of repair. Managers of the park decide to survey city residents to see if they believe the playground should be rebuilt. They hand out questionnaires to parents who bring children to the park. Describe possible biases in this sample.
39. **Another ride.** The survey of patrons waiting in line for the roller coaster in [Exercise 37](#) asks whether they think it is worthwhile to wait a long time for the ride and whether they'd like the amusement park to install still more roller coasters. What biases might cause a problem for this survey?
40. **Playground bias.** The survey described in [Exercise 38](#) asked,
Many people believe this playground is too small and in need of repair. Do you think the playground should be repaired and expanded even if that means imposing an entrance fee to the park?
Describe two ways this question may lead to response bias.
41. **(Possibly) Biased questions.** Examine each of the following questions for possible bias. If you think the question is biased, indicate how and propose a better question.
- Should companies that pollute the environment be compelled to pay the costs of cleanup?
 - Should a company enforce a strict dress code?
42. **More possibly biased questions.** Examine each of the following questions for possible bias. If you think the question is biased, indicate how and propose a better question.
- Do you think that price or quality is more important in selecting a tablet computer?
 - Given humanity's great tradition of exploration, do you favor continued funding for space flights?
43. **Phone surveys.** Anytime we conduct a survey, we must take care to avoid undercoverage. Suppose we plan to select 500 names from the city phone book, call their homes between noon and 4 p.m., and interview whoever answers, anticipating contacts with at least 200 people.
- Why is it difficult to use a simple random sample here?
 - Describe a more convenient, but still random, sampling strategy.
 - What kinds of households are likely to be included in the eventual sample of opinion? Who will be excluded?
 - Suppose, instead, that we continue calling each number, perhaps in the morning or evening, until an adult is contacted and interviewed. How does this improve the sampling design?
 - Random-digit dialing machines can generate the phone calls for us. How would this improve our design? Is anyone still excluded?
44. **Cell phone survey.** What about drawing a random sample only from cell phone exchanges? Discuss the advantages and disadvantages of such a sampling method compared with surveying randomly generated telephone numbers from non-cell phone exchanges. Do you think these advantages and disadvantages have changed over time? How do you expect they'll change in the future?
45. **Change.** How much change do you have on you right now? Go ahead, count it.
- How much change do you have?
 - Suppose you check on your change every day for a week as you head for lunch and average the results. What parameter would this average estimate?

- c. Suppose you ask 10 friends to average *their* change every day for a week, and you average those 10 measurements. What is the population now? What parameter would this average estimate?
- d. Do you think these 10 average change amounts are likely to be representative of the population of change amounts in your class? In your college? In the country? Why or why not?
46. **Fuel economy.** Occasionally, when I fill my car with gas, I figure out how many miles per gallon my car got. I wrote down those results after six fill-ups in the past few months. Overall, it appears my car gets 28.8 miles per gallon.
- What statistic have I calculated?
 - What is the parameter I'm trying to estimate?
 - How might my results be biased?
 - When the Environmental Protection Agency (EPA) checks a car like mine to predict its fuel economy, what parameter is it trying to estimate?
47. **Accounting.** Between quarterly audits, a company likes to check on its accounting procedures to address any problems before they become serious. The accounting staff processes payments on about 120 orders each day. The next day, the supervisor rechecks 10 of the transactions to be sure they were processed properly.
- Propose a sampling strategy for the supervisor.
 - How would you modify that strategy if the company makes both wholesale and retail sales, requiring different bookkeeping procedures?
48. **Happy workers?** A manufacturing company employs 14 project managers, 48 foremen, and 377 laborers. In an effort to keep informed about any possible sources of employee discontent, management wants to conduct job satisfaction interviews with a simple random sample of employees every month.
- Do you see any danger of bias in the company's plan? Explain.
 - How might you select a simple random sample?
 - Why do you think a simple random sample might not provide the best estimate of the parameters the company wants to estimate?
 - Propose a better sampling strategy.
 - Listed below are the last names of the project managers. Use random numbers to select two people to be interviewed. Be sure to explain your method carefully.

Barrett	Bowman	Chen
DeLara	DeRoos	Grigorov
Maceli	Mulvaney	Pagliarulo
Rosica	Smithson	Tadros
Williams	Yamamoto	

49. **Quality control.** Sammy's Salsa, a small local company, produces 20 cases of salsa a day. Each case contains 12 jars and is imprinted with a code indicating the date and batch number. To help maintain consistency, at the end of each day, Sammy selects three bottles of salsa, weighs the contents, and tastes the product. Help Sammy select the sample jars. Today's cases are coded 07N61 through 07N80.
- Carefully explain your sampling strategy.
 - Show how to use random numbers to pick the three jars for testing.
 - Did you use a simple random sample? Explain.
50. **Fish quality.** Concerned about reports of discolored scales on fish caught downstream from a newly sited chemical plant, scientists set up a field station in a shoreline public park. For one week they asked fishermen there to bring any fish they caught to the field station for a brief inspection. At the end of the week, the scientists said that 18% of the 234 fish that were submitted for inspection displayed the discoloration. From this information, can the researchers estimate what proportion of fish in the river have discolored scales? Explain.
51. **Sampling methods.** Consider each of these situations. Do you think the proposed sampling method is appropriate? Explain.
- We want to know what percentage of local doctors accept Medicaid patients. We call the offices of 50 doctors randomly selected from local Yellow Pages listings.
 - We want to know what percentage of local businesses anticipate hiring additional employees in the upcoming month. We randomly select a page in the Yellow Pages and call every business listed there.
52. **More sampling methods.** Consider each of these situations. Do you think the proposed sampling method is appropriate? Explain.
- We want to know if business leaders in the community support the development of an "incubator" site at a vacant lot on the edge of town. We spend a day phoning local businesses in the phone book to ask whether they'd sign a petition.
 - We want to know if travelers at the local airport are satisfied with the food available there. We go to the airport on a busy day and interview every tenth person in line in the food court.

Just Checking Answers

1

- It can be hard to reach all members of a population, and it can take so long that circumstances change, affecting the responses. A well-designed sample is often a better choice.
- This sample is probably biased—people who didn't like the food at the restaurant might not choose to eat there.
- No, only the sample size matters, not the fraction of the overall population.
- Students who frequent this website might be more enthusiastic about statistics than the overall population of statistics students. A large sample cannot compensate for bias.

e. It's the population "parameter." "Statistics" describe samples.

2 

- a. systematic
- b. stratified
- c. simple
- d. cluster

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

do not reproduce.

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Chapter 9 Data Sources: Experiments



9.1 Randomized, Comparative Experiments

9.2 The Four Principles of Experimental Design

9.3 Experimental Designs

9.4 Issues in Experimental Design

9.5 Displaying Data from Designed Experiments

Capital One

Not everyone graduates first in their class at a prestigious business school. But even doing that won't guarantee that the first company you start will become a Fortune 500 company within a decade. Richard Fairbank managed to do both. When he graduated from Stanford Business School in 1981, he wanted to start his own company, but, as he said in an interview with the *Stanford Business Magazine*, he had no experience, no money, and no business ideas. So he went to work for a consulting firm. Wanting to be on his own, he left in 1987 and landed a contract to study the operations of a large credit card bank in New York. It was then that he realized that the secret lay in data. He and his partner, Nigel Morris,

asked themselves, “Why not use the mountains of data that credit cards produce to design cards with prices and terms to satisfy different customers?” But they had a hard time selling this idea to the large credit card issuers. At the time all cards carried the same interest rate—19.8% with a \$20 annual fee, and almost half of the population didn’t qualify for a card. And credit issuers were naturally resistant to new ideas.

Finally, Fairbank and Morris signed on with Signet, a regional bank that hoped to expand its modest credit card operation. Using demographic and financial data about Signet’s customers, they designed and tested combinations of card features that allowed them to offer credit to customers who previously didn’t qualify. Signet’s credit card business grew and, by 1994, was spun off as Capital One with a market capitalization of \$1.1B. Twenty years later, Capital One’s revenues were \$24.1B and its market capitalization was \$33.7B.

Fairbank also introduced “scientific testing.” Capital One designs experiments to gather data about customers. For example, customers who hear about a better deal than the one their current card offers may phone, threatening to switch to another bank unless they get a better deal. To help identify which potential card-hoppers were serious, Fairbank designed an experiment. When a card-hopper called, the customer service agent’s computer randomly ordered one of three actions: match the claimed offer, split the difference in rates or fees, or just say no. In that way the company could gather data on who switched, who stayed, and how they behaved. When a potential card-hopper phones, the computer could give the operator a script specifying the terms to offer—or instruct the operator to bid the card-hopper a pleasant good-bye.

Fairbank attributes the phenomenal success of Capital One to their use of such experiments. According to Fairbank, “Anyone in the company can propose a test and, if the results are promising, Capital One will rush the new product or approach into use immediately.” Why does this work for Capital One? Because, as Fairbank says, “We don’t hesitate because our testing has already told us what will work.”

Capital One weathered the financial crisis of 2008 better than most banks, selling their mortgage business and repaying the federal bailout by July 2009. In February 2012, it acquired ING Bank and in November announced that ING Direct would become Capital One 360. The firm was named by *Fortune* as one of the top 100 companies to work for in 2013. However, the previous July, Capital One was fined by the Consumer Financial Protection Bureau for “misleading millions of customers” about paying for extra services like payment protection and credit monitoring when receiving a new card. The company agreed to refund \$140M to customers and pay \$55M in fines.¹

1business.time.com/2012/07/19/cfpb-orders-capital-one-to-return-140-million-to-customers/

Fairbank started his analysis at Signet by looking for factors—variables in the data—that could help identify profitable customers. For this search he used the most convenient data he could find—the records of customers’ transactions and interactions with customer service that had already been collected. As we have seen, found data such as these transactional data can be useful as an initial stage in understanding business circumstances. As we saw in the last chapter, this type of

backward-looking search in data that have already been collected is an example of a retrospective observational study.

Retrospective observational studies are common in business because they are so practical. The data on which they are based are already available. But they make a weak basis for business decisions for a number of reasons. First, markets and customers are always changing, so the business environment may have changed since the data were first collected. Second, even if the world were completely stable, it would be impossible to be sure that the relationships seen between variables were a sound basis for action. Fairbank might have found, for example, that people who carry more than one credit card were more profitable than those who used a single card. But would this correlation imply that he should entice customers who carry only one card to get a second one? Couldn't it simply be true that people who spend more (and are thus more profitable to the bank) tend to have more than one card?

Although observational studies cannot give us the answer to questions like these, they can help to generate hypotheses. Fairbank did gain some insights from his observational study. But instead of conclusions on which he could make a sound decision, the study provided new hypotheses—theories that he could test as he took his next steps. To establish a causal relationship between factors that he could control (such as attributes of the cards or a willingness to extend credit to a new segment of the population) and identified responses (such as profitability, willingness to accept a card offer, or likelihood of failing to pay debts) he needed to collect new data. And the best way to generate data that can establish causal relationships is through designed, randomized experiments.

A retrospective observational study uses found data, but an experiment randomly assigns participants (in Fairbank's case, customers or potential customers) to particular treatments (such as card contract offers or even card colors or pictures). Random assignment is almost magic. It makes it possible to compare the responses among participants assigned to different treatments, and to conclude that those treatments caused the differences that were found.

Fairbank revolutionized the credit card industry through his use and advocacy of designed experiments at Capital One. In the 1980s, Capital One ran thousands of experiments on their customers, testing products and combinations of offers to determine which features caused the best results. This continues to this day. In 2011, the company opened Capital One Labs in San Francisco, and now has opened offices in New York and Washington, DC, as well. At these offices, the company can run experiments and test products using the latest technology.

You see the evidence of banking experiments in your mailbox every day. Nearly every credit card offer you receive is a “treatment” made up of different factors. The choices of percentage rate, benefits, and fees are often part of an experiment to determine

their impact on card acquisition. On the Internet, vendors may randomly send you to a different site than other potential customers, to test which of two website designs encourages more “click-through” to the purchasing pages or which combination of price and special offer sells best.

You might think that the success of Capital One and a few other prominent firms, which have used experiments to beat the competition, would have made experiments the norm in business, but this is not the case. Despite their well-publicized success, experiments have been slow to penetrate many businesses.

In fact, recent developments are encouraging firms to go in the wrong direction! The lure of Big Data entices many firms to rely on observational data to hunt for correlations among variables. Some articles that you may have seen (and many software vendors) even suggest that analyses of massive datasets are all you need as a basis for business decisions. But, of course, this is nonsense. More data does not change the fact that correlation does not imply causation. And a designed experiment is the surest way to understand causal relationships between factors and responses. An understanding of the value and effectiveness of designed experiments is one of the most important lessons of this course.

Just Checking

In early 2007, a larger-than-usual number of cats and dogs developed kidney failure; many died. Initially, researchers didn't know why, so they used an observational study to investigate.

1. Suppose that, as a researcher for a pet food manufacturer, you are called on to plan a study seeking the cause of this problem. Specify how you might proceed. Would your study be prospective or retrospective?

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

9.1 Randomized, Comparative Experiments

Experiments are a critical tool for understanding what products and ideas will work in the marketplace. An **experiment** ⓘ is a study in which the experimenter *manipulates* attributes of what is being studied and observes the consequences. Usually, the attributes, called **factors** ⓘ, are manipulated by being set to particular **levels** ⓘ and then allocated or assigned to individuals. An experimenter identifies at least one factor to manipulate and at least one response variable to measure. Observed **responses** ⓘ can be quantitative measurements such as the amount of a product sold or categorical responses like whether the customer purchased the product or not. (However, we will restrict attention to quantitative responses in this text.) The combination of factor levels assigned to a subject is called that subject's **treatment** ⓘ.

The individuals on whom or which we experiment are known by a variety of terms. Humans who are experimented on are commonly called **subjects** ⓘ or **participants** ⓘ. Other individuals (rats, products, fiscal quarters, company divisions) are commonly referred to by the more generic term **experimental units** ⓘ.

You have probably already been the subject of marketing experiments. Every credit card offer you receive is actually a combination of various factors that specify a “treatment” assigned to you. It defines the specific offer you get. For example, at Capital One, the factors might include *Annual Fee*, *Interest Rate*, and *Communication Channel* (how the offer is delivered to you: by e-mail, direct mail, phone, etc.). A customer's treatment might be the combination of *no Annual Fee* and a *moderate Interest Rate* with the offer being sent by *e-mail*. Other customers receive different treatments. The possible responses might be categorical (do you accept the offer of that card?) or quantitative (how much do you spend with that card during the first three months you have it?).

Two key features distinguish an experiment from observational studies and other types of investigations. First, the experimenter actively and deliberately manipulates the factors to specify the treatment. Second, the experiment assigns the subjects to those treatments at *random*. The importance of **random assignment** ⓘ may not be immediately obvious. Experts, such as business executives and physicians, may think that they know how different subjects will respond to various treatments. In particular, marketing executives may want to send what they consider the best offer to their best customers, but this would make fair comparisons of treatments impossible. Without random assignment, we can't perform the analyses that allow us to conclude that differences among the treatments were responsible for any differences we observed in the responses. By using random assignment to ensure that the groups receiving different treatments are comparable, the experimenter can be sure that these differences are *due* to the differences in treatments. There are many stories of experts who were certain they knew the effect of a treatment and were proven wrong by a properly designed study. In business, it is important to get the facts rather than to just rely on what you may think you know from experience.

In Practice 9.1 Gaining insight from a marketing experiment

Manager You told me that most young riders of the Acela train use the Internet while on board (see **In Practice 8.1** ⓘ). I am looking for ways to get more young people to ride the Acela. Can you analyze

our customer database to figure out what to offer them?

Analyst We cannot use our customer database to answer your question because everyone has been treated the same way. We need to conduct a prospective designed experiment to see if the young riders can be induced to ride more often if we provide incentives.

I know that we recently purchased a mailing list of 16,000 college students. I've designed an experiment to manipulate the type of offer (the *factor*) at four different levels, each sent to 25% of the mailing list. The levels are (names in italics): (1) 10% off coupon for their next Acela ride (*Coupon*), (2) a 5000-mile Amtrak bonus card (*Card*), (3) a free Netflix download during their next Acela trip (*Movie*), and (4) no special offer (*No Offer*). We will monitor the four groups to see which group travels the most during the 12 months after sending the offer, and we will use *Miles Traveled* over the 12 months as the response variable. I'll use the results to assess whether the cost of the incentive is justified by their increased ridership.

9.2 The Four Principles of Experimental Design

When designing an experiment, there are four **principles of experimental design** ⓘ to keep in mind:

1. **Control.** Controlling external sources of variation helps ensure that any differences seen are due to the experimental factors being tested and not extraneous changes in the environment. Environmental conditions can't be controlled in the usual sense, but, by making conditions as similar as possible for all treatment groups, the experimenter reduces the influence of other changes. In a test of a new credit card, the effect of a possible change to the economic environment can be controlled by sending all offers to customers at the same time and in the same manner. Otherwise, events like a sudden spike or drop in gas prices or a terrorist attack could influence customers in different ways, making it difficult to distinguish the effects of the treatments from the other effects. By controlling extraneous sources of variation the experimenter reduces the variability of the responses, making it easier to discern differences among the treatment groups.
There is a second meaning of control in experiments. A bank testing the new creative idea of offering a card with special discounts on chocolate to attract more customers will want to compare its performance against one of their standard cards. Such a baseline measurement is called a control treatment, and the group that receives it is called the **control group** ⓘ.
2. **Randomize.** In any true experiment, subjects are assigned treatments at random. Randomization allows us to equalize the effects of unknown or uncontrollable sources of variation even when they aren't controlled. Although randomization can't eliminate the effects of these sources, it spreads them out across the treatment levels so that we can see past them. Randomization also makes it possible to use the powerful methods of inference to draw conclusions from your study. Randomization protects us even from effects we didn't know about. Perhaps women are more likely to respond to the chocolate benefit card. We don't need to test equal numbers of men and women—our mailing list may not have that information—but if we randomize, that tendency won't contaminate our results. There's an adage that says "Control what you can, and randomize the rest."
3. **Replicate.** Replication shows up in different ways in experiments. Because we need to estimate the variability of our measurements, we must make more than one observation at each level of each factor. Sometimes that just means making repeated observations. But, as we'll see later, some experiments combine two or more factors in ways that may permit a single observation for each *treatment*—that is, each combination of factor levels. When such an experiment is repeated in its entirety, it is said to be *replicated*. Repeated observations at each treatment are called **replicates** ⓘ. If the number of replicates is the same for each treatment combination, we say that the experiment is **balanced** ⓘ. An important advantage of balanced designs, especially those with more than one factor, is the ease of interpreting the effects. Unbalanced designs with more than one factor are more complicated, and their analysis is beyond the scope of this text.

A second kind of replication is to repeat the entire experiment for a different group of subjects, under different circumstances, or at a different time. Experiments do not require, and often can't obtain, representative random samples from an identified population. Experiments study the consequences of different levels of their factors. They rely on the random assignment of treatments to the subjects to generate the sampling distributions and to control for other possibly contaminating variables. When

we detect a significant difference in response among treatment groups, we can conclude that it is due to the difference in treatments. However, we should take care in generalizing that result too broadly if we've only studied a specialized population. A special offer of accelerated check-out lanes for regular customers may attract more business in December, but it may not be effective in July. Replication in a variety of circumstances increases our confidence that our results apply to other situations and populations.

4. **Block.** Sometimes we can identify a factor not under our control whose effect we don't care about, but which we suspect might have an effect either on our response variable or on the ways in which the factors we are studying affect that response. Perhaps men and women will respond differently to our chocolate offer. Or maybe customers with young children at home behave differently than those without. Platinum card members may be tempted by a premium offer much more than standard card members. Factors like these can account for some of the variation in our observed responses because subjects at different levels respond differently. But we can't *assign* them at random to subjects. So we deal with them by grouping, or **blocking** ⓘ, our subjects together and, in effect, analyzing the experiment accounting for each block. Such factors are called **blocking factors** ⓘ, and their levels are called **blocks**. Blocking in an experiment is like stratifying in a survey design. Blocking reduces variation by comparing subjects within these more homogeneous groups. That makes it easier to discern any differences in response due to the factors of interest. In addition, we may want to study the effect of the blocking factor itself. Blocking is an important compromise between randomization and control. However, unlike the first three principles, blocking is not required in all experiments.

Just Checking

Following concerns over the contamination of its pet foods by melamine, which had been linked to kidney failure, a manufacturer now claims its products are safe. You are called on to design the study to demonstrate the safety of the new formulation.

2. Identify the treatment and response.
3. How would you implement control, randomization, and replication?

In Practice 9.2 Understanding experimental design principles

Manager Can you explain to me how we know that your experiment about young riders will get us the information we want?

Analyst Of course. The experiment follows the three required principles of experimental design (blocking is optional). Specifically,

Control: It is impossible to control other factors that may influence a person's decision to use the Acela. I decided to restrict attention to college and university students, and I sent all offers out at the same time to reduce variation. I also used a control group—one that receives no offer—which will be used to compare to the other three treatment levels.

Randomization: Although I can't control the other factors (besides *Offer*) that may influence a person's decision to use the Acela, by randomizing which students receive which offer, I hope that the influences of all those other factors will average out, enabling me to see the "pure" effect of the four treatments.

Replication: I've replicated by sending each type of offer to 4000 students. I hope that the response is high enough that will be able to see differences in *Miles Traveled* among the groups. This experiment is balanced, since the number of subjects is the same for all four treatments.

Blocking: I have not used blocking in the experiment. Possible blocking factors might include demographic variables such as the region of the student's home or college, their sex, or their parent's income, but I have not used those factors in the design.

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

9.3 Experimental Designs

One Factor Designs

The most common experimental design used in business is the single factor experiment with two levels. They are widely used and have many names including **A/B tests**, **bucket tests** or *split-run tests*. Sometimes they are called **champion/challenger designs** when they're used to test a new idea (the challenger) against the current version (the champion). In this case, the customers offered the champion are the control group, and the customers offered the challenger (a special deal, a new offer, a new service, etc.) are the test group. A diagram of this design is shown in **Figure 9.1**. The single factor design is an example of a **completely randomized design**—a design where all the treatment combinations are assigned to the participants completely at random.

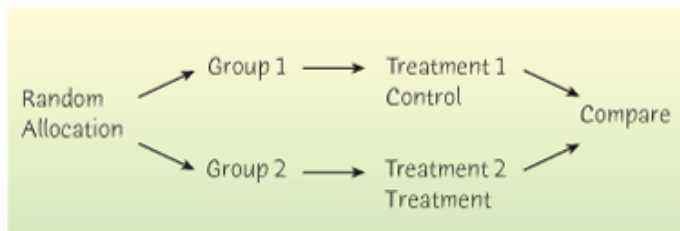


Figure 9.1

The simplest randomized design has two groups randomly assigned two different treatments.

Randomized Block Designs

It is not always possible to have completely random assignment of experimental units to treatments because some factors cannot be randomly assigned. For example, we can't randomly assign factors based on people's behavior, age, sex, and other attributes. Nonetheless, if we believe that these factors are important, we can break the population into groups based on the uncontrollable factor. The resulting groups are called blocks. Once we have a blocking factor, we then randomize subjects to the treatments *within each block*. This is called a **randomized block design**. In the experiment diagrammed in **Figure 9.2**, a marketer wanted to know the effect of two types of offers in each of two segments: a high spending group and a low spending group. The marketer selected 12,000 customers *from each group* at random and then randomly assigned the three treatments to the 12,000 customers *in each group* so that 4000 customers in each segment received each of the three treatments. A display makes the process clearer.

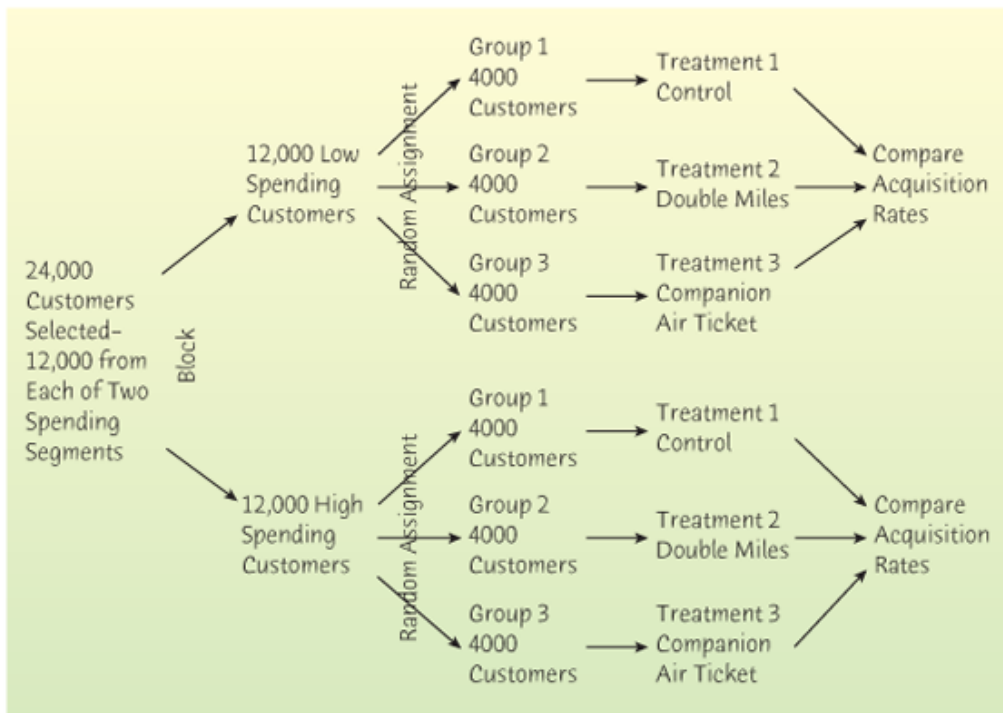


Figure 9.2

This example of a randomized block design shows that customers are randomized to treatments within each segment, or block.

Multi-Factor Designs

An experiment with more than one factor is called a **factorial design**. A full factorial design contains treatments that represent all possible combinations of all levels of all factors. Each specific combination of factors is called a treatment. The number of treatments can quickly become very large. With only three factors, one at 3 levels, one at 4, and one at 5, there would be $3 \times 4 \times 5 = 60$ different treatments.

Because of the possibility for huge numbers of treatments (and the requirement for very large numbers of experimental units, if we want appropriate replication), researchers typically limit the number of levels to just a few. Alternatively, advanced experimental designs allow researchers to use fewer treatments at the cost of some information loss.

It may seem that the added complexity of multiple factors is not worth the trouble. In fact, just the opposite is true. Testing multiple factors in a single experiment makes more efficient use of the available subjects. And testing factors together is the only way to see what happens at *combinations* of the levels.

An experiment to test the effectiveness of offering a \$150 rebate on a new credit card may find that it increases customer spending by 1%. Another experiment finds that lowering the interest rate increases spending by 2%. But unless some customers were offered *both* the \$150 rebate *and* the lower interest rate, the analyst can't learn whether offering both together would lead to still greater spending or less.

When the combination of two factors has a different effect than you would expect by adding the effects of the two factors together, that phenomenon is called an **interaction** . If the experiment does not contain both

factors, it is impossible to see interactions. That can be a major omission because such effects can have the most important and surprising consequences in business.

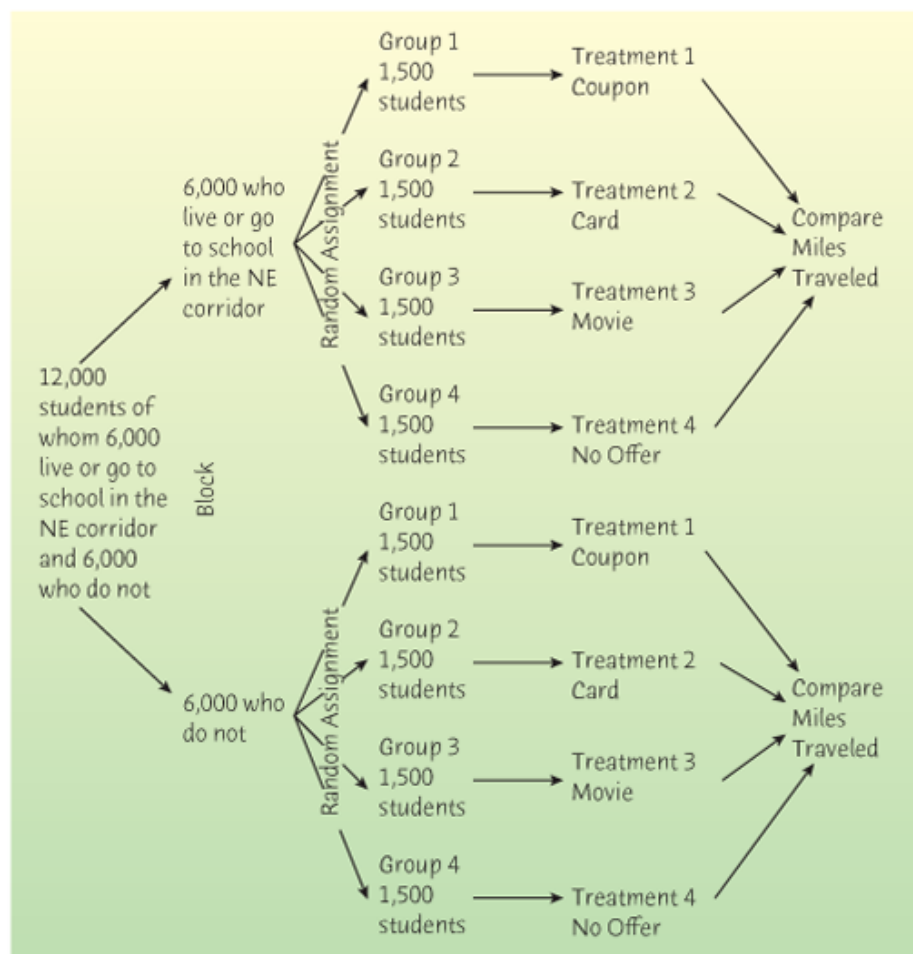
In Practice 9.3 Designing an experiment

Continuing In Practice 9.2 (page 285), you are considering splitting up the students to be experimented on into two groups before mailing the offers: those who live or go to school in the Northeast corridor, where the Acela operates, and those who don't. Using home and school zip codes, you split the original 12,000 students into those groups and find that 6000 live or go to school in the Northeast corridor and 6000 do not.* You plan to randomize the treatments within those two groups and you'll monitor them to see if this factor, *NE corridor*, affects their *Miles Traveled* as well as the type of offer they receive.

*There is no requirement that the block sizes must be equal. However, an analysis with unequal block sizes is more complicated.

Manager Can you explain your design to me in words and/or pictures?

Analyst I want to test the three incentive: *Coupon*, *Card*, and *Movie*, against what we currently do, which is none of the above. So I'll randomly send those three incentives to the same number of students and use a fourth group as a control. But first I'm going to split the students into two blocks, those who live or go to school in the NE corridor and those who don't. I could just restrict the experiment to the first group, but I want to see how well the incentives work (if at all) on the others. I'll treat that variable, *NE corridor*, as a blocking variable. Here's a diagram explaining the design:



Guided Example Designing a Direct Mail Experiment



At a major credit card bank, management has been pleased with the success of a recent campaign to cross-sell Silver card customers with the new SkyWest Gold card. But you, as a marketing analyst, think that the revenue of the card can be increased by adding three months of double miles on SkyWest to the offer, and you think the additional gain in charges will offset the cost of the double miles. You want to design a marketing experiment to find out what the difference will be in revenue if you offer the double miles. You've also been thinking about offering a new version of the miles called "use anywhere miles," which can be transferred to other airlines, so you want to test that version as well.

You also know that customers receive so many offers that they tend to disregard most of their direct mail. So, you'd like to see what happens if you send the offer in a shiny gold envelope with the SkyWest logo prominently displayed on the front. How can we design an experiment to see whether either of these factors has an effect on charges?

Define and state the problem.	We want to study two factors to see their effect on the revenue generated for a new credit card offer.
Response Specify the response variable.	Revenue is a percentage of the amount charged to the card by the cardholder. To measure the success, we will use the monthly charges of customers who receive the various offers. We will use the three months after the offer is sent out as the collection period and use the total amount charged per customer during this period as the response.
Factors Identify the factors you plan to test.	We will offer customers three levels of the factor <i>Miles</i> for the SkyWest Gold card: <i>Regular</i> (no additional) <i>Miles</i> , <i>Double Miles</i> , and <i>Double Use Anywhere Miles</i> . Customers will receive the offer in the <i>Standard Envelope</i> or the new <i>SkyWest Logo Envelope</i> (factor <i>Envelope</i>).
Levels Specify the levels of the factors you will use.	We will send out all the offers to customers at the same time (in mid September) and evaluate the response as total charges in the period October through December.
Experimental Design Observe the principles of design:	A total of 30,000 current Silver card customers will be randomly selected from our customer records to receive one of the six offers.
Control any sources of variability you know of and can control.	

Randomly assign experimental units to treatments to equalize the effects of unknown or uncontrollable sources of variation.

Replicate results by placing more than one customer (usually many) in each treatment group.

Make a Picture A diagram of your design can help you think about it.

Specify any other details about the experiment. You must give enough details so that another experimenter could exactly replicate your experiment.

It's generally better to include details that might seem irrelevant because they may turn out to make a difference.

Specify how to measure the response.

✓ *Regular Miles with Standard Envelope*

✓ *Double Miles with Standard Envelope*

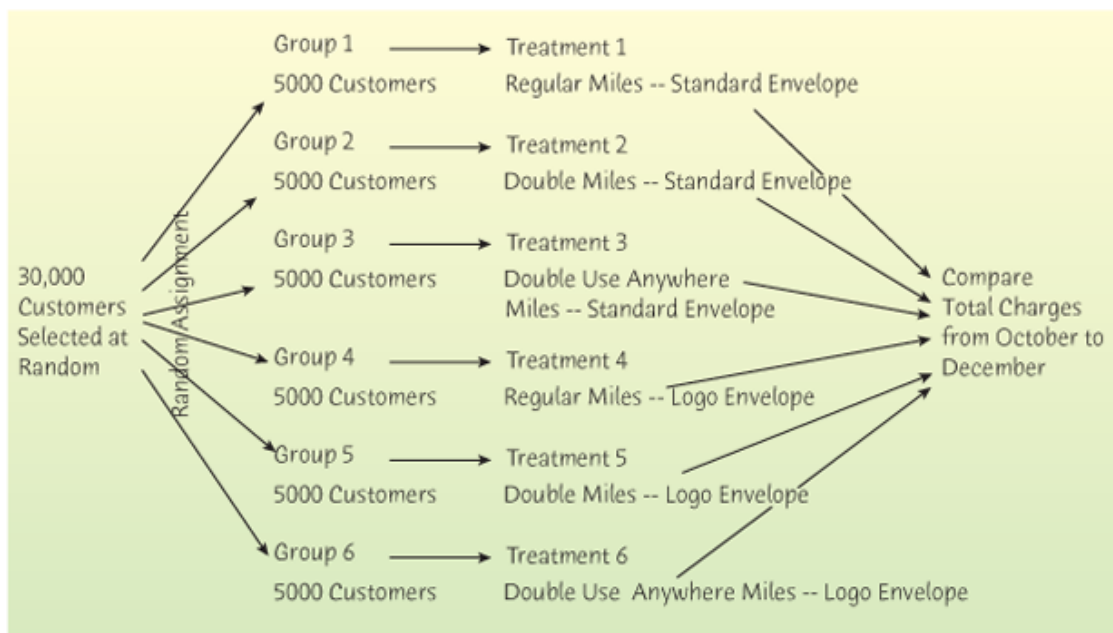
✓ *Double Use Anywhere Miles with Standard Envelope*

✓ *Regular Miles with Logo Envelope*

✓ *Double Miles with Logo Envelope*

✓ *Double Use Anywhere Miles with Logo Envelope*

On January 15, we will examine the total card charges for each customer for the period October 1 through December 31.



Write up the proposal to carry out the design and collect the data. At this stage of planning the do step is to get authorization to implement your proposal.

Communicate the basic outline of your proposal. Include a brief description of the experimental design and a plan for its implementation. Discuss how the experiment will address the original problem you set out to solve.

MEMO

Re: Test mailing for creative offer and envelope

We designed an experiment to test whether several creative ideas from marketing will generate enough revenue to justify their cost. We tested two factors: *Envelope* at two levels (*Logo* and *Standard*) and *Miles* at three levels (*Regular*, *Double*, *Use Anywhere*) to create a 2×3 factorial design. We will select 30,000 current Silver card members, who will be randomly assigned to one of the 6 cells of the design. We will evaluate the results using cumulative revenue for the 3 months post-mailing, which was sent out. When the results are in, we will reconvene to analyze the results. Of particular interest are:

✓ Whether offering *Double Miles* is worth the cost of the miles

✓ Whether the *Double Use Anywhere Miles* are worth the cost

✓ Whether the *Logo Envelope* increased spending enough to justify the added expense

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

9.4 Issues in Experimental Design

Blinding and Placebos

Humans are notoriously susceptible to errors in judgment—all of us. When we know what treatment is assigned, it's difficult not to let that knowledge influence our response or our assessment of the response, even when we try to be careful.

Suppose you were trying to sell your new brand of cola to be stocked in a school's vending machines. You might hope to convince the committee designated to make the choice that students prefer your less expensive cola, or at least that they can't taste the difference. You could set up an experiment to see which of the three competing brands students prefer (or whether they can tell the difference at all). But people have brand loyalties. If they know which brand they are tasting, it might influence their rating. To avoid this bias, it would be better to disguise the brands as much as possible. This strategy is called **blinding** the participants to the treatment. Even professional taste testers in food industry experiments are blinded to the treatment to reduce any prior feelings that might influence their judgment.

But it isn't just the subjects who should be blind. Experimenters themselves often subconsciously behave in ways that favor what they believe. It wouldn't be appropriate for you to run the study yourself if you have an interest in the outcome. People are so good at picking up subtle cues about treatments that the best (in fact, the only) defense against such biases in experiments on human subjects is to keep anyone who could affect the outcome or the measurement of the response from knowing which subjects have been assigned to which treatments. So, not only should your cola-tasting subjects be blinded, but also you, as the experimenter, shouldn't know which drink is which—at least until you're ready to analyze the results.

Blinding by Misleading

Social science experiments can sometimes blind subjects by disguising the purpose of a study. One of the authors participated as an undergraduate volunteer in one such (now infamous) psychology experiment. The subjects were told that the experiment was about 3-D spatial perception and were assigned to draw a model of a horse and were randomly assigned to a room alone or in a group. While they were busy drawing, a loud noise and then groaning were heard coming from the room next door. The *real* purpose of the experiment was to see whether being in a group affects how people reacted to the apparent disaster. The horse was only a pretext. The subjects were blind to the treatment because they were misled.

There are two main classes of individuals who can affect the outcome of the experiment:

- Those who could influence the results (the subjects, treatment administrators, or technicians)
- Those who evaluate the results (judges, experimenters, etc.)

When all the individuals in either one of these classes are blinded, an experiment is said to be **single-blind** ⓘ. When everyone in both classes is blinded, we call the experiment **double-blind** ⓘ. Double-blinding is the gold standard for any experiment involving both human subjects and human judgment about the response. Double-blind studies are the gold standard for a reason: You cannot overcome the subtle influence that results from experimenter knowledge of the treatments because it is subconscious, and the influence is quite significant. An experiment showed, for example, that the mere knowledge on the part of dentists that their (actually identically treated) patients might have received a placebo before tooth removal massively increased their patient's pain 1.5 hours post-procedure (Gracely et al., 1985, *Lancet*)!

Placebos and Authority

The placebo effect is stronger when placebo treatments are administered with authority or by a figure who appears to be an authority. "Doctors" in white coats generate a stronger effect than salespeople in polyester suits. But the placebo effect is not reduced much, even when subjects know that the effect exists. People often suspect that they've gotten the placebo if nothing at all happens. So, recently, drug manufacturers have gone so far in making placebos realistic that they cause the same side effects as the drug being tested! Such "active placebos" usually induce a stronger placebo effect. When those side effects include loss of appetite or hair, the practice may raise ethical questions.

Often simply applying *any* treatment can induce an improvement. Every parent knows the medicinal value of a kiss to make a toddler's scrape or bump stop hurting. Some of the improvement seen with a treatment—even an effective treatment—can be due simply to the act of treating. To separate these two effects, we can sometimes use a control treatment that mimics the treatment itself. A "fake" treatment that looks just like the treatments being tested is called a **placebo** ⓘ. Placebos are the best way to blind subjects so they don't know whether they have received the treatment or not. One common version of a placebo in drug testing is a "sugar pill." Especially when psychological attitude can affect the results, control group subjects treated with a placebo may show an improvement.

The fact is that subjects treated with a placebo sometimes improve. It's not unusual for 20% or more of subjects given a placebo to report reduction in pain, improved movement, or greater alertness or even to demonstrate improved health or performance. This **placebo effect** ⓘ highlights both the importance of effective blinding and the importance of comparing treatments with a control. Placebo controls are so effective that you should use them as an essential tool for blinding whenever possible.

Just Checking

The pet food manufacturer we've been following hires you to perform the experiment to test whether their new formulation is safe and nutritious for cats and dogs.

4. How would you establish a control group?
5. Would you use blinding? How? (Can or should you use double-blinding?)
6. Both cats and dogs are to be tested. Should you block? Explain.

The best experiments are usually:

- Randomized
- Double-blind
- Comparative
- Placebo-controlled

Confounding

A credit card bank wanted to test the sensitivity of the market to two factors: the annual fee charged for a card and the annual percentage rate charged. The bank selected 100,000 people at random from a mailing list and sent out 50,000 offers with a low rate and no fee and 50,000 offers with a higher rate and a \$50 annual fee. They discovered that people preferred the low-rate, no-fee card. No surprise. In fact, customers signed up for that card at over twice the rate as the other offer. But the question the bank really wanted to answer was: “How much of the change was due to the rate, and how much was due to the fee?” Unfortunately, there’s simply no way to separate out the two effects with that experimental design.

If the bank had followed a factorial design in the two factors and sent out all four possible different treatments—low rate with no fee; low rate with \$50 fee; high rate with no fee, and high rate with \$50 fee—each to 25,000 people, it could have learned about both factors and could have also learned about the interaction between rate and fee. But we can’t tease apart these two effects because the people who were offered the low rate were also offered the no-fee card. Whenever the effects of two (or more) variables on a response cannot be distinguished from each other, we say that the two factors are **confounded** .

Confounding can arise either because of lurking variables in an observational study experiment, or because of poor design in an experiment. When variables are confounded it can be difficult to know which variable is really responsible for the effect. A shock to the economic or political situation that occurs during a marketing experiment can overwhelm the effects of the factors being manipulated. In designed experiments, randomization will usually take care of confounding by distributing uncontrolled factors over the treatments at random. But be sure to watch out for potential confounding effects even in a well-designed experiment.

Just Checking

A researcher wanted to test how much familiarity helps memory. He designed an experiment and gave subjects (sophomores in a U.S. university) the following two groups of cities to memorize. He recorded the number of correct responses.

A: Paris, Rome, London, New York, Moscow

B: Bandar Seri Begawan, Nouakchott, Ulaanbaatar, Podgorica, Jayawardenepura Kotte

7. What other factor(s) did the researcher unwittingly confound with familiarity?

Personal use only
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

9.5 Displaying Data from Designed Experiments

Does the Difference Make a Difference?

A coworker has a new design for your startup's webpage that he claims will keep people more interested in your company than the old design. He decides to run an A/B test, sending half of people searching for your company to the old site (A) and half to his new design (B). After a week, you examine the results and find that the average amount of time for A is 80 seconds. The new design (B) resulted in an average of 90 seconds.

That sounds good. But was that increase really due to the new website, or just to random variation? If the website has only a small effect, it will be hard to tell the difference. If the amount of time people spend on your site is highly variable, that makes it hard to tell the difference as well. And if you try to make a decision based on only on a few subjects, that makes things even worse. These three factors—size of the effect, variability, and the number of runs—all affect your ability to draw conclusions.

Observations vary. Suppose we display the results using boxplots. **Figure 9.3** shows two possible pairs where the response variable is *Amount of Time* spent on the site. In the left pair, suppose the boxplot in green shows the results for the green A site and the orange boxplot shows results for site B. It looks pretty convincing. Now look at the pair on the right. The actual mean difference (10 seconds) is the same but there's so much variability in the times that the effect is hard to see.² Later, we'll see how to use the standard deviation of the difference as a measure to judge whether the difference is large enough to be fairly sure that it didn't occur just by natural variation. When an observed difference is large enough in standard deviations to be convincing, we say it's statistically significant. We'll put numbers on this concept later, but we'd say the difference is statistically significant in the left pair and not statistically significant in the right pair.

²Of course the boxplots show medians at their centers, and we're trying to find differences among means. But for roughly symmetric distributions like these, the means and medians are very close.

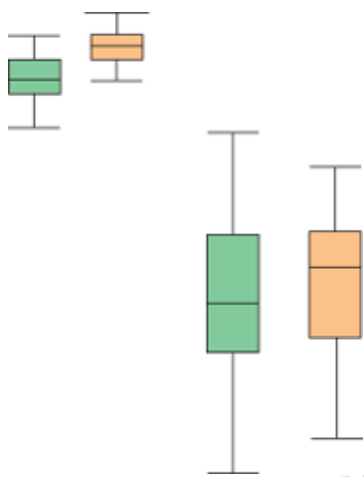


Figure 9.3

The boxplots in both pairs have centers the same distance apart, but when the spreads are large, the observed difference may be just from random fluctuation.

Later chapters show how statistical tests quantify this intuition. For now, the important point is that a difference is statistically significant if we don't believe that it's likely to have occurred only by chance.

Boxplots are great for visualizing changes in a response for different levels of a single quantitative factor, but when the experiment involves more than one factor, we have to be careful to think about all the factors together.

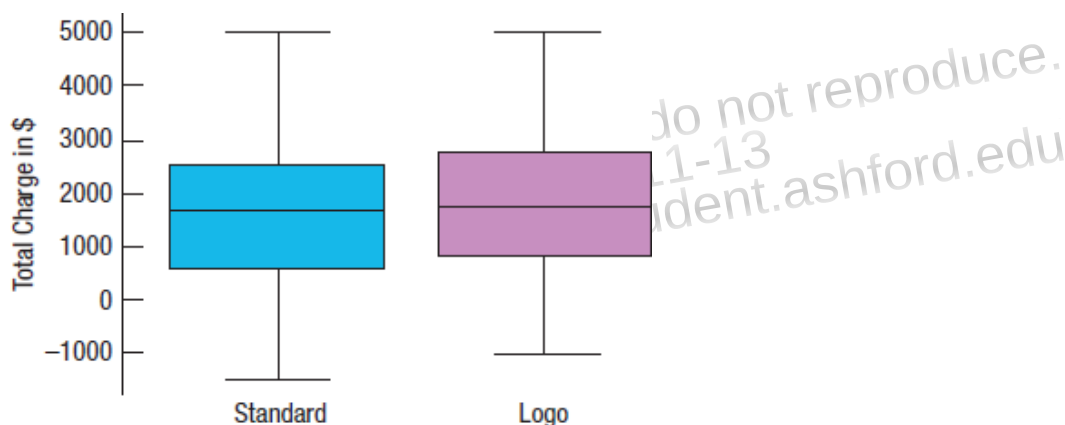
In our direct mail example from the Guided Example on page 289, we looked at two factors: *Miles* and *Envelope*. *Miles* had three levels: *Regular Miles*, *Double Miles*, and *Double Use Anywhere Miles*. The factor *Envelope* had two levels: *Standard* and new *Logo*. The three levels of *Miles* and the two levels of *Envelope* resulted in six treatment groups. Because this was a completely randomized design, 5000 customers were allocated at random to each treatment.

Three months after the offer was mailed out, the total charges on the card were recorded for each of the 30,000 cardholders in the experiment. The mean *Total Charges* for each treatment combination are shown in Table 9.1. Figure 9.4 shows boxplots of the six treatment groups' responses for each factor.

Table 9.1

Mean *Total Charges* by *Treatment* for the direct mail example on page 289. (Data in **Direct mail**)

		Envelope		
		Standard	Logo	Means
Miles	No Offer	1611	1762	1687
	Double Miles	1716	1874	1795
	Use Anywhere	1795	1979	1887
	Means	1707	1872	1789



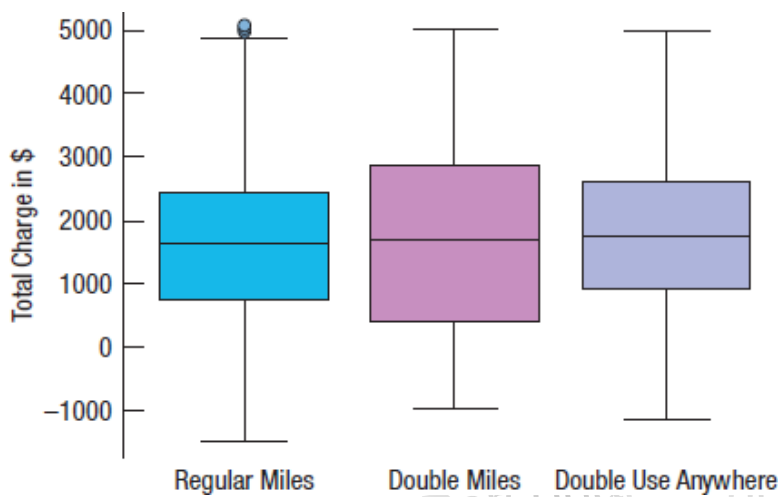


Figure 9.4

Boxplots of *Total Charge* by each factor. It is difficult to see the effects of the factors for several reasons. First, the other factor hasn't been accounted for; second, the effects are small compared to the overall variation in charges; third, an interaction between the predictor variables may mask the main effects; and finally, watch out for outliers in each treatment combination.

If you look closely at [Figure 9.4](#), you may be able to discern a very slight increase in the *Total Charges* for some levels of the factors, but it's very difficult to see. There are several reasons for this. First, the variation due to each factor gets in the way of seeing the effect of the other factor. For example, each customer in the boxplot for the *Logo Envelope* got one of three different offers. If those offers had an effect on spending, then that increased the variation within the *Logo* treatment group. Second, as is typical in a marketing experiment of this kind, the effects are very small compared to the variability in people's spending. That's why companies use such a large sample size. And that's why we'll study efficient methods for determining important effects later in the text. Third, the main effects of the factors in a multi factor experiment may be masked by the interaction effect. We'll see how to display the interaction in the next section. Finally, watch out for outliers. It's a good idea to display the data in each cell using box plots or histograms to check for outliers.

Interaction Plots

Whenever we have more than one factor, we should investigate whether the effect of one factor is constant for all levels of the other. In other words, we should explore whether an interaction between the two factors affects the response. An **interaction plot**, as shown in [Figure 9.5](#), is a plot of means for each treatment group. It is essential for visualizing interactions.

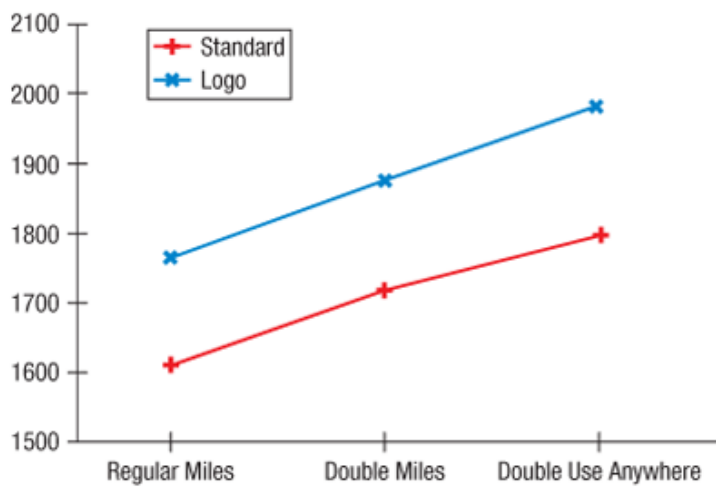


Figure 9.5

An interaction plot of the *Miles* and *Envelope* effects. The parallel lines show that the effects of the three *Miles* offers are roughly the same over the two different *Envelopes* and therefore that the interaction effect is small.

This plot displays means of the response at each treatment combination. The levels of one factor are shown on the x-axis. A separate line for each level of another factor connects the treatment means. When the effect of one factor on the response depends on the level of the other factor, the lines will not be parallel, giving evidence of an interaction between the factors.

This interaction plot shows the mean *Charges* at all six treatment groups. The levels of one of the factors, in this case *Miles*, are shown on the x-axis, and the mean *Charges* of the groups for each *Envelope* level are shown at each *Miles* level. The means of each level of *Envelope* are connected for ease of understanding. Notice that the effect of *Double Miles* over *Regular Miles* is about the same for both the *Standard* and *Logo* *Envelopes*. And the same is true for the *Double Use Anywhere* miles. This indicates that the effect of *Miles* is constant for the two different *Envelopes*. The lines are nearly parallel, which indicates that there is essentially no interaction effect.

To create the interaction plot in [Figure 9.5](#), start with the table of means as in [Table 9.1](#). Choose one factor to be on the x-axis. Here we have chosen *Miles*. Choose one level of the other factor to begin, say *Standard*, for *Envelope*. Plot the means for *Standard* for each level of *Miles*, connecting the points with a line. Here we've chosen red for *Standard*. Repeat with a different color for each level. Here there is only one other level, *Logo*, for which we've chosen blue. Many statistics packages offer interaction plots as part of advanced modeling methods such as multiple regression and ANOVA, but for now, it's important simply to understand what an interaction effect is and how to interpret interaction plots.

Later in the text we'll learn how to judge whether effects are statistically significant, meaning that they are enough so that we can be reasonably sure that the differences we see aren't just due to natural variation.

Just Checking

Suppose that you had run the randomized block design from [In Practice 9.3](#) (see page 288). You would have had two levels of the (blocking) factor *NE Corridor* (*NE* or *not*) and the same four levels of *Offer* (*Coupon*, *Card*, *Movie*, and *No Offer*).

8. What would a significant interaction effect between *NE Corridor* and *Offer*. mean? What would you tell the marketing group?

A marketing group at a small phone company designs an experiment to test two factors: 1. Data (100Mb or Unlimited). 2. Overseas calls (standard rates or free). They offer the 4 treatment combinations at the same price randomly to 10,000 customers who go to their website. Historically about 4 to 5% of customers looking at their website sign up for one of the plans. During the experiment they find that 5% of the customers opt for the 100Mb standard rate combination, 6% for the Unlimited data, standard rate combination, 5% for the 100Mb free international plan, and 30% for the Unlimited data, free international calls plan.

9. Draw an interaction plot and explain the results of the experiment to the marketing group.

Guided Example A Follow-up Experiment



After analyzing the data, the bank decided to go with the *Logo* envelope, but a marketing specialist thought that more *Miles* might increase spending even more. A new test was designed to test both the type of *Miles* and the amount. Again, total *Charge* in three months is the response. (Data in **Direct mail follow up**)

State the problem. Response Specify the response variable. Factors Identify the factors you plan to test. Levels Specify the levels of the factors you will use. Experimental Design Specify the design. Make a Picture A diagram of your design can help you think about it. We could also draw this diagram like the one on page 290 with 6 treatment groups, but now we are thinking of the design as having two distinct factors that we wish to evaluate	We want to study the two factors <i>Miles</i> and <i>Amount</i> to see their effect on the revenue generated for a new credit card offer. To measure the success, we will use the monthly charges of customers who receive the various offers. We will use the three months after the offer is sent out as the collection period and the total amount charged per customer during this period as the response. We will offer each customer one of the two levels of the factor <i>Miles</i> for the SkyWest Gold card: <i>SkyWest Miles</i> or <i>Use Anywhere Miles</i>. Customers are offered three levels of <i>Miles</i>: <i>Regular Miles</i>, <i>Double Miles</i>, and <i>Triple Miles</i>.
---	--

individually, so this form gives the right impression.

Specify any other experimental details. You must give enough details so that another experimenter could exactly replicate your experiment.

It's generally better to include details that might seem irrelevant than to leave out matters that could turn out to make a difference.

Specify how to measure the response and your hypotheses.

We will send out all the offers to customers at the same time (in mid March) and evaluate the response as total charges in the period April through June.

A total of 60,000 current Gold card customers will be selected from our customer records to be randomly assigned to receive one of the six offers.

✓Regular SkyWest Miles

✓Double SkyWest Miles

✓Triple SkyWest Miles

✓Regular Use Anywhere Miles

✓Double Use Anywhere Miles

✓Triple Use Anywhere Miles

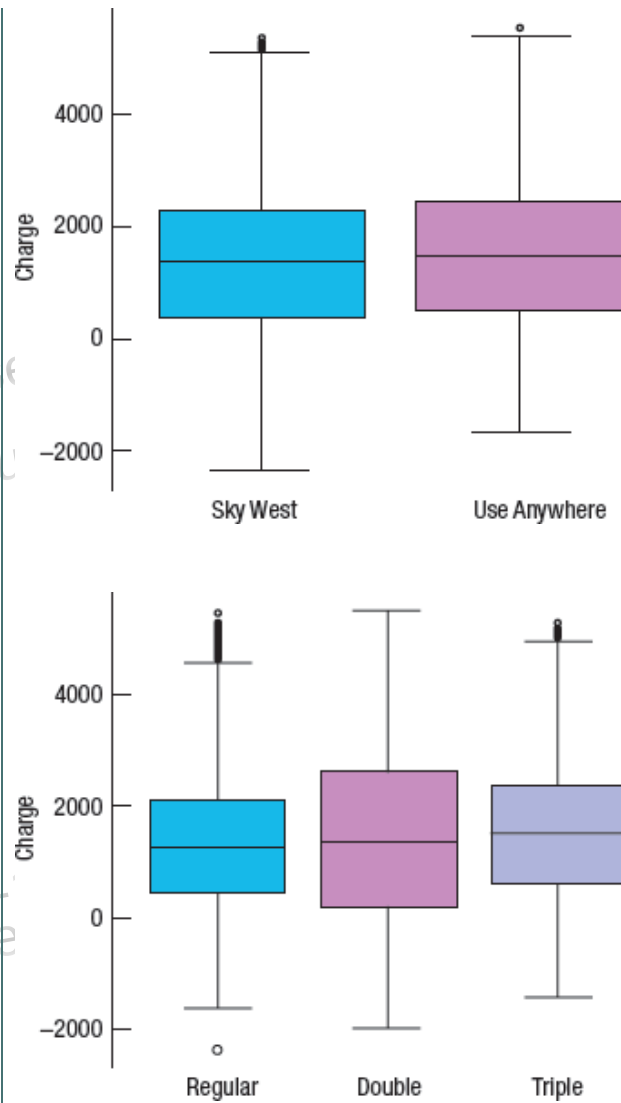


On June 15, we will examine the total card charges for each customer for the period April 1 through June 30.

We want to measure the effect of the two types of Miles and the three award amounts.

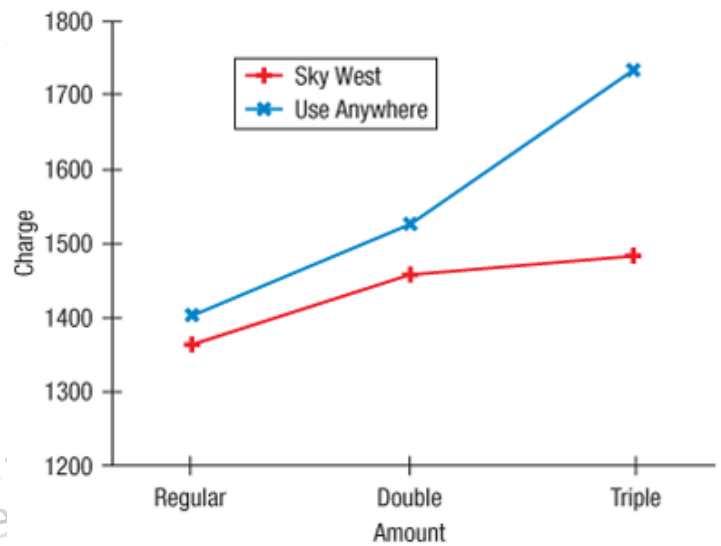
Plot Examine the boxplots and interaction plots, being careful to realize that an interaction effect may make the interpretation of the main effects problematic. Always be sure to examine the interaction plot before making

conclusions about which factors influence the response.



Boxplots by each factor show that there may be a slight increase in charges due to the *Use Anywhere Miles* and the *Amount of Miles* offered, but the differences are hard to see because of the intrinsic variation in *Charges*.

There are some outliers apparent in the boxplots, but none exerts a large influence on its group mean, so we will leave them in.



The interaction plot shows that offering *Triple Miles* may have a much larger effect for *Use Anywhere Miles* than for *Sky West Miles*.

To answer the initial question, we ask whether the differences we observe in the means of the groups are meaningful.

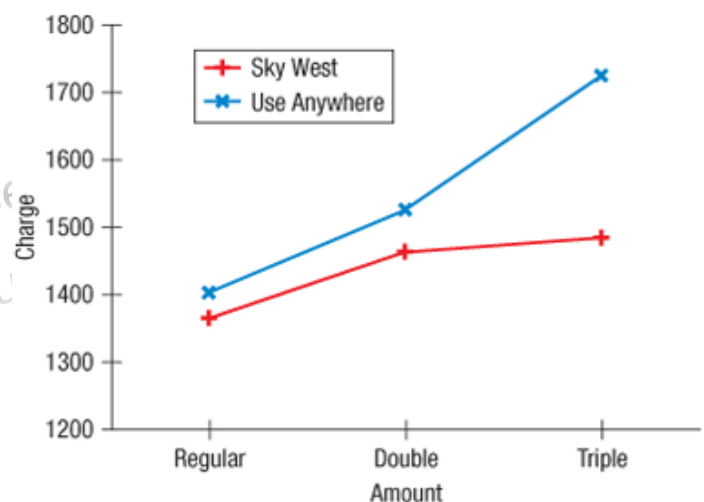
Because this is a randomized experiment, we can attribute significant differences to the treatments.

Be sure to make recommendations based on the context of your business decision.

MEMO

Re: Test mailing for creative offer and envelope

The mailing for testing the *Triple Miles* initiative went out in March, and results on charges from April through June were available in early July. We found that *Use Anywhere Miles* performed better than the standard *Sky West Miles*, but that the amount they increased charges depended on the amount offered.



As we can see, Triple Miles for the Sky West Miles didn't increase Charge significantly and is probably not worth the added expense. However, Triple Miles for the Use Anywhere Miles generated an average \$205 more in average Charge. Although we do not know whether the effects are statistically significant, we feel that the added revenue of the *Triple Miles* may justify their cost.

In summary, our preliminary recommendations are to offer *Triple Miles* for the *Use Anywhere Miles* offers for our Gold Card customers but keep the *Double Miles* offer for the *Sky West Miles*.

What Can Go Wrong?

- **Don't give up just because you can't run an experiment.** Sometimes we can't run an experiment because we can't identify or control the factors. Sometimes it would simply be unethical to run the experiment. (Consider randomly assigning employees to two environments—one where workers were exposed to massive amounts of cigarette smoke and one a smoke-free environment—to see differences in health and productivity.) If we can't perform an experiment, often an observational study is a good choice.
- **Beware of confounding.** Use randomization whenever possible to ensure that the factors not in your experiment are not confounded with your treatment levels. Be alert to confounding that cannot be avoided, and report it along with your results.
- **Bad things can happen even to good experiments.** Protect yourself by recording additional information. An experiment in which the air-conditioning failed for two weeks, affecting the results, was saved by recording the temperature (although that was not originally one of the factors) and estimating the effect the higher temperature had on the response.³ It's generally good practice to collect as much information as possible about your experimental units and the circumstances of the experiment. For example, in the direct mail experiment, it would be wise to record details of the general economy and any global events (such as a sharp downturn in the stock market) that might affect customer behavior.

³R. D. DeVeaux and M. Szelewski, "Optimizing Automatic Splitless Injection Parameters for Gas Chromatographic Environmental Analysis," *Journal of Chromatographic Science* 27, no. 9 (1989): 513–518.

- **Don't spend your entire budget on the first run.** Just as it's a good idea to pretest a survey, it's always wise to try a small pilot experiment before running the full-scale experiment. You may learn, for example, how to choose factor levels more effectively, about effects you forgot to control, and about unanticipated confounding.
- **Be wary of generalizing to situations other than the one at hand.** Think hard about how the data were generated to understand the breadth of conclusions you are entitled to draw.

- When there is an interaction between factors, be careful in interpreting the effect of the individual factors on the response variable. Look at this interaction plot:

The experiment was run at two temperatures and two pressure levels. High amounts of material were produced at high pressure with high temperature and at low pressure with low temperature. What's the effect of *Temperature*? Of *Pressure*? When looked at individually, each factor appears to have no effect on *Yield*, but it would be silly (and wrong) to say that neither *Temperature* nor *Pressure* was important. The real story is in the interaction. (See [Figure 9.6](#).)

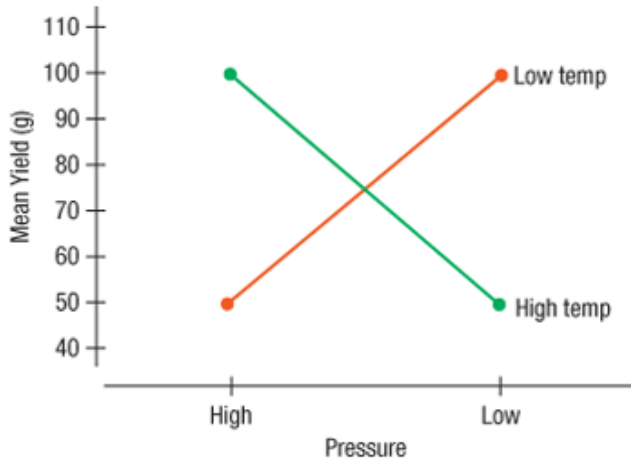


Figure 9.6

An interaction plot of *Yield* by *Temperature* and *Pressure*. There is no simple effect of *Pressure* because the average *Yield* at the two pressures is the same. That doesn't mean that *Pressure* has no effect on the *Yield*. In the presence of an interaction, be careful when interpreting the effect of individual factors.

Ethics in Action

A large U.S. multinational that provides field services to energy companies worldwide operates in approximately 50 countries and employs nearly 85,000 people. While very profitable, this company has had its share of controversies. The most publicized involved allegations of misconduct with government contracts and inattention to safety resulting in incidents where employees were injured and the environment negatively impacted. Consequently, its top level executives became quite concerned after several of its employees in the United States filed complaints with the human resources department about the safety issues in the company.

To be proactive and potentially avoid more public scrutiny, management appointed Beth Morgan-Brown, head of corporate human resources, to coordinate an effort to investigate this issue more broadly. In the first phase of the investigation, she conducted interviews and reviewed files. As Beth had suspected, the Southeast region of the U.S. was particularly problematic. Beth had always believed that HR in that regional corporate office did not provide adequate safety training for management.

So, she developed a set of videos to see if they would better inform management of the serious safety issues in the field. She randomly selected 100 managers in each of the South East and West regions, assigning half of them in each region to receive the video training and the other half to receive no training. At the end of the training period she gave each manager a set of questions about safety to test their knowledge.

To her surprise, the managers in the South East who received the training actually did worse than those in the South East who received no training, but in the West, the trained managers did much better than those in the Western control group. On average overall, however, the trained managers did not do significantly better than the control group, and so Beth reported to corporate that the training was unsuccessful.

- **Identify the ethical dilemma in this scenario.**
- **What are the undesirable consequences?**
- **Propose an ethical solution that considers the welfare of all stakeholders.**

Chapter 9 From Learning to Earning

Learning Objectives

Know the elements of a designed randomized experiment.

- *Experimental units* (sometimes called *subjects* or *participants*) are assigned at random to *treatments*.
- The experimenter manipulates *factors*, setting them to specified *levels* to establish the treatments.
- A quantitative *response variable* is measured or observed for each experimental unit.
- We can attribute differences in the response to the differences among the treatments.

State and apply the Four Principles of Experimental Design.

- *Control* sources of variation other than the factors being tested. Make the conditions as similar as possible for all treatment groups except for differences among the treatments.
- *Randomize* the assignment of subjects to treatments. *Balance* the design by assigning the same number of subjects to each treatment.
- *Replicate* the experiment on more than one subject.
- *Block* the experiment by grouping together subjects who are similar in important ways that you cannot control.

Work with *Blinding* and *Control groups*.

- A *single-blind* study is one in which either all those who can affect the results or all those who evaluate the results are kept ignorant of which subjects receive which treatments.
- A *double-blind* study is one in which both those classes of actors are ignorant of the treatment assignment.
- A *control group* is assigned to a null treatment or to the best available alternative treatment.
 - Control subjects are often administered a *placebo* or null treatment that mimics the treatment being studied but is known to be inactive.

Terms

A/B test

An A/B test design offers two alternative treatments to judge which is better.

Balanced

An experiment design is said to be *balanced* if an equal number of experimental units receive each treatment. Balance is a desirable feature because balanced designs are easier to analyze.

Blind, Blinding

Any individual associated with an experiment who is not aware of how subjects have been allocated to treatment groups is said to be blinded.

Blocking, Blocking Factor

When groups of experimental units are similar, it is often a good idea to gather them together into the same level of a factor. The factor is called a blocking factor and its levels are called blocks. By blocking we isolate the variability attributable to the differences between the blocks so that we can see the differences in the means due to the treatments more clearly.

Bucket test

Compares two alternative treatments, another name for an A/B test.

Champion/challenge design

An A/B test that compares a new treatment (the challenger) with an established one (the champion).

Confounded

When a factor is associated with another factor in such a way that their effects cannot be separated, we say that these two factors are confounded.

Control

When we limit the levels of a factor not explicitly part of the experiment design, we have controlled that factor. (By contrast, the factors we are testing are said to be *manipulated*.)

Control group

The experimental units assigned to a baseline treatment level, typically either the default treatment, which is well understood, or a null, placebo treatment. Their responses provide a basis for comparison.

Designs

- **Randomized block design:** The randomization occurs only within blocks.
- **Completely randomized design:** All experimental units have an equal chance of receiving any treatment.
- **Factorial design:** Includes more than one factor in the same design and includes every combination of all the levels of each factor.

Double-blind, Single-blind

There are two classes of individuals who can affect the outcome of an experiment:

those who could *influence the results* (subjects, treatment administrators, or technicians)

those who *evaluate the results* (judges, treating physicians, etc.)

When every individual in *either* of these classes is blinded, an experiment is said to be single-blind.

When everyone in *both* classes is blinded, we call the experiment double-blind.

Experiment

An experiment *manipulates* factor levels to create treatments, *randomly assigns* subjects to these treatment levels, and then *compares* the responses of the subject groups across treatment levels.

Experimental units

Individuals on whom an experiment is performed. Usually called subjects or participants when they are human.

Factor

A variable whose levels are controlled by the experimenter. Experiments attempt to discover the effects that differences in factor levels may have on the responses of the experimental units.

Interaction

When the effects of the levels of one factor change depending on the level of the other factor, the two factors are said to interact. When interaction terms are present, it is misleading to talk about the main effect of one factor because how large it is *depends* on the level of the other factor.

Interaction plot

A plot that shows the means at each treatment combination, highlighting the factor effects and their behavior at all the combinations.

Level

The specific values that the experimenter chooses for a factor are called the levels of the factor.

Placebo

A treatment that mimics the treatment to be studied, designed so that all groups think they are receiving the same treatment. Many subjects respond to such a treatment (a response known as a *placebo effect*). Only by comparing with a placebo can we be sure that the observed effect of a treatment is not due simply to the placebo effect.

Placebo effect

The tendency of many human subjects (often 20% or more of experiment subjects) to show a response even when administered a placebo.

Principles of experimental design

- **Control** aspects of the experiment that we know may have an effect on the response, but that are not the factors being studied.
- **Randomize** subjects to treatments to even out effects that we cannot control.
- **Replicate** over as many subjects as possible. Results for a single subject are just anecdotes.
- **Block** to reduce the effects of identifiable attributes of the subjects that cannot be controlled.

Random assignment

To be valid, an experiment must assign experimental units to treatment groups at random. This is called random assignment.

Replicates

One result is an anecdote; not data. Multiple cases should be treated in each treatment level of an experiment. These are referred to as *replicates*.

Response

A variable whose values are compared across different treatments. In a randomized experiment, large response differences can be attributed to the effect of differences in treatment level.

Subjects or Participants

When the experimental units are people, they are usually referred to as subjects or participants.

Treatment

The process, intervention, or other controlled circumstance applied to randomly assigned experimental units. Treatments are the different levels of a single factor or are made up of combinations of levels of two or more factors.

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Personal use only, do not reproduce.
2020-11-13
valerie.albury@student.ashford.edu

Chapter 9 Exercises

Section 9.1

1. For the following experiment, identify the experimental units, the treatments, the response, and the random assignment.

A commercial food lab compared recipes for chocolate chip cookies. They baked cookies with different kinds of chips (milk chocolate, dark chocolate, and semi-sweet). All other ingredients and amounts were the same. Ten trained tasters rated the cookies on a scale of 1 to 10. The cookies were presented to the tasters in a random order.

2. For the following experiment, identify the experimental units, the treatments, the response, and the random assignment.

An investment club decided to compare investment strategies. Starting with nine equal investment amounts, three invested in the “dogs of the Dow”—stocks in the Dow Industrial average that had been underperforming relative to the rest of the Dow average. The relative amounts to invest in each of the stocks were chosen randomly and differently for each fund. Three funds invested following the advice of a TV investment show host, again choosing the specific stocks and allocations randomly for the three funds. And three funds invested by throwing darts at a page from the *Wall Street Journal* that listed stocks on the NYSE, and invested in each of the stocks hit by a dart, throwing a different set of darts for each of the three funds. At the end of six months the funds were compared.

Section 9.2

3. For the cookie recipe experiment of [Exercise 1](#), identify how Control, Randomization, and Replication were used.
4. For the investment experiment of [Exercise 2](#), identify how Control, Randomization, and Replication were used.

Section 9.3

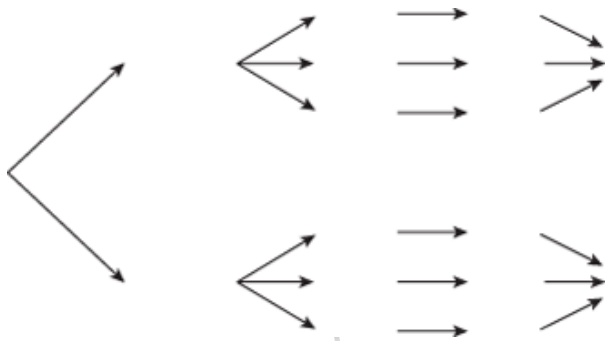
5. An Internet sale site randomly sent customers to one of three versions of its welcome page. It recorded how long each visitor stayed in the site.

Here is a diagram of that experiment. Fill in the parts of the experiment.



6. Analysts from the Internet company of [Exercise 5](#) are now concerned that customers who come directly to their site (by typing their URL into a browser) might respond differently than those referred to the site from other sites (such as search engines). They decide to block according to how the customer arrived at their site.

Here is a diagram of that experiment. Fill in the parts.



Section 9.4

7. For the following experiment, indicate whether it was single-blind, double-blind, or not blinded at all. Explain your reasoning.

Makers of a new frozen entrée arranged for it to be served to randomly selected customers at a restaurant in place of the equivalent entrée ordinarily prepared in the kitchen. After their meal, the customers were asked about the quality of the food.

8. For the following experiment, indicate whether it was single-blind, double-blind, or not blinded at all. Explain your reasoning.

Does a “stop smoking” program work better if it costs more? Smokers responding to an advertisement offering to help them stop smoking were randomly offered a program costing \$100 or the same program costing \$250. The offer was made individually to each client by presenting a sealed envelope so the clerk providing the offer did not know the details of the offer. At the end of the program (a course and films along with diet and smoking cessation aids), clients were followed for six months to see if they had indeed quit smoking.

9. Is the experiment of **Exercise 1** ☐ blind? Could it be made double blind? Explain.
10. Is the experiment of **Exercise 5** ☐ blind? Could it be made double blind? Explain.