Bell Work:

Determine whether you survey the population or a sample.

A. You want to know the average age of all teachers on campus.

B. You want to know whether Coke or Pepsi products are preferred on campus.

C. You want to know what the favorite hobbies are of the students at a school.

Introduction

The New York Times/CBS News Poll is a well-known regular polling organization that releases results of polls taken to help clarify the opinions of Americans on pending elections, current leaders, or economic or foreign policy issues. In an article entitled "How the Poll Was Conducted" that explains some of the details of a recent poll, the following statements appear[1]:

"In theory, in 19 cases out of 20, overall results based on such samples will differ by no more than three percentage points in either direction from what would have been obtained by seeking to interview all American adults."

"In addition to sampling error, the practical difficulties of conducting any survey of public opinion may introduce other sources of error into the poll. Variation in the wording and order of questions, for example, may lead to somewhat different results."

These statements illustrate two different potential problems with opinion polls, surveys, observational studies, and experiments. In this chapter, we will investigate these problems and more by looking at sampling in detail.

Census vs. Sample

A *sample* is a representative subset of a population. If a statistician or other researcher wants to know some information about a population, the only way to be truly sure is to conduct a census. In a *census*, every unit in the population being studied is measured or surveyed. In opinion polls, like the *New York Times* poll mentioned above, results are generalized from a sample. If we really wanted to know the true approval rating of the president, for example, we would have to ask every single American adult his or her opinion. There are some obvious reasons why a census is impractical in this case, and in most situations.

First, it would be extremely expensive for the polling organization. They would need an extremely large workforce to try and collect the opinions of every American adult. Also, it would take many workers and many hours to organize, interpret, and display this information. Even if it could be done in several months, by the time the results were published, it would be very probable that recent events had changed peoples' opinions and that the results would be obsolete.

In addition, a census has the potential to be destructive to the population being studied. For example, many manufacturing companies test their products for quality control. A padlock manufacturer might use a machine to see how much force it can apply to the lock before it breaks. If they did this with every lock, they would have none left to sell! Likewise, it would not be a good idea for a biologist to find the number of fish in a lake by draining the lake and counting them all!

The U.S. Census is probably the largest and longest running census, since the Constitution mandates a complete counting of the population. The first U.S. Census was taken in 1790 and was done by U.S. Marshalls on horseback. Taken every 10 years, a Census was conducted in 2010, and in a report by the Government Accountability Office in 1994, was estimated to cost $11 billion. This cost has recently increased as computer problems have forced the forms to be completed by $\mathbf{hand}^3$. You can find a great deal of information about the U.S. Census, as well as data from past Censuses, on the Census Bureau's website: http://www.census.gov/.

Due to all of the difficulties associated with a census, sampling is much more practical. However, it is important to understand that even the most carefully planned sample will be subject to random variation between the sample and the population. Recall that these differences due to chance are called *sampling error*. We can use the laws of probability to predict the level of accuracy in our sample. Opinion polls, like the *New York Times* poll mentioned in the introduction, tend to refer to this as *margin of error*. The second statement quoted from the *New York Times* article mentions another problem with sampling. That is, it is often difficult to obtain a sample that accurately reflects the total population. It is also possible to make mistakes in selecting the sample and collecting the information. These problems result in a non-representative sample, or one in which our conclusions differ from what they would have been if we had been able to conduct a census.

To help understand these ideas, consider the following theoretical example. A coin is considered fair if the probability, p, of the coin landing on heads is the same as the probability of it landing on tails $(p = 0.5)$. The probability is defined as the proportion of heads obtained if the coin were flipped an infinite number of times. Since it is impractical, if not impossible, to flip a coin an infinite number of times, we might try looking at 10 samples, with each sample consisting of 10 flips of the coin. Theoretically, you would expect the coin to land on heads 50% of the time, but it is very possible that, due to chance alone, we would experience results that differ from this. These differences are due to sampling error. As we will investigate in detail in later chapters, we can decrease the sampling error by increasing the sample size (or the number of coin flips in this case). It is also possible that the results we obtain could differ from those expected if we were not careful about the way we flipped the coin or allowed it to land on different surfaces. This would be an example of a non-representative sample.

At the following website, you can see the results of a large number of coin flips: http://www.mathsonline.co.uk/nonmembers/resource/prob/coins.html. You can see the random variation among samples by asking for the site to flip 10 coins 10 times. Our results for that experiment produced the following numbers of heads: 3, 3, 4, 4, 4, 4, 5, 6, 6, 6. This seems quite strange, since the expected number is 5. How do your results compare?

Bias in Samples and Surveys

The term most frequently applied to a non-representative sample is *bias*. Bias has many potential sources. It is important when selecting a sample or designing a survey that a statistician make every effort to eliminate potential sources of bias. In this section, we will discuss some of the most common types of bias. While these concepts are universal, the terms used to define them here may be different than those used in other sources.

Sampling Bias

In general, sampling bias refers to the methods used in selecting the sample. The *sampling frame* is the term we use to refer to the group or listing from which the sample is to be chosen. If you wanted to study the population of students in your school, you could obtain a list of all the students from the office and choose students from the list. This list would be the sampling frame.

Incorrect Sampling Frame

If the list from which you choose your sample does not accurately reflect the characteristics of the population, this is called *incorrect sampling frame*. A sampling frame error occurs when some group from the population does not have the opportunity to be represented in the sample. For example, surveys are often done over the telephone. You could use the telephone book as a sampling frame by choosing numbers from the telephone book. However, in addition to the many other potential problems with telephone poles, some phone numbers are not listed in the telephone book. Also, if your population includes all adults, it is possible that you are leaving out important groups of that population. For example, many younger adults in particular tend to only use their cell phones or computer-based phone services and may not even have traditional phone service. Even if you picked phone numbers randomly, the sampling frame could be incorrect, because there are also people, especially those who may be economically disadvantaged, who have no phone. There is absolutely no chance for these individuals to be represented in your sample. A term often used to describe the problems when a group of the population is not represented in a survey is *undercoverage*. Undercoverage can result from all of the different sampling biases.

One of the most famous examples of sampling frame error occurred during the 1936 U.S. presidential election. The Literary Digest, a popular magazine at the time, conducted a poll and predicted that Alf Landon would win the election that, as it turned out, was won in a landslide by Franklin Delano Roosevelt. The magazine obtained a huge sample of ten million people, and from that pool, 2 million replied. With these numbers, you would typically expect very accurate results. However, the magazine used their subscription list as their sampling frame. During the depression, these individuals would have been only the wealthiest Americans, who tended to vote Republican, and left the majority of typical voters under-covered.

Convenience Sampling

Suppose your statistics teacher gave you an assignment to perform a survey of 20 individuals. You would most likely tend to ask your friends and family to participate, because it would be easy and quick. This is an example of *convenience sampling*, or convenience bias. While it is not always true, your friends are usually people who share common values, interests, and opinions. This could cause those opinions to be over-represented in relation to the true population. Also, have you ever been approached by someone conducting a survey on the street or in a mall? If such a person were just to ask the first 20 people they found, there is the potential that large groups representing various opinions would not be included, resulting in undercoverage.

Judgment Sampling

Judgment sampling occurs when an individual or organization that is usually considered an expert in the field being studied chooses the individuals or group of individuals to be used in the sample. Because it is based on a subjective choice, even by someone considered an expert, it is very susceptible to bias. In some sense, this is what those responsible for the Literary Digest poll did. They incorrectly chose groups they believed would represent the population. If a person wants to do a survey on middle-class Americans, how would this person decide who to include? It would be left to this person's own judgment to create the criteria for those considered middle-class. This individual's judgment might result in a different view of the middle class that might include wealthier individuals that others would not consider part of the population. Similar to judgment sampling, in *quota sampling*, an individual or organization attempts to include the proper proportions of individuals of different subgroups in their sample. While it might sound like a good idea, it is subject to an individual's prejudice and is, therefore, prone to bias.

Size Bias

If one particular subgroup in a population is likely to be over-represented or under-represented due to its size, this is sometimes called *size bias*. If we chose a state at random from a map by closing our eyes and pointing to a particular place, larger states would have a greater chance of being chosen than smaller ones. As another example, suppose that we wanted to do a survey to find out the typical size of a student's math class at a school. The chances are greater that we would choose someone from a larger class for our survey. To understand this, say that you went to a very small school where there are only four math classes, with one class having 35 students, and the other three classes having only 8 students. If you simply choose students at random, it is more likely you will select students for your sample who will will say the typical size of a math class is 35, since there are more students in the larger class.

Here's one more example: a person driving on an interstate highway tends to say things like, “Wow, I was going the speed limit, and everyone was just flying by me.” The conclusion this person is making about the population of all drivers on this highway is that most of them are traveling faster than the speed limit. This may indeed be true, but let’s say that most people on the highway, along with our driver, really are abiding by the speed limit. In a sense, the driver is collecting a sample, and only those few who are close to our driver will be included in the sample. There will be a larger number of drivers going faster in our sample, so they will be over-represented. As you may already see, these definitions are not absolute, and often in a practical example, there are many types of overlapping bias that could be present and contribute to overcoverage or undercoverage. We could also cite incorrect sampling frame or convenience bias as potential problems in this example.

Response Bias

The term *response bias* refers to problems that result from the ways in which the survey or poll is actually presented to the individuals in the sample.

Voluntary Response Bias

Television and radio stations often ask viewers/listeners to call in with opinions about a particular issue they are covering. The websites for these and other organizations also usually include some sort of online poll question of the day. Reality television shows and fan balloting in professional sports to choose all-star players make use of these types of polls as well. All of these polls usually come with a disclaimer stating that, "This is not a scientific poll." While perhaps entertaining, these types of polls are very susceptible to *voluntary response bias*. The people who respond to these types of surveys tend to feel very strongly one way or another about the issue in question, and the results might not reflect the overall population. Those who still have an opinion, but may not feel quite so passionately about the issue, may not be motivated to respond to the poll. This is especially true for phone-in or mail-in surveys in which there is a cost to participate. The effort or cost required tends to weed out much of the population in favor of those who hold extremely polarized views. A news channel might show a report about a child killed in a drive-by shooting and then ask for people to call in and answer a question about tougher criminal sentencing laws. They would most likely receive responses from people who were very moved by the emotional nature of the story and wanted anything to be done to improve the situation. An even bigger problem is present in those types of polls in which there is no control over how many times an individual may respond.

Non-Response Bias

One of the biggest problems in polling is that most people just don't want to be bothered taking the time to respond to a poll of any kind. They hang up on a telephone survey, put a mail-in survey in the recycling bin, or walk quickly past an interviewer on the street. We just don't know how much these individuals' beliefs and opinions reflect those of the general population, and, therefore, almost all surveys could be prone to *non-response bias*.

Questionnaire Bias

Questionnaire bias occurs when the way in which the question is asked influences the response given by the individual. It is possible to ask the same question in two different ways that would lead individuals with the same basic opinions to respond differently. Consider the following two questions about gun control.

"Do you believe that it is reasonable for the government to impose some limits on purchases of certain types of weapons in an effort to reduce gun violence in urban areas?"

"Do you believe that it is reasonable for the government to infringe on an individual's constitutional right to bear arms?"

A gun rights activist might feel very strongly that the government should never be in the position of limiting guns in any way and would answer no to both questions. Someone who is very strongly against gun ownership, on the other hand, would probably answer yes to both questions. However, individuals with a more tempered, middle position on the issue might believe in an individual's right to own a gun under some circumstances, while still feeling that there is a need for regulation. These individuals would most likely answer these two questions differently.

You can see how easy it would be to manipulate the wording of a question to obtain a certain response to a poll question. Questionnaire bias is not necessarily always a deliberate action. If a question is poorly worded, confusing, or just plain hard to understand, it could lead to non-representative results. When you ask people to choose between two options, it is even possible that the order in which you list the choices may influence their response!

Incorrect Response Bias

A major problem with surveys is that you can never be sure that the person is actually responding truthfully. When an individual intentionally responds to a survey with an untruthful answer, this is called *incorrect response bias*. This can occur when asking questions about extremely sensitive or personal issues. For example, a survey conducted about illegal drinking among teens might be prone to this type of bias. Even if guaranteed their responses are confidential, some teenagers may not want to admit to engaging in such behavior at all. Others may want to appear more rebellious than they really are, but in either case, we cannot be sure of the truthfulness of the responses.

Another example is related to the donation of blood. Because the dangers of donated blood being tainted with diseases carrying a negative social stereotype increased in the 1990's, the Red Cross has recently had to deal with incorrect response bias on a constant and especially urgent basis. Individuals who have engaged in behavior that puts them at risk for contracting AIDS or other diseases have the potential to pass these diseases on through donated blood[4]. Screening for at-risk behaviors involves asking many personal questions that some find awkward or insulting and may result in knowingly false answers. The Red Cross has gone to great lengths to devise a system with several opportunities for individuals giving blood to anonymously report the potential danger of their donation.

In using this example, we don't want to give the impression that the blood supply is unsafe. According to the Red Cross, "Like most medical procedures, blood transfusions have associated risk. In the more than fifteen years since March 1985, when the FDA first licensed a test to detect HIV antibodies in donated blood, the Centers for Disease Control and Prevention has reported only 41 cases of AIDS caused by transfusion of blood that tested negative for the AIDS virus. During this time, more than 216 million blood components were transfused in the United States. The tests to detect HIV were designed specifically to screen blood donors. These tests have been regularly upgraded since they were introduced. Although the tests to detect HIV and other blood-borne diseases are extremely accurate, they cannot detect the presence of the virus in the 'window period' of infection, the time before detectable antibodies or antigens are produced. That is why there is still a very slim chance of contracting HIV from blood that tests negative. Research continues to further reduce the very small risk." [4] Source: http://chapters.redcross.org/br/nypennregion/safety/mythsaid.htm

Reducing Bias

Randomization

The best technique for reducing bias in sampling is *randomization*. When a *simple random sample* of size n (commonly referred to as an SRS) is taken from a population, all possible samples of size n in the population have an equal probability of being selected for the sample. For example, if your statistics teacher wants to choose a student at random for a special prize, he or she could simply place the names of all the students in the class in a hat, mix them up, and choose one. More scientifically, your teacher could assign each student in the class a number from 1 to 25 (assuming there are 25 students in the class) and then use a computer or calculator to generate a random number to choose one student. This would be a simple random sample of size 1.

Cluster Sampling

Cluster sampling is when a naturally occurring group is selected at random, and then either all of that group, or randomly selected individuals from that group, are used for the sample. If we select at random from out of that group, or cluster into smaller subgroups, this is referred to as *multi-stage sampling*. For example, to survey student opinions or study their performance, we could choose 5 schools at random from your state and then use an SRS (simple random sample) from each school. If we wanted a national survey of urban schools, we might first choose 5 major urban areas from around the country at random, and then select 5 schools at random from each of these cities. This would be both cluster and multi-stage sampling. Cluster sampling is often done by selecting a particular block or street at random from within a town or city. It is also used at large public gatherings or rallies. If officials take a picture of a small, representative area of the crowd and count the individuals in just that area, they can use that count to estimate the total crowd in attendance.

Stratified Sampling

In *stratified sampling*, the population is divided into groups, called strata (the singular term is 'stratum'), that have some meaningful relationship. Very often, groups in a population that are similar may respond differently to a survey. In order to help reflect the population, we stratify to insure that each opinion is represented in the sample. For example, we often stratify by gender or race in order to make sure that the often divergent views of these different groups are represented. In a survey of high school students, we might choose to stratify by school to be sure that the opinions of different communities are included. If each school has an approximately equal number of students, then we could simply choose to take an SRS of size 25 from each school. If the numbers in each stratum are different, then it would be more appropriate to choose a fixed sample (100 students, for example) from each school and take a number from each school proportionate to the total school size.

Lesson Summary

If you collect information from every unit in a population, it is called a census. Because a census is so difficult to do, we instead take a representative subset of the population, called a sample, to try and make conclusions about the entire population. The downside to sampling is that we can never be completely sure that we have captured the truth about the entire population, due to random variation in our sample that is called sampling error. The list of the population from which the sample is chosen is called the sampling frame. Poor technique in surveying or choosing a sample can also lead to incorrect conclusions about the population that are generally referred to as bias. Selection bias refers to choosing a sample that results in a subgroup that is not representative of the population. Incorrect sampling frame occurs when the group from which you choose your sample does not include everyone in the population, or at least units that reflect the full diversity of the population. Incorrect sampling frame errors result in undercoverage. This is where a segment of the population containing an important characteristic did not have an opportunity to be chosen for the sample and will be marginalized, or even left out altogether.

Points to Consider

- How is the margin of error for a survey calculated?
- What are the effects of sample size on sampling error?

Assignment: Worksheet