7.2: Sampling Methods

Learning Objectives

- Understand different methods of sampling
- Understand error from sampling

As we mentioned in a previous section, the first thing we should do before conducting a survey is to identify the population that we want to study. Suppose we are hired by a politician to determine the amount of support he has among the electorate should he decide to run for another term. What population should we study? Every person in the district? Not every person is eligible to vote, and regardless of how strongly someone likes or dislikes the candidate, they don't have much to do with him being re-elected if they are not able to vote.

What about eligible voters in the district? That might be better, but if someone is eligible to vote but does not register by the deadline, they won't have any say in the election either. What about registered voters? Many people are registered but choose not to vote. What about "likely voters?"

This is the criteria used in much political polling, but it is sometimes difficult to define a "likely voter." Is it someone who voted in the last election? In the last general election? In the last presidential election? Should we consider someone who just turned 18 a "likely voter?" They weren't eligible to vote in the past, so how do we judge the likelihood that they will vote in the next election?

In November 1998, former professional wrestler Jesse "The Body" Ventura was elected governor of Minnesota. Up until right before the election, most polls showed he had little chance of winning. There were several contributing factors to the polls not reflecting the actual intent of the electorate:

- Ventura was running on a third-party ticket and most polling methods are better suited to a two-candidate race.
- Many respondents to polls may have been embarrassed to tell pollsters that they were planning to vote for a professional wrestler.
- The mere fact that the polls showed Ventura had little chance of winning might have prompted some people to vote for him in protest to send a message to the major-party candidates.

But one of the major contributing factors was that Ventura recruited a substantial amount of support from young people, particularly college students, who had never voted before and who registered specifically to vote in the gubernatorial election. The polls did not deem these young people likely voters (since in most cases young people have a lower rate of voter registration and a turnout rate for elections) and so the polling samples were subject to **sampling bias**: they omitted a portion of the electorate that was weighted in favor of the winning candidate.

Sampling bias

Sampling bias occurs if every member of the population doesn’t have equal likelihood of being in the sample.

So even identifying the population can be a difficult job, but once we have identified the population, how do we choose an appropriate sample? Remember, although we would prefer to survey all members of the population, this is usually impractical unless the population is very small, so we choose a sample. There are many ways to sample a population, but there is one goal we need to keep in mind: we would like the sample to be *representative of the population*.

Returning to our hypothetical job as a political pollster, we would not anticipate very accurate results if we drew all of our samples from among the customers at a Starbucks, nor would we expect that a sample drawn entirely from the membership list of the local Kiwanis club would provide a useful picture of district-wide support for our candidate.

One way to ensure that the sample has a reasonable chance of mirroring the population is to employ *randomness*. The most basic random method is simple random sampling.

Simple random sample

A **random sample** is one in which each member of the population has an equal probability of being chosen. A **simple random sample** is one in which every member of the population and any group of members has an equal probability of being chosen.

Example 7.2.1

How could you obtain a simple random sample of 1000 likely voters in the state?

Solution

If we could somehow identify all likely voters in the state, put each of their names on a piece of paper, toss the slips into a (very large) hat and draw 1000 slips out of the hat, we would have a simple random sample.

In practice, computers are better suited for this sort of endeavor than millions of slips of paper and extremely large headgear.

It is always possible, however, that even a random sample might end up not being totally representative of the population. If we repeatedly take samples of 1000 people from among the population of likely voters in the state of California, some of these samples might tend to have a slightly higher percentage of Democrats (or Republicans) than does the general population; some samples might include more older people and some samples might include more younger people; etc. In most cases, this **sampling variability** is not significant.

Sampling variability

The natural variation of samples is called **sampling variability**.

This is unavoidable and expected in random sampling, and in most cases is not an issue.

To help account for variability, pollsters might instead use a **stratified sample**.

Stratified sampling

In **stratified sampling**, a population is divided into a number of subgroups (or strata). Random samples are then taken from each subgroup with sample sizes proportional to the size of the subgroup in the population.

Example 7.2.2

Suppose in a particular state that previous data indicated that the electorate was comprised of 39% Democrats, 37% Republicans and 24% independents. How would you obtain a stratified sample of 1000 voters?

Solution

In a sample of 1000 people, you would then expect to get about 390 Democrats, 370 Republicans and 240 independents. To accomplish this, you could randomly select 390 people from among those voters known to be Democrats, 370 from those known to be Republicans, and 240 from those with no party affiliation.

Stratified sampling can also be used to select a sample with people in desired age groups, a specified mix ratio of males and females, etc. A variation on this technique is called **quota sampling**.

Quota Sampling

Quota sampling is a variation on stratified sampling, wherein samples are collected in each subgroup until the desired quota is met.

Example 7.2.3

How would you implement quota sampling to obtain the stratified sample from the previous example?

Solution

Suppose the pollsters call people at random, but once they have met their quota of 390 Democrats, they only gather people who do not identify themselves as a Democrat.

You may have had the experience of being called by a telephone pollster who started by asking you your age, income, etc. and then thanked you for your time and hung up before asking any "real" questions. Most likely, they already had contacted enough people in your demographic group and were looking for people who were older or younger, richer or poorer, etc. Quota sampling is usually a bit easier than stratified sampling, but also does not ensure the same level of randomness.

Another sampling method is **cluster sampling**, in which the population is divided into groups, and one or more groups are randomly selected to be in the sample.

Cluster sampling

In **cluster sampling**, the population is divided into subgroups (clusters), and a set of subgroups are selected to be in the sample.

In both stratified and cluster sampling, the population is divided into non-overlapping subgroups. The difference between them is that in stratified sampling, subjects from each subgroup are randomly selected, but in cluster sampling, some subgroups are selected and everyone in the subgroup is included in the sample.

Example 7.2.4

How could you obtain a cluster sample of faculty from Fullerton College?

Solution

If the college wanted to do a cluster sample of faculty, since faculty are already divided into departments, they could randomly select 10 departments and give the survey to all the faculty in those departments. Each faculty member will be listed in only one department (their primary discipline), even if they are members of multiple departments.

Other sampling methods include **systematic sampling**.

Systematic sampling

In **systematic sampling**, every n^{th} member of the population is selected to be in the sample.

Example 7.2.5

How could a pollster select a sample using systematic sampling?

Solution

The pollster calls every 100th name in the phone book. Systematic sampling is not as random as a simple random sample (if your name is Albert Aardvark and your sister Alexis Aardvark is right after you in the phone book, there is no way you could both end up in the sample) but it can yield acceptable samples.

Perhaps the worst types of sampling methods are **convenience samples** and **voluntary response samples**.

Convenience sampling and voluntary response sampling

In **convenience sampling,** the sample is chosen by selecting whoever is convenient.

In **voluntary response sampling,** the sample is chosen by allowing the respondents to volunteer.

Example 7.2.6

How could a pollster select a sample using convenience sampling?

Solution

The pollster stands on a street corner and interviews the first 100 people who agree to speak to him. This is convenient for the pollster but does not generate a sample that is representative of the entire population.

Example 7.2.7

How could a researcher collect data on how voters feel about a tax proposal that will be on the ballot using voluntary response sampling?

Solution

A website has a survey asking readers to give their opinion on a tax proposal. This is a self-selected sample, or voluntary response sample, in which respondents volunteer to participate.

Usually voluntary response samples are skewed towards people who have a particularly strong opinion about the subject of the survey or who just have way too much time on their hands and enjoy taking surveys.

Try It 7.2.1

In each case, indicate what sampling method was used:

a. Every 4th person in the class was selected.
b. A sample was selected to contain 25 men and 35 women.
c. Viewers of a new show are asked to vote on the show's website.
d. A website randomly selects 50 of their customers to send a satisfaction survey.
e. To survey voters in a town, a polling company randomly selects 10 city blocks, and interviews everyone who lives on those blocks.

Answer

a. Systematic
b. Stratified or Quota
c. Voluntary response
d. Simple random
e. Cluster

This page titled 7.2: Sampling Methods is shared under a CC BY-SA 3.0 license and was authored, remixed, and/or curated by David Lippman & Jeff Eldridge (The OpenTextBookStore) via source content that was edited to the style and standards of the LibreTexts platform.

- **10.4: Sampling methods** by David Lippman & Jeff Eldridge is licensed CC BY-SA 3.0. Original source: http://www.opentextbookstore.com/mathinsociety.