

## 1.2: Sampling Methods

As stated before, if you want to know something about a population, it is often impossible or impractical to examine the whole population. It might be too expensive in terms of time or money. It might be impractical – you can't test all batteries for their length of lifetime because there wouldn't be any batteries left to sell. You need to look at a sample. Hopefully the sample behaves the same as the population.

When you choose a sample you want it to be as similar to the population as possible. If you want to test a new painkiller for adults you would want the sample to include people who are fat, skinny, old, young, healthy, not healthy, male, female, etc.

There are many ways to collect a sample. None are perfect, and you are not guaranteed to collect a representative sample. That is unfortunately the limitations of sampling. However, there are several techniques that can result in samples that give you a semi-accurate picture of the population. Just remember to be aware that the sample may not be representative. As an example, you can take a random sample of a group of people that are equally males and females, yet by chance everyone you choose is female. If this happens, it may be a good idea to collect a new sample if you have the time and money.

There are many sampling techniques, though only four will be presented here. The simplest, and the type that is strived for is a **simple random sample**. This is where you pick the sample such that every sample has the same chance of being chosen. This type of sample is actually hard to collect, since it is sometimes difficult to obtain a complete list of all individuals. There are many cases where you cannot conduct a truly random sample. However, you can get as close as you can. Now suppose you are interested in what type of music people like. It might not make sense to try to find an answer for everyone in the U.S. You probably don't like the same music as your parents. The answers vary so much you probably couldn't find an answer for everyone all at once. It might make sense to look at people in different age groups, or people of different ethnicities. This is called a **stratified sample**. The issue with this sample type is that sometimes people subdivide the population too much. It is best to just have one stratification. Also, a stratified sample has similar problems that a simple random sample has. If your population has some order in it, then you could do a systematic sample. This is popular in manufacturing. The problem is that it is possible to miss a manufacturing mistake because of how this sample is taken. If you are collecting polling data based on location, then a **cluster sample** that divides the population based on geographical means would be the easiest sample to conduct. The problem is that if you are looking for opinions of people, and people who live in the same region may have similar opinions. As you can see each of the sampling techniques have pluses and minuses. Include convenience

### Definition 1.2.1

A **simple random sample (SRS)** of size  $n$  is a sample that is selected from a population in a way that ensures that every different possible sample of size  $n$  has the same chance of being selected. Also, every individual associated with the population has the same chance of being selected

Ways to select a simple random sample:

Put all names in a hat and draw a certain number of names out.

Assign each individual a number and use a random number table or a calculator or computer to randomly select the individuals that will be measured.

### Example 1.2.1 choosing a simple random sample

Describe how to take a simple random sample from a classroom.

#### Solution

Give each student in the class a number. Using a random number generator you could then pick the number of students you want to pick.

### Example 1.2.2 how not to choose a simple random sample

You want to choose 5 students out of a class of 20. Give some examples of samples that are not simple random samples:

#### Solution

Choose 5 students from the front row. The people in the last row have no chance of being selected.

Choose the 5 shortest students. The tallest students have no chance of being selected.

#### Definition 1.2.2

**Stratified sampling** is where you break the population into groups called strata, then take a simple random sample from each strata.

For example:

If you want to look at musical preference, you could divide the individuals into age groups and then conduct simple random samples inside each group.

If you want to calculate the average price of textbooks, you could divide the individuals into groups by major and then conduct simple random samples inside each group.

#### Definition 1.2.3

**Systematic sampling** is where you randomly choose a starting place then select every  $k$ th individual to measure.

For example:

You select every 5th item on an assembly line

You select every 10th name on the list

You select every 3rd customer that comes into the store.

#### Definition 1.2.4

**Cluster sampling** is where you break the population into groups called clusters. Randomly pick some clusters then poll all individuals in those clusters.

For example:

A large city wants to poll all businesses in the city. They divide the city into sections (clusters), maybe a square block for each section, and use a random number generator to pick some of the clusters.

Then they poll all businesses in each chosen cluster. You want to measure whether a tree in the forest is infected with bark beetles. Instead of having to walk all over the forest, you divide the forest up into sectors, and then randomly pick the sectors that you will travel to. Then record whether a tree is infected or not for every tree in that sector.

Many people confuse stratified sampling and cluster sampling. In stratified sampling you use all the groups and some of the members in each group. Cluster sampling is the other way around. It uses some of the groups and all the members in each group.

The four sampling techniques that were presented all have advantages and disadvantages. There is another sampling technique that is sometimes utilized because either the researcher doesn't know better, or it is easier to do. This sampling technique is known as a convenience sample. This sample will not result in a representative sample, and should be avoided.

#### Definition 1.2.5

**Convenience sample** is one where the researcher picks individuals to be included that are easy for the researcher to collect.

An example of a convenience sample is if you want to know the opinion of people about the criminal justice system, and you stand on a street corner near the county court house, and questioning the first 10 people who walk by. The people who walk by the county court house are most likely involved in some fashion with the criminal justice system, and their opinion would not represent the opinions of all individuals.

On a rare occasion, you do want to collect the entire population. In which case you conduct a census.

## Definition 1.2.6

A **census** is when every individual of interest is measured.

## Example 1.2.3 sampling type

Banner Health is a several state nonprofit chain of hospitals. Management wants to assess the incident of complications after surgery. They wish to use a sample of surgery patients. Several sampling techniques are described below. Categorize each technique as simple random sample, stratified sample, systematic sample, cluster sample, or convenience sampling.

- Obtain a list of patients who had surgery at all Banner Health facilities. Divide the patients according to type of surgery. Draw simple random samples from each group.
- Obtain a list of patients who had surgery at all Banner Health facilities. Number these patients, and then use a random number table to obtain the sample.
- Randomly select some Banner Health facilities from each of the seven states, and then include all the patients on the surgery lists of the states.
- At the beginning of the year, instruct each Banner Health facility to record any complications from every 100th surgery.
- Instruct each Banner Health facilities to record any complications from 20 surgeries this week and send in the results.

**Solution**

- This is a stratified sample since the patients were separated into different strata and then random samples were taken from each strata. The problem with this is that some types of surgeries may have more chances for complications than others. Of course, the stratified sample would show you this.
- This is a random sample since each patient has the same chance of being chosen. The problem with this one is that it will take a while to collect the data.
- This is a cluster sample since all patients are questioned in each of the selected hospitals. The problem with this is that you could have by chance selected hospitals that have no complications.
- This is a systematic sample since they selected every 100th surgery. The problem with this is that if every 90th surgery has complications, you wouldn't see this come up in the data.
- This is a convenience sample since they left it up to the facility how to do it. The problem with convenience samples is that the person collecting the data will probably collect data from surgeries that had no complications.

## Homework

- Researchers want to collect cholesterol levels of U.S. patients who had a heart attack two days prior. The following are different sampling techniques that the researcher could use. Classify each as simple random sample, stratified sample, systematic sample, cluster sample, or convenience sample.
  - The researchers randomly select 5 hospitals in the U.S. then measure the cholesterol levels of all the heart attack patients in each of those hospitals.
  - The researchers list all of the heart attack patients and measure the cholesterol level of every 25th person on the list.
  - The researchers go to one hospital on a given day and measure the cholesterol level of the heart attack patients at that time.
  - The researchers list all of the heart attack patients. They then measure the cholesterol levels of randomly selected patients.
  - The researchers divide the heart attack patients based on race, and then measure the cholesterol levels of randomly selected patients in each race grouping.
- The quality control officer at a manufacturing plant needs to determine what percentage of items in a batch are defective. The following are different sampling techniques that could be used by the officer. Classify each as simple random sample, stratified sample, systematic sample, cluster sample, or convenience sample.
  - The officer lists all of the batches in a given month. The number of defective items is counted in randomly selected batches.
  - The officer takes the first 10 batches and counts the number of defective items.
  - The officer groups the batches made in a month into which shift they are made. The number of defective items is counted in randomly selected batches in each shift.
  - The officer chooses every 15th batch off the line and counts the number of defective items in each chosen batch.

- e. The officer divides the batches made in a month into which day they were made. Then certain days are picked and every batch made that day is counted to determine the number of defective items.
3. You wish to determine the GPA of students at your school. Describe what process you would go through to collect a sample if you use a simple random sample.
4. You wish to determine the GPA of students at your school. Describe what process you would go through to collect a sample if you use a stratified sample.
5. You wish to determine the GPA of students at your school. Describe what process you would go through to collect a sample if you use a systematic sample.
6. You wish to determine the GPA of students at your school. Describe what process you would go through to collect a sample if you use a cluster sample.
7. You wish to determine the GPA of students at your school. Describe what process you would go through to collect a sample if you use a convenience sample.

### Answer

1.
  - a. Cluster sample
  - b. Systematic sample
  - c. Convenience sample
  - d. Simple random sample
  - e. Stratified sample
3. See solutions
5. See solutions
7. See solutions

---

This page titled [1.2: Sampling Methods](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Kathryn Kozak](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- **1.2: Sampling Methods** by [Kathryn Kozak](#) is licensed [CC BY-SA 4.0](#). Original source: <https://s3-us-west-2.amazonaws.com/oerfiles/statsusingtech2.pdf>.