

Stratified Random and Cluster Sampling

by Sophia Tutorial



WHAT'S COVERED

This tutorial will cover the topic of stratified random sampling, which is a random sampling procedure that subdivides the population into groups. In addition, we will introduce cluster samples. This lesson will focus on:

1. Stratified Random Samples
2. Cluster Samples
3. Real-World Comparison

1. Stratified Random Samples

Suppose a high school has just adopted a new, healthy lunch provider, and they would like to solicit student opinion on the healthy lunch options. The school has a total of 420 students: 100 freshmen, 110 sophomores, 120 juniors, and 90 seniors.

How would a simple random sample look?

For a simple random sample of 42 students, think of ways that 42 students could be chosen, each having an equal chance of being selected. First, assign each student a unique number 1 to 420 (total number of students). Once this is done, you could:

- Use a random number generator to select 42 numbers, ignoring repeats. The students who corresponded to those numbers will be surveyed about the school's new, healthy options.
- Put the 420 student names in a hat and draw out 42.

Now, is there a way that the study might improve and guarantee an accurate cross-section of students between the grades? After all, freshman might feel differently about the healthy options than seniors so it will be important to have individuals from each grade weigh in on the lunch options.

This can be done with a **stratified random sample**. Stratified random sampling is a method where the population is subdivided into groups called strata. **Strata** are groups with homogeneous characteristic(s). They are separated by the characteristic that we think might affect the overall sample. This is to avoid having too many of the sample having this one characteristic that may affect the sample.

In the above example, it would look something like this: since 42 is 10% of the school's population, your survey should be 10% of each grade.

- 10% of the freshmen class of 100 is 10, so you would want to randomly select ten individuals from the freshman class to participate.
- 10% of the sophomore class of 110 is 11, so you would want to randomly select 11 individuals from the sophomore class to participate.
- 10% of the junior class of 120 is 12, so you would want to randomly select 12 individuals from the junior class to participate.
- 10% of the senior class of 90 is 9, so you would want to randomly select nine individuals from the senior class to participate.

Once the groups are in place, a simple random sample is carried out within each stratum, like putting names in a hat or assigning everyone a unique number and randomly selecting numbers. You can have as many strata as you please, but they must be roughly homogeneous.

Video Transcription

[MUSIC PLAYING] Pretend you've subdivided billiard balls into low, middle, and high numbers. To take a stratified random sample of the 15, this is what you do. Put all the low numbered balls in hat one. Put all the middle numbered balls in hat 2.

And finally, put all the high numbered balls in hat three. At that point, you'd randomly select two from each hat. The result would give you a stratified random sample of six billiard balls. You're guaranteed to have exactly two low numbers, exactly two middle numbers, and exactly two high numbers.

[MUSIC PLAYING]



TERM TO KNOW

Stratified Random Sample

A random sampling method where individuals are separated into homogeneous groups, then simple random samples are taken within each group.

Stratum/Strata

The homogeneous groups in a stratified random sample. All individuals in each stratum have something in common, and we would like to see how that affects the outcome of the sample.

2. Cluster Samples

When using a **cluster sample**, the population is divided into groups. These groups are called **clusters**. It's

important to note that these groups are natural groupings. They don't necessarily have anything in common, other than say, geography, typically. Therefore, we're going to take a random sample of clusters instead of a random sample of individuals.

Each individual in the cluster is going to be part of the sample if we select that cluster. So unlike the groups in a stratified random sample, the groups in a cluster sample aren't based on a characteristic or variable. The individuals in the cluster just happen to be near each other.

IN CONTEXT

Suppose you work at a potato chip company and it's your job to implement some quality control in the manufacturing department. Maybe you stand at the start of the assembly line and take a simple random sample of individual chips. That would work just fine.

However, it might be easier for you to sample some bags of chips. The bags of chips are clusters. You would then take a bag of chips off the assembly line and sample every chip in that bag for quality control. That's cluster sampling.

Similar to every sampling method, cluster sampling has pros and cons.

Advantages and Disadvantages for Cluster Sampling	
Advantages	Easier than a simple random sample, and often it doesn't cost as much
	Typically gives similar results because the clusters are fairly heterogeneous
Disadvantages	Risk that clusters are NOT heterogeneous--perhaps they do have some characteristic other than just being geographically different from each other that might affect the sample's findings.

 TERMS TO KNOW

Cluster Sample

A sampling method where the population is separated into groups, typically geographically, and a random selection of clusters is made. Each individual in the cluster becomes part of the sample.

Clusters

Smaller subgroups of the population, not necessarily similar in any way besides all being together in one place, making the individuals easier to sample together.

3. Real-World Comparison

Suppose a landlord of an apartment complex wants to know whether a new carpet he's considering is appropriate for all the apartments in the building. Each of the four floors has eight apartments.

What would a simple random sample look like? How might a cluster sample be different from a stratified random sample?

Well, he could randomly select eight apartments from the building, and that would be a simple random sample.

He could randomly select two apartments per floor, and that would be a stratified random sample.

Or, a third option would be a cluster sample. He could take a spinner like the one shown below and spin it.



Suppose it landed on three. That means that every apartment on the third floor would receive carpeting. He doesn't have to have the carpet installers going to all these different rooms on all these different floors. He can simply instruct everyone to go up to the third floor and install carpet in every room on that floor, which would be far easier for him and just as cost effective. This would be a cluster sample, as opposed to some other type of sample.

But what if all the floors were NOT heterogeneous? What if apartments on the third floor allowed pets? The carpet might not hold up as well. That's one of the disadvantages of cluster sampling in action. But typically, the clusters are fairly representative and very similar to a simple random sample.



SUMMARY

In a stratified random sample, the population is broken down into homogeneous groups called "strata." The reason for this is to separate an otherwise homogeneous group that exhibits characteristics that may misrepresent the population. The idea is to force them into groups and then take a simple random sample within each of the strata. Cluster sampling, on the other hand, is done by taking naturally-occurring--typically geographically--similar groups and taking a simple random sample of the clusters. Then, each member of the cluster becomes part of the sample. A couple of advantages of cluster samples are that they are more cost effective, and usually achieve the same results as a simple random sample. The disadvantage is that sometimes the cluster may not be heterogeneous, as seen in the landlord example with pets allowed on carpet.

Good luck!

Source: This work is adapted from Sophia author Jonathan Osters.



TERMS TO KNOW

Cluster Sample

A sampling method where the population is separated into groups, typically geographically, and a random selection of clusters is made. Each individual in the cluster becomes part of the sample.

Clusters

Smaller subgroups of the population, not necessarily similar in any way besides all being together in one place, making the individuals easier to sample together.

Stratified Random Sample

A random sampling method where individuals are separated into homogenous groups, then simple random samples are taken within each group.

Stratum/Strata

The homogenous groups in a stratified random sample. All individuals in each stratum have something in common, and we would like to see how that affects the outcome of the sample.