Probability sampling methods

coursera.org/learn/the-power-of-statistics/supplement/ghBkL/probability-sampling-methods

Earlier, you learned that there are two main types of sampling methods: probability sampling and non-probability sampling. **Probability sampling** uses random selection to generate a sample . **Non-probability sampling** is often based on convenience, or the personal preferences of the researcher, rather than random selection. The sampling method you use helps determine if your sample is representative of your population, and if your sample is biased. Probability sampling gives you the best chance to create a sample that is representative of the population.

In this reading, you'll learn more about the different methods of probability sampling, and the benefits and drawbacks of each method.

Probability Sampling Methods

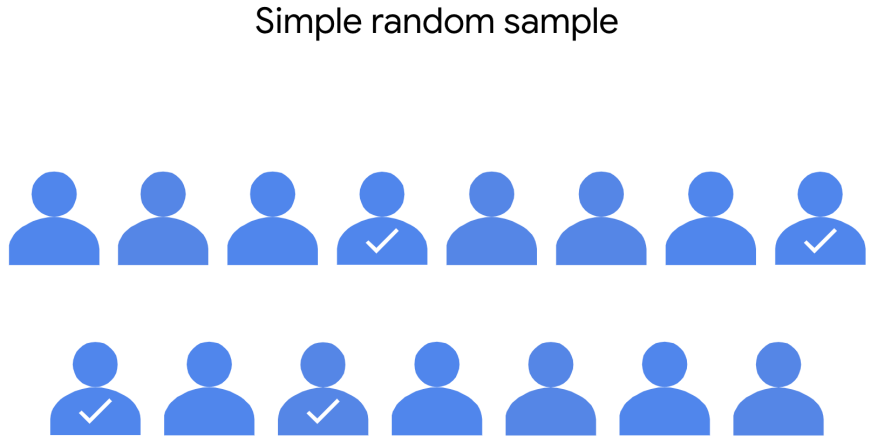
There are four different probability sampling methods:

- Simple random sampling
- Stratified random sampling
- Cluster random sampling
- Systematic random sampling

Let's explore each method in more detail.

Simple random sampling

In a **simple random sample**, every member of a population is selected randomly and has an equal chance of being chosen. You can randomly select members using a random number generator, or by another method of random selection.



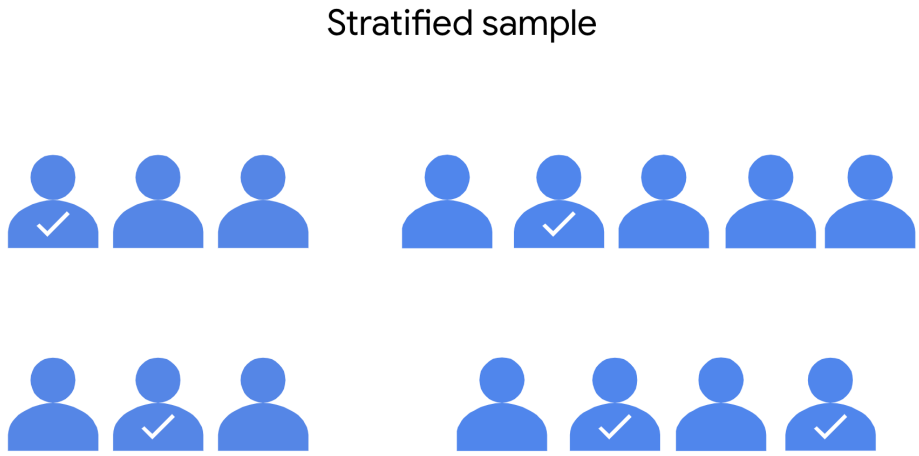
For example, imagine you want to survey the employees of a company about their work experience. The company employs 10,000 people. You can assign each employee in the company database a number from 1 to 10,000, and then use a random number generator to select 100 people for your sample. In this scenario, each of the employees has an equal chance of being chosen for the sample.

The main benefit of simple random samples is that they're usually fairly representative, since every member of the population has an equal chance of being chosen. Random samples tend to avoid bias, and surveys like these give you more reliable results.

However, in practice, it's often expensive and time-consuming to collect large simple random samples. And if your sample size is not large enough, a specific group of people in the population may be underrepresented in your sample. If you use a larger sample size, your sample will more accurately reflect the population.

Stratified random sampling

In a **stratified random sample**,you divide a population into groups, and randomly select some members from each group to be in the sample. These groups are called strata. Strata can be organized by age, gender, income, or whatever category you're interested in studying.



For example, imagine you're doing market research for a new product, and you want to analyze the preferences of consumers in different age groups. You might divide your target population into strata according to age: 20-29, 30-39, 40-49, 50-59, etc. Then, you can survey an equal number of people from each age group, and draw conclusions about the consumer preferences of each age group. Your results will help marketers decide which age groups to focus on to optimize sales for the new product.

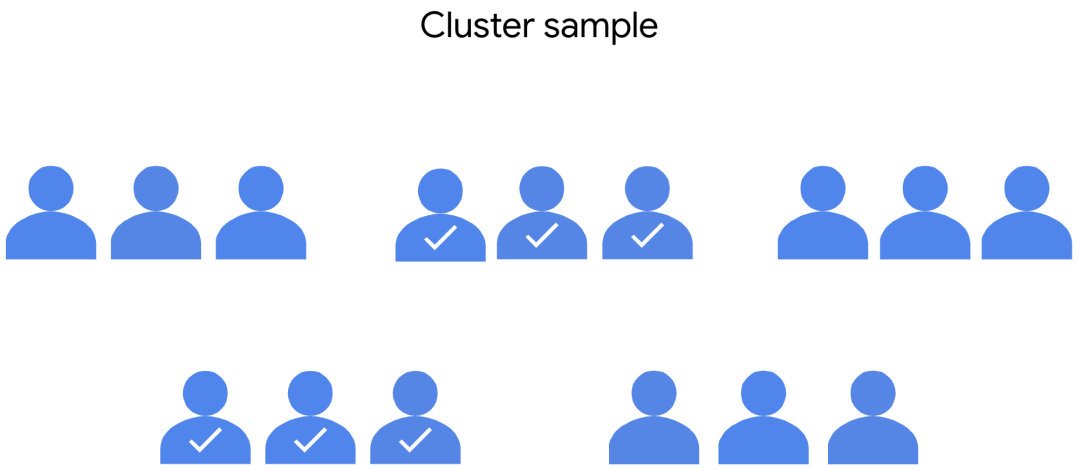
Stratified random samples help ensure that members from each group in the population are included in the survey. This method helps provide equal representation for underrepresented groups, and allows you to draw more precise conclusions about each of the strata. There may be significant differences in the purchasing habits of a 21-year-old and a 51-year-old. Stratified sampling helps ensure that both perspectives are captured in the sample.

One main disadvantage of stratified sampling is that it can be difficult to identify appropriate strata for a study if you lack knowledge of a population. For example, if you want to study median income among a population, you may want to stratify your sample by job type, or industry, or location, or education level. If you don't know how relevant these categories are to median income, it will be difficult to choose the best one for your study.

Cluster random sampling

When you're conducting a **cluster random sample**, you divide a population into clusters, randomly select certain clusters, and include all members from the chosen clusters in the sample.

Cluster sampling is similar to stratified random sampling, but in stratified sampling, you randomly choose *some* members from each group to be in the sample. In cluster sampling, you choose *all* members from a group to be in the sample. Clusters are divided using identifying details, such as age, gender, location, or whatever you want to study.



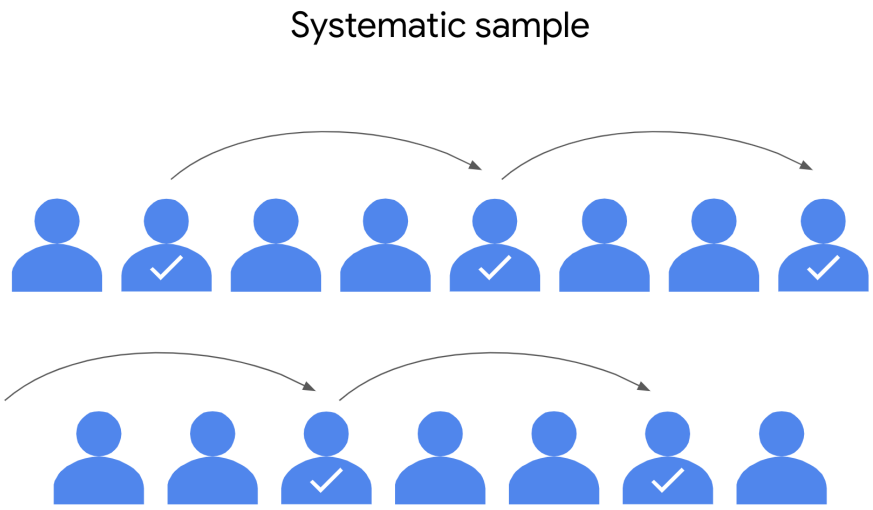
For example, imagine you want to conduct an employee satisfaction survey at a global restaurant franchise using cluster sampling. The franchise has 40 restaurants around the world. Each restaurant has about the same number of employees in similar job roles. You randomly select 4 restaurants as clusters. You include all the employees at the 4 restaurants in your sample.

One advantage of this method is that a cluster sample gets every member from a particular cluster, which is useful when each cluster reflects the population as a whole. This method is helpful when dealing with large and diverse populations that have clearly defined subgroups. If researchers want to learn more about home ownership in the suburbs of Auckland, New Zealand, they can use several well-chosen suburbs as a representative sample of all the suburbs in the city.

A main disadvantage of cluster sampling is that it may be difficult to create clusters that accurately reflect the overall population. For example, for practical reasons, you may only have access to restaurants in England when the franchise has locations all over the world. And employees in England may have different characteristics and values than employees in other countries.

Systematic random sampling

In a **systematic random sample**, you put every member of a population into an ordered sequence. Then, you choose a random starting point in the sequence and select members for your sample at regular intervals.



Imagine you want to survey students at a high school about their study habits. For a systematic random sample, you'd put the students' names in alphabetical order and randomly choose a starting point: say, number 4. Starting with number 4, you select every 10th name on the list (4, 14, 24, 34, …), until you have a sample of 100 students.

One advantage of systematic random samples is that they're often representative of the population, since every member has an equal chance of being included in the sample. Whether the student's last name starts with L or Q isn't going to affect their characteristics. Systematic sampling is also quick and convenient when you have a complete list of the members of your population.

One disadvantage of systematic sampling is that you need to know the size of the population that you want to study before you begin. If you don't have this information, it's difficult to choose consistent intervals. Plus, if there's a hidden pattern in the sequence, you might not get a representative sample. For example, if every 10th name on your list happens to be an honor student, you may only get feedback on the study habits of honor students – and not *all* students.

Key takeaways

The four methods of probability sampling we've covered—simple, stratified, cluster, and systematic—are all based on random selection, which is the preferred method of sampling for most data professionals. Probability sampling methods give you the best chance to create a sample that is representative of the population as a whole. And working with a representative sample allows you to make reliable inferences and accurate predictions about the population you're researching.