The stages of the sampling process

Recently, you've been learning about sampling. As a data professional, you'll work with sample data all the time. Often, this will be sample data previously collected by other researchers; sometimes, your team may collect their own data. Either way, it's important to know how the sampling process works, because it helps determine whether your sample is representative of the population, and whether your sample is unbiased.

In this reading, we'll go over the main stages of the sampling process in more detail. This will give you a better understanding of how the sampling process works and how each step of the process can affect your sample data.

The sampling process

First, let's review the main steps of the sampling process:

1. Identify the target population
2. Select the sampling frame
3. Choose the sampling method
4. Determine the sample size
5. Collect the sample data

Let's explore each step in more detail with an example. Imagine you're a data professional working for a company that manufactures home appliances. The company wants to find out how customers feel about the innovative digital features on their newest refrigerator model. The refrigerator has been on the market for two years and 10,000 people have purchased it. Your manager asks you to conduct a customer satisfaction survey and share the results with stakeholders.

Step 1: Identify the target population

The first step in the sampling process is defining your target population. The **target population** is the complete set of elements that you're interested in knowing more about. Depending on the context of your research, your population may include individuals, organizations, objects, events, or any other type of data you want to investigate.

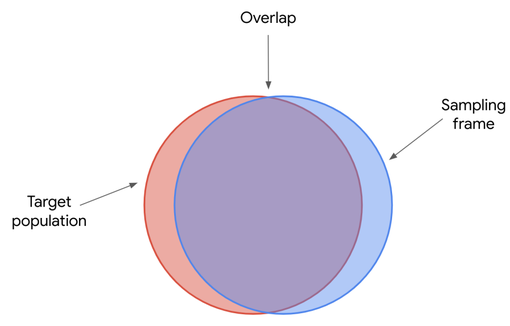
A well-defined population reduces the probability of including participants who do not fit the precise scope of your research. For example, you don't want to include all the company's customers, or customers who purchased the company's other refrigerator models.

In this case, your target population will be the 10,000 customers who purchased the company's newest refrigerator model. These are the customers you want to survey to learn about their experience with the newest model.

Step 2: Select the sampling frame

The next step in the sampling process is to create a sampling frame. A **sampling frame** is a list of all the individuals or items in your target population.

The difference between a target population and a sampling frame is that the population is general and the frame is specific. So, if your target population is all the customers who purchased the refrigerator, your sampling frame could be an alphabetical list of the names of all these customers. The customers in your sample will be selected from this list. Ideally, your sampling frame should include the entire target population. However, for practical reasons, your sampling frame may not exactly match your target population, because you may not have access to every member of the population. For instance, the company's customer database may be incomplete, or contain data processing errors. Or, some customers may have changed their contact information since their purchase, and you may be unable to locate or contact them.



Your sampling frame is the *accessible* part of your target population.

Step 3: Choose the sampling method

The third step in the sampling process is choosing a sampling method.

There are two main types of sampling methods: **probability sampling** and **non-probability sampling**. Later on, we'll explore the specific methods in more detail. For now, just know that 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. Often, probability sampling methods require more time and resources than non-probability sampling methods.

Ideally, your sample will be representative of the population. One way to help ensure that your sample is representative is to choose the right sampling method. Because probability sampling methods are based on random selection, every element in the population has an equal chance of being included in the sample. This gives you the best chance to get a representative sample, as your results are more likely to accurately reflect the overall population.

So, assuming you have the budget, the resources, and the time, you should use a probability sampling method for your survey.

Step 4: Determine the sample size

Step four of the sampling process is to determine the best size for your sample, since you don't have the resources to survey everyone in your sampling frame. In statistics, sample size refers to the number of individuals or items chosen for a study or experiment.

Sample size helps determine the precision of the predictions you make about the population. In general, the larger the sample size, the more precise your predictions. However, using larger samples typically requires more resources.

The sample size you choose depends on various factors, including the sampling method, the size and complexity of the target population, the limits of your resources, your timeline, and the goal of your research.

Based on these factors, you can decide how many customers to include in your sample.

Step 5: Collect the sample data

Now, you're ready to collect your sample data, which is the final step in the sampling process.

You give a customer satisfaction survey to the customers selected for your sample. The survey responses provide useful data on how customers feel about the digital features of the refrigerator. Then, you share your results with stakeholders to help them make more informed decisions about whether to continue to invest in these features for future versions of this refrigerator, and develop similar features for other models.

Key takeaways

Effective sampling ensures that your sample data is representative of your population. Then, when you use sample data to make inferences about the population, you can be reasonably confident that your inferences are reliable.

The decisions you make at each step of the sampling process can affect the quality of your sample data. Understanding the sampling process will make you a better data professional, whether you're analyzing data collected by other researchers or conducting a survey on your own.

Mark as completed