

What is the difference between sampling with replacement and sampling without replacement?

Authored by
stats writer

April 24, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the difference between sampling with replacement and sampling without replacement?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=138618>

Sampling with replacement refers to a method of selecting a sample from a population where each element has an equal chance of being chosen more than once. This means that after an element is selected, it is returned to the population before the next selection is made. On the other hand, sampling without replacement involves selecting a sample from a population where each element can only be chosen once. Once an element is selected, it is removed from the population and cannot be chosen again. This results in a smaller and more diverse sample compared to sampling with replacement. Ultimately, the main difference between the two methods is the potential for repeated selection of the same element in sampling with replacement, whereas sampling without replacement ensures each element is only chosen once.

Sampling With Replacement vs. Without Replacement

Often in statistics we're interested in collecting data so that we can answer some research question.

For example, we might want to answer the following questions:

- 1. What is the median household income in Cincinnati, Ohio?**
- 2. What is the mean weight of a certain population of turtles?**
- 3. What percentage of residents in a certain county support a certain law?**

In each scenario, we are interested in answering some question about a , which represents every possible

individual element that we're interested in measuring.

However, instead of collecting data on every individual in a population we typically just collect data on a sample of the population, which represents a portion of the population.

There are two different ways to collect samples: Sampling with replacement and sampling without replacement.

This tutorial explains the difference between the two methods along with examples of when each is used in practice.

Sampling with Replacement

Suppose we have the names of 5 students in a hat:

AndyKarlTylerBeccaJessica

Suppose we would like to take a sample of 2 students with replacement.

On the first random draw, we might select the name Tyler. We would then place his name back in the hat and draw again. On the second draw, we might select

the name Tyler again. Thus our sample would be: {Tyler, Tyler}

This is an example of obtaining a sample with replacement because we replace the name we choose after each random draw.

When we sample with replacement, the items in the sample are independent because the outcome of one random draw is not affected by the previous draw.

For example, the probability of choosing the name Tyler is $1/5$ on the first draw and $1/5$ again on the second draw. The outcome of the first draw does not affect the probability of the outcome on the second draw.

In each of these methods, sampling with replacement is used because it allows us to use the same dataset multiple times to build models as opposed to going out and gathering new data, which can be time-consuming and expensive.

Sampling without Replacement

Again, suppose we have the names of 5 students in a hat:

AndyKarlTylerBeccaJessica

Suppose we would like to take a sample of 2 students without replacement.

On the first random draw, we might select the name Tyler. We would then leave his name out of the hat. On the second draw, we might select the name Andy. Thus our sample would be: {Tyler, Andy}

This is an example of obtaining a sample without replacement because we do not replace the name we choose after each random draw.

When we sample without replacement, the items in the sample are dependent because the outcome of one random draw is affected by the previous draw.

For example, the probability of choosing the name Tyler is $1/5$ on the first draw and the probability of choosing the name Andy is $1/4$ on the second draw. The outcome of the first draw affects the probability of the outcome on the second draw.

Sampling without replacement is the method we use when we want to select a from a population.

For example, if we want to estimate the median household income in Cincinnati, Ohio there might be a total of 500,000 different households.

Thus, we might want to collect a random sample of 2,000 households but we don't want the data for any given household to appear twice in the sample so we would sample without replacement.

In other words, once we've chosen a certain household to be included in the sample we don't want there to be any chance of selecting that household to be included again.