Sampling With or Without Replacement

by Sophia

WHAT'S COVERED

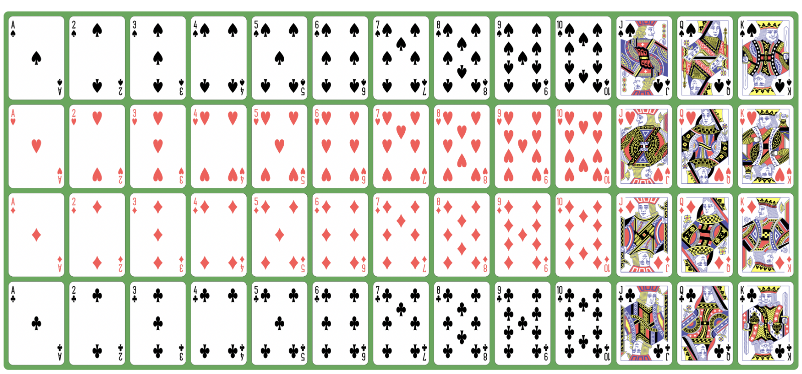
This tutorial will cover sampling, both with and without replacement. Our discussion breaks down as follows:

1. Sampling With Replacement

Sampling with replacement means that you put everything back once you've selected it.

Typically, one big requirement for statistical inference is that the individuals, the values from the sample, are independent. One doesn't affect any of the others. When sampling with replacement, each trial is independent.

EXAMPLE Consider a standard deck of 52 cards:



What is the probability that you draw a spade?

The probability of a spade on the first draw is 13 out of 52, or one-fourth.

$$P(\text{spade on first draw}) = \frac{13}{52} = \frac{1}{4}$$

Suppose you pull the 10 of spades, but then you put it back into the deck. Now, what's the probability of a spade on the second draw?

It's one fourth again. It's the same 52 cards. Therefore, you have the same likelihood of selecting a spade.

$$P(\textit{spade on second draw}) = \frac{1}{4}$$

BIG IDEA

When sampling with replacement, the trials are independent.

TERM TO KNOW

Sampling With Replacement

A sampling plan where each observation that is sampled is replaced after each time it is sampled, resulting in an observation being able to be selected more than once.

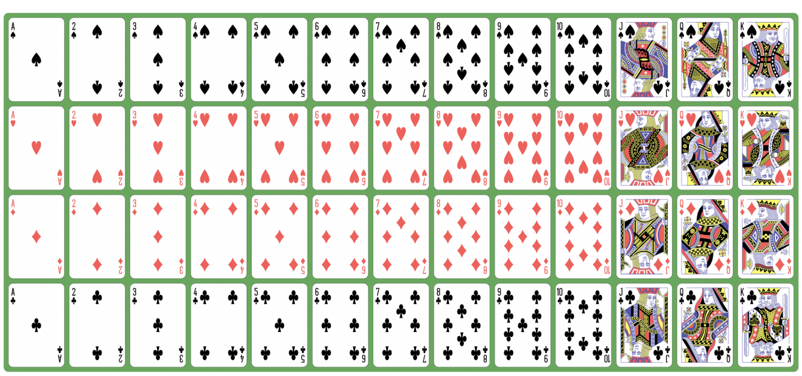
2. Sampling Without Replacement

Typically sampling with replacement will lead to independence, which is a requirement for a lot of statistical analysis. However, it's not often that you sample with replacement. It simply doesn't make sense to do this in real life.

EXAMPLE You wouldn't call a person twice for their opinion in a poll, so we don't put someone back into the population and see if you can sample them again.

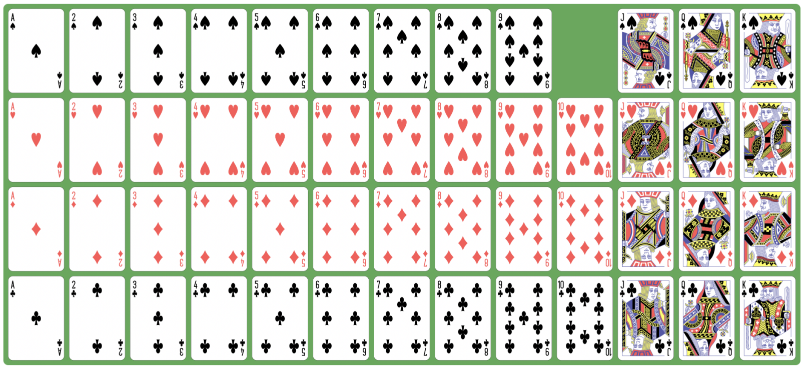
Most situations are considered **sampling without replacement**, which means that each observation is not put back once it's selected--once it's selected, it's out and cannot be selected again.

EXAMPLE Let's go back to the example with the standard deck of 52 cards. What is the probability that you select a spade on the first draw?



On the first draw, you have all 52 cards available, so the probability of drawing a spade is 13 out of 52, or one-fourth, as we had found before.

Suppose you drew the Ten of Spades and did not place it back in the deck of cards. Now, what's the probability of a spade on the second draw?



Now that there are only 12 spades left out of 51 cards, the probability of a spade on the second draw is not equal to one-fourth.

$$P(spade\ on\ second\ draw) = \frac{12}{51} \neq \frac{1}{4}$$

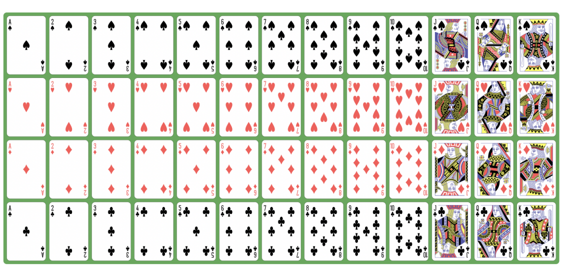
This means that the first draw and the second draw are dependent. The probability of a spade on the second draw changed after knowing that you got a spade on the first draw and did not replace it before drawing again.

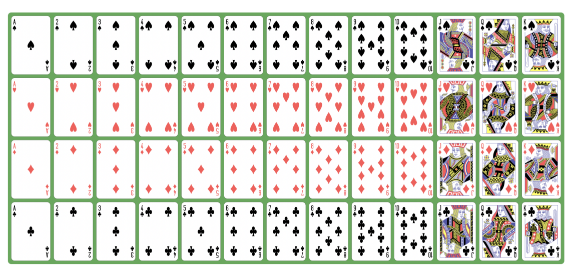
BIG IDEA

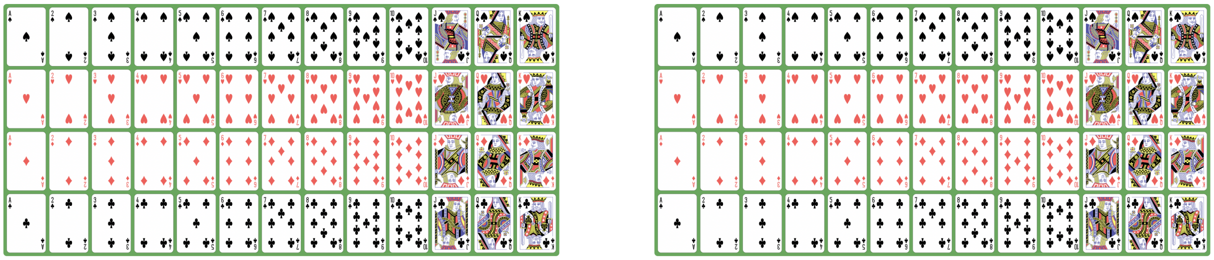
Even though the sampling that happens in real life doesn't technically fit the definition for independent observations, there's going to be a workaround.

EXAMPLE Suppose that your population was very large. Suppose you had four decks of cards, totaling 208 different cards.

What is the probability of drawing a diamond?



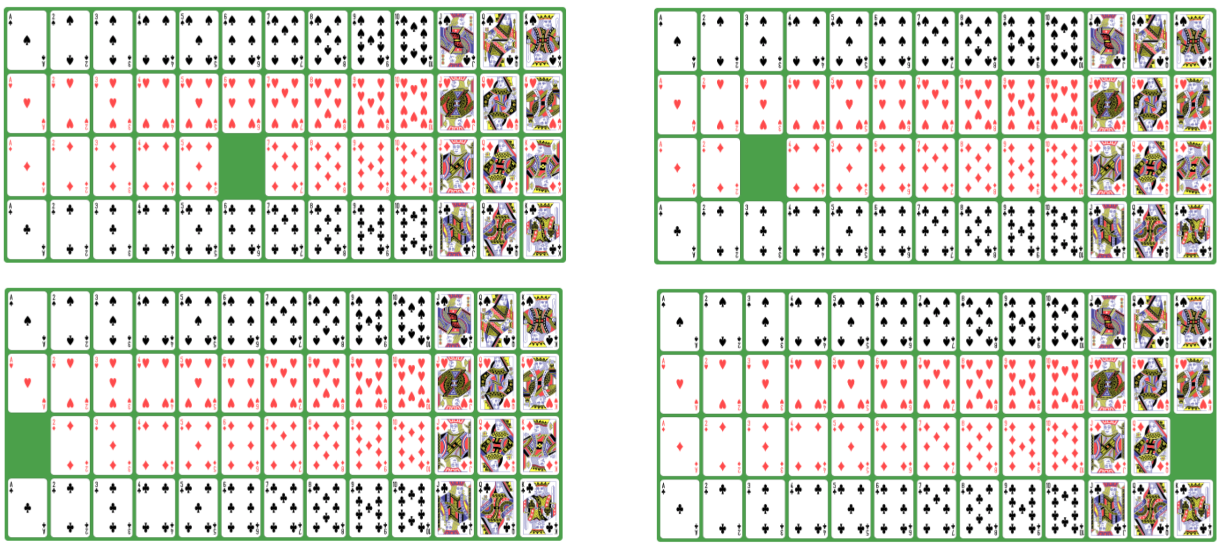




There are 52 diamonds out of 208 cards, so the probability of a diamond on the first draw is one-fourth probability, the same as if there were one deck.

$$P(\textit{diamond on first draw}) = \frac{52}{208} = 0.25$$

Suppose the worst case scenario happened in terms of independence and every card picked you have picked was the same suit. Take four diamonds from the group and do not replace them into the deck.



Now, what is the probability of drawing a diamond on the fifth draw?

There are now only 48 diamonds out of 204 cards remaining, so the probability of a diamond on the fifth draw is 48/204.

$$P(\textit{diamond on fifth draw}) = \frac{48}{204} = 0.235$$

The larger population actually has an effect now. The probability is about 0.24, which is different than 0.25, but not dramatically--even after five draws. The probability of a diamond didn't change particularly that much from the first to the last draw.

$$P(diamond\ on\ 1st\ draw) = \frac{52}{208} = 0.25$$

$$P(diamond\ on\ last\ draw) = \frac{48}{204} = 0.235$$

When you sample without replacement, if the population is large enough, then the probabilities don't shift very much as you sample. The sampling without replacement becomes almost independent because the probabilities don't change very much.

The question is, when is the population large enough? How large is considered a large population? You're going to institute a rule.

KEY CONCEPT

For independence, a large population is going to be at least 10 times larger than the sample.

$Population \geq 10n$

If that's the case, then you're going to say that the probabilities don't shift very much when you sample "n" items from the population. Therefore, you can treat the sampling as being almost independent.

TERM TO KNOW

Sampling without Replacement

A sampling plan where each observation that is sampled is kept out of subsequent selections, resulting in a sample where each observation can be selected no more than one time.

SUMMARY

Sampling with replacement is the gold standard, in a sense. It always creates independent trials. The probability of particular events doesn't change at all from trial to trial. However, in real life, when you sample without replacement, the probabilities do necessarily change. Your workaround is that if the population from which you're sampling is at least 10 times larger than the sample that you're drawing, the trials can be considered nearly independent.

Good luck!

Source: THIS TUTORIAL WAS AUTHORED BY JONATHAN OSTERS FOR SOPHIA LEARNING. PLEASE SEE OUR **TERMS OF USE**.

TERMS TO KNOW

Sampling With Replacement

A sampling plan where each observation that is sampled is replaced after each time it is sampled, resulting in an observation being able to be selected more than once.

Sampling Without Replacement

A sampling plan where each observation that is sampled is kept out of subsequent selections, resulting in a sample where each observation can be selected no more than one time.