Distribution of Sample Means

by Sophia

1. Distribution of Sample Means

A **distribution of sample means** is a distribution that shows the means from all possible samples of a given size.

⇨ EXAMPLE Consider the spinner shown here:

Suppose you spun it four times to obtain a sample mean. The first spin was a 2, the second spin was a 4, the third spin was a 3, and the fourth spin was a 1. The sample mean, then, would be 2.50. There's are many possible samples that could be taken of size 4 for this spinner, and there are many possible means that could arise from those samples, as shown below:

Sample	Mean

$S_1 = \{2, 4, 3, 1\}$	$\overline{x}_1 = 2.50$
$S_2 = \{1, 4, 3, 1\}$	$\overline{x}_2 = 2.25$
$S_3 = \{4, 2, 4, 4\}$	$\overline{x}_3 = 3.50$
$S_4 = \{2, 2, 3, 1\}$	$\overline{x}_4 = 2.00$
$S_5 = \{3, 1, 1, 1\}$	$\overline{x}_5 = 1.50$
$S_6 = \{1, 1, 1, 2\}$	$\overline{x}_6 = 1.25$

So how can we represent all these distributions?

⬡ STEP BY STEP

Step 1: First, take these sample means and graph them. Draw out an axis. For this one, it should go from 1 to 4 because this set can't average anything higher than four or lower than a one.

Step 2: Take the average value, for example, the mean of 2.5, and put a dot at 2.5 on the x-axis, much like a dot plot. Do this for all the sample means that you have found.
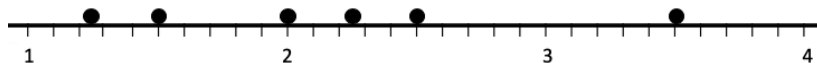
Step 3: You can keep doing this over and over again. Ideally, you would do this hundreds or thousands of times, to show the distribution of all possible samples that could be taken of size four. Once you've enumerated every possible sample of size four from this spinner, then the sampling distribution looks like this:
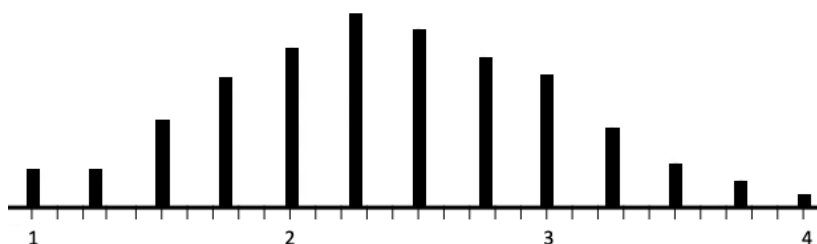
On the graph, the lowest number you can get is one, and the highest number you can get is four. On the far right of the graph is the point that represents a spin of 4 fours, {4, 4, 4, 4}. On the far left is the point that represents a spin of 4 ones, {1, 1, 1, 1}. Notice that 4 ones happens more than 4 fours. Why is that? If you take a look at the spinner, you'll see that there are more ones on the spinner than there are fours.

You can also notice that, since there are more ones, this actually pulls the average down a bit. The most

frequent average is 2.25, not 2.5, which would be the exact middle between 1 and 4. Therefore, this distribution is skewed slightly to the right because the numbers on the spinner are not evenly distributed.
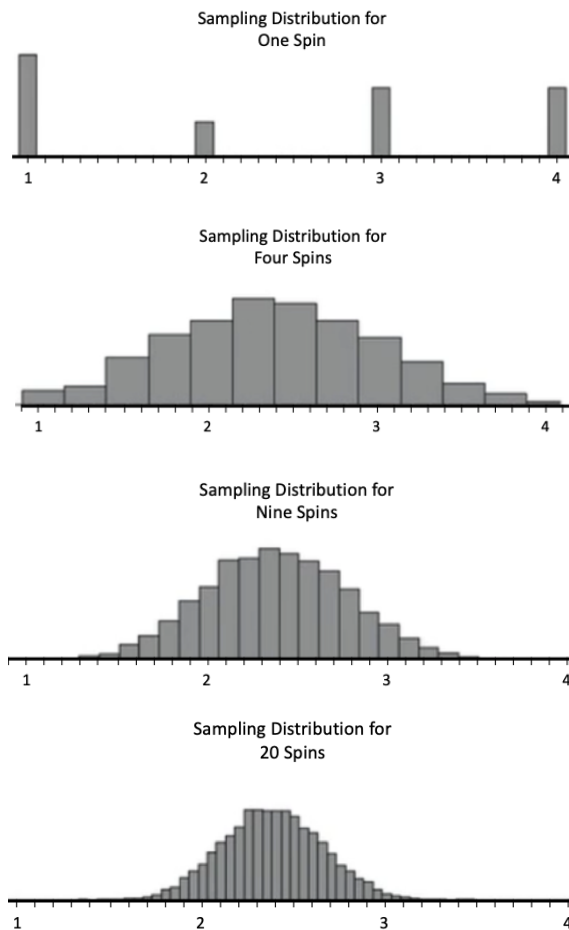
Distribution of Sample Means

A distribution where each data point consists of a mean of a collected sample. For a given sample size, every possible sample mean will be plotted in the distribution.

1a. Mean

⮑ EXAMPLE For the spinner above, the following are the histograms for a sample size of 1 spin, 4 spins, 9 spins, and 20 spins.
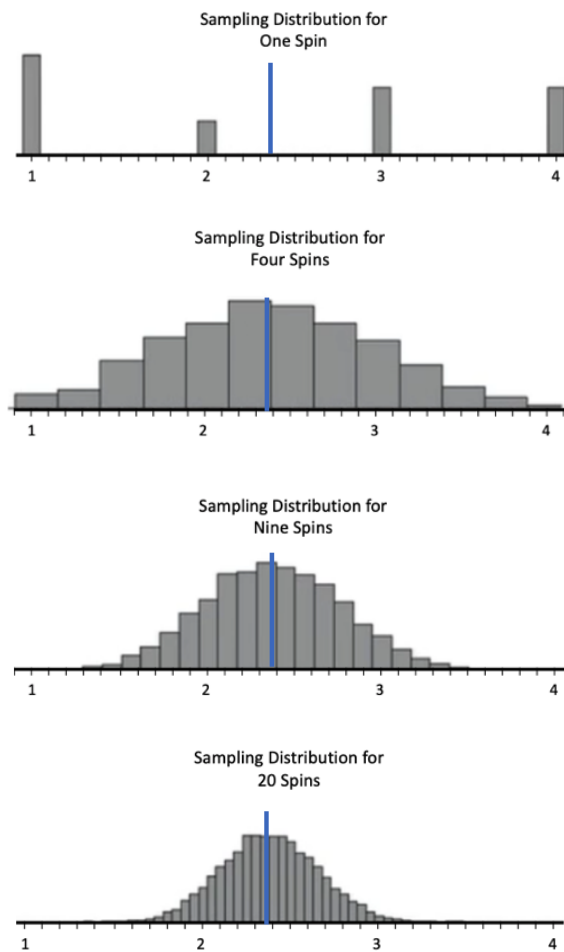
Sampling Distribution for One Spin

Sampling Distribution for Four Spins

Sampling Distribution for Nine Spins

Sampling Distribution for 20 Spins

With the sampling distribution when the sample size was 1, you'll see that 1 occurs about 3/8 of the time, 2 occurs about 1/8 of the time, 3 occurs about a fourth of the time, and 4 occurs about a fourth of the time. This produces a mean of 2.375:

$$mean = \frac{1+1+1+2+3+3+4+4}{8} = \frac{19}{8} = 2.375$$

You'll notice that when the sample size is 4, the shape of the distribution of sample means is significantly

different from when the sample size was 1. However, there some similarities and differences that you can recognize here about all four of these sampling distributions. The similarities are their centers--all of them are centered at 2.375. You'll notice that some of these are more tightly packed around that number--for instance, the samples of size 20 are more tightly packed around 2.375 than the samples of size 1--but they all are centered at that very same number.

Sampling Distribution for
One Spin

Sampling Distribution for
Four Spins

Sampling Distribution for
Nine Spins

Sampling Distribution for
20 Spins

What we can see here is that the mean of the sampling distribution of sample means is the same as the mean for the population. In this case, it was 2.375.

🧪 FORMULA TO KNOW

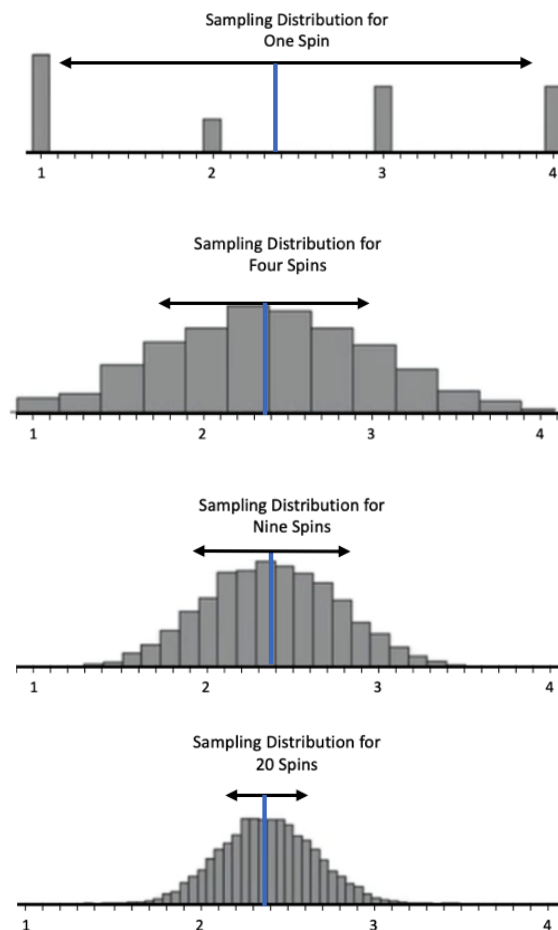**Mean of a Sampling Distribution of Sample Means**

$$\mu_{\bar{x}} = \mu_{population}$$

1b. Standard Deviation

↪ EXAMPLE How about the spread? The arrows on each of the histograms below indicate the standard deviation of each distribution.

Sampling Distribution for
One Spin

Sampling Distribution for
Four Spins

Sampling Distribution for
Nine Spins

Sampling Distribution for
20 Spins

Notice the arrows on the first distribution are very wide, and they seem to diminish in size as each distribution is graphed. When we get to the lowest distribution where the sample size was 20, its spread is much, much less.

The rule that's being followed is that the **standard deviation of a distribution of sample means** is the standard deviation of the population divided by the square root of sample size.

🧪 FORMULA TO KNOW

Standard Deviation of a Sampling Distribution of Sample Means

$$\sigma_{\bar{x}} = \frac{\sigma_{population}}{\sqrt{n}}$$

What that indicates is that when the sample size is 4, the standard deviation of that sampling distribution of sample means is going to be half as large as it was when the sample size was one. When the sample size is 9, it's going to be a third the size of the original standard deviation. And when n is 20, it's going to be the original standard deviation divided by the square root of 20.

🚩 HINT

The standard deviation of the sampling distribution is also called the **standard error**.

> 📄 TERMS TO KNOW
>
> **Standard Deviation of a Distribution of Sample Means**
> The standard deviation of the population, divided by the square root of sample size.
>
> **Standard Error**
> The standard deviation of the sampling distribution of sample means.

1c. Shape

↪ EXAMPLE Lastly, let's discuss the measured center, or measured spread, and describe the shape of these distributions. You'll notice that the shape is becoming more and more like the normal distribution as the sample size increases. There's a theorem that describes that, called the central limit theorem.

The **Central Limit Theorem** states that when the sample size is large (at least 30 for most distributions with a finite standard deviation), the sampling distribution of the sample means is approximately normal.

This means we can use the normal distribution to calculate probabilities on them, which is nice because normal calculations are easy to do.

Therefore, it's going to be normal, or approximately normal, with a mean of the same as that of the population, and a standard deviation equal to the standard deviation of the population divided by the square root of sample size.

> 📄 TERM TO KNOW
>
> **Central Limit Theorem**
> A theorem that explains the shape of a sampling distribution of sample means. It states that if the sample size is large (generally $n \geq 30$), and the standard deviation of the population is finite, then the distribution of sample means will be approximately normal.

> 📋 SUMMARY
>
> The distribution of sample means is called a sampling distribution of sample means. The sampling distribution of sample means has an approximately normal sampling distribution when the sample size is large. This is the Central Limit Theorem. The mean of the sampling distribution is the mean of the population. The standard deviation of the sampling distribution, which is also called the standard error, is the standard deviation of the population divided by the square root of the sample size.
>
> Good luck!

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

📄 **TERMS TO KNOW**

Central Limit Theorem

A theorem that explains the shape of a sampling distribution of sample means. It states that if the sample size is large (generally n ≥ 30), and the standard deviation of the population is finite, then the distribution of sample means will be approximately normal.

Distribution of Sample Means

A distribution where each data point consists of a mean of a collected sample. For a given sample size, every possible sample mean will be plotted in the distribution.

Standard Deviation of a Distribution of Sample Means

The standard deviation of the population, divided by the square root of sample size.

Standard Error

The standard deviation of the sampling distribution of sample means distribution.

🧪 **FORMULAS TO KNOW**

Mean of a Sampling Distribution of Sample Means

$$\mu_{\overline{x}} = \mu_{population}$$

Standard Deviation of a Sampling Distribution of Sample Means

$$\sigma_{\overline{x}} = \frac{\sigma_{population}}{\sqrt{n}}$$