

## Sampling Distribution of the Sample Mean, $\bar{x}$

**CO-6:** Apply basic concepts of probability, random variation, and commonly used statistical probability distributions.

### Behavior of the Sample Mean ( $\bar{x}$ )

#### Learning Objectives

**LO 6.22:** Apply the sampling distribution of the sample mean as summarized by the Central Limit Theorem (when appropriate). In particular, be able to identify unusual samples from a given population.

So far, we've discussed the behavior of the statistic  $\hat{p}$ , the sample proportion, relative to the parameter  $p$ , the population proportion (when the variable of interest is categorical).

We are now moving on to explore the behavior of the statistic  $\bar{x}$ , the sample mean, relative to the parameter  $\mu$  ( $\mu$ ), the population mean (when the variable of interest is quantitative).

Let's begin with an example.

#### ✓ EXAMPLE 9: Behavior of Sample Means

Birth weights are recorded for all babies in a town. The mean birth weight is 3,500 grams,  $\mu = \mu = 3,500$  g. If we collect many random samples of 9 babies at a time, how do you think sample means will behave?

Here again, we are working with a random variable, since random samples will have means that vary unpredictably in the short run but exhibit patterns in the long run.

Based on our intuition and what we have learned about the behavior of sample proportions, we might expect the following about the distribution of sample means:

**Center:** Some sample means will be on the low side — say 3,000 grams or so — while others will be on the high side — say 4,000 grams or so. In repeated sampling, we might expect that the random samples will average out to the underlying population mean of 3,500 g. In other words, the mean of the sample means will be  $\mu$  ( $\mu$ ), just as the mean of sample proportions was  $p$ .

**Spread:** For large samples, we might expect that sample means will not stray too far from the population mean of 3,500. Sample means lower than 3,000 or higher than 4,000 might be surprising. For smaller samples, we would be less surprised by sample means that varied quite a bit from 3,500. In other words, we might expect greater variability in sample means for smaller samples. So sample size will again play a role in the spread of the distribution of sample measures, as we observed for sample proportions.

**Shape:** Sample means closest to 3,500 will be the most common, with sample means far from 3,500 in either direction progressively less likely. In other words, the shape of the distribution of sample means should bulge in the middle and taper at the ends with a shape that is somewhat normal. This, again, is what we saw when we looked at the sample proportions.

#### Comment:

- The **distribution** of the values of the sample mean ( $\bar{x}$ ) in repeated **samples** is called the **sampling distribution of  $\bar{x}$** .

Let's look at a simulation:

#### Video

**Video:** [Simulation #3 \( \$\bar{x}\$ \)](#) (4:31)

**Did I Get This?:** [Simulation #3 \( \$\bar{x}\$ \)](#)

The results we found in our simulations are not surprising. Advanced probability theory confirms that by asserting the following:

## The Sampling Distribution of the Sample Mean

If repeated random samples of a given size  $n$  are taken from a population of values for a quantitative variable, where the population mean is  $\mu$  (mu) and the population standard deviation is  $\sigma$  (sigma) then the mean of all sample means ( $\bar{x}$ ) is population mean  $\mu$  (mu).

As for the spread of all sample means, theory dictates the behavior much more precisely than saying that there is less spread for larger samples. In fact, the standard deviation of all sample means is directly related to the sample size,  $n$  as indicated below.

$$\text{The standard deviation of all sample means } (\bar{x}) \text{ is exactly } \frac{\sigma}{\sqrt{n}}$$

Since the square root of sample size  $n$  appears in the denominator, the standard deviation does decrease as sample size increases.

### Learn by Doing: [Sampling Distribution \(x-bar\)](#)

Let's compare and contrast what we now know about the sampling distributions for sample means and sample proportions.

Variable	Parameter	Statistic	Sampling Distribution		
			Center	Spread	Shape
Categorical (example: left-handed or not)	$p$ = population proportion	$\hat{p}$ = sample proportion	$p$	$\sqrt{\frac{p(1-p)}{n}}$	Normal IF $np \geq 10$ and $n(1-p) \geq 10$
Quantitative (example: age)	$\mu$ = population mean, $\sigma$ = population standard deviation	$\bar{x}$ = sample mean	$\mu$	$\frac{\sigma}{\sqrt{n}}$	<b>When will the distribution of sample means be approximately normal ?</b>

Now we will investigate the shape of the sampling distribution of sample means. When we were discussing the sampling distribution of sample proportions, we said that this distribution is approximately normal if  $np \geq 10$  and  $n(1-p) \geq 10$ . In other words, we had a guideline based on sample size for determining the conditions under which we could use normal probability calculations for sample proportions.

When will the distribution of sample means be approximately normal? Does this depend on the size of the sample?

It seems reasonable that a population with a normal distribution will have sample means that are normally distributed even for very small samples. We saw this illustrated in the previous simulation with samples of size 10.

What happens if the distribution of the variable in the population is heavily skewed? Do sample means have a skewed distribution also? If we take really large samples, will the sample means become more normally distributed?

In the next simulation, we will investigate these questions.

### Video

**Video:** [Simulation #4 \(x-bar\)](#) (5:02)

### Did I Get This?: [Simulation #4 \(x-bar\)](#)

To summarize, the distribution of sample means will be approximately normal as long as the sample size is large enough. This discovery is probably the single most important result presented in introductory statistics courses. It is stated formally as the **Central Limit Theorem**.

We will depend on the Central Limit Theorem again and again in order to do normal probability calculations when we use sample means to draw conclusions about a population mean. We now know that we can do this even if the population distribution is not

normal.

How large a sample size do we need in order to assume that sample means will be normally distributed? Well, it really depends on the population distribution, as we saw in the simulation. The general rule of thumb is that samples of size 30 or greater will have a fairly normal distribution regardless of the shape of the distribution of the variable in the population.

### Applet: [Sampling Distribution for a Sample Mean](#)

#### Comment:

- For categorical variables, our claim that sample proportions are approximately normal for large enough  $n$  is actually a special case of the Central Limit Theorem. In this case, we think of the data as 0's and 1's and the "average" of these 0's and 1's is equal to the proportion we have discussed.

Before we work some examples, let's compare and contrast what we now know about the sampling distributions for sample means and sample proportions.

Variable	Parameter	Statistic	Sampling Distribution		
			Center	Spread	Shape
Categorical (example: left-handed or not)	$p$ = population proportion	$\hat{p}$ = sample proportion	$p$	$\sqrt{\frac{p(1-p)}{n}}$	Normal if $np \geq 10$ and $n(1-p) \geq 10$
Quantitative (example: age)	$\mu$ = population mean, $\sigma$ = population standard deviation	$\bar{x}$ = sample mean	$\mu$	$\frac{\sigma}{\sqrt{n}}$	Normal if $n > 30$ (always normal if population is normal)

### Learn by Doing: [Using the Sampling Distribution of x-bar](#)

#### ✓ EXAMPLE 10: Using the Sampling Distribution of x-bar

Household size in the United States has a mean of 2.6 people and standard deviation of 1.4 people. It should be clear that this distribution is skewed right as the smallest possible value is a household of 1 person but the largest households can be very large indeed.

- (a) What is the probability that a randomly chosen household has more than 3 people?

A normal approximation should not be used here, because the distribution of household sizes would be considerably skewed to the right. We do not have enough information to solve this problem.

- (b) What is the probability that the mean size of a random sample of 10 households is more than 3?

By anyone's standards, 10 is a small sample size. The Central Limit Theorem does not guarantee sample mean coming from a skewed population to be approximately normal unless the sample size is large.

- (c) What is the probability that the mean size of a random sample of 100 households is more than 3?

Now we may invoke the Central Limit Theorem: even though the distribution of household size  $X$  is skewed, the distribution of sample mean household size ( $\bar{x}$ ) is approximately normal for a large sample size such as 100. Its mean is the same as the population mean, 2.6, and its standard deviation is the population standard deviation divided by the square root of the sample size:

$$\frac{\sigma}{\sqrt{n}} = \frac{1.4}{\sqrt{100}} = 0.14$$

To find

$$P(\bar{x} > 3)$$

we standardize 3 to into a z-score by subtracting the mean and dividing the result by the standard deviation (of the sample mean). Then we can find the probability using the standard normal calculator or table.

$$P(\bar{x} > 3) = P\left(Z > \frac{3 - 2.6}{\frac{1.4}{\sqrt{100}}}\right) = P(Z > 2.86) = 0.0021$$

Households of more than 3 people are, of course, quite common, but it would be extremely unusual for the mean size of a sample of 100 households to be more than 3.

The purpose of the next activity is to give guided practice in finding the sampling distribution of the sample mean ( $\bar{x}$ ), and use it to learn about the likelihood of getting certain values of  $\bar{x}$ .

**Learn by Doing:** [Using the Sampling Distribution of  \$\bar{x}\$  #2](#)

**Did I Get This?:** [Using the Sampling Distribution of  \$\bar{x}\$](#)

[Sampling Distribution of the Sample Mean,  \$\bar{x}\$](#)  is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by LibreTexts.