

6.1: The Sampling Distribution of Means

Learning Objectives

- To learn what the sampling distribution of \bar{X} is when the sample size is large.
- To learn what the sampling distribution of \bar{X} is when the population is normal.

In Example 6.1.1, we constructed the probability distribution of the sample mean for samples of size two drawn from the population of four rowers. The probability distribution is:

\bar{x}	152	154	156	158	160	162	164
$P(\bar{x})$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

Figure 6.1.1 shows a side-by-side comparison of a histogram for the original population and a histogram for this distribution. Whereas the distribution of the population is uniform, the sampling distribution of the mean has a shape approaching the shape of the familiar bell curve. This phenomenon of the sampling distribution of the mean taking on a bell shape even though the population distribution is not bell-shaped happens in general. Here is a somewhat more realistic example.

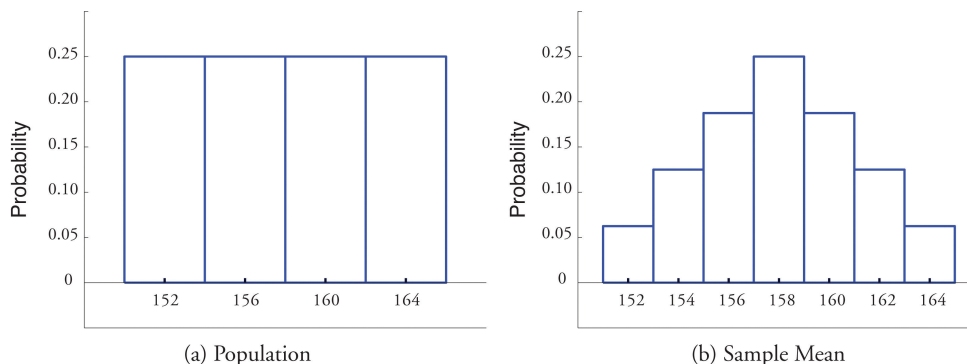


Figure 6.1.1: Distribution of a Population and a Sample Mean

Suppose we take samples of size 1, 5, 10, or 20 from a population that consists entirely of the numbers 0 and 1, half the population 0, half 1, so that the population mean is 0.5. The sampling distributions are:

$n = 1$:

\bar{x}	0	1
$P(\bar{x})$	0.5	0.5

$n = 5$:

\bar{x}	0	0.2	0.4	0.6	0.8	1
$P(\bar{x})$	0.03	0.16	0.31	0.31	0.16	0.03

$n = 10$:

\bar{x}	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$P(\bar{x})$	0.00	0.01	0.04	0.12	0.21	0.25	0.21	0.12	0.04	0.01	0.00

$n = 20$:

\bar{x}	0	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$P(\bar{x})$	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.07	0.12	0.16	0.18

and

\bar{x}	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1
$P(\bar{x})$	0.16	0.12	0.07	0.04	0.01	0.00	0.00	0.00	0.00	0.00

Histograms illustrating these distributions are shown in Figure 6.1.2.

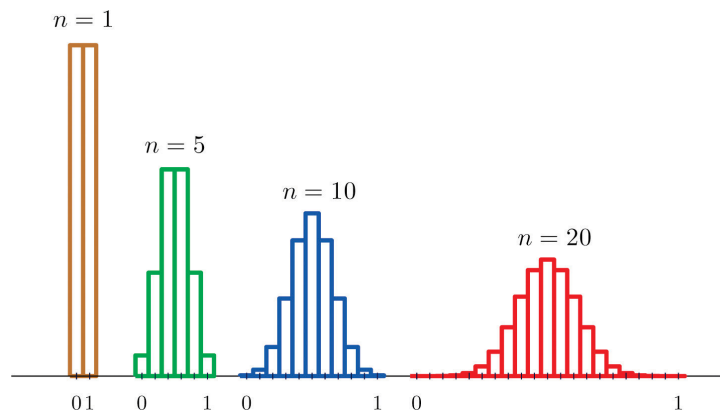


Figure 6.1.2: Distributions of the Sample Mean

As n increases the sampling distribution of \bar{X} evolves in an interesting way: the probabilities on the lower and the upper ends shrink and the probabilities in the middle become larger in relation to them. If we were to continue to increase n then the shape of the sampling distribution would become smoother and more bell-shaped.

What we are seeing in these examples does not depend on the particular population distributions involved. In general, one may start with any distribution and the sampling distribution of the sample mean will increasingly resemble the bell-shaped normal curve as the sample size increases. This is the content of the Central Limit Theorem.

The Central Limit Theorem

For samples of size 30 or more, the sample mean is approximately normally distributed, with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$, where n is the sample size. The larger the sample size, the better the approximation. The **Central Limit Theorem** is illustrated for several common population distributions in Figure 6.1.3.

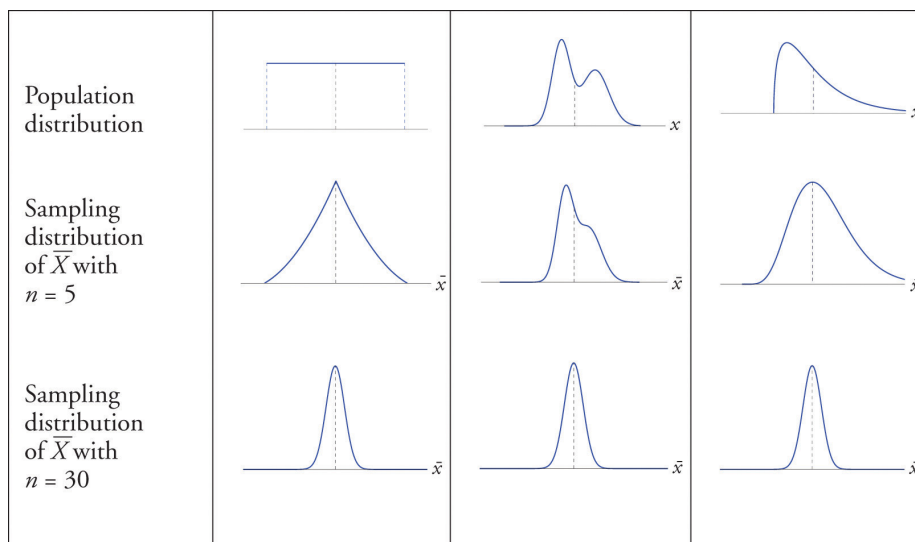


Figure 6.1.3: Distribution of Populations and Sample Means

The dashed vertical lines in the figures locate the population mean. Regardless of the distribution of the population, as the sample size is increased the shape of the sampling distribution of the sample mean becomes increasingly bell-shaped, centered on the population mean. Typically by the time the sample size is 30 the distribution of the sample mean is practically the same as a normal distribution.

The importance of the Central Limit Theorem is that it allows us to make probability statements about the sample mean, specifically in relation to its value in comparison to the population mean, as we will see in the examples. But to use the result properly we must first realize that there are two separate random variables (and therefore two probability distributions) at play:

1. X , the measurement of a single element selected at random from the population; the distribution of X is the distribution of the population, with mean the population mean μ and standard deviation the population standard deviation σ ;
2. \bar{X} , the mean of the measurements in a sample of size n ; the distribution of \bar{X} is its sampling distribution, with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

✓ Example 6.1.1

Let \bar{X} be the mean of a random sample of size 50 drawn from a population with mean 112 and standard deviation 40.

1. Find the mean and standard deviation of \bar{X} .
2. Find the probability that \bar{X} assumes a value between 110 and 114.
3. Find the probability that \bar{X} assumes a value greater than 113.

Solution

1. By the formulas in the previous section

$$\mu_{\bar{X}} = \mu = 112$$

and

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{50}} = 5.65685$$

2. Since the sample size is at least 30, the Central Limit Theorem applies: \bar{X} is approximately normally distributed. We compute probabilities using Figure 5.3.1 in the usual way, just being careful to use $\sigma_{\bar{X}}$ and not σ when we standardize:

$$\begin{aligned} P(110 < \bar{X} < 114) &= P\left(\frac{110 - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < Z < \frac{114 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) \\ &= P\left(\frac{110 - 112}{5.65685} < Z < \frac{114 - 112}{5.65685}\right) \\ &= P(-0.35 < Z < 0.35) \\ &= 0.6368 - 0.3632 \\ &= 0.2736 \end{aligned}$$

3. Similarly

$$\begin{aligned} P(\bar{X} > 113) &= P\left(Z > \frac{113 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) \\ &= P\left(Z > \frac{113 - 112}{5.65685}\right) \\ &= P(Z > 0.18) \\ &= 1 - P(Z < 0.18) \\ &= 1 - 0.5714 \\ &= 0.4286 \end{aligned}$$

Note that if in the above example we had been asked to compute the probability that the value of a single randomly selected element of the population exceeds 113, that is, to compute the number $P(X > 113)$, we would not have been able to do so, since we do not know the distribution of X , but only that its mean is 112 and its standard deviation is 40. By contrast we could compute

$P(\bar{X} > 113)$ even without complete knowledge of the distribution of X because the Central Limit Theorem guarantees that \bar{X} is approximately normal.

✓ Example 6.1.2

The numerical population of grade point averages at a college has mean 2.61 and standard deviation 0.5. If a random sample of size 100 is taken from the population, what is the probability that the sample mean will be between 2.51 and 2.71?

Solution

The sample mean \bar{X} has mean $\mu_{\bar{X}} = \mu = 2.61$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.5}{10} = 0.05$, so

$$\begin{aligned} P(2.51 < \bar{X} < 2.71) &= P\left(\frac{2.51 - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < Z < \frac{2.71 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) \\ &= P\left(\frac{2.51 - 2.61}{0.05} < Z < \frac{2.71 - 2.61}{0.05}\right) \\ &= P(-2 < Z < 2) \\ &= P(Z < 2) - P(Z < -2) \\ &= 0.9772 - 0.0228 \\ &= 0.9544 \end{aligned}$$

Normally Distributed Populations

The Central Limit Theorem says that no matter what the distribution of the population is, as long as the sample is “large,” meaning of size 30 or more, the sample mean is approximately normally distributed. If the population is normal to begin with then the sample mean also has a normal distribution, regardless of the sample size.

For samples of any size drawn from a normally distributed population, the sample mean is normally distributed, with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, where n is the sample size.

The effect of increasing the sample size is shown in Figure 6.1.4.

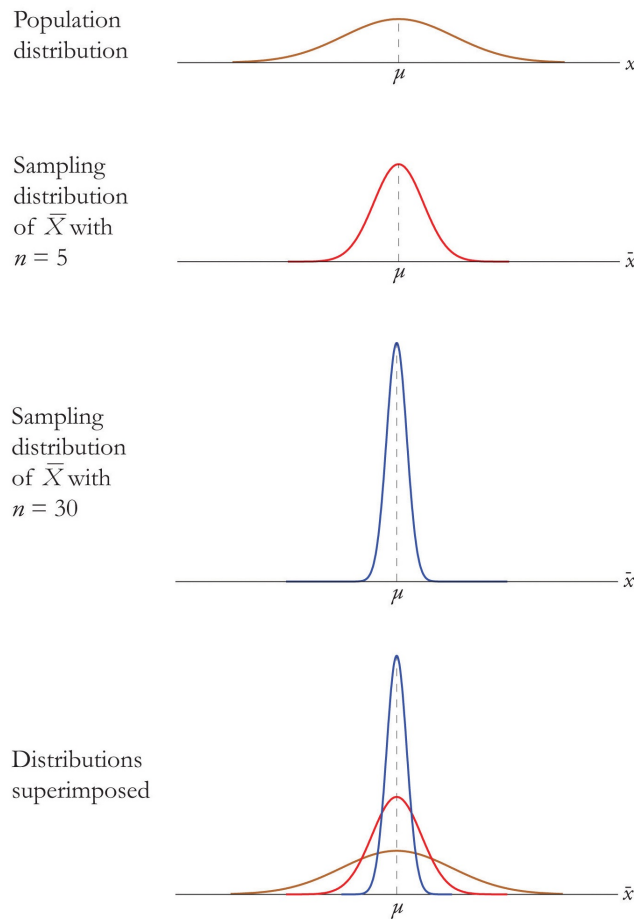


Figure 6.1.4: Distribution of Sample Means for a Normal Population

✓ Example 6.1.3

A prototype automotive tire has a design life of 38,500 miles with a standard deviation of 2,500 miles. Five such tires are manufactured and tested. On the assumption that the actual population mean is 38,500 miles and the actual population standard deviation is 2,500 miles, find the probability that the sample mean will be less than 36,000 miles. Assume that the distribution of lifetimes of such tires is normal.

Solution

For simplicity we use units of thousands of miles. Then the sample mean \bar{X} has mean $\mu_{\bar{X}} = \mu = 38.5$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{5}} = 1.11803$. Since the population is normally distributed, so is \bar{X} , hence

$$\begin{aligned} P(\bar{X} < 36) &= P\left(Z < \frac{36 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) \\ &= P\left(Z < \frac{36 - 38.5}{1.11803}\right) \\ &= P(Z < -2.24) \\ &= 0.0125 \end{aligned}$$

That is, if the tires perform as designed, there is only about a 1.25% chance that the average of a sample of this size would be so low.

✓ Example 6.1.4

An automobile battery manufacturer claims that its midgrade battery has a mean life of 50 months with a standard deviation of 6 months. Suppose the distribution of battery lives of this particular brand is approximately normal.

- On the assumption that the manufacturer's claims are true, find the probability that a randomly selected battery of this type will last less than 48 months.
- On the same assumption, find the probability that the mean of a random sample of 36 such batteries will be less than 48 months.

Solution

- Since the population is known to have a normal distribution

$$\begin{aligned}P(X < 48) &= P\left(Z < \frac{48 - \mu}{\sigma}\right) \\&= P\left(Z < \frac{48 - 50}{6}\right) \\&= P(Z < -0.33) \\&= 0.3707\end{aligned}$$

- The sample mean has mean $\mu_{\bar{X}} = \mu = 50$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{36}} = 1$. Thus

$$\begin{aligned}P(\bar{X} < 48) &= P\left(Z < \frac{48 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) \\&= P\left(Z < \frac{48 - 50}{1}\right) \\&= P(Z < -2) \\&= 0.0228\end{aligned}$$

Key Takeaway

- When the sample size is at least 30 the sample mean is normally distributed.
- When the population is normal the sample mean is normally distributed regardless of the sample size.

6.1: The Sampling Distribution of Means is shared under a CC BY-NC license and was authored, remixed, and/or curated by LibreTexts.

- 6.2: The Sampling Distribution of the Sample Mean by Anonymous is licensed CC BY-NC-SA 3.0. Original source: <https://2012books.lardbucket.org/books/beginning-statistics>.