

## 6.4: Sampling Distributions and the Central Limit Theorem

The sample mean, denoted  $\bar{x}$ , is the average of a sample of a variable  $X$ . The sample mean is an estimate of the population mean  $\mu$ . Every sample has a sample mean and these sample means differ (depending on the sample). Thus, before a sample is selected  $\bar{x}$  is a variable, in fact, if the sample is a random sample then  $\bar{x}$  is a random variable. For this reason, we can think of the “distribution of  $\bar{x}$ ,” called the “Sampling Distribution of  $\bar{x}$ ,” as the theoretical histogram constructed from the sample averages of all possible samples of size  $n$ .

### Mean and Standard Deviation of a Sample Mean

Let  $\bar{x}$  be the mean of a random sample of size  $n$  from a population having mean  $\mu$  and standard deviation  $\sigma$ , then

The mean of the sample means =  $\mu_{\bar{x}} = \mu$ .

The standard deviation (standard error) of the sample means =  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ .

This says that the mean of the sample means is the same as the population mean. The standard deviation of the sample means is the population standard deviation divided by the square root of the sample size. This is called the **sampling distribution** of the mean.

Let  $X$  be the height of men in the United States. Studies show that the heights of 15-year old boys in the United States are normally distributed with average height 67 inches and a standard deviation of 2.5 inches. A random experiment consists of choosing 16 15-year old boys at random. Compute the mean and standard deviation of  $\bar{x}$ , that is, the mean and standard deviation for the average height of a random sample of 16 boys.

### Solution

The mean of the sample means is the same as the population mean  $\mu_{\bar{x}} = 67$ .

The standard deviation in the sample means is the population standard deviation divided by the square root of the sample size,  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{16}} = 0.625$ .

Notice that the mean of a sample means is always the same as the mean of the population, but the standard deviation is smaller. See Figure 6-30.

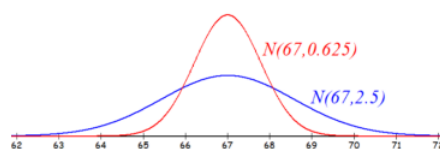
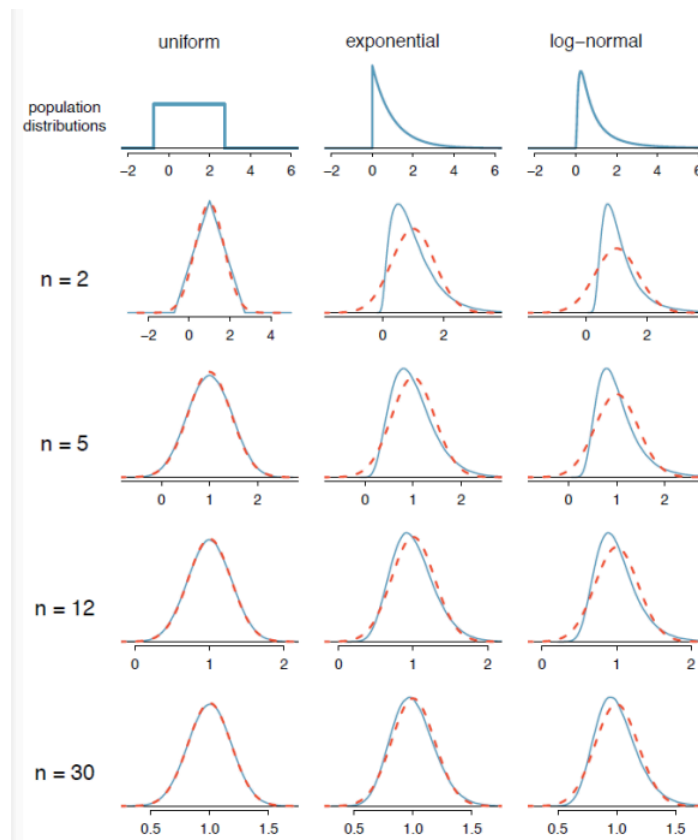


Figure 6-30

### Sampling Distribution of a Sample Mean

If a population is normally distributed  $N(\mu, \sigma)$ , then the sample mean  $\bar{x}$  of  $n$  independent observations is normally distributed as  $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ .

Figure 6-31 shows three population distributions and the corresponding sampling distributions for sample sizes of 2, 5, 12 and 30. Notice as the sample size gets larger, the sampling distribution gets closer to the dashed red line of the normal distribution. **Video explanation of this process:** [https://youtu.be/lsCc\\_pS3O28](https://youtu.be/lsCc_pS3O28).



Retrieved from [OpenIntroStatistics](#).

Figure 6-31

The Central Limit Theorem establishes that in some situations the distribution of the sample statistic will take on a normal distribution, even when the population is not normally distributed. This allows us to use the normal distribution to make inferences from samples to populations.

The Central Limit Theorem guarantees that the distribution of the sample mean will be normally distributed when the sample size is large (usually 30 or higher) no matter what shape the population distribution is.

### Finding Probabilities Using the Central Limit Theorem (CLT)

If we are finding the probability of a sample mean and have a sample size of 30 or more, or the population was normally distributed, then we can use the normal distribution to find the probability that the sample mean is below, above or between two values using the CLT.

Watch this video on using this applet for the Central Limit Theorem, and then take some time to play with the applet to get a sense of the difference between the distribution of the population, the distribution of a sample and the sampling distribution.

Watch the video on how to use the applet: <https://youtu.be/aIPvgiXyBMI>.

Try the applet on your own. Applet: [http://onlinestatbook.com/stat\\_sim/sampling\\_dist/index.html](http://onlinestatbook.com/stat_sim/sampling_dist/index.html).

The population of midterm scores for all students taking a PSU Business Statistics course has a known standard deviation of 5.27. The mean of the population is 18.07 and the median of the population is 19. A sample of 25 was taken and the sample mean was 18.07 and we want to know what the sampling distribution for the mean looks like. Figure 6-32 shows 3 graphs using the [Sampling Distribution Applet](#).

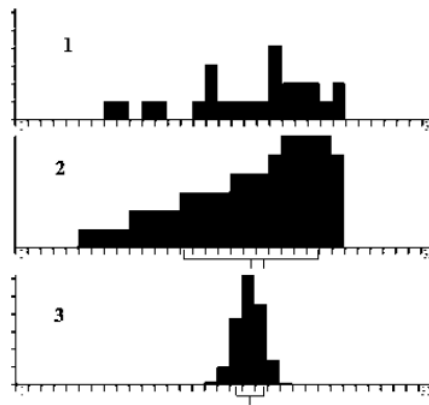


Figure 6-32

- What is the mean and standard deviation of the sampling distribution?
- Would you expect midterm exam scores to be skewed or bell-shaped?
- Which of these graphs in Figure 6-32 correspond to the distribution of the population, distribution of a single sample and the sampling distribution of the mean?
- Compute the probability that for next term's class they have a sample mean of more than 20.

### Solution

- By the Central Limit Theorem (CLT) the mean of the sampling distribution  $\mu_{\bar{x}}$  equals the mean of the population which was given as  $\mu=18.07$ . The standard deviation of the sampling distribution by the CLT would be the population standard deviation divided by the square root of the sample size  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5.27}{\sqrt{25}} = 1.054$
- The population mean = 18.07 is smaller than the median = 19 therefore the distribution is negatively skewed, the mean is pulled in the direction of the outliers.
- Using the [Sampling Distribution Applet](#) and the CLT, the sampling distribution will be bell-shaped therefore, graph 3 has to be the sampling distribution.

Graphs 1 & 2 in Figure 6-32 are both negatively skewed. A single sample of 25 should look similar to the entire population, but we would expect only 25 items and not every score possible would be received from the 25 students.

Graph 1 in Figure 6-32 fits this description and therefore the graph of the distribution of a single sample (which is not the same thing as the sampling distribution) is graph 1.

This leaves graph 2 as the distribution of the population. Figure 6-33 is a picture of the applet modeling the exam scores. Note the top picture is the population distribution, the second graph is simulating a single sample drawn and the bottom picture is a graph of all the sample means for each sample. This last graph is the sampling distribution of the means.

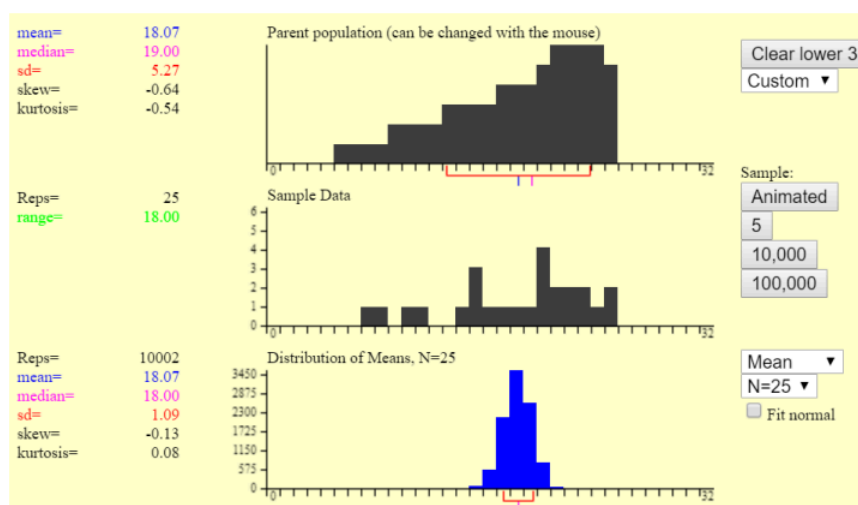


Figure 6-33

- The  $P(\bar{X} > 20)$  would be normally distributed with a mean  $\mu_{\bar{x}} = 18.07$  with a standard deviation of  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5.27}{\sqrt{25}} = 1.054$

Draw and shade the sampling distribution curve.

This calculator can be used to draw and shade the sampling distribution: <http://homepage.divms.uiowa.edu/~mbognar/applets/normal.html>, filling in the mean  $\mu$ , standard deviation  $\frac{\sigma}{\sqrt{n}}$  and x-value (in this case the sample mean) will find the probability. See Figure 6-34.

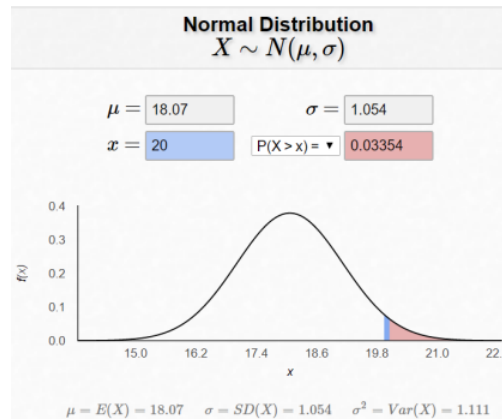


Figure 6-34

TI Calculator:  $\text{normalcdf}(20, 1E99, 18.07, 5.27/\sqrt{25}) = 0.0335$ .

Excel:  $P(\bar{X} > 20) = 1 - \text{NORM.DIST}(20, 18.07, 5.27/\text{SQRT}(25), \text{TRUE}) = 0.0335$ .

### Example 1

Let  $X$  be the height of 15-year old boys in the United States. Studies show that the heights of 15-year old boys in the United States are normally distributed with average height of 67 inches and a standard deviation of 2.5 inches. A random experiment consists of randomly choosing sixteen 15-year old boys. Compute the probability that the mean height of those sampled is 69.5 inches or taller.

### Solution

$$\mu = 67, \quad \sigma = 2.5, \quad n = 16 \quad (6.4.1)$$

$$SE_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{16}} = \frac{2.5}{4} = 0.625 \quad (6.4.2)$$

$$z = \frac{\bar{x} - \mu}{SE_{\bar{X}}} = \frac{69.5 - 67}{0.625} = \frac{2.5}{0.625} = 4 \quad (6.4.3)$$

Use the z-table to get the tail probability.

$$P(\bar{X} \geq 69.5) = P(Z \geq 4) = 1 - \Phi(4) \quad (6.4.4)$$

$$\Phi(4.00) \approx 0.99997 \Rightarrow P(Z \geq 4) \approx 1 - 0.99997 = 0.00003 \quad (6.4.5)$$

Therefore,

$$P(\bar{X} \geq 69.5) \approx 3 \times 10^{-5} \text{ (about 0.003\%)} \quad (6.4.6)$$

This should make sense because one would think that the likelihood of randomly selecting 16 boys that have an average height of 5'9.5" would be slim.

Figure 6-35 shows the density curves showing the shaded areas of  $P(X \geq 69.5)$  and  $P(\mu_{\bar{x}} \geq 69.5)$ .

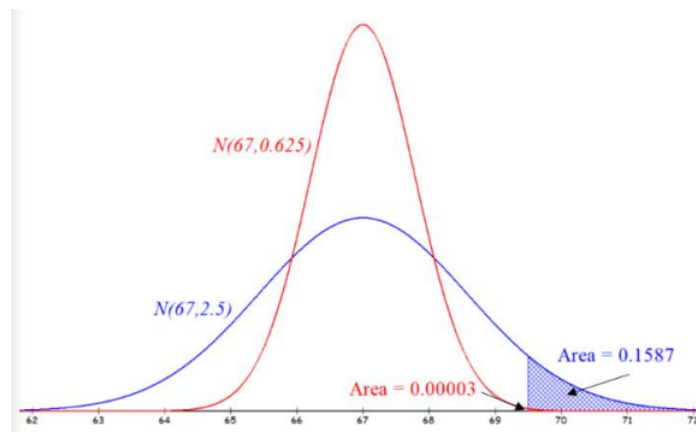


Figure 6-35

The sampling distribution has a much smaller spread (standard deviation) and hence less area to the right of 69.5.

In general, the Central Limit Theorem questions will use the same method as previous sections, however you will use a standard deviation of  $\frac{\sigma}{\sqrt{n}}$  and a z-score of  $z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$ .

### Example 2

The average teacher's salary in Connecticut (ranked first among states) is \$57,337. Suppose that the distribution of salaries is normally distributed with a standard deviation of \$7,500.

- What is the probability that a randomly selected teacher makes less than \$55,000 per year?
- If we sample 10 teachers' salaries, what is the probability that the sample mean is less than \$55,000?
- If we sample 100 teachers' salaries, what is the probability that the sample mean is less than \$55,000?

### Solution

- Probability that a randomly selected teacher makes less than \$55,000.

$$\mu = 57,337, \quad \sigma = 7,500, \quad x = 55,000 \quad (6.4.7)$$

$$z = \frac{x - \mu}{\sigma} = \frac{55,000 - 57,337}{7,500} = \frac{-2,337}{7,500} \approx -0.3116 \quad (6.4.8)$$

$$P(X < 55,000) = P(Z < -0.3116) \approx 1 - \Phi(0.3116) \quad (6.4.9)$$

$$\Phi(0.31) \approx 0.6217 \text{ (table)} \Rightarrow P(X < 55,000) \approx 1 - 0.622 \approx 0.378 \quad (6.4.10)$$

- Probability that the sample mean of  $n=10$  salaries is less than \$55,000.

$$n = 10, \quad SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{7,500}{\sqrt{10}} \approx 2,371.71 \quad (6.4.11)$$

$$z = \frac{55,000 - 57,337}{SE_{\bar{x}}} = \frac{-2,337}{2,371.71} \approx -0.985 \quad (6.4.12)$$

$$P(\bar{X} < 55,000) = P(Z < -0.985) \approx 1 - \Phi(0.99) \quad (6.4.13)$$

$$\Phi(0.99) \approx 0.8389 \Rightarrow P(\bar{X} < 55,000) \approx 1 - 0.839 \approx 0.161 \quad (6.4.14)$$

- Probability that the sample mean of  $n=100$  salaries is less than \$55,000.

$$n = 100, \quad SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{7,500}{10} = 750 \quad (6.4.15)$$

$$z = \frac{55,000 - 57,337}{750} = \frac{-2,337}{750} = -3.116 \quad (6.4.16)$$

$$P(\bar{X} < 55,000) = P(Z < -3.116) \approx \Phi(-3.12) \approx 0.0009 \quad (6.4.17)$$

As the sample size increase, the probability of seeing a sample mean of less than \$55,000 is getting smaller.

When you have a z-score that is less than -3 or greater than 3 we would call this a rare event or outlier. We will use this same process in inferential statistics in chapter 8.

---

This page titled [6.4: Sampling Distributions and the Central Limit Theorem](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Rachel Webb](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- **6.5: The Central Limit Theorem** by [Rachel Webb](#) is licensed [CC BY-SA 4.0](#). Original source: <https://mostlyharmlessstat.wixsite.com/webpage>.

 [Toggle block-level attributions](#)