

## 7.1: The Sample Mean and Sources of Error

Suppose  $X$  is a random variable with a distribution that may be known or unknown (it can be any distribution). Using a subscript that matches the random variable, suppose:

- $\mu_x$  = the mean of  $X$
- $\sigma_x$  = the standard deviation of  $X$

If you draw random samples of size  $n$ , then as  $n$  increases, the random variable  $\bar{X}$  which consists of sample means, tends to be normally distributed and

$$\bar{X} \sim N\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right). \quad (7.1.1)$$

The central limit theorem for sample means says that if you keep drawing larger and larger samples (such as rolling one, two, five, and finally, ten dice) and calculating their means, the sample means form their own *normal distribution* (the sampling distribution). The normal distribution has the same mean as the original distribution and a variance that equals the original variance divided by, the sample size. The variable  $n$  is the number of values that are averaged together, not the number of times the experiment is done.

To put it more formally, if you draw random samples of size  $n$ , the distribution of the random variable  $\bar{X}$ , which consists of sample means, is called the *sampling distribution of the mean*. The sampling distribution of the mean approaches a normal distribution as  $n$ , the sample size, increases.

The random variable  $\bar{X}$  has a different  $z$ -score associated with it from that of the random variable  $X$ . The mean  $\bar{x}$  is the value of  $\bar{X}$  in one sample.

$$z = \frac{\bar{x} - \mu_x}{\left(\frac{\sigma_x}{\sqrt{n}}\right)} \quad (7.1.2)$$

- $\mu_x$  is the average of both  $X$  and  $\bar{X}$ .
- $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$  = standard deviation of  $\bar{X}$  and is called the standard error of the mean.

### Example 7.1.1

An unknown distribution has a mean of 90 and a standard deviation of 15. Samples of size  $n = 25$  are drawn randomly from the population.

- Find the probability that the sample mean is between 85 and 92.
- Find the value that is two standard deviations above the expected value, 90, of the sample mean.

#### Answer

a.

Let  $X$  = one value from the original unknown population. The probability question asks you to find a probability for the **sample mean**.

Let  $\bar{X}$  = the mean of a sample of size 25. Since  $\mu_x = 90$ ,  $\sigma_x = 15$ , and  $n = 25$ ,

$$\bar{X} \sim N\left(90, \frac{15}{\sqrt{25}}\right).$$

Find  $P(85 < x < 92)$ . Draw a graph.

$$P(85 < x < 92) = 0.6997$$

The probability that the sample mean is between 85 and 92 is 0.6997.

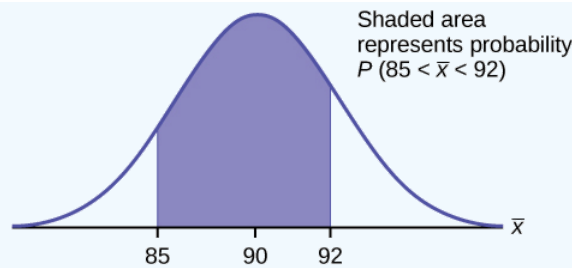


Figure 7.1.1.

`normalcdf` (lower value, upper value, mean, standard error of the mean)

The parameter list is abbreviated (lower value, upper value,  $\mu$ ,  $\frac{\sigma}{\sqrt{n}}$ )

$$\text{normalcdf} \left( 85, 92, 90, \frac{15}{\sqrt{25}} \right) = 0.6997$$

b.

To find the value that is two standard deviations above the expected value 90, use the formula:

$$\begin{aligned} \text{value} &= \mu_x + (\text{\#ofTSDEVs}) \left( \frac{\sigma_x}{\sqrt{n}} \right) \\ &= 90 + 2 \left( \frac{15}{\sqrt{25}} \right) = 96 \end{aligned}$$

The value that is two standard deviations above the expected value is 96.

The standard error of the mean is

$$\frac{\sigma_x}{\sqrt{n}} = \frac{15}{\sqrt{25}} = 3.$$

Recall that the standard error of the mean is a description of how far (on average) that the sample mean will be from the population mean in repeated simple random samples of size  $n$ .

### Example 7.1.2

The length of time, in hours, it takes an "over 40" group of people to play one soccer match is normally distributed with a **mean of two hours** and a **standard deviation of 0.5 hours**. A **sample of size  $n = 50$**  is drawn randomly from the population. Find the probability that the **sample mean** is between 1.8 hours and 2.3 hours.

**Answer**

Let  $X$  = the time, in hours, it takes to play one soccer match.

The probability question asks you to find a probability for the **sample mean time, in hours**, it takes to play one soccer match.

Let  $\bar{X}$  = the mean time, in hours, it takes to play one soccer match.

If  $\mu_x =$  \_\_\_\_\_,  $\sigma_x =$  \_\_\_\_\_, and  $n =$  \_\_\_\_\_, then  $X \sim N(\text{_____, ____})$  by the central limit theorem for means.

$$\mu_x = 2, \sigma_x = 0.5, n = 50, \text{ and } X \sim N \left( 2, \frac{0.5}{\sqrt{50}} \right)$$

Find  $P(1.8 < \bar{x} < 2.3)$ . Draw a graph.

$$P(1.8 < \bar{x} < 2.3) = 0.9977$$

$$\text{normalcdf} \left( 1.8, 2.3, 2, \frac{.5}{\sqrt{50}} \right) = 0.9977$$

The probability that the mean time is between 1.8 hours and 2.3 hours is 0.9977.

### Example 7.1.3

In a recent study reported Oct. 29, 2012 on the Flurry Blog, the mean age of tablet users is 34 years. Suppose the standard deviation is 15 years. Take a sample of size  $n = 100$ .

- What are the mean and standard deviation for the sample mean ages of tablet users?
- What does the distribution look like?
- Find the probability that the sample mean age is more than 30 years (the reported mean age of tablet users in this particular study).
- Find the 95<sup>th</sup> percentile for the sample mean age (to one decimal place).

#### Answer

- Since the sample mean tends to target the population mean, we have  $\mu_x = \mu = 34$ . The sample standard deviation is given by:

$$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = \frac{15}{10} = 1.5$$

- The central limit theorem states that for large sample sizes ( $n$ ), the sampling distribution will be approximately normal.
- The probability that the sample mean age is more than 30 is given by:

$$P(X > 30) = \text{normalcdf}(30, E99, 34, 1.5) = 0.9962$$

- Let  $k$  = the 95<sup>th</sup> percentile.

$$k = \text{invNorm}\left(0.95, 34, \frac{15}{\sqrt{100}}\right) = 36.5$$

### Exercise 7.1.3

In an article on Flurry Blog, a gaming marketing gap for men between the ages of 30 and 40 is identified. You are researching a startup game targeted at the 35-year-old demographic. Your idea is to develop a strategy game that can be played by men from their late 20s through their late 30s. Based on the article's data, industry research shows that the average strategy player is 28 years old with a standard deviation of 4.8 years. You take a sample of 100 randomly selected gamers. If your target market is 29- to 35-year-olds, should you continue with your development strategy?

#### Answer

You need to determine the probability for men whose mean age is between 29 and 35 years of age wanting to play a strategy game.

$$P(29 < \bar{x} < 35) = \text{normalcdf}\left(29, 35, 28, \frac{4.8}{\sqrt{100}}\right) = 0.0186 \quad (7.1.3)$$

You can conclude there is approximately a 1.9% chance that your game will be played by men whose mean age is between 29 and 35.

### Example 7.1.4

The mean number of minutes for app engagement by a tablet user is 8.2 minutes. Suppose the standard deviation is one minute. Take a sample of 60.

- What are the mean and standard deviation for the sample mean number of app engagement by a tablet user?
- What is the standard error of the mean?
- Find the 90<sup>th</sup> percentile for the sample mean time for app engagement for a tablet user. Interpret this value in a complete sentence.

d. Find the probability that the sample mean is between eight minutes and 8.5 minutes.

#### Answer

a.  $\mu = \mu = 8.2, \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{60}} = 0.13$

b. This allows us to calculate the probability of sample means of a particular distance from the mean, in repeated samples of size 60.

c. Let  $k$  = the 90<sup>th</sup> percentile

$k = \text{invNorm}\left(0.90, 8.2, \frac{1}{\sqrt{60}}\right) = 8.37$ . This value indicates that 90 percent of the average app engagement time for table users is less than 8.37 minutes.

d.  $P(8 < \bar{x} < 8.5) = \text{normalcdf}\left(8, 8.5, 8.2, \frac{1}{\sqrt{60}}\right) = 0.9293$

## WeBWork Problems

In a population whose distribution may be known or unknown, if the size ( $n$ ) of samples is sufficiently large, the distribution of the sample means will be approximately normal. The mean of the sample means will equal the population mean. The standard deviation of the distribution of the sample means, called the standard error of the mean, is equal to the population standard deviation divided by the square root of the sample size ( $n$ ).

## Formula Review

- The Central Limit Theorem for Sample Means:

$$\bar{X} \sim N\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right)$$

- The Mean  $\bar{X} : \sigma_x$
- Central Limit Theorem for Sample Means z-score and standard error of the mean:

$$z = \frac{\bar{x} - \mu_x}{\left(\frac{\sigma_x}{\sqrt{n}}\right)}$$

- Standard Error of the Mean (Standard Deviation ( $\bar{X}$ )):

$$\frac{\sigma_x}{\sqrt{n}}$$

## Glossary

### Average

a number that describes the central tendency of the data; there are a number of specialized averages, including the arithmetic mean, weighted mean, median, mode, and geometric mean.

### Central Limit Theorem

Given a random variable (RV) with known mean  $\mu$  and known standard deviation,  $\sigma$ , we are sampling with size  $n$ , and we are interested in two new RVs: the sample mean,  $\bar{X}$ , and the sample sum,  $\sum X$ . If the size ( $n$ ) of the sample is sufficiently large, then  $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$  and  $\sum X \sim N(n\mu, (\sqrt{n})(\sigma))$ . If the size ( $n$ ) of the sample is sufficiently large, then the distribution of the sample means and the distribution of the sample sums will approximate a normal distributions regardless of the shape of

the population. The mean of the sample means will equal the population mean, and the mean of the sample sums will equal  $n$  times the population mean. The standard deviation of the distribution of the sample means,  $\frac{\sigma}{\sqrt{n}}$ , is called the standard error of the mean.

### Normal Distribution

a continuous random variable (RV) with pdf  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ , where  $\mu$  is the mean of the distribution and  $\sigma$  is the standard deviation; notation:  $X \sim N(\mu, \sigma)$ . If  $\mu = 0$  and  $\sigma = 1$ , the RV is called a **standard normal distribution**.

### Standard Error of the Mean

the standard deviation of the distribution of the sample means, or  $\frac{\sigma}{\sqrt{n}}$ .

### References

1. Baran, Daya. "20 Percent of Americans Have Never Used Email." WebGuild, 2010. Available online at [www.webguild.org/20080519/20-...ver-used-email](http://www.webguild.org/20080519/20-...ver-used-email) (accessed May 17, 2013).
2. Data from The Flurry Blog, 2013. Available online at [blog.flurry.com](http://blog.flurry.com) (accessed May 17, 2013).
3. Data from the United States Department of Agriculture.

---

This page titled [7.1: The Sample Mean and Sources of Error](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- [7.2: The Central Limit Theorem for Sample Means \(Averages\)](#) by [OpenStax](#) is licensed [CC BY 4.0](#). Original source: <https://openstax.org/details/books/introductory-statistics>.