



# Introduction

By:  
OpenStaxCollege



*Have you ever wondered what the average number of M&Ms in a bag at the grocery store is?  
You can use confidence intervals to answer this question. (credit: comedy\_nose/flickr)*

## Chapter Objectives

By the end of this chapter, the student should be able to:

- Calculate and interpret confidence intervals for estimating a population mean and a population proportion.
- Interpret the Student's  $t$  probability distribution as the sample size changes.
- Discriminate between problems applying the normal and the Student's  $t$  distributions.
- Calculate the sample size required to estimate a population mean and a population proportion given a desired confidence level and margin of error.

Suppose you were trying to determine the mean rent of a two-bedroom apartment in your town. You might look in the classified section of the newspaper, write down several rents listed, and average them together. You would have obtained a point estimate of the true mean. If you are trying to determine the percentage of times you

make a basket when shooting a basketball, you might count the number of shots you make and divide that by the number of shots you attempted. In this case, you would have obtained a point estimate for the true proportion.

We use sample data to make generalizations about an unknown population. This part of statistics is called inferential statistics. **The sample data help us to make an estimate of a population parameter.** We realize that the point estimate is most likely not the exact value of the population parameter, but close to it. After calculating point estimates, we construct interval estimates, called confidence intervals.

In this chapter, you will learn to construct and interpret confidence intervals. You will also learn a new distribution, the Student's-t, and how it is used with these intervals. Throughout the chapter, it is important to keep in mind that the confidence interval is a random variable. It is the population parameter that is fixed.

If you worked in the marketing department of an entertainment company, you might be interested in the mean number of songs a consumer downloads a month from iTunes. If so, you could conduct a survey and calculate the sample mean,  $\bar{x}$ , and the sample standard deviation,  $s$ . You would use  $\bar{x}$  to estimate the population mean and  $s$  to estimate the population standard deviation. The sample mean,  $\bar{x}$ , is the point estimate for the population mean,  $\mu$ . The sample standard deviation,  $s$ , is the point estimate for the population standard deviation,  $\sigma$ .

Each of  $\bar{x}$  and  $s$  is called a statistic.

A confidence interval is another type of estimate but, instead of being just one number, it is an interval of numbers. The interval of numbers is a range of values calculated from a given set of sample data. The confidence interval is likely to include an unknown population parameter.

Suppose, for the iTunes example, we do not know the population mean  $\mu$ , but we do know that the population standard deviation is  $\sigma = 1$  and our sample size is 100. Then, by the central limit theorem, the standard deviation for the sample mean is

$$\frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{100}} = 0.1.$$

The empirical rule, which applies to bell-shaped distributions, says that in approximately 95% of the samples, the sample mean,  $\bar{x}$ , will be within two standard deviations of the population mean  $\mu$ . For our iTunes example, two standard deviations is  $(2)(0.1) = 0.2$ . The sample mean  $\bar{x}$  is likely to be within 0.2 units of  $\mu$ .

## Introduction

Because  $\bar{x}$  is within 0.2 units of  $\mu$ , which is unknown, then  $\mu$  is likely to be within 0.2 units of  $\bar{x}$  in 95% of the samples. The population mean  $\mu$  is contained in an interval whose lower number is calculated by taking the sample mean and subtracting two standard deviations  $(2)(0.1)$  and whose upper number is calculated by taking the sample mean and adding two standard deviations. In other words,  $\mu$  is between  $\bar{x} - 0.2$  and  $\bar{x} + 0.2$  in 95% of all the samples.

For the iTunes example, suppose that a sample produced a sample mean  $\bar{x} = 2$ . Then the unknown population mean  $\mu$  is between

$$\bar{x} - 0.2 = 2 - 0.2 = 1.8 \text{ and } \bar{x} + 0.2 = 2 + 0.2 = 2.2$$

We say that we are **95% confident** that the unknown population mean number of songs downloaded from iTunes per month is between 1.8 and 2.2. **The 95% confidence interval is (1.8, 2.2).**

The 95% confidence interval implies two possibilities. Either the interval (1.8, 2.2) contains the true mean  $\mu$  or our sample produced an  $\bar{x}$  that is not within 0.2 units of the true mean  $\mu$ . The second possibility happens for only 5% of all the samples (95–100%).

Remember that a confidence interval is created for an unknown population parameter like the population mean,  $\mu$ . Confidence intervals for some parameters have the form:

**(point estimate – margin of error, point estimate + margin of error)**

The margin of error depends on the confidence level or percentage of confidence and the standard error of the mean.

When you read newspapers and journals, some reports will use the phrase "margin of error." Other reports will not use that phrase, but include a confidence interval as the point estimate plus or minus the margin of error. These are two ways of expressing the same concept.

## Note

Although the text only covers symmetrical confidence intervals, there are non-symmetrical confidence intervals (for example, a confidence interval for the standard deviation).

## Collaborative Exercise

## Introduction

Have your instructor record the number of meals each student in your class eats out in a week. Assume that the standard deviation is known to be three meals. Construct an approximate 95% confidence interval for the true mean number of meals students eat out each week.

1. Calculate the sample mean.
2. Let  $\sigma = 3$  and  $n =$  the number of students surveyed.
3. Construct the interval  $\left( \left( \bar{x} - 2 \right) \left( \frac{\sigma}{\sqrt{n}} \right), \left( \bar{x} + 2 \right) \left( \frac{\sigma}{\sqrt{n}} \right) \right)$ .

We say we are approximately 95% confident that the true mean number of meals that students eat out in a week is between \_\_\_\_\_ and \_\_\_\_\_.