Chapter 8: Estimating with Confidence

Section 8.1

Confidence Intervals: The Basics

The Practice of Statistics, 4th edition – For AP*

STARNES, YATES, MOORE

Chapter 8
Estimating with Confidence

- **8.1 Confidence Intervals: The Basics**
- 8.2 Estimating a Population Proportion
- 8.3 Estimating a Population Mean

Section 8.1
Confidence Intervals: The Basics

Learning Objectives

After this section, you should be able to…

- ✓ INTERPRET a confidence level
- ✓ INTERPRET a confidence interval in context
- ✓ DESCRIBE how a confidence interval gives a range of plausible values for the parameter
- ✓ DESCRIBE the inference conditions necessary to construct confidence intervals
- ✓ EXPLAIN practical issues that can affect the interpretation of a confidence interval

Introduction

Our goal in many statistical settings is to use a sample statistic to estimate a population parameter. In Chapter 4, we learned **if we randomly select the sample, we should be able to generalize our results to the population of interest**.

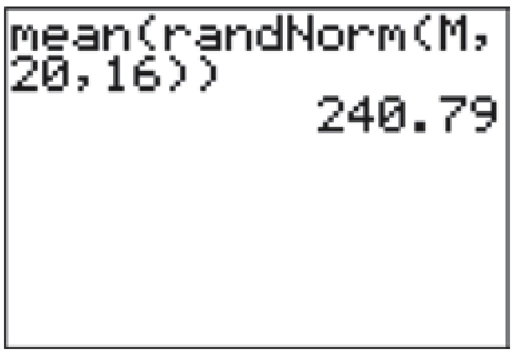
In Chapter 7, we learned that **different samples yield different results for our estimate**. Statistical inference uses the language of probability to express the strength of our conclusions by taking chance variation due to random selection or random assignment into account.

In this chapter, we'll learn one method of statistical inference – *confidence intervals* – so we may estimate the value of a parameter from a sample statistic. As we do so, we'll learn not only how to construct a confidence interval, but also how to report probabilities that would describe *what would happen if we used the inference method many times*.

Activity: The Mystery Mean

Your teacher has selected a "Mystery Mean" value μ and stored it as "M" in their calculator. Your task is to work together with 3 or 4 students to estimate this value.

The following command was executed on their calculator:
`mean(randNorm(M,20,16))`



The result was 240.79. This tells us the calculator chose an SRS of 16 observations from a Normal population with mean M and standard deviation 20. The resulting sample mean of those 16 values was 240.79.

Your group must determine an interval of *reasonable* values for the population mean μ. Use the result above and what you learned about sampling distributions in the previous chapter.

Share your team's results with the class.

Confidence Intervals: The Basics

If you had to give one number to estimate an unknown population parameter, what would it be? If you were estimating a population mean μ you would probably use $\bar{x}$. If you were estimating a population proportion p, you might use $\hat{p}$. In both cases, you would be providing a **point estimate** of the parameter of interest.

Definition:

A **point estimator** is a statistic that provides an estimate of a population parameter. The value of that statistic from a sample is called a **point estimate**. Ideally, a point estimate is our “best guess” at the value of an unknown parameter.

We learned in Chapter 7 that an ideal point estimator will have no bias and low variability. Since variability is almost always present when calculating statistics from different samples, we must extend our thinking about estimating parameters to include an acknowledgement that repeated sampling could yield different results.

Alternate Example – From golf balls to graphing calculators

Problem: In each of the following settings, determine the point estimator you would use and calculate the value of the point estimate.

(a) The makers of a new golf ball want to estimate the median distance the new balls will travel when hit by mechanical driver. They select a random sample of 10 balls and measure the distance each ball travels after being hit by the mechanical driver. Here are the distances (in yards): 285 286 284 285 282 284 287 290 288 285

Use the sample median as a point estimator for the true median. The sample median is 285 yards.

(**b) The golf ball manufacturer would also like to investigate the variability of the distance travelled by the golf balls by estimating the interquartile range.**

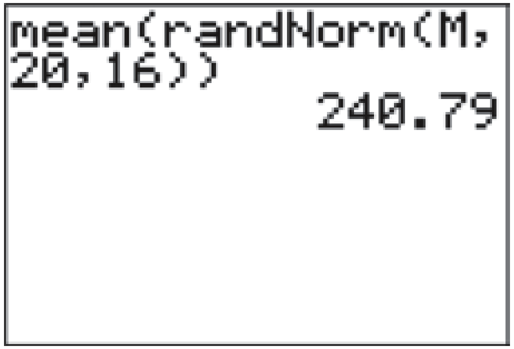
Use the sample *IQR* as a point estimator for the true *IQR*. The sample *IQR* is 287 – 284 = 3 yards.

(c) The math department wants to know what proportion of its students own a graphing calculator, so they take a random sample of 100 students and find that 28 own a graphing calculator.

Use the sample proportion $\hat{p}$ as a point estimator for the true proportion p. The sample proportion is $\hat{p} = 0.28$

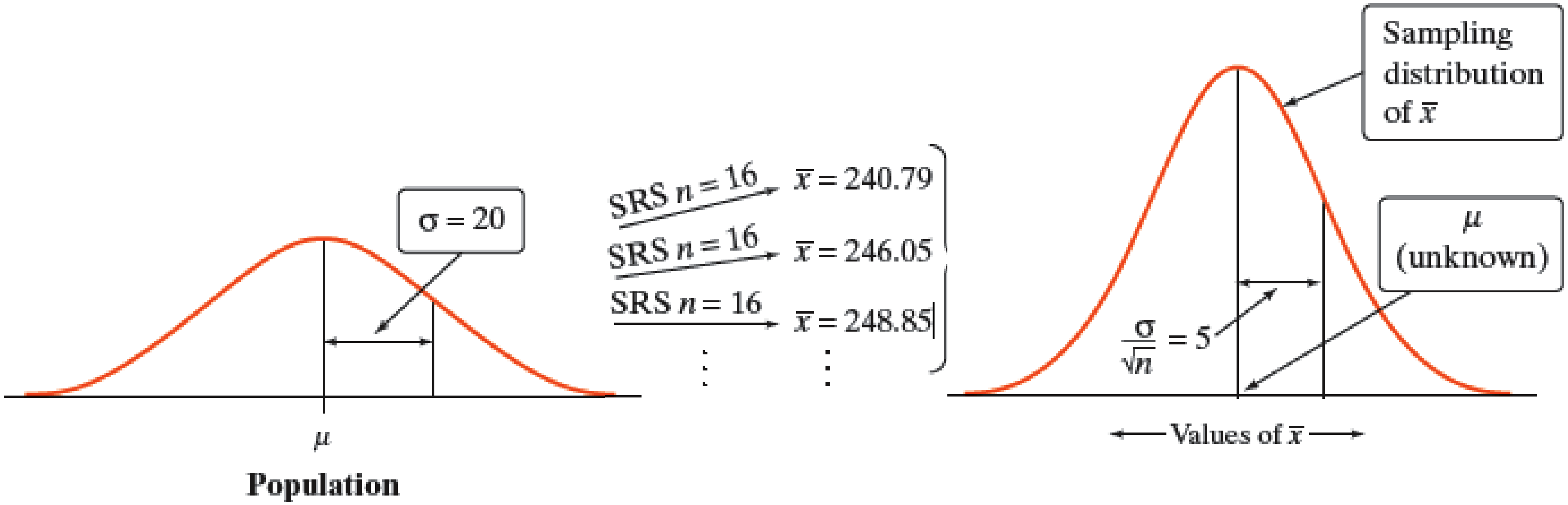
The Idea of a Confidence Interval

Recall the “Mystery Mean” Activity. Is the value of the population mean μ exactly 240.79? Probably not. However, since the sample mean is 240.79, we could guess that μ is “somewhere” around 240.79. **How close to 240.79 is μ likely to be?**

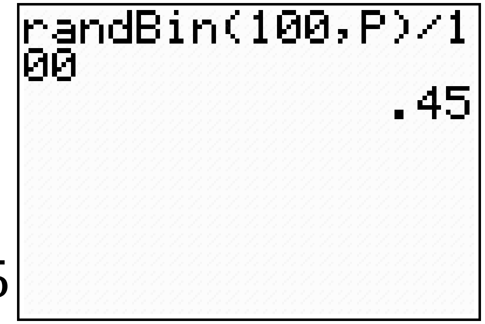


To answer this question, we must ask another:

How would the sample mean $\bar{x}$ vary if we took many SRSs of size 16 from the population?



Alternate Example – The mystery proportion



In "The Mystery Proportion" Alternate Activity, students should have used the following logic to come up with an "interval estimate" for the unknown population proportion *p*:

1. To estimate *p*, we use the sample proportion $\hat{p} = 0.45$

2. In repeated samples, the values of follow an approximately Normal distribution (since our sample size is relatively large) with mean *p* and standard deviation . $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{100}}$. Since we don't actually know the value of *p*, we can use $\hat{p}$ to help estimate the standard deviation of the sampling distribution of the sample proportion: $\sigma_{\hat{p}} = \sqrt{\frac{0.45(1-0.45)}{100}} = 0.05$

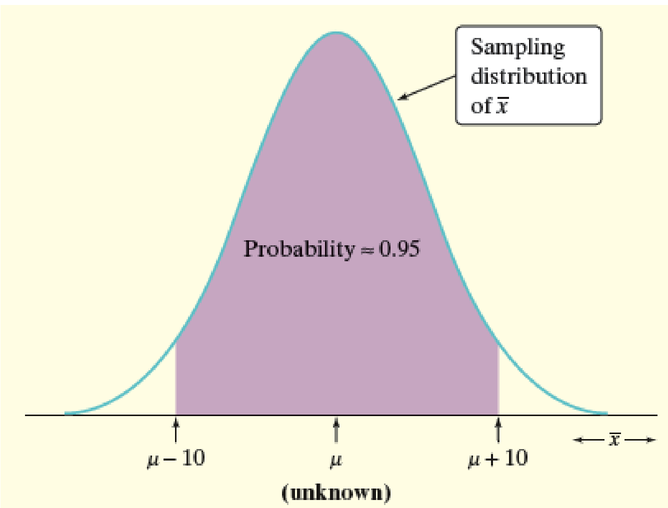
3. In about 95% of samples of size 100, $\hat{p}$ will be within 0.10 (2 standard deviations) of *p*.

4. Thus, in about 95% of all possible samples of size 100, *p* will be within 0.10 of $\hat{p}$. Thus, we estimate that *p* lies somewhere between 0.45 – 0.10 = 0.35 and 0.45 + 0.10 = 0.55.

5. The interval from 0.35 to 0.55 is an approximate 95% confidence interval for *p*. In other words, it wouldn't be surprising if we found out that the true value of *p* was any value from 0.35 to 0.55.

The Idea of a Confidence Interval

To estimate the Mystery Mean μ, we can use $\bar{x} = 240.79$ as a point estimate. We donÕt expect μ to be exactly equal to $\bar{x}$ so we need to say how accurate we think our estimate is.



- In repeated samples, the values of $\bar{x}$ follow a Normal distribution with mean and standard deviation 5.
- The 68 - 95 - 99.7 Rule tells us that in 95% of all samples of size 16, $\bar{x}$ will be within 10 (two standard deviations) of μ.
- If $\bar{x}$ is within 10 points of μ, then μ is within 10 points of $\bar{x}$.

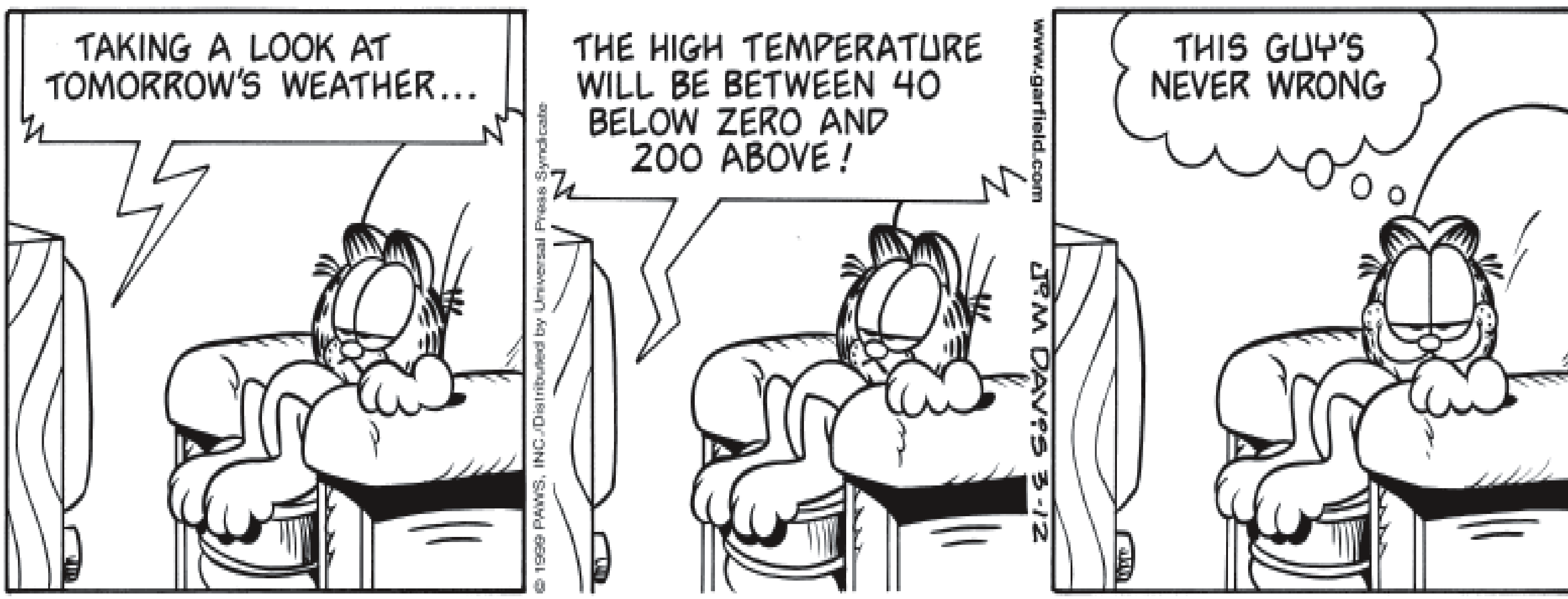
Therefore, the interval from $\bar{x} - 10$ to $\bar{x} + 10$ will "capture" μ in about 95% of all samples of size 16.

If we estimate that μ lies somewhere in the interval **230.79** to **250.79**, we'd be calculating an interval using a method that captures the true μ in about 95% of all possible samples of this size.

The Idea of a Confidence Interval

The big idea : The sampling distribution of $\bar{x}$ tells us how close to μ the sample mean $\bar{x}$ is likely to be. All confidence intervals we construct will have a form similar to this:

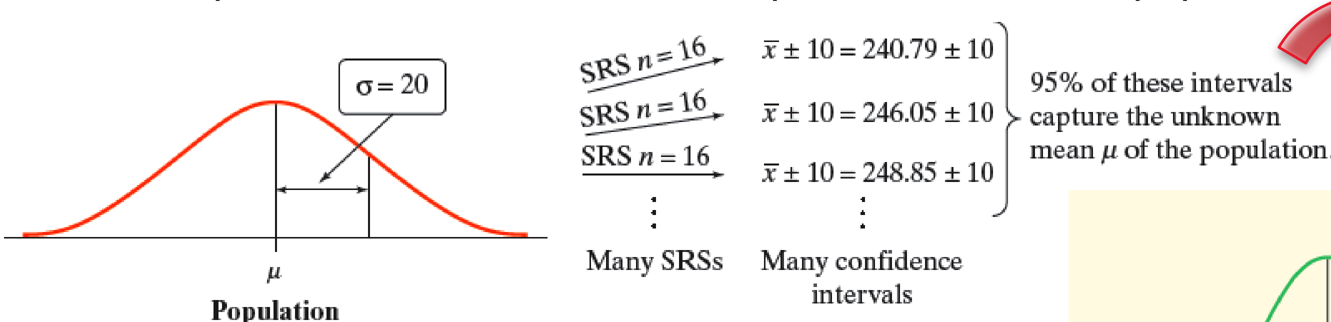
$$\text{estimate} \pm \text{margin of error}$$

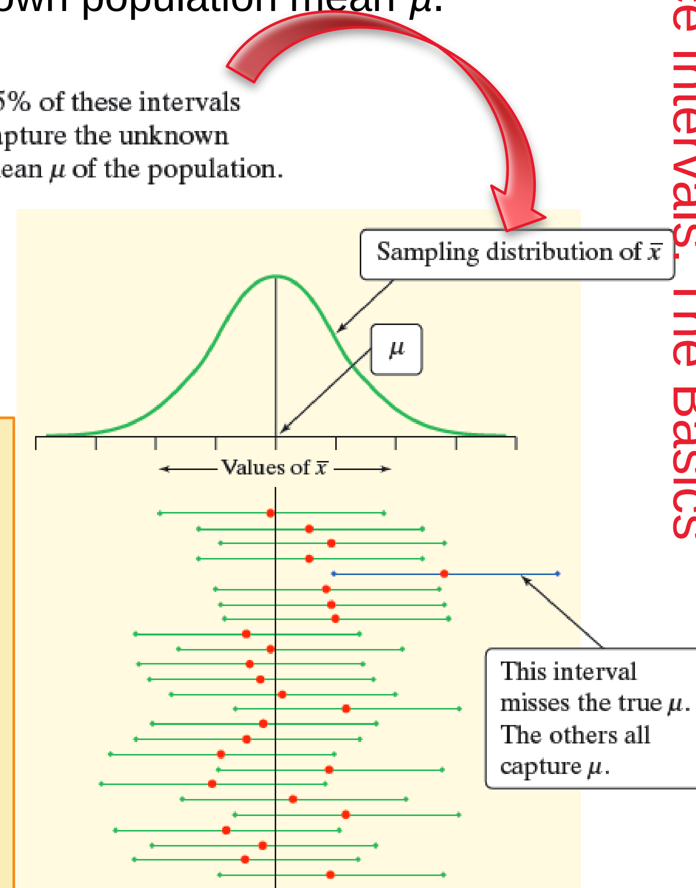


We usually choose a confidence level of 90% or higher because we want to be quite sure of our conclusions. The most common confidence level is 95%.

Interpreting Confidence Levels and Confidence Intervals

The confidence level is the overall capture rate if the method is used many times. Starting with the population, imagine taking many SRSs of 16 observations. The sample mean will vary from sample to sample, but when we use the method *estimate ± margin of error* to get an interval based on each sample, 95% of these intervals capture the unknown population mean μ.





Interpreting Confidence Level and Confidence Intervals

Confidence level*:* To say that we are 95% *confident* is shorthand for "95% *of* all possible samples of a given size from this population will result in an interval that captures the unknown parameter."

Confidence interval*:* To interpret a *C*% confidence interval for an unknown parameter, say, "We are *C%* confident that the interval from _____ to _____ captures the actual value of the [population parameter in context]."

- **Interpreting Confidence Levels and Confidence Intervals**

The confidence level tells us how likely it is that the method we are using will produce an interval that captures the population parameter *if we use it many times*.

The confidence level does not tell us the chance that a particular confidence interval captures the population parameter.

Instead, the confidence interval gives us a set of plausible values for the parameter.

We interpret confidence levels and confidence intervals in much the same way whether we are estimating a population mean, proportion, or some other parameter.

Alternate Example – The mystery proportion

- The 95% confidence *level* in the mystery proportion example tells us that in about 95% of samples of size 100 from the mystery population, the interval $\hat{p} \pm 0.10$ will contain the population proportion *p*. The actual value of $\hat{p}$ in the example was $\hat{p} = 0.45$, so when we interpret the confidence *interval* we say that "We are 95% confident that the interval from 0.35 to 0.55 captures the mystery proportion."

Alternate Example – Presidential Approval Ratings

According to www.gallup.com, on August 13, 2010, the 95% confidence interval for the true proportion of Americans who approved of the job Barack Obama was doing as president was 0.44 ± 0.03.

Problem: Interpret the confidence interval and the confidence level.

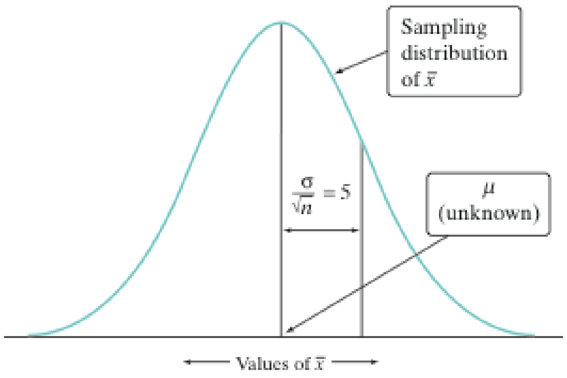
Solution: *Interval*: We are 95% confident that the interval from 0.41 to 0.47 captures the true proportion of Americans who approve of the job Barack Obama was doing as president at the time of the poll. *Level*: In 95% of all possible samples of the same size, the resulting confidence interval would capture the true proportion of Americans who approve of the job Barack Obama was doing as president.

Constructing a Confidence Interval

Why settle for 95% confidence when estimating a parameter? The price we pay for greater confidence is a wider interval.

When we calculated a 95% confidence interval for the mystery mean μ, we started with

estimate ± margin of error



Our estimate came from the sample statistic $\bar{x}$. Since the sampling distribution of $\bar{x}$ is Normal, about 95% of the values of $\bar{x}$ will lie within 2 standard deviations ($2\sigma_{\bar{x}}$) of the mystery mean μ. That is, our interval could be written as:

$$\mathbf{240.79 \pm 2 \cdot 5 = \bar{x} \pm 2\sigma_{\bar{x}}}$$

This leads to a more general formula for confidence intervals:

statistic ± (critical value) • (standard deviation of statistic)

Calculating a Confidence Interval

Calculating a Confidence Interval

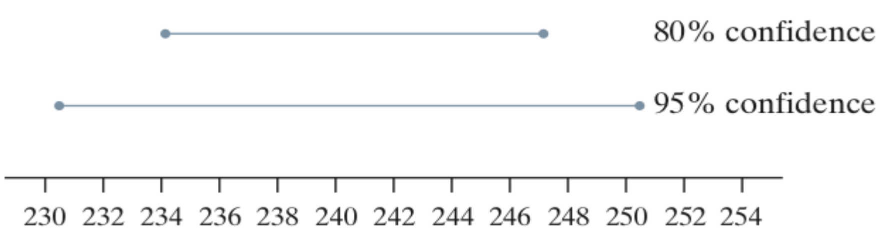
The confidence interval for estimating a population parameter has the form

statistic ± (critical value) • (standard deviation of statistic)

where the statistic we use is the point estimator for the parameter.

Properties of Confidence Intervals:

- The "margin of error" is the (critical value) • (standard deviation of statistic)
- The user chooses the confidence level, and the margin of error follows from this choice.
- The critical value depends on the confidence level and the sampling distribution of the statistic.
  - Greater confidence requires a larger critical value
  - The standard deviation of the statistic depends on the sample size *n*



The margin of error gets smaller when:

✓ The confidence level decreases
✓ The sample size *n* increases

Using Confidence Intervals

Before calculating a confidence interval for μ or p there are three important **conditions** that you should check.

1) **Random**: The data should come from a well-designed random sample or randomized experiment.

2) **Normal:** The sampling distribution of the statistic is approximately Normal.

For means: The sampling distribution is exactly Normal if the population distribution is Normal. When the population distribution is not Normal, then the central limit theorem tells us the sampling distribution will be approximately Normal if n is sufficiently large ($n \geq 30$).

For proportions: We can use the Normal approximation to the sampling distribution as long as $np \geq 10$ and $n(1 - p) \geq 10$.

3) **Independent**: Individual observations are independent. When sampling without replacement, the sample size n should be no more than 10% of the population size N (the *10% condition*) to use our formula for the standard deviation of the statistic.

Section 8.1
Confidence Intervals: The Basics

Summary

In this section, we learned that…

- ✓ To estimate an unknown population parameter, start with a statistic that provides a reasonable guess. The chosen statistic is a **point estimator** for the parameter. The specific value of the point estimator that we use gives a **point estimate** for the parameter.
- ✓ A **confidence interval** uses sample data to estimate an unknown population parameter with an indication of how precise the estimate is and of how confident we are that the result is correct.
- ✓ Any confidence interval has two parts: an interval computed from the data and a confidence level C. The interval has the form

  estimate ± margin of error

- ✓ When calculating a confidence interval, it is common to use the form

  statistic ± (critical value) · (standard deviation of statistic)

+

Section 8.1
Confidence Intervals: The Basics

Summary

In this section, we learned that…

- ✓ The **confidence level C** is the success rate of the method that produces the interval. If you use 95% confidence intervals often, in the long run 95% of your intervals will contain the true parameter value. You don't know whether a 95% confidence interval calculated from a particular set of data actually captures the true parameter value.

- ✓ Other things being equal, the **margin of error** of a confidence interval gets smaller as the confidence level C decreases and/or the sample size *n* increases.

- ✓ Before you calculate a confidence interval for a population mean or proportion, be sure to check conditions: **Random** sampling or random assignment, **Normal** sampling distribution, and **Independent** observations.

- ✓ The margin of error for a confidence interval includes only chance variation, not other sources of error like nonresponse and undercoverage.

Looking Ahead…

In the next Section…

We'll learn how to estimate a population proportion.

We'll learn about

- ✓ **Conditions for estimating *p***
- ✓ **Constructing a confidence interval for *p***
- ✓ **The four-step process for estimating *p***
- ✓ **Choosing the sample size for estimating *p***