CHAPTER 8
Estimating with Confidence

8.1
Confidence Intervals: The Basics

The Practice of Statistics, 5th Edition
Starnes, Tabor, Yates, Moore
Confidence Intervals: The Basics

**Learning Objectives**

After this section, you should be able to:

- DETERMINE the point estimate and margin of error from a confidence interval.
- INTERPRET a confidence interval in context.
- INTERPRET a confidence level in context.
- DESCRIBE how the sample size and confidence level affect the length of a confidence interval.
- EXPLAIN how practical issues like nonresponse, undercoverage, and response bias can affect the interpretation of a confidence interval.
Activity: The Mystery Mean

Suppose your teacher has selected a “Mystery Mean” value $\mu$ 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:

```plaintext
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 $\mu$. 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 $\mu$ 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.

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.

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.
The Idea of a Confidence Interval

Recall the “Mystery Mean” Activity. Is the value of the population mean $\mu$ exactly 240.79? Probably not. However, since the sample mean is 240.79, we could guess that $\mu$ is “somewhere” around 240.79. How close to 240.79 is $\mu$ 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?**
The Idea of a Confidence Interval

To estimate the Mystery Mean, we can use $\bar{x} = 240.79$ as a point estimate. We don’t expect $\mu$ 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 $\mu$.

If $\bar{x}$ is within 10 points of $\mu$, then $\mu$ is within 10 points of $\bar{x}$.

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

If we estimate that $\mu$ lies somewhere in the interval 230.79 to 250.79, we’d be calculating an interval using a method that captures the true $\mu$ 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 the sample mean $\bar{x}$ is likely to be. All confidence intervals we construct will have a form similar to this:

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

A C% confidence interval gives an interval of plausible values for a parameter. The interval is calculated from the data and has the form

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

The difference between the point estimate and the true parameter value will be less than the margin of error in C% of all samples.

The confidence level C gives the overall success rate of the method for calculating the confidence interval. That is, in C% of all possible samples, the method would yield an interval that captures the true parameter value.
Interpreting Confidence Levels and Intervals

The confidence level is the overall capture rate if the method is used many times. The sample mean will vary from sample to sample, but when we use the method \textit{estimate} \( \pm \) \textit{margin of error} to get an interval based on each sample, C\% of these intervals capture the unknown population mean \( \mu \).

\[
\begin{align*}
\text{SRS } n = 16 & \quad \bar{x} \pm 10 = 240.79 \pm 10 \\
\text{SRS } n = 16 & \quad \bar{x} \pm 10 = 246.05 \pm 10 \\
\vdots & \quad \vdots \\
\text{Many SRSs} & \quad \text{Many confidence intervals}
\end{align*}
\]

95\% of these intervals capture the unknown mean \( \mu \) of the population. This interval misses the true \( \mu \). The others all capture \( \mu \).
Interpreting Confidence Levels and Intervals

Interpreting Confidence Intervals

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

To say that we are 95% confident is shorthand for “If we take many samples of the same size from this population, about 95% of them will result in an interval that captures the actual parameter value.”
Interpreting Confidence Levels and 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.
Constructing Confidence Intervals

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 \( \mu \), we started with

\[ \text{estimate} \pm \text{margin of error} \]

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 (\( 2 \times \sigma / \sqrt{n} \)) of the mystery mean \( \mu \). That is, our interval could be written as:

\[ 240.79 \pm 2 \times 5 = \bar{x} \pm 2 \frac{\sigma}{\sqrt{n}} \]

This leads to a more general formula for confidence intervals:

\[ \text{statistic} \pm (\text{critical value}) \times (\text{standard deviation of statistic}) \]
Constructing Confidence Intervals

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 $n$
Using Confidence Intervals Wisely

Here are two important cautions to keep in mind when constructing and interpreting confidence intervals.

- Our method of calculation assumes that the data come from an SRS of size $n$ from the population of interest.

- The margin of error in a confidence interval covers only chance variation due to random sampling or random assignment.
Confidence Intervals: The Basics

Section Summary

In this section, we learned how to…

- **DETERMINE** the point estimate and margin of error from a confidence interval.
- **INTERPRET** a confidence interval in context.
- **INTERPRET** a confidence level in context.
- **DESCRIBE** how the sample size and confidence level affect the length of a confidence interval.
- **EXPLAIN** how practical issues like nonresponse, undercoverage, and response bias can affect the interpretation of a confidence interval.