+

# Unit 5: Estimating with Confidence

## Section 8.1

## Confidence Intervals: The Basics

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

**STARNES, YATES, MOORE**

# Unit 5
# 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

## Introduction

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

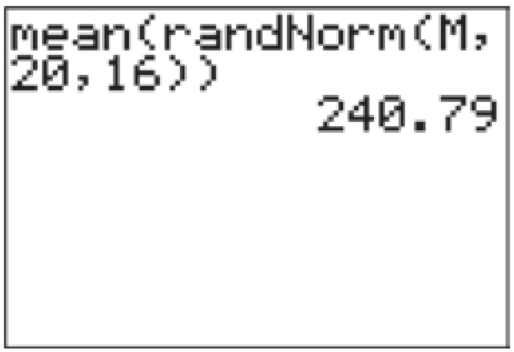
We also learned that **different samples yield different results for our estimate**. Statistical inference uses the language of probability to express the strength of our conclusions by taking chance variation due to random selection or random assignment into account.

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*.

## The Mystery Mean

A “Mystery Mean” value $\mu$ was stored as “M” in a calculator. The task is to estimate this value.

The following command was executed on a 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.

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.

# 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.

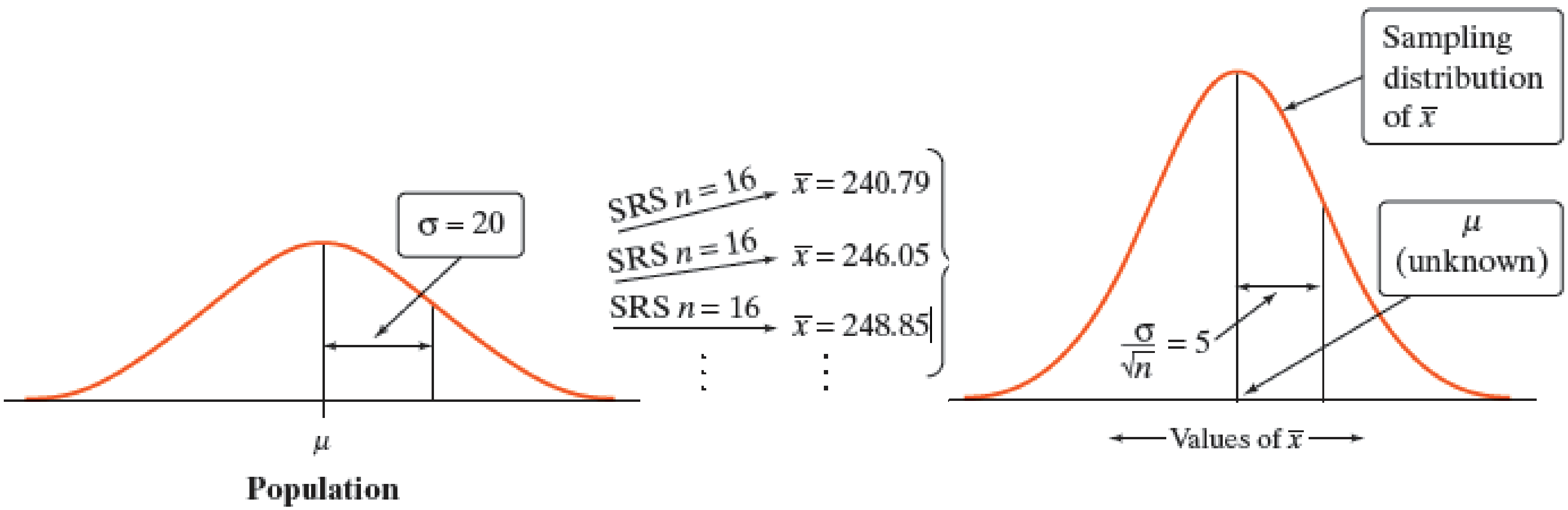
We previously learned that an ideal point estimator will have no bias and low variability. Since variability is almost always present when calculating statistics from different samples, we must extend our thinking about estimating parameters to include an acknowledgement that repeated sampling could 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 μ likely to be?**

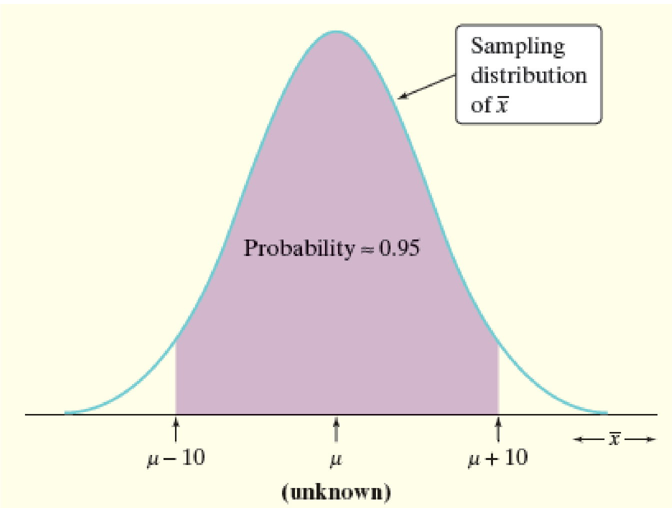

To answer this question, we must ask another:

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



## The Idea of a Confidence Interval

To estimate the Mystery Mean $\mu$, 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" $\mu$ 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 $\mu$ the sample mean $\bar{x}$ is likely to be. All confidence intervals we construct will have a form similar to this:

**estimate $\pm$ margin of error**

**Definition:**

A **confidence interval** for a parameter has two parts:

- An interval calculated from the data, which has the form:

  estimate $\pm$ margin of error

- The **margin of error** tells how close the estimate tends to be to the unknown parameter in repeated random sampling.
- A **confidence level *C***, 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.

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%.

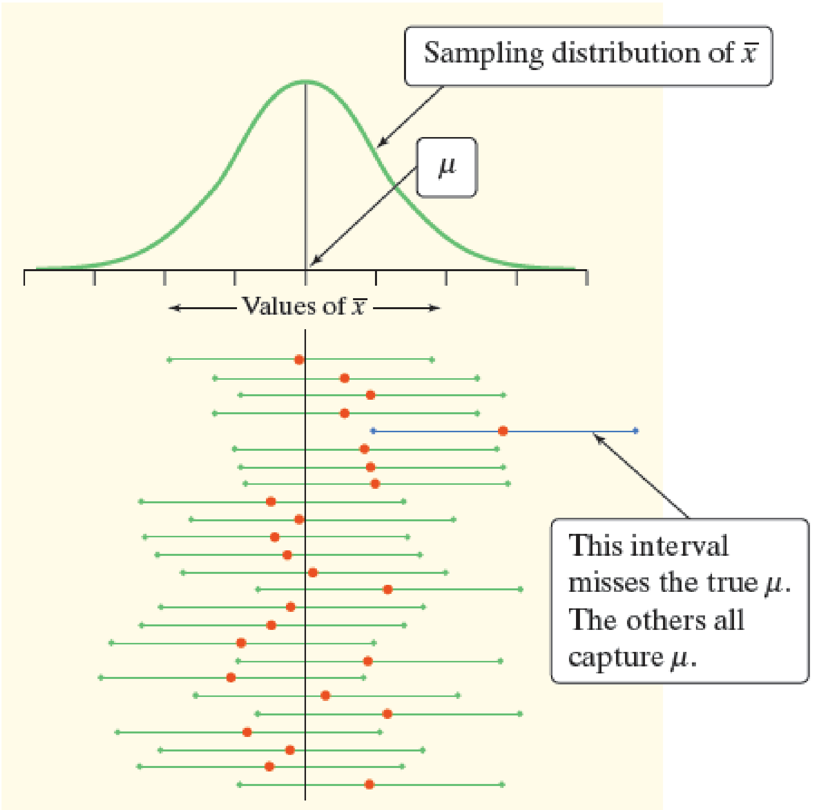
## Interpreting Confidence Levels and Confidence Intervals

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 $\mu$.

**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.

# 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 user chooses the confidence level, and the margin of error follows from this choice.
- The margin of error gets smaller when:
  - The confidence level decreases
  - The sample size n increases
- 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

Before calculating a confidence interval for $\mu$ 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.

# 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***

**Homework**

Textbook – Chapter 10 #'s, 1-3.

Additional notes packet – Read and answer all questions, pgs. 1-11