CHAPTER 8
Estimating with Confidence

8.3
Estimating a Population Mean

The Practice of Statistics, 5th Edition
Starnes, Tabor, Yates, Moore
Estimating a Population Mean

Learning Objectives

After this section, you should be able to:

✓ STATE and CHECK the Random, 10%, and Normal/Large Sample conditions for constructing a confidence interval for a population mean.

✓ EXPLAIN how the t distributions are different from the standard Normal distribution and why it is necessary to use a t distribution when calculating a confidence interval for a population mean.

✓ DETERMINE critical values for calculating a C% confidence interval for a population mean using a table or technology.

✓ CONSTRUCT and INTERPRET a confidence interval for a population mean.

✓ DETERMINE the sample size required to obtain a C% confidence interval for a population mean with a specified margin of error.
When σ Is Unknown: The $t$ Distributions

When the sampling distribution of $\bar{x}$ is close to Normal, we can find probabilities involving $\bar{x}$ by standardizing:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

When we don’t know $\sigma$, we can estimate it using the sample standard deviation $s_x$. What happens when we standardize?

$$?? = \frac{\bar{x}}{s_x / \sqrt{n}}$$
When $\sigma$ Is Unknown: The $t$ Distributions

When we standardize based on the sample standard deviation $s_x$, our statistic has a new distribution called a $t$ distribution.

- It has a different shape than the standard Normal curve:
  - It is symmetric with a single peak at 0,
  - However, it has much more area in the tails.

Like any standardized statistic, $t$ tells us how far $\bar{x}$ is from its mean in standard deviation units.

There is a different $t$ distribution for each sample size, specified by its degrees of freedom (df).
The *t* Distributions; Degrees of Freedom

When we perform inference about a population mean \( \mu \) using a *t* distribution, the appropriate degrees of freedom are found by subtracting 1 from the sample size \( n \), making \( df = n - 1 \).

We will write the *t* distribution with \( n - 1 \) degrees of freedom as \( t_{n-1} \).

### Conditions for Constructing a Confidence Interval About a Proportion

Draw an SRS of size \( n \) from a large population that has a Normal distribution with mean \( \mu \) and standard deviation \( \sigma \). The statistic

\[
t = \frac{\bar{X}}{S_X / \sqrt{n}}
\]

has the *t distribution* with **degrees of freedom** \( df = n - 1 \). When the population distribution isn’t Normal, this statistic will have approximately a \( t_{n - 1} \) distribution if the sample size is large enough.
The *t* Distributions; Degrees of Freedom

When comparing the density curves of the standard Normal distribution and *t* distributions, several facts are apparent:

- The density curves of the *t* distributions are similar in shape to the standard Normal curve.
- The spread of the *t* distributions is a bit greater than that of the standard Normal distribution.
- The *t* distributions have more probability in the tails and less in the center than does the standard Normal.
- As the degrees of freedom increase, the *t* density curve approaches the standard Normal curve ever more closely.

![Graph showing the comparison between standard Normal and *t* distributions with different degrees of freedom.](image)
Example: Using Table B to Find Critical $t^*$ Values

**Problem:** What critical value $t^*$ from Table B should be used in constructing a confidence interval for the population mean in each of the following settings?

(a) A 95% confidence interval based on an SRS of size $n = 12$.

**Solution:** In Table B, we consult the row corresponding to $\text{df} = 12 - 1 = 11$.

We move across that row to the entry that is directly above 95% confidence level on the bottom of the chart.

The desired critical value is $t^* = 2.201$. 

![Table B](image)
Example: Using Table B to Find Critical $t^*$ Values

**Problem:** What critical value $t^*$ from Table B should be used in constructing a confidence interval for the population mean in each of the following settings?

(b) A 90% confidence interval from a random sample of 48 observations.

<table>
<thead>
<tr>
<th>df</th>
<th>.10</th>
<th>.05</th>
<th>.025</th>
<th>.02</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>1.310</td>
<td>1.697</td>
<td>2.042</td>
<td>2.147</td>
</tr>
<tr>
<td>40</td>
<td>1.303</td>
<td>1.684</td>
<td>2.021</td>
<td>2.123</td>
</tr>
<tr>
<td>50</td>
<td>1.299</td>
<td>1.676</td>
<td>2.009</td>
<td>2.109</td>
</tr>
<tr>
<td>$z^*$</td>
<td>1.282</td>
<td>1.645</td>
<td>1.960</td>
<td>2.054</td>
</tr>
</tbody>
</table>

**Solution:** With 48 observations, we want to find the $t^*$ critical value for $df = 48 - 1 = 47$ and 90% confidence.

There is no $df = 47$ row in Table B, so we use the more conservative $df = 40$.

The corresponding critical value is $t^* = 1.684$. 
Conditions for Estimating $\mu$

As with proportions, you should check some important conditions before constructing a confidence interval for a population mean.

Conditions For Constructing A Confidence Interval About A Mean

- **Random**: The data come from a well-designed random sample or randomized experiment.
  
  - **10%**: When sampling without replacement, check that $n \leq \frac{1}{10}N$

- **Normal/Large Sample**: The population has a Normal distribution or the sample size is large ($n \geq 30$). If the population distribution has unknown shape and $n < 30$, use a graph of the sample data to assess the Normality of the population. Do not use $t$ procedures if the graph shows strong skewness or outliers.
Constructing a Confidence Interval for $\mu$

When the conditions for inference are satisfied, the sampling distribution for $\bar{x}$ has roughly a Normal distribution. Because we don’t know $\sigma$, we estimate it by the sample standard deviation $s_x$.

The **standard error of the sample mean** $\bar{x}$ is $\frac{s_x}{\sqrt{n}}$, where $s_x$ is the sample standard deviation. It describes how far $\bar{x}$ will be from $\mu$, on average, in repeated SRSs of size $n$.

To construct a confidence interval for $\mu$,

1. Replace the standard deviation of $\bar{x}$ by its standard error in the formula for the one-sample $z$ interval for a population mean.

2. Use critical values from the $t$ distribution with $n - 1$ degrees of freedom in place of the $z$ critical values. That is,

$$\text{statistic} \pm \text{(critical value)} \times \text{(standard deviation of statistic)}$$

$$= \bar{x} \pm t \times \frac{s_x}{\sqrt{n}}$$
One-Sample $t$ Interval for a Population Mean

The one-sample $t$ interval for a population mean is similar in both reasoning and computational detail to the one-sample $z$ interval for a population proportion.

One-Sample $t$ Interval for a Population Mean

When the conditions are met, a $C\%$ confidence interval for the unknown mean $\mu$ is

$$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$$

where $t^*$ is the critical value for the $t_{n-1}$ distribution with $C\%$ of its area between $-t^*$ and $t^*$. 
Example: A one-sample $t$ interval for $\mu$

Environmentalists, government officials, and vehicle manufacturers are all interested in studying the auto exhaust emissions produced by motor vehicles.

The major pollutants in auto exhaust from gasoline engines are hydrocarbons, carbon monoxide, and nitrogen oxides (NOX). Researchers collected data on the NOX levels (in grams/mile) for a random sample of 40 light-duty engines of the same type.

The mean NOX reading was 1.2675 and the standard deviation was 0.3332.

**Problem:** (a) Construct and interpret a 95% confidence interval for the mean amount of NOX emitted by light-duty engines of this type.
Example: Constructing a confidence interval for $\mu$

**State**: We want to estimate the true mean amount $\mu$ of NOX emitted by all light-duty engines of this type at a 95% confidence level.

**Plan**: If the conditions are met, we should use a one-sample $t$ interval to estimate $\mu$.

- Random: The data come from a “random sample” of 40 engines from the population of all light-duty engines of this type.
  - 10%??: We are sampling without replacement, so we need to assume that there are at least $10(40) = 400$ light-duty engines of this type.

- Large Sample: We don’t know if the population distribution of NOX emissions is Normal. Because the sample size is large, $n = 40 > 30$, we should be safe using a $t$ distribution.
Example: Constructing a confidence interval for $\mu$

**Do:** From the information given, $\bar{x} = 1.2675\text{ g/mi}$ and $s_x = 0.3332\text{ g/mi}$.

To find the critical value $t^*$, we use the t distribution with $df = 40 - 1 = 39$.

Unfortunately, there is no row corresponding to 39 degrees of freedom in Table B. We can’t pretend we have a larger sample size than we actually do, so we use the more conservative $df = 30$. 

![Upper-tail probability $p$](image)
Example: Constructing a confidence interval for $\mu$

\[
\bar{x} \pm t^* \frac{s_x}{\sqrt{n}} = 1.2675 \pm 2.042 \frac{0.3332}{\sqrt{40}} = 1.2675 \pm 0.1076
\]

\[= (1.1599, 1.3751)\]

**Conclude:** We are 95% confident that the interval from 1.1599 to 1.3751 grams/mile captures the true mean level of nitrogen oxides emitted by this type of light-duty engine.
Choosing the Sample Size

We determine a sample size for a desired margin of error when estimating a mean in much the same way we did when estimating a proportion.

Choosing Sample Size for a Desired Margin of Error When Estimating \( \mu \)

To determine the sample size \( n \) that will yield a level \( C \) confidence interval for a population mean with a specified margin of error \( ME \):

1. Get a reasonable value for the population standard deviation \( \sigma \) from an earlier or pilot study.
2. Find the critical value \( z^* \) from a standard Normal curve for confidence level \( C \).
3. Set the expression for the margin of error to be less than or equal to \( ME \) and solve for \( n \):

\[
z^* \frac{\sigma}{\sqrt{n}} \leq ME
\]
Example: Determining sample size from margin of error

Researchers would like to estimate the mean cholesterol level $\mu$ of a particular variety of monkey that is often used in laboratory experiments.

They would like their estimate to be within 1 milligram per deciliter (mg/dl) of the true value of $\mu$ at a 95% confidence level.

A previous study involving this variety of monkey suggests that the standard deviation of cholesterol level is about 5 mg/dl.

**Problem**: Obtaining monkeys is time-consuming and expensive, so the researchers want to know the minimum number of monkeys they will need to generate a satisfactory estimate.
Example: Determining sample size from margin of error

**Solution:** For 95% confidence, $z^* = 1.96$.

We will use $\sigma = 5$ as our best guess for the standard deviation of the monkeys’ cholesterol level.

Set the expression for the margin of error to be at most 1 and solve for $n$:

$$\frac{1.96}{\sqrt{n}} \leq 1$$

$$\frac{(1.96)(5)}{1} \leq \sqrt{n}$$

$$96.04 \leq n$$

Because 96 monkeys would give a slightly larger margin of error than desired, the researchers would need 97 monkeys to estimate the cholesterol levels to their satisfaction.
Estimating a Population Mean

Section Summary

In this section, we learned how to…

✓ **STATE** and **CHECK** the Random, 10%, and Normal/Large Sample conditions for constructing a confidence interval for a population mean.

✓ **EXPLAIN** how the t distributions are different from the standard Normal distribution and why it is necessary to use a t distribution when calculating a confidence interval for a population mean.

✓ **DETERMINE** critical values for calculating a C% confidence interval for a population mean using a table or technology.

✓ **CONSTRUCT** and **INTERPRET** a confidence interval for a population mean.

✓ **DETERMINE** the sample size required to obtain a C% confidence interval for a population mean with a specified margin of error.