This section presents methods for estimating a population mean when the population standard deviation is not known. With \( \sigma \) unknown, we use the Student t distribution assuming that the relevant requirements are satisfied.

**Sample Mean**

The sample mean is the best point estimate of the population mean.

You may use the Excel program to find critical values, confidence interval, and sample size when using the t-distribution. [http://cfcc.edu/~cmoore/0704mean.xls](http://cfcc.edu/~cmoore/0704mean.xls)

### Student t Distribution

If the distribution of a population is essentially normal, then the distribution of

\[
  t = \frac{\bar{x} - \mu}{s/\sqrt{n}}
\]

is a Student t Distribution for all samples of size \( n \). It is often referred to as a t distribution and is used to find critical values denoted by \( t_{a/2} \).

### Definition

The number of *degrees of freedom* for a collection of sample data is the number of sample values that can vary after certain restrictions have been imposed on all data values. The degree of freedom is often abbreviated \( df \).

\[\text{degrees of freedom } df = n - 1\]

in this section.
Margin of Error $E$ for Estimate of $\mu$ (With $\sigma$ Not Known)

Formula 7-6

$$E = t_{\alpha/2} \frac{s}{\sqrt{n}}$$

where $t_{\alpha/2}$ has $n - 1$ degrees of freedom.

Table A-3 lists values for $t_{\alpha/2}$

Notation

$\mu =$ population mean
$\bar{x} =$ sample mean
$s =$ sample standard deviation
$n =$ number of sample values
$E =$ margin of error
$t_{\alpha/2} =$ critical $t$ value separating an area of $\alpha/2$ in the right tail of the $t$ distribution

Confidence Interval for the Estimate of $\mu$ (With $\sigma$ Not Known)

$$\bar{x} - E < \mu < \bar{x} + E$$

where $E = t_{df} \frac{s}{\sqrt{n}}$, df = $n - 1$

$t_{df}$ found in Table A-3

Procedure for Constructing a Confidence interval for $\mu$ (With $\sigma$ Unknown)

1. Verify that the requirements are satisfied.
2. Using $n - 1$ degrees of freedom, refer to Table A-3 or use technology to find the critical value $t_{\alpha/2}$ that corresponds to the desired confidence level.
3. Evaluate the margin of error $E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$.
4. Find the values of $\bar{x} - E$ and $\bar{x} + E$. Substitute those values in the general format for the confidence interval: $\bar{x} - E < \mu < \bar{x} + E$
5. Round the resulting confidence interval limits.
Example:

A common claim is that garlic lowers cholesterol levels. In a test of the effectiveness of garlic, 49 subjects were treated with doses of raw garlic, and their cholesterol levels were measured before and after the treatment. The changes in their levels of LDL cholesterol (in mg/dL) have a mean of 0.4 and a standard deviation of 21.0. Use the sample statistics of $n = 49$, $\bar{x} = 0.4$ and $s = 21.0$ to construct a 95% confidence interval estimate of the mean net change in LDL cholesterol after the garlic treatment. What does the confidence interval suggest about the effectiveness of garlic in reducing LDL cholesterol?

Example:

Requirements are satisfied: simple random sample and $n = 49$ (i.e., $n > 30$).

95% implies $a = 0.05$.

With $n = 49$, the $df = 49 - 1 = 48$

Closest $df$ is 50, two tails, so $t_{a/2} = 2.009$

Using $t_{a/2} = 2.009$, $s = 21.0$ and $n = 49$ the margin of error is:

$$E = t_{a/2} \frac{s}{\sqrt{n}} = 2.009 \times \frac{21.0}{\sqrt{49}} = 6.027$$

Example:

Construct the confidence interval:

$$\bar{x} = 0.4, E = 6.027$$

$$\bar{x} - E < \mu < \bar{x} + E$$

$$0.4 - 6.027 < \mu < 0.4 + 6.027$$

$$-5.6 < \mu < 6.4$$

We are 95% confident that the limits of −5.6 and 6.4 actually do contain the value of $\mu$, the mean of the changes in LDL cholesterol for the population. Because the confidence interval limits contain the value of 0, it is very possible that the mean of the changes in LDL cholesterol is equal to 0, suggesting that the garlic treatment did not affect the LDL cholesterol levels. It does not appear that the garlic treatment is effective in lowering LDL cholesterol.

Important Properties of the Student $t$ Distribution

1. The Student $t$ distribution is different for different sample sizes (see the following slide, for the cases $n = 3$ and $n = 12$).
2. The Student $t$ distribution has the same general symmetric bell shape as the standard normal distribution but it reflects the greater variability (with wider distributions) that is expected with small samples.
3. The Student $t$ distribution has a mean of $t = 0$ (just as the standard normal distribution has a mean of $z = 0$).
4. The standard deviation of the Student $t$ distribution varies with the sample size and is greater than 1 (unlike the standard normal distribution, which has a $\sigma = 1$).
5. As the sample size $n$ gets larger, the Student $t$ distribution gets closer to the normal distribution.
Student t Distributions for \( n = 3 \) and \( n = 12 \)

Choosing the Appropriate Distribution

Use the normal (z) distribution when:
- \( \sigma \) known and normally distributed population
- \( \sigma \) known and \( n > 30 \)

Use the t distribution when:
- \( \sigma \) not known and normally distributed population
- \( \sigma \) not known and \( n > 30 \)

Use a nonparametric method or bootstrapping when:
- Population is not normally distributed and \( n \leq 30 \)

Finding the Point Estimate and E from a Confidence Interval

Point estimate of \( \mu \):
\[ \bar{x} = \frac{(\text{upper confidence limit}) + (\text{lower confidence limit})}{2} \]

Margin of Error:
\[ E = \frac{(\text{upper confidence limit}) - (\text{lower confidence limit})}{2} \]
Confidence Intervals for Comparing Data

As in Sections 7-2 and 7-3, confidence intervals can be used informally to compare different data sets, but the overlapping of confidence intervals should not be used for making formal and final conclusions about equality of means.

Recap

In this section we have discussed:
• Student t distribution.
• Degrees of freedom.
• Margin of error.
• Confidence intervals for μ with σ unknown.
• Choosing the appropriate distribution.
• Point estimates.
• Using confidence intervals to compare data.

Assume that we want to construct a confidence interval using the given confidence level. Do one of the following, as appropriate: (a) Find the critical value \( z_{\frac{\alpha}{2}} \), (b) find the critical value \( t_{\frac{\alpha}{2}} \), (c) state that neither the normal nor the t distribution applies.

371/6 99%; \( n = 25; \sigma \) is known; population appears to be normally distributed.

371/8 95%; \( n = 40; \sigma \) is unknown; population appears to be skewed.
Assume that we want to construct a confidence interval using the given confidence level. Do one of the following, as appropriate: (a) Find the critical value \( z_{\alpha/2} \), (b) find the critical value \( t_{\alpha/2} \), (c) state that neither the normal nor the \( t \) distribution applies.

371/10. 95% confidence; \( n = 9 \); \( \sigma \) is unknown; population appears to be very skewed.

Assume that we want to construct a confidence interval using the given confidence level. Do one of the following, as appropriate: (a) Find the critical value \( z_{\alpha/2} \), (b) find the critical value \( t_{\alpha/2} \), (c) state that neither the normal nor the \( t \) distribution applies.

371/12. 95% confidence; \( n = 38 \); \( \sigma \) is unknown; population appears to be skewed.

Use the given confidence level and sample data to find (a) the margin of error and (b) the confidence interval for the population mean. Assume that the sample is a simple random sample and the population has a normal distribution.

372/13. Hospital Costs 95% confidence; \( n = 20 \); \( \bar{x} = \$9004 \); \( s = \$569 \) (based on data from hospital costs for car crash victims who wore seat belts, from the U.S. Department of Transportation)

Use the given confidence level and sample data to find (a) the margin of error and (b) the confidence interval for the population mean. Assume that the sample is a simple random sample and the population has a normal distribution.

372/14. Car Pollution 99% confidence; \( n = 7 \); \( \bar{x} = 0.12 \); \( s = 0.04 \) (original values are nitrogen-oxide emissions in grams/mile, from the Environmental Protection Agency)
Birth Weights
A random sample of the birth weights of 186 babies has a mean of 3103 g and a standard deviation of 696 g (based on data from “Cognitive Outcomes of Preschool Children with Prenatal Cocaine Exposure,” by Singer et al., Journal of the American Medical Association, Vol. 291, No. 20). These babies were born to mothers who did not use cocaine during their pregnancies.

a. What is the best point estimate of the mean weight of babies born to mothers who did not use cocaine during their pregnancies?

b. Construct a 95% confidence interval estimate of the mean birth weight for all such babies.

c. Compare the confidence interval from part (b) to this confidence interval obtained from birth weights of babies born to mothers who used cocaine during pregnancy: 2608 g < μ < 2792 g. Does cocaine use appear to affect the birth weight of a baby?

Atkins Weight Loss Program
In a test of the Atkins weight loss program, 40 individuals participated in a randomized trial with overweight adults. After 12 months, the mean weight loss was found to be 2.1 lb, with a standard deviation of 4.8 lb.

a. What is the best point estimate of the mean weight loss of all overweight adults who follow the Atkins program?

b. Construct a 99% confidence interval estimate of the mean weight loss for all such subjects.

c. Does the Atkins program appear to be effective? Is it practical?

Ages of Oscar Winning Actresses and Actors
The ages of the 79 actresses at the time that they won Oscars for the Best Actress category have a mean of 35.8 years and a standard deviation of 11.3 years. The ages of the 79 actors at the time that they won Oscars for the category of Best Actor have a mean of 43.8 years and a standard deviation of 8.9 years. Assume that the samples are simple random samples.

a. Construct the 99% confidence interval estimate of the mean age of actresses at the time that they win Oscars for the Best Actress category.

b. Construct the 99% confidence interval estimate of the mean age of actors at the time that they win Oscars for the Best Actor category.

c. Compare the results.

Estimating Car Pollution
In a sample of seven cars, each car was tested for nitrogen-oxide emissions (in grams per mile) and the following results were obtained: 0.06, 0.11, 0.16, 0.15, 0.14, 0.08, 0.15 (based on data from the EPA). Assuming that this sample is representative of the cars in use, construct a 98% confidence interval estimate of the mean amount of nitrogen-oxide emissions for all cars. If the EPA requires that nitrogen-oxide emissions be less than 0.165 g/mi, can we safely conclude that this requirement is being met?