Section 3.3 – Estimating a Population Mean

Point Estimate of the Population Mean

The sample mean $\bar{x}$ is the best ***point estimate*** of the population mean μ.

Confidence Interval for Estimating a Population Mean (with σ Known)

Objective

Construct a confidence interval used to estimate a population mean.

Notation

μ = population mean
σ = population standard deviation
$\bar{x}$ = sample mean
n = number of sample values
E = margin of error
$z_{\alpha/2} = z$ score separating an area of $\alpha/2$ in the right tail of the standard normal distribution

Requirements

1. The sample is a simple random sample. (All samples of the same size have an equal chance of being selected.)
2. The value of the population standard deviation σ is known.
3. Either or both of these conditions is/are satisfied: The population is normally distributed or $n > 30$.

Confidence Interval

$$\bar{x} - E < \mu < \bar{x} + E \quad \textit{wher} \quad E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Or $\bar{x} \pm E$ Or $(\bar{x} - E, \bar{x} + E)$

Definition

The two values $\bar{x} - E$ and $\bar{x} + E$ are called ***confidence interval limits***.

Sample Mean

1. For all populations, the sample mean x is an unbiased estimator of the population mean μ, meaning that the distribution of sample means tends to center about the value of the population mean μ.
2. For many populations, the distribution of sample means x tends to be more consistent (with less variation) than the distributions of other sample statistics.

Procedure for Constructing a Confidence Interval for μ (with Known σ)

1. Verify that the requirements are satisfied.
2. Refer to Standard Normal Distribution Table or use technology to find the critical value $z_{\alpha/2}$ that corresponds to the desired confidence level.
3. Evaluate the margin of error $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
4. Find the values of $\bar{x} - E$ and $\bar{x} + E$. Substitute those values in the general format of the confidence interval: $\bar{x} - E < \mu < \bar{x} + E$
5. Round using the confidence intervals round-off rules.

Round-Off Rule for Confidence Intervals Used to Estimate μ

1. When using the *original set of data*, round the confidence interval limits to one more decimal place than used in original set of data.
2. When the original set of data is unknown and only the **summary statistics** (n, $\bar{x}$, s) are used, round the confidence interval limits to the same number of decimal places used for the sample mean.

Example

People have died in boat and aircraft accidents because an obsolete estimate of the mean weight of men was used. In recent decades, the mean weight of men has increased considerably, so we need to update our estimate of that mean so that boats, aircraft, elevators, and other such devices do not become dangerously overloaded. Using the weights of men from Data Set 1 in Appendix B, we obtain these sample statistics for the simple random sample: $n = 40$ and $\bar{x} = 172.55$ lb. Research from several other sources suggests that the population of weights of men has a standard deviation given by $\sigma = 26$ lb.

a) Find the best point estimate of the mean weight of the population of all men.

b) Construct a 95% confidence interval estimate of the mean weight of all men.

c) What do the results suggest about the mean weight of 166.3 lb that was used to determine the safe passenger capacity of water vessels in 1960 (as given in the National Transportation and Safety Board safety recommendation M-04-04)?

Solution

a) The sample mean of 172.55 lb. is the best point estimate of the mean weight of the population of all men

b) A 95% confidence interval or 0.95 implies $\sigma = 0.05$, so $z_{\alpha/2} = 1.96$.

Calculate the margin of error.

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{26}{\sqrt{40}}$$

$$\underline{= 8.0574835}$$

The confidence interval: $\bar{x} - E < \mu < \bar{x} + E$

$$172.55 - 8.0574835 < \mu < 172.55 + 8.0574835$$

$$\underline{164.49 < \mu < 180.61}$$

c) Based on the confidence interval, it is possible that the mean weight of 166.3 lb. used in 1960 could be the mean weight of men today. However, the best point estimate of 172.55 lb. suggests that the mean weight of men is now considerably greater than 166.3 lb. considering that an underestimate of the mean weight of men could result in lives lost through overloaded boats and aircraft, these results strongly suggest that additional data should be collected. (Additional data have been collected, and the assumed mean weight of men has been increased.)

Finding a Sample Size for Estimating a Population Mean

μ = population mean

σ = population standard deviation

$\bar{x}$ = population standard deviation

E = desired margin of error

$z_{\alpha/2}$ = z score separating an area of $\alpha/2$ in the right tail of the standard normal distribution

$$n = \left[\frac{z_{\alpha/2} \bullet \sigma}{E}\right]^2$$

Finding the Sample Size *n* when σ is Unknown

1. Use the range rule of thumb to estimate the standard deviation as follows: $\sigma \approx$ range/4.
2. Start the sample collection process without knowing σ and, using the first several values, calculate the sample standard deviation s and use it in place of σ. The estimated value of σ can then be improved as more sample data are obtained, and the sample size can be refined accordingly.
3. Estimate the value of σ by using the results of some other study that was done earlier.

Example

Assume that we want to estimate the mean IQ score for the population of statistics students. How many statistics students must be randomly selected for IQ tests if we want 95% confidence that the sample mean is within 3 IQ points of the population mean?

Solution

For 95% confidence interval, we have $\alpha = 0.05 \quad so \quad z_{\alpha/2} = 1.96$

Since we need the sample mean to be within 3 IQ points of μ, the margin of error is E = 3. Also, $\sigma = 15$

$$n = \left[\frac{z_{\alpha/2} \bullet \sigma}{E}\right]^2 = \left[\frac{1.96 \bullet 15}{3}\right]^2$$

$$\approx 97$$

With a simple random sample of only 97 statistics students, we will be 95% confident that the sample mean is within 3 IQ points of the true population mean μ.

Student *t* Distribution

Definitions

Suppose that a simple random sample of size *n* is taken from a population. If the population from which the sample is drawn follows a normal distribution, the distribution of

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

follows Student's *t*-distribution with $n - 1$ degrees of freedom where $\bar{x}$ is the sample mean and *s* is the sample standard deviation.

Distribution for all samples of size *n*. It is often referred to as a *t* distribution and is used to find criti***cal values*** denoted by $t_{\alpha/2}$.

The number of ***degrees of freedom*** (*df*).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.

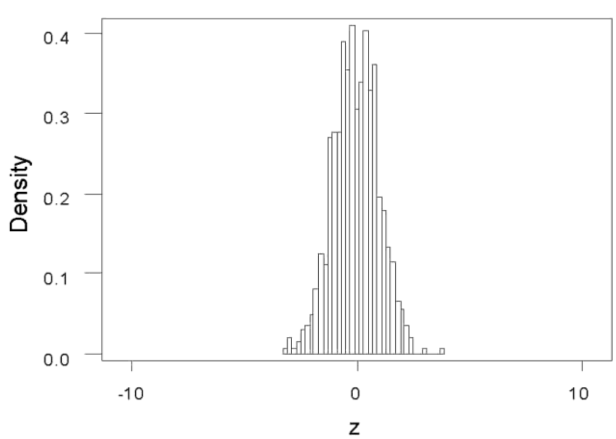
Degrees of freedom: *df* = *n* – 1

Example

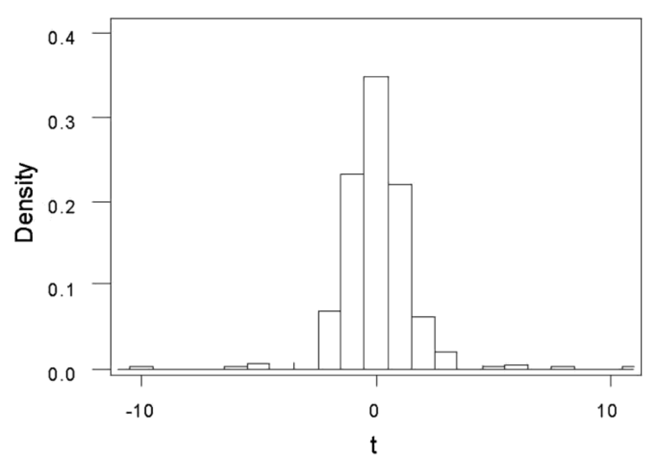
Obtain 1,000 simple random samples of size $n = 5$ from a normal population with $\mu = 50$ and $\sigma = 10$. Determine the sample mean and sample standard deviation for each of the samples.

Compute $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ and $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ for each sample. Draw a histogram for both *z* and *t*.

Solution



Histogram for z

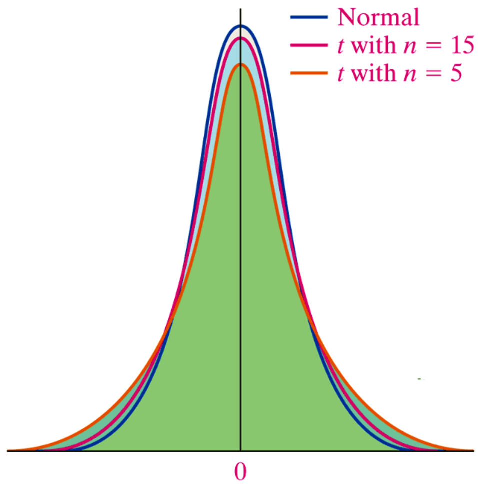


Histogram for t

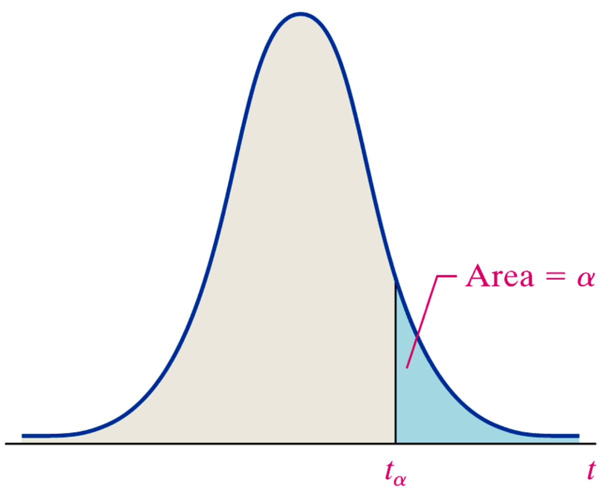
- ✓ The histogram for *z* is symmetric and bell-shaped with the center of the distribution at 0 and virtually all the rectangles between –3 and 3. In other words, *z* follows a standard normal distribution
- ✓ The histogram for *t* is also symmetric and bell-shaped with the center of the distribution at 0, but the distribution of *t* has longer tails (i.e., *t* is more dispersed), so it is unlikely that *t* follows a standard normal distribution. The additional spread in the distribution of *t* can be attributed to the fact that we use *s* to find *t* instead of *σ*. Because the sample standard deviation is itself a random variable (rather than a constant such as *σ*), we have more dispersion in the distribution of *t*.

Properties of the *t*-Distribution

1. The *t*-distribution is different for different degrees of freedom.
2. The *t*-distribution is centered at 0 and is symmetric about 0.
3. The area under the curve is 1. The area under the curve to the right of 0 equals the area under the curve to the left of 0, which equals 1/2.
4. As t increases or decreases without bound, the graph approaches, but never equals, zero.
5. The area in the tails of the *t*-distribution is a little greater than the area in the tails of the standard normal distribution, because we are using s as an estimate of σ, thereby introducing further variability into the *t*- statistic.
6. As the sample size n increases, the density curve of t gets closer to the standard normal density curve. This result occurs because, as the sample size n increases, the values of s get closer to the values of σ, by the Law of Large Numbers.



The notation t_α is the ***t*-value** such that the area under the standard normal curve to the right is α.



Example

Find the *t*-value such that the area under the *t*-distribution to the right of the *t*-value is 0.2 assuming 10 degrees of freedom. That is, find $t_{0.20}$ with 10 degrees of freedom.

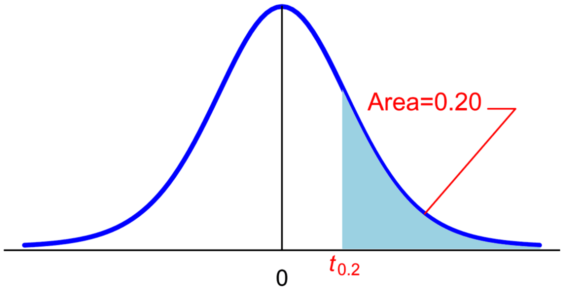
Solution

t distribution critical values

	Upper-tail probability *p*											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587

$df = 10$

$t_{0.20} = 0.879$



Example

A sample of size $n = 7$ is a simple random sample selected from a normally distributed population. Find the critical value $t_{\alpha/2}$ corresponding to a 95% confidence level.

Solution

Because $n = 7$, the number of degrees of freedom is: $n - 1 = 6$.

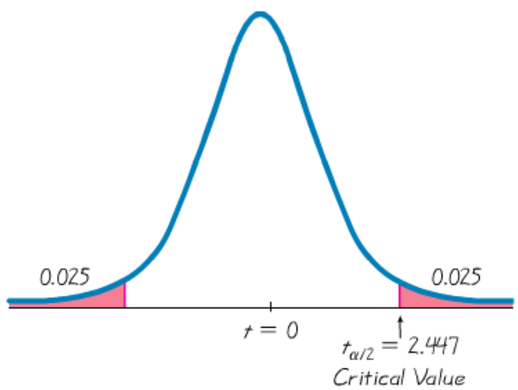
Using *t*–Distribution Table:

Degrees of Freedom	Area in Two Tails 0.01	0.02	0.05	0.10	0.20
6	**3.707**	3.143	**2.447**	1.943	1.440

The value is 2.447.

$t_{\alpha/2} = t_{0.05/2} = t_{0.025} = 2.447$

Such critical values $t_{\alpha/2}$ are used for the margin of error E and confidence interval.



Margin of Error *E* for Estimate of *m* (With σ Not Known)

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

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

Notation

μ = 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 μ (With σ Not Known)

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

Where $E = t_{\alpha/2} \frac{s}{\sqrt{n}}$ $\quad df = n - 1$

Procedure for Constructing a Confidence Interval for μ (With σ 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 $t_{a/2}$ that corresponds to the desired confidence level.
3. Evaluate the margin of error $E = t_{\alpha/2} \frac{s}{\sqrt{n}}$
4. Find the values of $\bar{x} - E < \mu < \bar{x} + E$. Substitute those values in the general format for the confidence interval:
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?

Solution

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 $t_{\alpha/2} = 2.009$

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

$$E = t_{\alpha/2}\frac{s}{\sqrt{n}}$$
$$= 2.009 \cdot \frac{21.0}{\sqrt{49}}$$
$$\boxed{= 6.027}$$

Construct the confidence interval: $\bar{x} = 0.4; \quad E = 6.027$

$$\bar{x} - E < \mu < \bar{x} + E$$
$$0.4 - 6.027 < \mu < 0.4 + 6.027$$
$$\boxed{-5.6 < \mu < 6.4}$$

We are 95% confident that the limits of –5.6 and 6.4 actually do contain the value of μ, 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.

Example

Construct a 99% confidence interval about the population mean weight (in grams) of pennies minted after 1982. Assume μ = 0.02 grams.

2.46	2.47	2.49	2.48	2.5	2.44	2.46	2.45	2.49
2.47	2.45	2.46	2.45	2.46	2.47	2.44	2.45	

Solution

$$t_{\alpha/2} = t_{.01/2} = t_{.005} \boxed{= 1.746}$$

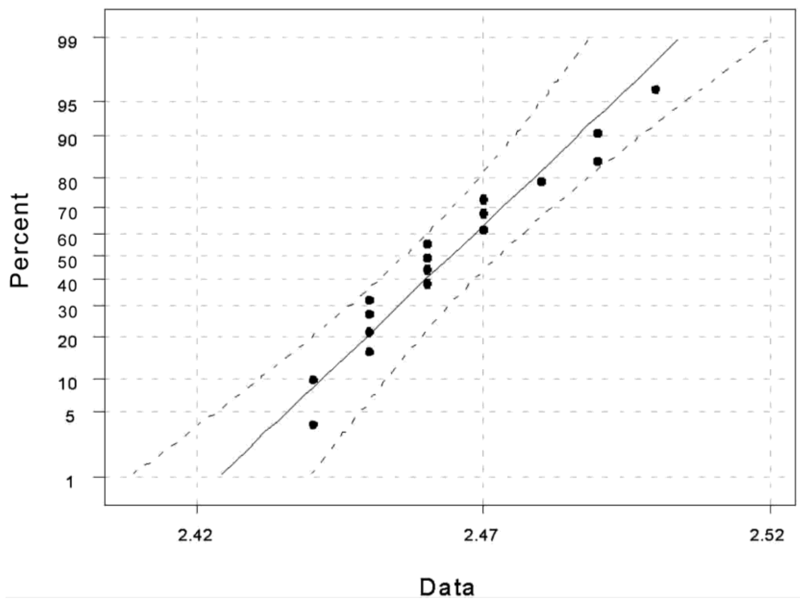
$$E = t_{\alpha/2}\frac{s}{\sqrt{n}} = 1.746\frac{.02}{\sqrt{17}} \boxed{= .008}$$

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

$$2.464 - .008 < \mu < 2.464 + .008$$

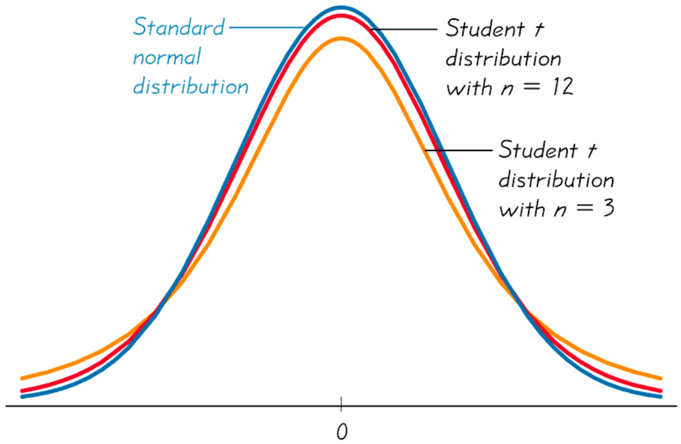
$$\boxed{2.456 < \mu < 2.472}$$

We are 99% confident that the mean weight of pennies minted after 1982 is between 2.456 and 2.472 grams.



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 s = 1).

5. As the sample size n gets larger, the Student *t* distribution gets closer to the normal distribution.

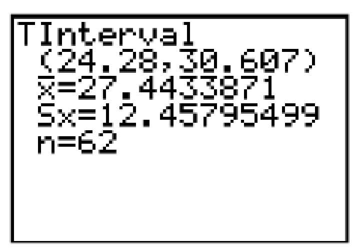
Finding the Point Estimate and *E* from a Confidence Interval

Point estimate of μ: $$\bar{x} = \frac{(upper\ confidence\ limit) + (lower\ confidence\ limit)}{2}$$

Margin of Error: $$E = \frac{(upper\ confidence\ limit) - (lower\ confidence\ limit)}{2}$$

TI-83/84 PLUS The TI-83/84 Plus calculator can be used to generate confidence intervals for original sample values stored in a list, or you can use the summary statistics n, $\bar{x}$, and s. Either enter the data in list L1 or have the summary statistics available, then press the **STAT** key. Now select **TESTS** and choose **TInterval** if σ is not known. (Choose **ZInterval** if σ is known.) After making the required entries, the calculator display will include the confidence interval in the format of ($\bar{x} - E$, $\bar{x} + E$).

TI-83/84 PLUS



Confidence Intervals for Comparing Data

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.

Determining the Sample Size *n*

The sample size required to estimate the population mean, μ, with a level of confidence $(1-\alpha)\cdot 100\%$ with a specified margin of error, E, is given by

$$n = \left[\frac{z_{\alpha/2} \bullet s}{E}\right]^2$$

where n is rounded up to the nearest whole number.

Example

How large a sample would be required to estimate the mean weight of a penny manufactured after 1982 within 0.005 grams with 99% confidence? Assume $\sigma = 0.02$.

Solution

Given: $s = 0.02 \quad E = 0.005$

$z_{\alpha/2} = z_{0.005} = 2.575$

$$n = \left(\frac{z_{\alpha/2} \bullet s}{E}\right)^2 = \left(\frac{(2.575)(0.02)}{0.005}\right)^2 = 106.09$$

$\therefore n = 107$

Exercises *Section* 3.3 – Estimating a Population Mean

1. A design engineer of the Ford Motor Company must estimate the mean leg length of all adults. She obtains a list of the 1275 employees at her facility; then obtains a simple random sample of 50 employees. If she uses this sample to construct a 95% confidence interval to estimate the mean leg length for the population of all adults, will her estimate be good? Why or why not?

2. Find the critical value $z_{\alpha/2}$ that corresponds to a 90% confidence level.

3. Find the critical value $z_{\alpha/2}$ that corresponds to a 98% confidence level.

4. Find $z_{\alpha/2}$ for $\alpha = 0.20$

5. Find $z_{\alpha/2}$ for $\alpha = 0.07$

6. How many adults must be randomly selected to estimate the mean FICO (credit rating) score of working adults in U.S.? We want 95% confidence that the sample mean is within 3 points of the population mean, and the population standard deviation is 68.

7. A simple random sample of 40 salaries of NCAA football coaches has a mean of \$415,953. Assume that σ = \$463,364.

a) Find the best estimate of the mean salary of all NCAA football coaches.
b) Construct a 95% confidence interval estimate of the mean salary of an NCAA football coach.
c) Does the confidence interval contain the actual population mean of \$474,477?

8. A simple random sample of 50 adults (including males and females) is obtained, and each person's red blood cell count (in cells per microliter) is measured. The sample mean is 4.63. The population standard deviation for red blood cell counts is 0.54.

a) Find the best point estimate of the mean red blood cell count of adults.
b) Construct a 99% confidence interval estimate of the mean red blood cell count of adults.
c) The normal range of red blood cell counts for adults is 4.7 to 6.1 for males and 4.3 to 5.4 for females. What does the confidence interval suggest about these normal ranges?

9. A simple random sample of 125 SAT scores has a mean of 1522. Assume that SAT scores have a standard deviation of 333.

a) Construct a 95% confidence interval estimate of the mean SAT score.
b) Construct a 99% confidence interval estimate of the mean SAT score.
c) Which of the preceding confidence intervals is wider? Why?

10. When 14 different second-year medical students measured the blood pressure of the same person, they obtained the results listed below. Assuming that the population standard deviation is known to be 10 mmHg, construct a 95% confidence interval estimate of the population mean. Ideally, what should the confidence interval be in this situation?

138 130 135 140 120 125 120 130 130 144 143 140 130 150

11. Do the given conditions justify using the margin of error $E = z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$ when finding a confidence interval estimate of the population mean μ?

a) The sample size is $n = 4$, $\sigma = 12.5$, and the original population is normally distributed

b) The sample size is $n = 5$ and σ is not known

12. Use the confidence level and sample data to find the margin of error E.

a) Replacement times for washing machines: 90% confidence; $n = 37$, $\bar{x} = 10.4\ yrs$, $\sigma = 2.2\ yrs$

b) College students' annual earnings: 99% confidence; $n = 76$, $\bar{x} = \$4196$, $\sigma = \$848$

13. Use the confidence level and sample data to find a confidence interval for estimating the population μ. A laboratory tested 89 chicken eggs and found that the mean amount of cholesterol was 203 milligrams with $\sigma = 11.4\ mg$. Construct a 95% confidence interval for the true mean cholesterol content μ, of all such eggs.

14. Use the confidence level and sample data to find a confidence interval for estimating the population μ. A group of 66 randomly selected students have a mean score of 34.3 on a placement test. The population standard deviation $\sigma = 3$. What is the 90% confidence interval for the mean score, μ, of all students taking the test?

15. Use the given information to find the minimum sample size required to estimate an unknown population mean μ. Margin error: \$139, confidence level: 99%, σ = \$522

16. What does it mean when we say that the methods for constructing confidence intervals in this section are robust against departures from normality? Are the methods for constructing confidence intervals in this section robust against poor sampling methods?

17. Assume that we want tic instruct a confidence interval using the given confidence level 95%; $n = 23$; $\boldsymbol{\sigma}$ is unknown; population appears to be normally distributed. Do one of the following

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.

18. Assume that we want tic instruct a confidence interval using the given confidence level 99%; $n = 25$; $\boldsymbol{\sigma}$ is known; population appears to be normally distributed. Do one of the following

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.

19. Assume that we want tic instruct a confidence interval using the given confidence level 99%; $n = 6$; $\boldsymbol{\sigma}$ is unknown; population appears to be very skewed. Do one of the following

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.

20. Assume that we want tic instruct a confidence interval using the given confidence level 90%; $n = 200$; $\boldsymbol{\sigma}$ = 15.0; population appears to be skewed. Do one of the following

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.

21. Assume that we want tic instruct a confidence interval using the given confidence level 95%; $n = 9$; $\boldsymbol{\sigma}$ is unknown; population appears to be very skewed. Do one of the following

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.

22. Given $95\%\ \mathit{confidence};\quad n = 20,\quad \overline{x} = \$9004,\quad s = \$569$. Assume that the sample is a simple random and the population has a normal distribution.

a) Find the margin error

b) Find the confidence interval for the population mean μ.

23. Given $99\%\ \mathit{confidence};\quad n = 7,\quad \overline{x} = 0.12,\quad s = 0.04$. Assume that the sample is a simple random and the population has a normal distribution.

a) Find the margin error

b) Find the confidence interval for the population mean μ.

24. In a test of the effectiveness of garlic for lowering cholesterol, 47 subjects were treated with Garlicin, which is garlic in a processed tablet form. Cholesterol levels were measured before and after the treatment. The changes in their levels of LDL cholesterol (in mg/dL) have a mean of 3.2 and standard deviation of 18.6.

a) What is the best point estimate of the population mean net change LDL cholesterol after Garlicin treatment?

b) Construct a 95% confidence interval estimate of the mean net change in LDL cholesterol after the Garlicin treatment. What does the confidence interval suggest about the effectiveness of Garlicin in reducing LDL cholesterol?

25. A random sample of the birth weights of 186 babies has a mean of 3103 g and a standard deviation of 696 g. 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 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: $2608\ g < \mu < 2792\ g$. Does cocaine use appear to affect the birth weight of a baby?

26. In a study designed to test the effectiveness of acupuncture for treating migraine, 142 subjects were treated with acupuncture and 80 subjects were given a sham treatment. The numbers of migraine attacks for the acupuncture treatment group had a mean of 1.8 and a standard deviation of 1.4. The numbers of migraine attacks for the sham treatment group had a mean of 1.6 and standard deviation of 1.2

a) Construct a 95% confidence interval estimate of the mean number of migraine attacks for those treated with acupuncture.

b) Construct a 95% confidence interval estimate of the mean number of migraine attacks for those given a sham treatment.

c) Compare the two confidence intervals. What do the results about the effectiveness of acupuncture?

27. 30 randomly selected students took the statistics final. If the sample mean was 79 and the standard deviation was 14.5, construct a 99% confidence interval for the mean score of all students.
Use the given degree of confidence and sample data to construct a confidence level interval for the population mean μ. Assume that the population has a normal distribution.