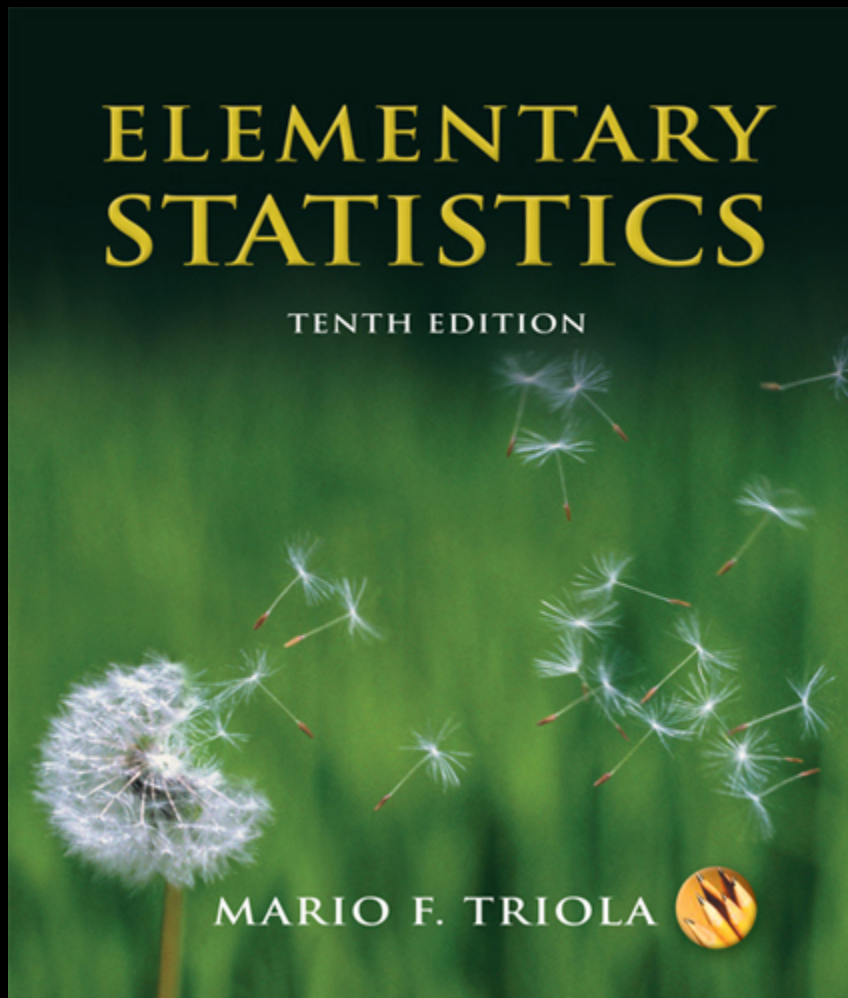


Lecture Slides



Elementary Statistics Tenth Edition

and the Triola Statistics Series

by Mario F. Triola

Chapter 7

Estimates and Sample Sizes

7-1 Overview

7-2 Estimating a Population Proportion

7-3 Estimating a Population Mean: σ Known

7-4 Estimating a Population Mean: σ Not Known

7-5 Estimating a Population Variance

Section 7-1 Overview



Created by Erin Hodgess, Houston, Texas
Revised to accompany 10th Edition, Tom Wegleitner, Centreville, VA



Overview

This chapter presents the beginning of inferential statistics.

- ❖ **The two major applications of inferential statistics involve the use of sample data to (1) estimate the value of a population parameter, and (2) test some claim (or hypothesis) about a population.**
- ❖ **We introduce methods for estimating values of these important population parameters: proportions, means, and variances.**
- ❖ **We also present methods for determining sample sizes necessary to estimate those parameters.**



Section 7-2

Estimating a Population Proportion

Created by Erin Hodgess, Houston, Texas
Revised to accompany 10th Edition, Tom Wegleitner, Centreville, VA



Key Concept

In this section we present important methods for using a sample proportion to estimate the value of a population proportion with a **confidence interval**. We also present methods for finding the size of the sample needed to estimate a population proportion.

Requirements for Estimating a Population Proportion

- 1. The sample is a simple random sample.**
- 2. The conditions for the binomial distribution are satisfied. (See Section 5-3.)**
- 3. There are at least 5 successes and 5 failures.**

Notation for Proportions

p = **population** proportion

$\hat{p} = \frac{x}{n}$ **sample proportion**
of x **successes** in a sample of size n
↑
(pronounced 'p-hat')

$\hat{q} = 1 - \hat{p} =$ **sample proportion**
of **failures** in a sample size of n

Definition

A **point estimate** is a single value (or point) used to approximate a population parameter.

Definition

The sample proportion \hat{p} is the best point estimate of the population proportion p .

Example: 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Using these survey results, find the best point estimate of the proportion of all adult Minnesotans opposed to photo-cop use.

Because the sample proportion is the best point estimate of the population proportion, we conclude that the best point estimate of p is 0.51. When using the survey results to estimate the percentage of all adult Minnesotans that are opposed to photo-cop use, our best estimate is 51%.

Definition

A confidence interval (or interval estimate) is a range (or an interval) of values used to estimate the true value of a population parameter. A confidence interval is sometimes abbreviated as CI.

Definition

A **confidence level** is the probability $1 - \alpha$ (often expressed as the equivalent percentage value) that is the proportion of times that the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times. (The confidence level is also called **degree of confidence**, or the **confidence coefficient**.)

Most common choices are 90%, 95%, or 99%.

$(\alpha = 10\%), (\alpha = 5\%), (\alpha = 1\%)$

Example: 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Using these survey results, find the 95% confidence interval of the proportion of all adult Minnesotans opposed to photo-cop use.

“We are 95% confident that the interval from 0.476 to 0.544 does contain the true value of p .”

This means if we were to select many different samples of size 829 and construct the corresponding confidence intervals, 95% of them would actually contain the value of the population proportion p .

Using Confidence Intervals for Comparisons

Do not use the **overlapping** of confidence intervals as the basis for making formal and final conclusions about the equality of proportions.

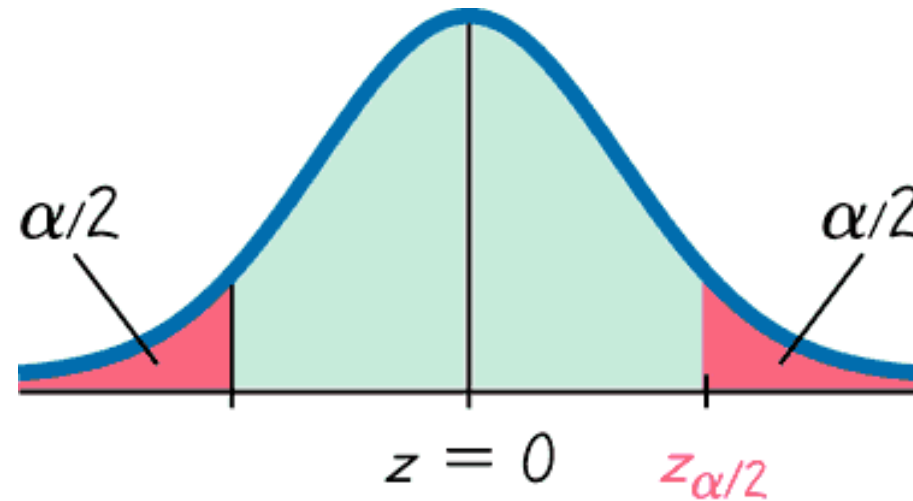
Critical Values

- 1. We know from Section 6-6 that under certain conditions, the sampling distribution of sample proportions can be approximated by a normal distribution, as in Figure 7-2, following.**
- 2. Sample proportions have a relatively small chance (with probability denoted by α) of falling in one of the red tails of Figure 7-2, following.**
- 3. Denoting the area of each shaded tail by $\alpha/2$, we see that there is a total probability of α that a sample proportion will fall in either of the two red tails.**

Critical Values

4. By the rule of complements (from Chapter 4), there is a probability of $1-\alpha$ that a sample proportion will fall within the inner region of Figure 7-2, following.
5. The z score separating the right-tail is commonly denoted by $z_{\alpha/2}$ and is referred to as a **critical value** because it is on the borderline separating sample proportions that are likely to occur from those that are unlikely to occur.

The Critical Value $z_{\alpha/2}$



Found from \uparrow
Table A-2
(corresponds to
area of $1 - \alpha/2$)

Figure 7-2

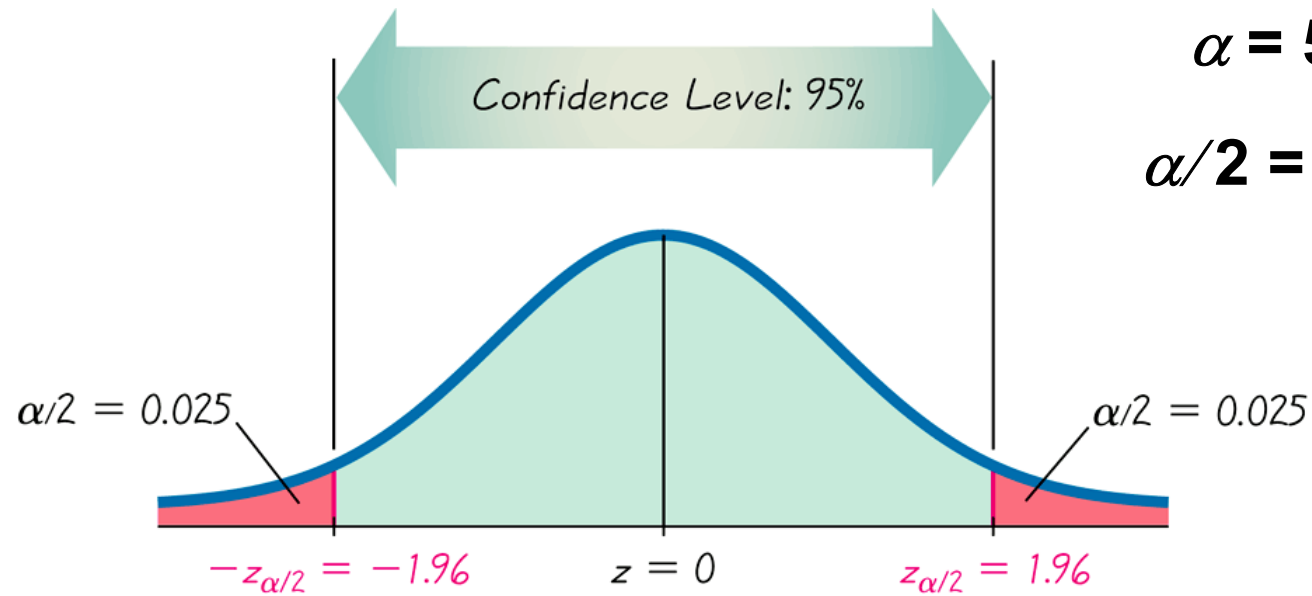
Notation for Critical Value

The critical value $z_{\alpha/2}$ is the positive z value that is at the vertical boundary separating an area of $\alpha/2$ in the right tail of the standard normal distribution. (The value of $-z_{\alpha/2}$ is at the vertical boundary for the area of $\alpha/2$ in the left tail.) The subscript $\alpha/2$ is simply a reminder that the z score separates an area of $\alpha/2$ in the right tail of the standard normal distribution.

Definition

A **critical value** is the number on the borderline separating sample statistics that are likely to occur from those that are unlikely to occur. The number $z_{\alpha/2}$ is a critical value that is a z score with the property that it separates an area of $\alpha/2$ in the right tail of the standard normal distribution. (See Figure 7-2).

Finding $z_{\alpha/2}$ for a 95% Confidence Level



$$\alpha = 5\%$$

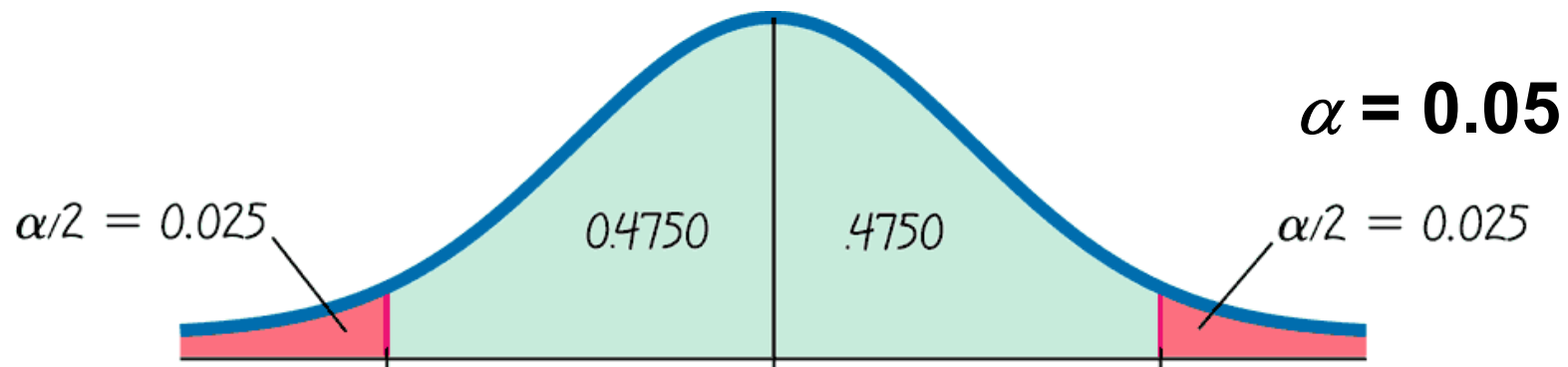
$$\alpha/2 = 2.5\% = .025$$

$$-z_{\alpha/2}$$

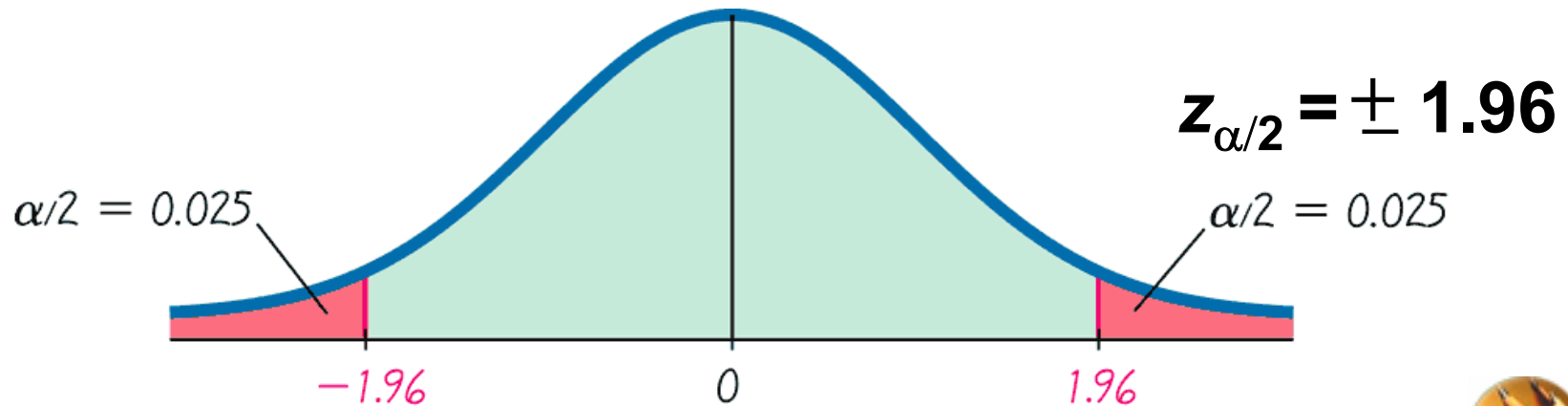
$$z_{\alpha/2}$$

Critical Values

Finding $z_{\alpha/2}$ for a 95% Confidence Level - cont



Use Table A-2 to find a z score of 1.96



Definition

When data from a simple random sample are used to estimate a population proportion p , the **margin of error**, denoted by E , is the maximum likely (with probability $1 - \alpha$) difference between the observed proportion \hat{p} and the true value of the population proportion p . The margin of error E is also called the **maximum error of the estimate** and can be found by multiplying the critical value and the standard deviation of the sample proportions, as shown in Formula 7-1, following.

Margin of Error of the Estimate of p

Formula 7-1

$$E = Z_{\alpha/2} \sqrt{\frac{\hat{p} \hat{q}}{n}}$$

Confidence Interval for Population Proportion

$$\hat{p} - E < p < \hat{p} + E$$

where

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Confidence Interval for Population Proportion - cont

$$\hat{p} - E < p < \hat{p} + E$$

$$\hat{p} \pm E$$

$$(\hat{p} - E, \hat{p} + E)$$

Round-Off Rule for Confidence Interval Estimates of p

**Round the confidence interval limits
for p to
three significant digits.**

Procedure for Constructing a Confidence Interval for p

1. Verify that the required assumptions are satisfied. (The sample is a simple random sample, the conditions for the binomial distribution are satisfied, and the normal distribution can be used to approximate the distribution of sample proportions because $np \geq 5$, and $nq \geq 5$ are both satisfied.)
2. Refer to Table A-2 and find the critical value $z_{\alpha/2}$ that corresponds to the desired confidence level.

3. Evaluate the margin of error $E = \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Procedure for Constructing a Confidence Interval for p - cont

- Using the value of the calculated margin of error, E and the value of the sample proportion, \hat{p} , find the values of $\hat{p} - E$ and $\hat{p} + E$. Substitute those values in the general format for the confidence interval:

$$\hat{p} - E < p < \hat{p} + E$$

- Round the resulting confidence interval limits to three significant digits.

Example: 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

- a) Find the margin of error E that corresponds to a 95% confidence level.
- b) Find the 95% confidence interval estimate of the population proportion p .
- c) Based on the results, can we safely conclude that the majority of adult Minnesotans oppose use the the photo-cop?

Example: 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

a) Find the margin of error E that corresponds to a 95% confidence level.

First, we check for assumptions. We note that $n\hat{p} = 422.79 \geq 5$, and $n\hat{q} = 406.21 \geq 5$.

Next, we calculate the margin of error. We have found that $\hat{p} = 0.51$, $\hat{q} = 1 - 0.51 = 0.49$, $z_{\alpha/2} = 1.96$, and $n = 829$.

$$E = 1.96 \sqrt{\frac{(0.51)(0.49)}{829}}$$
$$E = 0.03403$$

Example: 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

b) Find the 95% confidence interval for the population proportion p .

We substitute our values from Part a to obtain:

$$0.51 - 0.03403 < p < 0.51 + 0.03403,$$
$$0.476 < p < 0.544$$

Example: 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

c) Based on the results, can we safely conclude that the majority of adult Minnesotans oppose use of the photo-cop?

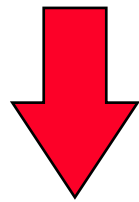
Based on the survey results, we are 95% confident that the limits of 47.6% and 54.4% contain the true percentage of adult Minnesotans opposed to the photo-cop. The percentage of opposed adult Minnesotans is likely to be any value between 47.6% and 54.4%. However, a majority requires a percentage greater than 50%, so we cannot safely conclude that the majority is opposed (because the entire confidence interval is not greater than 50%).

Sample Size

Suppose we want to collect sample data with the objective of estimating some population. The question is **how many** sample items must be obtained?

Determining Sample Size

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p} \hat{q}}{n}}$$



(solve for n by algebra)

$$n = \frac{(z_{\alpha/2})^2 \hat{p} \hat{q}}{E^2}$$

Sample Size for Estimating Proportion p

When an estimate of \hat{p} is known:

$$n = \frac{(Z_{\alpha/2})^2 \hat{p} \hat{q}}{E^2}$$

Formula 7-2

When no estimate of \hat{p} is known:

$$n = \frac{(Z_{\alpha/2})^2 0.25}{E^2}$$

Formula 7-3

Example: Suppose a sociologist wants to determine the current percentage of U.S. households using e-mail. How many households must be surveyed in order to be 95% confident that the sample percentage is in error by no more than four percentage points?

- a) Use this result from an earlier study: In 1997, 16.9% of U.S. households used e-mail (based on data from *The World Almanac and Book of Facts*).
- b) Assume that we have no prior information suggesting a possible value of \hat{p} .

Example: Suppose a sociologist wants to determine the current percentage of U.S. households using e-mail. How many households must be surveyed in order to be 95% confident that the sample percentage is in error by no more than four percentage points?

a) Use this result from an earlier study: In 1997, 16.9% of U.S. households used e-mail (based on data from *The World Almanac and Book of Facts*).

$$\begin{aligned}n &= \frac{[z_{\alpha/2}]^2 \hat{p} \hat{q}}{E^2} \\&= \frac{[1.96]^2 (0.169)(0.831)}{0.04^2} \\&= 337.194 \\&= 338 \text{ households}\end{aligned}$$

To be 95% confident that our sample percentage is within four percentage points of the true percentage for all households, we should randomly select and survey 338 households.

Example: Suppose a sociologist wants to determine the current percentage of U.S. households using e-mail. How many households must be surveyed in order to be 95% confident that the sample percentage is in error by no more than four percentage points?

b) Assume that we have no prior information suggesting a possible value of \hat{p} .

$$\begin{aligned}n &= \frac{[z_{\alpha/2}]^2 \cdot 0.25}{E^2} \\&= \frac{(1.96)^2 (0.25)}{0.04^2} \\&= 600.25 \\&= 601 \text{ households}\end{aligned}$$

With no prior information, we need a larger sample to achieve the same results with 95% confidence and an error of no more than 4%.

Finding the Point Estimate and E from a Confidence Interval

Point estimate of \hat{p} :

$$\hat{p} = \frac{(\text{upper confidence limit}) + (\text{lower confidence limit})}{2}$$

Margin of Error:

$$E = \frac{(\text{upper confidence limit}) - (\text{lower confidence limit})}{2}$$

Recap

In this section we have discussed:

- ❖ **Point estimates.**
- ❖ **Confidence intervals.**
- ❖ **Confidence levels.**
- ❖ **Critical values.**
- ❖ **Margin of error.**
- ❖ **Determining sample sizes.**



Section 7-3

Estimating a Population Mean: σ Known

Created by Erin Hodgess, Houston, Texas
Revised to accompany 10th Edition, Tom Wegleitner, Centreville, VA



Key Concept

This section presents methods for using sample data to find a point estimate and confidence interval estimate of a population mean. A key requirement in this section is that we know the standard deviation of the population.

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 satisfied: The population is normally distributed or $n > 30$.**

Point Estimate of the Population Mean

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

Sample Mean

1. For all populations, the sample mean \bar{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 \bar{x} tends to be more consistent (with **less variation**) than the distributions of other sample statistics.

Example: A study found the body temperatures of 106 healthy adults. The sample mean was 98.2 degrees and the sample standard deviation was 0.62 degrees. Find the point estimate of the population mean μ of all body temperatures.

Because the sample mean \bar{x} is the best point estimate of the population mean μ , we conclude that the best point estimate of the population mean μ of all body temperatures is 98.20° F.

Definition

The **margin of error** is the maximum likely difference observed between sample mean \bar{X} and population mean μ , and is denoted by E .

Formula

Margin of Error

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

Formula 7-4

Margin of error for mean (based on known σ)

Confidence Interval estimate of the Population Mean μ (with σ Known)

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

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**.

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

1. Verify that the requirements are satisfied.
2. Refer to Table A-2 and find the critical value $z_{\alpha/2}$ that corresponds to the desired degree of confidence.
3. Evaluate the margin of error $E = z_{\alpha/2} \cdot \sigma/\sqrt{n}$.
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: A study found the body temperatures of 106 healthy adults. The sample mean was 98.2 degrees and the sample standard deviation was 0.62 degrees. Find the margin of error E and the 95% confidence interval for μ .

$$n = 106$$

$$\bar{x} = 98.20^\circ$$

$$s = 0.62^\circ$$

$$\alpha = 0.05$$

$$\alpha/2 = 0.025$$

$$z_{\alpha/2} = 1.96$$

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{0.62}{\sqrt{106}} = 0.12$$

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

$$98.08^\circ < \mu < 98.32^\circ$$

$$98.20^\circ - 0.12 < \mu < 98.20^\circ + 0.12$$

Example: A study found the body temperatures of 106 healthy adults. The sample mean was 98.2 degrees and the sample standard deviation was 0.62 degrees. Find the margin of error E and the 95% confidence interval for μ .

$$n = 106$$

$$\bar{x} = 98.20^\circ \quad E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{0.62}{\sqrt{106}} = 0.12$$

$$s = 0.62^\circ$$

$$\alpha = 0.05$$

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

$$\alpha/2 = 0.025$$

$$98.08^\circ < \mu < 98.32^\circ$$

$$z_{\alpha/2} = 1.96$$

Based on the sample provided, the confidence interval for the population mean is $98.08^\circ < \mu < 98.32^\circ$. If we were to select many different samples of the same size, 95% of the confidence intervals would actually contain the population mean μ .

Sample Size for Estimating Mean μ

$$n = \left[\frac{(Z_{\alpha/2}) \cdot \sigma}{E} \right]^2$$

Formula 7-5

Where

$z_{\alpha/2}$ = critical z score based on the desired confidence level

E = desired margin of error

σ = population standard deviation

Round-Off Rule for Sample Size n

When finding the sample size n , if the use of Formula 7-5 does not result in a whole number, always **increase** the value of n to the next **larger** whole number.

Finding the Sample Size n When σ is Unknown

1. Use the range rule of thumb (see Section 3-3) to estimate the standard deviation as follows:
 $\sigma \approx \text{range}/4$.
2. Conduct a pilot study by starting the sampling process. Start the sample collection process and, using the first several values, calculate the sample standard deviation s and use it in place of σ .
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 professors. How many statistics professors must be randomly selected for IQ tests if we want 95% confidence that the sample mean is within 2 IQ points of the population mean? Assume that $\sigma = 15$, as is found in the general population.

$$\alpha = 0.05$$

$$\alpha/2 = 0.025$$

$$z_{\alpha/2} = 1.96$$

$$E = 2$$

$$\sigma = 15$$

$$n = \left[\frac{1.96 \cdot 15}{2} \right]^2 = 216.09 = 217$$

With a simple random sample of only 217 statistics professors, we will be 95% confident that the sample mean will be within 2 IQ points of the true population mean μ .

Recap

In this section we have discussed:

- ❖ Margin of error.
- ❖ Confidence interval estimate of the population mean with σ known.
- ❖ Round off rules.
- ❖ Sample size for estimating the mean μ .



Section 7-4

Estimating a Population Mean: σ Not Known

Created by Erin Hodgess, Houston, Texas
Revised to accompany 10th Edition, Tom Wegleitner, Centreville, VA



Key Concept

This section presents methods for finding a confidence interval estimate of a population mean when the population standard deviation is **not known**. With σ unknown, we will use the **Student t distribution** assuming that certain requirements are satisfied.

Requirements with σ Unknown

- 1) The sample is a simple random sample.
- 2) Either the sample is from a normally distributed population, or $n > 30$.

Use Student t distribution

Student t Distribution

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

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

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 $t_{\alpha/2}$.

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.

degrees of freedom = $n - 1$
in this section.

Margin of Error E for Estimate of μ (With σ 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 $t_{\alpha/2}$

Confidence Interval for the Estimate of μ (With σ Not Known)

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

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

$t_{\alpha/2}$ found in Table A-3

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 and find the critical value $t_{\alpha/2}$ that corresponds to the desired confidence level.
3. Evaluate the margin of error $E = t_{\alpha/2} \cdot s / \sqrt{n}$.
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 study found the body temperatures of 106 healthy adults. The sample mean was 98.2 degrees and the sample standard deviation was 0.62 degrees. Find the margin of error E and the 95% confidence interval for μ .

Example: A study found the body temperatures of 106 healthy adults. The sample mean was 98.2 degrees and the sample standard deviation was 0.62 degrees. Find the margin of error E and the 95% confidence interval for μ .

$$n = 106$$

$$\bar{x} = 98.20^\circ$$

$$s = 0.62^\circ$$

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 1.984 \cdot \frac{0.62}{\sqrt{106}} = 0.1195$$

$$\alpha = 0.05$$

$$\alpha/2 = 0.025$$

$$t_{\alpha/2} = 1.96$$

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

$$98.08^\circ < \mu < 98.32^\circ$$

Example: A study found the body temperatures of 106 healthy adults. The sample mean was 98.2 degrees and the sample standard deviation was 0.62 degrees. Find the margin of error E and the 95% confidence interval for μ .

$$n = 106$$

$$\bar{x} = 98.20^\circ$$

$$s = 0.62^\circ$$

$$E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 1.984 \cdot \frac{0.62}{\sqrt{106}} = 0.1195$$

$$\alpha = 0.05$$

$$\alpha/2 = 0.025$$

$$t_{\alpha/2} = 1.96$$

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

$$98.08^\circ < \mu < 98.32^\circ$$

Based on the sample provided, the confidence interval for the population mean is $98.08^\circ < \mu < 98.32^\circ$. The interval is the same here as in Section 7-2, but in some other cases, the difference would be much greater.

Important Properties of the Student t Distribution

1. The Student t distribution is different for different sample sizes (see Figure 7-5, following, 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 $n = 12$

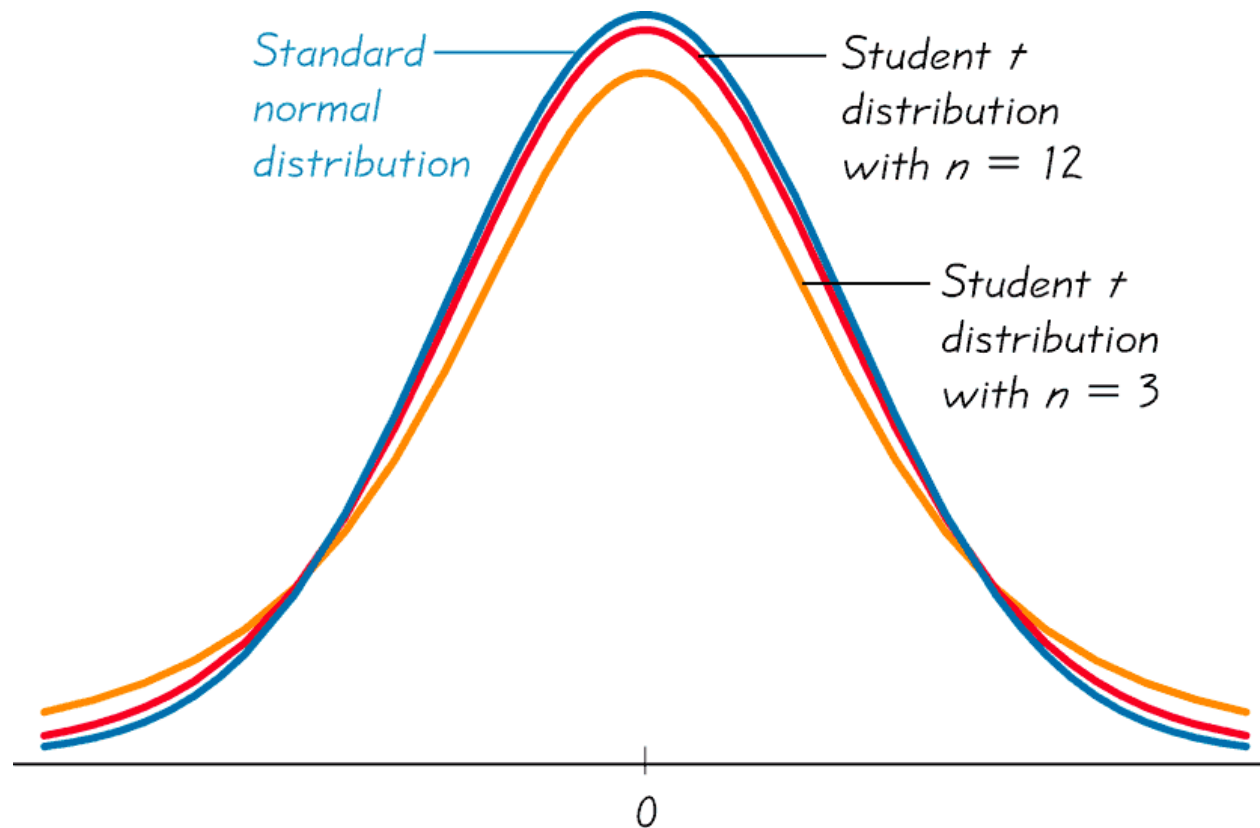


Figure 7-5

Choosing the Appropriate Distribution

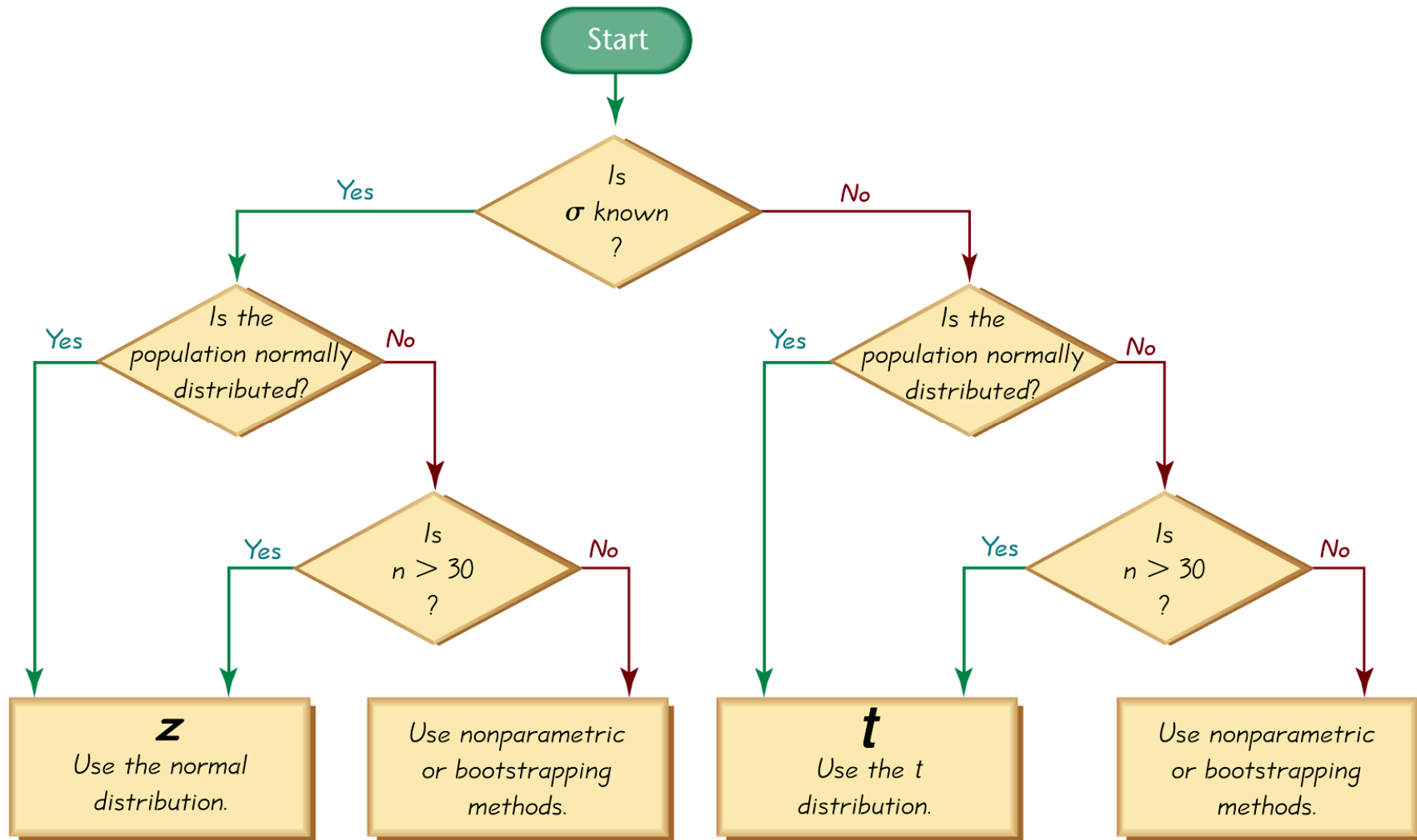


Figure 7-6

Example: Flesch ease of reading scores for 12 different pages randomly selected from J.K. Rowling's *Harry Potter and the Sorcerer's Stone*. Find the 95% interval estimate of μ , the mean Flesch ease of reading score. (The 12 pages' distribution appears to be bell-shaped with $\bar{x} = 80.75$ and $s = 4.68$.)

Example: Flesch ease of reading scores for 12 different pages randomly selected from J.K. Rowling's *Harry Potter and the Sorcerer's Stone*. Find the 95% interval estimate of μ , the mean Flesch ease of reading score. (The 12 pages' distribution appears to be bell-shaped with $\bar{x} = 80.75$ and $s = 4.68$.)

$$\bar{x} = 80.75 \quad E = t_{\alpha/2} \frac{s}{\sqrt{n}} \quad s = \frac{(2.201)(4.68)}{\sqrt{12}} = 2.97355$$

$$s = 4.68$$

$$\alpha = 0.05$$

$$\alpha/2 = 0.025$$

$$t_{\alpha/2} = 2.201$$

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

$$80.75 - 2.97355 < \mu < 80.75 + 2.97355$$

$$77.77645 < \mu < 83.72355$$

$$77.78 < \mu < 83.72$$

We are 95% confident that this interval contains the mean Flesch ease of reading score for all pages.

Finding the Point Estimate and E from a Confidence Interval

Point estimate of μ :

$$\bar{x} = \frac{(\text{upper confidence limit}) + (\text{lower confidence limit})}{2}$$

Margin of Error:

$$E = \frac{(\text{upper confidence limit}) - (\text{lower confidence limit})}{2}$$

Confidence Intervals for Comparing Data

As before in Sections 7-2 and 7-3, do not use the **overlapping** of confidence intervals as the basis for making final conclusions about the 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.**



Section 7-5

Estimating a Population Variance

Created by Erin Hodgess, Houston, Texas
Revised to accompany 10th Edition, Tom Wegleitner, Centreville, VA



Key Concept

This section presents methods for (1) finding a confidence interval estimate of a population standard deviation or variance and (2) determining the sample size required to estimate a population standard deviation or variance.

We also introduce the chi-square distribution, which is used for finding a confidence interval estimate for σ or σ^2 .

Requirements

- 1. The sample is a simple random sample.**
- 2. The population must have normally distributed values (even if the sample is large).**

Chi-Square Distribution

$$\chi^2 = \frac{(n - 1) s^2}{\sigma^2} \quad \text{Formula 7-7}$$

where

n = sample size

s^2 = sample variance

σ^2 = population variance

Properties of the Distribution of the Chi-Square Statistic

1. The chi-square distribution is not symmetric, unlike the normal and Student t distributions. As the number of degrees of freedom increases, the distribution becomes more symmetric.

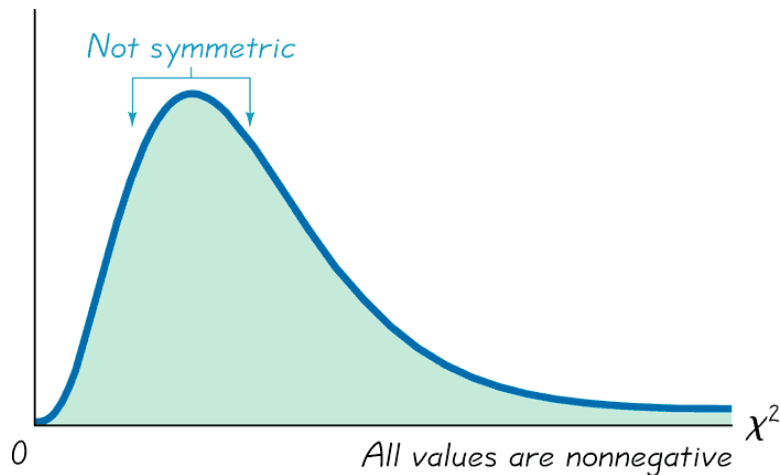


Figure 7-8 Chi-Square Distribution

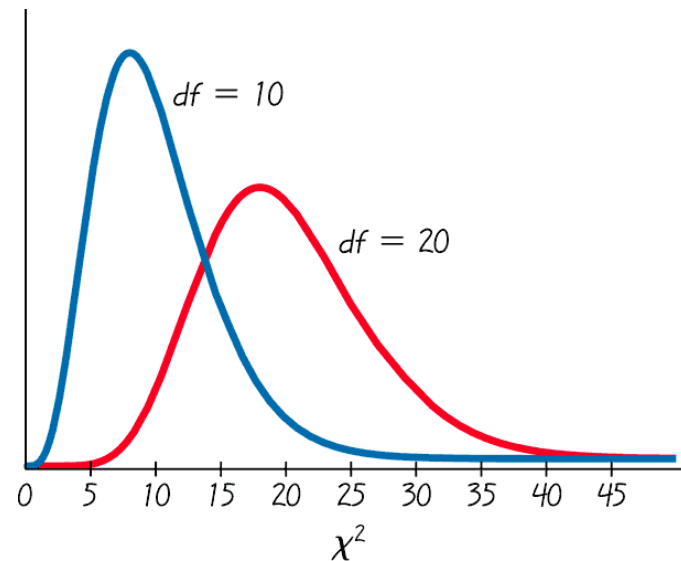


Figure 7-9 Chi-Square Distribution for $df = 10$ and $df = 20$

Properties of the Distribution of the Chi-Square Statistic - cont

2. The values of chi-square can be zero or positive, but they cannot be negative.
3. The chi-square distribution is different for each number of degrees of freedom, which is $df = n - 1$ in this section. As the number increases, the chi-square distribution approaches a normal distribution.

In Table A-4, each critical value of χ^2 corresponds to an area given in the top row of the table, and that area represents the **cumulative area located to the right** of the critical value.

Example:

Find the critical values of χ^2 that determine critical regions containing an area of 0.025 in each tail. Assume that the relevant sample size is 10 so that the number of degrees of freedom is $10 - 1$, or 9.

$$\alpha = 0.05$$

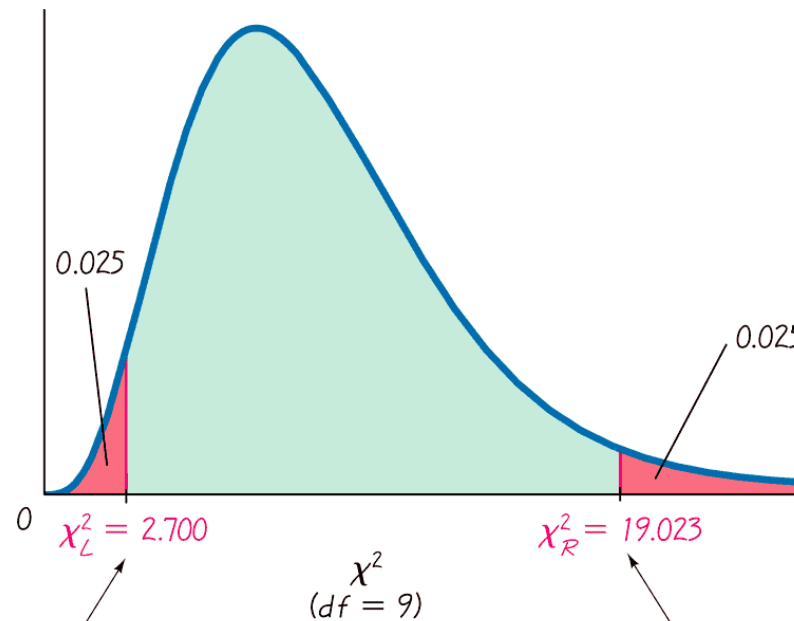
$$\alpha/2 = 0.025$$

$$1 - \alpha/2 = 0.975$$

Critical Values of the Chi-Square Distribution

Areas to the **right** of each tail

Figure 7-10



To obtain this critical value, locate 9 at the left column for degrees of freedom and then locate 0.975 across the top. The total area to the right of this critical value is 0.975, which we get by subtracting 0.025 from 1.

To obtain this critical value, locate 9 at the left column for degrees of freedom and then locate 0.025 across the top.

Estimators of σ^2

The sample variance s^2 is the best point estimate of the population variance σ^2 .

Confidence Interval (or Interval Estimate) for the Population Variance σ^2

Right-tail CV \nearrow

$$\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$$

Confidence Interval (or Interval Estimate) for the Population Variance σ^2 - cont

$$\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$$

Right-tail CV \nearrow χ^2_R \nwarrow χ^2_L Left-tail CV

Confidence Interval (or Interval Estimate) for the Population Variance σ^2 - cont

$$\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$$

Right-tail CV \nearrow χ^2_R \nwarrow χ^2_L Left-tail CV

Confidence Interval for the Population Standard Deviation σ

$$\sqrt{\frac{(n-1)s^2}{\chi^2_R}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_L}}$$

Procedure for Constructing a Confidence Interval for σ or σ^2

1. Verify that the required assumptions are satisfied.
2. Using $n - 1$ degrees of freedom, refer to Table A-4 and find the critical values χ^2_R and χ^2_L that correspond to the desired confidence level.
3. Evaluate the upper and lower confidence interval limits using this format of the confidence interval:

$$\frac{(n - 1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n - 1)s^2}{\chi^2_L}$$

Procedure for Constructing a Confidence Interval for σ or σ^2 - cont

4. If a confidence interval estimate of σ is desired, take the square root of the upper and lower confidence interval limits and change σ^2 to σ .
5. Round the resulting confidence level limits. If using the original set of data to construct a confidence interval, round the confidence interval limits to one more decimal place than is used for the original set of data. If using the sample standard deviation or variance, round the confidence interval limits to the same number of decimal places.

Confidence Intervals for Comparing Data

As in previous sections, do not use the overlapping of confidence intervals as the basis for making final conclusions about the equality of variances or standard deviations.

Example:

A study found the body temperatures of 106 healthy adults. The sample mean was 98.2 degrees and the sample standard deviation was 0.62 degrees. Find the 95% confidence interval for σ .

$$n = 106$$

$$\bar{x} = 98.2^\circ$$

$$s = 0.62^\circ$$

$$\alpha = 0.05$$

$$\alpha/2 = 0.025$$

$$1 - \alpha/2 = 0.975$$

$$\chi^2_R = 129.561, \chi^2_L = 74.222$$

$$\frac{(106 - 1)(0.62)^2}{129.561} < \sigma^2 < \frac{(106 - 1)(0.62)^2}{74.222}$$

$$0.31 < \sigma^2 < 0.54$$

$$0.56 < \sigma < 0.74$$

We are 95% confident that the limits of 0.56°F and 0.74°F contain the true value of σ . We are 95% confident that the standard deviation of body temperatures of all healthy people is between 0.56°F and 0.74°F.

Determining Sample Sizes

Table 7-2

Sample Size for σ^2		Sample Size for σ	
To be 95% confident that s^2 is within	of the value of σ^2 , the sample size n should be at least	To be 95% confident that s is within	of the value of σ , the sample size n should be at least
1%	77,207	1%	19,204
5%	3,148	5%	767
10%	805	10%	191
20%	210	20%	47
30%	97	30%	20
40%	56	40%	11
50%	37	50%	7
To be 99% confident that s^2 is within	of the value of σ^2 , the sample size n should be at least	To be 99% confident that s is within	of the value of σ , the sample size n should be at least
1%	133,448	1%	33,218
5%	5,457	5%	1,335
10%	1,401	10%	335
20%	368	20%	84
30%	171	30%	37
40%	100	40%	21
50%	67	50%	13

Example:

We want to estimate σ , the standard deviation off all body temperatures. We want to be 95% confident that our estimate is within 10% of the true value of σ . How large should the sample be? Assume that the population is normally distributed.

From Table 7-2, we can see that 95% confidence and an error of 10% for σ correspond to a sample of size 191.

Recap

In this section we have discussed:

- ❖ **The chi-square distribution.**
- ❖ **Using Table A-4.**
- ❖ **Confidence intervals for the population variance and standard deviation.**
- ❖ **Determining sample sizes.**