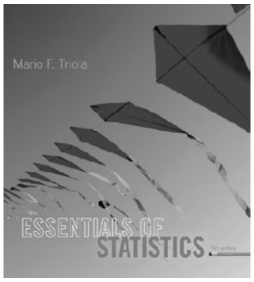
Lecture Slides



Essentials of Statistics
5th Edition

and the Triola Statistics Series

by Mario F. Triola

 Section 7.1-1

Chapter 7
Estimates and Sample Sizes

7-1 Review and Preview

7-2 Estimating a Population Proportion

7-3 Estimating a Population Mean

7-4 Estimating a Population Standard Deviation or Variance

 Section 7.1-2

Review

- In Chapters 2 & 3, we used *descriptive statistics* when we summarized data using tools such as graphs and statistics such as the mean and standard deviation.
- Chapter 6 we introduced critical values:

  z_α denotes the z score with an area of α to its right.
  If $\alpha = 0.025$, the critical value is $z_{0.025} = 1.96$.

  That is, the critical value $z_{0.025} = 1.96$ has an area of 0.025 to its right.

 Section 7.1-3

11

Preview

This chapter presents the beginning of inferential statistics.

- The two major activities of inferential statistics are (1) to use sample data to estimate values of a population parameters, and (2) to test hypotheses or claims made about population parameters.
- We introduce methods for estimating values of these important population parameters: proportions, means, and standard deviation / variances.
- We also present methods for determining sample sizes necessary to estimate those parameters.

 Section 7.1-4

Chapter 7
Estimates and Sample Sizes

7-1 Review and Preview

7-2 Estimating a Population Proportion

7-3 Estimating a Population Mean

7-4 Estimating a Population Standard Deviation or Variance

 Section 7.2-5

Key Concept

In this section we present methods for using a sample proportion to estimate the value of a population proportion.

- The sample proportion is the best point estimate of the population proportion.
- We can use a sample proportion to construct a confidence interval to estimate the true value of a population proportion, and we should know how to interpret such confidence intervals.
- We should know how to find the sample size necessary to estimate a population proportion.

 Section 7.2-6

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

The Pew Research Center conducted a survey of 1007 adults and found that 85% of them know what Twitter is.

The best point estimate of *p*, the population proportion, is the sample proportion:

$$\hat{p} = 0.85$$

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 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%.

($\alpha = 0.10$), ($\alpha = 0.05$), ($\alpha = 0.01$)

Interpreting a Confidence Interval

We must be careful to interpret confidence intervals correctly. There is a correct interpretation and many different and creative incorrect interpretations of the confidence interval $0.828 < p < 0.872$.

"We are 95% confident that the interval from 0.828 to 0.872 actually does contain the true value of the population proportion *p*."

This means that if we were to select many different samples of size 1007 and construct the corresponding confidence intervals, 95% of them would actually contain the value of the population proportion *p*.

(Note that in this correct interpretation, the level of 95% refers to the success rate of the process being used to estimate the proportion.)

Caution

Know the correct interpretation of a confidence interval.

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 proportions*.

Using Confidence Intervals for Hypothesis Tests

A confidence interval can be used to test some claim made about a population proportion *p*.

For now, we do not yet use a formal method of hypothesis testing, so we simply generate a confidence interval and make an informal judgment based on the result.

Critical Values

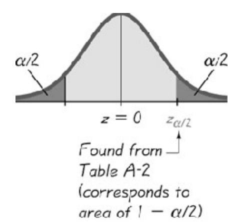
A standard *z* score can be used to distinguish between sample statistics that are likely to occur and those that are unlikely to occur. Such a *z* score is called a critical value. Critical values are based on the following observations:

1. Under certain conditions, the sampling distribution of sample proportions can be approximated by a normal distribution.

2. A *z* score associated with a sample proportion has a probability of $\alpha/2$ of falling in the right tail.

Critical Values

3. The *z* score separating the right-tail region is commonly denoted by $z_{\alpha/2}$ and is referred to as a critical value because it is on the borderline separating *z* scores from sample proportions that are likely to occur from those that are unlikely to occur.

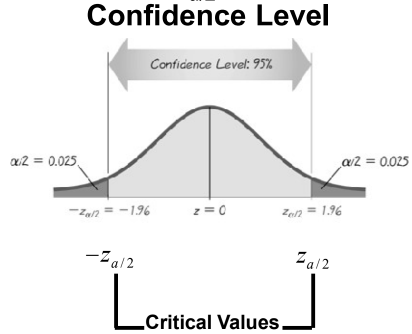


Definition

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.

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



Common Critical Values

Confidence Level	α	Critical Value, $z_{\alpha/2}$
90%	0.10	1.645
95%	0.05	1.96
99%	0.01	2.575

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 difference (with probability 1 – α, such as 0.95) between the observed proportion $\hat{p}$ and the true value of the population proportion *p*.

Margin of Error for Proportions

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:

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

Confidence Interval for Estimating a Population Proportion *p*

p = population proportion

$\hat{p}$ = sample proportion

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.

Confidence Interval for Estimating a Population Proportion *p*

1. The sample is a simple random sample.
2. The conditions for the binomial distribution are satisfied: there is a fixed number of trials, the trials are independent, there are two categories of outcomes, and the probabilities remain constant for each trial.
3. There are at least 5 successes and 5 failures.

Confidence Interval for Estimating a Population Proportion *p*

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

where

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

Confidence Interval for Estimating a Population Proportion *p*

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

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

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

Procedure for Constructing a Confidence Interval for *p* - cont

4. 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:

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

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

Example

In the Chapter Problem we noted that a Pew Research Center poll of 1007 randomly selected adults showed that 85% of respondents know what Twitter is. The sample results are $n = 1007$ and $\hat{p} = 0.70$.

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 more than 75% of adults know what Twitter is?
d. Assuming that you are a newspaper reporter, write a brief statement that accurately describes the results and includes all of the relevant information.

Example - Continued

Requirement check: simple random sample; fixed number of trials, 1007; trials are independent; two outcomes per trial; probability remains constant. Note: number of successes and failures are both at least 5.

a) Use the formula to find the margin of error.

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.96\sqrt{\frac{(0.85)(0.15)}{1007}}$$

$$E = 0.0220545$$

Example - Continued

b) The 95% confidence interval:

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

$$0.85 - 0.0220545 < p < 0.85 + 0.0220545$$

$$0.828 < p < 0.872$$

Example - Continued

c) Based on the confidence interval obtained in part (b), it does appear that more than 75% of adults know what Twitter is.

Because the limits of 0.828 and 0.872 are likely to contain the true population proportion, it appears that the population proportion is a value greater than 0.75.

Example - Continued

d) Here is one statement that summarizes the results:

85% of U.S. adults know what Twitter is. That percentage is based on a Pew Research Center poll of 1007 randomly selected adults.

In theory, in 95% of such polls, the percentage should differ by no more than 2.2 percentage points in either direction from the percentage that would be found by interviewing all adults in the United States.

Analyzing Polls

When analyzing polls consider:

1. The sample should be a simple random sample, not an inappropriate sample (such as a voluntary response sample).
2. The confidence level should be provided. (It is often 95%, but media reports often neglect to identify it.)
3. The sample size should be provided. (It is usually provided by the media, but not always.)
4. Except for relatively rare cases, the quality of the poll results depends on the sampling method and the size of the sample, but the size of the population is usually not a factor.

Caution

Never follow the common misconception that poll results are unreliable if the sample size is a small percentage of the population size.

The population size is usually not a factor in determining the reliability of a poll.

Sample Size

Suppose we want to collect sample data in order to estimate some population proportion.

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:

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

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

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

Round-Off Rule for Determining Sample Size

If the computed sample size *n* is not a whole number, round the value of *n* up to the next larger whole number.

Example

Many companies are interested in knowing the percentage of adults who buy clothing online.

How many adults must be surveyed in order to be 95% confident that the sample percentage is in error by no more than three percentage points?

a. Use a recent result from the Census Bureau: 66% of adults buy clothing online.

b. Assume that we have no prior information suggesting a possible value of the proportion.

Example - Continued

a) Use $\hat{p} = 0.66$ and $\hat{q} = 1 - \hat{p} = 0.34$

$\alpha = 0.05$ so $z_{\alpha/2} = 1.96$

$E = 0.03$

$$n = \frac{(z_{\alpha/2})^2\,\hat{p}\hat{q}}{E^2} = \frac{(1.96)^2(0.66)(0.34)}{(0.03)^2} = 957.839 = 958$$

To be 95% confident that our sample percentage is within three percentage points of the true percentage for all adults, we should obtain a simple random sample of 958 adults.

Example - Continued

b) Use $\alpha = 0.05$ so $z_{\alpha/2} = 1.96$

$E = 0.03$

$$n = \frac{(z_{\alpha/2})^2 \cdot 0.25}{E^2} = \frac{(1.96)^2 \cdot 0.25}{(0.03)^2} = 1067.1111 = 1068$$

To be 95% confident that our sample percentage is within three percentage points of the true percentage for all adults, we should obtain a simple random sample of 1068 adults.

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:

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

Chapter 7 Estimates and Sample Sizes

Key Concept

This section presents methods for using the sample mean to make an inference about the value of the corresponding population mean.

Key Concept

1. We should know that the sample mean $\bar{x}$ is the best point estimate of the population mean μ.
2. We should learn how to use sample data to construct a confidence interval for estimating the value of a population mean, and we should know how to interpret such confidence intervals.
3. We should develop the ability to determine the sample size necessary to estimate a population mean.

Requirements

The procedure we use has a requirement that the population is normally distributed or the sample size is greater than 30.

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

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

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

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

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.

The degree of freedom is often abbreviated df.

degrees of freedom = *n* – 1
for the methods in this section

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 $E = t_{\alpha/2} \frac{s}{\sqrt{n}}$ **df = *n* – 1**

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

Procedure for Constructing a Confidence Interval for *μ* (With *σ* Not Known)

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 $E = t_{\alpha/2} \cdot s / \sqrt{n}$.
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 - Continued

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

95% implies $\alpha = 0.05$.
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:

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

Example - Continued

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 μ, the mean of the changes in LDL cholesterol for the population.

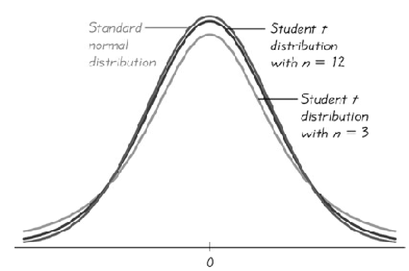
Because the confidence interval limits contain the value of 0, it is very possible that the mean of the changes in LDL cholesterol is equal to 0, suggesting that the garlic treatment did not affect the LDL cholesterol levels.

It does not appear that the garlic treatment is effective in lowering LDL cholesterol.

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 $\sigma = 1$).
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



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:

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

Finding a Sample Size for Estimating a Population Mean

μ = population mean
σ = population standard deviation
$\bar{x}$ = sample mean
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

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

Round-Off Rule for Sample Size *n*

If the computed sample size *n* is not a whole number, round the value of *n* up 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:

$$\sigma \approx range / 4$$

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 earlier study.

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?

α = 0.05
$\alpha/2$ = 0.025
$z_{\alpha/2}$ = 1.96
E = 3
σ = 15

$$n = \left[\frac{1.96 \bullet 15}{3}\right]^2 = 96.04 = 97$$

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*.

Part 2: Key Concept

This section presents methods for estimating a population mean. In addition to knowing the values of the sample data or statistics, we must also know the value of the population standard deviation, *σ*.

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

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

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

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.

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

$$\overline{x} - E < \mu < \overline{x} + E \quad \text{where} \quad E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

or $\quad \overline{x} \pm E$

or $\quad (\overline{x} - E, \overline{x} + E)$

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 a random sample, we obtain these sample statistics for the simple random sample: $n = 40$ and $\overline{x} = 172.55$ lb.

Research from several other sources suggests that the population of weights of men has a standard deviation given by $\sigma = 26$ lb.

Example - Continued

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)?

Example - Continued

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 implies that $\alpha = 0.05$, so $z_{\alpha/2} = 1.96$.

Calculate the margin of error.

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{26}{\sqrt{40}} = 8.0574835$$

Construct the confidence interval.

$$\overline{x} - E < \mu < \overline{x} - E$$
$$172.55 - 8.0574835 < \mu < 172.55 + 8.0574835$$
$$164.49 < \mu < 180.61$$

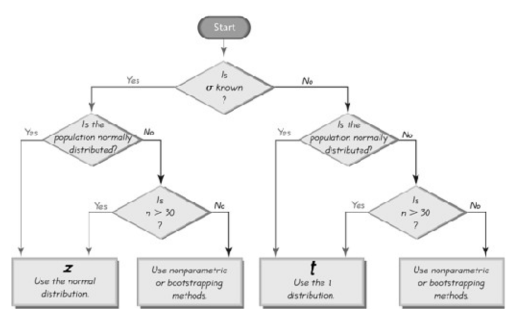
Example - Continued

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.

Choosing the Appropriate Distribution



Choosing the Appropriate Distribution

Use the normal (z) distribution	σ known and normally distributed population or $n > 30$
Use t distribution	σ not known and normally distributed population or $n > 30$
Use a nonparametric method or bootstrapping	Population is not normally distributed and $n \le 30$

Chapter 7
Estimates and Sample Sizes

7-1 Review and Preview

7-2 Estimating a Population Proportion

7-3 Estimating a Population Mean

7-4 Estimating a Population Standard Deviation or Variance

Key Concept

This section we introduce the chi-square probability distribution so that we can construct confidence interval estimates of a population standard deviation or variance.

We also present a method for determining the sample size required to estimate a population standard deviation or variance.

Chi-Square Distribution

In a normally distributed population with variance σ^2 , assume that we randomly select independent samples of size n and, for each sample, compute the sample variance s^2 (which is the square of the sample standard deviation s).

The sample statistic χ^2 (pronounced chi-square) has a sampling distribution called the chi-square distribution.

Chi-Square Distribution

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

where

s = sample size

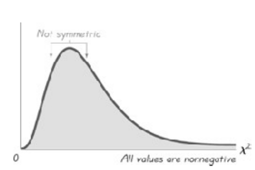
s^2 = sample variance

σ^2 = population variance

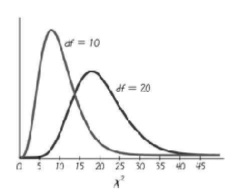
df = n – 1 degrees of freedom

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.



Chi-Square Distribution



Chi-Square Distribution for df = 10 and df = 20

Properties of the Distribution of the Chi-Square Statistic

2. The values of chi-square can be zero or positive, but it cannot be negative.
3. The chi-square distribution is different for each number of degrees of freedom, which is df = $n - 1$. As the number of degrees of freedom 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.

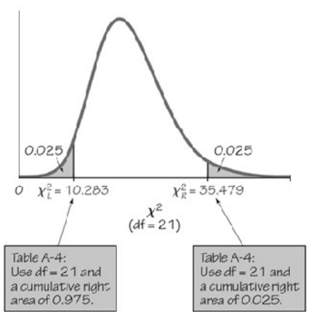
Example

A simple random sample of 22 IQ scores is obtained. Construction of a confidence interval for the population standard deviation σ requires the left and right critical values of χ^2 corresponding to a confidence level of 95% and a sample size of n = 22.

Find the critical χ^2 values corresponding to a 95% level of confidence.

Example - Continued

Critical Values of the Chi-Square Distribution



Estimators of σ^2

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

Estimator of σ

The sample standard deviation s is a commonly used point estimate of σ (even though it is a biased estimate).

Confidence Interval for Estimating a Population Standard Deviation or Variance

σ	=	population standard deviation
σ^2	=	population variance
s	=	sample standard deviation
s^2	=	sample variance
n	=	number of sample values
E	=	margin of error
χ^2_L	=	left-tailed critical value of χ^2
χ^2_R	=	right-tailed critical value of χ^2

Confidence Interval for Estimating a Population Standard Deviation or Variance

Requirements:

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

Confidence Interval for Estimating a Population Standard Deviation or Variance

Confidence Interval for the Population Variance σ^2

$$\frac{(n-1)s^2}{\chi_R^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_L^2}$$

Confidence Interval for Estimating a Population Standard Deviation or Variance

Confidence Interval for the Population Standard Deviation σ

$$\sqrt{\frac{(n-1)s^2}{\chi_R^2}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi_L^2}}$$

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 or use technology to find the critical values χ_R^2 and χ_L^2 that correspond to the desired confidence level.
3. Evaluate the upper and lower confidence interval limits using this format of the confidence interval:

$$\frac{(n-1)s^2}{\chi_R^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_L^2}$$

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

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.

Caution

Confidence intervals can be used informally to compare the variation in different data sets, but the overlapping of confidence intervals should not be used for making formal and final conclusions about equality of variances or standard deviations.

Example

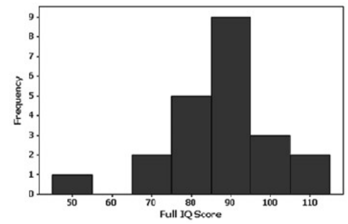
A group of 22 subjects took an IQ test during part of a study. The subjects had a standard deviation IQ score of 14.3.

Construct a 95% confidence interval estimate of σ, the standard deviation of the population from which the sample was obtained.

Example - Continued

We must first verify the requirements are met.

1. We can treat the sample as a simple random sample.
2. The following display shows a histogram of the data and the normality assumption can be reasonably met.



Example - Continued

n = 22 so df = 22 – 1 = 21

Use table A-4 to find:

$$\chi_L^2 = 10.283 \text{ and } \chi_R^2 = 35.479$$

$$\frac{(n-1)s^2}{\chi_R^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_L^2}$$

$$\frac{(22-1)(14.3)^2}{35.479} < \sigma^2 < \frac{(22-1)(14.3)^2}{10.283}$$

Example - Continued

Evaluation of the preceding expression yields:

$$121.0 < \sigma^2 < 417.6$$

Finding the square root of each part (before rounding), then rounding to two decimal places, yields this 95% confidence interval estimate of the population standard deviation:

$$11.0 < \sigma < 20.4$$

Based on this result, we have 95% confidence that the limits of 11.0 and 20.4 contain the true value of σ.

Determining Sample Sizes

The procedures for finding the sample size necessary to estimate σ^2 are much more complex than the procedures given earlier for means and proportions. Instead of using very complicated procedures, we will use Table 7-2.

STATDISK also provides sample sizes. With STATDISK, select Analysis, Sample Size Determination, and then Estimate St Dev.

Minitab, Excel, and the TI-83/84 Plus calculator do not provide such sample sizes.

Determining Sample Sizes

σ		σ^2	
To be 95% confident that s is within	of the value of σ, the sample size n should be at least	To be 95% confident that s^2 is within	of the value of σ^2, the sample size n should be at least
1%	19,205	1%	77,208
5%	768	5%	3,149
10%	192	10%	806
20%	48	20%	211
30%	21	30%	98
40%	12	40%	57
50%	8	50%	38
To be 99% confident that s is within	of the value of σ, the sample size n should be at least	To be 99% confident that s^2 is within	of the value of σ^2 the sample size n should be at least
1%	33,218	1%	133,449
5%	1,336	5%	5,458
10%	336	10%	1,402
20%	85	20%	369
30%	38	30%	172
40%	22	40%	101
50%	14	50%	68

Example

We want to estimate the standard deviation σ of all voltage levels in a home. We want to be 95% confident that our estimate is within 20% 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 20% for σ correspond to a sample of size 48.

We should obtain a simple random sample of 48 voltage levels form the population of voltage levels.