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: $\sigma$ Known
7-4 Estimating a Population Mean: $\sigma$ Not Known
7-5 Estimating a Population Variance
Section 7-2
Estimating 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.

3. There are at least 5 successes and 5 failures.
Notation for Proportions

\[ p = \text{population proportion} \]

\[ \hat{p} = \frac{x}{n} \]

(pronounced ‘p-hat’)

\[ \hat{q} = 1 - \hat{p} = \text{sample proportion of failures in a sample size of } n \]
Definition

A **point estimate** is a single value (or point) used to approximate a population parameter.
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 C.I.
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 $p$. 
Critical Values

1. 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 $\alpha$) 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 probability of $\alpha$ that a sample proportion will fall in either of the two red tails.
4. By the rule of complements 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 $Z_{\alpha/2}$

Figure 7-2

Found from Table A-2 (corresponds to area of $1 - \alpha/2$)
Finding $z_{\alpha/2}$ for a 95% Confidence Level

- $z_{\alpha/2} = -1.96$
- $z = 0$
- $z_{\alpha/2} = 1.96$

$\alpha = 5\%$
$\alpha/2 = 2.5\% = .025$

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

$\alpha = 0.05$

$\alpha/2 = 0.025$

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

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

Copyright © 2007 Pearson Education, Inc Publishing as Pearson Addison-Wesley.
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 difference (with probability $1 - \alpha$) 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

\[ \hat{p} - E < p < \hat{p} + E \]

\[ \hat{p} \pm E \]

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

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

$$0.476 < p < 0.544$$

The 99% confidence interval $0.465 < p < 0.555$
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)

\[ n = \left( \frac{Z_{\alpha/2}}{2} \right)^2 \frac{\hat{p} \hat{q}}{E^2} \]
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*).

\[
n = \frac{[z_{a/2}]^2 \hat{p} \hat{q}}{E^2}
\]

\[
= \frac{[1.96]^2 (0.169)(0.831)}{0.04^2}
\]

\[
= 337.194
\]

\[
= 338 \text{ households}
\]

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 $\hat{p}$.

\[ n = \frac{[z_{a/2}]^2 \cdot 0.25}{E^2} \]

\[ = \frac{(1.96)^2 (0.25)}{0.04^2} \]

\[ = 600.25 \]

\[ = 601 \text{ households} \]

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:

$$E = \frac{(\text{upper confidence limit}) - (\text{lower confidence limit})}{2}$$
Section 7-3
Estimating a Population Mean: $\sigma$ Known
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 $\sigma$ 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 $\mu$. 
Sample Mean

1. For all populations, the sample mean $\bar{x}$ is an unbiased estimator of the population mean $\mu$, meaning that the distribution of sample means tends to center about the value of the population mean $\mu$.

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 standard deviation was 0.62 degrees. Find the point estimate of the population mean $\mu$ of all body temperatures.

Because the sample mean $\bar{x}$ is the best point estimate of the population mean $\mu$, we conclude that the best point estimate of the population mean $\mu$ 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 $\mu$, and is denoted by $E$. 
Margin of Error

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

Margin of error for mean (based on known \( \sigma \))
Confidence Interval estimate of the Population Mean $\mu$ (with $\sigma$ 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 \( \mu \) (with Known \( \sigma \))

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

\[
\begin{align*}
    n &= 106 \\
    \bar{x} &= 98.20^\circ \\
    s &= 0.62^\circ \\
    \alpha &= 0.05 \\
    \alpha/2 &= 0.025 \\
    z_{\alpha/2} &= 1.96
\end{align*}
\]

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

\[
\bar{x} - E < \mu < \bar{x} + E
\]

\[
98.08^\circ < \mu < 98.32^\circ
\]

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 \( \mu \).
Sample Size for Estimating Mean $\mu$

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

Where

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

$E = \text{desired margin of error}$

$\sigma = \text{population standard deviation}$

Formula 7-5
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 $\sigma$ 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 $\sigma$.

3. Estimate the value of $\sigma$ 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.

\[
\begin{align*}
\alpha &= 0.05 \\
\alpha/2 &= 0.025 \\
z_{\alpha/2} &= 1.96 \\
E &= 2 \\
\sigma &= 15 \\
n &= \left[ \frac{1.96 \cdot 15}{2} \right]^2 = 216.09 = 217
\end{align*}
\]

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 $\mu$. 
Section 7-4

Estimating a Population Mean: $\sigma$ Not Known
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 $\sigma$ unknown, we will use the Student $t$ distribution assuming that certain requirements are satisfied.
Requirements with $\sigma$ 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
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 a 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 $\mu$
(With $\sigma$ Not Known)

Formula 7-6

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

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

Table A-3 lists values for $t_{\alpha/2}$
Confidence Interval for the Estimate of $\mu$ (With $\sigma$ Not Known)

\[ x - E < \mu < x + E \]

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

$t_{\alpha/2}$ found in Table A-3
Procedure for Constructing a Confidence Interval for $\mu$ (With $\sigma$ 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 $E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$.

4. Find the values of $\bar{x} - E$ and $\bar{x} + E$. Substitute those values in the general format for the confidence interval:

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

5. Round the resulting confidence interval limits.
Example: A 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 $\mu$.

$$n = 106$$
$$\bar{x} = 98.20^\circ$$
$$s = 0.62^\circ$$

$$\alpha = 0.05$$
$$\alpha/2 = 0.025$$
$$t_{\alpha/2} = 1.984$$

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

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

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

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

Start

Is \( \sigma \) known?

Yes

Is the population normally distributed?

Yes

Is \( n > 30 \) ?

Yes

Use the normal distribution.

Use nonparametric or bootstrapping methods.

Use the \( t \) distribution.

Use nonparametric or bootstrapping methods.

No

No

No

No

Copyright © 2007 Pearson Education, Inc Publishing as Pearson Addison-Wesley.
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 \( \mu \), 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: “Ease of reading” scores for 12 different pages randomly selected from *Harry Potter*. Find the 95% interval estimate of $\mu$, (The 12 pages’ distribution appears to be bell-shaped with $x = 80.75$ and $s = 4.68$.)

\[
\bar{x} = 80.75 \quad E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 2.201 \cdot \frac{4.68}{\sqrt{12}} = 2.97355
\]

\[
s = 4.68
\]

\[
\alpha = 0.05 \quad \alpha/2 = 0.025
\]

\[
t_{\alpha/2} = 2.201
\]

\[
\bar{x} - E < \mu < \bar{x} + E
\]

\[
80.75 - 2.97355 < \mu < 80.75 + 2.97355
\]

\[
77.77645 < \mu < 83.72355
\]

\[
77.78 < \mu < 83.72
\]

We are 95% confident that this interval contains the mean “ease of reading” score for all pages.
Finding the Point Estimate and \( E \) from a Confidence Interval

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

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

Copyright © 2007 Pearson Education, Inc Publishing as Pearson Addison-Wesley.
Section 7-5
Estimating a Population Variance
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 $\sigma$ or $\sigma^2$. 
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

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

*Formula 7-7*

where

\( n \) = sample size

\( s^2 \) = sample variance

\( \sigma^2 \) = 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 $df = 20$
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 $\chi^2$ 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 $\chi^2$ 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.

\[ \chi^2_L = 2.700 \]

\[ \chi^2_R = 19.023 \]

\( df = 9 \)
Estimators of $\sigma^2$

The sample variance $s^2$ is the best point estimate of the population variance $\sigma^2$. 
Confidence Interval (or Interval Estimate) for the Population Variance $\sigma^2$

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

Right-tail CV
Confidence Interval (or Interval Estimate) for the Population Variance $\sigma^2$ - cont

$$\frac{(n-1)s^2}{\chi^2_{R}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{L}}$$

Right-tail CV $\chi^2_{R}$

Left-tail CV $\chi^2_{L}$
Confidence Interval (or Interval Estimate) for the Population Variance $\sigma^2$ - cont

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

Right-tail CV

Left-tail CV

Confidence Interval for the Population Standard Deviation $\sigma$

$$\sqrt{\frac{(n-1)s^2}{\chi^2_R}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_L}}$$
Procedure for Constructing a Confidence Interval for $\sigma$ or $\sigma^2$

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 $\chi^2_R$ and $\chi^2_L$ that correspond to the desired confidence level.

3. Evaluate the upper and lower confidence interval limits using this format of the confidence interval:

$$\frac{(n - 1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n - 1)s^2}{\chi^2_L}$$
4. If a confidence interval estimate of $\sigma$ is desired, take the square root of the upper and lower confidence interval limits and change $\sigma^2$ to $\sigma$.

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 decimals places.
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 $\sigma$.

\[
\begin{align*}
\bar{x} &= 98.2^\circ \\
s &= 0.62^\circ \\
\alpha &= 0.05 \\
\alpha / 2 &= 0.025 \\
1 - \alpha / 2 &= 0.975
\end{align*}
\]

\[
\chi^2_R = 129.561, \quad \chi^2_L = 74.222
\]

\[
(106 - 1)(0.62)^2 < \sigma^2 < (106 - 1)(0.62)^2
\]

\[
\frac{129.561}{74.222}
\]

\[
0.31 < \sigma^2 < 0.54 \quad \quad \quad 0.56 < \sigma < 0.74
\]

We are 95% confident that the limits of 0.56°F and 0.74°F contain the true value of $\sigma$. 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

<table>
<thead>
<tr>
<th>Sample Size for $\sigma^2$</th>
<th>Sample Size for $\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>To be 95% confident that $s^2$ is within</td>
<td>of the value of $\sigma^2$, the sample size $n$ should be at least</td>
</tr>
<tr>
<td>1%</td>
<td>77,207</td>
</tr>
<tr>
<td>5%</td>
<td>3,148</td>
</tr>
<tr>
<td>10%</td>
<td>805</td>
</tr>
<tr>
<td>20%</td>
<td>210</td>
</tr>
<tr>
<td>30%</td>
<td>97</td>
</tr>
<tr>
<td>40%</td>
<td>56</td>
</tr>
<tr>
<td>50%</td>
<td>37</td>
</tr>
<tr>
<td>To be 99% confident that $s^2$ is within</td>
<td>of the value of $\sigma^2$, the sample size $n$ should be at least</td>
</tr>
<tr>
<td>1%</td>
<td>133,448</td>
</tr>
<tr>
<td>5%</td>
<td>5,457</td>
</tr>
<tr>
<td>10%</td>
<td>1,401</td>
</tr>
<tr>
<td>20%</td>
<td>368</td>
</tr>
<tr>
<td>30%</td>
<td>171</td>
</tr>
<tr>
<td>40%</td>
<td>100</td>
</tr>
<tr>
<td>50%</td>
<td>67</td>
</tr>
</tbody>
</table>
Example:

We want to estimate $\sigma$, the standard deviation of all body temperatures. We want to be 95% confident that our estimate is within 10% of the true value of $\sigma$. 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 $\sigma$ correspond to a sample of size 191.