
Chapter 9: Estimating the Value of a Parameter Using Confidence Intervals

- [9.1 Estimating a Population Proportion](#)
- [9.2 Estimating a Population Mean](#)
- [9.3 Confidence Intervals for a Population Standard Deviation](#)
- [9.4 Putting It Together: Which Procedure Do I Use?](#)

In [Section 8.2](#), saw the following excerpt from a report by the Pew Research Foundation:

Hispanics See Their Situation in U.S. Deteriorating

Half (50%) of all Latinos say that the situation of Latinos in this country is worse now than it was a year ago, according to a new nationwide survey of 2,015 Hispanic adults conducted by the Pew Hispanic Center. (Source: [Pew Research](#))

Further down that particular article, there's another interesting line:

The margin of error for the full sample is plus or minus 2.8 percentage points; for registered voters, the margin of error is 4.4 percentage points.

How did they determine that the margin was $\pm 2.8\%$?

This is the question we'll answer in Chapter 9. In fact, we'll talk about confidence intervals for the population mean, population proportion, and population variance.

:: start ::



Section 9.1: Estimating a Population Proportion

9.1 Estimating a Population Proportion

9.2 Estimating a Population Mean

9.3 Confidence Intervals for a Population Standard Deviation

9.4 Putting It Together: Which Procedure Do I Use?

Objectives

By the end of this lesson, you will be able to...

1. construct and interpret a CI for p
2. determine the sample size necessary for estimating p within a specified margin of error

Perhaps the most common confidence intervals that we see in the news are regarding proportions. Consider the following examples:

Hispanics See Their Situation in U.S. Deteriorating

Half (50%) of all Latinos say that the situation of Latinos in this country is worse now than it was a year ago, according to a new nationwide survey of 2,015 Hispanic adults conducted by the Pew Hispanic Center. [...] *The margin of error of the survey is plus or minus 2.8 percentage points at the 95% confidence level.* (Source: [Pew Research](#))

International Poll: No Consensus On Who Was Behind 9/11

A new WorldPublicOpinion.org poll of 17 nations finds that majorities in only nine of them believe that al Qaeda was behind the 9/11 terrorist attacks on the United States. [...] On average, 46 percent say that al Qaeda was behind the attacks while 15 percent say the US government, seven percent Israel, and seven percent some other perpetrator. One in four say they do not know. [...] *Margins of error range from +/-3 to 4 percent.* (Source: [WorldPublicOpinion](#))

Stem cell, marijuana proposals lead in Mich. poll

A recent poll shows voter support leading opposition for ballot proposals to loosen Michigan's restrictions on embryonic stem cell research and allow medical use of marijuana. The EPIC-MRA poll conducted for The Detroit News and television stations WXYZ, WILX, WOOD and WJRT found 50 percent of likely Michigan voters support the stem cell proposal, 32 percent against and 18 percent undecided. The telephone poll of 602 likely Michigan voters was conducted Sept. 22 through Wednesday. *It has a margin of sampling error of plus or minus 4 percentage points.* (Source: [Associated Press](#))

In [Section 8.2](#), saw the following excerpt from a report by the Pew Research Foundation:

Hispanics See Their Situation in U.S. Deteriorating

Half (50%) of all Latinos say that the situation of Latinos in this country is worse now than it was a year ago, according to a new nationwide survey of 2,015 Hispanic adults conducted by the Pew Hispanic Center. (Source: [Pew Research](#))

Further down that particular article, there's another interesting line:

The margin of error for the full sample is plus or minus 2.8 percentage points; for registered voters, the margin of error is 4.4 percentage points.

How did they determine that the margin was $\pm 2.8\%$?

The Logic of Confidence Intervals

The whole point of collecting information from a sample is to gain some information about the population. For

example, when the report says that half of all Latinos say that the situation is worse now than it was a year ago, it's not saying that they actually asked *every single Latino* living in the United States. Rather, it's based on a *sample*.

In a similar manner, consider one of the results from the [American Time Use Survey](#):

Employed persons worked an average of 7.6 hours on the days that they worked. They worked longer on weekdays than on weekend days - 7.9 versus 5.6 hours.

The news release isn't saying that the average of time spent for *all* employed persons is 7.6 hours per day - they're referring to those in the [sample of 12,250 individuals](#) in the study.

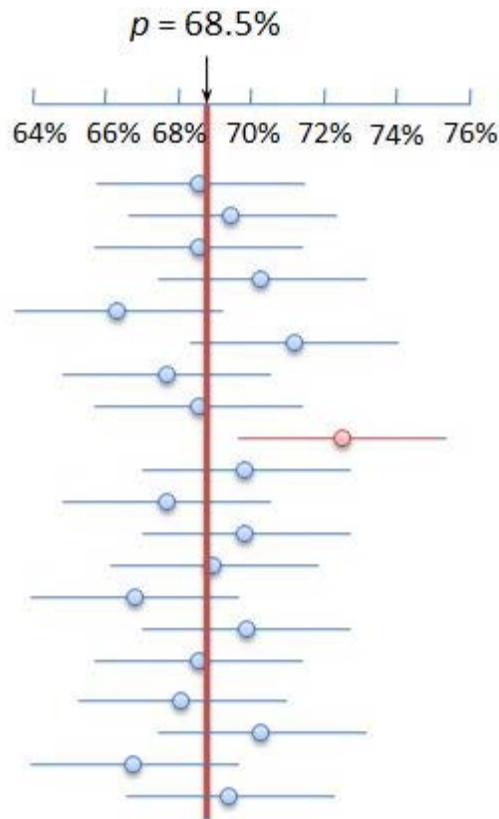
Both of these examples are called **point estimates**. 50%, for example, is the point estimate for the percentage of *all* Latinos who feel that way. Similarly, the average number of hours worked per day of 7.6 is a point estimate the average number of hours worked per day for *all* employed persons.

A **confidence interval estimate** is an interval of numbers, along with a measure of the likelihood that the interval contains the unknown parameter.

The **level of confidence** is the expected proportion of intervals that will contain the parameter if a large number of samples is maintained. The notation we use is $(1 - \alpha)100\%$ for the confidence interval. (This will make more sense a bit later!)

The idea of a confidence interval is this: Suppose we're wondering what proportion of ECC students are part-time. We might take a sample of 100 individuals and find a sample proportion of 65%. If we say that we're 95% confident that the real proportion is somewhere between 61% and 69%, we're saying that if we were to repeat this with new samples, and gave a margin of $\pm 4\%$ every time, our interval would contain the actual proportion 95% of the time.

See, every time we get a new sample, we have a new estimate, and hence a new confidence interval. Sometimes we'll be right. Sometimes we won't. The idea behind a certain % (like 95%) is that we're saying we'll be right 95% of the time. Here's a visual for 20 theoretical samples and corresponding confidence intervals for the proportion who work:



You can see that 1 of the 20 (5%) confidence intervals doesn't contain the actual value of 68.5% (based on data from elgin.edu). But the other 19 (or 95% of them) *do* contain the real value.

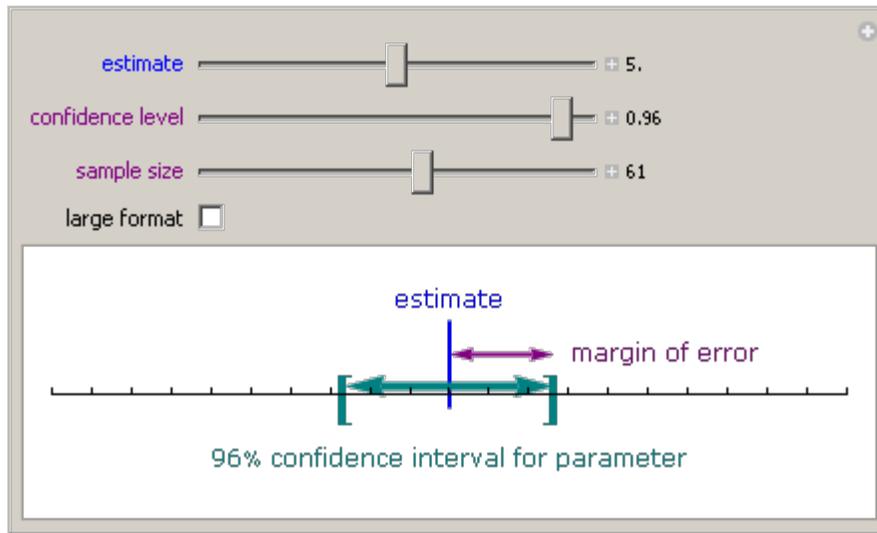
This is the idea for what we'll be doing. We'll be giving an interval of value for where we think the mean, proportion, or standard deviation is, and then a confidence level for what proportion of the intervals we believe will contain the true parameter.

Exploring Confidence Intervals about the Population Mean

To do some exploring yourself, go to the Demonstrations Project from Wolfram Research, and download the [Confidence Intervals demonstration](#). If you haven't already, download and install the player by clicking on the image to the right.



Once you have the player installed and the Confidence Intervals demonstration downloaded, move the sliders for the *estimate*, *confidence level*, and *sample size* on the confidence interval and margin of error.



Even though this is about a confidence interval for the mean (and not the proportion), the idea is the same. Hopefully you noticed a few keys:

- The *estimate* simply slides the interval left and right.
- A larger *confidence level* means a wider interval - to be more confident, we need to cast a wider net in order to "catch" the actual population mean.
- A larger *sample size* means a narrower interval - the more we have in our sample, the closer we are to having the actual population mean.

Constructing Confidence Intervals

Before we can start constructing confidence intervals, we need to review some of the theoretical framework we set up in [Chapter 8](#). In particular, the information about the distribution of \hat{p} .

Reviewing the Distribution of the Sample Proportion

In [Section 8.2](#), we introduced the idea of a proportion, along with its distribution.

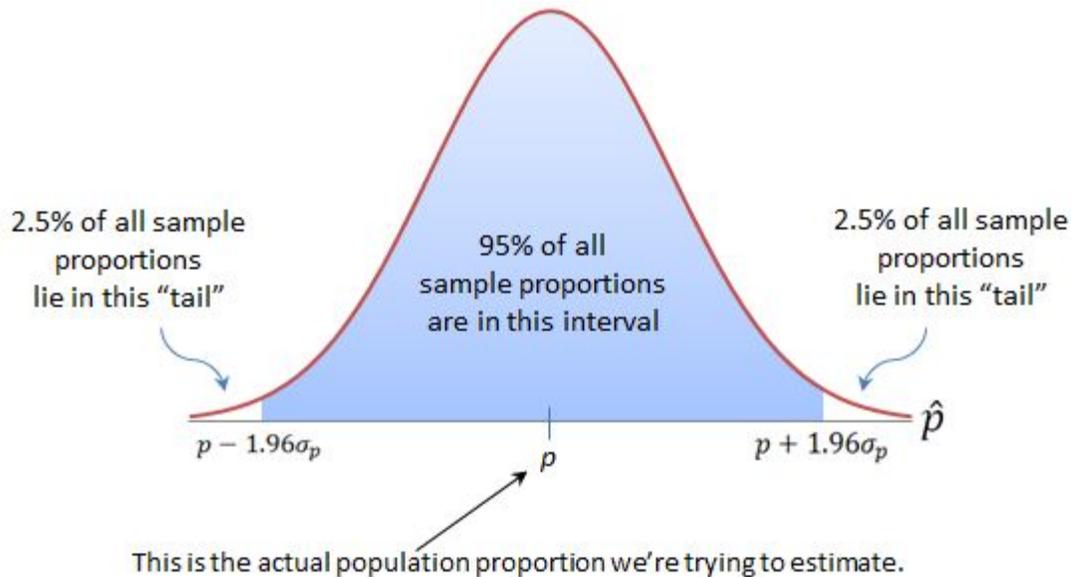
Sampling Distribution of \hat{p}

For a simple random sample of size n such that $n \leq 0.05N$ (in other words, the sample is less than 5% of

the population), and $np(1-p) \geq 10$, \hat{p} is approximately normally distributed, with

$$\mu_{\hat{p}} = p \text{ and } \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

So if $np(1-p) \geq 10$, \hat{p} will be approximately normally distributed, with the mean and standard deviation above. Using the properties of the normal distribution, that means about 95% of all sample proportions will be within 1.96 standard deviations of the mean (p).



In other words, 95% of all sample proportions will be in the interval:

$$p - 1.96\sigma_{\hat{p}} < \hat{p} < p + 1.96\sigma_{\hat{p}}$$

With a little algebraic manipulation, we get the following:

$$\hat{p} - 1.96\sigma_{\hat{p}} < p < \hat{p} + 1.96\sigma_{\hat{p}}$$

And since $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$, we can further get the 95% confidence interval for the proportion as:

$$\hat{p} - 1.96 \sqrt{\frac{p(1-p)}{n}} < p < \hat{p} + 1.96 \sqrt{\frac{p(1-p)}{n}}$$

This is a problem, though. How can we say that the population proportion, p , is in this interval, when the lower and upper bounds **contain** p ?! All is not lost! In general, we can say that as long as $n \leq 0.05N$ and $n\hat{p}(1-\hat{p}) \geq 10$, we can use \hat{p} in place of p in the standard error of the sample proportion, which gives us this for a 95% confidence interval:

$$\hat{p} - 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Or rephrased,

$$\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

So in general, this means that 95% of the time, an interval of that form would contain the population proportion - that's exactly what we want!

Finding a 95% Confidence Interval

Example 1

Polls are very common examples of confidence intervals, so let's look at a controversial topic in Illinois - concealed carry. Illinois was the last state to allow concealed carry in public places, and the issues has been brought up frequently through polling over the years. Two 2011 polls are interesting to consider.

First, a [poll sponsored by the Illinois Council Against Handgun Violence](#) conducted March 23-27, 2011 found that 56% of likely voters statewide oppose concealed-carry legislation. **Find a 95% confidence interval for the proportion of likely voters who oppose this type of legislation.**

Interesting, [another poll conducted at the same time by the Illinois State Rifle Association](#) found that 47% of voters in four state senate districts support concealed carry legislation. **Find a 95% confidence interval for the proportion of likely voters who oppose this type of legislation.**

Solution:

Before we can do any analysis, we need to consider if we are *able* to find a confidence interval. To do that, we need the sample sizes for these polls. From the links, we can see that 600 likely voters were included in the first sample, with 957 in the second. In both cases, $n\hat{p}(1-\hat{p}) \geq 10$ and $n \leq 0.05N$, so we can perform the confidence intervals.

For the ICAHV poll, $\hat{p} = 56\%$, so a 95% confidence interval is:

$$0.56 \pm 1.96 \sqrt{\frac{0.56(1-0.56)}{600}}$$
$$0.56 \pm 0.04$$

So we can be 95% confident that the proportion of likely Illinois voters who oppose concealed carry legislation is between 52% and 60%.

For the ISRA poll, $\hat{p} = 47\%$, so a 95% confidence interval is:

$$0.47 \pm 1.96 \sqrt{\frac{0.47(1-0.47)}{957}}$$
$$0.47 \pm 0.03$$

So we can be 95% confident that the proportion of voters in these districts who *support* concealed carry legislation is between 44% and 50%.

Wait... what is going on? Something is clearly wrong here. If these polls were similar, then they should have opposite results, not relatively similar for the *opposite* side of the issue!

So what happened? Well, the first issue is the samples - they're not drawn from the same population. Of greater concern is the questioning in the second survey. Look at the link and the question wording in general. What concerns would you have?

Constructing Confidence Intervals about a Population Proportion

What if we want to be more confident? Well, we can just replace the 1.96 with a different Z corresponding to a different area in the "tails". With that, we have the following result:

A $(1-\alpha)$ 100% confidence interval for p is

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Note: We must have $n\hat{p}(1-\hat{p}) \geq 10$ and $n \leq 0.05N$ in order to construct this interval.

The Margin of Error

Most of the time (but not always), confidence intervals look roughly like:

$$\text{point estimate} \pm \text{margin of error}$$

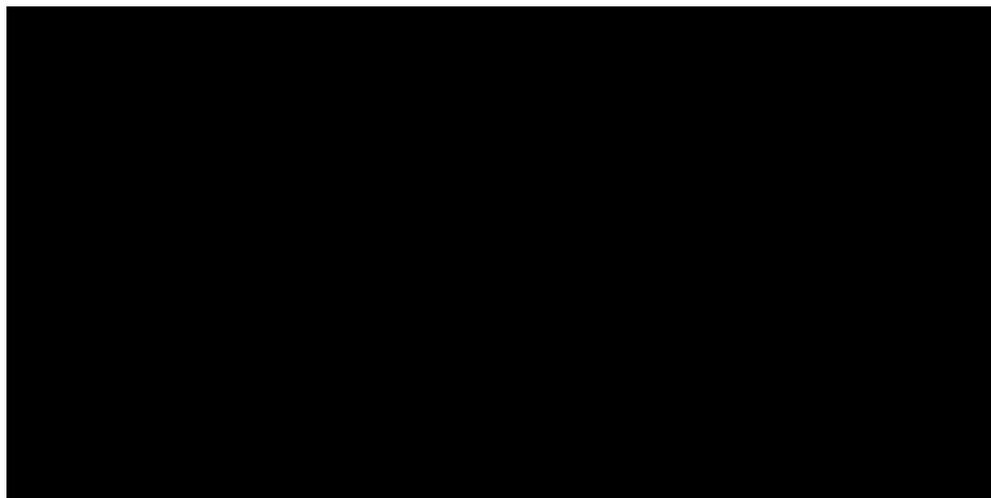
So in the case of a confidence interval for the population proportion shown above, the margin of error is the portion after the \pm , or..

The **margin of error**, E , in a $(1-\alpha)$ 100% confidence interval for p is

$$E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where n is the sample size.

For more on the margin of error, watch this YouTube video, from David Longstreet:



Let's look at one of the polls above and see how they found the margin of error for their confidence interval.

Example 2

Consider the excerpt shown below from a poll conducted by Pew Research:

Stem cell, marijuana proposals lead in Mich. poll

*A recent poll shows voter support leading opposition for ballot proposals to loosen Michigan's restrictions on embryonic stem cell research and allow medical use of marijuana. The EPIC-MRA poll conducted for The Detroit News and television stations WXYZ, WILX, WOOD and WJRT found 50 percent of likely Michigan voters support the stem cell proposal, 32 percent against and 18 percent undecided. The telephone poll of 602 likely Michigan voters was conducted Sept. 22 through Wednesday. **It has a margin of sampling error of plus or minus 4 percentage points.** (Source: [Associated Press](#))*

Using the confidence interval formula above, let's see if we can get the $\pm 4\%$ they got.

Since we want a 95% confidence level, $\alpha = 0.05$. The margin of error is then:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96 \cdot \sqrt{\frac{0.5(1-0.5)}{602}} \approx 0.040$$

And so the margin of error is $\pm 4.0\%$.

Why don't you try one:

Example 3

Read the following excerpt and calculate the margin of error for the indicated sample proportion.

Tories lead but voter volatility on the rise

OTTAWA—The Conservatives still hold a strong lead but shifting allegiances in Quebec and a sharp upsurge in ABC (Anybody But Conservative) thinking nationally could put a Tory majority victory out of reach, a new poll shows.

[...]

*The Toronto Star/Angus Reid survey, conducted after the televised leaders' debates, shows **support for Stephen Harper's Conservative party remains unchanged at 40 per cent**, while the Liberals are slightly closing the gap, with their backing among decided voters rising to 25 per cent, according to the survey.*

The poll on voting intentions, approval ratings and strategic voting surveyed 1,176 adult Canadians on Oct. 2-3 and has a margin of error of plus-or-minus [xx] percentage points, 19 times out of 20.

[reveal answer]

Finding Confidence Intervals Using StatCrunch

1. Go to **Stat > Proportions > One sample > with summary**.
2. Enter the number of successes and the number of observations, and press **Next**.
3. Click the Confidence Interval button and set the Confidence Level. Then press **Calculate**.

The results should appear.

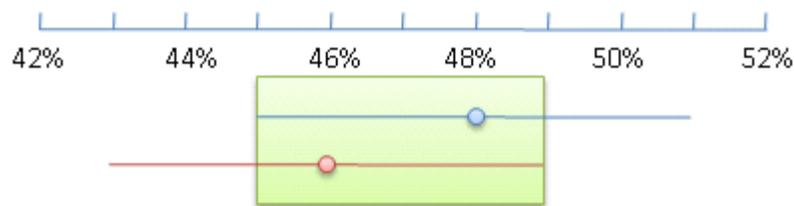
An interesting consequence of margins of errors in polls is the concept of a "statistical tie". Check out the following excerpt from a Wall Street Journal Article:

Financial Crisis Has Little Sway in Presidential Poll

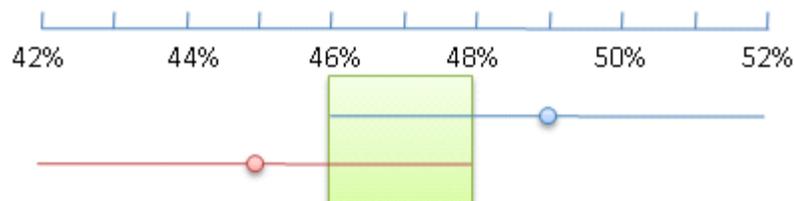
WASHINGTON -- The race between Barack Obama and John McCain remains very tight, despite financial turmoil that has turned the nation's attention to economic issues that tend to favor the Democrats, according to a new Wall Street Journal/NBC News poll.

[...] Overall, *the race remains a statistical tie*, with 48% favoring Sen. Obama and his running mate, Sen. Joe Biden, and 46% favoring Sen. McCain and his vice-presidential choice, Alaska Gov. Sarah Palin. In the latest Journal poll, two weeks ago, Sen. Obama had a one-point edge. *The new poll had a margin of error of plus or minus three percentage points.* (Source: [Wall Street Journal](#))

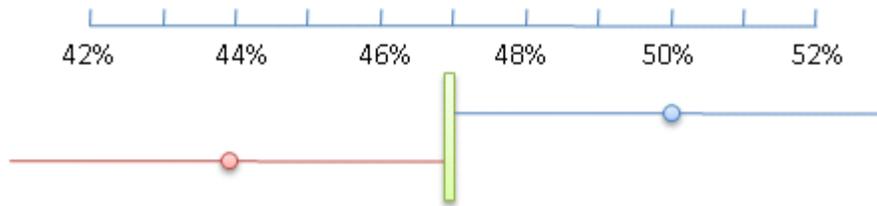
In this case, with a margin of error of $\pm 3\%$, the confidence intervals actually overlap.



In fact, they'd overlap even if they were 49% and 45%.



It wouldn't be until 50% and 44% that we could say Obama would be statistically ahead of McCain.



Determining the Sample Size Needed

We sometimes need to know the sample size necessary to get a desired margin of error. The way we answer these types of questions is to go back to the margin of error:

The **margin of error**, E , in a $(1-\alpha)100\%$ confidence interval p is

$$E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where n is the sample size.

If we're given the margin of error, we can solve for the sample size and get the following result:

The **sample size required** to obtain a $(1-\alpha)100\%$ confidence interval for p with a margin of error E is:

$$n = \hat{p}(1-\hat{p}) \left(\frac{z_{\alpha/2}}{E} \right)^2$$

where n is rounded up to the next integer and \hat{p} is a prior estimate of p . If no prior estimate is available, use $\hat{p} = 0.5$.

Let's try one.

Example 4

Suppose you want to know what proportion of Elgin, IL residents are registered to vote. You'd like your results to be accurate to within 2 percentage points with 95% confidence. What sample size is necessary..

- if you use results from the US Census stating that about 72.1% of all citizens were registered for the 2004 election? (Source: [US Census](#))
- you don't use a prior estimate?

[\[reveal answer \]](#)

<< [previous section](#) | [next section](#) >>

1 2 3 4 5 6 7 8 **9** 10 11 12 13



This work is licensed under a Creative Commons License.

Section 9.2: Estimating a Population Mean

- 9.1 Estimating a Population Proportion
- 9.2 Estimating a Population Mean**
- 9.3 Confidence Intervals for a Population Standard Deviation
- 9.4 Putting It Together: Which Procedure Do I Use?

Objectives

By the end of this lesson, you will be able to...

1. state properties of Student's t-distribution
2. determine t-values
3. construct and interpret a confidence interval for a population mean
4. find the sample size needed to estimate the population mean within a given margin of error

Similar to confidence intervals about proportions from Section 9.1, we can also find confidence intervals about population means. Using the same general set-up, we should have something like this:

$$\text{point estimate} \pm \text{margin of error}$$

What's our point estimate for the population mean? Why, the *sample mean*, of course! The margin of error will be similar as well:

$$\bar{x} \pm z_{\alpha/2} \cdot \sigma_{\bar{x}}$$

If you recall, we discussed the distribution of \bar{x} in [Chapter 8](#):

The Central Limit Theorem

Regardless of the distribution shape of the population, the sampling distribution of \bar{x} becomes approximately normal as the sample size n increases (conservatively $n \geq 30$).

The Sampling Distribution of \bar{x}

If a simple random sample of size n is drawn from a large population with mean μ and standard deviation σ , the sampling distribution of \bar{x} will have mean and standard deviation:

$$\mu_{\bar{x}} = \mu \text{ and } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

where $\sigma_{\bar{x}}$ is the **standard error of the mean**.

So substituting into the above formula, we get:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

There's another problem here, similar to the previous section - if we're looking for a confidence interval for the population mean, μ , how would we know σ ? So like before, we can introduce an estimate there - in this case, the sample standard deviation, s . This introduces a lot of variability, though, since s can differ quite a bit from σ . As a result, we need to introduce a new distribution that's similar to the standard normal, but takes into consideration some of this variability. It's called the **Student's t-Distribution**.

Student's t-Distribution

The so-called Student's distribution has an interesting history. Here's a quick summary taken from Wikipedia:

The "student's" distribution was actually published in 1908 by W. S. Gosset. Gosset, however, was employed at a brewery that forbade the publication of research by its staff members. To circumvent this restriction, Gosset used the name "Student", and consequently the distribution was named "Student t-distribution".

Source: [Wikipedia](#)

Gosset was trying to do research dealing with small samples. He found that even when the standard deviation was not known, the distribution of the sample means was still symmetric and similar to the normal distribution. In fact, as the sample size increases, the distribution approaches the standard normal distribution.

Student's t-Distribution

Suppose a simple random sample size n is taken from a population. If the population follows a normal distribution, then

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

follows a Student's t-distribution with $n-1$ degrees of freedom.*

* The concept of *degrees of freedom* is pretty abstract. One way to think of it is like this: suppose five students are choosing from five different lollipops. Each of the first four students has a choice, but the last student does not, so there are $5-1=4$ degrees of freedom.

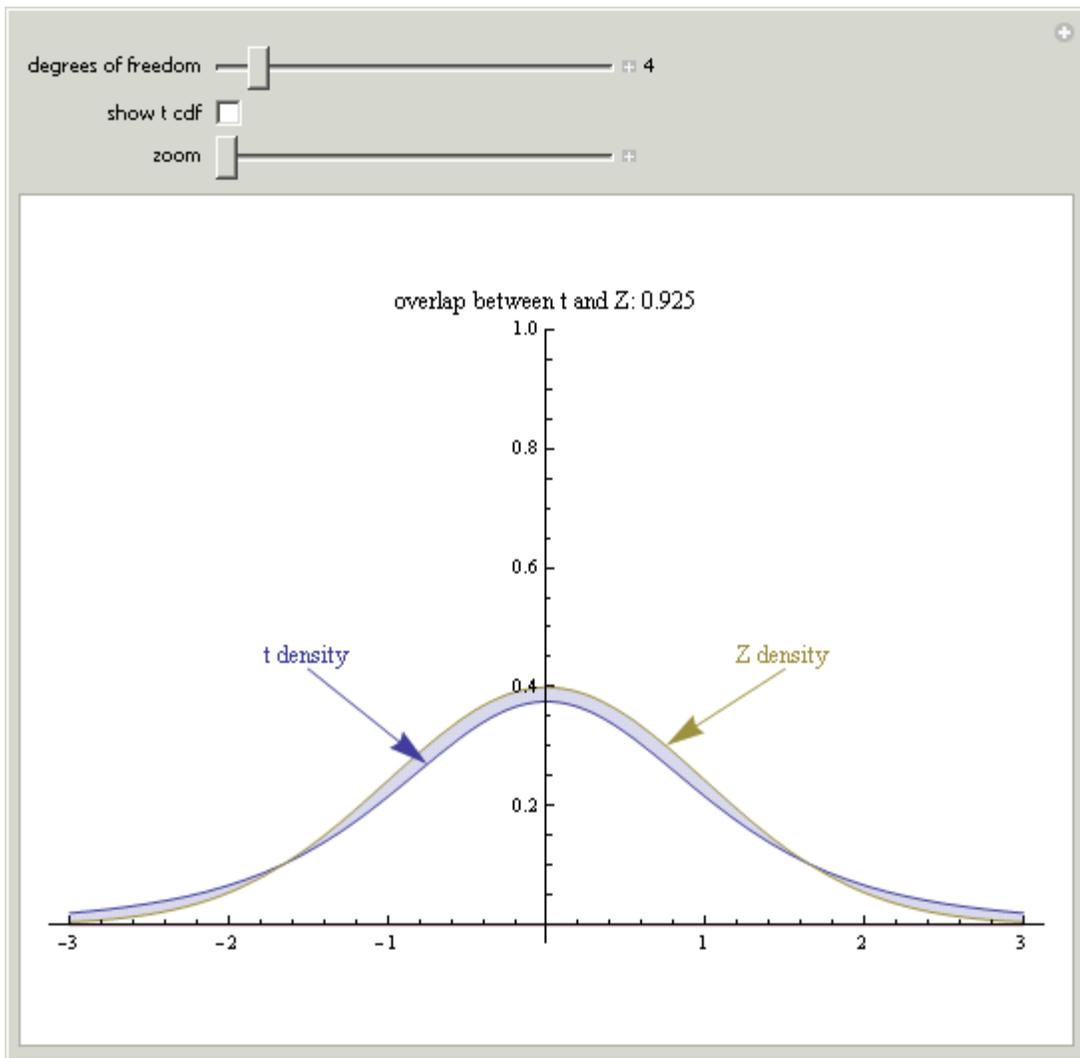
The key idea behind the t-distribution is that it has a similar shape to the standard normal distribution, but has more variability and is affected by n .

Exploring the t-Distribution

To do some exploring yourself, go to the Demonstrations Project from Wolfram Research, and download the [Student's t-Distribution demonstration](#). If you haven't already, download and install the player by clicking on the image to the right.



Once you have the player installed and the Student's t-Distribution demonstration downloaded, uncheck the **show t-cdf** box (see below) and move the slider for the **degrees of freedom** to see the relationship between the standard normal distribution and the t-distribution.



You should notice that as the degrees of freedom increase, the distributions become more and more similar.

Finding Critical Values

Find critical values in the t-distribution using a table will be a bit different from finding critical values for the standard normal distribution.

Before we start the section, you need a copy of the table. You can download a [printable copy](#) of this table, or use the table in the back of your textbook. It should look something like the image below (trimmed to make it more viewable).

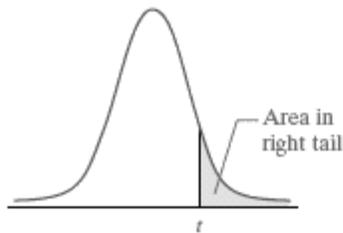


TABLE III
t-Distribution
Area In Right Tail

df	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005
1	1.000	1.376	1.963	3.078	6.314	12.706	15.894	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787

Notice that unlike with the [standard normal table](#), the t-table has probabilities along the top and critical values in the middle. This is because we primarily use the t-table to find t_{α} - the value with α area (probability) to the right.

Let's try an example.

Example 1

Find $t_{0.05}$ with 21 degrees of freedom. You can use the table above, or print one out yourself. Your textbook should also come with a copy you can use.

[\[reveal answer \]](#)

Finding Critical Values Using StatCrunch

Click on **Stat > Calculators > T**

Enter the degrees of freedom, the direction of the inequality, and the probability (leave X blank). Then press Compute. The image below

shows the t-value with an area of 0.05 to the right if there are 15 degrees of freedom.

DF: 15
Prob(X => 1.7530503) = 0.05
Close Compute

Example 2

Use the technology of your choice to find $t_{0.01}$ with 14 degrees of freedom

[\[reveal answer \]](#)

Constructing Confidence Intervals

Before we can start constructing confidence intervals, we need to review some of the theoretical framework we set up in [Chapter 8](#). In particular, the information about the distribution of \bar{x} .

The Central Limit Theorem

Regardless of the distribution shape of the population, the sampling distribution of \bar{x} becomes approximately normal as the sample size n increases (conservatively $n \geq 30$).

The Sampling Distribution of \bar{x}

If a simple random sample of size n is drawn from a large population with mean μ and standard deviation σ , the sampling distribution of \bar{x} will have mean and standard deviation:

$$\mu_{\bar{x}} = \mu \text{ and } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

where $\sigma_{\bar{x}}$ is the **standard error of the mean**.

Constructing a $(1-\alpha)100\%$ Confidence Interval about μ

In general, a **$(1-\alpha)100\%$ confidence interval for μ when σ is unknown** is

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

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

Note: The sample size must be large ($n \geq 30$) with no outliers or the population must be normally distributed.

Example 3

Suppose we'd like to know how many hours per week online students at ECC work. If we take a sample of 20 students and find a mean of 16.3

hours with a standard deviation of 5.4 hours, construct a 95% confidence interval for the average of all online ECC students.

[[reveal answer](#)]

Finding Confidence Intervals Using StatCrunch

1. Enter the data.
2. Go to **Stat > t-Statistics > One Sample**.
3. Select **with data** if you have the data, or **with summary** if you only have the summary statistics.
4. If you chose **with data**, click on the variable that you want for the confidence interval. Otherwise, enter the sample statistics.
5. Click on **Next**.
6. Click on the Confidence Interval Radio Button and enter your confidence level if it is not 95%.
7. Click on **Calculate**.

The Confidence interval should appear.

Example 4

In [Example 1](#) in Section 7.4, we looked the resting heart rates of 25 Statistics students.

heart rate				
61	63	64	65	65
67	71	72	73	74
75	77	79	80	81
82	83	83	84	85
86	86	89	95	95

([Click here](#) to view the data in a format more easily copied.)

Use the data to construct a 90% confidence interval for the true average resting heart rate of the students in this class.

Be sure to check that the conditions for creating confidence intervals are met.

[[reveal answer](#)]

Determining the Sample Size Needed

Suppose instead of trying to calculate a confidence interval given a sample mean and sample size, you are targeting a specific accuracy.

For example, say you'd like to know the average IQ of ECC students within 3 points. What sample size you would need?

The way we answer these types of questions is to go back to the margin of error definition:

The **margin of error**, E , in a $(1-\alpha)100\%$ confidence interval for μ is

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

where n is the sample size.

If we're given the margin of error, we can solve for the sample size and get $n = \left(\frac{t_{\alpha/2} \cdot s}{E}\right)^2$.

Uh-oh... another problem. How can we figure out the t -value, when we need the degrees of freedom... which depends on the sample size?! The key here is that for larger sample sizes, the t -distribution approaches the standard normal distribution (think about it - as n increases, the sample standard deviation gets closer to the population standard deviation). So what we can do instead is use z to approximate t .

The **sample size required** to estimate μ with a $(1-\alpha)100\%$ level of confidence and a margin of error, E , is:

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

where n is rounded up to the nearest whole number.

Example 5

Let's again refer to the IQs of ECC students. How many students would we need to sample if we wanted a 95% confidence interval for the average IQ of ECC students to be within 3 points of the true population mean? (Recall that $\sigma = 15$ for IQs.)

[\[reveal answer \]](#)

[<< previous section](#) | [next section >>](#)

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#)



This work is licensed under a Creative Commons License.

Section 9.3: Confidence Intervals for a Population Standard Deviation

- 9.1 Estimating a Population Proportion
- 9.2 Estimating a Population Mean
- 9.3 Confidence Intervals for a Population Standard Deviation**
- 9.4 Putting It Together: Which Procedure Do I Use?

Objectives

By the end of this lesson, you will be able to...

1. find critical values for the χ^2 distribution
2. construct and interpret CIs about σ^2 and σ

The last parameters we need to find confidence intervals for are the population variance (σ^2) and standard deviation (σ).

There are many instances where we might be interested in knowing something about the spread of a population based on a sample. For example, we might believe that a particular group of students seems to have a wider variation in their grades than those from the past. Or a part manufacturer may be concerned that one of the parts it's manufacturing is too inconsistent, even though the mean may be at specifications.

Before we can develop a confidence interval for the variance, we need another distribution.

The Chi-Square (χ^2) distribution

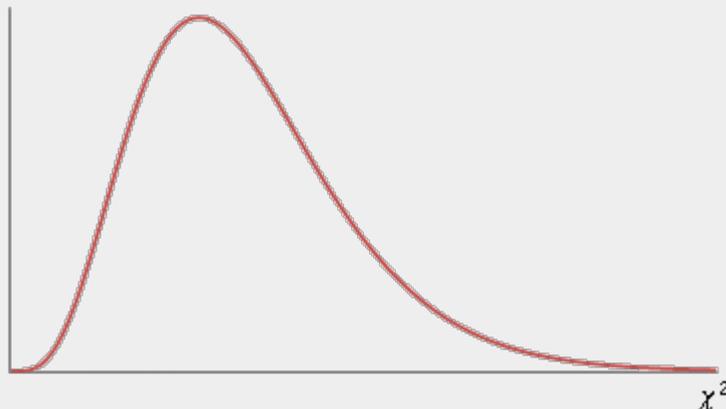
Note: "chi-square" is pronounced "kai" as in sky, not "chai" like [the tea](#).

The Chi-Square (χ^2) distribution

If a simple random sample size n is obtained from a normally distributed population with mean μ and standard deviation σ , then

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

has a **chi-square distribution** with $n-1$ degrees of freedom.



Properties of the χ^2 distribution

1. It is *not* symmetric.
2. The shape depends on the degrees of freedom.
3. As the number of degrees of freedom increases, the distribution becomes more symmetric.
4. $\chi^2 \geq 0$

Finding Critical Values

Find critical values in the χ^2 distribution using a table is done in the same manner we found critical values for the t-distribution.

Before we start the section, you need a copy of the table. You can download a [printable copy](#) of this table, or use the table in the back of your textbook. It should look something like the image below (trimmed to make it more viewable).

Degrees of Freedom	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892

Notice that the table is similar to the [t-table](#), with probabilities along the top and critical values in the middle. This is because we primarily use the chi-square-table to find critical values.

Let's try an example.

Example 1

Find the critical values in the χ^2 distribution which separate the middle 95% from the 2.5% in each tail, assuming there are 12 degrees of freedom.

You can use the table above, or print one out yourself. Your textbook should also come with a copy you can use.

[\[reveal answer \]](#)

Finding Critical Values Using StatCrunch

Click on **Stat > Calculators > Chi-Square**

Enter the degrees of freedom, the direction of the inequality, and the probability (leave X blank). Then press **Compute**.

Example 2

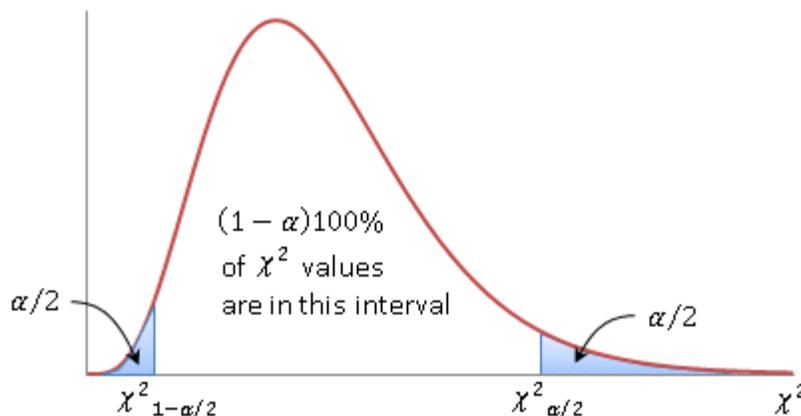
Use the technology of your choice to find $\chi^2_{0.01}$ with 20 degrees of freedom.

[\[reveal answer \]](#)

Constructing Confidence Intervals about σ^2 and σ

Now that we have the basics of the distribution of the variable χ^2 , we can work on constructing a formula for the confidence interval.

From the distribution shape on the previous page, we know that $(1 - \alpha)100\%$ of the χ^2 values will be between the two critical values shown below.



This gives us the following inequality:

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

If we solve the inequality for σ^2 , we get the formula for the confidence interval:

A **$(1 - \alpha)100\%$ confidence interval for σ^2** is

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

Note: The sample *must* be taken from a normally distributed population.

Note #2: If a confidence interval for σ is desired, we can take the square root of each part.

Now that we have the confidence interval formula, let's try a couple examples.

Example 3

Suppose a sample of 30 ECC students are given an IQ test. If the sample has a standard deviation of 12.23 points, find a 90% confidence interval for the population standard deviation.

Solution: We first need to find the critical values:

$$\chi^2_{1-\alpha/2} = \chi^2_{0.95,29} \approx 17.708 \text{ and } \chi^2_{\alpha/2} = \chi^2_{0.05,29} \approx 42.557$$

Then the confidence interval is:

$$\begin{aligned} \frac{(n-1)s^2}{\chi^2_{\alpha/2}} &< \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \\ \frac{(30-1)12.23^2}{42.557} &< \sigma^2 < \frac{(30-1)12.23^2}{17.708} \\ 101.9249 &< \sigma^2 < 244.9472 \\ 10.10 &< \sigma < 15.65 \end{aligned}$$

So we are 90% confident that the standard deviation of the IQ of ECC students is between 10.10 and 15.65 bpm.

Finding Confidence Intervals Using StatCrunch

1. Click on **Stat > Variance > One sample**
2. Select **with data** if you have the data, or **with summary** if you only have the summary statistics.
3. If you chose **with data**, click on the variable that you want for the confidence interval. Otherwise, enter the sample statistics.
4. Click on **Next**.
5. Click the Confidence Interval radio button
6. Enter the desired level of confidence and press **Calculate**

The confidence interval should be displayed.

Note: If you need a confidence interval about the population standard deviation, take the square root of the values in the resulting confidence interval.

Here's one for you to try:

Example 4

In [Example 3](#) in Section 9.1, we assumed the standard deviation of the

resting heart rates of students was 10 bpm.

heart rate				
61	63	64	65	65
67	71	72	73	74
75	77	79	80	81
82	83	83	84	85
86	86	89	95	95

([Click here](#) to view the data in a format more easily copied.)

Use StatCrunch to find a 95% confidence interval for the standard deviation of the resting heart rates for students in this particular class.

[\[reveal answer \]](#)

[<< previous section](#) | [next section >>](#)

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#)



This work is licensed under a Creative Commons License.

Section 9.4: Putting It Together: Which Procedure Do I Use?

- 9.1 Estimating a Population Proportion
- 9.2 Estimating a Population Mean
- 9.3 Confidence Intervals for a Population Standard Deviation
- 9.4 Putting It Together: Which Procedure Do I Use?**

Objectives

By the end of this lesson, you will be able to...

1. determine the appropriate CI to construct

In this chapter, we've learned three different confidence intervals. In order for them to be of any value, we need to know which one to apply. Let's do a quick review of the three.

Confidence Intervals about p

This is probably the easiest one - whenever we're looking at a proportion (percent), this is the confidence interval we want.

A $(1-\alpha)$ 100% confidence interval for p is

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Note: We must have $n\hat{p}(1-\hat{p}) \geq 10$ and $n \leq 0.05N$ in order to construct this interval.

Confidence Intervals about μ

This is the typical confidence interval for a mean. Use this when you're given information from a sample or if you're only given data. (In that case, you calculate the sample mean and standard deviation yourself, so you can't possibly know the population standard deviation.)

In general, a $(1-\alpha)$ 100% confidence interval for μ when σ is unknown is

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

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

Note: The sample size must be large ($n \geq 30$) or the population must be normally distributed.

Confidence Intervals about σ^2 and σ

The last confidence interval is for either the variance or standard deviation. You should be able to key on those words to help you recognize this confidence interval.

A $(1-\alpha)$ 100% confidence interval for σ^2 is

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

Note: The sample *must* be taken from a normally distributed population.

Some Examples

Let's try some examples. For each situation, simply determine which confidence interval should be used. You should specify the parameter and state the appropriate interval formula (because there are two for the population mean).

Example 1

We would like to know the fraction of ECC students who commute to school from their parents' homes. We send emails to students using their ECC email account until 100 have responded; 62 of the responders were commuters.

Find a 95% confidence interval for the fraction of ECC students who commute to school from their parents' homes.

[\[reveal answer \]](#)

Example 2

Suppose a student measuring the boiling temperature of a certain liquid observes the readings (in °C) 102.5, 101.7, 103.1, 100.9, 100.5, and 102.2 on 6 different samples of the liquid. He calculates the sample mean to be 101.82. If he knows that the standard deviation for this procedure is 1.2°, what is the confidence interval for the population mean at a 95% confidence level?

[\[reveal answer \]](#)

Example 3

To assess scholastic performance, a state administers an achievement test to a simple random sample of 100 high school seniors. The mean score of the students who took the exam is 99.7 points, with a standard deviation of 7.9 points. Find an approximate 90% confidence interval for the average of the population scores that would have been obtained had every high school senior in the state been administered the achievement test.

[\[reveal answer \]](#)

Example 4

We know from previous examples that the standard deviation of IQs is normally distributed with a standard deviation of 15. Suppose we wonder if the IQs of ECC students have more variation. To answer this question, we collect the IQs from a random sample of ECC students and find a standard deviation of 16.2.

Based on this information, do you believe with 95% confidence that the IQs of ECC students have more variation than the population?

[[reveal answer](#)]

[<< previous section](#) | [next section >>](#)

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#)



This work is licensed under a Creative Commons License.