

Chapter 7

Key Ideas

Confidence Interval, Confidence Level

Point Estimate, Margin of Error, Critical Value, Standard Error

t distribution, Chi-Square distribution

Section 7-1: Overview

All of the material in Chapters 4-6 forms a foundation of what is called *inferential statistics*. We already dealt with inferential statistics in Chapter 10 in a regression setting. Now, we will explore estimation.

Here is an outline of the main idea of inferential statistics:

1. There is a population of interest (i.e. the group we want to know something about)
2. We draw a random sample of size n from the population.
3. We compute statistics from the sample,
4. We use these statistics to *estimate* similar parameters in the population.

For example, suppose we want to know what percentage of Denison students own a red car. To estimate this, we take a random sample of 100 students and find out what percentage of those 100 students own red cars. Then, we say that the sample percentage should close to the actual percentage of *all* Denison students who own red cars. In this example:

Population of Interest: All Denison Students

Sample Size: $n = 100$

Statistic from the Sample: The percentage of students sampled who own red cars

Parameter in the Population: The percentage of all Denison students who own red cars

How do we actually estimate the parameter, though?

General Estimation Framework

Suppose we want to estimate a parameter (e.g. population proportion, population average, etc.).

The first thing to notice is that it would be impossible to exactly pinpoint the value with 100% accuracy without sampling every single member of the population, since there would always be some uncertainty. As a result, the best we can do is make a guess at the true value, and then include a margin of error based on a certain level of confidence we have in our results. The estimate and the margin of error form something called a **confidence interval**.

A confidence interval is made of 2 different parts.

1. The **point estimate** is the sample statistic (this is our best guess at the true parameter value given our sample).
2. The **margin of error** is added and subtracted from the point estimate to make the interval.
It can also be subdivided into two parts:
 - a. A **critical value** from a distribution (more to come on this later)
 - b. The **standard error** of the point estimate (more to come on this as well)

The confidence interval (CI) has this form:

$$\begin{aligned}\text{CI} &= (\text{Point Estimate}) \pm (\text{Margin of Error}) \\ &= (\text{Point Estimate}) \pm (\text{Critical Value}) * (\text{Standard Error})\end{aligned}$$

Of course there is no guarantee that the true population parameter will be in this interval, so we have to make some sort of statement about the chances that this will be true.

The **confidence level** is the probability that the interval actually covers the true population parameter. Often, the confidence level is denoted $(1 - \alpha)$, where α is the chance that it does *not* cover the true parameter. For example, if $\alpha = 0.05$, then the confidence level is 0.95, or 95%. Thus we would say that we are 95% confident that the interval covers the parameter.

Interpreting Confidence Intervals

Let's consider the example from before, where we want to estimate the percentage of Denison students who own a red car. Suppose that of our sample of size 100, 15 students owned a car. This gives a point estimate of 15%, or 0.15 for the population parameter. Suppose also that we calculated a critical value of 1.645 and a standard error of 0.0357, with a confidence level of 95%. In this case, the confidence interval will be:

$$\begin{aligned}\text{CI} &= (\text{Point Estimate}) \pm (\text{Margin of Error}) \\ &= (\text{Point Estimate}) \pm (\text{Critical Value}) * (\text{Standard Error}) \\ &= 0.15 \pm 1.645 * 0.0357 \\ &= 0.15 \pm 0.0587\end{aligned}$$

This gives the interval (0.0913, 0.2087). To interpret this interval, any of the following statements are equivalent:

1. We are 95% confident that the true percentage of all Denison students who own red cars is between 9.13% and 20.87%.
2. If we repeatedly took different samples of size 100 and computed a CI for each of those samples, 95% of the computed intervals would cover the true percentage of all Denison students who own red cars.
3. There is a 95% chance that the interval (0.0913, 0.2087) covers the true percentage of all Denison students who own red cars.

A General Note

Depending on the parameter you want to estimate, formulas for the point estimate, critical value, and standard error will change. However, the format of a confidence interval is always the same.

Section 7-2: Estimating a Population Proportion

To estimate a population proportion, we will be using Normal Approximation (see Sections 6-5 and 6-6 for more details). Because of this, there are 3 conditions that must be satisfied for us to estimate the population proportion.

1. The sample must be a simple random sample.
2. The Binomial conditions are satisfied –
 - a. Fixed Number of Trials (this is the sample size n)
 - b. Independent Trials (this follows from the simple random sample requirement)
 - c. There are two types of outcomes – success and failure
 - d. The probability of success/failure is the same in each trial (this should be true if the population is large)
3. There are at least 5 successes and 5 failures (this guarantees some theoretical assumptions that we shouldn't get into)

Some Notation:

p = the population proportion (unknown quantity that we want to estimate)

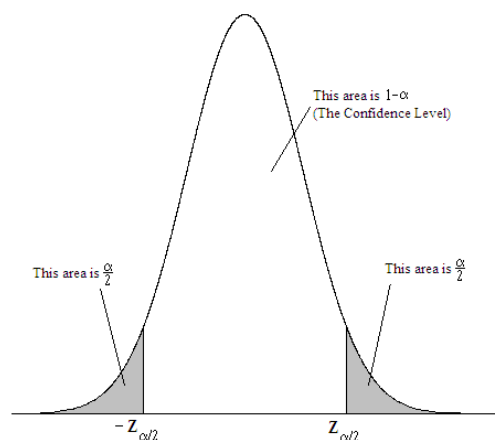
$\hat{p} = \frac{x}{n}$ = the sample proportion of x successes in a sample of size n

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

$Z_{\alpha/2}$ = the z-score with an area *above* of $\alpha/2$ (See figure on the right)

Common Values of $Z_{\alpha/2}$:

| Confidence Level | α | $Z_{\alpha/2}$ |
|------------------|----------|----------------|
| 90% | 0.10 | 1.645 |
| 95% | 0.05 | 1.96 |
| 99% | 0.01 | 2.575 |



Computing a Confidence Interval for The Population Proportion

For a sample size of n , confidence level $1 - \alpha$, and sample proportion \hat{p} , the parts of the confidence interval are:

Point Estimate: \hat{p}

Critical Value: $Z_{\alpha/2}$

Standard Error: $\sqrt{\frac{\hat{p}\hat{q}}{n}}$

So the confidence interval is:

CI = (Point Estimate) \pm (Margin of Error)

= (Point Estimate) \pm (Critical Value) \cdot (Standard Error)

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

Example

Engineers at a bottle-making factory are interested in estimating the percentage of defective bottles manufactured by the facility in general. To do this, they take a simple random sample of 100 bottles and find that 8 of them are defective. Find a 95% confidence interval for the percentage of defective bottles manufactured by the factory and interpret it.

Solution:

First, we note a few things:

- The 3 conditions above are met
- The sample proportion is $\hat{p} = \frac{x}{n} = \frac{8}{100} = 0.08$
- The confidence level is $95\% = 0.95 = 1 - \alpha$, so $\alpha = 0.05$
- The critical value is $Z_{\alpha/2} = Z_{0.05/2} = Z_{0.025} = 1.96$
- The standard error is $\sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{(0.08)(0.92)}{100}} = \sqrt{\frac{0.0736}{100}} = 0.0271$

Therefore, the confidence interval is:

$$\hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} \rightarrow 0.08 \pm (1.96)(0.0271) \rightarrow 0.08 \pm 0.0532 \rightarrow (0.0268, 0.1332)$$

Therefore, we can say that we are 95% confident that the percentage of defective bottles manufactured in this facility is between 2.68% and 13.32%.

Determining Sample Size

Clearly, the size of the sample has an impact on the confidence interval, since it affects the margin of error and the point estimate. The question is what sample size should be chosen at the start of the study to give a nice confidence interval. Often, what statisticians will do is decide they want a small margin of error, then find a sample size that gives them that margin of error.

Example

In the example above, for $\hat{p} = 0.08$ and $\alpha = 0.05$ (i.e. confidence level of 95%), find the sample size n required for the margin of error to be less than 0.02.

Solution:

We want the margin of error to be 0.02.

Also, we know the margin of error is $Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$, because this is the amount added and subtracted to the point estimate.

Therefore, figuring out n is as easy as solving the equation $Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.02$ for n !

$$Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.02 \Rightarrow 1.96 \cdot \sqrt{\frac{(0.08)(0.92)}{n}} = 0.02 \Rightarrow \sqrt{\frac{0.0736}{n}} = 0.0102 \Rightarrow \frac{0.0736}{n} = 0.00010404 \Rightarrow n = 707.42$$

Thus we would need a sample size of **at least 708** to get a margin of error that small.

Finding A General Formula for Sample Size

In the general case, we can solve for n as follows (here, E is the desired size of the margin of error):

$$Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} = E \Rightarrow \sqrt{\frac{\hat{p}\hat{q}}{n}} = \frac{E}{Z_{\alpha/2}} \Rightarrow \frac{\hat{p}\hat{q}}{n} = \frac{E^2}{Z_{\alpha/2}^2} \Rightarrow \frac{1}{n} = \frac{E^2}{\hat{p}\hat{q}Z_{\alpha/2}^2} \Rightarrow \boxed{n = \hat{p}\hat{q} \frac{Z_{\alpha/2}^2}{E^2}}$$

If the value of \hat{p} is known, the formula is as stated above.

If the value of \hat{p} is *unknown*, then use $\hat{p} = 0.50$.

Section 7-3: Estimating a Population Mean – σ Known

To estimate a population mean using the method in this section, 3 conditions must be met:

1. The sample is a simple random sample
2. The population standard deviation σ is known.
3. Either the population is normally distributed or $n > 30$ (so we can use the CLT)

As you may have already guessed, the best point estimate of the population mean is the sample mean, \bar{x} .

Also, from the CLT, we also know that \bar{x} is either exactly or approximately normally distributed (depending on which case is true in the 3rd condition) with a mean of μ and a standard deviation of $\frac{\sigma}{\sqrt{n}}$.

For this reason, we have the following.

Computing a Confidence Interval for The Population Mean (σ Known)

For a sample size of n , confidence level $1 - \alpha$, and sample mean \bar{x} , the parts of the confidence interval are:

Point Estimate: \bar{x}

Critical Value: $Z_{\alpha/2}$

Standard Error: $\frac{\sigma}{\sqrt{n}}$

So the confidence interval is:

CI = (Point Estimate) \pm (Margin of Error)

= (Point Estimate) \pm (Critical Value) \cdot (Standard Error)

$$= \bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Example

Suppose you want to estimate the average price of a home in Columbus, OH. From past information, you know that the standard deviation in housing prices is $\sigma = \$100,000$. You take a simple random sample of size $n = 50$ and compute a sample mean of $\bar{x} = \$246,000$. Find a 95% confidence interval for the population mean.

Solution:

First, we note a few things:

- The 3 conditions above are met (if you think about it, X is not normally distributed – it is skewed, but $n > 30$)
- The point estimate is $\bar{x} = 246,000$
- The confidence level is $95\% = 0.95 = 1 - \alpha$, so $\alpha = 0.05$
- The critical value is $Z_{\alpha/2} = Z_{0.05/2} = Z_{0.025} = 1.96$
- The standard error is $\frac{\sigma}{\sqrt{n}} = \frac{100,000}{\sqrt{50}} = 14142.14$

Therefore, the confidence interval is:

$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \Rightarrow 246,000 \pm (1.96)(14142.14) \Rightarrow 246,000 \pm 27,718.59 \Rightarrow (218,281.41, 273,718.59)$$

So with 95% confidence, we claim that the average price of a home is somewhere between \$218,281 and \$273,718, approximately.

Determining Sample Size

Now suppose that your supervisor asks you to choose a sample large enough to get a margin of error of \$10,000 or less. How do we determine the sample size?

As in the situation of estimating the population proportion, we just set the margin of error equal to 10,000:

$$Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 10,000 \Rightarrow 1.96 \frac{100,000}{\sqrt{n}} = 10,000 \Rightarrow 1.96 \frac{100,000}{10,000} = \sqrt{n} \Rightarrow \sqrt{n} = 19.6 \Rightarrow n = 384.16$$

Therefore, you need to take a sample of **at least 385** homes.

Finding A General Formula for Sample Size

In the general case, we can solve for n as follows (here, E is the desired size of the margin of error):

$$Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = E \Rightarrow \frac{\sigma}{\sqrt{n}} = \frac{E}{Z_{\alpha/2}} \Rightarrow \frac{1}{\sqrt{n}} = \frac{E}{\sigma \cdot Z_{\alpha/2}} \Rightarrow \frac{1}{n} = \left(\frac{E}{\sigma \cdot Z_{\alpha/2}} \right)^2 \Rightarrow n = \sigma^2 \frac{Z_{\alpha/2}^2}{E^2}$$

Notice that this is similar to the formula for sample size when estimating the population proportion.

We just replace $\hat{p}\hat{q}$ with σ^2 .

Note: Often in practice, σ is unknown. The common solution for sample size calculations is to stick in some sort of estimate for σ .

This could be based on a pilot study, or using the **Range Rule-of-Thumb**: $\sigma \approx \frac{\text{Range}}{4}$ (See p. 344 for more details).

Section 7-4: Estimating a Population Mean – σ Unknown

In the previous section, we estimated the population mean by using the fact that \bar{X} was normally distributed with a mean

of $\mu_{\bar{X}} = \mu$ and a standard deviation of $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

For this reason, Z had a standard normal distribution, where:

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Now, however, we no longer know σ . How do we get around that? It seems as though the most sensible approach would be to just estimate σ with the sample standard deviation s . The question, however, is if T has a standard normal distribution, where:

$$T = \frac{\bar{X} - \mu}{s / \sqrt{n}}$$

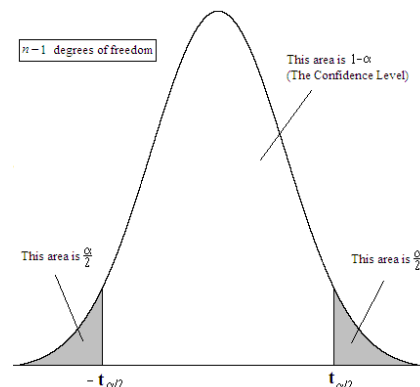
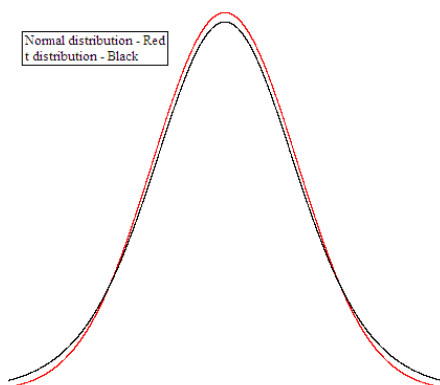
Unfortunately, T does *not* have a standard normal distribution, because we have to account for the fact that s will not be exactly equal to the true population standard deviation σ . However, through statistical theory it turns out that the distribution of T can be specified. The distribution of T is called the **Student-t distribution**, or just the **t distribution**.

Recall: The reason we divided by $n - 1$ instead of n in the sample standard deviation formula $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ was because of something called *degrees of freedom*. In this case, there were only $n - 1$ degrees of freedom because we already knew the sample mean \bar{x} (see Section 3-3 for this discussion).

It turns out that the t distribution changes shape depending on the degrees of freedom of s . For this reason, in finding critical values you have to know the degrees of freedom, which is always $n - 1$.

A Note About the t distribution

The t distribution is also symmetric, like the standard normal distribution, and has a similar shape. However, it is a bit narrower and tapers off less quickly to the right and left (see picture on the left). In computing the confidence interval, we will need to use the value $t_{\alpha/2}$. This is exactly like the value $Z_{\alpha/2}$ from before, except for the t distribution (see picture on the right).



To find the value of $t_{\alpha/2}$, we use Table A-3 (last page in the textbook, or p. 774). Choose the row with “Degrees of Freedom” = $n - 1$ and the column with “Area in One Tail” = $\alpha/2$ (or equivalently, “Area in Two Tails” = α). The critical value is listed at the intersection of that row and column.

(Note: If the degrees of freedom does not appear in the first column, just choose the closest value in the table)

Estimating the Population Mean with σ Unknown

To estimate a population mean in this new situation, 2 conditions must be met:

1. The sample is a simple random sample
2. Either the population is normally distributed or $n > 30$ (so we can use the CLT)

(Notice that these are conditions 1 and 3 from the case with σ known. The only condition that isn't repeated is the condition that σ is known, which is obviously false in this case.)

As before, the best point estimate of the population mean is the sample mean, \bar{x} .

Also, from the discussion above, we know that we will be using a t distribution with $n - 1$ degrees of freedom.

Computing a Confidence Interval for The Population Mean (σ Known)

For a sample size of n , confidence level $1 - \alpha$, and sample mean \bar{x} , the parts of the confidence interval are:

Point Estimate: \bar{x}

Critical Value: $t_{\alpha/2}$ (with $n - 1$ degrees of freedom)

Standard Error: $\frac{s}{\sqrt{n}}$

So the confidence interval is:

$$CI = (\text{Point Estimate}) \pm (\text{Margin of Error})$$

$$= (\text{Point Estimate}) \pm (\text{Critical Value}) \cdot (\text{Standard Error})$$

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

Example

For a class project, some students are doing a survey to find out the average number of hours a Denison student spends studying each week. They take a simple random sample of 91 Denison students and compute a sample mean of 5.4 hours per week with a standard deviation of 1.2 hours. Find a 95% confidence interval for the population mean.

Solution:

First, we note a few things:

- The 2 conditions above are met ($n > 30$)
- The point estimate is $\bar{x} = 5.4$
- The confidence level is $95\% = 0.95 = 1 - \alpha$, so $\alpha = 0.05$
- The degrees of freedom are $n - 1 = 90$
- The critical value is $t_{\alpha/2} = t_{0.05/2} = 1.987$ (Area in one tail is 0.025, or area in two tails is 0.05)
- The standard error is $\frac{s}{\sqrt{n}} = \frac{1.2}{\sqrt{91}} = 0.126$

Therefore, the confidence interval is:

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \Rightarrow 5.4 \pm (1.987)(0.126) \Rightarrow 5.4 \pm 0.250 \Rightarrow (5.15, 5.65)$$

So with 95% confidence, we claim that the average hours per week studied by Denison students is between 5.15 and 5.65 hours.

Determining Sample Size

Now suppose that the students want to choose a sample large enough to get a margin of error of 0.10 or less. How do we determine the sample size?

As in the other times we estimated sample size, we just set the margin of error equal to 0.10:

$$t_{\alpha/2} \frac{s}{\sqrt{n}} = 0.10 \Rightarrow 1.987 \frac{1.2}{\sqrt{n}} = 0.10 \Rightarrow 1.987 \frac{1.2}{0.10} = \sqrt{n} \Rightarrow \sqrt{n} = 23.844 \Rightarrow n = 568.54$$

Therefore, they would need to take a sample of **at least 569** Denison students.

Finding A General Formula for Sample Size

In the general case, we can solve for n as follows (here, E is the desired size of the margin of error):

$$t_{\alpha/2} \frac{s}{\sqrt{n}} = E \Rightarrow \frac{s}{\sqrt{n}} = \frac{E}{t_{\alpha/2}} \Rightarrow \frac{1}{\sqrt{n}} = \frac{E}{s \cdot t_{\alpha/2}} \Rightarrow \frac{1}{n} = \left(\frac{E}{s \cdot t_{\alpha/2}} \right)^2 \Rightarrow \boxed{n = s^2 \frac{t_{\alpha/2}^2}{E^2}}$$

Notice again that this is similar to the formulas for sample size when estimating the population proportion and mean with σ known.

Choosing Between Z and t

Just to summarize the difference between when you use Z and t, here is a useful table.

| If σ is known, and: | If σ is unknown, and: | If: |
|---|---|--|
| The population is normally distributed and/or $n \geq 30$ | The population is normally distributed and/or $n \geq 30$ | The population is <i>not</i> normally distributed and $n < 30$ |
| Then use Z | Then use t | You have to use other methods |

Effects of Sample Size and Confidence Level on Confidence Intervals

Notice that in all three estimates we discussed (proportion and mean with σ known and unknown), two things were always true:

1. In the margin of error, n appears in the denominator.
2. The larger α gets, the smaller the critical value gets.

This means that if n is increased, the margin of error will get smaller. Also, if the confidence level $1 - \alpha$ is decreased (i.e. α is increased), the margin of error will get smaller. A smaller margin of error means a tighter confidence interval, which means we are saying the range of values for the true population proportion/mean is much smaller. In general, it is a good thing to have a smaller confidence interval, as long as your degree of confidence is high.

This introduces the idea of a trade-off in the size of a confidence interval:

- For a fixed sample size n , to reduce the size of a confidence interval, we must reduce the confidence level.
- For a fixed confidence level, to reduce the size of a confidence interval, we must increase the sample size.

Thus to get better confidence intervals, you must either reduce your level of confidence or gather more data.

One More Note

Suppose you are given a confidence interval, but not the point estimate or margin of error. You can find these values using just the interval. The point estimate is at the center of the interval, and the margin of error is half the width of the interval.

In other words, for a confidence interval (a, b) :

$$\hat{p} = \frac{b+a}{2} \text{ and (Margin of Error)} = \frac{b-a}{2}$$

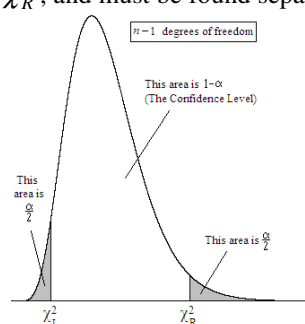
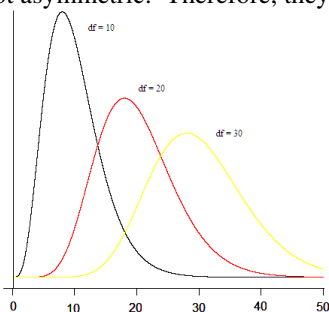
Section 7-5: Estimating a Population Variance σ^2

To come up with a confidence interval for σ^2 , we have to use yet another different distribution.

The quantity $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$ has what is called a **Chi-Square Distribution**. (Chi is pronounced “kigh”, like “high” with a k)

The Chi-Square Distribution is quite a bit different than the normal or the t distributions, because it is not symmetric. This distribution also depends on degrees of freedom, and the larger the degrees of freedom are, the more symmetric it gets (see picture below left). Similar to the normal and t distributions, the critical values for the Chi-Square Distribution are the two values that give the top and bottom areas of $\alpha/2$ (see picture below right). However, notice that these values are no longer opposites of each other, since the

distribution is not asymmetric. Therefore, they are called χ_L^2 and χ_R^2 , and must be found separately.



The critical values χ_L^2 and χ_R^2 can be determined with Table A-4 (p.775 in the textbook). To find these in the table, first find the degrees of freedom $n - 1$ in the left column. Then find the appropriate area to the right of the value (this is $\alpha/2$ for χ_R^2 and $1 - \alpha/2$ for χ_L^2) along the top row. The value is at the intersection of the degrees of freedom and area row/column.

Example: Let degrees of freedom be 20 and $\alpha = 0.05$ (95% confidence). Then the area to the right of χ_R^2 is $0.05/2 = 0.025$. Using the table, we get $\chi_R^2 = 34.170$. The area to the right of χ_L^2 is $1 - 0.025 = 0.975$. From the table, then, $\chi_L^2 = 9.591$.

Estimating the Population Variance

To estimate a population variance, 2 conditions must be met (notice especially requirement 2):

1. The sample is a simple random sample
2. The population *must* be normally distributed.

Computing a Confidence Interval for The Population Variance

For a sample size of n , confidence level $1 - \alpha$, and sample standard deviation s , the confidence interval is:

Left Endpoint of the CI: $\frac{(n-1)s^2}{\chi_R^2}$

Right Endpoint of the CI: $\frac{(n-1)s^2}{\chi_L^2}$

So the confidence interval is: $\left(\frac{(n-1)s^2}{\chi_R^2}, \frac{(n-1)s^2}{\chi_L^2} \right)$

Example (continued)

Recall last section's example: Some students are doing a survey to find out the average number of hours a Denison student spends studying each week. They take a simple random sample of 91 Denison students and compute a sample mean of 5.4 hours per week with a standard deviation of 1.2 hours (i.e. sample variance is $1.2^2 = 1.44$). Find a 95% confidence interval for the population variance. Assume the population is normally distributed.

Solution:

First, we note a few things:

- The 2 conditions above are met
- The confidence level is $95\% = 0.95 = 1 - \alpha$, so $\alpha = 0.05$
- The degrees of freedom are $n - 1 = 90$
- The area above χ_R^2 is $\alpha/2 = 0.05/2 = 0.025$ and the area above χ_L^2 is $1 - \alpha/2 = 1 - 0.05/2 = 1 - 0.025 = 0.975$
- From Table A-4, $\chi_R^2 = 118.136$ and $\chi_L^2 = 65.647$.
- The left endpoint of the CI is $\frac{(n-1)s^2}{\chi_R^2} = \frac{90 \cdot 1.2^2}{118.136} = \frac{129.6}{118.136} = 1.097$
- The right endpoint of the CI is $\frac{(n-1)s^2}{\chi_L^2} = \frac{90 \cdot 1.2^2}{65.647} = \frac{129.6}{65.647} = 1.974$

Therefore, the confidence interval is: (1.097, 1.974)

So with 95% confidence, we claim that the variance in hours per week studied by Denison students is between 1.097 and 1.974 hours.

Equivalently, we could say the standard deviation in hours per week studied by Denison students is between 1.047 and 1.405 hours. (Just take the square root of both endpoints)

Determining Sample Size

To find the proper sample size for estimating population variance, we cannot rely on the arguments presented for the proportion and mean, since there is no margin of error being added and subtracted here. To find the sample size, we will just use a table (see p. 371).

Suppose we want the sample variance s^2 to be within P percent of the value σ^2 . Here is the sample size required for given values of P and confidence levels 95% and 99%:

| Confidence Level 95% | |
|----------------------|--------|
| P | n |
| 1% | 77,207 |
| 5% | 3,148 |
| 10% | 805 |
| 20% | 210 |
| 30% | 97 |
| 40% | 56 |
| 50% | 37 |

| Confidence Level 99% | |
|----------------------|---------|
| P | n |
| 1% | 133,448 |
| 5% | 5,457 |
| 10% | 1,401 |
| 20% | 368 |
| 30% | 171 |
| 40% | 100 |
| 50% | 67 |

Alternate tables are available if you prefer working with standard deviation instead of variance.

Suppose we want the sample standard deviation s to be within P percent of the value σ . Here is the sample size required for given values of P and confidence levels 95% and 99%:

| Confidence Level 95% | |
|----------------------|--------|
| P | n |
| 1% | 19,204 |
| 5% | 767 |
| 10% | 191 |
| 20% | 47 |
| 30% | 20 |
| 40% | 11 |
| 50% | 7 |

| Confidence Level 99% | |
|----------------------|--------|
| P | n |
| 1% | 33,218 |
| 5% | 1,335 |
| 10% | 335 |
| 20% | 84 |
| 30% | 37 |
| 40% | 21 |
| 50% | 13 |

Example

Suppose, at 95% confidence, we want the sample standard deviation s to be within 10% of the population standard deviation σ . What sample size should we use?

Solution:

From the table, with P = 10% and Confidence Level 95%, this sample size is $n = 191$.

Summary

In conclusion, here are the formulas for computing confidence intervals of various parameters:

Confidence Interval for the Population Proportion

Confidence Level: $1 - \alpha$

Sample Proportion: $\hat{p} = \frac{\text{\# successes}}{n}$

Critical Value: $Z_{\alpha/2}$ (found in Table A-2, back cover of book or p. 772-773)

Confidence Interval: $\hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Sample Size Calculation: $n = \hat{p}\hat{q} \frac{Z_{\alpha/2}^2}{E^2}$

Confidence Interval for the Population Mean (σ Known)

Confidence Level: $1 - \alpha$

Sample Mean: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

Critical Value: $Z_{\alpha/2}$ (found in Table A-2, back cover of book or p. 772-773)

Confidence Interval: $\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

Sample Size Calculation: $n = \sigma^2 \frac{Z_{\alpha/2}^2}{E^2}$

Confidence Interval for the Population Mean (σ unknown)Confidence Level: $1 - \alpha$ Degrees of Freedom: $n - 1$

Sample Mean: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

Critical Value: $t_{\alpha/2}$ (found in Table A-3, last page of book or p. 774)

Confidence Interval: $\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$

Sample Size Calculation: $n = s^2 \frac{t_{\alpha/2}^2}{E^2}$

Confidence Interval for the Population VarianceConfidence Level: $1 - \alpha$ Degrees of Freedom: $n - 1$ Critical Value: χ_R^2 and χ_L^2 (found in Table A-4, last page of book or p. 775)

Confidence Interval: $\left(\frac{(n-1)s^2}{\chi_R^2}, \frac{(n-1)s^2}{\chi_L^2} \right)$

Sample Size Calculation: (tables on previous page)