7–2

Confidence Intervals for the Mean When σ Is Unknown

Objective 3

Find the confidence interval for the mean when σ is unknown.

When σ is known and the sample size is 30 or more, or the population is normally distributed if the sample size is less than 30, the confidence interval for the mean can be found by using the z distribution as shown in Section 7–1. However, most of the time, the value of σ is not known, so it must be estimated by using s, namely, the standard deviation of the sample. When s is used, especially when the sample size is small, critical values greater than the values for $z_{\alpha/2}$ are used in confidence intervals in order to keep the interval at a given level, such as the 95%. These values are taken from the *Student t distribution,* most often called the ***t* distribution.**

To use this method, the samples must be simple random samples, and the population from which the samples were taken must be normally or approximately normally distributed, or the sample size must be 30 or more.

Some important characteristics of the t distribution are described now.

Historical Notes

The *t* distribution was formulated in 1908 by an Irish brewing employee named W. S. Gosset. Gosset was involved in researching new methods of manufacturing ale. Because brewing employees were not allowed to publish results, Gosset published his finding using the pseudonym *Student;* hence, the *t* distribution is sometimes called *Student's t distribution.*

Characteristics of the *t* Distribution

The t distribution shares some characteristics of the normal distribution and differs from it in others. The t distribution is similar to the standard normal distribution in these ways:

1. It is bell-shaped.
2. It is symmetric about the mean.
3. The mean, median, and mode are equal to 0 and are located at the center of the distribution.
4. The curve never touches the x axis.

The t distribution differs from the standard normal distribution in the following ways:

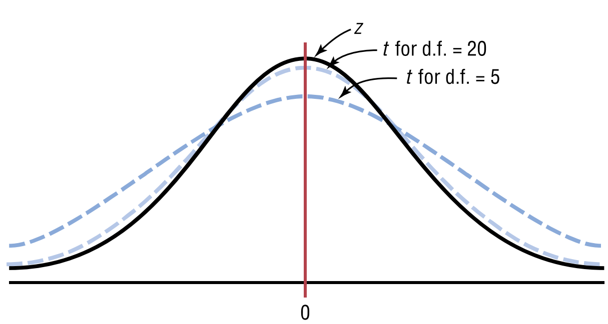
1. The variance is greater than 1.
2. The t distribution is actually a family of curves based on the concept of *degrees of freedom,* which is related to sample size.
3. As the sample size increases, the t distribution approaches the standard normal distribution. See Figure 7–6.

Many statistical distributions use the concept of degrees of freedom, and the formulas for finding the degrees of freedom vary for different statistical tests. The **degrees of freedom** are the number of values that are free to vary after a sample statistic has been computed, and they tell the researcher which specific curve to use when a distribution consists of a family of curves.

For example, if the mean of 5 values is 10, then 4 of the 5 values are free to vary. But once 4 values are selected, the fifth value must be a specific number to get a sum of 50, since 50 ÷ 5 = 10. Hence, the degrees of freedom are 5 − 1 = 4, and this value tells the researcher which t curve to use.

Figure 7–6

The *t* Family of Curves



The symbol d.f. will be used for degrees of freedom. The degrees of freedom for a confidence interval for the mean are found by subtracting 1 from the sample size. That is, d.f. $= n - 1$. *Note:* For some statistical tests used later in this book, the degrees of freedom are not equal to $n - 1$.

The formula for finding a confidence interval about the mean by using the t distribution is given now.

Formula for a Specific Confidence Interval for the Mean When σ Is Unknown

$$\bar{X} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{X} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

The degrees of freedom are $n - 1$.

The values for $t_{\alpha/2}$ are found in Table F in Appendix C. The top row of Table F, labeled Confidence Intervals, is used to get these values. The other two rows, labeled One tail and Two tails, will be explained in Chapter 8 and should not be used here.

Example 7–5 shows how to find the value in Table F for $t_{\alpha/2}$.

Example 7–5

Find the $t_{\alpha/2}$ value for a 95% confidence interval when the sample size is 22.

Solution

The d.f. $= 22 - 1$, or 21. Find 21 in the left column and 95% in the row labeled Confidence Intervals. The intersection where the two meet gives the value for $t_{\alpha/2}$, which is 2.080. See Figure 7–7.

Figure 7–7

Finding $t_{\alpha/2}$ for Example 7–5

Table F

The t Distribution

d.f.	Confidence Intervals	50%	80%	90%	95%	98%	99%
	One tail α	0.25	0.10	0.05	0.025	0.01	0.005
	Two tails α	0.50	0.20	0.10	0.05	0.02	0.01
1							
2							
3							
⋮							
21					2.080	2.518	2.831
⋮							
$(z)\infty$		0.674	1.282[a]	1.645[b]	1.960	2.326[c]	2.576[d]

When d.f. is greater than 30, it may fall between two table values. For example, if d.f. = 68, it falls between 65 and 70. Many textbooks say to use the closest value, for example, 68 is closer to 70 than 65; however, in this textbook a conservative approach is used. In this case, always round down to the nearest table value. In this case, 68 rounds down to 65.

Note: At the bottom of Table F where d.f. is large or ∞, the $z_{\alpha/2}$ values can be found for specific confidence intervals. The reason is that as the degrees of freedom increase, the t distribution approaches the standard normal distribution.

Examples 7–6 and 7–7 show how to find the confidence interval when you are using the t distribution.

Assumptions for Finding a Confidence Interval for a Mean When σ Is Unknown

1. The sample is a random sample.
2. Either $n \geq 30$ or the population is normally distributed if $n < 30$.

Example 7–6 Sleeping Time

Ten randomly selected people were asked how long they slept at night. The mean time was 7.1 hours, and the standard deviation was 0.78 hour. Find the 95% confidence interval of the mean time. Assume the variable is normally distributed.

Source: Based on information in *Number Freaking*.

Solution

Since σ is unknown and s must replace it, the t distribution (Table F) must be used for the confidence interval. Hence, with 9 degrees of freedom $t_{\alpha/2} = 2.262$. The 95% confidence interval can be found by substituting in the formula.

$$\bar{X} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{X} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

$$7.1 - 2.262\left(\frac{0.78}{\sqrt{10}}\right) < \mu < 7.1 + 2.262\left(\frac{0.78}{\sqrt{10}}\right)$$

$$7.1 - 0.56 < \mu < 7.1 + 0.56$$

$$6.54 < \mu < 7.66$$

Therefore, one can be 95% confident that the population mean is between 6.54 and 7.66 inches.

Example 7–7 Home Fires Started by Candles

The data represent a sample of the number of home fires started by candles for the past several years. (Data are from the National Fire Protection Association.) Find the 99% confidence interval for the mean number of home fires started by candles each year.

5460 5900 6090 6310 7160 8440 9930

Solution

Step 1 Find the mean and standard deviation for the data. Use the formulas in Chapter 3 or your calculator. The mean $\bar{X} = 7041.4$. The standard deviation $s = 1610.3$.

Step 2 Find $t_{\alpha/2}$ in Table F. Use the 99% confidence interval with d.f. $= 6$. It is 3.707.

Step 3 Substitute in the formula and solve.

$$\bar{X} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{X} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

$$7041.4 - 3.707\left(\frac{1610.3}{\sqrt{7}}\right) < \mu < 7041.4 + 3.707\left(\frac{1610.3}{\sqrt{7}}\right)$$

$$7041.4 - 2256.2 < \mu < 7041.4 + 2256.2$$

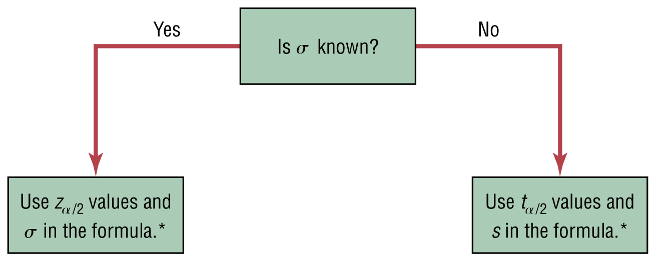
$$4785.2 < \mu < 9297.6$$

One can be 99% confident that the population mean number of home fires started by candles each year is between 4785.2 and 9297.6, based on a sample of home fires occurring over a period of 7 years.

Students sometimes have difficulty deciding whether to use $z_{\alpha/2}$ or $t_{\alpha/2}$ values when finding confidence intervals for the mean. As stated previously, when σ is known, $z_{\alpha/2}$ values can be used *no matter what the sample size is,* as long as the variable is normally distributed or $n \geq 30$. When σ is unknown and $n \geq 30$, then s can be used in the formula and $t_{\alpha/2}$ values can be used. Finally, when σ is unknown and $n < 30$, s is used in the formula and $t_{\alpha/2}$ values are used, as long as the variable is approximately normally distributed. These rules are summarized in Figure 7–8.

Figure 7–8

When to Use the *z* or *t* Distribution



*If $n < 30$, the variable must be normally distributed.

Applying the Concepts 7–2

Sport Drink Decision

Assume you get a new job as a coach for a sports team, and one of your first decisions is to choose the sports drink that the team will use during practices and games. You obtain a *Sports Report* magazine so you can use your statistical background to help you make the best decision. The following table lists the most popular sports drinks and some important information about each of them. Answer the following questions about the table.

Drink	Calories	Sodium	Potassium	Cost
Gatorade	60	110	25	$1.29
Powerade	68	77	32	1.19
All Sport	75	55	55	0.89
10-K	63	55	35	0.79
Exceed	69	50	44	1.59
1st Ade	58	58	25	1.09
Hydra Fuel	85	23	50	1.89

1. Would this be considered a small sample?
2. Compute the mean cost per container, and create a 90% confidence interval about that mean. Do all the costs per container fall inside the confidence interval? If not, which ones do not?
3. Are there any you would consider outliers?
4. How many degrees of freedom are there?
5. If cost is a major factor influencing your decision, would you consider cost per container or cost per serving?
6. List which drink you would recommend and why.

See page 398 for the answers.