8.1 INTERVAL ESTIMATION OF A POPULATION MEAN:
LARGE SAMPLES OR \( \sigma \) KNOWN

When the sample size is large \((n \geq 30)\), the margin of error of estimate for a confidence interval for \(\mu\) can be found with Excel’s CONFIDENCE function. This function has three inputs:

1) the confidence level subtracted from 100\%, entered as a decimal, \(\alpha = \text{Alpha}\)
2) the standard deviation of the population, \(\sigma = \text{Standard_dev}\)
   (use \(s\), the sample standard deviation, if \(\sigma\) is not known)
3) the sample size, \(n = \text{Size}\).
This margin of error of estimate can then be subtracted from and added to the sample mean in order to construct the interval. Indeed, recall that the endpoints of such a confidence interval are as follows:

\[ \bar{x} \pm z \sigma \sqrt{\frac{1}{n}} \]

Here, \( n \) is the sample size, \( \sigma \) is the population standard deviation, and \( z \) is the cut-off for confidence level \( = \alpha \). If \( n \) is large and \( \sigma \) is not known, then it can be replaced by \( s \), the sample standard deviation.

**Example 8-1** A publishing company has just published a new college textbook. Before the company decides the price at which to sell this textbook, it wants to know the average price of all such textbooks in the market. The research department at the company took a sample of 36 such textbooks and collected information on their prices. This information produced a mean price of $70.50 for this sample. It is known that the standard deviation of the prices of all such textbooks is $4.50.

a) What is the point estimate of the mean price of all such college textbooks?

b) What is the margin of error of estimate for a 95% confidence interval?

c) Construct a 90% confidence interval for the mean price of all such college textbooks.

**Solution:**

a) The point estimate is the sample mean, 70.50. Enter this value into cell A1 of a new Excel spreadsheet.

b) For the margin of error of estimate for a 95% confidence interval, click on cell A2, click on the \( fx \) icon, and select the CONFIDENCE function. The first input is 100% – 95% = 5% entered as a decimal. You can either enter .05 or just 1-.95. The second input is the standard deviation of the population. (Use the sample standard deviation, if the population one is not known). Enter 4.50. (This function will automatically divide this quantity by the square root of \( n \) for the standard deviation of the sampling distribution.) The third input is the sample size. Enter 36. Click on **OK** and you should see the margin of error of estimate is 1.469971, or ±$1.47.
c) For the 90% confidence interval, we first need the margin of error of estimate. Copy cell A2 (from part b) to cell A3, double-click on it in order to edit it, and change the first input from 1-0.95 to 1-0.90. Hit ENTER and you should see the margin of error of estimate change to 1.23364, or $1.23. Now click on cell A4, type “=” and click on cell A1 (containing the sample mean), type “-” and click on cell A3 (containing the margin of error of estimate), and hit ENTER. You should see the lower limit of the confidence interval appear: 69.26636 or $69.27. Click on cell A5, type “=” and click on cell A1, type “+” and click on cell A3, and hit ENTER. You should see the upper limit of the confidence interval appear: 71.73364 or $71.73. Format these to two-decimal-place currency if you wish (go to Format>Cells, click on the Number tab and select Currency).

Figure 8.2 Use Excel's CONFIDENCE function with addition and subtraction to get the upper and lower limits of the confidence interval. (Answers may differ slightly from those found using values from a normal distribution table due to round-off error.)
Example 8-2  According to a report by the Consumer Federation of America, National Credit Union Foundation, and the Credit Union National Association, households with negative assets carried an average of $15,528 in debt in 2002 (CBS.MarketWatch.com, May 14, 2002). Assume that this mean was based on a random sample of 400 households and that the standard deviation of debts for households in this sample was $4200. Construct a 99% confidence interval for the 2002 mean debt for all such households.

Solution:  Click on an empty cell in an Excel worksheet and insert the CONFIDENCE function. For a 99% confidence interval, use $\alpha = 1\%$ or .01 as the first input. The population standard deviation is unknown, but we can use the sample standard deviation of $\$4200$ for an approximation. This is the second input. The sample size is 400, which is the third input. Hit ENTER and you should see that the margin of error of estimate is 540.9252 or $\$540.93$. Click on another empty cell and type “=15528-” and click on the cell with the margin of error of estimate in it. Hit ENTER and you should see that the lower limit of the confidence interval is 14987.07. Click on another empty cell and type “=15528+” and click on the cell with the margin of error of estimate in it again. Hit ENTER and you should see that the upper limit of the confidence interval is 16068.93.

First and foremost, you cannot use Excel’s CONFIDENCE function for small samples, since it uses values based on the normal distribution. But, as is discussed in the text, we can still handle this situation by considering the $t$-distribution. Indeed, in such case, the endpoints for the confidence interval are as follows:

$$\bar{x} \pm t \frac{s}{\sqrt{n}}$$

Here, $n$ is the sample size, $s$ is the sample standard deviation, and $t$ is the cut-off for confidence level = $\alpha$. (Here, the margin of error of estimate is $t \frac{s}{\sqrt{n}}$.)

Thankfully, Excel has a built-in function, namely TINV, that can find the cut-off values based on the $t$-distribution. This function requires two inputs:
1) the confidence level subtracted from 100%, entered as a decimal = \textbf{Probability} (note that this is the area of the two tails in the \textit{t}-distribution)

2) one less than the sample size, $n - 1 = \textbf{Deg\_freedom}$.

\textbf{Example 8-3} Dr. Moore wanted to estimate the mean cholesterol level for all adult men living in Hartford. He took a sample of 25 adult men from Hartford and found that the mean cholesterol level for this sample is 186 with a standard deviation of 12. Assume that the cholesterol levels for all adult men in Hartford are (approximately) normally distributed. Construct a 95\% confidence interval for the population mean, $\mu$.

\textbf{Solution:} There are three steps involved:

1) Use $\text{TINV}(\alpha, n-1)$ to find the value of $t$ associated with a confidence level of 95\% ($\alpha = .05$) and sample size of 25 ($n-1 = 24$).

\[=\text{TINV}(0.05, 24)\]

\begin{center}
\begin{tabular}{|c|c|}
\hline
\textbf{Probability} & 0.05 \\
\hline
\textbf{Deg\_freedom} & 24 \\
\hline
\end{tabular}
\end{center}

\textit{Figure 8.4} Using Excel’s \textit{TINV} function to obtain the value of $t$ associated with a 95\% confidence interval with a small sample size of 25.

2) Calculate the margin of error of estimate using the formula $=t*\frac{s}{\text{SQRT}(n)}$ using the value of $t$ calculated in step 1, $s = 12$, and $n = 25$.

\begin{center}
\begin{tabular}{|c|c|c|c|c|}
\hline
\textbf{A} & \textbf{B} & \textbf{C} & \textbf{D} & \textbf{E} \\
\hline
1 & 2.063898 & & & \\
2 & 4.953356 & & & \\
\hline
\end{tabular}
\end{center}

\textit{Figure 8.5} Using the formula for the margin of error of estimate with a cell reference to the $t$-value calculated previously.

3) Add and subtract the value calculated in step 2 to and from the sample mean of 186 in order to get the upper and lower limits of the interval.
8.3 INTERVAL ESTIMATION OF A POPULATION PROPORTION: LARGE SAMPLES

For a sufficiently large sample, the endpoints for the confidence interval about a sample proportion are as follows:

$$\hat{p} \pm z \sqrt{\frac{pq}{n}}$$

Here, $n$ is the sample size, $p$ is the population proportion, $q = 1 - p$, and $z$ is the cut-off for confidence level $= \alpha$. (Here, the margin of error of estimate is $z \sqrt{\frac{pq}{n}}$.) (Recall that the criteria for a “sufficiently large sample” are that $n \hat{p}$ and $n \hat{q}$ are both greater than 5.) In order to construct a confidence interval for a population proportion, $p$, when we have a large sample, we can use Excel’s CONFIDENCE function again, as illustrated below.

Example 8-4 According to a 2002 survey by FindLaw.com, 20% of Americans needed legal advice during the past year to resolve such thorny issues as family trusts and landlord disputes (CBS.MarketWatch.com, August 6, 2002). Suppose a recent sample of 1000 adult Americans showed that 20% of them needed legal advice during the past year to resolve such family-related issues.

a) What is the point estimate of the population proportion?

b) What is the margin of error of estimate for a 95% confidence interval?

c) Find a 99% confidence interval for the percentage of all adult Americans who needed legal advice during the past year to resolve such family-related issues.

Solution:

a) The point estimate is the sample proportion, 20% or 0.20. Enter this value into cell A1 of a new Excel spreadsheet.

b) For the margin of error of estimate for a 95% confidence interval, click on cell A2, click on the $f_x$ icon, and select the CONFIDENCE function. The first input is 100% – 95% = 5% entered as a decimal. You can either enter .05 or just 1-.95. The second input is the standard deviation of the population, which we will
approximate with the sample standard deviation. Note that \( \hat{p} = .20 \), so \( \hat{q} = 1 -.20 = .80 \). So enter \( \text{SQRT}(.2 \times .8) \). The third input is the sample size, 1000. Click OK and you the margin of error of estimate 0.024792, or ± 2.5%, appears.

\[
\text{CONFIDENCE}(0.05, \text{SQRT}(0.2 \times 0.8), 1000)
\]

\( = 0.024791765 \)

Returns the confidence interval for a population mean. See Help for the equation used.

\( \text{Size} \) is the sample size.

Figure 8.7 Using Excel’s CONFIDENCE function to find the margin of error of estimate for a 95% confidence interval for \( p \). (This is also called the “margin of error.”)

c) For the 99% confidence interval, we first need the margin of error of estimate. Copy cell A2 (from part b) to cell A3, double-click on it in order to edit it, and change the first input from .05 to .01. Hit ENTER and you should see the margin of error of estimate change to .032582, or 3.3%. Now click on cell A4, type “=” and click on cell A1 (containing the sample proportion), type “-” and click on cell A3 (containing the margin of error of estimate), and hit ENTER. You should see the lower limit of the confidence interval appear: 0.167418 or 16.7%. Click on cell A5, type “=” and click on cell A1, type “+” and click on cell A3, and hit ENTER. You should see the upper limit of the confidence interval appear: 0.232582 or 23.3%. Format these to percentages with one decimal place if you wish (go to Format>Cells, click on the Number tab and select Percentage, and change the number of decimal places to 1).

\[
\begin{array}{cccccccc}
A & B & C & D & E & F & G \\
1 & 20.0\% & = & \text{sample proportion} & & & \\
2 & 2.5\% & = & \text{maximum error of estimate for 95\% confidence interval} & & & \\
3 & 3.3\% & = & \text{maximum error of estimate for 99\% confidence interval} & & & \\
4 & 16.7\% & = & \text{lower limit of 99\% confidence interval} & & & \\
5 & 23.3\% & = & \text{upper limit of 99\% confidence interval} & & & \\
\end{array}
\]

Figure 8.8 Use Excel’s CONFIDENCE function with addition and subtraction to get the upper and lower limits of the confidence interval for \( p \).
8.4 DETERMINING THE SAMPLE SIZE FOR THE ESTIMATION OF MEAN

In order to find the sample size required for estimating the mean with a given level of confidence and margin of error of estimate, you can either use trial-and-error with the CONFIDENCE function, or calculate it using the formula in your textbook. We discuss both approaches below.

For example, what would be the minimum required sample size if you want to be 99% confident that the sample mean is within two units of the population mean, given that $\sigma = 1.4$? Well, using trial and error, start off with $n = 30$, and see what you get. Then, suitably modify the value of $n$ to hone in on the best result:

CONFIDENCE(0.01, 1.4, 30) = 0.658. That’s well within 2 units, so try a smaller one, say $n = 15$.
CONFIDENCE(0.01, 1.4, 15) = 0.931. Still that’s less than 2. Try $n = 5$.
CONFIDENCE(0.01, 1.4, 5) = 1.612. That’s closer to 2. What about $n = 4$?
CONFIDENCE(0.01, 1.4, 4) = 1.803. That’s even closer. What about $n = 3$?
CONFIDENCE(0.01, 1.4, 3) = 2.082. That’s no longer within 2 units. Therefore, $n = 4$ is the minimum sample size required.

Alternatively, rather than using trial and error, recall that the formula for the margin of error of estimate in the mean (for large samples) is given by:

$$E = z \frac{\sigma}{\sqrt{n}},$$

where $E$ = margin of error (which coincides with one-half the length of the confidence interval), $z$ = cut-off for confidence level = $\alpha$, $\sigma$ = population standard deviation, and $n$ = sample size. Since we are given values of $E$, $z$, and $\sigma$, we need to solve this equation for $n$. Doing so yields the result:

$$n = \frac{z^2 \sigma^2}{E^2} = \left( \frac{z \sigma}{E} \right)^2$$

Now that you have the formula, you can easily use Excel to perform the arithmetic necessary. Keep in mind that the right-side will seldom be an integer. As such, to obtain the smallest sample size necessary to accompany the values of $E$, $z$, and $\sigma$, choose $n$ to be the integer just larger than the right-side of the above formula. For instance, if $\left( \frac{z \sigma}{E} \right)^2 = 67.12$, then choose $n$ to be 68.

Clearly, the benefit of using this method over trial and error is that it directly, in a single computation, reveals what the value of $n$ ought to be, whereas trial and error can take several iterations to hone in on the same value. We illustrate both approaches in the following example.
Example 8-5  An alumni association wants to estimate the mean debt of this year’s college graduates. It is known that the population standard deviation of the debts of this year’s college graduates is $11,800. How large a sample should be selected so that the estimate with a 99% confidence level is within $800 of the population mean?

Solution:

Trial and Error approach:  Start by typing =CONFIDENCE(0.01, 11800, 30) into an empty cell in an Excel worksheet, since 1-.99 = .01 and the standard deviation is 11800. This sample size of 30 gives a margin of error of estimate of 5549.314, which is way too big. So we must have a much larger sample in order to get it down around 800. Try changing the “30” to “1000.” This gives a value of 961.1695. This is much closer! Keep changing the sample size, n, in CONFIDENCE(0.01, 11800, n) in order to see where it changes from above 800 to below 800. You should find that n = 1444 gives a margin of error of estimate of 799.8644 whereas n = 1443 gives a margin of error of estimate of 800.1415. So to be within $800, the minimum sample size required is 1444.

Algebraic approach:  We use the formula \( n = \frac{z^2 \sigma^2}{E^2} = \left( \frac{z \sigma}{E} \right)^2 \). In this problem, we identify \( E = 800 \), \( \sigma = 11,800 \), and \( z = \text{NORMSINV}(0.005) = 2.576 \). (Remember to divide the confidence level by 2 as the input into the NORMSINV function.). Substitute these values into distinct cells in Excel and compute the right-side of the formula to obtain: \( \left( \frac{(2.576)(11,800)}{800} \right)^2 = 1443.696 \). So, we would choose n to be 1444, as above.

8.5 DETERMINING THE SAMPLE SIZE FOR THE ESTIMATION OF PROPORTION

Just as with the mean, we can determine the sample size required for estimating the proportion with a given level of confidence and margin of error of estimate either by calculating it using the formula derived in the textbook, or by using trial-and-error with the CONFIDENCE function. Regarding the former approach, recall that the formula for the margin of error in this case is \( E = z \sqrt{\frac{\hat{p} \hat{q}}{n}} \), where \( \hat{p} \) is the sample proportion and \( \hat{q} = 1 - \hat{p} \); if the actual population proportion \( p \) is known, then use it, along with \( q = 1 - p \), instead. Algebraically solving the above equation for \( n \) yields:

\[
 n = \frac{z^2 \hat{p} \hat{q}}{E^2}
\]

Both approaches are illustrated in the following example.
Example 8-6  Lombard Electronics Company has just installed a new machine that makes a part that is used in clocks. The company wants to estimate the proportion of these parts produced by this machine that are defective. The company manager wants this estimate to be within .02 of the population proportion for a 95% confidence level. What is the most conservative estimate of the sample size that will limit the margin of error to within .02 of the population proportion?

Solution:

Trial and Error approach: We assume that the probability of getting a defective part is .50 since we don’t know any better a priori. As such, start by typing =CONFIDENCE(0.05, SQRT(.5*.5), 30) into an empty cell in an Excel worksheet, since 1-.95 = .05 and we’re using \( p = q = .50 \) for the most conservative estimate. This sample size of 30 gives a margin of error of estimate of 0.178919, which is way too big. So we must have a much larger sample in order to get it down around 0.02. Try changing the “30” to “1000.” This gives a value of 0.03099. This is much closer! Keep changing the sample size, \( n \), in CONFIDENCE(0.05, SQRT(.5*.5), \( n \)) in order to see where it changes from above 0.02 to below 0.02. You should find that \( n = 2401 \) gives a margin of error of estimate of 0.019996, which is not quite there; also, \( n = 2400 \) gives a margin of error of estimate of 0.020004. So to be within 0.02, the minimum sample size required is 2402.

Algebraic approach: We use the formula \( n = \frac{z^2 \hat{p} \hat{q}}{E^2} \). In this problem, we identify \( E = 0.02 \), \( \hat{p} = \hat{q} = 0.50 \) (by assumption), and \( z = \text{NORMSINV}(0.025) = 1.96 \). (Remember to divide the confidence level by 2 as the input into the NORMSINV function.). Substitute these values into distinct cells in Excel and compute the right-side of the formula to obtain: \( \frac{(1.96)^2(0.5)(0.5)}{0.02^2} = 2401 \). So, we would choose \( n \) to be 2402, to be on the safe side.

Example 8-7  Consider Example 8-6 again. Suppose a preliminary sample of 200 parts produced by this machine showed that 7% of them are defective. How large a sample should the company select so that the 95% confidence interval for \( p \) is within .02 of the population proportion?

Solution:

Trial and Error approach: Start by typing =CONFIDENCE(0.05, SQRT(.07*.93), 30) into a an empty cell in an Excel worksheet, since 1-.95 = .01 and we’re using \( \hat{p} = .07 \) and \( \hat{q} = .93 \) from the preliminary sample. This sample size of 30 gives a margin of error of estimate of 0.091301, which is still too big. We must have a much larger sample in order to get it down around 0.02. Again change the “30” to “1000.” This gives a value of 0.015814. This is too small! So try a smaller sample, say \( n = 500 \). This gives a value of 0.022364, much closer. As before, keep changing the sample size, \( n \), in
CONFIDENCE(0.05, SQRT(0.07*0.93), n) in order to see where it changes from above 0.02 to below 0.02. You should find that \( n = 626 \) gives a margin of error of estimate of 0.019987, whereas \( n = 625 \) gives a margin of error of estimate of 0.020003. So, in order to be within 0.02, the minimum sample size required is 626.

**Algebraic approach:** We use the formula \( n = \frac{z^2 \hat{p} \hat{q}}{E^2} \). In this problem, we identify \( E = 0.02 \), \( \hat{p} = 0.07 \), \( \hat{q} = 0.93 \) (by assumption), and \( z = NORMSINV(0.025) = 1.96 \). (Remember to divide the confidence level by 2 as the input into the NORMSINV function.). Substitute these values into distinct cells in Excel and compute the right-side of the formula to obtain: \( \frac{(1.96)^2(0.07)(0.93)}{(0.02)^2} = 625.22 \). So, we would choose \( n \) to be 626.

**Exercises**

8.1 According to *Money* magazine, the average net worth of U.S. households in 2002 was $355,000, (*Money*, Fall 2002). Assume that this mean is based on a random sample of 500 households and that the sample standard deviation is $125,000. Use Excel’s CONFIDENCE function to construct a 99% confidence interval for the 2002 mean net worth of all U.S. households.

8.2 According to a 2002 survey by America Online, mothers with children under age 18 spent an average of 16.87 hours per week online (*USA TODAY*, May 7, 2002). Suppose that this mean is based on a random sample of 1000 such mothers and that the standard deviation for this sample is 3.2 hours per week. Use Excel’s CONFIDENCE function to construct a 95% confidence interval for the corresponding population mean for all such mothers.

8.3 A random sample of 16 airline passengers at the Bay City airport showed that the mean time spent waiting in line to check in at the ticket counters was 31 minutes with a standard deviation of 7 minutes. Use Excel’s TINV function and the formula for the margin of error of estimate in order to construct a 99% confidence interval for the mean time spent waiting in line by all passengers at this airport. (Assume that such waiting times for all passengers are normally distributed.)

8.4 A random sample of 20 acres gave a mean yield of wheat equal to 41.2 bushels per acre with a standard deviation of 3 bushels. Assume that the yield of wheat per acre is normally distributed, and use Excel’s TINV function and the formula for the margin of error of estimate in order to construct a 90% confidence interval for the population mean, \( \mu \).

8.5 In a Maritz poll of 1004 adult drivers conducted in July 2002, 36% said that they “often” or “sometimes” talk on their cell phones while driving (*USA TODAY*, October 23, 2002). Assume that these 1004 drivers make a random sample of all
adult drivers in the United States. Use Excel’s CONFIDENCE function in order to answer the following:

a) What is the point estimate of the corresponding population proportion?

b) What is the margin of error of estimate for a 95% confidence interval?

c) Construct a 99% confidence interval for the proportion of all adult drivers in the United States who “often” or “sometimes” talk on their cell phones while driving.

8.6 A marketing researcher wants to find a 95% confidence interval for the mean amount that visitors to a theme park spend per person per day. She knows that the standard deviation of the amounts spent per person per day by all visitors to this park is $11. Use both the trial and error approach involving Excel’s CONFIDENCE function and the algebraic approach to determine how large a sample the researcher should select so that the estimate will be within $2 of the population mean.

8.7 Tony’s Pizza guarantees all pizza deliveries within 30 minutes of the placement of orders. An agency wants to estimate the proportion of all pizzas delivered within 30 minutes by Tony’s. Use both the trial and error approach involving Excel’s CONFIDENCE function and the algebraic approach to determine the most conservative estimate of the sample size that would limit the margin of error to within .02 of the population proportion for a 99% confidence interval.

8.8 Refer to exercise 8.7. Assume that a preliminary study has shown that 93% of all Tony’s pizzas are delivered within 30 minutes. Use both the trial and error approach involving Excel’s CONFIDENCE function and the algebraic approach to determine how large the sample size should be so that the 99% confidence interval for the population proportion has a margin of error of .02.