CHAPTER 8 ESTIMATION

CONFIDENCE INTERVALS FOR THE MEAN – WHEN \( \sigma \) IS KNOWN (SECTION 8.1 OF UNDERSTANDABLE STATISTICS)

Excel’s function for computing confidence intervals for a population mean assumes a normal distribution, regardless of sample size. Therefore, as stated in the text, if the population distribution is not a normal distribution, then a large sample should be used. The command syntax is

\[
\text{CONFIDENCE}(\alpha, \text{standard deviation}, \text{sample size})
\]

where \( \alpha \) equals 1 minus the confidence level. In other words, an alpha of 0.05 indicates a 95% confidence interval. To generate a 99% confidence interval, use \( \alpha = 0.01 \). The standard deviation is the population standard deviation \( \sigma \).

Recall from the discussion in Understandable Statistics that a confidence interval for the population mean has the form

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

Excel uses the formula

\[
E = z_c \frac{\sigma}{\sqrt{n}}
\]

where \( z_c \) is the critical value for the chosen confidence level. For \( c = 95\% \), \( z_c = 1.96 \).

CONFIDENCE returns the value of \( E \). To find the lower boundary of the confidence interval, you must subtract \( E \) from the sample mean \( \bar{x} \) of your data; to find the upper boundary, you add \( E \) to the sample mean.

Example

Lucy decided to try to estimate the average number of miles she drives each day. For a three-month period, she selected a random sample of 35 days and kept a record of the distance driven on each of those sample days. The sample mean was 28.3 miles, and also assume that the standard deviation was \( \sigma = 5 \) miles. Find a 95% confidence interval for the population mean of miles Lucy drove per day in the three-month period.

For a 95% confidence level, alpha is 0.05. In our Excel worksheet, after putting in some information and labels, we enter =CONFIDENCE(0.05,5,35) in Cell B3 to represent \( E \). In Cell D4, we type =28.3–B3 to compute the lower value of the confidence interval, and we place the upper value of the confidence interval in Cell F4 using =28.3+B3.
Confidence Intervals for the Mean – When $\sigma$ is unknown (Section 8.2 of *Understandable Statistics*)

When the population standard deviation $\sigma$ is unknown, the sample standard deviation $s$ is used to replace $\sigma$, and the confidence intervals are computed using Student’s $t$ distribution.

The confidence interval is computed as

$$\bar{x} - E \to \bar{x} + E$$

where $E = t_{c} \frac{s}{\sqrt{n}}$.

Excel has two commands for Student’s $t$ distribution under the statistical options of Paste Function.

**TDIST(x, degrees of freedom, tails)**

returns the area in the tail of Student’s $t$ distribution beyond the specified value of $x$, for the specified number of degrees of freedom and number of tails (1 or 2).

**TINV(probability, degrees of freedom)**

returns the $t_{c}$ value such that the area in the two tails beyond the $t_{c}$ value equals the specified probability for the specified degrees of freedom.

We can use the TINV command to find the $t_{c}$ value to use in the confidence interval. For instance, if we have a sample of size 10 with mean $\bar{x} = 6$ and $s = 1.2$, then the $t_{c}$ value we use in the computation of a 95% confidence interval is 2.262. Notice that for a 95% confidence interval, 5% of the area is in the two tails. The number of degrees of freedom is $10 - 1 = 9$.

We can also find the $E$ value in the Describe Statistics dialog box found by selecting **Tools** ➔ **Data Analysis** ➔ **Descriptive Statistics**. If you have data entered into a worksheet, you can use these menu selections to automatically compute the sample mean $\bar{x}$ and the sample standard deviation $s$ for the data, as well as the value of $E$ for the confidence interval on the basis of Student’s $t$ distribution no matter what the sample size is.
Example

Open the Excel workbook Svls02.xls. This workbook contains the weights in pounds of fifty randomly selected professional football linebackers.

Use the menu selection ▶ Tools ▶ Data Analysis ▶ Descriptive Statistics to access the following dialog box. Since the weights are in Column A, we select the cells containing the range and use these for the input range. We place the upper left corner of the output in Cell C1 and check that we want summary statistics and a 95% confidence interval.

Next we widen cells in Column C and D to accommodate the output. Notice that the sample mean is in Cell D3 and the value of $E$ for a 95% confidence interval (again, based on Student’s $t$ distribution, not the standard normal distribution) is in Cell D16. We put the confidence interval in Cells F2 to H2, by entering $=D3-D16$ in Cell F2 and $=D3+D16$ in Cell H2. The display should be similar to the one shown at the top of the next page.
LAB ACTIVITIES FOR CONFIDENCE INTERVALS FOR THE MEAN

1. Retrieve the worksheet Svls03.xls. This contains the heights (in feet) of 65 randomly selected pro basketball players. Use the Descriptive Statistics dialog box to get summary statistics for the data and to create a 90% confidence interval.

2. Retrieve the worksheet Svls01.xls from the Excel data disk. This worksheet contains the number of shares of Disney Stock (in hundreds of shares) sold for a random sample of sixty trading days in 1993 and 1994. Use the Descriptive Statistics dialog box to get summary statistics for the data and to create the following confidence intervals.
   (a) Find a 99% confidence interval for the population mean volume.
   (b) Find a 95% confidence interval for the population mean volume.
   (c) Find a 90% confidence interval for the population mean volume.
   (d) Find an 85% confidence interval for the population mean volume.
   (e) What do you notice about the lengths of the intervals as the confidence level decreases?

3. There are many types of errors that will cause a computer program to terminate or give incorrect results. One type of error is punctuation. For instance, if a comma is inserted in the wrong place, the program might not run. A study of programs written by students in a beginning programming course showed that 75 out of 300 errors selected at random were punctuation errors. Find a 99% confidence interval for the proportion of errors made by beginning programming students that are punctuation errors. Next find a 90% confidence interval. Use the CONFIDENCE command to find the interval. Is this interval longer or shorter?