9.3: Confidence Intervals

Up to this point, we have learned how to estimate the population parameter for the mean using sample data and a sample statistic. From one point of view, this makes sense: we have one value for our parameter so we use a single value (called a point estimate) to estimate it. However, we have seen that all statistics have sampling error and that the value we find for the sample mean will bounce around based on the people in our sample, simply due to random chance. Thinking about estimation from this perspective, it would make more sense to take that error into account rather than relying just on our point estimate. To do this, we calculate what is known as a confidence interval.

A confidence interval starts with our point estimate then creates a range of scores considered plausible based on our standard deviation, our sample size, and the level of confidence with which we would like to estimate the parameter. This range, which extends equally in both directions away from the point estimate, is called the margin of error. We calculate the margin of error by multiplying our two-tailed critical value by our standard error:

$$\text{Margin of Error } = t^{*}(s/\sqrt{n}) \tag{9.3.1}$$

One important consideration when calculating the margin of error is that it can only be calculated using the critical value for a two-tailed test. This is because the margin of error moves away from the point estimate in both directions, so a one-tailed value does not make sense.

The critical value we use will be based on a chosen level of confidence, which is equal to 1 – α. Thus, a 95% level of confidence corresponds to α = 0.05. Thus, at the 0.05 level of significance, we create a 95% Confidence Interval. How to interpret that is discussed further on.

Once we have our margin of error calculated, we add it to our point estimate for the mean to get an upper bound to the confidence interval and subtract it from the point estimate for the mean to get a lower bound for the confidence interval:

$$\begin{array}{l}\text{Upper Bound} = \bar{X} + \text{Margin of Error} \\ \text{Lower Bound } = \bar{X} - \text{Margin of Error}\end{array} \tag{9.3.2}$$

Or simply:

$$\text{Confidence Interval } = \overline{X} \pm t^{*}(s/\sqrt{n}) \tag{9.3.3}$$

To write out a confidence interval, we always use soft brackets and put the lower bound, a comma, and the upper bound:

$$\text{Confidence Interval} = \text{ (Lower Bound, Upper Bound)} \tag{9.3.4}$$

Let’s see what this looks like with some actual numbers by taking our oil change data and using it to create a 95% confidence interval estimating the average length of time it takes at the new mechanic. We already found that our average was $\overline{X}$= 53.75 and our standard error was $s_{\overline{X}}$ = 6.86. We also found a critical value to test our hypothesis, but remember that we were testing a one-tailed hypothesis, so that critical value won’t work. To see why that is, look at the column headers on the t-table. The column for one-tailed α = 0.05 is the same as a two-tailed α = 0.10. If we used the old critical value, we’d actually be creating a 90% confidence interval (1.00-0.10 = 0.90, or 90%). To find the correct value, we use the column for two-tailed α = 0.05 and, again, the row for 3 degrees of freedom, to find $t*$ = 3.182.

Now we have all the pieces we need to construct our confidence interval:

$$95\% CI = 53.75 \pm 3.182(6.86)$$

$$\begin{aligned}\text{Upper Bound } &= 53.75 + 3.182(6.86) \\ UB &= 53.75 + 21.83 \\ UB &= 75.58\end{aligned}$$

$$\begin{aligned}\text{Lower Bound } &= 53.75 - 3.182(6.86) \\ LB &= 53.75 - 21.83 \\ LB &= 31.92\end{aligned}$$

$$95\% CI = (31.92, 75.58)$$

So we find that our 95% confidence interval runs from 31.92 minutes to 75.58 minutes, but what does that actually mean? The range (31.92, 75.58) represents values of the mean that we consider reasonable or plausible based on our observed data. It includes our point estimate of the mean, $\overline{X}$= 53.75, in the center, but it also has a range of values that could also have been the case based on what we know about how much these scores vary (i.e. our standard error).

It is very tempting to also interpret this interval by saying that we are 95% confident that the true population mean falls within the range (31.92, 75.58), but this is not true. The reason it is not true is that phrasing our interpretation this way suggests that we have firmly established an interval and the population mean does or does not fall into it, suggesting that our interval is firm and the population mean will move around. However, the population mean is an absolute that does not change; it is our interval that will vary from data collection to data collection, even taking into account our standard error. The correct interpretation, then, is that we are 95% confident that the range (31.92, 75.58) brackets the true population mean. This is a very subtle difference, but it is an important one.
