

8.3: 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:

$$\text{Margin of Error} = t^*(s/\sqrt{n}) \quad (8.3.1)$$

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 $1 - \alpha$. Thus, a 95% level of confidence corresponds to $\alpha = 0.05$. 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:

$$\begin{aligned} \text{Upper Bound} &= \bar{X} + \text{Margin of Error} \\ \text{Lower Bound} &= \bar{X} - \text{Margin of Error} \end{aligned} \quad (8.3.2)$$

Or simply:

$$\text{Confidence Interval} = \bar{X} \pm t^*(s/\sqrt{n}) \quad (8.3.3)$$

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

$$\text{Confidence Interval} = (\text{Lower Bound}, \text{Upper Bound}) \quad (8.3.4)$$

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 $\bar{X} = 53.75$ and our standard error was $s_{\bar{X}} = 6.86$. 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 $\alpha = 0.05$ is the same as a two-tailed $\alpha = 0.10$. 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 $\alpha = 0.05$ 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:

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

$$\text{Upper Bound} = 53.75 + 3.182(6.86)$$

$$UB = 53.75 + 21.83$$

$$UB = 75.58$$

$$\text{Lower Bound} = 53.75 - 3.182(6.86)$$

$$LB = 53.75 - 21.83$$

$$LB = 31.92$$

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

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, $\bar{X} = 53.75$, 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.

Hypothesis Testing with Confidence Intervals

As a function of how they are constructed, we can also use confidence intervals to test hypotheses. However, we are limited to testing two-tailed hypotheses only, because of how the intervals work, as discussed above.

Once a confidence interval has been constructed, using it to test a hypothesis is simple. The range of the confidence interval brackets (or contains, or is around) the null hypothesis value, we fail to reject the null hypothesis. If it does not bracket the null hypothesis value (i.e. if the entire range is above the null hypothesis value or below it), we reject the null hypothesis. The reason for this is clear if we think about what a confidence interval represents. Remember: a confidence interval is a range of values that we consider reasonable or plausible based on our data. Thus, if the null hypothesis value is in that range, then it is a value that is plausible based on our observations. If the null hypothesis is plausible, then we have no reason to reject it. Thus, if our confidence interval brackets the null hypothesis value, thereby making it a reasonable or plausible value based on our observed data, then we have no evidence against the null hypothesis and fail to reject it. However, if we build a confidence interval of reasonable values based on our observations and it does not contain the null hypothesis value, then we have no empirical (observed) reason to believe the null hypothesis value and therefore reject the null hypothesis.

Let's see an example. You hear that the national average on a measure of friendliness is 38 points. You want to know if people in your community are more or less friendly than people nationwide, so you collect data from 30 random people in town to look for a difference. We'll follow the same four step hypothesis testing procedure as before.

Step 1: State the Hypotheses We will start by laying out our null and alternative hypotheses:

H_0 : There is no difference in how friendly the local community is compared to the national average

$$H_0 : \mu = 38$$

H_A : There is a difference in how friendly the local community is compared to the national average

$$H_A : \mu \neq 38$$

Step 2: Find the Critical Values We need our critical values in order to determine the width of our margin of error. We will assume a significance level of $\alpha = 0.05$ (which will give us a 95% CI). From the t -table, a two-tailed critical value at $\alpha = 0.05$ with 29 degrees of freedom ($N - 1 = 30 - 1 = 29$) is $t^* = 2.045$.

Step 3: Calculations Now we can construct our confidence interval. After we collect our data, we find that the average person in our community scored 39.85, or $\bar{X} = 39.85$, and our standard deviation was $s = 5.61$. First, we need to use this standard deviation, plus our sample size of $N = 30$, to calculate our standard error:

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{5.61}{5.48} = 1.02$$

Now we can put that value, our point estimate for the sample mean, and our critical value from step 2 into the formula for a confidence interval:

$$95\%CI = 39.85 \pm 2.045(1.02)$$

$$\text{Upper Bound} = 39.85 + 2.045(1.02)$$

$$UB = 39.85 + 2.09$$

$$UB = 41.94$$

$$\text{Lower Bound} = 39.85 - 2.045(1.02)$$

$$LB = 39.85 - 2.09$$

$$LB = 37.76$$

$$95\%CI = (37.76, 41.94)$$

Step 4: Make the Decision Finally, we can compare our confidence interval to our null hypothesis value. The null value of 38 is higher than our lower bound of 37.76 and lower than our upper bound of 41.94. Thus, the confidence interval brackets our null hypothesis value, and we fail to reject the null hypothesis:

Fail to Reject H_0 . Based on our sample of 30 people, our community not different in average friendliness ($\bar{X} = 39.85$) than the nation as a whole, 95% CI = (37.76, 41.94).

Note that we don't report a test statistic or p -value because that is not how we tested the hypothesis, but we do report the value we found for our confidence interval.

An important characteristic of hypothesis testing is that both methods will always give you the same result. That is because both are based on the standard error and critical values in their calculations. To check this, we can calculate a t -statistic for the example above and find it to be $t = 1.81$, which is smaller than our critical value of 2.045 and fails to reject the null hypothesis.

Confidence Intervals using z Confidence intervals can also be constructed using z -score criteria, if one knows the population standard deviation. The format, calculations, and interpretation are all exactly the same, only replacing t^* with z^* and $s_{\bar{X}}$ with $\sigma_{\bar{X}}$.

This page titled [8.3: Confidence Intervals](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Foster et al.](#) ([University of Missouri's Affordable and Open Access Educational Resources Initiative](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.