Print

Constructing Confidence Intervals

For the impurity example, the 95% confidence interval for the mean is 5.82 to 6.42. This interval provides a range of likely values for the true mean of impurity, based on 100 observations and the point estimate of 6.12.

But, how is this confidence interval constructed? A confidence interval starts with the point estimate. In this case, the point estimate is the sample mean. Then you add a margin of error around the point estimate.

To construct the margin of error, you need three values: the standard error, the confidence level ($1-\alpha$), and a critical value from the t distribution.

The standard error is an estimate of the standard deviation of sample means. As you have learned, this is calculated from the sample data.

The confidence level is specified by you, the analyst. It's typically set at 95%, but you can change this. For any confidence interval, there is some risk that the interval doesn't capture the true parameter value. This risk is denoted as alpha (α). For a 95% confidence interval, the alpha risk is 5%.

If you are willing to accept only a 1% risk that the confidence interval doesn't capture the true parameter value, you'd construct a 99% confidence interval. Likewise, if you are willing to accept an alpha risk of 10%, you'd construct a 90% confidence interval. We explore the confidence level and alpha risk in the next video.

The third value that you need, in order to construct the margin of error, is a value from the t distribution, or, more formally, the Student's t-distribution.

Earlier, you learned that sample means are approximately normally distributed. It turns out that sample means actually follow the t-distribution instead of the normal distribution.

The t-distribution is a sampling distribution that is centered at zero, like the standard normal distribution, and it has a mounded shape. But, unlike the normal distribution, its shape changes slightly based on the sample size.

This is because the standard deviation of sample means is estimated using sample data. And, as you have seen, the spread of the distribution of sample means is based largely on the sample size.

If the sample size is small, the t-distribution is flatter, and more spread out. As the sample size increases, the t-distribution approaches the normal distribution. Here, df refers to the degrees of freedom. (See the Exploratory Data Analysis module for a discussion of degrees of freedom.)

For the t-distribution, the degrees of freedom is n (the sample size) minus 1. To construct a 95% confidence interval for the mean, we need to find values of the t-distribution that encompass the central 95% of the distribution.

To do this, we use the Distribution Calculator in JMP. You can launch this script from the Help menu under Sample Data, Teaching Scripts, then Interactive Teaching Modules.

We'll find the t value for a 95% confidence interval for the mean for the Impurity data. We change the distribution to t, and change the degrees of freedom to n-1, or 99. We'll input the probability, find the central probability, and change the probability to 0.95. The t value, used to construct the confidence interval, is 1.98.

Let's put this all together, and calculate the interval using the formula you saw earlier. For the Impurity data, the sample mean is 6.12, the standard error is 0.15, and the t value is 1.98.

First, you calculate the margin of error. The confidence interval, then, is 5.82 to 6.42. Checking our work, you can see that this matches the confidence interval reported by JMP.

Of course, we rely on software to do the work for us, and there's usually no need to look up the t value or to calculate these values yourself. But it helps to know where these values come from and how these intervals are constructed.

Close