6 CONFIDENCE INTERVALS

6.05 Confidence levels

The 95 percent confidence interval tells us that we can be 95 percent confident that our point estimate (which could be a mean or a proportion) falls within our confidence interval. Or, in other words, it tells us that if we would draw an infinite number of samples similar to our actual sample, and for every sample we would compute a 95 percent confidence interval with a similar margin of error, in 95 percent of the cases the population value would fall within this confidence interval. This, of course, also means that in 5 percent of the cases this method will produce an interval that does *not* contain the actual population parameter. If you would like to reduce the chance of an incorrect inference, you could go for a larger confidence level, such as, for instance, 99 percent. In this video I will tell you how you can change your confidence level and what the consequences are of doing so.

Imagine you asked a sample of 100 new parents if their babies like to answer nature's call during the diaper-changing process. 17 percent reported that this is the case. Our sample proportion p thus equals 0.17. The formula to compute the 95% confidence interval for a proportion is: p plus and minus the z-score for the 95% confidence level times the standard error, which equals the square root of p multiplied with one minus p, divided by n. You can look up the z-score for a 95% confidence level in the z-table. Look at *this* standard normal distribution here. When you have a 0.95 probability that your value falls within z standard errors from the mean, that means that 0.025 probability falls in the two tails. If we look up the z-scores which are displayed here in the z-table, we find values of plus and minus 1.96. You can see that *here*. We can now easily compute the interval. That's 0.17 plus and minus 1.96 times the standard error, which is the square root of 0.17 times 0.83 divided by 100. This leads to a confidence interval with the endpoints 0.10 and 0.24.

You can now imagine that it is not so difficult to construct intervals with other confidence levels. Let's first look at the 99% confidence interval. This is the formula: p plus and minus the z-score for the 99% confidence level times the standard error. The only difference is the different z-score. Look at this standard normal distribution. When you have a 0.99 probability that your value falls within z standard errors from the mean, that means that 0.005 probability falls in the two tails. If we look up the z-scores which are indicated here, we find values of plus and minus 2.58. You can see that *here*. We can now compute the interval. That's 0.17 plus and minus 2.58 times the standard error, which was 0.038. This leads to a confidence interval with the endpoints 0.07 and 0.27. For the 90% confidence level, we find a z-score of 1.645. This leads to a confidence interval of 0.17 plus and minus 1.645 times 0.038. That makes a confidence interval with the endpoints of 0.11 and 0.23.

I have here displayed the confidence intervals graphically. You can see that a higher confidence level leads to a wider confidence interval. In other words, the more confident we are that we draw a correct inference, the larger our margin of error. That means that we have to compromise between confidence and precision. As one gets better, the other gets worse. We never settle for a 100 percent confidence interval because the margin of error is then far too large, which means that our conclusions are not very informative. In most cases the 95 percent confidence level is used.

We can also use other confidence levels when we construct a confidence interval to estimate a population *mean*. Suppose we have asked a sample of 30 new parents in Amsterdam how much hours of sleep they have lost after their first child was born. The mean is 2.6 hours per night and the standard deviation is 0.9 hours per night. *This* is the formula we use to construct our 95% confidence interval: X-bar plus and minus the t-score for the 95% confidence level times the standard error, which equals the sample standard deviation divided by the square root of the sample size. Now, what is the t-score for the 95% confidence level? That's dependent on the degrees of freedom, which equals n minus 1. That is 30 minus 1 is 29. In the t-table we should look in the column of the 95% percent confidence level and the row of 29 degrees of freedom. That gives a t-score of 2.045. The confidence interval becomes 2.6 plus and minus 2.045 times 0.9 divided by the square root of 30. That gives an interval from 2.26 to 2.94.

If we would want to construct an interval with a confidence level of 99% we simply replace the t-score for the 95%-level with the t-score for the 99%-level. You can look it up in the table, it's 2.756. The confidence interval is 2.6 plus and minus 2.756 times 0.9 divided by the square root of 30. That leads to an interval from 2.15 to 3.05. You can also easily do that for other confidence levels.

Let me conclude this video by giving you a step-by-step plan for constructing a confidence interval. First, decide which confidence level you want to use. For instance, do you settle for the regular 95 percent level, or do you want to be more confident and less precise or more precise but less confident? Second, decide if you're dealing with a proportion or a mean. If you're interested in a proportion you work with the z-distribution and if you're interested in a mean you have to use the t-distribution. So, in the case of a proportion you look up the relevant z-score, and in the case of a mean you look up the relevant t-score. Don't forget that if you're interested in a mean, you should also compute the degrees of freedom, which is equal to n minus 1. Fourth, compute the endpoints of the confidence interval. And finally, interpret the results substantively.

That's it! If you're not 95 percent confident now that you can now construct a confidence interval yourself, re-watch the last couple of videos!