6 CONFIDENCE INTERVALS

6.04 CI for proportion

In the last couple of weeks, I have learned that newborn baby's like to poo. They like to poo a lot. And it happens to be the case that my daughter Lowis especially likes to poo under specific circumstances. That is, *exactly* when I'm changing her diaper. Seriously, once she decided to answer nature's call six times, I repeat, *six times* during the diaper-changing process… Is that normal? I'm not sure. But I know how I could check it. Suppose that I asked a simple random sample of 100 new parents if their baby likes to defecate while his or her diaper is being changed. In this video I'll tell you how we can, on the basis of such a study, construct a confidence interval to estimate a population proportion.

Say that 17 percent of my hundred respondents report that their baby's like to poo while the diaper is being changed. 83 percent report that their baby's don't do that. We thus have a proportion of 0.17 whose baby's like to poo while their diaper is being changed. When we construct a confidence interval for a proportion, we employ the sampling distribution of the sample proportion. We know that, as long as our sample is large enough, this sampling distribution is normally distributed with a mean that is equal to the population proportion π and a standard deviation that is equal to the square root of π multiplied with one minus π, divided by n. We also know that the probability of finding a sample proportion of less than about 2 standard deviations from the mean (which is the population proportion) is 0.95. More precisely, if we look up the z-score which corresponds to this probability, we'll find a value of 1.96. This means that we have a 95% chance that our sample proportion will fall within 1.96 standard deviations of our population proportion. This is what we call the margin of error.

The formula with which we can compute the 95% confidence interval looks like this: p plus and minus 1.96 times the standard deviation of the sampling distribution of the sample proportion. 1.96 is the z-score that corresponds to the 95% confidence level, so we could also write: p plus and minus the z-score for the 95% confidence level times the standard deviation of the sampling distribution of the sample proportion. We're talking about the 95 *percent* confidence interval here. That means that we can say that if we would draw an infinite number of samples from our population, in 95% of the cases our confidence interval would contain population proportion π.

However, as you might have noticed, we don't know the value of population proportion π, so it is impossible to compute the standard deviation of the sampling distribution of the sample proportion. We therefore substitute the population parameter, π, with an estimate, and this estimate is our sample statistic, p. This leads to the following formula: p plus and minus the z-score for the 95% confidence level times the *estimated* standard deviation of the sampling distribution of the sample proportion. Just like when we construct a confidence interval for a mean, we call this estimated standard deviation of the sampling distribution the standard error.

In contrast with the confidence interval for a mean, when it comes to constructing a confidence interval for a proportion, we don't make use of the t-distribution, and just stick with the standard normal distribution. However, your data need to satisfy one essential assumption: you should have at least fifteen successes and fifteen failures. In other words, n times p and n times (one minus p) need to be larger than or equal to fifteen. If this is not the case, you cannot compute a confidence interval on the basis of the discussed formula.

Okay, let's go back to our example. We have a proportion of 0.17 that reports that the baby poos while the diaper is being changed. *This* is the formula we use. Let's first compute the standard error:

it's the square root of 0.17 times 0.83 divided by 100. That makes about 0.038. The margin of error then is 1.96 times 0.038. That's about 0.07. 0.17 minus 0.07 equals 0.10 and 0.17 plus 0.07 equals 0.24. So, our confidence interval ranges from 0.10 to 0.24. This means that we can be 95% confident that the population proportion falls between 0.10 and 0.24. Or, in other words, if we would draw an infinite number of samples with n equals 100 from our population, and for every sample we would compute the confidence intervals with this margin of error, in 95% of the cases the population value would fall within the confidence interval. This 95 percent confidence interval demonstrates that most babies don't like to poo while their diaper is being changed. On the other hand it's not *that* exceptional if they do poo. We can be 95 percent confident that between 10 and 24 percent does poo during the changing process.

Thank god. Nothing's wrong with my little daughter Lowis. In fact, my theory is that my daughter's pooing habits are good news. They might well indicate that she will be toilet-trained soon. After all, she doesn't like to poo in a diaper… Now, let's hope that later on, she *does* like to do it in the toilet…