6 CONFIDENCE INTERVALS

6.07 Example: confidence interval for mean

Warning: this is a pretty filthy video. So for those of you who can't handle that, it might be a good idea to skip it… Okay. Five weeks ago, my daughter Lowis was born. A couple of days after her birth, I was changing her diaper. Just after I removed her old one, and *before* I put on a new one, I saw her gaze become blurry. One second later I heard a strange and loud sound coming from her bottom. Oh no…. But there was nothing. Strange. A couple of hours later, when I was changing her diaper again, by coincidence I looked at the wall next to the place where the diaper-changing process takes place. And suddenly I realized what had happened before. My daughter had re-painted the freshly painted wall with a couple of brown streaks. Apparently, the defecation process of little babies can be pretty explosive.

Of course, this made me curious about other babies. How far are babies able to spurt their poo? What is their maximum shooting distance? Suppose I asked a simple random sample of 150 new parents to measure how far their babies poo and to report to me their personal records. Let's say that it turns out that the mean pooing distance is 28 centimeters. The standard deviation is 21 centimeters. I would like to construct a 90 percent confidence interval to estimate the population mean. Let's first take a look at the step-by-step plan for constructing a confidence interval. This is it. We already know the confidence level (90%) and that we are dealing with a mean instead of a proportion. So, the next step is to look up the relevant t-score. To do that, we should first know the degrees of freedom. That's n minus 1 equals 150 minus 1 is 149. So, in the t-table we look in the column of the 90% confidence level and in the row of 100 df. After all, 149 is not in the table and 100 is the closest lower value. The relevant t-score is 1.66.

We can now complete the formula. The lower endpoint is 28 minus 1.66 times 21 divided by the square root of 150. That makes about 25.2. The higher endpoint is 28 plus 1.66 times 21 divided by the square root of 150. That is about 30.8. The 90% confidence interval thus ranges from 25.2 to 30.8. We can be 90% confident that this interval contains the actual population mean. More precisely, if we would draw an infinite number of samples with n equals 150 from our population, and for every sample we would compute the confidence interval with this margin of error, in 90% of the cases, the population value would fall within the confidence interval.

Now imagine I was not interested in the population mean of the pooing distance, but in the percentage of babies that have a 'personal poo-spurting distance record' of over 50 centimeters. Suppose that 18 percent of the babies spurts further than 50 centimeters, and that I would like to construct a 97% confidence interval. Now we're not interested in a population *mean* anymore, but in a population *proportion*. So *this* is the formula we're working with. First we have to look up the relevant z-score in the z-table. That we employ the 97% confidence level means that we're interested in *these* z-scores. That means that we have to look up in the z-table the scores that correspond to a cumulative probability of respectively 0.015 and 0.985. We find them here and here. That's minus 2.17 and 2.17.

So, if we go back to the formula, that makes 0.18 plus and minus 2.17 times the square root of 0.18 times 1 minus 0.18, divided by 150. That leads to an interval from 0.11 to 0.25. We can be 97% confident that this interval contains the actual population proportion of babies that can spurt further than fifty centimeters. So, my advice to new parents: *never* paint the walls of your baby's room before your son or daughter is toilet-trained…