6 CONFIDENCE INTERVALS

6.02 CI for mean with known population standard deviation

New parents like me usually lose a lot of sleep. In fact, often new parents become very much like nursing zombies. But despite my 'zombieness', I still want to know the numbers! To what extent do other new parents also lose sleep? Suppose that to answer that question I asked a sample of 60 new parents how much sleeping hours they lost after their first child was born. Now, imagine that the mean lost sleeping hours per night is 2.6 hours. The standard deviation in the sample is 0.9 hours. Suppose you also know the standard deviation in the population. It is 1.1 hours. In practice it is very unlikely that you'll know the value of this parameter, but for now just suppose you do.

In this video I will show you how you can construct a confidence interval based on the information from your sample and the population standard deviation. I'll explain how such a confidence interval should be interpreted, and I will also show you why being a new dad could have disastrous consequences for one's ability to… Focus!

To construct a confidence interval we make use of the sampling distribution of the mean. We're dealing, after all, with a sample from a population. We know that, as long as our sample is sufficiently large, the sampling distribution is normally distributed with a mean that is equal to the population mean Mu and a standard deviation that is equal to the population standard deviation divided by the square root of *n*. We also know that the probability of finding a sample mean of less than about 2 standard deviations from the mean is 0.95. More precisely, if we look up the z-scores which correspond to this probability, we'll find values of minus 1.96 and 1.96. This means that we have a 95% chance that our sample mean will fall within 1.96 standard deviations of population mean Mu.

This distance of 1.96 standard deviations is what we call the **margin of error**. The margin of error tells us how accurately our sample mean X-bar is likely to estimate our population mean Mu. Now, the formula of the 95% confidence interval is the following. It is the point estimate (or the sample mean) plus and minus the margin of error (which equals 1.96 standard deviations). Note that we're dealing with the sampling distribution of the sample mean here, so the standard deviation equals Sigma divided by the square root of *n*.

Now, pay close attention because this is a little complicated. Suppose you draw a sample. The mean of this sample is represented by *this* dot. The lines here represent the margins of error on both sides of the mean. Together they form the 95% confidence interval. If the sample mean falls within the red area, than the confidence interval contains the population mean Mu. That's the case *here*. However, if the sample mean does not fall within the red area, the confidence interval does *not* contain population parameter Mu. That's the case *here*. We're talking about the 95% confidence interval. That means that the probability that the confidence interval of a randomly selected sample contains the population parameter is 0.95. The probability that it does not contain the population mean is 0.05. In other words, if we would draw an infinite number of samples from our population, in 95% of the cases our confidence interval would contain population mean Mu.

Let's go back to our example. The sample mean of lost sleeping hours is 2.6. The population standard deviation is 1.1. The sample size is 60. We now have all the necessary ingredients to compute our confidence interval. This is the formula: X-bar plus and minus Sigma-X-bar. We know that Sigma-X-bar equals Sigma divided by the square root of n. Let's first compute Sigma-X-bar. That's 1.1 divided by the square root of 60. That makes 0.142. Now we compute the margin of error. That's 1.96 times 0.142. That's about 0.28. The sample mean equals 2.6. So the 95% confidence interval is the interval between 2.6 minus 0.28 and 2.6 plus 0.28 – which is the interval between 2.32 and 2.88. We say that we have 95 percent confidence that this interval contains the actual population mean. More precisely, if we would draw an infinite number of samples with n equals 60 from our population, and for every sample we would compute the confidence intervals with this margin of error, in 95 percent of the cases, the population value would fall within the confidence interval.

Okay, if this situation of new parents would last for a year, we would have 95 percent confidence that they would lose between 2.32 times 365 and 2.88 times 365 hours of sleep in that year. That's between 846.8 hours and 1051.2 hours. Or, in other words, between 35.3 and 43.8 full days….