6 CONFIDENCE INTERVALS

6.01 Statistical inference

Sorry, I'm a little confused. The thing is, me and my girlfriend got a baby a couple of weeks ago and I don't sleep that much anymore. Because I'm a little tired, or, a better word might be, exhausted, it could be the case that I'm a little less focused than I usually am… Let me before I start be a little more concrete about what's going on. I usually slept about eight hours every night. Now, I only sleep about five hours per night. That means a reduction of three hours per night. That's about 20 hours per week and about 1000 hours per year. That's about *40 days*. In other words, if my daughter continues with her present sleeping schedule, after a year I will have lost about forty full days of sleep!! Just to let you know…

But, sleep or no sleep, I'm still a scientist. And I was wondering how much sleep other new parents in my home city of Amsterdam lost after they had their first baby. Are me and my girlfriend the only ones losing so much sleep? Suppose I contacted the local authorities and asked them the contact details of all new parents in Amsterdam. Let's say that new parents are those who got a baby within the last six months. I drew a simple random sample of 60 new parents and asked them how much hours per night they slept less than before they had a baby.

In this module I will talk about **statistical inference**. I will, on the basis of sample information, draw conclusions about the entire population from which the sample was drawn. We can distinguish two types of statistical inference methods. We can **estimate** population parameters and we can **test hypotheses** about these parameters. In this module I will talk about the first type of inferential statistics. The second type, hypothesis testing, will be discussed in the next module. There are two ways in which we can estimate the value of a population parameter. The first one is the so-called **point estimate**. It is a single number that is our best guess for the population parameter. The second one is the **interval estimate**. It is a range of values within which we expect the parameter to fall.

Let's assume that the mean number of hours that the 60 respondents in my sample slept less after they had their first baby is 2.6. That means that a good point estimate for the mean number of lost sleeping-hours in the population is, well, 2.6. In other words, the statistic X-bar (which in our case is 2.6 hours) is a good point estimate for the parameter Mu. However, one individual point estimate doesn't tell us if this estimate is close to the population parameter we are interested in or not. Therefore, next to a point estimate, researchers often also want to know the likely *precision* of this point estimate.

They show this likely precision by also computing an interval estimate. An interval estimate is a range of numbers, which, most likely, contains the actual population value. On the basis of our sample mean of 2.6 hours we might predict, for instance, that the mean lost sleeping-hours of all new parents in Amsterdam lies somewhere between 2.3 and 2.9. The *probability* that the interval contains the population value is what we call the **confidence level**. The confidence level always has a value close to one. In most cases it's 0.95. In that case we talk about the **95 percent confidence interval**. In the next videos we'll discuss how we can construct such confidence intervals.