

module_stat4_lecture8

📅 Thu, 2/17 4:18PM ⌚ 11:42

SUMMARY KEYWORDS

confidence interval, standard error, sample size, square root, sample, population, estimate, compute, construct, point, remember, standard deviation, error, theorem, data, quantitative variable, clip, confidence, draw, talk

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:05

So I introduced you to the idea of confidence intervals and how to construct them. And the basic idea came from the central limit theorem, and the properties of normal distribution. So now I'll show you how to actually compute confidence intervals in a couple of examples. So let's go back to our example with the sleep data on teenagers how much they sleep on how much sleep they get on a weeknight. And this was our data from a sample of size six. And remember, earlier, we computed the point estimate of this, which was just the mean of these numbers. So that was nine, right. And we also computed the standard error, which was s over square root of six, that was .45. And now we're going to construct the confidence interval for the system. So remember, 95% confidence interval is given by this formula, which is $\bar{x}$ minus two times the standard error. That's the lower end, and the upper end, it's $\bar{x}$ plus two times the standard error. Right? So in this case, so this will be nine minus two times this, so two times point 45, that's point nine. And at the upper end, it will be nine plus two times the standard error. So that means the confidence interval for this is 8.1, to 9.9. So the confidence interval for this estimate of nine. So this is how you construct confidence interval. One bit of warning here is see the central limit theorem only applies when n goes to infinity. And as I said, there, you typically are good to apply it if n is greater than or equal to 30. So in this case, our sample size is only six. Right? So strictly speaking, I shouldn't be using these values to construct the confidence interval. But I just wanted to show you in a simple case, by hand how you could compute. But more generally, if you have a big large sample size with 30, or 100 data points, would you would use something like Excel, what I showed you in the previous clip, how to construct the mean, how to get the standard error, which is essentially the standard deviation of this divided by square root of n , right. And once you have that, then just use this to figure out the lower and upper point of the confidence. Now, now that you've got a hang of confidence intervals, so let me talk a little bit about its properties. Right? So see the confidence interval, it's centered around the mean, right? The $\bar{x}$, and it's plus minus two s square root of it. Right? So it's the same distance below as, as, as the distance of right, it's symmetric around that $\bar{x}$. Now, a couple of things to note here, right. So, see, as S goes up, remember what is S , S is the standard deviation of the sample, right. So, that's the sample standard deviation. So, if S becomes bigger, if S goes up, that means this moves further to the left and this moves further to the right. So that means your confidence intervals becomes bigger right. So, instead of a

tight confidence interval like this, it becomes more if S goes up right then that means this moves here and this moves here, right. So, in that case, to say that you have 95% confidence, you have to expand that into the second thing is this that see as your sample size n goes up, right, this s over square root of n this false because n is in the denominator of this so what that how would that means is that this upper point, it moves this way, the lower point it moves this way. Right? What happens is that means the confidence interval from here, this comes in, this comes in, so it becomes a much tighter confidence interval. So, in that case, your margin of error is smaller. Right. So that's another way of saying this, that your margin of error this shrinks as your sample size goes up. Now, what does this margin of error mean? Let me again, reiterate what it means. So see, remember, this is your population. Right? Now, you're drawing a random sample, this could be drawn from here, or it could be drawn from here, it could be drawn from here, it could be drawn from here. Right? Depending on if you're drawing this sample, you get a different $\bar{x}$. If you draw this one, you get a different $\bar{x}$. If you draw this one, you'll get another different $\bar{x}$, if you draw this one, you'll get a different export. Right? So none of these will exactly match up with the population parameter μ . Just because you're not drawing the whole population, you're drawing just a part of it. And depending on which part you draw, you could get one answer, depending on another part, if you draw, you'll get a different answer. Right. So none of these will match up, that this is not going to match up with me, right? So these two are never going to exactly match up unless you're tremendously lucky. So this sort of an error, which is induced by the fact that you're not looking at the entire population, but just by looking at a part of it. So that is called sampling. It it's not that you're doing a mistake deliberately, that it's coming from the fact that see, you cannot ask everyone in the population about this. So you're only asking a small segment of it. Right? And that's what introducing this error? And what statistics gives you is a handle on this, what is this margin of error? What is it influenced by? And how can we incorporate it into our confidence levels? About the answer that we get? Now, let me end this clip by talking about so far of the example that I've done is have a quantitative variable, right. Now, what if the underlying variable is categorical? And we're interested in figuring out what fraction of the population smoke or do this do that? Right? So remember, earlier, we, we talked about this, the point estimate of the fraction is the pot is the proportion of your sample who do who smoke or who do this or support policy or policy be. And so in deriving the confidence interval, we use the same formula, it's p minus two s over square root of n , that will be the bottom point. And the top point will be p plus two over square root of n , the only thing that will change is that now where the standard deviation, s is given by square root of P times one minus p . We also saw this earlier when we were talking about standard errors for this sort of estimates, right? So it's nothing new. I'm just reminding you of that. So let's do an example like this. So remember, the example we had done earlier, where we had a sample size 10. And whether people supported policy or not. So in this data, our point estimate for the fraction of the population whose support policy is point four, that was there were only these four yeses, four yeses, right? And the standard error that we got was point 155. And we had computed this earlier, right. Now we want to look at the confidence interval for this. So what we're going to use is again, this formula here, right? So that's going to be point four minus two times the standard error, which is point 2145. And point four plus the same thing. And what this becomes is 0.09 to 0.7. So this is my confidence interval for this estimate. So in this clip I showed you how to derive confidence intervals, in particular examples, whether it's a quantitative variable, or whether the underlying variable is categorical. A couple of warnings about both of these examples that we did well, in one, the sample size was six, the other the sample size was 10. So typically, you can use this way, the central limit theorem and this way of deriving confidence interval, so long as the sample size is greater than equal to 30. So in none of our examples, technically, should we be using this

procedure? But we still did because I just wanted to show you in a small sample size, how to compute these things by hand. So let's stop here. And in the next clip, I'll talk a little bit more about confidence interval. What does it mean other levels of confidence under the 95%