

module_stat4_lecture9

📅 Thu, 2/17 4:18PM ⌚ 15:37

SUMMARY KEYWORDS

confidence interval, normal distribution, point, canadians, lie, sample size, confidence, interval, fraction, weight, estimate, confident, sentence, true, distribution, answer, mew, survey, talk, standard deviation

SPEAKERS

Sumon Majumdar

Sumon Majumdar 00:05

So in this segment, we have been talking about the confidence interval. And I've shown you how to construct a Confidence Interval how it looks. But I want to once again talk about interpretation of the confidence interval, because I want to make sure we're all on the same page about this important concept. So let me do that using an example. So this is the same example we've been doing before, that suppose we're looking at the support for policy a among Canadians. And suppose we estimated it, and the point estimate was point four 40%. And the 95% confidence interval is that it ranges from 0.09, which is 9% to point 71, which is 71%. Right? So what does this mean? Now, can we use this to say that 40% of Canadians support policy a? Because our point estimate was point four, this is not true. Because see, all the we are trying to find out right, what fraction of Canadian support policy just because we got p equal to 0.4? It doesn't mean that if we were to survey all Canadians, even then the answer would be point four. Right? almost surely that answer would be different from point four. Right? Unless you're extremely, extremely lucky. That from the answer that you got from your small sample of Canadians, right, that would match up with what you got if you had surveyed all Canadians, right? So this sentence is certainly not true. What if we made it a little bit, we took away the complete certainty of that sentence. And we say that it's probably true that 40% of Canadian support policy. Again, we know that this is not going to be true. Because if we were to, again, survey all the Canadians, right, it would be improbable that that value would exactly match up with what we got in the survey. So again, this sentence is also not true. Okay, what else? We can say something like this, right? We don't exactly know what fraction of Canadian support policy A. So we don't exactly know, what fraction of Canadian support policy, but we know that it is between 9% and 71%. Right? So one thing we have now acknowledged is that we don't know exactly, and we'll never know that exactly. Right. We'll never know exactly what fraction of Canadian support policy. But so this gives you a sense that we have to look at some sort of an into that. But what they say is that we know that that it will lie between 9% and 71%. Again, that's not true. Because see, this is the 95% confidence interval. So there is always some chance that the actual value could lie below this, or it could lie above this. So this particular sentence, this last one, although it's close, but it's not. So what is correct? So let's try a few other iterations, at least one other right. So what if we saw say that we don't exactly know, missing? No, here? We don't exactly know what fraction of Canadian support policy. But we know that that fraction is probably between 9% and 71%. So this is pretty close it because we

we know, our 95% confidence interval is between 9% and 71%. Right? So this sentence is accurate, but we can make it a little bit better because this just says that this is probably between those two. Right? So it's very close. So to make the accurate statement, the right statement will be that we don't exactly know in missing a know here. We don't exactly know what fraction of Canadian support policy a but we are 95% confidence that that confident that that fraction is between 9% and 71% So this has the confidence of the researcher based on his or her methods that that he or she is 95% confident that if we were to serve all the Canadians, right? It's likely that that fraction, that answer that we would get in that case, would lie between nines. So that's the meaning of a confidence interval. Let me do one more. Again, just because you encounter this confidence intervals all the time when you do data work, right. So it's important to know exactly what it means. So let me do another example. So this is an example of mean weight of Americans. And my source of data is from a National Health and Nutrition exam survey, where they survey a group of men and women and ask them about various things, including their weight, their height today, waist, and BMI and lots of things like that. So here, let me just focus on the weight. So if you look at the data from 2015, and 16, so if you look at women, so they have a sample size of 2760. So these are all which are, who are 20 years or older. So the point estimate of the rate is 77.3. Ah, so that's the point estimate, the 95% confidence interval is 75.7 to 78.9 kg. So again, what do we mean by that? So the researcher is 95%, confident that the average weight of an American woman is between 75.7 and 78 point. Another way of saying the same thing you see from this particular sample of 2760 women, right, we got the number 77.3. So this is from this sample. Now, what if we did another sample of another 2760? It's a different sample, then we 95% confidence that the answer from this sample, this different sample would also lie in this interval. Or if we did another sample like this, right, then the answer from there would also lie between these two. In fact, like if you took 20 samples, right, so 95% of 20 is 90, right? So you would be you're confident that 99 of their samples, right? The answer, the mean way that you would find would lie between 75.7 and 78 point. That's the meaning of the confidence. So since I talked about women, let me do the same data for men, right, this is for men, the sample size was 2591. And the point estimate of the weight was 89.7. And the 95% confidence interval was 87.9 to 91.5. And if you compare these two, you can see that the bottom most point of the men's estimate, right? That is still higher than the highest point of the woman's esteem. So if you compare these two, you can certainly say that men, on average, the average man's weight is going to be higher than the average woman's weight. Right? You can say that with 95% confidence. Okay, so so far, we have just talked about the 95% confidence interval, right? And that is the most common one that's used everywhere. And that's the one that people always talk about. Sometimes they don't even mention that it's the 95% confidence interval. They just say the confidence interval is this. But what about other confidence levels? Right? What if we wanted to make it even more Sure? 99% Or maybe not as much 90% So we're going to show you that you can do this are the confidence intervals, pretty similarly to what we have been doing so far. So let me quickly show you this. So this again boils down to our normal distribution. And again, this, you can go back to probability module four, and you'll see this graph, right. So that normal distribution centered around knew that. So there's the normal distribution of $\bar{x}$. And if you go minus 1.62, standard deviation, or plus 1.6, standard deviation, you'll capture 90% of the observations. On the other hand, if you're interested in seeing, where does 99% of observations lie, it's mean minus 2.6, or mean, plus 2.6. So once you know this fact, about normal distributions, now you can invert it to get the confident 99% confidence. So what do I mean by that? So what I mean by this is this, right, so the probability that $\bar{x}$ right lies between $\mu + 2.6$, standard deviation, and $\mu - 2.6$ standard deviation, that's point. And now, if I invert this inequality to get from you, right, so this means $\bar{X}$ minus 2.6, the standard deviation new will be greater than equals. And similarly, if you look here, this will give you that μ must be less than or equal to $\bar{x}$ plus two. So what that means is that the 99%

confidence interval will be $\bar{X}$ minus 2.6 times s over square root of n , right, because remember, the standard deviation is s over square root of n . And the upper point will be $\bar{x}$ plus 2.6 times s over square root of that. So this will be your 90% 99% confidence. So remember, the 95% confidence interval was $\bar{x}$ minus two s over square root of n and $\bar{x}$ plus two s over square root of n . 90% is the same thing, but instead of this to replace it by 2.6. So what you get is a wider interval, right? So you get a wider interval as compared to the 95%. This would be the 95% here would be the 99. And this again stems back to this normal distribution, right, if you want to make sure that 99% of the observations are captured, what you need is a wider interval. And similarly, if you want to look at the 90% con, the 90% confidence interval instead of 2.6 you replace that by the 1.6. Okay, and this will be so here is 99% Right, this will be 95%. Right and 90% will be in between it will be even now. But so, if you want to compromise on your precision, right, you will get a narrower interval, it but if you want a higher level of precision, if you want to have the more definitely that it's 99% Then you have to catch it get a wider and let me end this by one little note right something which will not be covering in this course, is that if the sample size is less than 30. Remember I said you can use the central limit theorem and all the things that we did when the sample size is greater than or equal to 30. If the sample size is less than 30, then instead of the normal distribution, you need to use the t distribution. So we're not covering here, but in more in depth statistics courses, you will see the t distribution but the basic concept of confidence interval will be the same. We'll be looking at the point estimate and looking at an interval around it. But instead of the numbers that we used here 2.6 1.6, things like that, for the t distribution, it will be different numbers. So let me stop this clip about confidence interval here. And the important thing to take away from this clip is what exactly does a confidence interval