

module_stat4_lecture5

📅 Thu, 2/17 4:18PM ⌚ 12:50

SUMMARY KEYWORDS

sample standard deviation, estimate, standard error, square root, quantitative variable, central limit theorem, standard deviation, squared, population variance, probability, sigma, clip, sample, variance, bar, sample size, remember, average, weeknights, distribution

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:19

So the last clip, I introduced you to the law of large numbers, and secondly, the central limit theorem. And what the central limit theorem essentially said, is that as n becomes bigger and bigger, right, $\bar{x}$, which is your sample average, the distribution of that is normal and centered around the population mean of μ . So in this clip, what we're going to investigate are implications of that. And implications which are going to be pretty useful for our purposes in statistics. So the central limit theorem said that as n goes to infinity, $\bar{x}$ is normally distributed with a mean of μ , right? And a variance of σ^2/n . So first thing, do you may be thinking, Okay, this whole says n goes to infinity. Is it useful in our case, because you're never going to have a sample of size infinite. But what people have found as that typically, as soon as n is greater than equal to 30, that's in most cases enough to apply the central limit theorem. So so long as this your sample size is greater than equal to 30. In practice, you can use the central limit theorem. So we've already talked about it's centered around the mean. Right. So now let me talk about the about the second term, right. So this is the variance. Now in this variance, you see there is an n in the denominator, and there is a σ^2 on the top. So what is that σ^2 . So it's the population variance. That's the population variance. But typically, you don't even know the population mean. So typically, you don't know the population variance either. But we're going to see how we can use something in its place to get a sense of this. So coming back, right, so the variance is σ^2/n . So the standard deviation, remember what the standard deviation is, it's the square root of the variance. So in this case, this will be the square root of σ^2/n . So the square root of σ^2 is σ . And this is $1/\sqrt{n}$. So this is the standard deviation here. So this is the standard deviation of $\bar{x}$. Now one thing to notice is that as n becomes bigger rates as your sample size increases, so what happens? See the n is on the denominator of the standard deviation. So the standard deviation decreases. So what's its implication? So let me show it to you using a diagram. Remember, this is your $\bar{x}$. And what we have here is the distribution of this $\bar{x}$. So as n becomes larger, right, as your sample size grows, so what's the impact? Right? What's the impact is that this thing moves inwards, right? So from the blue, it becomes the black PDF, but this is the PDF, right? And in the center, this goes up. So in terms of probabilities, what that means is that the probability of getting $\bar{x}$ bars in these extremes right in is where it's far away from you that those probabilities come down. And

the probability of getting things which are close to μ , that probability goes up. So you see that having a larger sample is always useful. We also saw that with the law of large numbers in the last clip, but here it highlights for you even more, this importance of having a larger sample size. So what it means is the probability of having these extreme values of $\bar{x}$ that things which are far away from what you're trying to find out those points probabilities decrease. And the probability of getting things which are close to what you're trying to find out those that probability goes up. And the precise impact is, is given by, by the impact on the standard deviation. So remember the the what is standard deviation? Standard deviation gives you the spread of a distribution. Right? So coming back to here, it gives you a sense of how far away right how much it's spread out. So as the standard deviation shrinks, what it means is that the spread becomes lower right, so the, the distribution comes closer and closer to μ . Now, so the standard deviation here is given by $\sigma / \sqrt{n}$. Now, you have the estimate, right? So your estimate is just the mean of the axes, right? So for a quantitative variable, a numeric variable, like things like the amount of loan, the amount of sleep that a person has, right, so in that case, you just look at the sample, take the mean, that's your estimate of the mean. Now, the other thing is you want to know what's the spread, right? What's the possibility that this is far apart or close to what you're actually trying to find? So that's given by the standard deviation. The problem is that we don't know the σ , which is the population variance. So what we do is that we replace the σ instead by s , which is the sample standard deviation. So this is something that Professor McKeown talked about in the previous module as well. Right. So s is the sample standard deviation. So what we're going to do is we're going to replace the σ that we don't know, by our estimate of that, which is the sample standard deviation. And if you recall, again from Professor McKeown's last module, the formula for that sample standard deviation is that look at x_i minus the mean, take the square of that, right, add all of those up, and divided by $n - 1$. So that's, um, take the square root of that, that will give you the sample standard deviation. And this is called the standard error of your estimate. So when you replace the σ by s , and divided by square root of n , that's called the standard error of your estimate. Because see, you are never going to hit μ . So when you estimate this $\bar{x}$, right, it's never going to hit this μ . Exactly. Right. Sometimes it's great to be here. Sometime it's going to be here. Sometimes it's going to be there. But there's always an error, right? You're never going to exactly find the population μ . So if you want to get a sense of how big that error is. So that's given by the standard error, and this is given by this formula here, which is $s / \sqrt{n}$. So let's, let's do an example like this. So suppose one is interested in how much sleep does an average teenager get on weekends? So maybe our researcher and this was also embedded question in one of the previous clips, right? So suppose you draw a sample of size six. So you sample six teenagers asked them, How much sleep do they get on weeknights? Right. And their answers are a 10 a day. So these are hours. So first thing if you want to estimate how much sleep an average teenager gets on weeknights, right. So first thing, the estimate is just going to be the mean of these, right? So if we add all of these up, 8 plus 8 is 16, plus 10, 26, 34, and 10 and 10. So that's 54 divided by six, so that's 9. Right? So your estimate of how much sleep this teenager is getting on average is nine. So now that we have an estimate of, of $\bar{x}$, right, the next question is what's our standard error of estimate. So to get the standard error, so we know it s , right, the sample standard deviation divided by square root of n , so that square root of six. So the main thing is how to calculate this. This s , which is the sample standard deviation. So remember, here the mean is nine. Right? So what we have to look at is the difference of 8 minus 9, then this 8 minus 9, 10 minus 9, right, so there's $x_i - \bar{x}$, and then squared those, add those up. So first thing, read, in this case, we have to eight minus nine, that's minus one, squared, that's one. Next one, eight minus nine, that's another minus one squared, that's 110 minus nine, that's another one. This is 8 minus 9, again, that's 1, 10 minus 9, that's 1. And the last one is also okay. And this divided by $n - 1$, so there are n is six here, so this is five. So what we need is this, right? So that

square root of six over five. And if you use your calculator, this is going to be 1.095. And now to calculate the standard error, this is going to be 1.095 over square root of six. And again, if you use your calculator for this, this will be point four fourths. So this is how you for a quantitative variable rate a numeric variable. This is how you calculate the estimate. That's simple. It's just the average and standard error, which is the the sample standard deviation divided by square root of it. So let's stop this clip here. And in the next clip, I'm going to show you how you can do this things in Excel. And also do this if you didn't have a quantitative variable, but it was a categorical variable. What are you...