

module_stat4_lecture7

Thu, 2/17 4:18PM 16:19

SUMMARY KEYWORDS

confidence interval, estimate, standard deviation, normal distribution, trump, clinton, biden, bar, higher, polls, interval, votes, margin, error, square root, cnn poll, lie, election, liberals, voters

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:05

So so far I've talked about estimates and estimates from the sample as standing for the population parameter, right. Now the question is, how much confidence do we have in those estimates. So that will be our next subject. And we'll try and construct what are called Confidence intervals. So reviewing what we have done so far, so we have estimates of a parameter, it's either the mean, or the proportion, p . And then we have the standard error, which is basically the standard deviation of this estimate. And what we also know, is the central limit theorem, which says that as n becomes really big, then this $\bar{X}$ follows the normal distribution with the mean μ and variance σ^2/n . Now, let's try and think back about how the normal distribution looks. And this is something we covered in probability module four in the last few lectures. And that's what we are going to use to answer this question that how much confidence do we have in this particular estimate $\bar{x}$ that we came up with. Now, let me draw back your memory to probability module four. So this is the PDF of the normal distribution. So remember, so this is the PDF of the normal distribution. And remember, according to the central limit theorem, our $\bar{x}$ our estimate is normally distributed. And it is normally distributed and centered around this population parameter μ . Now, one great property of the normal distribution is that firstly, it's, it has the highest point at the μ , right, and then it's built shaped on either side. And the second fact that we know is that if you look at so that means that most of the most of the values are close to the center. But if you look around it, you know that if you look at two standard deviations away from μ , minus two standard deviations, $\mu - 2\sigma$, plus two standard deviations, $\mu + 2\sigma$, right, that's where 95% of the observations lie. So the 95% of the observations lie in this range started from this $\mu - 2\sigma$ and this is $\mu + 2\sigma$. And this is what we're going to make use of in constructing our confidence interval y . So see the sample that we have x and the mean that we have from that $\bar{x}$, right? It could have be here, it could be here, it could be here, it could be anywhere, right, depending on where your sample is, right, but what the central limit theorem and this particular property of the of the normal distribution tells us is that there is 95% chance that it will lie in this interval, right somewhere from here, up to here. So now let's make use of this fact. So what's the fact that we have is that the probability that this $\bar{x}$ lies between $\mu - 2\sigma$ and $\mu + 2\sigma$ is point nine five $\bar{x}$ we know knew we don't, but if we look at this inequality, right, so that means $\bar{x}$ must be less than $\mu + 2\sigma$. Now, if I take this

to the other side, so this means $\bar{x}$ minus two standard deviation to μ must be greater than equal to this. So what this gives you is kind of a lower bound on μ . Similarly, if you look at this part of the inequality, you know that $\bar{x}$ must be greater than or equal to μ minus two standard deviation. Right. Now, if I take this to the other side, what that means is μ must be less than or equal to $\bar{x}$ plus two times standard deviation. So what this gives you is an upper bound on μ . So now you'll see you have an upper bound on μ , you have a lower bound on μ . So what you have done is that using your estimate, now what you've come up with is what's called a confidence interval for μ . Right? So what do you know is that from the previous slide, right, we know that μ must be between two things, right? So see here, μ must be greater than or equal to $\bar{x}$ minus two standard deviation. And from here, μ must be less than or equal to $\bar{x}$ plus two times the standard deviation. So we know that is 95% chance that μ must be between this and this. So that's our confidence interval for μ . So the confidence interval for μ is $\bar{x}$ minus two times the standard deviation. So remember, the standard deviation here, right, which we talked about earlier, is equal to $s / \sqrt{n}$. So the bottom part of the confidence interval is $\bar{x}$ minus two $s / \sqrt{n}$. And the top part is $\bar{x}$ plus two $s / \sqrt{n}$. And this here is called the 95% confidence interval for μ . So second, two things, right. So my point estimate of μ , right, if I had to pick a single number for μ , my estimate would be $\bar{x}$. But if we had to pick an interval, where I'm 95% sure, based on my calculations, right, that I'm 95% sure where μ lies. That's the confidence interval. And that is my estimate, which is $\bar{x}$, right? Plus minus two times $s / \sqrt{n}$, right? So see, it's the bottom part is minus two $s / \sqrt{n}$, the upper part is plus two $s / \sqrt{n}$. And this thing here, is called the margin of error. So anytime you talk about an estimate, you talk about its margin of error, right? How low or how high can it be? So the 95% confidence interval, for μ is the estimate, right? Minus this margin of error, up to the upper point is the estimate, plus the margin. So I'm repeating myself, but I just want you to get this idea, because this is a very important idea. And it comes up all the time, even in real life. Right? So your estimate, you have to estimate the point estimate. And if you want to look at what's the margin of error, it's plus minus this. Now, let me show you some real examples of this confidence interval. Right. And you all, and one of the most popular things is electoral polls, right? So in the 2020 US presidential elections, you know, it was Donald Trump versus Joe Biden. And there are lots of polls were conducted before the election. So this is a report from a CNN poll of 886 voters, which was conducted in October of 2020. Right, so that means in this case, n is equal to 886. And what they reported is that 54% of these voters say they would vote for Biden and 42% would say that they would be voting for Trump. Right and the margin of error is plus minus 3.6%. Right? So what it means is that for Trump, this range extends from 54 minus 3.6. So that's 50.4. And what's the highest point highest, according to this margin of error? Is 57.6. Right? So that's the confidence interval for Biden. And for Trump, similarly, it's 42 minus 3.6. Right? So that's 38.4. Up to 42 plus 3.6. That's 45 points. So based on the CNN poll, these are the confidence intervals, the 95% confidence interval for Biden and Trump. In reality, once the election took place, what would what was the results? Biden got 51.3% of the popular vote. Right, so this is reality. And Trump got 46.8% of the vote. So this is what happened later. Remember, these polls were conducted before the elections took place. But all of these polls, they come with a margin of error. Let me give you another example. So this was the previous election to that 2016, in which case, it was Clinton versus Trump. Right. So this is a Bloomberg poll of 799 voters, which was conducted in November 2016, in which 46% of the voters said that they would be voting for Clinton 43% for Trump, and the margin of error was plus minus 3.5%. Right? So according to this, right, so the confidence interval for Clinton is 42.5. That's the lower point, the high point is 49.5. Right? And for Trump, the low point is 39.5. And the high point is 46.5. Now, if you compare these two now you'll see that in the worst case scenario of for Clinton, she could get less votes than the best case scenario for Trump. Right? So in this case, there are scenarios where based on my 95%

confidence interval for the to where Clinton could get less than votes than Trump, although the point estimates are 46% for Clinton, and 43%. For Trump, the margin of error is big enough so that in the worst case scenario, Clinton could get lesser votes than what Trump is. Right. So in a case like this, one would say that, based on this poll, the election is too close to we do not have enough confidence to say that in this case, after we take into account this interval, that Clinton is likely to get more votes than Trump. This was not the case with the CNN poll between Biden and Trump. See, even in the worst possible case, right? Biden would get 50.4 That would be better than the best possible scenario for Trump, which is 45.6. So in based on the CNN poll between Biden and Trump, one would say that Biden definitely has a higher chance of winning. Right? That wouldn't be the case in in this one where it's Clinton versus Trump. And you get similar sort of confidence intervals it with, even with the recent 2021 Canadian federal election, projection, number of seats between the liberals, conservatives and NDP. Right? This was the confidence interval, the liberals would get somewhere between 115 to 186. Conservatives between 90 to 148, an NDP between 19 and 54. Right? So this is based on a collection of polls from this website, 338 characters and if you compare it to the actual outcome, right liberals got 160, which was exactly somewhere in the middle of that interval. Conservatives got 119, which was again in the middle of this interval, NDP got 25. So which was again in the middle of the center. But you know, if for the NDP it was closer to the lower end than it was to the higher end. Right? For liberals. It was 160. So I think that's slightly higher than it was closer to the higher end, right, of this projection. But the important thing is these are the confidence intervals for the seats for the projected number of seats based on election. Okay, so these are examples of real life confidence intervals. But what I want to take you to take away from this lecture is this idea of a confidence interval, right, and how to construct it. And in the next group, we'll be doing examples with data of constructing such confidence intervals.