

module_prob4_lecture12

Fri, 2/18 12:51PM 12:06

SUMMARY KEYWORDS

probability, cdf, standard deviation, normal distribution, figure, symmetric, point, delivery, distributed, sigma, rule, lying, cumulative distribution function, dot, dist, birth weight, equal, clicker, norm dist, remember

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:05

So in the last clip, I introduced you to the idea of the 68, 95, 99.7 rule, and we were applying it to this particular example. So in this example, this is a Door Dash drivers delivery time, this is normally distributed with a mean of 20 minutes, and a standard deviation for. So that means, so this 20 As the high like most of the values lie around 20. But exactly where do they lie between what's the probability of different values. So that's what we are interested in figuring out. So remember, so here is the mean 20, if you look at one standard deviation, plus minus of it, so that 16, right, so 20 minus four, or 20 plus four, remember, the probability of it lying in this interval is 68%. So this is point six, eight. Now, if we look at two standard deviations, so that's two times four, that's eight, right? So 20 minus eight is 12, and 20 plus eight, that's 28. So the probability of lying between 12 and 28. Remember, that was the 95%. So this is point nine, five. And if you look at three standard deviations, so three times sigma, right, so that's 12. Right? So 20 minus 12 is eight, and 20 plus 12 is 32. The probability of that is point nine, nine. So these are the this is using the game to 68, 95 99.7. Now, this can also be used to calculate other probability. So suppose you want to just figure out what's the probability that the delivery time will be less than or equal to 24 minutes? Sometimes, like DoorDash would surely be interested in figuring out such abilities. So remember, 24. So that mean, plus one sigma? So how do we figure out what's the probability of less than or equal to 24? Now see, you know that the probability between 16 and 24 Right, so one sigma less than one sigma more than 68%. So that means, and remember the the unit the normal distribution is symmetric, right? So that means on this side will be 34. And this side will be 34. Right? So between mean and 2024, that will be 34%. Right? So up to between this and 24. That will be 34%. Right? And below the mean, right? All of this? We know, because the uniform distribution is symmetric, right? So that area is going to be 50%. So what's the probability total probability that the delivery time is less than or equal to 24? There will be point five, right? So there's this 50% plus the 34%, which is point three, four. So together, this will be point eight, four. So this is how you can figure out other probabilities once you know the probabilities of one standard deviation, two standard deviation, three standard deviation. What if we wanted to, for example, figure out what's the probability of it lying between 12 and 24 minutes? Remember, 12 is two standard deviations away, right? And 24 is one standard deviation. So how do we figure out what's the probability of lying between this and this? Again, we will make use of our rule and the fact that it's

symmetric. So first thing if you look at so here's my mean 20. So what's the probability of it lying between 20 and 24? So see, between this and this at 68%, right? So half of that right, so the probability of line between this and this that's going to be 34%. Similarly, if you look at this, the 12 which is different standard deviation. And here that's 28, the two standard deviation up, right? So together, that's 95%. Right? So between here and here is half of that. So the half of 95, that will be equal to, what is that 47.5%. So the total probability of it lying between 12 and 24, will be this 34%, which is point three, four plus this 47.5475. So that'll be five, a point eight. Right, so this is the probability of a line between 12 and 24. Again, we're making use of that 68-95-99.7 rule. And the fact that the normal distribution is symmetric to figure out this. Now, these were great if these numbers here, right, so these are very special numbers, right? So these are one standard deviation, two standard deviation, three standard deviation away from the mean. What if we wanted to figure out some other number like the probability of it being less than or equal to 22? So this is not one of these is not a number, which is one, two or three standard deviation way. Right? So this is 22 is half a standard deviation. So but we don't have an exact rule for that. So what we have to figure out, if we had to figure out what's the probability that the delivery time will be less than or equal to 22? That means we need to figure out the CDF at 22. Or if we want to figure out what's the probability that this delivery time detail there will be between, let's say, 22, and 26. So this will be the CDF at 26 Minus the CDF. So how do we figure out this? So for this, we need software. And Excel gives you exactly a way to do that using the norm dot Dist. formula. So let me show it to you how to do that. So here, I now have an Excel sheet. And see this is my normal parameters mu of 20. That's the mean of the DoorDash. Drivers. delivery times is sigma is four, right? And I want to find out what is the CDF at 22. So what I'm going to do is use equal to and norm. See, this already comes up norm dot Dist. And its parameters are this X. So my x is at C 3, then I have to put in the mean, My mean is here at A 3, so I want to put a three, then I have to put in the standard deviation, which is a v3. So that's B, three. And then after put in true gives cumulative distribution function false gives the probability mass function of the density function, right. So I need the CDF here. So I'm going to throw that so this is going to be point six nine, that's the probability the CDF at 22. The other CDF I need is at 26. So what I'm going to do is just copy this and paste it here. So let me check this, right. So this is at C 4. So that means a 26. Sorry, a 26 a three. So right, it's that mean? And that sigma and C are put dollar signs, so this two doesn't change, only this changes. So that so that gives me this probability, it's point nine three. So now once I know this, let me go back to my slides, right. So remember, this is point nine, three, the other one is point six, nine. Back in my slide, so I know this is point six, nine, here, this is point nine, three minus point six, nine. So that's going to be point two, four. So this is how you can use Excel. Same thing in Google Sheets, instead of the dot it's just NORM DIST. This is how you can figure out the CDF and the probabilities for a normal distribution. So now, here's a clicker question. Again, based on the normal distribution, so you may wish to just stop the video here and adapt this. So hope you're able to do this clicker question What it says is that suppose a baby's birth weight is normally distributed with mean 3.5 kg and variance of point five. And these are actually true numbers. In the sense if we look at the weight of birth weight of babies, those are normally distributed with a mean of 3.5 kg. So a family is expecting a baby, what's the probability that the baby's weight will be between three and four kg? So see, this three is the mean, minus one standard deviation. And the four is the mean, plus one standard deviation. So the probability of it lying within that one standard deviation is 68%, or point six, eight. So the correct answer here is point six. So I've introduced you to the normal distribution, and shown you how you can calculate various probabilities of the normal distribution. In the next clip, I'm going to show you some more properties of variables which are normally distributed