

module_prob3_lecture15

Mon, 2/7 10:37AM 15:28

SUMMARY KEYWORDS

pmf, cdf, binomial, probability, equal, google sheets, graph, binomial probability, excel, parameters, paste, change, cumulative distribution function, dist, distribution, talked, similarly, minus lambda, compute, true

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:04

So next, as promised, we'll be using Excel and Google Sheets to look at the binomial and poisson distributions, probabilities. So if you recall, the binomial formula is given by this. And the choice of formulas this. Now these are pretty complicated things to calculate, especially if n is big, and x is big, right. And in the poisson case, it requires e to the power minus lambda, and E is this special number, it's e , it's especially helpful if you can get software to calculate this for you. So there are already built in statistical packages, both in Excel and in Google Sheets, to do these computations very, so let's do that. So here's my Excel file, on the first page of the Excel file was the computation of the mean and variance that we did in the previous clip. And in the second sheet, is the binomial distribution that we're about to do now. And on the third sheet will be the poisson. This file will be available for download from the website. So I would encourage you all to download the file and try these out on your own. Because as I've said before, in learning software, there is no better practice than to get hands on trying to do this on your own. So let's look at the binomial, right, so what we have computing is the probability of various values of x , right, so on this column, so what I've done is written as x equal to 012, so on, and the way I have done it, is by using the formula, so I started out with the zero, and to do one, it's just this plus one, the next one will be the previous one, plus the next one, and so on. So now suppose we want to compute the binomial probability mass function. Now the binomial has two parameters. It's n , the number of trials and P the probability of success in each trial. So first of all, let's look at the binomial probabilities when n is equal to 10, and p is equal to two. So here, we want to calculate what's the probability that x is equal to zero. For x equal to zero, what's the binomial probability that x is equal to zero under these parameters, so what we're going to do in Excel, we have something called the BINOM dot DIST. BINOM.DIST, it's a, that gives you the binomial probabilities, the first parameter in so the first element of that is the number S . So that's the x . So I'm going to put in C 3 here. The second is the number of trials. So the number of trials is, is here is a tree. Let's see, I don't want, I want this to be fixed, I don't want this to be changing as I cut and paste the formula into other cells, right. So to avoid that, I'm going to put dollar signs in front of a and dollar signs in front of three. So that tells Excel that that cell is fixed, right. And similarly for p , this is here, right? And I'm going to put dollar signs in front of B . And in front of three to make sure that it's, it's fixed there. And the last component here is if you put in true, that gives you the cumulative distribution function, if you

put in the false then it gives you the probability mass function. So if I want to get the probability mass function, I'm just going to put in false ID, this gives me the PMF for that particular x . It's as simple as that. And then if I want to plug it in for the other values of x , so see here n is 10. So I can only go from x equal to zero up to 10. So if I plug these in, this gives me the PMF. Now if I wanted to do exactly the same thing, but instead of the probability mass function, I want to do the CDF. Right? So that's here. So what I'm going to do is I'm going to copy and paste this here. So I want to get binomial distribution. So first thing the X is it's because I've cut and pasted here, it has changed that first thing, it's I wanted actually to be x equal to zero. So I want to sorry, I want this to be instead of G three, I want this to be C three. Rest of it is fine. Because then A and three, that's fixed here, VM , three, that's fixed here. And the last component, remember, if it's false, then it gives you the PMF. If it is true, then it gives you the CD. So if that this is now the CDF, and then if I copy and paste here, it will give me the cumulative distribution. And so see, cumulative distribution, it goes up to one at the end, right? Because if we want to look at the probability that x takes values less than or equal to 10, that probability is one because all the possible values of x are from zero to 10. So this is my CDF. Now similarly, if I want to compute the PMF of the binomial, for a different set of parameters, I want to do it for this two set of parameters. So what I'm going to do is I'm going to call copy this, right, paste it here. Now see, this has changed, but I wanted to be C three, it is changed to E threes, right? So to change that, right, now I needed a new set of parameters, my set of parameters are this and this, right, so I'm going to change those, so I'm going to change this to a four that will make it 30. And then I'll make it to be for this false, that will give me PMF. And then if I copy and paste, now, n is equal to 30. So it should go down up to 30. So this gives you the probability for the various rounds. It is as easy as this. Now, see, if you wish to graph this red, you can also grab this. So if I just select these two columns here, and I want to plot them out. So sorry, let me start from the top from x , and PMFs. So suppose I want to graph it. So what I'm going to go to is insert and see if you go to charts, one of them is the scatterplot. Right, and the different types of scatter plot, this just gives you the dots, this joins the lines, this joins the lines, this one joins it smoothly, this doesn't join it smoothly, right? So let's choose this right. So, this gives you a graph of the PMF 's right and you can change various things. So you can put in a chart title saying binomial PMF. See, it already tells you that the blue thing is the PMF for n equal to 10 p equal to point two degree one is the PMF for n equal to 30 people now you can change lots of other elements of the graph right so for example, if you want to change let's say you want to add an x axis titles. So here is it adds in the two axis titles and you can change this so for example, this will be x if you enter change this so let's put in their space and put in PMF if you want to change the style of the graph, see it gives you lots of different styles this this you can choose according to your what do you like? So this is how you can get the binomial PMF CDF. And it could be one block. So what I would like you to try on your own is to derive the binomial CDF for this different set of parameters n equal to 30 p equal to point two. And then I would like to try graphing the two cities. And let me end the day by showing you how you can do the same thing also for the poisson. So again, for poisson, first generated the axes, so these are my x s for boys are the only there's only one parameter, which is λ , right, so I'm choosing two different λ s, one is five, the other is 12. So if I want to, again, generate the points of PMF, for x equal to zero, and λ equal to five, what I'm going to do is again, equal to and see if this package called poisson distribution, POISSON DIST. So the first one is the x to put in C 3, then as the mean mean is meaning the λ , so we'll put into a three button that's fixed. And then again, if you want, just the PMF you write false. And then, if you want to copy and paste a love read, you can go as long as you want, right, it will generate all the probabilities. Similarly, if you want to instead, instead of the PMF, if you want to do the CDF, it's similar, it's just pasted in there, to change this from F three to C three. And I have to change the false to true, that will make it better to give you the seedlings. And, and if I cut and paste and keep going that you give you the CTFs. So this is how you can generate the PMF the

CDF for the binomial for the poisson. And there are lots of other distributions that you can do as well in Excel. And it's almost the same procedure in Google Sheets. And I've written it down in my slide. So there instead of buying on dog dest that Excel has, it just says by noticed. Similarly, it's just true for softest. So. So these, these are several differences, but between Excel and Google Sheets, but the procedure is the same. So hopefully, you will try, you'll download this file and try it out on your own. And this will give you a good idea of how to derive this probabilities, how to draw the charts, play around with the charts, because that's how you will learn how to create these things in Excel sheets. So let me end the module with a quick recap of what we did. So I introduced you to the idea of random variables. And here we focus mostly on discrete random variables. In the next one, we'll be talking about continuous random variables. And I talked about probabilities in two ways. One is the probability mass function, which is the probability of the random variable taking particular value. And secondly, the CDF or the cumulative distribution function, where which, which is the probability that the random variable takes values less than or equal to particular. And then I talked about expectation and variance. And in the next module, I'm going to do a related kind of measure, which is called correlation. And then I ended it by talking about three common discrete distributions. One is the Bernoulli, which is the success or failure in a single trial. Binomial the number of successes in n trials and poisson is a count variable where it looks at the number of occurrences within a given time period. So hope this was got you a good introduction to it. two random variables and in particular discrete random variables. And please do remember the concepts of the PMF the CDF and expectation and variance, because we're going to need these concepts also in the next module. And these are very important concepts in the future. So let's stop