TRANSCRIPT MY CLASS NOTES

JIGSAW ACADEMY
THE ONLINE SCHOOL OF ANALYTICS

Now, let's spend some time in looking at different kinds of distributions – discrete and continuous. In this topic we will start with a very simple discrete distribution which is called a binomial distribution.

>>>

Remember that if a random variable can have only discrete outcomes, we have a discrete probability distribution. For example a flip of a coin, it can have only two outcomes - heads or tails. The number of customers walking into a store on any given day; the number of times a machine breaks down in a year; the number of people claiming fire insurance in a month - all of these are examples of random variables whose outcomes can only be discrete.

Now, there are many kinds of discrete distributions, so if we have a random variable whose outcome is discrete, there are ways of describing those random variable outcomes, and these are essentially what are called discrete distributions.

There are many types of discrete distributions:

- Binomial
- Negative Binomial
- Geometric
- Hyper- geometric
- Poisson

We will start with first understanding a simple kind of a discrete distribution which is a Binomial Distribution.

>>>

The Binomial Distribution is an example of a probability distribution of a discrete random variable. The simplest example is a coin toss, but there are many other examples:

- The gender of babies delivered in a hospital. It is a random variable, and it can have only two outcomes – male or female.
- Whether or not a drug has fatal side effects. Again, either the drug has fatal side effects or not.

Now a common feature across all these examples is that the random variable can have only two possible outcomes: could be one or zero, male female, win or loss, black or white. There are no external factors influencing the probability of each outcome over time. What this means is supposing we take a look at the coin flip. Now what is the probability that when you flip a coin you see heads - fifty percent? Now over time, the probability of seeing a head does not change. Whether I flip a coin today or I flip a coin ten years from now, the likelihood of getting heads is still fifty percent.

Also, the chances of each outcome ore independent of previous results which means on any throw of a coin the chances of getting a head is still fifty percent. Now it maybe in the last five throws you've seen five heads in a row, but on the sixth throw – just the sixth throw – the probability of getting heads is still fifty percent. If these conditions are met for the random variable outcome, then we can say that the random variable outcome is an example of a binomial distribution.

>>>

Let's start with a very simple example to understand binomial distribution better and then we will look at more business examples. The simplest example of a binomial distribution is a coin toss. Remember when we toss a coin its outcome is random because we don't whether we are going to get a heads or a tails. The number of possible outcomes is only two – heads or tails. Are there any external factors influencing the outcome of probability over time: No. Are the chances of each outcome independent of previous trials: Yes.

>>>

When we toss a coin once this is actually an example of what is called a Bernoulli trial. Then you have multiple Bernoulli trials, you have what is called a binomial distribution. What that means is, supposing we take a coin and we toss it ten times. We are interested in measuring the number of heads that we see when we toss a coin ten times. That distribution of the number of heads that we see when we toss a coin ten times follows what is called a binomial distribution.

In fact, let's think about this for a minute. Supposing we toss a coin ten times, how many heads do you expect to see? Will you always see five heads? No, sometimes you may see zero heads. You many see one head. You may see two heads. You may see eight heads. You may see ten heads. Are all of these outcomes equally likely? No. What is more likely? It's certainly more likely that we will see five heads. It is certainly very, very unlikely that we will see zero heads. It is very, very unlikely that we will see ten heads. But actually, rather than just talking about it, it is highly likely that we will see five heads, and that it is highly unlikely that we will see ten heads.

>>>

We can actually calculate these probabilities using mathematical formulae. It turns out that if you have a random variable that follows a binomial distribution there exist a mathematical formula for calculating the probability of any outcome. For a binomial distribution, that formula is this. The probability of an outcome, let's call it x, equals to factorial n by factorial x times n minus x factorial p raised to the power x times 1 minus p raised to the power n minus x. Now there are a lot of things here. What is n, x and p. x is the outcome that we are interested in. n is the number of trials of your Bernoulli process.

In this example when we are tossing a coin the number of times we tossed a coin is the number of trials. p is the probability of success on each trial. Remember we are interested in seeing heads, so the probability of success is point five. Now I also said that this formula has a factorial. What is factorial? Factorial is simply, if I say n factorial, it is the multiplication of n times n minus one, times n minus two, all the way to one. If for example, if n was five, then five factorial would be five times four times three times two times one.

For a binomially distributed random variable the probability of any outcome x is given by this formula. Now imagine I asked you what is the probability that if I throw a coin ten times I see four heads?

>>>

Remember the formula says simply substitute the values in this formula and I will get the probability of four heads. What is n? n is the number of trials. Remember we are flipping a coin ten times, so n equals to ten. What is x? x is the success or the outcome that we are interested in. We are interested in four heads, so x equals to four. When I substitute these values, this becomes factorial ten by factorial four times factorial six times probability of success. p is the probability of success, so point five raised to the power four times one minus p which is point six raised to the power six.

Actually, if you do this you will see that you will get point two zero five. What is this point two zero five? It says that if you flip a coin ten times the likelihood that you will see four heads is twenty percent. Remember this is a fair coin, meaning the probability of getting heads is fifty percent. But if you flip a coin ten times it is not necessary that every single time you will see five heads. Sometimes you may see six, sometimes you may see seven, sometimes you may see three and sometimes you may see zero.

We are calculating mathematically, how likely is it that you are going to see, for example, four heads when you throw a fair coin ten times. It turns out twenty percent of the time. Similarly if I asked you to calculate probability of seeing eight heads you would say in this formula this would become factorial ten by factorial eight times factorial two times point five raised to the power eight times point two raised to the power two. Now should you expect the probability of seeing eight heads to be higher or lower than twenty percent?

The good news for us is every time we want to calculate this probability we don't really need to do these calculations by hand. In whatever tool that we use for statistical analysis there will be built in functions that help us do these calculations. For example, in excel, to calculate the binomial distribution probability there is a function called BINOM.DIST. If you go to excel and type in equals to binomial distribution excel will give you this suggestion which is the parameter of this function.

There are four parameters: the number of successes, the number of trials, the probability of success and cumulative. In our example I want to see the probability of four heads in ten throws of a fair coin. The number of successes will be four. The number of trials will be ten. The probability of success is the probability of seeing head on any throw of one fair coin - fifty percent. In cumulative equals to false. I will explain cumulative in a little while, for now just remember cumulative equals to false. If I type in these values in the binomial distribution formula what will I get? Let's go to excel and see.

>>>

Imagine that I'm interested in the probability of seeing four heads in ten throws of a fair coin. I just have to say binomial distribution and then excel will ask me number of successes - four successes, number of trials – ten trials, probability of success – point five and cumulative for now we are simply going to say false. The probability of getting four heads in ten throws of a coin is twenty percent. Now, of course, we can do this for all possible outcomes. So what is the number of heads that you can see when you throw a coin ten times, the range of possible outcomes, we can say zero heads, one, two, all the way to ten.

Now for each of these outcomes I can calculate a probability. It is simply binomial distribution, the number of successes, the number of trials always ten, the probability of success is always point five, and cumulative for now we are simply going to use false. So the probability that if you throw a coin ten times you will see zero heads, in other words all tails, is zero point zero zero zero nine seven six six. If I drag this formula down I will get the probabilities for each of these outcomes. Which of these outcomes has the maximum probability? Five – twenty four percent. Is that reasonable? Yes, because we would expect that if you throw a coin ten times I would expect in general to see five heads. There is a twenty four point six percent chance that when I throw a coin ten times, I will see five heads. And notice something interesting here. The further you move away from this five, which is the expected value of this distribution, the lower the probability, which again makes sense.

I should expect to see very low probabilities for zero heads and ten heads. I should expect to see low probabilities for one head, nine heads and so on. Notice here that these are symmetric because the probability is fifty percent, so these are symmetric values. But it is not necessary that you will always see that. In fact, we can do a probability distribution chart. All we have to do is insert a chart like this and take off Series 1. This is a probability distribution for the random variable which is the number of heads in ten throws of a coin. The range of possible values is on the x axis and the associated probabilities are on the y axis.

>>

Let's do another example. Imagine that you have a production process and in that production process the probability of seeing a defective item is twenty percent. Now supposing you're doing quality control and you have a lot of five pieces. What is the probability of seeing defects in that lot? If you have five pieces, you could potentially see zero defects, one defect, two defects, three, four or five. What about the probabilities? In this case it would be binomial distribution, the number of successes is zero- what is the probability of seeing zero defects? The number of trials is five.

The probability of success is point two. And cumulative equals to false. We are still using false. And if I drag this down, this is the probability of seeing defects. And if we do a chart you can see what the distribution looks like.

This is a probability distribution chart of seeing defects in a lot of five, when the probability of seeing a defect is twenty percent. Notice a couple of interesting things. What will be the sum of all these probabilities? Should be hundred percent because these are the only possible outcomes of this random variable - hundred percent.

>>>

Similarly if you go back to our coin toss example, the sum of all these probabilities should be again one hundred percent, because remember these are the only possible outcomes of this random variable. Let's also take a look at cumulative. In both the examples, we've simply looked at cumulative equals to false. But cumulative actually, in the function, can have two values. It could be either false or it could be true. If we use cumulative equals to false, we are calculating what is called a point probability.

What is a point probability? It is simply the probability of outcome being exactly equal to x. But if we say cumulative equals to true, we are calculating what is called a cumulative probability. Remember cumulative binomial distribution function, that's exactly what we are calculating here. We are calculating, if we use cumulative equals to true, the probability of outcomes less than equal to x.

For example here, instead of in this formula using false, I say true. I say binomial distribution, outcomes is four, the number of trials is ten, the probability of success is point five and cumulative equals to true. What I'm calculating is remember, the probability of outcomes less than equal to four, which is the likelihood of seeing less than equal to four heads when I throw a coin ten times, which is nothing but the sum of seeing zero heads plus one head plus two, three and four. This gives me the probability of seeing less than equal to four heads.

How do I know this is true? Let me sum these up. In fact, let's do that as a check. Supposing we sum up these probabilities, zero, one, two, three, four.
We should get exactly the same number.
Cumulative equals to two gives us the cumulative probability of outcomes less than equal to x. Cumulative equals to false gives us point probability of outcome exactly equal to x. If I now just drag this formula up and down, by the time we get to ten, the cumulative probability has to be equal to one because we are summing up all possible outcomes and this is the final possible outcome.

We have looked at calculating probabilities of an outcome when we have a random variable that is distributed with a binomial distribution. Remember that if we know that our random number is distributed binomially there is a mathematical formula that we can use to calculate probability of any outcome. And it turns out this is true for variety of distributions. You could have distributions that are binomial; but you could also have distributions that are different. For example, Poisson, for example geometry, and as long as we know what kind of a distribution we are dealing with, there exist mathematical formulae that help us calculate probabilities of outcomes that we are seeing.

One final thing about these probabilities, what exactly are these probabilities? For example if we say that the probability of seeing five heads is twenty-four percent, what does that mean? It means that because this is a random variable the outcome is subject to randomness, random variation that no one can control. The probability of seeing five heads in ten throws of a coin is the random chance probability because these outcomes are subject to random variation. If the coin were fair we would expect to see heads fifty percent of the times. But we will not see every time we throw a coin ten times, five heads, because that outcome is random.

Therefore on the long run there is a twenty-four percent chance that when you throw a coin ten times you see five heads. This is the random chance probability of an outcome given a certain probability of distribution. Now in the next topic we will take a look at how we use binomial distributions in the context of a business problem.

>>
Now takes a look at a binomial distribution application to a business problem. Quality control is

something that is very important when you have a manufacturing process. Supposing we work in a manufacturing unit and the process has a defect rate of ten percent. What that means is that one out of ten products is potentially defective. Let's say we work in quality control and as a part of the quality control process we evaluate products to see what the defect rates are like. Since we have a defect rate of ten percent, ideally we should expect to see that in every ten products, one is defective.

If our products are stored in lots of ten and we pick a random lot and we get two defects. How likely is this outcome because of random chance? How do we do that? Remember that if the defect rate is ten percent, then ideally in every ten products, one is defective. However, it is possible that in a particular lot of ten products, you may have two defective simply because of randomness – meaning your manufacturing process still has a defect rate of ten percent but you just happen to have two defective products in one lot, or, potentially, three defective products in one lot.

What we want to calculate is how likely is it that you will end up with two defective products in one lot of ten? This is a binomial probability that we are going to calculate because either a product is defective, or it is not defective. If I wanted to calculate this probability I could put in the parameters in the actual binomial probability function or I could use excel. In excel, remember, we need to specify the number of successes, the number of trials, the probability of success and cumulative equals to false. In our example the number of successes – success is defined as finding a defective product – the number of successes that we are interested in is two defects. The number of trials is ten because we are looking at lot s of ten.

The probability of success – the probability of any particular product being defective is ten percent, and false. It turns out that there is a nineteen percent probability that if you take any random lot and you find two defects that this could happen because of random chance.

>>>

Let's look at the slightly more complex example of again applying a binomial distribution to a business problem. Imagine that you are working in the Finance Department for a particular company and you are reviewing your Accounts Receivable balances. What are Accounts Receivables? Many times, especially if you are a manufacturing company, it turns out that when you sell your product to a customer you don't receive payment immediately. You raise an invoice and the customer will pay you in thirty days, forty-five days, based on whatever the contract agreement is. It is very rare that you will receive your money right away. The money that is owed to you for product that has already been delivered is what is called Accounts Receivables.

Supposing you are trying to do some cash flow planning for your company, now, what is cash flow? The working capital, essentially how much money is there in your hand at any given day for you to use for the regular day-to-day running of your company. So you are looking at cash flow, you are trying to figure out your working capital balances. One of the components of working capital is your Accounts Receivables balances. You are expected to get a certain amount of money in thirty days. You are expected to get another certain amount of money in forty-five days depending on when you made sales to your customers and what that agreement is around the payment cycle.

Let's say that you are trying to figure out working capital requirements, and obviously one of the inputs for your working capital sources is going to be your Accounts Receivables balances. Another complication you may have is sometimes your customers may not pay on time. So based on historical data of your Accounts Receivables balances and customer payment, you know that on an average forty percent of your customers are more than sixty days late with payments. So they are supposed to pay say today, forty percent of your customers are more than sixty days late with payments.

Imagine that you had a total of hundred fifty customers, and when you're doing your cash flow working capital planning, you want to create a contingency fund for situations where more than fifty percent of your customers are late, because remember if more customers than expected are late, then you are going to ha a problem with your cash flow. You may need to have a contingency fund. Now, what will be the size of the contingency fund? How much money should you put in that contingency fund? It can't be an arbitrary amount. You have to justify that, because obviously there is an opportunity cost of keeping money in an account, right, in a contingency fund?

If you think about it, one critical input into the size of your contingency fund is going to be how likely is it that on any given month more than fifty percent of my customers will be late, if historical data tells me that forty percent is the average number of customers that are late with payments? I am just going to repeat this, because this is easy to understand. Remember we are looking at Accounts Receivables balances because that is money that is due to the company. If that money comes on time we can use it for working capital. However, not all the money always comes on time. Looking at historical data we know that on average forty percent of our customers are more than sixty days late with payments.

Remember that forty percent is an average, so some months it could be more than forty percent, and some other months it could be less than forty percent. If it is less than forty percent we don't really have a problem. However, if more than forty percent of our customers are late with payments then we may have issues with working capital because we haven't got the money that we expected to receive. Therefore we are creating a contingency fund; some cash that we want to keep in that fund for dealing with working capital situations where more than expected numbers of customers are late with payments.

So now the question is what should be the size of the contingency fund? The size of the contingency fund should depend on what is the average Accounts Receivables balance across our customers, but it should also depend on how likely is it that we will have a situation where more than fifty percent of our customers are late with payments. If you think about it, again, this is a binomial distributed random variable. Why is that, because every customer can either be late or not late, and if customer A is late that does not impact the chances of customer B being late? If we use this binomial distribution to solve this.

>>>

The probability of more than fifty percent of my customers being late in any month – how do we calculate that? Remember that we have hundred fifty customers, so more than hundred fifty customers being late, as I said initially, more than seventy-five customers being late in any given month, which is what - the probability of more than seventy-six customers being late, seventyseven customers being late plus seventy-eight being late and so on, all the way to probability of all my hundred fifty customers being late. We could do this and we could use excel to do this but this is still a tedious calculation because we have to calculate seventy-five different probabilities.

>>>

We want to check the calculation that we've seen earlier which was, what is the probability that more than fifty percent of my customers will be late in any given month given that the average probability is forty percent of being late. Now remember greater than seventy-five, when I have hundred and fifty customers, are the probability of seventy-six customers being late, plus the probability of seventy-seven being late, and so on. We can calculate these individual probabilities using binomial distribution and we can say binomial distribution, I'm interested in seventy-six customers being late out of hundred and fifty when the late percentage or the probability of a customer being late is forty percent. And I'm going to say false.

When I do that, what I get is the probability that in any given month, say seventy-six of my customers will be late with payments when I'm expecting forty percent to be late with payments on average. If I just drag this formula down I'm going to get the probability of seventy-six being late, the probability of seventy-seven being late, the probability of seventy-eight being late, and to answer the question of what is the probability of greater than seventy-five customers being late, I just need to sum these probabilities up. So let’s do that and see what we get. We get 0.005. However, there is a faster way of doing this, just to calculate cumulative probability. Remember cumulative probability is the same function that we use in excel, except instead of using false we say true.

If I wanted to do cumulative probability of less than equal to seventy-six I would say binomial distribution, the number of successes is seventysix, the number of trials is one fifty, the probability of success point four, and true. What I have here is the probability that I will see less than equal to seventy-six customers being late in any given month. Now what are we interested in? We are interested in the probability of greater than seventy-five customers being late. The probability of greater than seventy-five is the same as one minus the probability of less than equal to seventy-five. Why is that? Because remember the sum of probabilities across all the possible outcomes has to be equal to one.

What is probability of less than equal to seventyfive? That we can calculate using a binomial distribution formula, right? We can simply say binomial distribution one fifty, seventy-five, hundred fifty, four and true. Now the probability of greater than seventy-five, I'm going to say equals to one minus binomial distribution, the number of successes is seventy-five, the number of trials is hundred fifty, the probability of success is point four and true. The reason I'm using true is because I want to calculate less than equal to seventy-five and then I want to do one minus that to get the probability of greater than seventy-five customers being late in any given month. You can see this number is the same as this number which we had calculated.

Cumulative equal to true gives you the probability of all outcomes less than equal to x where x is the outcome that we are interested in.

>>>

There is a faster method to do this, and that is to use the concept of cumulative probability. What is cumulative probability?

>>

Remember in excel we were looking at the binomial distribution function and we said binomial distribution has some parameters: the number of successes, the number of trials, the probability of success and something called cumulative which we have listed as false. When you say false in excel for a binomial distribution function then what we are calculating is the random chance probability of seeing exactly S successes in T trials when the probability of success in any trial is P. If we say binomial distribution four, ten, point five, false in the coin toss example, we are calculating the probability of exactly seeing four heads in ten throws of a fair coin, where the probability of head is point five.

What happens if you replace the false with true? In the distribution function in excel the cumulative can have two values: false or true. What happens when we set this to true? When we set this to true, essentially what happens is you are calculating now the random chance probability of seeing less than equal to S successes, in other words if we had said binomial distribution four, ten, point five, true, we were calculating the probability of seeing less than equal to four heads in ten throws of a fair coin. What is less than equal to four heads? It's the probability of seeing zero heads plus the probability of seeing one head, plus the probability of seeing two heads, plus the probability of three heads plus the probability of four heads. So what this is called is a cumulative probability.

>>>

How do we use this cum probability in our situation where we want to calculate the probability of seeing greater than fifty percent customers being late? Now remember greater than fifty percent of our customers being late essentially mean seventy-six plus seventy-seven plus seventy-eight plus seventy-nine, the probability of all these being late. The short cut to this calculation is to say the probability of greater than seventy-five customers being late should be equal to one minus the probability of seeing less than equal to seventy-five customers being late. Why is that? Remember the probability across all outcomes should add up to one.

The probability of seeing greater than seventy-five customers being late has to be equal to one minus the probability of seeing less than equal to seventy-five customers being late. And the probability of seeing less than equal to seventyfive customers being late is directly available to us in excel using cumulative equals to true. In this particular example, the probability of seeing greater than seventy-five of my customers being late is the same as one minus the binomial distribution probability of seeing less than equal to seventy-five customers being late.

If we do that in excel, we get a probability of 0.005. It turns out it is highly unlikely that in any given month you will see more than of your customers being late with payments. Therefore the size of your contingency fund who probability of seeing less than equal to seventy-five customers being late should not be very, very large, because it is fairly unlikely that you will be required to use it. But let's just review the same calculation in excel to confirm that greater than seventy-five is the same as one minus less than equal to seventyfive.

>>

We were looking at discrete distributions and we've reviewed binomial distribution. Now what we will do is take a look at other discrete distributions including some special cases of binomial. These are hypergeometric, negative binomial and geometric distributions. Then we will also take a look at a different kind of a discrete distribution which is a Poisson distribution.

>>>

Now what is a hypergeometric distribution? A hypergeometric distribution is generated when we have Bernoulli trials, like a binomial distribution but selections are not replaced. What does that mean? Imagine that we had a population and we were picking units from that population. Now if I don't replace the unit that I have picked back into the population, the probability of successive outcomes is changed, because your population base is changing. If we however replace the unit that we have picked back into the population, then your probability is going to be different because your population now has the same denominator in every calculation. If you have a situation where selections are not replaced, then you can't really have a binomial distribution, what you have is a hypergeometric distribution.

Again, you can see there is a formula here that helps us calculate the probabilities of a hypergeometric distribution. We are simply going to use a tool to calculate this for us.
>>>

So we'll do that with an example, this is an HR example. Supposing that a company wants to encourage diversity in its management ranks. There are eighteen employees eligible for promotion and of these, nine are women. When the promotions are announced, it is seen that eight people have been promoted of which three are woman. Is there a problem with discrimination? Now in this example, as and when you select these people, so when the first person has been promoted, the second person has been promoted, you're taking them out of that eighteen which is the population.

The probability of successive outcomes which is who is getting promoted and is that a woman is going to change, because you are not replacing people back into the pool. What we should be using here is a hypergeometric distribution. In excel we simply have to type in the distribution formula and excel will ask us for some parameters. It turns out that there are parameters for hypergeometric distribution that depend on the sample outcome and the population outcome.
What are they?

If we wanted to calculate a hypergeometric distribution we would type in hypergeometric distribution in excel and excel will tell us that it needs the following parameters. First the number of sample successes, the second the total sample size, third the number of population successes, fourth the population size and then cumulative. In our particular example, remember our population total is eighteen people. These are people eligible for promotion out of which nine are women. We are interested in women, so the success is basically choosing a woman. The number of population is eighteen, the population success is nine. When the promotions are announced, that is a sample of the eighteen, so the total sample size of promotions was eight, of which three were women.

Those are the numbers that we want to input here. Three, eight, sample size successes in the population are nine out of eighteen, and we are going to say cumulative equals to false. There is a twenty-four percent chance that three out of eight people that have been selected for promotion are women, and that could have happened only due to random chance variation, because ideally if the same proportion had been maintained, you should have expected to see four women out of the eight people that have been selected for promotion.

>>>

Let’s look at other binomial type distribution. We will start with something called a negative binomial. A negative binomial distribution is used when we are interested in the number of trials we will need to get to X successes. For example, supposing we have a store and the probability that any random customer walking into my store and making a purchase is twenty percent. What is the probability that the thirtieth purchase in my store will happen with the hundredth customer?

How do we do that? We use what is called a negative binomial distribution. In excel the negative binomial distribution requires the following parameters: the number of failures, the number of successes, the probability of success and cumulative. What is the number of failures? Remember that if you want the thirtieth purchase in your store to happen with the hundredth customer, there have to be seventy customers who don't make a purchase. So the number of failures is seventy. The number of successes is thirty. The probability of success is twenty percent, point two. And because I'm interested in the probability of exactly my hundredth customer making the thirtieth purchase, I will use cumulative equals to false. What we get is a probability value of 0.001.

Let's just see this in excel again. We say a negative binomial distribution. Excel asks us for number of failures, number of successes, probability of success and cumulative. In our particular example we want the hundredth customer to make the thirtieth purchase which means there have to be seventy customers that don't make a purchase – number of failures, thirty customers who make a purchase – number of successes, probability of success is twenty percent – any one customer making a purchase, and cumulative equals to false. 0.001 which is a very low probability.

>>>

Another binomial type distribution - geometric distribution. Geometric distribution is used when you are interested only in the probability of the first success. Let's take a look at an example. Supposing that we have a production process and it produces components with a defect rate of two percent.

What that means is two out of hundred components that it produces will be defective. What is the probability that a Quality Control Inspector will need to review at the most twenty items before finding a defect? Now let's take a look at this question carefully, because what it's asking is what is the probability that a QC Inspector will need to review at the most twenty items, which is less than equal to twenty items. What should be the answer? Right away you understand that because we are asking for less than equal to twenty items, we have to say cumulative equals to true. And because we are looking at the probability of finding the first defect we will use geometric distribution.

Now in excel, geometric distribution is a special case of a negative binomial distribution. We will use the same negative binomial distribution function that we had just reviewed for geometric distributions as well. But, what is the difference? The number of successes will be one because we are looking at the first success. Therefore in this particular example the number of failures has to be nineteen, right? The number of failures is nineteen, the number of successes is one, the probability of success is 0.02 – because there is a two percent defect rate and we use cumulative equals to true because we are looking at at the most twenty items.

So negative binomial distribution 19, 1, 0.02, true which gives us 0.33

Let's replicate that in excel. Negative binomial distribution, remember for geometric we still use negative binomial distribution, number of failures is nineteen, the number of successes is one, the probability of success is 0.02, and cumulative equals to true. If we had changed this question to say what is the probability that the Quality Control Inspector would have to give you exactly twenty items for finding the first defect. We would say negative binomial distribution, number of failures is nineteen, number of successes is one, probability of success is 0.02, and false. One percent chance that you would find the first defect in the twentieth item.

>>>

We have been looking at probability distributions and we started with discrete distributions. Within discrete distributions, we've looked at binomial distributions and we've looked at three other distributions that are special cases of binomial – hypergeometric, negative binomial and geometric.

>>>

What we will look at in this topic now is what is called a Poisson distribution which is another kind of a discrete probability distribution. A Poisson distribution is used to model the number of events that occur in a specified time frame

>>>

For example, the number of insurance claims received by a company in a month; the number of people infected by a new disease in a day; the number of telephone calls received by a call centre in an hour; the number of patients needing emergency services in one particular hospital in one day. All of these are examples of a Poisson distributed random variable outcome.

>>>

In order for us to know that we are correctly using a Poisson distribution, these conditions have to apply.

1. The events have to be counted as whole numbers.
2. The events are independent. Remember what does independence mean? If one event

occurs it does not impact the probability of the second event occurring.

3. The average frequency of occurrence for the given time period is known.
4. We have a way of counting the number of events that have already happened.

If all of these conditions apply then we can model the outcome of this random variable using a Poisson distribution. Just like a binomial distribution or any other distribution there is a mathematical formula that allows us to calculate probabilities if we know that the random variable outcome is distributed as a Poisson distribution.

>>>

That is essentially this formula here: lambda raised to the power x, e to the power minus lambda by factorial x. x is the outcome that we are interested in. Lambda is the mean number of occurrences or the average number of occurrences that we already know happened in a given interval of time. Essentially we get lambda from historical data. One important thing to notice here is that there is no sample size number in the formula for Poisson probability. It only depends on the mean number of occurrences that have happened in a given interval of time in historical data.

>>>

Let’s use an example to understand how we apply Poisson distributions. Imagine that you work as a Manager in a call centre, and you have a staff of fifty-five people who on average handle three hundred and thirty calls in an hour. Now a major holiday is coming up and five of your resources want to take leave on that day. You estimate that the remaining fifty resources can manage twenty percent greater calls, but what happens if there are more than twenty percent greater calls on that particular day? Then you may need an additional resource because otherwise you will have calls that will go unanswered.

Now there is a cost to the resource so obviously you want to make sure that if you do take a resource that that person will be used. In other words you need to be sure about whether or not you are actually going to use that person. How do we know whether we are actually going to use that person? You have to figure out how likely is it that on that particular day you get more than twenty percent increase in call volume.

How do we do that? Remember the number of calls that come to a call centre in an hour is a Poisson distributed random variable; it is discrete and it can take any number of values from zero to infinity. If we want to use a Poisson distribution we first have to calculate lambda. Now remember there are three hundred and thirty calls on hour that are managed by fifty-five resources. In other words, the average number of calls an hour per resource is six. The fifty resources can manage twenty percent greater call volume which means essentially you are looking at the fifty resources managing seven calls an hour because three hundred and thirty times twenty percent increase divided by the fifty numbers of resources is 7. 2.

Now of course you can’t have 7.2 calls an hour so we are going to make that equal to seven calls an hour.

Essentially your fifty resources can stretch and manage seven calls an hour, but if you have greater than seven calls an hour, then you are going to be in trouble. What we want to calculate actually is how likely is it that on that particular day you are going to see eight or more calls an hour when historically you’ve been seeing six calls in an hour. Of course we can calculate this using a Poisson distribution formula and we will just implement this in excel.

What we need is the probability of seeing greater than equal to eight calls an hour when the historical data says average is six. Greater than equal to eight – if we want to calculate the probability of greater than equal to an outcome in excel we will do one minus the cumulative probability of less than equal to that outcome. How do we do that? We have to say one minus
Poisson distribution.

When you say Poisson distribution excel is going to ask you for some parameters. The first parameter is the outcome that we are interested in. Remember that if we greater than equal to eight, we have to say one minus less than equal to seven. The outcome I'm going to say is seven. The average that we have observed in historical data which is the mean is six, and cumulative should be equal to true. We get twenty-five percent. What does this 25% mean? Remember the 25% essentially is interpreted like this. It says there is a 25% chance that on any given day you will end up seeing more than eight calls in an hour when you're expecting six simply because of random variation. What does this mean for our question?

>>>

Should we hire a resource because now we know that there is a 25% chance that we will actually use the resource? Now as a Manager this is a call that you have to take. You may have some estimate of what is the loss for every call that you don't handle. You can balance against the cost of actually using this resource and the 25% chance that you are actually going to use the resource and then decide whether or not you want to hire the resource.

>>>

Here's another example of a Poisson distribution. Imagine that there is an ATM machine in a particular location, and you work for a bank, and the bank notes that there is an average of eighty withdrawals a day with an average transaction amount of $40. The bank needs to stock the machine with appropriate levels of cash. How do they decide how much money to put in the machine? Should it be equal to eighty times forty per day? Not necessarily, because remember an average of eighty withdrawals a day doesn't mean every day eighty withdrawals. Some days it could be greater than eighty and some days it could be less than eighty.

Assuming that this ATM has starting with a zero balance, how do you decide what is the most appropriate amount of cash that needs to be stocked for a four day period. Here's another input that will be useful to you when you make this decision. A customer service KRA, which is a key result area, something that the bank is very particular about. One of the things that the bank is very particular about is that it wants to keep customer complaints to less than 10%. You need to stock the ATM machine. How do you decide how much cash to keep, knowing that you want to reduce customer complaints to less than 10%? How do you think we can do this?

One way to do this is to calculate what is the probability of seeing a certain number of withdrawals that have a probability of less than 10%? How do we ensure that we stock the ATM machine with the right level of cash and make sure that customer complaints are less than 10%? When will customers complain - if the ATM machine is out of cash? Imagine that for the ATM machine to be out of cash essentially more number of people than expected withdraw money. We can't just keep an infinite amount of money in the ATM machine. It is a finite number.

When we say that the average is eighty withdrawals a day and the average transaction amount is $40, then on average there is $3,200 worth of withdrawal every day. But we know that some days when there are eighty-five withdrawals, ninety withdrawals or hundred withdrawals. How do we decide what is that cut-off so that we stock for more than eighty but in a way that optimizes the cash balance for the bank, which is there will be some customers who will be dissatisfied because some days the ATM machine may run out of money, but we want to keep that dissatisfaction rate to less than 10%.

Let's think about this again using a Poisson distribution. Let's go to excel and see how we might be able to solve this problem. The historical data tells us that there are eighty withdrawals a day. If I was interested in finding out how likely is it that I will get more than eighty withdrawals in one particular day. We can do that using Poisson distribution, right? We can say one minus Poisson distribution. My number of withdrawals is eighty, the outcome that I'm interested in. The average is also eighty and cumulative equals to true. If we do this, there is a 47% chance that on any given day you will see eighty or greater number of withdrawals.

We want to figure out what is that number where the greater than that number of withdrawals is less than 10%. The chances of seeing greater than that number of withdrawals should be less than 10%. Let's just do a trial and error method. We have eighty withdrawals, let's try eighty-five. What happens? What is the probability that on any given day you will see eighty-five or greater number of withdrawals? Again we can do one minus Poisson distribution, eighty-five; eighty is the average, and true. This is the probability that you will see greater than eighty-five number of withdrawals on any given day.

Similarly let's try ninety. We can just drag the formula down. What is interesting here? The probability of seeing greater than ninety withdrawals in any given day is 12%. The probability of seeing greater than ninety-five withdrawals a day is 4%. We are interested in restricting that probability to 10% so it is some number between ninety and ninety-five. Let's just insert a couple of rows here and let's try doing ninety-one and ninety-two. Now if we just drag this formula down at greater than ninety-one withdrawals there is a 10% probability.

Again, how are we going to use this information? Remember the average is eighty, but it doesn't mean every day you are going to see eighty withdrawals. And if you want to make sure that you're dissatisfaction rate from your customers is restricted to 10% or lower then you need to budget for ninety-one withdrawals a day.

>>>

The question is asking, if you started with zero balance, what is the most appropriate amount of cash that needs to be stocked for a four day period. What will be the answer? It will be ninetyone withdrawals times $40 per withdrawal times four days. The appropriate amount of cash that you should be stocking is $14,560 in that ATM.
>>>

Many times there is some confusion about whether a distribution is binomial or Poisson. If the mean or average probability of an event happening per unit of time is given, and you are asked to calculate a probability of n such events happening in a given time, then the Poisson distribution is used. However, if you know the exact probability of an event happening in any point in time, and you are asked to calculate the probability of this event happening k times out of n then you will use the binomial distribution.

>>>

Remember the binomial distribution describes the distribution of binary data from a finite sample. Therefore it gives us the probability of getting r events out of n trials. The Poisson distribution, on the other hand, describes the distribution of binary data from an infinite sample. Therefore it gives us the probability of getting r events in a population.

>>>

We have looked at discrete distributions and we looked at five different kinds of discrete distributions.

- Binomial distributions – when you have outcomes that can have only two values – one zero, yes no.

- Hypergeometric distributions which are a special case of binomial where take a unit out of a sample and you don't replace it.
- Negative binomial – when you are interested in the number of trials that it takes to get to r outcomes. What is the probability that my rth event happens in the nth trial.
- Geometric distribution where we are interested in the probability of the first success.
- Poisson distribution when you are modelling events that can take any value from zero to infinity.

For all of these distributions we looked at how to calculate probabilities of outcomes that we are interested in using mathematical probability density functions or probability distribution formulae. We also looked at how all of these distributions – binomial, geometric, negative binomial, Poisson, etc. can be used in the context of business problem solving.

>>>

We've looked at probability distributions and we started with discrete distributions, now let's look at continuous distributions. And within continuous distributions we will focus pretty much on normal distributions and a special type of normal distribution which is called a standard normal distribution.

>>

Let's start with first understanding what a continuous distribution is? A continuous distribution is generated when you have a random variable that can take any value, not restricted to whole number values but can also have decimal values. Examples of continuously distributed random variables include the height of males in Bangalore; the average waiting time per patient at a particular hospital; the per capita income for a particular location.

The most common kind of a continuous probability distribution is the normal distribution, but there are many kinds of continuous distributions. In statistics normal kinds of continuous distributions are very pervasive, they are used extensively. We will talk about why normal distributions are so popular at a later stage. But for now let us just understand what a normal distribution is.

>>

You can see three different curves here: A, B and C. Which of these is a normal distribution? It turns out all of them are normal. When we say a normal distribution we are not talking about the single curve. In fact if you look at the properties of a normal distribution, a normal distribution is a probability distribution that is symmetric about the single mean, it's mean. Mean equals to the Median equals to the Mode, so the central tendency is very strong. And the two tails extend indefinitely and never touch the axis.

What does the third point mean? Remember in a probability distribution the X axis has the possible outcomes of a random variable. The Y axis has the probability of each of those outcomes. In a continuous distribution the data points can be any value, they're not restricted to whole numbers. Imagine that we had a normal distribution that had values between say 10 and 100. If I were to ask you how many data points are there between 99 and 100, what would be your answer? Actually the answer is infinite. Why? Because between 99 and 100 you could say 99.1, 99.2; and between 99.1 and 99.2 you could say 99.11, 99.12 and so on. In fact there really are an infinite number of points between any two points in a continuous probability distribution.

As long as an outcome is possible can its probability be zero? No, and therefore if the tail actually touch the axis, that means you have a possible outcome whose probability is zero, which we know cannot be the case. That's what it means to say that two tails of a normal probability distribution extend indefinitely never touch the axis.

But what is the difference between these three curves? All of them are normal but clearly they are not the same. In fact if you think about it, to describe a normal distribution what are the parameters that you will need. You will need two parameters. You will need the mean of the normal distribution and you will need the standard deviation. As long as you have the mean and the standard deviation you have all the information that you require to describe the normal distribution. Think about A, B and C. If you had the mean for C and the standard deviation for C, you know exactly what the distribution of C looks like.

>>

If you look at the normal distribution formula, the probability density function you can see that it is actually a function of two parameters, mean and standard deviation. Even though the formula here looks a little scary: one by sigma, square root of two pi, e raised to the power minus x minus mu by two sigma square, the parameters in this formula are still only mu and sigma. We know two is a constant, pi is a constant, e is a constant and x is an outcome that we are interested in.

Let’s use this formula to calculate the normal probability of an outcome that we are interested in.

>>>

Before we do that, let’s just cover one more concept which is the area under the curve.

Remember the total area under a normal probability curve is always one. Why is that? In a probability distribution you have all possible outcomes on the X axis and their associated probabilities on the Y axis. The sum of all possible outcomes, the probabilities of all the possible outcomes has to be one. Therefore we can think of the area under the curve as a probability and you can always compute probability between two values on the curve.

If you wanted to calculate the probability that you had an outcome x that was between one and two how would we do that? It is represented by this area. And how would we actually calculate this? You could calculate the probability of x between one and two as the probability of less than equal to two minus the probability of less than equal to one. That would give us this shaded area.

One more thing that we should emphasize here, which is true for a normal distribution but also for any continuous distribution, in a continuous distribution we never calculate a point probability. You should never say cumulative equals to false in a continuous distribution because there is no such thing as a point probability on a continuous distribution. We always talk about probabilities of less than equal to an outcome or greater than an outcome but never the probability of exactly being equal to an outcome.

>>

Let's look at an example where we apply a normal distribution. Let us say that we are interested in the IQ of Jigsaw students and we test 100 random students and find that the IQ is normally distributed with an average score of 108 and a standard deviation of 7. Supposing we were to pick a random student, how likely is it that that random student has an IQ of greater than 115?

What do we know? We know that the IQ score is a random variable that is distributed continuously and when we look at the score data we can see that it is distributed normally. The average is 108, the standard deviation is 7. We need to calculate the probability of the score being greater than 115. In fact the shaded area in this visualization is the probability that we are interested in. What is the probability that if you pick a student their score is going to be greater than 115?

Now we can use the normal distribution formula and substitute the values of mean and sigma or we could use a tool like excel to actually do this calculation. Remember we are interested in greater than 115 which is essentially nothing but one minus the cumulative probability of less than equal to 115. If you use excel, excel will ask us for some parameters. The parameters are the outcome, the mean, the standard deviation and the cumulative value.

>>>

If we wanted to do this in excel we would simply say one minus normal distribution and excel is listing out the parameters. The outcome that we are interested in is greater than 115, the mean in the data is 108, the standard deviation is 7, and remember cumulative can only be true for a continuous distribution. There is a 15% chance or a 15.8% chance that if you pick a random student their IQ will turn out to be greater than 115 when across Jigsaw students the average IQ score is 108.

>>>

Let's look at a slightly more complex example. Imagine that you are manufacturing a product, say a light bulb and you want to guarantee a performance for this product. When you say guarantee, you essentially say that look if the product doesn't meet the number of performance hours we will replace the product for free. Obviously you want to make sure that the product replacement rate is restricted. Let's say you want to restrict it to less than 5%. What is the number of hours that you should guarantee for your product?

You could take, let's say a sample of 1000 products and you find that the average performance hours in those 1000 products is distributed normally with 71,450 hours being the average and a standard deviation being 2700 hours. Given this information what should be the number of performance hours that you should guarantee in order to make sure that your replacement rate because of failures are less than 5%?

One way to do this is to actually solve this equation. Remember the y is the probability of the outcome, right? The probability of the outcome has to be less than 0.05 and we know the value of mu, we know the value of sigma, we need to calculate the value of x. We could do that using an arithmetic calculation.

>>>

We could also do a trial and error method. Remember what we want is the value of x where the probability of less than equal to x is 0.05. We can do that in excel using a manual method - a trial and error method.

>>>

We know that the average number of hours is 71,450. What is the probability that any random component or any random component or product that you pick will last less than 71,450 hours? We can calculate that in excel using the normal distribution function. The probability that any product that you pick will last less than 71,450 hours can be calculated using the normal distribution function in excel and the parameters that we have from the data. This should ideally be 50% because remember 71,450 is the average. So half your products are going to last less than 71,450 hours and half your products are going to last more than 71,450 hours.

But what we want is the number of hours that make this probability less than 0.05. Let’s just reduce the number of hours by a certain arbitrary number of hours. Say reduce it by a thousand each, so 70000, 69000, 68000, 67000 and calculate the associated probabilities. I'm simply going to drag this formula down because I've linked it up to the number of hours on the left column.

Notice here at 67000 hours the probability that any random product will last less than 67000 hours is 0.049 or very close to 0.05. What does this mean? It means that if the manufacturer guarantees that his product is going to last 67,000 hours, only five times out of hundred he should expect to see a complaint. Only for five products out of hundred will he see performance of less than 67,000 hours which means that he will have to replace the product.

>>>

The value of x that will guarantee a failure rate of less than 5% is 67,000 hours.

>>>

Even this is a tedious way of calculation because it is a trial and error method. There is a better way of calculating that x value that will give us a specific probability and that is done using what is called a standard normal probability table.

>>>

What is a standard normal distribution? We will see that in the next topic.

We were reviewing continuous distributions and within continuous distributions we focused on normal distributions. We will now take a look at one special kind of a normal distribution – the standard normal distributions.

>>>

When we say a normal distribution it is not one curve, it is an infinite number of curves. In fact for any combination of a mean and a standard deviation we can come up with a normal distribution. And so for any outcome x, we can compute probabilities given a mean and a standard deviation from a normal distribution. This is a normal distribution, so is this, so is this, so is this, so is this and so on.
There are an infinite number of normal distributions.

>>>

What we will look at now is one very special case of a normal distribution the standard normal. A standard normal distribution is a distribution with a mean equal to zero and a standard deviation equals to one. Why is it useful? Imagine that we were looking at three different normal distributions.

The first distribution A has a mean of 250 and a standard deviation of 20.
The second distribution, again normal, B has a mean of 1890 and a standard deviation of 135. And the third normal distribution, C has a mean of 0.26 and a standard deviation of 0.03.

These are three different normal distributions. Imagine that I wanted to calculate the probabilities of some very specific outcomes from each of these distributions.

For A I want to calculate the probabilities of less than equal to 210, for B outcomes less than equal to 1620, and for C outcomes less than equal to 0.2.

Now of course we can use a tool like excel or any other statistical software tool to calculate these probabilities using a normal distribution function. For A if I wanted to calculate the probability of seeing an outcome less than equal to 210 from a distribution with a mean of 250 and a standard deviation of 20, I simply in excel have to say normal distribution which is the distribution function for normal, the outcome that I'm interested in 210, the mean which is 250, the standard deviation 20 and cumulative equals to true. This will give me the probabilities of seeing outcomes less than equal to 210. It turns out that probability is
0.02.

What about the second distribution, outcome less than 1620? Again, I simply have to say normal distribution, 1620, the mean is 1890, the standard deviation is 135 and true. This is also 0.02. What about the third distribution probability for outcomes less than 0.2 when dealing with a mean of 0.26 and a standard deviation of 0.03? Outcome 0.2, mean is 0.26 standard deviation 0.03 and true. This is also
0.02.

>>>

The probability of each of those outcomes is 0.02 – do you think that is a coincidence? Actually it is not, because there is a pattern to the outcomes relative to the mean and the standard deviation.

>>>

Can you think of what that pattern is? It turns out each outcome is mathematically the mean of that distribution minus two times the standard deviation of that distribution.

That is for distribution A, 210 is the mean for distribution A, 250 minus two times the standard deviation for distribution A, 40. Similarly for distribution B, 1620 is the mean for distribution B minus two times the standard deviation for distribution B, and similarly for C. In other words each outcome is two standard deviations less than its mean; therefore for any normal distributions you can express the outcome in terms of standardized distance. How many standardized deviations away is it from its mean?

>>>

As long as an outcome is the same standard deviation away from its mean, its probability will be the same. How do we convert to standardized units? You take the outcome minus the mean of the distribution; divide it by the standard deviation of that distribution. And this essentially will convert any normal distributions into a normal distribution with a mean of zero and a standard deviation of one. We are essentially converting a normal distribution to a standard normal distribution.

A standard normal distribution is a normal distribution with a mean of zero and a standard deviation of one. All we are doing is expressing outcomes in terms of distance away from the mean. The distance is in terms of standard deviations of that distribution. As we saw, as long as an outcome is the same standard deviation away from its mean the probabilities of those outcomes will be the same, irrespective of the real mean and the real standard deviation.

>>

One of the reasons that we do standard normal distributions is that the calculation of probabilities becomes a lot easier. In fact the probability distribution function, the probability density function for a standard normal distribution is listed here.

It is one by sigma, square root of two pi, e raised to the power of minus z square by two. If you remember the original normal distributions density function that was a lot more complicated. Arithmetically calculating a distance is a lot more straightforward and then using that in this formula is a lot more straightforward than calculating from a regular normal distribution.

Another reason why standard normal distributions are popular is because of distance measures. Many times statisticians talk in terms of distance from the mean. If you tell a statistician that you have an outcome that is 2.5 distance away from the mean, they understand what it means in terms of probability. The further a distance away from the mean the lower the probability of that outcome. Outcomes that are close to the mean will have high probabilities. Outcomes that are far away from the mean will have low probabilities.

In fact if you remember the normal distributions properties 95% of all observations will lie within two standard deviation of the mean. And 99.6% of all observations will lie within three standard deviations of the mean. If you are told of an outcome which is more than three standard deviations away from the mean you automatically know that this is a very highly unlikely outcome.

>>>

Finally standard normal distributions allow us to use distribution tables. When we calculate probabilities of outcomes we now use tools like excel or other statistical software, but earlier, many years ago before there was easy access to excel, the probability calculations had to be done using the formula. And the formula for normal distribution is fairly complex. Instead what was used was you calculate the probabilities of all outcomes for a standard normal distribution, remember mean is zero, standard deviation is one. And you publish that as a table. And you can convert any normal distribution to a standard normal distribution and then look up the probabilities in this table. Instead of doing the calculations you could simply refer to a probability table.

However, it is easier for us currently to just use a tool than look at a table, but I am going to show you how to use a table so you understand what the values mean. In the table, the majority of the numbers – the main numbers that you can see in rows and columns are essentially probabilities. Remember Z is the outcome minus mean divided by standard deviation, so Z is distance.

The first column shows distance from the mean. 0.1 distance from the mean, 0.2 distance from the mean, 2.3 distance from the mean and so on. The other headings on these columns show the distance in the second decimal place. The first column is one decimal place. If you want to calculate Z to the second decimal place then you will use these columns. These probabilities are cumulative probabilities of less than equal to Z. For example 0.8643 is the cumulative probability of outcomes less than equal to 1.1 standard deviation away from the mean. If you look at this figure here the probabilities are essentially from minus infinity to Z, in other words less than equal to Z.

How do we use this for calculations? >>

Let’s take the IQ example. Average IQ scores for Jigsaw students is 108 with a standard deviation of

7. What is the probability that a random student will have a score of less than 120? What is the probability that a random student will have a score between 110 and 115? What is the probability that a student will have a score of less than 105?

Of course we can do all of these, we can generate the answers to all of these simply by using distribution function formulae in excel or any other tool, but we are going to see how to use a table to come up with these answers. Let’s start with the first question. What is the probability that if you pick a random Jigsaw student their score will be less than equal to 120?

>>

Remember what we want to do here is calculate a distance. We want to convert this normal distribution to standard normal distributions and calculate a Z distance. The Z distance is nothing but x minus mu by sigma. x is the outcome, mu is the mean, sigma is the standard deviation. In our example Z will be equal to 120 minus 108 which is 12, divided by 7 which gives us a Z score of 1.71. Now we use the table to identify the cumulative probability associated with Z equal to less than 1.71. You can see 1.7 is here; second decimal place is 0.01 so 1.71 and the probability of Z less than equal to 1.71 is
0.9564.

Remember Z is nothing but the distance from the mean. When we say 1.71 we are saying a score of

120 is 1.71 standard deviations away from the mean for this distribution. The mean for this distribution was 108, the standard deviation was 7. Therefore this score is 1.71 standard deviations higher than the mean. And the probability of seeing outcomes less than equal to 1.71 standard deviations away from the mean is 0.9564.

We can validate this using excel. If I wanted to calculate the probability of seeing scores less than equal to 120 I will use normal distribution 120, the mean is 108, the standard deviation is 7 and cumulative equals to true. I will get 0.956 which is exactly the value that we see in the table.

>>

If you don't want to use a tool or you don't have access to a tool, and you want to quickly calculate probabilities associated with a normal distribution, one simple way is to calculate a Z score and look up that Z score associated probability in a distribution table of a standard normal distribution.

Let's do the second question – what is the likelihood that a random student will have a score between 110 and 115? We want the Z scores between 110 and 115. So first we convert 110 to a Z score and that will be 0.28, 110 minus 108 divided by 7, so that's 0.28. And 115 converted to a Z score which is 1, 115 minus 108 divided by 7. We want the probability of Z between 0.28 and 1. How do you think that can be done?

>>>

If you visualize this – if this is your distribution, this is Z equals to 1. And this is Z equals to 0.28. Remember these are distances from the mean. The first 0.28 is 0.28 distance away from the mean of 108, and the second outcome is one standard deviation away from the mean which is 108. We want this probability. How do I actually get that? I simply do probability of Z less than equal to 1 minus the probability of Z less than equal to 0.28. That will give me the in between probability. If I look it up in the table I need probability of Z less than equal to 1 minus the probability of Z less than equal to 0.28.
>>>

Less than equal to 1 is 0.8413 minus the probability of less than equal to 0.28 is
0.6103.

>>

Therefore the probability that I am looking for is essentially 0.84 minus 0.61 which is 0.23.

>>

Finally, last question, which is slightly more complex. What is the probability that a random student will have a score of less than 105? Now of course first step, calculate the distance and the distance turns out to be 0.42. What does it mean to have a negative Z score? Actually it simply means that this outcome is to the left of the mean, lower than the mean. The negative distance is simply telling you that you have an outcome that is less than the mean. We have an outcome that is -0.42 distance away from the mean on the left side of the distribution, left of the mean.

Now how do I calculate probability of Z less than equal to -0.42. In the table we only see positive values of Z. In order to calculate that we remember the fact that a normal distribution is a symmetric distribution. The probability of Z less than -0.42 should be the same as the probability of Z greater than 0.42. How do we calculate the probability of Z greater than 0.42? It is imply 1 minus the probabilities of Z less than equal to 0.42.

To put this all in one line the probability of Z less than –0.42 is the same as the probability of Z greater than 0.42 which is the same as the probability of Z less than equal to 0.42. So what is the probability of Z less than equal to 0.42 – I can look that up in the table 0.6628. So one minus that is 0.3372 which is the probability of Z less than -0.42 or in other words the probability that a random student that you pick will have a score of less than 105 when Jigsaw students on average have means of 108 with a standard deviation of 7 is

33%. This is how a standard normal distribution table is used. But remember, most of us will end up using a tool because it's very convenient and we can do the calculation much faster with access to a tool. If you don't have access to a tool, or you want to use the distribution table using a distance measure then you will look up a table like this.

>>>

We have looked at a variety of probability distributions including discrete distributions like binomial and Poisson, and continuous distributions like normal. What we are going to do now is to understand how probability distributions can apply to business situations using the case study that we've been reviewing, the telecom churn data set.

>>>

The telecom churn data set that we've looked at includes data on subscribers of telecom services. It has data on their usage. It has data on their service plan. It has data on their promotion participation. And it has some demographic data about the customer including when they started the service, what their birth date is, where is there location, and so on.

I'm going to use the same data set to illustrate how distributions are used to solve some specific business questions. We'll start with binomial.

>>>

The average promotion participation rate for these customers is 10%. This 10% that we have got is from data.

>>>

If we go back to the data set remember we have promotion data on customers. How does the promotion data work? In any month, if this customer participated in a promotion they would have a 1 value otherwise they would have a zero value.

One way for me to calculate average participation rate is to add up the number of 1's in each row, which is the number of times they have participated in a promotion divided by 21 which is the total months of participation tag is allowed. So we have 21 months of data. This is a rough calculation because you can see that not every customer was included in every month – they may have left. But as a rough calculation this is 14% participation rate, essentially means that this particular customer participated in only a few promotions in the 21 months, and so on.

If I take that for every customer and I just do an average for all customers I will get an average promotion participation rate across all my customers which are equal to 10%. I simply took an average of this column, so that's where the 10% is coming from for our problem.

>>>

The average promotion participation rate across these customers is 10%. On a new promotion that was sent to 15000 customers you received 1650 people signing up for the promotion. This is higher than the 10% that you expect which should have been 1500 customers. The question is, is this likely due to random chance variation? Or does this imply that the sign-ups for your new campaign are really higher than expected? Remember this is a binomial distribution problem because the random variable that we are dealing with, which is whether or not a customer signs up has only two outcomes – yes they sign up, or no they don't sign up.

Given a binomially distributed random variable we should be able to calculate the probability of an observed outcome. In this case we want to find out how likely is it that if we were expecting a 10% success rate we will see 1650 sign-ups in 15000 customers simply because of random chance. The way to do that is to simply say binomial distribution, the number of successes is 1650, the number of trials is 15000, the probability of success is 0.1 and cumulative equals to false.

Remember this will give us the exact probability of seeing 1650 successes. That may not be very useful actually because what we ideally want to look at is how likely is it that I will see more than 1650 or more sign-ups, in which case I will do cumulative probability of less than 1650 sign-ups. One minus binomial distribution, 1649, 15000, 0.1 and true. Remember this will give me the probability of seeing 1650 or greater sign-ups, which even is a very, very low number. Remember this number tells us that the random chance probability of seeing 1650 or higher sign-ups is very, very, very, very low. Therefore it is likely that your new campaign is a success, because more than your average number of participants has actually signed up for this promotion.

>>
Let's look at a Poisson distribution application now. For the same data set imagine that the number of

applications that were received in a day for new subscriptions in your most popular office is 150. Remember a Poisson distribution is a discrete distribution where you know the average number of events that are occurring in a given time frame. In this case, the number of applications received in a day on average is 150 in your most popular office.

Remember an average of 150 does not mean everyday 150. Some days a lot less than 150 - 140, 145, and 130 maybe, and in some days there could be a lot more than 150. If there are a lot of customers that come to your office for applications, you may run out of application forms. And you don't want to turn away more than 10% of your customers. You don't want to turn away customers more than 10% of the time because you don't have application forms.

So the question is: what is the number of application forms that you should have in the office in order to make sure that the chances of turning away your customer is less than 10%? Maybe there is a lead time to order application forms and therefore you don't want to run into a situation where you're constantly out of forms. So what is the number of forms that should be in your office in order to make sure that you turn away customers only 10% of the time?

We can do that using a Poisson distribution formula because we want to calculate how likely is it for x number of customers the chances of seeing that number of customers is less than 10%. We can calculate this for Poisson distribution using a

Poisson distribution function. So for example, what is the likelihood that I will see more than 151 customers in any day? Simply one minus Poisson distribution, 151 when expecting 150 and true, because I'm doing more than 151 customers. So there is a 44% chance that on any given day I will see more than 151 customers.

What do I do now? I will just drag this formula down for all these total number of customers that I have listed, which is simply all the way up to about 180 and I will try and find that number where the probability of seeing greater than that many customers is less than 10%.

So that is here. What does this mean? This essentially tells us that the probability of seeing greater than 166 customers in any given day is less than 10%, which means that if I stock a 165 form sin my office then I will turn away my customers less than 10% of the time. The chance of seeing less than equal to 165 customers in any given day is 90%, the chances of seeing greater than 166 customers is 9%. So if I stock 165 forms in my office then I am good. I will not turn away more than 10% of the time my customers because I don't have an application form.

>>>

Finally let's look at a normal distribution application. You look at the distribution of total usage across your customers, which are remember we have usage minutes. Now supposing you add up all the usage minutes and you look at a distribution of that usage minutes. You find that it is normally distributed with the mean of 280 and standard deviation of 65. You are trying to identify an outlier customer that your CEO would like to talk to. They want to talk to customers who are very, very different from your regular customers – either a lot higher than normal customer usage or a lot lower than normal customer usage.

So how do you identify these outlying customers? Remember an outlier is a customer whose behaviour is unlikely because of random chance variation. Their behaviour is very, very different from your regular customers. If you remember normal distributions, you will remember that approximately 95% of all data points for a normally distributed variable will lie within 2 standard deviations of the mean. What does that mean? It means only 5% of observations will be more than 2 standard deviations away from the mean. The chances of this happening due to random chance variation will only be less than 5%.

How do I identify this cut-off? Mean plus 2 standard deviation is 410, mean minus 2 standard deviation is 150. In other words, the probability of seeing customers with less than 150 minutes of usage or greater than 410 minutes of usage cannot be more than 5%. How do I know this? I can validate that – what is the probability of seeing customers whose usage is less than 150. Simply have to do normal distribution, 150, mean is 280, standard deviation 65 and true. This is the probability that customers will have less than 150 minutes of usage because of random chance variation – 2.27%. Similarly the probability of seeing greater than 410 minutes of usage – one minus normal distribution, 410, 280, 65 and true.

If we add this up, this will add up to less than 5% because this data is normally distributed and therefore it has to be that less than 5% of the data will lie outside of the 2 standard deviation minutes. If you want to identify outlying customers for your CEO to talk to you should look for customers whose usage is greater than 410 minutes or less than 150 minutes.

>>>

A quick recap of what we've covered so far in this statistics module. We started with an understanding of what is statistics and we looked at two kinds of statistics – summary statistics or descriptive statistics and inferential statistics. Descriptive statistics are used to help us understand data at a high level. You get some basic trends and information contained in data especially when we have data sets that are very, very large. Common descriptive statistics that we looked at include the mean, the median, the mode, the standard deviation, the shape of data – whether it is skewed or symmetric, and
[inaudible].

We then started looking at inferential statistics which is using a sample to make inferences about a population. In order to understand inferential statistics we reviewed concepts of samples and populations, what is a representative sample and how to choose a representative sample.

We then looked at probability theory and simple probability calculations, because we will work with random variables and random variables have probability distributions. We looked at what a probability distribution is, we also looked at types of distributions, essentially discrete distributions and continuous distributions.

The discrete distributions that we reviewed include binomial and Poisson and other examples. The continuous distribution that we reviewed was primarily normal distribution which is very, very popular in statistics because of some properties of the normal distribution which make it useful for inferential statistics.

>>>

In this section we'll take a look at how we do probabilistic computations inside R. There are two types of probability functions inside R which are relevant for our course. The first set of functions are called the density functions, the second set of functions are called the cumulative density functions. The density functions are like the excel functions, where we provide FALSE as the input to the cumulative parameter. The cumulative functions on the other hand in R are similar to the excel functions where cumulative is passed as TRUE in the form of a parameter.

Let us start our discussion with first probability distribution which is the binomial distribution and let us take a look at how we compute densities using the relevant density function. As far as binomial distribution is concerned the relevant distribution function is termed as dbinom – the d here stands for density, binom here stands for binomial distribution. dbinom takes these parameters as inputs. There are three parameters which are taken as inputs. The first parameter is the number of successes, the second parameter is the number of trials and the third parameter is the probability of success.

Suppose we toss our coin ten times and we are interested in finding out the number of heads that show up. Since this is a binomial process so we can use binomial distribution to model the probabilities. Let us assume we are interested in finding out the probability that exactly three heads show up when we toss a coin ten times. We will use this particular function dbinom to compute this probability. Since seeing heads is termed as success here, so there are 3 successes, 10 total number of trials, and the probability of success in each trial is 0.5 so we supplied 0.5 for the probability of success parameter.

Let us execute this command; this gives us the probability that we'll see exactly three heads. This highlighted part would be the excel counterpart for this R function. In excel we use binom.dist to compute the probability when we are talking about data following binomial distribution. You can see, to do the same computation in excel I'll have to supply the parameter false. The reason is I am computing a density here.

One of the good things about using R is that we can supply vectors as inputs to the function. Now suppose I want to find out the probability that I'll get one head, two heads, three heads, and so on till ten heads, then what I can do is I can supply a vector of first ten numbers and this particular command will give me the density at each of those points. Let us execute this command. As you can see now I have ten values corresponding to probability that I'll see exactly one head, exactly two heads, exactly three heads, and so on.

Let us know take a look at how we would compute the cumulative probability distribution, or how we would make cumulative probability computations. The way to compute cumulative probability is to use a function which starts with the letter p. Since we're interested in finding out cumulative probability for binomial distribution the relevant function will be pbinom. The parameters of this function are similar to those of the parameters of dbinom; both take similar arguments – number of successes, number of trials and probability of success in each trial.

Suppose I am interested in finding out the probability that I'll observe up to three heads. The way to compute this probability is to use the function pbinom, number of successes is 3, number of trials is 10, and probability of success in each trial is 0.5. Let us execute this command. We can see that the probability is 17%.

Again we can supply a vector to this particular function. Let us take a look at that. What this output tells me is that there is around 1% chance that I'll see up to 1 head, there is around 5% chance that I'll see up to 2 heads, there is around 17% chance that I'll will see up to 3 heads and so on. Also note that the excel counterpart for cumulative probability functions take a parameter true in the place of cumulative.

Let us know discus another probability distribution which is the negative binomial distribution. Now the density function for negative binomial distribution as I have reiterated in the previous examples starts with the letter d. So when I say dnbinom it means that I am talking about density of negative binomial distribution. The parameters that this function takes are number of failures, number of successes and probability of success. These parameters are very similar to the parameters passed to the excel counterpart of this particular function.

Suppose I toss five coins and I am interested in the probability that the second head occurs in the fifth trial, then I can use this particular function to compute the probability. Since the second head occurs in the fifth trial, this means when I conduct my fifth trial there would have been three failures. Since there are only two successes I am applying two as the number of successes. The probability of success in each trial is going to remain 0.5. Let us execute this command. This gives me the probability that I'll see second head in the fifth trial.

This is the excel counterpart for this particular command and you can see I have supplied a parameter false because I am finding the density here. The cumulative probability function would start with p and for negative binomial distribution the function would be pnbinom. The parameters remain the same – number of failures, number of successes and probability of success. This particular command would tell me the probability of observing up to two heads in fifth trial. The excel counterpart for this command will be negbinom. dist and as you can notice I have supplied the parameter true.

Let us discuss the next probability functions which belong to the category of probability distributions termed as hypergeometric distributions. The hypergeometric distributions are used whenever we do sampling without replacement. In order to compute density, the relevant density function will be dhyper. These are the parameters that this function takes. The first parameter is the sample number of successes, the second number is the population number of successes, the next parameter is the population number of failures, and the last parameter is the sample size.

Let us assume you want to find the probability that we see 2 reds when 5 cards are drawn at random from a deck without replacement. Since we want to see 2 reds, the sample number of successes would be 2. There are 26 red cards so population number of successes is 26. Also there are 26 black cards, so population number of failures is also 26, and since you are taking 5 cards at random the sample size is 5. Let us execute this statement. This gives us the probability that we'll see 2 reds when 5 cards are drawn at random from a deck without replacement.

The relevant cumulative distribution function would start with p and for hypergeometric distribution the function would be called phyper. The parameters remain the same both for dhyper as well as phyper. Let us execute this statement. This probability represents the probability of observing up to 2 reds when 5 cards are drawn at random from a deck without replacement.

Let us discuss the next distribution which is the Poisson distribution. The density function for the Poisson distribution is called dpois and it takes two parameters. The first parameter is x, the second parameter is mean. The x here refers to the observed rate. Suppose we want to find out the probability that we will observe a rate of 4 when on an average the rate is 6. We will use this command dpois, 4 is the observed rate, 6 is the on an average population rate. As you can see this is the excel counterpart for this function. Again you would notice that I have supplied false as the parameter because I am computing the density.

The cumulative probability function for Poisson distribution would be ppois – p stands for cumulative probability. Again if I am interested in finding out the probability of observing up to a rate of 4, when on an average the rate is 6, then I will use this command ppois(4,6). This probability is what I get. The excel counterpart would be this particular function and as you can notice I have supplied true as the parameter.

Let us know discuss how we compute cumulative probability distributions when a data is normally distributed. Since we want to find out cumulative probability distribution for normal data our function should start with prefix p. The relevant function will be pnorm. p stands for cumulative distribution and norm stands for normal distribution. Let us say we are interested in finding out the probability that we will observe a number less than or equal to 1.65 when our mean is zero and standard deviation is one. The first argument for this function is the observed data value. The second argument is the mean of our population or of a data. And the last argument is the standard deviation of a data. Let us execute this command and we get a probability of 95%. This is the excel counterpart for this particular function.

>>

In this section we've seen how we make probabilistic computations inside R and we figured out that there are two types of functions. The first type is the density functions and the second type is the cumulative density functions. We've also seen what are the density and cumulative density functions for binomial distribution, negative binomial distribution, hypergeometric distribution, Poisson distribution and normal distribution.

>>>

Inferential Statistics

Let's now look at how inferential statistics could be applied to the Bank telemarketing case study.
Inferential statistics deals with understanding random chance variation and looks at how to quantify amount of random chance variation in a variables outcome. And it uses this knowledge to make inferences about a larger population based on smaller samples, because a sample is like a random variable from an underlying population.
In the Bank telemarketing dataset, what are random variables?
We can argue that the sample itself is like a random variable because this is not all the customers of the Bank. This is a subset of the Bank's customers. But within the sample some of the attributes are also random. For example, the outcome variable which is whether or not a customer signs up for a term deposit offer. While there are many factors that influence the outcome, we can never be 100% sure why one customer decides to sign up and another customer doesn't decide to sign up. There is some randomness that is still inherent in that outcome. Similarly call duration. When an agent makes a call, he cannot be sure how long the call will last. He may have a fair idea that it will last 30 seconds, 2 minutes, more than 3 minutes, less than 10 seconds. But he will never be sure exactly how long the call will last.
So these are all examples of random variables in the data set.
Remember in this case study, we are trying to determine attributes that influence the outcome of whether or not a customer will sign up for a term

deposit. So to some extent, we are trying to quantify those impact of variables that allow us to have a good handle on what the outcome is. But remember even if we include all possible factors, we will never be 100% sure or 100% accurate in our prediction because there will still be some amount of random chance variation. The idea is to capture as much of the non-random chance variation as possible so that we can be reasonably sure of the outcome, but we should not expect 100% accuracy. What sort of questions can we answer by applying the inferential statistics concepts that we have just learnt?

For example:

- What is the probability that a random customer will end up signing for a deposit?
- What is the probability that if an agent makes a call, it will last less than 200 seconds?
- What is the probability that if a customer is chosen at random, that customer will end up having both a housing loan and a personal loan?
- What is the probability that a random customer will have a personal loan given that he already has a housing loan?
- What is the likelihood that a customer will sign up for a term deposit given that he has responded negatively to the last campaign?

Again, all of these can be easily calculated in excel.

So the first question says- what is the probability that a random customer will end up signing for a deposit?

Remember probability of an outcome is in the numerator the number of times the outcome can happen over total number of outcomes. In this case, the number of times the customer has said yes over the total number of customers. This as we saw when we ran the descriptive statistics as well is 12%.

What is the probability that a random call will last for less than 200 seconds?

Now remember we have calls that last anywhere between 0 seconds to many minutes. We want to count how many times in this dataset has the call lasted less than 200 seconds and divide that by 45211. So here, we will use the countif function to count how often the duration variable has the value of less than 200.

So countif(L2:L45211) which is the duration variable has a value of less than 200 and divide that by 45211. It turns out that this probability is 55%.

What is the probability that a random customer will have both a housing loan and a personal loan?

The probability that a customer will have a housing loan is 56%. Here we count the number of customers that have a housing loan divided by total number of customers- turns out to be 56%. Then what is the probability that a customer will have a personal loan? This is the number of times a customer has a personal loan 'yes' divided by 45211. Remember these two variables – personal loan and housing loan essentially have yes/ no values. Now we want the combined probability that a customer will have both a housing loan and a personal loan. So this is the probability of housing times the probability of personal which is 9%. We can check this as well. If you just go here and choose with a filter option housing 'yes' and personal loan 'yes' and we look at the number of observations, we get 4103. And if you divide 4103, we will get 9%.

What is the probability that a random customer will have a personal loan given that he already has a housing loan?

This is an example of conditional probability.

The number of customers with housing loans is 25130. Of the 25130, the number of customers that have personal loans is 4368. So the conditional probability of a person have a personal loan given that he already has a housing loan is 17%. It is 4368 divided by 25130. Remember the denominator is not 45211 because this is a conditional probability problem.

What is the probability that a customer will sign up for a term deposit given that he has responded negatively to the last campaign? Again conditional probability. The number of customers that responded negatively to the last campaign is 4902. The number of customers with success in this campaign from the customers that responded negatively to the last campaign is 619, which translates to a 13% conditional probability that a customer has said 'no' in the last campaign and has said 'yes' in the current campaign.

We have looked at random variables and calculated probabilities. But we have also talked about random variable distributions. If we know what kind of a random variable we are dealing with, we can figure out the distribution of that random variable and then we can calculate outcome probabilities.

Let's look at applications of random variable distributions in the Bank dataset.

Now if we look at the outcome variable itself, it's a yes/no variable. So what kind of a random variable is that? It's a discrete, binomially distributed random variable. And we already know that the binomial distribution has a formula that we can use to calculate outcome probabilities. Applications of this, for example, could be:

What is the probability that if we call 10 random customers, 5 will sign up?

What is the probability that if we choose 15 customers, at least 10 of them will be younger than 40?

So how do we calculate this?

We will use excel, because excel has built in distribution functions. So we don't have to manually calculate these probabilities.

Let's look at the first question- what is the probability that if we call 10 random customers, 5 will sign up?

Remember the probability of success here is 12%. We calculated that by saying overall 12% of customers in this dataset have said yes. So the probability of 5 successes in 10 trials can be calculated using the binomial distribution formula as binomial distribution 5, number of successes out of 10, the number of trials 0.12, the probability of success and false which is 0.003.

The next question says- what is the probability that if I choose 15 customers, at least 10 of them will be younger than 40?

First we should find - what is the likelihood of customers being less than 40 in the dataset. That is equal to 0.517. Remember I get this probability from the dataset. I simply count the number of customers whose age is less than 40 and divide that by 45211. So that probability I get as 0.517.

Now I want to find the likelihood of at least 10 customers out of 15 being less than 40. Remember at least 10 essentially means 10 or 11 or 12 or 13 or 14 or 15 out of the 15 be younger than 40. So the probability of 10+11+12+13+14+15 is the same as 1- the probability of less than equal to 9. And we can do that as 1 – binomial distribution (9,15,0.517 which is the probability of success and true). Because the probability of <=9 is a cumulative probability. So if I do this in excel I will get 17%. So if we choose 15 customers, there is a 17% likelihood that at least 10 of them will be younger than 40 years old.

Slide 1

In this second module of statistics we have reviewed probability distributions and how to calculate a probability of a random variable outcome based on the appropriate probability distribution.

Slide 2

There are two kinds of probability distributions.

- Discrete distributions where the outcomes can only have whole number values
- Continuous distributions where the random variable can have continuous outcomes

Slide 3

The examples of discrete distributions that we have reviewed include

Binomial or the Bernoulli distribution which can have only two possible outcomes. For e.g. will a customer buy my product or will not buy my product. Will a customer default on a loan or will not default on a loan. You have random variables that can only have two possible outcomes that are discrete.

In that case the probability distribution of the outcomes of this random variable will follow a binomial distribution. There is a probability distribution function defined for a binomial distribution, which is essentially

$$P(x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

Where *n* is the number of trials for the random variable, *x* is the outcome that we are interested in, *p* is the probability of success and (1-p) is the probability of failure.

We could use this formula or we could use statistical tool like excel or SAS or R to calculate the outcome of a binomial random variable. We also looked at special cases of binomial distributions including negative binomial and hypergeometric and finally we also looked at what is called a poisson distribution which is a discrete distribution but the outcomes can range from zero to infinity.

For examples the number of claims filed in a month, the number of calls that a call centre use in a day. The probability distribution function of the poisson distribution is

$$P(X=x)=\frac{\lambda^x e^{-\lambda}}{x!}$$

Well λ is the average occurrence in the given time period and *x* is the outcome that we are interested in. Again we can calculate this poisson probabilities very easily using a tool.

Slide 4

We finally also look at a continuous distributions. There are many kinds of continuous distributions we focused on a normal distribution. This is because a normal distribution occurs very frequently in natural life. It is also very frequently applied in inferential statistics for another reason. Normal distributions are symmetric. The probability function for the normal distribution function is

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Where μ is the mean and σ is the standard deviation.

Remember in a continuous distribution we do not actually calculate point probability of an outcome. We can only calculate and report cumulative probabilities. Probabilities of all outcomes <= *x*. so we can do that using the normal distribution function or using a tool. We also looked at standard normal distribution which is essentially a way of standardizing a normal distribution to express it in terms of distant from the mean.

It is done to make calculations easy for using a table and it also intuitively appealing because you are express the outcome in terms of distance from the mean and further the distance from the mean, we know the lower the probability. So it gives us the good sense of how likely or not a particular outcome is given its distribution.