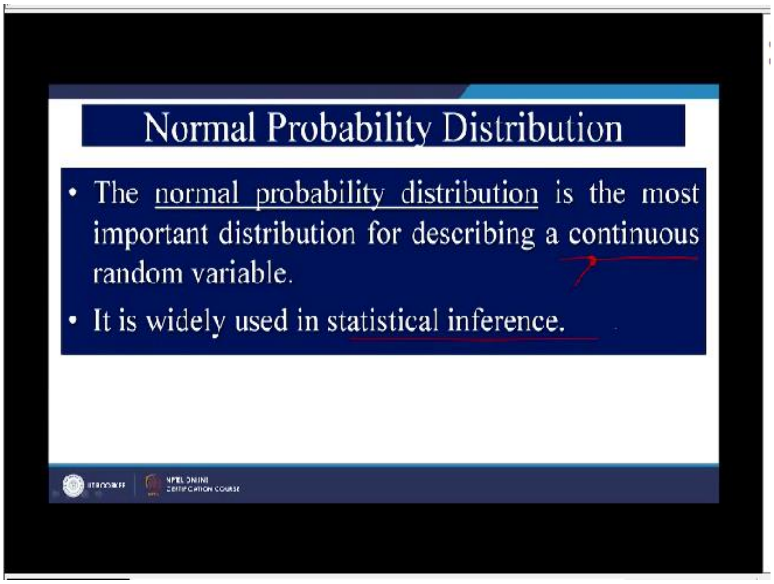
Business Statistics
Prof. M. K. Barua
Department of Management Studies
Indian Institute of Technology-Roorkee

Lecture-18
Probability Distributions-part III

Good morning friends as you are aware in previous session we discussed discrete probability distribution wherein we have seen binomial distribution, poisson distribution, hypergeometric distribution and we have seen continuous distribution as well and the distribution which we have looked at was uniform continuous distribution. Today will see one more continuous distribution which is normal distribution, out of all the distributions available I would say that this the most important one.

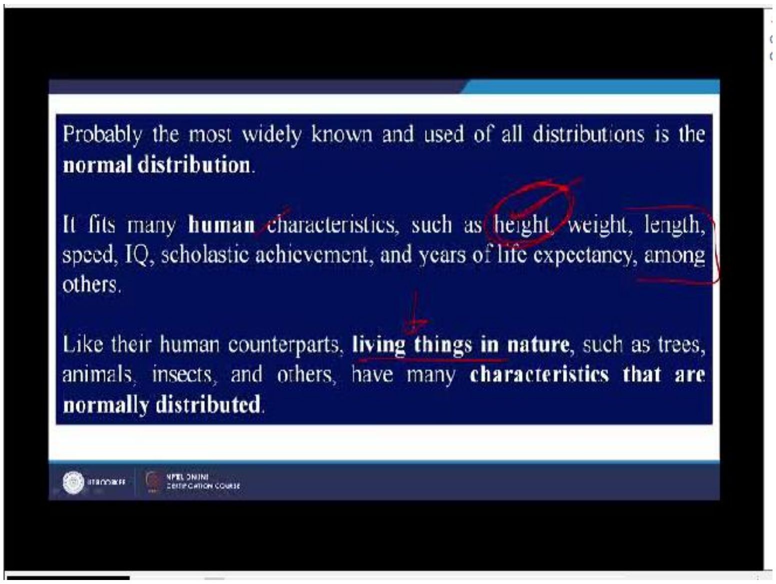
(Refer Slide Time: 01:10)



It has got larger applicability compared to other distributions and it is a distribution which is applicable in variety of situations. So, yes we know that normal distribution is most important distribution for describing continuous random variable right, in discrete distribution we have seen that a discrete random variable was taking a discrete value right. It is widely used in statistical inferences, so as I have already said as most widely known and used distribution.

It fits several human characteristics be it height, be it weight, be it income, be it length, be it intelligent quotient are a scholastic achievement and years of life expectancy and there are many more.

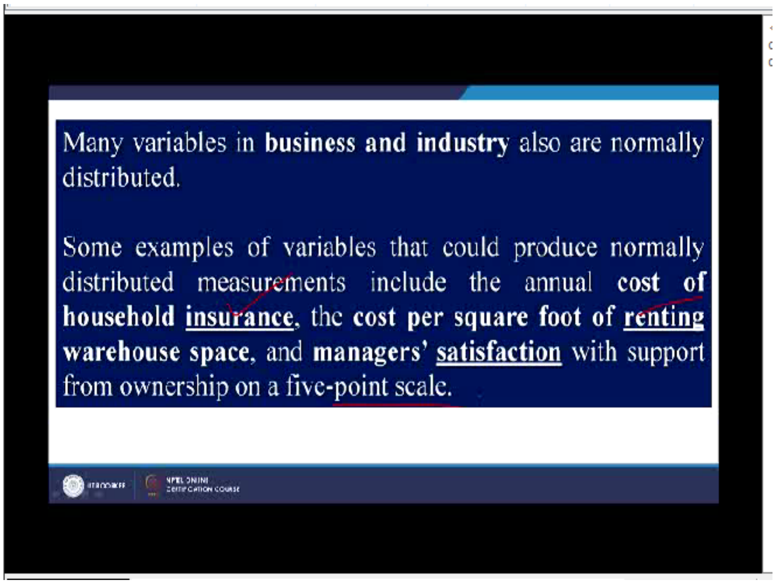
(Refer Slide Time: 02:09)



You take any human characteristics this binomial this normal distribution will fit the particular characteristics and this not only true in just one country or one region, its applicable across the world. And if you want to let say apply normal distribution height of let say height of Indians and if you want to know whether that height will follow normal distribution or not yes it will follow.

If you take height of all the people in the word then will it follow normal distribution yes it will follow right. So, it is applicable everywhere is not only applicable to human characteristics but to the characteristics of living things as well, be it life of tree or let say let say life of insects or animals. So, this normal distribution it exhibits characteristics of living things as well apart from human characteristics and living things characteristics.

(Refer Slide Time: 03:41)

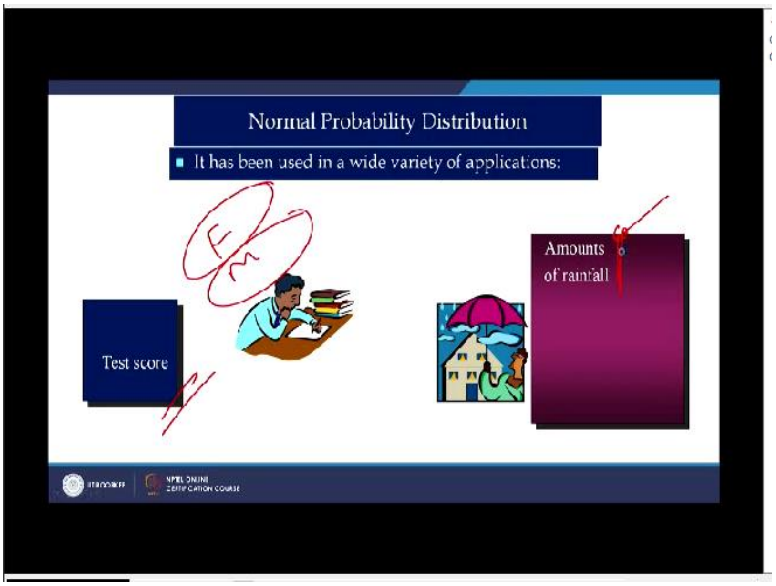


Many variables related to business and industry are also normally distributed, again let say the income of companies let say the life of companies. So, all these will fit normal distribution, apart from the let say the insurance premium, let say the cost per square footage of renting warehouse space, manager's satisfaction with support from ownership on a five-point scale. So, all these variables and other variables related to industry will also fit in normal distribution. It has got wide variety of applications, we have already talked about heights of people.

(Refer Slide Time: 04:38)

So we will know that within 2 sigma limits approximately 95% population will fall right or the height of the population will be 95% within 2 sigma limit. Similarly scientific measurements test scores.

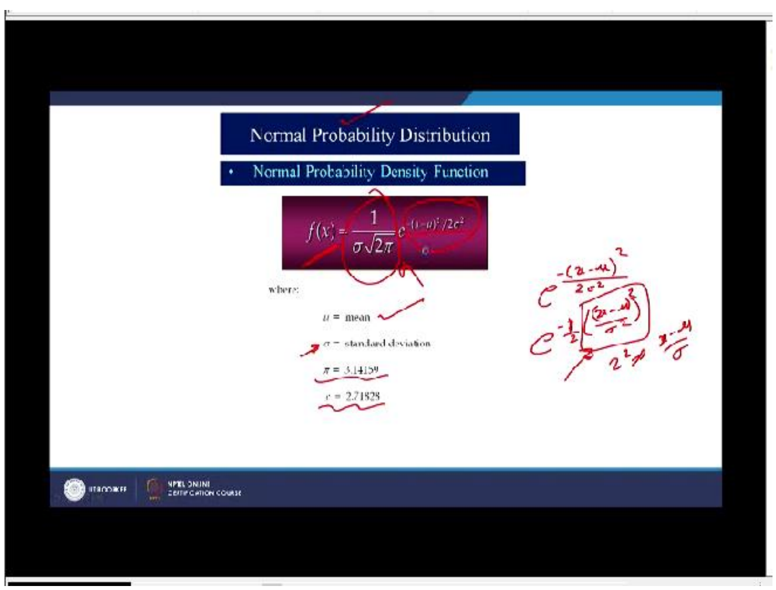
(Refer Slide Time: 04:58)



Let say there are 12 lakh students taking one particular examination and if you compare their marks once you have got the marks of all those students. Then you will find that their marks would also be normally distributed, whether and if you do it let say for female students and for male students. Then also you will see that the marks of female students would be normally distributed and again for male students as well.

Amount of rainfall again normally distributed be it in a country be it in district be any region or worldwide. So, you will find that the amounts of rainfall is also normally distributed.

(Refer Slide Time: 05:57)



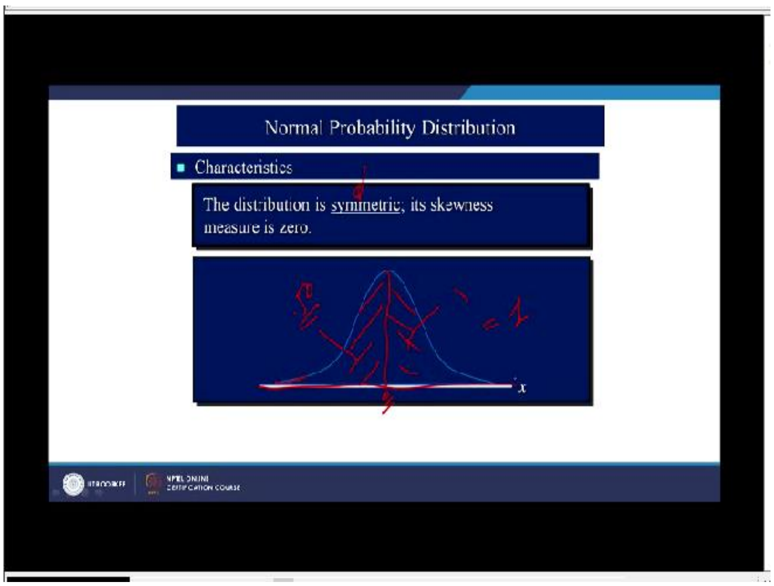
Let us look normal probability density function similar to other density functions which we have seen, so will have density function mean and variance for normal probability distribution as well.

So, normal probability density function, $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$

So, here µ is mean, σ standard deviation, pie 3.14, and e is 2.71.

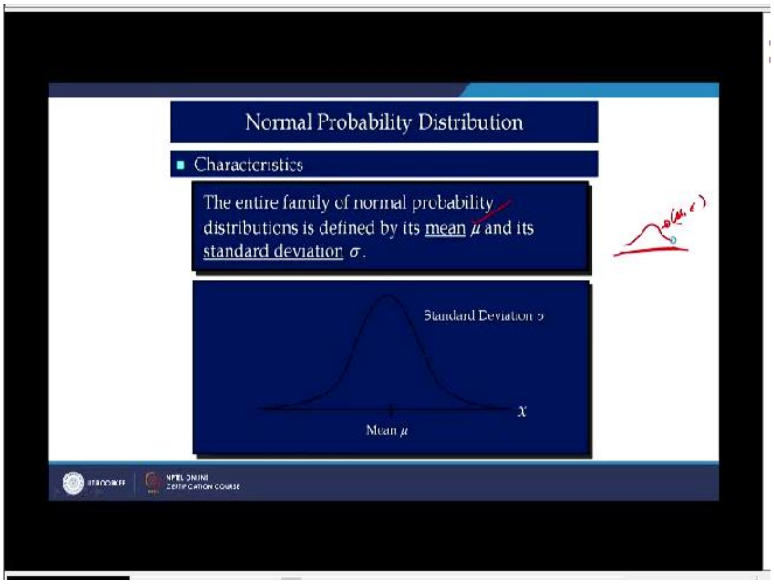
Now if you look at this term carefully, so this is e x- µ whole square is 2 Sigma square right this is nothing but z score also this is you can write it like this 1/2 (x- µ) Sigma square right. In other words this is z square is not it this is the formula for z, z = (x- µ)/Sigma. So, we will see this expression little later, so normal probability density function is given by this equation.

(Refer Slide Time: 07:37)



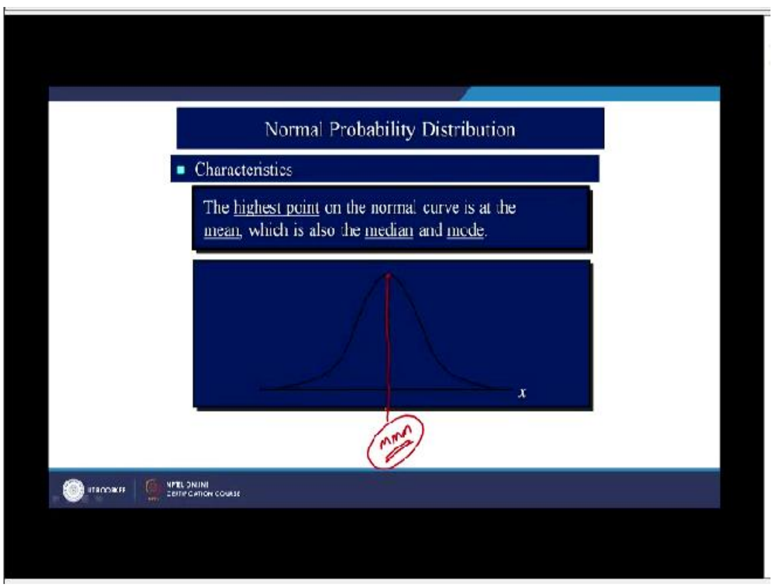
Now let us workout couple of examples and there are some more characteristics of normal probability distribution. The distribution is symmetric its skewness measure is 0, so if you look at this its measure is 0, so this is a symmetric curve it is not a skewed curve right. So, area in this region is 50% or 0.5 in this region is also 0.5 area under this curve is unity right, 1 right. So, it is a symmetric distribution one of the characteristics of normal probability distributions. So it is symmetric right, but this distribution, this curve never touches this x-axis ok.

(Refer Slide Time: 08:32)



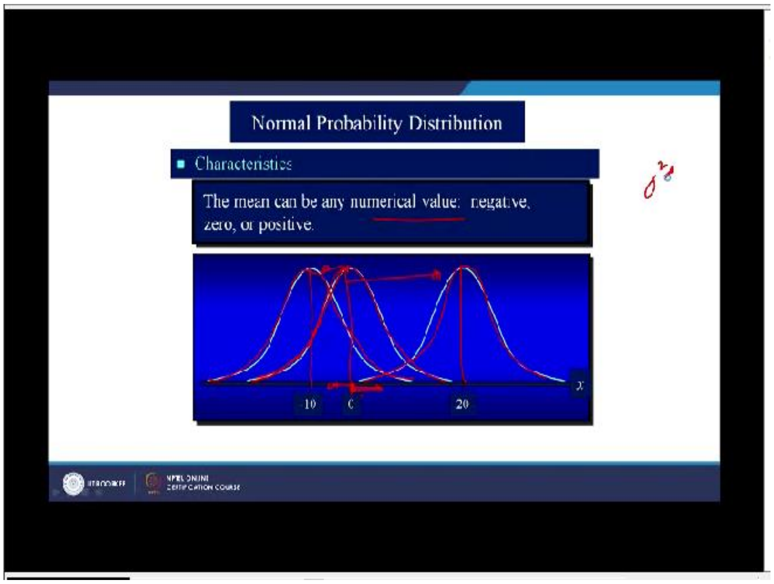
The entire family of normal probability distribution is defined by its mean and standard deviation, so this is a distribution a normal distribution and if you want to define normal distribution then it should always be define in terms of mu and standard deviation right. So, this is how you can define a standard distribution curve right.

(Refer Slide Time: 09:05)



The highest point on the normal curve is at the mean which is also median and mode, this another characteristics of normal distribution. So, this is the highest point and it will be so mean, mode and median all will be same and that is one of the properties of normal distribution.

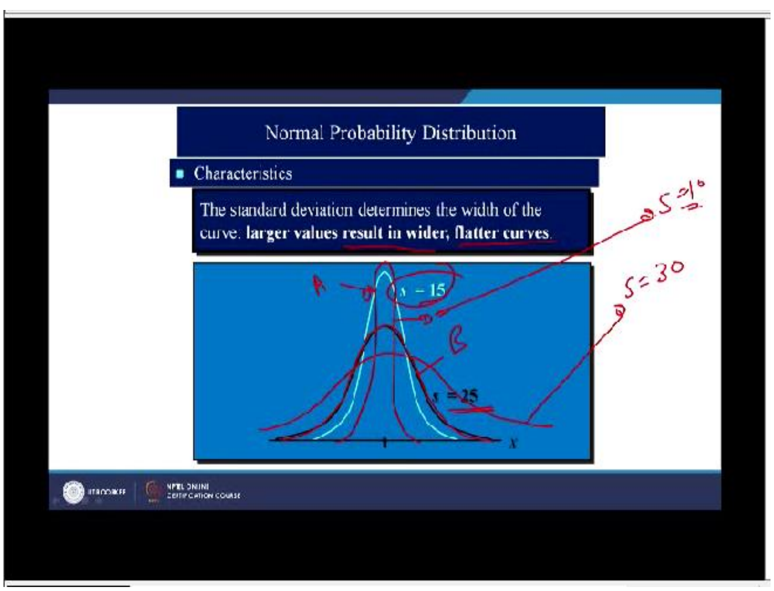
(Refer Slide Time: 09:33)



The mean can be any numerical value it can be negative 0 or positive. So, let say this distribution right, this distribution, this mean is 0 is not it. So, all the values towards right 20 would be positive and values towards left hand side would be negative. So, this particular distribution mean is 0 right now if you change the distribution from here to here, then its mean is now negative -10 right.

Similarly if shifted in this direction, so this is now mean is 20 ok, so it can have mean 0 positive or negative right. But its variance can never be negative because you calculate variance by just squaring standard deviation right.

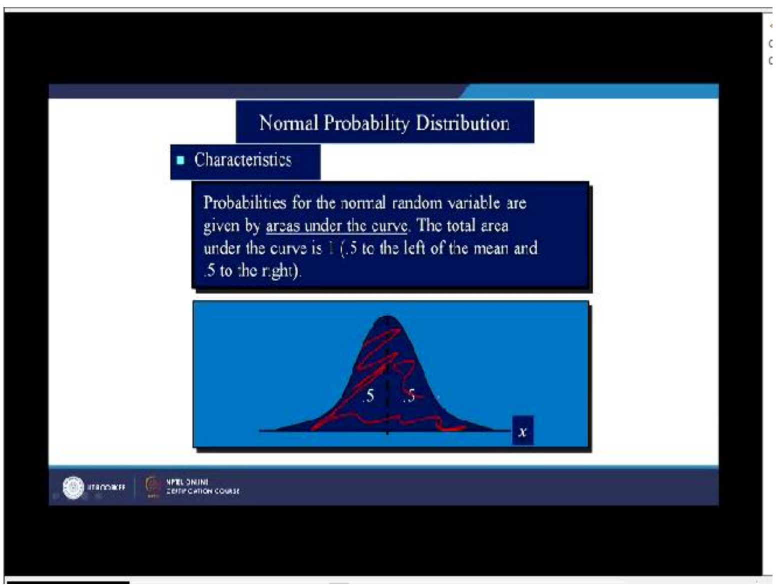
(Refer Slide Time: 10:46)



The standard deviation determines the width of the curve larger values result in wider flatter curves, so this is standard deviation 15 for let say this is curve A this curve B. So, for curve B standard deviation is higher than curve A. So, it will always be wider curve and flatter curve right, let us draw one more curve, curve C right and standard deviation is let say 30. So, how that curve would look like, so that would be like this isn't it.

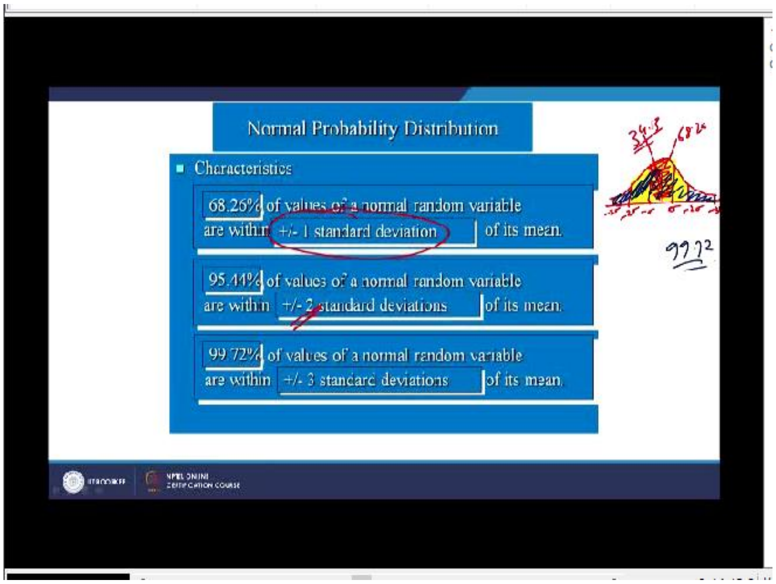
If S is let say 10 how that curve would look like that would be like this is not it, so this is let say curve D and having standard deviation 10 right. So, this is how you can have a curve related to its standard deviation and when say this S here I am talking about standard deviation of the sample.

(Refer Slide Time: 12:13)



The probabilities for the normal random variable are given by area areas under the curve the total area under curve is 1 we have already seen that right. So, this entire area is 1.

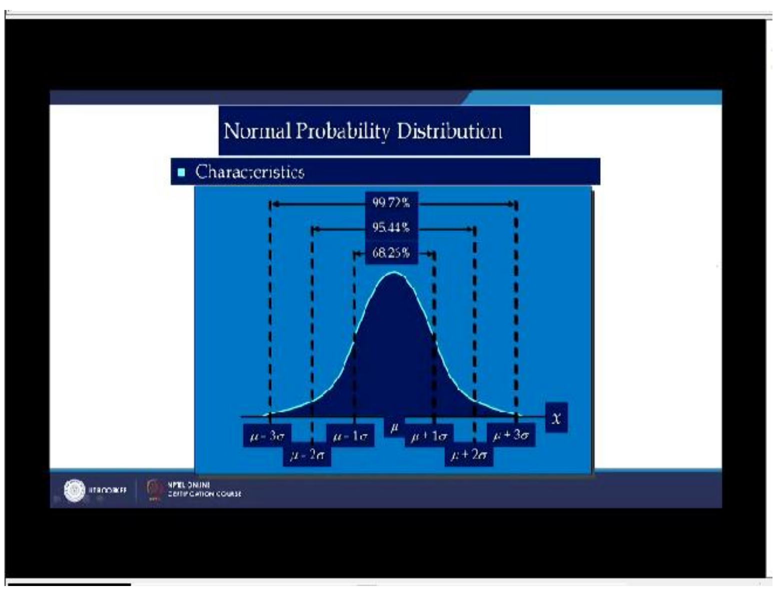
(Refer Slide Time: 12:33)



Some more characteristics, you will always have 68.26% of values within 1 sigma limit. So, let us look at this curve this is your mean, this 1 sigma limit is -1 this +1 right. So, -1, +1 is -2 sigma, -2 sigma, so within 1 sigma limit you will have this much area 68.26 it means one side how much it would be 34.13 is not it. So, within one sigma limit 68.26% of the area within 2 sigma limits 98.45.

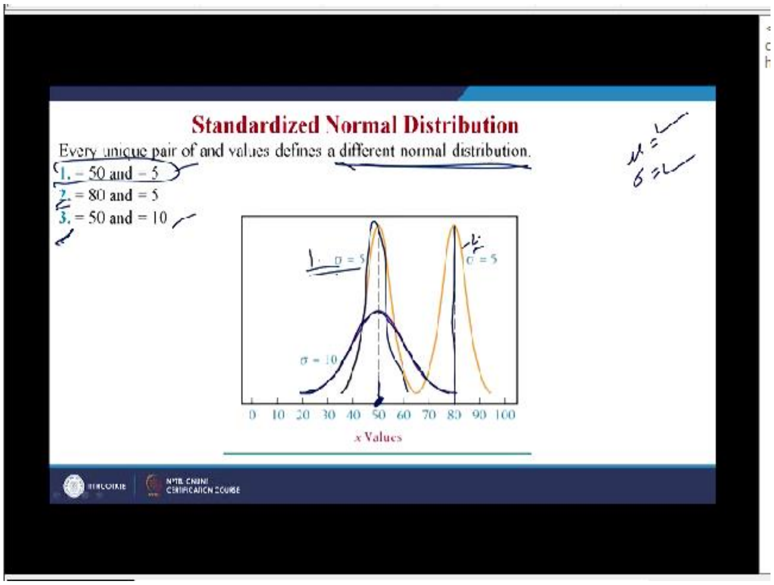
So, let me this area this total area this area is nothing but 95.44 now this is -3 sigma this and this is -3 sigma. Now this area is this area would be 99.72 right.

(Refer Slide Time: 14:13)



So, this is again one of the important characteristics of normal distribution right within 1 sigma limits, 2 sigma limits and 3 sigma limits.

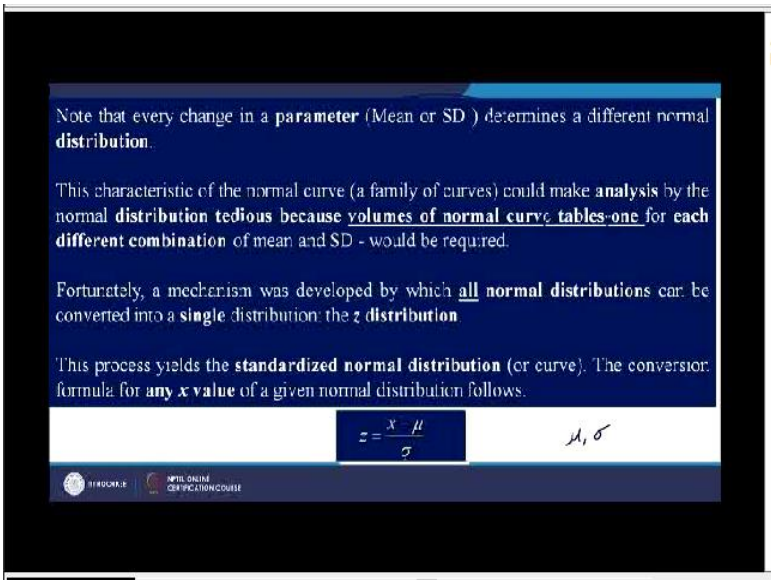
(Refer Slide Time: 14:24)



Now look at what is standardized normal distribution, so if you look at if you are given let say mu and sigma then you will have distribution for a different value of mu and Sigma you will have another distribution. So, for every pair of mu and sigma you will have different normal distribution, for let say first case when mean is 50 standard deviation is 5 right this is 5 is not it.

So, this is first case for second case mean is 80, so this mean is 80 is not it and sigma is 5, so this is case number 2 right point number 2. And third one is mean is 50 yes the same one and standard deviation is 10, so just see in this case first and third the means are same but standard deviations are different. So, standard deviation is 10 in case of third that is why you have got a flatter curve right is not it.

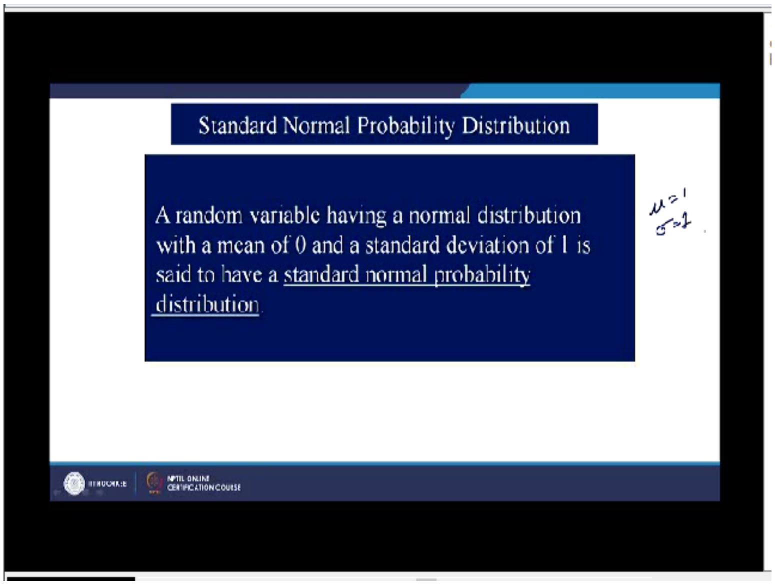
(Refer Slide Time: 15:41)



Now we know that for every parameter you will have a normal distribution, for every set of mean and standard deviation you have different normal distribution. Now this characteristics of normal curve would make the analysis very tedious because you will have to prepare volumes of normal curve tables for each set of mean and standard deviation. So, if you want to let say prepare normal curve tables are z tables then you will have to have lots of tables for each pair of mean and standard deviation.

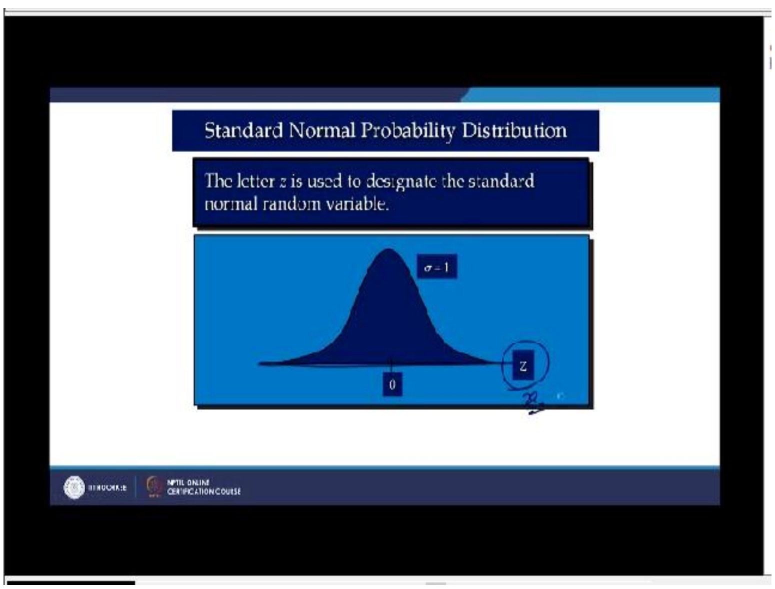
So, this problem can be overcome by using something called z distribution or standardized normal distribution. So, a mechanism was developed by which all normal distributions can be converted into single distribution, so if you got out let say hundreds of you know pairs of mu and standard deviation. Then you can convert all of them into z value or standardized normal distribution. So, the conversion formula for this is yes this is given here, so this $z = x- \mu /\sigma$ right.

(Refer Slide Time: 17:20)



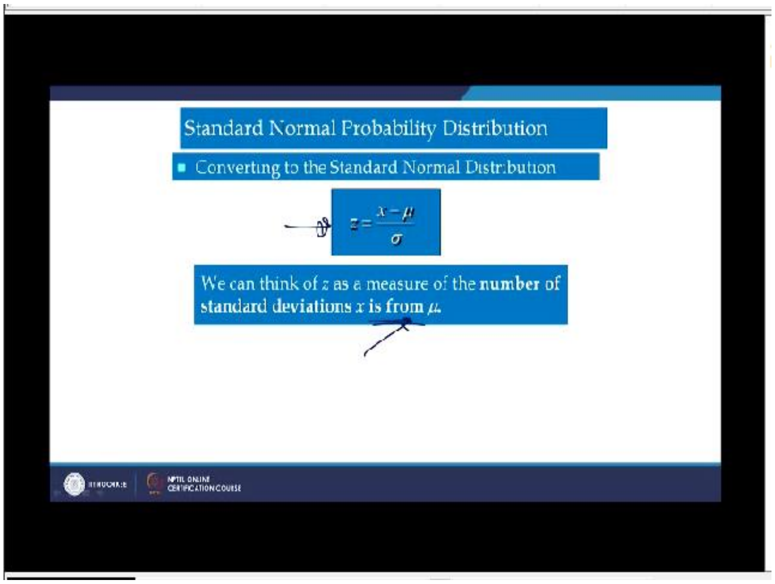
So, a random variable having standardized normal distribution will always have mean=0 and sigma= 1 ok. So you can represent any given distribution either in terms of its value let us call it x and it can easily be converted into z.

(Refer Slide Time: 17:49)



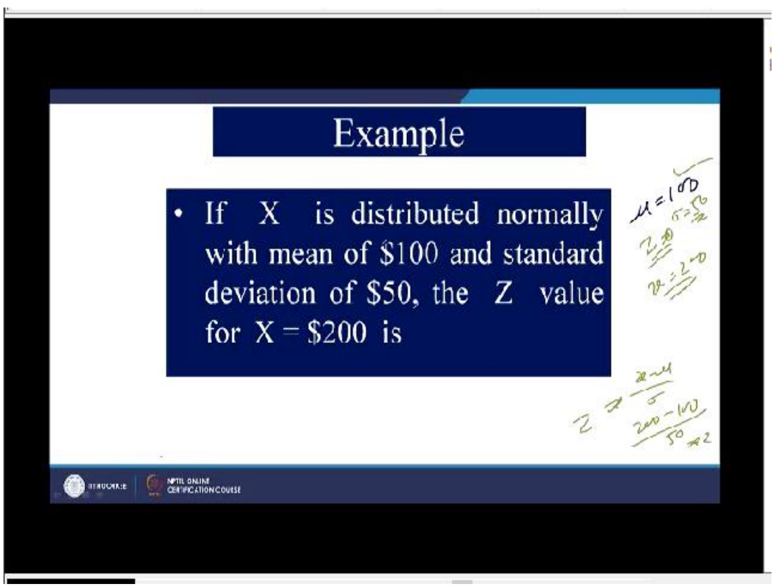
So, this is now z scale right rather than having x scale, so you can convert x scaling to z scale and vice versa.

(Refer Slide Time: 18:02)



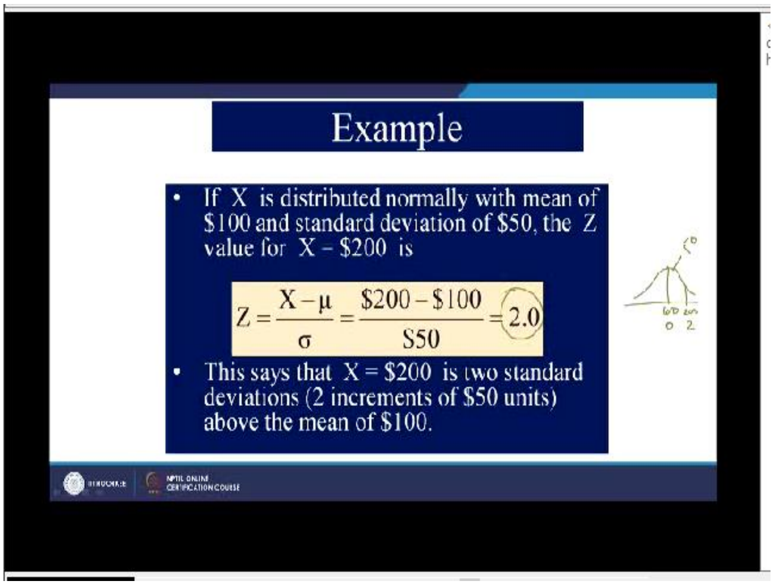
So, this how you can convert x value into you can easily calculate z value by using x- µ /σ. So, you can think of z is a measure of number of standard deviation x is away from mu is what we have calculated when we did find out outlets using two methods for calculating outlets The first one was a IQR and the second one was z value. So, this is the same z value, now let us look at couple of examples.

(Refer Slide Time: 18:51)



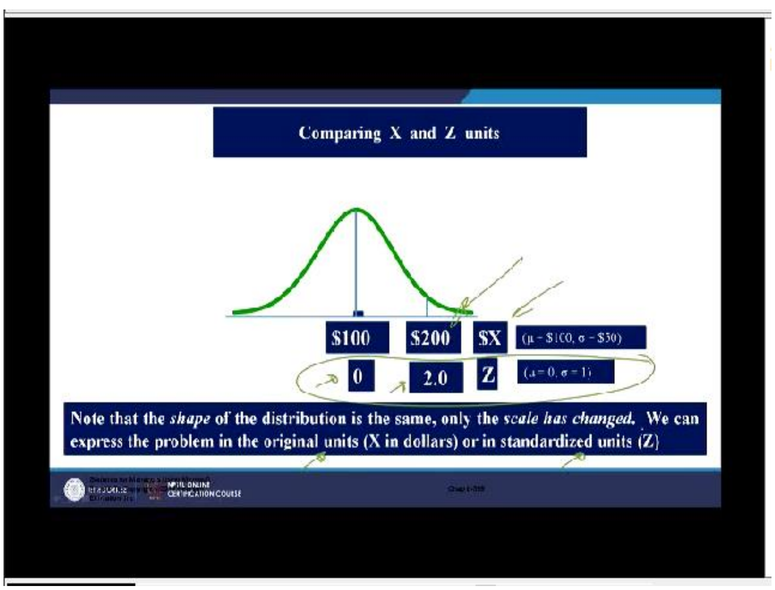
If x is distributed normally with mean of 100, so mean is let us look at mean value 100 and z value as ok if we will calculate and standard deviation=50 right. Now what would be the z value when axis is 200, so if x is distributed normally with mean 100 and standard deviation 50 z value for x would be what? So, you can easily calculate you know that is x-mu/z sorry sigma this z. So, x is 200- mean is 100 this is 50, so this becomes 2 right.

(Refer Slide Time: 19:47)



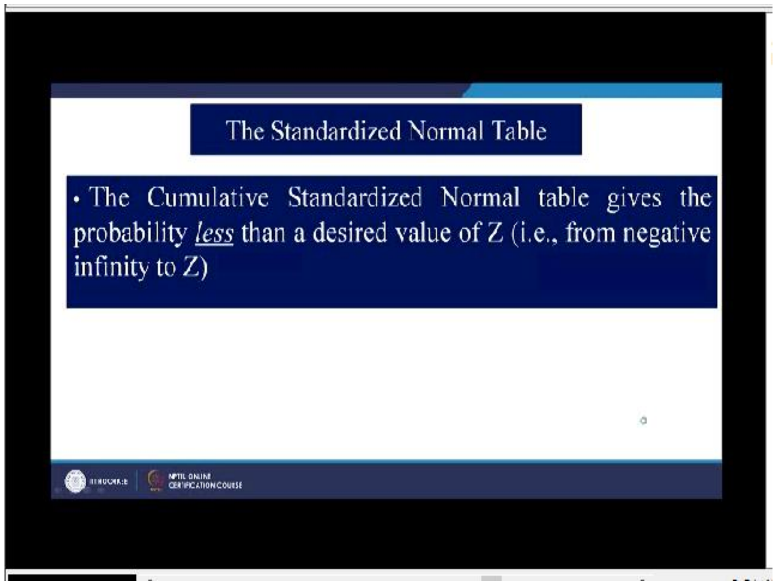
So, this how you can calculate z value, now this says that x = 200 is 2 standard deviation above the mean, above the mean means towards right side or left side. So this mean, mean is 100 standard deviation is 50 and this is z value, let say this is 100, this is 200, z is equal to if you convert have if you have z square, so this is 0 and this is 2 right.

(Refer Slide Time: 20:27)



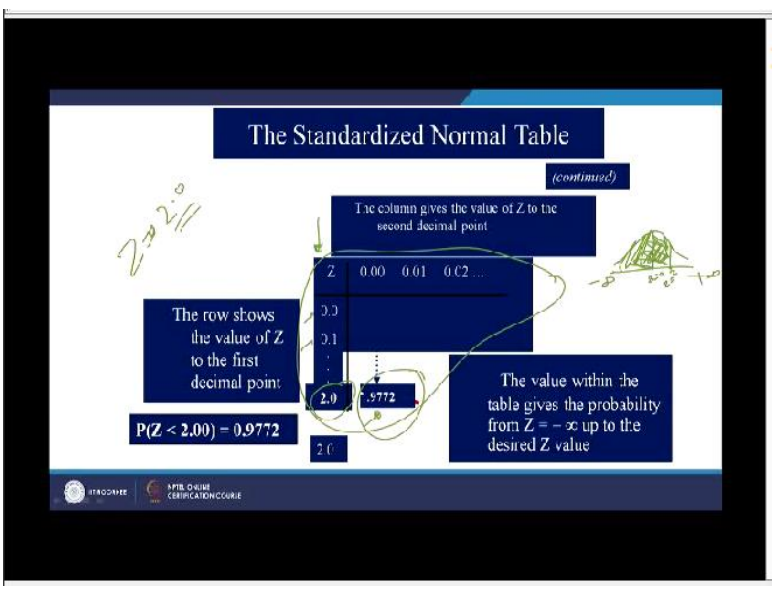
So, we will say that this x = 200 is 2 standard deviation towards right side of the 0 value right and this is you, so you have got 2 scales over here this is z scale and this is your original variable. So, as we have already said and we have always said that the mean=0 and sigma=1 for standardized normal distribution. So the shape of the distribution is same only what has changed the scale has changed from the original x unit to standardized units right which is z. So this how you will convert x value into z value.

(Refer Slide Time: 21:26)



The cumulative standardized normal table gives probability less than desired value of z, so you can have a z value from negative to positive value.

(Refer Slide Time: 21:47)

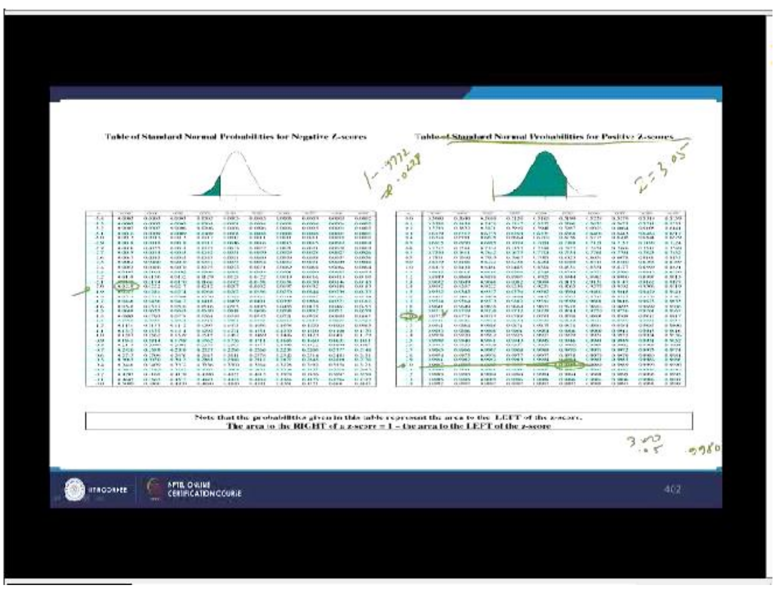


Let us look at the area under curve when z=2 right we are solving the same example where mean was 100, standard deviation was 50 and x was 200 right. So, what is the probability or what is the area under the curve when z is less than or equal to 2, this is z and this is 2 right. So z less than 2.00, this probability you want to find out, so if you look at any book on statistics you will find a standardized normal table along with other tables which we have seen in previous sessions like binomial table, Poisson table.

So, look at z table, so this how z table looks like here is just a part of that z table, so in this column if you look at z = 0.0, 0.1 and2 and 3 so and so on 3.13.2 and so on right 3.1, 31 so on. Similarly here these are the decimal values, so let say if you want to know what is the value of z or what is the area under curve when z = 2.0 right. So, this is 2 and this 0.0**,** so this area under curve is 0.9972.

So, how it will look like if I draw a curve let us look at let me draw a curve varying this is so this your distribution. And this is your z = 0 here z = 2 here right this is 2, now we want to know this area and which is 0.9772. So, this area is 0.9772 from minus infinity this -infinity and this +infinity right.

(Refer Slide Time: 24:22)

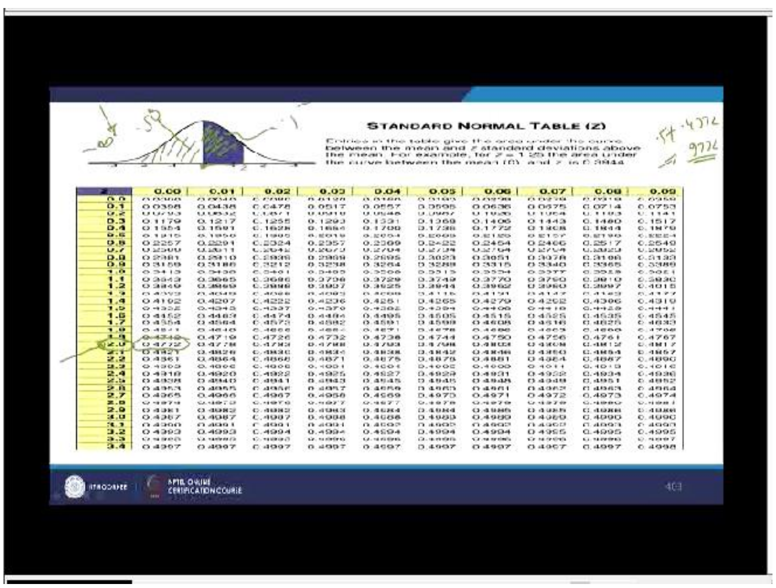


Now let us look at z table, how it looks like, so this is standard normal probability for positive z score. So, let say z = 2, so z = 2 right here it is and probabilities are 0.9972. so, what would be the area under curve when z is equal to let say 3.05 ok, so 3.05 is where first of all look at 3 right. So, this is 3.00 and then in it this is 6.05 0.05 is here right. So, if you look at here this is, so this is the value 0.9980.

So, you can calculate area under the curve from -infinity to z value, now let say, so this is let say at z the area under curve is 0.9972, what is the area under curve when z is equal to -2 look at this -2 is somewhere here right. This point this is the area under curve and z = -0.0228. So, there is no need to remember it you will you just subtract 1- 0.9772 you will get 0.0228, so if you know the value of z.

If you know the value of z you can easily calculate probabilities and if it is negative value then you just subtract that probability from 1, you will get the negative value of z. so, this how you should read z tables, so this is one of these z tables.

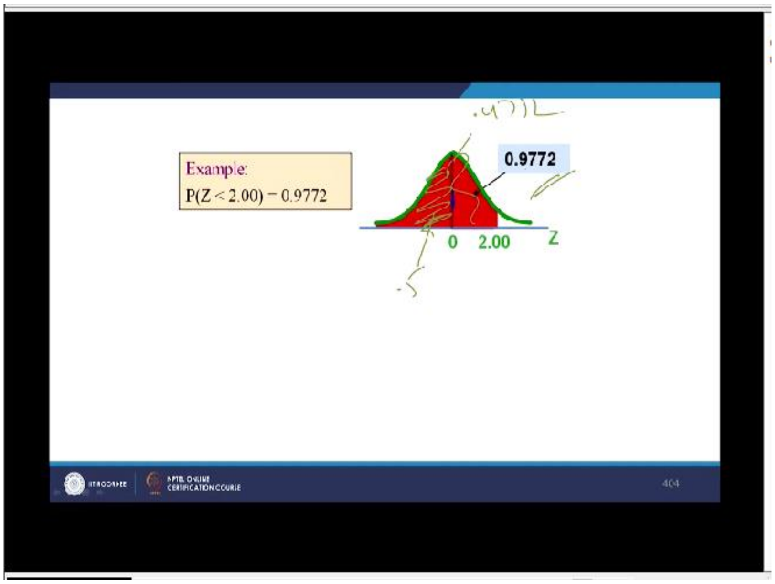
(Refer Slide Time: 26:55)



You can have a different type of z table but you will get the same information, so in this type of z table you have got area under the curve from 0 to z value not from -infinity to some z value right. So, it gives you only probabilities from 0 to either some positive direction or negative direction, direction of Z value right. So, let say what is the value of z what is the area under the curve and z=2 right.

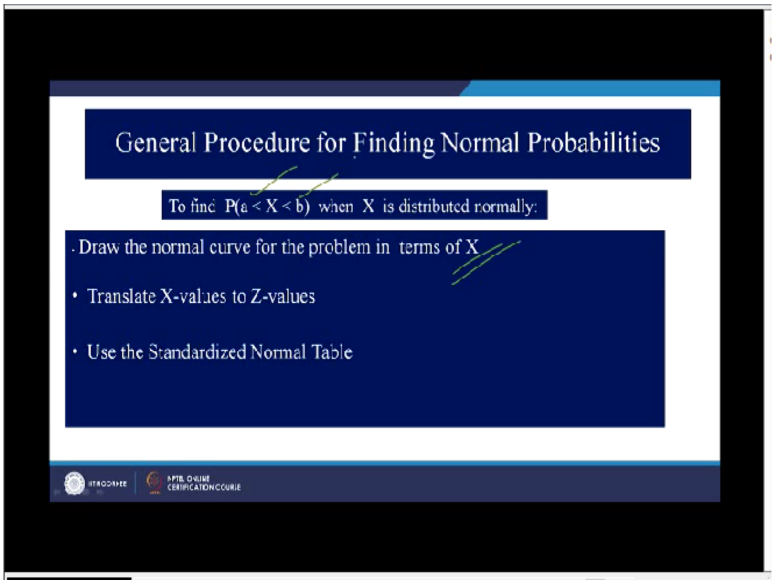
So this is, so z = 2 is here right and area under the curve is 0.4772 but in earlier case it was 0.9772 how come, why there is a difference here. Because here we are calculating z value from 0 to 2, in previous slide we had area from minus infinity to 2. And we also know that the 50% area is this is not it or this is 0.5 the remaining is 0.5. So, you just it 0.5 in 0.4772, so you will get 0.9772 which is the same answer which was there in previous table. So, either you use this table or the other one tables this one and the same thing but you should know whether the area under the curve is from mean or from minus infinity.

(Refer Slide Time: 28:48)



So, this how you just add this area which is 0.5 and this area which is 0.4772, so it becomes 0.9772.

(Refer Slide Time: 29:01)



Let us look at general procedure for finding normal probabilities, let say if you want to find out probability between A and B value when a variable is normally distributed having some mean and some standard deviation. So, first of all you should draw the normal curve by taking X into account then convert x value into z value by using that formula and use standardized normal table to know the area under the curve. So, we will look at how to find out probabilities between A and B, in next session for the time being let me sign off, thank you very much.