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

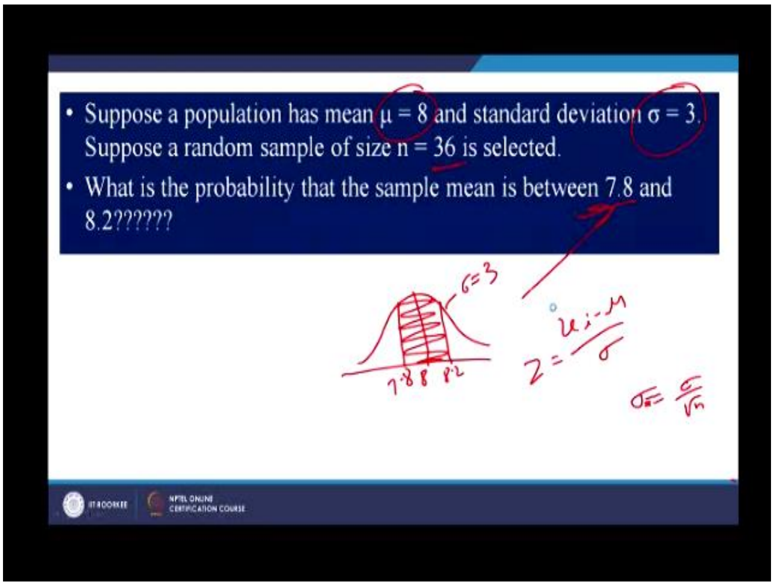
Lecture-24
Sampling Distribution-II

Good afternoon friends, I welcome you all in this session, as you aware in previous session we were discussing about sampling and sampling distribution and we have seen that if there is a population distribution available, then if we take a sample from the population and if we find out its mean, if you take one more sample again we find out its mean, then both these means would be different.

And similarly when you keep on taking samples from population each sample mean would be different. So when you draw distribution of these different sample means it becomes sampling distribution. We have seen that the population mean and sampling mean was same, but so far as standard deviation is concerned, the standard deviation of population and the standard deviation of sampling distribution they were different.

And there was a relationship as well between the standard deviation of population distribution and standard deviation which we call standard error in case of sampling distribution and their relationship was it was like this is not it is not it.

(Refer Slide Time: 01:55)

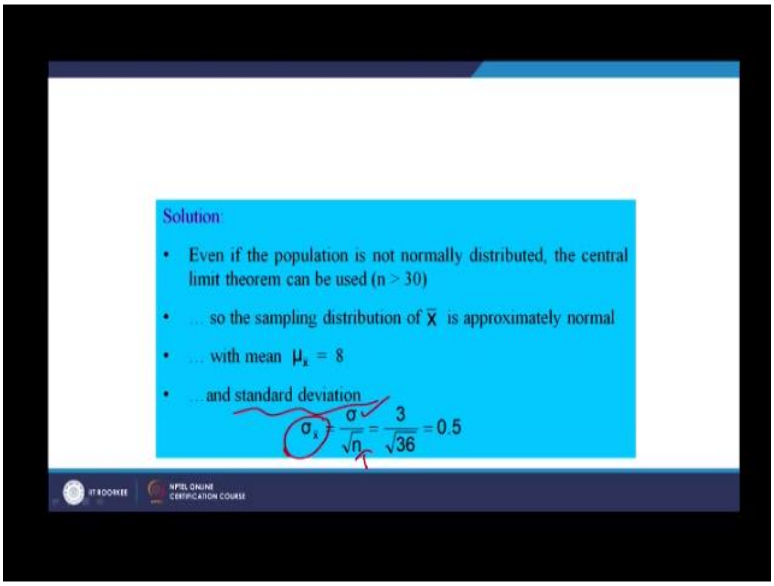


So and we have also seen that the as you increase sample size the distribution of the sampling distribution approaches towards normality and this is known as central limit theorem, what is central limit theorem in central limit theorem whatever is the shape of population distribution whether it is normal or non normal as we increase sample size, the shape of sampling distribution approaches towards normality.

We did workout 1 example as well in previous session, in today's session we will see some more examples. So let us say have been given population mean as it and standard deviation is 3. Now if we take a random sample of size 36, now what is the probability that the sample mean is between 7.8 and 8.2. So how the distribution would look like first of all it will be like this, is not it. So you have got this mean isn’t it, standard deviation is 3, isn’t it, probability of sample mean between 7.8 and 8.2. So let us say 7.8 and let us say this is 8.2 isn’t it.

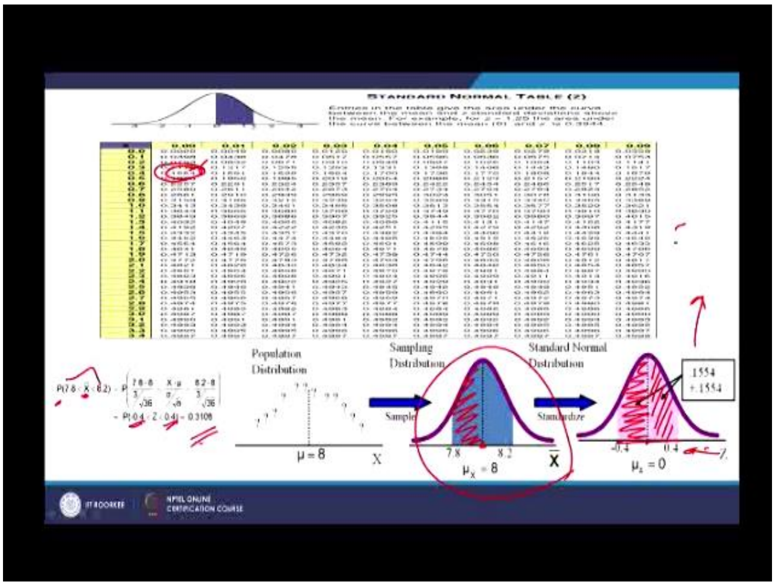
So we need to calculate this area isn’t it, so how to proceed for this first of all we should get Z value right. It is Xi-mu isn’t? it divided by sigma. So x1 is 7.8 and x2 is 8.2 isn’t it. So let us find out Z value.

(Refer Slide Time: 03:55)



And before Z value we you can calculate standard deviation as well for sampling distribution, so this is the relationship between standard error of mean and population standard deviation and sample size. So this value is 0.5, let us look at this. So for we have to find out probability between X that is mean X, X bar varies between 7.8 to 8.2.

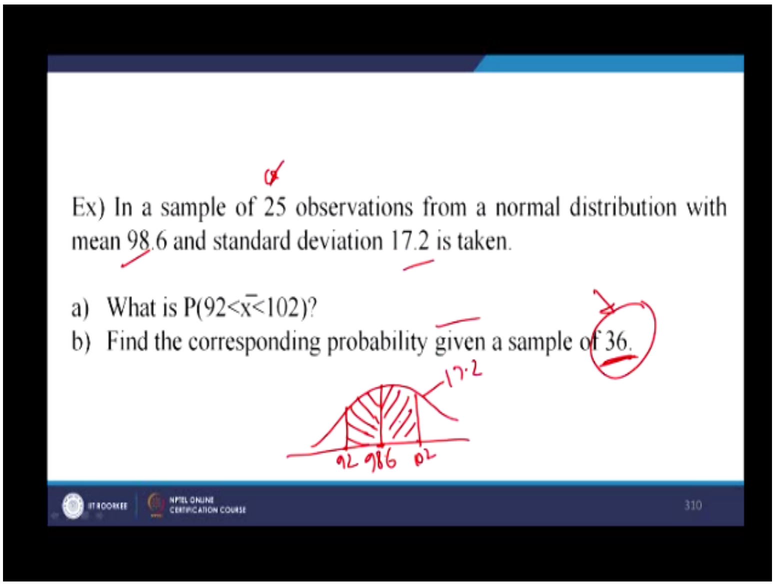
(Refer Slide Time: 04:36)



So this comes around -0.4 and +0.4, so Z is if you look at this you z scale this is now your z scale, this is X scale is not it, this Z scale so from mean this side it is -0.4 and this side it is +0.4. Now let us look at the area under curve from mean to this, we let us look at what is this area.

So this 0.4, Z=0.4 is this is the one, is not it, 0.1554, so this area is 0.1554, similarly Z=4 towards right side again same area 0.1554. So you this is now again 0.1554, so if you add this 2 it becomes 0.3108. So the probability that the mean would be between 7.8 and 8.2 is approximately 31%.

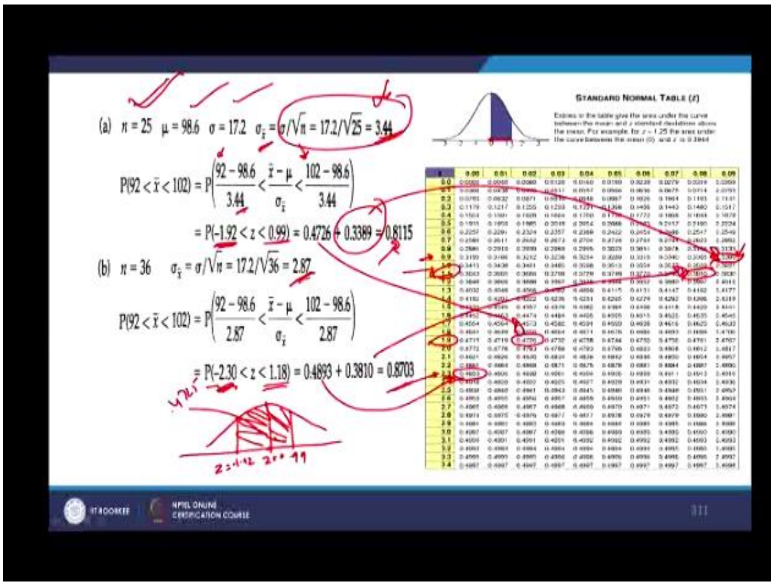
(Refer Slide Time: 06:06)



Let us look at this question in a sample of 25 observations from a normal distribution with mean is this standard deviation is this taken, what is the probability that mean is between 92 to 102, find the corresponding probability given n=36. So we have changed n from 25 to 36. So we have increased sample size, so what would happen to the standard error it would reduce is not it.

If you look at this formula then as we increase n, this value would decrease is not it. So how the distribution would look like for this means 98.6, standard deviation 17.2 and probability it is between 92 let us say this 92 and this is 102. So you need to calculate first of all z value from here to here, here to here and then this area and then this area. And then we led these 2 areas ok. So let us look at this.

(Refer Slide Time: 07:29)

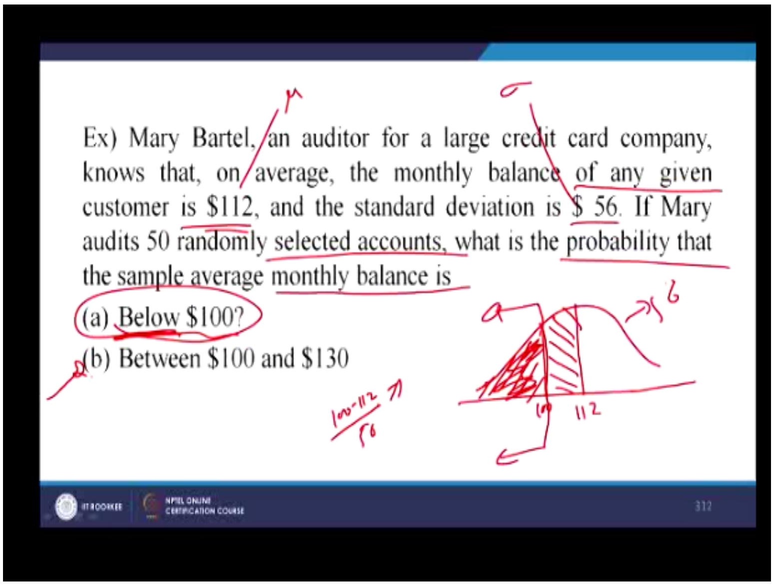


So when X is 92, so (x-μ)/3.44 which is this because we have you been given n=25, μ is equal to this and standard deviation is equal to this, so this is 3.44, so z when X is 92 which is 1.92 and Z when X is 102 which is 0.99. So let us macro environment redraw this Z value, so Z mean=0, here Z is -1.92 and here Z is let us 0.99 is not it. So area from here to here Z=0-1.92 right. So 1.92, so this is 1.9 and 2 is this 0.4725.

So this area is 0.4725, what about this area again Z from here to 0.99 right. So this 0.9 is this 0.99 is this, 0.3810 this is here right 0.3889 this is the area and the first one was 0.4726 right this is the 1. So add these 2 areas we will get this value is approximately 81%. Now the same question if we increase sample size. Then what should happen this value should decrease is not it.

So it was 3.44, now it is 2.87, you calculate again Z values so for Z values would be is 2.3, now if you look at this particular table values from here to right hand side all these are positive values and left hand side negative values. So Z=1 and Z=-1, the area under curve would remain same is not it. So when let us say Z=2.3 right, so 2.3 is fare this is 2.30 right. So this is 0.4893 here it is.

(Refer Slide Time: 10:43)

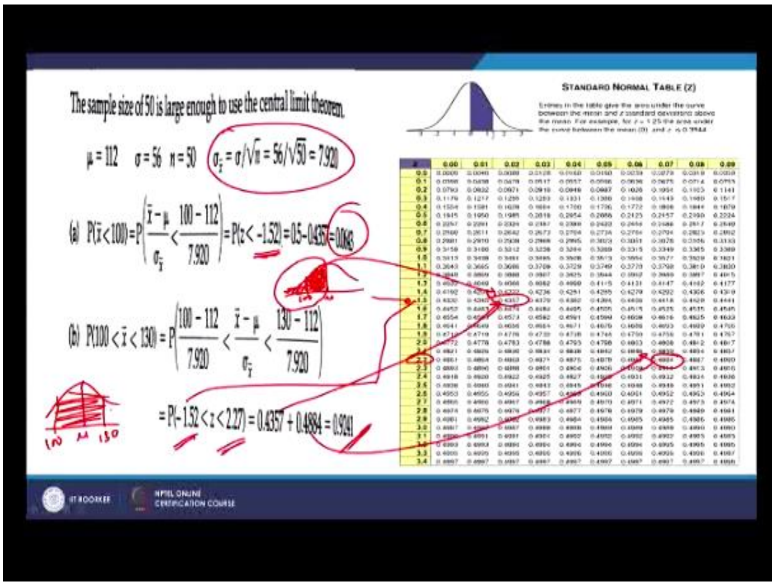


What about 1.18, 1.1 is this is not it and 8 is here, so 0.3810 which is here right, this 81, so just add these 2 values right and this your answer 87%. Let’s look at one more question Mary Bartel an auditor for large credit card company knows that on an average the monthly balance of any given customer is 112 dollars which is standard deviation of 56 dollar. So mean is this and standard deviation is this is not it, if is Mary Bartel Rs. 50 randomly selected accounts.

What is the probability that sample average monthly balance is below 100, so how it should be calculated, so this is your distribution right mean is 112, standard deviation is 56 is not it and you to see whether it is below 100, so this is 100 right. Now we have to see below 100 means which areas should we calculate this area or this area. Just read the question carefully, it is below 100 right. So this is your 100 pound right, this is your 100.

So all this area which is below 100, so we need this particular area right ok, so how to do it, how to solve this question, first convert it into Z value so this X-X bar right, so 100-112/56 is not it. So will get some negative value of Z and for this we will see this second part little later.

(Refer Slide Time: 12:26)

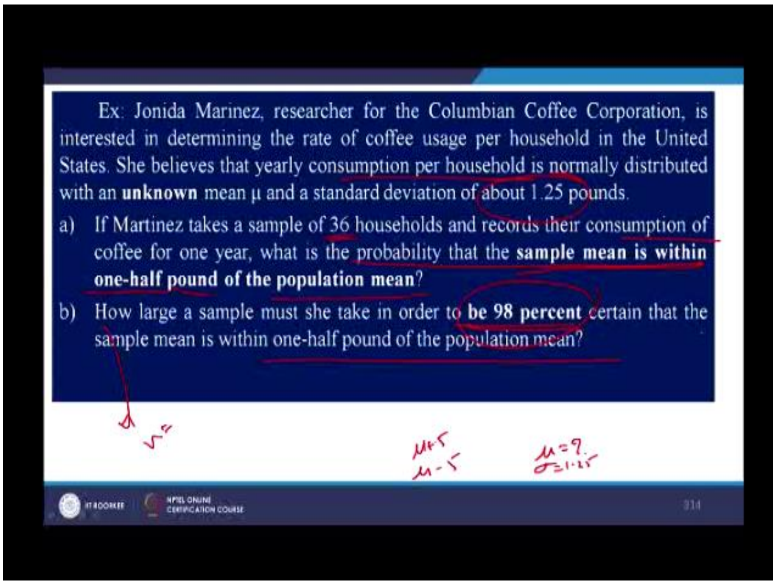


So part A, so first of all calculate standard error which is 7.92 ok and Z values now 1.52-1.52. So let us look at what is that area, so the area which would we would be getting is let us look at minor 1.52, so 1.5 and this is 0.4357. So what is this area, this area is let me draw distribution, this is your mean and this is your X value which is 100 and we want this area and what we have got is this area this is this 0.4357 is this area is not it.

We are interested in calculating this area, but we know that this entire area is 0.5, so we will subtract this area form 0.5, so we will get remaining area which is 0.643. So probability is 6.43% right. Now what is the probability that the monthly balance is between 100 and 130 100 and 130 is calculate Z values so -150, -1.52, and 2.27 area just look it 1.5, so 1.5 and 2 is this 0.43, so this value is here and 2.27**.**

So this is 2.2 and 7 is this 0.4884 which is here ok, we just add these 2 values, why we are adding these 2 because we want to know this is your mean and this is 100 and this is 130 is not it, so we have to add a centre area ok.

(Refer Slide Time: 14:51)

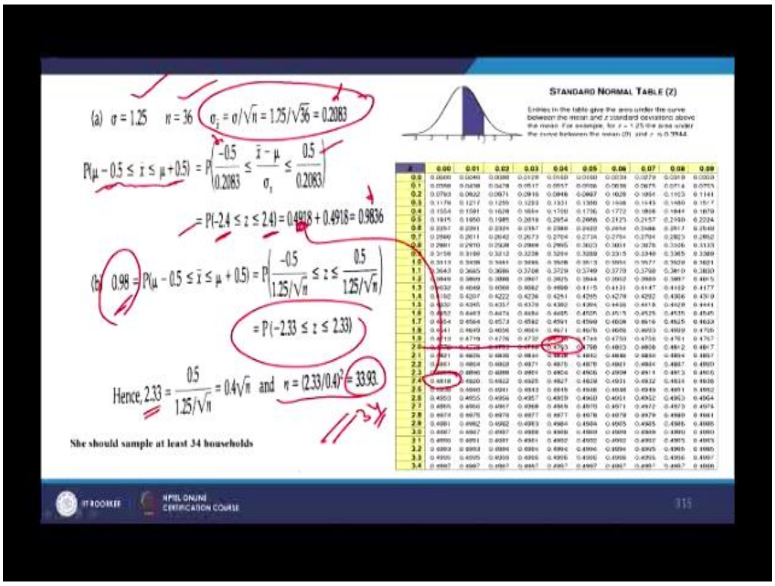


Let us look at one more question Jonda Marinez researcher for the Columbia coffee Corporation is interested in determining the rate of coffee usage per household in United States, she believed that yearly consumption per household is normally distributed with unknown mean but standard deviation is known which is 1.25. So this is not known and standard deviation is 1.25.

If Marinez takes a sample of 36 households and because their consumption of the coffee for 1 year probability that the sample mean is within 1 half pound of the population mean which is not known. So let us look at let us say this is let us assume that population mean is this. So we have to calculate what is the probability that is sample mean is within one-half pound it means 0.5, so +0.5 of this and -0.5 of this.

So in this range we have to find out probability, second part is how large sample must she take in order to be 98% certain that this sample mean is within one-half pound of population mean, so how large we need to find out sample size here is not it, so what will be the sample size, so that the 90% no 98% of the sample mean is within one-half pound of population mean, let us look at this.

(Refer Slide Time: 17:01)

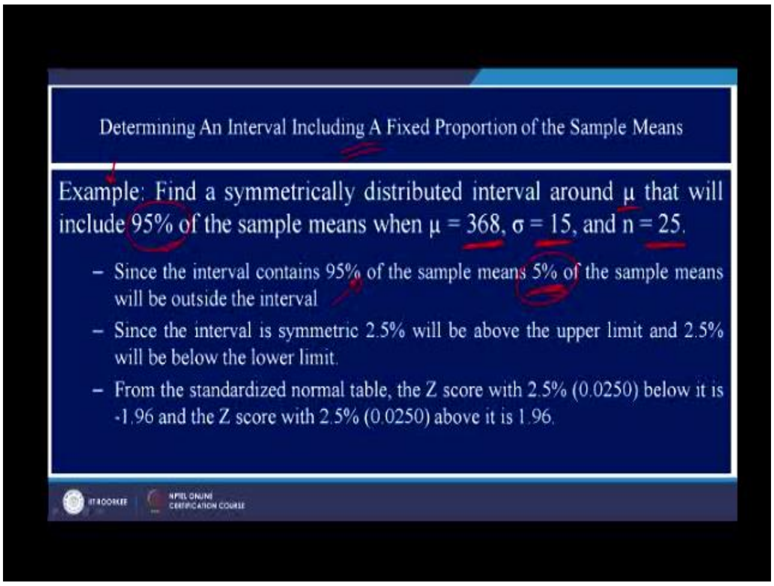


Since we do not know the population mean we have information about standard deviation and sample size is first of all calculate standard error this standard error of mean right which is 0.208, now within we have to find out probability where the mean is one-half pound right, so this is μ+0.5 and μ-0.5 So it is -0.5, -0.5/sigma X bar and +0.5/ $\sigma_{\bar{x}}$ 8. So we get Z as -2.4 and +2.4, just look at what was the probability and Z=-2.4.

So this is 2.1, 0.2 and 0.3 this 0.48, this 0.47 2.4 no this 2.04 this value we have to look at 2.4 exactly so this is 2.4, so this value is here ended 0.4918, 2.4 positive side same values you just add this 2 is not it, now the second part is how large is sample must she take in order to be 98% confident that the sample mean is within one-half pound of the population mean. So we know that this value is 98%.

Now since this value is known and this is already known from this equation, so we just calculate the value of n, so 2.33 and when you solve this question for n it becomes 33.93. So as good as 34 right. So these are couple of examples on sampling distribution. Now let us look at extension of sampling distribution and will work out couple of examples where in we have been given the probabilities and we need to find out the interval around mean.

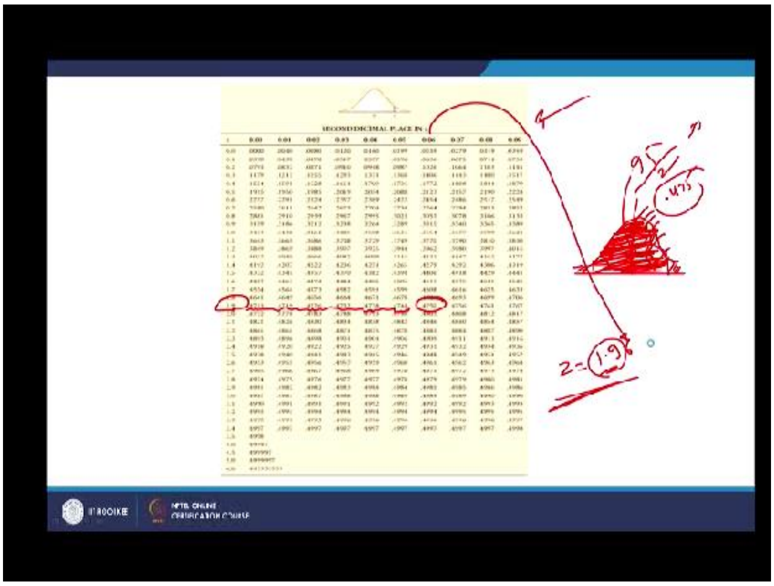
(Refer Slide Time: 19:49)



Till now what we were given we were given intervals and we were finding probabilities, but here the probabilities are given right, so determining an interval including a fixed proportion of sample mean, just look at this example find a symmetrically distributed interval around mean that will contain 95% of the sample means. So mean is 368 standard deviation is this and sample size 30 25.

So what we have to calculate here we have to calculate the upper and lower limit within 95% of confidence level ok. So let us look at this since interval is 95% so the remaining area is 5%. So if you look at the 5% it is in lower tell it is 2.5 % and in a patel it is 2.5%, now from Z score table we can calculate what is the value of Z when certain probabilities are given.

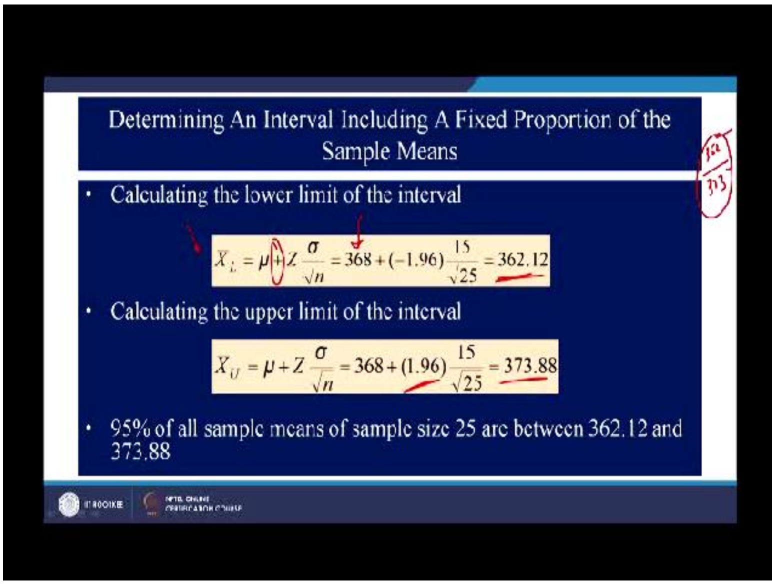
(Refer Slide Time: 21:06)



So let us look at this, so what we are saying here is this is your mean let us say this is your mean ok and this is your 95% area ok. So this is 95% or 0.95, this side is 2.5%, percentage is 2.5% is not it, so this becomes 100% or 1, now when this area is 0.95 this centre what is this area 50% of this, this point .475 isn't it. Now let us look at what is the value of Z so just divide this by 2 you will get 0.475 right.

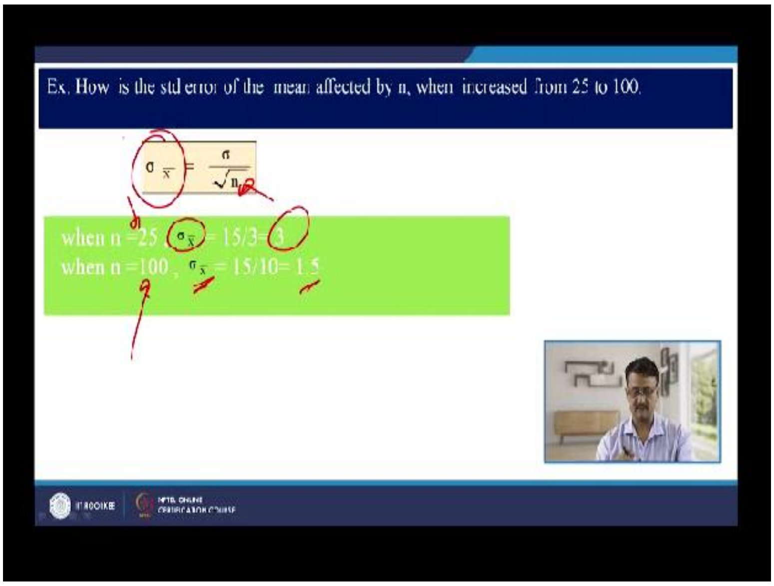
Now look at this table look at this Z table and see where that area is 0.475, let us look the 0.475, 0.47, 0.47 yeah this right 0.475, 0.474 is 0.475 isn't it. So what is the value of Z, 1.96 right this is 6 and this is 1.9, 1.9 is here isn't it. So till now what you are doing we used to have used to have Z values and we were finding probabilities. Now in this case we have probability and we are finding Z value right.

(Refer Slide Time: 23:07)



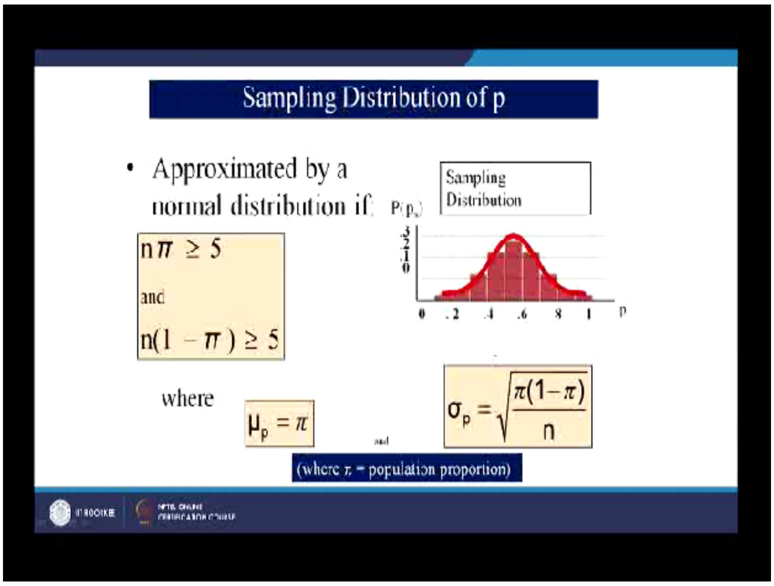
So Z is 1.96 ok so this how we can calculate Z value from Z table right, so lower limit = µ + this and upper limit is this, actually this is this is minus, µ minus ok. So excel is we know that mean has already been given 368 in our question means 368 is not it n is 25 and sigma is 15 right. So n 25 and sigma is 15. So this becomes 362 and upper limit is this 371. So what we say let between 362 and 373 we this is the mean around which when the areas 95% these are the X values is not it.

(Refer Slide Time: 24:41)



So 95% of all sample means of size this are between 362 to 373 right, now let us look at what is the effect of though we have already seen the effect of n on standard error. So let us look at this when n= 25 standard error is 3, but when we increase n then how standard error decreases right this is the relationship. So as we increase sample size this standard error decreases. Now let us look at population proportion.

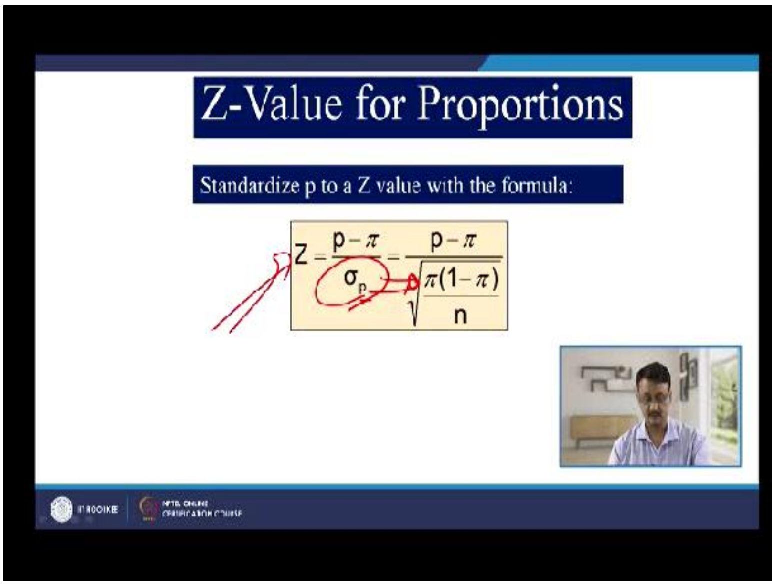
(Refer Slide Time: 25:17)



So far we have talked about population mean right, so let us look at the population proportion, the proportion of the population having some characteristics let us say in a class there are 30% students who are writing with left handle or let us 3% writing with left hand right or let say there are 10% students who have got height more than 6 feet right. So that is nothing but population proportion.

And we can always estimate population proportion from sample proportion, so this sample proportion p is an estimator of population proportion ok. So sample proportion is this, this number of items in the sample having characteristics of interest/sample size. So p will always be between 0 to 10% and p is approximately distributed is normal distribution of course when n is large.

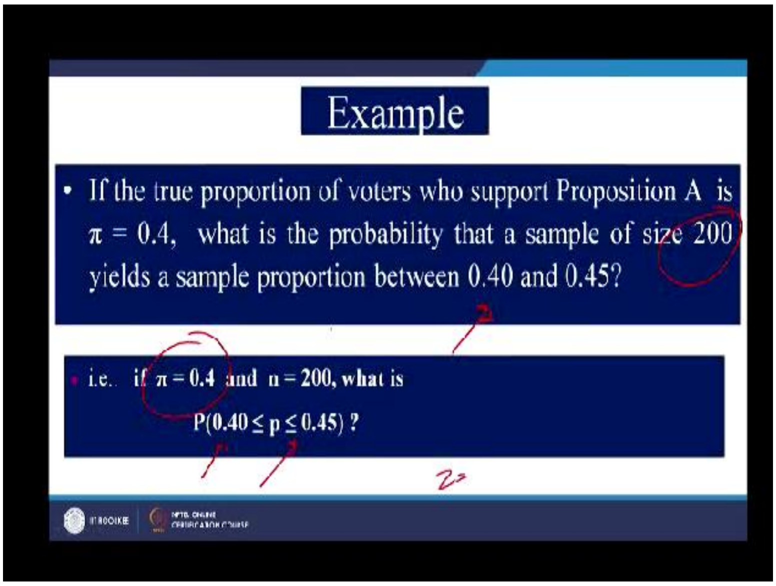
(Refer Slide Time: 26:17)



So sampling distribution of p earlier we have seen the sampling distribution of mean here sampling distribution of proportion, so this is sampling distribution of proportion, it is a normal distribution right similar to the sampling distribution of mean. Now if you look at approximated by normal distribution, so this is $n\pi$ is greater than or equal to 5 and $n(1-\pi)$ is greater than equal to π.

So we will say that the population proportion is this and the standard deviation of proportion is this ok. So almost relationship is same here it is also however it is denominator it is n is not it, in case of let us what it was sigma/root n is not it, but here it is n right and this whole under root. So the value of Z can be determined in a way in which we determine the value of Z or means right.

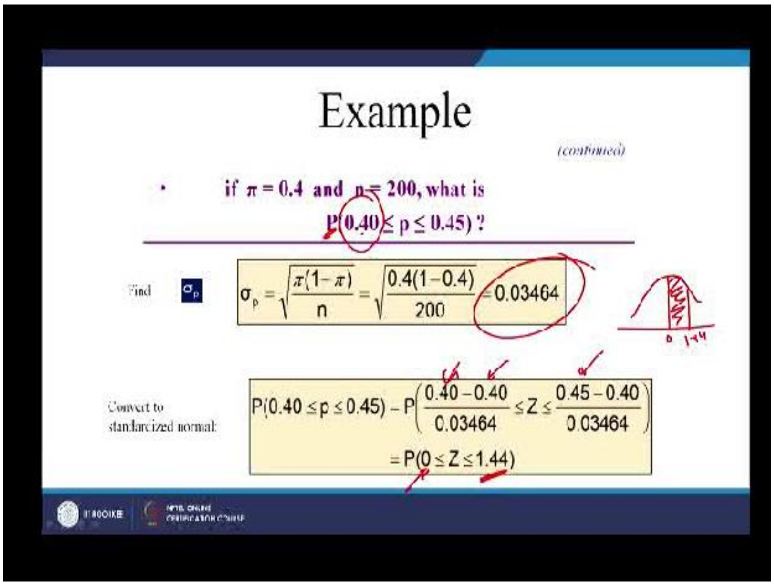
(Refer Slide Time: 28:07)



So $z=(p-\pi)/\sigma_p$ right we know sigma p is this is not it, so let us work out an example if the true proportion of voters who support a proposition A is 40% right, so we are saying that in a population there are 40% people who are supporting to let us say a candidate A, what is the probability of what is the probability that sample of size 200, what is the probability that a sample of size 200 yields a sample proportion between 40-45%.

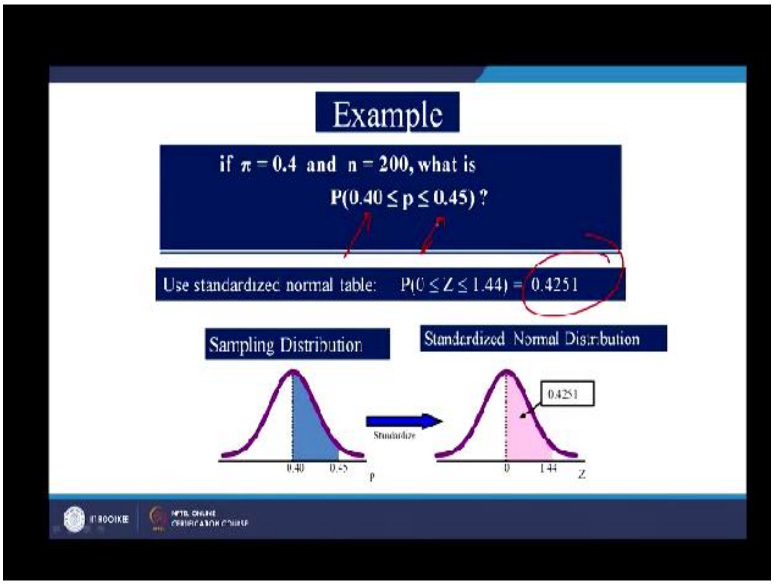
So here population proportion is known you have been given the p values let us call p1 and p2 just like we have seen earlier x1 and x2. So first of all calculate Z value and before that Z even you should calculate sigma p right or the standard deviation of proportion.

(Refer Slide Time: 29:08)



So standard deviation of proportion is 0.034, now you are supposed to convert it into Z value right, so when small p is 0.4 right when p is 0.4 this is your z value 0 because pi is 4 right, now when p is 0.45 Z values 1.44, so if you look at how would you draw this curve which area we want to find out this is 0 and this is 1.44. So we are interested in this area.

(Refer Slide Time: 30:04)



Now this can be calculated by looking at table, so what it was 1.44 right, so 1.4 is this and 44 is this so 0.4251. So will say that this be probability for p ranging between 0.4 to 0.45, so in today's session we have seen how to solve questions related to proportion as well as population mean, in next class we will discuss some more related to sampling and sampling distribution, thank you.