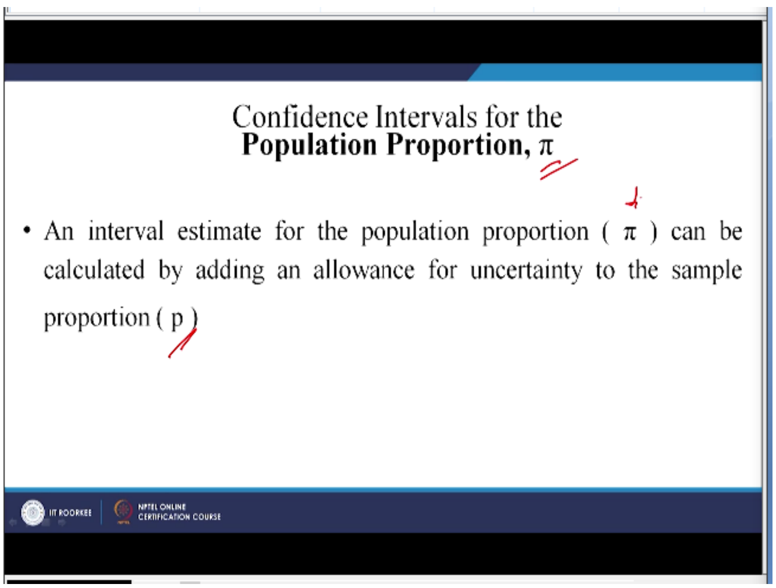
Business Statistics
Prof. M. K. Barua
Department of Management Studies
Indian Institute of Technology-Roorkee

Lecture-28
Confidence Interval–I

Hello friends, I welcome you all in this session as you are aware in previous session we did talk about confidence intervals for population means in two situations. The first situation was where population standard deviation was known and the population standard deviation was unknown. Let me briefly remind what we did in previous session, when we did not know the standard deviation of the population we actually calculated sample standard deviation by going for pilot kind of survey.

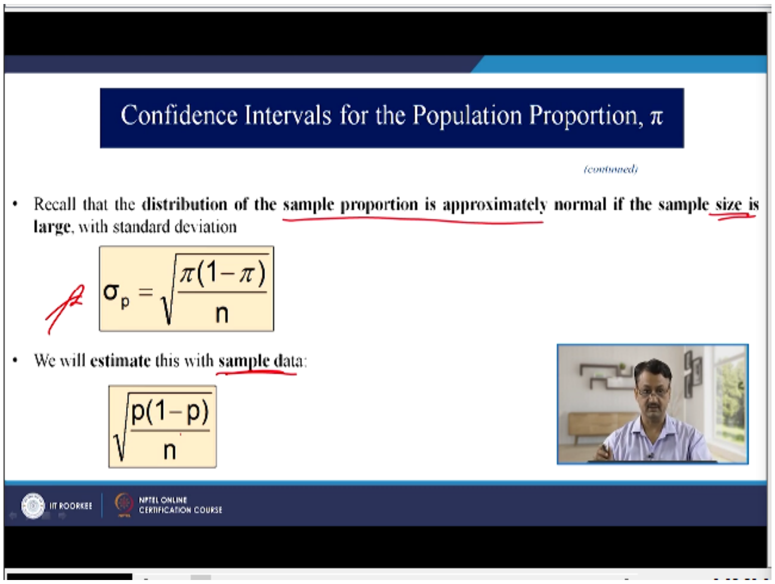
And then we found out standard deviation of the sample and we use that sample standard deviation instead of population standard deviation. In today's session we are going to talk about population proportion will come up with confidence interval for population proportion. Generally we represent population proportion by pi.

(Refer Slide Time: 01:32)



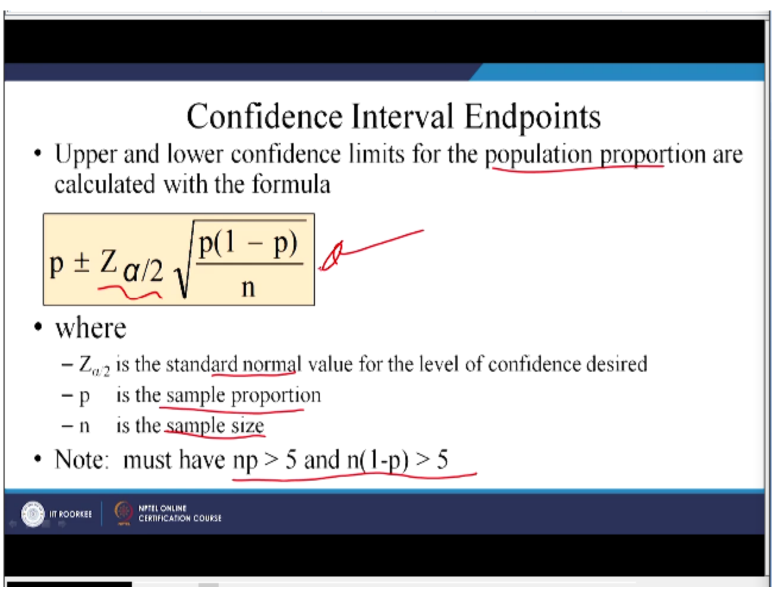
And an interval estimate for this pi can be calculated by adding an allowance for uncertainty to this sample proportion. So, this p sample proportion and pi is population proportion.

(Refer Slide Time: 01:52)



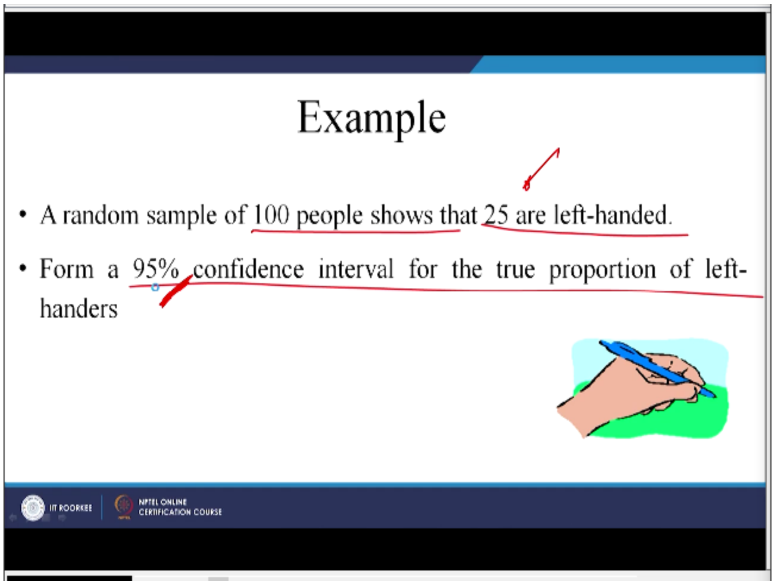
So, we know that the distribution of sample proportion is approximately normal when the sample size is sufficiently large. So this is the standard deviation it is $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$ and we will estimate this sample data. So, this is your value $\sqrt{\frac{p(1-p)}{n}}$. So we will calculate upper and lower limits of population proportion using this particular formula.

(Refer Slide Time: 02:27)



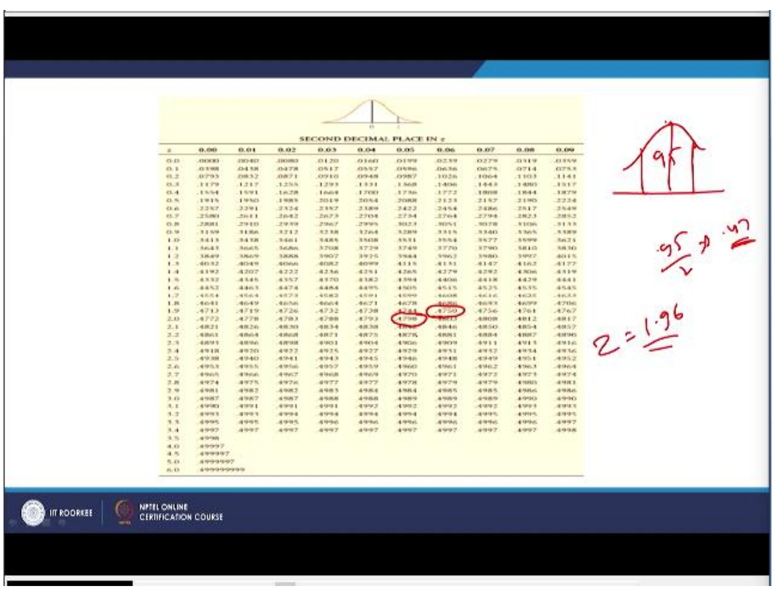
So, P± this Z alpha/2 to which is nothing but the standard normal value which would be getting from table, p is the sample proportion and n is the sample size. So we must have the situation np and nq what should be greater than 5. So, this formula gives you upper and lower limits, let us look at this example.

(Refer Slide Time: 03:01)



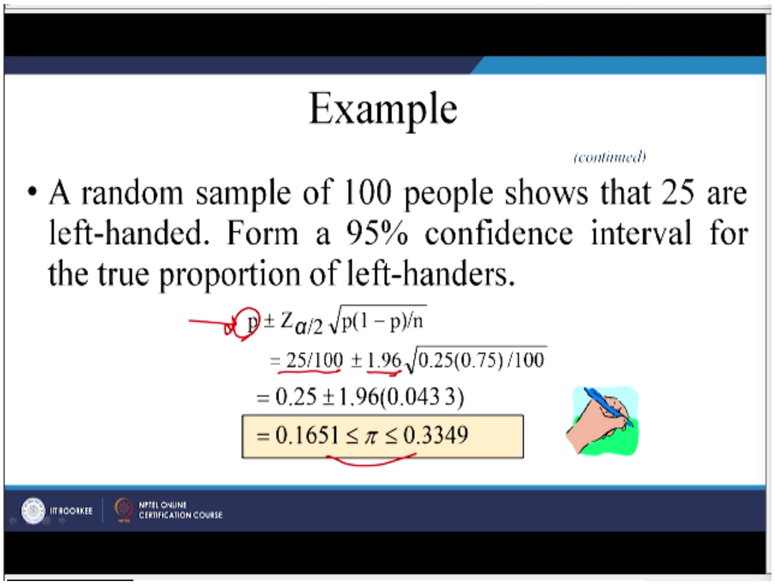
You have taken a random sample of 100 people and the sample shows that 25% are left-handed, their dominant hand is left, form a 95% confidence interval for true proportion of the left-handers. So, what you have to find out we want to find out what is the population proportion or the confidence interval of population proportion when the sample proportion is 25%, so what you need for this first of all you need the table right. So this is your Z table, so our confidence interval is 95 right.

(Refer Slide Time: 03:51)



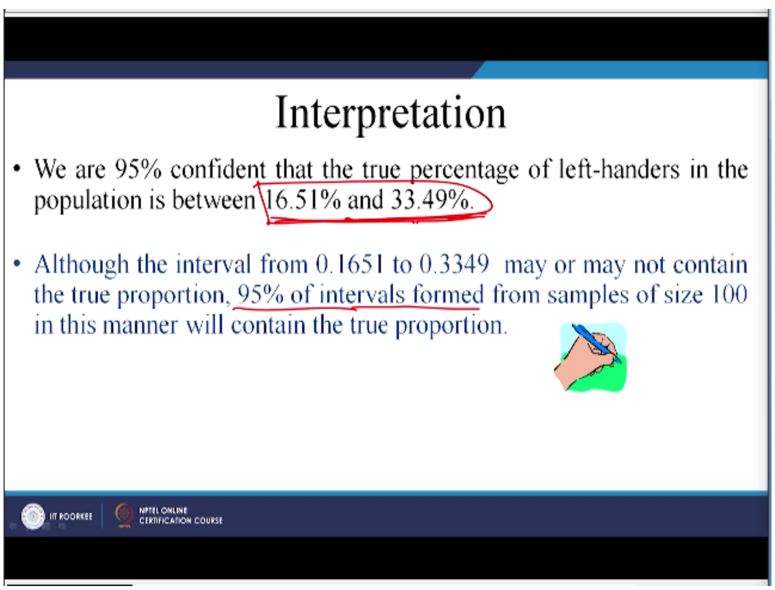
So, 95 is this, this is your 95 right, so 50% of both sides would be 0.95 let say divided by 0.2. So, it would be around 0.475, so look at this Z value where probabilities 0.475 so let us look at this, this is right. So, Z value is 1.96 right, so if you put these values in this formula.

(Refer Slide Time: 04:36)



This is sample proportion which is 0.25 ± $Z_{\alpha/2}$ this is your Z value from table, p is 0.25, (1-P) is 0.75 and n is 100. So, you are saying that the population proportion is between 0.1651 to 0.3349.

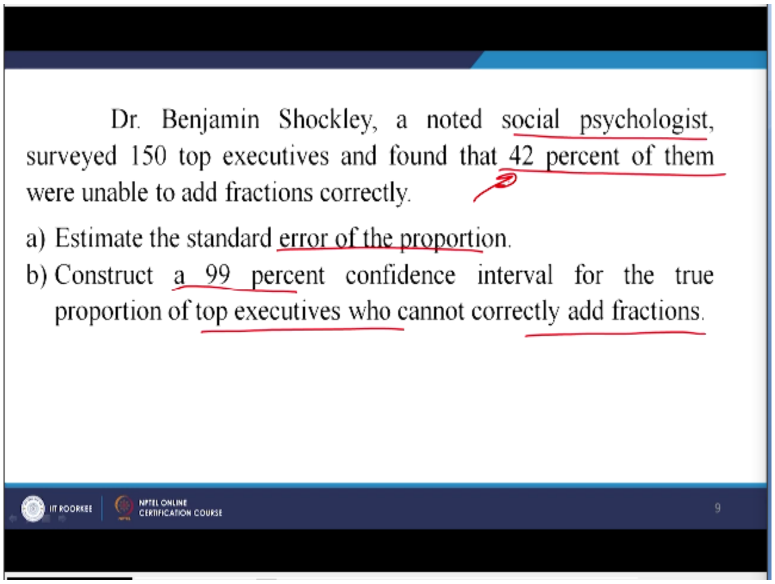
(Refer Slide Time: 05:11)



It is not necessary that the sample proportion would be lying in this or population proportion would be lying in this what we have said when we discussed confidence interval of population mean we have said that the we are 95% confident or 95% of the intervals so formed from different samples of sample size 100 the population proportion would fall in this range is the meaning of 95% confidence interval.

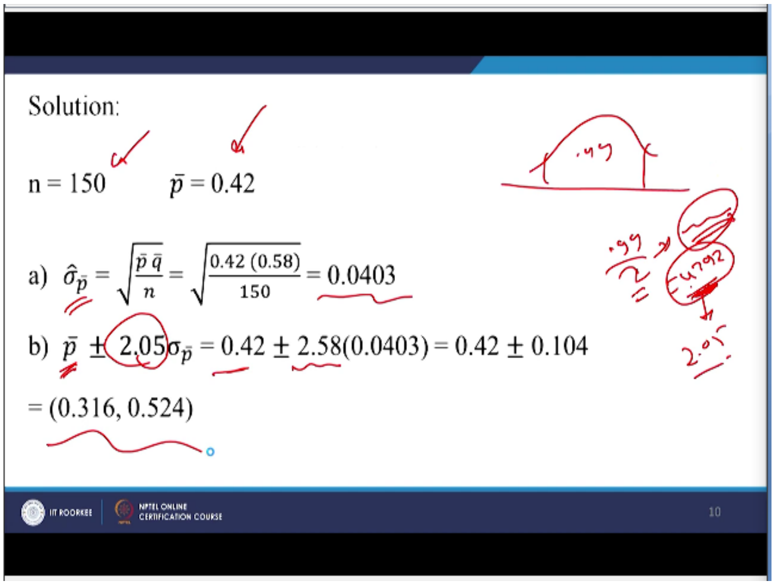
It is not necessary that the population proportion in the first sample would be in this range or in second sample. It would be this range but 95% of the intervals so formed will have population proportion.

(Refer Slide Time: 06:12)



Let us look at another example Benjamin Shockley a noted social psychologist surveyed 150 top executives and found that 42% of them were unable to add fractions correctly. Now the question is estimate the standard error of the proportion, so standard error would be what is sigma divided by under root of n. construct a 99% confidence interval for the true proportion of top executives who cannot correctly add fractions.

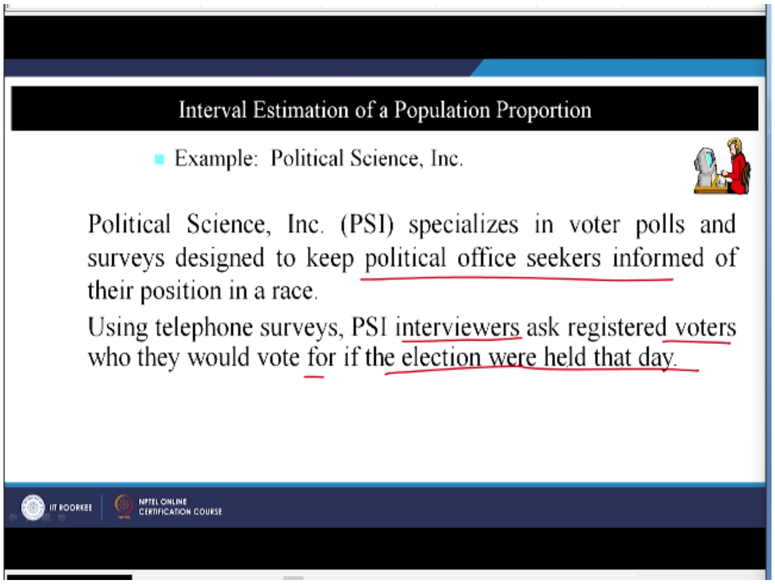
(Refer Slide Time: 06:54)



Let us workout this example, so we know that sample size is 150 and P bar is 42% which is given here standard error would be under root of P bar q bar divided by n. So, this is 0.0403 standard error for this particular part of the question. Now if we look at second part construct and 99% confidence interval for the true proportion of top executives who cannot add fractions. So this is given right 42% ±2.05 from where did you get this 2.05 this value 2.05.

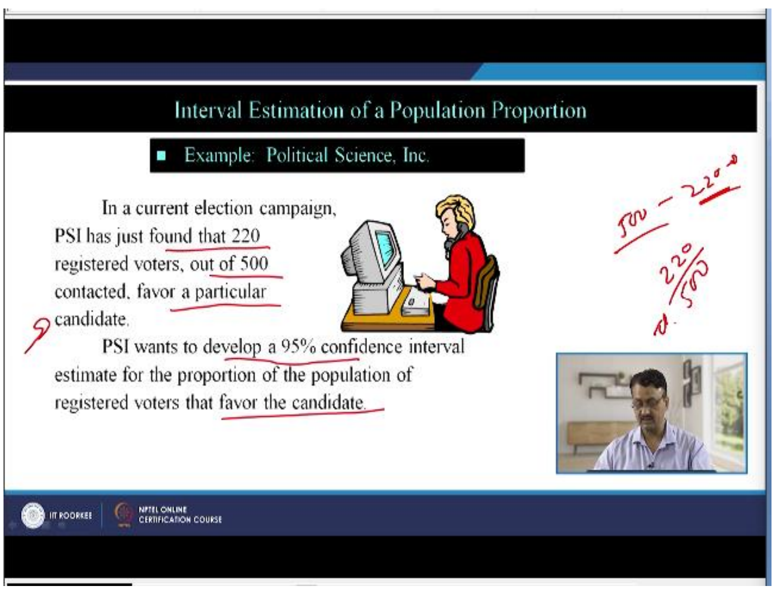
We got it from Z table right and how did we get that value this is 0.99, 50% of this would be you just divided by 2 right. So, that value would be whatever is that value over here, so that probability would be a Z value of 2.05, let us look at this 2.05, this 2.0 and this 5 right. So, 0.4792 is this 0.4792 right when you divide 0.99/2 you will get this and this is the probability at Z value 2.05 right.

(Refer Slide Time: 08:37)



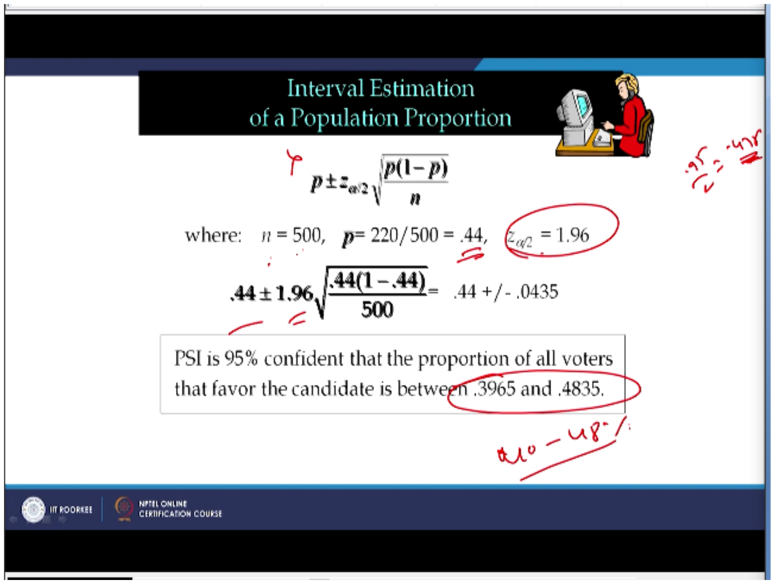
So, this is your confidence interval, let us look at one more example political science in corporations specializes in voter polls and surveys designed to keep political offices, political office seekers informed of their position in race in political race. Using telephone surveys PSI interviewers ask registered voters who they would vote for if the elections were held that day.

(Refer Slide Time: 09:12)



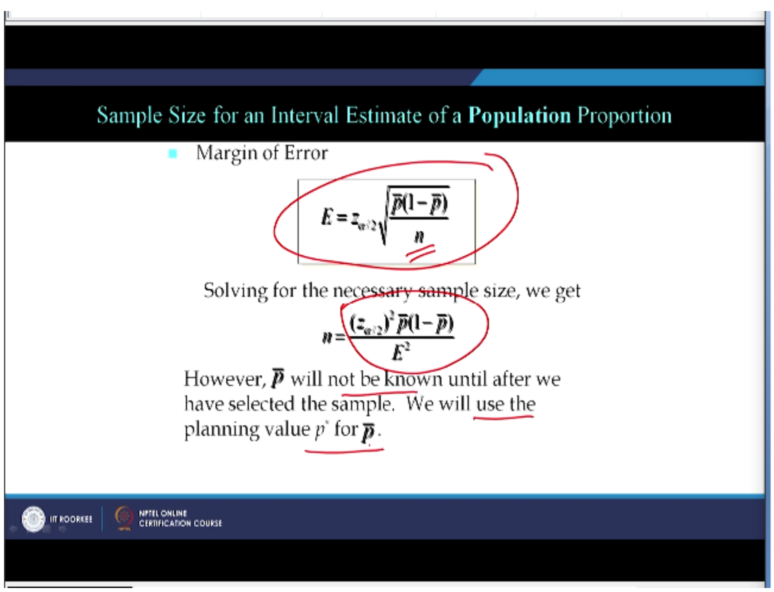
And the answers which the PSI got PSI found that 220 registered voters out of 500 contacted favored a particular candidate. So out of 500, 220 favored a particular candidate PSI wants to development 95% confidence interval estimate for the proportion of population of registered voters that favored the candidate. So out of 500, 220 favored a particular candidate, out of 100 what it would be, it would be 2 let say this is the percentage right.

(Refer Slide Time: 09:59)



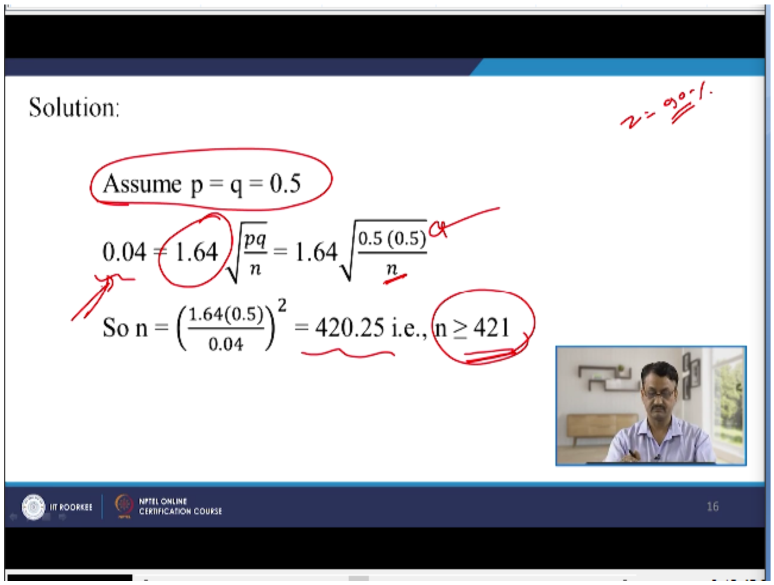
So, this value is 0.44, n is given right it is 500, Z alpha/2 or Z value is 1.96 again from table right. So 0.95/2.475 right look at Z value when probabilities 0. when area under curve is 0.7 0.475, so that would be 1.96, so this is p bar 0.44 ± 1.96 * P bar-q bar/100 right, so the range is this, this is the range. So, we say that approximately 40% to 48% people would favor that particular candidate. Now let us look at sample size determination so far we have seen confidence interval for population mean and population proportion.

(Refer Slide Time: 11:06)



Let us look at how to estimate or how to calculate sample size, we know that the margin of error is represented by this formula. So we can solve this equation for n and n is equal to this, so this will not be known until after we have selected the sample right. So we will use the planning values of p star or p bar, so this is a question wherein in fact what we did so far was we were given sample size we were given a standard deviation of population we were given confidence level what we were doing. We were finding confidence intervals, so here what we have to do here we have to find out sample size that is n value, let us look at this question.

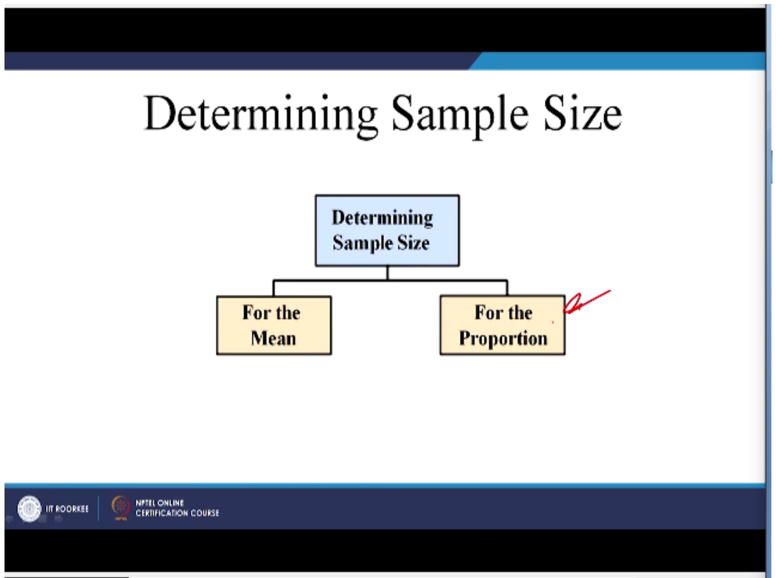
(Refer Slide Time: 12:14)



For a test market find the sample size needed to estimate the true proportion of consumer satisfied with certain new product in this range at 90% confidence level. Assume you have no strong feeling about what the proportion is, so in this case from the proportion we have to calculate sample size and we have been given precision of 0.04. Since we would not been given initial P value.

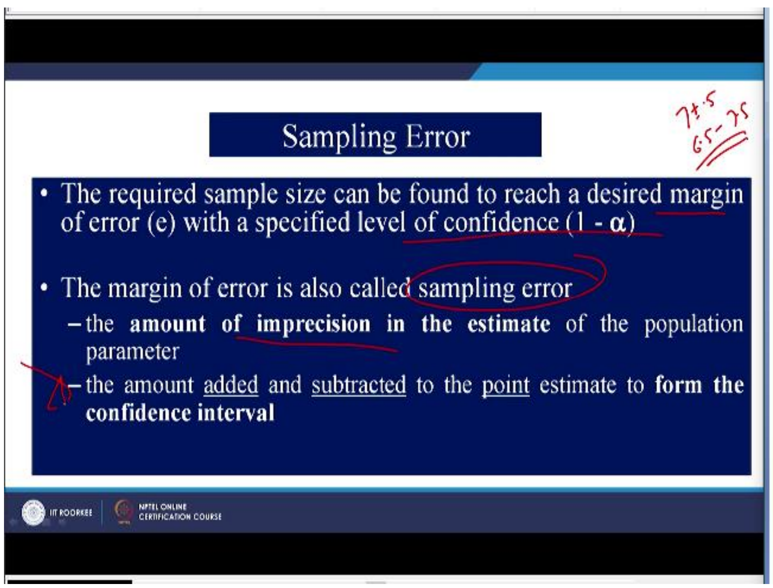
So, you can assume it 50% or 0.50. Now this is precision value this is equal to 1.6 for from where did you get this 1.64, this is Z value at 90% significance level. So 0.04 = 1/1.64 under root of pq/n since we have not been given p and q you can assume them equal and n is something which is to be calculated right. So, n=421 approximately 421, this how you should determine sample size but for this what you need?. You need precision right or what is the level of error you are accepting which determines sample size for the mean right.

(Refer Slide Time: 13:52)



We determine sample size for proportion, so we know that sampling error would have to be there for calculation of sample size.

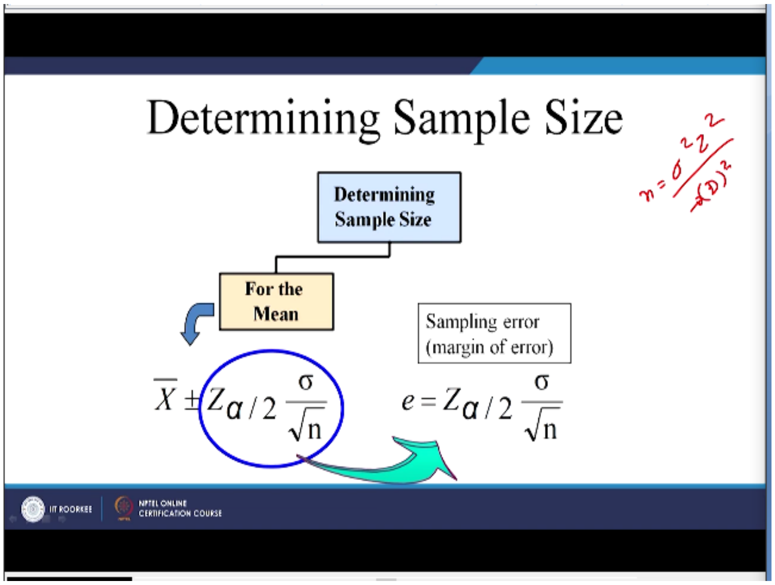
(Refer Slide Time: 14:08)



So, it can be found to reach a desired margin of error with specified level of confidence, the margin of error is also known as sampling error. The amount of imprecision in the estimate of population parameter or the sampling error, sampling error is the amount of imprecision in the estimation the amount added or subtracted to the point estimate to form the confidence interval right. So, let us say if you are saying that the point estimate of next year GDP 7 right.

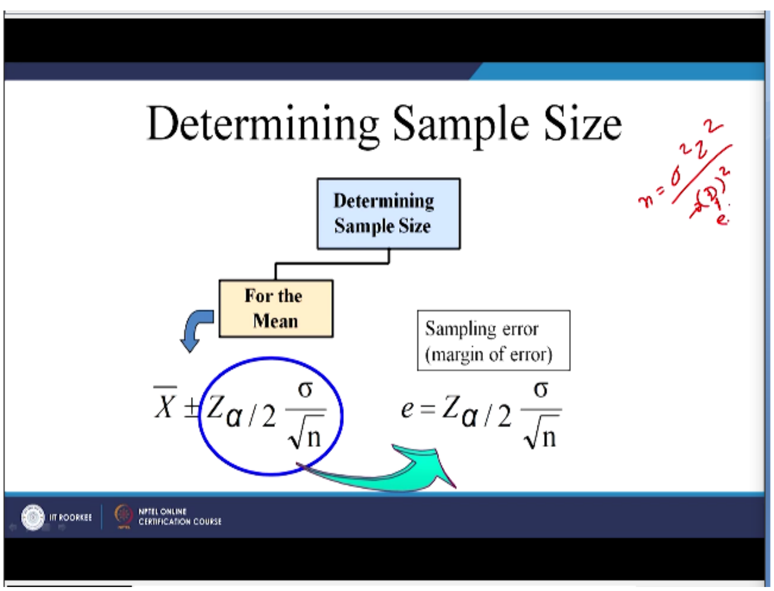
But if you let say add 0.5 and subtract 0.5 then it becomes a confidence interval. So, this is the meaning of this particular sentence.

(Refer Slide Time: 14:55)



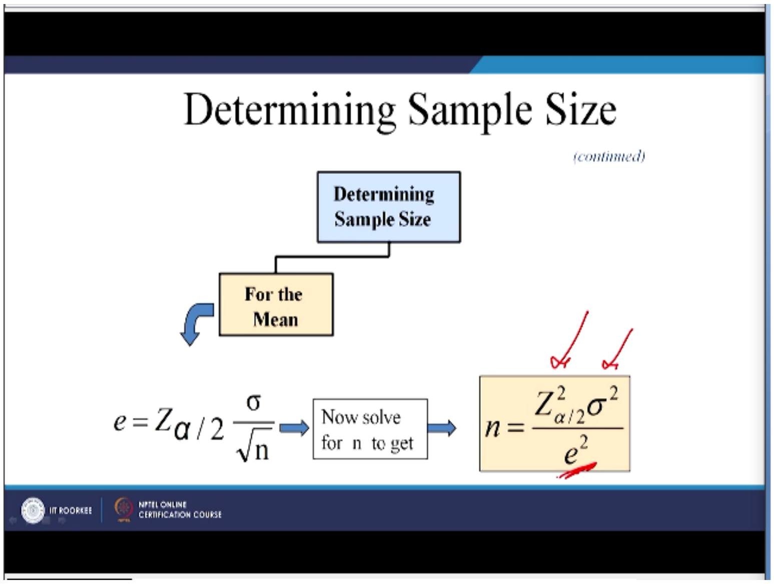
So, sample size is again using this formula you can find out it is when you take a square of this then it becomes a n=sigma square it is basically sigma square Z square divided by the precision whatever is your precision you can call it D or whatever is the symbol for precision right.

(Refer Slide Time: 15:27)

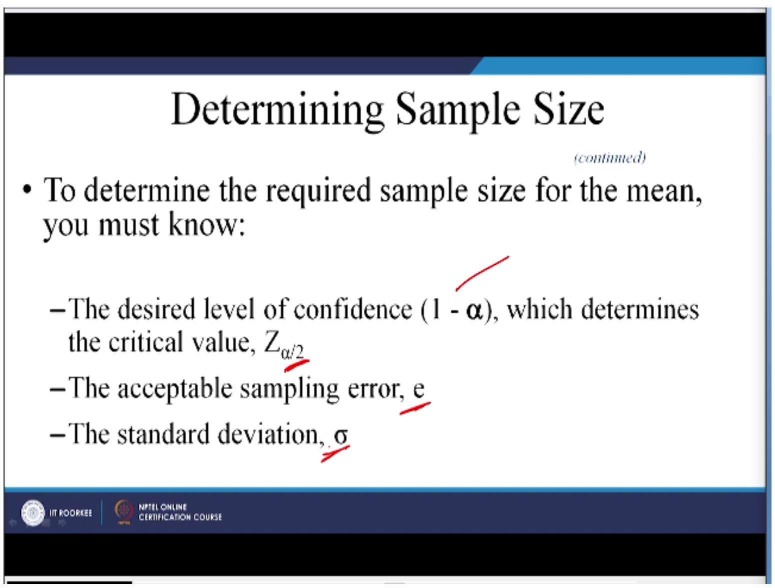


So, this is or in fact e also can be used instead of D right. So, $\frac{Z_{\alpha/2}{}^2\sigma^2}{e^2}$ ok.

(Refer Slide Time: 15:33)

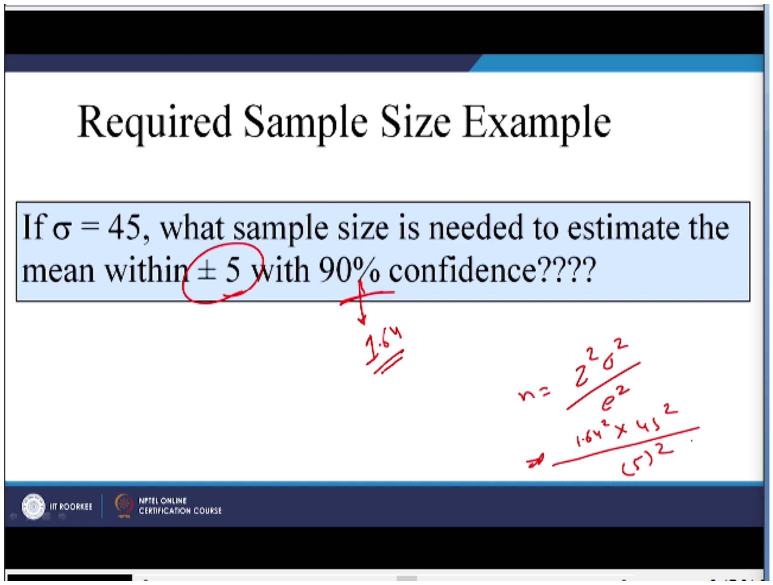


(Refer Slide Time: 15:45)



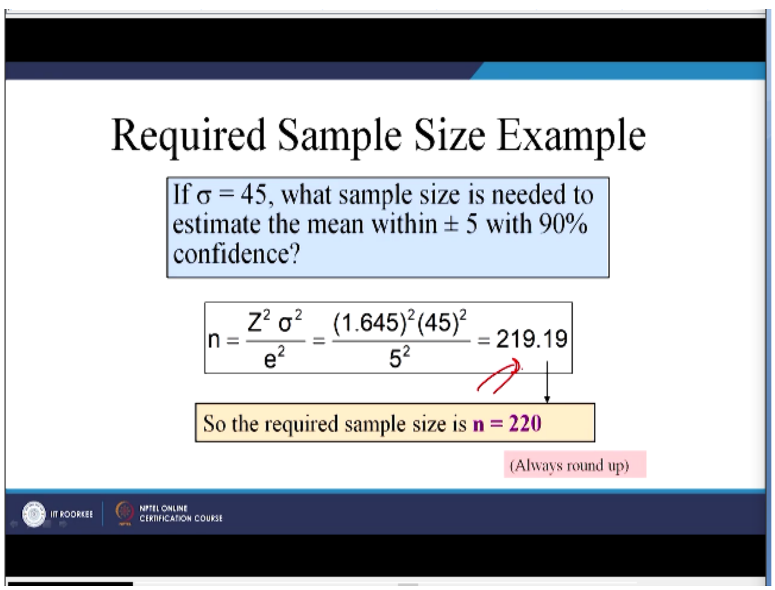
So, let us look at this question, so for finding simple size unit or this confidence level then you will be calculating you will be finding Z value from table error is required and standard deviation is required.

(Refer Slide Time: 16:12)



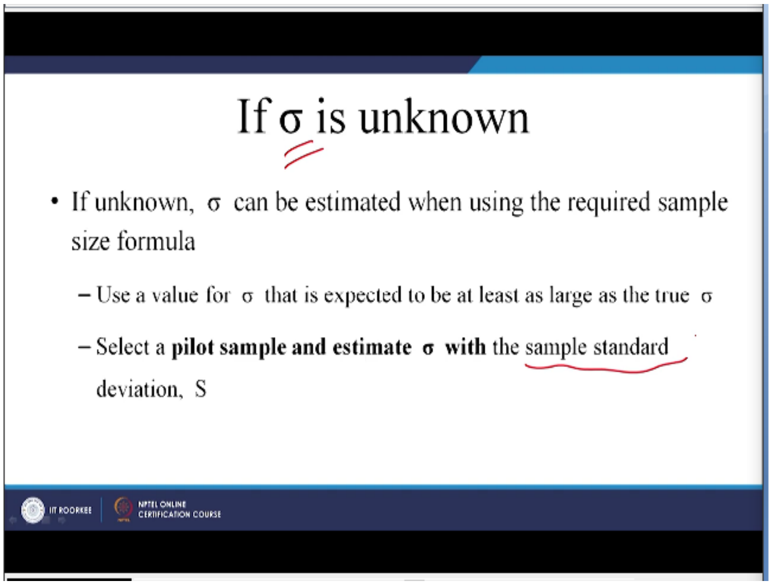
Let us look at this particular question where standard deviation is 45 what sample size is needed to estimate the mean with ±5 with 90% confidence level, so we know that at 90% confidence level Z=1.64 right sigma is given. So, our n formula is what is the value of n sigma square Z square/error term square right. So, it would be 1.64 square into 45 square/error term which is 5 square right.

(Refer Slide Time: 17:02)



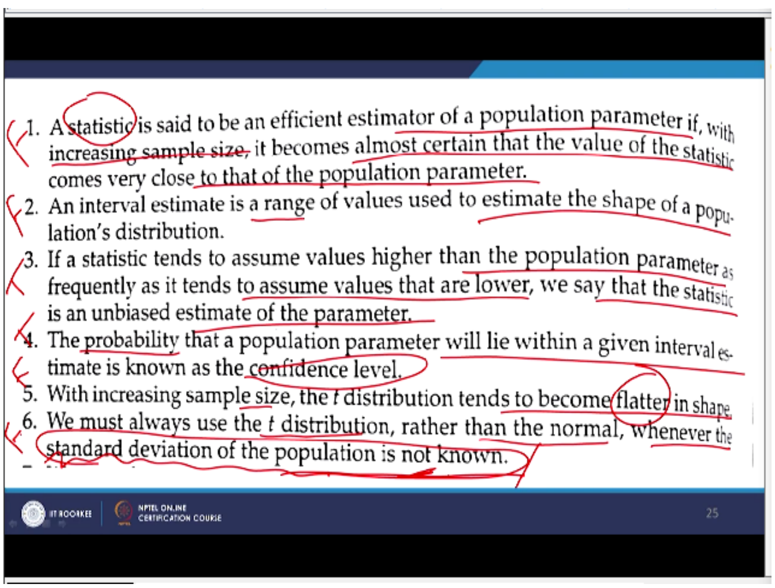
And the answer would be this approximately 220, so it is always good to roundup the fractional value. So, this how you should be calculating sample size for population mean now if standard deviation is unknown.

(Refer Slide Time: 17:26)



Then what is to be done, will be using sample standard deviation is what we have done earlier as well. In case if population standard deviation is unknown then you need to use sample standard deviation.

(Refer Slide Time: 17:46)



So, now let us workout couple of exercises, these are true and false statements, so let us look at the first statement a statistics is said to be an efficient estimator of population parameter. A statistics is said to be an efficient estimator of population parameter if with increasing sample size it becomes almost certain that the value of the statistics comes very close to the population parameter is this statement true or false just try to find out.

A statistics is said to be an efficient estimator of population parameter let us call it sample mean or sample proportion with increasing sample size it becomes almost certain that the value of statistics comes very close to the population parameter. This statement is false, this is not a correct statement, it is nothing to do with increasing sample size. The only point which you need to keep in mind is that the sample you are taking should be random.

Let us look at second statement and interval estimate is a range of values used to estimate the shape of a population distribution. An interval estimate is a range of values used to estimate the shape of population distribution, is this correct do you use this interval estimate this interval or the range to estimate shape of population distribution no this is false. We do not use range or interval estimate to estimate shape of the population distribution but we estimate population parameter right.

Let us look at next question, if a statistic tends to assume values higher than population parameter as frequently as it tends to assume values that are lower. We say that the statistics is unbiased estimator of population parameter, is it true or false, this statement is true statement. Because there is a possibility that the sample statistics may sometimes be less than population parameter and sometimes more than population parameters it is possible.

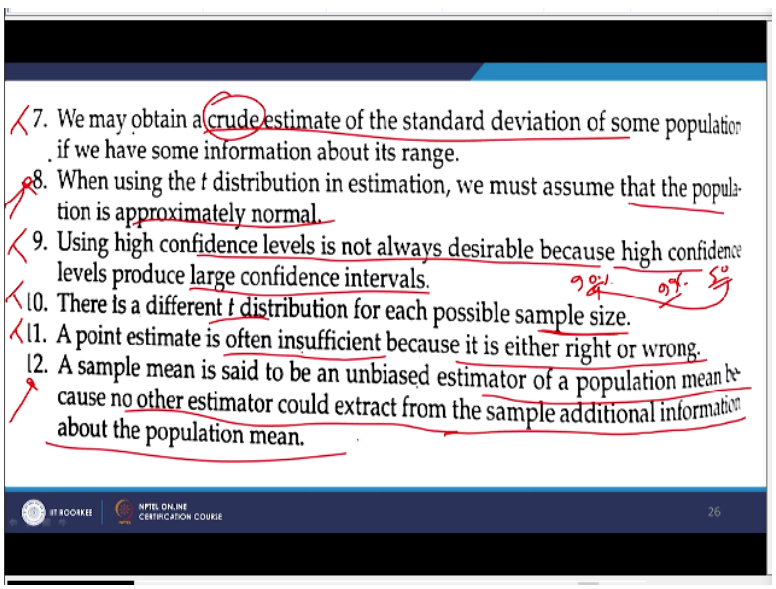
Let us look at the next one the probability that population parameter will lie within a given interval estimate is known as confidence level. What is this the probability that a population parameter will lie within a given interval estimate is known as confidence level is correct or not just think for a while, the probability that a population parameter will lie within a given interval estimate is known as confidence level, this is true because confidence level is the probability.

Probability that the population parameter will lie within confidence interval. So do not get confuse between confidence level and confidence interval. Confidence interval is the range in which population parameter will lie, well confidence level is the probability that the population parameter will lie within confidence interval. Let us look at next one with increasing sample size the t distribution tends to become flatter in shape.

When to use t distribution first of all tell me t distribution is to be used when 2 conditions are fulfilled, first is sample size is less than 30 and population standard deviation is unknown only under these 2 conditions you should use t distribution. Now the question is with increasing sample size t distribution tends to become flatter in shape is it true or false, this is absolutely false, is the other way round as we increase sample size the t distribution tends to become normal in shape not flatter in shape ok.

Let us look at 6th we must always use t distribution rather than normal whenever the standard deviation of population is not known tell me is this true or false. This statement is true or false I have just mentioned 2 conditions for t distribution the first is sample size has to be less than 30 and population standard deviation is not known right. So in this statement there is nothing like there is no mention about sample size, only population deviation is unknown. So this statement is false, had there been a statement about sample size then this statement would have become true right.

(Refer Slide Time: 24:04)



Let us look at the next one, we may obtain a crude estimate of standard deviation of some population if we have some information about it is range is it possible can we obtain crude estimate of the standard deviation if some information about it is range is given, yes this statement is true. If we know the range then there is a possibility of finding out it is standard deviation and that would be very crude estimate not the perfect estimate.

Let us look at next point, when using the t distribution in estimation we must assume that the population is approximately normal. When using t distribution in estimation we must assume that the population is approximately normal, so this statement is a true statement right. Let us look at 9th one using high confidence level is not always desirable because high confidence level produce large confidence interval this is true or false.

We always use confidence level of let say a 90% or 95% or 99% we do not use confidence level of let say 30% or 40%. So, this statement is true or false, this statement is true statement because as we increase confidence level the confidence interval will also increase. I will give an example, let say you go for buying washing machine the shopkeeper tells you that the shopping machine would be delivered within 1 hour and the confidence level is 50.

Now he also says that confident the washing machine would be reaching his home within let say 5 hours and confidence is 90%. He also says that the washing machine would be reaching home within 1 day and confidence level is 99%, so what is happening just with the increasing in confidence level the confidence interval is also increasing right. So, this statement is a true statement.

Let us look at the next one there is a different t distribution for each possible sample size is it correct or incorrect, there is a different t distribution for each possible sample size, yes this is a true statement there is a different t distribution for each possible sample size, so this is true. Let us look at the next one, a point estimate is often insufficient because it is either right or wrong this statement is true or false.

A point estimate is often insufficient because it is either right or wrong, this statement is 100% true. Because what happens when you go for point estimation and if the estimation is not correct then it would be wrong. So it is always good to go for interval estimation rather than point estimation, so point estimation is insufficient because it is either true or false right.

Let us look at this try to answer this question, a sample mean is said to be an unbiased estimator of population mean because no other estimator could extract from the sample additional information about population mean. A sample mean is said to be an unbiased estimator of population mean because no other estimator could extract from the sample additional information about population mean.

So this statement is a true statement, so let me summarize what we have done in today's session, we have looked at how to find out confidence interval for population proportion. We have looked at sample size determination in case of population proportion. So from the error term value we easily calculated n value. Similarly we have also calculated the sample size for population mean and the sample size for population mean is it sigma square Z square divide byerror from s square.

Apart from these we worked out couple of examples on this topic on estimation and confidence interval. In the next lecture will discuss some more questions on estimation that is point estimation and confidence an interval estimation with this let me complete today's session, thank you very much.