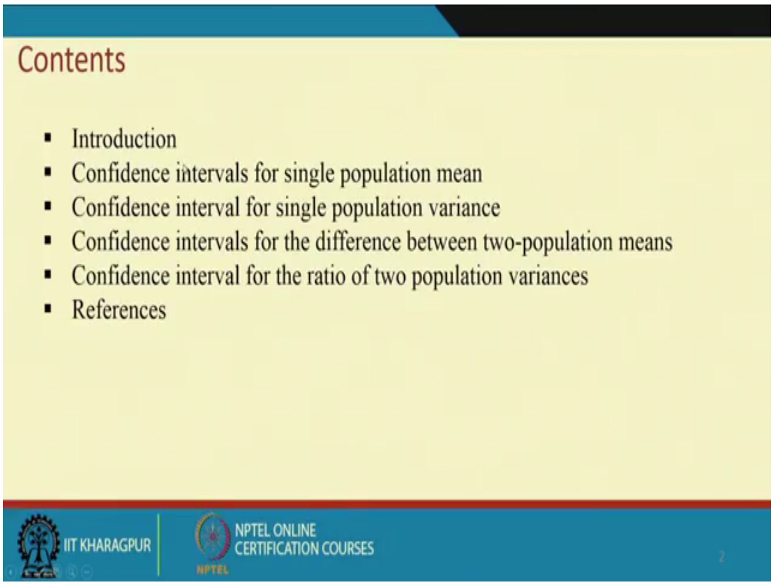
Design and Analysis of Experiments
Prof. Jhareswar Maiti
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture - 09
Estimation

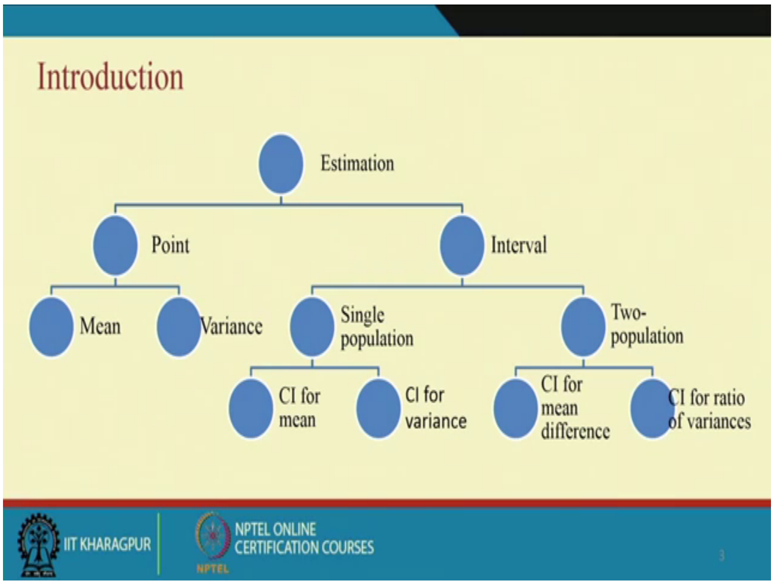
Welcome to the lecture 09 of Design Analysis of Experiment; today's topic is estimation.

(Refer Slide Time: 00:29)

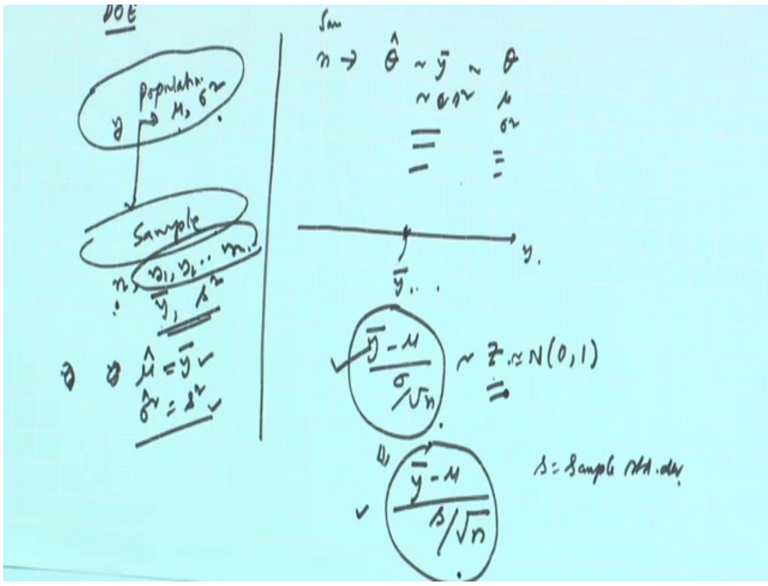


The content of today presentation will first introduce the estimation concepts and then we will find out the confidence interval for single population mean. Confidence interval for single population variance, confidence interval for the difference between two population means and confidence interval for the ratio of two population variances.

(Refer Slide Time: 00:56)



(Refer Slide Time: 01:07)



Now, I told you in earlier lectures that the concept of population and that of sample, they are very much related. Sample is taken from the population in order to estimate the population parameter.

For example, if population for a particular characteristics y having mean mu and variance sigma square and if you collect a sample of size n and that is data y 1, y 2 to y n; then you can calculate statistics y bar and a square and these statistics. If n is representative n; of for this population then these statistics can be used to estimate or as an estimate of that parameters.

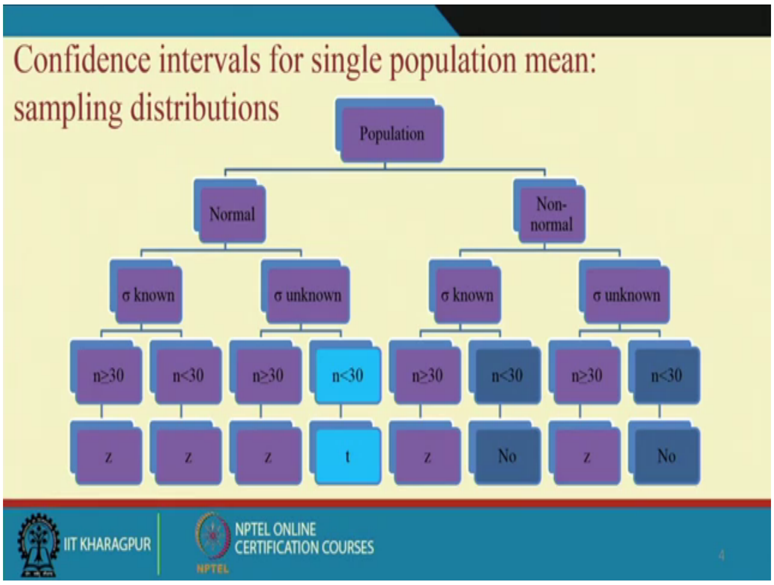
For example mu estimate equal to y bar and similarly your sigma square estimate will be a square. So, now another concept is that we said that if you go for a second sample; the estimate value will differ; go for third sample, estimate value again differs. So, at as a result what happen? There will be a range of values on within which the population parameters may lie.

This two concept will be discussed under estimation, as a result the estimation is of two types point estimation and interval estimation. When we say point estimation it is nothing, but suppose your sample average is 10; then we say this 10 is the point estimate of the population mean mu. Similarly, if sample average sample variance is 0.4 or let it be 4; then the estimate of the population variance will be 4.

But as sample statistics is a random variable; so it can take a larger number of values. So, what we require to do? We require to have a distribution of that statistics and from that distribution, we will find out some interval with certain confidence level that the population parameter will lie within this interval; with a confidence level has this as specified.

So, by estimation there will be point and interval estimation and interval estimation; it can be for single population, can be for two population and in this lecture we will be a discussing on confidence interval for mean and confidence interval for variance; for single and double populations.

(Refer Slide Time: 04:31)



So, single population part you know that you have n number of data sample data which is y 1 to y n. And then you calculate a statistics theta cap; this can be y bar can be a square can be something else. And then corresponding population parameters, suppose theta which can be mu, can be sigma square; can be others for which the statistic which computed this is 0.1 value; you get, suppose my data spread y 1 to y n here y; then if you calculate average of this, you may find out the average value is this.

So, this is y bar from a sample; this is point estimate. Now this y bar can represent mu that mean the population mean or may not and as a result; what happened we require some amount of confidence when we saying that the estimated value of population mean is this that is come us under interval estimation.

Now another important concept in estimation is that; when you are talking about the population, whether the population is normal or not normal. If the population is normal and we are talking about confidence interval for single population mean; then there maybe two situations, that is sigma is known and sigma is unknown and that is true for normal, non normal population also sigma known and sigma unknown.

Now, if sigma is known and n the sample size is large that is greater than equal to 30, then the statistics here; it is the mean will follow with certain manipulation will follow Z distribution. You have seen in central limit theorem that y bar minus mu by sigma by root n inside the limit theorem using central limit theorem last class. I have shown sigma by route n that follows Z distribution and Z distribution is nothing, but unit normal distribution with mean 0 variance one last class I told you. So, in this case sigma is known and a sample size is sample size is large suppose sample size not large less than 30. What will happen as the population is normal even if the sample size is less small; then also Z distribution these statistics, this one will follow Z distribution.

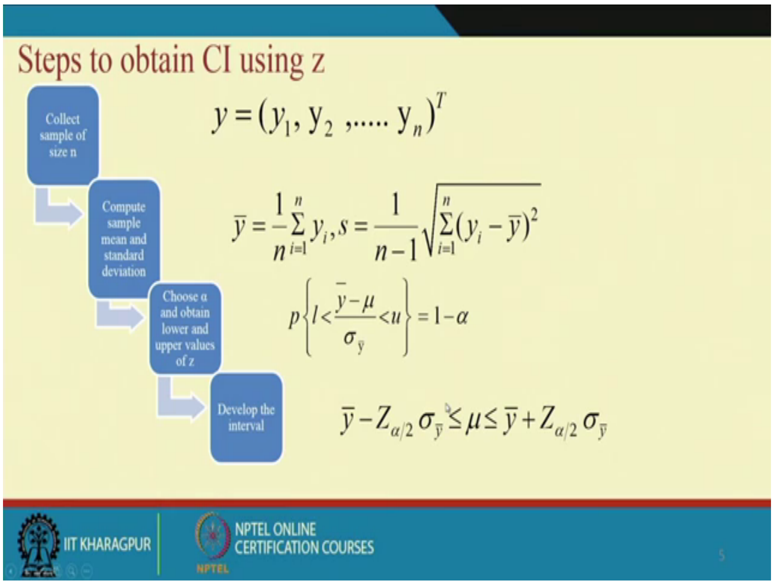
So what is the message here? If you are sampling from a normal population with known population variance, then irrespective of the sample size; the statistic y bar minus mu by sigma by root n follow Z distribution. Now what will happen when sigma is unknown? When sigma is unknown, these statistics will be read; will be transform to this; that means, what is this transformation sigma will be replaced by s. What is s here? s is the sample variance sample standard deviation.

So, that mean qualitative there is qualitative change between these statistics to these statistics, here sigma population variance is known here is population variance is not known as a result sample estimate is used that is s by root n. So, these statistics will follow normal distribution; if n is large a greater than or equal to 30. If n is small which will be the case in most of the times in this particular design of experiment situations or when you experiment; we may not have this amount of observations applications.

So, in that case; what is the issue is that, you required to use t distribution and when you will be using t distribution; I already explained you in last class. So, it is basically Z by square root of chi square by its degrees of freedom and this quantity satisfy this condition. Now, what will happen for non normal situation if sigma is known n greater than equal to 30; you see sigma known n greater than equal to 30, then you can use Z distribution this quantity follow Z distribution irrespective of whether it is coming from normal or non normal population, when n is large.

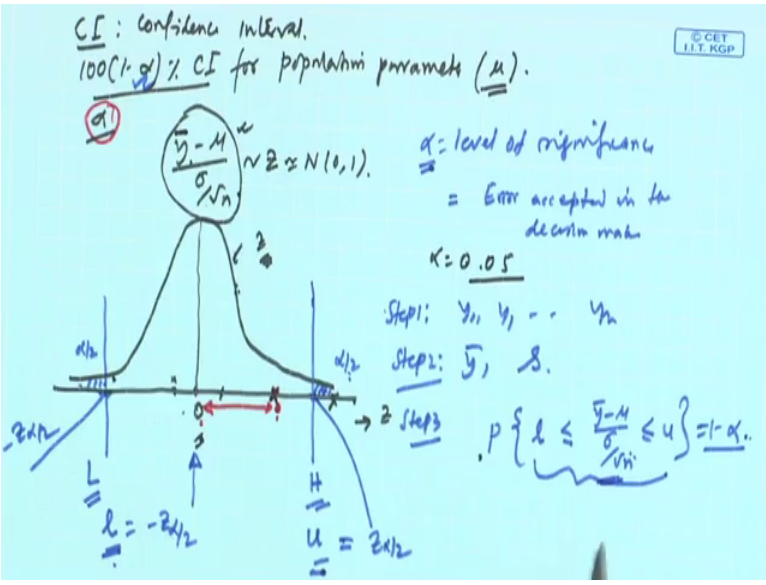
Suppose n is small; coming from normal population n is small sigma is known. So, we do not know what will be the resultant distribution of this. It is not known in case of sigma unknown, if sigma is unknown you use this and take in a large quantity; large sample size. This also follows Z distribution that is your n normal distribution; if m n less than 30; that means, it is small sample size and then distribution is not known.

(Refer Slide Time: 10:50)



Now, I will give you the steps; so how to obtain confidence interval? So, confidence interval CI; we will error paper to write down 100 into 1 minus alpha percent CI.

(Refer Slide Time: 11:05)



So, this confidence interval CI means confidence interval for whom confidence interval for population parameter.

Here we are using mu, mean mu, population parameter mu. So, few things then what is alpha here? So, I told you that y bar minus mu by sigma by root n is a statistics that follows Z distribution; Z is nothing, but your normal Z 0, 1. So, assume that this is your 0, then your normal distribution something like this.

This is Z distribution; now what this is the quantity follow? Z distribution and here one important statistics is there, y bar which is the sample average erase. If I say that this curve is the distribution of this, then if you take first sample; the value maybe somewhere here; if you take second sample value maybe somewhere here, third sample value maybe again here.

So, like this there will be n number of samples. So, n number of values and accordingly this distribution is coming. So, any value y bar value along this line; this Z line it is true, but you know that from theory that this mean theoretical mean is 0. Suppose you collect a sample and this quantity falls here, this statistics falls here is it equivalent to 0 or is it; e other way, the difference between 0 to these is sufficiently large or it is with certain confidence. It is equal to similar to 0; it is just because of randomness this difference is coming. So, if then what happened? Your basically making a distance from this to the estimated value the distance is this. So, what is the distance that is accepted?

So, that will be determined by alpha. So, alpha is known as level of significance alpha is known a level of significance. Now our; as we are interested to the target that is mean 0 here; so if you deviate from 0, both side the either right or left; so that and depending on the level of significance. So, some distance will be take it considered as acceptable beyond which it cannot be acceptable.

So, at both side; it is accept to be considered total case. So, this alpha will be divided by 2 parts; one is alpha by 2, this side the area under the curve and also this side alpha by 2. So, this alpha is level of significance mean that is the error accepted in accepting that the sample average is representative to the population mean.

So, x error in accept error accepted in the decision making; what does it mean? Suppose you take another sample and the mean value is such that the average sample; average is such that this quantity falls here. So, you are saying that this value is far away from this; so it is not representative of 0. So, you will not accept it, but please keep in mind that this axis is Z axis; where any value is true. So, by not accepting this; you may commit an error and that error is alpha; either this side or that other left side, you will commit that

that is the alpha. So, level of significance is nothing, but the error accepted in decision making.

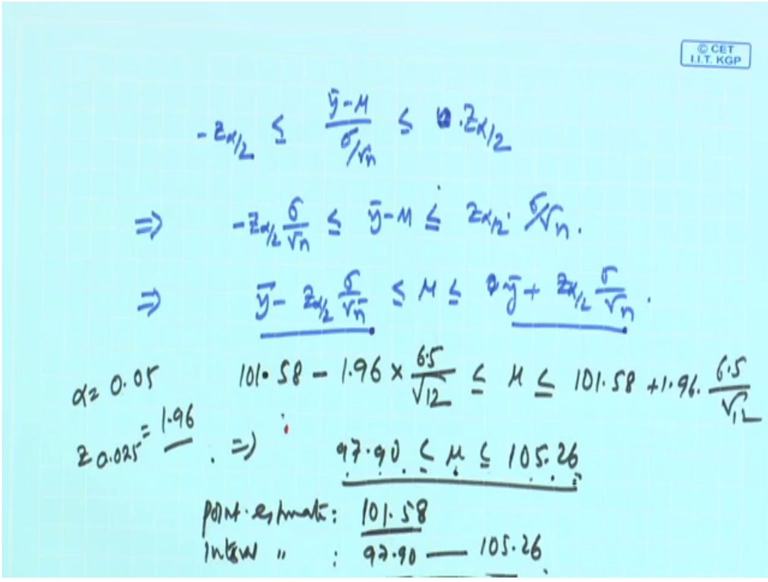
Usually alpha equal to 0.05 what does it mean? That means, 5 percent error, what does it mean? Suppose, if you take such 100 such decisions; 95 percent time, you will correct and 5 times; 95 times, 5 times it will be wrong, so that sense also you can use. So, in order to get this, what are the steps? You see the steps here, steps this first collect data, second is the compute the mean a sample average and sample standard deviation and then and what we are doing?

So, first you have data y 1 to y 2 to y n; then compute y bar and s step, that is step ome; data collection, step two computation of sample statistics. Now step three is what we say that; if the computed value fall within this range. Suppose, this is the lower range; this is the high, this is the lowest lower and this is the higher one.

Or other way, we can say this is l and this is u lower limit and upper limit and then what does it mean? If this is alpha by 2, this is alpha by 2 area under this curve is 1 minus alpha. So, we are accepting that probability l lower value less than equal to y bar minus mu by sigma by root n less than equal to u; this will be 1 minus alpha. So, this is that mean; what is the probability that the y bar that estimate of mean population mean lies between the lower specified limit and upper specified limit is 1 minus alpha.

If alpha is 0.05; this is 0.95 or 95 percent, you are confident that the estimated sample average or the sample average is the estimate of population mean. Now, as this quantity follow Z distribution; so and we all know if alpha; even alpha what is this Z value? This Z value is Z alpha by 2 alpha by 2; for what that is for the area under the curve right to this point. Similarly, this will be; this point will be minus Z alpha by 2. So, you are looking for the interval means there l is minus Z alpha by 2 and u is plus Z alpha by 2 you want this interval.

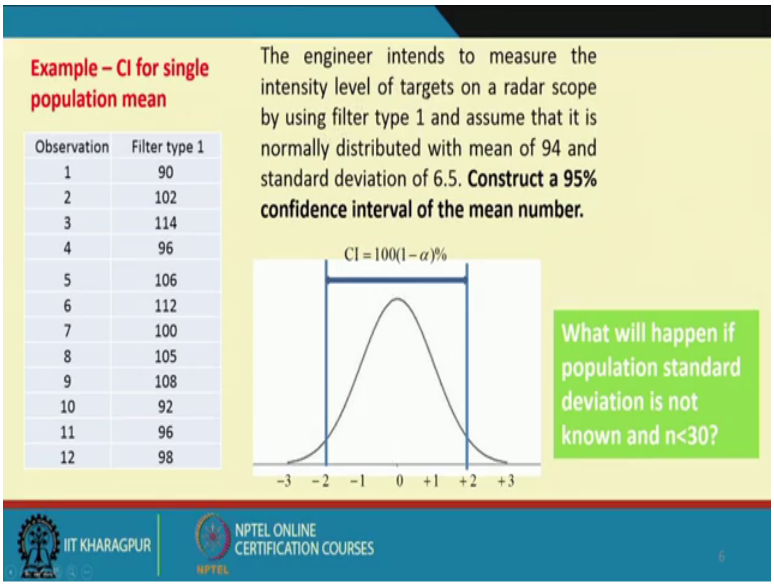
(Refer Slide Time: 19:24)



So, what you write? Then you write like this; l is minus Z alpha by 2 less than equal to y bar minus mu by sigma by root n less than equal to u. What is u? u is Z alpha by 2. So, now you manipulate what will happen? This will be minus Z alpha by 2 sigma; by root n less than equal to y bar minus mu less than equal to Z alpha by 2 sigma by root n.

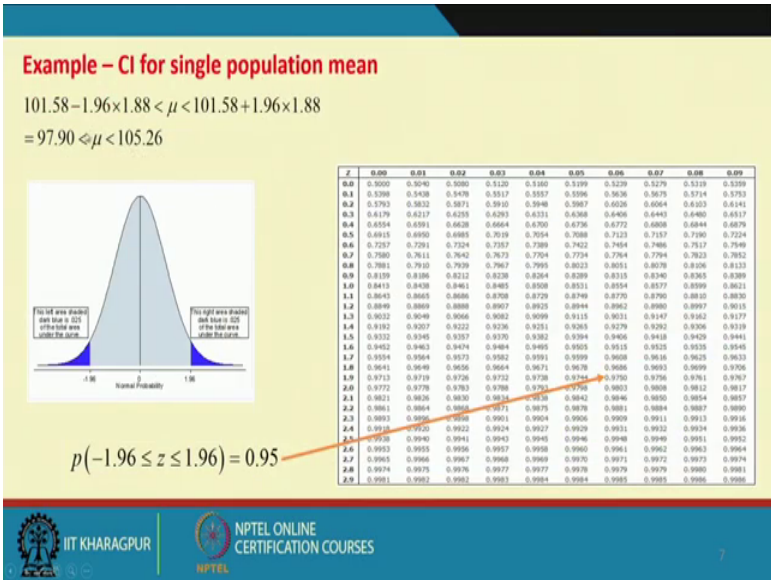
So, if you further manipulate you will find out y bar minus Z alpha by 2 sigma by root n less than equal to mu less than equal to y bar plus Z alpha by 2 sigma by root n. So, what is the confidence interval for mu? Then y bar minus this to y bar plus this.

(Refer Slide Time: 20:32)



Now, see one example; the engineer intends to measure the intensity level on target on radar scope by using filter type 1 and assume that it is normally distributed with mean 94 and standard deviation 6.5; construct 95 percent confidence interval. So, just you put what is the sample? These are the data.

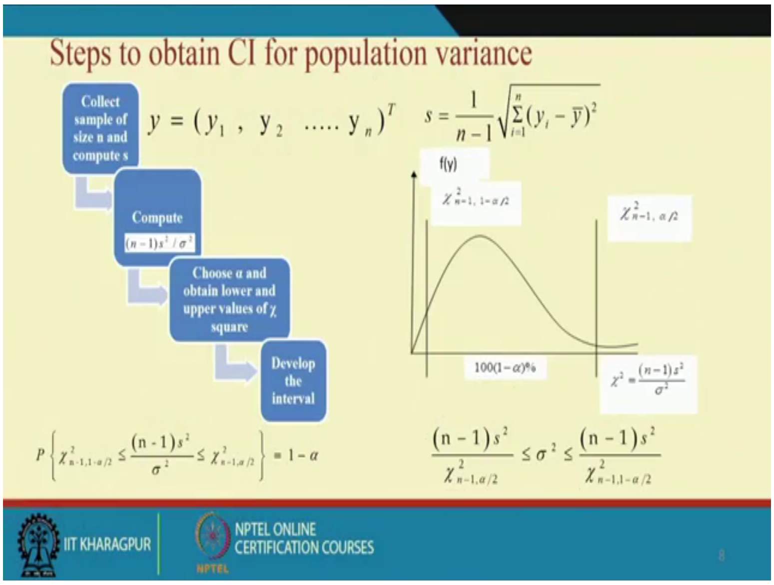
(Refer Slide Time: 20:56)



You find out the average from the data, the average from the data y bar; what is y bar? y bar is 101.58. Now, consider alpha equal to 0.05; so then Z 0.025; which is alpha by 2 which this is nothing, but 1.896. So, then 1.896 into what is sigma? Sigma is 6.5 divided by root n; that is 12 less than equal to mu less than equal to 101.58 plus 1.898 into 6.5 by 12. So, this quantity ultimately will give you 97.90 less than equal to mu less than equal to 15.26. So, this is the confidence interval for mu; what does it mean? This interval 99.90 to 905.26; this interval contains mu with a confidence of 95 percent.

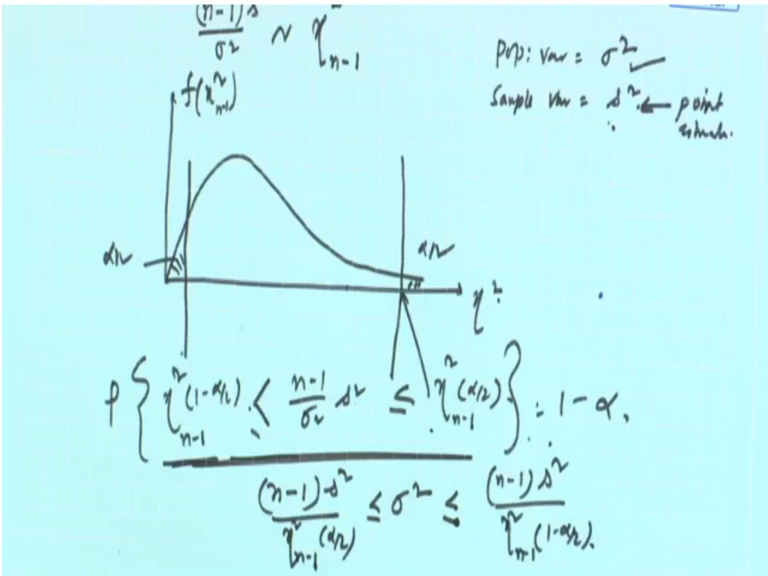
This interval contains mu, so then what is the point? Estimate here point; estimate is 101.58; interval estimate is 97.90 to 105.26. Now, see what happen; this one falls within this definitely. So, this is the concept of point estimation and interval estimation.

(Refer Slide Time: 22:46)



Now, come to the second one that population variance; so single population variance. So, here what you will do basically?

(Refer Slide Time: 23:08)



So if you can recall the last lecture where we have explained that we have created n minus 1 by sigma square into a square; this quantity and we say that this quantity follows chi square distribution within minus 1 degrees of freedom; chi square distribution with n minus 1 degrees of freedom.

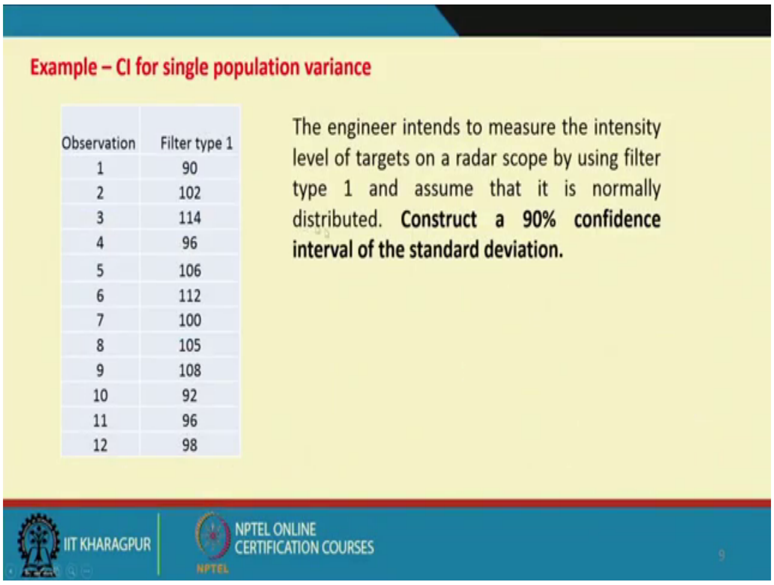
Now, you do one thing; you assume that this is the theoretical chi square distribution. So, chi square then this is the PDF of chi square and it is n minus 1. So, suppose what happened that this is the chi square distribution, I do not know what will be the exact nature, but this is the case. So, in that case what happened? We are interested to know; what happen, actually I am repeating.

So, your population variance is sigma square; you do not know it; you have collected a sample; sample variance is s square, this value is point estimate. So, if you collect several sample s will change; so you require interval. So, now you may create a interval like this here and here.

So, this side is alpha by 2; this side is also alpha by 2, then what is this quantity? chi square n minus 1 alpha by 2 right to this, what is this quantity chi square n minus 1; right to this right to this 1 minus alpha by 2. So, again what you do? Then probability at this less than equal to n minus 1; by sigma square into s square less than equal to; this is 1 minus alpha, so now you manipulate this. So, what you want from here? You will find out that sigma square you want less than equal to less than equal to and that will be n minus 1 s square by chi square n minus 1 alpha by 2 and then this will be n minus 1; s square by chi square n minus 1; 1 minus alpha by 2.

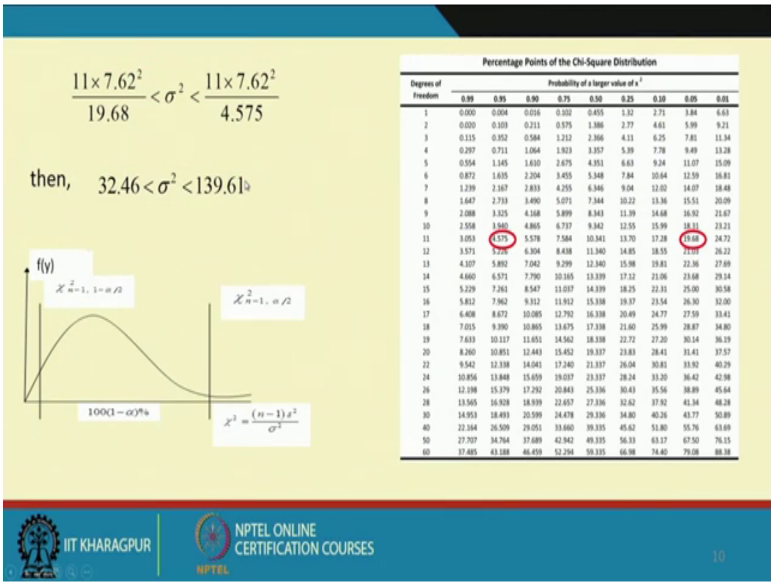
What is happening here? Denominated n minus 1; s square n minus 1, s square a that is numerator, then denominator here chi square alpha; alpha by 2 and this value, this is large value chi square 1 minus alpha by this value, this is smaller than this value. So, this quantity will be smaller than this quantity; obviously, and this is the interval with.

(Refer Slide Time: 26:23)



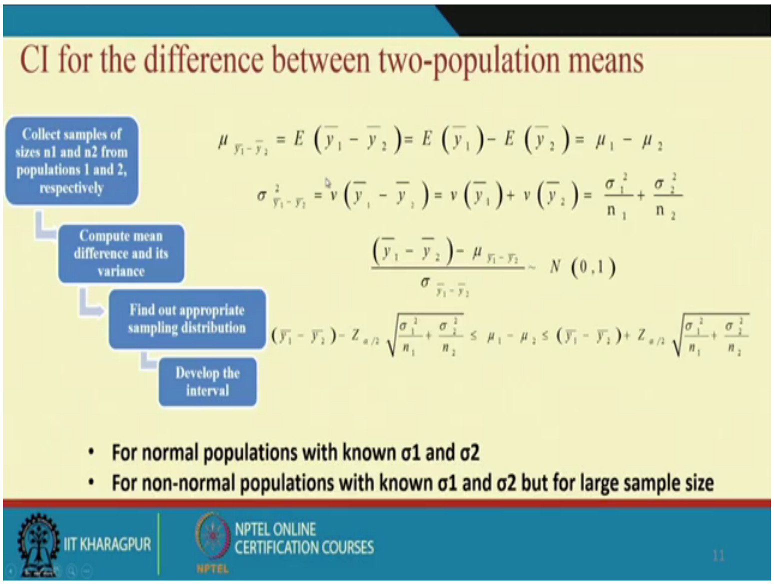
So, now see this for the same example; we are interested to know the confidence interval for population standard deviations.

(Refer Slide Time: 26:35)



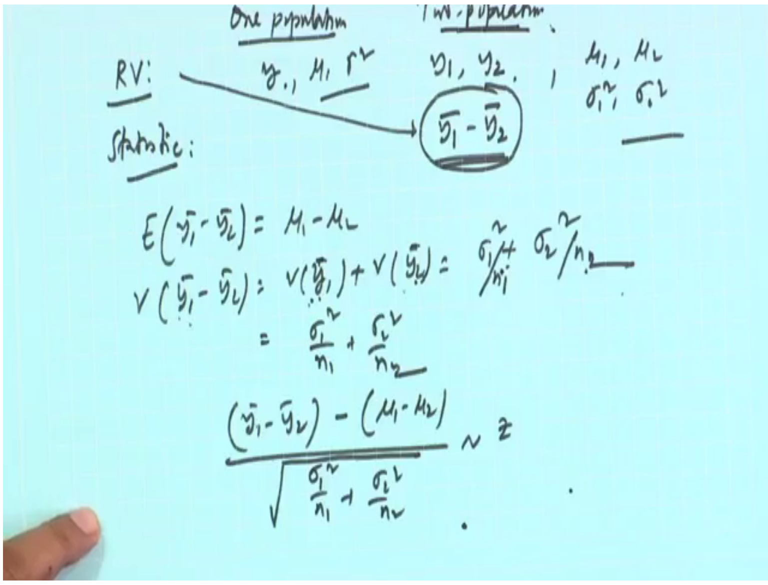
We just put this all those values here, you see 11, seven point; these two by this, so this is the interval for variance. Now, if you take the square root; you will get the confidence interval for standard deviation.

(Refer Slide Time: 26:55)



Now, very quickly I will just go for two population means; so please remember here.

(Refer Slide Time: 27:04)



We are interested to know the random variable of interest the statistic, so what happens? So, when it is one population random variable interest is one population random variable is y 2 population random variable y 1, y 2. So, now in two population means our statist interesting 1 is we want to test whether y 1 bar minus that this is what is our; I can say that resultant random variable from the objectives point of view, we want to see that the difference between two population means.

So, this is our random variable of interest; what do we want to know? We want to know the expected value of y 1 bar minus y 2 bar; that will be mu 1 minus mu 2, you also require to know variance of y 1 bar; y 2 bar, it will be variance of y 1 bar plus variance of y 2 bar; because y 1 bar, y 2 bar; at 2 coming from 2 different population, they are independent. So, this will be your sigma 1 square plus sigma 2 square; when the two population mean is mu 1 and mu 2 and variance is sigma 1 square and sigma 2 square, here you have seen mu and sigma square for one population.

So, is it correct that y 1 bar and y 2 bar will be this; if it is y 1, y 2; this is correct. If it is y 1 bar, y 2 bar; it is not correct, this will be by n 1 by n 2; because I told you in last class that if y is normally distributed with mu and sigma square or no y; then what will happen? If y is; y variability is sigma square then the y bars variability will be sigma square by n that is what you are writing.

So; that means, this is sigma 1 square by n 1 plus sigma 2 square by n 2. Now from central limit theorem, you can write that y 1 bar minus y 2 bar minus, its expected value is mu 1 minus mu 2; divided by the sigma by root n component, that is sigma square by n 1 plus sigma 2 square by n 2 or square root this follows Z distribution.

This follows Z distribution for large sample size; small sample size, if whatever maybe if it is coming from if the sigma 1, sigma 2 are known; coming from normal population. Both the population are normal and then using the concept that is the confidence interval that probability low less than this value lies, within the lower limit and upper limit that will give you the interval. But these two population means is very important concept and that the confidence interval calculation is also very important.

Because in end of an doe; this subject, what I will see that, it is every time; we will see that there will be concept hypothesis testing and confidence interval estimation a confidence interval for one population, then two population, then more than two population kind of things that mean one label power that; that means, two factor with two label it is two population factor with three levels three population.

So, many come when more than three that is the multi level case. So, every time we will go for confidence interval and as well as; what is the other one? That is the pair wise comparison simultaneous confidence interval, many things will be discussed. So, for the timing this much as although I said that we want to cover up to two population variances, but in next class I will explain; I will start this one two population confidence level; two population mean and ratio between two population variances.

Thank you very much.