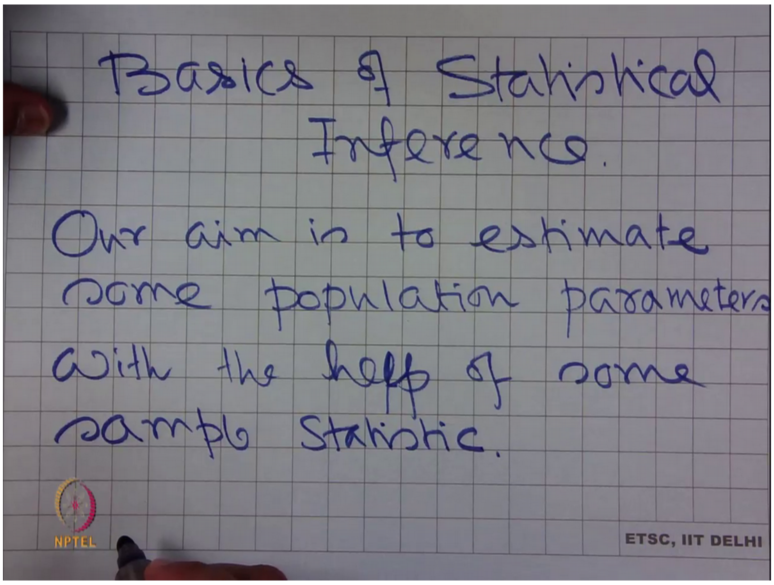
Statistical Inference
Prof. Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 07
Statistical Inference

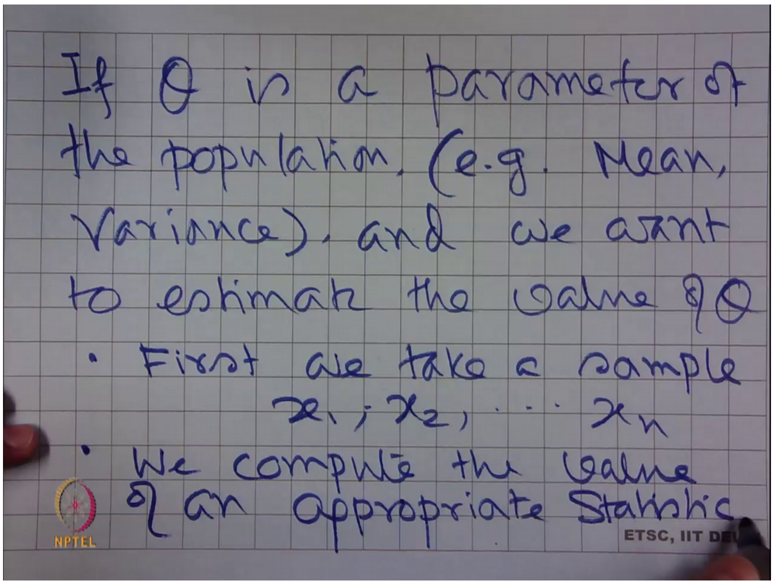
Welcome students to the 7th lecture of the MOOC's series on Statistical Inference. In the last 6 lectures, I have covered some very basics of statistical inference including simple random sampling, with replacement and without replacement, and also some probability distributions namely chi square t and f which are very important from the inference point of view as we will see later.

(Refer Slide Time: 01:12)



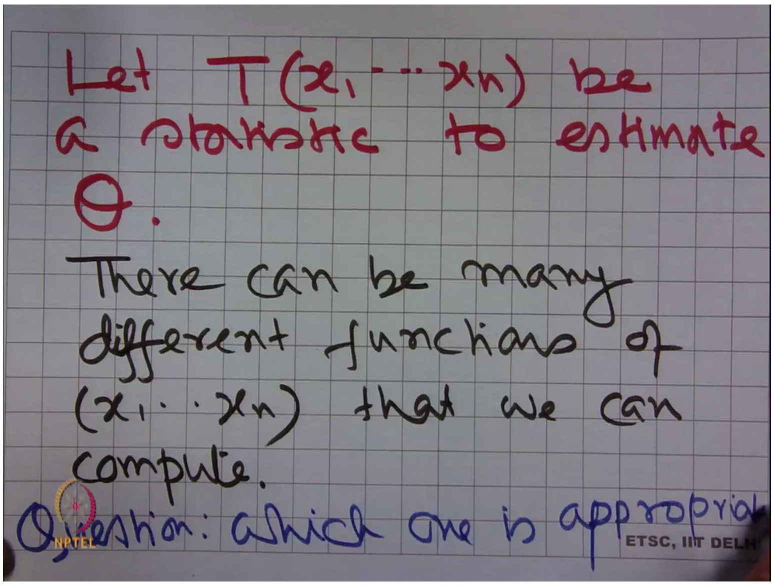
Now, let us look at some basics of statistical inference. You know that our aim is to estimate some population parameters with the help of some sample statistic.

(Refer Slide Time: 02:14)



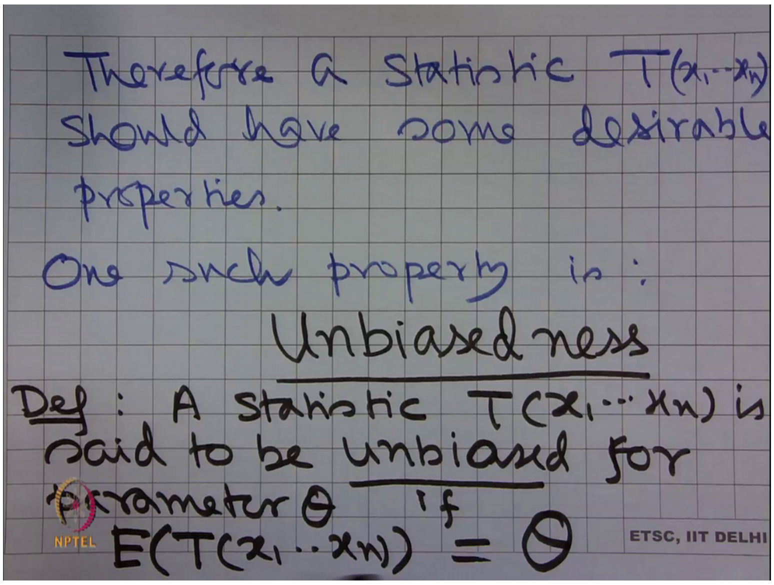
So, if theta is the parameter of the population such as mean variance. And we want to estimate the value of theta, then what we do? First we take a sample x 1 x 2 up to xn.

(Refer Slide Time: 03:48)



Then we compute the value of and appropriate statistic. Let T x 1 x 2 x n be a statistic to estimate theta. Now there can be many different functions of x 1 x 2 x n that we can compute. Therefore, question is how do you choose an appropriate statistic.

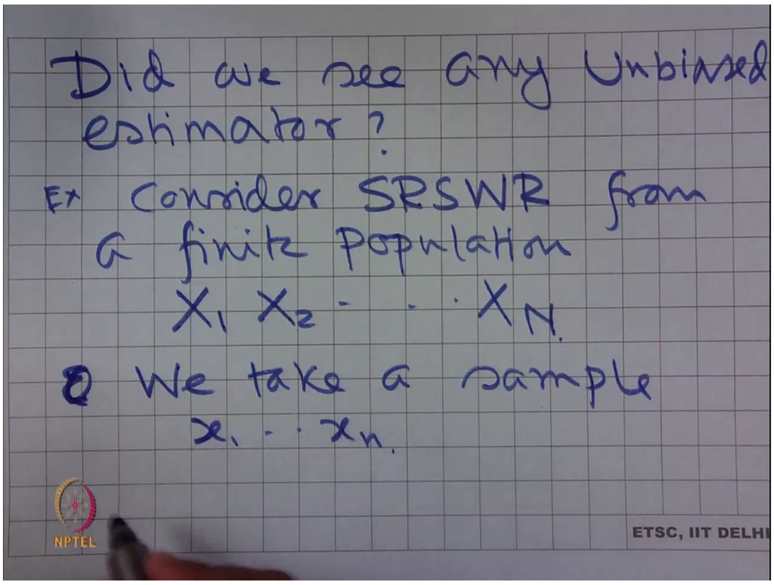
(Refer Slide Time: 05:08)



Therefore a statistic T of course, of x 1 of x n should have some desirable properties. I talk about the simplest of the property and easiest of the property to understand.

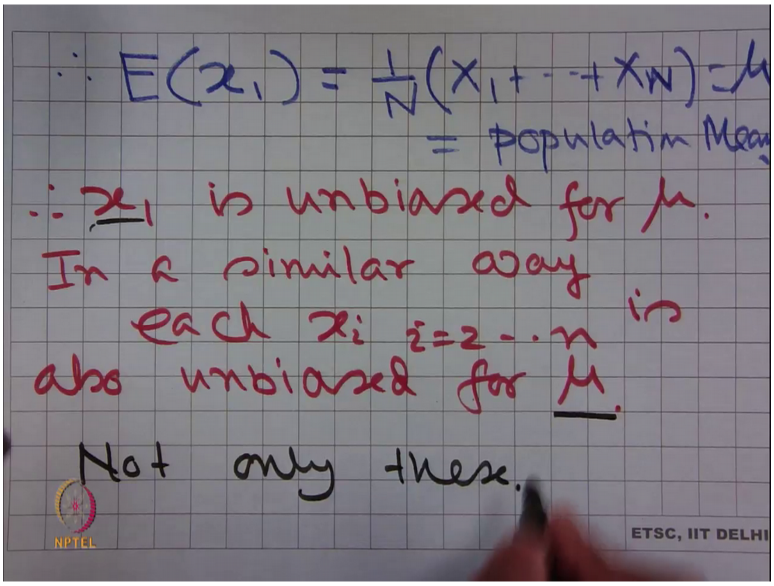
And many of you might have rightly guessed, the property that I am talking about is unbiased ness. What does it mean? A statistic T x 1 up to x n is said to be unbiased to for parameter theta if expected value of T x 1 to x n is equal to theta; that is, our aim is to estimate the parameter theta, we have taken a sample x 1 x 2 up to x n, we compute the value of T, the expected value of that statistic is unbiased for theta, then we call that particular statistic to be unbiased for theta, and that value that we compute out of the sample taken may be considered an appropriate value for theta or an estimate for theta.

(Refer Slide Time: 07:52)



Have we seen such unbiased estimator? What do you think? Yes, we did. For example, consider simple random sampling with replacement from a finite population X 1 X 2 X N. What we know that we take a sample x 1 x 2 xn.

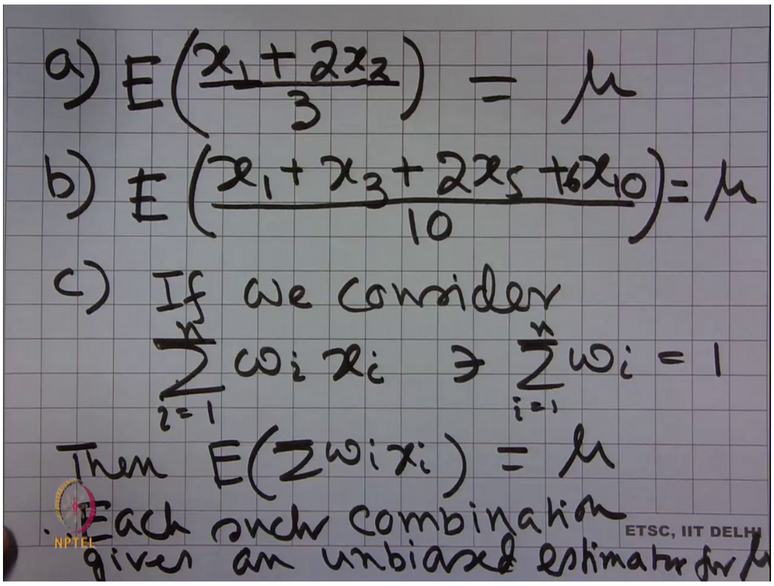
(Refer Slide Time: 09:02)



Therefore, expected value of x 1 is equal to 1 by n x 1 plus xn is equal to mu is equal to population mean. Therefore, x 1 is unbiased for mu. In a similar way, each xi i is equal to 2 to n is also unbiased for mu. Because expected value of each xi is going to be population mean.

And this is true not for simple random sampling with replacement, even if we do without replacement, we have seen earlier that each one of them will actually be an unbiased estimator for mu. Because each xi irrespective of whether it is with replacement or without replacement. We will take the values x 1 x 2 and x capital N each with probability 1 by N. Not only this, can you find some other unbiased estimator for mu? So, I give you a few consider x 1 plus 2 x 2 by 3.

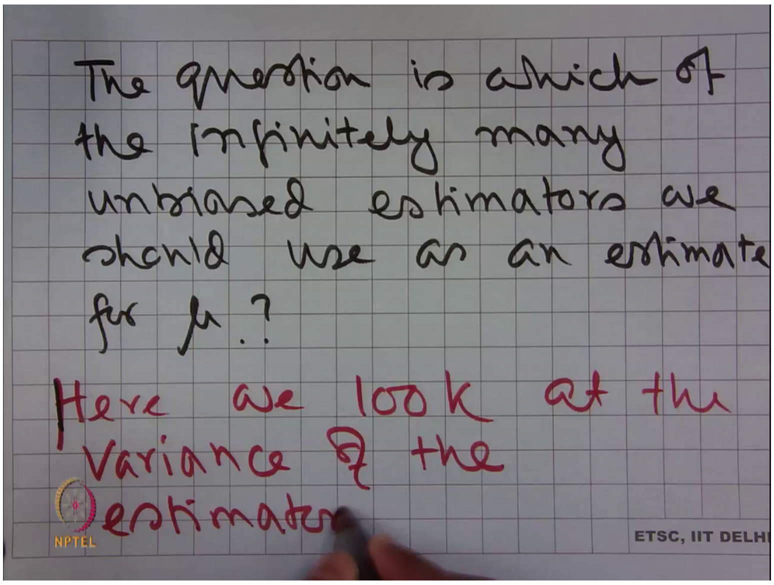
(Refer Slide Time: 11:10)



What is going to be it is expectation? It is expectation is going to be for this x 1 we will get mu from this x 2 we will get 2 mu. So, the sum is 3 mu divided by 3 is equal to mu. In a similar way, let us consider x 1 plus x 3 plus 2 x 5 plus x 10 divided by 10. What will be it is expectation? So, expectation of x 1 will be mu, this will give another mu, this will give 2 mu, and this will give some 6 mu. So, the sum is going to be 10 mu, that divided by 10 is equal to mu.

In fact, if we consider sigma wi xi i is equal to 1 to n; such that sigma wi is equal to 1, then expected value of sigma wi xi is equal to mu, that is very clear. Therefore, each such combination gives an unbiased estimator for mu..

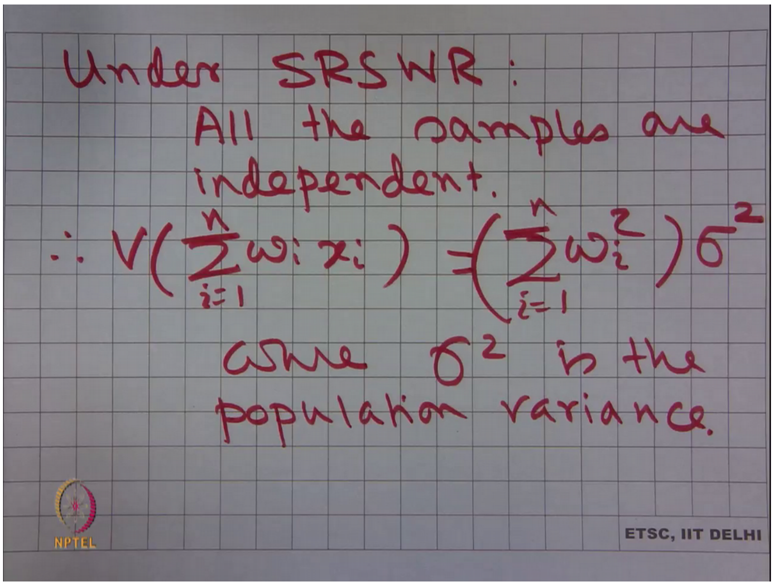
(Refer Slide Time: 13:44)



Thus we can see, we can find many different unbiased estimator. Therefore, the question is which of the infinitely many unbiased estimators we should use as an estimate for mu.

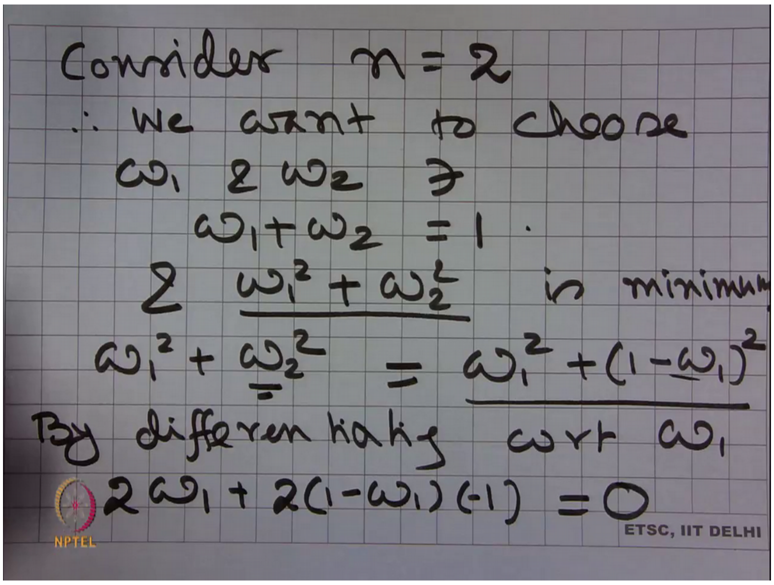
The most important concept here is the variance of the estimator.

(Refer Slide Time: 14:59)



Under SRSWR all the samples are independent, therefore, variance of sigma wi xi i is equal to 1 to n is equal to sigma wi square i is equal to 1 to n into sigma square where sigma square is the population variance. So, the question is how do we choose the weights of the different samples. So, that the linear combination wi xi is an unbiased estimator for mu, and variance of sigma wi xi is the minimum.

(Refer Slide Time: 16:34)

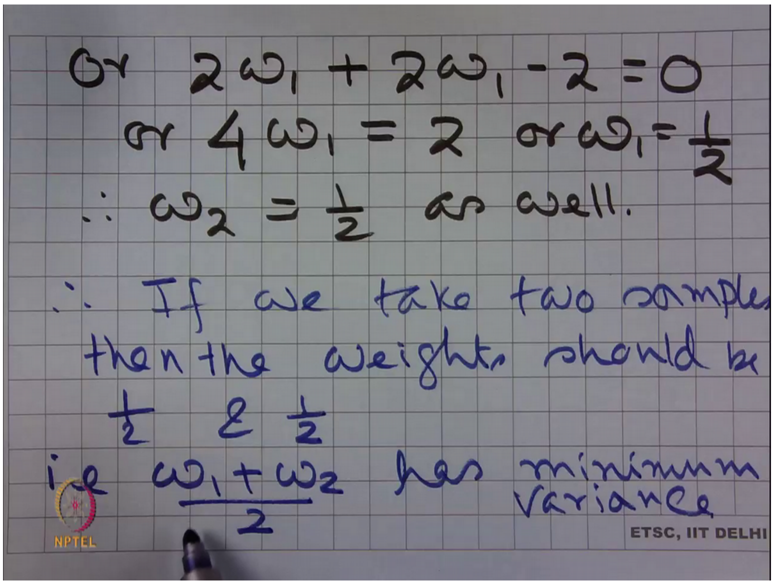


Consider n is equal to 2.

Therefore we want to choose w 1 and w 2 such that w 1 plus w 2 is equal to 1. And w 1 square plus w 2 square is minimum. Do we know the answer? We know, but let me still work it out. So, we are minimizing w 1 square plus w 2 square. By putting w 2 is equal to 1 minus w 1 from here is equal to w 1 square plus 1 minus w 1 square.

In order to minimize this, let us differentiate it with respect to w 1, what we get? 2 times w 1 plus 2 times 1 minus w 1 into minus 1 because of this minus sign is equal to 0.

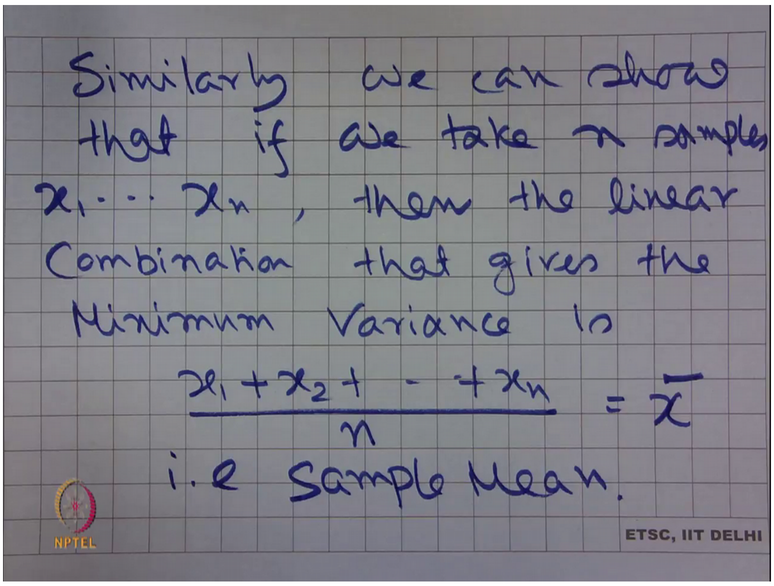
(Refer Slide Time: 18:24)



Or 2 times w 1 plus 2 times w 1 minus 2 is equal to 0. Or 4 times w 1 is equal to 2 or w 1 is equal to half. Therefore, w 2 is equal to half as well.

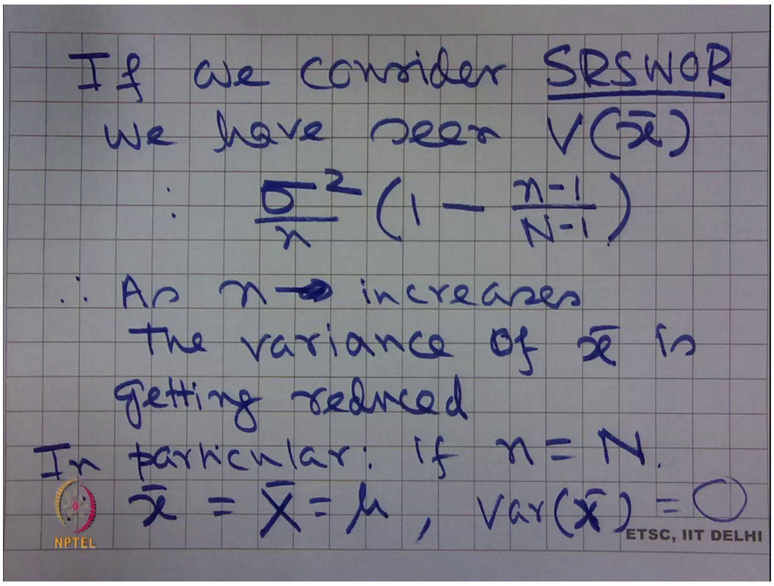
What does it tell you? It tells us that if we take 2 samples, then the weights should be 1 by 2 and 1 by 2; that is, w 1 plus w 2 by 2 has minimum variance. And what is going to be that variance, we know that that variance is going to be sigma square by 2. Or in other words if I take 2 sample, then arithmetic mean of the sample values is going to be the minimum variance unbiased estimator for mu.

(Refer Slide Time: 20:15)



Similarly, we can show that if we take a n samples x 1 x 2 up to xn, then the linear combination that gives the minimum variance is x 1 plus x 2 plus xn by n; that is, x bar that is the sample mean. Therefore, sample mean is not only unbiased estimator for mu; it is also having minimum variance among all linear combinations of the sample values x 1 x 2 xn.

(Refer Slide Time: 21:55)

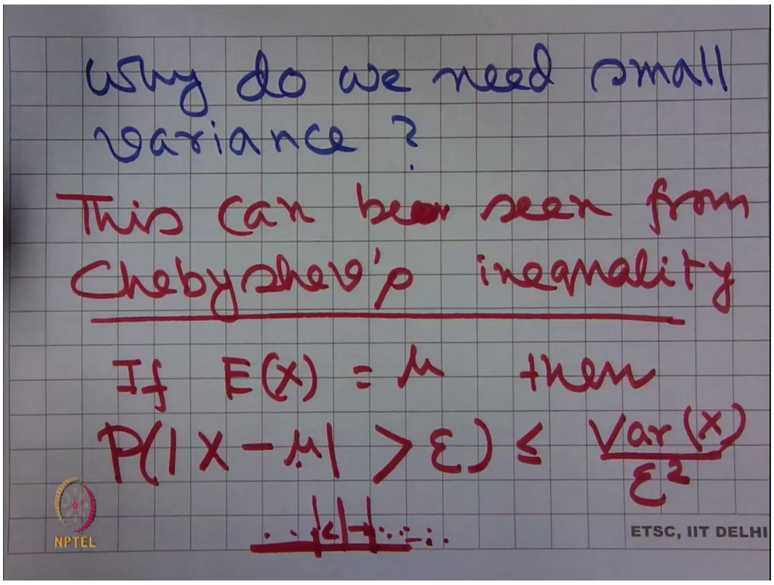


If we consider SRSWOR we have found what is the variance of x bar? Do you remember? It is sigma square by n into 1 minus n minus 1 upon n minus 1. What does it mean? It means that as n increases the variance of x bar is getting reduced. Because as n increases this value decreases, and not only these value decreases as n minus 1 increases, 1 minus n minus 1 upon capital N minus 1 also decreases. Therefore, the overall variance keeps on reducing as we keep on taking more and more samples.

In particular, if n is equal to capital N; that means, my population size is capital N, I am taking small n many samples, but basically here I am seeking, here I am saying that I am taking capital N many samples, and since it is SRSWOR. What does it mean? It means I have taken the entire population as my sample. And therefore, the sample mean is same as the population mean, which is mu and the variances is; this is n minus 1. So, this part becomes 0 therefore, variance becomes 0, which is understandable.

Because if I consider the entire population, and take it is mean then it has to be same as the population mean, therefore, there is no deviation from the population mean and therefore, variance is going to be 0; that is, therefore, there is no dispersion and we get the sample mean equal to population mean and therefore, variance of sample mean is 0.

(Refer Slide Time: 25:08)



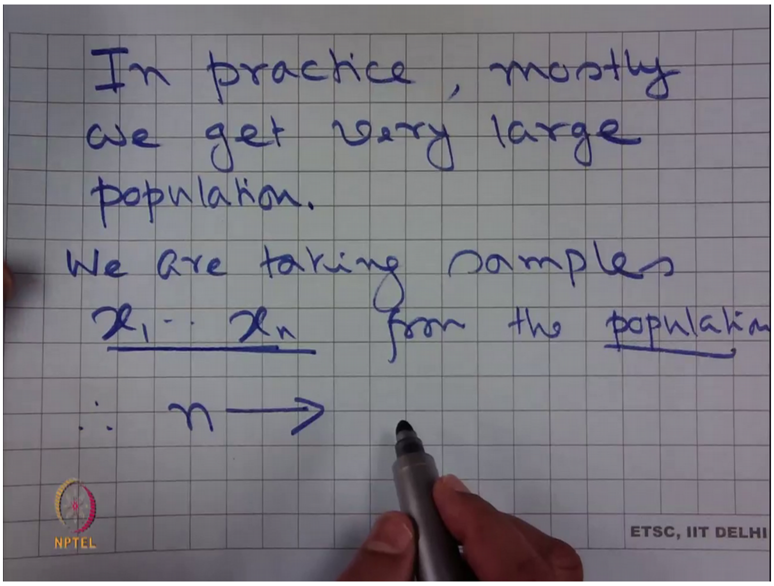
Now, you may ask me why do we need small variance because variance is a measure of dispersion.

So, smaller is the variance; that means, my estimate is very close to the parameter that it is estimating. This can be seen from chebyshevs inequality, are you familiar with chebyshevs inequality? I hope all of you have done in the first course that you might have done on probability the concept of chebyshevs inequality. If you do not know I am keeping it as a practice problem in the tutorial sheet one, you should try and prove chebyshevs inequality. This is not to be graded by us it is for your own knowledge that you try to prove chebyshevs inequality by studying some material.

If we cannot, we will upload the solution. But the chebyshevs inequality says that if expectation of x is equal to mu, then probability modulus of X minus mu greater than epsilon is less than equal to variance of X upon epsilon square; that means, mu is the expectation that probability X minus mu greater than epsilon so, if we take this epsilon; that means, that the probability that x will lie outside this, outside this limit, that is going to be less than equal to variance of x upon epsilon square.

Therefore, as variance of x gets smaller, the probability that it is going beyond that gets smaller, or in other words it says that that x will remain within epsilon distance of the mean that probability increases if the variance of x gets smaller.

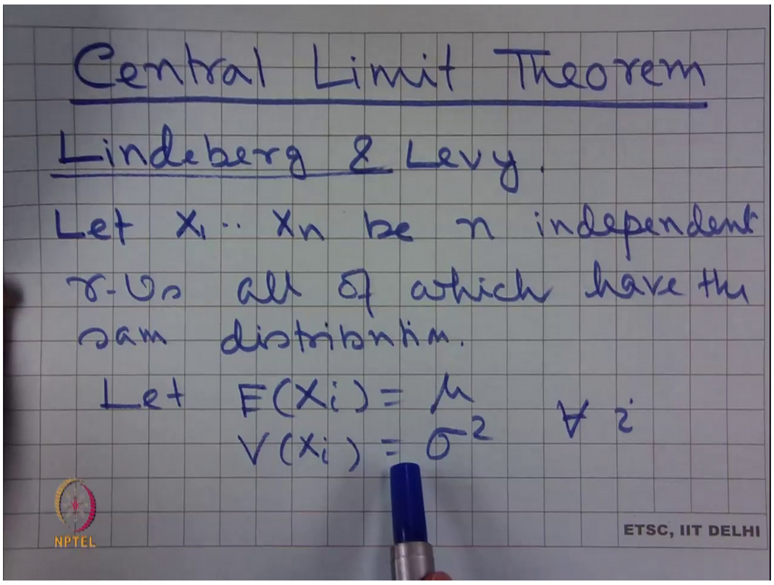
(Refer Slide Time: 28:06)



Now, in practice, mostly we get very large population we are sampling x 1 x 2 xn from the population. Obviously, if the population is large we cannot take a very small sample. We have to take larger sample otherwise; we cannot get meaningful estimate of the population parameter.

For example, if you want to compute the average income of the people of Delhi, we cannot just take few samples and based on the average of the sample we can say that is going to be the average income of Delhi population; no, that will not work we have to take meaningful size sample. So, that it represents the population. Therefore, what will happen the sample size is going to increase, n increases, right? And if n increases we have something called central limit theorem.

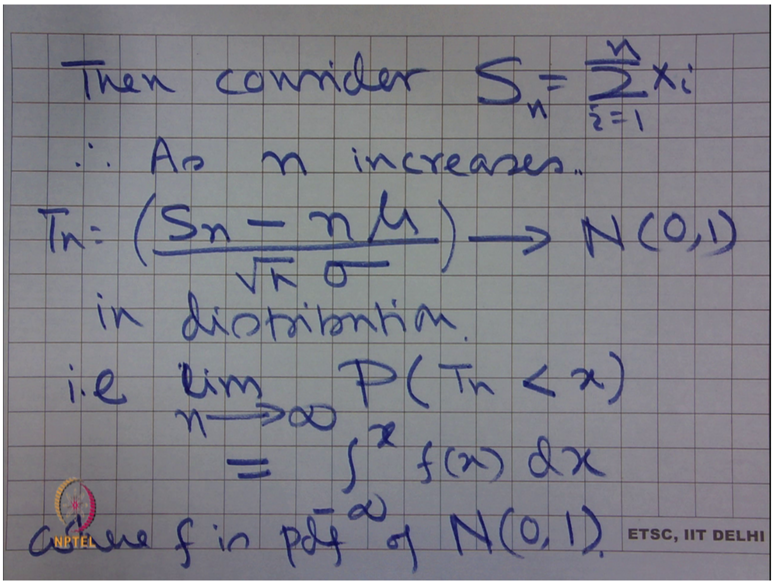
(Refer Slide Time: 30:02)



Again I expect all of you have some basic idea of central limit theorem, I am stating one simple version of this theorem there is not a single central limit theorem there are different versions. But the simple version that we will be using is by Lindeberg and Levy. And it says that let X 1 Xn be n independent random variables, all of which have the same distribution. Let expected value of Xi is equal to mu and variance of Xi is equal to sigma square for all i.

What does it mean? It means that each of the Xi has the same expectation each of the Xi has the same variance.

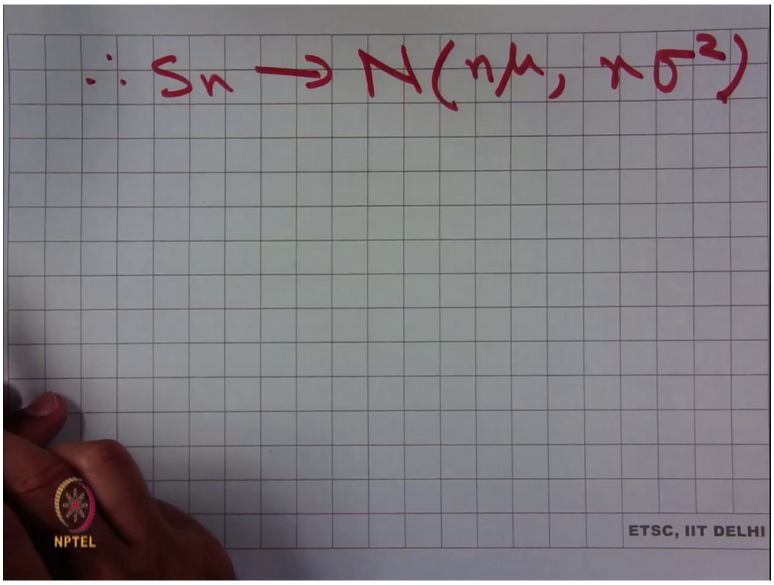
(Refer Slide Time: 31:58)



Then consider s is equal to sigma xi i is equal to 1 to n. Therefore, as let me call it Sn what is Sn? Sn is the sum of n samples. Now if n increases, central limit theorem suggests that Sn minus n mu upon root over n sigma converges to normal 0 1 in distribution.

That is, if we call these to be Tn probability limit n going to infinity probability Tn less than x is equal to minus infinity to x fx dx, where f is pdf of normal 0 1. So, what does it mean? That if I take sample of size n, then as n increases Sn minus n mu upon root n sigma converges to normal 0 1.

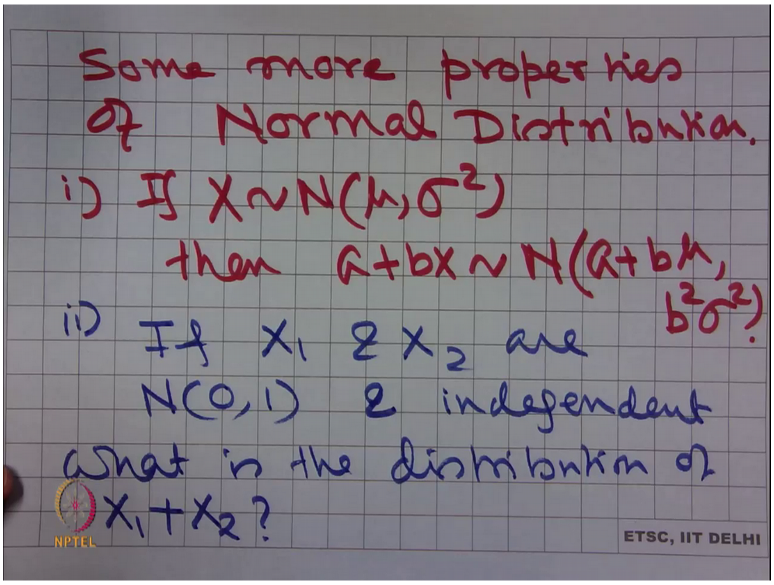
(Refer Slide Time: 34:13)



Or in other words, Sn therefore converges to normal with mean n mu, and variance n sigma square that is the advantage.

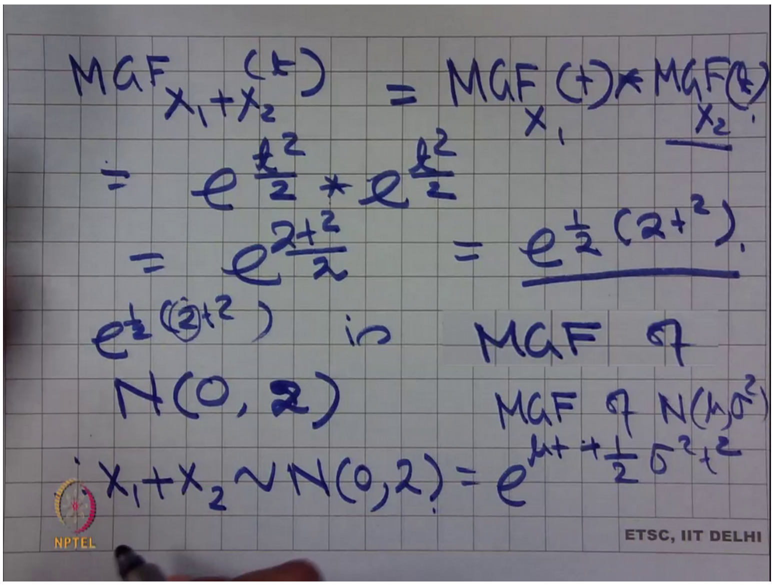
Therefore if we take more and more samples we can approximate the distribution of the sum of the random variable using normal distribution. This has a lot of convenience for us, because then we can approximate many of the statistic using normal, and that is very important because in the last few classes, we have studied normal distribution in depth, and we have also obtained a family of distributions like chi square n Tn F m comma n which all depend upon normal distribution. That is why in large sample theory the normal distribution is of prime importance.

(Refer Slide Time: 35:32)



So, let us examine some more properties of normal distribution. What we have seen? We have seen that if x is normal with mu sigma square, then a plus bx is normal with a plus b mu comma b square sigma square. This is something that we have already proved. Now let us consider a different problem if X 1 and X 2 are normal 0 1 and independent what is the, distribution of X 1 plus X 2? This is very simple.

(Refer Slide Time: 37:09)

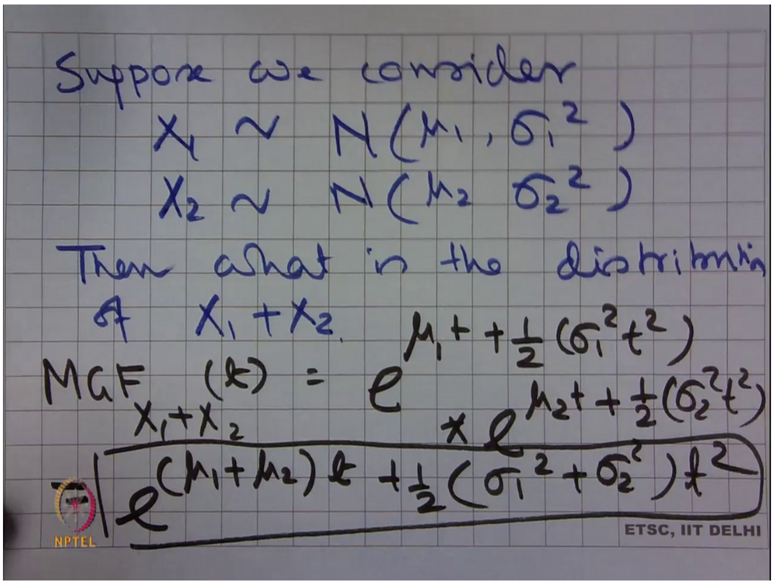


Because, we know that moment generating function of X 1 plus X 2 is equal to moment generating function of X 1 multiplied by moment generating function of X 2.

And we know that moment generating function of x 1 is e to the power t square by 2. Similarly, MGF of x 2 e to the power t square by 2. Therefore, this is equal to e to the power 2 t square upon 2, is equal to e to the power half into 2 t square. Now this is the MGF of a normal population, normal with mean 0, variance is equal to 2, right? Because we know that MGF of normal mu sigma square is equal to e to the power mu t plus half sigma square t square.

Here mu is equal to 0 so, you are looking at only e to the power half sigma square t square and that sigma square is coming out to be 2. Therefore, X 1 plus X 2 distributed as with mean 0 and variance 2. We can even elaborate it more further.

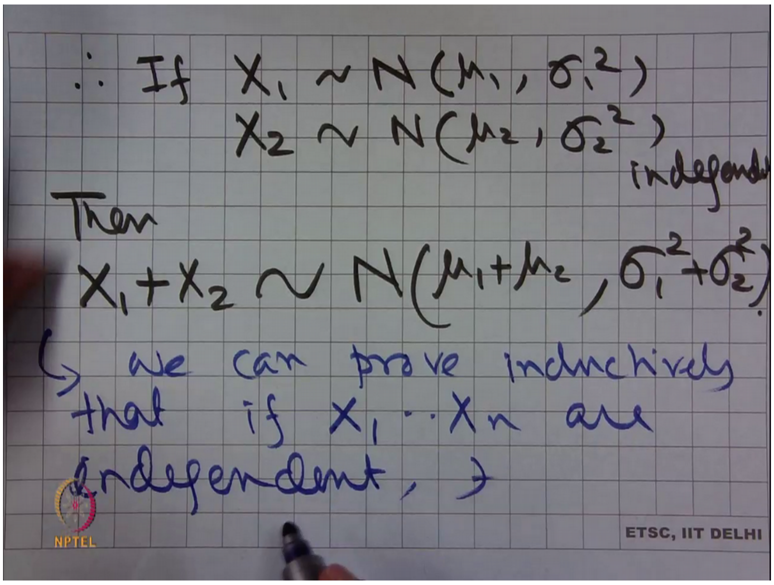
(Refer Slide Time: 39:35)



Suppose we consider X 1 to be normal with mu 1 sigma 1 square X 2 is normal with mu 2 sigma 2 square, then what is the distribution of X 1 plus X 2.

As before MGF of X 1 plus X 2 at t is equal to e to the power mu 1 t plus half sigma 1 square t square multiplied by e to the power mu 2 t plus half sigma 2 square t square. Therefore, this is equal to e to the power mu 1 plus mu 2 t plus half sigma 1 square plus sigma 2 square into t square.

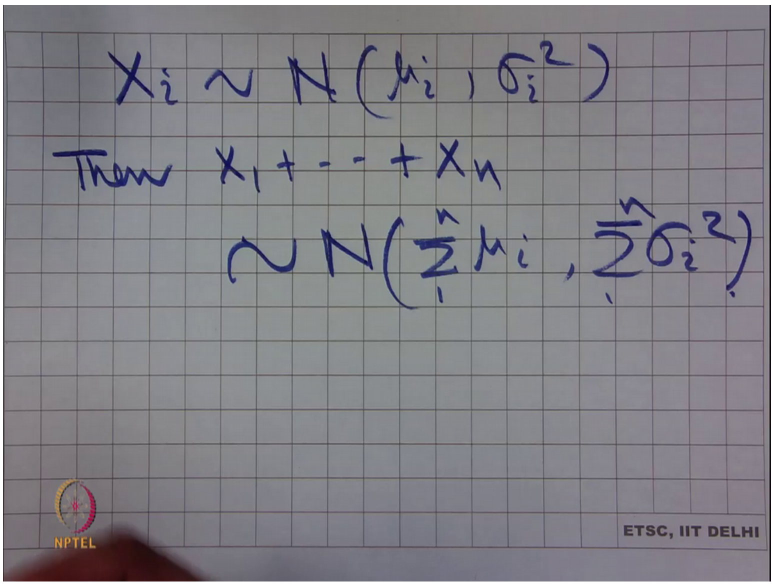
(Refer Slide Time: 41:10)



Therefore, it says that if x 1 is normal with some arbitrary mean and arbitrary variance, and x 2 is normal with some other arbitrary mean and arbitrary variance, then if x 1 and x 2 are independent, then x 1 plus x 2 is distributed as normal with mu 1 plus mu 2 and variance with sigma 1 square plus sigma 2 square.

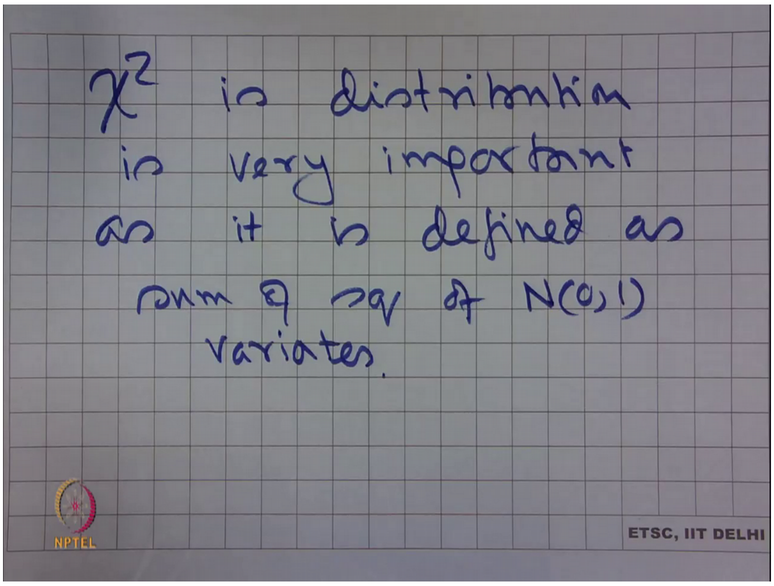
I have proved it for 2, we can use induction that if X 1 X 2 Xn are independent such that Xi is normal with mu i and sigma i square, then x 1 plus x 2 up to xn is distributed as normal with sigma mu i and variance is equal to sigma i square 1 to n..

(Refer Slide Time: 42:18)



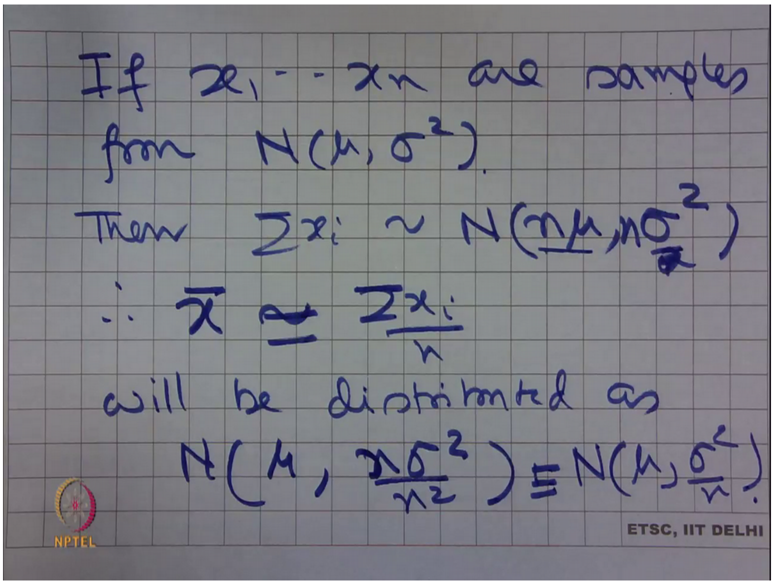
So, this is a very strong result, and that helps us a lot, because when we are taking samples from a population, we know that if the sample size is large we can sort of approximate it with normal, and not only that we can find the distribution of the sum of the samples, and therefore, from here we can calculate the sample mean and we can see that sample mean will be distributed as normal as well.

(Refer Slide Time: 43:39)



Ok students now, we can understand the utility of normal distribution. And we have already seen that chi squared T F are all derived from normal distribution. In particular, the chi squared distribution is very important as it is defined as sum of square of normal 0 1 variants. In the next lecture, I will be talking about estimation of sigma square, which is the population variance. And we will be seen that we can use chi square distribution there for estimating the population variance.

(Refer Slide Time: 44:50)



So, so far what we have studied that if x 1 x 2 xn are samples from normal distribution, then sigma xi is normal with mean is equal to n mu and variance is equal to sigma square by variance is equal to n sigma square. Therefore, x bar which is sigma xi by n will be distributed as normal with since I am dividing it by n, I can divide it by n here. That is going to be mu, and since I am dividing by n the variance is going to be divided by n square.

Therefore, sample mean or expectation of sample mean is going to be mu. Or sample mean is going to be an unbiased estimator for population mean, and as n increases the variance will be decreasing. So, so far we got an estimator for population mean. Our next target is population variance that is sigma square. And therefore, we need to find estimate for sigma square. So, I stop here now, in the next lecture I shall talk about how to find unbiased estimator for sigma square which is the population variance.

Thank you.