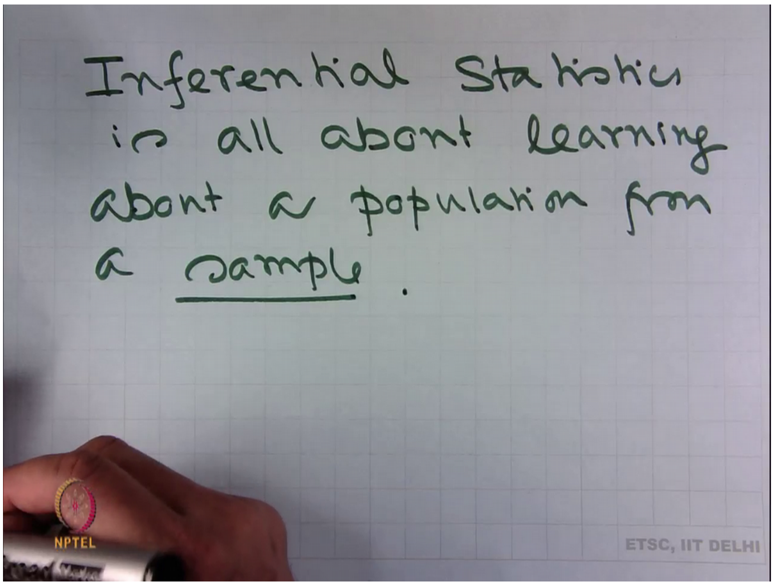
Statistical Inference
Prof. Niladri Chatterjee
Departments of Mathematics
Indian Institute of Technology, Delhi

Lecture – 02
Statistical Inference

Welcome students to the second lecture on the MOOC's series of lectures on Statistical Inference. In the first lecture, I have given a brief introduction to what statistics, or in particular statistical inference means to us, and also I have revised some very popular and commonly used probability distribution. Among them in discreet we had binomial for as so and geometric distribution, for continuous case we have used or we have learnt normal uniform exponential and gamma distribution.

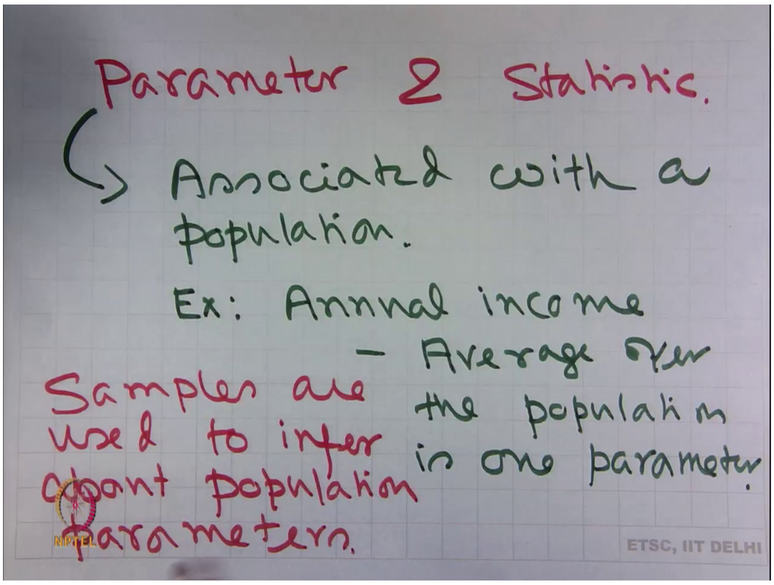
We will be using the main future. In today's term, I will give you some brief idea about sampling and sampling statistics.

(Refer Slide Time: 01:45)



We know that the inferential statistics is all about learning about population from a sample. At this juncture I like to distinguish between 2 terms parameter and statistic; a parameter associated with a population.

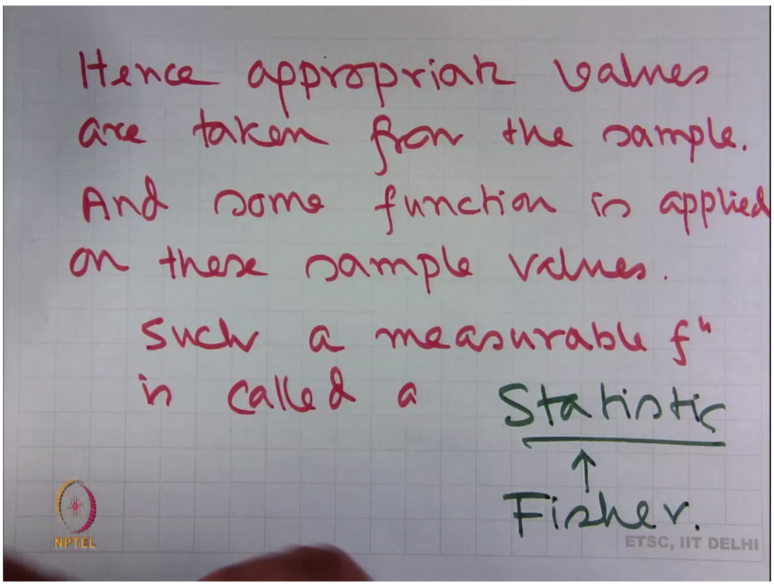
(Refer Slide Time: 02:28)



So, this is property of the population we can say; for example, if you are considering the annual income, then it is average over the population is one parameter. That is population parameter that is fixed, but often this is unknown to us. And the whole purpose of inferential statistic is, to understand the population parameter on the basis of some sample.

So, samples are used to infer about population parameter.

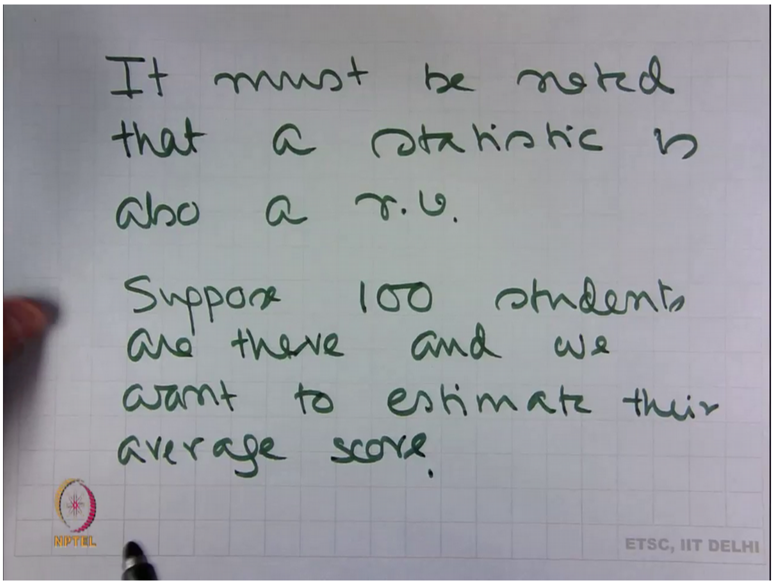
(Refer Slide Time: 04:33)



You know that to do that appropriate values are taken from the sample and some function is applied on those sample values; such a measurable function is called a statistic. The name is given by fisher. So, we have the population, we want to know about some of it is parameter. We have taken sample from the sample observation we have measured certain values, and applied a function on that one in such a way that from that function value, we can get an idea about the population parameter.

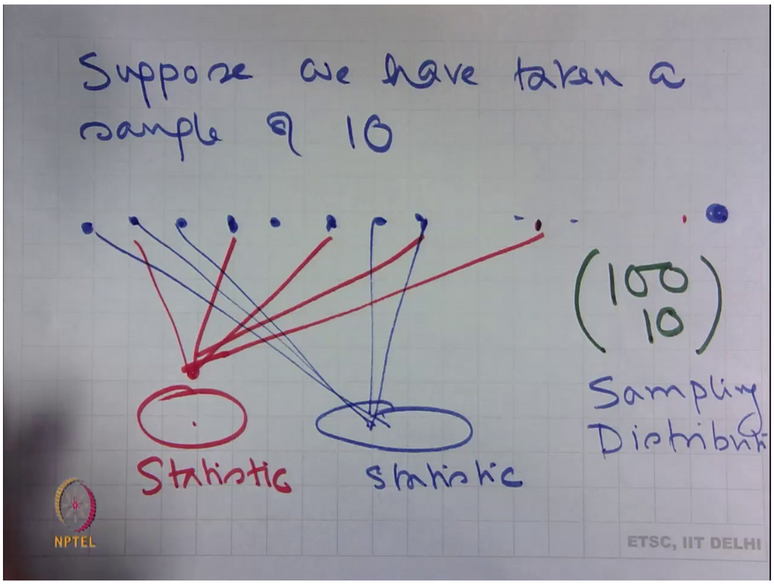
Such a function of the sample values is called statistic, ok.

(Refer Slide Time: 06:27)



It must be noted the statistic is also a random variable; for elastration suppose 100 students are there, and I want to know the average score of the students in statistics, o, what we do?

(Refer Slide Time: 07:40)

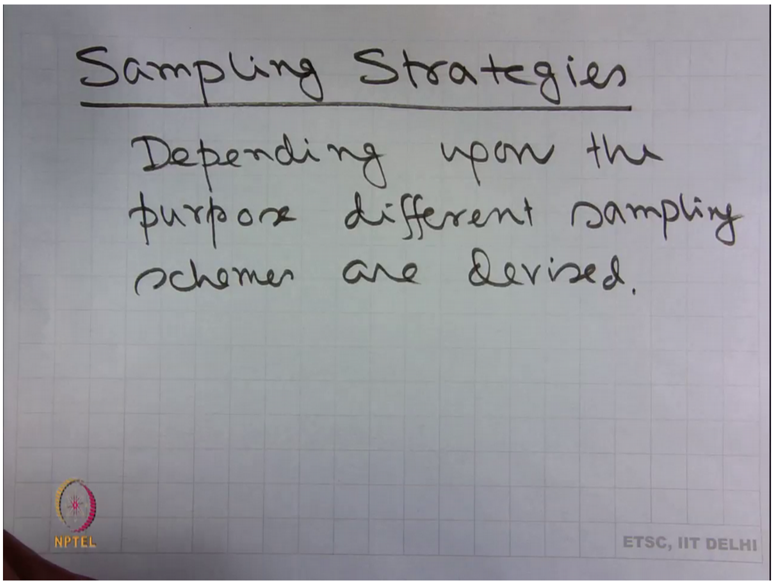


Suppose we have taken a sample of 10. So, suppose this is the entire population. And we have taken tens samples out of it, and suppose like that 10 samples are taken, and suppose I compute the statistic, why this is random variable because if I do the repeat the same experiment, it is not mandatory that I will be selecting the same state of samples.

Now, I will be selecting may be some different samples, and therefore, I compute it is corresponding statistic. It is clear that if there are 100 students, and I take 10 out of it, then I can get 100 see 10 so many difference sample. Each one of them we produce as statistics. This is the statistic that I am going to use to estimate the population parameter; however, these statistic is random as there can be as many as 100 see 10 so many different statistic. Therefore, it make sense to talk about the distribution of that statistic which is called sampling distribution.

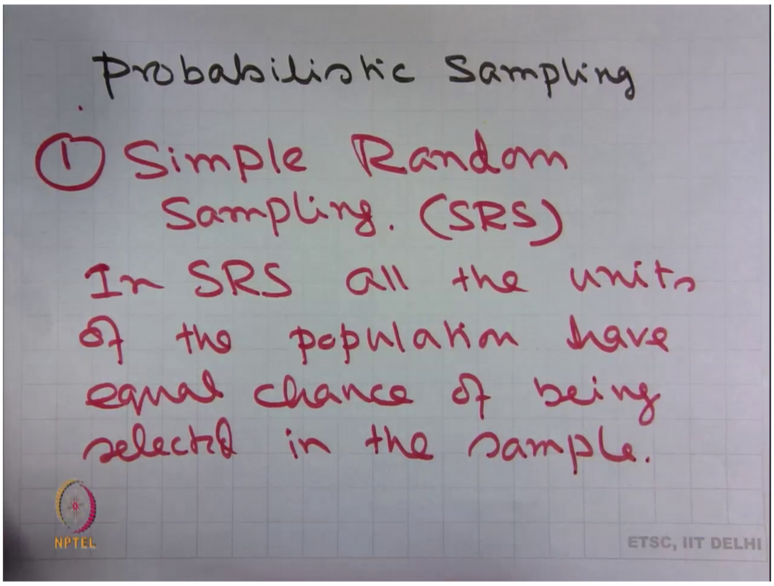
In today's lecture I will be talking about some sampling distributions which will be helpful in estimating population parameters.

(Refer Slide Time: 10:06)



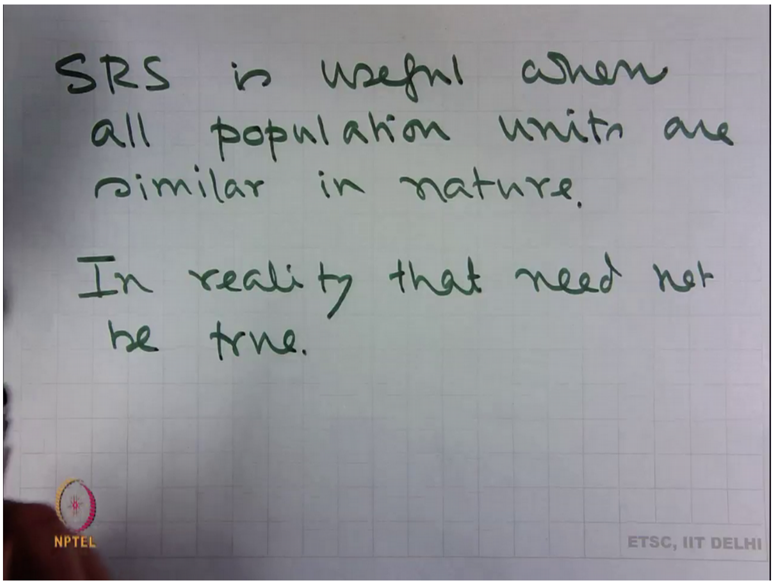
Before that sampling strategies depending upon the purpose different sampling schemes are devised. First of all, it can be probabilistic or non-probabilistic.

(Refer Slide Time: 10:55)



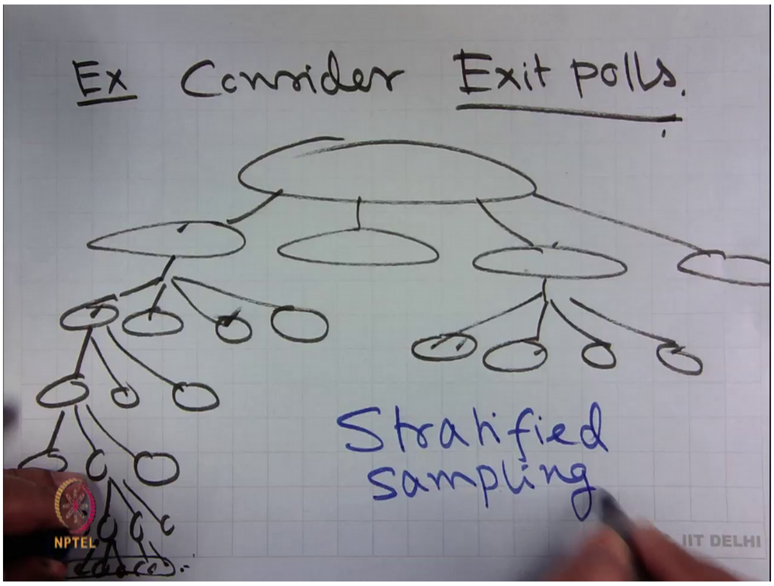
Let us focus on probabilistic sampling. The simplest form is simple random sampling. In simple random sampling, which we often write has SRS all the units of the population have equal chance of being selected in the sample. It is sometimes difficult to design and time consuming to execute, but it is very simple in understanding and it gives equal chance to all the members of the population.

(Refer Slide Time: 12:29)



Simple random sampling is useful when all population members or all population units are similar in nature, but in reality that need not be true.

(Refer Slide Time: 13:19)



For example, consider exit polls. There are so many voters in terms of millions of voters scattered all over the country in different states different district different blocks different polling booths. It is impossible to consider all of them with equal probability.

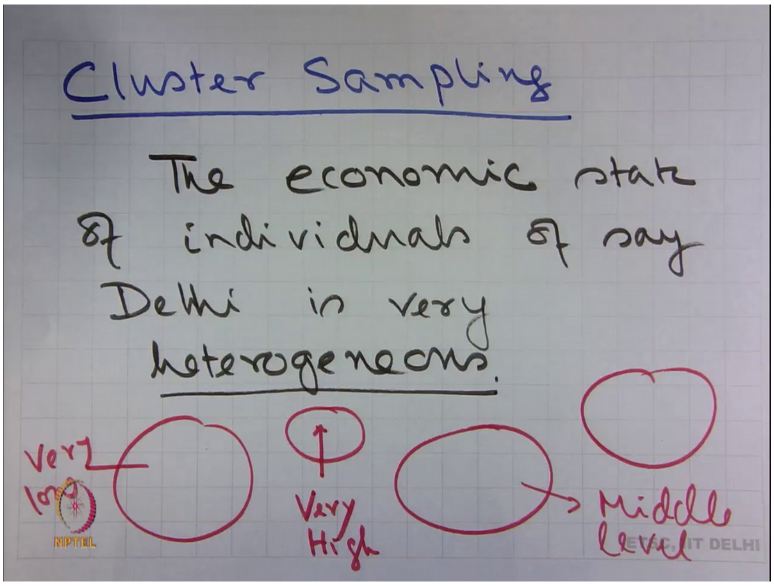
So, what can be done? If this is the country, we can break it into different states. You may choose some of the states randomly. With in those states they are may be many districts.

Again it is not possible to consider all the voters of one district. So, what is done suppose some of the district are chosen randomly. Then with in a district, there may be different sub division. Within the sub division there may be different blocks, within each block there may be several booths, within each booth there may be several voters.

We need to take the opinion of some of these voters to predict what is going to happen therefore, giving equal probability to all the members or to all the voting population of the countries impossible. So, what is done the population is divided into different stratum, right. At each level it is subdivided into lower order stratum like that until we reach to the very basic units, and from home the actual data has to be collected.

Such way of sampling is called stratified sampling.

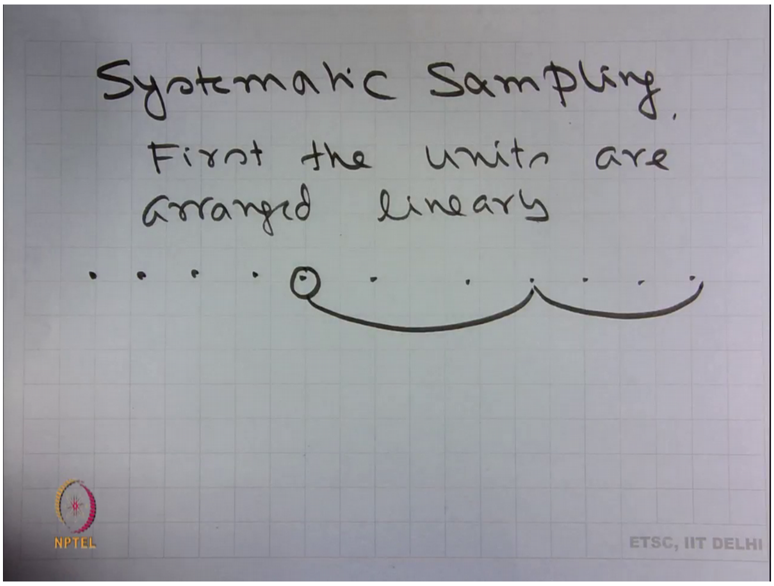
(Refer Slide Time: 16:02)



Another way of sampling is called cluster sampling. For example, your making an economic survey on the population of Delhi. Delhi has huge population, but the problem here is all the elements of that population and not somewhat homogeneous in nature. In fact, they are highly heterogeneous, therefore, application of simple random sampling will lead to very erroneous result or a result with a huge variance because the outcome of the survey will be very different based upon the samples that you have chosen. Therefore, it is more meaningful to divide the entire population into different clusters such that the clusters are homogeneous in nature.

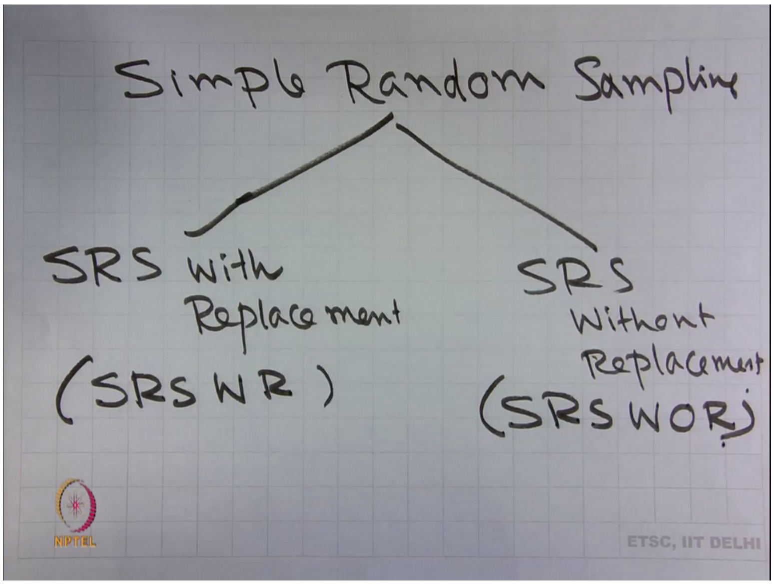
Say for example, this may be a cluster of people with very high income. This may be a cluster of very low income; this may be a cluster of say middle level income. Therefore, choosing the sample over the entire population randomly, first we divide it into some clusters which are more homogeneous. And then depending upon the size of cluster we can decide how many samples will be taken from each of the clusters.

(Refer Slide Time: 18:44)



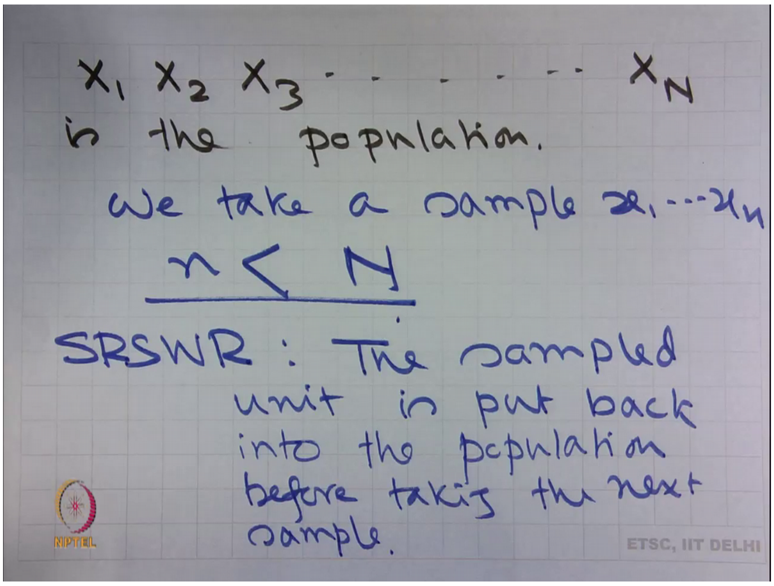
By sampling in such fashion is called cluster sampling similarly, there is something called systematic sampling; it is a mixture of strategies. So, first the units are arranged linearly say by sorting them according to their names. Then instead of applying simple random sampling over the entire population one of the, this starting point is randomly chosen over the entire population. After that the members are chosen according to some fixed interval. This is also very useful technique for sampling as it is much less time consuming and costly in comparison with a simple random sampling.

(Refer Slide Time: 20:00)



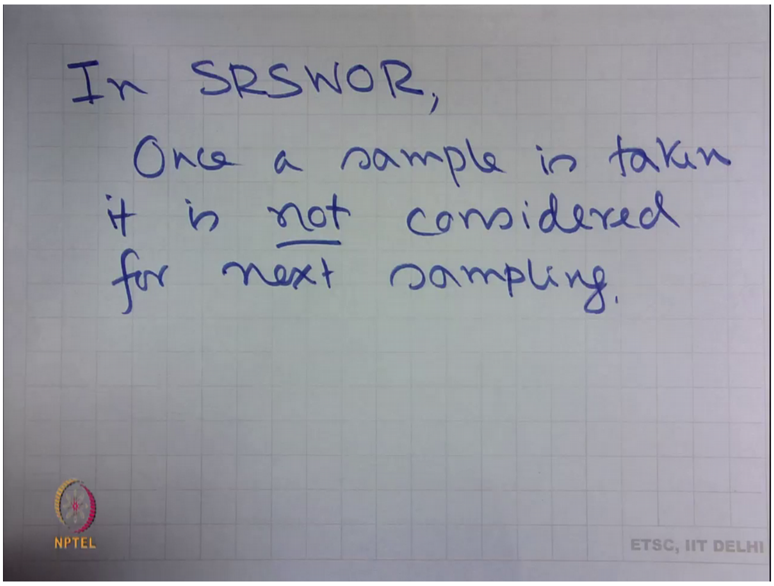
So, in today's class we will be focusing on simple random sampling, which again can be subdivided into 2 ways simple random sampling, which replacement and simple random sampling without replacement. In short we call it SRSWR and this is called SRSW or what is the difference? The difference is as follows.

(Refer Slide Time: 20:58)



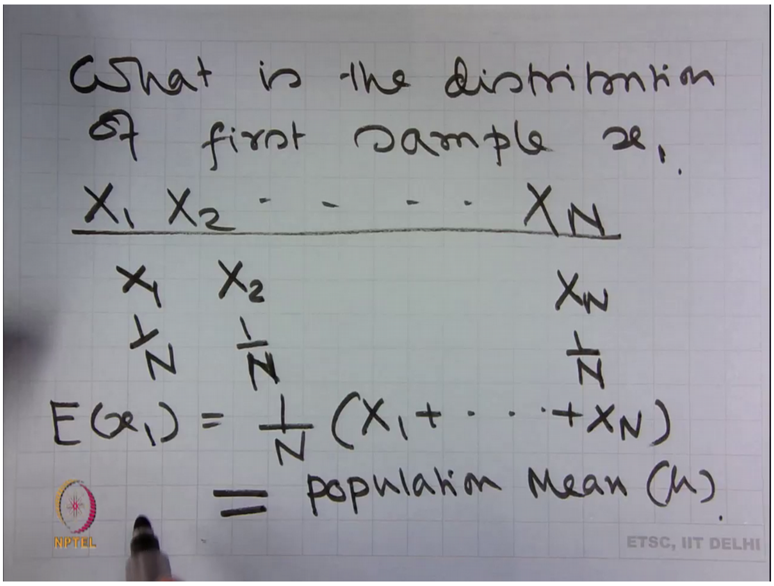
Suppose, X 1 X 2 X 3 XN is the population and we take a sample X 1 X 2 X N where N is much less than N. If after taking one sample, we put it back again into the population, before choosing the next sample this is called simple random sampling with replacement, the sampled unit is put back into the population before taking the next sample.

(Refer Slide Time: 22:37)



On the other hand,, in SRS WOR once sample is taken it is not considered for next sampling. Therefore, in SRSWOR one individual unit can be represented only once at the most in the sample. On the other hand, in SRSWR theoretically in all the N samples there is a possibility that you can choose only the same element small N number of times.

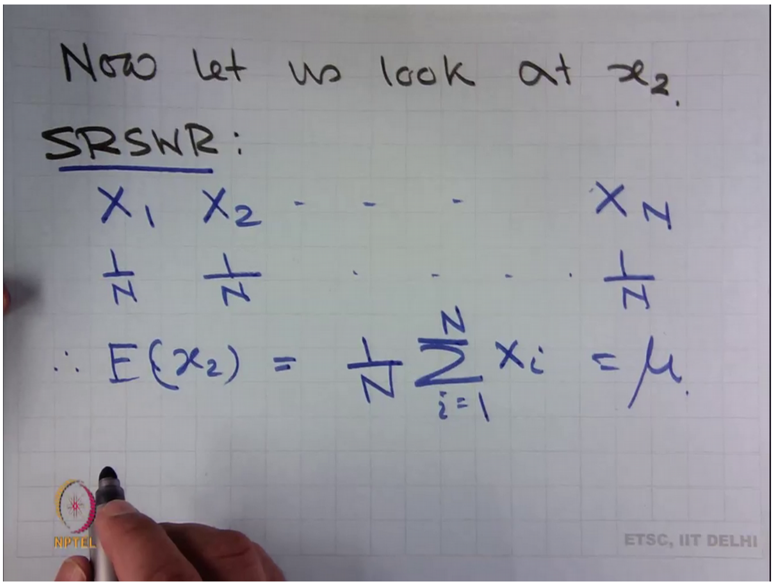
(Refer Slide Time: 23:43)



Therefore, let us look at what is the distribution of the first sample. So, you have X 1 X 2 X N as the population. X 1 is the very first sample taken out of it, and each one of them has equal probability. So, it can be X 1 with probability 1 by N. It can be X 2 with probability 1 by N and it can be Xn with probability 1 by N. Therefore, what is the expected value of x 1 which is equal to the population mean; let us denote it by mu.

Therefore, the first sample has expected value equal to population mean. That is irrespective of whether it is with replacement or without replacement. Now let us look at X 2 under a SRSWR.

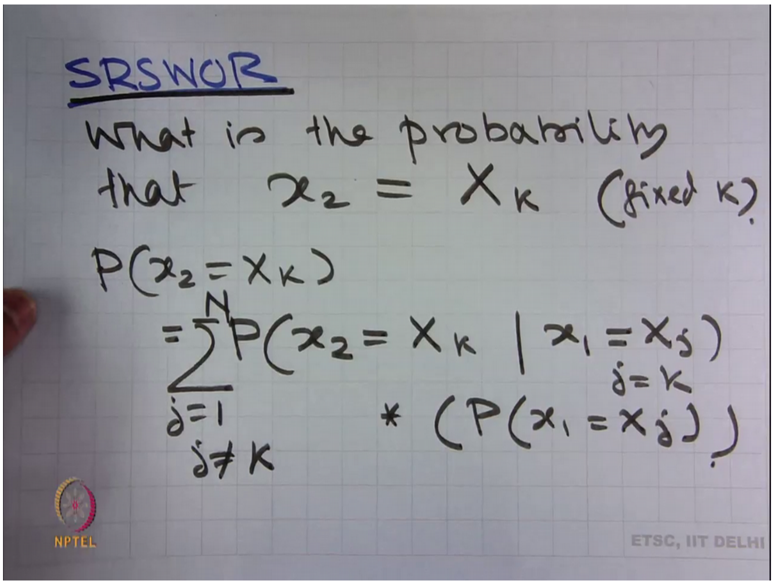
(Refer Slide Time: 25:24)



So, if you are doing simple random sampling with replacement, then whichever element has been taken in the first sample or as the first unit of the sample, it is put back into the population again.

So, for SRSWR again I have all X 1 X 2 Xn, and they are all equally likely. So, all of them will have the same probability 1 by N 1 by N up to 1 by N. And therefore, expected value of X 2 is equal to again 1 by N into sigma x i i is equal to 1 to N is equal to mu.

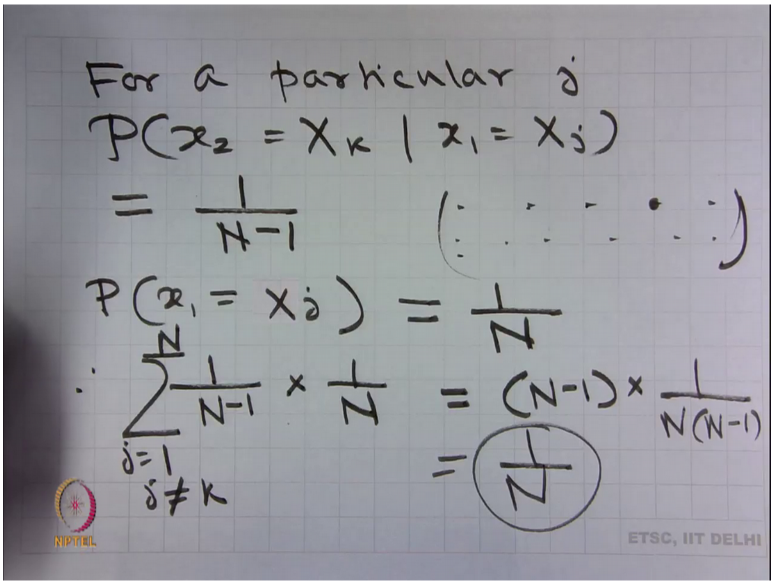
(Refer Slide Time: 26:46)



But now things change with respect to SRSWOR. So, what will happen in this case? What is let us look at what is the probability that X 2 is equal to Xk fixed k, right. Say what is the probability that the second unit taken as a sample is actually the k th member of the population.

So, probability X 2 is equal to Xk is same as probability X 2 is equal to Xk given that the first sample is equal to Xj j not equal to k multiplied by probability X 1 is equal to Xj, right. And this we need to sum over, all j j is equal to 1 to N j not equal to k. Because it is without replacement, there for second sample can be Xk if the first sample is something different from Xk and given that that then the second sample is actually x k and this has to be multiplied by what is the probability of first sample is Xj.

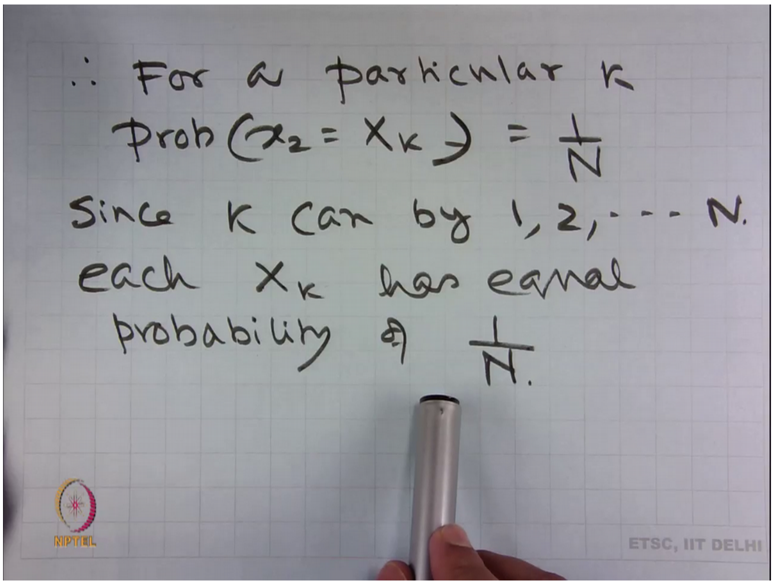
(Refer Slide Time: 28:55)



For a particular j probability X 2 is equal to Xk given X 1 is equal to Xj is equal to 1 upon N minus 1 because there are N many capital N many elements in the population. The jth one has been taken now there are N minus 1 element, out of this we are choosing one element with simple random sampling. And therefore, all the N minus 1 of them have the same probability therefore, it is 1 upon N minus 1.

Now what is the probability x 1 is equal to Xj probability x 1 is equal to Xj is 1 upon N, right? Therefore, what we are getting 1 upon N minus 1 into 1 upon N, and this we are summing over j is equal to 1 to N j not equal to k. Therefore, actually we are summing it for N minus 1 times therefore, it is N minus 1 into 1 upon N into N minus one is equal to 1 upon N.

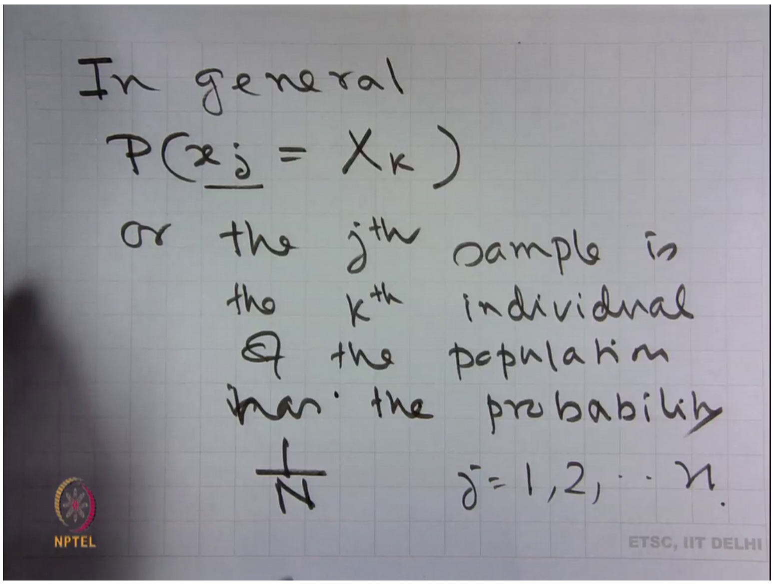
(Refer Slide Time: 30:54)



If you remember we have selected a particular Xk, probability that X 2 is equal to Xk is equal to 1 upon N.

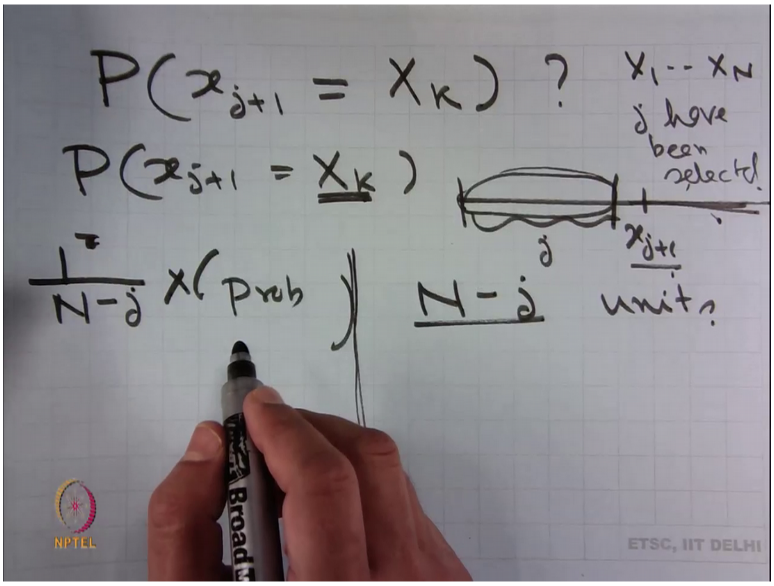
Since k can be anything between 1 to capital N, each Xk has equal probability of 1 upon N. Therefore, we see that whether we have chosen it with replacement or without replacement it does not matter, thus second unit that is being chosen also has equal probability of getting one of the capital N many units as the second selection.

(Refer Slide Time: 32:09)



In general, probability Xj is equal to Xk or the jth sample is the kth individual of the population is has the probability 1 by N; that is, each one of the original population members will have equal chance of getting selected in the sample at the jth instance for j is equal to 1 2 up to N. Because we are choosing N samples from the population. This can be shown in the following way, what is the probability that the j plus 1 it is selection is equal to the k th member of the population?

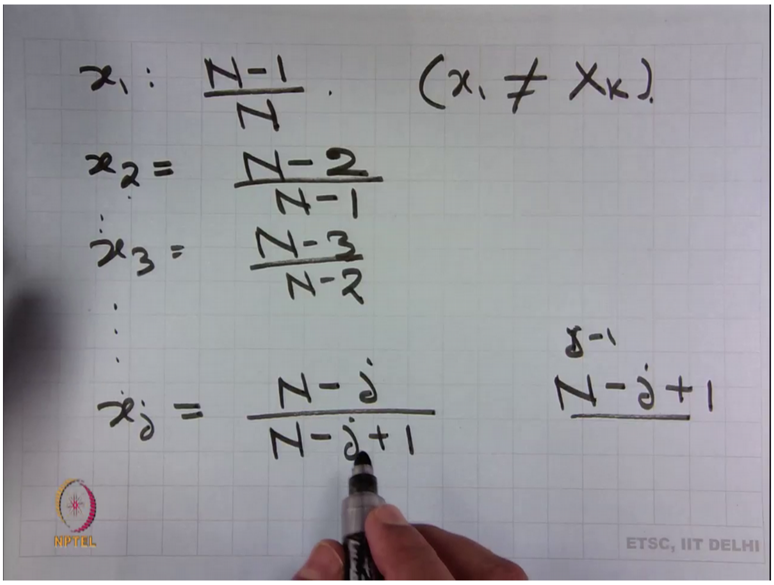
(Refer Slide Time: 33:24)



So, this we can calculate in the following way probability x j plus 1 is equal to Xk is equal to; consider the following first j samples have been chosen, and I am looking at the j plus 1 at sample. In the first j samples out of X 1 X 2 X N, j have been selected, since j of them have been selected, how many are left ? You are left with N minus j units, right?

So, in the j plus 1 th instance, I am choosing a particular element x k. First of all, Xk has to be here, it has not been chosen yet in the already selected j many samples. And now I am going to choose Xk as the j plus 1th unit in the sample. So, that is of equal probability therefore, probability of selecting Xk as that one is 1 upon N minus j, multiplied by the probability that it has not been chosen at all right so, we do it in the following way.

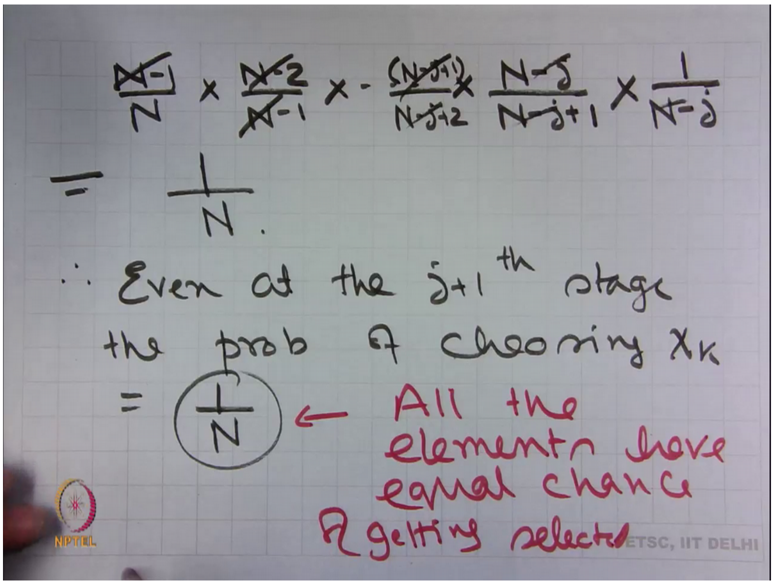
(Refer Slide Time: 35:25)



So, in the first instance, we can choose something other than the Xk that has the probability N minus 1 upon N. Because out of N one of them is Xk and so, that we cannot select at the first instance I can select any of the remaining N minus 1 of them. In the second instance therefore, what I will choose? I will choose one from among the remaining N minus 1, but not k and that probabilities N minus 2 upon N minus 1.

Similarly, x 3 will have the probability N minus 3 upon N minus 2. And when I am selecting the xj I have already selected j minus 1. So, how many are left? The number of elements left is N minus j plus 1, out of that I will be choosing one of them except Xk. So, I have the selections possibility of N minus j of them out of N minus j plus 1. Therefore, what we are getting is that we have not chosen Xk till the jth sample that probability is N minus 1 upon N multiplied by N minus 2 upon N minus 1 multiplied by up to N minus j upon N minus j plus 1.

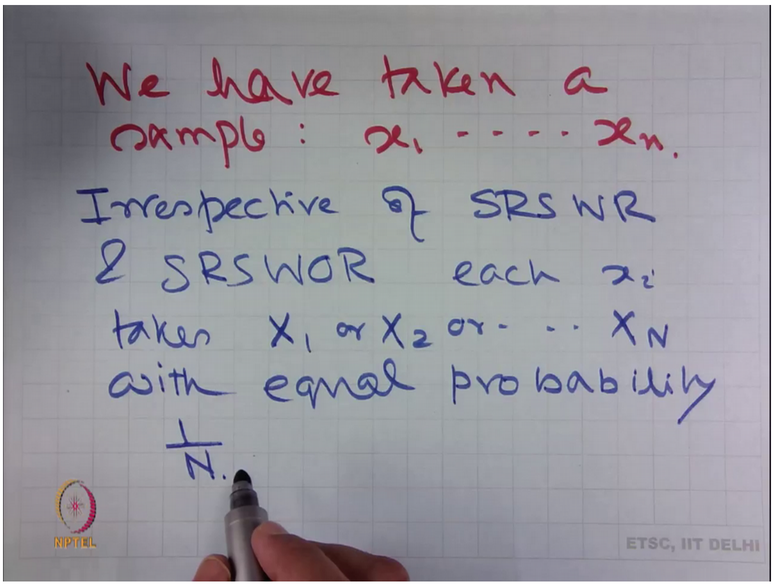
(Refer Slide Time: 37:41)



So, this one has to be N minus j plus 1 upon N minus j plus 2, multiplied by now that out of that N minus j element I am choosing Xk that probabilities 1 upon N minus j. So, what we are getting? These terms cancel each other, and what we are left with is 1 upon N. Therefore, even at the j plus 1 th stage, the probability of choosing xk is equal to 1 upon N.

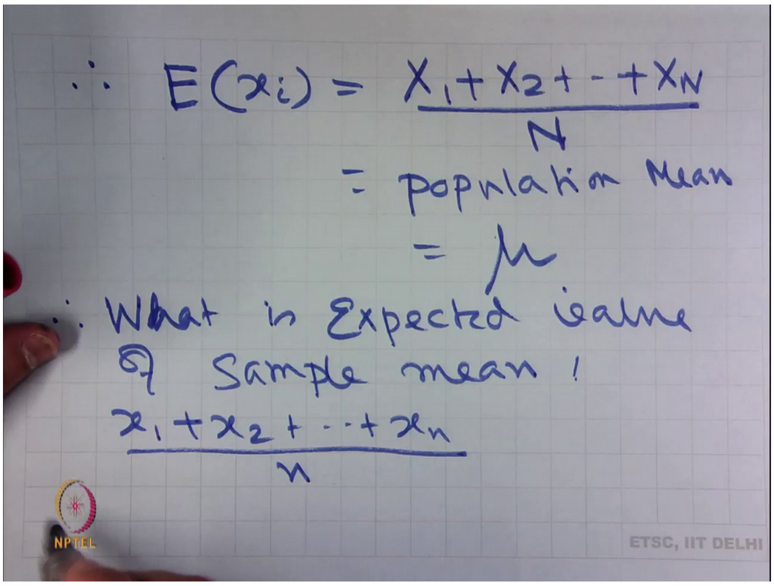
Therefore, it suggests that all the elements have equal chance of getting selected. In all the 1 2 3 or N th selection, what is the effect of it?

(Refer Slide Time: 39:16)



The effect of is the following so, we have taken a sample irrespective of SRS or and SRSWOR, each xi takes X 1 or X 2 or X N with equal probability 1 upon N.

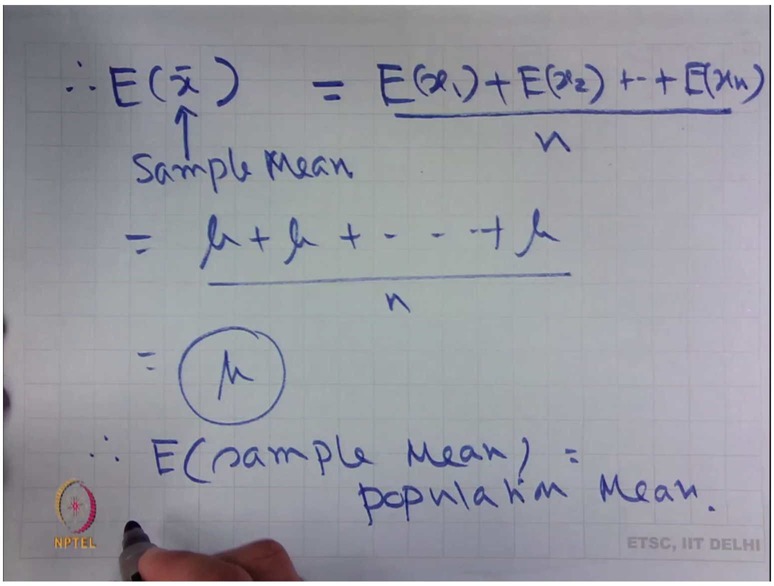
(Refer Slide Time: 40:27)



Therefore, expected value of x i is equal to X 1 plus X 2 plus X N upon N; is equal to population mean is equal to mu. And this is irrespective of SRSWR or SRSWOR.

Therefore, what is the expected value of sample mean; which is X 1 plus X 2 plus X N upon N. We know that from the theory of expectation that expectation of sum is equal to sum of expectation.

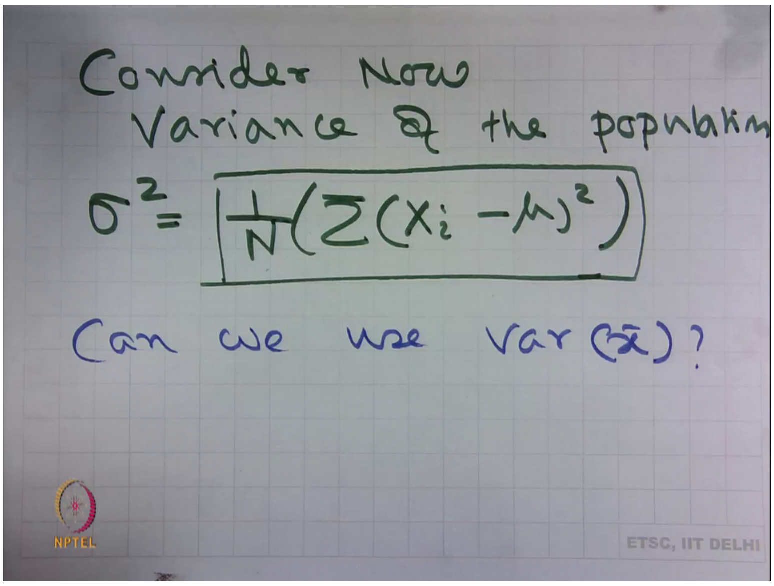
(Refer Slide Time: 41:48)



Therefore, expected value of x bar which is the sample mean, as I already said that this is a random variable is equal to expected value of X 1 plus expected value of X 2 plus expected value of Xn divided by N is equal to mu plus mu plus mu.

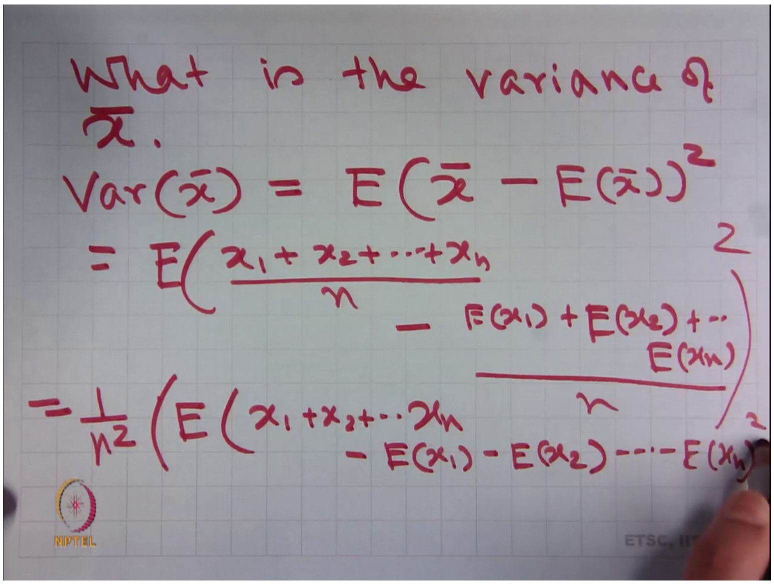
Therefore, expected value of sample mean is equal to population mean. Thus we can see that sample mean can be used to estimate the population mean. Let us now consider variance of the population.

(Refer Slide Time: 43:00)



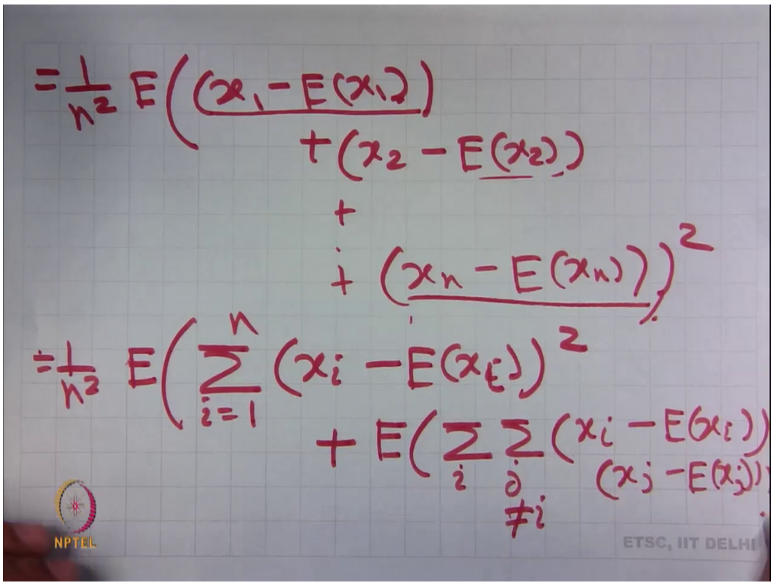
The variance of the population is sigma square is equal to 1 by N sigma X i minus mu whole square. Suppose you want to estimate the population variance; can we use variance of x bar for that purpose? Let us see. So, question is what is the variance of x bar, as x bar is a random variable itself this makes sense.

(Refer Slide Time: 44:11)



Now, variance of x bar is equal to expected value of x bar minus expected value of x bar whole square is equal to expected value of x 1 plus x 2 plus x n by n minus expected value of x 1 plus expected value of x 2. Now 1 by n we can take out as commons. So, it is 1 by n square into expected value of x 1 plus x 2 plus Xn minus expected value of x 1 minus expected value of x 2 minus expected value of xn whole square.

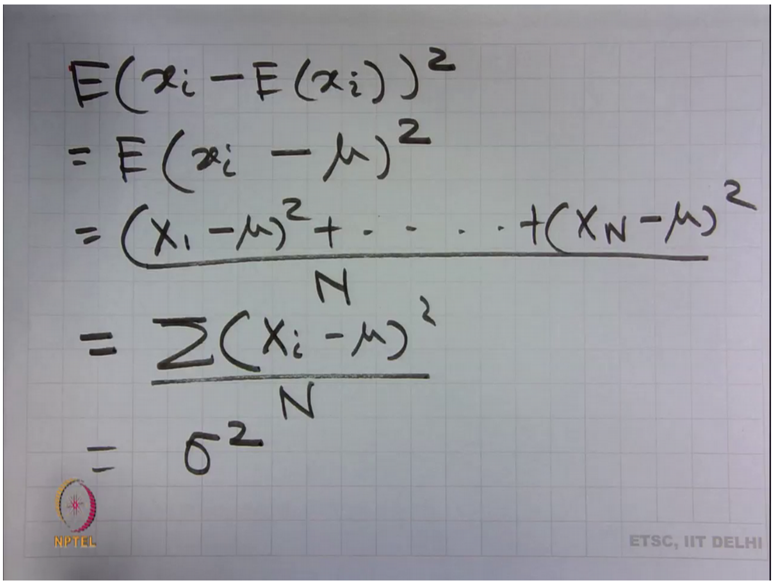
(Refer Slide Time: 45:48)



We rewrite it as expected value of X 1 minus expected e of x 1, plus x 2 minus e of x 2 plus xn minus e of x n whole square. So, it is n terms and their square it is expectation.

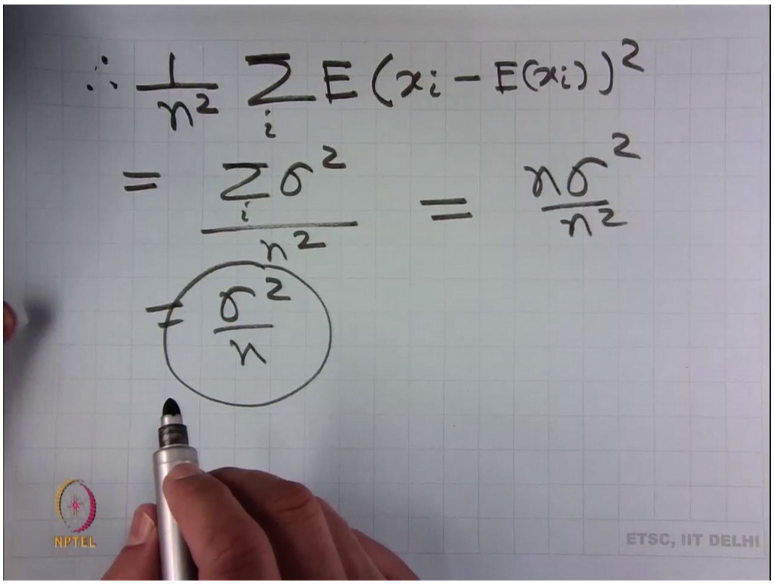
This N terms square can be written as so, e square plus these square plus these square up to Xn minus Exn whole square. So, this we can write it as x i minus expected value of x i whole square i is equal to 1 to n. Plus, expected value of sigma over i sigma over j, j not equal to i x i minus expected value of x i multiplied by xj minus expected value of xj.

(Refer Slide Time: 47:32)



Now, expected value of x i minus expected value of x i whole square is equal to expected value of x i minus mu whole square; is equal to X 1 minus mu whole square plus Xn minus mu whole square divided by N, is equal to sigma X i minus mu whole square by N this is nothing but the population variance which is sigma square.

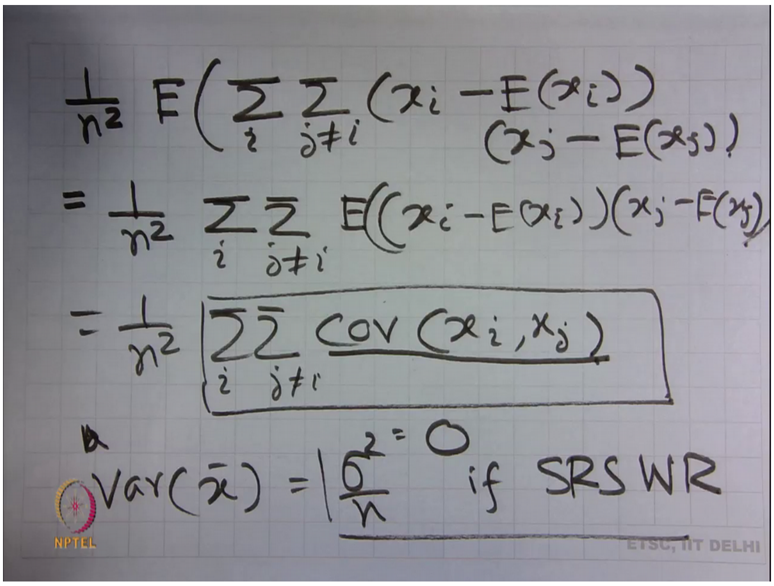
(Refer Slide Time: 48:21)



Therefore 1 by N square into sigma over i expected value of x i minus expected value of x i whole square is equal to summation over i sigma square divided by n squared; is equal to n sigma square upon n square is equal to sigma square by n.

So, that is the first part of the expansion.

(Refer Slide Time: 49:13)



The second part is 1 by n square into expectation of sigma over i sigma over j not equal to i x i minus expected value of x i into xj minus expected value of xj is equal to 1 upon n square sigma over i sigma over j not equal to i, expected value of x I minus expected value of x i; is equal to 1 upon n square sigma. So, this is giving us 1 upon N square into summation over xi xj, covariance of xi xj for all possible pairs i n j.

Now, if it is a simple random sampling with replacement, then we can see that x i and xj they behave in the same way. In fact, they are independent of each other. Therefore, if we are considering simple random sampling with replacement, then each of the covariance term will become 0. Therefore, we are left with variance of x bar is equal to sigma square by n if simple random sampling with replacement.

But when we are using simple random sampling without replacement, this covariance will not be 0. Because what is we are getting in the jth term will depend upon what will what we have got at the ith time. And therefore, this will have some contribution therefore, the variance of x bar will not be sigma square by n rather it has to be adjusted by a quantity. In the next class I will start with this point and then I will go further with sampling distribution.

Thank you so much.