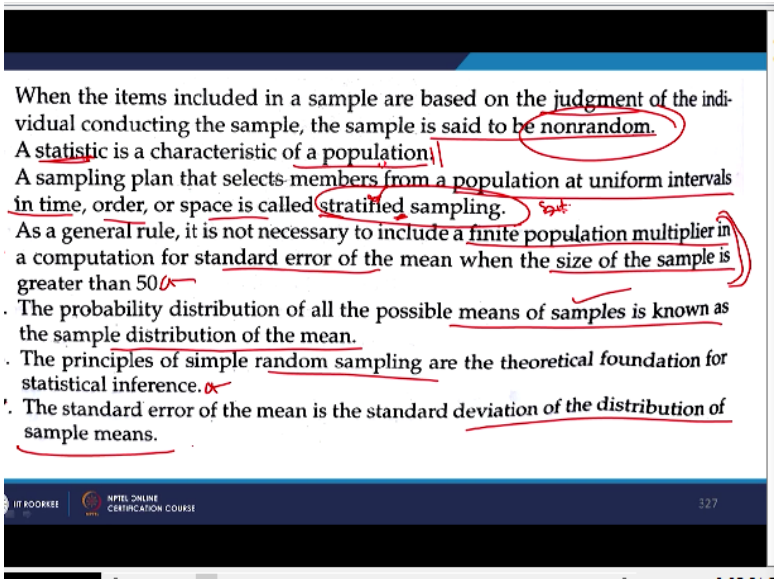
Business Statistics
Prof. M. K. Barua
Department of Management Studies
Indian Institute of Technology-Roorkee

Lecture-25
Sampling Distribution–III

Good afternoon friends, I welcome you all in this session as you are aware in previous session we did work out couple of examples on sampling and sampling distribution. And we have seen how we calculated probabilities within even range of x values. And in one of the examples we have also found out the value of Z where the probabilities were given. In today' session we will work out couple of exercises, now this particular exercise will help you in understanding the concepts of sampling and sampling distribution. So these are couple of statements and you have to tell me whether these statements are true or false.

(Refer Slide Time: 01:22)



So, let us begin with first statement, when the items included in a sample or based on the judgment of the individual conducting the sample. This sample is said to be non-random, now there is an individual person means choosing samples from a population. So, what it would be, it would be random or non-random it won't be random. So, this statement is true it is said to be non-random right.

A statistic is characteristics of a population is it so? We have seen characteristics of population as well as characteristics of sample. So what is this, statistic is what is about population or sample. So, this statement is false because statistic is a characteristics of sample not population what is the characteristic of population is called parameter right not the statistic. Let us look at third one a sampling plan that select members from a population at uniform intervals either in time order or a specific or a space is called stratified sampling is it so.

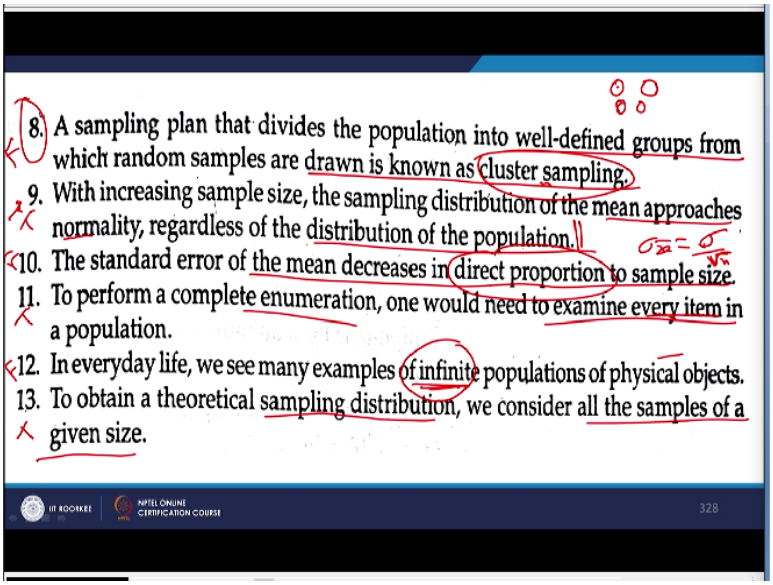
First of all you should know that whether this is random or random sampling or non-random sampling. In other words whether this one is a probabilistic sampling or non-probabilistic sampling, so we know that stratified random sampling is probabilistic one. But in probabilistic sampling we do not select items from population on the basis of these things right. So, this statement is actually false one right.

By the way what that sampling is called it is not stratified that is known as systematic sampling right. Of course systematic sampling is again a is a probabilistic sampling. Let us look at 4th one is a general rule it is not necessary to include a finite population multiplier, in computation of standard error of the mean when size of the sample is greater than 50 is it correct. In fact I did not cover much about population finite multiplier.

But it is used when your sample fraction is more than 0.05, sample fraction is nothing but the ratio of sample size to population ok. We can skip this but this statement is false one. Let us look at the next one the probability distribution of all possible means of sample is known as sampling, sample distribution of the mean is correct or incorrect. The probability distribution of all the possible means of samples is known as sample distribution of the mean. yes, this is true right.

Let us look at the next one the principles of simple random sampling or the theoretical foundation of statistical inference is it true yes this one is true. Because on the basis of sample itself we infer about population. Let us look at 7th one, the standard error of the mean is the standard deviation of the distribution of sample mean is it correct. The standard error of mean is the standard deviation of the distribution of sample means yes this is true right.

(Refer Slide Time: 05:32)



Let us look at next slide statement number 8, a sampling plan that divides the population into well-defined groups from which random samples are drawn is known as cluster sample or cluster sampling is it true or false. We have seen that this is a type of probabilistic sampling right and in cluster sampling and in its stratified sampling in both the samplings we first divide we first make different sub-groups or strata right.

And then from each of depending on which method we have to apply we select samples from each of this groups are strata right. So, this is not cluster sampling, so this statement is false, a sampling plan that divides the population to well-defined groups from which random samples are drawn is known as cluster sampling, no this is known as in fact stratified sampling right.

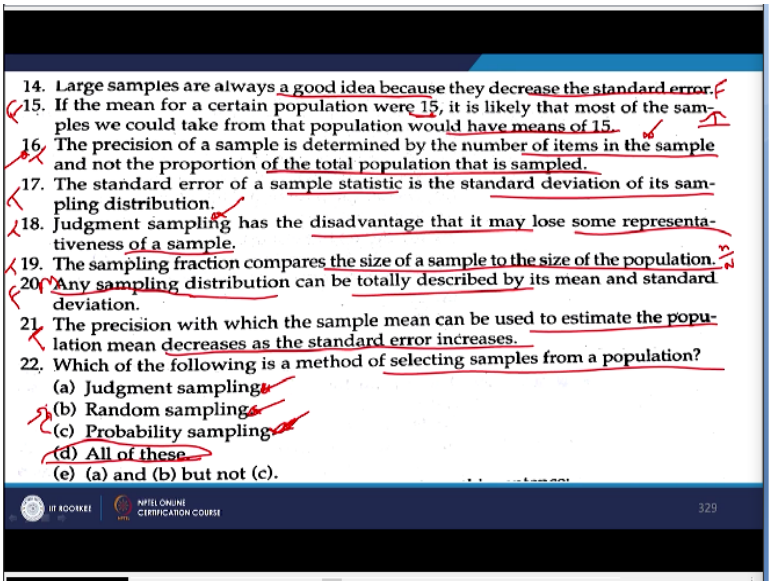
Let us look at 9th one with increase in sample size the sampling distribution of the mean approach is normality regardless of the distribution of the population, this is true right. And this is what you have seen and this is one of the properties of sampling central limit theorem that whatever is the shape of the population distribution the shape of the sampling distribution will be approaching normal when we increase sample size.

The standard error of the mean decreases in direct proportion to the sample size is it correct, the standard error of mean right this what this divided by under root, so it is not directly proportional to the sample size it is inversely and this under root of n ok. So, this statement is false right, to

perform a complete enumeration one would need to examine every item in a population is it correct yes this one is correct.

This one is true right, if you do not evaluate every item then it becomes if it takes few of them then it become sampling. In everyday life we see many examples of infinite populations of physical objects true or false, do we see infinite populations of physical object can we not count them yes we can count them, so we see finite, so this is false. Let us look at the next one to obtain a theoretical sampling distribution we consider all the samples of given size is this correct, to obtain a theoretical sampling distribution we consider all the samples of a given size this is true ok.

(Refer Slide Time: 09:01)



Let us look at next one large samples are always a good idea because they decrease the standard error would it be right to say, so no waste is a false statement right not a correct one ok. This is not the reason why we take large sample because it decreases the standard error, we take large sample of course large sample size minimizes standard error. But that is not the reason for taking large samples, so this one is false.

Let us look at the next one, if the mean for certain population were 15 it is likely that most of the samples we could take from the population would have mean of 15. So, let say this is a population with mean 15 right, if you take few samples then would you be getting sample means

as 15 no it is not correct is false. You will have for one or two samples it is possible but is not possible for all of them right.

This mean would be varying look at 16th one, the precision of a sample is determined by the number of items in the sample and not the proportion of the total population that is sampled. The precision of the sample is determined by the number of items in the sample and not the proportion of the total population that is sample, yes this one is correct. So, 16 number is correct one. Let us look at 17th the standard error of sample is statistic is this standard deviation of its sampling distribution is this correct.

The standard error of sample is statistic is the standard deviation of its sampling distribution, so this is correct, this one is true statement. Let us look at the 18th one, judgment sampling has the disadvantage that it may lose some representativeness of sample, is this correct do you think so. Judgment sampling has the disadvantage that it may lose some representativeness of sample, yes that is why this is not a good method of sampling.

This is a non-problastic sampling method judgmental sampling right it is look at the 19th one. The sampling fraction comprise compares the size of the sample to the size of the population the sampling fraction compares size of the sample small and to the size of the population capital and yes this correct this is sampling fraction right. Any sampling distribution can be totally described by its mean and standard deviation is it possible.

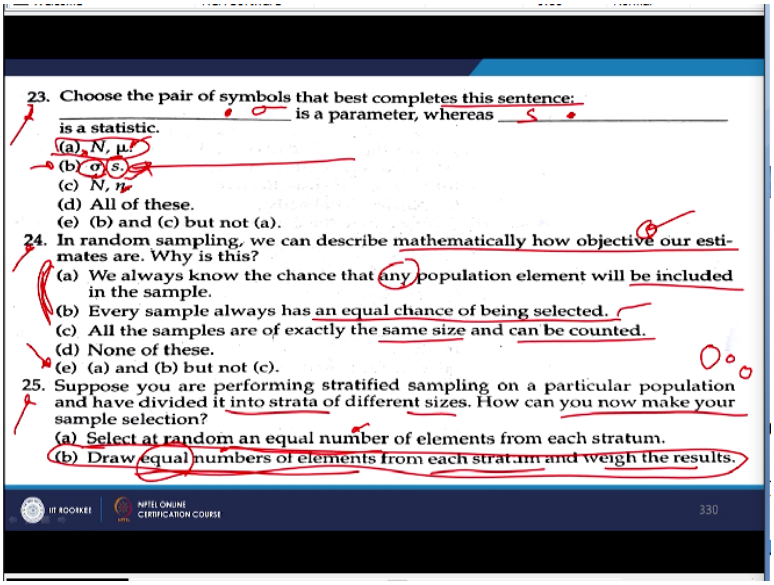
Any sampling distribution though describe couple of distributions by its mean and standard deviation but is not necessary that any sampling distribution can be totally described by mean and standard deviation, so this is false. Though had it been like this many then this statement would have become true, 21st the precision with which the sample mean can be used to estimate the population mean decreases as the standard error increases.

The precision with which the sample mean can be used to estimate the population mean decreases it means precision decreases as the standard error increases, yes this is true. Of course when there is more error how precision would increase precision would 100% decrease right. It

is it will always decrease, which of the following is a method of selecting samples from a population, let us look at these methods.

So, do we select sample from population using judgmental sampling yes we do random sampling yes we do this probability sampling yes, so will say that in my in fact random sampling is nothing but a type of probability sampling. And of course judgmental sampling is non-problastic sampling, so all of this abcd abc a and b but not c no this is there in the answer, so right answer is d all of these right.

(Refer Slide Time: 14:48)



Let us look at this one, choose the pair of symbols that best complete the sentence and sentence is dash is a parameter whereas dash is a statistic. Parameter is always about population and statistics is about sample, so of course this is sample size sorry population size and this population mean. So we have to write here some parameter and here statistic right, so these 2 are both are about population right what about this capital N and small n.

Now small n is not actually statistics it is sample size right, all of these let us look at second first, this is population standard deviation and this is sample standard deviation b and c but not a no, so the right answer is b. So, standard deviation and standard deviation of sample right standard deviation of population and standard deviation of sample right. So, answer b correct, in random sampling when we describe mathematically how the objective our estimates are why is this?.

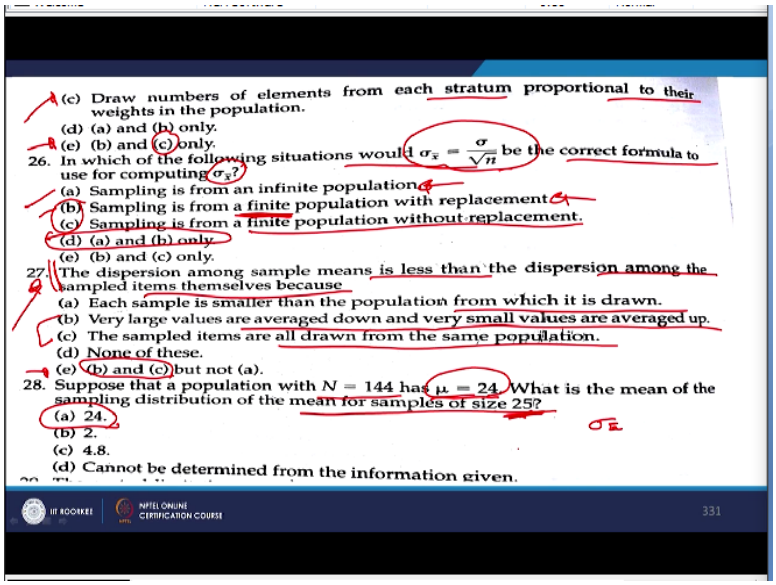
Let me read it once again in random sampling we can describe mathematically how objective our estimates are why is this we always know the chance that any population element will be included in the sample any just focus on this right. Second every sample always has an equal chance of being selected, all the samples are of exactly the same size and can be counted, none of this a and b, but not c.

In random sampling we can describe mathematically how objective are estimates are, so why what are the reasons for that. If you look at this 24th then e is correct, so every sample always has an equal chance of being selected this is there right this is one of the basic feature of probabilistic sampling right. And second is we always know that chance that any population element will be included in the sample right.

So, we know these two things let us we always say that the random sampling our objective estimate should be our estimates would be objectively correct right. Let us look at 25th suppose you are performing stratified sampling on a particular population and have divided into strata of different sizes. Now how can you now make your sample selection, so you got different strata let say this one this one this one right.

Now how to select elements from each one of these now strata selected random and equal number of elements from each strata should we go for equal number of elements from each one of them. Draw equal number of elements from each stratum and weigh the result first of all select at random and equal number of elements, draw equal number of elements from each strata and weigh the result.

(Refer Slide Time: 19:32)



Draw number of draw numbers of elements from each stratum proportional to their weights in the population. So, this is there in the answer sheet, so c is here in part a it means b also correct right. So, draw equal number of elements from each strata and weigh the results not necessarily of course there is nothing wrong in selecting those data those elements randomly. But we have to have equal number of elements from each strata.

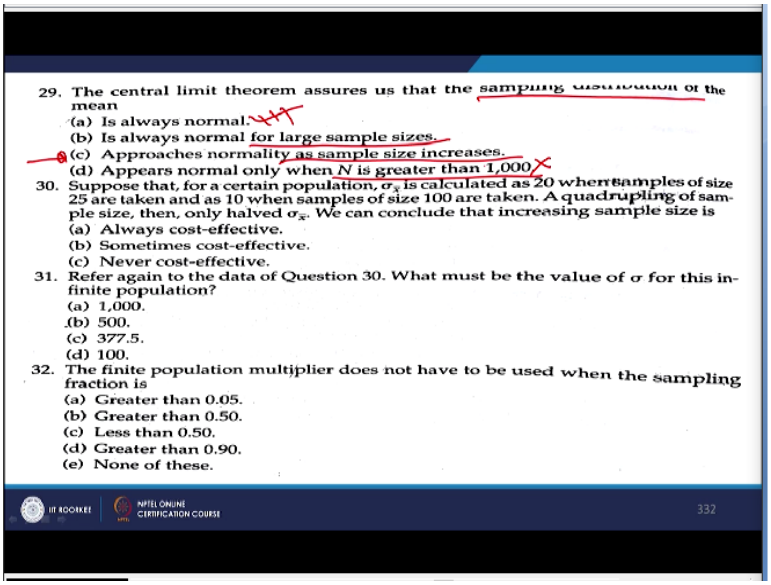
Let us look at this just once again keep in mind draw equal number of elements from each stratum this is an important point you should remember ok. Let us look at 26th, in which of the following situations would be the correct formula to use for computing population standard error. Sampling is from an infinite population, sampling is from a finite population with replacement, sampling is from a finite population without replacement a and b only it means this and this and b and c on it means these two.

So, for 26th option d is correct which is a and b, so sampling is from an infinite population of from finite population with replacement would you view a condition wherein this formula would be appropriate to apply. Let us look at 27th the dispersion among sample mean is less than the dispersion among the sample items themselves because the dispersion among sample mean is less than the dispersion among the sample items themselves.

Because each sample is smaller than the population from which it is drawn very large values are average down and very small values are averaged up. This sampled items are all drawn from the sample from same population, none of these and b and c but not a. So these 2 are the correct options it means answer e is correct. So, very large values have average down and very small values are averaged up.

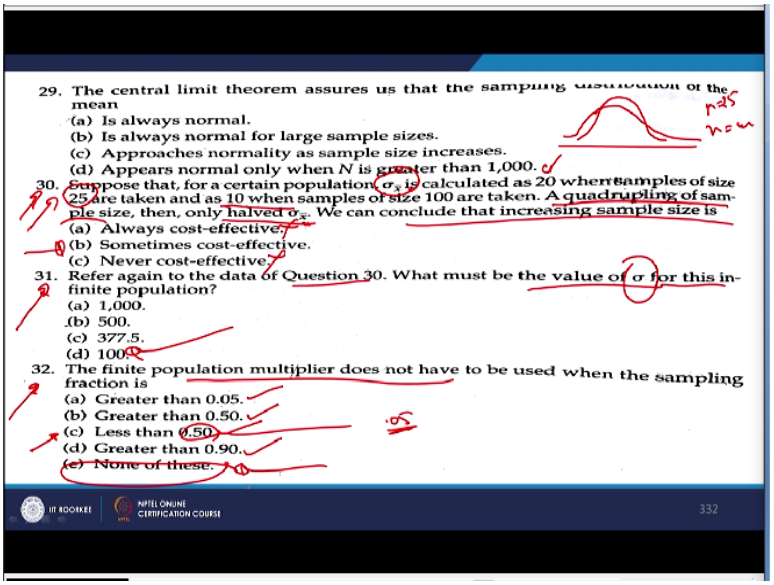
And the sample items are all drawn from same population that is why this statement is like this ok. let us look at 28th, suppose that a population with size 144. And has got mean 24 what is the mean of the sampling distribution of the mean for sample of size 25, so in fact the answer would be 24 remain as it is here right. What is the mean of the sampling distribution of the mean for sample of size 25. Because this sample size affects this not the mean ok.

(Refer Slide Time: 23:57)



The central limit theorem assures that the assumption distribution that sampling distribution of the mean is always normal, is always normal for large sample sizes approaches normality as sample size increases. Appears normally when n is greater than 1000, this is not correct this is **is** always normal no actually it depends on sample size. So, for 29th option c is correct, yes approaches normality as sample size increases right ok.

(Refer Slide Time: 24:39)



So, in fact in one of the classes we have seen that our let say this was skewed distribution right let us say population distribution. And when we increase sample size let say this is the distribution or sampling distribution with n is equal to let say 5 let say 25. But if we take n is equal to 100 then the same it would become exactly normal right, let us look at 30th number statement suppose that for certain population this is calculated as 20.

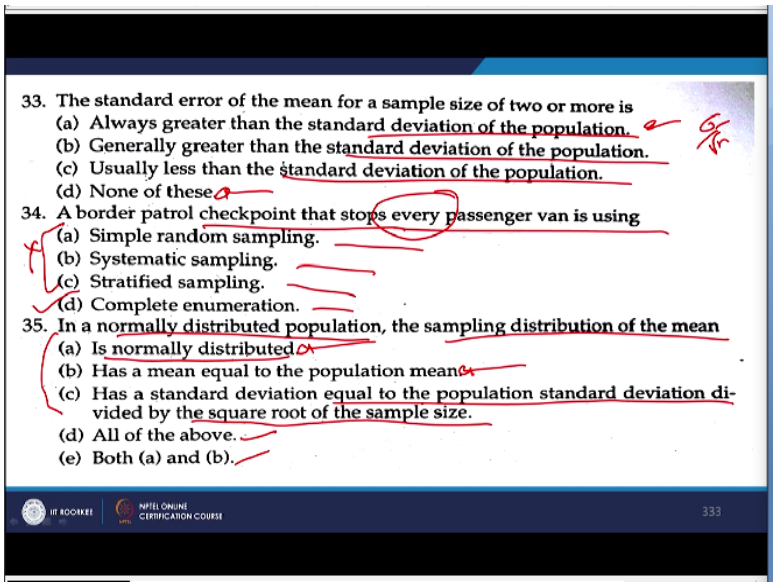
When sample of size 25 are taken and s 10 when sample size of 100 are taken, so this value the population standard error is 21 when sample size is 25 and it is 10 means sample size 100 right. A quadrupling of sample size right we have just increase sample size by 4 times right it was 25 now it is 100. Then how this half of this we can conclude that increasing sample size is always cost effective.

Sometimes cost effective near so these 2 are extreme answer, so the final one is this one right sometimes cost effective for question number 30th. Let us look at 31st one, refer again to the data in question 30th what must be the value for value of sigma for this infinite population. So the answer would be 100, what must be the value of sigma for this infinite population and the answer would be 100 ok.

Let us look at 32nd the finite population multiplier does not have to be used when the sampling fraction is greater than 0.05, greater than 0.5, lesser than 0.5, greater than 0.9, so answer is none

of these. Because it has to be less than 0.05 here it been less than 0.05 then you would have chosen this option, but for the time being for this statement this the answer right none of these.

(Refer Slide Time: 28:01)



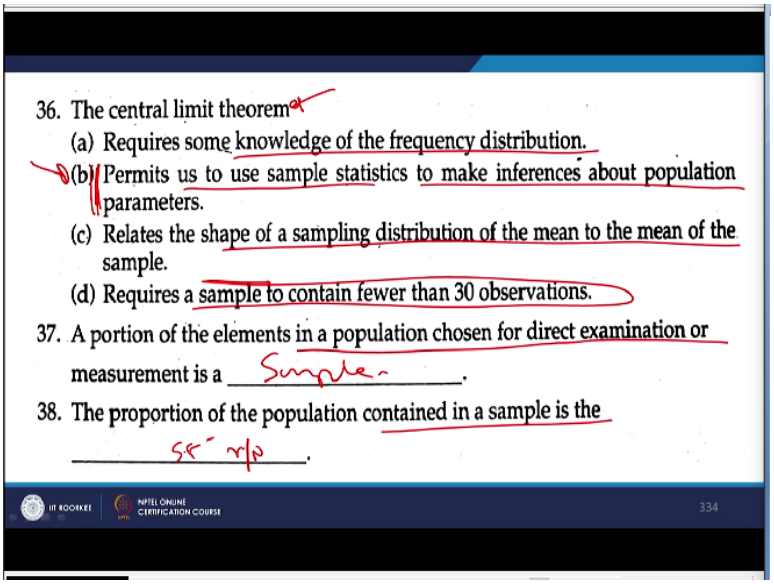
The standard error of the mean for a sample size of 2 or more is always greater than the standard deviation of the population. Generally greater than the standard deviation of the population usually less than standard deviation of population, none of these or it would be the standard error of the mean of a sample size of two or more is always greater than the standard deviation of population generally no, it is d none of these right.

Because we have seen in sampling distribution that the mean of population distribution and mean of sampling distribution are same. But the standard deviation of the and we do not call it standard deviations, so for a sampling distribution is can be we called standard error right. But the standard error will always be less than population standard deviation by this formula.

A border patrol checkpoint that stops every passenger van is using what simple random sampling, systematic sampling, stratified sampling or complete enumeration, just read this carefully. A border patrol checkpoint that stops every passenger, it means there is nothing like sampling right, so answer is d, it is complete enumeration. In a normally distributed population the sampling distribution of the mean is what?.

In a normally distributed population the sampling distribution of mean is normally distributed has a mean equal to the population mean yes this is true, has a standard deviation equal to the population standard deviation divided by square root of sample size, yes this is also true. So, what is the wrong answer what is the correct answer all of the above both a and b, answer is all of the above right all these 3 points are correct. Because it is always normally distributed of course we always say that sample size increases it becomes normal right.

(Refer Slide Time: 31:00)

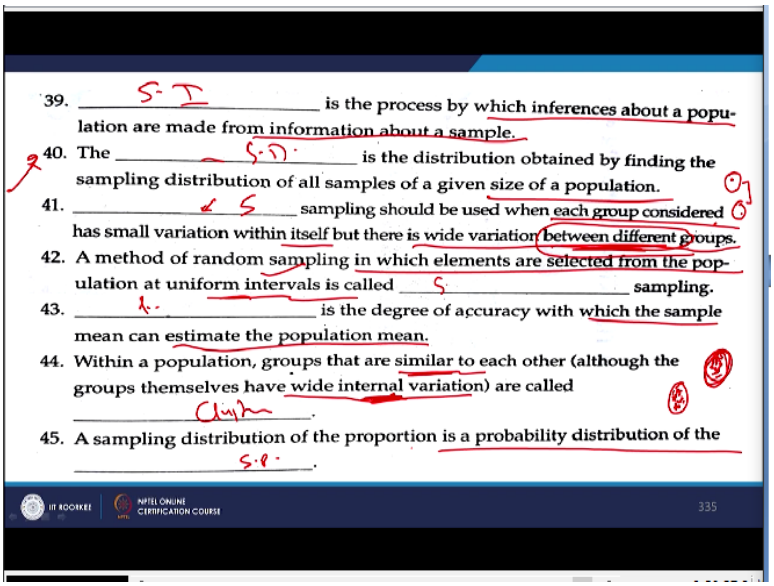


Let us look at this slide, the central limit theorem, the central limit theorem requires some knowledge of the frequency distribution permits us to use sample statistics to make inferences about population parameter relates the shape of the sampling distribution of the mean to the mean of the sample requires a sample to contain fewer than 30 observation. What is the correct answer?.

Central limit theorem actually permits us to use sample statistic to make inference about population parameter, for example with sample mean will calculate population mean. Let us look at this a portion of the elements in a population chosen for direct examination or measurement is what, what it is called a portion of the element in a population chosen for direct examination or measurement is called sample.

The proportion of the population contained in a sample is what?, the proportion of population containing in sample is sample fraction this is what we have seen or sampling fraction this is the one right 39th.

(Refer Slide Time: 32:39)



Is the process by which inferences about a population are made from information about the sample. So statistical inference is the process by which inferences about population are made from information about sample, so this is statistical inference. 40th the dash is the distribution obtained by finding the sampling distribution of all samples of a given size which the theoretical sampling distribution right or sampling distribution simply.

Dash sampling should be used when each group considered has small variation within itself but there is a wide variation between different groups, just see which sampling should be used when each group considered has small variation within itself but there is a wide variation between groups, this is stratified random sampling 41 right. Because let say there are 2 groups of let say boys and girls right.

So, these 2 are different groups right however in this group boys and girls there is little variation among themselves right. A method of random sampling in which elements are selected from the population at uniform interval is called systematic right, either uniform interval or let say after

some number of units or after some specific time and so on right. So, either time or let say some number or any logic should be there before selecting samples right.

So, that would be called systematic one. 43rd is the degree of accuracy with which the samples mean can estimate the population mean. So, this is the precision, precision is the degree of accuracy with which the sample mean can estimate the population mean right, within a population groups that are similar to each other although the group themselves have wide internal variation are called clusters, there is wide internal variation right.

So you will have let say a group of let say this one cluster second cluster. So, in this you will have let say if you are finding mean height of students studying in an educational institute, so all students are here right. Right from 1st year to final year PG and so on, so all of them are let say sitting in a canteen right. So, this is called a cluster right, let say many of them again one more cluster there in a bank right.

So there is wide internal variation right because you will have final year students as well as first year students right. A sampling distribution of the proportion is a probability distribution of the sample proportion right, so with this let me finish this session, next session will start some of the concepts related with estimation, thank you very much.