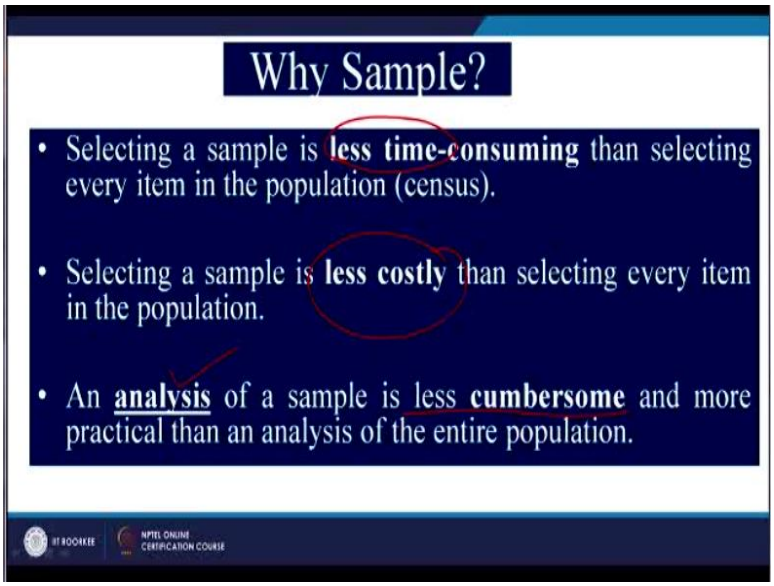
**Business Statistics**
**Prof. M. K. Barua**
**Department of Management Studies**
**Indian Institute of Technology-Roorkee**

**Lecture-22**
**Sampling and Sampling techniques**

Good afternoon friends I welcome you all in this session, as you were aware in previous session we discussed about probability distributions and we have worked out couple of exercises. In today's session we are going to have next lecture which is a logical in the series and the title of this lecture is sampling and sampling distribution. In one of the sessions we have seen what is sampling, what is sample, what are the benefits of sampling.

So, just quickly let me tell you once again that sampling helps you in making decisions about population and what are the benefits of sampling, there are couple of benefits.

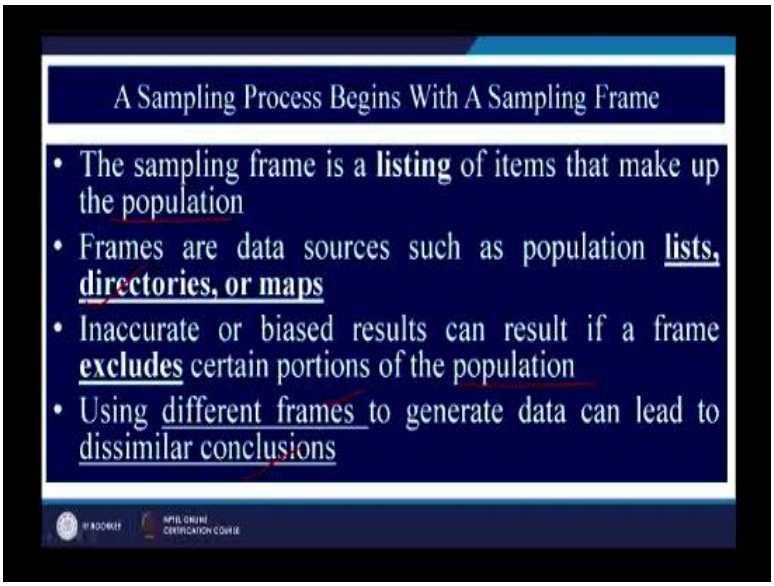
**(Refer Slide Time: 01:28)**




The first one is it saves time, it is less time consuming compare to complete enumeration, it is less expensive as a compare to complete enumeration. And many times analysis which you perform on sample or less cumbersome compare to the analysis on the entire population. The one more example of sampling is that let us say if the material which you are going to test is destructive in nature then it is good to go for sampling rather than complete enumeration.

So, sampling begins with something called sampling frame, so initially you will have a population and from that population you have to select representative samples. So, before selecting representative samples from the population, now there is something called sampling frame from where you will select these samples.

**(Refer Slide Time: 02:41)**



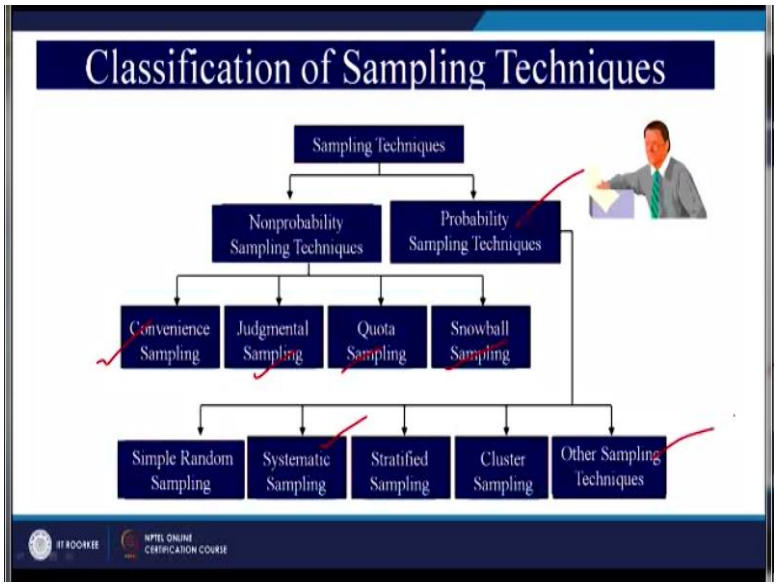

Sampling frame is nothing but a list of all the elements which are there in the population right. So, let us say telephone directory is one of the examples of sampling frame, so let us say you have got a telephone directory of all those homes of a particular city. And you have got names of those people who have got telephone numbers registered in the telephone directory. So, from the telephone directory you can select some of the samples or let us say maps or let us say if there is a class of 80 student's right.

So, that role is nothing but a sampling frame which will help you in selecting some of the students from population 80 student's right. Now inaccurate and biased results can result if a frame excludes certain portion of the population. So, try to have a frame in such a way that it should include all the population, no part of the population should be excluded from frame otherwise you will get wrong results.

Using different frames to generate data can lead to dissimilar conclusion yeah it it will be actually isn't it. So, **so** try to have an updated frame from where you are selecting samples. Let

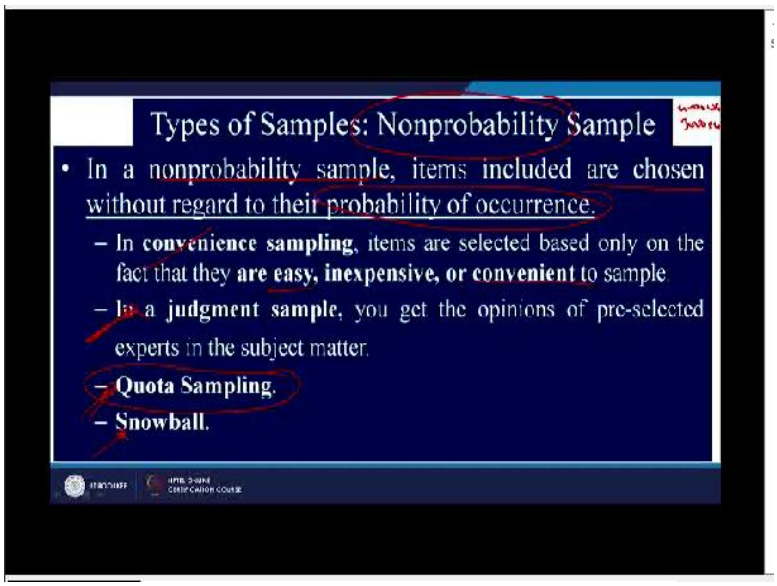
us look at how to sampling there are different sampling techniques broadly you have got 2 types of sampling techniques, you have got probabilistic sampling techniques and non-probability sampling techniques.

**(Refer Slide Time: 04:53)**




When you talk about non probability sampling techniques you have got four Convenience, judgmental, quota and snowball sampling. We will see each of these samplings in coming slides. So, for probability sampling techniques are concern you have got simple random sampling, you have got stratified sampling, you have got cluster sampling, you have got systematic sampling and other sampling techniques. So, let us look at what these sampling techniques are.

**(Refer Slide Time: 05:22)**



If you look at non probability sampling techniques, now the major difference in probability and non-probability sampling technique is this. In probability sampling the chance of selection of every element is equal while in case of non-probability sampling. The chance of an item getting selected is not equal right. So, just look at this in a non-probability sample items included are chosen without regard to their probability of occurrence.

So, that is one of the biggest drawback of non-probability sampling and if you let us say want to measure height of let us say average height of students in any educational institute. Then what you should do you can either use probability sampling and non-probability sampling let say if you are using convenient sampling then let say you have got 1000 student in institute.

So, you just select students according to your convenience, so let say if a student is at restaurant or it's walking on road or in sitting at some place then you just go and give him questionnaire for which you are collecting data right. So, that would be convenience sampling now many a times the sample which you select may not even be a representative one. So, that is a drawback of convenient sampling or non-probability sampling.

But the plus points of convenient sampling are easy, its inexpensive, and convenient it saves time is not it. But the problem is it is a non problasitic right you do not the probability of chance of a student getting selected is not equal right. If you look at judgmental of sampling here the samples are selected on the basis of experience of the person who selecting sample. Now this is quite a biased one because every person will have biasness towards selecting samples.

So, again it is not a good method, so for sampling is concerned right but there are you know several types of assignments research and projects where in initially you go for non-probability sampling. In initial stage of your researcher during framing of your initial hypothesis you should go for non-probabilistic sampling. Quota sampling is quite an interesting one.

So, let say in an institute there are 4000 undergraduate students and let say 3000 post graduate in student's right PG right and this are UG right. Now let say you want 100 samples out of the 7000, so what you would do 100 samples you want. So, you will select how many from here its

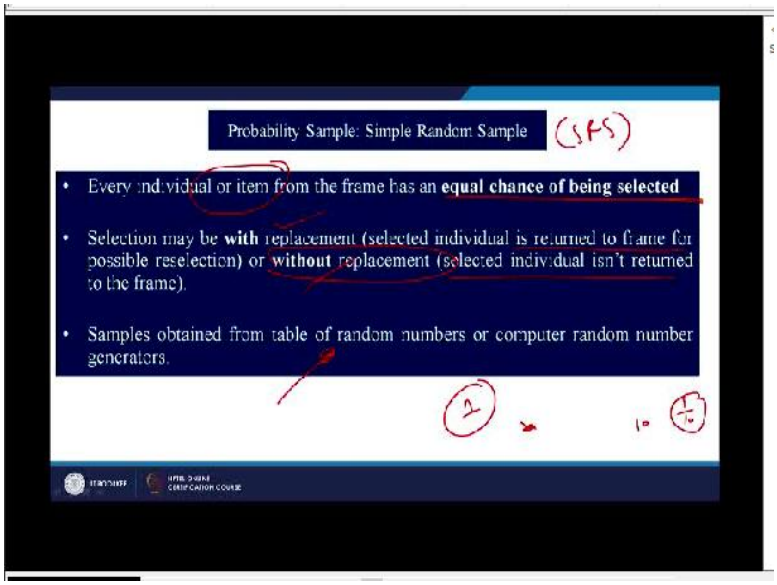
4000 by 7000 into 100 right similarly 3000 by 7000 into 100. So, you first have quota right, quota can be on several criterion it can be on the basis of let us gender male, female; rural, urban; UG students, PG students; teachers, students right.

So, first of all you will have 2 groups and then from each group you select samples either conveniently or using judgmental sampling. This is this quota sampling is somewhat similar to stratified random sampling which is a probabilistic sampling the major differences that in quota sampling you have got let say two groups. But this samples you select from those two groups are not using simple random sampling.

They are just selected on the basis of convenient sampling or judgmental sampling which 4th one is snowball sampling. A snowball sampling is again a non-probabilistic sampling technique where in what you do you select a sample and then you take reference from him. You ask a similar kind of sample from him and then you approach the second one and so on right. For example if you are looking for a male a widowed male who is 80 years old right.

So, first you talk to one person who is 80 years of age and widow as well then you talk to him and ask him is there any other person similar to him who is also more than of 80 years of age. Then he will give a certain references right, so you just talk to them, so over a period of time you will have lots of sample right. So, this known as snowball sampling if you look at probabilistic sampling techniques the first one is simple random sampling most important right.

**(Refer Slide Time: 11:51)**

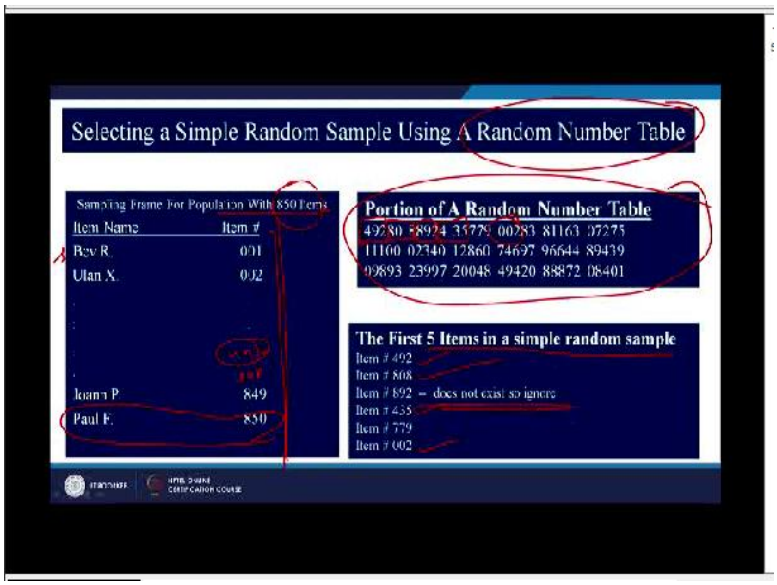



SRS simple random sampling, so every individual item from the frame has an equal chance of being selected. So let see if there are 10 students and if you select one then probability of each one getting selected is 1 by 10 right. So, this is not there in case of non-probabilistic sampling now here the selection can be with replacement and without replacement. So with replacement means the selected individual is return to the frame for possible reselection.

Here the selected individual is not return to the frame, so every time let see if you go for without replacement, so your population will reduce by let say a one unit isn't it. The samples obtained from table and there are several ways in which you can you perform simple random sampling or you can select samples using simple random sampling and one of them is using random number table's right.

The another one is using a roll it feels right, so you can use a roulette wheel as well for selecting random samples and there are computer generated random numbers and there are certain methods available for generating random numbers and so on right.

**(Refer Slide Time: 13:25)**

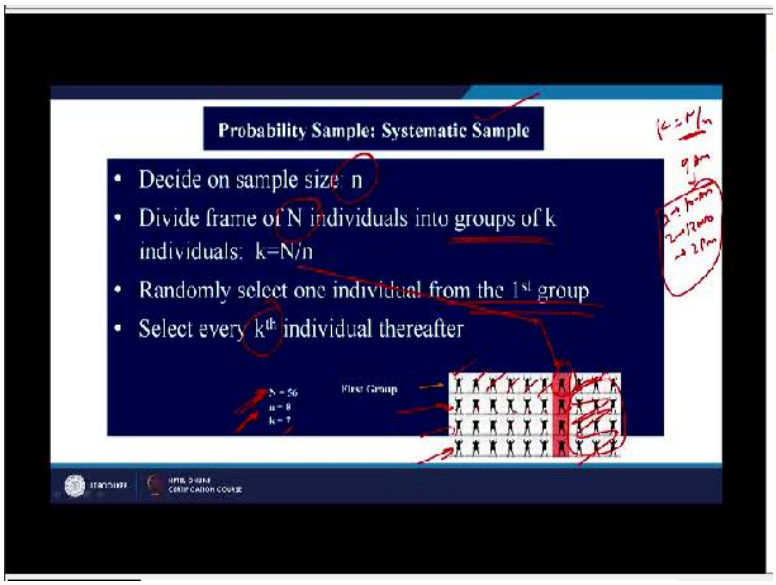



Let us look at random number table generating random numbers, so let say in a sampling frame for population you have got 850 items or 150 names. So, item number series 1, 2 and so on right. So, this one is 850 paul right and you have been given random number table this is your random number table. So, let say the first random number is and you are looking for 3 digit random number.

So, first random number is 492, so select item number 492 somewhere here right which is here right it is 492 is here. So, this is first element selected second is again 3 digits we have to take, so 80 and this 8, so 808. So, the next item would be 808 right, so the second fellow, then third random number of 3 digit is 892, 892 is not there in the sampling frame. So, this fellow, so does not exist, so we have to ignore this right.

The next one is 435 isn't select 435 and so on right and finally you would let say 002, so item number is 002. So, this how you are selected 5 items from frame of 850 items right simple random sampling.

**(Refer Slide Time: 15:17)**




And there is something was systematic sampling now we select the sample on the basis of certain logic. Now it could be let say time interval or let say you are selecting every let say population is 100, so you are selecting let say 10th number sample then 28th, 30th and so on right. So, you can have some logic right either time or some other criterion right. So, let say you have got population N and sample size is small and right.

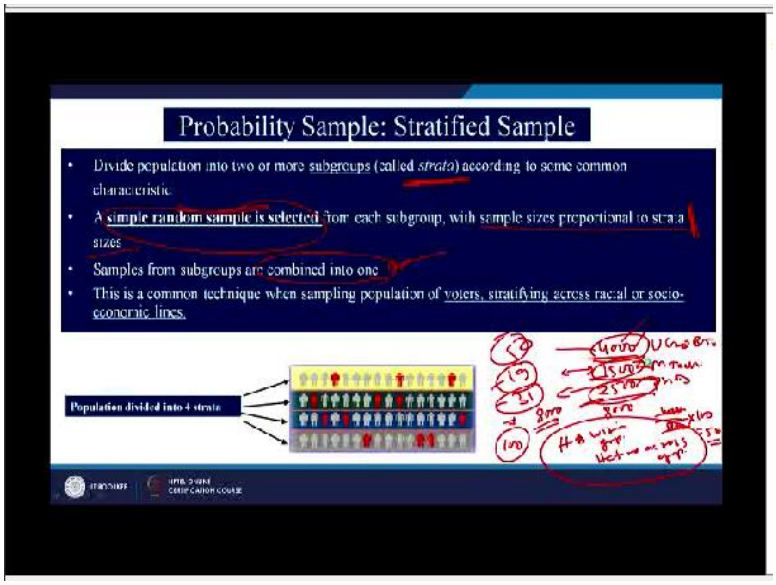
So, we you want to know the how many groups would be there, so k=N/n right, so divide frame of N individuals into groups of k individuals using N/n. So, randomly is select one individual from first group and then select every case individual. So, let say N is 56 and small n is 8, so N/n is 7, so you will have groups of so every 7th person will be selected. So, let say 1st, 2nd, 3rd, 4th, 5th, 6th and this 7th right from first group right.

So, this your second group third group fourth group right, so after this again 1st, 2nd, 3rd, 4th, 5th, 6th, 7th and so on right. You can ignore for the time being always right ok. So, this is known as systematic random sampling, so I will give you one more example let say you up you are a quality control engineer and on an assembly line some parts are being assembled and you want to know whether qualities is being quality parts are being produced or not.

So, let say shift starts at 9:00 a.m. in the morning and you take a part which comes out at 10:00 a.m. right. So, this is the first part you have selected the second part let say 12:00 noon right.

Then 2 p.m. so, after every 2 hours you are selecting samples right, so this is nothing but you are selecting samples on the basis of time interval. Let us look at stratified random sampling, stratified random sampling is basically a 2 stage sampling. So, here we divide our population into 2 groups are more than 2 groups.

**(Refer Slide Time: 18:25)**




So, each of these groups the subgroups are called starter right, so first of all will prepare several strata 2 strata or more than 2 strata. And from each stratum we will select few items and that would form our sample right, so from each stratum will select samples using simple random sampling. In fact in quota sampling also we did have 2 or more strata but the selection from those strata was not using simple random sampling ok.

So, with you can have a with sample size is proportional to stratum strata sizes is not necessarily but maybe non-proportional as well right. So, sample from subgroups are combined into one this is the second stage this is common technique when sampling evolution of voter stratifying across ratio or social economic lines. So, let us take an example educational institute let say there are 8000 students right, out of 8000 you got 4000 undergraduate students right. Now let us call them B.tech students.

You got 1500 M tech student's right and you got 2500 PhD student right and you to select let say 100 students out of this total ok. So, this is total 8000 students right you to select how many 100,

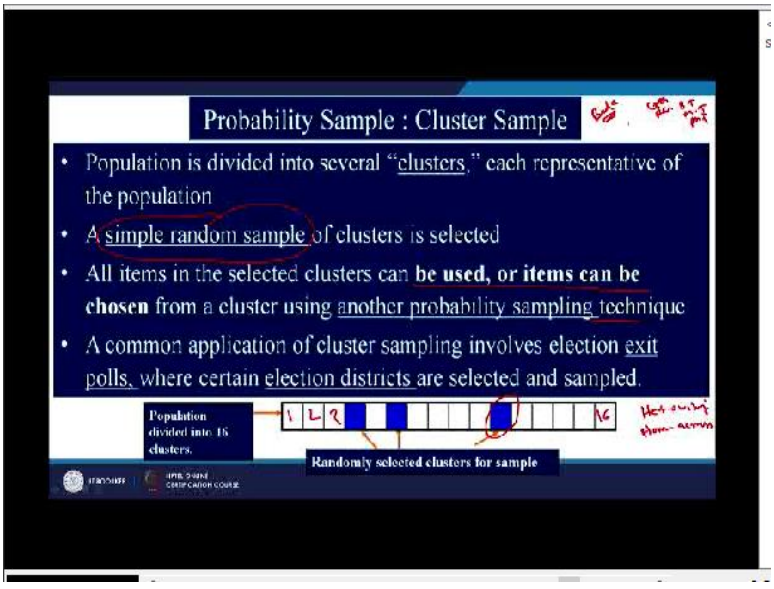
so how should we go ahead which is 4000 by 8000 into 10 right. So, this is 1 by 2 and this 100, so this is 50 isn't it. Then 1500 by 8000 into 100, so that would be approximately let say 19 students right.

So, select 50 from here select 19 from here and let say select from select 31 from this, so this total becomes 100. So, what we are saying we have selected the samples proportion to their original numbers in total population right. So, this known as stratified random sampling this is 2 stage sampling which are those 2 stages. In first stage we just prepared this strata and in second stage we selected samples right.

Now keep in mind whenever you select these Strata in stratified random sampling you are ensuring that there is homogeneity all these are homogenous students right homogeneities to be maintained. So, homogeneity within group and heterogeneity a cross group or a cross strata, so you have got 4000 all UG, all B tech students, right all M tech students, all PhD students ok. So, what is the point to be noted here there is homogeneity within groups.

Homogeneity means all would be of similar items right and heterogeneity across groups means these 2 groups are heterogeneous right. So, you all are UG all are PG isn't it. So, this how you can select samples using stratified random sampling.

**(Refer Slide Time: 23:22)**



If you look at cluster sampling again two stage sampling, so he population is divided into several clusters each representative of the population. So, again we divide our population into different groups and then either you select one particular group and perform analysis or you can use simple random sampling from each of those clusters. All items in the selected process can be used or items can be chosen from a cluster using another probability sampling technique.

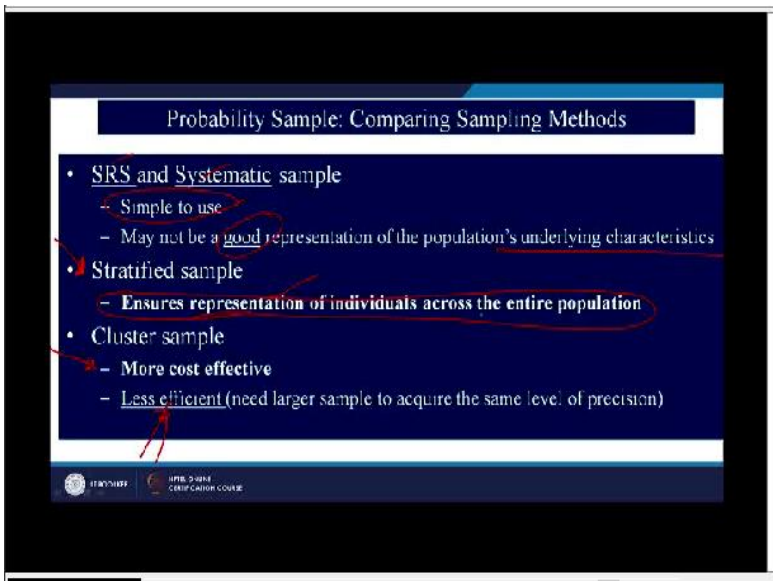
So, either you choose because you have initially prepared several clusters. So, either you choose one cluster and perform analysis on it or you take samples from each of those clusters and then you are sample and then perform analysis. A common application cluster sampling involves election exit polls where certain election districts are selected and sample. So let say these are different clusters their 16 plus 1, 2, 3 and so on right this is sixteenth for right.

So, randomly selected clusters for samples, so you just take one cluster and perform analysis or if you wish you take samples from each of the 16 clusters now what is the difference between previous one and this one. Keep in mind here in cluster sampling there is heterogeneity, heterogeneity within groups and homogeneity across groups. So, let say again in an education institute you are measuring.

Let say you want to know average height of students right, so rather than making different groups of UG, PG or let say B tech, M tech and PhD rather than making all this things what you should do let us you should go to a bank in the campus. So, in bank you will have all students B tech, M tech and PhD all will be there to where. So, that would be a cluster let us look at you go to restaurant in the campus.

So, in restaurant again you will find B tech M tech and PhD students, so that would be another cluster isn't it let say you go to let us a playground isn't it. So, you again you will find B tech, M tech and PhD students, so this how you can select one cluster and you can perform analysis. Now let us compare the sampling techniques which is better at the end of the day or let say out of these 2 which is better SRS and systematic sampling.

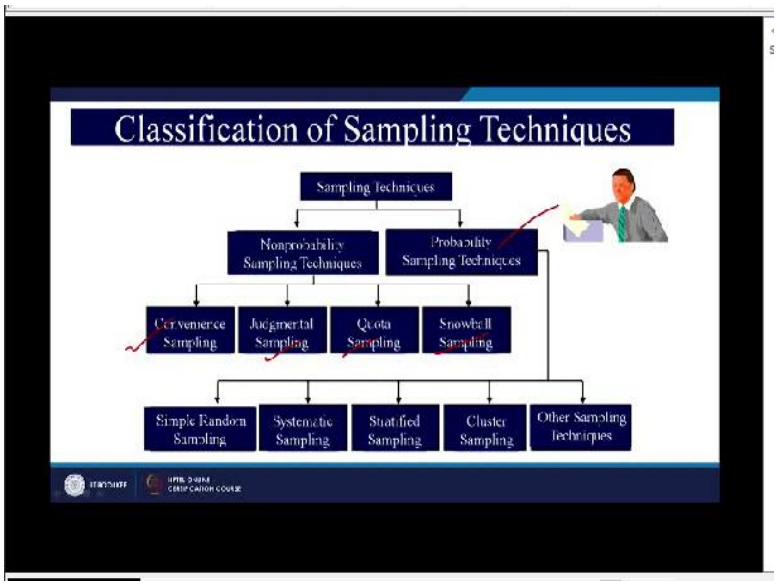
**(Refer Slide Time: 26:41)**




Of course the advantages are there simple to use right may not be good representation of the populations underlying characteristics, so that is a drawback. If you come if you look at stratified sample stratified sampling it ensures representation of individual across entire population. So, this is out of all these 4 this one would be profiled. Because it ensures representation of individual cluster across entire population.

Though cluster sampling is also good it is not a bad one however the advantage of cluster sampling is it. It is more cost effective compared to stratified sampling but less efficient, so cluster sampling is less efficient more cost effective but stratified sampling is accurate one accurate and it is presents representation across entire population. So, if someone ask you which is the best out of all these 4 then you can say stratified random sampling.

**(Refer Slide Time: 28:13)**




So, let me summarize what you done in today's session we have looked at different sampling techniques. In initial minutes of this session we discussed about advantages of sampling and then we will probabilistic and non-probabilistic sampling. In non-probabilistic sampling we have seen convenient sampling, judgmental sampling, quota Sampling and snowball sampling. In probabilistic sampling we have seen 4 methods we have seen snow simple random sampling.

We have seen stratified sampling we have seen cluster sampling and using systematic sampling. And when we compared all these sampling we found that stratified random sampling is the best one to use. So, if you are doing any survey then initially you should go for non-probabilistic sampling because those techniques will help you in getting some initial idea about the hypothesis which you are going to use in your project or in your research.

So, with this let me complete the session, in next session we will talk about sampling and sampling distribution which is again quite an important topic, thank you very much.