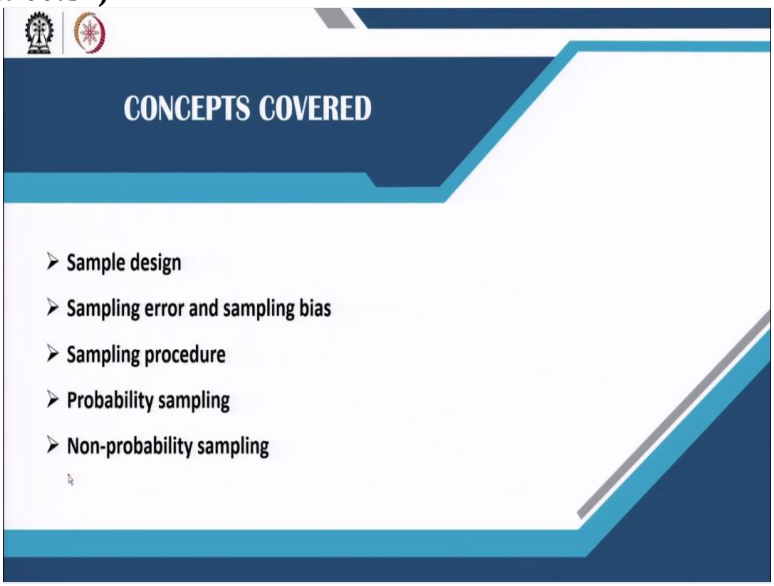
Urban Landuse and Transportation Planning
Prof. Debapratim Pandit
Department of Architecture and Regional Planning
Indian Institute of Technology - Kharagpur

Lecture-11
Sampling Theory-1

(Refer Slide Time: 00:32)



Welcome back. Module 3 will cover data collection and survey techniques and lecture 11 is on sampling theory part 1. So, the different concepts covered in this lecture are sample design, sampling error and sampling bias, sampling procedure, probability sampling and non-probability sampling.

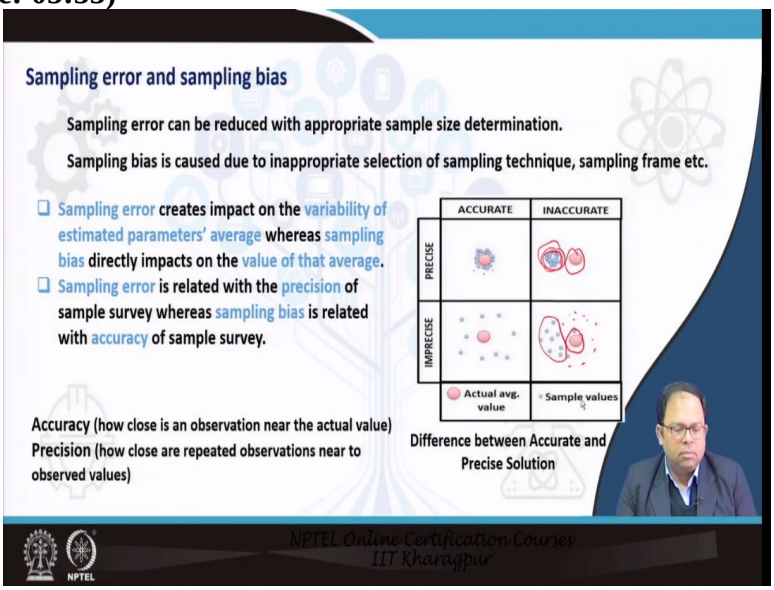
(Refer Slide Time: 00:51)

Background

Studies, like the one used in landuse and transportation, require surveys. But due to time and monetary constraints an entire population cannot be surveyed. For this a subset of the population, called sample, is studied. Sample design involves choosing a sample of appropriate size that is representative of the entire population. The important components of sample design are enumerated below.

Sample design is based on the **characteristics of the population**. A population can be finite, like the population in an urban area or infinite, in case of sampling water. The **sampling unit** is based on geographical location, socio-economic status, and ethnicity. For carrying out probability sampling one may need to start off with a comprehensive and correct list of all the elements of the population. This is called the **source list**, like the list of all persons in an area. From this a sample can be randomly selected. In the absence of such a list, non-probability based techniques are used. The next important determinant for survey design is the **parameter of interest**. For example, if the survey concerns bicyclists in an area, then the parameter of interest is a person having a bicycle. So, people with a bicycle become part of the sampling unit only. **Sampling procedure** deals with the technique of surveying to be used. And **sample size** depends on the precision required (sampling error), acceptable confidence level, parameters of the population, the size of the population, and population variance.

(Refer Slide Time: 05:55)



Sampling error and sampling bias

There are two types of error associated with survey sampling such as sampling error and sampling bias. Sampling error crops up because instead of studying the entire population, a small sample is studied. Every time a new sample is drawn from the population, the sampling error varies. Also, the larger the size of the sample, lower is the sampling error, generally.

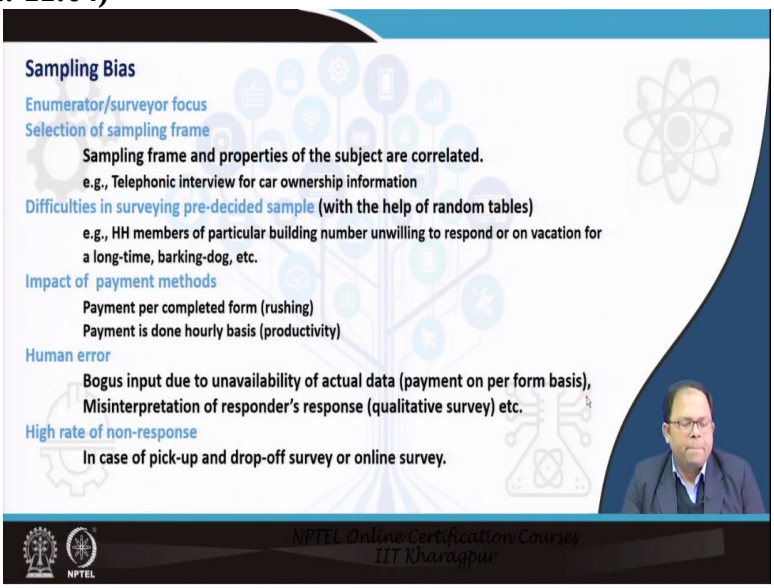
Sampling bias is caused due to inappropriate selection of sampling technique, sampling frame etc. Sampling frame is the method of collecting the samples. For example, somebody is conducting the telephonic interviews to collect samples. So, that is the sampling frame. So, sampling bias is caused by incorrect sample collection technique and has nothing to do with the sample size. Sampling error creates impact on the variability of the estimated parameters average, whereas sampling bias directly impacts the value of the average. Sampling bias results in a sampling mean very different from the population mean. Sampling error, on the other hand, results in variability of the estimated parameters.

So, an average of all these draws for the mean value will show the mean to be near to the population mean, but the variability would be high, i.e. the standard error or the standard deviation of this observation would be high. So, sampling error is related with precision and precision means how close are repeated observations near to the observed values.

A sampling bias is related with accuracy of sample survey. So, when sampling bias is there, then the estimates are not accurate. For example, in this case, one can see even though the readings are precise, that estimate is wrong.

If a study is so designed, that it is studying only a certain type of people in the entire population and not all the different kinds of people, it will lead to an inaccurate result. This is a sampling bias. This is a case of both inaccuracy and imprecision.

(Refer Slide Time: 11:04)



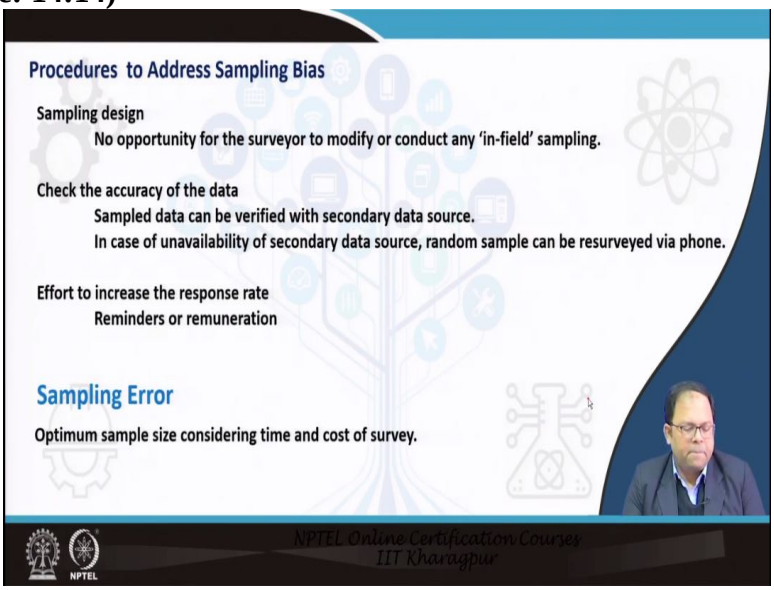
Sources of sampling bias

Sampling bias can be a result of enumerator or surveyor focus. That means, a surveyor, may choose to focus on certain groups and he/she misses out on the other people. So, that leads to some amount of bias from the enumerator's point of view.

Depending on the sampling frame, there can also be a bias. For example, if a telephonic interview is conducted for car ownership, it results in a bias. This is because telephone ownership and car ownership are correlated.

Sometimes a pre-decided sample cannot be surveyed or the enumerator is unable to fill up the entire questionnaire due to unavoidable reasons. When enumerators are engaged on an hourly basis, quality of data collected can go down. There is also a possibility of misinterpreting the response on the part of the enumerator. So, these are the different forms of sampling bias that can occur.

(Refer Slide Time: 14:14)

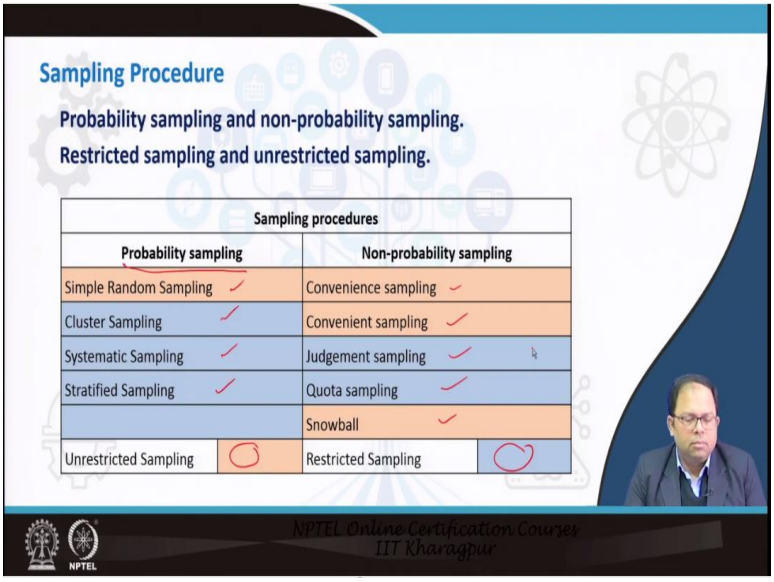


Removing Sample bias

To remove sample bias, the survey questionnaire should be unambiguous. This will weed out surveyor discretion. And then the accuracy of the data should be verified with a secondary source. If secondary sources are not available then for part of the survey can be rechecked via telephonic interviews. Also incentives can be offered to the respondents to improve response rate. So, these are the different ways we can address sample bias.

An optimum sample size can reduce both sample error and the time and cost of the survey. This will be taken up at the next lecture.

(Refer Slide Time: 15:42)

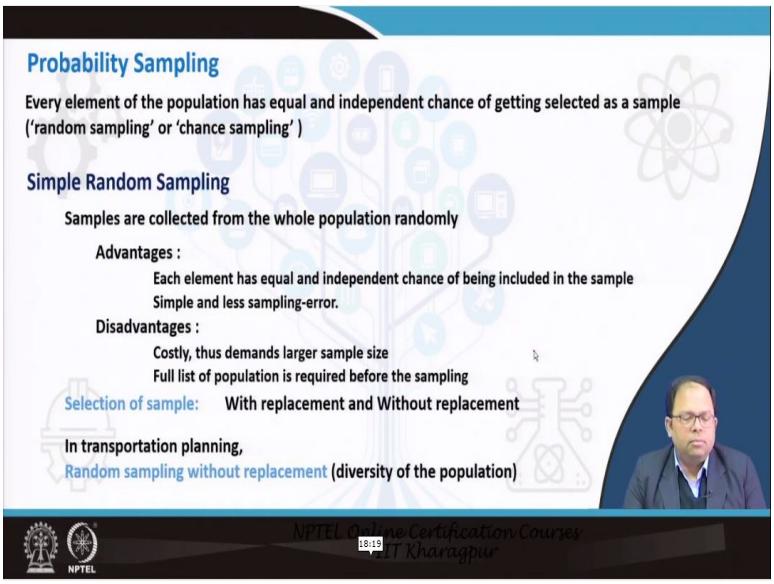


Sampling procedure

One way of categorizing sampling procedure can be in terms of probability and non-probability sampling. In probability sampling, unlike non-probability sampling, every member has an equal chance to be drawn into a sample from a particular population. Another categorization can be in terms of restricted sampling and unrestricted sampling. In restricted sampling there are certain rules or certain restrictions that are first put in and then a survey is conducted.

The table shows different types of probability and non-probability sampling techniques. The restricted ones are marked in gray, whereas non-restricted ones are marked in yellow. Under probability sampling there are different kinds of probability sampling like simple random sampling, cluster sampling, systematic sampling, stratified sampling and so on. Whereas in nonprobability sampling, there is convenience sampling, convenient sampling, judgment sampling, quota sampling and snowball sampling.

(Refer Slide Time: 17:44)

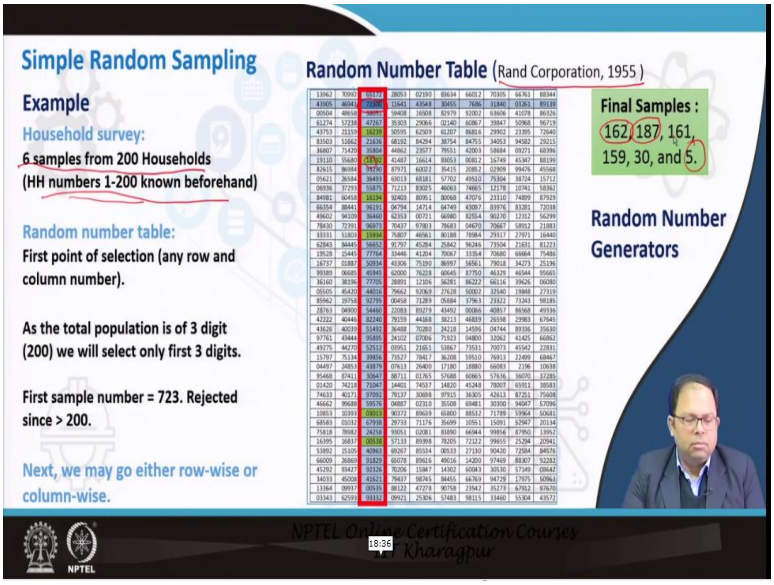


Probability sampling

In probability sampling every element of the population has equal and independent chance of getting selected as a sample and it is either called random sampling or chance sampling. Random sampling or chance sampling is the most common form of sampling that people undertake.

Probability sampling has got certain advantages. There each element has an equal and independent chance of being included in the sample. Thus, the chance of sampling bias and sampling error is less and it is a simple survey. Whereas, the disadvantages are, that it requires a much larger sample size. So, it is also costly. However, the main problem is that, the full list of population is required before we start the sampling. Simple random sampling can also be done with replacement and without replacement. When the sampling is done with replacement, a person has the chance of getting drawn again in the next draw also. When sampling is done without replacement a person is not getting another chance to be drawn in the sample. In transportation planning, most of the random sampling is done without replacement to improve the heterogeneity or the diversity of the population.

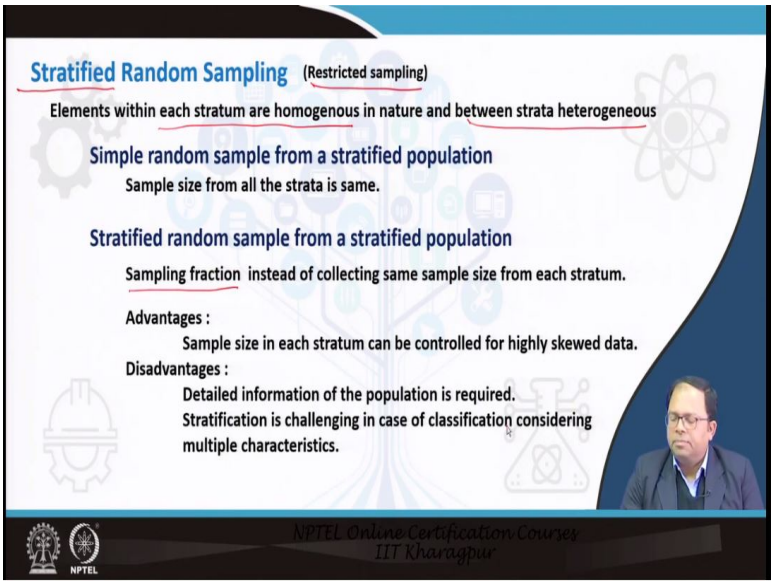
(Refer Slide Time: 19:57)



Simple random sampling

This is an example of how random sampling is done. In this case, a random number table prepared by Rand Corporation is used. One can also use random number generators in the computer, like in Excel. This kind of table can be carried in the field and can be used to decide whom to survey. Let us suppose there is a requirement to collect 6 samples from 200 households. Households numbered 1 to 200 are known. Starting from any row or column a number is selected. Since, the households are numbered between 1 and 200 (i.e. a three digit number) the first three digits of the selected cells are looked into. The first number is 723 which are more than 200 and this is not considered. The next numbers are 380 and then 472, which are similarly rejected. The next cell is 162, and the household numbered 162 is selected. This process is repeated till six households are randomly identified.

(Refer Slide Time: 23:04)

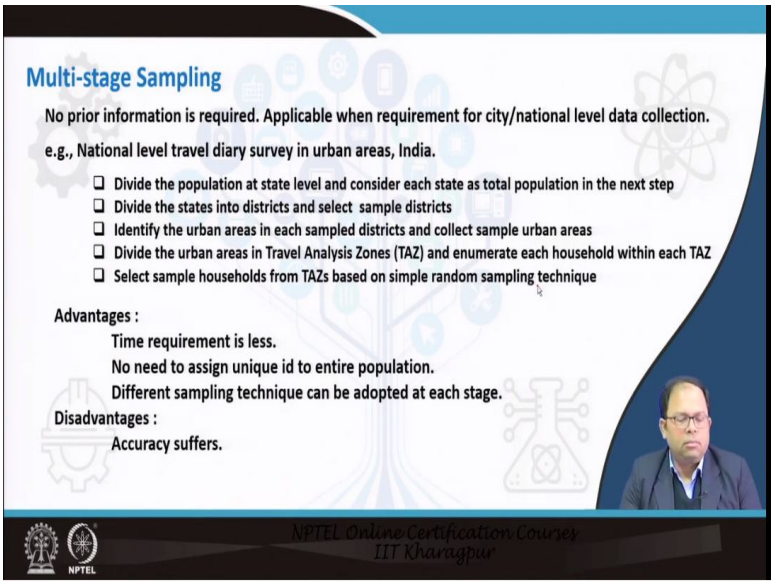


Stratified random sampling

The next procedure is stratified random sampling, which is a restricted sampling procedure. It is restricted because the population or sample is stratified. There are two types of stratified random sampling. The first type is simple random sampling from a stratified population and the other is stratified random sampling from a stratified population.

In stratified random sampling the entire population is divided into different groups having different characteristics. Thus, elements within each stratum are homogeneous in nature and there is heterogeneity between strata. And a survey of each group is carried out, with equal number of samples from each stratum. But if the strata have unequal populations this can result in sampling errors. In such a case, stratified random sample from a stratified population takes care of the problem. In this case the number of samples collected from each stratum is equivalent to its share in the population. Thus, sampling fraction is considered in stratified random sampling from a stratified population, which helps when the data is highly skewed. Stratification becomes challenging when multiple characteristics are used for stratification of the population instead of one.

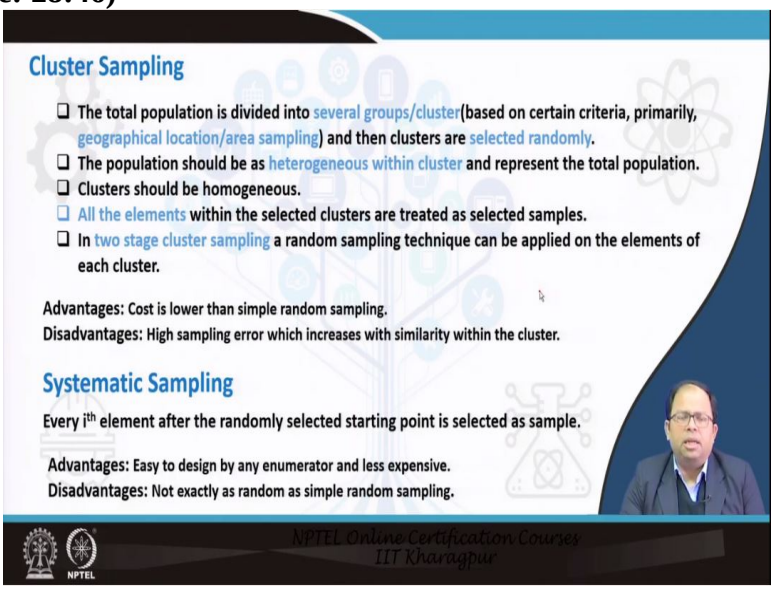
(Refer Slide Time: 26:17)



Multi-stage sampling

In multistage sampling no prior information is required. But this is applicable in case when the data collection requirement is at a very high level, like a travel diary survey at the national level. It is very difficult to survey a sample from the entire population for a huge sample size. For this the population at the state levels is considered which are further divided into districts. A few sample districts are considered. From the selected districts a few urban areas are sampled. And then within each of the sampled areas, one can determine the TAZ (travel analysis zones), where sampling can be done based on simple random sampling technique. As the population is broken into so many levels, the technique suffers from a lack of accuracy.

(Refer Slide Time: 28:46)



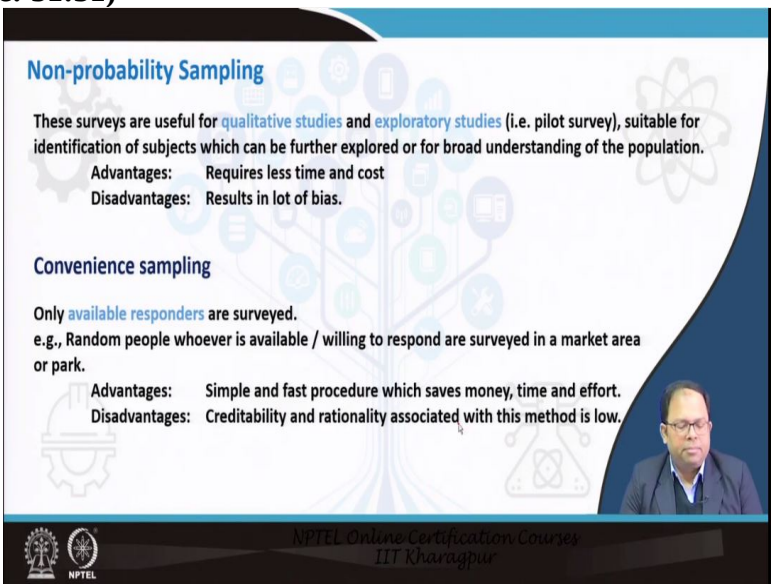
Cluster sampling

In this case the total population is divided into several groups or clusters based on some criteria, and then some of these clusters are selected randomly. The population within a cluster displays the heterogeneity of the entire population. That is why only a few clusters can be selected for studying in detail. All the elements within the selected clusters are then treated as selected samples and one can determine the different parameters based on this particular sample.

Systematic sampling

In this case, every element after a randomly selected starting point is selected as sample. For example, if 5 households are to be selected from a list of 200 households and a starting household is randomly selected at 30. To this number we add 40 (200/5 = 40) to obtain the subsequent samples, i.e. 70, 110, 150, and 190. This technique is easy to design and conduct and less expensive, though the degree of randomness is reduced.

(Refer Slide Time: 31:31)



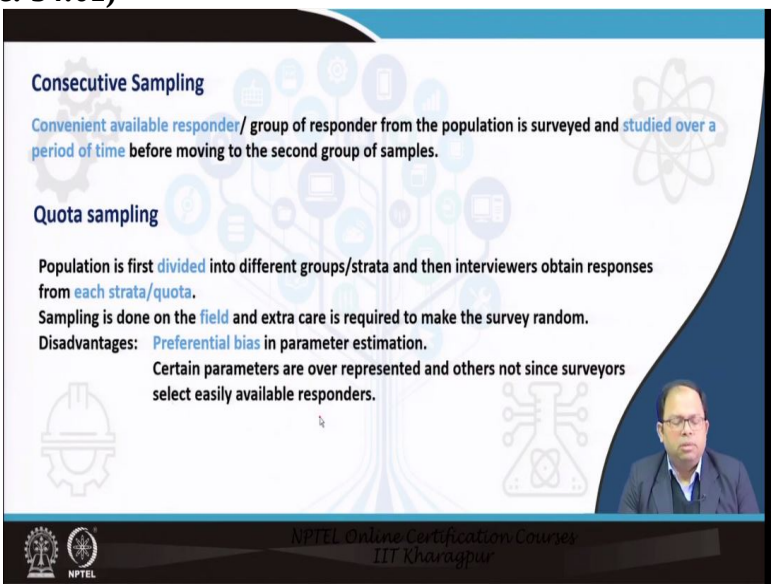
Non-probability sampling

These surveys are mostly useful for qualitative studies and exploratory studies like a pilot survey (a survey before the actual survey to study the characteristics of the population). This helps in developing a broad understanding (within the limits of time and cost) of the population so that one can actually design the sampling procedures. As samples are not randomly selected, there is a high chance of bias.

The first type of non-probability sampling is convenience sampling, where only available responders are surveyed. For example, if one wants to find out the people who use both Para

transit and bus, the enumerator can go directly to a Para-transit stop and bus stop and randomly survey people instead of going to a household, where many people do not even use Para transit or bus. While the credibility and rationality associated with this method is of course lower, the procedure is simple, fast and cost effective.

(Refer Slide Time: 34:01)



Consecutive sampling

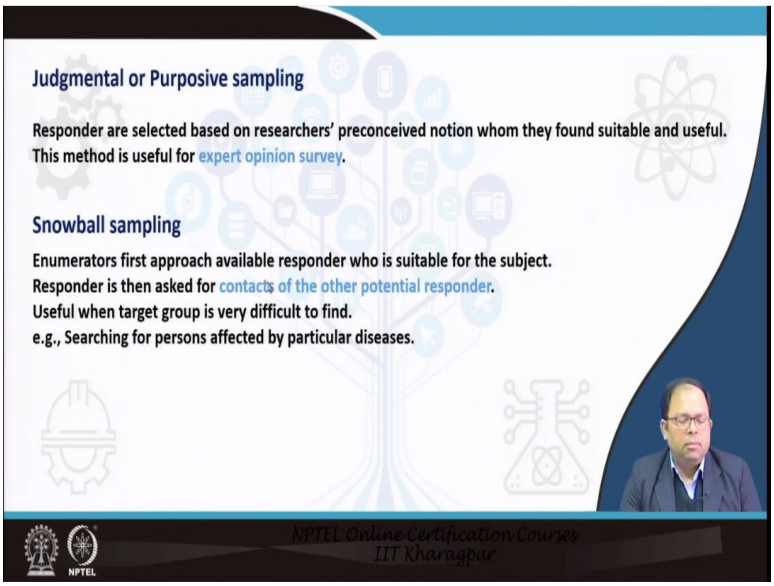
Consecutive sampling is almost similar to convenient sampling. But here, available responders or groups of responders are studied over a period of time before the next group of samples is taken up. So certain characteristics are determined with respect to the first sample before moving onto the next set of samples, where certain modifications can be adopted based on the study of the first group of samples.

Quota sampling

In quota sampling, the population is divided into different groups and samples are taken from each group to meet a quota. The sampling is done on the field and extra care is required to make the survey random.

There may be a preferential bias in parameter estimation, where the enumerator may prefer certain groups or certain service or certain household type and ignore other kinds of households. With this kind of bias creeping in certain parameters are over represented and others are not.

(Refer Slide Time: 35:28)



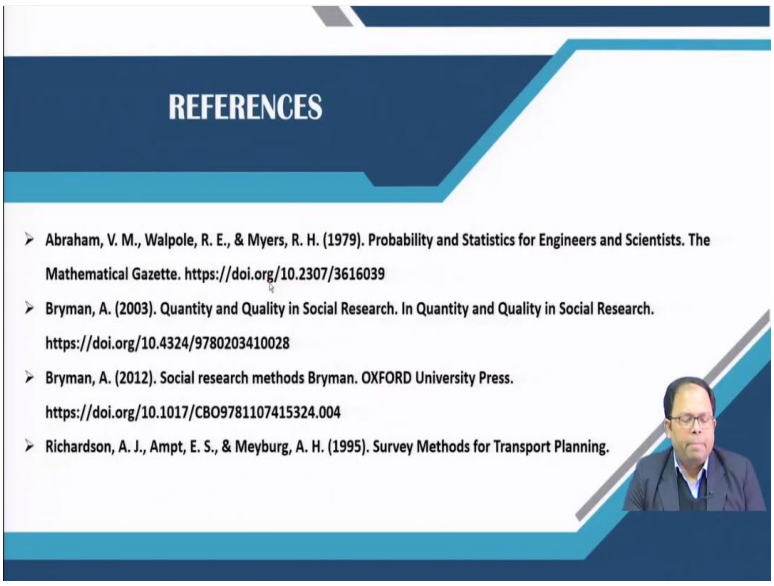
Judgmental or purposive sampling

Here responders are selected based on the researcher's preconceived notion of suitability and usefulness. For example, in expert opinion survey, the surveyor determines who is an expert in their particular field and enumerates his/her opinion.

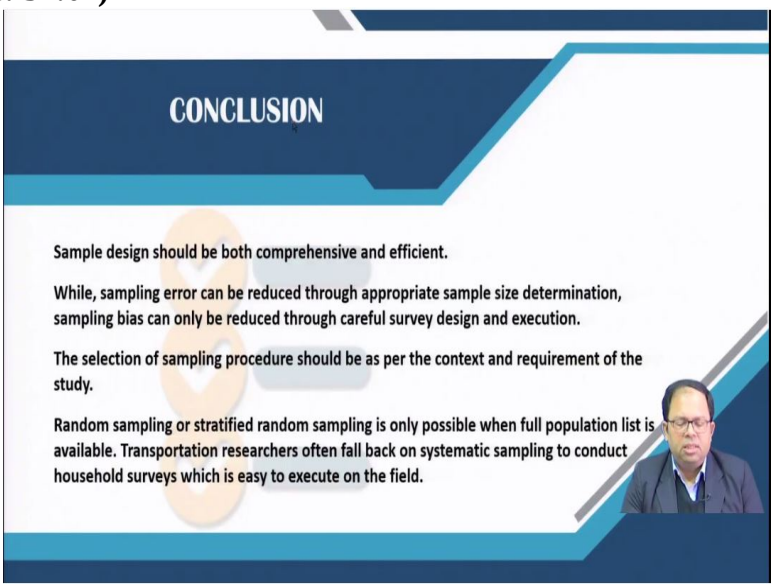
Snowball sampling

In snowball sampling enumerators approach available responders suitable for the particular study and then the responder is asked for contacts of other potential responders. This technique is very useful when the target group is very difficult to find. For example, people suffering from a particular disease can be traced through doctor's appointment or various support groups.

(Refer Slide Time: 36:51)



(Refer Slide Time: 37:02)



Conclusion

Sample design should be both comprehensive and efficient. While sampling error can be reduced through appropriate sample size determination, sampling bias can only be reduced through careful survey design and execution. The selection of sampling procedures should be as per the context and the requirement of the study. Random sampling or stratified random sampling is only possible when full population list is available. Transportation researchers often fall back on systematic sampling to conduct household survey which is easy to execute on the field.