INDIAN INSTITUTE OF TECHNOLOGY ROORKEE

NPTEL

NPTEL ONLINE CERTIFICATION COURSE

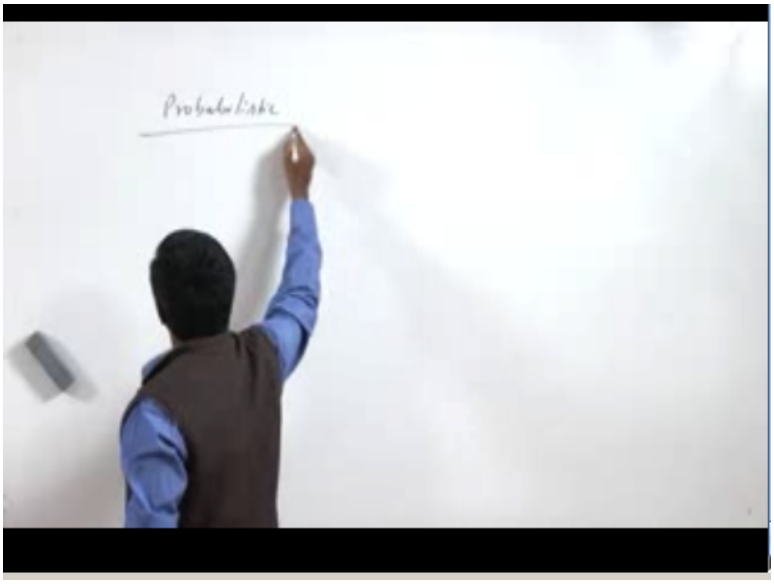
Marketing Research

Lec-16
Sampling & Sample Size Determination

Dr. Jogendra Kumar Nayak
Department of Management Studies
Indian Institute of technology Roorkee

Welcome friends into the session of sampling so in the last session so we were continuing with that so in the last session we discussed about the non projective sampling basically what is the example and how do we start the sampling and what is the non projective sampling and then we started with the projective sampling methods in which improbably we said there are a few.

(Refer Slide Time: 00:46)

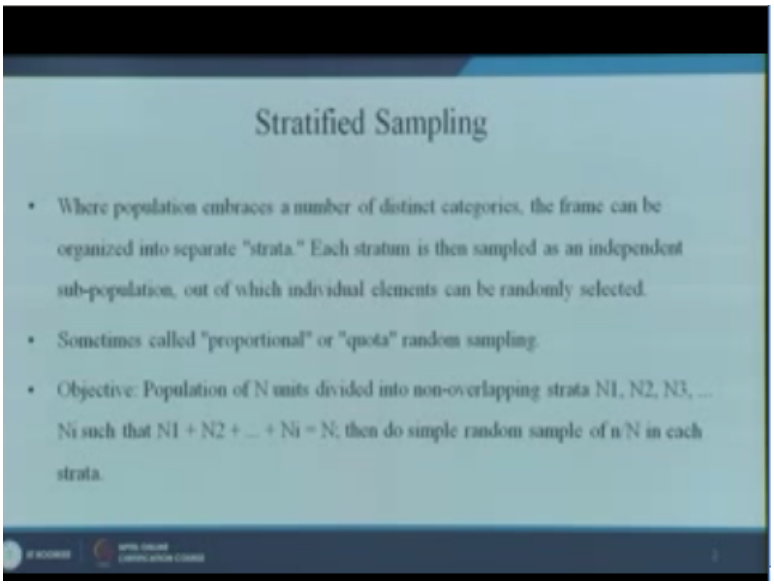


For example we started with a simple random sampling okay in the first simple sampling in which we said that every item right every respondent is an equal chance of selection so that is the duty of a probable sampling where the buyer sense is not there so there Is no buyer sense because everybody is getting so in my classes I said there are 50 students and I want to pick up the new

five so I pulled up randomly and this five are the people who also there is no buyer isolation process.

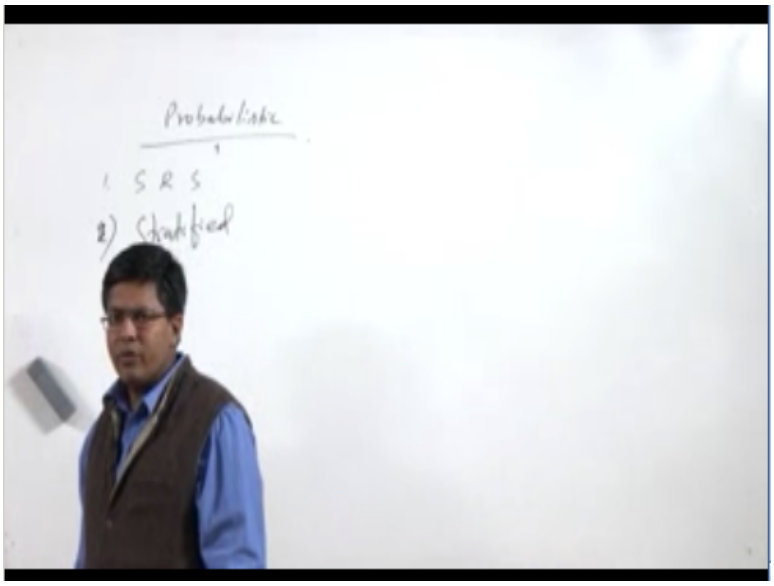
Yes one thing is for that they could be questionable whether this five are actually adequate or they are justified because above them might the five were students actually who were the low attendance in the class whenever visited the class most of the time though in that case this becomes a you know difficulty but so be it but the objective of a projects you know sampling is to reduce the buyers right. Second let us see once we finish that the next was the stratified sampling

(Refer Slide Time: 01:55)



So stratified sampling as you can understand from the name right stratified.

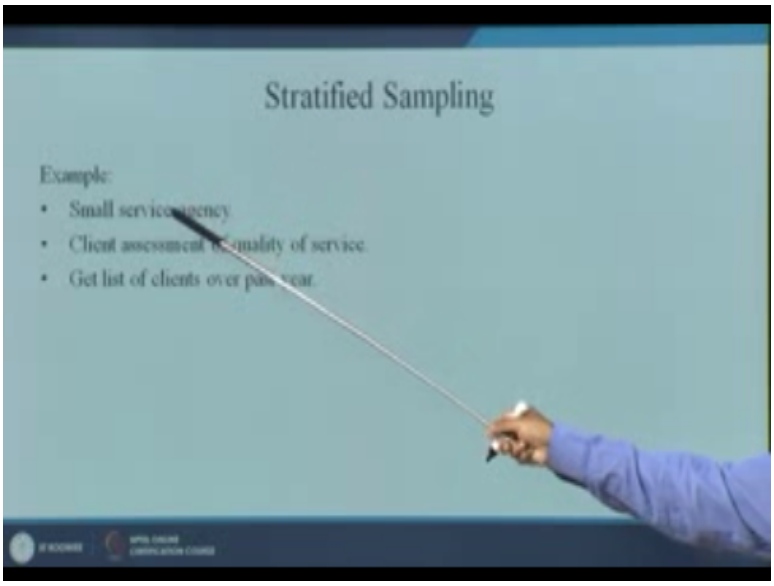
(Refer Slide Time: 02:03)



Which itself is a word like strata right so this is what emerged from the word strata right what does the strata mean the art is the example the crust of the earth the layer is you know spread up into different strata's okay the atmosphere where all this are strata right basically we divide what he is saying population stratified sampling is a number of distinct categories the free can be organized into separate strata so whatever is a population say.

Now each that is why I have been divided into several strata okay these strata has actually they are completely defining the population so sometimes they are propositional or quota or sampling also but it is not necessary that it should be proposition ate all the time propositional it could be dis propositional dis stratified all the time is also there stratified sampling are of two types proposinate stratified sampling and disproposinate stratified sampling but one thing is there which are I am telling right. Let us see so example if for example.

(Refer Slide Time: 03:19)



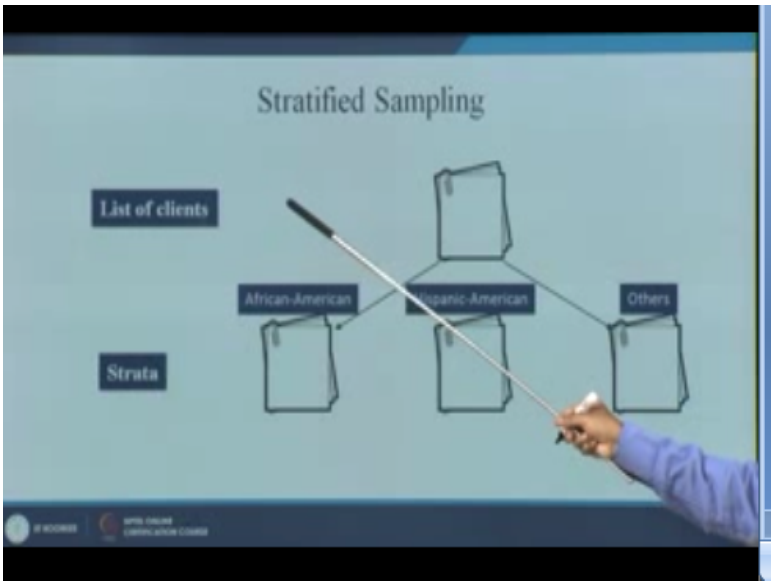
Starts are divided into small you see and somebody is using a client assessment of quality of service we can use stratified sampling get to so when you n want to make a study right about let us say right about let us say for example the population about its behavior or something let us say what is the kind of the cigarette smokers like the habits of cigarette smokers right.

So in such a condition stratified sampling could be used to reduce to increase three accuracy and reduce the cause of or in a study now what he is doing basically so as I said the strata the stratified sampling traits the population into several strata's.

(Refer Slide Time: 04:06)

To its several variables right so each strata now from each strata there would be some representation right there will be some representation for or from each strata so let us see what is deemed.

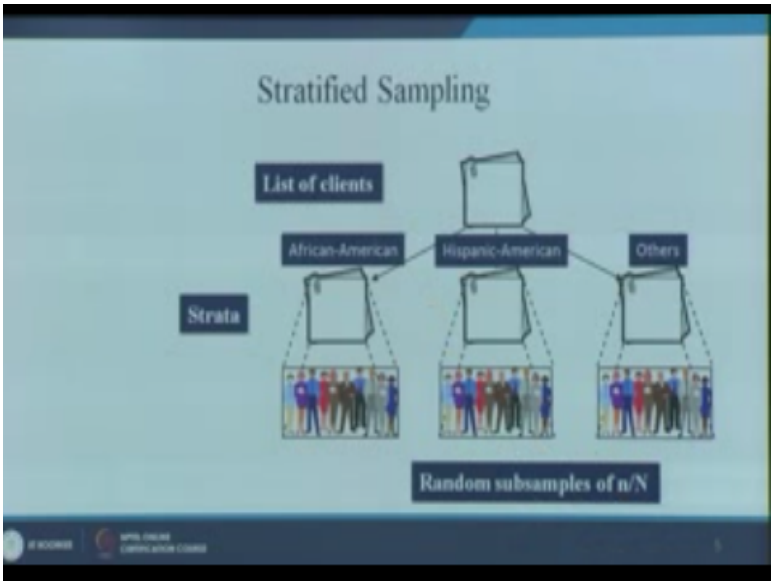
(Refer Slide Time: 04:21)



The list of clients sampling for you right so from where three different kinds of clients have been three strata's have been built African American Hispanic American and others the difference between you can see something is similar to you like quota but here the difference between the quota a and the at ratified sampling is that the coating sampling you were not using probabilistic method whether you are using a convent or judgmental method right.

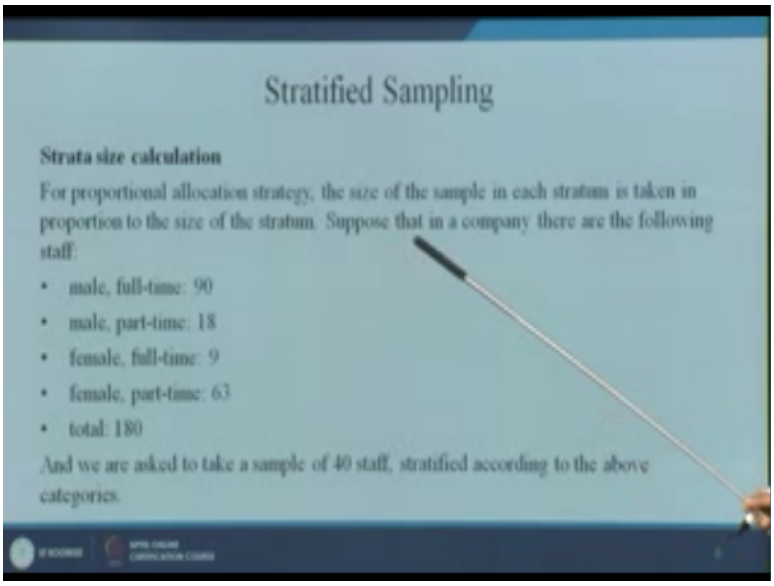
In the case of stratified sampling the method of lifting the segment is the prpoblastic method it is the only difference so now let us say from each stratum there is different presentation okay then now randomly short samples are pulled from the each strata okay.

(Refer Slide Time: 05:13)



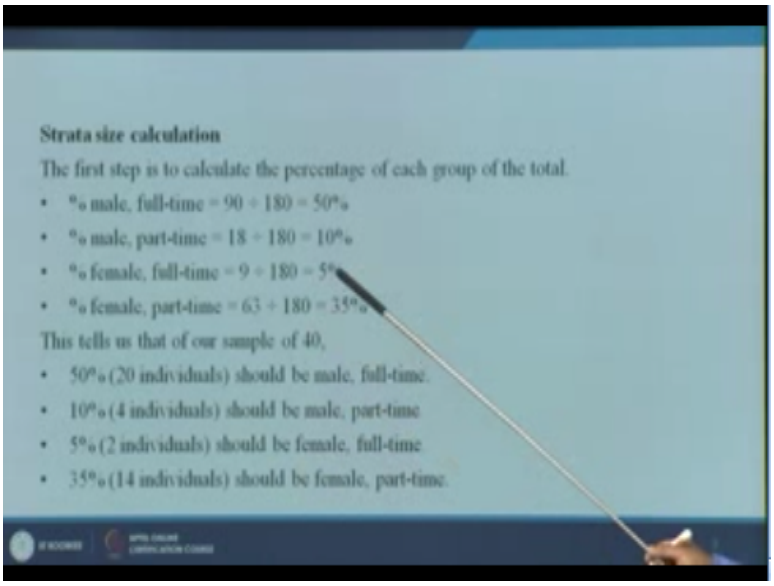
So when you pull up the strata when you broke up into each strata and from each strata you randomly you know select few samples right if you repondence you can say whatever. So the thing is that you have divided when the population into several clear cut stratus and this strata's are reprentating right there is at least some representation it cannot be zero that is a very important thing okay so for example it is a case.

(Refer Slide Time: 05:48)



For propositional strategy for you know for stratified sampling what did you saying the size of the sample in each Stratton can be proposition ate to the size of the stratum suppose if you company the following staff male 90 full time 90 part-time fully 18 female full time 18 full time4 is 9 female is part-time is 63 okay so the total is 180 now we are asked to take a sample of 40 so we need only 40 people as our sample now how do we use that stratified sample let us see.

(Refer Slide Time: 06:27)



Now first into the male full time is total is 180 so 90 out of 180 is of 50% male part-time was 10% female fulltime was 180 out of 180 it is only 50% but female part-time it was 35% so now once we have the percentages what we do next is we went for a sample of 50 we take 50% of that is 20 right should b3e male full time 10 individuals 10% that is it should be male part-time which is of 18 there 5% that is two individuals female fulltime and 35 that is 15 individuals must be stratified.

In stratified sampling the biggest advantage of stratified sampling is the stratified sampling helps you to bring in you know involve all the strata's into the study right you are not omitting any study any part of the study so finally this is a seen okay. But let me tell you stratified sampling which I have list out in the slide is that when you make a stratified sampling please keep in mind do not take to many variables the maximum number of variables viewers one should consider generally we consider the demographic variable so we should be a variable only one or two.

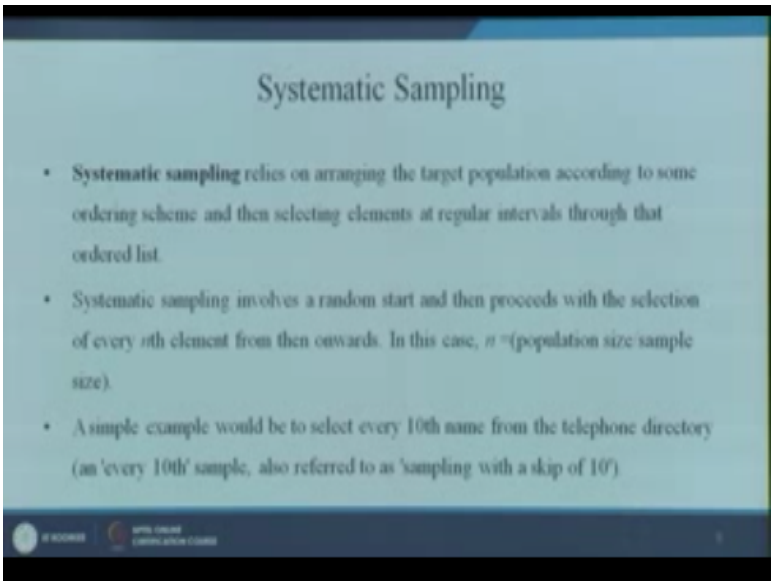
It is not the you cannot do more than that but it should take more than one or two variables to create a strata's then what happens is it becomes highly complex okay so should divide and there other thing is the number of strata to make should not be more than six ideally it should be less. Then six strata's okay so for example four five and the sixth okay so we can say male let's say Indian male American female Indian female and American.

So let us say a male let us say a Italian female so what I am doing is if I am dividing the population I should divide the population maximum up to six strata's because the whole

objective of the stratified sampling is to equally divide into the equal numbers so that you did not have the spend too much of time and effect in getting your results or you know proper finalists so that it is it will be at the object will be lost so let us not take more than two very few variables.

Maximum and six strata's maximum okay so this is what to remember so after we are done with the stratified with the sampling you can understand now I hope stratified sampling is a very useful sampling it can be used in most useful research studies right so now let's say the next system it is resembling the name suggests.

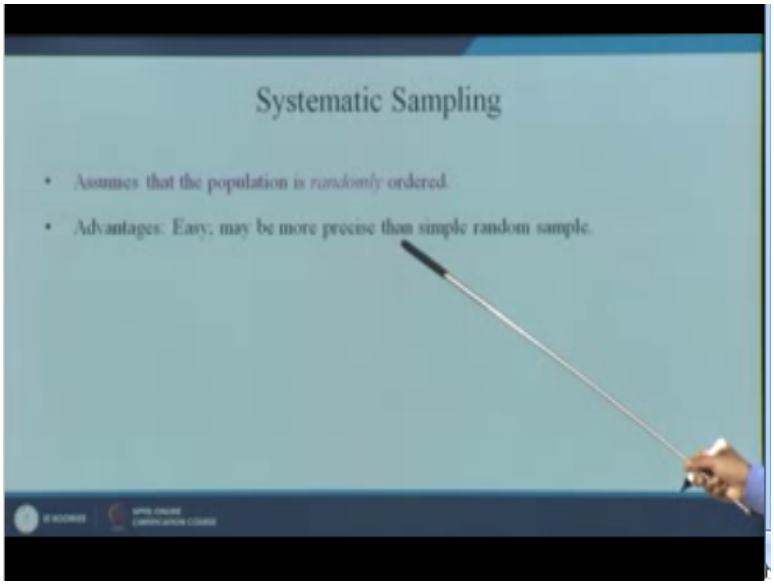
(Refer Slide Time: 09:34)



Systematic sampling there is a system so something is between the systematic sampling you say what is it relates anything with some modeling's team right so for example in this case let's say there are six strata's example let's consider you say. Now I want three samples out of it so what I can do is I can either take like every second alternatives so first I will take this one then I will leave this I will leave this I will take this one so I have got three samples I say similarly you must see people come from a any census study if they are going to systematic sampling with a colony what they do is divide the colony into how many people's you want to stay out of hundred one ten or twenty samples right.

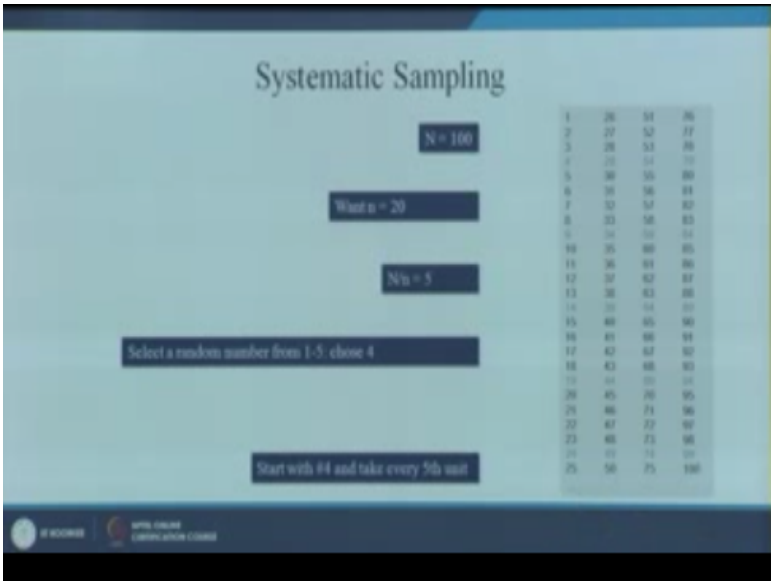
We have similar example so you just have to divide that every 20 suppose you need that model 20 samples you require is 20 so if you require 20 out of 100 so every fifth element becomes your choice right let's see this. So the population is randomly ordered.

(Refer Slide Time: 10:51)



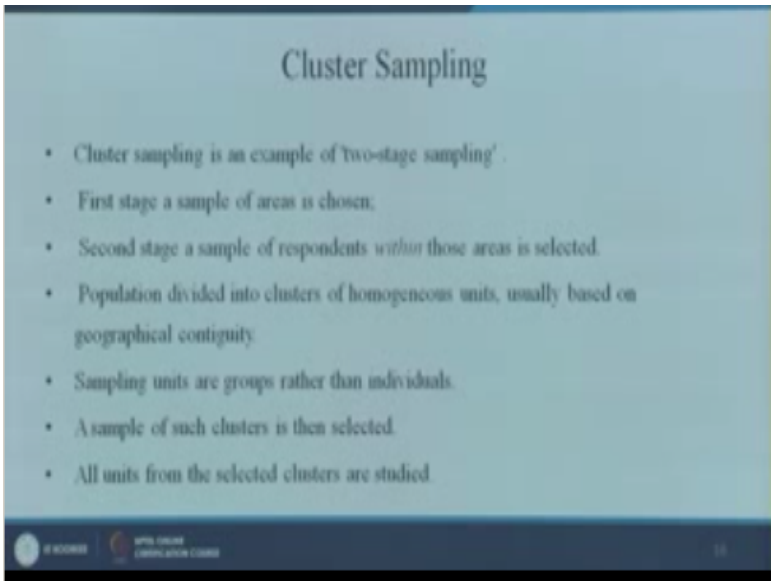
Precise okay actually this generally comes after the systematic simple random sampling right it is the issue if you asked me about how to take the valuation second one right.

(Refer Slide Time: 11:03)



So let us take this case now 100 total number of respondents are there in the library the study wants to be twenty out of respondent should be pulled out as a sample study this population is 100 right so if you do that every fifth element has to be selected now how it can be done now let us see if you go to this so fourth ninth fourth fifteenth nineteenth 24 29 every five gap so they sleeted 50 so this is the way the systematic sampling is done right. The way the systematic sampling is done right.

(Refer Slide Time: 11:40)



The last sampling that is called the cluster sampling right under the provistic this is the last one so in cluster sampling model what is happening there are cluster sampling is like area sampling basically so what is the basically a cluster cluster is something like this is a cluster okay

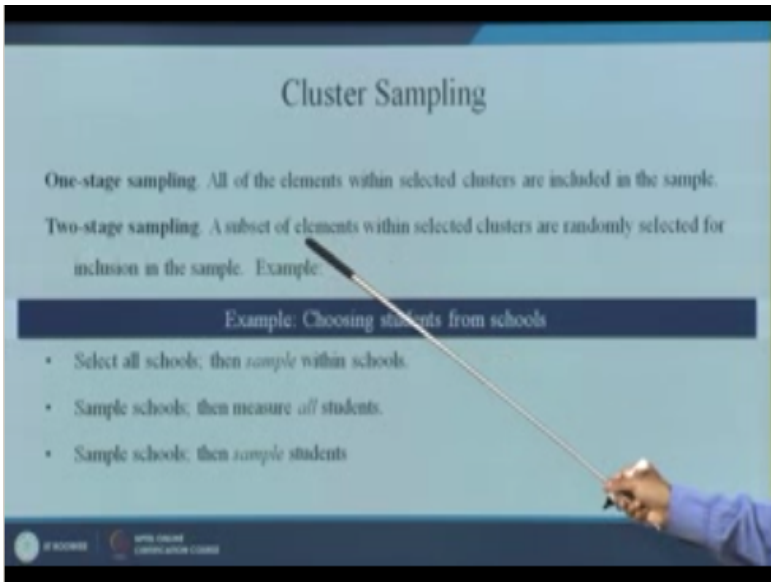
(Refer Slide Time: 11:59)

These are the clusters there are the data points are the repondence are very similar in nature okay are very similar in which this is cluster A this is cluster B as assumed so what are you saying it is like a two stage sampling it says the first stage is sample which has chosen as I said as a area sampling. So in the second stage the repondence within those areas are selected in a prpoblastic manner okay so the population is evaluated into clusters of homogenous units usually based on geographical contiguity or similarity you can understand so sample units are grouped groups rather than.

So someone has to make a study on the Indian populating right the government wants to make some study some understanding about people's behavior why people leave a tax or something right so what it is done it has got several state and now it wants to select randomly it pulls out a state anyone one state must be done randomly and it completely does a study on the population of this state right.

So when you does it is called a one stage right so let's see one stage two stage multiple stage right so in the one stage all the elements within the selected clusters are included so what does it says the elements all the complete the all the elements within the selected clusters could be the state could be the obvious state in the large condition may be the district or the village or the town whatever it is you say are included in the sample. In the two stages sampling what it is doing is.

(Refer Slide Time: 14:15)



A subset of the element is within the selected clusters are randomly selected example is let's see this example so select all schools then sample within the school's first there are several schools so you to call the schools and then you randomly pull out a few schools okay then from the schools you take sample students sop first in the two stage what we did so it is just like what you know the explanation has become too more simpler.

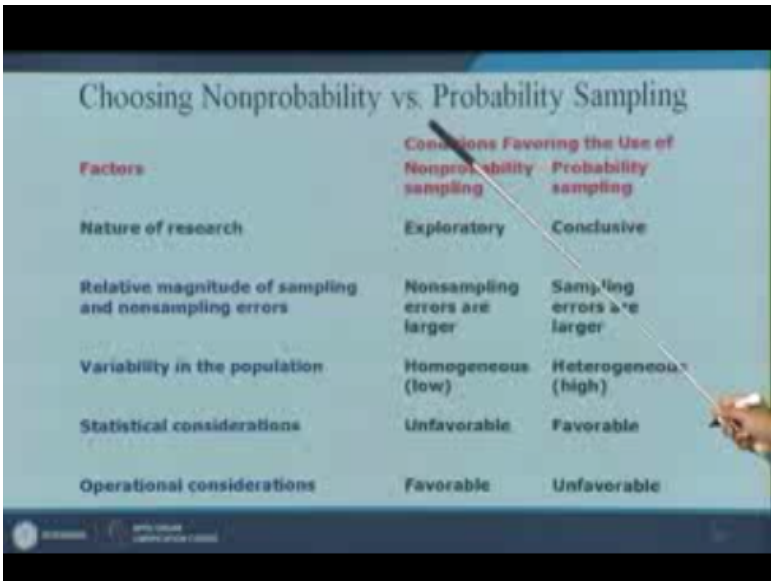
Out of suppose we understand the population is or the study has to be conducted among the students or schools have taken out of which randomly sums of students some schools are selected and after that again randomly from those schools again students are selected randomly right.

So in a two stage way we are doing so this is some I think which we say is a two stage sample in a multi stage only the third stage that goes little deeper right one stage deeper you can understand so sampling is complex from the two or more levels of units are embedded in one and a another so example you say is first stage random number of district chosen in all the states.

So from all the states randomly one from one state and two from other state the distinct are chosen then randomly the villages are chosen after the distinct next so may be might choosing such villages may be you might have missed with only few villages then the third stage is selecting the house is now when this is all ultimate units the houses also selected to the last step as survey.

So this is what is done basically in a multiple stage multiple stage is nothing but the two stage only a advance version of the two stage you are going little more deeper so first stage you selected all the districts or all the states there in all the in some other districts then from some other towns villages and then finally the houses okay. So this is basically the last slide.

(Refer Slide Time: 16:29)



Where we are saying the difference between the nonprobablity and probability sampling so if you see this condition favoring the use of this two so nature of research as I associated exploratory this is conclusive the magnitude f sampling error non sampling errors are which is not related to the sample exactly so here this is when there are larger so if it is non sampling are larger go for a explorative study the sampling errors are larger then you go for a conclusive study non profanity sampling and sampling is larger you go for a probabilistic sampling.

So variability in the populations it is homogenous so the when there is low differences there is a small differences in the variability it is good to go for a non probability if it is high if the samples are different from each other you should never go for non probably sampling so that is what I sais in most of the experiments we use the provistic sampling right.

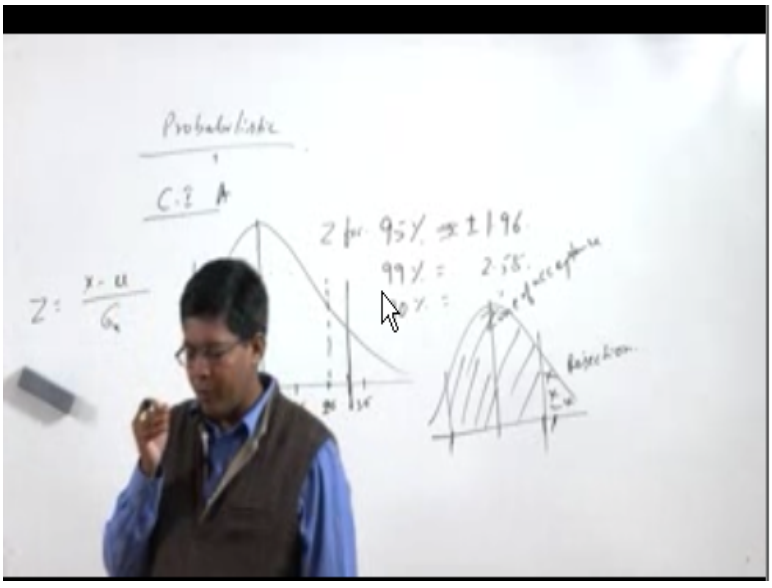
So these are some of the things right then I will come from her to the sample size right okay now once you understood okay what to use first you understood whether you use the non prpoblastic sampling when you right a research paper or when you studying a study it entirely depends not upon you right what kind of as method you want to use s you can select your sample randomly or

select your sample no randomly so if you want if you feel that all the people do not understand this subject then it is better to go for a non prpoblastic method if you feel that everybody has got a equal knowledge in the same subject.

So for example students in my class then I use the prpoblastic method but now the question comes what is the right sample size okay what is the right sample size now how do I know the right sample size is there any method that somebody can tell me because everybody the research students have been come and ask what should be the sample size how many sample should be taken then how do I know you should know about your study.

So there somebody has to be told what is the method let us start with the simple methods so simple method is either you go for an mathematical approach right we say the confidence interval approach right.

(Refer Slide Time: 19:07)



So what is the confidence interval and approach and how it is used so let us understand it is proven normal distribution first let me explain this then I will go the other one so this is the normal distribution so what is happening is we are saying that if you remember which as you do not let us go to the shebisos theorem to understand that okay.

So what is the normal and the bells cowry we say is this is the point where we mean median and mode line right and this is not squid for a moment okay so there are three standard deviation to the right to the left okay so one standard deviation two standard deviation three standard deviation now this standard deviation is nothing but the exact value.

So we say should have seen the test z test distributing it he z value we talk about the t value and the z value is basically that value which is nothing but the standardized standard deviation you can understand it is the standard deviation from the mean that means it is the standardized value right it is the standardized value and it is the unit less.

So what is happening here now suppose for example my confidence interval is suppose for example I am saying let say that where does it lie confidence level of 95 percent at a confidential level confidence level is basically something like the researcher wants to be confident right so if he is confident at 95 percent so there is a chance of error of only 5 percent right which is also termed as the alpha basically okay.

So let me start from beginning may be okay let me explain this otherwise I wil, finish this first so what I am doing here is so when there is 95 percent confid3ence level right the z value is generally +1+2 to -1.96 because the z value for z for 95 percent is equal to + or -1.96 similarly there is three most common intervals that we used are 95% 99% and 90% okay.

So you have three different values for them I think it is 99 2.58 and this I do not remember which we can check the table I think so when this value this what so this are nothing but the standard deviation so 1.96 somewhere here 2.58 is somewhere here so let’s say this to this is where the value falls okay now the question comes okay if I have the 95% interval what I mean so what do I mean I say it is 95% or 90% the z value and for 96 so the area under the cause it falls and this becomes the complete my expectance zone this is my zone of acceptance okay.

And this portion is my deduction zone so if the calculated value if my zee value falls beyond this at a90% level suppose somewhere here then mu null hypothesism is which I will explain that my null hypothesis is rejected let’s assume that so how why it is I am talking about this now this confidence interval method is used to calculate seven the end now how is that now zee is basically equal to calculate as X be the observed value mew be the mean from the standard level so what is the standard error has to be understood.

Standard error is nothing but is please this is very important standard error is the distribution of the standard samples and the deviation it is a distribution of the sample is the simple standard deviation now number of samples are there their sample means the distribution of this sample means gives me the standard error right.

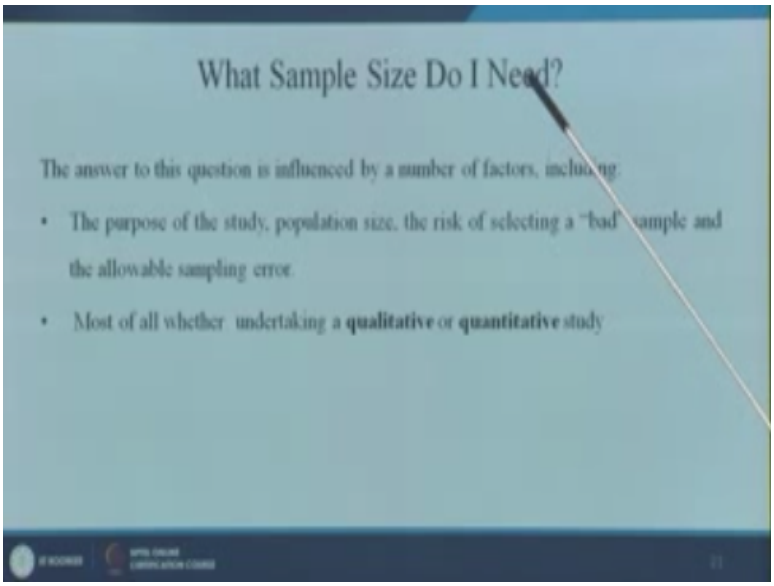
So if the sample error standard error is less or more that will have an effect on the study if it is less then I would assume that this sample are actually representing the population and they are good samples that truly represent the populating but if the samples have a high standard error the distribution is too much dispersed and then we will say this is a problem so lower the standard error better the study higher the standard error more the study okay.

So let us see how do you do this now this standard error is nothing but this standard error sigma x is equal to sigma to the power on route of end but sigma b in the population standard deviation okay now if I have this now can I write the new formula the new formula will look like that means z zee is equal to some people say z some people say zee so upon sigma upon route over of n so if I need n so that will this will come so that means sigma by route over n is equal to in mew upon z.

Or this is equal to let's say it is equal to this implies a route over of n it is equal to zee sigma upon x-mew right it is a +n is equal to z sigma square by x-mew square or we can say n is equal to x sigma by x-mew whole square right at the top so if you use this method and you can see that means what did I say earlier my standard error is low this standard error will depend upon the number of sample size if your sample size increases what happens to your standard error it goes on decreasing but if you have a smaller sample size.

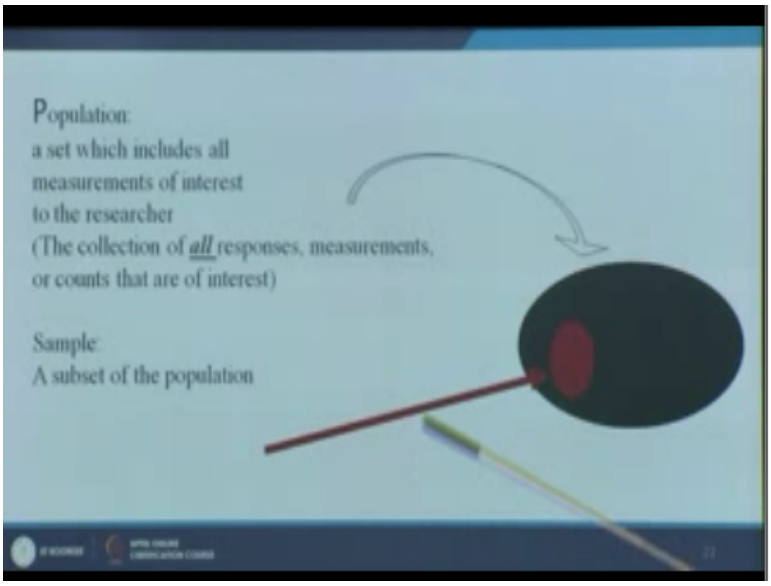
Then your standard error will increase so in order to improve the distribution of the samples standard deviation you should have a higher sample okay right now what is higher how do you find out now this is the formula now it is simply the possession this is the position right so what position level of study on that is basically got the x-mew right this is how you find out the lets go the slide.

(Refer Slide Time: 26:32)



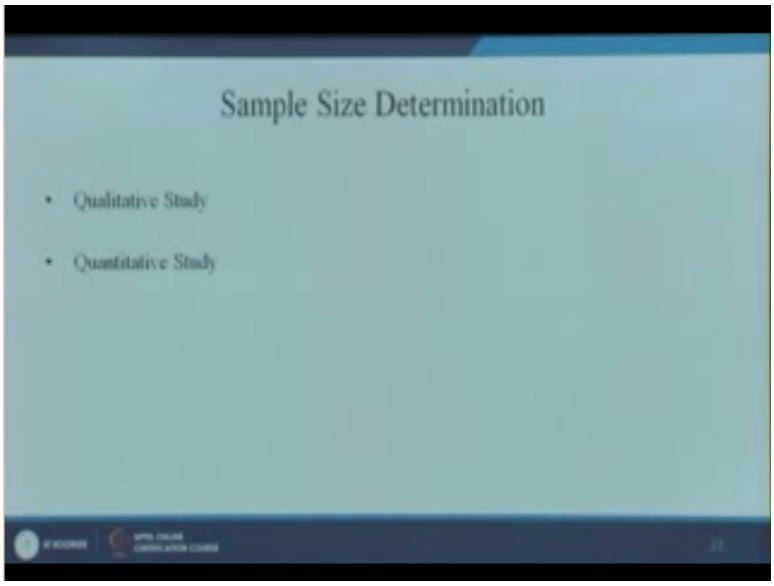
What samples do I need so the risk of selecting the sample and the purpose of selecting bad samples and allowable sampling error whether it is a qualitative or quantitative study that also has an effect now.

(Refer Slide Time: 26; 48)



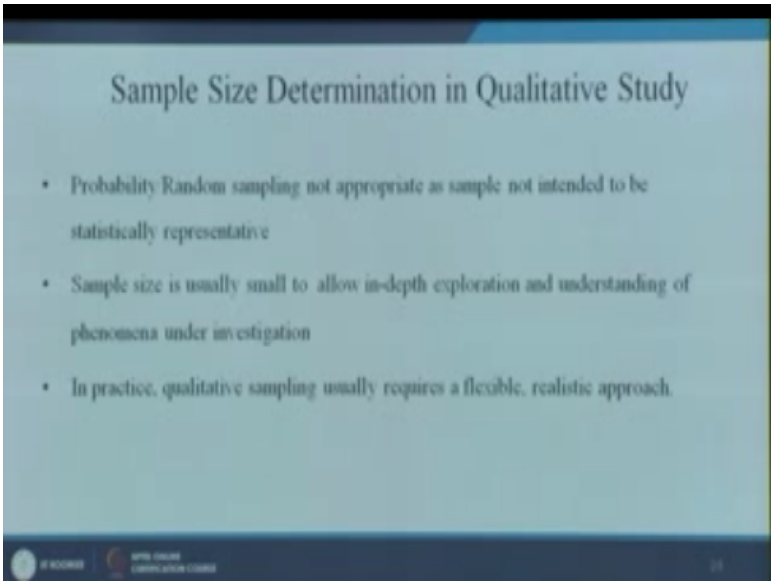
As I said the population this is the sample right as I have said.

(Refer Slide Time: 26:52)



For example you see.

(Refer Slide Time: 26:55)



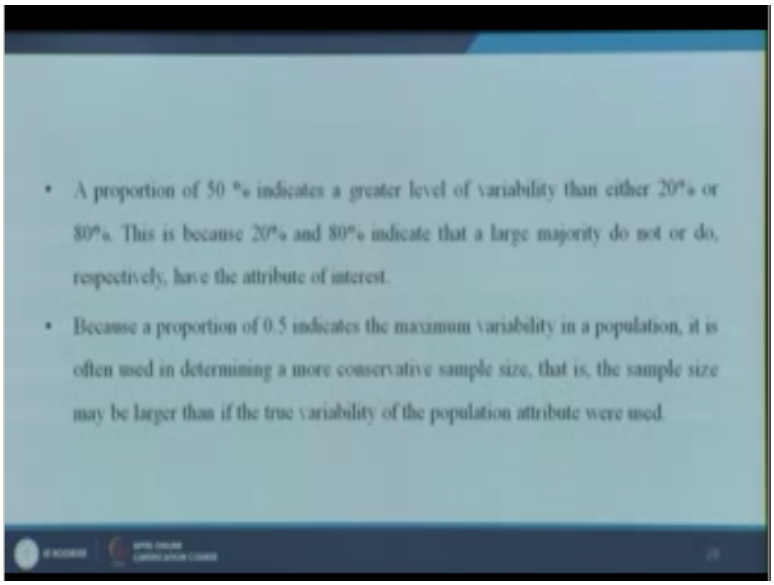
Sample size determination qualitative study you require in depth exploration method of understanding so there a large samples are make a meaning so when you are doing a focus group you know you need not to hundred focus groups you will be need 6 to 5 or 6 7 something like this or around that so if you do induct study which is in qualitative study so you have number of samples to be less.

So right or the number of n is less there right so in a quantitative study we determine the samples as level of precision the level of confidence or risk what level of confidence 100 or 99% 90% so it depends right degree of variability in the tributes been measured so what is the how frequently the variable would change right what is the behavior of the variable that we are using that also will determine so and then finally external validity now what do you mean by external validity you have remembered that I have explained you in the last session I believe that external variables generalize ability.

That means if you have conducted a study this study should be able to you should be able to generalize study across so tomorrow if is say somebody else have started a study you will be able to refer or your study can be used in some other place also so the level of precision is a true value of the population is to estimate it to be the range of often expressed percentage points okay so what is the level of precision we are seeing exactly how precisely we want right the information confidence level is the probability that a confidence level will include the populating parameter.

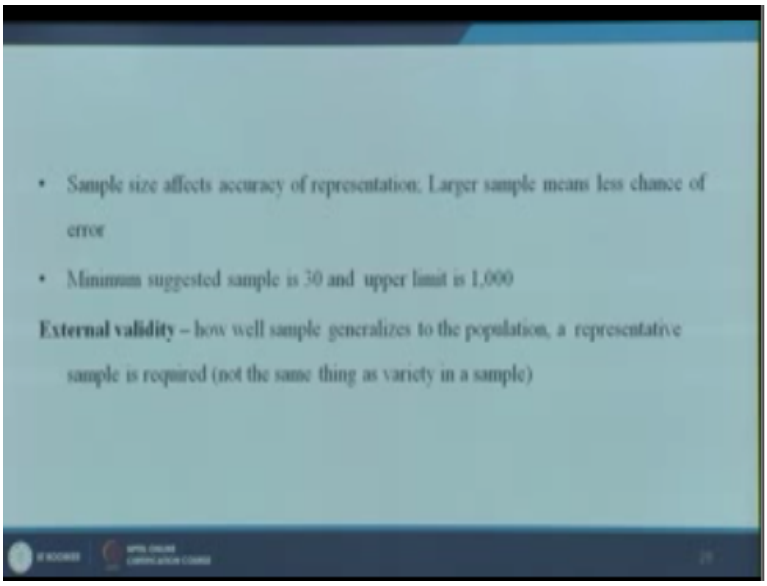
So 95% confidence level is written 95 out of 100 samples will have the true population value in the ratio in the end of the precision right and 5 will not have this five is what we call the alpha or the type one error okay so this is the variability so it refers to the distribution of the tributes the more heterogeneous of the largest sample size is required to obtain a given level of precision okay. So let us take a problem.

(Refer Slide Time: 29:11)



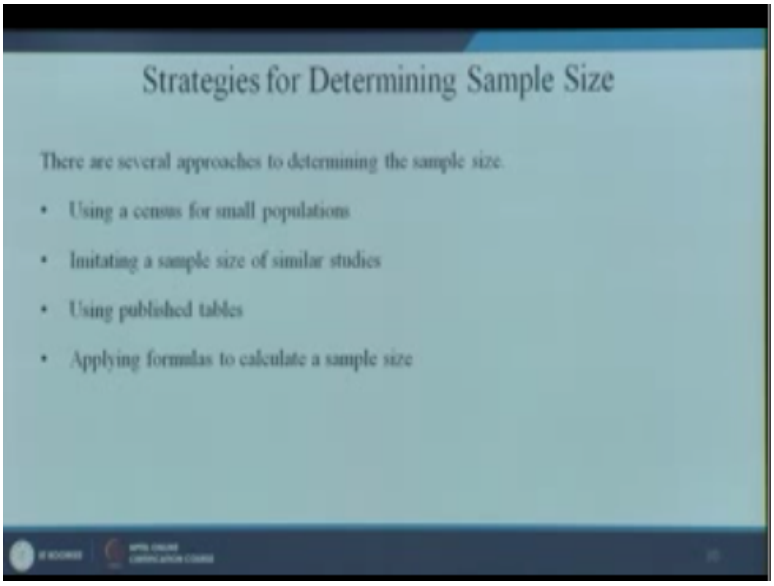
I think I have a problem let us go to this.

(Refer Slide Time: 29:14)



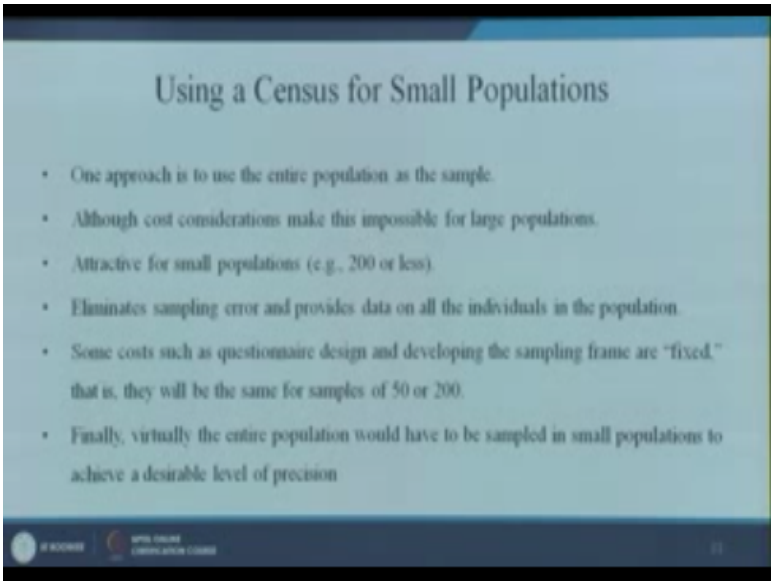
Okay.

(Refer Slide Time: 29:17)



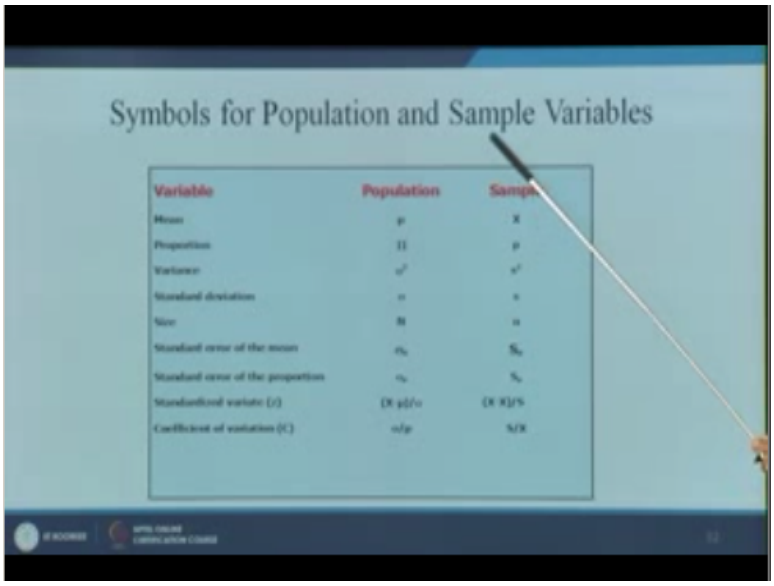
Let us go to this one.

(Refer Slide Time: 29:18)



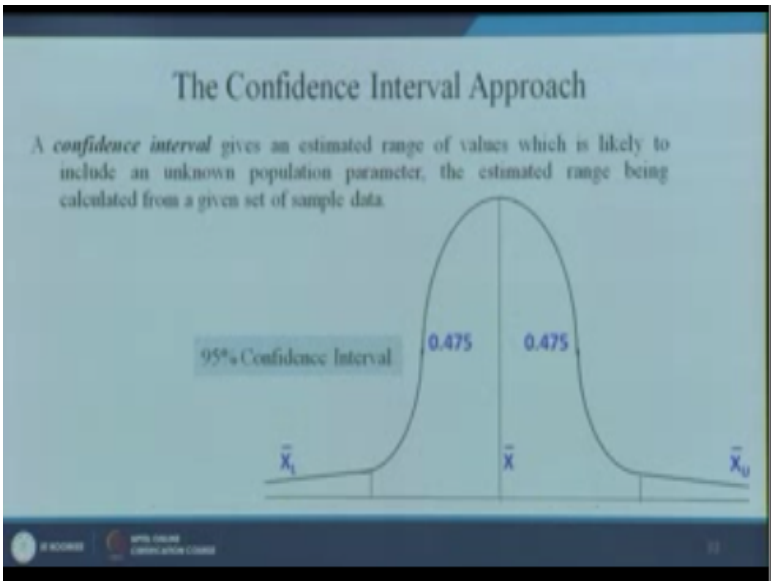
Other side using a census this is not required because the census is something everybody understand very costly thing.

(Refer Slide Time: 29:29)



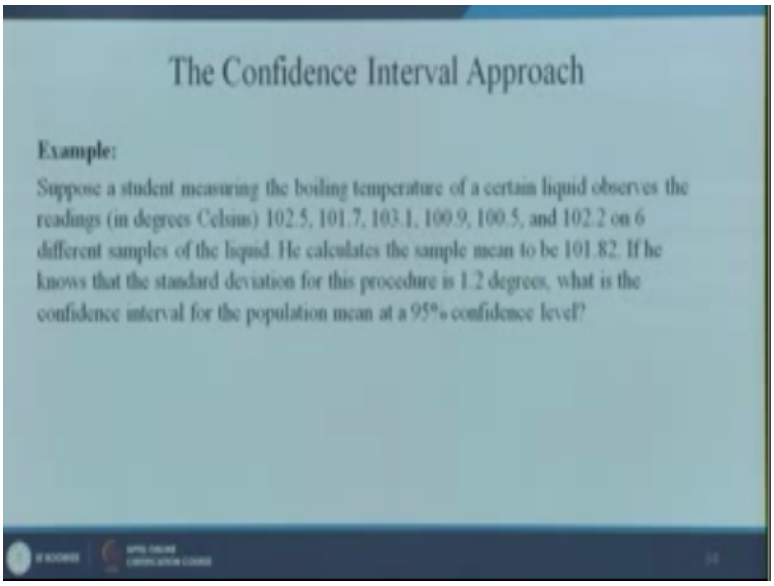
So let us see this are the notations for usual population and sample first of all so in populating in terms of mean we say μπ so these are some of the notifications which you can later on go through it.

(Refer Slide Time: 29:42)



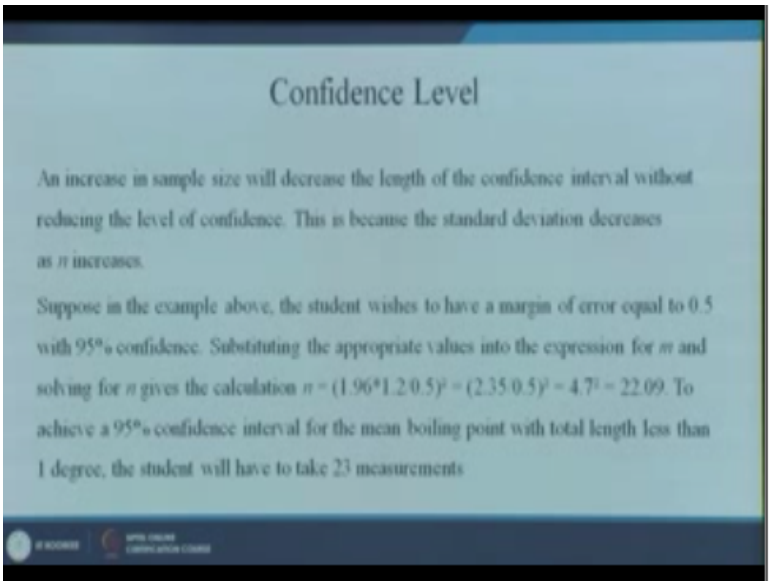
Now this are the confidence interval now this are the confidence level looks like here are the confidence interval looks like the area of the crop if you can see for a 95% confidence level is divide into 95 is divide into .475 because the entire area is 1 okay so this is .475 and the rejection there it is here .025 here .025 it is a two tail test if is a one till test all the points zero for 5 will either lie this side or this side that we will see in the when we will do hypothesis okay. Now let us take.

(Refer Slide Time: 30:11)



The confidence interval approach along for example suppose a student measuring the temperature of a certain liquid observe the variance this okay on six different samples f liquid we calculate the sample to be this much if the standard deviation is known to be 1.2 degree what is the confidence interval for the population men had then 95% confidence level okay.

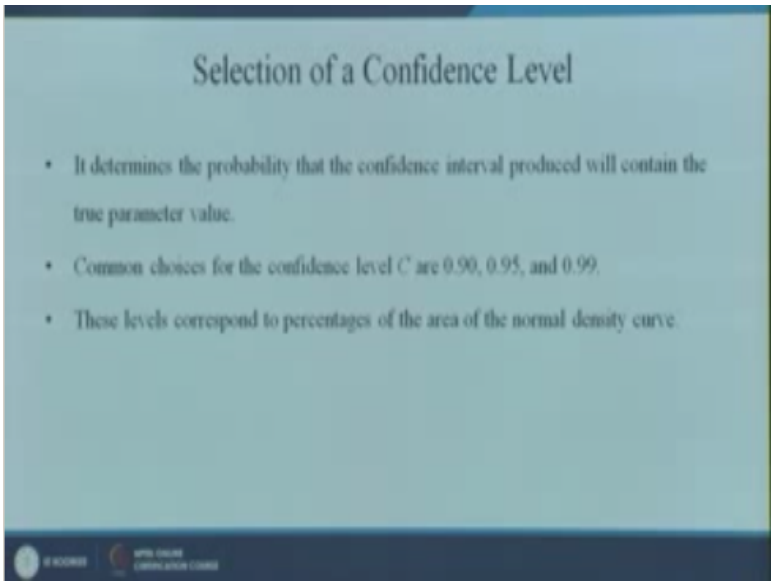
(Refer Slide Time: 30:33)



So basically we are no interested in the sample size t calculate the sample size so to calculate sample size what it is done is it you see n is equal to the same formula which have developed here right so we have used here is 95%.

So 1.95 and this is equal t 22. 09 so achieve A 95% confidence interval right for the mean boiling point the total length then 1 degree the student have to take how many 23 measurements need to be 22.09 bring the 23^{rrd} so is a 23rd measurement right so this is how the you know samples has done.

(Refer Slide Time: 31:13)



Let me tell you to do something the easier way to remember is apart from the calculation that is have in your study ten variables please have a general rule is every variable should we having at least 20 respondents or 20 samples so if you are having 10 variables you should be 100 samples size if you are having 15 variables then we have 300 samples.

So this is the thumb rule that we generally use the number of questions or items or variables whatever your views so multiplied by 20 but all the time it is possible so sometimes we also bring it down it is only 10 per question at 10 per item is also good enough. So if you have a variable 20 even if you collect 200 it is not bad right but going below that is highly objectionable it is questionable when you do a sample size keeps in mind these things.

And then you can calculate the sample size and then go ahead with your studies but if you do a over simple size that is a only a waste of money and time but if you do a poor sample size as I showed you the standard error will increase the sample distribution will have a problem and then the study will be a weak study. So it is always advisable to have a adequate sample size thanks for the moment.

For Further Details Contact

Coordinator. Educational Technology Cell
Indian Institute of Technology Roorkee
Roorkee 247 667
E-Mail Etcellitrke@gmail.com etcell@itr.ernet.in

Website: www.itr.ac.in/centers/ETC. www.nptel.ac.in

Production Team
Sarath Koovery
Mohan Raj. S
Jithin. K
Pankaj saini
Graphics
Binoy. V.P
Camera
Arun. S

Online Editing
Arun.S
Video Editing
Arun.S

NPTEL Coordinator
Prof B.K Gandhi

An Educational Technology Cell
IIT Roorkee Production