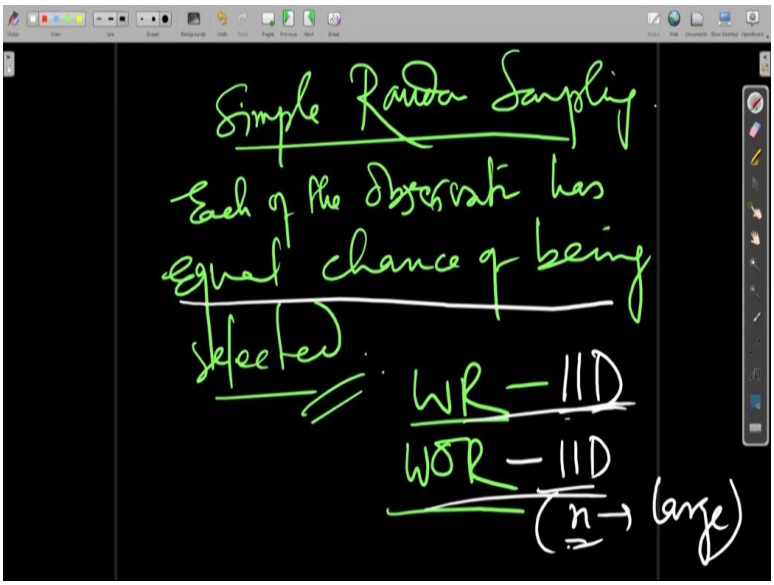
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology - Kharagpur

Module - 3
Lecture - 24
Sampling (Contd.)

Hello and welcome back to the lecture on Applied Econometrics. We are talking about sampling. And in the previous lecture, we discussed in detail about the processes of sampling. And essentially, what you have spoken, all the processes of random number table or RANDBETWEEN function or doing a lottery or using a RAND function; they are essentially dealing with simple random sampling. So, what is a simple random sampling?

And, what is only random sampling, let us say? You know, there are other types of random sampling. So, indeed, there are different types of random sampling and simple random sampling is one of them. And by simple random sampling, what I mean is that all the observations in your population has equal chance of being selected. So, let me write down simple random sampling.

(Refer Slide Time: 01:18)



So, each of the observation has equal chance of being selected. Now, for random sampling, there are other types of random sampling as well, and we may not always have a simple random sampling. So, in simple random sampling, whereas you have to actually ensure that

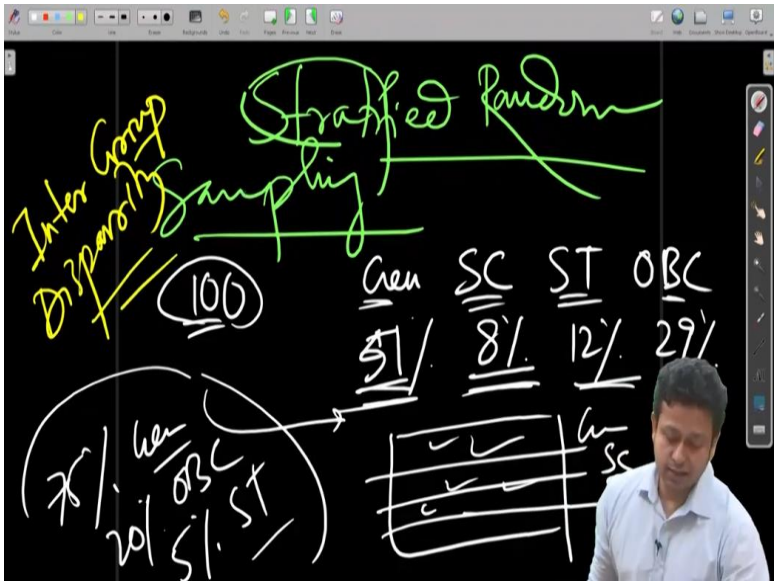
all the observations have a number assigned to them, and only then you can ensure that they have equal chance.

When you put all the numbers in a lottery bowl, then also every individual observation has an equal chance of being selected. Now, when it comes to other types of random sampling, and we will see that things might not just be like that; but in this context of simple random sampling, I would also say, there are little differences of simple random sampling. So, it could be simple random sampling with replacement; it could be simple random sampling without replacement.

And previously, we have spoken about it. So, when we are dealing with a with replacement sample, that becomes essentially IID sample, IID random variable. And here also, we have IID random variable if n is large. If n is significantly large, then what happens is that you can actually consider this like almost IID distributions, because, if you pick out some of the sample, then it really does not matter, it really does not influence the occurrence of the next sample that we draw.

So, essentially, they are independent. So, that is why, we can have with replacement, without replacement. And if the sample size is large, though without replacement also behaves like a IID random variable, similar to what we see with replacement sample, which is IID. Now, simple random sampling has this property, equal chance of being selected, because they are all IID random variables. Now, when we go to some other types of random sampling; let us say stratified random sampling.

(Refer Slide Time: 03:56)



This sounds like a big name, but if we explain what stratified random sampling is, it will basically make sense. So, let us talk about what is a strata in a stratified random sampling. Now, for example, let us say we are talking about a political election here. And we have a Panchayati Raj system; we have this Panchayat, local Panchayats. And there, always we have some sort of reservation.

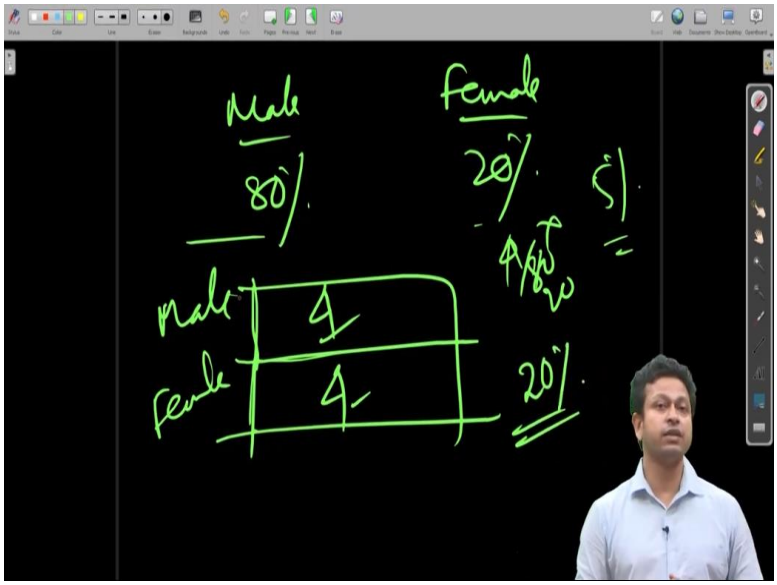
So, let us say, in my Panchayat, there are 10 positions or maybe 100 positions. And these 100 positions are to be distributed among different castes, let us say General Caste, Scheduled Caste, Scheduled Tribe and let us say Other Backward Class. Now, so, let us say there is already something laid down in the constitution. The seats are to be reserved according to their percentage in the population.

So, let us say General Caste has maybe, all over the country, it is like 49; I do not know how much; 51% maybe. And Scheduled Caste; I really do not have any idea; let us say 8%. Let us say this one is 12%, and this one is, let us say 29%. So, that will make 100%. Now, so, essentially, whenever you select a candidate, so, you have to kind of follow this rule. So, you have to have like 51 seats.

If there are 100 candidates, 51 candidates are to be from General Caste; and 8 candidates from Scheduled Caste; 12 candidates from Scheduled Tribe; and 29 candidates from the OBC. Now, let us say, in a particular constituency where you were actually doing that, you have maybe 75% General Caste and maybe 20% OBC; and let us say there are only ST which is 5%.

So, essentially, you just follow this rule to ensure that the reservation policy is sort of followed, but that will not really be representative of that particular constituency. This is one example, how we do stratified random sampling. So, you basically select individuals, just to ensure that this is strata. This strata, we have for General candidate, for SC, for ST or for OBC; you just fill up those strata. So, this is one example. Another example of stratified random sampling.

(Refer Slide Time: 06:47)



So, let us say you have in your MBA class, let us say you have male and female students; and you have 80% male students and 20% female students; of course quite biased towards male; but the administration has decided that at least in the class representative group, they want to see equal representation of male and female. So, how do they do that? So, let us say they have decided that for male and female, there has to be equal representation.

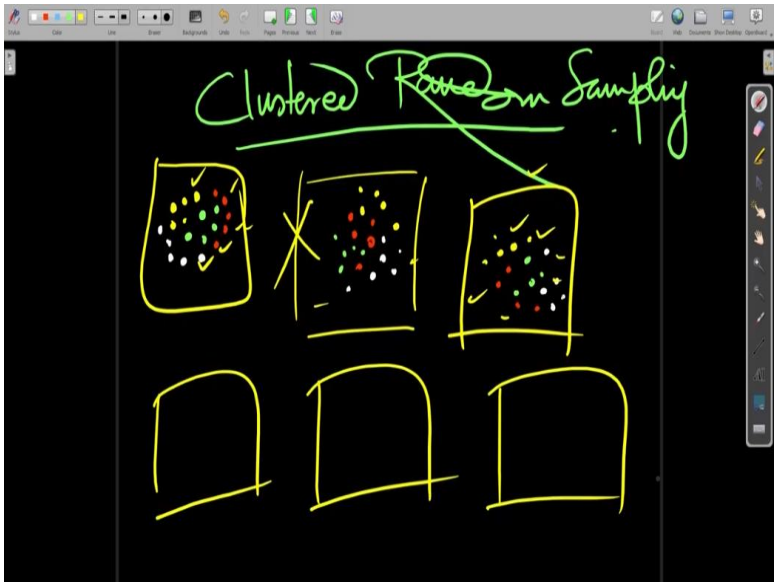
And let us say there has to be maybe 4 from male and 4 from female. So, which means that it will be constituting how much? 4 out of 80; 4 out of 80 is how much? 5%. Whereas, 4 out of 20, if there are 20 students, would be 20%. So, essentially, you can see that, depending on how you want to have your samples, your strata, the way you design your strata and the way you decide to actually select individuals in the strata, have a composition in the strata, your weights could be quite different from what the population is or basically it is not.

So, basically, you are actually putting weight here. And when you are putting weight, so, it is no longer IID, because for IID, mean is going to be constant across the different random variables. So, here, you are not really doing that, because you are assigning weight on your

own. So, stratified random sample, when we say, we are not really following the IID random variable here. So, that is something we need to remember.

There could be another type of sampling; of course, I am talking about a few types, and there could be some other types also, but we will just share with you some of the most important techniques of sampling. So, the third type of random sampling is clustered random sampling.

(Refer Slide Time: 08:42)



Now, what is a clustered random sampling? So, as the name suggests, of course, there are some clusters. And why do we really need cluster random sampling? I think it is always good to explain these kind of concepts with some examples. And one very good example could be Indian village. Let us say you just want to understand something about Indian village. And let us say you want to see the income disparity or something like that in Indian village.

And, you know as an economist that Indian village is a very unique sort of composition. So, what is a composition? So, you see that every village has some pattern repeated. What is that pattern? Let us say the caste people; let us say, yellow is my upper caste; so, they kind of huddle together. And let us say my green is my, say lower caste; they kind of huddle together. Then you have your white. Let us say they are from some different religion; they will huddle together.

Then there are some, let us say; some socioeconomic criteria; there are some very rich people, they might be in another side, they could be some business caste. So, this way, a village will have this kind of composition. And you will see that kind of pattern is being

repeated. So, as such, the villages are, there, the different villages are more similar in nature, like one village is more like the other village, but the intra-village disparity is quite startling.

So, within a village, you can see lot of disparity. So, then you can have this kind of clustered random sampling. So, I am just drawing few more of these, just to give you an idea that this kind of observations are made about Indian villages, where the disparity that we see is more within the village than between different villages. So, in those cases, what we do is, we actually select some of the clusters.

Let us say, this one and this one, we have selected. And once we select the clusters, we actually do the survey for each and every individual in that cluster, because we are more interested to understand the intra-village disparity here. So, that is why we select each of these individual unit, and then we look into every individual within that cluster. So, it really does not matter, does not really make a lot of difference if I include all the villages, because these villages are kind of same.

So, this is basically the cluster random sampling. And whereas, we used the stratified random sampling, there we sort of had intergroup disparity. So, when we had intergroup disparity, like male, female, and then, let us say different caste for reservation. So, then, we use something like stratified random sampling. In the clustered random sampling, it is kind of IID, it is almost IID, because, here what you are doing is, of course, you are selecting individual villages, and then you are selecting all the individuals.

When you are selecting the clusters, so, here, there you are basically doing the same thing as you have done for simple random sampling; you give them numbers, then you select them randomly. So, it behaves like an IID random variable. So, that is basically the idea of sampling. And we dedicated this time just to have an idea how the different sampling techniques are.

And we are not going to touch upon clustered random sampling or a stratified random sampling until in the third module, when we talk about randomisation; but till that time, whenever we are dealing with any econometric problem, we are just going to talk about simple random sampling unless otherwise mentioned. So, we will kind of assume that all the

random variables are IID and it follows, the sample is actually representative of the population and so forth.

So, with this, I will end the lecture on sampling. And going forward, we are going to talk about the two very important natural laws; one is central limit theorem and the other is law of large number. Thank you.