

This tutorial is going to talk about multi-stage sampling. Now, multi-stage sampling is a very, very, very common sampling procedure when the population is very, very large. So for instance, suppose that you wanted to sample from the entire United States as a whole.

An SRS would be completely out of the question. It would be way too difficult to do. The reason it would be too difficult is because you'd have to somehow account for every person in the United States, and maybe assign them a number, and pull numbers out of a hat, or use some kind of random sampling procedure. And that would be too difficult to assign to everyone.

How about a stratified random sample stratifying by state? Well, the strata in that case are still too big where you would take a few people from Maine, and a few people from Minnesota, and a few people from North Dakota. And it would still be too large, and it really wouldn't be cost effective if you were actually going to perform this.

Now what about a cluster sample? The clusters are too big. If you recall what cluster sampling was, if we called the states clusters, you would randomly select some of the clusters and then sample everyone within that cluster. You'd be sampling entire states. Like everyone in North Carolina would be in the sample if you select North Carolina as a cluster.

So none of those really make any sense. The way out of the box here is a multi-stage design. You can perform a combination of some of these strategies that we've talked about. So first, select the clusters.

In this case, the geographic simplicity makes it so states make the most sense as clusters. So randomly, we selected California, Tennessee, Minnesota, Massachusetts, and Oklahoma. Those were just randomly selected states, five of them.

So what can we do now with those clusters? We're not going to sample everyone within that state. We're only going to sample some.

We're only going to take a random sample. And not every state needs to be represented, as would be the case with the stratified random sample. So for instance, let's take Minnesota specifically.

So what we could do with Minnesota is randomly select counties in Minnesota. So maybe we'll take Carver County, and Marshall County, and maybe a few other ones. And from within those counties, if that's a small enough basis that we can get everyone within the county, then we'll stop.

But if we need to, we can keep going. And maybe within Carver County here, say we choose to sample towns

within the county. So randomly within the county of Carver here, there's Hollywood, Watertown, Waconia, Hamburg, Cologne, East Union, Chaska, and Chanhassen.

So maybe we randomly select three of those towns. Chanhassen, Waconia, and Chaska. And then maybe from within that, if those are small enough units, then we can stop. Or we can continue on.

And maybe within Chaska, we can sample some neighborhoods. And usually, by the time you get to neighborhoods within a town, it's easy enough to just sort of walk around the neighborhood and get almost everybody within that neighborhood. But the idea is you sort of continue zooming in from larger areas to smaller and smaller areas until you can find the people that you need.

So to recap, multi-stage sampling is used when the population is so big and the groups or strata or clusters so large that it makes more sense to sort of zoom in and take small groups. And so you start with certain clusters, but then you sample within those clusters instead of taking the full cluster. So it combines cluster sampling, stratified designs, and simple random designs.

So the terms that we used here are multi-stage sampling. But we used some of the other ideas of cluster sampling, stratified random sampling, and simple random sampling, each of which came in its own tutorial. So if you want to watch those, you certainly can. Good luck, and we'll see you next time.