

---

There are many ways of collecting your random sample. The method that this tutorial will look at is a cluster sample. With a cluster sample, you start by dividing the population into roughly equal heterogeneous groups. The groups are mixed up. There's all kinds of people contained-- people or units, objects, whatever it is you're studying contained within each thing-- each group. And then, you randomly select a couple of the groups. And then, everyone in that group is selected for the sample.

So if we, for example, broke everyone up into six groups. And then within the groups, things are roughly the same. We would then randomly select a couple of the groups. So let's say we select two of the groups randomly and then interview everyone within that group or observe every subject within that group.

An advantage to doing this is that it's less costly. You don't have to travel all around to talk to every group. You don't have to interview every single person in the population. But a disadvantage is that it could be less precise or less accurate. Here's an example.

Example 1 says that all nurses are assigned to primarily one hospital. Randomly select several hospitals, then interview all the nurses at those hospitals. This is a cluster sample. Because the hospitals have a roughly heterogeneous distribution of nurses by perhaps age or race or gender or length of time worked in a hospital, it's heterogeneously grouped across the hospitals.

And then, we're randomly selecting a couple of those groups, randomly selecting a couple of hospitals. Here, the advantage is that the interviewer can stay at one hospital and interview all of the nurses there instead of having to bounce all around town to get to nurse A at hospital 1 and then the next nurse at hospital 2 and then the next nurse at hospital 3. So this is a really advantageous savings on time and then on cost as well here.

Second example says all houses are randomly assigned to one zip code. Randomly pick several zip codes and survey all the houses in those locations. Again, same set of advantages. The interviewer doesn't have to bounce around to different towns or even different states. He's staying within that one zip code rather than going from place A to place B to place C. He's only making a couple of jumps to visit each of the groups.

But again, the disadvantage is that he might be losing precision or accuracy in not visiting everyone. If the zip code that he first goes to is really, really different from the one next to it that he doesn't end up going to because it wasn't selected, then the opinions of the people in that second place

aren't being represented in the sample, and the sample isn't as accurate or as precise. This has been your tutorial on cluster samples.