POLI 343
Introduction to Political Research

Session 11-Probability Sampling

Lecturer: Prof. A. Essuman-Johnson, Dept. of Political Science
Contact Information: aessuman-johnson@ug.edu.gh

UNIVERSITY OF GHANA
College of Education
School of Continuing and Distance Education
2014/2015 – 2016/2017

godsonug.wordpress.com/blog

Probability Sampling

A **probability sampling** method is any method of sampling that utilizes some form of ***random selection***. In order to have a random selection method, you must set up some process or procedure that assures that the different units in your population have equal probabilities of being chosen. Humans have long practised various forms of random selection, such as picking a name out of a hat, or choosing the short straw. These days, we tend to use computers as the mechanism for generating random numbers as the basis for random selection.

Definition of Terms

Before we explain the various probability methods we have to define some basic terms. These are:

N = the number of cases in the sampling frame

n = the number of cases in the sample

That is it. With those terms defined we can begin to define the different probability sampling methods. These include **simple**, **systematic**, **stratified**, **cluster** or **area random** and **multi-stage** sampling.

Simple Random Sampling

The simplest form of random sampling is called **simple random sampling**. Here's the quick description of simple random sampling:

Objective: To select *n* units out of *N* such that each unit has an equal chance of being selected.

Procedure: Use a table of random numbers, or a lottery
to select the sample.

A somewhat stilted, if accurate, definition.

How to select a Simple Random Sample

Let's assume that we are doing some research with a small service agency that wishes to assess client's views of quality of service over the past year. First, we have to get the sampling frame organized. To accomplish this, we'll go through agency records to identify every client over the past 12 months. If we're lucky, the agency has good accurate computerized records and can quickly produce such a list. Then, we have to actually draw the sample. Decide on the number of clients you would like to have in the final sample.

SelectiŶg a Siŵple RaŶdoŵ Saŵple ;CoŶt'd:

For the sake of the example, let's say you want to select 100 clients to survey and that there were 1000 clients over the past 12 months. Then, the sampling fraction is f = n/N = 100/1000 = 0.10 or 10%. Now, to actually draw the sample, you have several options. You could print off the list of 1000 clients, tear then into separate strips, put the strips in a hat, mix them up real good, close your eyes and pull out the first 100. But this mechanical procedure would be tedious and the quality of the sample would depend on how thoroughly you mixed them up and how randomly you reached in.

SelectiŶg a Siŵple RaŶdoŵ Saŵple ;CoŶt'd�:

Perhaps a better procedure would be to use the kind of ball machine that is popular with many of the state lotteries. You would need three sets of balls numbered 0 to 9, one set for each of the digits from 000 to 999 (if we select 000 we'll call that 1000). Number the list of names from 1 to 1000 and then use the ball machine to select the three digits that selects each person.

SelectiŶg a Siŵple RaŶdoŵ Saŵple ;CoŶt'd⍰:

The obvious disadvantage here is that you need to get the ball machines. For the lottery method, you get all the 1000 names into a box and randomly pick the first 100 names out of the box to form the sample. Its like picking 5 numbers from the lotto drum in the national weekly lottery.

Advantages and Disadvantages of Simple Random Sampling

Simple random sampling is simple to accomplish and is easy to explain to others. Because simple random sampling is a fair way to select a sample, it is reasonable to generalize the results from the sample back to the population.

Simple random sampling is not the most statistically efficient method of sampling and you may, just because of the luck of the draw, not get good representation of subgroups in a population. To deal with these issues, we have to turn to other sampling methods.

Systematic Random Sampling

It is a type of sampling where the units of the population are ordered in some way and randomly select one of the first **k**th units in the ordered list.

Steps needed to achieve a Systematic Random Sample:

- ❖Number the units in the population from 1 to N
- ❖Decide on the n (sample size) that you want or need
- ❖k = N/n = the interval size
- ❖Randomly select an integer between 1 to k then take every kth unit
- ❖All of this will be much clearer with an example. Let's assume that we have a population that only has N=100 people in it and that you want to take a sample of n=20.

Steps needed to achieve a Systematic Random Sample:

- To use systematic sampling, the population must be listed in a random order. The sampling fraction would be f = 20/100 = 20%. in this case, the interval size, k, is equal to N/n = 100/20 = 5.
- Now, select a random integer from 1 to 5. In our example, imagine that you chose 4.
- Now, to select the sample, start with the 4th unit in the list and take every k-th unit (every 5th, because k=5). You would be sampling units 4, 9, 14, 19, and so on to 100 and you would wind up with 20 units in your sample.

Systeŵatic RaŶdoŵ SaŵpliŶg ;CoŶt'd⍰:

- ❖For this to work, it is essential that the units in the population are randomly ordered, at least with respect to the characteristics you are measuring. Why would you ever want to use systematic random sampling? For one thing, it is fairly easy to do.
- ❖You only have to select a single random number to start things off. It may also be more precise than simple random sampling. Finally, in some situations there is simply no easier way to do random sampling. For instance, I once had to do a study that involved sampling from all the books in a library.

Systeŵatic RaŶdoŵ SaŵpliŶg ;CoŶt'd⍰:

❖Once selected, I would have to go to the shelf, locate the book, and record when it last circulated. I knew that I had a fairly good sampling frame in the form of the shelf list (which is a card catalog where the entries are arranged in the order they occur on the shelf). To do a simple random sample, I could have estimated the total number of books and generated random numbers to draw the sample; but how would I find book
#74,329 easily if that is the number I selected? I couldn't very well count the cards until I came to 74,329. Stratifying wouldn't solve that problem either.

Systeŵatic RaŶdoŵ SaŵpliŶg ;CoŶt’d�:

For instance, we could have stratified by card catalog drawer and drawn a simple random sample within each drawer. But we would still be stuck counting cards.

Instead, we did a systematic random sample. we estimated the number of books in the entire collection. Let's imagine it was 100,000. We decided that we wanted to take a sample of 1000 for a sampling fraction of 1000/100,000 = 1%. To get the sampling interval k, we divided N/n = 100,000/1000 = 100. Then we selected a random integer between 1 and 100.

Systeŵatic RaŶdoŵ SaŵpliŶg ;CoŶt'd⍰:

Let's say we got 57. Next we did a little side study to determine how thick a thousand cards are in the card catalog (taking into account the varying ages of the cards). Let's say that on average we found that two cards that were separated by 100 cards were about 0.75 inches apart in the catalog drawer. That information gave me everything we needed to draw the sample. We counted to the 57th by hand and recorded the book information. Then, we took a compass. (Remember those from your high-school math class?

Systeŵatic RaŶdoŵ SaŵpliŶg ;CoŶt’d⍰:

They are the funny little metal instruments with a sharp pin on one end and a pencil on the other that you used to draw circles in geometry class.) Then we set the compass at 0.75", stuck the pin end in at the 57th card and pointed with the pencil end to the next card (approximately 100 books away).

Systeŵatic RaŶdoŵ SaŵpliŶg ;CoŶt'd⍰:

In this way, we approximated selecting the 157th, 257th, 357th and so on. We were able to accomplish the entire selection procedure in very little time using this systematic random sampling approach. We would probably still be there counting cards if we had tried another random sampling method.

Stratified Random Sampling

Stratified Random Sampling, also sometimes called *proportional* or *quota* random sampling, involves dividing your population into homogeneous subgroups and then taking a simple random sample in each subgroup. In more formal terms:

Objective: Divide the population into non-overlapping groups (i.e., *strata*) N_1, N_2, N_3, ... N_i, such that $N_1 + N_2 + N_3 + ... + N_i = N$. Then do a simple random sample of $f = n/N$ in each strata.

Why Stratified Random Sampling?

- First, it assures that you will be able to represent not only the overall population, but also key subgroups of the population, especially small minority groups. If you want to be able to talk about subgroups, this may be the only way to effectively assure you will be able to. If the subgroup is extremely small, you can use different sampling fractions (f) within the different strata to randomly over-sample the small group (although you'll then have to weight the within-group estimates using the sampling fraction whenever you want overall population estimates).

Why Stratified RaŶdoŵ SaŵpliŶg? ;CoŶt’d:

When we use the same sampling fraction within strata we are conducting *proportionate* stratified random sampling. When we use different sampling fractions in the strata, we call this *disproportionate* stratified random sampling.

❖Second, stratified random sampling will generally have more statistical precision than simple random sampling. This will only be true if the strata or groups are homogeneous. If they are, we expect that the variability within-groups is lower than the variability for the population as a whole. Stratified sampling capitalizes on that fact.

Example of Stratified Random Sampling

Let us say that the population of clients for our agency can be divided into three groups: Akan, African-American and Hispanic-American. Furthermore, let's assume that both the Hausas and Nzemas are relatively small minorities of the clientele (10% and 5% respectively). If we just did a simple random sample of n=100 with a sampling fraction of 10%, we would expect by chance alone that we would only get 10 and 5 persons from each of our two smaller groups. And, by chance, we could get fewer than that. If we stratify, we can do better.

Example of Stratified Random Sampling ;CoŶt'd:

First, let us determine how many people we want to have in each group. Let us say we still want to take a sample of 100 from the population of 1000 clients over the past year. But we think that in order to say anything about subgroups we will need at least 25 cases in each group. So, let's sample 50 Akans, 25 Hausas, and 25 Nzemas. We know that 10% of the population, or 100 clients, are African-American. If we randomly sample 25 of these, we have a within-stratum sampling fraction of 25/100 = 25%.

Example of Stratified Random Sampling ;CoŶt'd:

Finally, by subtraction we know that there are 850 Akan clients. Our within-stratum sampling fraction for them is 50/850 = about 5.88%. Because the groups are more homogeneous within-group than across the population as a whole, we can expect greater statistical precision (less variance). Because we stratified, we know we will have enough cases from each group to make meaningful subgroup inferences.

Problem with Stratified Random Sampling

The problem with stratified random sampling methods when we have to sample a population that's disbursed across a wide geographic region is that you will have to cover a lot of ground geographically in order to get to each of the units you sampled. Imagine taking a simple random sample of all the residents of New York State in order to conduct personal interviews. By the luck of the draw you will wind up with respondents who come from all over the state. Your interviewers are going to have a lot of traveling to do.

Cluster/Area Sampling and The Steps

A type of sampling whereby the units of the population are grouped into clusters where one or more clusters are selected at random. Thus, if a cluster is selected, all of the units that form the cluster are included in the sample.

The steps to follow in Cluster Sampling are as follows:

- ❖Divide population into clusters (usually along geographic boundaries)
- ❖Randomly sample clusters
- ❖Measure <u>all</u> units within sampled clusters

Example of Cluster/Area Random Sampling

Let us say that we have to do a survey of town governments that will require us going to the towns personally. If we do a simple random sample state-wide, we will have to cover the entire state geographically. Instead, we decide to do a cluster sampling of five counties. Once these are selected, we go to *every* town government in the five areas. Clearly this strategy will help us to economize on our mileage.

Merits/Demerits of Cluster/Area Random Sampling

- Cluster or area sampling is useful in situations like this, and is done primarily for efficiency of administration. Note also, that we probably do not have to worry about using this approach if we are conducting a mail or telephone survey because it doesn't matter as much (or cost more or raise inefficiency) where we call or send letters to.
- On the other hand, it is not free from error and not also comprehensive.

Multi-Stage Sampling

In most real applied social research, we would use sampling methods that are considerably more complex than these simple variations. The most important principle here is that we can combine the simple methods described earlier in a variety of useful ways that help us address our sampling needs in the most efficient and effective manner possible. When we combine sampling methods, we call this **multi-stage sampling**.

Example of Multi-Stage Sampling

Consider the idea of sampling New York State residents for face-to-face interviews. Clearly, we would want to do some type of cluster sampling as the first stage of the process. We might sample townships or census tracts throughout the state. But in cluster sampling, we would then go on to measure everyone in the clusters we select. Even if we are sampling census tracts, we may not be able to measure ***everyone***. So, we might set up a stratified sampling process within the clusters. In this case, we would have a two-stage sampling process with stratified samples within cluster samples.

Example of Multi-Stage SaŵpliŶg ;CoŶt'd⦆:

Or, consider the problem of sampling students in grade schools. We might begin with a national sample of school districts stratified by economics and educational level. Within selected districts, we might do a simple random sample of schools. Within schools, we might do a simple random sample of classes or grades. And, within classes, we might even do a simple random sample of students. In this case, we have three or four stages in the sampling process and we use both stratified and simple random sampling. By combining different sampling methods we are able to achieve a rich variety of probabilistic sampling methods that can be used in a wide range of social research contexts.

Merits and Demerits of Multi-Stage Sampling

Merits

- It is a good representative of the population.
- Multi-stage sampling is an improvement over the earlier methods.

Demerits

- It is a difficult and complex method of sampling.
- It is a subjective phenomenon.