Session 02

Survey Sampling Methods

Why Is This Important?
Who is in our study and how we sample them are critical aspects of psychological research. The sample that we draw for our study determines the generalizability of our findings. If we obtain a small sample of rare individuals, then our finding cannot apply to a broader population. We want to draw as unbiased a sample as we can from the population of interest.

Population

In most situations, it is impossible to study an entire population. We typically study a subset of people drawn from a larger population and use inferential statistics to make an inference from the sample back to the population. The validity of that inference depends on how representative the sample or subset is of the population from which it is drawn. Our goal, as researchers, is to obtain the most representative sample that we can. Some sampling strategies can get us pretty close to the population, others have problems that might result in a biased sample. As we learn about different sampling strategies, let's use the population drawn below as our starting point.

Sampling method refers to the way that observations are selected from a population to be in the sample for a sample survey.

Population Parameter vs. Sample Statistic

The reason for conducting a sample survey is to estimate the value of some attribute of a population.

- **Population parameter**. A population parameter is the true value of a population attribute.
- **Sample statistic**. A sample statistic is an estimate, based on sample data, of a population parameter.

Consider this example. A public opinion pollster wants to know the percentage of voters that favor a flat-rate income tax. The *actual* percentage of all the voters is a population parameter. The *estimate* of that percentage, based on sample data, is a sample statistic.

The quality of a sample statistic (i.e., accuracy, precision, representativeness) is strongly affected by the way that sample observations are chosen; that is., by the sampling method.

Sampling Frame
where do we start? When we use probability sampling, we begin by defining our population. Once we have done this, we must have some sort of record or directory to use to select individual participants from the target population. The sampling frame is the population as it is defined and available through records.

If our desired population is everyone living in Srilanka, the census that is conducted every 10 years is our best record or directory of the population. Census data would be our sampling frame. If our target population is everyone on a college campus, the campus directory would be our best choice for a sampling frame.

Probability vs. Non-Probability Samples

As a group, sampling methods fall into one of two categories.

- **Probability samples**. With probability sampling methods, each population element has a known (non-zero) chance of being chosen for the sample.

- **Non-probability samples**. With non-probability sampling methods, we do not know the probability that each population element will be chosen, and/or we cannot be sure that each population element has a non-zero chance of being chosen.

  Non-probability sampling methods offer two potential advantages - convenience and cost. The main disadvantage is that non-probability sampling methods do not allow you to estimate the extent to which sample statistics are likely to differ from population parameters. Only probability sampling methods permit that kind of analysis.

Probability Sampling Strategies

- Probability sampling strategies typically use a random or chance process, although there are important exceptions to this rule. Random sampling is a strategy for selecting study participants in which each and every person has an equal and independent chance of being selected. What does it mean to be independent? The researchers select each person for the study separately.
- Let's say you were asked to participate in an experiment, enjoyed it, and told your friends to contact the researcher to volunteer for the study. This would be an example of non-independent sampling. We assume that friends share similar values and by recruiting your friends to be in the study, the sample might represent you and your friends but not the whole population of interest. The "equal chance" and "independent" components of random sampling are what makes us confident that the sample has a reasonable chance of representing the population.

Simple random sampling

Simple random sampling is the most straightforward of the random sampling strategies. We use this strategy when we believe that the population is relatively homogeneous for the characteristic of interest.

For example, let's say you were surveying first-time parents about their attitudes toward mandatory seat belt laws. You might expect that their status as new parents might lead to similar

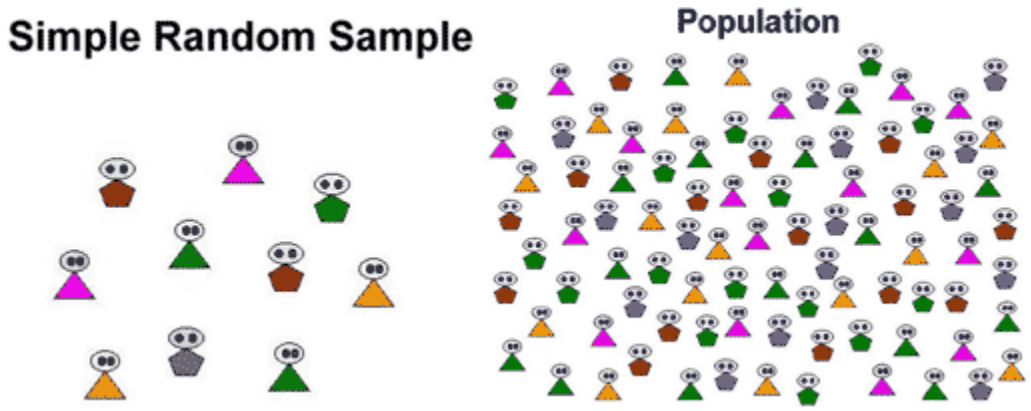
concerns about safety. On campus, those who share a major might also have similar interests and values; we might expect psychology majors to share concerns about access to mental health services on campus.

Procedures

How do we actually take a random sample? The specific procedures that you follow may vary depending on your resources, but all involve some type of random process. If you are using the campus directory as your sampling frame, find a random number table (they are available in the appendix of many research methods and statistics textbooks) or use the random number generator on your scientific calculator. Look at the directory and find out how many pages it has (ours has 265 pages). This directory would require 3 columns from the random number table (or a range of 1-265 on your calculator random number generator).

If your directory had 85 pages, it would require 2 columns (or range of 1-85). Look across the columns (or use the first number generated) to find the "start" page in the directory for the first participant. Once you find the page, use a similar process to select the first participant. Keep selecting pages and participants until you have the desired sample size. Your next step is to contact everyone selected and ask them to participate in your study. If you are using a list generated by the registrar, you can use a spreadsheet program to randomly select your participants from the list.

Let's say we used simple random sampling to select 10 participants from our population. This is what our sample would look like.



Let's compare our sample to the population. How well did we do? Is this a representative sample?

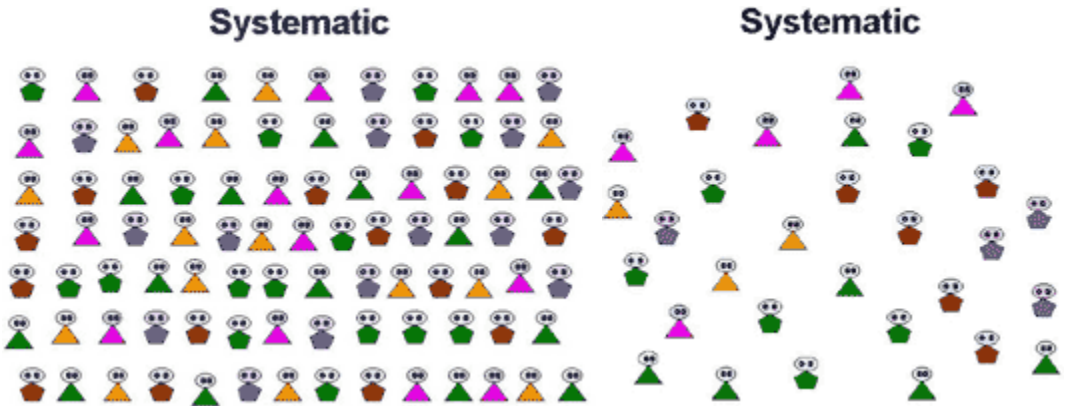
Answer: Simple random sampling seems to have to drawn a pretty good sample. At least one of each type of person is represented in our sample.

Systematic sampling

Systematic sampling yields a probability sample but it is not a random sampling strategy (it is one of our exceptions). Systematic sampling strategies take every nth person from the sampling

frame. For example, you choose a random start page and take every 45th name in the directory until you have the desired sample size. Its major advantage is that it is much less cumbersome to use than the procedures outlined for simple random sampling.

Let's say that we lined up our population into a nice and neat sampling frame and selected every 3rd member. What would our sample look like?



How well did we do? Is this a representative sample?

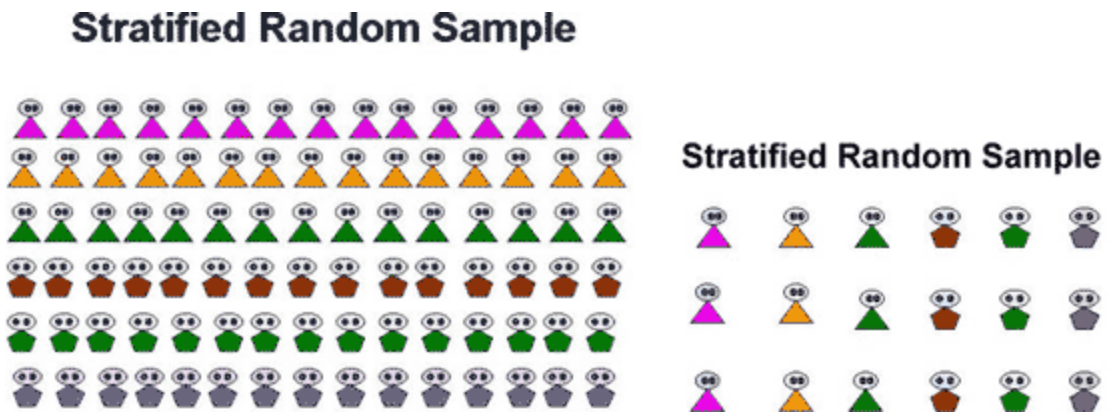
Answer: Now we have a good cross section of the population. But it looks like with this strategy, we ended up with a higher proportion of “reds” than there may be in the population.

Stratified random sampling

Stratified random sampling is used when we have subgroups in our population that are likely to differ substantially in their responses or behavior. This sampling technique treats the population as though it were two or more separate populations and then randomly samples within each.

For example, you are interested in visual-spatial reasoning and previous research suggests that men and women will perform differently on these types of task. So, you divide your sample into male and female members and randomly select equal numbers within each subgroup (or "stratum"). With this technique, you are guaranteed to have enough of each subgroup for meaningful analysis.

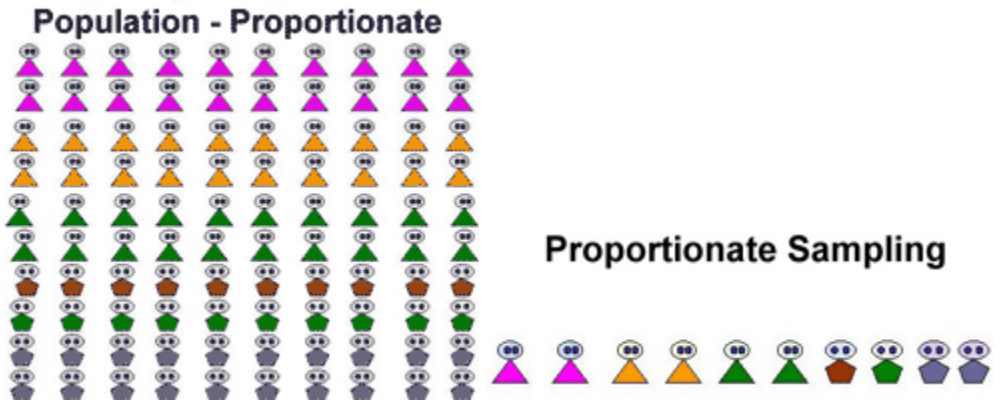
Let's see what a sample might look like if we used this technique with our population.



Proportionate sampling

Proportionate sampling is a variation of stratified random sampling. We use this technique when our subgroups vary dramatically in size in our population. For example, we are interested in risk taking among college students and suspect that risk taking might differ between smokers and nonsmokers. Given increasing societal pressures against smoking, there are many fewer smokers on campus than nonsmokers. Rather than take equal numbers of smokers and nonsmokers, we want each group represented in their proportions in the population.

Proportionate sampling strategies begin by stratifying the population into relevant subgroups and then random sampling within each subgroup. The number of participants that we recruit from each subgroup is equal to their proportion in the population.



Cluster sampling

With cluster sampling, every member of the population is assigned to one, and only one, group. Each group is called a cluster. A sample of clusters is chosen, using a probability method (often simple random sampling). Only individuals within sampled clusters are surveyed.

Cluster sampling is useful when it would be impossible or impractical to identify every person in the sample. Suppose a college does not print a student directory. It would be most practical in this instance

to sample students from classes. Rather than randomly sample 10% of students from each class, which would be a difficult task, randomly sampling every student in 10% of the classes would be easier.

Sampling every student in a class is not a random procedure. However, by randomly selecting the classes, you have a greater probability of capturing a representative sample of the population. Many students believe that it is not possible to gather a representative sample for a class project or a thesis. However, this type of cluster sampling is easily done, especially since all colleges publish lists of classes for registration.

Let's see what cluster sampling for our population looks like.

Note the difference between cluster sampling and stratified sampling. With stratified sampling, the sample includes elements from each stratum. With cluster sampling, in contrast, the sample includes elements only from sampled clusters.

Multistage sampling

With multistage sampling, we select a sample by using combinations of different sampling methods.

For example, in Stage 1, we might use cluster sampling to choose clusters from a population. Then, in Stage 2, we might use simple random sampling to select a subset of elements from each chosen cluster for the final sample.