

1.2 Sampling

Undercoverage = Omitting population members from the sampling frame

Two kind of Error o statistic:

-Sample error

-the population mean will probably be different from your sample mean

-the population percentage will probably be different from your sample percentage

-->Caused by your sample will not perfectly represent the population

-Non-sampling error

-using a bad list

-make sure that you pay careful attention that everyone in the population who is supposed to be represented in your sampling frame is in there.

-->Caused by poor sample design, sloppy data collection, inaccurate measurement instruments, bias in data collection, other problem introduced by the researcher.

Simulation = representing of a real-world phenomenon.

5 type of Sampling

SRS

SRS = Simple Random Sample

One Method of SRS

- Number all of the individuals in the population with a unique number.
 - Like student ID number
- Put all the student ID numbers in a place from which you can draw randomly without looking (like a hat)
- Draw 5 ID's and use those students as your sample.



Another Method of SRS



- Generate a list of random numbers as long as the list of the population.
- Randomly assign these numbers to the population in the list.
- Take the first 5 numbers (whoever gets assigned 1 through 5).

Stratified Sample

-First, the list is divided into groups or strata

this is a way to make it so that there are certain proportions of groups in the final sample.

-Next, the SRS takes place for each of the strata.

Steps in Stratified Sampling

1. Divide entire population into distinct subgroups called strata.
2. The strata are based on a specific characteristic, such as age, income, education level, and so on.
3. All members of a stratum share this specific characteristic.
4. Draw an SRS from each stratum.

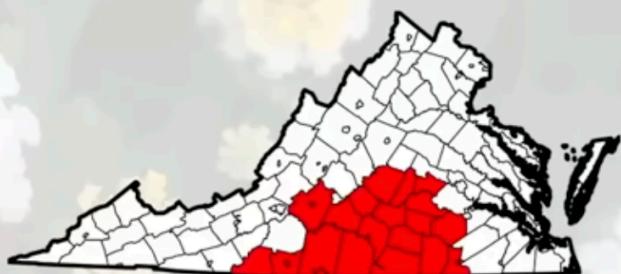
Limitations of Stratified Sampling

- Oversampling one group means your summary statistic is unbalanced
- It is not possible to do without a list beforehand (like with SRS)
- It also is hard because you have to split the list into groups (“strata”) then SRS from the strata

Cluster Sampling

Why use Cluster Sampling?

- The problem is localized to a particular location
- In cluster sampling, we begin by dividing the map in geographic areas.
- Then we randomly pick clusters, or areas, from the map. We take all the people in the cluster.



Problems with Cluster Sampling

- Sometimes, the people located in a cluster are all similar in a way that makes the problem hard to study.
- If cancer rates are high all over the clusters, it's hard to see if a geographic location is causing higher rates.



Photo courtesy of Hansueli Krapf

- Cluster sampling is used when geography is important in sampling.
 - the map is divided into areas, and all the people in a particular area are sampled
 - Biased toward type of people living in the area
- Convenience Sampling

Convenience Sampling

- Convenience sampling can be used under low risk circumstances
- What ice cream is the best from the restaurant next to the hospital?
- However, often results are not reliable



Photo by Managementby

What is Convenience Sampling?

- Using results or data that are conveniently or readily obtained.
- Can be useful if not a lot of resources allocated to the study.
- Use an already-assembled group for surveys.
- Ask patients in the waiting room to fill out a survey, or students in a class.

What are the Problems with Convenience Sampling?

- There is a bias in every group.
- Often miss important subpopulations (what stratified sampling addresses).
- Results can be severely biased



Systematic Sampling

Every member in the population is given a number. After the first member is chosen at random, the remaining members are chosen from a given interval.

SAMPLING METHOD	DESCRIPTION	EXAMPLE
Random sampling (aka simple random sampling)	Gathering a representative sample from a population where each member in the population has an equal chance of being selected.	Using a random number generator to select students in a class to complete a task.
Stratified sampling	Smaller groups or strata within the sample are represented proportionally to the population.	Finding out a favourite soap opera from different age categories of people in a year group.
Systematic sampling	Every member in the population is given a number. After the first member is chosen at random, the remaining members are chosen from a given interval.	A list of people with their first names in alphabetical order are numbered. The 5th person is chosen randomly, followed by every subsequent 8th person.
Non random sampling	Convenience sampling is used for ease of data collection. Volunteers usually collect data.	Asking people at a given location about how long their commute to work is.

SAMPLING METHOD	DESCRIPTION	EXAMPLE
Capture recapture	Collecting a sample of data from one location at different points in time, marking the individuals to estimate a population size.	A sample of woodlice were captured, marked and released. Another sample of woodlice was captured 5 days later and the number of marked woodlice was counted.