

SAMPLING

Categorical (or qualitative) variable - places an individual into one of several sub-groups.

- **nominal variable (by name)**: No natural ranking or ordering of the variable exist. eg hair color
- **ordinal variable (by order)**: Ordering exists. Eg heat (low, medium, high)

Methods of random sampling

1. Simple Random Sampling
2. Systematic Sampling
3. Stratified Random Sampling
4. Cluster Sampling
5. Multistage Sampling

Quantitative (or numerical) variable - measured on a numeric scale, where the arithmetic operations such as adding and averaging make sense.

- **continuous variable** : can take on any value in an interval.
- **discrete variable** : can take on only particular values (eg shoe size)

The **population** in a statistical study is the entire group of individuals about which we want information.

A **sample** is a part of the population from which we actually collect information, using it to draw conclusions about the entire population

A **sampling design** describes exactly how to choose a sample from the population

A numerical value associated with a **population** is called a **parameter**

- population mean
- population standard deviation
- population proportion

A numerical value computed from a **sample** is called a **statistic**

- sample mean
- sample standard deviation
- sample proportion

Simple random sampling

1. All units in population have the same chance of being in the sample, and
2. Every conceivable group of units of the required size has the same chance of being the selected sample.

- Disadvantages: With a simple random sample it is possible (but not very likely) to be biased towards one section of population

- SRS doesn't favor any part of the population

- **Two ways to select a SRS:**

- Using random digits (**table B**)

A **simple random sample (SRS)** of size n consists of n individuals from the population chosen in such a way that every set of n individuals has an equal chance to be the sample actually selected.

Properties: Each entry in the table is equally likely to be any of the 10 digits 0 through 9.

The entries are independent of each other. That is, knowledge of one part of the table gives no information about any other part.

How to use: 1. Give each member of the population a numerical label of the same length.

2. Read from Table B successive groups of digits of the length you used as labels. Your sample contains the individuals whose labels you find in the table.

Ignore any group of digits that was not used as a label or that duplicates a label already in the sample

- Using random number generator (**Excel**)

EXAMPLE: population of 17935. We want a random sample of size 100 ($n=100$).
Using Excel's Data Analysis: "random number generator"

Systematic sampling

- **Systematic sampling** occurs when one individual is randomly selected in a population and then every n th on list there to achieve the desired total sample size
- Use it when no list of population exist or the list is in roughly random order.

Properties

1. Similar to but different from a SRS. If the population is in random order, the systematic sample will be much like an SRS. However, SS **does not** have the property that every sample has the same chance of being the selected sample.
2. Systematic sampling does not necessarily give a representative sample, especially when there is a cycle in the list of a population.

Stratified random sampling

- Need to ensure all sub-groups are represented IF there is likely a difference in the outcome of a study between the subgroups (population is heterogeneous in the question of interest)
- There is better representation when there are subgroups for the feature of interest that are:
 - different to each other (heterogeneous), but
 - similar within each subgroup (homogeneous)
- The sample size must be proportional to the size of each stratum in the population

To select a **stratified random sample**, first classify the population into groups of similar individuals, called strata. Then, choose a separate SRS in each stratum and combine these SRSs to form the full sample.

Procedure:

1. Sort population into the sub-groups of similar units (called strata).
2. Select a simple random sample from each stratum.
3. Combine these samples from the sub-groups to make the total representative sample.
 - Strata might be age groups in a survey of people.
 - Strata do not overlap and they constitute the whole population.

Benefits:

- SRS can be biased towards one section of a population.
- Give more precise estimate for the whole population