⊙ Harnessing Data

15-July-2025
Tuesday

- steps involved in Descriptive Statistics
  1. Collecting the data
  2. Preparing the data (Using matplotlib/Seaborn)
  3. Summarizing the data

- Making sense of the data

A sample which is drawn from the population should have same characteristics as the population

Sampling can be:

- with replacement: a member of the population may be chosen more than once
- without replacement: a member of the population may be chosen only once (lottery ticket)

- Collecting the Data.

step 1: Define the object or aims of the experiment
ie. Estimate the average life of electronic component

step 2: Define the variable and population of interest
ie. usage, power rating, battery life etc

step 3: Defining the data collection scheme and data measuring scheme
ie. sampling Procedure, sample size, data measuring device

step 4: Defining the appropriate descriptive and inferential analysis techniques.

- Methods used to collect data.

Experiment: The investigator controls or modifies the environment and observes the effect on the variable

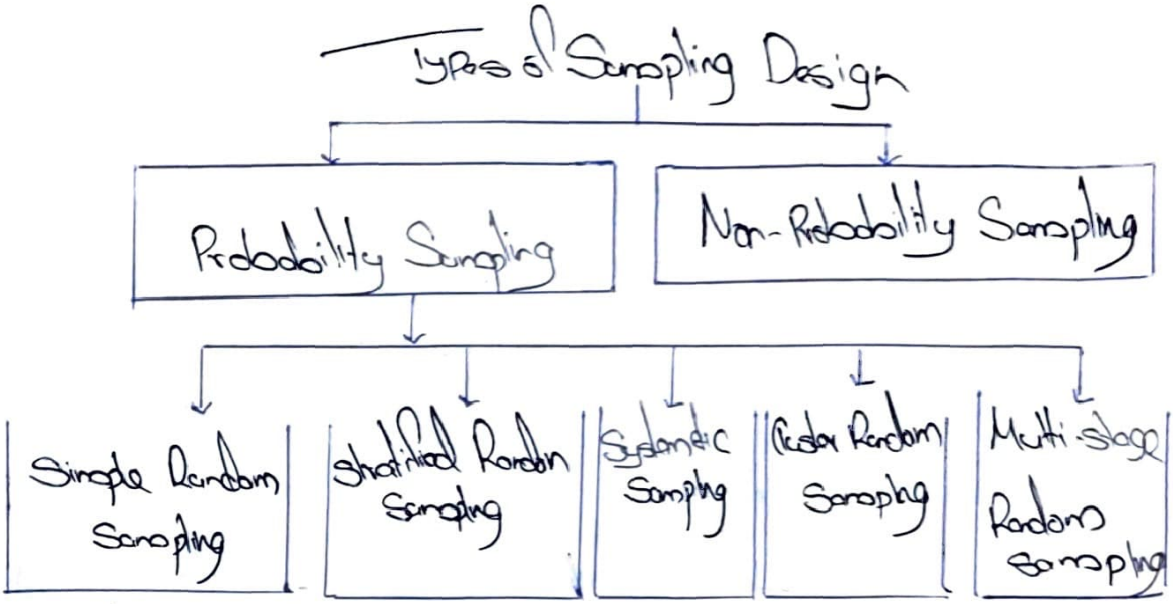
Survey: Data are obtained by sampling some of the population of interest. The investigator does not modify the environment

Census: A 100% Survey, Every element of the population is listed
Seldom used: difficult and time-consuming to compile, and expensive

Judgment Samples: It is a non-probability sampling technique in which the sample members are chosen only on the basis of the researcher's knowledge and judgment

Probability Samples: Sample in which the elements to be selected are drawn on the basis of probability, Each element in a population has a certain probability of being selected as part of the sample.

- Types of Sampling



- Probability Sampling.

1. Simple Random Sampling:

Every member of the population has an equal chance of being selected. Selection is done completely at random.

- ~~Example: You have a list of 1,000 people. You pick every 10th person (10, 20, 30, ...) after starting from a random number, say 5~~
- Example: You have a class of 100 students. You randomly pick 10 names using a random number generator.

2. Systematic Sampling

You select members from a population at regular intervals after choosing a starting point randomly

Example: You have a list of 1,000 people. You pick every 10th person (10, 20, 30)... after starting from a random number, say 5.

3. Stratified Sampling

The population is divided into subgroups (strata) based on shared characteristics, and you randomly sample from each subgroup.

Example: You want to survey employees in a company. You divide them into departments (HR, Tech, Sales) and then randomly select a few from each department

4. Cluster Sampling

The population is divided into clusters usually based on geography or structure, and entire clusters are randomly selected. All individuals in those clusters are surveyed.

Example:

You want to survey students across Kerala. You randomly select 3 schools (clusters) and survey all students in those 3 schools

5. Multistage Sampling.

A combination of sampling methods, applied in multiple stages - often starts with cluster sampling, followed by simple or stratified sampling within the chosen clusters

Example:

Step1: Randomly choose 5 districts in Kenya (cluster)

Step2: From each district, randomly select 3 schools (stage 2)

step3: From each school, randomly choose 10 student to survey (stage 3).

_Non-Probability Sampling methods

1. Convenience sampling.

Pick observers easiest to reach.

2. Judgmental/Purposive sampling.

Pick people based on your judgment or purpose

3. Snowball Sampling.

Existing subjects recruit future subjects

Great for hidden/hard-to-reach populations

Example: You interview drug users

→ they refer you to others

4. Quota Sampling.

You divide your population into groups and choose a fixed number from each group

Ensure representation of groups