

# UNIT 1

Statistical Concepts: Variables, Branch  
of Statistics, Univariate, Bivariate and  
Multivariate data, Sampling

Presented By :

Aditya Joshi

Assistant Professor

Graphic Era Deemed to be University, Dehradun

# Variables

**Traits** (distinguishing quality) or **characteristic** that can change values from case to case.

Example : Age, Gender, Income,

It is a causal relationship

CAUSE-----> EFFECT

Independent Variable----->dependent Variable

It is a characteristic, number or quantity that increase or decrease over time, or takes different values in different situations.


**Independent Variable:** it is a variable that can be controlled or manipulated

**Dependent Variable:** It is a variable that can not be controlled or manipulated.  
Its value are predicted from the independent variables

For example: loan application approval

Independent variables

Dependent variable



The diagram shows a table with 9 columns. A blue bracket above the first eight columns is labeled 'Independent variables'. An orange bracket above the last column is labeled 'Dependent variable'.

Application ID	State	Gender	Age	Marital Status	Own House	Job	Income/month	Approved
1	UK	M	35	Y	Y	Y	100000	Y
2	UK	M	32	N	Y	N	NA	N
3	UK	F	20	N	N	N	NA	N
4	UK	F	50	Y	Y	Y	50000	Y
5	UK	M	39	N	N	Y	10000	N

# Branch of Statistics

There are two main branches of statistics

1. Descriptive Statistics
2. Inferential Statistics

## **Descriptive Statistics : (Describing data) वर्णात्मक**

- Includes the techniques that are used to summarize & describe numerical data for the purpose of easier interpretation.
- Gather, sort, summarize data from sample.
- Organize and summarize data using number & graphs.

## **Tools :**

Data Summary(bar graphs, histogram, pie charts, etc shapes of graph and skewness)

Measure of central tendency (mean/ mode/ median)

Measures of variability (range, variance and standard deviation)

Bar graph	In a form of bars
Histograms	Bars are attached to each other
Pie chart	Based on percentage
Line graphs	In a form of line
Tables	Frequency table, frequency distribution table
skewness	Data is symmetrical, skewed to left, skewed to right
Mean	Mean is the average
Mode	Mode is the number that occurs most frequently
Median	Median is a middle number of a dataset

## **Inferential statistics : (make predictions) अनुमान**

- Using sample data to make an inference or draw a conclusion of the population.
- Uses probability to determine how confident we can be that the conclusions made are correct. (Confidence Intervals, Margins of Error)

**Tools** : Hypothesis test, analysis of variance etc.

# Univariate, Bivariate and Multivariate data and its Analysis

## Univariate Data:

- This type of data consists of **only one variable**.
- The analysis of univariate data is thus the simplest form of analysis, since the information deals with only one quantity that changes.
- It does not deal with causes or relationships and the main purpose of the analysis is to describe the data and find patterns that exist within it.

## Example:

Height of seven students of a class is recorded

Height (in cm)	164	167.3	170	174.2	178	180	186
----------------	-----	-------	-----	-------	-----	-----	-----

There is only one variable i.e. height it is not dealing with any cause or relationship .

The description of patterns can be made by drawing conclusions using **central tendency measures** (mean, median, mode) **dispersion or spread of data** (range, min, max, quartile, standard deviation) by using frequency distribution tables, histograms, pie charts, bar charts.

## **Bivariate data:**

- This type of data involves two different variables.
- This type of data deals with cause and relationships and the analysis is done to find out the relationship among the two variables.

## **Example:**

Temperature and ice-cream sale in summer season

Temp in Celsius	Ice Cream sales
20	2000
25	2500
35	5000
43	7800



- Temperature and Sales are directly proportional to each other because temperature increases to sales also increases.
- It involves comparisons, relationships, causes and explanations, these are **plotted on X and Y axis** of graph for better understanding. One of these variables are independent and other is dependent.

**One more example:** Traffic Accident with weather on a particular day.

## **Multivariate Analysis:**

**When data involves three or more variables**

### **Example :**

- An advertiser wants to compare the popularity of four advertisement on a website, then their click rates could be measured for both men and women and relation between variable can be examined.
- It is similar to bivariate but contains more than one dependent variables. some of the techniques are regression analysis, path analysis, factor analysis and Multivariate analysis of variance (MANOVA)

# Sampling

- It is the process of selecting a sample from the population. For this population is divided into a number of parts called Sampling units.
- The Process of using a small number of items or parts of larger population to make a conclusions about the whole population.

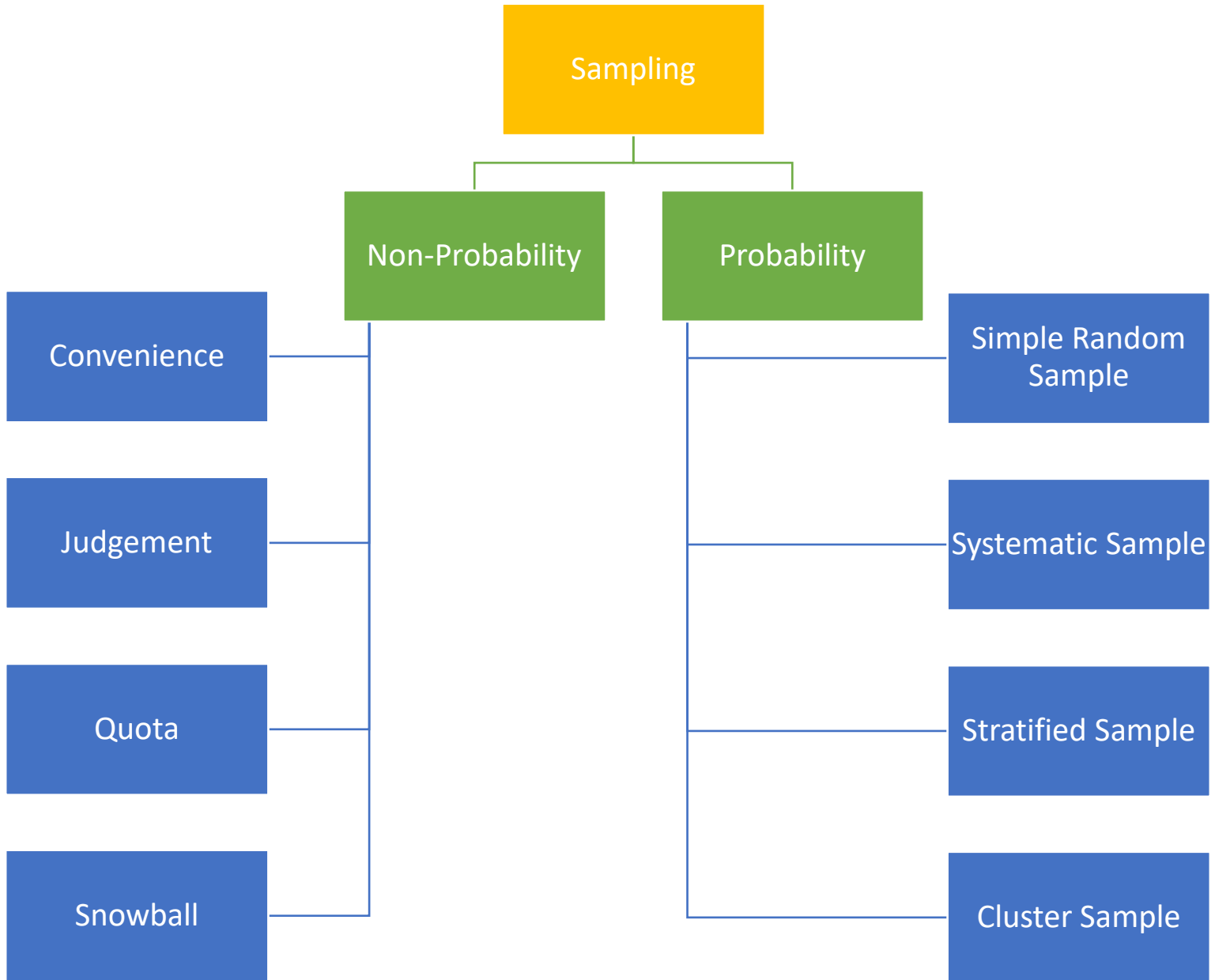
## **Two major categories of sampling :**

### **1. Probability Sampling**

Known, non zero probability for every element, each member of the population has equal chance of being selected, randomness is the element of control.

### **2. Non-Probability Sampling**

Probability of selecting any particular member in population is unknown based on personal judgement.



# Probability Sampling

## 1. Simple Random Sampling

Here all members have the same chance (probability) of being selected. Random method provides an unbiased cross selection of the population.

Example: we wish to draw a sample of 50 students from a population of 400 students. Place all 400 names in a container and draw out 50 names one by one

## 2. Systematic Sampling

Each member of the sample comes after an equal interval from its previous member.

**Example:** for a sample of 50 students the sampling fraction is  $50/400=1/8$

i.e. select one student out of every eight students in the population.

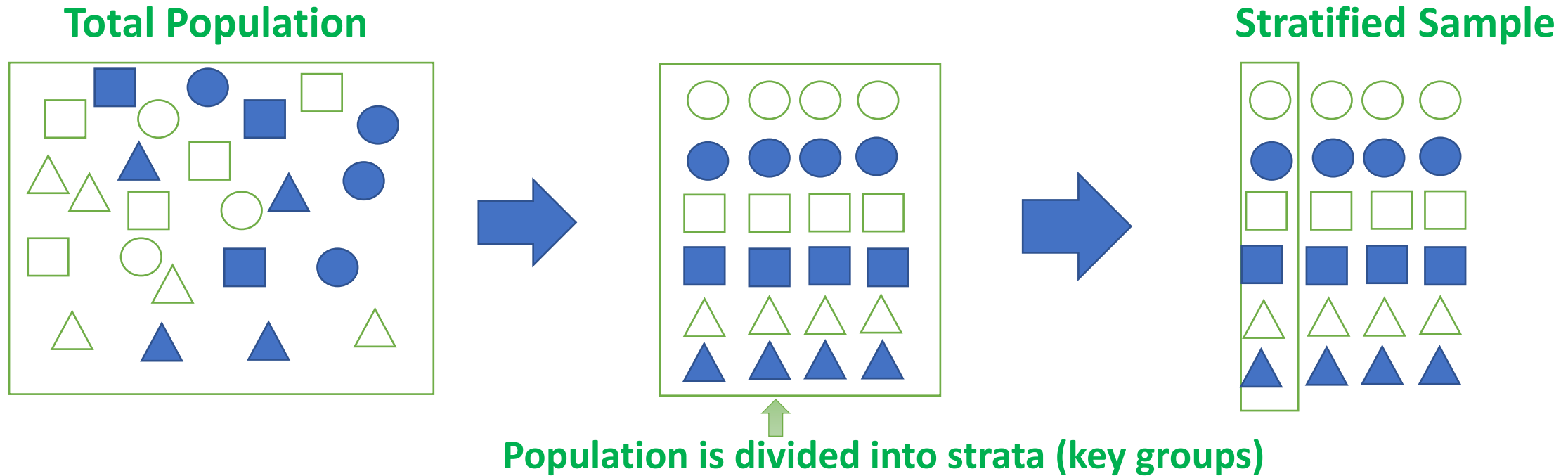
The starting point is chosen at random.



Houses to survey

### 3. Stratified Sampling

The population is divided into smaller homogeneous group or strata by some characteristic and from each of these strata members are selected randomly.



### 4. Cluster Sampling (Area Sampling)

A researcher selects sampling units at random and does complete observation of all units in the group.

**Example:** The study involves Primary Schools select randomly 15 schools. Then study all the children of 15 schools.

In cluster sampling the unit of sampling consists of multiple cases it is also called area sampling.

# Non-Probability Sampling

## 1. Purposive or Judgemental(based on his/her judgement)

In this sampling method the researcher selects a “**typical group**” of individual who might represent the larger population and then collects data from this group.

## 2. Convenience sample:

Obtaining the members who are most conveniently available.

In selecting the incidental sample, the researcher, determine, the required sample size and then simply collects data on that number of individuals who are available easily

Example: people responding to an advertisal invitation “on the street” interview

### **3. Quota Sampling:**

The selection of the sample is made by the researcher, who decides the quotas for selecting sample from specified subgroups of the population.

#### **Example:**

An interviewer might need data from 40 adults and 20 adolescents in order to study students television viewing habits.

Selection will be

- 20 adult man and 20 adult woman.
- 10 adolescent girls and 10 adolescent boys.

## 4. Snowball Sampling:

The researcher identify and selects available respondents who meet the criteria for inclusion. After the data have been collected from the subject, the researcher ask for a referral of other individuals, who would also meet the criteria and represent the population of concern.

**Chain sampling, Chain referral Sampling, Referral sampling**

