CHAPTER 1: BASIC CONCEPTS OF STATISTICS

OBJECTIVES :

After reading this chapter, you should be able to:

- Recognize statistics.
- Explain statistics.
- Differentiate the descriptive and inferential statistics
- Classify sampling techniques.
- Differentiate random sampling and non-random sampling.
- Distinguish levels of measurements

What is Statistics?

STATISTICS is the science of

- planning studies and experiments,

- collecting,

- organizing,

- presenting,

- analyzing,

- interpreting, and

- drawing conclusions based on the data.

LESSON 1.1 Definitions of Terms

• **Collection of data** refers to the process of obtaining information.

• **Organization of data** refers to the determining/ascertaining (after a calculation, investigation, experiment, survey, or study) manner of presenting the data into tables, graphs, or charts so that logical and statistical conclusion can be drawn from the collected measurements.

• **Analysis of data** refers to the process of extracting from the given data relevant information from which numerical description can be formed.

• **Interpretation of data** refers to the task of drawing conclusions from the analyzed data.

BRANCHES OF STATISTICS

• **Descriptive statistics** is the branch of statistics that involves the collection, organization, presentation, summarization or analysis of data.

• **Inferential statistics** is the branch of statistics that involves using a sample to interpret, and draw conclusions based on the data or about a population. A basic tool in the study of inferential statistics is probability. An area of inferential statistics called hypothesis testing is a decision-

making process for evaluating claims about a population, based on information obtained from samples.

LESSON 1.2 Classifications of Variables and Data

CLASSIFICATIONS OF VARIABLES AND DATA
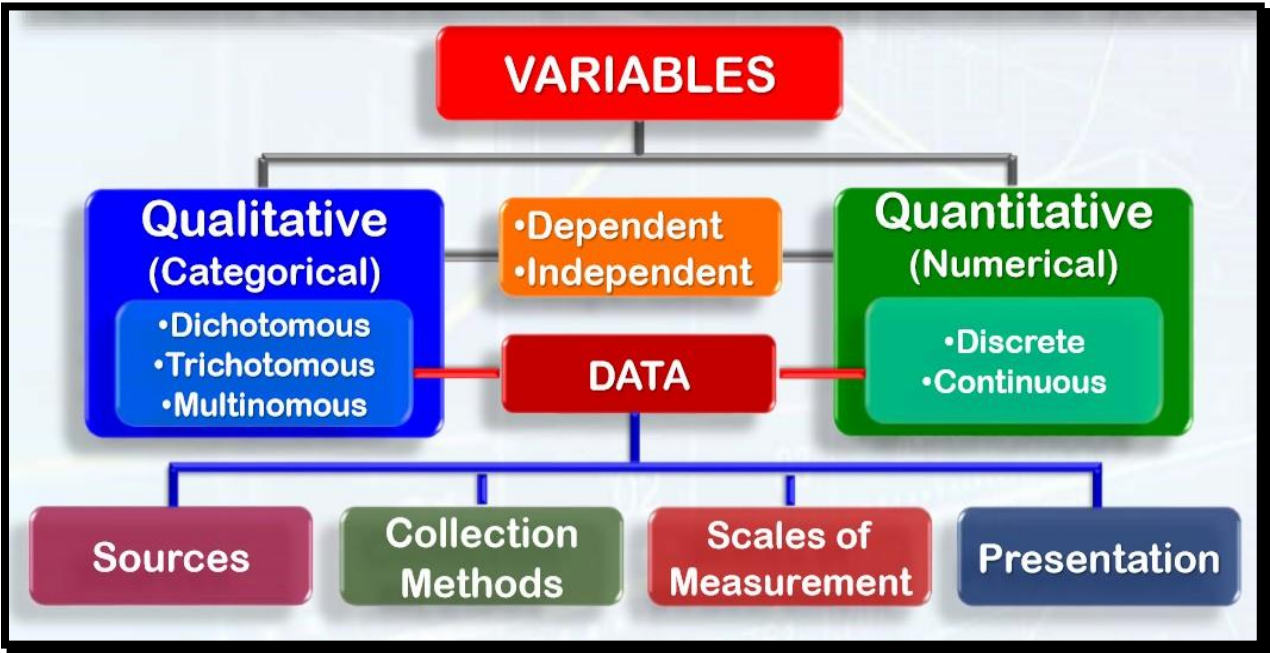
Variable (or Response Variable)

- A characteristic or attribute of interest about each individual element of a population or sample that can assume different values.

Example #1:

A student's age at entrance into college, the color of the student's hair, the student's height, and the student's weight are four variables.

Data

- It is the collection of observations.
- It consists of information coming from observations (realized value of a variable), counts, measurements, or responses.
- It is the set of values collected from the variable from each of the elements that belong to the sample. Once all the data are collected, it is common practice to refer to the set of data as the sample.

Example #2:

The set of 30 heights gathered from 30 students is an example of a set of data.

Data Value

- The value of the variable associated with one element of a population or sample. This value may be a number, a word, or a symbol.

Example #3:

Angelo entered college at age "23," his hair is "brown," he is "71 inches" tall, and he weighs "183 pounds." These four data values are the values for the four variables as applied to Angelo.

Data sets are called **populations** and **samples.**

Population

- Collection of all outcomes, responses, measurements, or counts that are of interest.
- Consists of all subjects (human or otherwise) that are being studied.

Sample

- A sample is a subset, or part, of a population.
- A sample is a group of subjects selected from a population.

Elementary unit or Element

- It is a member of the population whose measurement on the variable of interest is what we wish to examine.
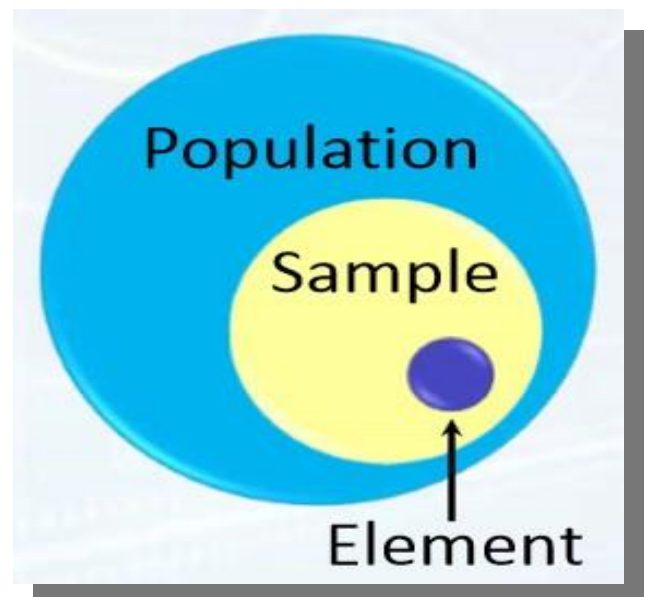
Experiment

- A planned activity whose results yield a set of data.
- An experiment includes the activities for both selecting the elements and obtaining the data values.

EXPERIMENTAL CLASSIFICATION\

A researcher may classify variables according to the function they serve in the experiment.

• **Independent variables** are variables controlled by the experimenter/researcher, and expected to have an effect on the behavior of the subjects. The independent variable is also called explanatory variable.

• **Dependent variable** is some measure of the behavior of subjects and expected to be influenced by the independent variable. The dependent variable is also called outcome variable.

Example #4:

In the sit-up study, the researchers gave the groups two different types of instructions, general and specific. *Hence, the **independent variable** is the type of instruction. The **dependent variable**, then, is the resultant variable, that is, the number of sit-ups each group was able to performafter four days of exercise.*

Parameter

- A numerical description of a population characteristic.
- A numerical value summarizing all the data of an entire population.

Statistic

- A numerical description of a sample characteristic.
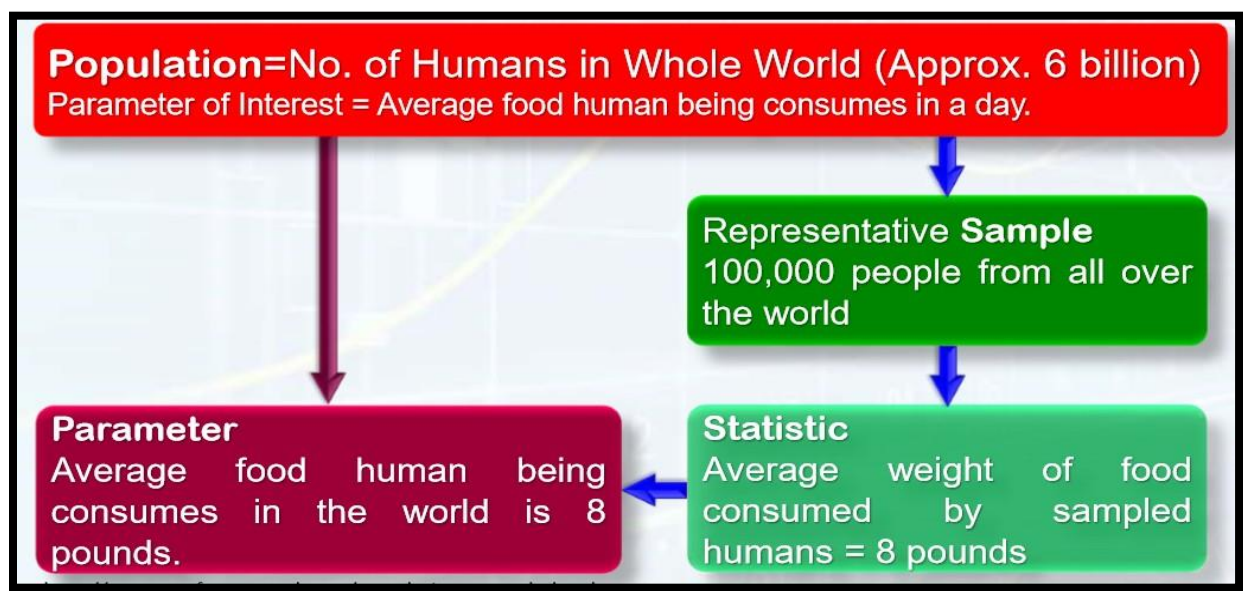- A numerical value summarizing the sample data.

SYMBOLIC NOTATION FOR SAMPLE AND POPULATION MEASURES

Statistical Measure	Sample Statistic	Population Parameter
Mean	$\overline{x}$	μ
Standard deviation	s	σ
Variance	s^2	σ^2
Size	n	N
Proportion	p	π
Correlation	r	ρ

RELATIONSHIPS AMONG PROBABILITY, STATISTICS, POPULATION, AND SAMPLE
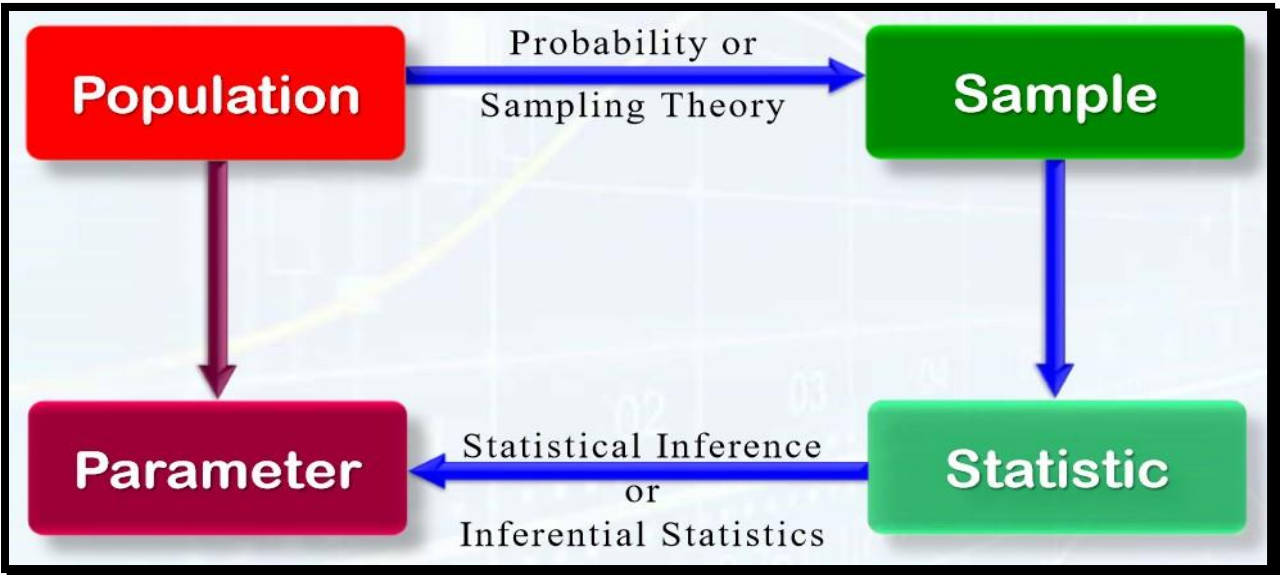
Example #5:

SOURCE OF DATA

Primary data are date documented by the primary source. The data collectors themselves documented this data.

Example #6: census, sample survey, experiment

Secondary data are data documented by a secondary source. An individual/agency, other than the data collectors, documented this data.

Example #7: books, journals, magazines, theses

DATA COLLECTION METHODS

1. SURVEYS

- It is a method of collecting data on the variable of interest by asking people questions. When data came from asking all the people in the population, then the study is called a *census**. On the other hand, when data came from asking a sample of people selected from a well-defined population, then the study is called *sample survey.*

(**Census* or *Registration* requires the enactment of law to take effect for it needs the participation of a large, if not the entire, population.)

DIFFERENT METHODS OF COMMUNICATION

a) Personal Interview - It refers to as the direct method of gathering data since this requires a face-to-face inquiry with the respondent.

b) Self-Administered Questionnaire - It is an inventory of information listed down to which a respondent answers. There is no face-to-face confrontation.

2. OBSERVATION

- It is a method of collecting data on the phenomenon of interest by recording the observations made about the phenomenon as it actually happens.

- It makes use of the different human senses in gathering information.

- It is useful in studying the reactions and behavior of individuals or groups of persons/objects in a given situation or environment as it happens.

3. EXPERIMENTATION

- It is a method of collecting data where there is direct human intervention on the conditions that may affect the values of the variables of interest.

- It is conducted in laboratories where specimens are subjected to some aspects of control to find out cause and effect relationships.

TYPES OF DATA

Qualitative, or Attribute, or Categorical Variable

• consist of attributes, labels, or nonnumerical entries.

• A variable that describes or categorizes an element of a population.

▪ **Dichotomous**

▪ **Trichotomous**

▪ **Multinomous**

Quantitative, or Numerical Variable

• A variable that quantifies an element of a population.

• It consists of numerical measurements or counts and can be ordered or ranked.

DISCRETE VARIABLES

- Assume values that can be counted.

- Can be assigned values such as 0, 1, 2, 3 and are said to be countable.

Example #8:

Examples of discrete variables are the *number of children in a family, the number of students in a classroom, and the number of calls received by a switchboard operator each day for a month.*

CONTINUOUS VARIABLES

- Can assume an infinite number of values in an interval between any two specific values. They are obtained by measuring. They often include fractions and decimals.

Example #9:

Variable	Recorded value	Boundaries
Length	15 centimeters (cm)	14.5–15.5 cm
Temperature	86 degrees Fahrenheit (°F)	85.5–86.5°F
Time	0.43 second (sec)	0.425–0.435 sec
Mass	1.6 grams (g)	1.55–1.65 g

SCALE OF MEASUREMENTS

Measurement - It is the process of determining the value or label of the variable based on what has been observed.

Nominal Level of Measurement

    • Data are qualitative only.
    • Data at this level are categorized using names, labels, or qualities. No mathematical computations can be made at this level.

Example #10 (on the right side)

Grade (A, B, C, D, F)
Judging (first place, second place, etc.)
Rating scale (poor, good, excellent)
Ranking of tennis players

Ordinal Level of Measurement

    - Data are qualitative or quantitative.

    - Data at this level can be arranged in order, or ranked, but differences between data entries are not meaningful.

Example #11: (on the right side)

NATscore
IQ
Temperature

Interval Level of Measurement

    - Data can be ordered, and meaningful differences between data entries can be calculated.

    - At the interval level, a zero entry simply represents a position on a scale; the entry is not an inherent zero. Note: An inherent zero is a zero that implies "none."

Example #12: (on the right side)

Height
Weight
Time
Salary
Age

Ratio level of measurement

    - Data are similar to data at the interval level, with the added property that a zero entry is an inherent zero.

    - A ratio of two data values can be formed so that one data value can be meaningfully expressed as a multiple of another.

Example #13: (on the right side)

Zip code
Gender (male, female)
Eye color (blue, brown, green, hazel)
Political affiliation
Religious affiliation
Major field (mathematics, computers, etc.)
Nationality

Level of measurement has all of the following properties:

a) The numbers in the system are used to classify a person/object into distinct, nonoverlapping, and complete/exhaustive categories.

b) The system arranges categories according to magnitude/degree.

c) The system has a fixed unit of measurement representing a set of size throughout the scale; and

d) The system has an absolute zero.

• **Ratio** level of measurement satisfies a, b, c, and d

• **Interval** level of measurement satisfies only a, b, and c

• **Ordinal** level of measurement satisfies only a, and b

• **Nominal** level of measurement satisfies only a

TYPES OF DATA AND MEASUREMENT SCALES
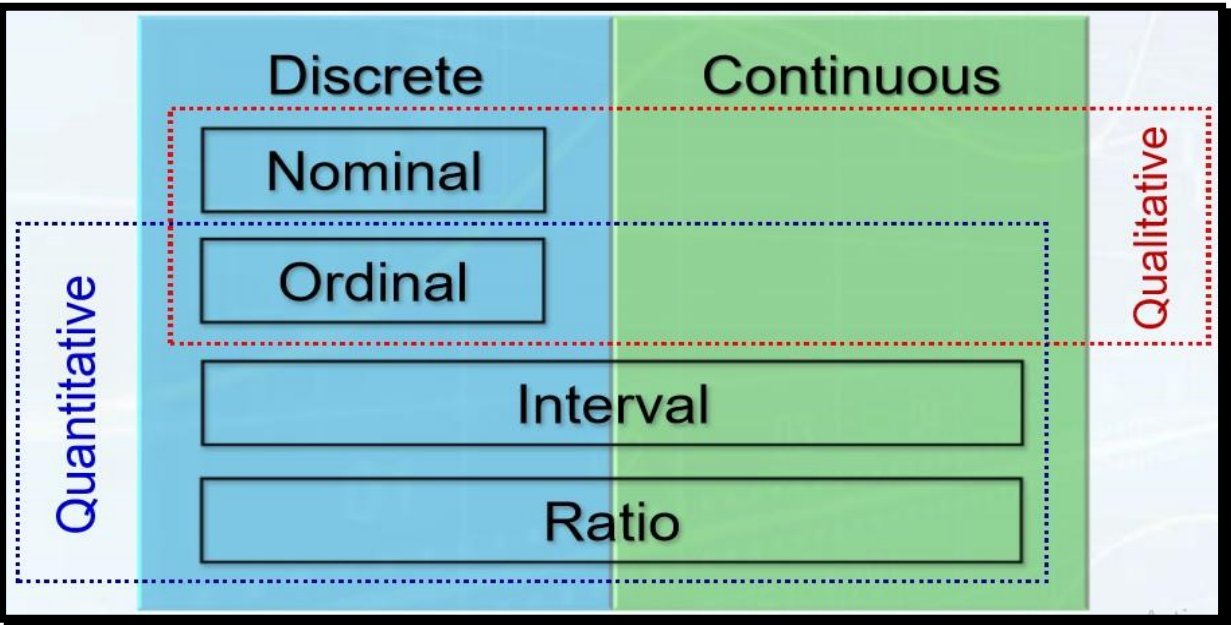
METHODS OF PRESENTATION OF DATA

1. Textual method

- This method presents the collected data in narrative and paragraphs forms.

2. Tabular method

- This method presents the collected data in table which are orderly arranged in rows and columns for an easier and more comprehensive comparison of figures.

3. Graphical method

- This method presents the collected data in visual or pictorial form to get a clear view of data (e.g. histogram, pie chart, pareto chart, pictograph, etc.).

LESSON 1.3 SAMPLING TECHNIQUES

Census - It is a count or measure of an entire population. Taking a census provides complete information, but it is often costly and difficult to perform.

Sampling - It refers to the process of selecting individuals from target population.

Sampling frame - A list of all elements or other units containing the elements or members in a population.

SAMPLING TECHNIQUES

> ➢ **Probability Sampling**
> ➢ **Nonprobability Sampling**

Probability Sampling or Random Sampling is a process whose members had an equal chance of being selected from the population.

• **Types of Probability sampling**

     ○ Simple Random Sampling   ○ Cluster Sampling

     ○ Systematic Sampling   ○ Multistage Sampling

     ○ Stratified Sampling

Simple Random Sampling - It is a process of selecting n sample size in the population via random numbers or through lottery.


Systematic Sampling - A systematic sample is a sample in which each member of the population is assigned a number. The members of the population are ordered in some way, a starting number is randomly selected, and then sample members are selected at regular intervals from the starting number. (For instance, every 3rd, 5th, or 100th member is selected.)
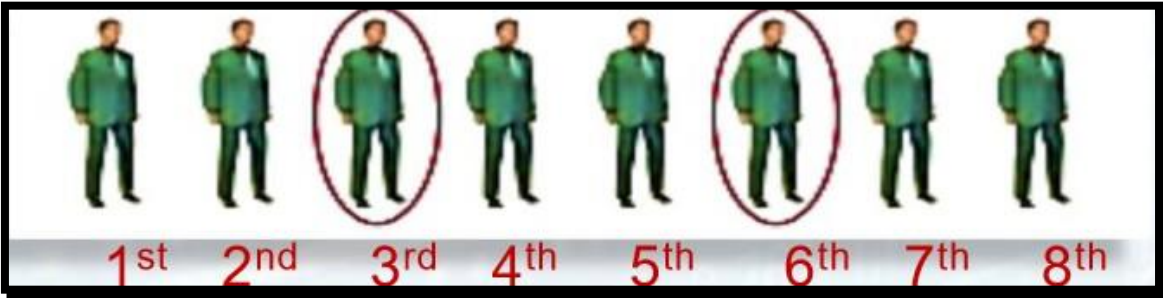
Stratified Sampling - A stratified sample is a sample obtained by dividing the population into subgroups, called strata, according to various homogeneous characteristics and then selecting members from each stratum for the sample.

Cluster Sampling - Here the population is divided into groups called clusters by some means such as geographic area or schools in a large school district, etc. Then the researcher randomly selects some of these clusters and uses all members of the selected clusters as the subjects of the samples.
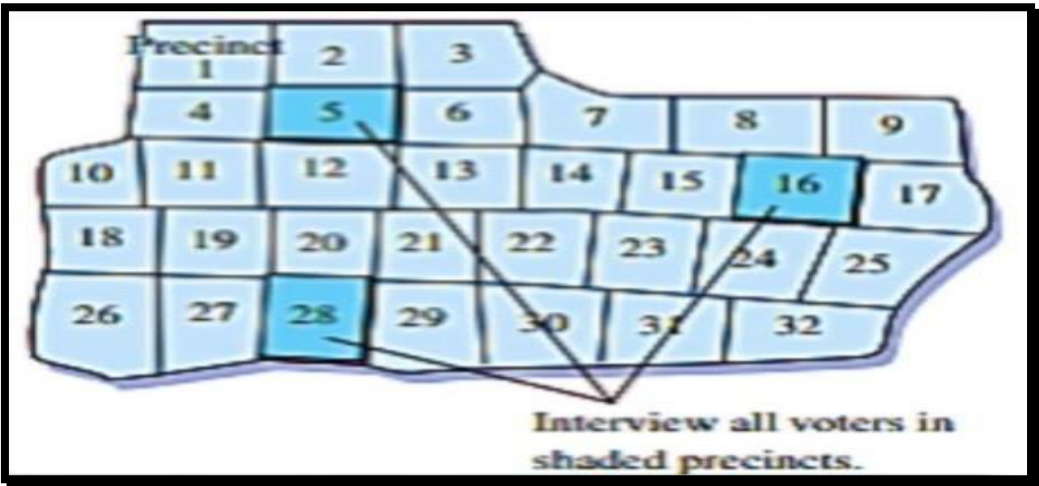
Multistage Sampling - A sample design in which the elements of the sampling frame are subdivided and the sample can be obtained by using combination of methods. This is usually used for national, regional, provincial or country level studies.

1st level	: 4 provinces/region	= 3
2nd level	: 3 municipalities per province	= 12
3rd level	: 2 barangays per municipality	= 24

**Nonprobability sampling or nonrandom sampling**

- Is a sampling procedure where samples selected in a deliberate manner with little or no attention to randomization.

- Some segments of the population do not have a chance of being selected or included in the sample or cannot be specified

Types of Nonprobability sampling

○ Convenience Sampling   ○ Snowball Sampling

○ Purposive Sampling   ○ Networking Sampling

○ Quota Sampling

Convenience Sampling - A convenience sample consists only of available members of the population.
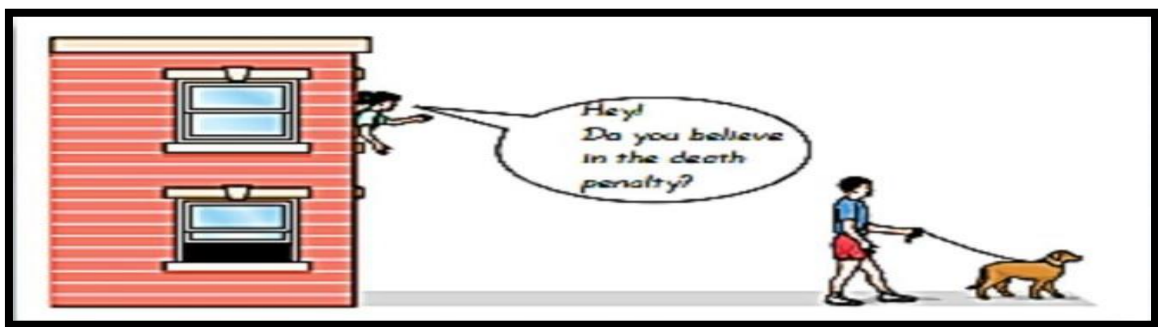

Purposive Sampling - It is also called judgment sampling. The sampling units are selected personally or subjectively by the researcher, who attempts to obtain a sample that appears to be representative of the population.

A human resource director interviews the qualified applicants in supervisory position. (Note: Qualified applicants are selected by the HRD which is based from his own judgment.)

Quota Sampling - in this method, the researcher determines the sampling size which should be filled up. The basic idea is to set a target number of completed interviews with specified subgroups of the population of interest.

For example, a researcher might ask for a sample of 100 females, or 100 individuals between the ages of 20-30.

Snowball Sampling - it involves starting a process with one individual or group and using their contacts to develop the sample, hence "snowball".
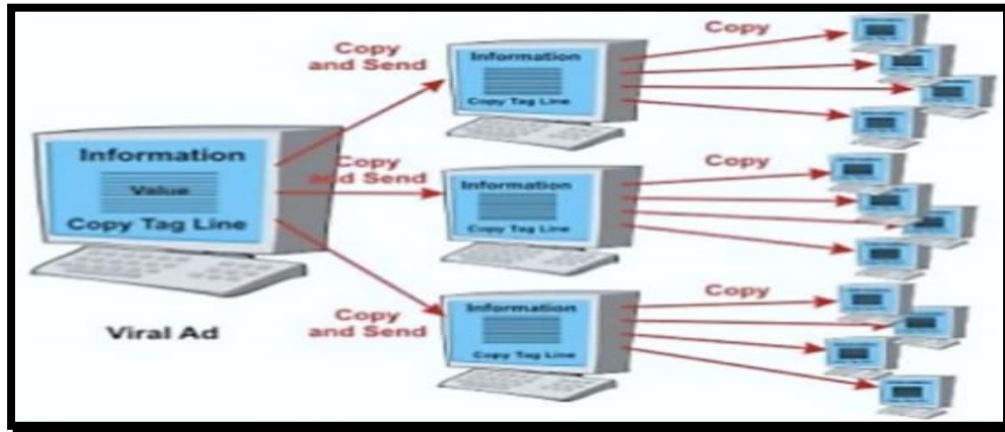
Networking Sampling - This is used to find socially devalued urban populations such as addicts, alcoholics, child abusers and criminals, because they are usually "hidden from outsiders."

For further understanding and example check the following link below:

- https://youtu.be/SFPGVTThJNk
- https://youtu.be/ZxV-kf0yBss
- https://youtu.be/hZxnzfnt5v8
- https://youtu.be/saO1yLxd1p8

Reference:

- https://youtu.be/SFPGVTThJNk
- https://youtu.be/ZxV-kf0yBss
- https://youtu.be/hZxnzfnt5v8
- https://youtu.be/saO1yLxd1p8