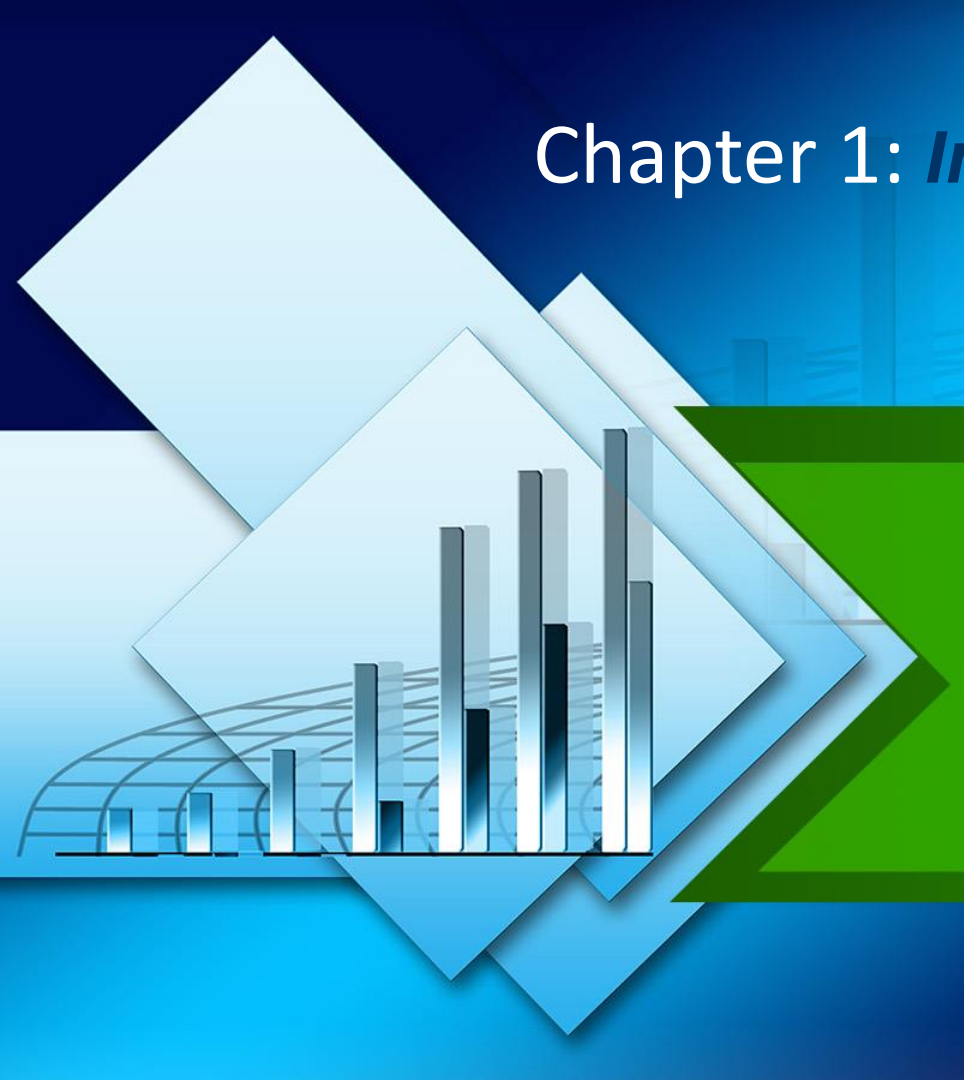
Chapter 1: *Introduction to Statistics*




1 2 3

Learning Outcomes

When you will complete this chapter, you would be able to-

- Understand the purpose of statistics.
- Know the differences between descriptive and inferential statistics.
- Understand the differences between a sample and a population.

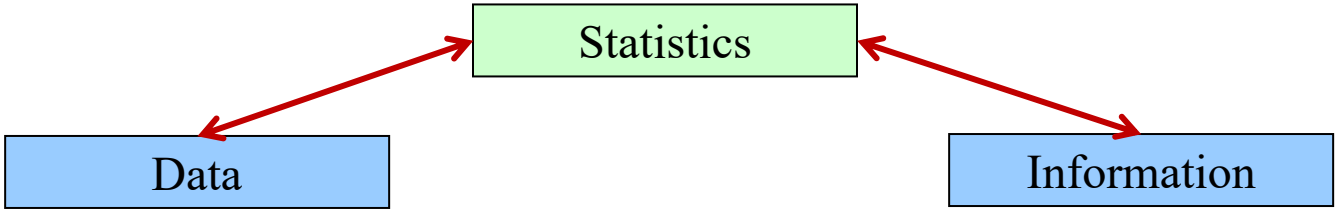
Contents

From this lecture, you are going to learn…

- ❑Statistics and types of statistics
- ❑Population and sample
- ❑Parameter and statistic
- ❑Sampling Techniques

What is Statistics?

"Statistics is a way to get information from data"



Statistics is a ***tool*** for creating ***new understanding*** from a set of numbers.

Definitions: Oxford English Dictionary

What is Statistics?

STATISTICS is the science of collecting, organizing, presenting, analyzing, and interpreting data to assist in making more effective decisions.

Collect	Gathering of facts or data
Organize	Arranging data for the presentation
Present	Summarizing data in textual, graphical, or tabular forms.
Analyze	Describing the data by using statistical methods and procedures.
Interpret	Making conclusions based on the analyzed data.

Application of Statistics in Pharmacy:

Pharmaceutical statistics is the application of statistics to matters concerning the pharmaceutical industry. This can be from issues of design of experiments to the analysis of drug trials to issues of commercialization of medicine.

- Evaluate the activity of a drug; e.g.; effect of caffeine on attention; compare the analgesic effect of a plant extract.
- To explore whether the changes produced by the drug are due to the action of the drug or by chance
- To compare the action of two or more different drugs or different dosages of the same drug are studied using statistical methods.
- To find an association between disease and risk factors such as Coronary artery disease and smoking
- Public health, including epidemiology, health services research, nutrition, environmental health, and healthcare policy & management.
- Design and analysis of clinical trials in medicine.

- Population genetics, and statistical genetics in order to link variation in genotype with a variation in phenotype. In biomedical research, this work can assist in finding candidates for gene alleles that can cause or influence predisposition to disease in human genetics
- Analysis of genomics data. Example: from microarray or proteomics experiments. Often concerning diseases or disease stages.
- Systems biology for gene network inference or pathways analysis
- Demographic studies: Age, gender, height, weight, BMI
- Epidemiology: deficiency of iron in anemia, iodized salt and goiter, hygiene and microbial disease

Limitations of Statistics:

The limitations of statistics are being pointed out as follows:

 Statistics deals with population and statistical methods deal with a mass of data and not with a single figure.

 Statistics can be misused. The data placed in the hands of an inexpert may lead to being compiled inaccurately and interpreted unscientifically.

 Statistical results are true only on average.

Branches of Statistics or Types of Statistics

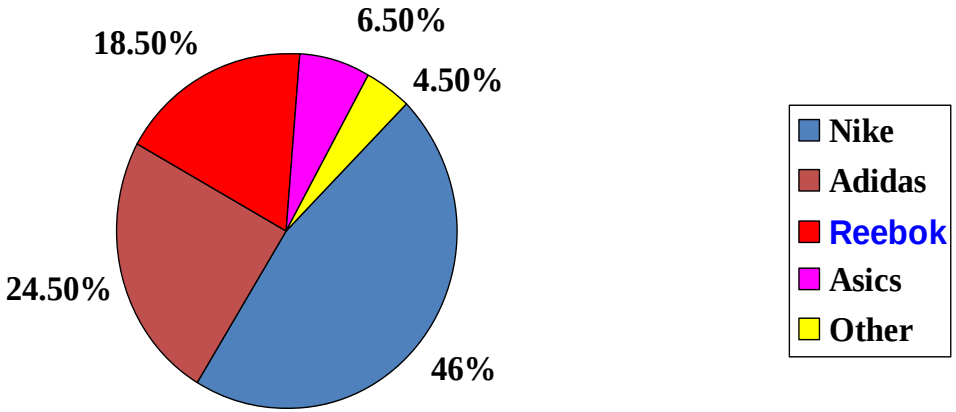
Branches of Statistics or Types of Statistics

Descriptive Statistics Involves organizing, summarizing, and displaying data.

e.g. Tables, charts, averages .

Inferential Statistics Involves using **sample data** to draw conclusions about a **population.**

Descriptive Statistics



EXAMPLE : Pie Chart For showing favorite type of running shoes of 200 runners.

Inferential Statistics

Example:

The accounting department of a large firm will select a sample of the invoices to check for accuracy for all the invoices of the company.

Exercise: Descriptive Vs Inferential Statistics

A follow up study was conducted among male who were aged 48, for 18 years. For men who took unhygienic food, approximately 70% were alive at age 65. For men who took hygienic food, 90% were alive at age 65.

For this example

Descriptive Statistics

- For men who took unhygienic food, approximately 70% were alive at age 65.
- For men who took hygienic food, 90% were alive at age 65.

A possible conclusion that means Inference is:

- Taking hygienic food is associated with a longer life for men.

Population VS Sample

Population:

The collection of all possible individuals, objects, or measurements of interest.

"ALL"

Sample:

Representative part of the population

"SOME"

Example:

Population – Total number of DIU students during the year 2011

Sample – Few selected students of DIU during the year 2011

Census, Parameter VS Statistic

- **Census:** Collection of data from every member of a population.

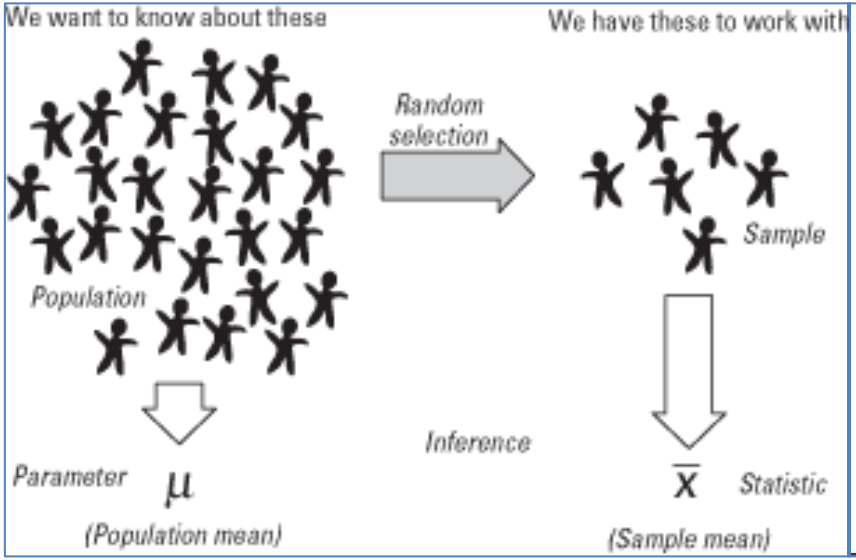
- **Parameter**: measurable characteristic of a population. It is usually referred to true or actual value.

  Example: *average* CGPA from all the DIU students.

- **Statistic**: measurable characteristic of a sample.

  Example: *average* CGPA from few students of DIU

Population, Sample, Parameter, Statistic
At a glance….



population of interest:

students on the robotics team

variable:

number of siblings

2 3 2 1 1 0 1 2 1 0 4 1

sample statistics:

mean:

$\bar{x} = 1.5$

standard dev:

$s_x = 1.3$

population parameters:

mean:

$\mu = 1.5$

standard dev:

$\sigma_x = 1.1$

Variable:

- A variable is a characteristic, often but not always quantitatively measured, containing two or more values or categories that can vary from person to person, object to object or from phenomenon to phenomenon.
- A variable is a measurable quantity that can vary within its domain.

Example:

- Sex is a variable which is composed of two categories, male and female.
- Again age is the variable because it may assume 10 years, 20 years, and so on. Other examples of variables are hair color, family size, weight, monthly income, etc.

N.B. A *constant* is a particular type of variable, which does not vary from one member of a group to another.

Types of Variable:

There are two types of variable such as:

(1) Qualitative Variable:

- A qualitative variable is one for which numerical measurement is not possible, such as hair color (brown, black, white, etc.) religion (Muslim, Hindu, Christian, etc.).
- The observations can be neither meaningfully ordered nor physically measured, only classified and then enumerated.

(2) Quantitative Variable:

A quantitative variable is one for which the resulting observations are numeric and thus possesses a natural ordering. Examples of such variable include the weight of a student, the height of the people, family size, yearly rainfall, etc.

Types of Quantitative variable

Discrete: A quantitative variable which can assume the only certain or isolated value is called a discrete variable.

e.g. family size, number of road accidents per day, etc.

Continuous : A quantitative variable which can assume any value within a specific range is called a continuous variable.

e.g. height, the weight of a student, etc.

Types variables (according to scale)

Data and Information:

Any measurements of one or more variables recorded either from population units or from sample units are called Data.

- So data are facts or figures from which conclusions can be drawn.

When data that have been recorded, classified, organized, related, or interpreted within a framework so that meaning emerges is called Information.

Types of Data:

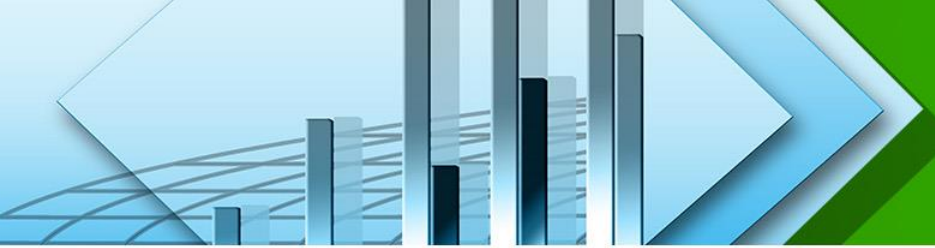
(i) Primary Data: The data which are obtained by direct observations from the population or sample.

(ii) Secondary Data: The data which are already obtained by some other persons or organizations.

Types of Data (Based on the Source):

(i) Primary Data: The data which are obtained by direct observations from the population or sample.

(ii) Secondary Data: The data which are already obtained by some other persons or organizations.



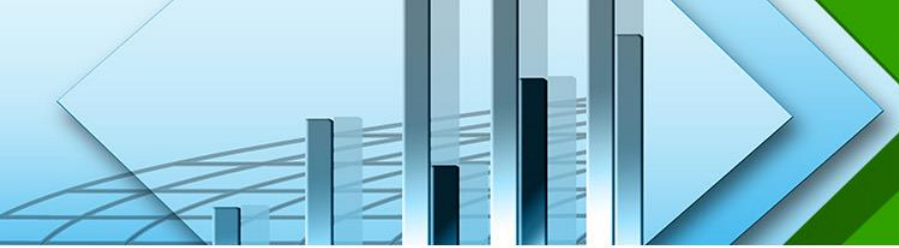
Types of Data (Based on the arrangement):

(i) Raw Data or Ungrouped Data:

- Raw data is the data which just collected with no sorting or other processing.
- Ungrouped data which is also known as raw data is data that has not been placed in any group or category after collection.
- Data is categorized in numbers or characteristics therefore, the data which has not been put in any of the categories is ungrouped.

(ii) Grouped Data:

- Grouped data is the type of data which is classified into groups after collection.
- The raw data is categorized into various groups and a table is created. The primary purpose of the table is to show the data points occurring in each group. For instance, when a test is done, the results are the data in this scenario and there are many ways to group this data.



Types of Data
(According to the level of measurement):

Data can be classified according to levels of measurement.

- Statistical data may be broadly classified as ***categorical (or Qualitative)*** and ***numerical (or Quantitative).***
- Categorical are of two types: ***nominal*** and ***ordinal***
- Numerical data are measured in ***interval scale*** and ***ratio scale***.

Types of Data (Based on Variables)

The following section provides the identification of various types of data.

(i) Nominal Data:

- All qualitative measurements are nominal, regardless of whether the categories are designated by names (red, white, male) or numerals (June 20, room 19).
- There is no particular order.
- For example, Jersey numbers of football players, name, religion, eye color, sex, etc.

(ii) Ordinal Data:

- When there is an ordered relationship among the categories, the variable is said to be an ordinal variable.
- In other words, the categories of the ordinal variable may have the typical relations “higher”, “greater”, “more desired”, “less difficult” and so on.
- So all data are categories as their ranked and there is an ordered relationship among categories.
- For example, official designation, academic degrees, socio-economic status, etc.

(iii) Interval Data:

- Data generated through the measurement of an interval variable are called interval data.
- That is a meaningful difference between values. Distance between the two categories is known and constant.
- Zero is just a point on this scale. It does not represent the absence of the condition. Zero degree Celsius does not represent the absence of heat.
- For example, temperature, students IQ ratings, dress size, etc.

(iv) Ratio Data:

- Data generated through the measurement of a ratio variable are called ratio data.
- Meaningful zero points and the ratio between values.
- For example, a number of patients seen, distance to class, height, weight, etc.

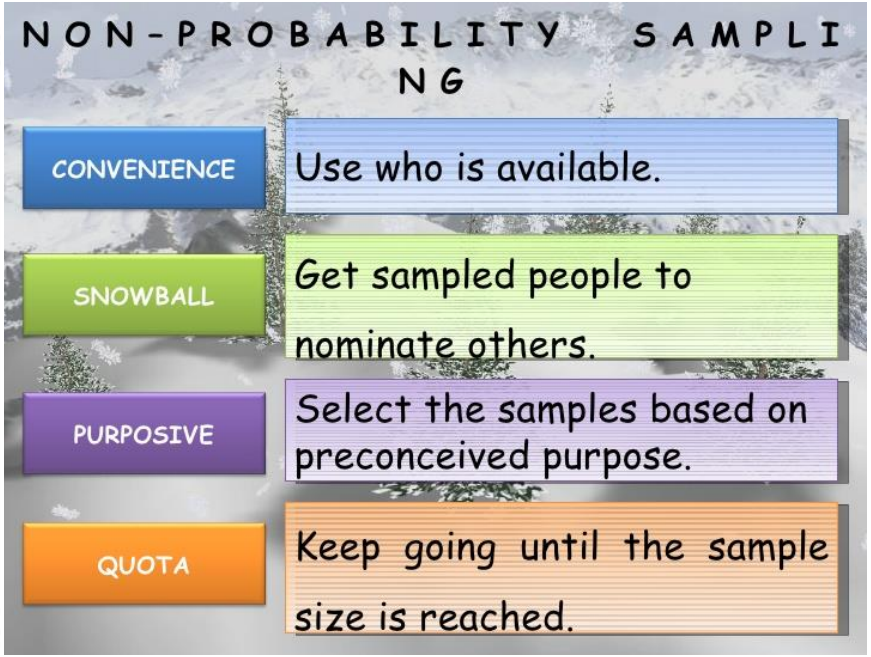
Comparison of Characteristics Between Different Level of Measurement:

Scale	Characteristics	Examples
Nominal	Categories are homogeneous, mutually exclusive, and no assumptions about ordered relationships between categories made.	Sex of subject, eye color (black, brown), Religion, Political affiliation, Place of residence(urban-rural), Class roll numbers (12, 53), etc.
Ordinal	All of the above plus the categories can be rank- ordered.	Examination grade (A, B, C), Health status (poor, average, good), Level of education (Illiterate, primary, secondary), etc.
Interval	All of the above plus exact differences between categories are specified and an arbitrary zero point is assumed.	Temperature (980C), IQ test score, Calendar time (6 AM, 5 PM), etc.
Ratio	All of the above with the expectation that a true zero point is assumed.	Height, Weight, Fat consumed (in gm), Distance (in km), etc.

Sampling Techniques

Probability Sampling

Simple Random	Select from a full list of the population (sampling frame). Can use a random number table to do this.
Systematic	Start at random, at a point on the sampling frame, and choose every tenth case (or some other frequency) depending on sampling frame size.
Stratified	Sampling frame stratified (for example by class, race, sex) then random sampling
Cluster	Population divided into units or clusters each containing individuals in a range of circum-stances (for example, different types of young offender institution could be sampled).
Multi-stage	An extension of the cluster sample, in which samples are drawn from within clusters (for example sampling by age, sex and ethnicity from within young offender institutions identified through cluster sampling).



CLASS EXERCISE

Classify the following as quantitative or qualitative data

- **Color of the eye**
- **Number of typewriters in a room**
- **Civil status**
- **Address**
- **Telephone numbers**
- **Age of teachers**
- **Rank of students**
- **Speed of a car**
- **Birth rates**
- **Score in mathematics examination**

Identify each of the following as continuous or discrete

1. Weight of a body
2. Length of a rod
3. Number of chairs in the room
4. Dimensions of a table
5. Number of possible outcomes in throwing a die
6. Number of hairs on your head
7. Amount of sales in a business firm
8. All rational numbers
9. Speed of light
10. Area of a land
11. Lifetime of television tubes and batteries
12. Life span of a person
13. Number of passengers in a plane.


Thank
You