Statistics:

It is difficult to define statistics in a few words. No formal definition thus has emerged so far and no definition is perhaps beyond controversy.

According to Sir R. A Fisher, "The science of statistics is essentially a branch of applied mathematics applied to observational data.

According to Croxton and Cowden, “Statistics is the science, which deals with the collection, analysis and interpretation of numerical data.”

American Heritage dictionary defines statistics as, “The mathematics of the collection, organization and interpretation of numerical data especially the analysis of population characteristics by inference from sampling.”

Scope and use of statistics:

Statistics has been useful in researches of almost all disciplines. We can say in a word, where numbers or numerical values play a part, statistical methods are useful and sometimes indispensable. A few fields where statistics is applicable or useful are listed below:

1. Statistics in planning: In fact statistics originated from the concept of planning. In the ancient time a state chief felt the necessity of statistics in the planning of public issues. Nowadays statistics is indispensable in the planning of any issue in any field.
2. Statistics in population, health and family planning.
3. Statistics in biology.
4. Statistics in business and commerce.
5. Statistics in agriculture
6. Statistics in physical science
7. Statistics in environment
8. Statistics in medicine
9. Statistics in psychology and education
10. Statistics in production industry etc.

Most. Salma Yeasmin Jannaty
Lecturer of Statistics, MNS Department,
BRAC University.
E-mail: salma.yeasmin@bracu.ac.bd

Classification of Statistics :

Statistics can broadly classified into two branches:

i) Descriptive statistics

ii) Inferential statistics

i) Descriptive statistics:

In descriptive statistics we analyze the data without considering the theory of probability. Descriptive statistics includes the methods of-

i) Data collection

ii) Construction of data file

iii) Data classification

iv) Tabulation

v) frequency distribution

vi) Graphs and diagrams

vii) Measure of central tendency

viii) measure of dispersion or variation

ix) measures of skewness and kurtosis

x) Moments

xi) Some measure of attributes etc.

ii) Inferential statistics

In this statistics we use powerful tool of statistical methods based on the theory of probability. It utilizes sample data to make estimates, decisions, predictions or other generalization about a larger set of data.

Most. Salma Yeasmin Jannaty
Lecturer of Statistics, MNS Department,
BRAC University.
E-mail: salma.yeasmin@bracu.ac.bd

Some Basic Definitions:

Population:

An aggregate of all individuals or items defined on some common characteristics is called a population. It is denoted by 'N'.

Example: If we want to study average weight of the students of first semester BBA then the set that consists of all the weights of the student of first semester BBA will be the population in this case

Parameter:

A parameter is a numerical measure that describes a characteristic of a population.

Example: Population mean, population variance.

Sample:

A small but representative part with finite number of individuals or items of population is called a sample.

In many particular situations, it is impossible or even impractical to study the whole population, in such case only a small and representative part of population taken under consideration to draw inferences about the population by analyzing that part of population. Such a part of population is known as sample. Sample size is denoted by 'n'.

Statistic:

Statistic is a numerical measure that describes a characteristic of sample. Example: sample mean, sample variance.

Most. Salma Yeasmin Jannaty
Lecturer of Statistics, MNS Department,
BRAC University.
E-mail: salma.yeasmin@bracu.ac.bd

Estimator and estimate:

Estimator: Any function of sample values, a statistic, when used to estimate a parameter is called an estimator.

Ex- sample mean and sample variance are estimator.

Estimate: If we get a numeric value of the estimator, then it is called estimate.

Ex- mean of a specific sample, variance of a specific sample.

Operations with notations

Summation:

1. $x_1 + x_2 + x_3 + . \ . \ . + x_n = \sum_{i=1}^{n} x_i$
2. $f_1 x_1 + f_2 x_2 + . \ . \ . + f_n x_n = \sum_{i=1}^{n} f_i x_i$
3. $ax_1 + ax_2 + ax_3 + . \ . \ . + ax_n = a \sum_{i=1}^{n} x_i$
4. $a + a + a + . \ . \ . + a = \sum_{i=1}^{n} a = na$

Product:

$$x_1 * x_2 * x_3 * . \ . \ . * x_3 = \prod_{i=1}^{n} x_i$$

Most. Salma Yeasmin Jannaty
Lecturer of Statistics, MNS Department,
BRAC University.
E-mail: salma.yeasmin@bracu.ac.bd

Scale of measurement:

It is belief of some researchers that if it is to be researchable then it must be measurable. Measurement deals with the quantifying procedure of any phenomenon substantial or insubstantial and it entails comparison with standard.

Measurement is a process of assigning numbers to some characteristics or variables or events according to scientific rules. There are four scales of measurement.

a) Nominal scale:

At this scale or level numbers are assigned to identify, label, or classify individuals or objects. The scale is used for variables which can be divided into categories that are qualitatively different but no difference can be made on magnitude.

Ex- Region, gender, Religion etc.

b) Ordinal scale:

The measurement scale in which numbers are assigned to the categories or variable values for identification as well as ranking is called ordinal scale.

Ex- economic status

c) Interval scale:

The measurement scale in which numbers are assigned to the variable values in such a way that the level of measurement is broken down on a scale of equal units and the zero value on the scale is not absolutely zero, is called an interval scale.

Ex- Temperature, IQ score etc.

d) Ratio scale:

The measurement scale in which numbers are assigned to the variable values in such a way that the level of measurement is broken down on a scale of equal units and the zero value on the scale is absolutely zero, is called an ratio scale.

Ex- Height, family size etc.

Most. Salma Yeasmin Jannaty
Lecturer of Statistics, MNS Department,
BRAC University.
E-mail: salma.yeasmin@bracu.ac.bd

Variable:

Variable is a measurable quantity which can very within its domain. For example, family size is a variable, because it is measurable quantity within it's domain.

Example: family size, height weight

Types of variables:

There are two basic types of variables

i) Qualitative variable (Category variable or attribute): Qualitative variable is one for which numerical measurement is not possible.

Example: Hair color, religion, home district.

ii) Quantitative variable: Quantitative variable is one for which numerical measurement is possible.

Example: Height, weight, family size etc.

Quantitative variable can be subdivided into two variable.

a) Discrete variable
b) Continuous variable

Discrete variable:

A quantitative variable which possesses isolated or integral value is called discrete variable.

Example: family size.

Continuous variable:

A quantitative variable with takes value within a range or limit is called continuous variable.

Example: height, weight.

Most. Salma Yeasmin Jannaty
Lecturer of Statistics, MNS Department,
BRAC University.
E-mail: salma.yeasmin@bracu.ac.bd

Exercise:

Classify each variable as qualitative or quantitative:

i) Marital status of nurses in a hospital
ii) Time it takes to run a marathon
iii) Color of automobiles in a shopping center parking lot.
iv) Age of people living in a personal care home.

Discrete or continuous:

ii) Number of pizzas sold by Pizza express each day.
iii) Lifetime (in hours) of 15 iPod batteries
iv) Number of students in STA201section 20.
v) Age of the students of a class.

Data: Data are numerical facts gathered from a statistical investigation.

Sources of data:

a) Primary data

Primary data is a type of data that is collected by researchers directly from main sources through interviews, surveys, experiments, etc. Primary data are usually collected from the source where the data originally originates from and are regarded as the best kind of data in research.

Ex- An organization doing market research about a new product (say phone) they are about to release will need to collect data like purchasing power, feature preferences, daily phone usage, etc. from the target market. The data from past surveys are not used because the product differs.

b) Secondary data

Secondary data is the data that has already been collected through primary sources and made readily available for researchers to use for their own research. It is a type of data that has already been collected in the past.

A researcher may have collected the data for a particular project, then made it available to be used by another researcher. The data may also have been collected

Most. Salma Yeasmin Jannaty
Lecturer of Statistics, MNS Department,
BRAC University.
E-mail: salma.yeasmin@bracu.ac.bd

for general use with no specific research purpose like in the case of the national census.

Sources of secondary data include books, personal sources, journals, newspapers, websites, government records etc. Secondary data are known to be readily available compared to that of primary data. It requires very little research and needs for manpower to use these sources.

Methods of primary data collection:

Here are the top six data collection methods:

1. Observations
2. Mail questionnaires
3. Interviews
4. Documents and records
5. Focus groups
6. Oral histories

i) **Observation:** Observation is such a method where we have to use eyes rather than ears and voice. Observation is way of gathering data by watching behavior, events, or noting physical characteristics in their natural setting. Observations can be overt (everyone knows they are being observed) or covert (no one knows they are being observed and the observer is concealed).

ii) **Mail questionnaire:**

Mailed questionnaires.is a tool of research where an answerer sends his answers through mail to the researcher. It is administered by mail to designated respondents under an accompanying cover letter and is returned, by mail, by the respondent to the research organization.

Most. Salma Yeasmin Jannaty
Lecturer of Statistics, MNS Department,
BRAC University.
E-mail: salma.yeasmin@bracu.ac.bd

iii) **Interview:**

Interviews are a method of data collection that involves two or more people exchanging information through a series of questions and answers. The questions are designed by a researcher to elicit information from interview participants on a specific topic or set of topics. Some methods of taking interviews are-

- Telephone interview
- Indirect inquiry
- Personal interview

Basic points of making a questionnaire:

1. Be comprehensible

Use a clear and comprehensible language to ease the cognitive burden for the respondents. Each and every question reduces the respondent's capability of concentration. Therefore, if you want to keep their attention, the questions should be as comprehensible as possible. This is especially true if you have less educated people in your sample.

2. Be clear

It sounds obvious, but questions need to be clear and unambiguous. Using vague buzz words, unfamiliar terms or everyday language can blur your results. Even though respondents may think they understand what you mean, everyone will have something different in mind, when answering your question.

But be careful. Sometimes being clear runs contrary to being comprehensible, especially if you try to be overly precise. Being clear should never lead to these extremely long and awkward questions that nobody will read thoroughly, especially when completing the questionnaire on a mobile device. Always keep the respondent in mind.

3. Be neutral

Avoid suggestive questions or unbalanced answering options. The respondents may not necessarily mind or even notice, but your results may then lean towards one or

Most. Salma Yeasmin Jannaty
Lecturer of Statistics, MNS Department,
BRAC University.
E-mail: salma.yeasmin@bracu.ac.bd

another answering option. In this case you are not measuring the objective facts, but implicitly asking for approval of your subjective standpoint. Your data will be biased. Hence, you should always take a neutral standpoint and try to be as objective as possible when writing a questionnaire.

4. Operationalize

Very often, you will want to find out about attitudes and behaviors that can't be evaluated directly. Try to operationalize these concepts and translate them into clear and tangible indicators. Instead of asking directly whether someone is "lifestyle oriented", rather ask for specific products or activities, the respondent may have had contact with during the last weeks. Not only will it be easier for the respondent to find an answer but also lead to much more accurate results.

5. Mind the order

Any clues given at the beginning of the questionnaire may affect the answers to questions that follow. Or the first statements presented to a respondent may affect the respondent's choice of an answering option. This is what psychologists call priming, an effect of the short-term memory on our decision making. Therefore, if possible, try to randomize the order of your questions and statements. If you can't do that, at least try to optimize the order to get natural, unbiased feedback.

Most. Salma Yeasmin Jannaty
Lecturer of Statistics, MNS Department,
BRAC University.
E-mail: salma.yeasmin@bracu.ac.bd