Introduction to Statistics

❖ Origin of Statistics

Statistics has been derived from the Latin word 'Status' or Italian word 'Statista' or the French word 'Statistique' and the German word 'Statistik'. Status and Statistik means political state and Statista means activities of political state. At first Professor G. Achenwall used the word Statistics in 1749. According to Professor G Achenwall "Statistics is the political science of the several countries".

PC Mahalanobish name the word of Statistics in Bengali word Porisongkkhan.

❖ Definition of Statistics

Statistics is concerned with scientific methods for collecting, organizing, summarizing, presenting and analyzing sample data as well as drawing valid conclusions about population characteristics and making reasonable decision on the basis of such analysis.

❖ Some Definition of Statistics

- Statistics can be defined as the collection presentation and interpretation of numerical data.- Croxton and Crowed.
- Statistics are measurement, enumerations or estimates of natural or social phenomena systematically arrangement to exhibit their inner relation.- Conner.
- The science of Statistics is essentially a branch of applied mathematics and can be regarded as a mathematics applied to observation data.- R.A Fisher.

❖ Types of Statistics

There are two types of Statistics, such as descriptive statistics and inferential statistics.

- **Descriptive Statistics**

Descriptive statistics deals with collection, tabulation, presentation and analysis of data. The study of frequency distribution, measures of central tendency, measures of dispersion, correlation, regression etc. are included in descriptive statistics.

- **Inferential Statistics**

The descriptive statistics are used for making predictions or decisions relating to unobserved characteristic. The methods of taking decision is known as inferential statistics or statistical inference. The inference is made by sampling, sampling distribution, estimation of parameter and test regarding any hypothesis on parameter.

❖ **Parameter**

Any characteristics of population about which inference are to be made is called parameter. Population mean (μ) and population variance (σ^2) are examples parameter.

❖ **Statistic**

Any characteristic of sample is usually known as statistic. Sample mean ($\bar{x}$) and sample variance (s^2) are examples of statistic.

❖ **Scope and importance of Statistics**

- *Statistics and planning:* Statistics in indispensable into planning in the modern age which is termed as "the age of planning". Almost all over the world the govt. are re-storing to planning for economic development.

- *Statistics and economics:* Statistical data and techniques of statistical analysis have to immensely useful involving economical problem. Such as wages, price, time series analysis, demand analysis.

- *Statistics and business:* Statistics is an irresponsible tool of production control. Business executive are relying more and more on statistical techniques for studying the much and desire of the valued customers.

- *Statistics and industry:* In industry statistics is widely used inequality control. In production engineering to find out whether the product is confirming to the specifications or not. Statistical tools, such as inspection plan, control chart etc.

- *Statistics and mathematics:* Statistics are intimately related recent advancements in statistical technique are the outcome of wide applications of mathematics.

- **Statistics and modern science:** In medical science the statistical tools for collection, presentation and analysis of observed facts relating to causes and incidence of dieses and the result of application various drugs and medicine are of great importance.

- **Statistics, psychology and education:** In education and physiology statistics has found wide application such as, determining or to determine the reliability and validity to a test, factor analysis etc.

- **Statistics and war:** In war the theory of decision function can be a great assistance to the military and personal to plan "maximum destruction with minimum effort."

1.14 Population

Simply, population means the aggregate of human individuals in a defined area or region. In statistic population refers the totality of all the items or individuals having some specific characteristics. e.g. All the students of university of Dhaka constitute a population.

A population can be classified into one of two groups-

(a) Finite population : A population having a finite number of units (or individuals or items) called a finite population.

Example : The population consisting the students of Dhaka University.

(b) Infinite population : A population having an infinite number of units (or individuals items) is called an infinite population.

Example : The population consisting of all possible outcomes (here head and tail) in successi tosses of a coin.

Sample : A representative and considerably small part of a population is known as a sample of th population.

Example :

(i) A group of 1050 students from 31,000 students of Dhaka University consistute a sample.

(ii) A spoonful boiling rice from a cooking pot of boiling rice.

1.15 Variable

The characteristic which varies over the units (or from unit to unit) is called variable.

Example : Father's occupation of the students of Dhaka University. Here, the unit or population-u is-a student and their father's occupation varies. Say, one father's occupation is teaching, another fathe occupation is business, else other father's occupation is driving and so on.

Classification of variable

We can classify the variables according to the characteristics into two groups :

(a) Qualitative variable

(b) Quantitative variable

(a) Qualitative variable : There are many characteristics that can not be ex.pressed in a numerical form (i. g., color, sex, occupation etc) but we can arrange them according to their quality attribute. These types of characteristic are known as qualitative variable.

Example : Father's occupation of the students of Dhaka University.

(b) Quantitative variable : The characteristics of an unit or item that are expressed in numeric form or in numbers are called quantitative variable.

Example : Monthly family income of the students of Dhaka University. Quantitative variable c further be classified into two sub-groups :

(i) Discrete variable; (ii) Continuous variable

(i) Discrete variable

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

Example : Family size, population size. Number of road accidents per day in Bangladesh.

(ii) Continuous variable

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

Example : Height, weight etc.

1.16 Constant

A numerical characteristic which does never change or vary it's value is termed as constant.

Example : The ratio of circumference and diameter of a circle (π). And the value of $\pi = 3.1416$ is some for all the circles.

1.17 Scale of Measurements

To measure a variable there are various ways. We classify the ways into four scales of measurement.

(a) Nominal scale

(b) Ordinal scale

(c) Interval scale

(d) Ratio scale

(a) Nominal Scale : A scale that measures a variable nominally (or by name without any order) is called nominal scale. Here, we can set numerical values for the names but can not order them meaningfully. By setting values we only can categorize them. The variable which is measured in nominal scale is known as nominal variable.

Example : Religion, color, sex.

The nominal scale has the following properties :

(i) Data categories are mutually exclusive and exhaustive.

(ii) Data categories have no logical order.

(b) Ordinal Scale : A scale that identifies the values of a variable and arranges the values meaningfully in order of magnitude is called ordinal scale. Here, the magnitude is used only for comparison and not for any mathematical operation. The variable which is measured in nominal scale is known as ordinal variable, e.g., economic status. We can order the economic status in high, middle and low as follows :

Low = 1, middle class = 2, higher class = 3.

The properties of ordinal scale are :

(i) The data classifications are mutually exclusive and exhaustive.

(ii) Data classifications are ranked or ordered according to the particular trait they possess.

(c) Interval Scale : A scale which includes a definition of distance between the categories in terms of fixed and equal units is called interval scale. In this scale the concept of zero (0) is not included. So we can not perform any arithmetical or mathematical operation. The variable which is measured in interval scale is known as interval variable, e.g., temperature, I. Q. score. Here, thermometer records temperature interms of degrees and a 1 degree change (increase or decrease) in temperature implies the some amount of heat.

The properties of interval scale are :

(i) Data classifications are mutually exclusive and exhaustive.

(ii) Data classifications are scaled according to the amount of the characteristic they possess.

(iii) Equal differences in the characteristic are represented by equal differences in the measurements.
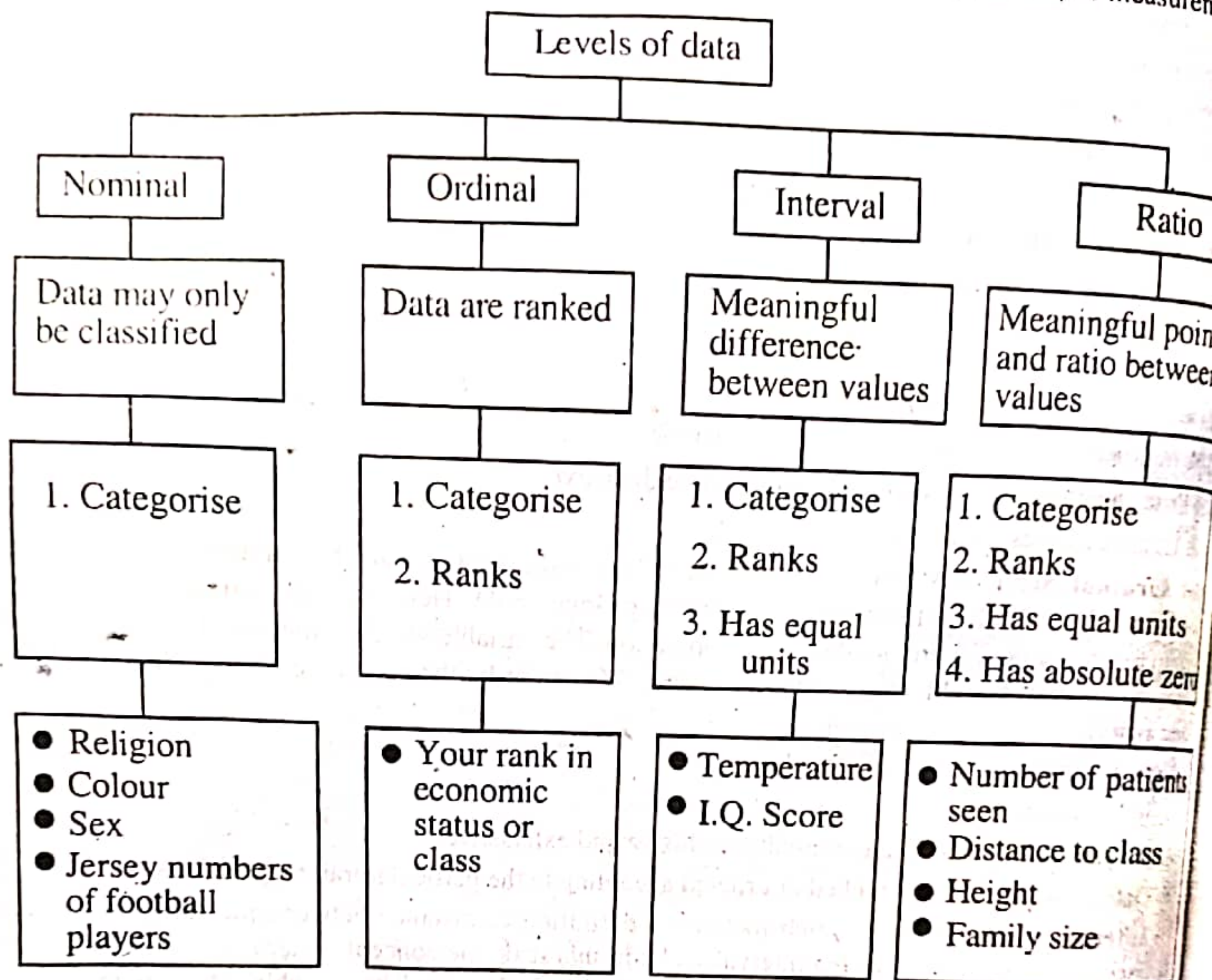
(d) Ratio Scale : The scale of measurement which includes all the properties of an interval scale including the concept of zero (0) is called ratio scale. In ratio scale we can perform the arithmetic operation (i.e, summation, substraction etc.). The variable that is measured in ratio scale is known as ratio variable, e.g., height, weight etc.

The properties of the ratio scale are :

Data classifications are mutually exclusive and exhaustive.

(ii) Data classifications are scaled according to the amount of the characteristics they possess.

(iii) Equal differnces in the characteristic are represented by equal differences in the numbers assig the classifications.

(iv) The zero point is the absence of the characteristic.

The following chart summerizes the major characteristics of the various levels (scale) of measuren

```
                          ┌──────────────────┐
                          │  Levels of data  │
                          └──────────────────┘
```

Nominal	Ordinal	Interval	Ratio
Data may only be classified	Data are ranked	Meaningful difference between values	Meaningful poin and ratio betwee values
1. Categorise	1. Categorise 2. Ranks	1. Categorise 2. Ranks 3. Has equal units	1. Categorise 2. Ranks 3. Has equal units 4. Has absolute zer
• Religion • Colour • Sex • Jersey numbers of football players	• Your rank in economic status or class	• Temperature • I.Q. Score	• Number of patients seen • Distance to class • Height • Family size

1.18 Origin and Scale of a Variable

Manytimes it is troublesome and time consuming to deal with the values of a variable. So we a rearrange the values of the variables for some easier calculation. In such rearrangement we change origin by subtracting (or adding) a constant value and the scale by dividing (or multiplying) anot constant value. Normally, we consider 'o' as origin and 'a' one-unit as scale.

For example, if $x_1, x_2, \cdots x_n$ be the values of a variable x then

$$y_1 = \frac{x_1 - a}{b}, \; y_2 = \frac{x_2 - a}{b}, \cdots, y_n = \frac{x_n - a}{b}$$ are the values where the origin and scale have been shifted.

Lecture notes on Data & Data Collection

Course Teacher:
Md. Shahajada Mia
Assistant Professor
Department of Statistics
Pabna University of Science and Technology

❖ Data

Data is the plural form of datum. Data are the collection of raw facts or information's and figures from any sects of inquiry for the purpose of statistical analysis. For example, to know the average height of the students of PUST. We have to collect the information on all the students or representative part of the students of PUST. Here the aggregate information would be called data.

❖ Types of data

According to nature or characteristics, data can be classified into two ways:

(i) Qualitative data and (ii) Quantitative data

• Qualitative data

The data which can not be measure by the numerical form but can be categorized into two or more groups is called qualitative data.

Example: sex, religion, level of educational, color etc.

Qualitative data can be classified into two ways: (a) nominal data and ordinal data.

> **Nominal data:** Nominal data are used to label variables where there is no quantitative value and has no order. So, if you change the order of the value then the meaning will remain the same. Example: sex, religion, color etc.

> **Ordinal data:** Ordinal data is almost the same as nominal data but not in the case of order as their categories can be ordered like 1st, 2nd, etc. However, there is no continuity in the relative distances between adjacent categories. Example: Economical status, time of day etc.

• Quantitative data

The data which can be expressed in numerical form or in numbers are called quantitative data.

Example: Height, weight, GRE score, income etc.

Quantitative data also two types. Such as (a) Interval and (b) Ratio data.

> **Interval data:** Interval Data are measured and ordered with the nearest items but have no meaningful zero. Example: Temperature, IELTS score etc.

> **Ratio data:** Ratio Data are measured and ordered with equidistant items and a meaningful zero and never be negative like interval data. Example: Income, Expenditure, weight etc.

❖ **Source of Data**

According to source of data can be classified into two ways:

(i)   Primary data

(ii) Secondary data

(i)   **Primary data:** the data which are obtained by direct observations from the population or sample is called primary data. primary data are normally collected for first time.

(ii) **Secondary data:** the data which are already obtained by some other persons or organisations and are already published or utilized are called secondary data.

❖ **Methods of Primary Data Collection**

There are several method to collect primary data which are as follows:

1. **Observation Method**

Observation method is used when the study relates to behavioural science. This method is planned systematically. It is subject to many controls and checks. The different types of observations are:

> Structured and unstructured observation

> Controlled and uncontrolled observation

> Participant, non-participant and disguised observation

2. **Interview Method**

The method of collecting data in terms of oral or verbal responses. It is achieved in two ways, such as:

> **Personal Interview:** In this method, a person known as an interviewer is required to ask questions face to face to the other person. The personal interview can be structured or unstructured, direct investigation, focused conversation, etc.

➢ **Telephonic Interview:** In this method, an interviewer obtains information by contacting people on the telephone to ask the questions or views orally.

3. Questionnaire Method

In this method, the set of questions are mailed to the respondent. They should read, reply and subsequently return the questionnaire. The questions are printed in the definite order on the form. A good survey should have the following features:

- Short and simple
- Should follow a logical sequence
- Provide adequate space for answers
- Avoid technical terms
- Should have good physical appearance such as colour, quality of the paper to attract the attention of the respondent.

4. Schedules

This method is similar to the questionnaire method with a slight difference. The enumerations are specially appointed for the purpose of filling the schedules. It explains the aims and objects of the investigation and may remove misunderstandings, if any have come up. Enumerators should be trained to perform their job with hard work and patience.

❖ **Methods of Secondary Data Collection**

Secondary data is data collected by someone other than the actual user. It means that the information is already available, and someone analyses it. The secondary data includes magazines, newspapers, books, journals, etc. It may be either published data or unpublished data. Published data are available in various resources including

- Government publications
- Public records
- Historical and statistical documents
- Business documents
- Technical and trade journals

Unpublished data includes

- Diaries
- Letters
- BBS, WHO, BARDEM, ICDDRB
- Unpublished biographies, etc.

❖ **Difference between primary and secondary data**

Differences between Primary Data and Secondary Data:

Basis of Difference	Primary Data	Secondary Data
1. Definition	1. They are collected for the first time.	1. They have already been collected by some other persons.
2. Accuracy	2. They are accurate.	2. They lack accuracy.
3. Nature of Data	3. They are in the form of raw data to which statistical methods are applied.	3. They are in the form of finished products.
4. Editing	4. They do not require editing.	4. They require editing.
5. Time and money	5. Collection of primary data requires a lot of time and money.	5. They dp not require a lot of time and money.
6. Precaution	6. Precaution is not required in the use of primary data.	6. Precautions are required in the use of secondary data.