[Academic Script]

Data types and scale; Sample and population

Subject:	Business Economics
Course:	B. A. (Hons.), 1st Semester, Undergraduate
Paper No. & Title:	Paper – 102 Statistics for Business Economics
Unit No. & Title:	Unit – 1 Introduction
Lecture No. & Title:	Lecture – 1 Data types and scale; Sample and population

Academic Script

1. Introduction

To take any kind of decision or to forecast in business you need data. Success of any statistical investigation depends on the accuracy of the available data. Therefore data collection is very important activity in decision making.

In this talk we shall study the different methods and sources of data collection. Data may be classified into "primarydata" or "secondary data". The amount of information contained in the data may be determined by different types of scales of measurements.

2. Data Types

i.Consider the following data regarding the outcomes of results in the corporation election -2015.

MahanagarPalika	BJP	Congress	Others
Ahmedabad	142	49	1
Vadodara	58	14	4
Surat	80	36	0
Rajkot	38	34	0

ii. During June 2014 to November 2015 Mutual funds have investedRs. 100523 crorein equity.

(The above data is Published in the news paper 'Gujarat Samachar' Dt. 3 Dec, 2015.)

The numerical facts in the above data are called statistics, which help us to understand a variety of situations.

According to Croxton and Cowden:

"Statistics is the Science which deals with the collection, analysis and interpretation of numerical data."

Now a day statistics is much more then the numerical datalike averages, medians and percentages, whichare useful in business and usefulto understand economic situations.

It is consideredas important part of scientific methods. Particularly in business and economics the collected facts and figures gives a better understanding to the manager and decision makers of their business and economic environment. Here we consider the use of statistics as an important tool for business and economic decision-making.

For example,

(i) To analyze the data regarding sales and profits of a grocery stores.

(ii) To know compound growth rate of production of a product.

(iii) To forecasts and predict cost of living index numbers, employment rate etc.

To make any kind of statistical analysis the required raw material is 'data'. To make a decision in any business or economic environment a data is required. Any facts and figures collected for analysis and summarized for presentation in raw form or tabulation form can be termed as 'data'.

All the data collected in particular study we referred to as the data set for the study. For example, data regarding book value, earning per share, market price of different securities of different companies. We called such collection as data set.

Here we consider book value, earning per share, market price as variables and measurements obtained for each variable for a particular security is called an observation.

If for a particular security, its book value is Rs. 150, earning per share is 6.5 and market price is Rs. 160 then the set of measures (150, 6.5, 160)) becomes an observation for the first security. We may have such information for number of securities of a particular sector like: banking, pharma, etc.

Qualitative and Quantitative Data.

The units in the data possessdifferent types of characteristics. The measurement of a characteristic vary from unit to unit, such characteristic of a unit is called variable characteristic. It can be either numerical or non numerical.

If the variable characteristic is non numeric then it is called qualitative variable and collection of such observations on such variable is called qualitative data.

Qualitative Data can be grouped byspecific categories. Such data aresometimes referred to as categorical data. Categorical data use either the nominalor ordinal scale of measurement.

We can summarize categorical data by countingthe number of observations in each category or by computing the proportion of theobservations in each category.

For example, 1.types of security: banking, pharma, FMCG. 2. Gender of employee. 3. Education level of employee etc.

If the variable characteristic is numeric then it is called numeric variable and the collection of observations on such variable is

called quantitative data. For example, price, EPS, book value of a security.

Arithmetic operations can be made for quantitative variables. For example,to compute the average value quantitative data may be added and then divided by the number of observations. This average is usually meaningful and easily interpreted. Ingeneral, more alternatives for statistical analysis are possible when data are quantitative.

Primary Data and Secondary Data.

Data used in statistical analysis may be obtained by self study or self investigation or from the other available sources. When the data was collected under the control and supervision of the investigator or by himself for the first time are called "primary data".

When the data was not collected by the investigator, but it derived from the other sources or used from the data collected by other persons then such data is referred to as "secondary data".

In short the data which is primary in the hands of one becomes secondary in the hands of another. For example, Suppose an investigator wants to study the working condition of workers of a big factory. If he collects the data himself by personally or by his agent contacting each worker of the factory and collect the data, then this data is called primary data but if such collected data is used by someone else, then this data becomes secondary data.

The data collected by NSSOI (National Sample Survey Organization), the data of the population census, the data published by office of the registrar general of India, New Delhi regarding birth and death rates in India are primary data. The

same data of birth and death rates is reproduced by UNO (United Nation Organization) in the UN Statistical Abstract then for UNO the same data becomes secondary data.

a) Sources for Primary Data.

Primary data can be collected by either conducting inquiry or by experiments.

1. By Inquiry.

Usually inquiry can be made by three ways (i) direct inquiry (ii) indirect inquiry (iii) method of questionnaire.

(i) Direct Inquiry:

In this method an investigator or his enumerator visits personally to the field of data collection and collect the necessary information. For example, In the earlier example, investigator or his enumerator visits the factory and meet each of the workers of the factory and collect the information regarding their working conditions.

(ii) Indirect Inquiry:

In this method the information is obtained with the help of the third party. For example, in the earlier example, an investigator contacts the manager of the factory and collects the information about the workers directly from the manager.

(iii) Method of Questionnaire:

A list of logically arranged questions relevant to the object of the study, keeping space between the questions for the answer, is called questionnaire. A method of obtaining information using such type of questionnaire is called a method of questionnaire. This method is very much useful when the field under investigation is very wide and respondents are spread over the vide area. Questionnaire may be sent by post, e-mail, or by enumerators. The success of collection of the data depends on

the design of the questionnaire, so questionnaire must be short, simple and ideal for collecting the information.

There is no hard and fast rule for the adoption of a particular method. Depending on time, cost, availability of expert enumerators, the purpose of inquiry etc. an appropriate method may be decided.

2. Experimental Studies.

By conducting actual practical the result or information can be obtained. For example, the data about the effect of a new drug can be obtained by conducting clinical trials. In such experiments, the dose level, effect of drug on different age grouped people such information can be collected.

b) Sources of Secondary Data.

There are two main sources of secondary data: published and unpublished.

1. Published Sources.

There are number of national and international organizations which collect statistical data and publish their findings in their reports or journals periodically. For example, Central Statistical Organization (CSO), National Sample Survey Organization(NSSO), Office of the Registrar General and Census Commissioners of India, Federation of Indian Chamber of Commerce and Industry, The Economic Times, The Financial Express, World Health Organization(WHO), International Monetary Fund(IMF), World Bank etc. From the suchpublisheddata the information about index numbers and exports, vital statistics, agricultural statistics and results of population census can be obtained.

2. Unpublished Sources.

Sometimes the data may be drawn on request from the unpublished internal records of private and public organization which are prepared for their reference. For example, data available from internal company records about employs, production, inventory, sales, customer profile; Ph. D. thesis of various universities etc.

But sometimes the unpublished data may contain error due to their bias computational error, lack of proper sample size etc. Therefore based on suitability, reliability and adequacy of the unpublished data investigator should be proceed with such data for further analysis.

3. Scale of Measurement

To determine the amount of information contained in the data various types of scales on measurements are required. Usually following types of scales of measures are used during the data collection.

(i) Nominal (ii) Ordinal (iii) Interval (iv) Ratio

(i) Nominal Scale:

When the data for a variable consists of labels of names or any qualitative characteristics used to identify an attribute of element a nominal scale is considered. For example, types of the security, name of the security etc. or in the example of working condition of workers of a factory sex, education level, working shift of workers etc.

Usually for data entry into a computer database, a numeric code is used for such category or labels. Like: 1 for male, 0 for female; 1 for banking sector, 2 for pharma sector, 3 for FMCG

sector etc. Here the scale of measurement is nominal even though the data appears as numerical values.

(ii) Ordinal Scale:

Such scale used for the properties of nominal data based on order or ranks if meaningful. For example, working status of the company, the rating may be given as excellent, good or poor such data have the properties of nominal scale, but in addition, the data can be ordered with respect to their quality. The excellent indicate the best working condition of the company, followed by good and then poor. Thus the scale of measurement is now ordinal. Similarly ordinal scale may be used for the risk factor of the company.

(iii) Interval Scale:

If the data have all the properties of ordinal data and the interval between values is expressed in terms of a fixed unit of measure interval scale may be used. For example, salary of worker is an example of interval- scaled data. Since if Rs. 8000, Rs. 12000 and Rs. 25000 are salary of the three workers of a company, they can be ranked or ordered according to higher to low salary. In addition, the difference between the salaries ismeaningful. For instance salary of third worker is 25000 – 12000 = 13000 more than the salary of worker second, while the worker 2 has 12000 – 8000 = 4000 salary more than the first one. In this sense interval scale is used for salary.

(iv) Ratio Scale:

If the data have all the properties of interval data and the ratio of two values is meaningful. For example, salaries, units produced, number of customers, costs etc. This scale requires that a zero value be included to indicate that nothing exists for the variable at the point zero. For example, consider the sales of the units of

the product type-I. A zero value for the sales would indicate that the unit has no sales. In addition, if we compare the sales of 5000 units of the product type I with sales of 2500 units of the product type- II the ratio property shows that the sales of the first product type is double then the sales of the product type-II.

4. Population

A group of all the units under study is called population or universe. For example, a group of all the workers of a factory becomes the population to study the working condition of the workers.

The data representing the complete enumeration of the study is known as population data. For example, the data regarding age, education level, length of service, salary etc. of all the 2000 workers of the factory is called a population data.

The number of units in the population is called size of the population, denoted by N. In case of our example N = 2000. Here the value of N is finite so such population is called the finite population; otherwise we shall call it infinite population.

5. Sample

A set or group of units selected from the population on the basis of some definite criterion is called sample. For example, from the population of workers of the factory, if we select 100 workers by any statistical method then such group of 100 workers is called sample.

The number of units in the sample is called sample size, denoted by n. Here in our case n = 100. The set of observations on such 100 workers is called sample data.

6. Conclusion

Data is a set of facts and figures collected for analysis and summarized for presentation in raw form or tabulation form. The data based on observations of numeric characteristic of variable under study is called quantitative data. The data based on observations of non-numeric characteristic of variable under study is called qualitative data. Data can be obtained through primary source or secondary source. When the data is collected by the investigator himself, it is called primary data. When the data has been collected by others and used it by investigator is known as secondary data. The most important method for primary data collection is through questionnaire. A questionnaire refers to a device used to secure answers to questions from the respondents.Usually four types of scales of measurements are used to determine the amount of information contained in the data. Viz.:

(i) Nominal Scale (ii) Ordinal Scale(iii) Interval Scale

(iv) RatioScale.