"Statistics is a Science that concerns itself with experimentation and the collection, description and analysis of data...Stastical methods are tools for examining data."

– R. A. Hultquist

1

Collection of Data

Content :

1.1 Origin and Growth of Statistics

Ain – I – Akbari

Indian contribution in the field of statistics has been quite significant from the time of Mauryan Empire (321-296 BC). The book 'Ain-I-Akbari' written by Abu Fazal during the time of Mughal Emperor, Akbar (1596-1597 AD), also mentions a highly developed statistical system for administrative and revenue services of the Mughal empire.

The German word 'Statistik' was first used by Gottfried Achen Wall in 1749 for analysis of data about the state. The contributions to statistics prior to 1750 were based mainly on the examples of data analysis without any use of explicit probabilistic ideas.

Abul Fazal

By 18th century, the term statistics was used for systematic collection of data by states. Statistics was formally introduced in Encyclopedia Britanica in 1797. The initial results on probability theory were found in 17th and 18th centuries, whose two giants were Laplace (1749-1827) and Gauss (1772-1855). The first statistical body "The Royal Statistical Society" was founded in1834 in London.

The advanced field of statistics was developed in the late 19th and early 20th century. Galton and Karl Pearson used mathematical statistics in Science, Industry and Politics. Karl Pearson was the founder of mathematical statistics. He founded the advanced statistical laboratory in England and in 1911, the world's first University statistics department at University College, London. In 1910 and 1920 Gosset and Fisher initiated to develop modern statistical science for the data having small samples. Fisher applied statistics to a variety of diversified fields such as genetics, biometry, psychology, education, agriculture, etc. He is well known as father of statistics. During 1930, the refinement and expansion of earlier development was found in the collaborative work between E. Pearson and J. Neyman. Advanced methods of statistics were developed day by day after that.

Growth of Statistics in India

The Indian Statistical Institute (ISI) was founded by the well known Indian Statistician P. C. Mahalanobis in 1931 at Calcutta. He started the first post graduate course in Statistics at Calcutta University in 1941. The Government had approved National Sample Survey proposed by Mahalanobis in 1950 and the first round of data collection took place in October, 1950. Indian Agriculture Statistics Research Institute (IASRI) is another institute which contributed a lot in the development of statistics in India. In I935 W. F. Willcox listed over a hundred definitions of statistics based on the meaning, scope and limitations of the subject. The definition of Statistics according to Croxton and Cowden is as under.

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

Nowadays, statistics is not only useful for quantitative data but also for qualitative data and considered as an independent part of scientific methods. Industrial statistics and a branch of statistics Operations Research (O.R.) was used in the military projects during the second world war. In India, O.R. came into existence in 1949 with the opening of an O.R. unit at the Regional Research Laboratory in Hyderabad. For national planning and survey in I953 at the ISI, Calcutta, P. C. Mahalanobis formulated the second five year plan with the help of O.R. techniques.

O.R. techniques can also be useful for government to maximize the per capita income with minimum resources, in the industry to decide optimum allocation of various resources such as men, machines etc., in marketing to decide the size of the stock to meet the future demand.

Thus, considering the use of Statistics in recent times, its importance cannot be ignored. The constructive development in statistical studies has considerably increased its scope and importance.

1.2 Quantitative Data and Qualitative Data

Population

Sample

Meaning and Differences

In statistics, a group of all the units under study is called a **population**. For example, to study the standard of living of workers of a factory, a group of all the workers of the factory becomes a population for this study. The total number of units contained in the population is called the **size of the population**. If the total number of units of the population is finite, say N, we call it a **finite population**. For example, if 700 workers work in a factory then for this population N = 700 and we shall call such a population as a finite population. A set or group of units selected from the population on the basis of some definite criterion is called sample and the number of units in the sample is termed as the sample size (n). For example, if we select 150 workers from the above population of workers by any statistical method then the group of selected workers is called sample and sample size n = 150. A unit selected in the sample is called a sample unit. The selected workers here are considered as sample units. Inspection of all the units of the population or of the sample is done for a statistical study of a problem covering different aspects. The set of all the observations obtained by such inspection is called **data**.

The units of the population or a sample possess different types of characteristics. It is possible that the measurement of a characteristic varies 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**. We shall call such qualitative characteristic as an attribute. The collection of observations on the attribute is called **qualitative data**. If the variable characteristic is numerical then it is called **numeric variable** and the collection of observations on the numerical variable is called **quantitative data**. In the above example, the data related to attributes like gender of workers, their education level is called the qualitative data and the data related to numeric variables like monthly income of workers, their age is called the quantitative data.

Thus, we shall say that the qualitative characteristic simply can be observed but cannot be measured on a scale, while the quantitative characteristic can be measured on a scale, e.g. income in rupees, age in years etc. Other examples could be the information related to the characteristics like religion, vegetarian or non-vegetarian, education level, etc are called qualitative data whereas the information related to characteristics like profit of a company, cost of advertising etc are called quantitative data.

1.3 Primary and Secondary Data

Meaning

The data obtained by any statistical inquiry must be correct and proper; otherwise no useful and valid conclusions can be drawn.

The data can be primary data or secondary data.

Primary Data

The data originally collected by any authorized agency or investigators for the first time are called **primary data**. For example, the data collected by NSSO (National Sample Survey Organization), the data of the population census of India, the data relating to death rates and birth rates in India published by office of the Registrar General of India, New Delhi are called **primary data.**

Secondary Data

When an investigator or agency uses the data already collected by any other agency or investigator, then such data becomes secondary data for the users. For example, the data regarding birth rate and death rate in India published by office of the Registrar General of India, New Delhi are reproduced by UNO (United Nations Organization) in the UN Statistical Abstract. For UNO, the data becomes secondary data. The data collected by an agency regarding some industries of Gujarat, which are later used by a research student for statistical analysis, become secondary data for the research student.

Difference Between Primary and Secondary Data

(1) Primary data are obtained for the first time and it is original, whereas secondary data are not original but it is the collected data reused by others.

(2) Primary data are collected from the units under inquiry, whereas secondary data are obtained from the primary data.

(3) Primary data are usually in raw form. Hence, they are required to be classified and tabulated, whereas secondary data are usually in classified and tabulated form.

(4) Collection of primary data requires large amount of money, time and energy, whereas secondary data require comparatively very less amount of money, time and energy.

(5) Primary data are almost reliable; whereas secondary data are obtained by individuals or institutions for their own interest and hence may not be reliable or relevant from the point of view of user's interest or purpose.

1.4 Methods of Collecting Primary Data

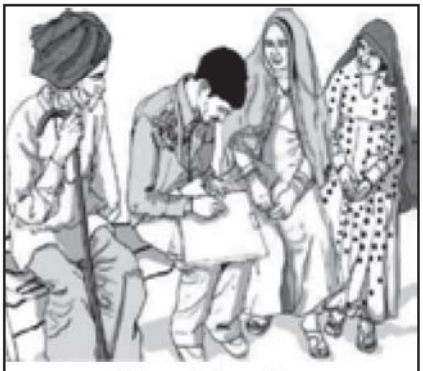
There are many popular methods for the collection of primary data. Commonly used methods are direct inquiry, indirect inquiry, inquiry by telephone or E-mail, method of questionnaire, inquiry by local news reporters.

We shall study the following three methods of collecting primary data :

(1) Direct inquiry (2) Indirect inquiry (3) Method of questionnaire.

1.4.1 Direct Inquiry

Meaning :



Direct inquiry

In this method an investigator himself or an enumerator appointed by him visits personally to the field and collects the necessary information. For example, if a person wants to collect data about the condition of farmers of Sanad taluka of Ahmedabad district, he must go to Sanand taluka, contact the farmers and obtain the data regarding their condition. The information obtained by this method is called data collected by direct inquiry. Similarly, a person interested in knowing the health status of the students of municipality schools of Ahmedabad city, personally meets the students of the municipality schools and asks questions pertaining to their health and collects the necessary information given by them. This is also called direct inquiry.

Advantages:

(1) The data obtained by this method are accurate and reliable.

(2) Due to face to face interviews with respondents, enumerators can solve any question or doubt or any confusion by giving proper explanation.

(3) By personal interview enumerator sometimes gets supplementary information of the respondents which may be useful at the time of interpretation of the results.

(4) This method is much better when the field of inquiry is limited.

(5) Sometimes when respondents are unwilling to give particular type of information, the enumerator can get such information indirectly by asking supplementary questions and giving assurance against the misuse of the information provided by them.

Disadvantages:

(1) This method requires more time, more number of enumerators when the field of inquiry is very wide.

(2) This method is more expensive.

(3) The information collected by this method may be less reliable if the enumerators are untrained.

Activity

Collect the primary data to study any five health characteristics of the students of your school. Identify the qualitative data and quantitative data from this collected data.

1.4.2 Indirect Inquiry

Meaning :

In any of the following situations, the investigator contacts an organization or a person capable of conducting inquiries and gets relevant information from such organization or person.

(1) When direct source does not exist.

(2) When the field of direct inquiry is very large and there is a lack of time for personal contact with the respondents.

(3) When the information to be obtained is complex.

(4) When the cost associated with direct inquiry is unbearable.

The method of collecting information in such a way is called indirect inquiry. In this method, the information is obtained with the help of the third party instead of enumerators. For example, in the earlier illustration, a person interested in getting the information about the conditions of farmers of Sanand taluka, approaches the talati of Sanand taluka who has the entire information of all the farmers in Sanand taluka, and gets information from him, instead of personally visiting each farmer.

A person interested in information of the educational status of the workers of an industry, having 2000 workers, approaches the manager of the industry who has complete records of the workers of the industry, instead of making personal visit to each worker. Similarly, advocate can get information regarding murders, thefts from the police station when needed.

Thus, the method of indirect inquiry is quite popular in practice; however the reliability of the information obtained by this method depends on the ability, honesty and experience of the enumerators who collect the information.

Advantages :

(1) This method is most suitable when the field of inquiry is very wide.

(2) When a respondent associated with the inquiry is not willing to give information, this is the only useful method to collect the information.

(3) It involves less cost, time and energy as compared to direct inquiry method.

(4) This method is very much useful for various departments of government to collect various types of information.

Disadvantages :

(1) When individual or organization associated with the task of providing information have any prejudices or a biased attitude, the information becomes less reliable.

(2) The method becomes useless when the third party from whom the information is obtained, is dishonest, inefficient or unable to provide correct information.

1.4.3 Method of Questionnaire

A list of logically arranged questions relevant to the object of the study is made. The space between the questions is provided for the answers. A list of questions prepared in such a way is called questionnaire. The method of obtaining information using such type of questionnaire is called a method of questionnaire.

The questions in the questionnaire should be short and simple so that respondent can understand the questions and easily answer them. This method is quite useful when the field or area from which the information is to be collected is very wide. Since there is considerable saving of time and cost in this method, it is the most economical method of inquiry. Questionnaire may be used in direct inquiry as well as indirect inquiry.

There are two ways of collecting information by questionnaire : (1) By post (2) By enumerators. Before discussing the above two methods in the next section, we shall discuss the points which should be kept in mind for drafting an ideal questionnaire.

1.4.3.1 Characteristics of an Ideal Questionnaire

The success of collecting data depends mainly on the design of questionnaire. A well designed questionnaire is known as an ideal questionnaire. The following points should be kept in mind while drafting an ideal questionnaire:

(1) Every questionnaire must have a covering letter or an appropriate title so that the reader gets clear idea about the purpose of the study.

(2) Number of questions should be as minimum as possible. The fewer number of questions have greater chance of getting better response from the respondent. Long questions discourage the respondent to provide accurate answer particularly towards the end of questionnaire. There is no specific rule for exact number of questions. If the number of questions is large, say 25 or more, it is advisable to divide the questionnaire into various parts to ensure clarity of questions.

(3) The number of questions in the questionnaire should be consistent with the purpose of the inquiry.

(4) Questions should be simple, short and understandable. They should not be ambiguous and dual in meaning. For example, if the question is ' Are you educated?', the respondent cannot understand the meaning of 'educated'. Does it mean the education up to secondary or higher secondary or graduate? Instead, the question should be 'What is your education level ?' Following five options can be considered for the answer to this question.

(a) Upto primary (b) Upto secondary (c) Upto higher secondary (d) Upto graduate (e) Uneducated.

(5) The arrangement of the questions should be logical. The order of the questions should be from general to specific. For example, (a) What is your opinion about the 'smart phone'? (b) Do you use 'smart phone'? This is the improper arrangement of the questions. The correct order is question (b) followed by question (a).

(6) Questions of sensitive nature or personal life should be avoided. For example, questions relating to marital status, other sources of income of respondent, etc. If such questions are necessary in the survey, an assurance should be given to respondent that the information provided by him/her will be kept confidential and will not be used for any other purpose.

(7) Answers to the questions should be free from any calculations. For example, avoid the question like 'What is the Average income of the earning members of your family ?'

(8) The questions having answers YES/ NO or multiple alternatives should be asked. But the questions whose answers do not have specific options should be avoided. For example, what is your opinion about the semester system against the yearly system in the education? Since there are a lot of variations in the answers like: (a) Strongly in favour (b) In favour (c) Neutral (d) Against (e) Strongly against. They are difficult to tabulate or interpret. Hence if necessary, there must be minimum number of such questions in the questionnaire.

Let us see an example of a questionnaire.

A specimen questionnaire for collecting information of the study habits of 12-th standard students of Gujarat state :

1. Name of the student : _____
2. Sex : ☐ Female ☐ Male
3. Place : _____ Village/ city : _____ District : _____
4. School : ☐ Not taking government grant ☐ Taking government grant
5. Stream : ☐ Arts ☐ Commerce ☐ Science ☐ Others
6. Medium of study : ☐ English ☐ Gujarati ☐ Others
7. School Time : _____
8. How much time do you spend for your daily school home work ? _____
9. Average time spent daily for reading and preparation : _____
10. Mode of entertainment :
    ☐ Sports ☐ Film or TV ☐ Music ☐ Spending time with family members
11. How much time do you spend for entertainment ? _____
12. Average hours of daily sleeping time : _____

Activity

Prepare a suitable questionnaire to collect the data regarding the popularity of different types of chocolates among the students of your school.

1.4.3.2 Questionnaire by Post

Meaning :

After preparing an ideal questionnaire, it is dispatched by post to respondents from whom the information is to be collected along with a request to complete and return it in a given period of time. Therefore, it is necessary to send a blank envelope along with the questionnaire to all the respondents bearing the address of investigator and a proper postal stamp. Since the answers to the questions are to be given by respondents themselves, it is essential that the questions in questionnaire are clear, short, simple, relevant and self explanatory. Now a days, questionnaire may be sent through computer (E-mail) and mobile phone also.

Advantages :

(1) This method of collecting data is quite useful when the field of investigation is very vast.

(2) This method is simple and it provides large amount of information with less expense.

(3) By this method, investigator can get information from respondents of those areas where it might be difficult to reach personally or by telephone.

Disadvantages :

(1) When respondents are illiterate or non cooperative, this method becomes useless.

(2) Sometimes information provided by the respondents may not be correct or they may return incomplete or blank questionnaire to the investigator.

(3) There may be a loss of questionnaire or the delay in getting the information because of the laziness of the respondent.

(4) There is a lack of assistantance or a person who clarifies the instructions or gives explanation of questions that may arise. Hence, there is a possibility of misinterpretation of questions.

1.4.3.3 Questionnaire by Enumerator

Meaning :

In this method, enumerators themselves contact the respondents and get response to the questions. Enumerator himself fills the questionnaire. Thus, the difference between the earlier method and this method is that in earlier method the questionnaire is sent to the respondents by post, whereas in this method the enumerators carry the questionnaire and personally meet the respondents. Here the enumerators should be enthusiastic, polite, honest and efficient in their work so that they can extract correct answers to the questions by providing relevant and supplementary information to the respondents. The enumerators are also instructed to create friendly atmosphere without entering into any controversy or showing any disrespect towards the respondents.

Advantages :

(1) The complete, correct and more relevant information can be obtained from the respondents by selecting proper enumerators.

(2) Enumerator can get proper information by giving proper explanations to the respondents who are illiterate or not cooperative.

(3) In this method, there is no problem of loss of questionnaire or incomplete answers to the questions.

(4) The information obtained is more reliable.

Disadvantages :

(1) Sometimes it is quite difficult to get expert enumerators in sufficient number.

(2) Due to the large number of enumerators, the remuneration paid to enumerators increases the total cost of inquiry.

(3) It is a very difficult task to train the untrained enumerators for the inquiry. Even after training, it is not proper to believe that they will work skillfully, honestly or efficiently. Hence, sometimes the information collected by them may not be reliable.

(4) This method becomes unsuitable when the respondents are spread over a wide area. In these circumstances, the cost and time of inquiry increases.

(5) Enumerators have to adjust with the convenient time of respondents or sometimes enumerators have to visit two to three times to the same respondent. As a result, it becomes difficult to finish the inquiry within the stipulated time.

Among the three methods of collecting primary data discussed earlier, there is no definite rule for using a particular method. The selection of a method of inquiry mainly depends on the following aspects:

(1) Spread of the area under inquiry (2) The purpose of inquiry (3) The monetary provision (4) The time frame (5) Possibility of getting expert enumerators and (6) The standard of accuracy.

1.5 Secondary Data

1.5.1 Sources of Secondary Data

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

Secondary data from published sources :

(1) Government publications : There are number of central and state government organizations which collect statistical data and publish their findings in the interest of the public. Some government publishers provide data on a periodical and regular basis. For example, Central Statistical Organization (CSO), National Sample Survey Organization (NSSO), Office of the Registrar General and Census Commissioners of India, Indian council of Agricultural Research (ICAR), Statistical Abstract of Gujarat State, Statistical Outline; Gujarat State, Socio-Economic Review; Gujarat State, etc.

From the published data of such government bodies the information about wholesale price index numbers, import-export, vital statistics, agricultural statistics, results of population census can be obtained.

(2) Semi-Government Publications: : Semi-Government organizations such as Life Insurance Corporation of India, State Electricity Boards, etc. regularly publish various types of important data.

(3) International Publications : International organizations like United Nations Organizations (UNO), International Monetary Fund (IMF), International Labour Organization (ILO) publish their important data.

(4) Reports of Research Organizations : Research institutes like Ahmedabad Textile lndustry's Research Association (ATIRA), Physical Research Laboratory (PRL); Ahmedabad, Salt and Marine Research Laboratory; Bhavnagar, Institute of Economic Growth; Delhi, National Council of Applied Economic Research; New Delhi provide data in their publications.

(5) Local Self Government Institutions and Autonomous Educational Institutions, Municipal Corporation, Jilla panchayats and agricultural universities publish their annual reports.

(6) Publications of Business and Commerce Organizations : Federation of Indian Chambers of Commerce and Industry publishes the journal 'Economic Trends', Institute of Chartered Accountants publishes the journal 'The Chartered Accountant' and The Institute of Foreign Trade publishes the journal 'Foreign Trade Review.'

(7) News papers and Periodicals : The data regarding economics, commerce, business, sports, etc. can be collected from the different news papers and periodicals like Economic and Political Weekly (EPW), Commerce, Business Today, Financial Express etc. These are also important sources of getting secondary data.

Secondary Data from Unpublished Sources :

Some of the statistical data may not be published. Sometimes the data may be drawn on request from the unpublished internal records of private and public organizations which are prepared for their reference. For example, salary of employees, their length of service, their educational level, data regarding investment of mutual funds in public and private sector companies, Ph.D. theses of various universities, etc.

1.5.2 Precautions while using Secondary Data :

A careful scrutiny must be done before using secondary data. The reason is that such data may be erroneous of inadequate sample size or may not be suitable with the purpose of the inquiry. Such secondary data may not be useful for statistical analysis, drawing conclusions and inferences from the analysis. Therefore, before using such data, the following precautions should be taken :

(1) Before using the secondary data, it should be verified as to who has collected the data and from where the data is collected. The data obtained from the records of private organizations may be less reliable as private organizations collect data according to their own ideology and prejudices.

(2) The purpose of collecting data must be relevant to the purpose of the study; otherwise the data will be of no use.

(3) The collected data must not be too old. It should be relevant with the current time period. For example, price of grains, gold, petrol etc.

(4) Before using the secondary data, the matters regarding scope of the data, region for which the data are collected and definition of the terms used in the data should be ascertained.

(5) Direct use of the estimated data should be avoided. The estimates given in the data may be wrongly calculated.

(6) The method of collection of the data should be known so that investigator becomes familiar with its advantages and disadvantages. To get proper benefits of secondary data, it should be used with the above precautions.

Summary

- Data is a set of observations expressed in quantitative or qualitative form.
- 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 by investigator then it is known as secondary data.
- The most important method for collecting primary data is by questionnaire method.
- A questionnaire refers to a tool used to get answers of questions from the respondent.

EXERCISE 1

Section A

For the following multiple choice questions choose the correct option.

1. Who used the German word 'Statistik' for the first time ?

(a) John Graunt (b) Willian Patty (c) Gottfried Adren Wall (d) Gauss

2. Who was one of the giants of initial results of probability theory among the following ?

(a) John Graunt (b) Lapalce (c) Fisher (d) J. Neyman

3. Who was the founder of mathematical statistics ?

(a) Karl Pearson (b) Laplace (c) Mahalanobis (d) Gosset

4. Out of the following, which one is an example of primary data ?

(a) Data collected from the records of Municipality.

(b) Data Collected from a published journal of an industry.

(c) Data collected from website.

(d) Data collected by NSSO.

5. Which one of the following is an example of qualitative data ?

(a) Income category (b) Production (in tons)

(c) Age of workers (in year) (d) Height of persons (in meter)

6. Which one of the following is true for secondary data ?

(a) Should never be used.

(b) Use after careful verification.

(c) It is not necessary to check while using it.

(d) Secondary data itself is a primary data.

7. Which one of the following is true for primary data ?

(a) Primary data is always more reliable as compared to secondary data.

(b) Primary data is less reliable as compared to secondary data.

(c) Primary data depends on whether the data is collected carefully or not.

(d) Primary data can be obtained from the govemment publications.

8. Which of the following statements is true ?

(a) The data collected by direct inquiry may be more accurate.

(b) The data collected by direct inquiry may be less accurate.

(c) The data collected by direct inquiry may not be reliable.

(d) The data obtained through e-mail is known as the data obtained by direct inquiry.

9. Which of the following is a proper method of getting supplementary information about the personal characteristics of the respondents ?

(a) Questionnaire by post (b) Direct inquiry (c) Indirect inquiry (d) From the news papers

10. Which method will be costly when the number of respondents are more and spread over the large area ?

(a) Questionnaire by post (b) Indirect inquiry (c) Direct inquiry (d) By telephone

Section B

Answer the following questions in one sentence :

1. Who was the founder of Indian Statistical Institute?
2. Define population.
3. Define sample.
4. Define qualitative data.
5. Define quantitative data.
6. Define primary data.
7. Define secondary data.
8. State the methods of collecting primary data.

Section C

Answer the following questions.

1. State the definition of statistics given by Croxton and Cowden.
2. What is data ?
3. What is questionnaire ?
4. What is unpublished data ?
5. What is a variable characteristic ?
6. What is an attribute ?

Section D

Answer the following questions.

1. What is the role of P. C. Mahalanobis in development of statistics in India?
2. State the difference between qualitative and quantitative data.
3. Give some examples of primary data.
4. Discuss the method of questionnaire.
5. Discuss questionnaire by post.
6. Discuss questionnaire by enumerators.
7. Describe the method of collecting secondary data from unpublished sources.
8. Discuss some applications of statistics.

Section E

Answer the following questions.

1. State the difference between primary and secondary data.
2. Discuss the method of collecting primary data by direct inquiry.
3. Discuss the method of collecting primary data by indirect inquiry.
4. Discuss the method of collecting secondary data.

5. Discuss origin and growth of statistics.
6. Discuss advantages and disadvantages of direct inquiry.
7. Discuss advantages and disadvantages of indirect inquiry.
8. Discuss the characteristics of an ideal questionnaire.
9. Discuss advantages and disadvantages of questionnaire by post.
10. Discuss advantages and disadvantages of questionnaire by enumerators.
11. Discuss what types of precautions should be taken while using secondary data.

● Historical Note ●

The term ' Statistics' is derived from the Latin word " Statisticum Collegium" (means council of State) and the Italian word "' Statista" (means Statesman or politician). John Graunt and William Petty developed statistical and census methods in 1662 to analyze the bills of mortality.

Some Indian Statisticians who have made a significant contribution in the development of Statistics are : Prof. C. R. Rao, Prof. R. R. Bahadur, Prof. D. Basu, Prof. D. Lahiri, Prof. K. R. Nair, Prof. P. V. Sukhatme, Prof. S. K. Mitra, Prof. R. C. Bose, Prof. S. N. Roy, Prof. N. M. Bhatt, Prof. C. G. Khatri, etc.

P. C. Mahalanobis
(1893 - 1972)

Prasanta Chandra Mahalanobis, popularly known as P. C. Mahalanobis, was an Indian statistician. He devised the Mahalanobis distance, a measure of distance between two populations. It is a fundamental concept in multivariate analysis. He was instrumental in formulating India's strategy for industrialization in the Second Five–Year Plan (1956–61). He founded the Indian Staistical Institute in Kolkata on December 17, 1931.

With the objective of providing comprehensive socioeconomic statistics, Mahalanobis became the pioneer of the establishment of the National Sample Survery in 1950 and also of the Central Statistical Organization to coordinate statistical activities in India. He served as the chairman of the United Nations' Sub-Commission on sampling from 1947 to 1951 and was appointed as the honorary statistical advisor to the Government of India in 1949.

For the pioneering work, he was awarded the Padma Vibhushan, one of India's highest honours, by the Indian government in 1968. A postage stamp was issued by Goverment of India with his picture.

●