

Statistics

Contribution of Mahalanobis to the development of sample survey theory

1. Introduction

By the end of this session, you will be able to:

- ♦ **Explain about Prasanta Chandra Mahalanobis and his early life**
- ♦ **Explain his contribution to sample survey and its impact**
- ♦ **Explain the contributions to human welfare**
- ♦ **Explain the later life of P C Mahalanobis**

Prasanta Chandra Mahalanobis was an Indian scientist and applied statistician. He is best remembered for the Mahalanobis distance, a statistical measure. He made pioneering studies in anthropometry in India. He contributed to the design of large-scale sample surveys.

He was the chief architect of the post-independence statistical system in India. He was appointed as the Honorary Statistical Adviser to the Government of India in 1949. A Central Statistical Unit was set up in 1949 to function under his technical guidance. In 1951, the Central Statistical Organization (CSO) was established to coordinate the statistical activities in various ministries and other governmental agencies.

Around the same time, the Standing Committee of the Departmental Statisticians as well as the National Income Committee felt the urgent need for quality data and recommended a sample survey covering rural areas of India.

Thus, the National Sample Survey started its operations in October 1950 under the leadership of Mahalanobis, the technical expertise being provided at the Indian Statistical Institute (ISI). Thus, both the apex statistical bodies of independent India, the Central Statistical Organization (CSO) and the National Sample Survey (NSS) and the State Statistical Bureaus are the creations of Mahalanobis.

2. Early Life of Mahalanobi

Mahalanobis belonged to a family of Bengali landed gentry who lived in Bikrampur (now in Bangladesh). His grandfather Gurucharan (1833–1916) moved to Calcutta in 1854 and built up a business, starting a chemist shop in 1860.

Gurucharan was influenced by Debendranath Tagore (1817–1905), father of the Nobel-prize-winning poet, Rabindranath Tagore. Gurucharan was actively involved in social movements such as the Brahmo Samaj, acting as its Treasurer and President. His house on 210 Cornwallis Street was the center of the Brahmo Samaj.

Gurucharan married a widow against social traditions. His elder son Subodhchandra (1867–1954) was the father of P. C. Mahalanobis.

He was a distinguished educationist who studied physiology at Edinburgh University. He became a Professor at the Presidency College and later became the head of the department of Physiology.

Subodhchandra also became a member of the Senate of the Calcutta University. Born in the house at 210 Cornwallis Street, P. C. Mahalanobis grew up in a socially active family surrounded by intellectuals and reformers.

Mahalanobis received his early schooling at the Brahmo Boys School in Calcutta graduating in 1908. He then joined the Presidency College, Calcutta and received a B.Sc degree with honors in physics in 1912.

He left for England in 1913 to join the University of London. He however missed a train and stayed with a friend at King's College, Cambridge. He was impressed by the Chapel there and his host's friend M. A. Candeth suggested that he could try joining there, which he did. He did well in his studies, but also took an interest in cross-country walking and punting on the river. He interacted with the mathematical genius Srinivasa Ramanujan during the latter's time at Cambridge.

After his Tripos in physics, Mahalanobis worked with C. T. R. Wilson at the Cavendish Laboratory. He took a short break and went to India and where he was introduced to the Principal of Presidency College and was invited to take classes in physics.

He went back to England and was introduced to the journal *Biometrika*. This interested him so much that he bought a complete set and took them to India. He discovered the utility of statistics to problems in meteorology, anthropology and began working on it on his journey back to India.

In Calcutta, Mahalanobis met Nirmalkumari, daughter of Herambhachandra Maitra, a leading educationist and member of the Brahmo Samaj. They married on 27 February 1923 although her father did not completely approve of it. The contention was partly due to Mahalanobis opposition of various clauses in the membership of the student wing of the Brahmo Samaj, including restraining members from drinking and smoking.

3. Contributions to Sample Surveys

According to C R Rao, *"the fame of Mahalanobis as a scientist will rest largely on his contributions to statistics. He viewed Statistics, or more generally collection and processing of information, as essential in seeking truth"*.

On the work done at the ISI, Fisher remarked: *"... The ISI has taken the lead in the original development of the technique of Sample Surveys, the most potential fact finding process available to the administration"*.

In particular, Mahalanobis innovative techniques and methodology for large scale sample surveys are widely acknowledged throughout the world. On the complexity of the multipurpose, multi-subject framework of NSS, Deming commented thus: *"... No country, developed, under-developed or over-developed, has such a wealth of information about its people as India...."*

During the early twenties, a civil servant, J A Hubback was dissatisfied with the inadequate and defective methods of collection of data on crop acreage and yields and devised crop cutting experiments which he termed 'a random sampling method'.

Mahalanobis was influenced by Hubback's method and adopted random cuts for estimating the acreage under jute crop in Bengal during the early forties. Mahalanobis's work was assessed by H Hotelling in his report submitted to the Indian Central Jute Committee as follows: *"... no technique of random*

sample has, so far as I can find, been developed in the United States or elsewhere, which can compare in accuracy or in economy with that described by Professor Mahalanobis...”

Mahalanobis sample survey estimate of jute production was 7 thousand 540 bales (1 bale is equal to 400 pounds). This was obtained at a cost of 8 lakh rupees with a work force of 600 while the plot to plot enumeration yielded a figure of 6 thousand 304 bales at an expenditure of 82 lakh rupees and 33 thousand employees which turned out to be an underestimate by 16 point 6 per cent. This was evidenced by the alternative customs and trade figure of 7 thousand 562 bales.

During the early period 1937 to 45, Mahalanobis introduced several innovative ideas and methodologies which he named as ‘experiments in statistical sampling’.

He mainly dealt with the problems of organization which arise when a sample survey has to be carried out on a very large scale. Just as the large-scale commercial production of a chemical is a matter of chemical engineering rather than of pure chemistry, the organization of large scale sample surveys was equated to ‘statistical engineering’ rather than pure theory of sampling by Mahalanobis.

These experiments covered wide ranging areas such as:

- ◆ Acreage and total production of important food and fiber crops
- ◆ Economic or demographic factors relating to indebtedness, unemployment, birth and death rates etc. of rural families
- ◆ Cost and level of living, consumption of food, clothes etc.
- ◆ Preferences to particular commodities
- ◆ Public opinion
- ◆ After-effects of Bengal famine
- ◆ Traffic surveys
- ◆ Demand for currency coins and average life of currency notes
- ◆ Study of bark yield using regression estimates etc.

The above mentioned experiments in sampling stressed the importance of ‘cost function’ and ‘variance function’. Obviously, as the sample size increases the variability in the estimates decreases but the cost of the survey tends to increase. Thus, a balance has to be struck which leads to an ‘optimum’ sample size. ‘Pilot surveys’ played an important role in determination of sample size as well as for testing the schedules and field conditions, estimating the time and cost for survey and the variability.

When the units are of varying size, Mahalanobis was aware of sampling of units with probabilities of selection proportional to their sizes instead of equal probabilities in 1937 itself. He, however, abandoned this method of cumulating the sizes and Recording the cumulative totals due to constraints of work load. The mathematical theory for Probability Proportional to Size Sampling (PPS) method was later given by Hansen and Hurwitz in 1943.

Mahalanobis in collaboration with D B Lahiri of the NSS presented a detailed analysis of errors in censuses and surveys in the Indian context. The technique of Inter Penetrating Network of Subsamples (IPNS) developed by Mahalanobis during the thirties, consists of drawing the sample in the form of two or more sub-samples, selected according to the same sampling scheme so that each subsample provides a valid estimate of the parameter of interest.

This technique helps in providing ‘a means of control (i.e. appraisal) of the quality of the information’, by way of securing information on non-sampling errors. This technique used by Deming as ‘replicated sampling’ was acknowledged to have ‘simplicity in calculation of the standard error of the estimate’ besides ability ‘to detect gross blunders in selection, recording and processing’.

4. The Impact

The three notable contributions to sample survey methodology by Mahalanobis, namely ‘pilot surveys, concept of optimum survey design, and inter penetrating network of subsamples (IPNS)’ had a great impact on the present day sampling techniques in particular and statistical methods in general. For example, pilot surveys are acknowledged as a prelude to Abraham Wald’s ‘sequential analysis’, which relates to decision making sequentially.

‘Optimum survey design’ stresses the Mahalanobisian philosophy that all the resources provided for a survey should be used optimally going beyond the mathematical propositions such as ‘sampling error should be minimized for a fixed cost’, or ‘cost should be minimized for a fixed sampling error’. This can be considered to be a precursor to the present day ‘operations research’ philosophy.

The IPNS technique, while being a tool for assessing and controlling the non-sampling errors in the survey also ‘permits evaluation of variances between investigators, coders and other workers in the various stages of processing’. Thus, IPNS technique could be considered really as the curtain-raiser for ‘re-sampling procedures’ like Bootstrap.

It is interesting to note that as early as in 1937 itself, Mahalanobis had considered the possibility of air surveys ‘using specially sensitized films’ for estimation of crop acreage. This is exactly what is done through ‘remote sensing’ satellites nowadays.

The method of collection of data for the ‘United Provinces Anthropometric Survey’ of 1941 organized by Mahalanobis, D N Majumdar and C R Rao was described as follows:

“The random selection of a sample is not an easy task. To pick up a sample in a demonstrably rigorous fashion requires elaborate preliminary arrangements, which are not often possible in practice. In this situation, rough and ready methods have to be used”.

The anthropologist, D N Majumdar ‘collected all healthy males between the ages of 18 and 48 (belonging to the caste or tribe under survey) who happened to be available, arranged them in serial order just as they came and picked up either the odd or the even numbered individuals for measurement.

Summing up, Mahalanobis, Majumdar and Rao in their paper in *Sankhya*, 1948 say “*The present samples may therefore be treated as having been drawn, for all practical purposes, at random. As far as one can judge, the assumption of randomness is more true of the present material than any other series of anthropometric measurements so far available in India*”.

Majumdar’s selection involved a ‘random start’ from 1 and 2 and selection of a ‘systematic sample’ with the ‘sampling interval’ equal to two, a technique which has been developed by Madow and Madow in 1944. The ‘arrangement’ of individuals by Majumdar ‘in a serial order just as they came in’ makes systematic sampling equivalent to ‘simple random sampling’ as stressed by the authors.

Later this was also formally shown by Madow and Madow and led to the concept of 'random permutations model'. The use of random permutations model in drawing inferences in sampling from finite populations is very well known.

Following the U P Anthropometric Survey, D N Majumdar and C R Rao carried out the Bengal Anthropometric Survey in 1945 on similar lines. In the foreword to the paper on the statistical study published in *Sankhya* in 1958 by D N Majumdar and C R Rao, Mahalanobis defines 'group' as individuals belonging to the same caste, religion or tribe and living in the same district, giving rise to a two-way classification.

He then points out that any further sub-division on the basis of sub-caste, clan, endogamy etc. would have resulted in a large number of such groups for which a much larger survey would be necessary to obtain sufficient number of individuals under each sub-group for *proper statistical analysis*.

This also arises in the problems of *small domain estimation, a technique which is of prime importance* in the present day context in view of decentralized planning and data dissemination in several countries.

Even from late 40's, controversies existed regarding the size of cuts to be used for crop cutting experiments followed by the Indian Council of Agricultural Research (ICAR) and ISI. Mahalanobis experimented on different sizes of cuts and preferred circular cuts of radius 4 feet for yield surveys, while ICAR under Panse's guidance was using rectangular cuts of size 33 feet \times 16.5 feet for surveys conducted by state agencies.

Panse emphasized that any sampling method must fit well into the existing administrative set up, while Mahalanobis stressed on well-trained investigators specially recruited for the survey.

B P Adhikari connects this to the observation that Panse and Mahalanobis belonged to two different social backgrounds – one from a place where the Moghul system of revenue collection had its influence and the other where the British system of revenue collection was in vogue.

Also, joint studies of the Ministry of Agriculture, CSO and ISI conducted in 1960-61 and the studies by a technical committee set up by the Planning Commission conducted on four crops during 1963-66 did not reveal any significant differences between the two types of cuts.

5. Contributions towards human welfare and later life of P C Mahalanobis

Mahalanobis believed that the ultimate analysis of statistics has one single aim: '*to improve the efficiency of action programmes for the welfare of humanity*'. In 1971 he observed: "*The use of sample surveys is spreading rapidly in underdeveloped countries But the danger still remains of much waste of resources in work which is highly imitative of advanced countries ...*"

He was always conscious of costs in a survey and insisted that "*from the statistical point of view our aim is to evolve a sampling technique which will give, for any given total expenditure, the highest possible accuracy in the final estimate.*"

In contrast to Neyman, Mahalanobis derived the optimum allocation in stratified sampling when costs varied from stratum to stratum. Lahiri comments that Mahalanobis was of the view that Neyman was not conscious about costs because the latter was possibly not thinking of stratification for purposes of field

surveys, where costs might differ from stratum to stratum, but only for sample selection in office for sample tabulation in which case the costs would not vary from stratum to stratum.

On the occasion of the birth centenary of Mahalanobis in 1993, the Government of India released a postage stamp bearing his picture and the Institute he founded in a fitting recognition to his fundamental contributions to statistics towards human welfare and national development.

C R Rao thinks that, perhaps, this is the only instance where a statistician has been honored with the issue of a postal stamp bearing his or her picture. As per the norms, government buildings are rarely named after individuals but for a few rare exceptions; it is indeed a worthy exception to name the building of the National Sample Survey Organisation at its Calcutta Head Quarters as 'Mahalanobis Bhavan'.

6. Later life of P C Mahalanobis

In later life, Mahalanobis was a member of the planning commission contributed prominently to newly independent India's five-year plans starting from the second. In the second five-year plan he emphasised industrialization on the basis of a two-sector model.

His variant of Wassily Leontief's Input-output model, the Mahalanobis model, was employed in the Second Five Year Plan, which worked towards the rapid industrialization of India and with other colleagues at his institute, he played a key role in the development of a statistical infrastructure.

He encouraged a project to assess deindustrialization in India and correct some previous census methodology errors and entrusted this project to Daniel Thorner.

Mahalanobis also had an abiding interest in cultural pursuits and served as secretary to Rabindranath Tagore, particularly during the latter's foreign travels, and also worked at his Visva-Bharati University, for some time. He received one of the highest civilian awards, the Padma Vibhushan from the Government of India for his contribution to science and services to the country.

Mahalanobis died on 28 June 1972, a day before his seventy-ninth birthday. Even at this age, he was still active doing research work and discharging his duties as the Secretary and Director of the Indian Statistical Institute and as the Honorary Statistical Advisor to the Cabinet of the Government of India.