JOURNAL OF
HUMAN ECOLOGY

J. Hum. Ecol., 4(3): 233-235 (1993)

# P.C. Mahalanobis' Contributions to Biometry

J.K. Ghosh and Partha P. Majumder
*Anthropometry and Human Genetics Unit, Indian Statistical Institute, Calcutta 700 035, West Bengal, India*

Prasanta Chandra Mahalanobis was born on 29 June 1893 in Calcutta (India). He graduated with honours in Physics from Presidency College, Calcutta, in 1912. He went to England in 1913 and completed the Tripos in Mathematics and Natural Science (Physics) from King's College, Cambridge, within a short period of a year and eight months. In Part II of the Tripos, he was the only candidate to get a first class in Physics. The Cambridge University awarded him a research scholarship. Mahalanobis decided to work with C.T.R. Wilson (of Cloud Chamber fame) at the Cavendish Laboratory. Before starting his research work, Mahalanobis returned to India for a short vacation in July 1915. Just before he left England for this vacation, Mahalanobis' tutor, W.H. Macaulay, drew his attention to the journal *Biometrika*. Mahalanobis found some of the papers, in *Biometrika* very interesting, and purchased a whole set of volumes and brought them back with him to India. These volumes were later to play a decisive role in transforming Mahalanobis-the-physicist to Mahalanobis-the-statistician.

Upon request from the Principal of the Presidency College, Mahalanobis started to teach Physics during his vacation in Calcutta. Soon he found plenty of things in India to hold his interest and absorb all his working time. he decided not to return to England.

Early on, one of his mentors, Acharya Brojendranath Seal, said to him "Prasanta, .. you have to do work in India similar to that of Karl Pearson in England. In today's world, whether it is science or social service, without statistical methods there is no way. This is your job". (Translated from a note in Bengali by P.C. Mahalanobis dated 17 April 1945). Mahalanobis took up this challenge seriously, and started to carefully read Karl Pearson's papers published in *Biometrika*. He developed an interest in biometry, which was to last throughout his life and to which he was to make very profound contributions.

During the annual meeting of the Indian Science Congress Association in 1920, Mahalanobis met the Director of the Zoological and Anthropological Survey of India, N. Annandale. Annandale requested Mahalanobis to analyse some anthropometric data on a group of Anglo-Indians of Calcutta collected by him. Mahalanobis performed the analyses and published his first paper on statistics (Mahalanobis, 1922). He continued analyses of the other anthropometric measurements on this group, and presented a synthesis of the results in his Presidential Address to the Anthropology section of the Indian Science Congress in 1925. In this address titled "Analysis of race-mixture in Bengal", Mahalanobis sought to provide statistical answers to anthropological questions such as: Do the Anglo-Indians show a greater affinity with the higher castes of Bengal or with the lower castes ? Is there any appreciable admixture with the aboriginal tribes ? For answering such questions, obviously a measure of distance between population groups based on anthropometric measurements was necessary.

This note is written on the occasion of P.C. Mahalanobis' birth centenary. For an account of his many other contributions to statistics, see Rao (1974).

Address all correspondence to Partha P. Majumder

At that time, the only statistic for comparing the resemblance between populations was Karl Pearson's "Coefficient of Racial Likeness" (CRL) (Tildesley, 1921; Pearson, 1936). Mahalanobis realized that CRL was a test of divergence between samples drawn from two populations rather than a measure of the actual magnitude of the divergence, because the magnitude of CRL was dependent on the sample sizes. In this study on Anglo-Indians, Mahalanobis proposed and used a measure of the actual magnitude of divergence which he called "first (provisional) measure of caste distance", D, (At the time of publication of this Presidential Address (Mahalanobis, 1927), Mahalanobis proposed another "theoretically preferable" distance measure, D′, in an appendix). The inferences arrived at by Mahalanobis in this study have been found to be largely valid from his own work conducted later in the United Provinces (Mahalanobis et al., 1949) and in Bengal, as also from studies conducted later by others, using more extensive data and more sophisticated statistical techniques.

During 1926-'27 Mahalanobis spent about six months in Karl Pearson's laboratory in the University College London. During this period, he undertook an extensive analysis of anthropometric data of various European population groups, and closely examined the CRL for measuring population relationships. In the process, statistical shortcomings of the CRL became clearer to him. After his return to India, his ideas on the problem of incorporating the observed correlations among anthropometric measurements used in measuring distance took a more concerete shape. In 1930, he published his seminal paper "On Tests and Measures of Group Divergence" proposing his famous $D^2$-statistic (Mahalanobis, 1930a). In the same year, based primarily on the work he did in Karl Pearson's laboratory, Mahalanobis published a paper in *Biometrika* (Mahalanobis, 1930b) which was the "first application of CRL to the discrimination of racial differences to be ascertained from measurements on the living". In this paper, which dealt with population groups of Sweden, Mahalanobis presented a very innovative graphical display of anthropometric interrelationships among the populations, taking two additional extrinsic variables into account—geographical location of habitat and occupation. Thus, the concept of forming clusters of populations began to take shape in his mind.

In 1936, Mahalanobis proposed the "natural" generalized distance $D^2$ for correlated variates, as well as its Studentized form using sample values of parameters (Mahalanobis, 1936). In retrospect it is clear that both these measures play a fundamental, extremely important role in statistics and data analysis.

Mahalanobis was also not satisfied with just providing a valuable tool ($D^2$) for cluster analysis. His logical mind led him to raise some fundamental issues connected with the application of the $D^2$-statistic. He argued that inferences on affinities among populations may depend on particular measurements chosen for study, in which case conclusions would not have the desired practical significance. The affinity configurations may change if one set of measurements is replaced by another. Mahalanobis, therefore, laid down an important axiom for the validity of cluster analysis called "Dimensional convergence of $D^2$" (Mahalanobis et al., 1937).

The formulation of the $D^2$-statistic, derivation of its properties and its applications are undoubtedly the most profound contributions of Mahalanobis to biometry. However, Mahalanobis made many other interesting contributions to biometry. He possessed an uncanny sense for numbers and could quickly point out recording mistakes in data coming under his scrutiny. In two papers entitled "Revision of Risley's anthropometric data" (Mahalanobis, 1933, 1934), he reconstructed a large series of anthropometric measurements

published by H. Risley, which were earlier condemned as faulty and unsuitable for statistical analysis. This work was highly praised by Sir Ronald Fisher. mahalanobis also argued for standardisation of anthropometric measurements. He examined anthropometric data available from different sources and showed how a comparative study was not possible because of differences in definitions of measurements in different investigations (Mahalanobis, 1928).

Apart from anthropometric studies, Mahalanobis conducted many other biometrical investigations on dextrality of snail shells, correlates of disease prevalence in humans and plants, demography, etc. In most of these studies, Mahalanobis developed novel statistical methods or made innovative applications of known methods. For example, in one of his early statistical studies conducted in 1920's on prevalence of dysentery and its correlates, he developed some useful smoothing techniques for time-series data using Fourier series. Such techniques are now commonly used.

As a scientist, Mahalanobis was, above all, a great applied statistician. To him, statistics was to be used for better understanding of scientific data, and for decision making for the welfare of society. In his biometrical work, it is the first aspect that dominates, while in his work on large-scale sample surveys and on planning, both aspects come together. He believed that theory grows out of a practical need and thus influences subsequent practical work. Innovation, systematisation and very concrete applications are the hallmarks of the sort of applied statistics that Mahalanobis practised.

## REFERENCES


Mahalanobis, P.C. : Anthropological observations on the Anglo-Indians of Calcutta. Part I : Analysis of male stature. *Rec. Ind. Museum, 23*: 1-96 (1922).

Mahalanobis, P.C.: Analysis of race-mixture in Bengal. *J. Asiatic Soc. Bengal, 23* : 301-333 (1927).

Mahalanobis, P.C.: On the need for standardization in measurements on the living. *Biometrika, 20* : 1-31 (1928).

Mahalanobis, P.C.: On tests and measures of group divergence. *J. Asiatic Soc. Bengal, 26* : 541-588 (1930a).

Mahalanobis, P.C.: A statistical study of certain anthropometric measurements from Sweden. *Biometrika, 22* : 94-108 (1930b).

Mahalanobis, P.C.: Revision of Risley's anthropometric data relating to tribes and castes of Bengal. *Sankhyā, 1* : 76-105 (1933).

Mahalanobis, P.C. : Revision of Risley's data relating to Chittagong hill tribes. *Sankhyā, 1* : 267-276 (1934).

Mahalanobis, P.C. : On the generalized distance in statistics. *Proc. Natl. Inst. Sci., 2* : 49-55 (1936).

Mahalanobis, P.C., Bose, R.C. and Roy, S.N.: Normalization of statistical variates and the use of rectangular coordinates in the theory of sampling distributions. (Appendix). *Sankyā, 3*: 35-40 (1937).

Mahalanobis, P.C., Majumder, D.N. and Rao, C.R.: Anthropometric survey of the United Provinces, 1941: A statistical study. *Sankhyā, 9* : 90-324 (1949).

Pearson, K.: On the coefficient of racial likeness. *Biometrika, 13* : 105-117 (1936).

Rao, C.R.: Prasanta Chandra Mahalanobis, 1893-1972. *Biog. Mem. Fellows Roy. Soc., 19*: 455-492 (1974).

Tildesley, M.L.: A first study of the Burmese skull. *Biometrika, 13*: 247-251 (1921).