

PROFESSIONAL TRAINING IN STATISTICS

by P. C. Mahalanobis



Reprinted from the Bulletin of the International Statistical Institute

VOL. XXXIII : PART V

PROFESSIONAL TRAINING IN STATISTICS

by P. C. Mahalanobis

Indian Statistical Institute, Calcutta (India)

The demand for trained statisticians is steadily increasing in India, but the statistical education given in the universities has not been found adequate for professional work. It is recognized that there is a great need of providing professional training of a high quality for a statistical career.

The nature of professional work in statistics varies widely, and it is convenient to break it up into 4 or 5 special fields (each of which, of course, may be further sub-divided into more specialized sectors).

- (i) Official and administrative statistics in Government and semi-Government Departments or on the management side of large commercial and industrial concerns.
- (ii) Statistical work in connexion with economic analysis and planning in Government, semi-Government, or commercial and industrial offices.
- (iii) Applications of mathematical or analytic statistics in scientific and technological work and research.
- (iv) Statistical sampling in various kinds of surveys.
- (v) Statistical education, training and theoretical research.

A comprehensive programme for professional training must include three things. First, training in certain statistical "crafts" or "skills" which would be same for all five groups; secondly, a basic educational course also same for all groups; and, thirdly, specialized courses for each group separately.

The first group, official and administrative statistics, offers the largest volume of employment in India, and although a large number of posts are being created and filled, the quality of work is still low for lack of trained personnel.

The demand for specialists in economic analysis and planning has just begun and the shortage of competent statisticians is most acute in this sector.

In the third sector (that is, in the use of mathematical or analytic statistics in scientific and technological work) the position in India is appreciably better. Present arrangements for education and training are directed mainly or almost exclusively to this particular group.

(In consequence little or no attention has been given to statistical craftsmanship with the result that statisticians with a great deal of knowledge of mathematical statistics have often been unable to deal with the professional work in official and administrative statistics).

There has been a rapid growth in the use of statistical sampling in India, and the supply of sampling statisticians falls far short of the demand because facilities for training in this subject are very meagre.

Training and opportunities for research in mathematical and analytic statistics have, however, expanded in India during the last 10 or 15 years and the position is not unsatisfactory. The supply of teachers for the usual academic type of statistical education has also expanded fairly rapidly.

But basic training in what may be called the "statistical crafts" have been neglected with unsatisfactory consequences. I shall consider these "crafts" under a number of heads.

Preparation of questionnaires and schedules. Training and skill in preparing questionnaires and schedules are essential requirements for professional work. Unfortunately, in India, this subject is almost completely neglected. I have continually found a kind of feeling among mathematical statisticians that the framing of a questionnaire or a schedule is not their concern. In consequence, questionnaires and schedules which have been used in the past or are still current are often of a poor quality. Usually there is much waste of paper together with lack of space needed for entering the data in a logical manner. The arrangement is often haphazard and without any reference to the procedure to be adopted in collecting the material. Concepts and definitions are not properly worked out, and so on. My own experience has been that it takes a good deal of time and practice for a junior statistician to learn how to prepare good forms and schedules.

Collection of primary data. It is necessary, in my opinion, for every professional statistician to have some personal knowledge and experience of actual methods of collection of data. Unfortunately this important aspect of statistical work usually has no place in the teaching programme. Most of the statisticians working in Government offices have, therefore, no idea of the difficulties or the conditions under which the data are actually collected.

Scrutiny of primary data. R. A. Fisher has stated somewhere that the first duty of a statistician is to cross-examine his data. In India so far very little attention has been given to this important subject. One reason is that, as mentioned above, statistical developments in India have been heavily concentrated in the third sector in which the statistician is usually associated with experienced scientific workers who undertake

the responsibility for making the scientific observations or collecting the statistical data. For example, in meteorology a great deal of attention is given, as a routine measure, to the continuing calibration of meteorological instruments. In agricultural and biological experiments the scientific worker is much concerned with controls and the validity of the observational data. In consequence, at least in India, it has not been necessary for the statistician to give attention to the accuracy of the data and he has been free to use his analytic methods without much primary scrutiny. Because of the collaboration of other scientific workers who have undertaken the responsibility of producing reliable data, this arrangement has worked satisfactorily in the case of meteorological observations and controlled experiments in the laboratory or the field.

Lack of training and experience in cross-examining data has had however most unsatisfactory consequences in the field of economic and social statistics, and also in large-scale surveys where the data have to be collected by ordinary investigators without scientific training, who are usually scattered over large areas and over whom the statistician has little or practically no control. In official and administrative statistics the result has been the growth of a "ritualistic" attitude of mind (to borrow the phrase aptly used by Stuart A. Rice). Most of the official and administrative statistics which comes to Government of India is collected through agencies at the periphery over which the Central Government often have no direct control. The statistician who deals with the material in the Central Government usually has no knowledge about the methods of collection of the primary material or the agency or men employed for this purpose. He simply accepts the data as something given. He has no curiosity or desire to question the accuracy (except in the very formal or trivial sense of reconciling misprints and discrepancies in totals or the results of similar arithmetic calculations). A senior statistician in Delhi told me, for example, that he did not know of a single instance in which the validity of a figure was ever challenged in his whole experience.

A second consequence has been the growth of a somewhat legal or formalistic attitude. A figure is received from a State Government or other agency, and it is usually assumed that the Centre has no right (or in any case, it is not proper) to challenge its validity. It is something like the record of a first or original Court of Law which has come up to a Court of Appeal where the evidence cannot be re-opened but arguments can be made only on points of law. (It is gratifying to note, however, that the present Census Commissioner is conducting, for the first time in India, a check on the census enumeration).

What is true of the Central Government is also generally true of the State Government because here also the statisticians usually deal with

data which come up to them from sources and agencies at the periphery. The State statistician also deals with the material as something given. In this way a great deal of statistics is collected and sent up to the Central Government the reliability of which is not questioned or examined at any stage.

Because of this legalistic or ritualistic attitude towards primary data, statisticians themselves have sometimes tended to take a purely formal view of statistical work. The statistician puts up some statistical data for analysis, and finds that this is not to the liking of his immediate superior officer or somebody higher up in the hierarchy (possibly even the Minister himself) who desires to have some other kind of analysis or even data which in his intuitive judgment is more realistic. Usually (but, of course, not always,) the statistician feels that, as a good civil servant, it is his duty to change his analysis or even his data.

In sample surveys my greatest difficulty in India has been to make highly trained statisticians (with a great deal of knowledge of mathematical statistics) realize the importance of investigator bias and response error. All their attention is focussed on the sampling error. In fact, the general attitude is to look upon the non-sampling error as something which does not concern the statistician, or in any case is a kind of dirty job which a highbrow statistician need not bother about. This is why the use of inter-penetrating samples have been persistently opposed in India as sheer waste of money.

I have dealt at some length with this particular point because in my opinion the lack of interest and skill in cross-examining the primary data constitutes the biggest single source of weakness of Indian statistics at the present time, and makes a great deal of the work in social and economic statistics more or less formal and unreal.

Computational work. A certain amount of laboratory practice in computation work is included in many (but not all) statistical courses in India. Sufficient emphasis is, however, not always given. One of my greatest difficulties has been to persuade a young M.Sc. in statistics to take up himself any serious computational work. Statistical education (and training) in India is often somewhat formal so that the student tends to think that his real concern is with mathematical and algebraic symbols, and that computations and calculations can be done by clerks (who belong to an entirely lower order of beings). I have often quoted (but usually without much effect) R. A. Fisher's statement that he learnt all his statistics through computations. In India, this negative attitude towards computational work is also partly conditioned by the general resistance against manual work which is considered to be derogatory to the dignity of educated or professional workers.

Organization of statistical processing. Even when the statistician has done a good deal of computational work in the practical class, he usually has no experience of the organization of processing of statistical work by computers. He has no knowledge of proper programming or of making even rough estimates of the time likely to be required to complete a job. In consequence, trained statisticians with high academic or research qualifications are often unable to cope with large scale processing work. My impression is that there is great deal of inefficiency and waste of money in such work in India.

Costing and budgeting. Because of the lack of knowledge about time requirements, the statistician is often unable to prepare realistic estimates of cost. I have seen even senior statisticians in India giving estimates of cost sometimes less than one-fourth or one-fifth of that actually required in practice or sometimes even four or five times higher.

One serious consequence has been the lack of reality in the scrutiny of estimates of statistical projects by the Finance Ministry or Finance Departments of State Governments. Sometimes the budget is cut down in such a way as to seriously hamper the work. Sometimes much larger grants are sanctioned than are really necessary. Estimates are prepared in a more or less standard administrative pattern. For example, if there is one senior statistician it would automatically follow that he would have so many assistants, so many research workers, so many investigators, etc., irrespective of the nature of the work. This is quite natural as estimates tend to be judged by yardsticks which have sole reference to work of a more or less routine type. Also, quite naturally, there is a tendency to put up much higher estimates than are considered really necessary on the view that the Finance Department would cut it down in any case by 50% or some such margin. This has led to a great deal of waste in both men and money at least in the Government of India.

Preparation of statistical tables and statements. Like the preparation of questionnaires and forms, the preparation of tables and statements does not find any place in most teaching programmes, and in consequence tends to be looked upon as no concern of the highbrow mathematical statistician. Usually tables and statements in Government offices are prepared by junior clerks, and often hide those very aspects which require to be brought out. And yet the tables and statements are the final end-products on which depends to a large extent the impression created on the mind of the administrator. My experience has been that it takes a great deal of time and practice to acquire skill in the presentation of statistical data.

Preparation of statistical reports. The statistical report contains the final results of a long process of statistical work of various kinds. If the

statistics are to be used in any way, such use must be based on the statistical report. It is necessary and desirable that a great deal of attention should be given to this part of the professional work. Unfortunately, usually this again has no place in the teaching programme. I have seen many statistical reports which are lacking in organization of thought, logical order of presentation of facts and conclusions, and written in a slovenly style. When care is given to the writing of the report, the tendency sometimes is to make it a kind of imitation of a technical paper suitable for publication in a scientific journal. Technical terms and jargons are used in abundance which simply annoy the administrator and create resistance. On the other hand, sometimes the reports are written in a highly polished administrative style, in which the entire attention is given to the form and the surface phrasing without bringing out the significant statistical facts or conclusions. A third type of report, which also is unfortunately too common, deals in a very elaborate and theoretical manner with small changes in say a statistical series which may be entirely due to chance fluctuations or simply to the unreliability of the primary material. Economists without professional training in statistics are often inclined to handle the statistical material in this superficial way which has created a good deal of unfavourable impression in administrative circles.

I have drawn attention to eight handicrafts of statistics which are much neglected in teaching programmes in India. Basic training in these subjects should be given to all students who intend to take up statistics as a professional career. These statistical crafts are of particular importance for work in official and administrative statistics. Professional workers desiring to enter these sectors should, in my opinion, have a great deal of advanced training and practice (besides the basic course) in these subjects.

Basic training in these subjects would improve the quality of work of statisticians engaged in the application of mathematical statistics in science and technology. Such training should also be obligatory for those who wish to take up statistical education and research as their profession. I am prepared to agree, however, that such training is not necessary for a person capable of doing "intuitive" work at the highest level. If, for example, a Ramanujan is discovered in the field of mathematical statistics he may be given full freedom to work in his own way.

Much thought has been given during the last 4 or 5 years in many countries of the world to the basic educational course dealing with the general theory and practice of statistical methods. A good deal of this basic course can be conveniently and profitably integrated with general education. I have discussed in a separate paper (recently prepared for a

UNESCO publication) the possibility of introducing mathematical statistics as a subject of study in secondary schools.

At a higher level, that is, at the stage of bachelor's or the master's degree it should be possible of course to cover a good deal of the more general part of the formal and mathematical theory together with a broad general knowledge of statistical methods and applications. General educational courses in statistics thus, in my opinion, can be conveniently given (in secondary schools in the first instance, and then) in the universities.

Professional training in statistics, however, requires a different type of institution. The position is somewhat analogous to medical science. Training in some of the basic subjects such as physics, chemistry, botany, zoology, physiology, etc. can be given in the universities. Professional training in medical subjects, however, requires specialized institutions working in close connection with clinical hospitals and surgeries. In the same way professional training in statistics can be given in an adequate manner only in close touch with statistical projects. The skills and craftsmanship which I have described in this paper, for example, can be more conveniently acquired in statistical workshops rather than in university class rooms.

But professional training is not confined merely to the acquiring of statistical craftsmanship but must go much further. I have no time to discuss this point at length. In the ultimate analysis professional competence is determined by the ability—

(a) to judge when the solution of a particular problem or the making of an administrative decision requires the help of statistical information;

(b) to analyse, integrate and make a critical appreciation of relevant information if already available; and

(c) when such information is not available, to prepare a project for the collection of the required information and their subsequent analysis and integration; also

(d) when a certain policy or action programme has been decided to develop statistical controls for the assessment of the progress of implementation; and

(e) to recommend changes in the policy or action programmes in the light of experience based on statistical information of the required degree of validity and precision.

These are the higher skills or techniques which a statistician must acquire in order to do fruitful work in statistics. Training and experience in professional work at this level can only be given in a specialized institution where such work is being actually done.

Résumé

Le travail statistique peut être classifié en cinq groupements principaux, à savoir, la statistique officielle et administrative, le travail statistique dans l'analyse de l'économie nationale et dans les projets de développement national, l'application de l'analyse statistique aux recherches scientifiques, l'échantillonnage et l'enseignement et la recherche statistique.

Un programme compréhensif d'enseignement professionnel doit comprendre un cours élémentaire de statistique théorique et l'enseignement de certaines techniques statistiques pour les cinq groups, ainsi que des cours spéciaux pour chaque groupe séparément.

L'enseignement de techniques statistiques doit comprendre la préparation de questionnaires, des données de base, les calculs, l'estimation de frais et la préparation d'un budget, la préparation des tableaux, et de rapports.

En dernière analyse le critère de l'aptitude professionnelle du statisticien doit être sa faculté de se servir des informations statistiques dans les problèmes qui exigent des décisions administratives, de développer les contrôles statistiques pour évaluer le progrès dans l'exécution d'une politique et d'en recommander des changements à la lumière des expériences basées sur des informations statistiques.

Les cours généraux de statistique peuvent être avantageusement introduits dans les écoles supérieures d'abord, et ensuite dans les universités. Pourtant la propre place de l'enseignement statistique professionnel est le laboratoire statistique qui est le centre des projets statistiques.

Dans l'état actuel de l'enseignement statistique dans l'Inde, il est pourvu suffisamment à l'enseignement de l'analyse statistique et de la recherche théorique, mais quant aux matières plus pratiques, l'enseignement est insuffisant.