

SOME RETHINKING ON STATISTICAL TRAINING AND RESEARCH IN INDIA*

BY

P. K. BOSE

Pro-Vice-Chancellor for Academic Affairs, Calcutta University.

INTRODUCTION

Training and research in any particular subject should be oriented generally according to the needs of the country. It is no doubt a fact that there will be always a small group who will engage themselves in fundamental research in the subject but the majority of the trained personnel will be engaged in routine work and problem oriented research. Accepting the above view I propose to discuss some aspects of statistical training and research as prevalent in India to-day.

1. WHY RETHINKING IS NECESSARY

Statistics in the sense of political arithmetic has been known to India since early days. As a scientific discipline, however, it crept its way in the thoughts and ideas of a few enthusiasts in the second decade of this century and has emerged half a century later as the key technology of the country. The annals of statistics in India divide itself into three distinct periods :

- (i) the period prior to 1931,
- (ii) the period 1931-1946 and
- (iii) the post-independence period from 1947.

During the earlier period very little interest was shown to statistics by Indian Universities. Official statistics developed as a

*Technical Address delivered at the inaugural session of the 24th annual conference of the Indian Society of Agricultural Statistics held at Madras on 31st December, 1970.

state craft as a by-product of the administration and not much of scientific methods were required or used for collection or analysis of data; however, in a few Universities where particular individuals were interested in this new discipline elementary courses were taught. However, the Government of India established two important statistical divisions during the earlier period. To scrutinise and collate the huge mass of agricultural data flowing in from provincial departments and to look after statistics of foreign trade, a central 'Statistical Bureau' was set up in 1895 to coordinate both agricultural and foreign trade statistics. Later on this division is known as the Department of Commercial Intelligence and Statistics. On the recommendation of the Royal Commission on Agriculture the Imperial Council of Agricultural Research came into existence in 1930 to promote, guide and co-ordinate agricultural research and to compile and distribute statistical information concerning agriculture and animal husbandry.

Milestones in the history during the second period are the establishment of the Indian Statistical Institute in 1931, setting up of Bowley-Robertson Committee in 1934, holding of the first Indian Statistical Conference at Calcutta in 1938, initiating the first large-scale sample survey on Jute Crop in 1937, the introduction of the first full-fledged post-graduate course on Statistics by Calcutta University in 1941, establishment of the Indian Standards Institution in 1945, setting up of the Gregory Committee in 1946, and recognition of statistics as a separate section by the Indian Science Congress Association. At first statistics courses in the University curricula were framed mainly due to the hope and expectation that it would gradually find useful applications in the physical, biological and social sciences. However, the earliest demand for trained statisticians that gave impetus to the courses in the Universities, came from government departments. It was only after the post-graduate courses were started in the Universities that the modern view of statistics has come to be emphasised.

With the dawn of independence the scope of teaching and research in statistics increased to a great extent, various national development schemes stimulated the application of statistical techniques to researches in various branches of science and technology and to problems of national planning. A number of Universities, Bombay, Gauhati, Patna, Banaras, Delhi, Poona,

Andhra, Baroda, Madras, etc., started teaching statistics from 1948 onwards.

Statistical Units were created in almost all important government departments, State Statistical Bureaus were established and the Central Statistical Organisation started functioning as a co-ordinating body. Besides Government Units, a number of institutions in India started researches in statistics (especially in the applied field). The All-India Institute of Hygiene and Public Health, Calcutta, and the Demographic Training and Research Centre, Bombay, carried out researches involving the application of statistical techniques in the fields of demography, public health and biological assay. The Forest Research Institute, Dehra Dun, made some statistical investigations in the fields of forestry, entomology and agriculture. The Gokhale Institute of Politics and Economics, Poona, carried out some important statistical work, relating to educational, demographic and economic problems. The Indian Meteorological Department, Poona, applied diverse statistical techniques in a very wide range of problems relating to meteorology and geophysics. These applications include among other things studies relating to seasonal forecasting, relationship between climate and geographical factors, crop production and weather, rainfall analysis, etc. The Indian Standards Institution, New Delhi, introduced new statistical concepts and techniques in the Indian Standards and through them to the industries. In particular, mention may be made of its work on evolving suitable sampling inspection plans, methods of sampling from bulk materials, introduction of quality control technique, etc.

The Department of Statistics, Reserve Bank of India, conducted statistical studies of the problems relating to India's foreign trade in diverse commodities, industrial production and input and output analysis. Thus during the last fifteen to twenty years in more and more new areas of science, technology, industry, commerce and agriculture statistical techniques were utilised. These give rise to an increasing demand of trained statisticians. Gradually it is felt that treatment of statistics as a basic science should be mathematically more rigorous than before. Finally a new feature entering into the picture is computing science which also leans to a considerable extent on modern mathematics for its base.

Thus the Universities and Research Institutions are now on the threshold of having to make judicious decisions as to the

proportions in which modern abstract mathematical basis of statistics, the contents of subject and its application, and the computer science have all to be blended in the curriculum at the different levels of teaching.

Further a stimulus has to be given to the new directions of research both on the theoretical and on the applied side so that freshly trained statistical personnel may be profitably engaged. At present less than 2% of the trained personnel are engaged in teaching profession and about 98% are engaged in other statistical jobs. Present day training given by the Universities and Institutions and also the evaluations made by them are not fully acceptable to the Government. This is evident from the fact, "Realising the role which statistics will have to play in the developing economy the Government of India recently constituted a cadre of professional statisticians known as Indian Statistical Service (Selection is done by a public examination conducted by the Union Public Service Commission). For entrants at various levels 'In Service Training Programmes' were drawn by an *ad hoc* committee of expert statisticians and approved by the Government of India for implementation. A staff college for training of statistical personnel is proposed to be established shortly."

'Besides the continuation of existing training facilities a comprehensive extension and intensification of training facilities is envisaged during 1966-71. Training centres for intermediate and primary level personnel are proposed to be strengthened adequately in pursuance of the national statistical training programme envisaged during the five year period.'

The above points lead us to think 'Whether rethinking is necessary'?

2. GENERAL REVIEW OF EXISTING CONDITION

Education in statistics in India is imparted in most of the Universities. 28 Universities are offering courses in Statistics up to post-graduate level. At the under-graduate level Statistics can be offered as a principal subject with one or two subsidiary subjects like Mathematics, Economics, Physics, etc. It can also be offered as a subsidiary subject at the undergraduate level with another subject as major or along with one or two other subjects of equal standard. At the post-graduate level facilities are available for pursuing a

master's degree in statistics or offer a paper or two in statistics as part of the M.A./M.Sc. course. Provision for teaching this subject at the post-graduate level is there in the faculties of Commerce, Agriculture, Education and Engineering. A few Universities are running a diploma course in Statistics in addition to their degree programmes. The important institutions imparting training in Statistics besides Universities are : The Indian Statistical Institute in Calcutta and the Institute of Agricultural Research Statistics in New Delhi. The Indian Statistical Institute in 1960 started courses of study for the degree of Bachelor of Statistics and Master of Statistics. The Institute of Agricultural Research Statistics has developed into a centre of training in Agricultural and Animal Husbandry Statistics and conducts post-graduate courses for the award of certificates and diplomas in Agricultural research statistics. The main burden of teaching still rests on the Universities of India, as annually about 1,700 scholars are trained by them. They belong to different categories : Diploma/Certificate recipients, under-graduate, post-graduate and research scholars. Approximately 400 scholars take their Master's degree and about 80 scholars receive diploma or certificate in statistics.

In the Indian Statistical Institute about 300 scholars get their training in the under-graduate and post-graduate and post-doctoral research. In the Institute of Agricultural Research Statistics about 20 students receive their training in the post-graduate class and about 25 more students are trained in the professional statistician's course.

These trained personnel are absorbed in the different Government offices (both Central and State), industries (both public and private), teaching profession and other miscellaneous jobs. Only about 200 trained statisticians are directly employed in the Universities and another similar number is employed in colleges and other research institutions for teaching Statistics (both ancillary and main). Remaining trained statistical personnel are absorbed in other professions.

Realising the role which statistics will have to play in the developing economy the Government of India recently constituted cadre of professional statisticians known as the Indian Statistical Service. In the Statistical Service there are four grades and in the junior cadre there are ten types of posts such as Senior/Junior

investigator, computer, machine supervisor, etc. Total staff, and expenditure of statistical personnel on all India basis are given below :

	<i>Expenditure (Rs. '000)</i>	<i>Total staff Excluding class IV as on 31-3-1966</i>
Central Ministries	50,847	8,190
States and Union territories	47,532	13,698
Public Sector Undertakings	1,317	249
Total All India :	99,696	22,137

Statisticians are also employed in trade, commerce and industry in the private sector. It is not possible to specify this number exactly but it will be approximately half of the public sector. The percentage of persons engaged in the teaching profession is negligible when compared against the number of persons working in other capacities. Among the latter about 20,000 are intermediate and primary workers.

Recently a team of statistical experts visited a number of industrial units, Government departments and research institutions throughout India. During their visits, discussions were held with the heads of the organisation and other technical personnel. Some of their viewpoints are recorded below :

(a) *Industrial undertakings/units.*

The most important issue which was stressed by almost all the industrial undertakings which the members visited was the fact that the present day training given to our students at the M.A./M.Sc. level was not sufficiently oriented for practical work. Fresh products from the Universities are good so far as the theoretical aspects of the subject are concerned but are not able to cope straight away with practical situations. It has become

necessary for the statisticians entering industrial undertakings to receive a preliminary 'on the job training' for about six months to one year usually for getting into grips with the problems which they have to tackle in the particular field concerned.

(b) Government departments.

The main issue which was stressed in most of the government departments which the committee visited was the fact that the students coming out with a degree or a post-graduate degree in Statistics are not very often acquainted with facts like :

- (i)* Sources of official statistics
- (ii)* Government publications dealing with statistical information
- (iii)* Organisations/Agencies set up by the government for collection/compilation of data of various kinds
- (iv)* Procedural details relating to mechanical tabulation.

The heads of the statistics divisions of the government departments with whom the members came into contact were strongly of the view that sufficient training could be given by the Universities and other institutions in various types of factual information like *(i)* to *(iv)*. Moreover students should be taught how to handle statistical projects and live data.

We have to consider carefully the viewpoints of two important potential employers of the statistically trained personnel. Out of the total employed persons 1.5 per cent is engaged in teaching and research and 98.5 per cent is engaged in other statistical jobs.

Let us now consider the level of teaching in the undergraduate and postgraduate stages. The level of teaching of statistics at the undergraduate stage in the different Universities is very uneven. In arts and science faculties only a few Universities offer the course as a main subject for study. Several Universities arrange for one paper in elementary statistical methods and few an additional paper on applied statistics, including official statistics, economic statistics and occasionally statistical quality control methods. These different practices make it difficult to plan a uniform syllabus of a fairly high standard at the post-graduate level. Moreover at this stage a considerable percentage of students are admitted without any background of statistics in some of the

Universities. Naturally the teaching should start from elementary stage leaving little time for advanced courses. It is often found that fresh students who were not entrusted with practical problems in administration, industry, often find it difficult to correlate their theoretical knowledge with the problems of handling new data and applying appropriate methods of analysis to draw conclusions. The practicals which the students do in the classes are often taken out of text books written by foreign authors. This is a very serious handicap. An additional problem arises in the case of students of economics and commerce courses. The present courses at undergraduate level are elementary. The main reason is that majority of students give up Mathematics at the school stage.

Although statistics has been introduced rather recently in the curricula of Indian Universities, still research workers have made significant contributions in almost all major areas of the subject. Researchers cover a variety of aspects including measure theory and probability, sampling techniques, design of experiments, multivariate analysis, statistical inference, stochastic process, econometrics, statistical quality control, demography, genetics and operations research. Research programmes which are at present in progress at the Institute of Agricultural Research Statistics can broadly be classified as

1. Statistical research in agricultural experimentation
2. Statistical research in animal husbandry experimentation
3. Sample surveys
4. Cost of production studies
5. Basic research in design of experiments and sampling.

This institute has made significant contributions in the fields of agricultural experimentation and sample surveys. It has published 'National index of field experiments' which is considered a very important publication. The subjects of research carried out in the Indian Statistical Institute are given below :

1. Mathematical statistics and probability
2. Mathematics and numerical analysis
3. Natural sciences
4. Computer science and computer development

5. Economics, econometrics, national income and regional survey
6. Demography and sociology
7. Psychometry
8. Phonetics and linguistics.

During the year 1969-70 research workers published more than 200 papers. Other research institutions of India also are carrying out researches in the different areas of applied statistics.

3. SOME SUGGESTIONS

It has been accepted that there is need for effecting some changes in our research and training programmes. It will be difficult to state the exact nature of this change but some indications may be given.

Government is our biggest employer, let us study the functions and composition of some big statistical units.

(i) Central Statistical Organisation

A. FUNCTIONS

- (a) To undertake research in development of standards and methodological studies.
- (b) Co-ordination and dissemination of statistical intelligence including graphical presentation.
- (c) To undertake statistical work relating to planning.
- (d) To organise and conduct training courses in official statistics.
- (e) To compile national income estimates and undertake studies in the field of National accounts.
- (f) Construction of Prices and Cost of Living indices for middle classes.

B. DISTRIBUTION OF STATISTICAL PERSONNEL BY NATURE OF DUTY

Professional	59	
Intermediate	101	
Primary	66	
<hr/>		
Total	226	Expenditure Rs. 24 lakhs.

(ii) *Directorate of National Sample Survey*

A. FUNCTIONS

- (a) Collection of comprehensive data on a continuing basis relating to socio-economic and demographic conditions, prices and area and yield of different crops, etc.
- (b) Conduct of annual surveys and related enquiries in the organised industrial sector.
- (c) Co-ordination of crop estimation surveys on principal crops in different states.
- (d) Programming for supervision of crop cutting experiments and participation in the training imparted to the primary field staff and tabulation of data obtained through these experiments.

B. DISTRIBUTION OF STATISTICAL PERSONNEL BY NATURE OF DUTY

Professional	93	
Intermediate	315	
Primary	936	
<hr/>		
Total	1344	Expenditure Rs. 83 lakhs.

(iii) *Department of Commercial Intelligence and Statistics*

A. FUNCTIONS

- (a) To collect and furnish commercial information, required by Government and the trade.
- (b) To extend assistance to commercial concerns with a view to stimulating foreign trade of India particularly the export of Indian produce and manufactures.
- (c) To compile and publish statistics of trade, shipping, etc., in the publications issued by the department.

B. DISTRIBUTION OF STATISTICAL PERSONNEL

Professional	—	7	
Intermediate	—	56	
Primary	—	138	
<hr/>			
Total	—	201	Expenditure Rs. 24 lakhs.

(iv) *Reserve Bank of India, Department of Statistics*

A. FUNCTIONS

- (a) Collection, compilation, analysis and dissemination of statistics relating to Indian economy.
- (b) Computation of the official series of index numbers of industrial profits and profitability, etc.
- (c) To design, conduct, and supervise large scale sample survey.

B. DISTRIBUTION OF STATISTICAL PERSONNEL BY NATURE OF DUTY

Professional	—	35	
Intermediate	—	34	
Primary	—	128	
<hr/>			
Total	—	197	Expenditure Rs. 18 lakhs.

From these specimen units it may be said that the bulk of statistical personnel (mostly undergraduates and certificate recipients) are employed as intermediate and primary statistical workers and their main job is to handle 'live data' and to make them intelligible to the common men. Thus a graduate course in statistics should be so framed that it should make a student fully equipped to be useful in whatever profession he chooses; whereas a post-graduate course should concentrate on specialised fields with emphasis on theoretical aspects. As far as possible teaching should be 'problem oriented' and the practicals should be on 'live examples.' At the present moment teaching is not 'problem oriented'; even practicals are based on imaginary data. There are reasons for this. Firstly, most of the teachers are not conscious of the fact that majority of their products will be, ultimately, employed as intermediate and primary workers. Secondly, statistical data collected by the different agencies are not available to the Universities for the purpose of teaching as there is a very big gap between the Universities and the collecting agencies (both private and the Government) and thirdly, there is dearth of competent teachers. As specific examples I may cite that very few Universities or Institutions make use of 'National index of field experiments' prepared by the Institute of Agricultural Research Statistics or the reports of the large scale sample surveys conducted by the Indian Statistical Institute and I.A.R.S. for the purpose of teaching.

My suggestion is that courses both in the undergraduate and postgraduate should be flexible. A large number of topics on mathematics, mathematical statistics and applied statistics should be included in the curriculum and students should be given the liberty of choosing the topics. Of course the number of topics for each student should be the same. At present the rigidity of the present syllabus should be relaxed gradually. More emphasis should be given on handling of 'live data' than at present. Periodically (every three years) the syllabus should be scrutinised and according to the needs new topics should be added or omitted. Greater collaboration should be established between the university teachers and the professional statisticians.

In the past in India we made original contributions in three major areas and each of the areas were developed while trying to solve some practical problems.

The first area I may mention 'Design of Experiments.' The first initiative was taken by the Indian Council of Agricultural Research in the early thirties of this century. On the recommendation of Royal Commission of Agriculture, Imperial Council of Agricultural Research came into existence in 1930. This Council in order to promote, guide and co-ordinate agricultural research and to compile and distribute statistical information concerning agriculture and animal husbandry published a series of research notes known as 'statistical notes for agricultural workers.' This paved the way for advanced training and research in this subject. A major breakthrough in this field was achieved by Prof. R.C. Bose in 1938 when his important paper 'On the application of Galois fields to the problem of construction of Hyper-Graeco Latin Square' was published. Later on a series of fundamental papers were published by Indian Statisticians. Their contributions are well known throughout the World.

The second area is 'Multivariate Analysis.' The concept of 'Tests and Measures of Group Divergence' was first given by Prof. P. C. Mahalanobis in 1930, while he was engaged in studying statistically some Anthropometric measurements. He also gave the notion of 'Distance Function.' The first important paper on this subject namely 'Normalisation of Statistical Variates and the Use of Rectangular Co-ordinates in the Theory of Sampling Distributions' was published in 1937 by him and the other two collaborators. A new technique was developed and a series of papers on 'Distance

Function', 'p. statistic' etc., were published by research workers which gained universal recognition.

The third area is 'Large Scale Sample Survey.' The Indian Statistical Institute, Calcutta, conducted the first Bengal Crop Survey in 1937 to estimate the total acreage of Jute and also the yield per acre. This work was continued for four years and in 1941 the whole jute tract of Bengal measuring about 60,000 square miles was covered for the same purpose. This was a national problem and it was solved by evolving new statistical techniques. Out of this project new concepts of 'cost function', 'variance function', 'optimum design for sampling' etc. were developed. Later on this technique has been utilised by other research organisations in India and abroad. I have tried to illustrate some specific cases of 'problem oriented' research, which later on opened up new lines of theoretical work.

I shall now indicate some new areas which should attract the attention of our research workers. These areas are : (a) Operations Research, (b) Information and Coding Theory, (c) Computer Science and (d) Statistical Models. After independence there has been a rapid growth in the size and complexity of industries. The mere size of modern industries means that administrative decisions can affect vast quantities of capital and large number of people. The pace of modern industries and business is such that decisions are required quickly and delay may prove dangerous for them. Methods of operations research help us to arrive at such decisions objectively. In the country a large number of heavy and medium scale industries have been established and it is essential that the students should be trained adequately in replacement theory, queueing theory, linear programming, dynamic programming, etc. Therefore these topics should form a part of our training and research programmes. Information coding theory and computer science have been found to be indispensable tools for the statisticians in this modern age of technology. It is necessary to lay more stress on the developments of these new ideas in the curriculum of the Universities and Institutions. In our country very little work on these topics has been done. Statistical models now play a fundamental role in many applied branches such as engineering education, etc. Statistical models are used in life testing and reliability. An important application of probability distribution is as time-to-failure models for components or systems. The number of hours or years

or cycles of satisfactory operation is a random variable whose exact value depends on many factors such as variations due to manufacturing tolerances and materials and changes in the environmental condition. Once an appropriate probabilistic model for time-to-failure has been constructed and its parameters estimated, this information may be used to predict life, develop optimum initial burn-in procedures, establish part-replacement schedules, etc.

'The need for explicit educational planning is increasingly gaining acceptance in recent years so much so that educational planning and economic planning are being treated as being complementary to each other. Consequently the present emphasis in educational planning is on determining the optimum amount of education or skills of each 'kind' and 'level' required for achieving specified economic objectives. For different objectives, we can construct different models.'

Statistical models are also used to study behaviour of individuals. For this purpose four different types of models have been proposed. In two cases the problem of statistical estimation is involved and in the other two cases we have to determine the factors and their loadings directly from the observed data. The study of these models has opened up a new line of research in multivariate analysis which is known as 'Factor Analysis' and 'Principal Component Analysis.'

It is desirable to introduce 'on the job training' for all students in the under-graduate classes for two months and for all students in the post-graduate classes (excepting those specialising in mathematical group) for four months during vacation. This is to be arranged in collaboration with the Government and the private industries and business. If we can arrange this then there will be no need for 'in service training' later on.

4. CONCLUSION

I have given some of my own thoughts on the problem. In order to draw up a suitable plan for this purpose a high powered committee should be set up consisting of University teachers, Professional statisticians, Government administrators and experts from the private sector. Sooner the flexible pattern of training and useful 'problem oriented research' are introduced in our University and Research Institutions, it is better for the future of Statistics in India,