Contributions of Statistics for Applied Research and National Development[1]

M.N. Das
Former Director, IASRI, New Delhi

1. Statistical Development as Product of Necessity

Statistics originated to solve certain types of problems faced in different countries. Initially methodology used was based on common sense and convenience. As social activities and thinking became more and more complex use of mathematics became necessary. Gradually methodology based on mathematical rigour and applicational convenience and validity was evolved to solve the problems. Science of Statistics had its foundation on such methodology.

There was a severe famine in Bengal in 1943. While probing the cause of the famine it was found that the concerned administration had no authentic information on crop production in different provinces of the country. The Indian Council of Agricultural Research, a premier organisation in the Ministry of Agriculture in Centre was entrusted with the task of evolving suitable methodology for estimation of annual production of different crops in different provinces on regular basis. At that time Dr. P.V. Sukhatme was Statistician in ICAR. He was entrusted with the job. He rose to the occasion and evolved an efficient technique for estimation of production of different crops using data obtained through crop-cutting experiments on randomly selected crop fields. He actually evolved and used what is now known as multistage random sampling technique. He was assisted in this job by persons like Dr. V.D. Thawani, Dr. R.D. Narain, Dr. A.R. Roy, Mr. V.R. Rao and several others. Success of the venture was also due to its willing and cooperative use by the provincial governments. Prominent among the persons who took lead and keen interest in using and popularising the method of sampling for estimation of crop production in different provinces were Dr. Ranga Rao in Maharashtra, Dr. K. Kishen in U.P., Dr. C. Misra in Orissa, Dr. D.N. Lal in Bihar and several others in other provinces. The method has been in use since then throughout India. In support of these activities, Dr. Sukhatme organised courses on sample

---

survey techniques and the lecture notes thereon formed the basis of the book on Sampling by Dr. P.V. Sukhatme. For publicising the activities and methodology regarding agricultural statistics Dr. P.V. Sukhatme and Dr. V.G. Panse formed the Indian Society of Agricultural Statistics and a journal of the Society was also being published from 1948. Several courses viz. Professional Statisticians Certificate Course, Diploma in Agricultural Statistics, Senior Certificate Course and Junior Certificate Course were also started. The first two courses were advanced courses and the other two were for those who used statistics for their activities.

It is thus seen how a social problem led to development of research and other related activities in Agricultural Statistics. It created awareness of need and use of statistics and statistical methods for collection of necessary data in objective manner and of appropriate methodology for processing the data to arrive at accurate and unbiased estimates of crop production.

When Dr. Sukhatme joined Food & Agriculture Organization as Head of Statistics Division, Dr. V. G. Panse, a close associate of Dr. Sukhatme took the position of Dr. Sukhatme in ICAR. He specialised in plant breeding methodology and was a student of Sir R.A. Fisher in Rothamstead. He was instrumental in developing the Institute in its present form. Earlier the organisation was functioning as a section of Statistics of ICAR. During Dr. Panse's time it was converted to an Institute with a separate building in Pusa. The Institute was named Institute of Agricultural Research Statistics. Through his efforts the Institute got in 1965 a main-frame computer IBM-1620. He created a unit for research in genetic and animal husbandry. Divisioning of the activities in the Institute was made during his time. The Institute started M.Sc and Ph.D. courses in association with Indian Agricultural Research Institute, Pusa.

Dr. Panse was an applied worker but at the same time he did not discourage research in and study of basic theoretical statistics. His able guidance and advice and a strict sense of discipline created a band of devoted workers in the Institute who brought international fame and recognition of the Institute and placed it on international map. From different parts of the country persons used to come to the Institute for help to get their data analysed, for data collection, for computer use and similar other activities.

Once an agronomist from Saurashtra region came to Dr. Panse for a design for one of his experiments. Dr. Panse asked one of his staff members to help him. The staff member after discussing the problem wrote some blocks of a Randomised block design after randomisation just like designs written in books. When it was shown to Dr. Panse he said that was not the way to help an

agronomist. He advised to first draw the plots with thick lines forming squares, show the irrigation channels, show the drainage channels, and then write the treatment numbers inside the squares as required. When the layout of the design was presented as above, he asked to indicate the precautions needed during the course of the experiment and to write the purpose and method of analysis of the resulting data, manner of presentation of the results and their interpretation taking into account alternate types of results. He further asked to take some fictitious data as if from the experiment and analyse them showing the steps of calculations, present the results and finally interpret the results. This shows how an agronomic problem need be solved and how subject-matter specialist in this case an agronomist can be helped and attracted towards statistics.

It should not be forgotten how I A S R I served the nation in times of need and contributed greatly for development of the science of Statistics and its applications to provide social services. The importance of the Institute should not be undermined and its functioning should not be diluted in any way by any type of discouraging actions or decisions.

The Indian Society of Agricultural Statistics must go on encouraging and supporting applied statistical activities serving some social need directly. Possibly some University departments may be entrusted with such activities providing financial support. The projects should be monitored closely from start and need not end with a report but with media publicity to attract public attention and enhance public awareness regarding utility of statistics.

2. Development of Official Statistics

Though statistics made significant contributions by streamlining an objective method for estimation of crop production and to help agricultural and other scientist in problems requiring statistical methodology, its contribution in other sectors are no less important and useful. Prof. P.C. Mahalanobis, a Professor of Physics in Presidency College, Calcutta through his contact with Sir R.A. Fisher got interested in Statistics and was convinced of its importance and utility for national development. He then established Indian Statistical Institute in Calcutta and started popularising Statistics and its applications in different social sectors. Many brilliant personalities including Dr. P.V. Sukhatme, Prof. R.C. Bose, Prof. S.N. Roy, Dr. K. Kishen, Dr. P.K. Bose, Dr. M.C. Chakraborty and other became associated with this Institute in some form or other. To spread knowledge, awareness and uses of statistics a department of statistics was opened in Calcutta University in 1940. Prominent scholars like Dr. C.R. Rao, Dr. H.K. Nandy, Dr. A.R. Roy, Dr. V.M. Dandekar and several others were the initial students of this department. Subsequently

realising the importance of statistics almost all the universities in the country opened departments of statistics either independently or jointly with department of mathematics. The syllabi for M.A./M.Sc. courses laid stress on applied statistics and subjects like Design and Analysis of experiments, Sampling Techniques, Statistical Genetics,   Statistical Quality Control, Economic Statistics, Official Statistics and Demographic Statistics were included in the syllabus of Calcutta University M.Sc course.

The then Prime Minister, Pandit Jawaharlal Nehru realised the importance of Statistics for effective planning for development of the country. He inducted Prof. Mahalanobis in the Planning Commission and put him in charge of Planning. Prof. Mahalanobis found that availability of data required for efficient planning was much below requirement in most sectors of social activities. He realised that strengthening of statistical system in the country was very much necessary. Through his efforts a department of Statistics was created at the Centre with its two main wings, Central Statistical Organisation and National Sample Survey Organisation. To meet manpower needs in Government Departments, The Indian Statistical Service was created through his efforts. Alongside Indian Economic Service was also created. The idea was that the Statistics Service is to provide service for social activities while Economic Service is to look after the economic aspects of Government functioning.

As a result of above organisational set up statistical activities increased considerably in the country and its awareness and utility increased. Major departments engaged in statistical activities are at present agricultural departments in the states, Directorate of Economics & Statistics at the Centre for collection and processing different types of agricultural statistics. Central Statistical Organisation coordinates all statistical activities in the country and controls statistical service personnel. It also brings out useful statistical publications. National Sample Survey Organisation conducts survey on one suitably chosen social activity each year and brings out reports on such surveys. Registrar General of India conducts decennial population census and this census is a very rich source of data on several aspects of the society. Central Water Commission and Central Pollution Control Board collect data from selected locations on different rivers in the country. These data are used to ascertain water availability and quality of water in different rivers monthwise. Several series of index numbers are published by different government departments at regular intervals. Prominent among these are Wholesale Price Index, Cost of Living Index, Consumer Price Index, Inflation Rate as an index indicating weekly fluctuations in wholesale price index. These are used for government policy decisions.

A number of government departments both in the States and Centre are collecting different types of statistics. Central Statistical Organisation brings out a publication called Guide to Official Statistics containing brief reports of such activities. In addition most of the statistical offices bring out reports and other publications regularly.

3. Applications of Statistics

It appears in private sector statistics has not made much impact. Its use in this sector is very much limited. In commercial and industrial activities potentiality of statistics has not yet been appreciated. However, certain pharmaceutical houses have statistical units. In medical fields also statistics is being used to aid research and assess hospital activities and performance and incidence rates of diseases. Some industries are using statistical quality control techniques to keep watch on quality of products and take corrective measures when necessary. This practice is not yet widespread. There is a number of statistical techniques for industrial experimentation to improve product quality and production techniques. Usually in India such techniques are not yet exploited though in developed countries these are widely used. Some commercial houses and their agents have started quality preference surveys though in a small way by sending field workers for collection of appropriate information from randomly selected clusters of houses. Doordarshan audience research unit also collects from randomly selected persons data on ratings of serials. For such surveys, it is necessary to adopt appropriate statistical techniques to remove any type of bias and to get information objectively and accurately. There is enough scope for exploitation of statistical methodology for commercial activities for collecting information on quality improvement and quality preference. For example, Dominos, a multinational company conducted an experiment to collect data on consumer preference of their individual products like pizza. Several variations of pizza were prepared using different proportions of necessary ingredients. A requisite number of judges were each served with each of the varieties and were asked to rate each variety taking into account their taste and other desirable parameters. Analysing these data appropriately the company took a decision as to the most wanted type of pizza. A similar study was once conducted by a student of IASRI in association with IARI. During green revolution days several new varieties of wheat were being cultivated. A question arose regarding consumer preference of the varieties. As a part of student research an experiment was conducted by IASRI and IARI, Pusa.

Here also following a proper design chapatis were prepared using different varieties of wheat and fed to different judges, each judge testing each of the

varieties and putting rank to each taking into account taste and other desirable properties. Through an appropriate analysis of the resulting data the varieties were graded. This study was followed by other similar studies and reports were prepared. The unfortunate thing with such studies which are of direct public interest is that a culture has not yet developed to publicise such findings among public. Appropriate publicity of such works not only serves a social need but also creates appreciation of the role of statistics for social service.

There are many organisations which are collecting data routinely year after year. Mostly, such data are collected compulsorily but their processing for different possible uses is optional and casual though their investigations can provide useful social service. As an example, Central Pollution Control Board collects every month voluminous data on water quality from different locations on most of the rivers in the country. The data are in form of values of some 25 different water quality parameters obtained from analysis of water samples collected from every location each month. Through proper statistical processing of such data in association with the concerned subject matter specialist useful indices of direct interest to the society can be obtained. For example, through appropriate combination of the more important quality parameters for specific water use suitable index can be worked out for some important locations and publicised regularly or displayed in concerned sites. Culture of such statistical activities has developed in regard to some weather parameter and some air quality parameters in some cities. Use of such statistics is not yet wide-spread and awareness regarding their uses is insignificant. Actually, some university departments or research institute may be entrusted with such activities for initial processing including development of appropriate methodology and for arrangement for their regular use and publicity.

There are many such areas where data can be used keeping in view some social needs and uses. National Sample Survey Organisation is conducting survey on selected subject areas every year. Data collected are mostly for future use if and when needed. Generally, it takes considerable time for reporting the data. Processing is mostly in form of summarisation of different variables for which data are collected. It appears that data from such surveys are not put to any immediate social use. Rather as part of such surveys certain type of data associated with the main purpose of the survey may also be collected so that these can be used to get some information of direct social value. For example, medical science is at present trying to substantiate the hypothesis that with advancing old age eating less food leads to longer life. Some experiments are being conducted on rats to be followed on humans. When NSSO conducts a survey on food consumption and food habit a schedule may be included to get data on food habit and some other variables of old persons in the selected

households, their age, number of deaths of old persons during last two years, their food habit and age at death. The data so collected can throw light on the hypothesis being pursued presently. Actually, such data have to be investigated jointly by the statisticians and appropriate medical persons.

In statistics considerable research findings are there which can be used profitably for applied work of scientist/technologist working in other areas. Proper interaction between these two main groups of workers is lacking. May be some applied workers want to use statistical methods but find that the available statistician cannot provide required service. Such situations are quite likely because statistical education at present is theoretically oriented and the students mostly do not gain applicational competence covering different areas of activities. I propose to discuss more about statistical education in a subsequent section. As against such situation often applied scientist do not come forward for statistical help or for application of research findings in statistics.

Further, statistical research workers often find it difficult to get collaboration of appropriate applied workers for applying their findings. To site an example, a research worker in a statistical research institute under ICMR got a very interesting method of group sampling. Often blood or similar other material is collected from a large number of persons for some project activities. Usually such samples are analysed individually to ascertain if each sampled person is positive or not in respect of some parameter. When the number of persons is large and trait under investigation is somewhat rare the individual samples show mostly negative response. A question then arose can the individual samples be grouped suitably and mixed so that from tests of such mixed samples for positive or negative response it is possible to find out uniquely the persons with positive response. The research worker mentioned above obtained a very convenient method of group formation together with an ingenious method of decoding the response data to sort out the persons with positive response. Though some medical persons were inclined to try the technique, the study was discontinued before conclusion. But the interaction proved useful to extend the research to find sampling technique to control the group sizes, as large group size may lead to the risk of too much dilution of mixed samples for detection when positive.

Similar situtions exist in areas of industrial experimentation and statistical quality control activities also.

4. *Statistical Education*

Teaching of Statistics started in India as early as 1940. These days it is being taught even at school level. Possibly teaching Statistics at school level

is not serving any useful purpose. It should be introduced at a more mature level of mental development. It is only serving to increase the burden of the students forcing them to be memory machines rather than thinking, visualising beings. It is a trend these days to introduce at school level anything and everything, set theory, new mathematics, calculus, elements of analysis, base operations, probability, statistics but not arithmetic and Vedic mathematics. We boast of our ancient culture but do not put before the young minds some of ancient wonderful contributions. For example, Vedic arithmetical operations are quite easy, useful and very interesting. This subject should find a place in school curriculum as number theory. The Vedic mathematical procedures can facilitate statistical calculations and can provide efficient algorithms for writing useful computer programs in Statistics, Astronomy and other scientific activities. At graduate and post- graduate levels emphasis is laid on theory and methodology but not on applications of methodology for solving problems facing the society. Even in research and projected activities in the universities the aim is theoretical achievements and publication of papers in mathematically inclined journals. This type of attitude is alienating statistics from problems in social sector for which the science of statistics originated. This attitude is also alienating statisticians from subject-matter specialist who can profit by applying statistical methodology. It is thus necessary to correct the situation by including in the syllabi applied topics including official statistics, economic statistics, demography, statistical genetics, statistical quality control techniques, industrial statistics and such other topics. The university departments have to collaborate with government and other agencies and institutions specialising in these topics and undertake student projects jointly on topics of current social need. Collaboration of commercial and industrial establishments and government agencies is also needed for success in such efforts.

These days universities are suffering from lack of personalities. Even a decade ago there was a very good market for Indian statisticians in USA and there was a drain of competent statisticians from this country. Even these days scholars are always anxious to visit US universities. The result is contributions of statistics are not making any mark on the society in India.

Indian statisticians have made good contributions in statistical research though mainly theoretical. A number of statistical journals are being published in the country. Some of these are Sankhya, Journal of the Indian Society of Agricultural Statistics, Calcutta Statistical Bulletin, Journal of the Indian Statistical Association, Assam Statistical Review and Gujarat Statistical Review. These journals also publish papers received from other countries. One drawback in these journals is the lack of application oriented papers. Some journals realised the situation and made special efforts to attract applied papers but it

seems that the attempt did not succeed. It is necessary to reverse the directions and that has to be done by those who work with subject-matter specialist and through collaborative research with arrangement to apply techniques obtained and create awareness of such techniques suitably. The present day aim of statisticians is to publish papers in reputed journals. But the aim should also be to research on applied topics that serves some social need and present through media the findings to the public. At present the universities and the statistical organisations do not have systematic arrangements for publicising project findings for popular appeal. There should be publicity officers in these organisations to undertake such functions in collaboration with concerned workers.

5. *Applied Research and Computer Use*

Because of lengthy data and complex calculations involving such data calculators and computers are extremely useful for statistical activities and research. This work is at present done by using program packages. These packages come mostly from outside the country particularly those required for scientific and technological activities including statistical activities. It appears Indian intelligentsia and concerned commercial establishments have not taken seriously the job of computer package making in the country. There are some small statistical packages brought out by some IASRI workers. Efforts should be made to monitor their performance, get feed back information to improve them where necessary. To establish such packages these need be reviewed by competent persons. Security considerations should not shadow efficiency of packages particularly in those organisations where profit is not the sole motive.

As a step for helping applied workers a package was worked out by me. It provides randomised layout plans and their analysis with customised data entry and editing facility for most types of designs including BIB, some series of PBIB designs, designs for factorial experiments, use of designs for obtaining diallel crossing plans and similar other topics.

Recently, IASRI has brought out a similar package dealing in incomplete blocks. It is hoped that other organisations and individuals shall encourage formation of such packages covering other areas of statistical/technical activities.

Knowledge of programming is quite helpful to statisticians. It helps them greatly in situations where problems used to be abandoned. Two examples from my personal experience are cited below.

A. *Example 1* : Estimation of Parameters in Weibull Distribution

Weibull distribution is widely used these days for different types of applied activities though it was evolved originally as a suitable distribution of breaking strength of materials. The form of the distribution is

$$f(x) = abx^{b-1} \ e^{-ax^b}, (x = 0... \infty)$$

where a and b are its parameters.

Mean of $x = p^q \ \Gamma (1 + q)$

Second moment about origin $= p^{2q} \ \Gamma (1 + 2q)$

where $p = 1/a$ and $q = 1/b$

There is no easy method of estimation of the parameters as the equations for solving out the parameters by the method of moments involve the unknown, b under gamma notation.

The problem was brought to my notice by a technologist who wanted a suitable distribution to fit his data on aresenic concentration in water samples. After consulting relevant literature and scrutinising the data it appeared that the Weibull or Gamma distribution may fit the data. It appeared to me that available methods for using Weibull distribution were either graphic or approximate. It seemed that a suitable computer program may provide exact solutions of the above two equations giving estimates of a and b.

Denoting the first and second moments about origin by m_1 and m_2 respectively we get from the above equations the following equation in q.

$$R \ (\Gamma \ (1 + q)^2) = \Gamma \ (1 + 2q)$$

where $R = m_2/m_1^2$

Taking suitable trial values of q, finding the corresponding gamma values on both sides of the equation and taking into account which side of the equation (LHS or RHS) is greater first a solution is obtained correct upto one place of decimal by trying for different values of q. Next, by refinement results can be corrected upto more and more decimal places. One difficulty here is that there is no easily available table of gamma values for fractional numbers. Hence first gamma values of all 3 digited decimal fractions were obtained and stored in a file and this file was used in the program for estimating parameters in the Weibull distribution. The same table was also used for investigation of the data using Gamma distribution. It was found that Weibull distribution fitted well to the data as judged from the closeness of fit test. The required percentile

points were then obtained using the distribution with the estimated parameters and conclusions were drawn from the percentile points.

B. *Example 2* : To obtain a series of Asymmetrical Rotatable Response Surface Designs

In course of trying some sets with suitable unknown parameters as factorial combinations for obtaining Rotatable Designs that have some more desirable property, two biquadretic equations were faced. Instead of abandoning the problem a suitable program was written for solution of such equations. With the program in hand any number of designs could be obtained, tabulated and tested for efficiency.