

THIRD PAKISTAN STATISTICAL CONFERENCE

LAHORE, FEBRUARY 1956

“STATISTICS MUST HAVE PURPOSE”

PRESIDENTIAL ADDRESS

By P. C. MAHALANOBIS

I am grateful to the convenors of the Third Pakistan Statistical Conference for the honour they have shown by inviting me to preside over this Conference and for the opportunity they have given me to meet the statisticians of Pakistan. I offer my greetings to all my friends here on behalf of my colleagues in India. My pleasure in being here today is all the greater because many of the statisticians in Pakistan are my old students and also because my own ancestral home was in Vikrampur, near Dacca. I am glad Dr. Quazi Mutabar Hussain (Head of the Institute of Statistics at Dacca) and Dr. M. Zia-ud-Din (Head of the Institute of Statistics at Lahore) have been able to come to Calcutta on several occasions to give lectures and participate in the work of the Indian Statistical Institute. I am glad young statisticians from Pakistan have been coming to Calcutta for training every year since 1950. In India we shall always look forward to the visit of our friends from Pakistan.

2. I am aware the Pakistani statisticians look upon me as their genuine friend; and I am deeply touched. But this cannot be your only reason for asking me to preside over this Conference. You will probably like to know about our experience in India and to learn from our successes and failures. I shall, therefore, tell you something about recent developments in India; and also try to convey to you the one lesson we have learned, namely, that statistics must have a purpose.

3. The basic problems of life are similar all over this sub-continent of ours. Poverty and unemployment are only different aspects of the low level of production and are the two most serious problems of both India and Pakistan. The solution of these twin problems of poverty and unemployment will require much hard thinking and positive action based on factual information. It has been clearly realized in India that economic policy and planning must be based on reliable statistics and that it is essential to build up an efficient system of statistical institutions and agencies.

EARLIER DEVELOPMENTS

4. Statistical activities are intimately associated with social and economic development. In the British period the collection of trade statistics began very early, on account of the trading interests of British merchants, and gradually became well organized. The population census was started in 1872. As crops are mostly rain-fed, agricultural out-turn is largely determined by the monsoon rainfall: and quite naturally rainfall and meteorological data began to be collected fairly early. Modern statistical methods (for example, multiple correlation) began to be used in the Meteorological Department under Sir Gilbert Walker more than 40 years ago and are still being used to prepare forecasts of seasonal rainfall over large regions. Statistical methods were used with much success

about 25 or 30 years ago in studying rainfall and floods which led to successful measures of flood prevention and control in many regions. These methods also began to be used in the hydrological research laboratory in Lahore many years ago.

5. Just about a third of a century ago Sir Ronald A. Fisher had developed the design of experiments which made it possible for the first time to study critically the variations in the yield of crops associated with different factors (varieties, manures, methods of cultivation etc.). These methods were introduced in India very early and are being used on an extensive scale since then to improve the efficiency of agricultural production. Some excellent work has been done on the design of experiments by my old student Quazi Mutahar Hussain of the Institute of Statistics at Dacca.

6. Jute was a most important cash crop in Bengal and naturally it was a matter of great concern to have reliable forecasts of the out-turn of jute. Estimates based on village returns were found entirely unreliable. Exactly twenty years ago I was asked by Government whether a sample survey could supply reliable forecasts of jute out-turn. This led to extensive researches spread over five years (1937-1941) which made it possible to develop a method by which estimates of the jute out-turn could be prepared with a margin of uncertainty of about two per cent. Since then sample surveys are being used in India on an increasing scale.

RECENT DEVELOPMENTS IN INDIA

7. This was broadly the position at the time of independence in 1947. Since then, in India there has been rapid developments in statistical activities mainly in response to the increasing demand for information required for purposes of planning. A Central Statistical Unit was established in New Delhi in 1949 and was converted into the Central Statistical Organization (CSO) in 1951. A National Income Committee was appointed in 1949 which laid the foundation for an orderly development of national income statistics. The National Sample Survey (NSS) was initiated in 1950; and is collecting since then social and economic information from forty or fifty thousand households all over India in the form of two "rounds" of survey every year. The Planning Commission was established in 1950. The Statistical Bureaus in the different States are being steadily strengthened and an annual conference as well as special conferences of Central and State statisticians are being organized every year.

8. I must also say a few words about the Indian Statistical Institute which was started as a non-profit scientific society in 1931. The Institute now has a Research and Training School; a Projects Branch which is looking after the statistical work of the National Sample Survey; a laboratory for electronic computers with a large workshop; Statistical Quality Control Units in Calcutta, Bombay, Bangalore; a rapidly growing section for planning; and an official organ *Sankhyā*: the Indian Journal of Statistics. There is also a cooperative enterprise, the International Statistical Education Centre which was established under the sponsorship of the UNESCO in 1950 and is being maintained by collaboration between the International and the Indian Statistical Institute with the support of United Nations and its Specialized Agencies (ILO, FAO, WHO, etc.) and the Government of India. Trainees come to the ISEC every year from Pakistan and other Asian countries.

STATISTICS MUST HAVE A PURPOSE

9. From our own experience in India I do not think I can give any better advice to my colleagues in Pakistan than to ask them always to remember that statistics must

have a purpose. Statistical data may be required to solve problems of a practical nature or for purely scientific or theoretical investigations. Some of the data may be required for current use; some may have to be collected for future utilization. The point to be stressed, however, is that statistics must have some purpose, practical or theoretical, and for immediate or future use.

10. Before starting to collect any new statistics it is useful to pause and enquire "for what purpose?" When a statistician is requested to collect some statistics his first responsibility is to ask the person or authority making the request to explain as clearly as possible the purposes for which the information would be used. One may find, for example, that statistical data are already available which would serve the purpose. Or, it may turn out that information of some other kind than that suggested at first would be more suitable. Also, when a complete census is suggested, a sample survey would be often found fully adequate.

11. Once the purpose for which the statistics is required has been made clear it becomes possible to construct suitable concepts, definitions, and standards. Also, as information begins to be collected and utilized it becomes possible to make necessary improvements in the methods of collection and of processing the data and in the utilization of the information more and more effectively. This is not an easy task; and sometimes much experimentation is needed before appropriate definitions can be set up. I can give an instructive example. When the National Sample Survey first started collecting information on unemployment in India the concepts and definitions adopted were broadly the same as those used in the western countries. It was, however, soon found that the western concept of unemployment was inadequate because millions of persons in India work on their own account in household enterprises. They may have to sit partly idle for lack of work; but they can neither gain nor lose paid jobs. It became necessary, therefore, to think not only of paid jobs which one may get or lose but also of the amount of work one is doing or the amount of money he is earning by such work. It was necessary to introduce the concept of "normal period of work" and "normal earnings"; and to collect information on the actual period of time during which the person is gainfully engaged in work and express this period of time as a fraction of the normal period of work; or to ascertain the actual amount earned in a given period and express this amount as a fraction of normal earnings; and to use these fractions as measures of employment. Sometimes it is not possible to speak of the degree of employment or unemployment (even in terms of period of work or of earnings) of a single individual; and it is necessary to refer to the household enterprise as a whole. It is necessary continually to construct new concepts and definitions to suit the needs of underdeveloped countries.

12. In underdeveloped countries there is a special danger of wasting effort and money in collecting statistics in imitation of more advanced countries of the world. Some of the information which are regularly collected in western countries (because of social and economic needs within their own socio-economic frame-work) may serve no useful purpose in a country like India; in such cases I would unhesitatingly refrain from collecting them. As this point became more and more clear much deadwood was got rid of in Indian statistics. There is also danger of wasting money in statistical processing and publications of a lavish but meaningless type in imitation of more advanced countries. It is not necessary to use, for example, large coloured charts or very expensive printing in publications meant for the use of statisticians. I am referring to such seemingly trivial points only because I have seen much waste of money in such matters in India.

STATISTICS FOR POLICY DECISIONS

13. In India we look upon official statistics as essentially a tool for economic policy. Statistical information has been of decisive significance in making important policy decisions. Rationing of food grains had been introduced during the war; there was decontrol for a short period in 1948 during which the price structure was completely distorted and food control had to be quickly restored. This, however, created many difficulties; and there was a sharp division of opinion when proposals for decontrol began to be considered in 1952. By this time, information on the consumption of cereals was being directly collected from several thousands of households scattered over hundreds of villages and a large number of urban areas throughout the country. The sample estimates based on the NSS showed a much higher total consumption of food-grains (about 25 or 30 per cent higher) than the estimates of out-turn of food crops compiled from the *patwari* and similar records. After a heated controversy the Government of India accepted a policy of gradual decontrol of food rationing which was entirely successful. There is no doubt that the Government decision was influenced by the sample estimates.

STATISTICS IN RELATION TO PLANNING

14. More recently, statistics has been steadily growing in importance in connexion with economic planning in India. In the Second Five Year Plan (1956-61) it is intended to expand as much as possible the basic industries (heavy machineries, mining, steel, electricity, transport etc.) and also expand education, health, and social services which would create a great deal of new purchasing power and, hence, fresh demand for food and consumer goods. It would be also necessary to estimate what would be the likely additional demand for different commodities and services due to the newly created income and purchasing power. The production of food and consumer goods must be expanded to meet the additional demand. This would ensure that there would be no inflation.

15. As the production of a given quantity of one commodity would require appropriate quantities of other materials, the production targets would be all inter-locked through inter-industry relations. It is necessary therefore that the production targets should be internally consistent in the sense that requirements at each stage would be balanced by the supply. If there is any lack of balance then there would be shortages and bottlenecks which would create delays and hamper the process of production or can even bring it to a standstill. Similar balances would have to be worked out for requirements of manpower. The investments required to achieve the targets of production would have to be worked out and balanced with the financial resources.

16. To formulate such a plan of production would obviously require a very large amount of detailed information of many kinds such as inter-industry relations; capital required for each unit of production; increment of income and of employment per unit of additional investment; pattern of consumption of different goods and services and how this pattern changes with increase of income; requirements of transport and many other things.

17. The responsibility of the statisticians does not end with the preparation of the plan. Suitable methods must be devised to assess the progress of the plan in physical terms, to ascertain to what extent the production targets are being realized, unemployment is decreasing, levels of income and consumption are rising, and social and cultural benefits are accruing. The need of a wide variety of detailed information for purposes of planning has given a great stimulus to the statistical services in India. Our greatest difficulty has been the acute shortage of statistical staff at different technical levels. Arrangements are being made, therefore, to expand facilities for training and research in statistics with

all possible speed. We realize, however, that it is a great handicap that adequate arrangements had not been made in advance.

SAMPLE SURVEYS

18. In India we have found that the use of sample surveys would enable a great deal of information being collected with speed and with sufficient accuracy for most practical purposes at a fraction of the cost of a complete census. The National Sample Survey is an agency of the Central Government with a permanent field staff of about 1100 persons (NSS) working under the direct control of the Ministry of Finance and a statistical section of about 800 persons (including office staff) working in the Indian Statistical Institute. The central sample covers about 1600 villages and 250 towns and the 4 big cities, Calcutta, Bombay, Madras and Delhi. In addition, the State Governments have been invited to set up a sampling organization in each State which would function independently of the Central NSS in organizational matters but would work in a fully co-ordinated manner at the technical level. Very broadly, the arrangement is that each State (or group of States) would try roughly to match the field staff of the central agency within its own region. Both the Central and the State agencies would use identical concepts, definitions, and standards so that the results would be comparable and, hence, compilable. There would be a common programme of work amounting to roughly 80 or 85 per cent of the total for which the same schedules would be used; the remaining 15 or 20 per cent of capacity would be used for special enquiries in which each agency (but not the other) is interested. Several States (U.P., Bombay, Travancore-Cochin) have already started work on an integrated programme; other States are likely to join in the near future.

19. The design of each round of survey is in the form of, at least, two and, at present, four independent and inter-penetrating net-work of sub-samples (drawn with replacement) each of which supplies a valid estimate. The differences between the estimates consist of both the sampling errors (arising from the fluctuations of sampling) and the non-sampling errors (arising from differences between different investigators, different conditions and methods of work and other sources of variation). Differences between sub-sample estimates can thus supply the total margin of uncertainty and not merely the sampling error. This is a great advantage of the method of inter-penetrating sub-samples drawn with replacement.

20. We have found the method of independent inter-penetrating sub-samples extremely convenient. We have been using this design from 1936 or 1937 and have now accepted it as a "standard" for all large scale sample surveys. Estimates are usually required of average values the distribution of which would be practically symmetrical in most cases. The probability that the estimate based on any one of the two or more sub-samples is less than (or, alternatively, greater than) the "true" or "population" value is half. The probability of two of the sub-sample estimates being both less than (or greater than) the "true" value of the average is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$; in the same way, the probability of n sub-sample estimates being *all* less than (or greater than) the true value is $(\frac{1}{2})^n$. The probability that the true value lies within the range of n sub-sample estimates is $[1 - (\frac{1}{2})^{n-1}]$. This method thus supplies without any calculations a good idea of the margin of uncertainty of the estimates.

21. The point to be emphasized is that the above result would be true not only of the average values but of any parameters or coefficients which are functions of these average values. In fact, if two or more sub-samples are subjected to the same processing operation (including linear or non-linear transformations of all possible

kinds) even then the above result will hold. We can use sample designs for which the formula for calculating the error based on one single sub-sample is not even known because we can find (almost without calculation, so to say) the over-all sampling variances from the differences between sub-sample estimates (irrespective of the methods by which the data are processed provided these methods are the same for all sub-samples). It is particularly convenient in an under-developed country like India to take advantage of the above result because it is possible to dispense with elaborate methods for calculating the standard error which may even require the use of complicated electronic computers. Interpenetrating sub-samples can serve as an automatic computer for the purpose of calculating the margin of variation of sample estimates.

22. I should also like to stress that inter-penetrating sub-samples supply most valuable controls over the work of the field investigators as well as on the processing and tabulation work. A comparison of sub-sample estimates would often reveal gross mistakes and enable these being corrected at an early stage. Finally, inter-penetrating sub-samples offer an excellent method for the study of differences arising from different types of schedules and questionnaires or different methods of conducting an enquiry by associating different types of work (or "treatments") with different sub-samples. From a statistical point of view, the inter-penetrating sub-samples can be considered exactly equivalent to replicated sub-samples in the Fisherian design of experiments and are amenable to analysis of variance in exactly the same way. (The only difference is that, in sample surveys, it is possible to calculate the sampling variance separately for each sub-sample estimate; and, hence to make a deeper analysis than is possible in agricultural field trials.)

23. Not only this. Inter-penetrating sub-samples would also supply a valid method for the comparison of the results from one round to another. If the observed differences between rounds are found to be of the same order as the sub-sample differences within the same round then one would infer that there was no significant change from one round to another. On the other hand, when between-round differences are significantly greater than within-round differences then it would be possible to assert with confidence that some real change has taken place. The results for different rounds would be clearly amenable to analysis of variance but usually no complicated calculations would be necessary as between-round differences can be directly compared with within-round differences.

24. I have been dwelling at such great length on sample surveys because, from our experience in India, I am inclined to think that in under-developed countries well-organized sample surveys on a country-wide scale would offer the most economical, the speediest, the most practical, and the most scientific way of advance in statistics. The Sample Survey should be set up as a permanent agency preferably with a whole-time staff (although there can be no objection to using a part-time staff under certain conditions). The design should be probability samples arranged in the form of inter-penetrating sub-samples drawn with replacement. There should be at least one survey each year; but two rounds per year would be better and can be easily arranged even in a large country like India. It would be still better to have three or even four rounds per year which should not present serious difficulties in countries smaller than India.

25. Such a Sample Survey would be able to collect continuing information on practically all phases of household life covering demography, employment and unemployment, income and savings, agricultural production, cottage industries, trade, transport, services, consumption (in terms of both expenditure and physical quantities in many cases), housing, education, simple information on health etc. It can also supply information on area sown, yield per acre, and total outturn of important crops. In addition to a fairly standard list of items which would be repeated from one round to another the Sample Survey

would be able to undertake at short notice *ad hoc* surveys of many kinds. For example, in India various special enquiries have been made during the last few years on incidence of existing taxation, pattern of small savings, land utilization (broadly on the lines of the FAO Agricultural Census), the distribution of land holdings by size and land tenancy etc.

26. I should like to stress the great flexibility of the sample survey. It is possible to start on a very small scale with a small nuclear staff working with a sample which would give total estimates for the country as a whole for a few important items. The scope and size of the survey can be then gradually increased either by including more and more items of information in the programme of work or by increasing the size of the sample (by including more villages and towns or more households in the sample) or by expanding both the scope and the size of the sample in a balanced way at the same time. In this approach it would be possible to give training to the field and the statistical staff and to increase their number gradually and in this way to build up the whole organization on sound lines.

27. In addition to a large-scale sample survey it would be most useful to have one or more small but efficient technical teams for "diagnostic" or "type" studies for investigations relating to structure rather than for sample estimates. For example, a type or diagnostic study can be made of the hand-loom industry or rural credit or agricultural marketing or technological norms or productivity in organized factories and so on. Such type or diagnostic studies are often of great help in developing appropriate methods for the collection of statistics through sample surveys or other suitable agencies.

28. I venture to think that an well organized sample survey at the national level, however small, and one or two expert teams for type studies would supply a most effective base for the expansion of statistical services in under-developed countries. Once such an agency is established it would be possible to make a rapid advance in response to increasing demands; and gradually to build up a sound statistical system to serve the country's needs.

RESEARCH AND TRAINING IN STATISTICS

29. I have already mentioned that in India the lack of trained personnel is the real bottleneck in statistics. We are trying to expand as quickly as possible different types of training at different technical levels. Statistics is a subject for the Master's Degree at 9 or 10 universities. We want such university training to continue but we do not think it to be sufficient. We attach much importance to professional and technical training in statistics (as much as we would do in engineering or medicine). We believe such training can be given or received only when the trainees are working on professional or technical jobs. We are now trying to provide professional training in the Indian Statistical Institute, the Central Statistical Organization or within the National Sample Survey and other agencies.

30. Here also I feel it my duty to throw out a warning. There is a tendency in certain quarters in India (and in other countries also) to try to make out as if "mathematical statistics" is a subject by itself without any reference to any applications to scientific research or practical problems. I must confess I do not understand this position. To me statistics must always have a purpose which may be, of course, to help in drawing valid conclusions from scientific observations for purely theoretical researches (for example, the distribution of galaxies of stars or other problems in astrophysics); or, to help in selecting best varieties for agricultural production; or, to help in industrial or economic development. There is plenty of scope for mathematical researches in connexion with statistical problems. I should like to give all possible encouragement to really able mathematicians to enter the field of statistics. But I should still hold that statistics is not a branch of mathematics but a field for the application of mathematics. Pure mathematics has its own justification independently of applications. But immediately one thinks of applications there must be a reference to some purpose, that is, to some contingent problem requiring to be solved.

One must then work on contingent relations which are capable of being observed or measured and on the basis of which it would be possible either to draw valid inferences in scientific matters, or to take decisions in practical problems.

31. Pursuing the question of training and research in statistics, I think that there would be many advantages in basing the teaching programme on actual professional or technical work. On this view, it would be a good plan to use a sample survey organization itself as a centre for professional and technical training in statistics. I do not, of course, have sufficient knowledge of present arrangements in Pakistan to be able to offer any advice. I may, however, make one or two general suggestions for your consideration. I should think it would be of great help to establish a Sample Survey organization in Pakistan (if necessary, on a small scale) at an early date. I have already mentioned that in India we have found that one central agency is not enough; we have already started building up decentralized agencies in different States or groups of States; we believe we should ultimately have 6 or 8 well-organized regional centres in India each covering a population of from 30 to 50 million. In the same way, it may be convenient to have two regional sample survey organizations, one in West and another in East Pakistan. It may be of great advantage if the two Institutes of Statistics in Lahore and Dacca can be closely associated with the statistical part of the work. I can not, of course, say anything about the details. But if any arrangement of this kind is feasible then it may also be possible to make professional and technical training an integral part of the teaching programme of the two Institutes of Statistics in Lahore and Dacca. Besides work on sample surveys, the two Institutes can help effectively in the design of experiments in agriculture and industry and take up enquiries relating to irrigation and floods. It would seem desirable to establish close links, at the scientific and technical level, between the two Institutes of Statistics and the statistical work in government offices.

INTERNATIONAL CO-OPERATION

32. I have continually stressed the importance of recognizing national needs and purposes in all statistical activities. There is also much need and scope of active collaboration at the international level. There are many problems which are common to different countries where the same methods and approaches would be useful. There is a continuing need of communication of information between different countries and also of making international comparisons. On the theoretical side there are methods which are in universal application. It is necessary and desirable that we should continually foster active international co-operation and collaboration in the field of statistics.

33. Just as I believe in the reality of national needs I also recognize the special needs of particular regions. We have much in common in climate, geography, plants and animals, habits of food, clothing, housing, and the manner of living in India and Pakistan. It should be particularly useful, therefore, for us to exchange our ideas and experience. This is why I have ventured to speak at such length of Indian experience in the hope that this will be of some help to you. We should also like to profit by your experience. It would be very useful if small groups of statisticians from Pakistan and India can meet for technical and scientific discussions from time to time. The meetings can be arranged in either Pakistan or India according to convenience. It would be useful to get working papers or scientific communications printed and circulated in advance.

34. I should like you kindly to consider this suggestion carefully. If you approve of this idea we can try to invite statisticians from Pakistan to attend a Joint Statistical Conference in India at a convenient date to consider technical and scientific problems of common interest. Discussions at such conferences would be of course at a personal and not at an official level but it would be proper to look to our respective Governments for general support. We have enjoyed your hospitality. We hope you will now give us a return visit.