STATISTICS FOR ECONOMIC DEVELOPMENT*

By P. C. MAHALANOBIS
Indian Statistical Institute

Friends,

I am glad to have this opportunity of being with you today. I consider it a great honour to follow on the lecture given by Sir Ronald Fisher on "Statistics for the Acquisition of Knowledge of the Real World." I shall make certain observations which may perhaps be considered to be supplementary or complimentary to some of his observations this afternoon.

I have selected the topic, "Statistics for Economic Development." The title I think indicates the point which I shall be continually stressing, namely, that statistics has two aspects one of which is utilitarian or economic; and the other, which is scientific, or as Sir Ronald has called it, aesthetic, in the sense of giving intellectual satisfaction.

The utilitarian or economic phase of statistics had its origin in time immemorial, from the collection of information relating to social and economic conditions to help in making policy and administrative decisions. The English word "statistics" is indeed connected with statecraft, the business of the State or of Government or Administration (1).

In every country whenever there has been an upsurge of social and economic and political activities, there has also been a rapid advance of statistics. In my own country, in India, during a period of great national prosperity in the 3rd or the 4th century B. C., in the time of the Emperor Asoka, we find a most elaborate system of statistics of which an account is given in the *Arthasastra* of *Kautilya* (2). Nearly two thousand years later in the 16th century, in the time of the great Muslim Emperor Akbar, we find again a very elaborate system of statistics in the comprehensive survey of the country called *Ain-i-Akbari* (3). In the British period in India, at the time of renewed countrywide economic activities, there were great advances in statistics of which one striking example was the survey of East India by Francis Buchanan in the first decade of the 19th century (4). Again, during the last ten or twelve years since independence, there has been a great deal of expansion of statistical activities in India.

I have given a few examples from my own country to illustrate the point that the advance of statistics has accompanied or has been an essential factor in social and economic progress. This is equally true for other countries (5).

The concept of random events emerged much later, as Sir Ronald Fisher has pointed out today. It came in connexion with games of chance, for example, in rules for the equitable division of stakes which I believe were given by Cardan and by Galileo in the 16th century.

---

*Presented at the Special Meeting on May 23, 1960, sponsored by the Asahi Shimbun under the joint collaboration of the Mathematical Society of Japan, the Research Association of Statistical Sciences and the Operations Research Society of Japan and originally published in the *Journal of the Operations Research Society of Japan*, Vol. 3, Number 3, January 1961 is reprinted here with the kind permission of the Editor.

A series of Bibliographical References are given in an Appendix; the reference numbers (1), (2), (3), etc. correspond to the serial number of these Bibliographical References.

The theory of 'mathematical probability' in fact was developed in connexion with problems of chance; this, I venture to suggest, had both a motivation which is utilitarian or economic (in estimating the chance of winning a stake) and another, which in Sir Ronald's sense, is aesthetic and which may perhaps be also called mathematical or logical.

This dual aspect of motivation in statistics led, later on, to a dichotomy in the organization of teaching and research in statistics, one portion being included under 'economic' and the other portion under 'pure mathematics.' This has had some unfortunate consequences. The attempt to teach and treat statistics as a branch of pure mathematics has led to sterile exercises in abstract mathematical symbols which have no relation to the real world.

This point deserves a little elaboration. I shall accept Bertrand Russell's description of pure mathematics, as all reasoning of the form : "If P, then Q", with the essential condition that one must not enquire what P and Q are, or whether they are real. And yet, my friends, I think you will not dispute that statistics is very much concerned with observations, classification, measurement, and experimentation in the world of physical reality, in all the sciences, both natural and social. Statistics in fact, has to deal with the raw data of all the sciences, natural and social, and therefore must deal with the world of reality. I submit, statistics cannot possibly be considered to be a branch of pure mathematics (in the sense of "if P, then Q").

Also, we do know that the statistical approach has had a tremendous impact on the progress of science. Let me give some examples. Consider the concept of a normal distribution of statistical variates. It had its origin in the mathematical theory of games of chance, but it also supplied the foundation for the theory of errors of observations in the hands of Gauss and other famous scientists. This still remains the basic method for the adjustment of physical observations in all the sciences.

Again, the same normal distribution was applied with great success in the kinetic theory of gases, in studying the distribution of velocities of molecules in three dimensions, which led to spectacular results in physics. I may also mention the development of "statistical mechanics" with the emergence of the concept of entropy linked with probability and the level of randomness of distribution of physical states.

In this way the great second law of thermodynamics has a statistical foundation. And it is this second law alone, among the so-called laws of physics developed in the 19th century or even during the last half a century, which still survives in its original form. Newton's laws of gravitation had to be modified, even Einstein's theories are being modified; but the second law still stands, perhaps, may I venture to suggest, because of its essentially statistical nature ?

Sir Ronald has referred to the need of some kind of axiom of ignorance in statistics. We find this also in physics in the principle of uncertainty enunciated by Heisenberg.

I shall not try to give further examples from physics. In biology I should like to mention that it was Sir Francis Galton who first got the concept of statistical correlation between biological variates such as the height of fathers and the height of sons. Now, this concept was entirely new in a physical sense, although much earlier the great Gauss himself had dealt with the mathematics of correlation in the form of the cross product of errors of observation without, however, being aware of the physical notion of correlation (6). Thus even Gauss missed the statistical concept of correlation although he had handled its pure mathematics. I think this supports the point which I made earlier, namely, that statistics deals with the world of reality while mathematics is concerned with the world of abstraction in which the question of physical reality is entirely meaningless.

However, mathematics has to be used continually in dealing with the real world, but as a tool. There is a wide gulf between the "world of reality" and the "world of pure mathematics," and Sir Ronald Fisher has shown the way to bridge this gulf with the help of the concept of "fiducial probability."

Perhaps I may give a very simple example. Let us consider the tossing of a coin. If it can be assumed axiomatically that the coin is unbiassed then, of course, statements about results of repeated tossing of the coin would be a part of deductive mathematics, in which the concept of "repeated sampling" would be of crucial importance.

But, in the world of reality when we have a real coin, the axiom is not written on it that it is unbiassed. It now becomes a problem of statistics to find out, from an observed series of observations of tossings of the coin, whether the coin is unbiassed; or, if it has a bias, to estimate the bias. This is a problem of statistics, that is, of inductive inference and not of deductive mathematics.

On this view, statistics must be treated as an applied science or a technology with a dual motivation, one utilitarian and economic and the other, I am using the convenient word aesthetic, in the sense of something which gives intellectual satisfaction.

On this view, statistics must always have a purpose, either to help in making policy and administrative decisions in economic and social affairs or in making inductive inferences in all the sciences, both natural and social. The statistical method now emerges as a technique for the collection of data and for extracting information from such data for a given purpose with maximum efficiency, that is, at minimum cost.

Let us look to the last three or four decades of statistics itself. Sir Ronald Fisher introduced the concept of design of experiments and the analysis of variance in order to improve the efficiency and usefulness of agricultural field trials (7). We see here the dual motivation, one utilitarian and economic and the other aesthetic, for example, in the beautiful concept of equi-partition of variance which is the basis of the analysis of variance. The normal distribution of statistical variates had supplied the basic concept of distribution of velocities of molecules in the kinetic theory of gases which reached its highest generalization in the principle of equal distribution of energy among the different modes of motion or different degrees of freedom of the gas particles. I venture to suggest that perhaps this theorem again, in its turn, supplied to statistics the basic principle of equi-partition of variance in the analysis of variance.

It is about thirtyfive years ago or a little more, in the early 1920's, that the concept of design of experiments was introduced. At about the same time, Walter A. Shewhart developed the method of SQC or Statistical Quality Control, which he very accurately described in his classic work as the *Economic Control of the Quality of Manufactured Products* (8). It is a beautiful method with which we are all familiar; it has something which satisfied scientific or aesthetic values, and yet it is basically motivated by utilitarian or economic considerations.

All these methods involved the application and use of a good deal of mathematics. It is very proper that we should use mathematics to solve practical problems. We cannot have too much of mathematics for statistics, but as a tool, not as a master. Statistics must have a purpose and mathematics must serve this purpose.

I may now turn to the field of economics. It was unfortunate in many ways that one part of statistics was included under pure economics, and got somewhat divorced from reality in a different way.

As time is short, I shall draw attention to only two points. Firstly, among economists, there developed a somewhat naive faith in the infallibility of a so-called complete enumeration or census, with almost complete lack of awareness of the fact that the collection of information about economic and social conditions are as much, or I should say, even more, subject to errors of ascertainment than observations in natural sciences.

It is only during the last decade or two that there is increasing recognition that censuses, even when they are called complete, are often incomplete and subject to errors. It is interesting to note that at the session of the International Statistical Institute which would be held in Tokyo next week there would be a special meeting to consider errors in censuses.

During the last twenty years, appreciable advances have also been made in the design of sample surveys based essentially on the fundamental principles of randomization, replication, and local control, first introduced in the Fisherian design of experiments. It has been shown that a properly conducted sample survey can supply information of sufficient accuracy for most practical purposes with greater speed and at a fraction of the cost of a so-called complete census. The accuracy of a sample survey is often greater with the added advantage that it is possible to calculate the margin of error on a logically valid basis. The method of interpenetrating net work of samples (IPNS), that is, two or more independent samples drawn with replacement, furnishes a very powerful tool for the study and comparison of both sampling and non-sampling errors (9).

In recent years there have been also great advances in operational research in which statistical methods have played an important role. In all these developments the motivation has been both utilitarian or economic and scientific or aesthetic. These two aspects, the economic and the scientific (or technological) aspects of statistics, like two sides of the same piece of paper, may perhaps be distinguished but cannot be separated.

There is continuing need of the use of mathematics in statistics; as I have already mentioned we cannot have too much of it. There is also continuing need of economic thinking in dealing with practical affairs. But I submit that statistics cannot any longer be dichotomized and included partly, under pure mathematics and partly, under pure economics, or wholly under one or the other.

The time has come to recognize statistics as an integrated discipline or technology in its own right. Teaching and research in statistics must therefore have an intimate relation to all the sciences, both natural and social. The teaching of statistics should be looked upon as something analogous to the teaching or training in engineering or medical sciences. I am glad to mention that, in India, in the Indian Statistical Institute, we hope to be able to introduce, very soon, full professional courses leading to the degree of Bachelor of Statistics or Master of Statistics; the degree of Bachelor of Statistics being exactly comparable to a professional degree in engineering or medicine.*

*These courses were inaugurated in Calcutta on 16 August 1960.

I should now like to consider some of the problems of economic development in India where the statistical approach, or perhaps you may call it, the approach of operational research with the help of statistics, was extremely useful.

I may start by saying that sophisticated economic theories, which may be appropriate for advanced countries, had acted for a long time as a formidable thought barrier to economic progress in India. India had started making steel in 1908. But in 1950, India was still making less than one million ton of steel per year, although India had a population of about 380 million, and had more iron ore of better quality than any other country of the world, more than USA plus USSR.

In 1950, Government had decided and announced that a plant to produce a second million ton of steel (a second and only a second million ton of steel) would be installed but this was immediately followed by a survey of current demand by economists and expert committees who pointed out that India was making one million ton of steel and was importing about 300,000 tons while the additional demand could be for only 200,000 or 300,000 tons. That is, the total current demand was only 1½ million tons; and the question was raised would it not be foolish to produce two million tons of steel ? This was the subject of a beautiful exercise in economic theory which led to the proposal for a new factory for a second million ton steel being dropped from the First Five Year Plan of India. We were assured that this decision was in accordance with the most refined economic theory, and that we should not produce two million tons of steel but purchase steel from abroad. This has cost India hundreds of million rupees subsequently for the import of steel.

Looking up the statistics of UK, USA, and USSR, I found that they were producing a great deal more steel than India, although the population in all those countries was much less, even the population of the three countries put together was only somewhat greater than that of India. They were not only producing a great deal more but they were going on increasing the production of steel. Japan also was doing the same thing; and I thought then all of them must be either bigger fools than my countrymen or there was something wrong in our own outlook.

Economic theory was originally developed for advanced countries; and is appropriate for such countries, because short range decisions can be effective in highly developed countries. If the demand goes up but there is a shortage of supply, the production process would get into operation very quickly. In an advanced country, if there is an export subsidy, motor cars would roll out of the factory in large numbers. But in a country which has no fatories, even if you give a billion dollars not a single motor car can physically come out until the tools of production and the technical personnel become available.

The heart of the problem of an underdeveloped country is this factor of time. It takes time to acquire the ability to manufacture goods. It is necessary to establish factories. It is necessary to have the scientific and technical personnel, and this takes time. Therefore, the problem of economic development has to be viewed over a long span of time. One must see where we want to go and how we should go there. And, in trying to do this, it is essential, continually, to use quantitative information, that is, statistics.

I may perhaps give an example to illustrate this point. In India, population may be increasing, we are not very sure, but it is quite possible that population is now increasing at the rate of seven million persons per year. Now, roughly, we need one ton of foodgrains for seven persons. On this basis, the additional foodgrains required for seven million persons would be one million ton per year, in five years this would come to 15 million tons. If we have to import this, at the world price of about 90 dollars per ton, we would have to spend about 1300 or 1400 million dollars in foreign currency in five years.

Instead of foodgrains, we can import fertilizers in the form of, say, ammonium sulphate. In India we have found that one ton of ammonium sulphate would usually give two or a little more than two tons of foodgrains; to grow 15 million tons of foodgrains we would require roughly half or say 7.5 million tons of ammonium sulphate. At a price of 50 dollars per ton, the total cost would be less than 400 million dollars in five years. This would call for long range planning because orders have to be placed in advance for the fertilizers and arrangements for distribution have also to be made in advance. But, if we can look ahead and plan for the future, then the savings in foreign currency would come to one thousand million dollars in five years.

We can look further ahead. Instead of importing fertilizers, we can manufacture fertilizers in India. The cost of installing a factory to produce a million ton of fertilizer per year, we find from our own experience, would be about 150 million dollars; and if we install five such factories in five years, the cost would come to 750 million dollars in five years. But a good part of this amount would be wages of domestic labour, and cost of domestic material; the foreign exchange required would be about 50 or 60 million dollars for each factory, or say 300 million dollars in all in five years which would be much less than cost of imported fertilizers. Also, once these factories are established, they would continue to produce for a very long time beyond the five year plan. But we must now plan eight or ten years ahead, because the first factory would take three or four years to start production, and so on for each of the factories which are to be erected in successive years.

We can go a step further and establish a factory to make the machinery to install new fertilizer factories, and the cost of foreign exchange may be only fifty or sixty million dollars, once for all. If we do this then the machine building factory would go on producing machinery to install a new fertilizer factory every year. In this way, only fifty or sixty million dollars of foreign currency can serve exactly the same purpose as 300 or 400 million or even 1400 million dollars in five years. This would, however, depend entirely on our ability to plan fifteen years ahead. I have given a simple example of the kind of thinking which is needed in India. In Japan, you are already familiar with this.

I should like quickly to sum up. In India we are now looking on the problem of economic development as having a five fold structure, with five aspects or five phases. It is easy to establish consumer goods in four or five years. We can import the machinery, work the factory with a diesel engine, and produce the goods. This does not take very long. The second phase or aspect consists of the large scale production of electricity, development of modern communications, mining, and light engineering. This would take eight or ten years to make an impact. The third aspect is of a more long term nature with the large scale production of steel, metals, and heavy machinery which would take, at least, fifteen years. The fourth aspect is the training of scientific and technical personnel in adequate numbers; this would require planning twenty years ahead. And the most basic and fundamental is the organization of scientific research which would take one whole generation or more.

Rapid progress would be possible only if we start, as soon as possible, laying the foundations of scientific and technological research; the training of scientific and technical personnel; economy building up of the steel and heavy machines industries; and proceeding with electricity, communications and consumers goods etc., as much as possible out of domestic resources. This is the only way to make the economy self sustaining in the course of a single generation.

The problem of planning then, as we see it in India, is to have a time horizon of ten, fifteen, or twentyfive years. The basic principle is that the right quantity of goods must be made available at the right moment, either for consumption or for investment, and the right number of scientific and technical personnel must also be made available to produce and utilize the same at the right moment.

Whatever targets we set up, there must be internal consistency, or balance, between requirements and supply of all important commodities and of labour and scientific manpower; and such balances must be achieved not only over short periods but also over a time horizon of five, ten, fifteen or twentyfive years.

This has given a new challenge to statistics, or, if you like, to operational research with the help of statistics, firstly, for the continuing collection of essential information required to prepare the plans for economic development, not in a rigid way, but retaining a good deal of flexibility, and with a time-horizon as I have mentioned, of five, ten or twenty years or more; secondly, of assessing the progress of the plan; and thirdly, of arranging a feed back of the information on the assessment of plan implementation so that rolling adjustments can be made or drastic changes can be introduced in the plan in the light of experience (10). This is what I mean by "statistics for economic development", at a national level.

But I should like to refer briefly, to something even deeper or more comprehensive. A big country like USA, USSR or China can bring about economic development almost on the basis of the resources available within the country itself, without much concern about foreign trade, although for every country, however big, foreign trade can be of great help in economic progress. For a smaller country like India, it is necessary to give greater attention to the development of foreign trade. The importance of foreign trade would be still greater for countries which are smaller than India.

The long range planning of international trade cannot, however, be done by any country, however big, by itself. There is an urgent need of long range planning at the international level. I am convinced that this is a problem of supreme importance for the future of the world.

One of the important factors in international tensions is the very existence of under-developed countries. It is my conviction that a rapid transformation of the under-developed countries is an essential condition for an enduring and world-wide peace (11).

Statistics and operational research are important tools for this purpose. The work has scarcely started. It would require a long educative process to prepare our mind for the shape of the one world which we must achieve if human civilization or the human species is to survive.

With the progress of science and technology, ancient barriers of space and time are rapidly vanishing. Whether we welcome it or resist it, the world is coming closer and becoming more integrated. International cooperation in economic development of the world, *as a whole*, is inescapable to save the human species from annihilation. I am convinced, therefore, that there would be a continuing and increasing demand for statistics and also for operational research to bring in the new age of peace, prosperity, and progress of the world, *as a whole*. International cooperation at a scientific and technical level is the first step in this direction.

I, therefore, thank you for having given me the opportunity to give this lecture.