THE ROLE OF MATHEMATICAL STATISTICS IN SECONDARY EDUCATION

by P. C. Mahalanobis

Reprinted from the Bulletin of the International Statistical Institute

VOL. XXXIII : PART V

THE ROLE OF MATHEMATICAL STATISTICS IN SECONDARY EDUCATION*

by P. C. Mahalanobis

Indian Statistical Institute, Calcutta (India)

My object in writing this paper is to point out the advantages of including mathematical (or, as I should prefer to call it, theoretical) statistics as a basic subject of study in the upper classes of secondary schools and in the first two years of college or university work. I shall first explain what I mean by mathematical (or theoretical) statistics.

The word statistics is usually used in three different senses. First, it means observational data in a numerical form relating to all kinds of things such as births and deaths, agricultural crops, rainfall, exports and imports, wages and prices, etc. Secondly, it means methods of reducing such observational data by appropriate numerical calculations to obtain derived figures such as averages, proportions, rates of change, index numbers, etc. Thirdly, it is the study of aggregates or "populations" of individuals in which the interest lies in the properties of the aggregate or the populations as a whole rather than in the properties of the individuals as such. Although the three meanings appear at first sight to be somewhat disconnected, the third (population) aspect supplies the unifying principle.

The number of human beings in a city or a country, price quotations, rainfall, or individual figures relating to almost any thing we can think of can be viewed as forming appropriate aggregates or populations. When a simple measurement (like the length of a rod) is repeated indefinitely, the results constitute a population of measurements. If the individuals or elements in a population are identical in all respects then a description of a single element (together with the total number of elements) would supply complete information about the population. A population which forms the subject of statistical study, however, always shows variation among its elements. Statistics may be therefore also viewed as the study of variation. In fact the variability of the elements is an important characteristic of the population.

* This article was written for a forthcoming publication of the United Nations Educational, Scientific and Cultural Organization, and has been reprinted for the International Statistical Conferences: India 1951 by the kind permission of UNESCO.

When the population consists of an infinite or an indefinitely large number of elements, it is not possible to study all the elements individually. All that can be done is to study a portion or a "sample" of the population. For example, any finite set of measurements of say the length of a rod has to be viewed as a sample out of the indefinitely large number of possible measurements. The number of births in a given place in one day or in one month may in the same way be looked upon as a sample out of the population of births during successive days or successive months in the past or in the future.

Two important consequences follow. First, the samples must obviously be drawn in such a manner as to be fully representative of the population. The adoption of (what is called) a random procedure secures this objective. In such a procedure the selection (of the sample) is made without any reference whatsoever to the properties of the individuals included in the sample. Secondly, the adoption of a random procedure of selection immediately makes it possible to calculate the probability of occurrence of the sample. The general procedure is to construct an appropriate mathematical or statistical model of the population (that is, to set up a suitable hypothesis), and then calculate the probability of occurrence of any given sample (or of any statistic or statistics based on the sample) in accordance with the mathematical theory of probability. Provided the relevant mathematical problems can be solved, it is then possible, on the basis of such calculations, to infer whether the observed sample may or may not be reasonably considered to have been drawn from the postulated population, that is, to judge whether or not the properties of the observed sample is in agreement with expectations based on the assumed hypothesis. If any particular model of a population turns out to be unsatisfactory (in the sense that observed samples do not show the expected properties), it is possible to try other models. In this way, if and when a suitable population model or hypothesis can be constructed, it becomes possible to predict the properties of samples which can be drawn from it. If necessary, fresh experiments can be conducted (that is, fresh samples can be drawn) to test the hypothesis. In this way, statistics (as the study of populations and random samples) supplies the general method of scientific or inductive inference through the familiar cycle of (1) observation (sample); (2) hypothesis (population); (3) prediction (properties of samples); and (4) fresh observations with testing of the hypothesis leading to its verification (within the limits of errors of sampling) or to its rejection requiring further investigations.

The object of collecting statistics (in the sense of raw data, for example, the price of wheat) is to compare these with other data (of the same or similar kind, for example, the price of wheat or of some other cereal) to be collected at different times or at different places. Although a sampling theory is often not explicitly used (or is sometimes not capable of being formulated for technical reasons), such comparisons are possible, in principle, only on the basis of a postulated hypothesis (for example, that wheat or cereal prices are remaining steady) which may include a dynamic factor which is changing with time (for example, that prices are rising at a given rate) in which the samples are to be viewed as having been drawn from the same or different postulated populations. Secondly, the reduction of the primary data (that is, the calculation of averages, rates of changes, etc.) also clearly have in view the making of such comparisons (which, in the ultimate analysis, means the study of samples drawn from hypothetical populations). The first two meanings of the word statistics, namely, (1) raw data and their comparison; and (2) methods of reduction of raw data for purposes of comparison are thus, in principle, included in the third sense of the word, namely, mathematical or theoretical statistics which is the study (based on the theory of probability) of populations and random samples of observational data.

There has been a rapid increase in the applications of mathematical statistics in recent times. Measurements in astronomy, land surveys, and experimental physics formed the subject field for fruitful developments in the classical theory of errors in late eighteenth and early nineteenth century. A little later, statistical methods were applied with great success in the kinetic theory of gases, and more recently in statistical mechanics and the quantum theory. Statistical methods began to be used in social and anthropological studies more than a century ago. With the emergence of the theory of natural selection and evolution, attention was focussed on the study of biological variation and correlation which led to rapid advances in biometric methods including bio-assays and the quantification of qualitative data. During the last thirty years many important developments have occurred (primarily in response to the needs of agricultural and biological investigations) in the foundations of theoretical statistics on one hand, and in the analysis of variance and the design of experiments on the other. Another recent line of advance has been the statistical quality control of manufactured products in industry and various applications of statistical methods in industry and commerce. During the last 10 or 15 years there has also been a rapid growth in the use of sample surveys for the collection of economic, social, and demographic statistics of all kinds.

The statistical method has thus become an indispensable tool in the natural and social sciences as well as in industry and technology. It is also being increasingly used in economic and social analysis and planning. In consequence, the demand for competent statisticians has been increasing rapidly but the supply still remains inadequate all over the world. A sound development of statistical education and training has therefore become an urgent necessity in modern society.

In this situation the study of mathematical statistics in the upper school or at the junior college or university stage would have important advantages. It would impart a proper understanding of statistical methods which would be of value to those students who subsequently pursue research in the natural and social sciences or who enter the field of commerce and industry. It would provide a good foundation of statistical knowledge to those who wish to have further education and training with a view to becoming professional statisticians. It would also provide the non-specialist layman a broad appreciation of statistics and statistical methods which would give him a deeper understanding of social and economic problems and help him in the making of policy decisions. I may also add that an acquaintance with the methods of statistical logic and inductive inference would provide to those interested in such questions a fundamental approach to the philosophy of science. The study of statistics can thus help in developing the capacity for critical thinking and in creating interest in the pursuit of basic problems. Intellectual value as well as practical usefulness, the two traditional reasons which usually decide the choice of academic subjects, are both in favour of statistics.

I may now make some observations on the scope and nature of the proposed course. R. A. Fisher has pointed out that "the science of statistics is a branch of Applied Mathematics, and may be regarded as mathematics applied to observational data. As in other mathematical studies, the same formula is equally relevant to a widely different group of subject-matter".* It is desirable, therefore, to emphasize in the basic course the underlying unity of the general methods and principles. Secondly, the peculiar position of statistics as a link between mathematics and the contingent world requires to be clearly brought out. In a purely logical sense, Klein's dictum is, of course, true that it is never possible to bridge the gulf between pure mathematics and the physical world. Pure mathematics is deductive logic in its most abstract and powerful form. Once the premises are accepted, the conclusion must follow inevitably. This is why Bertrand Russell defined pure mathematics as all propositions of the form: "If *P*, then

---

* *Statistical Methods for Research Workers*, 9th edition, 1944, p. 1.

Q". In a significant sense, probability statements do not conform to the above formula. Secondly, it is never possible to be certain that any given physical situation is being adequately represented by a particular mathematical model or hypothesis; and it is, therefore, never possible to invest statistical inferences with the categoric certainty of mathematical deduction. In principle, statistical (or inductive) inference must always remain uncertain in an absolute sense. The role of mathematics in statistics is to provide methods for the precise evaluation of the extent of such uncertainty. And once this margin of uncertainty is known, statistical results provide a truly scientific basis for practical decisions and action.

The inherent difficulty of all statistical inferences comes up clearly in the logical disjunction between the two different definitions of probability, one based on a purely abstract theory of arrangements, and the other from the essentially empirical concept of regularity or stability of frequency ratios. The operational connexion between the two definitions of probability is the crucial point in the statistical method. For practical applications, the concept of the stability of the frequency ratio is indispensable. For calculations of probability, the mathematical model is inescapable. It is the dual use of both which gives mathematical statistics its great flexibility and strength. The ultimate justification of this dual approach is not logical but lies in the fact that it works in practice.

In the basic course it is, therefore, necessary to develop statistics as a branch of applied and *not* of pure mathematics. A good deal of practice in applications and the working out of numerical examples is therefore desirable. This would serve to emphasize that the use of statistical models is essentially empirical in nature. L. Hogben has pointed out that a *proof* of a statistical theorem is not a statement about the way a mathematician arrives at a generalization. A proof "sets out the connection between a result already surmised to be true for reasons commonly concealed, and often forgotten, with a view to exhibiting its connections with other known results and hence with a view to recognition of its *legitimate scope and limitation*". He has further observed: "As the number of mathematical generalisations increases, the status which a particular result occupies in the entire corpus of current knowledge perenially invites reconsideration, and the pathway of discovery becomes less and less retraceable in a jungle of more and more rigorous demonstration".*

* *Chance and Choice*, 1950, p. 179.

A teaching programme can be organized in many different ways. It is not necessary to lay down a hard and fast syllabus. In fact a great deal of discretion should be left to the individual teacher. For purposes of illustration I shall describe one possible line of development which I would probably choose if I were asked to give a course myself. My chief aim would be to make the students thoroughly familiar with the two different definitions of probability and the connection between the two. On one hand, I would develop the purely logical definition of probability based on the abstract theory of arrangements; on the other, I would introduce sampling and frequency ratios on an experimental basis. The two definitions can be then linked up through suitable statistical models or hypotheses; and sampling distributions can be discussed with the help of the theory of probability with a view to testing of such hypotheses. It may be useful to give some further details.

One can start with the definition of 'population' as a collection of a number of elements or elementary units; and define a 'sample' as a collection made up of elements drawn from the population. Each element (in the population or in a sample) can be viewed as possessing one or more properties in the abstract. The composition of any sample can be then defined in terms of elements possessing particular properties. The different ways of extracting a sample of a given composition would then correspond to 'combinations', and the different linear arrangements of the elements in a sample to 'permutations' in algebra. One can next introduce the concept of a 'mutually exclusive' as well as 'exhaustive enumeration' of the properties of all the elements in a given population. It is then possible to give a purely formal definition of probability of a sample of a given size and composition as the ratio of the number of ways of extracting a sample of the given size and composition to the number of ways of extracting all possible samples of the same size (but of differing compositions) from the given population. This, however, is a purely abstract concept; and involves, as Hogben has pointed out, an assumption that each element has an equal opportunity to be associated with each remaining item in choosing a sample (which in effect secures conformity to the principle of randomization). In order to emphasize the formal nature of the definition, Hogben has coined a new word, namely, the 'electivity' (or 'choosability') of the sample for the more usual phrase, mathematical probability. The use of this new word 'electivity' will be probably convenient for teaching purposes. Having defined electivity (or mathematical probability) in the above way, and having shown its relationship to the theory of permutations and combinations in algebra, it is possible to lay down the two fundamental addition and product rules of electivity (or mathematical probability). The binomial and the multi-nomial theorems can be then used, and the students can be given a good deal of practice in algebraic manipulation and numerical calculations of mathematical probability on usual lines.

At the same time, students should be made familiar with experimental sampling. One can start with the tossing of coins, throwing of dice, or drawing cards from a pack or balls from urns containing balls of two or more different colours. It can be pointed out that the tossing of coins or dice represents sampling with replacement, while the drawing of balls from an urn or cards from a pack of cards represents sampling without replacement. (If the balls or cards are put back before each fresh drawing of a sample, then of course these would represent sampling with replacement). After a little practice with the tossing of coins and dice, it would be possible to introduce the use of a set of random numbers for sampling purposes. Odd and even numbers can be made to represent heads and tails respectively; or the frequency of different digits can be experimentally counted and studied; and so on. It is convenient, at the same time, to start taking measurements with an ordinary scale of the length of a small piece of wood or metal rod; and in this way secure samples of observations as raw material for the study of the empirical distribution of "errors" or deviations from the mean value.

It is also possible to give simple illustrations of area (or two dimensional) sampling. One can use a sheet of squared paper say about 32 centimetres each way and containing something like, say, 100,000 small squares (each one square millimetres each). An assigned proportion (for example, 5%, or 10%, or 20%, and so on) of the small squares may be then selected at random (with the help of two random numbers used as coordinates) and marked in black or coloured ink which would immediately supply a two dimensional field for experimental study. Each small square (one sq. mm.) would represent an element (or elementary unit) and would have the property of being either coloured or white. A sample-unit can be set up as either of the same size as an elementary unit (that is, one sq. mm.) or of a larger size, such as 2 adjoining elementary units, or 4 or 9 or 16 elementary units arranged in the form of a square; and so on. Sampling units can be then located at random (with the help of two random numbers as coordinates), and the experimentally observed proportion in the sample-units can be compared with the proportion of coloured elements in the field. Mean values, standard deviations (or higher moments) can also be calculated; and the distinction between random-like and patterned fields can be briefly indicated if desired.

Another line of work can be to take measurements of natural objects of all kinds. For example, one can collect a large number of leaves from a tree, and take measurements of the length and width of each leaf. Similarly, it is possible to take 50 or 100 books from a library, and measure the length, the width and the thickness of each book. Height and weight of students, marks secured by the students in different subjects in examinations etc. provide easily available material. In this way the students can collect without difficulty and within a short time a great deal of sample measurements or sample figures of all kinds.

Experimental data obtained in this way (through the tossing of coins and dice; drawing of cards from a pack of cards or of balls from an urn containing balls of different colours; sampling from a two dimensional field; measurements of a piece of rod or other articles; measurements of the length and width of leaves from a tree, etc.) would supply abundant material for constructing frequency distributions and histograms, and the calculation and study of important statistics like the average, the standard deviation, or the coefficient of variation. When measurements relating to two or more variates (like the length and width of leaves) are available, it is possible to prepare two-way correlation charts, draw the two regression lines, and calculate the coefficient of correlation. The sampling distribution of a number of important statistics (including Studentized forms of distribution, e.g. t-statistic) in repeated sampling can be studied experimentally.

The main drive in such experimental studies should be continually to direct the attention of the students to frequency ratios and of sampling variations. For example, they should study the observed proportion of heads and tails in repeated throws of a coin and compare the results with those expected from calculations based on the binomial theorem. They can also have an experimental demonstration of the clustering of observed mean values in successive samples. Once the students have become familiar with the stability of certain observed proportions in successive sampling it is possible to define the empirical probability of occurrence of an element of an assigned type (such as of a 'head' in tossing a coin, or a coloured ball in drawing balls from an urn etc.) as the statistical limit of the observed proportion of elements with the assigned property in successive samples drawn from a population. This approach is entirely empirical, and it should be emphasized that the stability of the observed proportion (within the limits of errors of sampling) is entirely a property of the system used in the sampling experiments. Having done this, the crucial step would be to establish a link between the empirical definition of probability (based on frequency ratios) with the electivity or mathematical probability (based on the abstract theory of arrangements). It should be pointed out that this linking up is not a purely logical process but is empirical in the sense that its real justification rests on the success with which the mathematical model makes it possible to calculate the properties of subsequent samples.

At this stage it is possible to use with great advantage R. A. Fisher's illustration of the experiment with tea-cups.* He considers the problem of designing an experiment to test the claim of a lady "that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup. The experiment consists in mixing eight cups of tea, four in one way and four in the other, and presenting them to the subject for judgement in a random order. Her task is to divide the 8 cups into two sets of 4, agreeing, if possible, with the treatments received. There are 70 ways of choosing a group of 4 objects out of 8. A subject without a faculty of discrimination would in fact divide the 8 cups correctly into two sets of 4 in one trial out of 70, or, more properly, with a frequency which would approach 1 in 70 more and more nearly the more often the test is repeated". This and similar illustrations can be used to explain the notion of a 'null hypothesis', and a 'test of significance'; and to emphasize the essential need and effectiveness of randomisation. In this way it would be possible to make the students familiar with the basic ideas of the design of experiments and analysis of variance.

Once the concept of random sampling and of sampling distributions have been made familiar, the subject can be developed in various ways. The choice must be left to the teacher and the requirements of the students. If the teacher or the students are interested in applications in physics, it is possible to give a simplified treatment with easy calculations of some of the results in the kinetic theory of gases. The basic differences in the mathematical models or hypotheses leading to the three important (namely, the classical, the Bose-Einstein, and the Fermi-Dirac) distributions in statistical mechanics can be explained by simple illustrations. For example, one can imagine a situation with three students A, B and C competing for two prizes X and Y. When the two prizes are entirely different and distinguishable, and the same student may receive one or both prizes, for example, when X is a prize in cash, and Y is selection for admission to an institution of higher studies, then it is possible to distribute each prize among the three students in three different ways so that for two prizes the total number

* *The Design of Experiments*, 4th edition, 1947, p. 11.

of different ways of making the distribution is $3 \times 3 = 9$ which corresponds to the classical distribution. If the two prizes are both cash prizes of the same amount and are thus indistinguishable, but it is possible to give both prizes to the same student, the total number of distributions would be 6 which corresponds to the Bose-Einstein statistics. Finally, if the two prizes are both vacancies in an institution of higher study, the prizes are indistinguishable, and also it is not possible for the same student to have both (as admission to an institution twice over has no meaning) then the total number of distributions is only 3 which is of the Fermi-Dirac type. This is merely a simple illustration to indicate how the students can be made familiar with basic concepts which he would find useful in his later studies.

If the students and the teachers are primarily interested in biological studies, then it would be useful to give a good deal of practice in examples and calculations relating to Mendelian segregation, linkage, blood group studies, the use of the design of experiments, etc. Birth and death rates, reproductive rates, and other demographic examples or applications in intelligence and aptitude testing or psychological experiments can also be illustrated with numerical calculations.

In still another direction, experimental and numerical work in statistical quality control can be arranged without any difficulty with the help of simple appliances. For example, it is possible to set up experiments in drawing samples from urns containing balls of different colours to illustrate acceptance sampling. Measurements on a large number of match sticks, or nails, or small balls etc. would supply material for constructing simple quality control charts. The important point here would be to keep a strictly chronological record of the measurements so that the student would become familiar with the notion of homogeneous runs. If desired, some heterogeneous material can be introduced from time to time to represent variations due to assignable causes. The students can be encouraged to keep records of their own performances in class exercises, or the hour at which they get up from bed, the hours of sitting down to meals, or the time taken in walking or coming by conveyance to the school etc., as raw material for preparing quality control charts. In this way the important basic concepts can be made thoroughly familiar.

If one is interested in sample surveys, it is possible to construct two dimensional fields in various ways to represent fairly realistic models or sampling situations which occur in practice. Any good map of a rural area or a town or a city can be used to illustrate the basic ideas of 'area sampling'. A list of villages or of households can also serve as a convenient 'frame' to illustrate different types of sample design.

The list of students in the school can be used to draw random samples for measuring height and weight or for ascertaining other characteristics. The advantages of stratification can be illustrated by treating each class as a separate stratum. Finally, economic data and social statistics of all kinds can be used for numerical calculations of moving averages, index numbers etc. as well as for comparative studies to illustrate the application of statistical methods in economics.

The above suggestions are purely illustrative. Each teacher must choose his material and his methods of presentation in accordance with his resources and the requirements of his students. The wide range of statistical data and methods is itself a difficulty. Any attempt at complete coverage is out of question. It is essential to make a careful selection. Whatever be the material or topics selected for teaching purposes, the aim must be to develop the general principles and methods and to emphasize their underlying unity. If this can be done, the students would acquire, through the study of theoretical statistics, a sound foundation for scientific and technical work, a critical appreciation of economic and social issues, and a fascinating intellectual tool for the pursuit of basic problems of logic and philosophy.

Résumé

On a recommandé l'introduction d'un cours élémentaire de la statistique aux classes supérieures des écoles et à l'étage collégiale, vu qu'il y a encore une manque de statisticiens bien que la méthode statistique s'utilise aujourd'hui dans les études d'un nombre croissant des branches de la science, naturelle comme sociale.

Après avoir dilaté quelque peu sur les concepts fondamentaux de la statistique, et éclaircissant quelques définitions, le communicateur a été d'avis que la science statistique doit être développée, dans le cours élémentaire, comme une branche de la mathématique appliquée. Quant

au cours d'étude, on n'a recommandé rien de rigide, mais a rélégué beaucoup à la discrétion de l'instructeur dont le but principal doit être, à l'avis du communicateur, de familiariser les lecteurs, non seulement avec la définition pure de la probabilité et des plans d'expérience, mais aussi avec la technique de les correlater au moyen des modèles ou des hypothèses statistiques. On a fourni librement des illustrations des branches différentes de la science pure et appliquée et on a présenté des suggestions ingénieuses pour faire des plans d'expérience utilisant des éléments qui se trouvent même dans la classe.