PRESIDENTIAL ADDRESS

By

Professor R. A. FISHER, Sc.D., F.R.S.

In the original sense of the word, 'Statistics' was the science of Statecraft ; to the political arithmetician of the eighteenth century, its function was to be the eyes and ears of the central government. It could tell the Prince how many able-bodied men he might mobilise, how many would be needed for the essentials of civil life ; how numerous or how wealthy, were sectarian minorities who might resent some contemplated change in the laws of property, or of marriage ; what was the taxable capacity of a province, his own, or of his neighbours.

This aspect of statistical information seems to have been long neglected in England, and we should have to look as far as Italy for a centralized statistical service capable of ensuring that Il Duce is the best informed potentate in Europe. It is certain that the power which such knowledge gives, like the power of lethal weapons, may be abused. On the other hand it *is* in the public interest that statesmen responsible for political decisions should be able, so clearly as is possible to foresee the consequences of their public actions.

In democratic countries the emphasis has long fallen upon the important task of providing public information. This function of official statistics is of value both for the aid it gives to private enterprise to use its initiative wisely, and for allowing public opinion an opportunity, at least, of acquiring such a knowledge of quantitative facts as will enable it to pursue aims which are not only worthy, or desirable, but are also within the bounds of practical possibility. It enables public opinion, *if it will*, to size up its own problems. The Socratic dictum "know thyself" is applicable even more to peoples than to individuals

Statistical work is also largely employed as a means to internal efficiency in each department of Government. I need not stress this aspect of official statistics, which is not I believe commonly neglected ; save to point out that here also the official service requires a supply of men, who, if they are to be competent, should have at least some specialised statistical training.

More recently, the great centres of scientific research have discovered that statistical science has a vital part to play in all quantitative problems. I will later give some account of this development, which has revolutionized the technical methods of statistics, and has greatly extended the public functions it can fulfil. Heads of research institutions in all countries are feeling the advantage of having as right-hand men statisticians technically capable of drawing up a research programme or of applying to a projected research a rational and objective criticism.

I have formed an opinion, which, be it right or wrong, I should like you to bear in mind throughout this address, that statistics in England has suffered severely from the wide separation, due to our long political history, which has grown up between official and academic statistics ; or, to speak functionally, between the duties of collection, enumeration, tabulation and publication, which absorb the time of official statisticians, and the duty of study, analysis and interpretation which falls to the lot of the mathematical or theoretical statisticians. A body such as the Royal Statistical Society does something to bridge this gap.

How wide the gap may seem may be illustrated by some witty comments made by A. L. Bowley in criticism of the presentation of official data. He was moved to draft an ideal footnote, which might find a place in any government publication. I cannot recall his exact words but the following is an attempt to recapture the spirit of his criticism.

Footnote:—(*a*) The terms used in the headings and margins of the table are all employed in a technical sense, known only to officers who compiled it, and which they are unable for official reasons to divulge.

(*b*) The sub-divisions of the table and the region to which it refers have been changed since the last return was published.

(*c*) Before tabulation the data have been subjected to numerous adjustments, allowances and other corrections, of a kind to vitiate any tests of significance which the reader may be tempted to apply to them.

The academic mind, as we know, is sometimes capable of assuming an aggressive attitude. The official mind, on the contrary, is and has to be, expert in the art of self-defence. If half of Bowley's indictment is true, it will always be easy, when analytic methods are rashly applied to official material, to show how frequently the material falsifies the assumptions made in the course of analysis ; to sneer at the elaboration of the calculations employed, or at the refinements which have been introduced, always without emphasizing the obvious fact that the more imperfectly the work of compilation has been done the more elaborate and difficult must be the task of extracting objective information. My point here is that in developing her statistical services India might learn from the difficulties which England had encountered, and somehow contrive neither to allow official statisticians to be blinded by ignorance of method, nor to allow academic statisticians to be sterilized by lack of responsible experience.

So recently as twenty years ago the task of the statistician, faced with an accumulated mass of observational data was understood to consist in the calculations of certain average values ; by that time also it had become usual to employ the quantities of the second degree, squares and products, to elicit from the data what are known as "probable errors," regarded as appropriate to the average values obtained. In putting the matter in this way I mean a wide interpretation to be given to the term average for these are of many different kinds and any large body of data will yield an immense variety of them. Which to choose must depend partly on what sort of information it is desired to obtain, and this part is not properly a statistical question. But expert statistical knowledge was recognized to be necessary for a comprehension of the arithmetical processes sufficient to ensure that they were appropriate to the meanings to be placed on them. This narrow measure of responsibility was however enlarged by the fact that all extensive bodies of data are liable to contain information on points which were not in view when they were collected ; to recognize such information, and to find the means of eliciting it, was most stimulating part of the statisticians' task.

As an illustration I may point to the widespread and important type of data known as time-series, such as are provided by annual figures of births, deaths, exports, prices and the whole range of vital and commercial statistics, the proper treatment of which has been a classical problem to all who are interested in national or sociological questions. Data of the same kind are amassed by meteorologists, and arise when agricultural experimentation is carried out continuously on the same land. In studying such data as this at Rothamsted it was found that a series of averages, using weights determined by position in the sequence, would provide, *first*, the linear trend representing the greater part of soil deterioration, *next*, a few terms representing other slow changes ascribable to changes in

variety, cultural practice and intensity of weed infestation, and *finally*, a residuum of more rapid fluctuations dominated by the variation of the weather from season to season. Moreover, the separation effected in this was such as to allow the year to year fluctuation to be studied in relation to meteorological causes, just as if the series were not affected at all by the large slow changes actually present.

The classical theory of errors due to the great German mathematician Gauss had shown how the interpretation of such estimates could be decided by taking account of the whole body of discrepancies. The residue of unexplained variation provide, in fact, the material for the calculation of probable errors, or in more modern usage, of standard errors. In 1908 the late W. S. Gossett, whose untimely death last year still casts a shadow on our discussions, wrote under the pseudonym of "Student" a paper, little appreciated at the time, which opened the door to a refinement of exactitude of which the theory of errors had hitherto seemed incapable. The concept of standard error is in fact insufficiently exact for application without reserve to the small samples of observations which the experimenter can command. Nevertheless, it was shown possible in an important range of cases to develop rigorously exact tests of significance from which the concept of standard error may be entirely eliminated. That the importance of Student's suggestion was not at once recognized illustrates all too clearly how ignorant academic statisticians were of the important decisions which experimental scientists, like statesmen, must take, always on the basis of a limited knowledge of fact. Biological research, however, throughout the world bristled with problems to which only exact tests adapted to a very wide variety of logical situations that had come into existence could be applied. In a few years more they were in general use among experimentalists. The period which followed has shown the somewhat ludicrous spectacle of entomologists, foresters, plant physiologists and others with no trace of mathematical pretensions, applying freely and with understanding in their daily work mathematical refinements, which most official statisticians could not understand, and which too many teachers in mathematical departments were unable to expound.

The development of exact test of significance had been based on the mathematical solutions of a number of what are known as *problems of distribution*. The solutions found were not very difficult mathematically. Certainly they were not beyond the power of the great series of mathematicians who, from the time of Laplace, have given their minds to the theory of probability. Had these minds been put into direct contact with the problems which beset the Laboratory worker we cannot doubt that they would have been solved by them some sixty years earlier. I want to insist on the important moral that responsibility for the teaching of statistical methods in our universities must be entrusted, certainly to highly trained mathematicians, but only to such mathematicians as have had sufficiently prolonged experience of practical research, and of responsibility for drawing conclusions from actual data, upon which practical action is to be taken. Mathematical acuteness alone is not enough. My revered teacher Prof. Whitehead of Cambridge used to say in one of his courses: "The essence of applied mathematics is to know what to ignore." And when I read current publications in mathematical statistics I am continually and forcibly reminded of the wisdom of this remark.

The solution of problems of distribution, although stimulated by the practical needs of chemists and biologists, had at once important mathematical consequences. For, if the statistician asks himself, as he ought, "What is my best method of averaging or combining the data before me," so as to elicit information on some definite issue, it is evident that he must be guided by magnitude, and the nature of the sampling errors to which different estimates are liable. This, in a few words, is the genesis of the theory of estimation. In

this theory we discuss what advantages each of innumerable possible estimates possesses, and what procedure will enable us to combine these advantages in our estimates. I shall hope to develop the theory more fully in a series of lectures at the University of Calcutta next week. For the moment, I want to note only the consequences of the progress so far made in this field as they affect the future of statistical research, and the position and responsibilities of the statistician.

A few years ago no statistician who had completed his work on an extensive mass of data could feel any rational assurance that some other statistician, using the same data, could not by some new method calculate estimates much more precise than his own. In these circumstances it was natural that if his results were actually rather precise he should display them with some satisfaction, if not he might in his chagrin even suppress them. High correlations and strongly significant comparisons were displayed with considerable pride, for they seemed to indicate not merely that the data were good, but that they had been brilliantly handled. But now, observe the change. So soon as any young graduate has mastered the theory, he can lay down the equations for the most efficient estimates possible. The whole tone of the subject has been altered. The statistician is no longer an alchemist expected to produce gold from any worthless material offered him. He is more like a chemist capable of assaying exactly how much of value it contains, and capable also of extracting this amount, and no more. In these circumstances, it would be foolish to commend a statistician because his results are precise, or to reprove because they are not. If he is competent in his craft, the value of the result follows solely from the value of the material given him. It contains so much information and no more. His job is only to produce what it contains.

The mathematical measurement of quantity of information has another and more fruitful consequence. If we want more than has been found to be present we are no longer tempted to harass the statistician to work miracles. On the contrary, since he can evaluate the material, we may require him to explain what was wrong with the data that more did not come of it. What supplementary information which could also have been collected would have been worth collecting? At what points could the material be most profitably amplified? In what respects also has the labour of collection been unprofitable? So that our resources may be employed in future where the yield is highest. Get the statistician to redesign the enquiry, and to justify his new plan by the methods of costings accountancy. Here the door is opened on work very well worth doing indeed. Immensely laborious calculations on inferior data may increase the yield from 95 to 100 per cent. A gain of 5 per cent. of perhaps a small total. A competent overhauling of the process of collection, or of the experimental design, may often increase the yield ten or twelve fold, for the same cost in time and labour. To consult the statistician after an experiment is finished is often merely to ask him to conduct a *post mortem* examination. He can perhaps say what the experiment died of. To utilise this kind of experience he must be induced to use his imagination, and to foresee in advance the difficulties and uncertainties with which, if they are not foreseen, his investigations will be beset.

I have tried to show that India, in the construction of her national organization, has a great opportunity to utilise the newly discovered possibilities of statistical science. This Conference is itself the strongest evidence that the will to seize this opportunity is not lacking. When I see too the widespread interest developing in this country in all branches of statistical work, and above all the brilliant school of workers that Prof. Mahalanobis has gathered round him in this University, I suggest that we can put every confidence that the work will be carried forward with the intellectual integrity that such a task requires.