Theory and Practice in Statistics

[The Address of the PRESIDENT, Dr F. YATES, C.B.E., F.R.S., delivered to the ROYAL STATISTICAL SOCIETY on Wednesday, June 19th, 1968]

INTRODUCTION

IT is a privilege and pleasure to address you this evening as your President. To me the occasion is particularly apposite as it is almost fifty years since R. A. Fisher began his professional scientific career with his appointment as statistician at Rothamsted. I in my turn began statistics under him in 1931. This I regard as one of the most fortunate chances of my life. It also marked the beginning of a lifelong friendship.

It is a tradition of our Society to muzzle our Presidents—at least, I understand this to be so, for when I intimated that I might wish to contribute to the recent discussion on regression methods I was told that as I would be in the Chair this would raise great difficulty. However, in his Presidential Address the President has his one opportunity of unburdening himself, and moreover, also by tradition, is protected from the cut and thrust of subsequent discussion. But do not think, if some of you find some of the ensuing remarks not to your liking, that I wish to shelter behind this immunity. I very much hope that my address will stimulate further discussion at future meetings, and that those who disagree will come forward boldly and say so. For it is by argument and discussion that we arrive at the truth, and at the best course for future action.

Two years ago I had the honour to be invited to give the first Fisher Memorial Lecture, and chose as my title "Computers: the second revolution in statistics". I do not intend to say much about computers this evening, but I am convinced that because of them we are faced with a revolutionary situation, and it therefore behoves us to take a long, hard critical look at the present state of our subject, and to ask ourselves whether statisticians are fulfilling their functions in the community, and to what extent current lines of activity are likely to meet future needs. These questions, I feel, should particularly be the concern of our Society, and I therefore have no hesitation in devoting my Presidential Address to them.

I find myself, however, in some slight difficulty. For earlier this year, Dr M. G. Kendall, whose outstanding achievements have just been recognized by the Society by the award of a Gold Medal, presented a paper "On the future of statistics—a second look". This covered much of the ground that I had hoped to cover; indeed, when I heard it I said to myself: "There goes your Presidential Address." Consequently, instead of discussing broad generalities, which have been ably dealt with by Dr Kendall, I propose to examine in a little more detail how far promise in our subject is matched by performance, and in so far as performance is found wanting, to see how we should endeavour to improve matters.

1. CROSS-FERTILIZATION BETWEEN THEORY AND PRACTICE

May I emphasize at the outset that in my view the theory underlying statistical methodology is a branch of *applied* mathematics. Some parts of it involve advanced and subtle mathematics, others little more than elementary algebra and a sense of structure. If our subject is to develop rapidly and fruitfully, in a manner that will effectively satisfy the growing demands on it, it is important that the mathematicians engaged in it should have constant contact with reality by being concerned also with practical applications. Not only will this suggest to them what problems are really worth tackling, it will serve as a constant corrective to lines of development, which though theoretically correct on the premises on which they started, do not correspond sufficiently closely to reality to be useful.

Certainly in the two branches of statistics with which I am most familiar, namely the design and analysis of experiments, and the design and analysis of surveys, both of which are now well developed, and both of great practical importance, the cross-fertilization of ideas between theory and practice was very great.

The basic principles of modern experimental design are rightly associated with the name of Fisher, but brilliant theoretician though he was, he did not evolve them by a process of pure mathematical reasoning. They were evolved over a period of years, during which he was engaged, in response to requests from biological and agricultural research workers at Rothamsted, in designing and interpreting actual experiments, both in the field and in the laboratory. In fact, the development of methods of analysis came first and design followed, as the history of the subject shows (Yates, 1964).

Nor should the contributions of practical workers who were not mathematicians be overlooked. Replication, arrangement in blocks or Latin squares and factorial design had long been practised in field experiments. Estimation of error from high-order interactions and fractional replication were later evolved in response to pressure from practical experimenters to get more out of their experiments. What had defeated practical workers was the correct estimation of error, but even here Gosset ("Student") had arrived at the correct solution for experiments arranged in blocks, though it was Fisher who showed him how to simplify the arithmetic ("Student", 1923).

What Fisher did was to provide a sound theoretical basis for this rather confused body of practice, introducing the principle of randomization to ensure validity in the estimates of error, formalizing the arithmetic by the analysis of variance, and providing the z-test of significance. He also contributed the idea of confounding, which proved to be of great practical value, though here again it was erroneous use of confounding in actual experiments that directed attention to the need for orthogonality in the corresponding analysis of variance (Yates, 1933).

The contributions of practical workers, sociologists and others, to survey design were even more striking. All the basic ideas, random sampling, stratification, use of a variable sampling fraction, multi-stage and multi-phase sampling were used in practice before they had been codified by mathematicians. What did defeat the non-mathematicians, as in experiment analysis, was the proper estimation of sampling errors. Bowley, for example (in 1930, I think), observed: "It is impossible, or at least very difficult, to calculate the sampling errors of a stratified sample."

2. Irrelevance of Much Present-day Statistical Literature

The two branches of statistics just mentioned demonstrate the need for the unifying hand of the mathematician. Only when the theory is well understood, and the consequences of the complexities that arise in practical applications explored, can the methods be safely and widely used. But once this stage is reached, the role of the mathematician *qua* mathematician becomes secondary, and the main task of mathematically trained statisticians engaged in practical applications of the methods is to understand them thoroughly and see that they are correctly and imaginatively applied. Only when radically new situations arise are extensions of theory necessary.

Unfortunately, mathematical statisticians, especially those who teach in universities, are expected from time to time to publish new mathematical contributions to their subject, and their competence tends to be judged by the number of their theoretical papers, rather than by the excellence of their teaching, or the usefulness of their practical statistical work. The process is self-generating. A publishes a paper which is irrelevant, or based on unrealistic premises. This stimulates B and C to further thoughts on the same lines. And in a few years a body of literature is built up.

This in itself is bad enough. It results in our statistical journals being cluttered up by vast amounts of rubbish. But it has worse effects, because some at least of the writers of the papers are sufficiently intrigued by them to include the topic in their courses, and set Ph.D. students to work on further investigations on the same lines.

In case you think I exaggerate, let me give you an example, that of "fixed" and "random" effects in experimental design. This dichotomy has now a considerable vogue in America, and is spreading to this country. When I first came across it, I concluded that it was irrelevant and misguided, but only for the Fifth Berkeley Symposium did I muster enough time and energy to go into it thoroughly, and set out my views in print (Yates, 1967). A few brief extracts from the paper will suffice here. For the arguments on which these conclusions are based, the original paper may be consulted.

> Be that as it may, the hare, once started, could not be stopped. Differences in formulae, arising from differences in definition, soon intruded, and before long it was represented that the tests of significance which could be correctly applied would differ for the two models. (For later developments see a review by Plackett (1960); the discussion on this paper is also worth reading.)

> A paper by Harter (1961) on the analysis of split plot designs exemplifies the extreme state of confusion that can arise from these differences of definition and from treating replicates as an additional factor R more or less on a par with the treatment factors.

> If these [Harter's] were indeed the appropriate tests it might, as he states, be a "crucial" question whether or not S_{BR} and S_{ABR} should be pooled to make up subplot error. Far from being crucial, during the many years I have been responsible for the analysis of split plot experiments, I have never considered the two components worth separation.

> Incidentally, it may be noted that Harter's arguments have really little to do with split plots. The same partition of error d.f. can be made in an ordinary $A \times B$ factorial experiment, and the same confused situation would arise if his arguments were valid.

As a further example of misleading triviata I commend to your attention a paper on experiment analysis entitled "Some aspects of the statistical analysis of the 'mixed' model", which recently occupied 22 pages of *Biometrics* (Koch and Sen, 1968). In it the authors, apparently not satisfied with the ordinary analysis of variance procedure, develop several further parametric and non-parametric tests for two-way tables. They illustrate these on two examples, both 4×8 tables, for which they also give the ordinary 4×8 analysis of variance, not recognizing that in neither example is this form of analysis strictly appropriate, the first example being 8 replicates of a single measurement of each of 4 variates, the second having a "split plot" 2×2 structure of treatments. That they think their alternative tests are worthy of general application is evinced by the fact that not only have they programmed them on a computer, but also offer to make the program generally available!

This sort of thing is not confined to the more complex parts of methodology. New solutions are proposed for the simplest problems, such as the estimation, from a single observed value x, of the mean μ of a normal population with known variance. Healy and I (1964) drew attention to a novel confidence interval solution of this problem which appeared in the *Journal of the American Statistical Association* and found its way into an examination paper for the Cambridge Diploma of Mathematical Statistics; and on which we forthrightly commented: "This last is the sort of thing that is bringing the whole science of statistics into disrepute." Another, equally remarkable, recently appeared in Series B of our *Journal* (Clutton-Brock, 1965). In it the author postulates a prior "bell-shaped" distribution of μ about zero with unknown σ. He states, incorrectly, that "the classical approach to this situation is to apply a significance test", taking $\hat{\mu} = 0$ if x is not significantly different from 0, and $\hat{\mu} = x$ otherwise. Ultimately, in 11 pages of the *Journal*, he recommends an estimate $\hat{\mu} = \gamma x$, where $\gamma = 0$ if $x^2 \leqslant 1$ and (in some circumstances at least) $= 1 - x^{-2}$ when $x^2 > 1$.

3. MATHEMATICS AND THEORIES OF PROBABILITY AND INFERENCE

The two examples quoted above both stem from the misuse of mathematical reasoning, primarily because the practical problems to the solution of which they are supposed to contribute are incorrectly defined. In basic statistical theory—probability, inference, decision making—the confusion is much more serious, but it stems, I believe, from much the same cause. Take Bayesianism. There are some situations in which a well-defined and estimable prior distribution exists, and for which a Bayesian solution is therefore appropriate. Thus, for example when estimating the yield of a crop for a region, year by year, from reasonably accurate acreage figures and inexact sampling data on average yield per acre, we may well, from a run of years, or even from independent statistics such as market returns, be able to determine with fair certainty the distribution over the years of the average yield per acre. High sample estimates may then be sensibly decreased, and low ones increased; but the statistician must be on the lookout for a time trend in the mean of the distribution, such as might arise from improved agricultural practice, and for this the unadjusted yields per acre should be used. But to use the Bayesian approach for problems in which there is no such underlying distribution, or in which the knowledge of the distribution is vague and indefinite, is forcing the problem into a mathematical mould it does not fit. In such circumstances, if in the light of other knowledge an estimate seems too extreme, it is much better to report the undoctored value and add a warning, with a suggested modified value if this is considered necessary.

I am not alone in feeling disquiet at the way the more fundamental aspects of our subject are being treated. Fisher, whose own contributions to the subject have been sadly neglected since his death, was alarmed at the lack of contact with reality. A passage from a lecture delivered at Michigan State University in 1958, which I have quoted elsewhere, but which is worth repeating, admirably sums up his views:

> Of course, there is quite a lot of continental influence in favour of regarding probability theory as a self-supporting branch of mathematics, and treating it in the traditionally abstract and, I think, fruitless way . . .. Perhaps we were lucky in England in having the whole mass of fallacious rubbish put out of sight until we had time to think about probability in concrete terms and in relation, above all, to the purposes for which we wanted the idea in the natural sciences. I am quite sure it is only personal contact with the business of the improvement of natural knowledge in the natural sciences that is capable to keep straight the thought of mathematically-minded people who have to grope their way through the complex entanglements of error, with which at present they are very much surrounded. I think it's worse in this country [the U.S.A.] than in most, though I may be wrong. Certainly there is grave confusion of thought. We are quite in danger of sending highly trained and highly intelligent young men out into the world with tables of erroneous numbers under their arms, and with a dense fog in the place where their brains ought to be . . ..

Tukey (1963) expressed similar alarm:

> The moral of all these remarks is quite simple, and . . . quite clear and firmly binding. It is that we must expect mathematical statistics to become more and more logically complex and more and more visibly pathological; that we must press forward along the paths of its development; that we must not let these complexities and pathologies interfere with what is to be done in practice; and that we must use the results of mathematical statistics as guidance —not as a code of law.
>
> Who can believe that the fundamental process of learning about the world—the analysis of experience—can rest upon theorems unflavoured by wisdom? Only those, I submit, who have either not yet come into contact with the real world, or who have found themselves wholly unable to meet the challenge of the real world's uncertainties.

My only comment on this is that I can see no reason why we should tolerate the theoretical branches of the subject becoming pathological.

4. Effect on Practice

How does all this confusion reflect on the practice of our subject? I think there are two main baleful influences. First, it creates a yawning gap between theoretical and practical statisticians. Second, it distorts teaching, so that many of our ablest new recruits start their statistical careers without any clear idea of how to apply important and quite elementary methods. Many of them become teachers in their turn, and have to teach not only budding professional statisticians, but students in other subjects who wish to acquire a working knowledge of the more elementary parts of statistics.

This last category I regard as of great importance. For much day-to-day practical statistics is, and always will be, done by workers who are not professional statisticians —and rightly so. But it is most important that they should be able to avoid elementary blunders, know when to consult a professional statistician, and know enough of the subject to be able to communicate with him intelligently and not be "blinded with science".

In consequence of all this, the standard of much day-to-day statistical work is regrettably low. All too frequently there are elementary errors in the presentation and analysis of data by biologists, economists and social scientists, which lead to completely erroneous conclusions. In spite of the great increase in the numbers of professional statisticians in universities, research institutes, government deparments and industry, there seems to have been little improvement in recent years. This in itself is an indictment of the training we give to our statisticians and to students of other subjects who take ancillary courses in statistics, and of our statistical textbooks. These matters, as I see it, are primarily the responsibility of the university statistical departments and are a direct consequence of the present-day obsession with advanced theory, largely divorced from practice.

A few examples, encountered in the course of my own work, will indicate the sort of thing I have in mind.

4.1. *Confusion between Association and Causation*

A well-known economist, asked to estimate the reduction in crop yields on the continent of Europe that might be expected from cutting off phosphate supplies by the invasion of North Africa, plotted the average wheat yields per acre of the various European countries against the phosphate consumption per acre. He obtained a beautiful graph, which showed that those countries where little phosphate was used (those of Eastern Europe) had much smaller yields than those countries where much was used, and deduced that there would be a dramatic fall in yield. This, of course, ignores that (a) phosphate is merely one of a closely associated complex of factors, soil, climate, other fertilizers, general standard of agricultural practice, etc. that together affect the average yield of a country, (b) stored reserves of phosphate in the soil will prevent any sudden fall in yield if phosphate is withheld. Fortunately we at Rothamsted had already made a very thorough study of fertilizer responses in the course of evolving a scientific wartime fertilizer policy for this country (Crowther and Yates, 1941) and were able to predict with confidence that the immediate loss of yield of cereals would be trivial. Fortunately also we were shown the economist's memorandum before it received any wide circulation.

4.2. *Use of Inappropriate Measures*

To assess the effects on industrial production of bombing raids on German cities, the percentage of buildings destroyed in the various towns was assessed (at great labour) from aerial photographs. This was a largely meaningless statistic, as by and large it was the centres of the towns that were destroyed and factories are on the outskirts. Instead I was instrumental in getting a similar statistic constructed for factory buildings only. We already had, from detailed surveys in this country of losses of production of bombed factories and area of damage, a measure for converting this latter statistic into loss of production. The estimated loss turned out to be quite small. This was a crucial piece of information for the reconsideration of bombing policy. Investigations in Germany after the war confirmed our pessimism, and showed that the Germans had in fact succeeded in increasing their production.

4.3. *Reckless use of Regression Analysis*

The following example, reported by Baker (1965), illustrates the misuse of multiple regression analysis in a way that is only possible now that computers are available. There were two experiments, treated similarly. The first "was designed primarily to test the influence of air temperature on the net assimilation rate [roughly photosynthesis] of cotton", grown in an environment controlled for air temperature and CO_2 concentration, but with natural illumination. In this experiment CO_2 concentration was held constant and a temperature of 20°, 25°, 30°, 35° or 40°C was selected at random for each day, being held constant during the day.

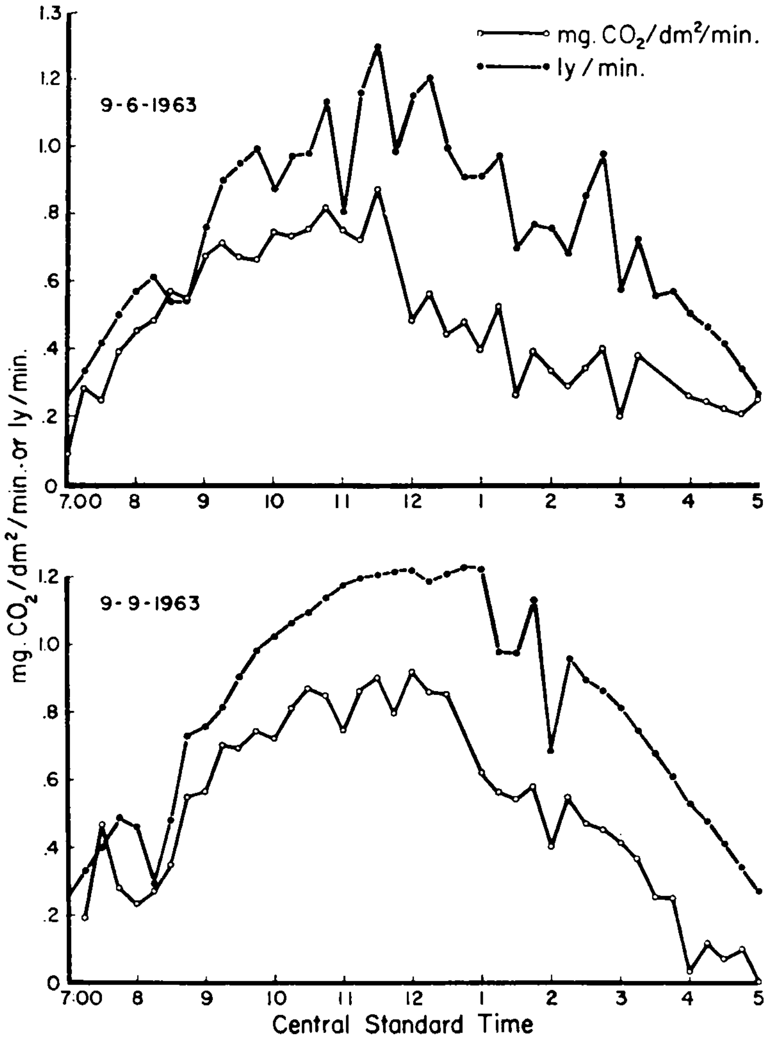

FIG. 1. Time course of apparent photosynthesis (mg CO_2 dm^2 ground area/min.) and light intensity (ly/min.).

Values of the apparent photosynthesis (y), light intensity (x_1), and vapour pressure deficit between an illuminated leaf and the air (x_2) were determined for each 15-minute interval. Fig. 1, reproduced from the paper, shows "two days' typical data" for y and x_1.

The data for all days were then thrown together and a multiple regression analysis was made of y on x_1, x_2, time of day (x_3) and air temperature (x_4), including linear and quadratic terms for each dependent variable and all possible linear interactions (product terms). The program used first computed the full regression and then eliminated successively the least significant term of those remaining. The reported result was the equation

$$Y = a + b_8 x_4^2 + b_{10} x_1 x_3 + b_{11} x_1 x_4 + b_{15}^* x_1 x_2 x_3 + b_{17} x_1 x_3 x_4,$$

where the * indicates the next least significant term (dropping which gave the "first significant reduction in goodness of fit"). Perhaps wisely, the author did not give the numerical values of the coefficients. Instead he gave graphs illustrating the relation of Y to x_1, x_2, x_3 separately at various chosen values of the other three variates.

Those familiar with multiple regression will doubtless be horrified at this maltreatment of the data. But is the author to blame? He is a plant physiologist, not a statistician. He was merely using a ready-made all-purpose statistical tool provided by the computer. The biggest crime, of course, is the failure to recognize that there is a hierarchy of terms in his original equation; consequently, in an empirical model of this kind, if a quadratic term is included the corresponding linear term should also be included, and if an interaction is included the linear terms for the separate factors should also be included. Had he adopted an exploratory approach, and worked forwards, not backwards, he might well have found that few if any of the quadratic or interactions terms were significant, or if formally significant, of any consequence. As temperature was held constant for the whole of each day he should also, I think, have separated the data and calculated regressions of y on x_1, x_2 and x_3 for each temperature separately; the differences between the coefficients of these regressions would then have revealed the effects of temperature on photosynthetic activity. A plot of the residuals after fitting linear terms only might also be revealing, particularly as the customary tests of significance are always suspect in time series data of this kind.

I will not comment here on the design of these experiments, beyond saying that with such elaborate apparatus something more informative might have been attempted.

Failure to adjust for disproportionate frequencies or empty cells in the analysis of tabular material

This is a very common fault. The following example is from the *Royal Society Report* (1963) on the emigration of scientists (the "brain drain"). A survey was made of Ph.D. candidates who were awarded their degrees in British universities in the years 1952–61, and the year of emigration of those who emigrated in each of these years was recorded. The data were then tabulated in a two-way table, year of award × year of emigration. This table is roughly triangular, because, apart from a few individuals awarded their degrees after emigration, entries referring to years of emigration preceding the year of award will be zero. But to measure the trend in emigration the marginal totals for each year of emigration were expressed as a percentage of the number of degrees awarded in that year. This greatly exaggerates the percentage increase over the years. A measure of the trend based on the numbers emigrating within five years of their degree (with estimates of future emigration for those awarded their degrees in 1958–61), gave an increase of 44 per cent over the 10 years, compared with the increase of 122 per cent given in the report. Over the years 1952–57 a similar measure (for which no estimates are required) gave a percentage increase of 41 per cent, compared with 158 per cent by the method of the report.

Confusion on principal components

A material has easily measured properties x_1, x_2; these are related to a further property y, which is a measure of its suitability, but which can be determined only by elaborate tests. In a search for something more fundamental than regression, the first principal component of x_1 and x_2 was proposed (by a well-qualified professional statistician) as a "general measure" of the property y! This is obvious nonsense, as can be seen from the fact that the same general measure will be obtained whatever the property y.

5. COMPUTERS

Many statisticians have been curiously reluctant to exploit computers. This attitude is now changing—it would indeed be surprising if it were not—but there is still a tendency to expect others to do the hard thinking necessary to develop techniques for their effective use in statistics. Programming is regarded either as an esoteric art or as a mundane task beneath the dignity of the statistician. It is still common, in papers on problems the solution of which depends on the use of a computer, for there to be no discussion of computer techniques and a mere acknowledgement at the end to someone who did the programming.

I am convinced that, whether we like it or not, almost all statistical computations of any magnitude will shortly be done on computers, and that the statistical methods used will in large part be conditioned by the programs that are readily available. It behoves us, therefore, to concern ourselves much more seriously than we have done hitherto with developing programs and programming systems that will enable sound statistical methods to be effectively and critically applied, not only by professional statisticians, but also by those trained in other disciplines who use statistics in the course of their day-to-day work. This is not by any means easy, as we have found by experience at Rothamsted, but it is an interesting and rewarding task, demanding knowledge of both the theory and practice of statistics and of the capabilities and limitations of computers. I am very glad that the Society has now given prominence to this activity, by publishing, in Series C (1967), a complete report of the proceedings of a recent conference held at the Atlas Computer Laboratory, Harwell, organized by Mr J. A. Nelder, who succeeds me as Head of the Statistics Department at Rothamsted, and Mr B. E. Cooper; and by initiating the publication of computer algorithms.

One of the troubles with computers is that they are developing so rapidly that requirements are in a constant state of flux. We are now on the eve of multi-access machines, with a console, if not on every desk, at least on many desks, and answers, if the central library contains an appropriate program, while you wait. This demands programming systems very different from those suitable for batch processing.

It is generally recognized that the programs available for routine statistical work on many machines are unsatisfactory. Many seem to have been written by programmers and systems analysts with only a smattering of statistics acquired from some outdated textbook. It is up to us, the professional statisticians, to devote part of our energies to remedying this state of affairs. Only we can see to it that the statistical content is satisfactory.

I discussed the subject at length in my Fisher Memorial Lecture. Here I will only make the following points:

1. There should be co-operation between different teams of statisticians in the construction of general programs, so that each team covers a branch of statistics in which it is really expert.

2. Programs should be designed so they can be made available on different computers, and every effort should be made to get them working on the computers at our universities and technical colleges. This must be primarily the job of the statistical departments of the universities and colleges, rather than the computer departments. I hope that manufacturers may also be persuaded to improve and extend the "statistical packages" they are beginning to provide as part of their software.

3. Students must be taught how to use general programs for the solution of their problems, as well as learning how to write *ad hoc* programs. This can be done only if there is a good statistical program library on their univeristy or college computer.

4. We should avoid striving for perfection or excessive generality. Something useful this year is better than something more nearly perfect in two years' time.

On one thing let us be clear. If we do not tackle these jobs, others will. Already there is a tendency to set up "data processing" as an activity independent of statistics, whereas much of it should form an integral part of statistical methodology.

6. What should be the Aims of our Society for the Future?

Dr Kendall, at the beginning of his paper on the future of statistics, expressed concern that the subject was "at present in danger of losing its identity", and listed various subjects such as operations research, actuarial science, econometrics, all of which have a large statistical content, but which are developing independently. But this is no new thing. Different branches of statistics have in the past developed for long periods of time in more or less water-tight compartments, and only much later has the common ground been recognized. Sir Ronald Fisher started his Presidential Address to this Society in 1952 by pointing out the independent development of probability theory, actuarial science and the theory of errors. His remarks on Gauss's development of the theory of errors are worth repeating:

> The Theory of Errors sprang with equal independence from the practical need to grapple with the problems, at first sight fantastically difficult, confronting astronomers and surveyors, in the combination of observations. The extent to which Statistical Science was broadened and enlarged by this development has not been appreciated by writers on the Theory of Probability. Perhaps only a genius of the magnitude of Gauss could have introduced and consolidated so many new ideas. Unknowns, other than probabilities, were seen to require estimation. Observations had to be discussed that were not merely counts of frequencies, but quantitative measurements on a conceptual continuum. The normal distribution was introduced, not as the necessary consequence of a deductive mathematical argument, but by a choice resting on penetrating understanding of its essential properties. It was amid such extensions, and as part of a much larger whole, that the Theory of Probability began to acquire a genuine importance for human thought.

What I think worried Dr Kendall was the tendency for the separate branches of activity to develop their own modes of thought and methodology, often supported by specialized journals and societies. Specialization we must accept. It is an inevitable consequence of the rapid expansion of research activities. Nevertheless, there is a common thread running through the practice of statistics. In any investigation that involves quantitative phenomena having any degree of complexity, there is need for an expert in the handling of numerical material and interpreting it correctly and in a manner relevant to the objectives in view. This expertise is the central core of what we mean by statistics. But because data must be collected, by experiment or observation, or past data must be used, and because the conclusions reached must be communicated to others if they are to form a basis for action, a statistician's mandate extends in both directions; at one end to the planning of experiments, etc., at the other to policy decisions. To be effective in this wider role, the statisticians must be not only experts in their own subject, but must also acquaint themselves with the technicalities of the subject in which they are working. And, like all experts, they must be in constant communication with the other members of the team, must in fact consider themselves as members of the team.

They must also be in constant touch with statistical developments in other branches of research, for such developments may well be relevant to their own special branch. Thus the techniques of design and analyses of experiments, first developed for agricultural field trials, proved of great use in biology, medicine and industry; the estimation of sampling errors for samples taken from plots of agricultural field trials clarified thought on the more complex problems of estimating error in surveys. Professional statisticians can do much to further this cross-fertilisation of ideas, and to eliminate barbarous and outmoded methods which tend to develop and persist in branches of research which are isolated from the main stream.

I believe that our Society can play a leading part in preserving the unity of our subject, and in preventing divergence of practice where no such divergence is required, but only if we regard the integration of theory and practice as one of our main tasks, and devote more of our attention to the development and dissemination of methods appropriate to real-life problems, and less to spinning webs of theoretical fantasy that have little relation to reality.

In many ways our Society is well placed to influence developments in this country. But, for historical reasons, we have so far not made much effort to unify our various activities. Our Journals bear evidence of this. Series A is primarily devoted to economic and sociological applications, but contains some of the more general papers on methodological aspects. Series B, which originally started as a medium for the publication of strictly practical papers on the newer methods, has become more and more theoretical, and now threatens to go the way of the *Annals of Mathematical Statistics*. Its original function has now been passed to *Applied Statistics* (Series C). This is the slimmest, but in many ways the liveliest and most useful of our Journals.

There has for some time been dissatisfaction with our journal structure and proposals for alteration are at present under consideration, but these mainly affect Series A and C. I think we should also consider whether Series B, in its present form, is performing its proper function, and whether it could not be greatly improved both by a more rigorous selection of papers and by making clearer the practical relevance of those published. I suspect, also, that much of the lengthy mathematical exposition in many of the papers could be substantially cut with benefit to the readers. Certainly I should like to see more papers on the problems that face us in making good use of computers. Some of the more theoretical and general of these might find a home in Series B.

Meetings are another activity directly under our control. Here again I am doubtful whether some of the more theoretical papers are suitable for presentation at meetings; equally whether some of the papers on economic and sociological subjects are not more suitable for more specialized journals. The recent decision, arising from the difficulty of securing enough papers for the General Meetings of the Society, to have fewer of these meetings, seems to me a move in the right direction. Except for papers on economic and sociological subjects, however, papers on applications are still not accorded the prestige I think they deserve.

In the matter of education we have a less direct part to play. The universities rightly insist on preserving freedom in what they shall teach. But the Society can perform a most useful function in furthering exchange of views on this most important aspect of our subject, both in discussion meetings and in the publication of papers on the topic. At a less exalted level we may have an even more important part to play. For the last three years I have been a member of the Mathematics Board of the Council for National Academic Awards, and have had the task of examining statistical syllabuses put forward by colleges of advanced technology. I will not comment on these here beyond saying that in my opinion many could be greatly improved. The Board has recently asked the Society to appoint an official representative, and Professor Tocher has kindly consented to serve in this capacity. We may well be asked in the future for our views on what should be taught. If so I hope we will endeavour to formulate them. I know this is more difficult than it sounds—witness the trouble we have had to get some measure of agreement on an experimental scheme for teaching in schools—but it may well greatly strengthen this vital sector of our educational system.

Finally, may I make it clear that these comments are not intended as a criticism of the present editors of the Society's Journals. The editors are merely following policy which has grown up over the years, and which is very similar to the policy adopted for other statistical journals. What I want to convey is that the time has now come to examine this policy to see if there is need for change, and if change is considered necessary, to implement it.

REFERENCES

BAKER, D. N. (1965). Effects of certain environmental factors on net assimilation in cotton. *Crop. Sci.*, **5**, 53–56.

CLUTTON-BROCK, M. (1965). Using the observations to estimate the prior distribution. *J. R. Statist. Soc.* B, **27**, 17–27.

CROWTHER, E. M. and YATES, F. (1941). Fertilizer policy in war-time. The fertilizer requirements of arable crops. *Emp. J. exp. Agric.*, **9**, 77–97.

FISHER, R. A. (1953). The expansion of statistics. *J. R. Statist. Soc.* A, **116**, 1–9.

—— (1958). The nature of probability. *Centennial Rev.*, **2**, 261–274.

HARTER, H. L. (1961). On the analysis of split-plot experiments. *Biometrics*, **17**, 144–149.

KENDALL, M. G. (1968). On the future of statistics—a second look. *J. R. Statist. Soc.* A, **131**, 182–204.

KOCH, G. G. and SEN, P. K. (1968). Some aspects of the statistical analysis of the 'mixed' model. *Biometrics*, **24**, 27–48.

PLACKETT, R. L. (1960). Models in the analysis of variance. *J. R. Statist. Soc.* B, **22**, 195–217.

ROYAL SOCIETY (1963). *Emigration of Scientists from the United Kingdom.* Report of a Committee appointed by the Council of the Royal Society. London: Royal Society.

"STUDENT" (1923). On testing varieties of cereals. *Biometrika*, **15**, 271–293.

TUKEY, J. W. (1963). Contribution to discussion following paper by Bennett, G. W. and Cornish, E. A.: A comparison of the simultaneous fiducial distributions derived from the multivariate normal distribution. *Bull. Inst. Int. Statist.*, **40**, 902–939.

YATES, F. (1933). The principles of orthogonality and confounding in replicated experiments. *J. agric. Sci.*, **23**, 108–145.
—— (1964). Sir Ronald Fisher and the design of experiments. *Biometrics*, **20**, 307–321.
—— (1966). Computers, the second revolution in statistics. *Biometrics*, **22**, 233–251.
—— (1967). A fresh look at the basic principles of the design and analysis of experiments. *Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability*, **4**, 777–790.
YATES, F. and HEALY, M. J. R. (1964). How should we reform the teaching of statistics? *J. R. Stastist. Soc.* A, **127**, 199–210.

PROCEEDINGS OF THE MEETING

Professor M. S. BARTLETT (Oxford University): It seems a fitting moment to be speaking about Frank Yates, who has been doubly honoured by this Society, first by the award of its Gold Medal and now by its Presidential office, and also doubly honoured by the Royal Society, first by his election to Fellowship and, much more recently, by the award of a Royal Society Medal. He also retired this year from the post he has held wit'ı distinction for so many years, head of the statistics department at Rothamsted, to which I think he was appointed when R. A. Fisher moved to the Galton Laboratory in 1933.

It cannot have been an easy task to follow such an illustrious predecessor, especially if, and here I am merely guessing, Fisher was ready to give unsolicited advice "from the wings" if the occasion seemed to him to require it. Be that as it may, Dr Yates has maintained the Fisherian tradition at Rothamsted of statistical work firmly based on practice, as he himself emphasizes in his paper tonight. In particular, some of his first and most important contributions were in the areas of sampling theory and of experimental design, two areas which have been closely associated with the Rothamsted department. Dr Yates's earlier familiarity with least-squares techniques enabled him to make sure that analysis of variance was kept on a sound basis and was able to cope with the complications of unequal class frequencies, missing plots and other sources of non-orthogonality. His remarks on the history of analysis of variance and experimental design are especially to be noted; the historical interplay of theory and practice in Fisher's development of analysis of variance may also be discerned from the account of analysis of variance to be found in *Statistical Methods for Research Workers*, where, owing to Fisher's reluctance to rewrite, the technique is still, I believe, described there at times in the rather obsolete language of intra-class correlations. Like Dr Yates, I would also regret to see the essential simplicity of the principles of analysis of variance lost behind any excessive theoretical superstructure, though—and here I must be rather careful what I say, for if the President has been unmuzzled this evening, the muzzling of other discussants extends in spirit, on this one occasion, to the proposer and seconder of the vote of thanks, as the President reminded us at the beginning of his Address—I am not entirely clear how much Dr Yates is intending to cover by his remarks in Section 2 of his paper. I am sure he would agree that there are problems in statistics where excessive emphasis on the overall null hypothesis is misplaced; as one example, if in a randomized block experiment treatment effects have been established, it may certainly be relevant to examine particular treatment contrasts with their own treatment times replication error term, when other treatments have behaved so differently that the inclusion of such plots could be misleading.

Returning to the history of the statistics department at Rothamsted, I would recall further developments associated with Dr Yates and his colleagues, who have included birds of passage like David Finney, Frank Anscombe and Michael Healy. I would mention the systematic development of the use of computers, emphasized by our President in his paper tonight, but more extensively in his Fisher Memorial Lecture, applications of multivariate analysis and other valuable practical developments in statistical methodology. I would support the plea for more papers on problems associated with computers, and also call the attention of Fellows to the introduction in *Applied Statistics* of an information service on computer algorithms.