THE INFLUENCE OF *STATISTICAL METHODS FOR RESEARCH WORKERS* ON THE DEVELOPMENT OF THE SCIENCE OF STATISTICS

F. YATES
Rothamsted Experimental Station

IT IS now twenty-five years since R. A. Fisher's *Statistical Methods for Research Workers* was first published. These twenty-five years have seen a complete revolution in the statistical methods employed in scientific research, a revolution which can be directly attributed to the ideas contained in this book, and which has spread in ever-widening circles until there is no field of statistics in which the influence of Fisherian ideas is not profoundly felt.

Statistical Methods for Research Workers is a peculiarly personal production. It was written after five years work at Rothamsted, the largest and oldest of the British agricultural research stations, where Fisher had been appointed in 1919. At that time the idea of employing a statistician in such a field was a novel one. It was thought by the Director, Sir John Russell, that the accumulated results of the Rothamsted experiments would repay further examination by a mathematical statistician. In the event the appointment had much more far-reaching effects, as it resulted in the evolution of new statistical methods suitable for dealing with experimental material, and in the radical improvement of experimental design.

Statistical Methods embodies the results of Fisher's researches during his early years at Rothamsted. The methods put forward are largely those developed by the author himself to deal with the novel problems encountered as a result of his contacts with agricultural and biological research workers. They are based on the results of his own researches in mathematical statistics, the more important of which have recently been published in collected form [1]. The book is brief—the first edition contained only 239 pages of large type (350 words to the page), and the present edition (the 11th) contains only 354 pages of slightly smaller type. No mathematical proofs are included, and the discussion of the various subjects is by no means exhaustive. Apart from the addition of a chapter on estimation in the second edition, there have only been relatively minor additions to subsequent editions.

To appreciate fully the achievement which the book represents, we must recall the statistical atmosphere of the time. It was the age of correlation and curve fitting. In *Tables for Statisticians and Bio-*

metricians, for example, first published in 1914, 37 per cent of the tabular matter was concerned with curve fitting, and a further 18 per cent with various forms of correlation. The normal and Poisson distributions occupied 17 per cent, χ^2 and "Student's" z 5 per cent, the remaining 23 per cent being devoted to tables of basic mathematical functions and miscellaneous statistical tables.

It was also the age of coefficients of all kinds. In attempts to assess the degree of association in 2×2 contingency tables, for example, such measures as the coefficient of association, the coefficient of mean square contingency, the coefficient of tetrachoric correlation, equiprobable tetrachoric correlation, and the coefficient of colligation, were proposed. The way in which these coefficients were used revealed considerable confusion between the problems of estimating the degree of association, and testing the significance of the existence of an association. In the field of regression and correlation we find, in addition to the ordinary partial regression and correlation coefficients, the multiple correlation coefficient, the correlation ratio, and Blakeman's criterion. Even such a simple concept as the percentage standard deviation was termed the coefficient of variation.[1]

Statistical Methods cut through this jungle, and broke fresh ground in a number of entirely distinct ways. It recognized, and emphasized, the difference between the problems of estimation and tests of significance. It set out methods for the exact treatment of sampling problems of the type that arise in the commonly required tests of significance and introduced a unity of approach into these problems which was previously lacking; apart from tests involving only the "classical" distributions, the normal, binomial, and Poisson, the whole of the tests discussed are shown to be dependent on three fundamental distributions, χ^2, t and z, of which the first two are special cases of the last. It showed how, by the use of exact methods, many of which are of importance even with quite large samples, the small samples that occur so frequently in experimental work, but somewhat rarely in observational data, can be treated statistically. It recognized for the first time the importance of efficiency in estimation processes, and described a method (the method of maximum likelihood) for obtaining efficient statistics in practical cases. (This aspect was more fully developed in the second edition.) Finally, it laid the foundations of sound experimental design and analysis.

In the following sections of this paper, I shall endeavor to describe in a little more detail those features of the book which were of particular

[1] Readers interested in the multiplicity of coefficients, etc., current at this time may consult Kendall [9], which contains a very full account of a great number of them.

novelty at the time, and which exerted most influence on the subsequent development of statistical science.

GENERAL PRINCIPLES

The introductory chapter, which is chiefly concerned with concepts and basic principles, is a masterpiece of brevity and clarity. Nowhere, I think, is it possible even today to find such a lucid outline of the scope of statistical science, together with a record, in non-mathematical terms, of the ideas that Fisher had developed on estimation, maximum likelihood, tests of significance and the like. The concepts of consistency, efficiency, and sufficiency of estimates are introduced. The necessity of using efficient statistics when testing for goodness of fit is stressed. Inverse probability is rejected. This chapter must indeed have had a profound influence in spreading the new concepts among those who were chiefly concerned with statistics as a practical tool in research enquiry.

DISTRIBUTIONS

The chapter on distributions sets a brisk pace which is maintained throughout the book. A lot of junk is cleared away. The lengthy discussions of measures of the "central tendency" and of dispersion found in most of the statistical textbooks of the time are dispensed with—the median is not mentioned, and the probable error is dismissed with the characteristic Fisherian phrase: "The common use of the probable error is its only recommendation." The use of the mean and the mean square estimate of the variance are justified by their sufficiency properties for the normal curve, and the reader is introduced to the use of the normal probability integral, the fitting of a normal curve, grouping corrections and errors, the use of $n-1$ instead of n as divisor of the sum of the squares of the deviations, and tests of departure from normality by third and fourth moments, all in the space of 14 pages. In the next 17 pages he is expected to become familiar with the Poisson and binomial distributions, including tests of the variability of small samples from those distributions by means of the χ^2 distribution.

χ^2

The chapter on the use of the χ^2 tests of goodness of fit, independence and homogeneity likewise covers a great deal of ground, some of it simple and familiar to the statisticians of the time, and other parts which represented novel and difficult applications. Its most important immediate influence was to make available to the ordinary non-mathematical statistician a coherent account of the uses of χ^2, freed from the confusion that had existed regarding the number of degrees of freedom which were appropriate in different cases. Although the point which had been under discussion for 10 years and had been very fully treated by Fisher in scientific papers (papers 5, 7 and 8 of [6]) and had, as far as 2×2 tables were concerned, been correctly treated (though by approximate methods) by Yule in his *Introduction to the theory of statistics* [14] from its first publication in 1911, it was still the subject of controversy, and the reviewer of *Statistical Methods* in the *British Medical Journal* [2] in 1926 felt it necessary to write:

> The trained statistician interested in Mr. Fisher's researches will miss a detailed justification of his conclusions. . . . Even if the statement that Professor Pearson's treatment of a fundamental problem contained a "serious error" had not been disputable, and therefore improper in a work addressed to elementary students, it would have reminded anyone of Macaulay's remark on a similar occasion—"just so we have heard a baby, mounted on the shoulders of its father, cry out, 'how much taller I am than Papa!' "

Actually, the point was discussed by Fisher in a passage (Section 20) which for clarity of statement and convincingness of argument would be difficult to better:

> It was formerly believed that, in entering the χ^2 table, n was always to be equated to one less than the number of frequency classes; this view led to many discrepancies, and has since been disproved with the establishment of the rule stated above. On the old view, any complication of the hypothesis such as that which in the above instance admitted differential viability, was bound to give an apparent improvement in the agreement between observation and hypothesis. When the change in n is allowed for, this bias disappears, and if the value of P, rightly calculated, is many fold increased, as in this instance, the increase may safely be ascribed to an improvement in the hypothesis, and not to a mere increase of available constants.

The t distribution, which is dealt with in Chapter V, was less controversial, but many of its applications were of greater novelty than χ^2. The distribution was first deduced by Gosset ("Student") [13] in 1908 for the purpose of testing the significance of the mean of a small sample. Gosset himself, in his original paper, expresses very clearly the reasons why such a test was required in experimental work:

> There are other experiments, however, which cannot easily be repeated very often; in such cases it is sometimes necessary to judge of the certainty of the results from a very small sample, which itself affords the only indication of the variability. Some chemical, many biological, and most agricultural and large-scale experiments belong to this class, which has hitherto been almost outside the range of statistical inquiry.

> Again, although it is well known that the method of using the normal curve is only trustworthy when the sample is "large," no one has yet told us very clearly where the limit between "large" and "small" samples is to be drawn.
>
> The aim of the present paper is to determine the point at which we may use the tables of the probability integral in judging of the significance of the mean of a series of experiments, and to furnish alternative tables for use when the number of experiments is too few.

Nevertheless it is to Fisher, I think, that credit must be given for first recognizing the fundamental nature of the advance that the t test represented. Fisher also established with certainty the form of the distribution (Gosset had obtained the correct form by approximate methods), and he replaced Gosset's z by the more convenient $t=z\sqrt{n}$ $=$ estimate/estimated standard error of estimate.

Gosset's table (of z) was included in *Tables for Statisticians and Biometricians*, but it was only with the publication of *Statistical Methods* that the wide applicability of the test, which covers the whole class of problems in which an estimate is tested by means of an estimate of its standard error based on a small number of degrees of freedom, was brought to the attention of research workers.

Another feature of Chapter V was to exert a major influence for the better in the application of statistical methods. This is the fact that regression is considered, as it should be, in its own right, and not as an offshoot of correlation. Nothing had bedevilled the interpretation of statistical data involving a number of variates so much as the use of the correlation coefficient, and in particular partial correlation. In cases in which the influence of one or more variates on another is under consideration, regression analysis is almost always more appropriate—the very terms "dependent" and "independent" variates indicate this. Regression analysis provides coefficients and equations which are immediately interpretable in real physical terms and which are unaffected (except for precision) by the distribution of the values of the independent variates. The method is therefore immediately applicable to experimental situations in which the values of the independent variates are deliberately chosen by the investigator, and to comparative work in which the distribution of the independent variates differs from group to group of the data owing to natural causes.

CORRELATION AND THE ANALYSIS OF VARIANCE

The remaining three chapters of the first edition deal with inter- and intra-class correlation, the analysis of variance, and the design of experiments. It is here that the historical influence is most apparent. The chapter on the correlation coefficient originally opened with the sentence:

> No quantity is more characteristic of modern statistical work than the correlation coefficient, and no method has been applied successfully to such various data as the method of correlation.[2]

though even at this point the reader is warned that

> In experimental work proper its position is much less central; . . . it is seldom, with controlled experimental conditions, that it is desired to express our conclusion in the form of a correlation coefficient.

Because of the importance that correlation analysis had assumed it was natural that the analysis of variance should be approached via correlation, but to those not trained in the school of correlation analysis (of which I am fortunate to be able to count myself one) this undoubtedly makes this part of the book more difficult to comprehend, as is admitted by Fisher in the preface to the 9th edition. Of all the statistical methods of analysis that Fisher has introduced, the analysis of variance has probably had the most profound and far-reaching influence. It is not, however, until the reader has mastered the subject of intra-class correlation that the possibility of an alternative approach, that of the analysis of variance, is revealed with the words (Section 40)

> A very great simplification is introduced into questions involving intra-class correlation when we recognise that in such cases the correlation merely measures the relative importance of two groups of factors causing variation.

Fisher himself has stated that the analysis of variance is "merely a way of arranging the arithmetic." This seems to me undue modesty. The concept of additive components of variance, and its concomitant, the possibility of expressing the values of a variate in terms of an additive set of parameters with the parameter for a given classification having a fixed value for every member of each single class of this classification, marked a major break with tradition, and provided the essential link between least squares and regression analyses and the problems previously treated by intra-class correlation. It also directed attention to the features of the data that mattered, namely the differences between the means of the different classes, instead of concentrating attention on the usually relatively unimportant aspect of the degree of similarity within classes. Furthermore it provided a method of eliminating more than one source of variation (in those cases common in planned experiments, in which the data are what is now known as orthogonal), and also automatically, as it were, provided a pooled estimate of error by means of which the individual class means might be compared.

[2] Significant changes were made in the wording of this passage in subsequent editions. In the fourth edition "modern statistical" was replaced by "biometrical" and the word "successfully" was deleted. In the fifth edition, "is" was changed to "has been."

In addition to developing the analysis of variance procedure, the first edition of *Statistical Methods* made available for the first time the relevant exact test of significance by providing a table of the z distribution for the 5 per cent level of significance. This was expanded and a 1 per cent level added in subsequent editions. The analysis of variance would have been of very considerable value for many purposes even had the z test not been available—the comparison of the means of a specific pair of classes, for example, can be made by means of the t test—but the provision of the exact test for variation between the means of all classes, or any group of them, introduced a logical completeness and exactitude into the whole structure of the methods described in *Statistical Methods* and enabled the book to be written without any substantial reference to approximate methods.

THE DESIGN OF EXPERIMENTS

The development of the analysis of variance opened the way to the whole of the modern technique of the design and analysis of experiments. The state of experimental design at that time, and the deductions that it was considered reasonable to draw from the results, are well illustrated by an extract from the Rothamsted Report for 1918–1920. Under the heading of "The amount of fertilizers to use," and after a discussion of the law of diminishing returns, and mention of the fact that on Broadbalk (the long-term wheat experiment) "the largest return is given not by the first dressing but by the second," it is stated:

> . . . a new experiment has been started to see if under ordinary conditions of farming the highest rate of profit is given by good rather than by small dressings of fertilizers. The results of the first year (1920) suggest that this may be so.

INCREASE IN WHEAT CROP, 1920, FROM SPRING DRESSINGS OF SULPHATE OF AMMONIA AND SUPERPHOSPHATE

	Grain: Bushels per acre			Straw: Cwts. per acre		
Date of application of manure	Feb. 10	March 6	May 10	Feb. 10	March 6	May 10
Single dressing	Nil*	0.9	2.7	2.7	6.9	9.4
Double dressing	7.0	—	3.7	11.7	—	12.7

* The correct value from the plot yields is −0.2. Presumably the alteration was made because the presence of a negative value in the table would have made the results appear less trustworthy to the average reader.

> While the single dressing (100 lbs. sulphate of ammonia per acre) gave no appreciable increase in grain, and only a few cwts. of additional straw, the double dressing gave increases of no less than 7 bushels of grain and 12 cwts. of straw. Late application of the double dressing, however, was risky, giving an unhealthy straw liable to lodge and prone to disease.

The experiment referred to consisted of 6 plots, one for each of the treatments shown in the table, together with a control receiving no nitrogen. The variation between the yields of grain of the plots receiving nitrogen is equivalent to a standard error of 9% per plot, which we now know is about what would be expected from variations in fertility and other sources of experimental error. The results, far from demonstrating that the response to the double dressing is more than double that to the single dressing, are not inconsistent with the hypothesis that there is little additional response to the double dressing.

The subject of experimental design is only dealt with very briefly in *Statistical Methods*. There are two sections at the end of the chapter on application of the analysis of variance, and an example on the analysis of the results of an agricultural field trial in the previous chapter. These passages, however, contain between them all the basic principles which govern modern experimental design.

It is of interest to note that the principles of design are expounded after the methods of analysis had been illustrated in a somewhat complex example. This was in fact the historical order in which the subject was developed. It was by applying the methods of the analysis of variance to the results of experiments which did not conform to the principles of good design that their defects became apparent.

Apart from the new method of analysis provided by the analysis of variance procedure, Fisher's really novel contribution to experimental design was his insistence on the necessity of randomization, in order to ensure that the estimates of error and tests of significance should be fully valid. Any form of systematic arrangement casts doubt on the estimates of error and tests of significance. In cases such as agricultural field trials, in which the variation of the yields from plot to plot itself exhibits systematic features, it may wholly vitiate them. The adoption of randomization does not preclude the possibility of imposing restrictions such as, for example, arranging all the treatments of each replicate in a compact block with allocation at random within the block. At the same time the types of restriction which are in fact capable of giving an unbiased estimate of error are quite limited. Thus an arrangement in randomized blocks in which the positions of the treatments are balanced so as to eliminate the effect of a fertility gradient across the plots is not capable of giving a fully valid estimate of error. On the other hand, Latin squares, which are double restricted arrangements in which each treatment occurs once only in each row and in each column, are capable of furnishing valid estimates of error provided that a random selection from all possible Latin squares of the given size is made.

These are the essentials of modern experimental design. In the succeeding years many detailed refinements have been introduced, both by Fisher himself and many others working in association with him. But it is on the foundations outlined above, which were expounded in the first edition of *Statistical Methods*, that all these refinements rest. And it is to the influence of *Statistical Methods* that much of the credit must be given for the rapid adoption of the new methods by practical agricultural and biological research workers.

The illustrative example of Chapter VII is itself worth careful study. The experiment, which was carried out at Rothamsted in 1922, was one on 12 varieties of potatoes with three plots (patches) of each variety, each of these plots being split into three for two types of potash and a control without potash.

The analysis of variance given in *Statistical Methods* is as follows:

Variance due to	Degrees of Freedom	Sum of Squares	Mean Square
Between varieties	11	43.6384	3.967
Between patches for same variety	24	17.4401	.727
Within patches			
Potash dressing	1	.2911	.2911
Sulphate v. chloride	1	.0584	.0584
Differential response of varieties	22	2.1911	.0996
Differential response in patches with same variety	48	8.0798	.1683
Total within patches	72	10.6204	
Total	107	71.6989	

It will be noted that no component corresponding to blocks was included in the analysis. Had the three replicates been considered as constituting blocks the whole-plot part of the analysis would have become:

	Degrees of freedom	Sum of Squares	Mean Square
Between replicates	2	10.1280	5.064
Between varieties	11	43.6384	3.967
Remainder	22	7.3120	0.3324
Total	35	61.0784	

Replicates are clearly significant, indicating some form of blocking on the ground (though such records as are available indicate that this blocking was somewhat imperfect). The whole-plot error is now reduced from 0.727 to 0.332. One may hazard the guess that the latter is a better, though doubtless somewhat imperfect, estimate of error. The fact that this point was not discussed is an interesting indication of the tentative character of the early statistical analyses of experimental results.

In other respects the analysis of this experiment exhibits a remarkable degree of development. Points to notice are: the method of dealing with split plots, working in terms of sub-plot units throughout; the introduction of an interaction term for the interaction between the different factors, varieties and potash; and the sub-division of the two degrees of freedom for potash into the average effect of potash and the difference between the two forms.

Incidentally, the derivation of the sums of squares for this sub-division provides an example of the way in which Fisher sometimes left points for his readers to worry out for themselves, considering, doubtless, that it would stimulate thought. The relevant passage runs as follows:

> The sum of the squares of the three deviations, divided by 36, is .3495; of this the square of the difference of the totals for the two potash dressings, divided by 72, contributes .0584, while the square of the difference between their mean and the total for the basal dressing, divided by 54, gives the remainder, .2911.

No clue is provided as to the derivation of the divisors 72 and 54, the second of which must undoubtedly have defeated many biologists and agriculturalists unversed in the formal algebraic theory of errors.

SUBSEQUENT ADDITIONS

As already mentioned the only real structural alteration to the book is the addition to the second edition of a chapter on estimation. This chapter replaced and expanded the very brief account of the method of maximum likelihood which is given as an Example in Chapter I in the first edition.

This expansion increased somewhat the emphasis on the problems of estimation, but the chapter is more specialized and difficult than the rest of the book and it remained true that the main emphasis of the more elementary parts of the book lay in the direction of tests of significance. I refer again to this point in the last section.

Numerous other additions have been made from time to time, giving accounts of new developments which appeared to the author to be of interest or importance. These, however, for the most part fit closely into the original framework of the book, and may be regarded as extensions of the structure already laid down, rather than as radical innovations. Probably the only two additions that can really claim this distinction are those on the analysis of covariance (fourth and fifth editions) and discriminant functions (seventh edition). Other additions of particular interest in revealing the development of the subject are those on orthogonal polynomials (third and seventh editions), the exact test of significance of 2×2 contingency tables (fifth edition) and the extension of the t test to give fiducial limits for the ratio of means and regression coefficients (tenth edition).

The sections on experimental design and analysis, the branch of the subject which has probably shown the greatest growth, have (apart from the addition on the analysis of covariance) been left almost without alteration. It is, I think, an interesting reflection on the historical sense of the author, and on his inclination to leave the field open to others to make their contributions, that although he was himself actively working in the field at the time he thought it better to leave them wholly unaltered, until he felt the time was ripe for a completely new book on the subject. Certainly *The Design of Experiments* can be regarded as a worthy offshoot of *Statistical Methods.*

It is also interesting to note that the parts of the book dealing with correlation have remained almost without alteration, and now stand as a monument to a bygone age of statistics.

RECEPTION OF THE BOOK

As is only to be expected with a book that marks such a fundamental break with tradition, its full significance was not immediately recognized. Nevertheless the reviewers of the first edition did perceive that the book was an important one, and they confined their criticisms mainly to lack of due deference to authority and to questions on intelligibility and presentation. As might be expected, the absence of mathematical proofs was felt by many to be a defect, either because it made the book difficult to follow, or, more strangely, because their absence would prevent the reader from verifying for himself the validity of the methods proposed. Even sixteen years later there still existed a desire for the mathematical "proof." (See, for example, M. G. Kendall [8].)

Fisher himself commented on this point in the prefaces to the later editions. Thus he states in the preface to the 9th edition:

> The practical application of general theorems is a different art from their establishment by mathematical proof. It requires fully as deep an understanding of their meaning, and is, moreover, useful to many to whom the other is unnecessary.

That such understanding does not flow from the mathematical proof is sufficiently demonstrated by the number of advanced textbooks in mathematical statistics in existence today which establish the procedure of the analysis of variance appropriate to replicated experiments and analogous material without reference to randomization.

Apart from this the main defect of the early reviews was their assumption that the applications of small sample theory were solely confined to small samples, and their consequent implication that the book was of limited interest to the general statistician. Thus the review in *Nature* (Anon.) [1] states: "It treats of the interesting and important subject of small samples in statistical work." That in *Science Progress* ("E.S.P.") [3] states: "The book is chiefly concerned with the best methods of handling small samples." That in the *Journal of the Royal Statistical Society* ("L.I.") [10], though it indicates that many of the methods were in fact not confined to small samples, concludes: "The book will undoubtedly prove of great value to research workers whose statistical series necessarily consist of small samples. . . . " In the *British Medical Journal* (Anon.) [2] it is stated: "Since in the kind of biological research with which Mr. Fisher has had to deal practically small samples only are usually available, he has given more attention to the particular methods applicable to small samples than authors of most textbooks have deemed necessary." Yule in his 8th edition (1927) refers to *Statistical Methods* as "a laboratory handbook rather than a textbook, [which] brings together in convenient form for the research worker the numerous methods developed, mainly by [the author], with special reference to small samples."

In actual fact many of the methods described in *Statistical Methods* are relevant to the treatment of large samples. The essential point is that even when the data consist of observations on a large number of separate individuals these often require to be grouped according to a relatively small number of classes. Tests of significance involving these classes frequently involve small sample theory. The examples included in the book are, indeed, fairly evenly distributed over what would be described as small and large samples, as is shown by the following table:

Size of sample	*No. of examples*
8–20	8
21–100	16
100–500	6
501–1000	4
Over 1000	16

The continued and growing demand for the book is best indicated by the numbers of copies printed for the various editions. These have been kindly supplied by the publishers, and are as follows:

1st Edition	1925	1050	7th Edition	1938	2000
2nd Edition	1928	1250	8th Edition	1941	2250
3rd Edition	1930	1500	9th Edition	1944	2000
4th Edition	1932	1500	10th Edition	1946	3000
5th Edition	1934	1500	Reprinted	1948	1500
6th Edition	1936	2000	11th Edition	1950	7500

In all nearly 20,000 copies have been sold during the first 25 years of the book's existence. The rate of sale during the latter half of this period has been very constant at about 1000 copies a year. The book has also been translated and published in French, Italian, and Spanish. It is being translated into German and into Japanese and publication should take place within one or two years. No figures for the distribution of sales of the English editions over the different countries are available, but the publishers state that the early editions were sold mainly, if not entirely, in the United Kingdom, and that it would be reasonable to assume that at present practically half of each new edition is sold abroad to various countries, principally to the United States of America.

Many requests (which have always been granted) have also been made for permission to reproduce in whole or in part tables and other matter first published in *Statistical Methods*. The basic tables have been reproduced almost without alteration in *Statistical Tables for Biological, Agricultural and Medical Research*, now in its third edition.

These facts provide additional evidence, if any is needed, of the wide influence of the book.

PRESENT TRENDS

In conclusion we may ask, in the light of present-day statistical teaching and practice, how far the methods embodied in *Statistical Methods* have found acceptance, whether they have been rightly understood and applied, and whether in the light of experience of their use extension or modification is required.

A full discussion of this point is beyond the scope of this article, but in the matter of experimental design, which, as I have tried to indicate, is the most novel contribution of *Statistical Methods*, it can be said without hesitation that the new methods have been completely accepted by biological and agricultural research workers, and that they are rapidly spreading through other branches of scientific and technical research in which the variability of the experimental material necessitates refined techniques. Their introduction has resulted in an immense gain in the accuracy and certainty of experimental results, and there is no reason to doubt that the development, which is still continuing, of technqiues appropriate to the very varied problems and situations met with in scientific experimentation, is on the right lines. As examples, we may instance work in recent years on long-term change-over trials in agriculture, biological assay, and industrial experimentation and quality control.

On the other hand the emphasis given to formal tests of significance throughout *Statistical Methods*, and to a great extent also in *The Design of Experiments*, has had two consequences which are not wholly satisfactory. In the first place it has resulted in what seems to me to be an undue concentration of effort by mathematical statisticians on investigations of tests of significance applicable to problems which are of little or no practical importance. Second, and more important, it has caused scientific research workers to pay undue attention to the results of the tests of significance they perform on their data, particularly data derived from experiments, and too little to the estimates of the magnitude of the effects they are investigating.

Historically this situation is understandable. When *Statistical Methods* was written the methods used for testing significance were, as we have seen, in the utmost confusion. In the interpretation of their results research workers in particular badly needed the convenience and the discipline afforded by reliable and easily applied tests of significance. The example, quoted above, of an early Rothamsted experiment, shows how important this discipline is. Nevertheless the occasions, even in research work, in which quantitative data are collected solely with the object of proving or disproving a given hypothesis are relatively rare.

Usually quantitative estimates and fiducial limits are required. Tests of significance are preliminary or ancillary.

The emphasis on tests of significance, and the consideration of the results of each experiment in isolation, have had the unfortunate consequence that scientific workers have often regarded the execution of a test of significance on an experiment as the ultimate objective. Results are significant or not significant and that is the end of it.

Research workers, therefore, have to accustom themselves to the fact that in many branches of research the really critical experiment is rare, and that it is frequently necessary to combine the results of numbers of experiments dealing with the same issue in order to form a satisfactory picture of the true situation. This is particularly true of agricultural field trials, where in general the effects of the treatments are found to vary with soil and meteorological conditions. In consequence it is absolutely essential to repeat the experiment at different places and in different years if results of any general validity or interest are to be obtained. In such circumstances a number of experiments of moderate accuracy are of far greater value than a single experiment of very high accuracy.

The combination of the results of groups of experiments on the same issue introduces problems of statistical technique, particularly in the estimation of errors, but also to some extent in the estimation of the effects themselves, which are not met with in the analysis of a single experiment. Uncritical application of the analysis of variance procedure is likely to give uninformative and sometimes misleading results.

The same situation is met with in the analysis of observational data. Multiple classifications are frequently met with in such data, which at first sight appear amenable to the analysis of variance technique. However, lack of orthogonality introduces many complications, and although the theory of the subject is well understood, practical methods of analysis which do not involve excessive computation are not available. Nevertheless the model provided by the simpler case where the data are orthogonal is of immense value in indicating the objectives to strive for, and we may confidently expect rapid development in this field.

We may expect also to see considerable developments in the practical applications of multivariate analysis. Here again practice has lagged behind theory because of the large amount of computation required. Just as the practical development of experimental design was made possible by the introduction of the desk calculating machine, which enabled the results to be analyzed without undue labor, so we may expect that recent developments of electronic and relay calculators, and the wider availability of punched card apparatus, will result in a corresponding development of multivariate analysis, so that in a few years' time the sections of *Statistical Methods* on covariance and discriminant functions will bear the same relation to this branch of the subject as do the sections on experimental design to present practice in that field.

REFERENCES

[1] Anon., Review of Fisher's Statistical Methods (First Edition), *Nature*, Vol. 116 (1925) 815.

[2] Anon., Review of Fisher's Statistical Methods (First Edition), *British Medical Journal*, Vol. 1 (1926) 578–9.

[3] "E.S.P.," Review of Fisher's Statistical Methods (First Edition), *Science Progress*, Vol. 20 (1926) 733–4.

[4] Fisher, R. A., *Statistical Methods for Research Workers* (Eleventh Edition). Edinburgh: Oliver and Boyd. 1950.

[5] Fisher, R. A., *The Design of Experiments* (Fourth Edition). Edinburgh: Oliver and Boyd. 1947.

[6] Fisher, R. A., *Contributions to Mathematical Statistics.* New York: John Wiley and Sons. 1950.

[7] Fisher, R. A. and Yates, F. *Statistical Tables for Biological, Agricultural and Medical Research* (Third Edition). Edinburgh: Oliver and Boyd. 1948.

[8] Kendall, M. G., *Nature*, Vol. 149 (1942) 451.

[9] Kendall, M. G., *The Advanced Theory of Statistics, Volume I* (Third Edition). London: Griffin. 1947.

[10] "L.I." Review of Fisher's Statistical Methods (First Edition), *Journal of the Royal Statistical Society*, Vol. 89 (1926) 145–6.

[11] Pearson, K., *Tables for Statisticians and Biometricians*, Cambridge University Press. 1914.

[12] Rothamsted Experimental Station. Report for 1918–1920.

[13] "Student," "The Probable Error of a Mean," *Biometrika*, Vol. 6 (1908) 1–25.

[14] Yule, G. U., *An Introduction to the Theory of Statistics.* London: Griffin. 1911.