Reproduced from the Annual Report Rothamsted Experimental Station, 1933, p. 43-50.

103

THE CONTRIBUTIONS OF ROTHAMSTED TO THE DEVELOPMENT OF THE SCIENCE OF STATISTICS

R. A. Fisher.

In 1919, the year in which the Statistical Laboratory was founded, the function of the statistician was understood to consist in the determination from the data presented to him of certain average values, more or less capable of scientific interpretation, and also by the use of quantities of the second degree, squares and products, of " probable errors " regarded as adherent to the averages obtained. The term averages in this expression is to be interpreted somewhat widely, as will be more fully explained, and any extensive body of data is capable of yielding an unending variety of such quantities. Exactly what averages to obtain from the data depends inevitably on what kind of information it is desired to elicit, and is to this extent not a statistical question. A comprehension of the arithmetical processes is, however, needed to ensure that the averages obtained shall be appropriate to the meanings which it is hoped to place upon them ; moreover, extensive bodies of data usually contain information on points which were not in mind when the observations were made, and some knowledge is needed to determine what types of information are available, and by what methods they can be elicited.

As a very simple example, it was shown in 1921 (1) that when the two methods yield appreciably different results, it was preferable to calculate the relative growth rate of plants or animals, not as some plant physiologists had maintained, by means of a formula analogous to that of simple interest, but on one analogous to compound interest. Again, with a sequence of annual figures, of the type prevalent in economic and vital statistics, and in the records of meteorological observations, and the " classical fields " at Rothams-

(1) R. A. Fisher—" Some remarks on the methods formulated in a recent article on ' The Quantitative Analysis of Plant Growth,' " Ann. App. Biol., VII, 367-372.

Annual Report of the Rothamsted Experimental Station, 43-50, (1933).

ted, it was shown in the same year (2) that by means of a series of averages related to the temporal order of the sequence, the greater part of the slow changes, ascribable to soil deterioration, weed infestation, changes in variety or cultural practice, could be separated from the annual fluctuations, ascribable to weather variations and "experimental error," so as to enable these two classes of variation to be studied without serious mutual interference and confusion. The most extensive work of this class undertaken in the laboratory was the calculation of the average effects of meteorological factors such as rainfall and sunshine at all periods of the year, on the yield of crops grown on the classical fields. This was first done for rainfall and wheat (3), and later the method was applied to sunshine and wheat, to rainfall and barley, and more recently to rainfall and mangolds. A later series of papers is concerned with the experimental evaluation of the constants of formulae, expressing the increase in yield produced by successive additions of one or more fertilisers (4).

The interpretation of all such estimates, formed by the combination of inexact observations, requires that the discrepancies also should be taken into account. This is the purpose of the calculation of a probable error, or standard error, but recent research commencing with the work of "*Student*" in 1908 has shown not merely that the concept of a probable error is insufficiently exact for application to the small numbers of observations usually available from experimental work, but that it is possible in suitable cases to develop exact tests of significance from which the notion of a standard error may be eliminated, and in which it plays therefore only a formal part. Neither the theoretical nor the practical significance of this advance was readily appreciated, partly because academic statisticians were not aware of the serious decisions which experimenters must take on the basis always of limited data, partly because "*Student*" only treated an isolated and especially simple case, and it was not understood that the exact mathematical treatment of other more complicated cases arising in practical research was at all practicable. The work of developing exact methods appropriate to the actual nature of the experimental data is probably that aspect of the work of the statistical laboratory which is best known. This is partly due to the fact that many biologists were in a position to appreciate the advantage of introducing such methods as ancillary to their own studies, as is exemplified in (5); partly to the fact that the exact

(2) R. A. Fisher—"Studies in Crop Variation. I. An Examination of the Yield of Dressed Grain from Broadbalk." J. Agri. Sci., XI, 107-135; and later W. A. Mackenzie—"Studies in Crop Variation III. An Examination of the Yield of Dressed Grain from Hoos Field," Journ. Agric. Sci., vol. XIV, 1924, pp. 434-460; R. J. Kalamkar—"A Statistical Examination of the Yield of Mangolds from Barnfield at Rothamsted," Journ. Agric. Sci., vol. XXIII, part II, 1933. pp. 161-175.

(3) R. A. Fisher—"The Influence of Rainfall on the Yield of Wheat at Rothamsted," Phil. Trans., 213, 89-142; L. H. C. Tippett "On the Effect of Sunshine on Wheat Yield at Rothamsted," J. Agri. Sci., XVI, 159-165; W. A. Mackenzie and J. Wishart—"The Influence of Rainfall on the Yield of Barley at Rothamsted," J.Agri. Sci., XX, 417-439; R. J. Kalamkar—"The Influence of Rainfall on the Yield of Mangolds at Rothamsted," Journ. Agric. Sci., XXIII, Part IV, 1933. pp. 571-579.

(4) B. Balmukand—"The Relation Between Yield and Soil Nutrients," J. Agri. Sci., XVIII, 602-627; R. J. Kalamkar—"An Application of the Resistance Formula to Potato Data," J. Agri. Sci., XX, 440-454.

(5) R. A. Fisher.—"The Accuracy of the Plating Method of Estimating the Density of Bacteria Populations" (1922), Ann. App. Biol., IX, 325-359; "Statistical Study on the Effect of Manuring on Infestation of Barley by Gout Fly" (1924), Ann. App. Biol., XI, 220-235; "Tests of Significance in Harmonic Analysis" (1929), Proc. Roy. Soc., A, 125, 54-59; J. B. Hutchinson—"The application of the 'method of maximum likelihood' to the estimation of linkage" (1929), Genetics, 14, 519-537.

mathematical treatment of problems hitherto regarded as insoluble, opened a new field of study in mathematical statistics (6), which supplied the foundation upon which simple and exact practical procedures were based. Much discussion was naturally engendered by the fact that many procedures widely used and believed to be satisfactory approximations were found to be wholly misleading. As might be expected both mathematical statisticians, and biologists other than agronomists, whose work involves numerical tests, such as geneticists, entomologists, marine biologists, etc., have been strongly represented among voluntary workers.

Among types of data to which particular attention has been paid is that presented by series of annual figures, such as those obtained in meteorological records, drain gauges, and the classical experiments. This type of data, to which the bulk of official statistics belongs, has offered the greatest difficulty to economists and sociologists, and it was inevitable that such progress as had been made at Rothamsted in the development of methods of analysis should have attracted interest outside the sphere of agricultural science.

Although the solution of the problems of statistical distribution was primarily necessitated by the immediate requirements of practical research, it has brought with it theoretical consequences in the development of the mathematical theory of estimation. The best method of averaging, or of combining the observations, for any defined purpose, may be inferred from the nature of the errors to which different types of estimate are liable. The practical importance of this step is that it enables the computer to go ahead with confidence that he is getting the whole of the value out of the material being analysed. Its theoretical importance is that it gives the qualities of coherence and exactitude to the processes of inductive reasoning, by which conclusions of general application are deduced from particular observations (arguments from the sample to the population); while other branches of mathematics are applied only to deductive reasoning. A whole series of papers deals with this development (7).

Any sweeping theoretical advance, simply because it affects the way in which people are thinking of their problems, is likely to have unexpected consequences. One striking effect, due to a too close pre-occupation with academic ideas, rather than with the practical purposes for which these ideas were developed, is that the scientific interest of the subject is thought to be exhausted. It is, indeed, true that schools of thought, whose whole horizon has been occupied for a generation by problems of "curve fitting," should find themselves

(6) R. A. Fisher.—"On the 'probable error' of a coefficient of correlation deduced from a small sample" (1921) Metron. 1 (4), 1-32; "The Goodness of Fit of Regression Formulae and the Distribution of Regression Coefficients" (1922), Jour. Roy. Stat. Soc., LXXXV, 597-612; "The Conditions Under which χ^2 measures the Discrepancy between Observation and Hypothesis" (1924), Jour. Roy. Stat. Soc., LXXXVII, 442-449; "The General Sampling Distribution of the Multiple Correlation Coefficient" (1928), Proc. Roy. Soc. A, 121, 654-673; "Tests of Significance in Harmonic Analysis" (1929), Proc. Roy. Soc., A, 125, 54-59; "The Moments of the Distribution for Normal Samples of Measures of Departure from Normality" (1930), Proc. Roy. Soc., A, 130, 16-28; "The Sampling Error of Estimated Deviates, Together with other Illustrations of the Properties and Applications of the Integrals and Derivatives of the Normal Error Function," Brit. Ass., Math. Tables, Vol. I (1931) (Introduction), pp. xxvi-xxxv.

(7) R. A. Fisher.—"A Mathematical Examination of the Methods of Determining the Accuracy of an Observation by the Mean Error, and the Mean Square Error" (1920), Monthly Notices of the Roy. Astrom. Soc. LXXX, 758-770; "On the Mathematical Foundations of Theoretical Statistics," Phil. Trans. Roy. Soc. London, A, CCXXII, 1922. Pp. 309-368; "Theory of Statistical Estimation," Proc. Cam. Phil. Soc., vol. XXII, 1925. Pp. 700-725; "Inverse Probability," Proc. Cam. Phil. Soc., vol. XXVI, 1930. Pp. 528-535; "Inverse Probability and the Use of Likelihood" (1932), Proc. Cam. Phil. Soc., XXVIII, 257-261.

faced with vacuity, when the methods of estimation appropriate to any particular case can be set down at once by any novice who knows the theory ; but the fact remains that had curve fitting been regarded solely as a means to the practical purpose of eliciting the scientific facts derivable from the data, the predominant feeling would have been merely that a troublesome obstacle had been removed. The ground has at the same time been cleared of another misapprehension derived from the same source. In the period when highly inefficient methods of estimation were habitual, the amount of information extracted by the statistician depended largely on his personal skill and acumen. "High correlations," and significant results, when obtained, were displayed with some pride, as in some way implying personal competence. When, on the other hand, methods known to be fully efficient are used, the amount of information which the data are capable of yielding has also been assessed, and it is useless either to commend the statistician if it is much, or to reproach him if it is little. The statistician must be treated less like a conjurer whose business it is to exceed expectation, than as a chemist who undertakes to assay how much of value the material submitted to him contains. The idea of treating "amount of information" as a mathematical quantity, like the ideas of likelihood, and of intrinsic accuracy, is itself derived from the theory of estimation. An essential part of the statistician's task is how to evaluate the limitations of the data in hand ; as is stated emphatically by E. B. Wilson :

> We must expect that in many cases the statistical indications will lead us temporarily to abandon our problem because of a realisation of the fact that material adequate to its solution cannot be had.—*American Journ. of Cancer*, 1932.

The exhaustion of the task of improving on inefficient methods of calculation brings us at once face to face with the defects in experimental technique or in observational procedure, to which the intrinsic limitations of the data are due.

The work of this type to which the Statistical Laboratory has given most attention is the improvement of field experiments (8). At the time this work was started, it was customary to carry out the operations of field plot experimentation with great care as to the measuring of the land, separation and weighing of the crop, etc, without the experimental results attaining to anything like comparable precision. This was due to the heterogeneity of the soil, which was always found to be considerable, however much

(8) R. A. Fisher and W. A. Mackenzie—"The Manurial Response of Different Potato Varieties" 1923, Journ. Agri. Sci., XIII, 311-320 ; R. A. Fisher—"The Arrangement of Field Experiments," Journ. Min. Agric., vol. XXXIII, 1926. pp. 503-513 ; T. Eden and R. A. Fisher—"Studies in crop variation : IV. The Experimental Determination of the Value of Top Dressings with Cereals," Journ. Agric. Sci., vol. XVII, 1927, Pp. 548-562 ; T. Eden and R. A. Fisher—"Studies in Crop Variation : VI. Experiments on the Response of the Potato to Potash and Nitrogen." Journ. Agric. Sci., vol. XIV, 1929, pp. 201-213 ; J. Wishart and H. J. G. Hines—"Fertilizer Trials on the Ordinary Farm," Journ. Min. Agric., 1929, pp. 524-532 ; H. G. Sanders—"A Note on the Value of Uniformity Trials for Subsequent Experiments," Journ. Agric. Sci., vol. XX, 1930, pp. 63-73 ; R. A. Fisher and J. Wishart—"The arrangement of Field Experiments and the Statistical Reduction of the Results," Imperial Bureau Soil Science, Tech. Com. No. 10, 1930; F. E. Allan and J. Wishart—"A Method of Estimating the Yield of a Missing Plot in Field Experimental Work," Journ. Agric. Sci., vol. XX, 1930, pp. 399-406 ; J. O. Irwin—"Precision Records in Horticulture," Journ. of Pomology and Horticultural Sci., vol. IX, 1931, pp. 149-194 ; B. G. Christidis—"The Importance of the Shape of Plots in Field Experimentation," Journ. Agric. Sci., vol. XXI, 1931, pp. 14-37 ; F. R. Immer—"Size and Shape of Plot in Relation to Field Experiments with Sugar Beets," Journ. Agric. Res. 44, 1932, pp. 649-668 ; S. H. Justesen—"Influence of Size and Shape of Plots on the Precision of Field Experiments with Potatoes," Journ. Agric. Sci., vol. XXII, 1932, pp. 366-372 ; R. J. Kalamkar—"Experimental Error and the Field Plot Technique with Potatoes," Journ. Agric. Sci., vol. XXII, 1932, pp. 373-383 ; F. Yates—"The Analysis of Replicated Experiments when the Field Results are Incomplete," Emp. Journ. Expt., Agric. I, 1933, pp. 129-142.

trouble had been given to choosing a " uniform piece of land." In these circumstances subdivision of the experimental area into small replicated plots performed the double service of diminishing the experimental error, and of providing an estimate of the error that remained. The rationale and implications of this procedure were, however, for some time the subject of a certain amount of misunderstanding.

In the first place, whereas the object of diminishing the experimental error is much aided by replication, replication is only one of many methods of furthering this aim. Care in ensuring that in all points the experimental area is treated as in agricultural practice, the elimination of border rows, accuracy in seed rates, spacing, and measurement of the experimental area, as well as care in the separation, weighing and analysis of the produce, all make their contribution to the amount of information which the experiment finally gives. It is only when the " working errors " are reduced to unimportant quantities that soil heterogeneity becomes the major cause of error, and that greatly increased precision can be attained by improving the replication and arrangement of the experiment. On the other hand replication is the sole source of the estimate of error, by which the value and significance of the experiment is to be assessed. The estimate of error is not created by the statistician out of nothing, but is inferred from the observations by a process of estimation analogous to that used in the estimation of any other quantity, and requiring the same care in experimental design if the estimate is to be a valid one.

Owing to the fact, however, that the material conduct of an experiment had been regarded as a different business from its statistical interpretation, serious lacunae had been permitted between what had, in fact, been done, and what was to be assumed for mathematical purposes. In consequence methods of statistical analysis had been widely used, which gave definitely misleading estimates of error ; and, on the other hand, methods of field experimentation had been employed which were inherently incapable of yielding a valid test. It was necessary to treat the question of the field procedure, and that of statistical analysis as but two aspects of a single problem, and an examination of the relationship between these two aspects showed that once the practical field procedure was fixed, only a single method of statistical analysis could be valid, and, what was of more practical importance, that its validity depended on the introduction of a random element in the arrangement of the plots. The specification of the particular process of randomisation carried out, determined in advance the correct statistical analysis of the results. The logical structure of each of the possible types of randomisation is easily sorted out by the arithmetical arrangement known as the analysis of variance.

When not more than eight varieties, or treatments, or combinations of these are to be compared, a very complete elimination of the errors due to soil heterogeneity is possible by means of the Latin square, in which the number of replications is equal to the number of treatments, each of which appears once in each row and once in each column of the square. If, apart from this restriction, the plots are arranged at random, a valid and usually much diminished estimate of error is available, as was inferred theoretically, and later demonstrated experimentally by Tedin, on the basis of 92 uniformity trials (9); systematic arrangements in a square may give consistently either an over- or an under-estimate. A simple and more flexible arrangement adaptable to any numbers of treatments and replications is that known as randomised blocks, in which each block contains one plot of each treatment, these being distributed at random within it. Both these arrangements are now used all over the world with the exception of France, Italy, and parts of Germany, where other methods believed to be satisfactory had been previously adopted.

A very considerable advance in precision which has been demonstrated at Rothamsted, but has not as yet been so widely adopted abroad, is made possible by a factorial arrangement of treatments, so that if, for example, some plots receive phosphate and others none, these sets will be equal in number, and similar in the manurial and cultural contrasts within them. A large number such as 24 or 48 treatment combinations are thus tested simultaneously with comparatively little replication, the loss of which is made good by the inner or implicit replication, which the factorial arrangement makes possible. Thus though 48 such treatments may be replicated only 3 times, the whole information of 144 plots is available for every single contrast among the treatments tested, and equally for the differential effects of each treatment in the presence or absence of others. The increase in precision obtained by combining several different questions in the same experiment is due to the fact that, with a factorial arrangement, every plot contributes equally to answering each of them, whereas had 144 plots been distributed in 3 experiments for 3 separate questions there would have been only 48 plots available for each. A very important further advantage is gained by constructing large and complex experiments, beyond the gain of precision, namely that each question is examined in a considerable variety of subordinate circumstances, so that all results are given a much wider inductive basis than is possible with simple experiments. There have been examples of such factorial experiments at Rothamsted and Woburn since 1927.

A principle of undoubted value in the arrangement of field experiments, the practical possibilities of which are still being explored (10), consists in sacrificing information on interactions of subordinate interest, which it may often be confidently foreseen will be unimportant, by confounding them with soil heterogeneity, so eliminating a larger proportion of the latter from the more important comparisons. "Confounding" has been successfully employed in several experiments, and it is certainly capable of yielding for equal labour, a much needed increase in precision. It has, however, the real disadvantage, as has appeared on several occasions, that later workers, not realising the purpose and intentional limitations of the experiment, have been tempted to draw illegitimate conclusions involving the contrasts which have been

(9) O. Tedin (1931)— "The Influence of Systematic Plot Arrangement upon the Estmate of Error in Field Experiments," Journ. Agric. Sci., vol. XXI, pp. 191-208; F. Yates—"The Formation of Latin Squares for Use in Field Experiments," Empire Journ. Expt. Agric., vol. I, No. 3, 1933 pp. 235-244.

(10) F. Yates—" The Principles of Orthogonality and Confounding in Replicated Experiments," Journ. Agric. Sci., vol. XXIII, Part I, 1933, pp. 108-145.

deliberately set aside. With increasing knowledge of the principles of experimentation it will perhaps be possible to utilise its advantages more freely.

A great deal of attention has been given to experiments involving the particularly important and particularly intricate problems raised by residual effects, requiring repeated experimentation on the same land, and especially to evaluating such effects upon land under a normal agricultural rotation. It is possible to treat such experiments as replicated in time, and designed to eliminate errors due to temporal as well as local fluctuations. A four course rotation of 100 plots has been laid down to examine the availability of the nutrients in farmyard manure, in artificially rotted straw, and in straw rotted in in the ground, in addition to a comparison between superphosphate and rock phosphate, in the year of application and in subsequent years of the rotation. The experiment should also demonstrate whether or not the humus manures produce, relatively to artificial fertilisers, a gradual amelioration in the condition of the soil. A six-course rotation of 90 plots has been established, both at Woburn and at Rothamsted, with a view to assessing the seasonal fluctuations in the response of six crops to the three chief manurial nutrients. In this experiment all treatments progress continually over all the plots of the experiment, so that as time goes on, more and more of the soil heterogeneity is eliminated from the averages, and from other comparisons. A three-course rotation on a similar plan, involving both humus and green manures has now been laid down.

The department has been much concerned with the development of an adequate sampling technique, fit for studies in plant physiology, evaluation of damage due to insect infestation or plant disease, agricultural meteorology, yield determination, and the provision of qualitative samples for analysis (11). The key to this whole group of problems seems to lie in knowing how to sample the growing crop, with a precision known roughly in advance, and accurately determinable from the sample data. The official programme in agricultural meteorology has been much reduced, but sampling observations on wheat organised by the department are now carried out at eight centres (12), with the result that at these places at least there is someone who knows how to sample a crop in a reliable and comparable manner. The same principle has been later applied to the meteorological researches connected with forestry and horticulture. The method has been applied in a number of cases to experimental yields, and is especially useful where the plots have been sub-divided below the limit of economical harvesting by agricultural methods. Its other applications may be expected to develop as the possibilities

---

(11) A. R. Clapham—" The Estimation of Yield in Cereal Crops by Sampling Methods," Journ. Agric. Sci., vol. XIX, 1929, pp. 214-235 ; J. Wishart and A. R. Clapham—" A Study in Sampling Technique : The Effect of Artificial Fertilisers on the Yield of Potatoes," Journ. Agric. Sci., vol. XIX, 1929, pp. 600-618 ; A. R. Clapham—" Studies in Sampling Technique : Cereal Experiments : I. Field Technique," Journ. Agric. Sci., vol. XXI, 1931, pp. 367-371 ; T. W. Simpson—" Studies in Sampling Technique ; Cereal Experiments, II. A small-scale Threshing and Winnowing Machine," Journ. Agric. Sci., vol. XXI, 1931, pp. 372-375 ; A. R. Clapham—" Studies in Sampling Technique : Cereal Experiments. III. Results and Discussion," Journ. Agric. Sci., vol. XXI, 1931, pp. 376-390 ; R. J. Kalamkar—" A Study in Sampling Technique with Wheat," Journ. Agric. Sci., vol. XXII, 1932, pp. 783-792 ; F. R. Immer—" Study of Sampling Technique with Sugar Beets," Journ. Agric. Res. 44, 1932, pp. 633-647.

(12) Journ. Min. Agric., vol. XXXIX, No. 12 (March, 1933), pp. 1,082-1,084; vol. XL, No. 3 (June, 1933), pp. 206-208 ; vol. XL, No. 7 (October, 1933), pp. 591-593 ; vol. XL, No. 10 (January, 1934), pp. 903-906

of the method, and the procedure of its correct execution, become better known.

The use of statistical methods in the design of experiments is, of course, applicable in laboratory as well as in field experiments, and the field technique developed is applicable to other than manurial problems ; many voluntary workers are concerned with these other fields of work, at home or overseas. By applying statistical methods not only to the interpretation but also to the design of experiments it is not uncommon for the value of the experiment to be increased five or tenfold, a result which could not be obtained from improved methods of interpretation only, unless previous methods had been excessively inefficient.