

*R. A. Fisher*

DEPARTMENT OF GENETICS  
UNIVERSITY OF CAMBRIDGE

CONTRIBUTIONS TO  
MATHEMATICAL STATISTICS

INDEX PREPARED BY

*John Tukey*

PRINCETON UNIVERSITY

NEW YORK 1950

John Wiley & Sons, Inc.

Chapman & Hall, Ltd. LONDON

## EDITOR'S PREFACE

Like so many others, for more than a quarter of a century I have found inspiration and guidance in the original contributions of Professor R. A. Fisher. Some of my reprints of his early papers have been almost worn out through constant use, and for several years I have wished that I had a bound volume of his papers, particularly if he could be persuaded to preface each article with a brief note setting it in the proper perspective as of today.

After some persuasion, Professor Fisher agreed in the fall of 1946 to undertake the selection of what he now considers to be his most outstanding contributions and to provide brief introductory notes wherever he felt that such information would be of help to students of his work. The present book is the result. Typographical and other errors in the original papers have also been corrected by the author and some obscure passages have been clarified.

A word of explanation for the method of reproduction is perhaps in order. I had come to have a feeling not only for the content but also for the typographical form in which many of the articles first appeared in the journals of the many different learned societies. One of my principal reasons, therefore, for suggesting the photographic reproduction of these papers just as they appeared, even as to page size, was to preserve for posterity the flavor of the original papers as well as their content, both of which I had come to cherish. Fortunately this method of reproduction also reduced the cost of production so that this book could be put out at a price within the reach of the many students who will want to own a copy.

Then came the question of providing a thoroughgoing subject matter index that would make it possible for a student to look up a particular topic in Fisher's work. I am indebted to Professor John Tukey for the preparation of such an index.

No comprehensive survey of Fisher's work would be complete without a brief biographical note outlining the highlights in his development as a statistician whose work has carried him into so many fields. Just such a biographical note had been prepared by Professor P. C. Mahalanobis with the cooperation of Professor Fisher on another occasion and published in *Sankhyā*. With the gracious consent of Professor Mahalanobis, this note has been reproduced.

As editor, I wish to express to the editors of the many different scientific journals represented herein my appreciation of their courtesy in permitting us to reproduce the articles contained in this volume.

WALTER A. SHEWHART

April, 1950

## PREFACE

In mathematics, more than in other disciplines, original sources often tend to be neglected, and sometimes are scarcely accessible to students. In more leisured times the best method of making original papers accessible has been the voluminous "collected works" often published by learned societies in memory of their distinguished members. There is, I believe, a great deal to be said for the suggestion, which John Wiley & Sons, Inc., made to me, that editorial notes made by the author in his lifetime might do much to aid the student, especially in understanding the original purpose in relation to the common knowledge, or common misapprehensions, current at the time the paper was written.

I have tried, in annotating each paper, to recall these circumstances accurately, and to give cross references to related papers. The diversity of subject matter will be forgiven, I believe, by readers who realise that deep-seated misapprehensions can only be removed by clarification over a wide front, and from many points of view. Correct computational technique; accurate mathematical analysis; the development of criteria and abstract concepts related to aims and choice, in terms of which the analysis may be unified; and the beginnings of an operational inductive logic, have each required exposition and exemplification, sometimes not unattended by controversy. In each of these fields there is still much to be done. I am still too often confronted by problems, even in my own research, to which I cannot confidently offer a solution, ever to be tempted to imply that finality has been reached (or to take very seriously this claim when made by others!). The formal mathematical treatises on the subject, of which several good examples are available, cannot easily cope with the variety of aspects and interests, each of which may afford opportunities for far-reaching advances. Therein lies the particular importance of studying the points of view and the methodological approaches by which advances have been made in the past. In some cases the title is sufficiently explanatory, and no further note will be needed.

I have, of course, chosen for reprinting the papers most likely to be of service. A fuller bibliography has been published from time to time in the successive editions of the general introductory textbook, *Statistical Methods for Research Workers*.

R. A. FISHER

Cambridge, England  
April, 1950

## BIOGRAPHY

The following biography of Professor Fisher is reprinted from *Sankhyā*  
(Vol. 4, Pt. 2, pp. 265–272, December, 1938).

# PROFESSOR RONALD AYLMER FISHER

## EARLY DAYS.

Ronald Aylmer Fisher was born on the 17th of February, 1890, in East Finchley, one of the northern suburbs of London, where his father had a large establishment. He was the youngest of seven children which would have been eight if his twin brother had lived. His father, G. Fisher, of the well-known firm of auctioneers, Robinson and Fisher in King Street, St. James's, was a man of extraordinary energy and had a wide and detailed knowledge of the fine arts. The father's family were mostly businessmen; but an uncle, the younger brother of his father, was placed high as a Cambridge Wrangler and entered the Church. The mother's father was a successful London solicitor noted for his social qualities. Besides sturdy common sense and interest in practical affairs, taste for mathematics and the fine arts, Fisher inherited a spirit of adventure from his family. His mother's only brother threw up excellent prospects in London to collect wild animals in Africa, and one of his brothers returned from the Argentine to serve in the European War and was killed in 1915. Extremely defective vision saved Fisher himself from a similar fate, for, though in 1914 he had qualified himself for commission, he was refused throughout the war the opportunity of military service.

As with many mathematicians, Fisher's special ability showed at an early age. Before he was six, his mother was reading him a popular book on Astronomy, an interest which he followed eagerly during boyhood. Love of mathematics dominated his educational career; and he was fortunate in coming under the tuition of a brilliant mathematical teacher, W. N. Roe of Stanmore Park, also well-known in England as a Somersetshire cricketer. At Harrow he worked under C. H. P. Mayo and W. N. Roseveare. The peculiar circumstances of the teaching by the latter will be of interest to statisticians familiar with Fisher's geometrical methods. On account of his eyes he was forbidden to work by artificial light, and when he would go to work with Mr. Roseveare in the evenings the instruction was given purely by ear without the use of paper and pencil or any other visual aid. To this early training may be attributed Fisher's tendency to use hypergeometrical representation, and his great power in penetrating into problems requiring geometrical intuition. It is, perhaps, a consequence of this that throughout life his solutions have been singularly independent of symbolism. He does not usually attempt to write down the analysis until the problem is solved in his mind, and sometimes, he confesses, after the key to the solution has been forgotten.

## CAMBRIDGE.

Though born to sufficiently wealthy parents, owing to financial difficulties of the family at the time, it would have been difficult for him to proceed to the University without the aid of scholarships which he succeeded in winning. He joined the Gonville and Caius College, Cambridge, in 1909; and passed the Mathematical Tripos Part II in 1912 as a Wrangler with distinction in the optical papers in schedule B. He spent another year at Cambridge with a Studentship in Physics, and studied statistical mechanics and quantum theory under James Jeans, and the theory of errors under F. J. M. Stratton.

At this period he was consciously out of sympathy with certain tendencies prevalent at the time in the University. Under the influence of Bateson it was widely believed that recent work on genetics had discredited the evolutionary principles developed by Darwin. Fisher became keenly interested in Mendelian evidence, and was convinced that this furnished a basis for a quantitative and mathematical treatment of the theory of evolution. He began at this early date a series of studies of which the outside world knew nothing until the publication in 1930 of his book on "The Genetical Theory of Natural Selection."

The second influence at Cambridge with which Fisher found himself out of sympathy was the recent passage of control in mathematical teaching from the earlier tradition

of mathematical physicists to a school of pure mathematicians of largely continental derivation. The explicit statement of a rigorous argument interested him, but only on the important condition that such explicit demonstration of rigour was needed. Mechanical drill in the technique of rigorous statement was abhorrent to him, partly for its pedantry, and partly as an inhibition to the active use of the mind. He felt it was more important to think actively, even at the expense of occasional errors from which an alert intelligence would soon recover, than to proceed with perfect safety at a snail's pace along well-known paths with the aid of the most perfectly designed mechanical crutches. The real interpretation of his state of mind at this time is probably that he was fitting himself for mathematical research rather than for mathematical teaching. Fisher himself thinks that he was merely a very wilful and impatient youngman. This is no doubt true, but he was impatient not because he was young but because he was a creative genius.

#### EARLY WORK IN STATISTICS.

Although he had a successful academic career at Cambridge he had not yet found his vocation in life. I have noted before that for some time he studied mathematical physics. He also attended, it is said, one lecture on statistics under G. Udny Yule. He read at this time very carefully and very critically Karl Pearson's "Mathematical Contributions to the Theory of Evolution." Shortly before the War, in 1914, he first came into contact with Karl Pearson. H. E. Soper had discussed in a paper in the *Biometrika* the distribution of the coefficient of correlation in samples from an infinite, normally distributed, bi-variate population, but had succeeded in giving only approximate results by a tedious process. The problem, which was clearly stated by Soper, was one after Fisher's own heart, and in a week he had sent to Karl Pearson the exact solution but in a more hastily and roughly sketched form than even that in which it was subsequently published in the *Biometrika*.<sup>1</sup> This first paper, which was the starting point of the modern theory of exact sampling distributions, shows in a characteristic manner the working of Fisher's genius.

It was only after he had completed this important work that Fisher read the earlier papers of "Student",<sup>2</sup> and noticed how the representation in many-dimensional space so perfectly supplied the completion of "Student's" tentative ideas. A little later he solved the problem of the exact distribution of the intra-class correlation. This solution together with his  $z$ -transformation for coefficients of correlation of both kinds was published in 1921 in *Metron*.<sup>3</sup>

Fisher was now seriously interested in theoretical statistics, and any invitation to engage in academic work of a statistical character would have been no doubt welcome, but no such opportunity occurred for some considerable time. During the war he was at first engaged in statistical work in the office of the Mercantile and General Investment Company from 1913 to 1915, and from 1915 to 1919 in teaching physics and mathematics in public schools. It cannot be said that school teaching was congenial to him, but the few years spent in such work was far from wasted. Experience in teaching physics supplied what Fisher has always regarded as the most serious omission of his studies at Cambridge, namely, actual touch with problems of practical experimentation. On the whole therefore it was probably fortunate that the opportunity of independent statistical work came at a time when his outlook was more mature.

#### ROTHAMSTED EXPERIMENTAL STATION.

At the close of the War he was offered the post of the chief statistician under Karl Pearson at the Galton Laboratory. At about the same time Sir John Russell, Director of the Rothamsted Agricultural Station, offered him the post of statistician with opportunities for building up a Statistical Laboratory at Rothamsted. Neither post had much financial attraction, but he unhesitatingly accepted the Rothamsted offer as he thought that facilities for independent research would be greater there.

1. *Biometrika*, x, 507-521 (1915).

2. W. S. Gosset, whose untimely death in November, 1937, took away a great pioneer in statistics.

3. *Metron*, i, part 4, 1-32.

At Rothamsted Fisher found his real vocation for life, and there followed a series of years of intense activity. On the theoretical side he continued his work on sampling distributions and in 1921 completed the classical memoir on "Mathematical Foundations of Theoretical Statistics"<sup>18</sup> which was intended to supply the framework for modern statistical theory. On the applied side the method for separating the slow changes in time series was given in the same year,<sup>4</sup> and a little later in 1924, the important memoir on the influence of rainfall on the yield of wheat.<sup>5</sup> In 1923 was published the first paper on field trials<sup>6</sup> which led to a revolution in the technique of agricultural trials throughout the world, and was the starting point of the work on the design of experiments. The z-test of significance and the arithmetical procedure known as the 'Analysis of Variance' took their present form at about the same time.

"Statistical Methods for Research Workers" was first published in 1925, and has since then run to six editions. This book has probably done more than anything else to make research workers in most diverse fields of study familiar with the practical applications of modern statistical methods, and to create a statistical attitude of mind among the younger generation of scientists.

During the next few years we have a number of important papers on exact sampling distributions, the theory of estimations, the Chi-square measure of contingency, Bayes' theorem and inverse probability on the mathematical side, and the arrangement of field experiments on the applied side.

#### GENETICAL THEORY OF NATURAL SELECTION.

Fisher had been working on Mendelism and genetics for a long time and occasional papers had been published since 1918, but from about 1925 this subject began to engage increasing attention. The book on "Genetical Theory of Natural Selection" was published in 1930, and constituted a landmark in the history of the subject. A mathematical theory was developed on the basis of recent genetical researches to establish the principle of Natural Selection, on a more rigorous basis than Darwin had claimed, as the effective cause of evolutionary change. This principle is considered, not merely as a qualitative hypothesis in connexion with the classes of facts which it is capable of explaining, but rather in the manner of theoretical physics, as an agent which is bound to cause definite changes at a calculable rate. Factors recognized by Darwin, such as sexual selection, can then be examined on a quantitative rather than a speculative basis; and phenomena unknown to Darwin, such as genetical 'dominance' and the selection of fertility in human societies are found to illustrate unexpected effects of the evolutionary mechanism. Like Clerk-Maxwell who translated Faraday's concepts into mathematical language and developed the electromagnetic theory of light, Fisher gave a quantitative form to Darwin's views and built up a statistical theory of evolution.

#### ACADEMIC HONOURS.

The course of events in his life was in the meantime following a normal channel. He was married in 1917 to Ruth Eileen, daughter of H. Grattan Guinness, M.D., and several children were born to them. In 1920 he was elected a Fellow of Gonville and Caius College, Cambridge, and was awarded the Sc.D. degree of the same University in 1926. The influence of his work was in the meantime spreading rapidly, and he was elected a Fellow of the Royal Society in 1929, and was awarded the Weldon Medal in the same year. He was elected Honorary Fellow of the American Statistical Association in 1930, and Foreign Member of the American Academy of Arts and Science in 1934.

#### GALTON PROFESSOR IN LONDON.

Fisher was appointed Galton Professor in the University of London on the retirement of Karl Pearson in 1933. He has always been interested in eugenics, and has been

---

4. *Jour. Agri. Sc.* xi, 107-135.

5. *Phil. Trans. Roy. Soc. B.* cccxiii, 89-142.

6. *Jour. Agri. Sc.* xiii, 311-320.

for a long time associated with the Eugenics Society first as Honorary Secretary and later as Vice-President. Besides the importance of his work in genetics it was therefore peculiarly appropriate that he should succeed to the Chair created by an endowment from the great founder of the science of eugenics. He also took over from Karl Pearson the editorial charge of the *Annals of Eugenics* from 1933. During the last few years Fisher has devoted a great deal of attention to the application of statistical theory to genetic research and a large number of papers have been published in this subject.

The methods first developed in connexion with agricultural experiments have found in recent years increasing application in industry, medicine and public health, education and psychology, and scientific experiments of all kinds. In 1935 Fisher published a book on "The Design of Experiments" in which the theoretical principles were developed with a great variety of illustrative examples.

In 1936 Fisher visited the United States and received an honorary degree from the Harvard University on the occasion of its Tercentenary celebrations. Early in 1937 he accepted the Honorary Fellowship of the Indian Statistical Institute, and the invitation to preside over the first session of the Indian Statistical Conference in January, 1938.

#### MAIN CURRENTS OF FISHER'S WORK.

Fisher's work falls naturally into three main streams:—(a) contributions to the mathematical theory of statistics; (b) application of statistical theory to agriculture and the design of experiments; and (c) contributions to genetics.

On the mathematical side, for the first time, Fisher has supplied a unified, and general theory for drawing rigorous conclusions from statistical data. The logical and philosophical consequences of Fisher's theory still require to be fully worked out. Its importance in the theory of knowledge can be easily appreciated from the claim that it supplies the only rigorous logical foundation for all inductive inferences in science.

The theory of design of experiments is intended to supply an adequate technique for collecting the primary data in such a way that valid inferences may be drawn from them, and for extracting the maximum amount of information contained in the data in the most efficient way.

The object of the statistical theory of evolution is to supply a quantitative and mathematical basis for biology in general, and eugenics, the science of man, in particular.

#### THE THEORY OF SAMPLING DISTRIBUTIONS.

A brief account is given below of Fisher's work on the mathematical theory of statistics to indicate its importance in recent developments in statistics.

The idea of the random sampling distribution of statistics is fundamental in modern statistical theory. The problem of finding such distributions is, however, one of great mathematical difficulty and very little progress was made until Fisher started working on the subject. The earliest example of the modern type of distribution was that of the Chi-square found by Karl Pearson in 1900. Several years later "Student" gave the correct distribution of the sample variance and of his now famous t-statistic or the mean divided by its estimated standard deviation.

I have already mentioned however that Fisher was not acquainted with "Student's" work when he wrote his 1914 paper on the exact distribution of the correlation coefficient. Here he introduced for the first time the brilliant technique of representing a sample of size  $n$  by a point in a space of  $n$ -dimensions. Such representation has proved extremely useful in subsequent work not only in the theory of distribution, but also in other fields of statistical theory such as the work of J. Neyman and E. S. Pearson. The same representation was used by Fisher in finding the exact distribution of the mean error and the mean square error in 1921,<sup>7</sup> the regression coefficient in 1922,<sup>8</sup> the partial correlation coefficient in 1924,<sup>9</sup> and the coefficient of multiple correlation in 1928.<sup>10</sup>

7. *Metron*, i, Part 4, 1-32.

9. *Metron*, iii, 329-332.

8. *Jour. Roy. Stat. Soc.* lxxv, 597-612.

10. *Proc. Roy. Soc. A*, cxxi, 654-673.



In 1923<sup>11</sup> he gave a rigorous proof of "Student's" result for the t-statistic, and a little later showed how it could be used for testing various statistical hypotheses.<sup>12</sup> In 1924 Fisher generalized "Student's" t to the well-known z-statistic or half the logarithm of the ratio of two estimated variances based on different degrees of freedom, and was able, with the help of this distribution, to give a unified treatment of practically all the important distributions involved in testing null hypotheses.<sup>13</sup> The distribution of the discriminant function<sup>14</sup> recently introduced by him to test the significance of the difference, as a whole, of two sets of means of samples drawn from a normal multi-variate correlated population, belongs to the same general class.

The work on the limiting forms of frequency distributions of the largest or smallest member of a sample,<sup>15</sup> of the error of an interpolated value,<sup>16</sup> and the exact distributions which arise in connexion with various tests of significance in harmonic analysis<sup>17</sup> may also be mentioned here.

#### THEORY OF ESTIMATION AND STATISTICAL INFERENCE.

In science we are always faced with the problem of arguing from the particular to the general, or in statistical language, from the sample to the population. The task of statistical estimation is to find, on the basis of an observed sample, the values of the unknown parameters of the population from which the sample has been derived. Fierce controversy has raged over this subject ever since the publication in 1763 of Bayes' posthumous memoir "An essay towards solving a problem in the doctrine of chances" (*Phil. Trans.* liii, p. 370) in which he proposed to solve this problem with the help of the principle of equal distribution of ignorance. After more than one century and a half new light was shed on this problem by Fisher in his remarkable memoir "On the Mathematical Foundations of Theoretical Statistics."<sup>18</sup> This paper laid the foundations of statistical inference by emphasizing the importance of exact solutions of sampling problems, and by supplying the elements of a valid theory of estimation.

Any statistic<sup>19</sup> which estimates a certain parameter must of course satisfy the criterion of 'consistency', that is, its value must tend to the estimated parameter as the sample size is indefinitely increased. Fisher however gives a second criterion, namely that of 'efficiency', which requires that the variance of the estimating statistic (at least for large samples) should not exceed that of any other consistent statistic estimating the same parameter.

This idea had occurred in an earlier paper of 1920, but a great achievement of the 1921 memoir consisted in providing a machinery for calculating such efficient statistics, in the absence of which we would have been compelled to grope with different consistent statistics, and to compare their variances. This machinery is supplied by Fisher's now well-known method of 'maximum likelihood', a maximum likelihood statistic being always an efficient solution.

In the same memoir Fisher introduces for the first time the idea of a 'sufficient' statistic T, which is inevitably also efficient, and which incorporates the whole of the information available in the sample in regard to a given parameter, in the sense that if T' is any other statistic estimating the same parameter, the joint distribution of T and T' is such that, for a given value of T, the distribution of T' does not involve the parameter estimated. When a sufficient statistic exists, it can always be obtained by the method of maximum likelihood. Many interesting applications of these ideas were given in the same paper, such

11. *Proc. Camb. Phil. Soc.* xxi, 655-658.

12. *Proc. Int. Math. Congress, Toronto 1924*, 805-813.

13. *Econometrica*, iii, 353-365 (1935).

14. *Annals of Eugenics*, vii (1936), 179-188.

15. *Proc. Camb. Phil. Soc.* xxiv (1928), 180-190.

16. *Proc. Camb. Phil. Soc.* xxiii (1927), 912-921.

17. *Proc. Roy. Soc. A*, cxxv (1929), 54-59.

18. *Phil. Trans. Roy. Soc. A*, cccxxii, 309-368.

19. The word "statistic" in the singular number has been reserved by Fisher to denote a quantity which is capable of being expressed only in terms of actual observations in the sample. This concept was originally introduced by "Student", and has been given especial emphasis in Fisher's theory.

as errors of grouping, inefficiency of estimation of parameters by the method of moments in Pearson's system of frequency curves, and the estimation of the number of micro-organisms in a sample of water or soil.

The concept of 'information', first tentatively introduced in the 1921 paper, was more fully developed in a later paper<sup>20</sup> in 1925, enabling the theory of estimation to be freed from the large sample assumption. The concept of ancillary statistics, which serve to enhance the precision of estimation in the case when no sufficient statistic exists, was also introduced. In 1934 the general method of finding the distribution of any sufficient statistic was indicated,<sup>21</sup> and it was shown that in certain cases when no sufficient statistics exists, the whole of the information contained in the sample may be recovered by using as ancillary what may be termed the configuration of the sample. The important properties of 'information', as technically defined by Fisher, and 'the maximum likelihood estimate', were clearly summarized in a paper in the same year,<sup>22</sup> and a revised proof of the two main theorems in the theory of estimation, together with illustrative examples on the use of ancillary statistics, were given in a paper in 1935.<sup>23</sup>

Another important contribution of Fisher to the theory of statistical inference is his idea of 'fiducial probability' introduced in 1930, which is designed to cover the case when we do not want a single estimate of our unknown parameter, but require an interval in which our unknown parameter may be expected to lie.<sup>24</sup> Fisher has laid emphasis on the fact that the concept of fiducial probability, though entirely different from that of ordinary probability, is equally rigorous; and has brought out the importance of the fiducial concept in statistical theory in other papers.<sup>25</sup>

The Chi-square test was perhaps Karl Pearson's greatest single discovery; but Fisher has done much to make its use popular. In 1922 he showed that in the case of contingency tables, for which the margins of the expected table are reconstructed from those of the observed table, the distribution of Chi-square as given by Pearson's formula can be used validly if we take for  $n'$  a number exceeding the degrees of freedom by unity.<sup>26</sup> This result was discussed in other papers and fully corroborated by material derived from random sampling experiments.<sup>27</sup> Fisher also brought the Chi-square test in relation with his general theory of statistical inference, showing in particular that the method of minimising Chi-square agrees for large samples with the method of maximum likelihood.<sup>28</sup>

There are many other papers in mathematical statistics<sup>29</sup> among which the most important are probably those on the use of combinatorial methods for calculating moment coefficients,<sup>30</sup> and on the number of reduced Latin Squares of different orders.<sup>31</sup>

#### STATISTICAL THEORY IN AGRICULTURE.

Fisher's influence in India has been of the greatest importance in connexion with the application of statistical theory to agriculture. The basic principles of the new method are now well-known and need not be discussed in detail. In order to appreciate the revolutionary advance brought about by the introduction of the new technique, let us, however, consider for a moment the contrast between experiments of the old and the new type.

- 
- |  |  |
|--|--|
| 20. <i>Proc. Camb. Phil. Soc.</i> xxii, 700-725.   | 21. <i>Proc. Roy. Soc. A</i> , cxliv, 285-307. |
| 22. <i>Proc. Roy. Soc. A</i> , cxlvi, 1-8.   | 23. <i>Jour. Roy. Stat. Soc.</i> xcvi, 39-82.  |
| 24. <i>Proc. Camb. Phil. Soc.</i> xxvi, 528-535.   |  |
| 25. <i>Proc. Camb. Phil. Soc.</i> xxviii (1932), 257-261; <i>Proc. Roy. Soc. A</i> , cxxxix (1933) 343-348; <i>Annals Eugenics</i> , vi (1935), 391-398.   |  |
| 26. <i>Jour. Roy. Stat. Soc.</i> lxxv, 87-94.  |  |
| 27. <i>Economica</i> , iii (1923), 139-147; <i>Eugenics Review</i> , xviii (1926), 32-33.  |  |
| 28. Ref. nos. (18) and (12) and <i>Bologna. Atti del Congresso Int. dei Matematici</i> , vi (1928), 94-100.  |  |
| 29. Among other miscellaneous work may be mentioned the method for testing the randomness of a sequence ( <i>Quar. Jour. Met. Soc.</i> LII, 250, 1926), and the properties and applications of the integrals and derivatives of the normal error function ( <i>Brit. Assoc. Mathematical Tables</i> , Vol. 1). |  |
| 30. <i>Proc. Lond. Math. Soc.</i> XXX, 199-238 (1929); <i>Proc. Roy. Soc. A</i> , cxxx, 16-28 (1930). <i>Proc. Lond. Math. Soc.</i> xxxiii, 195-208 (1930).  |  |
| 31. <i>Proc. Camb. Phil. Soc.</i> xxx (1934), 492-507.   |  |

Suppose we wish to compare the yield of, say, six varieties or the effect on yield of six kinds of manures. In the old type of experiment the field would be divided into six plots, and a single plot would be allotted to each treatment. As Fisher explains, "the treatment giving the highest yield would of course appear to be the best, but no one could say whether the plot would not in fact have yielded as well under some or all of the other treatments."<sup>32</sup> It is known that within the same field wide differences exist in the fertility of the soil. Even when the soil fertility is uniform, there are innumerable other causes which affect the yield. How can we be sure that the observed differences in yield are due to the difference in the treatments, and not to soil heterogeneity? How can we be sure that they are not due to chance fluctuations? This is the basic problem. In order to solve it we must eliminate the effect of soil heterogeneity, and make an unbiased estimate of the magnitude of errors due to chance so that we may be sure that the observed effect is significant in comparison with the size of such chance errors.

Let us now see how Fisher solved the problem. Consider the same experimental field which had been originally divided into six portions. Fisher simply further subdivided each portion<sup>33</sup> into a number of plots of smaller size; and within each portion (or block as he called it) he assigned one plot to each treatment but strictly in a random manner. We have now the randomized block in its modern form. Using the principle of block division in two directions symmetrically we get the well-known Latin square.

The important point to be noticed is that the results will be now governed entirely by the laws of chance. There are innumerable causes which produce differences between the plots, and we know from the conditions of the experiment that it is impossible in practice to secure that the plots will be all alike. But the validity of the estimate of error is now guaranteed by the process of randomization, namely "the provision that any two plots, not in the same block, shall have the same probability of being treated alike, and the same probability of being treated differently in each of the ways in which this is possible." The calculus of probability and the apparatus of the statistical theory of sampling distributions can be now used with complete confidence. The logical foundations of scientific inference were thus made secure, and agricultural experiments were placed for the first time on the same footing as experiments in other sciences.

The second point to be observed is that by the technique of block division the problem of soil heterogeneity was solved at the same time. As each block contains all the treatments, differences between the total yields of the different blocks could be safely ascribed, apart from errors of sampling, to soil differences; and could be eliminated by suitable statistical methods. This of course led to a great improvement in the precision of the comparisons. When we remember that in particular experiments in India as much as ninety per cent. of the total variation is sometimes caused by soil differences, the importance of eliminating the effect due to such differences will be easily appreciated.

Replication, randomization and block division or local control are thus the fundamental principles of design introduced by Fisher. Replication is essential because it is the sole source of the estimate of error, while randomization is necessary to guarantee the validity of the estimate, that is, to ensure that the estimate will be unbiased. The purpose of block division is to increase the precision of the comparisons by elimination of soil differences, while replication is also useful in securing the same object by diminishing the experimental error. Finally, the analysis of variance gives a convenient and valid method of extracting the information contained in the observations. As Wishart has pointed out, the Fisherian technique "was something in the nature of a revolution," and altered the subsequent course of agricultural experiments throughout the world.

---

32. *The Design of Experiments*, 1935, p. 69.

33. I need scarcely add that the experimental field may be divided into any number of convenient portions each of which is further sub-divided into a number of plots.

## THE INTRODUCTION OF THE FISHERIAN TECHNIQUE IN INDIA.

The modern period of field experiments began in India with the foundation of the Imperial Council of Agricultural Research in 1929, which, from its very inception, laid great emphasis on the use of statistical methods, created a statistical section at its headquarters, and gave a grant to the Statistical Laboratory, Calcutta, for work in this connexion. The earliest analysis of an experiment of the Latin square type was published from the Laboratory in 1931. The first complex (variety-manurial) experiment in 1932, the first split-plot experiment in 1933, and the earliest experiment using the principle of 'confounding' in 1935 were all designed and introduced from the same Laboratory. The use of the new technique has spread rapidly, and no important experiment in India is now laid out on a design of the old type. The ultimate beneficial effect of this movement on Indian agriculture is difficult to estimate or exaggerate.

### GENETIC STUDIES.

Fisher's method of maximum likelihood is now well-known and is being increasingly adopted in genetical work in India. More important than these contributions to 'formal questions' as Fisher himself calls it are his contributions to the problems of quantitative characters which depend on a large number of mendelian factors of small individual effects. Almost all economic characters in which the animal or plant breeder is interested are of this nature and here the usual genetic analysis completely breaks down. If, therefore, genetics is to exert its influence on breeding, geneticists and breeders must concentrate on the study of quantitative characters. Fisher has shown how statistics of the second and third degree can be utilised for this purpose, and has given methods for the estimation of various genetic quantities from which genetic variances can be calculated and the effects of dominance and environmental causes eliminated.

A knowledge of genetic variance and of the possible number of factors involved in a character is of fundamental importance to the breeder in assessing the selection potentiality of his material. Fisher has shown the way to attain this knowledge. The experimental study, however, requires a sound lay out and elaborate recording of observational data to make statistical analysis possible. Work has recently been started on cotton at the Institute of Plant Industry at Indore on the new lines, and it may be confidently asserted that the influence of Fisher will ere long become increasingly important in this field of study also.

### THE INFLUENCE OF FISHER ON STATISTICAL WORK IN INDIA.

In recent developments in statistics in India the most powerful influence has been that of R. A. Fisher. The movement for the study of analytic statistics started in this country at a time when Fisher's work was just coming into prominence, and the younger workers have been all brought up in what may be called the Fisherian tradition. The methods developed by him are being used not only in agriculture, but also in all kinds of investigations of importance to national welfare. On the theoretical side, what little is being done in India is based largely on his work. He has been the accepted and acknowledged leader for a long time, but his presence here as the first President of the Indian Statistical Conference has established personal relations which are highly cherished by all statistical workers in India.\*

P. C. MAHALANOBIS.

CALCUTTA,  
6th January, 1938.

---

\* Written on the occasion of the opening of the First Session of the Indian Statistical Conference presided over by Prof. R. A. Fisher in Calcutta on the 7th January, 1938.

## CONTENTS

- Paper 1.* On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample
- Paper 2.* A Mathematical Examination of the Methods of Determining the Accuracy of an Observation by the Mean Error, and by the Mean Square Error
- Paper 3.* Studies in Crop Variation. I. An Examination of the Yield of Dressed Grain from Broadbalk
- Paper 4.* The Accuracy of the Plating Method of Estimating the Density of Bacterial Populations
- Paper 5.* On the Interpretation of Chi Square from Contingency Tables, and the Calculation of  $P$
- Paper 6.* The Goodness of Fit of Regression Formulae and the Distribution of Regression Coefficients
- Paper 7.* Statistical Tests of Agreement between Observation and Hypothesis
- Paper 8.* The Conditions under which Chi Square Measures the Discrepancy between Observation and Hypothesis
- Paper 9.* On a Property Connecting the Chi-Square Measure of Discrepancy with the Method of Maximum Likelihood
- Paper 10.* On the Mathematical Foundations of Theoretical Statistics
- Paper 11.* Theory of Statistical Estimation
- Paper 12.* On a Distribution Yielding the Error Functions of Several Well Known Statistics
- Paper 13.* The Mathematical Distributions Used in the Common Tests of Significance
- Paper 14.* The General Sampling Distribution of the Multiple Correlation Coefficient
- Paper 15.* Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample
- Paper 16.* Tests of Significance in Harmonic Analysis
- Paper 17.* The Arrangement of Field Experiments
- Paper 18.* Studies in Crop Variation. VI. Experiments on the Response of the Potato to Potash and Nitrogen
- Paper 19.* The Distribution of Gene Ratios for Rare Mutations
- Paper 20.* Moments and Product Moments of Sampling Distributions
- Paper 21.* The Moments of the Distribution for Normal Samples of Measures of Departure from Normality
- Paper 22.* Inverse Probability
- Paper 23.* 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

- Paper 24.* Two New Properties of Mathematical Likelihood
- Paper 25.* The Fiducial Argument in Statistical Inference
- Paper 26.* The Logic of Inductive Inference
- Paper 27.* Uncertain Inference
- Paper 28.* A Test of the Supposed Precision of Systematic Arrangements
- Paper 29.* Professor Karl Pearson and the Method of Moments
- Paper 30.* Moments and Cumulants in the Specification of Distributions
- Paper 31.* The Wave of Advance of Advantageous Genes
- Paper 32.* The Use of Multiple Measurements in Taxonomic Problems
- Paper 33.* The Statistical Utilization of Multiple Measurements
- Paper 34.* The Precision of Discriminant Functions
- Paper 35.* The Comparison of Samples with Possibly Unequal Variances
- Paper 36.* The Sampling Distribution of Some Statistics Obtained from Non-linear Equations
- Paper 37.* On the Similarity of the Distributions Found for the Test of Significance in Harmonic Analysis, and in Stevens's Problem in Geometrical Probability
- Paper 38.* The Negative Binomial Distribution
- Paper 39.* The Theory of Confounding in Factorial Experiments in Relation to the Theory of Groups
- Paper 40.* A System of Confounding for Factors with More than Two Alternatives, Giving Completely Orthogonal Cubes and Higher Powers
- Paper 41.* Some Combinatorial Theorems and Enumerations Connected with the Numbers of Diagonal Types of a Latin Square
- Paper 42.* The Likelihood Solution of a Problem in Compounded Probabilities
- Paper 43.* The Relation between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population

## I

## ON THE "PROBABLE ERROR" OF A COEFFICIENT OF CORRELATION DEDUCED FROM A SMALL SAMPLE \*

## AUTHOR'S NOTE

This is the second of three papers dealing with the sampling errors of correlation coefficients covering the cases (i) "The frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population," *Biometrika*, Vol. 10, pp. 507-521, 1915.

Here the method of defining the sample by the coordinates of a point in Euclidean hyperspace was introduced, and it was shown that the exact sampling distribution could be obtained. The practical application of these results appears in the second paper in 1921, here referred to. It was concerned primarily to show that the sampling distribution was different for intraclass and interclass correlations, and to give the exact solution for the former, comparable with that given for the latter in 1915. The special simplicity of the solution in the intraclass case was one of the foundations of the recognition of the  $z$ -distribution, and the analysis of variance, in terms of which it would now be treated.

The third of these papers appeared in 1924, also in *Metron*: (iii) "The distribution of the partial correlation coefficient," *Metron*, Vol. 3, pp. 329-332, 1924.

It shows that the effect of the elimination of variates by partial correlation is simply to reduce the effective size of the sample by unity for each independent variate eliminated.

This group of three papers is part of a larger series appearing from 1915 to 1928, in which exact solutions were found for a variety of problems of distribution, and the corresponding tests of significance developed. Although many of these problems had been approached

\* This paper appeared in *Metron*, Vol. 1, No. 4, pp. 3-32, 1921. The established policy of *Metron* does not permit the republication of articles from that journal.

## 1.2b

using statistics called correlation coefficients (e.g., Biserial  $r$ , Biserial  $\eta$ , etc.), yet it appears that the term was widely misapplied, and the problems themselves are now simply treated as comparisons of means, tests of heterogeneity, or regression problems.

In the final section this paper contains a discussion of the bearing of the new exact solutions of distribution problems on the nature of inductive inference. This is of interest for comparison with the paper on "Inverse probability," published 1930, nine years later. In 1930 the notion of fiducial probability is first introduced using as example the distribution found in this paper. In view of this later development the statement, "We can know nothing of the probability of hypotheses or hypothetical quantities," is seen to be hasty and erroneous, in the light of the different type of argument later developed. It will be understood, however, as referring only to the Bayesian probabilities *a posteriori*.



# A MATHEMATICAL EXAMINATION OF THE METHODS OF DETERMINING THE ACCURACY OF AN OBSERVATION BY THE MEAN ERROR, AND BY THE MEAN SQUARE ERROR \*

## AUTHOR'S NOTE

This early paper arose from an examination of a statement by A. S. Eddington in his book, *Stellar Movements*, page 147. It concerns the relative precision of two estimates of the variance of a normal distribution: (a) using Bessel's formula, based on the mean square; and (b) using Peter's formula, using the mean deviation.

The results include the exact sampling distribution of Bessel's estimate and the mean square of Peter's. The variance of the latter is, in large samples, the larger in the ratio  $(\pi-2)$ .

Next is considered the class of estimates based on powers of the deviation in general, showing that the precision is maximised when  $p = 2$ , the variance being 14 per cent greater for  $p = 1$  and 9 per cent greater for  $p = 3$ . For continuous variation of  $p$  the precision of the mean square is a true maximum. These results had all been obtained before, without the knowledge of the author or of Eddington.

The most important point of the paper is the consideration of the simultaneous distribution of two estimates. This is examined in detail for the case of four observations, but the more general point is established that for a given value of  $\sigma_2$  the distribution of  $\sigma_1$  is independent of  $\sigma$ . Consequently when  $\sigma_2$ , the estimate based on the mean square, is known, a value of  $\sigma_1$ , the estimate based on the mean deviation, gives no additional information as to the true value. It is shown that the same proposition is true if any other estimate is substituted for  $\sigma_1$ , and consequently that the whole of the information respecting the variance which a sample provides is summed up in the single estimate  $\sigma_2$ . I believe this is the first occasion on which attention has been called to this property characteristic of a *sufficient* estimate.

\* Reprinted from *Monthly Notices of the Royal Astronomical Society*, Vol. LXXX, No. 8, pp. 758-770, 1920.