

Thank you for using eradoc, a platform to publish electronic copies of the Rothamsted Documents. Your requested document has been scanned from original documents. If you find this document is not readable, or you suspect there are some problems, please let us know and we will correct that.



ROTHAMSTED
RESEARCH

Report for 1927-28

[Full Table of Content](#)



Statistical Methods and Results

Rothamsted Research

Rothamsted Research (1928) *Statistical Methods and Results* ; Report For 1927-28, pp 58 - 64 -
DOI: <https://doi.org/10.23637/ERADOC-1-85>

This increase is quantitatively sufficient to account for the increased yield in grain and straw, without postulating the aid of any stimulus to plant growth.

- X. T. EDEN AND R. A. FISHER. "*Studies in Crop Variation, IV. The Experimental Determination of the Value of Top Dressings with Cereals.*" *Journal of Agricultural Science*, 1927. Vol. XVII., pp. 548-567.

A simple account of a top dressing experiment carried out at Rothamsted in 1926, with especial reference to the design of such experiments in general, to the statistical analysis of the data, and to the precision attained. The experiment was of 96 plots of winter oats and designed to test with precision the return from top dressings applied early or late, in single or double quantities, and using sulphate or muriate of ammonia. All possible combinations of these conditions were used, the whole having eight-fold replication. The results possessed a higher level of precision than on any previously attained in conditions which allow of a valid estimate of error, the standard error of each comparison being only 1.4 per cent. It is, then, of an order which allows discussion of the monetary return to the industrial farmer in relation to the cost of manure and labour. The experiment is one of a programme of research into top dressings, which it is hoped can be maintained at the same level of precision.

- XI. BHAI BALMUKAND. *Studies in Crop Variation, V. "The Relation between Yield and Soil Nutrients."* *Journal of Agricultural Science*, 1928. Vol. XVIII., pp. 602-627.

It is shown (a) that it is possible to fit Maskell's Resistance formula (in which the reciprocal of the yield is expressed as the sum of terms each dependent on a specific manurial factor) to experimental data involving the simultaneous variation of two numerical factors by a sufficiently rapid process of approximation, (b) that in every case discussed the formula fits the facts within the limits of experimental error estimated from the experiments themselves although formulæ of other types fail strikingly to do so, (c) that the parameters appropriate to each nutrient are therefore independent of other conditions and are capable of direct physical interpretation. The interpretation suggested supplies a direct measure of the quantity of each soil nutrient actually available to the plant, and of its specific importance in determining yield.

II.—STATISTICAL METHODS AND RESULTS.

(Statistical Department.)

- XII. R. A. FISHER. "*The General Sampling Distribution of the Multiple Correlation Coefficient.*" *Proceedings of the Royal Society (A)*, 1928. Vol. 121, pp. 654-673.

By an appropriate linear transformation of the independent variates it may be shown that the sampling distribution of the multiple correlation coefficient does not depend on the whole

matrix of correlations between these variates, but solely upon the multiple correlation in the population sampled.

The actual distribution (A) may then be easily obtained by similar methods to those by which the distribution of the simple correlation coefficient has been obtained.

The frequency function involves a hypergeometric function of P^2R^2 which is a rational function when n_1 and n_2 are both even, algebraic when n_2 only is even, and reducible to circular functions when n_1 and n_2 are both odd.

The case of large samples yields a series of distributions (B) of great interest, involving Bessel functions, which connect the X^2 distributions with the Gaussian, and are intimately related to a double Poisson summation. Owing to the practical importance of this limiting form, a table of its 5 per cent. points is given up to seven independent variates.

When n_2 is even, the probability integral of the general distribution is expressible in finite terms which are developed in Section 6.

The (B) distribution of Section 5 replaces the X^2 distribution in the analysis of variance if the squares summed are non-central. An analysis of variance so extended leads to a third group of distributions (C), closely related to (A), and tending like it to a common limit (B). The distinction between (A) and (C) arises from the fact that in cases proper to the multiple correlation the central displacements will vary from sample to sample owing to variations in the second order moment coefficients of the independent variates, and for such cases (A) is the correct distribution. The type (C), however, is of frequent occurrence owing to the absence or irrelevance of such variation.

XIII. R. A. FISHER. "*On a Distribution Yielding the Error Functions of several Well-known Statistics.*" Proceedings of the International Mathematical Congress, Toronto, 1924, pp. 805-813.

When the exact sampling distributions of a number of the statistics in most general use came to be worked out, it appeared that apart from avoidable differences in notation, nearly all were examples of three related types of distribution. Two of these had been previously found and their numerical values made in part available. In 1900, Pearson had established the distribution of his measure of discrepancy X^2 used in testing goodness of fit. Although as given this distribution was incorrect, the correction, so long as efficient methods of fitting are used, does not change the form, but only the particular member of the series to be employed. An identical distribution was subsequently found by "Student" for the variance as estimated from a normal sample, and by the author for the index dispersion for the Poisson and Binomial series. The distribution of t first found by "Student" in studying the mean of a unique sample, is also exact over a much wider range, and gives the distribution also of regression coefficients of all orders. The third distribution, that of z , which may be regarded as a generalisation either of that of X^2 or of t , completes this group of theoretical distributions and supplies the solution for intraclass correlations, the goodness of fit of regression

formulæ, the comparison of variances and the significance of multiple correlation and correlation ratios.

- XIV. R. A. FISHER AND J. WISHART. "*On the Distribution of the Error of an Interpolated Value, and on the Construction of Tables.*" Proceedings of the Cambridge Philosophical Society, 1927. Vol. XXIII., pp. 912-921.

The development of simple interpolation formulæ involving only even differences, has favoured the increased use of formulæ of high order; while the theoretical study of the remainder term makes it possible to design tables for which such formulæ are rigorously valid. This paper develops the theory of the distribution of the error of interpolated values, and shows that these will always have a higher precision than the tabular entries, and if the errors of the latter are normally distributed and independent, those of the former will be normally distributed also, with a variance which for high order formulæ tends to equality. This will not be the case for tables "correct to the nearest figure" in which the error distributions of the tabular entries will be rectangular, and those of the interpolate will have a complicated distribution. The advantages both in convenience and in precision of not cutting down tables as originally calculated, so as to be correct to the nearest figure are therefore to be considered in the publication of tables.

- XV. J. WISHART. "*On Errors in the Multiple Correlation Coefficient due to Random Sampling.*" Memoirs of the Royal Meteorological Society, 1928. Vol. II., No. 13, pp. 29-37.

The use of the multiple correlation in meteorological and agricultural problems is common where the effect of a number of independently varying factors on, say, the weather of a particular locality, is investigated. An experimental study is made of the kind of values, with their frequency of occurrence, which would arise from chance factors which had in reality no influence on the phenomenon studied. The mathematical theory for this particular case is now complete, and it is shown how the probability of occurrence of any value can be calculated. This method of testing the significance is recommended in place of the more usual probable error, for the distribution is far from normal.

- XVI. J. WISHART. "*Table of Significant Values of the Multiple Correlation Coefficient.*" Quarterly Journal of the Royal Meteorological Society, 1928. Vol. LIV., pp. 258-259.

This table gives the values of the multiple correlation coefficient that would occur in random sampling from uncorrelated material for the 1 and 5 per cent. levels of significance, and is based on the theory outlined in the preceding memoir. It is carried as far as six independent variates and up to samples of 100.

- XVII. J. WISHART. "*Le traitement correct des Problèmes de Corrélation Multiple en Météorologie et Agriculture.*" Report of the Association Française pour l'avancement des Sciences, 1928.

The use of the multiple correlation coefficient and its testing by the preceding significance tables is explained, and examples, taken from the weather forecasting work of Sir Gilbert Walker, are worked out. Comment is made on the recent correspondence in "Nature" on the use of the probable error in correlational work.

- XVIII. J. WISHART. "*The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population.*" *Biometrika*, 1928. Vol. XXA., pp. 32-52.

The distribution of the variance was first given by "Student" in 1908. The next advance was in 1915, when R. A. Fisher, for a two-variate population, gave the simultaneous distribution of the three second order moment coefficients, namely the two variances and the cross product moment (or co-variance). In this paper the problem is generalised to include any number of variates, and the multiple distribution of all second order product moment coefficients is deduced. A table follows giving the moment coefficients of this distribution, as far as the fourth order and eight variates.

- XIX. J. WISHART. "*A Problem in Combinatorial Analysis giving the Distribution of Certain Moment Statistics.*" *Proceedings of the London Mathematical Society*, 1929. Series 2, Vol. XXIX., pp. 309-321.

The method outlined in the previous paper for deducing the moments of the distribution from a particular generating function is tedious, and it is here shown that it is, after all, only a particular case of a more general problem whose solution can be reached through the theory of combinatorial analysis. The correspondence between the theories is indicated, and the special problem, which can be considered as a ring arrangement of rods, is worked out in full. Finally, an operational solution is demonstrated, and the arithmetical procedure for building up any required result is illustrated in an example.

- XX. J. WISHART. "*Sampling Errors in the Theory of Two Factors.*" *British Journal of Psychology*, 1928. Vol. XIX., Part 2, pp. 180-187.

It is a mathematical consequence of the theory that any ability can be resolved into two factors, one general and the other specific, that the tetrad difference of correlation coefficients between any four abilities should vanish, within the limits of random sampling error. A modified definition of the tetrad is introduced in order that the distribution reached in Paper No. XVIII. should be capable of application to this problem. An exact formula is then deduced for the standard error of the tetrad, and this is applied to some published results of psychological experiments.

- XXI. T. N. HOBLYN. "*A Statistical Analysis of the Daily Observations of the Maximum and Minimum Thermometers at Rothamsted.*" Quarterly Journal of the Royal Meteorological Society, 1928. Vol. LIV., pp. 183-202.

Daily records of maximum and minimum temperature at Rothamsted are available for 49 years. This paper gives the means, variances and covariances for each month, and analyses the variance and covariance into portions ascribable to variation from day to day and from year to year.

- XXII. R. A. FISHER AND T. N. HOBLYN. "*Maximum and Minimum-correlation Tables in Comparative Climatology.*" Geografiska Annaler, 1928. Vol. III., pp. 267-281.

In connection with Paper No. XXI. on Maximum and Minimum Temperature Variations at Rothamsted, the opportunity was taken to prepare and publish for the purposes of comparative climatology two-way tables of these variates for each month of the year for the first 49 years of experience at Rothamsted. The tables are supplemented by Analyses of Variance in which the year to year variation is distinguished from the day to day variation within the same year.

- XXIII. R. A. FISHER AND BHAI BALMUKAND. "*The Estimation of Linkage from the Offspring of Selfed Heterozygotes.*" Journal of Genetics, 1928. Vol. XX., pp. 79-92.

Five methods of solution are given of the statistical problem presented by typical linkage data. The example chosen shows the various errors into which the use of inefficient statistics leads. Of the efficient methods, the method of maximum likelihood possesses the advantage that it may be applied directly to any analogous problem, and is related in a previously unsuspected way to the measure of discrepancy χ^2 . The product ratio method, for using which a table is provided, enjoys the practical advantages of other efficient solutions, and is in addition unaffected by differential viability, if this is caused by one factor only. The method of minimum χ^2 , unlike the other two, is laborious in computation and seems to possess no special theoretical interest.

- XXIV. R. A. FISHER. "*The Possible Modifications of the Response of the Wild Type to Recurrent Mutations.*" American Naturalist, 1928. Vol. LXII., pp. 115-126.

The reaction of the wild type to mutations is known in many cases to be capable of a somewhat rapid modification in experimental conditions, by the selection through differential viability of factors capable of modifying this response.

It may be calculated that with mutation rates of the order of one in a million the corresponding selection in the state of nature, though extremely slow, cannot safely be neglected in the case of the heterozygotes.

The observed behaviour of multiple allelomorphs largely

supports, though that of specific modifiers seems to oppose, the view that complete dominance generally may be regarded as a product of such selective modification.

XXV. R. A. FISHER. "*Two Further Notes on the Origin of Dominance.*" *The American Naturalist*, 1928. Vol. LXII., pp. 571-574.

In connection with the previous paper (No. XXIV) the evidence of the behaviour of the cotton mutant, *Crinkled Dwarf*, investigated by Dr. S. C. Harland, is cited as demonstrating the natural evolution of dominance. This mutant is a clear recessive in the Sea Island cottons in which it occurs, but on crossing with other species the dominance of the wild type is found to be conditioned by a group of probably numerous modifiers in which the Sea Island has come to differ from other cottons.

It is suggested that the anomalous dominance of several breed characteristics in domestic poultry may be explained by the incidence of selection in early stages of domestication when the domestic flocks were frequently sired by wild jungle fowls.

XXVI. R. A. FISHER AND L. H. C. TIPPETT. "*Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample.*" *Proceedings of the Cambridge Philosophical Society*, 1928. Vol. XXIV., pp. 180-190.

The distribution of the greatest or least of a sample of n may be derived from that of the population sampled. If it tends to any limiting form as the size of the sample is increased, its distribution must obey a functional relation the solution of which is here given, with a discussion of the criteria which determine the limiting form, and of the gradual approach to the limit shown in normal samples.

XXVII. R. A. FISHER AND E. B. FORD. "*The Variability of Species in the Lepidoptera, with Reference to Abundance and Sex.*" *Transactions of the Entomological Society of London*, 1929. Vol. LXXVI., pp. 367-384.

The frequency distribution of depth of pigment in the ground-colour of the fore-wings of 35 species of British moths has been obtained by comparison of over 5,000 specimens with a standard colour scale.

For comparison of variabilities of groups of different average tint the standard deviations have been adjusted to eliminate any arbitrary elements which might have been introduced by the scale employed.

The mean tint is darker in the females than in the males, and is also darker in the more abundant than in the less abundant species.

Even after adjustment the mean variance is about 30 per cent. higher in females than in males, and is in both sexes greatest in the abundant species, and at least in those which are less than common.

It is also possible, though the difference is not in this material statistically significant, that the species with wider range are, in any one locality, the more variable.

The association of variability with abundance accords with an early generalisation of Darwin's, and with the theory that variability is determined by a balance between the influences of mutations and selection. This theory is insufficient numerically to account for the large differences in variability between the sexes.

In view of the frequency of polymorphism, and other marked variations, in the females as opposed to the males in Lepidoptera, it is suggested that the male sex hormones may inhibit the action of a number of the factors influencing the development of pigment, as in the well-known sex-controlled variation. The suggestion of Goldschmidt that there exist pigmentation factors in the Y chromosome capable of interaction with outosomal factors to cause pigmentary differentiation is an alternative view which may account for a few cases. This should result in purely female unisexual polymorphism (except for the possibility of occasional crossing-over between the X and Y chromosomes), but it is almost certainly an infrequent phenomenon. It is possible that sexual selection may, in part, be responsible for the complete inhibition of mimetic patterns in the males of certain mimetic species.

XXVIII. R. A. FISHER. "*Triplet Children in Great Britain and Ireland.*" Proceedings of the Royal Society (B), 1927. Vol. 102, pp. 286-311.

Measurements taken at a fixed age of 115 surviving triplet children, are reported upon in respect of the average growth attained, which is not appreciably different from that of children by single births; of the degree of resemblance between pairs of like and unlike sex, which confirm in entirely independent material the conclusion drawn from Lauterbach's measurements of twins; and of the inheritance of the twinning tendency which, in opposition to the view developed by Weinberg, indicates inheritance of diembryony on the paternal side.

XXIX. R. A. FISHER. "*On Some Objections to Mimicry Theory, Statistical and Genetic.*" Transactions of the Entomological Society of London, 1927. Vol. LXXV., pp. 269-278.

The statistical reasoning which led Marshall to dispute the applicability of Müller's theory to the mimetic approach of a more numerous to a less numerous form, is shown to be unsound, and the validity of Müller's argument is verified. The contention of Punnett that in certain cases mimetic forms must have arisen by saltations falls with Marshall's argument on which it is based. The more recent study of modifying factors shows that the Mendelian inheritance observed in polymorphic mimics does not show that these forms were not gradually evolved by natural selection; while the stability of the gene ratio of these factors implies selective action.