Rothamsted Report for 1932

Statistical Methods and Results

Rothamsted Research

than the nitrate level. There was a very rapid disappearance of added nitrate and a rather less rapid removal of added ammonia. The rate of removal of the latter, however, was such as to suggest that some of the ammonia might be taken up directly without nitrification. This was made practically certain by a study of the rate of nitrification, which was very low in certain soils from which ammonia was rapidly removed.

The equilibrium between ammonia and nitrate production in these soils is discussed.

xv. J. G. SHRIKHANDE. " *The Degree of Humification in Manures Measured by the use of Hydrogen Peroxide.*" Soil Science, 1933, Vol. XXXV, pp. 221-228.

It is known that humified organic matter can be distinguished from non-humified by the action of hydrogen peroxide. The action of three per cent. hydrogen peroxide has been used for measuring the degree of humification which appears to be a useful measure of the decomposition undergone by any one kind of plant material under different treatments. The loss after extraction with peroxide is not an infallible guide to the value of organic manures in general. A comparison has also been made between the extractive properties of water and peroxide.

STATISTICAL METHODS AND RESULTS

(Department of Statistics)

(*a*) MATHEMATICAL THEORY

XVI. R. A. FISHER. " *Inverse Probability and the Use of Likelihood.*" Proceedings of the Cambridge Philosophical Society, 1932, Vol. XXVIII, pp. 257-261.

An explanation of the distinction between these two methods of reasoning from experience, with a correction of some allusions to likelihood in which they are confused.

XVII. R. A. FISHER. " *The Concepts of Inverse Probability and Fiducial Probability Referring to Unknown Parameters.*" Proceedings of the Royal Society, A, 1933, Vol. CXXXIX p.p. 343-348.

The argument of Jeffreys in favour of a particular frequency distribution *a priori* for the precision constant of a normally distributed variate rests on the fallacy that the probability of the last of three observations, lying between the previous two, should be one-third, *irrespective of the distance apart of the two previous observations*.

The apparent simplicity of the results of assuming this particular distribution *a priori* rests on the fact that the *inverse* and the *fiducial* probability statements about the unknown parameter are thereby made to coincide, though logically they are entirely distinct. This particular distribution *a priori* is, however, not only hypothetical but unacceptable as such, since it implies that all ranges of values of the parameter covering finite ratios, however great, are infinitely improbable.

F

74

XVIII. F. YATES. " *The Principles of Orthogonality and Con-founding in Replicated Experiments.*" Journal of Agricultural Science, **1933**, Vol. XXIII, pp. 108-145.

The procedure of confounding certain treatment effects, either direct effects or high order interactions, with fertility differences, has been utilised in certain agricultural field trials of the complex type at Rothamsted and elsewhere for some years. The present paper describes the principles underlying this procedure, and the appropriate methods of analysis, and draws attention to certain dangers which must be guarded against when designing or analysing experiments of this type.

The discussion of confounding necessarily involves the consideration of the independence, or *orthogonality*, of the various sets of degrees of freedom appearing in the analysis of variance. The concept of orthogonality is therefore discussed, and the modifications which are necessary in the ordinary procedure of the analysis of variance when dealing with non-orthogonal data are illustrated by application to an example where there is a double classification (in this case classification by sex and treatments in a poultry feeding trial) and where the numbers in the various sub-classes are unequal, treatments not being equalised for sex, or sex for treatments. This type of data is of frequent occurrence where observation rather than planned experiment is the source of information. In addition to the accepted method of fitting constants by the method of least squares, various shorter methods of analysis are described, and their validity and applicability discussed.

XIX. F. R. IMMER. " *The Efficiency of the Correlation Coefficient for Estimating Linkage Intensities.*" American Naturalist, **1931**, Vol. LXV, pp. **567-572**.

Takezaki and Owen had independently derived a method of estimating linkage by means of the correlation coeffcient.

Takezaki derived a formula for the standard error of his estimate of p from the assumption that the standard error of r, obtained from the fourfold table, could be equated to the standard error of a correlation coefficient derived from a normal frequency surface having the same number of observations. This mistaken assumption has led to the precision of this method of estimating linkages being greatly over-estimated.

It is found that the curve for the actual efficiency of the correlation method calculated from the correct formula, does not exceed 100 per cent. for any possible values of p, from 0 to 1, in accordance with the general theory. The correlation method is fairly efficient in the coupling phase, and for loose linkage in repulsion. For close linkage in repulsion it is not efficient. Since there are other formulæ such as the maximum likelihood method, and the product ratio method, which are efficient for all values of p, it would seem preferable to use these formulae in most cases.

XX. R. S. KOSHAL. " *Application of the Method of Maximum Likelihood to the Improvement of Curves Fitted by the Method of Moments.*" Journal of the Royal Statistical Society, **1933**, Vol. XCVI, pp. **303-313**.

A method is given for the improvement of inefficient statistics obtained by the method of moments. It consists in the evaluation of

a number of L's directly from the equation $L = S$ ($n_s \log p$). For the estimation of S parameters it requires the calculation of $\frac{1}{2}(S+1)(S+2)$ values of L. These values of L provide simultaneous equations from which the corrections to be added to the moment estimates of the parameters can be calculated. The method is illustrated by its application to a coarsely-grouped skew distribution to which Pearson's Type I was fitted by the method of moments. It is shown that the calculation of additional fourteen values of L is not laborious, as most of the material for this calculation is provided by the moment solution.

XXI. T. EDEN AND F. YATES. *" On the Validity of Fisher's z Test when Applied to an Actual Example of Non-normal Data."* Journal of Agricultural Science, **1933**, Vol. XXIII, pp. **6-17**.

The validity of Fisher's z test depends theoretically on the assumption of normally distributed data. Since certain types of agricultural and other data to which the test may be usefully applied are decidedly non-normal in distribution practical tests with data of this nature are of importance in order to establish that the departures from normality ordinarily met with do not, in fact, invalidate the test. In this paper a test is described on data from the observation of height measurements on wheat. The data were arranged as an 8 block uniformity trial of 4 plots per block, and the distribution of the values of z obtained for a thousand random arrangements of the treatments. This distribution was found to agree satisfactorily with the theoretical distribution for normal data.

(*b*) TECHNIQUE OF FIELD EXPERIMENTS

XXII. S. H. JUSTESEN. *" Influence of Size and Shape of Plots on the Precision of Field Experiments with Potatoes."* Journal of Agricultural Science, **1932**, Vol. XXII, pp. **366-372**.

A uniformity trial with potatoes was used for investigating the effect of size and shape of plots on the precision of field experiments. Up to a certain limit the standard deviation in per cent. of the mean decreases when the size of plots is increased ; further increase of plot size increases the errors as a lesser part of the soil variation can be removed.

Two-row plots show less variation than either 1 or 3-row plots. This may be explained by row competition.

When the area to be used is fixed, smaller plots are more efficient than larger, owing to the greater number of replications in the former case. One exception occurs in the case where border rows are not harvested ; here 4-row plots are more efficient than 3-row plots, owing to the fact that a larger part of the area is included in the calculation when 4-row plots are used.

Long and narrow plots are more efficient than shorter and wider of the same size. The only exception is again explained by row competition.

In field experiments with potatoes, fairly large plots should be used ; at least two rows wide and preferably long and narrow strips.

XXIII. R. J. KALAMKAR. " *Experimental Error and the Field Plot Technique with Potatoes.*" Journal of Agricultural Science, 1932, Vol. XXII, pp. 373-385.

The present investigation consists of the statistical analysis of a uniformity trial with potatoes, conducted by Dr. Kirk. In this study the standard error in per cent. of the mean decreased slightly with the increase in the widths up to plots five rows wide, but any further increase in the width of the plot resulted in the higher standard error. The fertility contour map of the field is given to show graphically the effect of soil heterogeneity on the yield. The increased size of the plot resulted in decreased efficiency in the use of the land when the entire plot was harvested ; in other words, given a piece of land of certain size, it is advantageous to have a greater replication of smaller plots than a smaller number of larger plots. Four-row plots proved to be the most efficient when the border rows are discarded. The superiority of long and narrow plots over shorter and wider ones is demonstrated.

XXIV. R. J. KALAMKAR. " *A Study in Sampling Technique with Wheat.*" Journal of Agricultural Science, 1932, Vol. XXII, pp. 783-792.

The edge rows give significantly higher yields than the inside rows, indicating thereby the inadvisability of using edge rows in yield trials.

The variation between rows is very much greater than within rows. Different parts of the same drill row should therefore not be regarded as subject to independent error. The present investigation emphatically confirms Clapham's conclusions on this point.

A slight advantage may be gained by the subdivision of the area to be sampled, without additional labour.

In order to study the effect of the structure of a sampling unit of given size, five types of unit have been examined. Of these, method (1), in which the " sampling unit " consists of four parallel half-metre lengths on adjacent rows, appears to be the most precise, and may be recommended on the basis of this trial. The half-metres within such sampling units appear to be negatively rather than positively correlated, and a significantly lower sampling error is obtained in consequence.

Effect of competition between the rows is suggested as the probable explanation for the smaller variation between the sampling units than within them in method (1). Similar analysis on the ear number for the same method, moreover, showed that variation within sampling units was significantly less than the variation between sampling units. This is regarded as additional evidence that there is a competition effect in samples obtained by method (1).

Eighteen such complex units amounting to 36 metres of drill from one-fortieth acre plots would give about 5 per cent. sampling error.

A significant correlation of 0.73 between yield and ear number is obtained. This fact can be used to obtain increased precision for the prediction of yield when the number of ears is known.

XXV. F. R. IMMER. *" A Study of Sampling Technic with Sugar Beets."* Journal of Agricultural Research, 1932, Vol. XLIV, pp. 633-647.

Sampling technique was studied in relation to the determination of sugar percentage in sugar beets.

Regression of sugar percentage on weight of roots was not entirely linear. Ninety-two per cent. of the quadratic regression could be explained in terms of the linear function.

Soil heterogeneity between plots was found to affect sugar percentages significantly, even when the effect of weight was held constant by means of the regression relationship.

Tables are given showing the number of beets per plot needed to reduce the standard error of the mean to 0.3, 0.2, and 0.1 per cent. sugar for various sizes of plots and numbers of replications.

Variability in sugar percentage between plots and within plots must be considered in estimating the size of sample required and the number of replications needed to reduce the standard error to a given level.

The standard error of the mean of total sugar per beet was somewhat lower than the standard error for weight and much higher than that for sugar percentage.

Variability in sugar percentage between plots was essentially the same whether calculated from the mean of 10 beets analyzed individually or from a composite sample of the same number.

XXVI. F. R. IMMER. *" Size and Shape of Plot in Relation to Field Experiments with Sugar Beets."* Journal of Agricultural Research, 1932, Vol. XLIV, pp. 649-668.

Studies of size and shape of plot in relation to field experiments with sugar beet have been made, and the relationship determined between weight, sugar percentage, and apparent purity.

Standard errors, expressed in percentage of the mean, decreased in general with increased size of plot. An explanation is offered to account for a greater standard error from 6-row plots than from 3 or 4-row plots, when the entire plot is harvested.

Efficiency in use of land decreased with increased size of plot when the entire plot was harvested. When the border rows of the plots were removed, 4-row plots were most efficient.

Weight of beets was significantly correlated (negatively) with sugar percentage, but not with apparent purity. Sugar percentage was highly correlated (positively) with apparent purity. Intra-plot regression and correlation coefficients were given.

Contour maps for weight of roots, sugar percentage, and apparent purity were drawn from data on one hundred 6-row plots 2 rods long.

Sugar percentage varied significantly from plot to plot apart from its relation to weight. Fifty-four per cent. of the variability in apparent purity between plots was due to factors that affected sugar percentage as well.

The sampling error was calculated for sugar percentage and apparent purity determination for 4-row plots 2 rods long. The manner in which the standard error between plots may be reduced by replication and size of sample has been demonstrated.

(c) GENETICS

XXVII. R. A. FISHER. " *The Social Selection of Human Fertility.*" (The Herbert Spencer Lecture, delivered at Oxford, June 8th, 1932.) Oxford : The Clarendon Press, 1932, 32 pp.

Lecture delivered at Oxford in commemoration of Herbert Spencer. It is argued that the existence of natural law and the reliability of scientific prediction has the same basis in the physical, the biological and the social sciences.

XXVIII. R. A. FISHER. " *The Bearing of Genetics on Theories of Evolution.*" Science Progress, 1932, Vol. XXVII, pp. 273-287.

Lecture delivered before the Royal Society of Dublin. The biological phenomena of the recessiveness of mutations, of loss of vigour through inbreeding, and of adaptation to ensure cross-fertilisation, are interpreted as evidence that evolution is opposed rather than promoted by the mutations which occur.

XXIX. R. A. FISHER. " *On the Evidence Against the Chemical Induction of Melanism in Lepidoptera.*" Proceedings of the Royal Society, B, 1933, Vol. CXII, pp. 407-416.

A method is given of assessing by calculation the value of evidence of the non-occurrence of recessive mutations under experimental conditions. It appears that the evidence against the induction of melanic mutations in moths by feeding with lead, is insufficient to disprove the existence of mutation rates up to 5 per cent. or 8 per cent., according to the stage at which mutation is postulated.

Mutation rates of this magnitude would be far greater than those which can be certainly induced by any other agency.

The use of back-crosses instead of inbreeding would increase the value of experimental data of this kind by approximately thirty-fold.

THE SOIL

(Departments of Chemistry, Physics and Statistics.)

(a) SOIL CLASSIFICATION

XXX. L. L. LEE. " *The Possibilities of an International System for the Classification of Soils. Being a Consideration of the Influence of Geology and Climate on Soil Types. A Comparative Study of South-East England and Central New Jersey, U.S.A.*" Journal of the South-Eastern Agricultural College, 1931, No. 28, pp. 65-114.

An examination of the soils of South-East England, using the American methods of classification as modified for the New Jersey area of the United States. In both areas the soils occur in belts corresponding to the geological formations. Since geology has been the dominant factor in the soil formation of these areas, geological factors therefore receive first consideration in classifying the soils. Climatic factors have altered soil reaction more in Central New Jersey than in South-Eastern England, and, in general, podsolization is further advanced in the soils of Central New Jersey. A similar regional profile is developed in both districts, and consists typically