Introduction to "The Arrangement of Field Experiments"

by R.A. Fisher

By

T.P. Speed

Department of Statistics
University of California
Berkeley, California

Introduction to "The Arrangement of Field Experiments"

by R.A. Fisher

(J. Min. Agric. Gr. Br., **33,** 503-513; *Collected Papers* II, #38: 83-94)

T.P. Speed

Department of Statistics

University of California, Berkeley.

Background and significance

In 1919 the Director of Rothamsted Experimental Station, Sir John Russell invited Ronald Aylmer Fisher, a young mathematician with interests in evolution and genetics, to join the small group of scientists at Rothamsted in order that, Russell (1966, p.327), "after studying our records he should tell me whether they were suitable for proper statistical examination and might be expected to yield more information than we had extracted." Fisher accepted the invitation and in a very short time Russell realized (loc. cit.) "that he was more than a man of great ability; he was in fact a genius who must be retained." In the few years which followed, Fisher introduced the subdivision of sums of squares now known as an analysis of variance (anova) table (1923), derived the exact distribution of the (log of the) ratio of two independent chi- squared variates (1924), introduced the principles of blocking and randomization, as well as the randomized block, Latin square and split plot experiments, the latter with two anova tables (1925), promoted factorial experiments and foreshadowed the notion of confounding (1926). Of course Fisher made many contributions to theoretical statistics over this same period, see [*], but the above relate directly to the design and analysis of field experiments, the topic of the paper which follows. It was an incredibly productive period for Fisher, with his ideas quickly transforming agricultural experimentation in Great Britain and more widely, and in major respects these ideas have remained the statistical basis of agricultural experimentation to this day. Other fields of science

and technology such as horticulture, manufacturing industry, and later psychology and education, also adopted Fisher's statistical principles of experimentation, and again they continue to be regarded as fundamental today.

The story of Fisher and the design and analysis of experiments has been told many times before, and at greater length than is possible here, see Mahalanobis (1938), Hotelling (1951), Youden (1951), Yates (1964, 1975), Yates and Mather (1965), Russell (1966, pp.325-332), Neyman (1967), Cochran (1976) and Box (1978). In this short introduction to Fisher's delightful 1926 essay, we will briefly review the background to the paper and comment on its immediate and longer-term impact.

The main topics discussed in Fisher's essay are significance testing, replication, randomization, local control (the elimination of heterogeneity) and factorial (there called complex) experimentation. Replication, randomization and local control are all linked together in Fisher's discussion of the dual tasks of reducing the error in comparisons of interest, and in obtaining a valid estimate of that error. Most of these ideas had been published in the 2-3 previous years, either in articles on particular field experiments, or in the book *Statistical Methods for Research Workers*, see Pearce (this volume). What distinguishes this essay from these earlier works is its clarity, its comprehensiveness, its compactness - the material in the book was somewhat scattered - and the total absence of numerical or algebraic calculations. It is truly a discussion of principles. This form was undoubtedly dictated by the journal in which the essay appeared; indeed it is probable that the paper by Russell listed as reference (3) was the direct cause of this paper being written. Russell's essay was aimed at farmers interested in the results of field experiments, and others interested in methods of carrying them out, and no doubt Fisher saw this as an excellent opportunity to explain his new ideas to their natural audience.

Fisher does not appear to have had any direct contact with field experiments prior to his joining Rothamsted in 1919, but his Cambridge teacher F.J.M. Stratton had written on the application of the theory of errors to agriculture, Wood and Stratton (1910),

and it seems likely that Fisher would have been familiar with this paper and a later one, Mercer and Hall (1911), on the same topic. Sir A. Daniel Hall was Russell's predecessor as Director of Rothamsted and published an expository paper, Hall (1925), on the principles of agricultural experiments in the same volume in which Russell (1926) appeared.

Fisher's interest in the correlation coefficient and in evolutionary theory and genetics came together in his earlier paper, Fisher (1918), on the correlation between relatives, and this provided the foundation for his work on the design and analysis of field experiments. In that paper the modern term variance was introduced as the square of the standard deviation in a normal population. Fisher was considering measurements such as stature in humans, and his interest was in ascribing to various causes, fractions of the total variance which they combine to produce. After explaining the role of Mendelian inheritance in his analysis, Fisher derived a partition of the variance of such a measurement which, in it simplest form, was into two parts: an additive part and a second part, described as the effect of dominance (within locus interaction). His definitions involved what we would now describe as the weighted least-squares fitting of a linear model, with the first part being that "explained" by the model, and the second the residual. Later Fisher introduced a further term due to epistacy (between locus interaction) and he also considered multiple alleles at each locus. Thus by 1918 he already had a rather more complex than usual form of linear model for a number of factors, and a clear formulation of the idea of partitioning variation into ascribable causes.

On March 20, 1923 the paper, Fisher and Mackenzie (1923), was received by the Journal of Agricultural Science in Cambridge. This paper contained the analysis of a series of field experiments begun two years earlier at Rothamsted, and was probably the first such in which Fisher was involved. The innovations in this paper have been discussed elsewhere, Seal (1967), Yates (1975), and its main interest to us lies in the fact that we find there the first published analysis of variance table, including terms

due to the main effects and interactions of varieties and the manuring factor. This experiment was later reanalysed as a split-plot design, with two analysis of variance tables, in Fisher (1925); see Pearce (this volume) for a detailed analysis. Nine days after the Fisher and Mackenzie paper was received, Gosset ("Student") wrote to Fisher, Gosset (1970, letter 20), requesting assistance on a problem concerning the appropriate error when comparing a number of varieties each replicated an equal number of times in a field experiment. Fisher's solution — one of the few cases in which he presents an explicit linear model — appeared as a footnote in Gosset (1923), and was identical to that embodied in his paper with Mackenzie.

At this point it is clear that Fisher knew how to analyse field experiments: by a combination of (usually implicit) linear models for the estimation of effects, together with an analysis of variance table for the estimation of error, although it should be noted that he also tried a multiplicative model in Fisher and Mackenzie (1923). It is not clear when Fisher hit upon the idea of eliminating heterogeneity through blocking, or through the use of Latin squares. Interestingly, it appears that he did not get an opportunity to design and conduct an experiment at Rothamsted along his own lines until after the publication of his 1925 book and the paper which follows. It is stated in Box (1978, pp.156-158) that Fisher was working on Latin squares in 1924, and that in the same year he had used a Latin square design for a forestry experiment. The results of such an experiment would naturally not become available for some time afterwards, and as a consequence the numerical illustrations of the increase in precision achievable through blocking and the use of Latin squares offered in his book, refer to layouts randomly imposed on the mangold uniformity data of Mercer and Hall (1911).

It is equally unclear when (and how) Fisher hit upon the idea of randomization. The role which randomization plays in validating the comparison of treatment and error mean squares was touched upon, but not explained very fully in the 1925 book, and so the present paper is the first careful discussion of this issue. On its own, the

argument is clear and compelling, but the whole topic continues to puzzle students of the subject, even now. A possible explanation of this lies in the absence of any direct connection between the linear models according to which such experiments are usually analysed, and the randomization argument. The connection was made in the work of Kempthorne (1952) and his students, but in most treatments of the design and analysis of experiments, the problem remains.

Summarizing the discussion so far, we see that the principles of field experimentation outlined by Fisher in the paper below, were based in part on the result of his analysing experiments designed by others, and in part by numerical experiments he carried out on the uniformity trials of Mercer and Hall (1911). At the time of the publication of this paper, he could not have pointed to a single experiment successfully designed and analysed according to the principles expounded. It is clear that this is no way inhibited him from vigorously expounding his ideas, but it is interesting to note that at the time, they were just that: ideas.

The impact of Fisher's principles for designing and analysing field experiments was dramatic. His methods spread quickly in Great Britain and were also rapidly taken up by research workers in other countries and in other fields. In a sense the 1935 publication of *Design of Experiments* signalled the end, not the beginning of the Fisherian revolution in field experimentation, for by 1935 many national agricultural research organisations were using Fisher's ideas. A major factor in this rapid transformation was undoubtedly the central role played by Rothamsted Experimental Station in agricultural research in Great Britain, see Russell (1966), and in the British Empire more generally.

Box (1978, p.157) describes the enquiries and requests from agricultural workers running experiments, the world travels of Sir John Russell, the establishment of the Commonwealth Agricultural Bureaux, and the rapidly increasing number of voluntary workers attaching themselves to the Statistics Department at Rothamsted in this period. A report by one such person can be found in Goulden (1931). Later (1931) Fisher

also visited the U.S. For further remarks and details on early experiments run according to Fisher's principles we refer to Kerr (1988, sugar in Australia in 1928), Haines (1929, 1930a, 1930b, rubber in Malaya), Kirk (1929, potatoes in Canada), Richey (1930, agronomy in the U.S.A.), Gregory et al. (1932, cotton in the Sudan), Arnold (1975) for references to New Zealand work, Harrison et al. (1935, tea in India), Tippett (1935, textiles in Great Britain), Wishart (1935, silk in China). We also draw attention to the photograph given in Box (1978, Plate 6) of the Latin square experiment Fisher designed in 1929 for the Forestry Commussion of Great Britain. Shortly before the publication of *Design of Experiments*, Snedecor (1934) appeared, and this work contributed greatly to the spread of Fisher's ideas in the U.S.A. and more widely, not only in the sphere of agriculture, but in other fields as well.

Randomized block experiments, both complete and incomplete have continued to be popular in field experimentation, because of their flexibility, simplicity and robustness, see Patterson and Silvey (1980). In field experimentation, at least, the many special designs (Latin and Graeco-Latin squares, lattice designs, etc.), see e.g. Pearce (1983) and Mead (1988), are being replaced by the more flexible though less symmetric α-designs, Patterson *et al* (1978), and related row and column designs, Williams and John (1989). See also Seeger et al (1987). Essential use is made of computers to generate and test such designs (for efficiency), as well as to analyse data from the resulting experiment.

Factorial experiments have become widely used in most areas of applied science and technology. In these fields, at least, Fisher's view on whether Nature should be asked only one question at a time, has clearly won the day. The idea of confounding foreshadowed at the end of the paper, was quickly developed by Yates, M.M. Barnard, R.C. Bose and others in the 1930s and 1940s, and remains an important component of field and industrial experimentation today.

Although Fisher's approach to the design and analysis of field experiments was quickly and widely adopted, it was not accepted in every detail. Discussion

concerning the relative merits of systematic and randomized designs began with Gosset (1931), and this debate continued until after Gosset's death in 1937, see e.g. Gosset (1936a, 1936b, 1937-8), Arnold (1985), with Fisher, Yates, Neyman, E.S. Pearson, Jeffreys and others participating, see e.g. Barbacki and Fisher (1936), Yates (1939) and references therein.

Two questions concerning randomization which Fisher did not adequately address were a) what, if anything, should be done if an obviously undesirable arrangement arises as a result of randomization, and b) whether the analysis of a field experiment should be conditional upon the actual arrangement obtained through randomization, as that is an obvious ancillary. The relationship between model-based and randomization-based inference continues to be debated in the context of field experiments, as it does in other areas of statistics. The whole subject of randomization continues to puzzle some Bayesians, see Savage (1976), and others, and it seems safe to say that the recent resurgence of interest in neighbour methods sparked by the publication of Wilkinson et al. (1983), raises the topic once more in the context of field experiments.

For a quite different approach to field experiments, and references to a Bayesian view of randomization, see Neyman (1990) and the remarks of D. Rubin which follow that translation.

The paper which follows, far better than the scattered outline in Fisher (1925), or the more detailed manual Fisher and Wishart (1930), or the book Fisher (1935) which followed, can be viewed as outlining in simple terms Fisher's statistical principles of field experimentation. The fact that every one of the ideas so lucidly expounded here, for essentially the first time, remains central to the subject even today, is clear evidence that it deserves to be included in any collection entitled "Breakthroughs in Statistics".

Notes on the paper

Fisher asks in the section entitled *When is a result significant?* how one should interpret the result of an acre plot with manure yielding 10% more than a similarly located plot which did not get manure, but which was treated similarly in all other respects. He describes the classical interpretation of a significance level, and then goes on to say that about 500 years' experience would be required to estimate the upper 5% point of the distribution of such ratios, giving a simple argument using the order statistics for so doing. How did Fisher come up with the number 500, and why did he focus on the 5% point? His answer to the second question is clearly spelled out in the last paragraph of p.504, see also Hall and Selinger (1986) for further discussion of this point. As for the number 500, we can be sure that Fisher was well aware of the fact that the (asymptotic) variance of the [Np]-th order statistic in a random sample of size N is approximately

$$\frac{p(1-p)}{N} \cdot \frac{1}{f(x_p)^2}$$

where f is the common density and x_p is its upper 100p% point. With p = .05 and N = 500, the square root of the first factor is .01, and of course the second factor is unknown when f (not to mention x_p) is unknown. It seems reasonable to surmize that Fisher fixed upon N = 500 in order to make this standard deviation of the order of 1%, well aware that he could not know the other factor. (For the normal density this term is about 10.)

Having obtained this unrealistically large number, Fisher goes on to explain how the experimenter can use the t-distribution if the ten previous years' records are available. He argues as follows: the difference d in the year of an experiment should be divided by the estimated standard error s based upon the previous ten years of trials with a uniform treatment, and compared with the upper 5% point of the t-distribution on 10 degrees of freedom, i.e. with 2.238. Fisher measured the difference of yields and the standard error as percents of the mean yield, and concludes that if the standard

error based on ten years' data was 3%, an observed difference of 10% exceeds $3 \times 2.238 = 6.684$, and so is significant at the 5% level. He concludes by explaining how to calculate s.

In effect Fisher has argued that we do not need 500 years of uniformity data as he originally suggested,; some number of the order of 10 will suffice "if we put our trust in the theory of errors." His next and boldest step is to argue that no uniformity trials are necessary at all, provided that the experiment is properly randomized. In so doing, he is implicitly assuming that the magnitude of the year-to-year variation he had just been discussing, coincides with that of the error provided by the experiment itself, i.e. of the plot-to-plot variation, within a year. Of course this is not generally true, and a valid criticism of Fisher's emphasis on randomization, replication and local control in the context of agricultural trials is that the focusses on the component of variation of least practical importance. The within-experiment error with which he is so concerned in this paper is frequently far smaller than the variation observed from year to year, at one location, from location to location, in one year, and sizeable location-by-year interaction components of variation, in one year, are not uncommon, see Patterson and Silvey (1980).

In the paragraph entitled *Errors Wrongly Estimated* Fisher (footnote, p.506) disclaims responsibility for the design of an experiment discussed by Russell (1926). There must have been a misunderstanding here, because in the article Russell (pp.996-7) introduces and displays a balanced (i.e. systematic) plan, with the sentence "An instance is afforded by the experiment designed by Mr. R.A. Fisher for a detailed study of the effect of phosphates and of nitrogen on crop yield."

The paragraph beginning near the top of p.507 finds Fisher explaining the sense in which the estimate of error obtained by randomization is *valid*. Much effort has been devoted to the justification of these and similar remarks, and so it seems worthwhile to note some of the details. Suppose that $2r$ plots are labelled $i = 1, \ldots, 2r$, and that a subset T are of r plots are assigned to a treatment, and the remainder C to be controls,

the assignment being at random, so that all $\begin{bmatrix} 2r \\ r \end{bmatrix}$ possible assignments are equally likely. In what follows $\mathbf{E}_R$ denotes the expectation taken over the set of all such assignments, equally weighted, and under the additional assumption (the *null hypothesis*) of no differences between treated and control plots. If the yields are $y_1, \ldots, y_{2r}$, and $\bar{y}^T$, $\bar{y}^C$ and $\bar{y}$ denote the averages of the treated, control and all plots, respectively, then the treatment and error sums of squares are given by

$$SST = r\{(\bar{y}^T - \bar{y})^2 + (\bar{y}^C - \bar{y})^2\},$$

$$SSE = \sum_{i \in T} (y_i - \bar{y}^T)^2 + \sum_{i \in C} (y_i - \bar{y}^C)^2.$$

Write $\sigma^2 = (2r)^{-1} \sum_{i=1}^{2r} (y_i - \bar{y})^2$. Then it is an easy calculation based on the symmetry of the covariance matrix induced by the random assignment, to prove that

$$\mathbf{E}\{MST\} = \mathbf{E}\{MSE\} = \sigma^2,$$

where $MSE = SSE/2(r - 1)$ is Fisher's "estimated error", and σ^2 his "real" error in this case.

Fisher assert (p.88) that "..., the ratio of the real to the estimated error, calculated afresh for each of these arrangements, will be actually distributed in the theoretical distribution by which the significance of the result is tested." This will be trivially true if the actual randomization distribution is used, which can be done these days, but Fisher undoubtedly had in mind a chi-squared approximation to the distribution of SSE/σ^2, and a corresponding F approximation to the distribution of MST/MSE. Eden and Yates (1933) carried out a simulation study which supported this conclusions. Wald and Wolfowitz (1944) proved a theorem which justifies the chi-squared approximation in the simple case discussed above, whilst Welch (1937) and Pitman (1938) presented results which supported, to some extent, the use of the F distribution to approximate the randomization distribution of MST/MSE in the case of randomized complete block designs and Latin squares. The matter is not a simple one, see Davis and Speed (1988) for some related calculations, and it seems fair to say that the assertion of

Fisher quoted above has been shown to be a reasonable approximation to the truth, under certain conditions, for certain designs. The matter is really academic now, for we can actually calculate (or sample) the exact randomization distribution if we wish.

The discussion on p.508 of what is known as Beavan's half-drill strip method (see "Student" (2), cited by Fisher for fuller details), concerns the dispute between Fisher and "Student" mentioned on page x above. Fisher felt that the estimate of error used by "Student" in this type of experiment was not valid, whilst "Student" maintained to his death that the problem was more theoretical than practical, that the gain in precision which resulted was more than offset by any theoretical lack of validity. There seems to be little doubt that "Student" was correct on this point.

In his discussion of the Latin square Fisher notes that the two types of square he illustrates — the diagonal and Knight's move (also called Knut Vik) squares — had been used previously for variety trials in Ireland and Denmark. Both of these countries were carrying out co-operative agricultural research before the turn of the century, see Gossett (1936) for some historical remarks on their early activities, and Cochran (1976) and references therein. It is not hard to see that the diagonal square generally underestimates the error, whilst the Knight's move square generally overestimates it, see the discussion in §34 of Fisher (1935), and an interesting historical sidelight is the following remark concluding the section just cited: "It is a curious fact that the bias of the Knut Vik square, which was unexpected, appears to be actually larger than that of the diagonal square, which all experienced experimenters would confidently recognise." As Yates (1965-66) points out, this assertion was based on a small arithmetical error committed by Tedin in some numerical experiments; in fact the "bias" or loss of precision of the diagonal square is exactly equal to the gain in precision of the Knut Vik square.

So that the reader can better appreciate Fisher's remark about asking Nature few questions, it is worth quoting from Russell (1926, p.989, our italics):

"A committee or an investigator considering a scheme of experiments should first ..., and ask whether each experiment or question is framed in such a way that a definite answer can be given. The chief requirement is simplicity: *only one question should be asked at a time.*"

Biographical notes

Ronald Aylmer Fisher was born on 17 February 1890 in a suburb of London, England. He showed a special ability in mathematics at an early age, and received a good mathematical education at school prior to entering Gonville and Caius College, Cambridge in 1909. At Cambridge he studied mathematics and physics, and he also developed an interest in genetics and evolutionary theory, an interest which was to remain with him throughout his life. He graduated with first-class honours in 1912 and spent a further year at Cambridge on a studentship in Physics.

Upon leaving Cambridge, Fisher successively worked in a finance office, on a farm in Canada, and as a public school mathematics and physics teacher. His first paper, on the fitting of frequency curves by the method we now know as maximum likelihood, was published in 1912, and not long afterwards he obtained the exact (null) distribution of the normal correlation coefficient. This paper also drew attention to the 1908 paper of "Student" [see this volume], which had been neglected up to that time. Over this period Fisher was developing his ideas on the statistical aspects of genetics, and in 1918 published his pathbreaking synthesis of ideas from the Mendelian and the biometric schools: "The correlation between relatives on the supposition of Mendelian inheritance:"

Shortly after the publication of this paper Fisher joined Rothamsted Experimental Station, and went on to make a number of fundamental contributions to statistics and genetics. Further details of his life and work can be found in Yates and Mather 91965) and Box (1978).

REFERENCES

ARNOLD, G.C. (1985). The Hudson-Gosset correspondence. *The New Zealand Statistician*, **20**, 20-25.

BARBACKI, S. and FISHER, R.A. (1936). A test of the supposed precision of systematic arrangements. *Ann. Eugen.*, **7**, 189-193.

BOX, Joan Fisher (1978). *R.A. Fisher. The Life of a Scientist*, New York: John Wiley & Sons.

COCHRAN, W.G. (1976). Early development of techniques in comparative experimentation. In *On the History of Statistics and Probability*, D.B. Owen (Ed.), pages 3-25. New York: Marcel Dekker.

DAVIS, A.W. and SPEED, T.P. (1988). An Edgeworth expansion for the distribution of the F-ratio under a randomization model for the randomized block design. *Statistical Decision Theory and Related Topics IV*, **2**, eds S.S. Gupta and J.O. Berger. New York: Springer-Verlag. 119-130.

EDEN, T. and YATES, F. (1933). On the validity of Fisher's z test when applied to an actual sample of non-normal data. *J. Agric. Sci*, **23**, 6-17.

FISHER, R.A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinb.*, **52**, 399-433. [Reprinted with a commentary by P.A.P. Moran and C.A.B. Smith as Eugenics Laboratory Memoir XLI, 1966. London: Cambridge University Press.]

FISHER, R.A. (1923). Appendix to Gosset (1923).

FISHER, R.A. (1924). On a distribution yielding the error functions of several well known statistics. *Proc. Int. Congress. Math. Toronto*, **2**, 805-813.

FISHER, R.A. (1925). *Statistical Methods for Research Workers*, Edinburgh: Oliver and Boyd.

FISHER, R.A. (1935). *The Design of Experiments*, Edinburgh: Oliver and Boyd.

FISHER, R.A. and MACKENZIE, W.A. (1923). Studies in crop variation. II. The manurial response of different potato varieties. *J. Agric. Sci.*, **13**, 311-320.

FISHER, R.A. and WISHART, J. (1930). The arrangement of field experiments and the statistical reduction of the results. *Imperial Bureau of Soil Science, Technical Communication*, **10**, pp23.

GOSSET, W.S. ("Student") (1923). On testing varieties of cereals. *Biometrika*, **15**, 271-293.

GOSSET, W.S. ("Student") (1931). Yield trials. *Baillière's Encyclopedia of Scientific Agriculture*, London. [Reprinted as paper 15 in Gosset (1942)].

GOSSET, W.S. ("Student") (1936a). Cooperation in large-scale experiments. *Suppl. J.R. Statist. Soc.*, **3**, 115-122.

GOSSET, W.S. ("Student") (1936b). The half-drill strip system. Agricultural experiments. Letter to *Nature*, **138**, p.971.

GOSSET, W.S. ("Student") (1937-8). Comparison between balanced and random arrangements of field plots. *Biometrika*, **29**, 191-208.

GOSSET, W.S. (1942). *"Student's" Collected Papers*. Edited by E.S. Pearson and John Wishart. With a foreword by Launce McMullen. London: Biometrika

Office, University College.

GOSSET, W.S. (1970). *Letters from W.S. Gosset to R.A. Fisher, (1915-1936) Summaries by R.A. Fisher with a foreword by L.McMullen.* Private circulation.

GOULDEN, C.H. (1931). Modern methods of field experimentation. *Scientific Agriculture*, **11**, 681.

GREGORY, F.G., CROWTHER, F. and LAMBERT, A.R. (1932). The interrelation of factors controlling the production of cotton under irrigation in the Sudan. *J. Agric. Sci.*, **22**, 617.

HAINES, W.B. (1929). Block 6. Manuring experiment, RRI Experiment Station. First Report. *Quarterly Journal of the Rubber Research Institute of Malaya*, **1**, 241-244.

HAINES, W.B. (1930a). Block 6. Manuring experiment, RRI Experiment Station. Second Report. *Quarterly Journal of the Rubber Research Institute of Malaya*, **2**, 31-35.

HAINES, W.B. (1930b). Manuring of Rubber - II. Technique of plot experimentation. *Quarterly Journal of the Rubber Research Institute of Malaya*, **2**, 51-60.

HALL, A.D. (1925). The principles of agricultural experiments. *J. Min. Agric. Gr. Br.*, **32**, 202-210.

HALL, P. and SELINGER, B. (1986). Statistical significance: balancing evidence against doubt. *Austral. J. Statist.*, **28**, 354-370.

HARRISON, C.J., BOSE, S.S., and MAHALANOBIS, P.C. (1935). The effect manurial dressings, weather conditions and manufacturing processes on the quality of tea at Tocklai experimental station, Assam. *Sankhya*, **2**, 33-42.

HOTELLING, H. (1951). The impact of R.A. Fisher on statistics. *J. Amer. Statist. Assoc.*, **46**, 35-46.

HUDSON, (1985). *The New Zealand Statistician*, **20**, 20-25.

KEMPTHORNE, O. (1952). *The design and analysis of experiments.* New York: Wiley.

KERR, J.D. (1988). Introduction of statistical design and analysis by the Queensland Board of Sugar Experiment Stations. *Austral. J. Statist.*, **30B**, 44-53.

KIRK, L.E. (1929). Field plot technique with potatoes with special reference to the Latin square. *Scientific Agriculture*, **9**, 719-729.

MAHALANOBIS, P.C. (1938). Professor Ronald Aylmer Fisher. *Sankhya*, **4**, 265-272.

MEAD, R. (1988). *The design of experiments: Statistical principles for practical applications.* New York: Cambridge University Press.

MERCER, W.B. and HALL, A.D. (1911). The experimental error of field trials. *J. Agric. Sci.*, **4**, 107-132.

NEYMAN, J. (1967). R.A. Fisher (1890-1962): An appreciation. *Science*, **156**, 1456-1460.

NEYMAN, J. (1990). On the application of probability theory to agricultural experiments. Essay on principles. (Translation from the original Polish of pp.27-42 from *Rocz. Nauk. Roln* X (1923) 1-51 by D. Dabrowska, edited by T.P. Speed). *Statistical Science* (To appear).

PATTERSON, H.D. and SILVEY, V. (1980). Statutory and recommended list trials of crop varieties in the United Kingdom. *J. Roy Statist. Soc. Ser. A.*, **143**, 219-252.

PATTERSON, H.D. WILLIAMS, E.R. and HUNTER, E.A. (1978). Block designs for variety trials. *J. Ag. Sci. Camb.*, **90**, 395-400.

PEARCE, S.C. (1983). *The agricultural field experiment. A statistical examination of theory and practice.* New York: John Wiley & Sons.

PITMAN, E.J.G. (1938). Significance tests which may be applied to samples from any populations: III. The analysis of variance tests. *Biometrika*, **29**, 322-335.

RICHEY, F.D. (1930). Some applications of statistical methods to agronomic experiments. *J. Amer. Statist. Assoc.*, **25**, 269-283.

RUSSELL, E.J. (1926). Field experiments: How they are made and what they are. *J. Min. Agric. Gr. Br.*, **32**, 989-1001.

RUSSELL, E.J. (1966). *A history of agricultural science in Great Britain.* London: George Allen & Unwin.

SAVAGE, L.J. (1976). On rereading R.A. Fisher. (edited by J.W. Pratt). *Ann. Statist.*, **4**, 441-500.

SEAL, H. (1967). The historical development of the Gauss linear model. *Biometrika*, **54**, 1-24.

SEEGER, P., KRISTENSEN, K. and NORELL, L. (1987). Experimental designs in two dimensions for official sugar beet variety trials in Scandinavia. Report 286, Sveriges Lantbruksuniversitet (Swedish University of Agricultural Sciences) Uppsala, Department of Economics and Statistics.

SNEDECOR, G.W. (1934). *Computation and Interpretation of Analysis of Variance and Covariance.* Ames, Iowa: Collegiate Press.

TIPPETT, L.H.C. (1935). Some applications of statistical methods to the study of variation of quality in the production of cotton yarn. *Suppl. J. Roy. Statist. Soc.*, **2**, 27-62.

WELCH, B.L. (1937). On the z-test in randomized blocks and Latin squares. *Biometrika*, **29**, 21-51.

WILKINSON, G.N., ECKHART, S.R., HANCOCK, T.W. and O. MAYO (1983). Nearest neighbour (NN) analysis of field experiments (with discussion). *J. Roy. Statist. Soc. Ser. B,* **45**, 151-211.

WILLIAMS, E.R. and JOHN, J.A. (1989). Construction of row and column designs with contiguous replicates. *Appl. Statist.*, **38**, 149-154.

WISHART, J. (1935). Contribution to the discussion of Tippett (1935), pp58-59.

WOOD, T.B. and STRATTON, F.J.M. (1910). The interpretation of experimental results. *J. Agric. Sci.*, **3**, 417-440.

YATES, F. (1939). The comparative advantages of systematic and randomized arrangements in the design of agricultural and biological experiments. *Biometrika*, **30**, 440-466.

YATES, F. (1964). Sir Ronald Fisher and the design of experiments. *Biometrics*, **20**, 307-321.

YATES, F. (1965-66). A fresh look at the basic principles of the design and analysis of experiments. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* IV ed. J. Neyman and E.L. Scott, 777-790.

YATES, F. (1975). The early history of experimental design. In *A survey of statistical design and linear models,* ed. J.N. Srivastava, pp581-592. Amsterdam: North-Holland.

YATES, F. and MATHER, K. (1965). Ronald Aylmer Fisher. *Biographical Memoirs of Fellows of the Royal Society*, **9**, 91-120.

YOUDEN, W.J. (1951). The Fisherian revolution in methods of experimentation. *J. Amer. Statist. Assoc.*, **45**, 47-50.