NON-ADDITIVITIES IN A LATIN SQUARE DESIGN[1]

M. B. WILK[2]
Iowa State College and Princeton University[2]

AND

OSCAR KEMPTHORNE[3]
Iowa State College

A brief history of the Latin square design is given. A generalization of the design to the case in which the rows, columns and treatments represented in the experiment are samples from populations of rows, columns and treatments respectively is studied. A possible frame of reference for the interpretation of the experimental results is described. The leads to what is termed a "population model," various population parameters, means and components of variation which are of interest to the experimenter. No assumptions are made about additivity of experimental units and treatments. Results on expectations of mean squares in the analysis of variance are given in terms of quantities (denoted by Σ's) which result in a concise description and in terms of the components of variation. Biases in the estimation of components of variation by means of the analysis of variance are discussed and assessed. Comparisons of randomized block designs and Latin square designs are given for the general case of non-additive treatments for random, mixed, or fixed population of rows and columns. A generalization of the design is discussed. The mathematical machinery which is used to derive the results is presented briefly. Finally linear estimates and errors of estimates are discussed. One main conclusion of the study is that the Latin square analysis of variance may overestimate the error of treatment comparisons and underestimate the component of variation associated with treatment main effects.

INTRODUCTION

THE present paper gives results and discussion concerning the statistics of experiments involving the Latin square design, and is part of a series of investigations by one or both of the present authors on the meaning and role of linear models in the analysis of variance of randomized experiments.

The line of investigation stems back to the basic writings of Fisher [5, 6] and Yates [24, 25]. The procedure employed is related to the work of Neyman et al. [12]. A recent step was the "finite model analysis" of Kempthorne [10, 11[4]] in which the basic designs were considered under the assumption of additivity of treatment and experimental unit. A step in the direction of studying non-additivities was made for the case of randomized blocks by Kempthorne [10] and more general treatment of this case was given by Wilk [19, 20]. Research sponsored by the Wright Air Development Center on the "mixed model" controversy led the authors to the combination of the methods of

[1] Paper presented at a joint session of the Institute of Mathematical Statistics and the American Statistical Association at New York December, 1955.

[2] Paper prepared partly under Analysis of Variance project of the Statistical Laboratory, Iowa State College and partly in connection with research sponsored by the Office of Naval Research.

[3] Journal paper No. J-2952 of the Iowa Agricultural Experiment Station, Ames, Iowa. Project 890.

[4] Originally written and presented in December 1952.

finite model analysis" with general sampling of treatments and experimental material [23]. This development made use of the design random variables of Kempthorne [10, 11] and the sampling dummy variables of Cornfield [2]. The relationship of our early work on the completely randomized design to the 1949 work of Tukey [17] (unpublished and unknown to us at the time), Cornfield [3], and Bennett and Franklin [1], and Cornfield and Tukey [4] has been discussed elsewhere (Wilk and Kempthorne [22], Cornfield and Tukey [4]). Other relevant investigations are those of Scheffé [15] and Smith [16]. Work indirectly related to the present context is that of Fisher [7], Welch [18] and Pitman [14] on randomization tests.

The Latin square was first proposed as an experimental design by Fisher [5]. The early history is best given by the following quotation from Fisher [6]:

"Systematic arrangements in a square, in which the number of rows and of columns is equal to the number of varieties such as

A	B	C	D	E		A	B	C	D	E
E	A	B	C	D		D	E	A	B	C
D	E	A	B	C		B	C	D	E	A
C	D	E	A	B		E	A	B	C	D
B	C	D	E	A		C	D	E	A	B

have been used previously for variety trials in, for example, Ireland and Denmark; but the term 'Latin square' should not be applied to any such systematic arrangements. The problem of the Latin square, from which the name was borrowed, as formulated by Euler, consists in the enumeration of *every possible* arrangement, subject to the conditions that each row and column shall contain one plot of each variety. Consequently the term Latin square should only be applied to a process of randomization by which one is selected at random out of the total number of Latin squares possible: or, at least, to specify the agricultural requirement more strictly, out of a number of Latin squares in the aggregate, of which every pair of plots not in the same row or column belongs equally frequently to the same treatment."

The actual randomization procedure for some sizes of the square was given by Yates [24]. The enumeration of 6×6 Latin squares was done by Fisher and Yates [8]. Yates [25] gave a detailed discussion of Latin squares in which a row, column or treatment is missing and showed that "incomplete Latin squares of these types give unbiassed estimates of error and are therefore valid experimental arrangements." The validity of the Latin square was discussed by Neyman et al. [12], using a mode of reasoning related to that of the present paper. The 7×7 Latin squares were enumerated by Norton [13]. (The problem of orthogonal squares is not relevant in our present context.) The distribution of a criterion equivalent to the variance ratio (treatment sum of squares divided by treatment plus error sum of squares) was examined for the Latin square by Pitman [14] and Welch [18]. An evaluation of the basic Latin square design on the basis of a derived linear model with the assumption of additivity of units and treatments and no additional sampling was given by Kempthorne [10, Section 10.5].

Following Fisher [6] we shall take the term Latin square design to imply *both* a geometrical pattern *and* a suitably restricted randomization procedure. For our immediate purposes, which involve first and second moments only, we take Fisher's rule in the last part of the last sentence of the quotation above as defining the randomization restrictions.

The structure of the paper is as follows. First an experimental situation and design, which is in fact a generalization of the basic Latin square design, is described formally. A frame of reference for analysis is then formulated and some general results on expectations of mean squares given. Some comparisons of the Latin square design and the randomized block design are given and extension of the design discussed. The machinery by which the results were obtained is described briefly. The paper closes with a short discussion of estimates both of components of the population model and of errors of estimates.

One main conclusion of the study is that the Latin square analysis of variance may seriously overestimate the error of treatment comparisons and underestimate the component of variation associated with treatment main effects. In spite of this, the Latin square will, in many circumstances, be more advantageous than the randomized block design.

We would like to thank John Tukey for helpful comments in connection with the preparation of this paper.

EXPERIMENTAL UNITS AND TREATMENTS

Our analysis is centered around two basic notions, namely *treatment* and *experimental unit.* A treatment is a combination of stimuli or operations imposed by the experimenter, who is usually able to identify the most relevant or essential features. While it is usually reasonable to regard a treatment as perfectly reproducible in principle, treatments are ordinarily not perfectly reproducible in practice. For the analysis given here it will suffice to assume that to each treatment there corresponds a population of combinations of stimuli or operations from which one combination appears at random in each use of the treatment. This assumption is likely to be a satisfactory approximation to many real situations, and it does cover to a certain extent (i) variabilities arising from inevitable limitations in technique in, say, attempting to apply a pressure of exactly 10 p.s.i. and (ii) such treatments as variety of corn where we employ individual, somewhat variable, seeds and not an "abstract variety."

In most cases, an experimental unit may be thought of as quite apart from the experiment, but we use the term to denote the complex of experimental material and circumstances associated with possible treatment applications in the experiment. The relevant background for the experiment, so far as statistical inferences are concerned, is tied up with the appropriate population of experimental units. The latter is determined by such random sampling and allocation as is built into the experimental procedure, implicitly or explicitly. Just how identifiable experimental units are will vary from experiment to experiment. The association of units with individual animals of a litter may be practically sufficient in comparisons of rations. In the case of comparative psychological "tests" on the same person at different times, the experimental unit is a particular combination of person and test environment.

The experimenter is usually confident or at least hopeful that only certain features of the whole complex of conditions and stimuli involved in the application of a treatment are essential. For example, in noting the response of a plastic to a pressure of 50 p.s.i. we might ignore such details as the material of the pressure plates and whether the mechanism is hydraulic or not. On the other hand, if the rate of pressure increase is greater for a final pressure of 100 p.s.i. than that for 50 p.s.i., then the description of a treatment by the final pressure would probably be quite inadequate. Just how incomplete the description of a treatment may be is not in general a statistical question. The similar ambiguities related to experimental units are usually of less importance if constancy of their properties, rather than complete description, can be assured and if an adequate randomization procedure is followed.

We shall take the point of view throughout that an experimental unit is not reproducible (can be used only once). Usually experimental units are "given" in the sense that the experimenter cannot or does not feel it worthwhile to control their variation by means other than randomization.

The distinction between treatment and experimental unit is the essential of the difference between a Latin square experimental design, involving as a basis t^2 units and t treatments, and the fractional factorial 1-in-t of a $t\times t\times t$ experimental situation where t^2 treatments are selected according to a Latin square type pattern, at random, from the possible t^3 factor combinations. The latter procedure may be called Latin square sampling of factor combinations. The actual experiment with the selected treatments will involve experimental units as well, and some experimental design specifying the randomization of treatments to units.

THE EXPERIMENTAL SITUATION AND PROCEDURE

We suppose that there are RC experimental units classified into R rows and C columns with one unit at each row-column intersection, for example plots of land or portions of a roll of paper, classified by distance along and distance from an edge, or units of person-hours classified by person and by hour. We suppose also that there is a population of T treatments, which are reproducible at will. The notation and experimental procedure are then as follows:

(1) denote row in unit population by $i=1, 2, \cdots, R$,
(2) denote column in unit population by $j=1, 2, \cdots, C$,
(3) denote treatment in treatment population by $k=1, 2, \cdots, T$,
(4) select at random t rows and let $i^*=1, 2, \cdots, t$ denote the selected rows in the (random) order of selection ($t\leq R$),
(5) select at random t columns and let $j^*=1, 2, \cdots, t$ denote the selected columns in the (random) order of selection ($t\leq C$),
(6) select at random t treatments and let $k^*=1, 2, \cdots, t$ denote the selected treatments in the (random) order of selection ($t\leq T$),
(7) select a Latin square at random from the totality[5] of $t\times t$ Latin squares and use it to assign selected treatments to selected experimental units,
(8) let $f=1, 2, \cdots, t$ denote replicates of a given treatment.

[5] For our purposes (and from practical necessity for the larger squares) this can be related to "any grouping of $t\times t$ squares which fulfill the requirement of Fisher's [6] rule." Any transformation set has this property.

We may now characterize the resultant t^2 observations in two ways both of which are useful

(1) $z_{i^*j^*}$ is the observation on the unit corresponding to selected row i^* and selected column j^*

(2) x_{k^*f} is the observation on the fth replicate of selected treatment k^*.

We use the $z_{i^*j^*}$ when we work with row and column totals, and the x_{k^*f} when we work with the treatment totals. It may be recalled that some writers use a notation equivalent to $x_{i^*j^*(k^*)}$ indicating that once i^* and j^* and the plan are given k^* is fixed.

The examination we shall make includes the case when the totality of rows and columns are used which corresponds to the usual Latin square design. The possibility of a broader experiment involving a sampling of rows and columns does not appear to have been discussed in the literature, but a more inclusive examination seems to be of practical as well as theoretical interest.

In the case of an experiment involving $t=6$, $T=100$, $R=200$, $C=6$ we may be concerned with one or more of the following inference problems: (i) obtaining an estimate of variability of members of the population of 100 treatments relative to the entire population of 1,200 units; (ii) a similar estimate as in (i) but relative only to the actual 6 rows and 6 columns which were employed in the experiment; (iii) estimates concerning differences among a subset of treatments actually employed relative to any subset of the relevant experimental unit population; etc. In general the possible scope of the inference will be dictated by the experimental design and procedure. Any analysis suggested by the data itself is subject to the usual criticism but may be very useful and important nonetheless. Extension of inferences beyond the statistical populations, as determined by the random sampling and allocation procedures, cannot have a statistical justification or assessment. Such extensions are of the utmost scientific importance but must lean on many factors, only one of which is the statistical analysis.

FRAME OF REFERENCE FOR ANALYSIS; POPULATION MODEL

A prerequisite for formal analysis of the experimental data is a relevant conceptual "frame of reference" which we now develop, based on some explicit elementary or primitive assumptions. We utilize the notion of a true or typical yield for each particular combination of treatment and experimental unit. It is supposed that if one could use the same unit over and over again and one could apply the same treatment each time one would get a population of yields which would have a "representative" (often chosen to be the mean) value Y_{ijk} (which we shall refer to as the "true response") dependent *only* on the particular unit and treatment considered and not on the whole plan of treatments as assigned to units. This assumes that there is no competition or collaboration between units. It is easy to think of agronomic experiments for which this is not the case, for example an effective insect repellent on a particular plot of land causes the insects to move on to a plot of land which is being treated with an ineffective repellent.

The conceptual array of RCT numbers $\{Y_{ijk}\}$ represents the totality of relevant empirical information that is all that experimentation *alone* can give (as opposed to summary cause and effect, or mechanistic, notions). This is frequently a small part of the knowledge the experimenter wants, which is usually laws or patterns of behavior, or generalization to a wider background than the experiment provided.

The scale of observation will be assumed to be given. The procedure to be followed is not dependent on the scale of observation. Some scales are however better than others, as indeed is obvious without statistical considerations. One important aspect of our conclusions is that the choice of scale is an important problem from the point of view of statistical inference alone.

The statistical problem is what our observations tell us about certain parameters of (functions of the elements of) the population of RCT true responses. A useful decomposition of the "true" responses, which we will call the population model, is

$$Y_{ijk} = \mu + r_i + c_j + t_k + (rc)_{ij} + (rt)_{ik} + (ct)_{jk} + (rct)_{ijk} \qquad (1)$$

where, with replacement of a suffix by a dot denoting averaging over the suffix,

$$\mu = Y_{\cdot\cdot\cdot},\ r_i = Y_{i\cdot\cdot} - \mu,\ c_j = Y_{\cdot j\cdot} - \mu,\ t_k = Y_{\cdot\cdot k} - \mu$$
$$(rc)_{ij} = Y_{ij\cdot} - Y_{i\cdot\cdot} - Y_{\cdot j\cdot} + Y_{\cdot\cdot\cdot},\ (rt)_{ik} = (Y_{i\cdot k} - Y_{i\cdot\cdot} - Y_{\cdot\cdot k} + Y_{\cdot\cdot\cdot})$$
$$(ct)_{jk} = Y_{\cdot jk} - Y_{\cdot j\cdot} - Y_{\cdot\cdot k} + Y_{\cdot\cdot\cdot}, \quad \text{and}$$
$$(rct)_{ijk} = Y_{ijk} - Y_{ij\cdot} - Y_{i\cdot k} - Y_{\cdot jk} + Y_{i\cdot\cdot} + Y_{\cdot j\cdot} + Y_{\cdot\cdot k} - Y_{\cdot\cdot\cdot}.$$

This is an algebraic identity involving *main effects* and *interactions*. Each of the terms of (1) has a definite physical interpretation. For example, t_k is the mean difference which would be observed, averaged over the totality of experimental units, between treatment k and the mean of all treatments in the treatment population. As in all cases where main effects and interactions are used, it is important to note that a term depending on certain of the factors of treatment or classification depends on the totality of the other factors also; for example, $(rc)_{ij}$ is based on an average over the population of treatments considered. The number of parameters in (1) is $(1+R)(1+C)(1+T)$, the excess over RCT being due to the relationships:

$$\sum_i r_i = 0;\ \sum_j c_j = 0;\ \sum_k t_k = 0;\ \sum_{i \text{ or } j} (rc)_{ij} = 0;\ \sum_{i \text{ or } k} (rt)_{ik} = 0;$$
$$\sum_{j \text{ or } k} (ct)_{jk} = 0;\ \sum_{i,j \text{ or } k} (rct)_{ijk} = 0.$$

The question may be raised as to whether concern with interactions is justified. The answer is emphatically "yes" since we can make them large by an appropriate transformation of the scale of observation (and by the same token may be able to make them small). If there are no interactions, (1) reduces to

$$Y_{ijk} = \mu + r_i + c_j + t_k \qquad (2)$$

which states, for instance that treatments k and k' always differ by exactly $t_k - t_{k'}$, regardless of the row or column in which the difference is evaluated. In this case, a basis (usually non-statistical) for inferring from the experimental population to another (usually broader) population of real interest is that additivity holds also over the "target" population. This can, of course, only be an assumption.

To take account (in our formal specification of a conceptual frame of reference or population) of additional variabilities such as errors of measurement of responses, variabilities in treatment application and variation in the state of units (independent of treatment randomization), we make the assumption that the observable response if treatment k were applied to unit (ij) is a random variable y_{ijk},

$$y_{ijk} = Y_{ijk} + \epsilon_{ijk}, \tag{3}$$

where the ϵ_{ijk} are assumed to be uncorrelated random variables with means 0 and constant variance σ^2. The ϵ_{ijk} contain all technical error (error of technique, measurement, and sampling if it is used). We have thus introduced an array of RCT populations, represented by the RCT random variables $\{y_{ijk}\}$. (The distributional assumptions of the ϵ's can be weakened, at the cost of somewhat more awkward formulas, for instance we might assume that $E(\epsilon_{ijk}) = m_{ijk}$ (known) or $V(\epsilon_{ijk}) = \sigma_{ij}^2$. Because of the randomization of treatments to experimental units, the latter assumption, that the variance of technical errors depends on the experimental unit, will not complicate our statistical assessment of treatment differences, at least with regard to their estimated variances.)

DEFINITIONS

We have now described the RCT conceptual populations which we take as the frame of reference for analysis of the experimental data and proceed to the definition of additional parameters which we shall use. As measures of dispersion for the various sets of defined effects and interactions we use components of variation as follows:

$$\begin{aligned}
\sigma_r^2 &= \frac{1}{R-1}\sum_i r_i^2; \quad \sigma_c^2 = \frac{1}{C-1}\sum_j c_j^2; \quad \sigma_t^2 = \frac{1}{T-1}\sum_k t_k^2; \\
\sigma_{rc}^2 &= \frac{1}{(R-1)(C-1)}\sum_{ij} (rc)_{ij}^2; \quad \sigma_{rt}^2 = \frac{1}{(R-1)(T-1)}\sum_{ik} (rt)_{ik}^2; \\
\sigma_{ct}^2 &= \frac{1}{(C-1)(T-1)}\sum_{jk} (ct)_{jk}^2; \quad \sigma_{rct}^2 = \frac{1}{(R-1)(C-1)(T-1)}\sum_{ijk} (rct)_{ijk}^2.
\end{aligned} \tag{4}$$

In general each of these is a sum of squares divided by the number of quantities less the number of linear dependencies. These components of variation are related to Gini's mean squared difference and its generalizations. We have avoided the term components of variance because of possible confusion with the conventional definition of variance of a random variable.

RESULTS ON EXPECTATIONS OF MEAN SQUARES

We use the term "expectation" in the standard statistical sense, meaning average (weighted by probabilities if unequal) over the population of repetitions, and specifically this includes the population of randomizations.

We shall first give the expectations of mean squares of the usual analysis of variance in a condensed form which displays the inherent mathematical structure (though displaying *not at all* the structure of greatest experimental interest). In Table 1 they are given in terms of subscripted quantities Σ (read as cap sigma), the definitions of which are given in the lower part of the table.

TABLE 1

EXPECTATIONS OF MEAN SQUARES (condensed form)

	Mean Square	Expectation
Rows	R^*	$\Sigma_0 + t\Sigma_r$
Columns	C^*	$\Sigma_0 + t\Sigma_c$
Treatments	T^*	$\Sigma_0 + t\Sigma_t$
Discrepance	D^*	Σ_0

Definitions of the Σ's

$$\Sigma_r = \sigma_r^2 - \frac{1}{C}\sigma_{rc}^2 - \frac{1}{T}\sigma_{rt}^2 + \frac{1}{CT}\sigma_{rct}^2$$

$$\Sigma_c = \sigma_c^2 - \frac{1}{R}\sigma_{rc}^2 - \frac{1}{T}\sigma_{ct}^2 + \frac{1}{RT}\sigma_{rct}^2$$

$$\Sigma_t = \sigma_t^2 - \frac{1}{R}\sigma_{rt}^2 - \frac{1}{C}\sigma_{ct}^2 + \frac{1}{RC}\sigma_{rct}^2$$

$$\Sigma_{rc} = \sigma_{rc}^2 - \frac{1}{T}\sigma_{rct}^2$$

$$\Sigma_{rt} = \sigma_{rt}^2 - \frac{1}{C}\sigma_{rct}^2$$

$$\Sigma_{ct} = \sigma_{ct}^2 - \frac{1}{R}\sigma_{rct}^2$$

$$\Sigma_{rct} = \sigma_{rct}^2$$

$$\Sigma_0 = \sigma^2 + \Sigma_{rct} + \Sigma_{rc} + \Sigma_{rt} + \Sigma_{ct}$$

The Σ's have an obvious and simple structure in terms of the components of variation. They are not necessarily positive and cannot therefore be measures of dispersion of any population.

The equations which give the components of variation in terms of the Σ quantities are also simple: for example

$$\sigma_t^2 = \Sigma_t + \frac{1}{R}\Sigma_{rt} + \frac{1}{C}\Sigma_{ct} + \frac{1}{R_c}\Sigma_{rct}$$

and

$$\sigma_{rt}^2 = \Sigma_{rt} + \frac{1}{C}\Sigma_{rct}.$$

Writing the expectations of mean squares in terms of the Σ quantities makes it immediately obvious just what can be estimated unbiassedly by linear combinations of mean squares, namely linear combinations of Σ_0, Σ_r, Σ_c and Σ_t.

The "natural" or common way of estimating σ_t^2 is by $1/t(T^*-D^*)$ which in fact estimates Σ_t and as an estimate of σ_t^2 is subject to a bias of

$$-\frac{1}{R}\sigma_{rt}^2 - \frac{1}{C}\sigma_{ct}^2 + \frac{1}{RC}\sigma_{rct}^2$$

which will usually be negative. It is noteworthy that the magnitude of the bias depends on the absolute magnitudes of R and C and not their magnitudes relative to t.

An alternative form for the expectations of mean squares is also useful. Table 2 gives them in terms of the components of variation, which are the quantities of experimental interest.

The results for the special case $t=R=C=T$, which corresponds to the "usual" Latin square design, are given in Table 3, for comparison with other work (e.g. Neyman et al. [12]).

TABLE 2

EXPECTATIONS OF MEAN SQUARES

Mean Square	Expectation
R^*	$\sigma^2 + \left(\lambda + \frac{t}{CT}\right)\sigma_{rct}^2 + \frac{(C-t)}{C}\sigma_{rc}^2 + \frac{(T-t)}{T}\sigma_{rt}^2 + \sigma_{ct}^2 + t\sigma_r^2$
C^*	$\sigma^2 + \left(\lambda + \frac{t}{RT}\right)\sigma_{rct}^2 + \frac{(R-t)}{R}\sigma_{rc}^2 + \sigma_{rt}^2 + \frac{(T-t)}{T}\sigma_{rt}^2 + t\sigma_c^2$
T^*	$\sigma^2 + \left(\lambda + \frac{t}{RC}\right)\sigma_{rct}^2 + \sigma_{rc}^2 + \frac{(R-t)}{R}\sigma_{rt}^2 + \frac{(C-t)}{C}\sigma_{rt}^2 + \sigma_t^2$
D^*	$\sigma^2 + \lambda\sigma_{rct}^2 + \sigma_{rc}^2 + \sigma_{rt}^2 + \sigma_{ct}^2$
	$\lambda = \left(1 - \frac{1}{R} - \frac{1}{C} - \frac{1}{T}\right)$

TABLE 3

A SPECIAL CASE ($t = R = C = T$)

Mean Square	
R^*	$\sigma^2 + \left(1 - \frac{2}{t}\right)\sigma_{rct}^2 + \sigma_{ct}^2 + t\sigma_r^2$
C^*	$\sigma^2 + \left(1 - \frac{2}{t}\right)\sigma_{rct}^2 + \sigma_{rt}^2 + t\sigma_c^2$
T^*	$\sigma^2 + \left(1 - \frac{2}{t}\right)\sigma_{rct}^2 + \sigma_{rc}^2 + t\sigma_t^2$
D^*	$\sigma^2 + \left(1 - \frac{3}{t}\right)\sigma_{rct}^2 + \sigma_{rc}^2 + \sigma_{rt}^2 + \sigma_{ct}^2$

DISCUSSION OF EXPECTATIONS

(1) When we consider the Latin square as a design for comparing treatments, we see that if in fact the treatments are identical then σ_{rct}^2, σ_{rt}^2, σ_{ct}^2 and σ_t^2 will be zero and the expectations of treatment mean square and discrepance mean square are equal, regardless of σ_{rc}^2, which reflects the interactions of rows and columns. This was first stated by Fisher in 1926 and is the basis for use of the design to evaluate treatments.

(2) If the treatments and experimental units are additive then σ_{rct}^2, σ_{rt}^2 and σ_{ct}^2 will be zero and the quantity $1/t(T^*-D^*)$ estimates σ_t^2 unbiassedly. (The case of additivity for $t=R=C=T$ was treated by Kempthorne [10, Section 10.5].)

(3) Neyman et al. [12] in studying the case $t=R=C=T$ came to the conclusion that the Latin square was always positively biassed, in the sense that, if $\sigma_t^2=0$ then $E(T^*)$ is always greater than $E(D^*)$. Our results indicate that the reverse will ordinarily be true. The origin of the discrepancy lies in the fact that they did not include the row-treatment and column-treatment interactions in their manipulations, and this led to the incorrect conclusion that $1/tE(T^*-D^*)$ equals $\sigma_t^2+(1/RT)\sigma_{rct}^2$. This is not the form of the result of Neyman et al. because they made additional homogeneity assumptions.

(4) As an error term for treatments D^* is usually too big on the average, but if R and C are large the overestimation of error is unimportant.

(5) There appears to be no reasonable case for considering the usual agronomic Latin square experiment to have an origin such as the sampling of rows and columns here considered, even if the actual field in which the experiment is performed is a very small part of the population of interest. It must be remembered that in the situation examined here the whole population of interest is stratified by rows and columns *before* the sample of experimental material for the experiment is drawn.

(6) It would appear to be rare that, with sufficient heterogeneity of experimental material to warrant a two-way classification, treatments would act additively with respect to both rows and columns. It may be that the three-way interactions of rows, columns and treatments is often small but the usual way a Latin square experiment is done, i.e. taking rows and columns in such a way as to account for appreciable variability of the units, is such as to make row and column interactions with treatments very likely, because the widespread dependence of response to treatment on yield with any particular treatment would result in such interactions.

(7) The Latin square design does in general reduce the actual error associsociated with the evaluation of treatments (as may be seen from Table 2 by using the fact that the average variance of treatment differences is proportional to the treatment mean square after deleting the σ_t^2 term) but at the same time may seriously over-estimate the error if R or C

are not large. Most agronomic uses of the Latin square design fall in th category. The implications of this in specific situations should be considered. The design will certainly be good where anticipated treatment differences are large and their accurate "point" estimation is the main concern.

(8) In most statistical experimental situations it is desirable to consider tests of significance and tests of hypotheses about treatment effects. We accept the view that tests of significance are evaluatory procedures leading to assessments of strength of evidence against particular hypotheses, while tests of hypotheses are decision devices. We are here concerned with the former, and in this connection it should be noted that

(a) the expectations of mean squares are in some degree irrelevant to the exact (permutation) test of significance of the null hypothesis that the treatments are identical. The procedure by which this null hypothesis would be evaluated, in strict theory at least, would be to superimpose on the observations obtained with a particular plan all possible randomizations of plans (but not of selection of rows or columns) and determine the frequency with which the value of a chosen criterion for the plan actually observed is equalled or exceeded in the totality of possible plans. The effect of non-additivities under such circumstances is essentially unknown. (Some more detailed discussion of this is given by Kempthorne [10, p. 149 and Section 12.6].)

(b) if the observed treatment mean square is less than the observed error mean square and the variance ratio criterion is to be used in a randomization test it is difficult to see how a high level of significance can possibly be reached since, with given observations, the average content of treatment and discrepance mean squares are equal over the set of all randomizations. If one thinks of the sensitivity of an experiment as the average of squared treatment differences divided by the average error in estimation of these, then it appears that if R and C are not large and row-treatment and column-treatment interactions are important then the overestimation of error may occasion a serious underassessement of the sensitivity (or value) of the experiment. The desirable direction to move in is toward a scale of analysis where interactions are less important. (While the overestimation of error is less serious for the randomized block design the increase in *actual error* will often more than outweigh this. This matter is discussed further below.)

(9) What we have said above has the consequence that if experimenters would follow a rule of discarding Latin square experiments in which the treatment mean square is not significantly greater than the error mean square, say, they might well be discarding very good estimates of treatment main effect comparisons.

(10) In view of the fact that it is desirable for an experimenter not only to

have "good" estimates but also to know how "good" the estimates are, the strong recommendation is made to attempt to transform to a scale where additivity more nearly obtains. How to proceed is an outstanding problem, and, in our opinion, one of the most important in the analysis of experiments. Offhand, it seems that the practice of transforming to get a high value of F, the variance ratio, a procedure which has fallen into disrepute (perhaps because experimenters were using it for the wrong reasons) has some commendable aspects.

(11) In circumstances where row-treatment and column-treatment interactions are important and R and C not large it may be that $1/tT^*$ is a better estimate of σ_t^2 than $1/t(T^*-D^*)$. Also if for example R is large and C equals t the expression $1/2t(2T^*-D^*)$ may be a better estimate of σ_t^2 than the usual one. In evaluating such alternate estimations for σ_t^2 the possible difficulties or advantages associated with under-or-over-estimation must be most carefully considered.

COMPARISON WITH RANDOMIZED BLOCKS

We now consider the relative merits of the randomized block design (Wilk [21, 23 II and VI]) and the Latin square design. We do not consider the randomized block and Latin square analyses of variance for a particular set of observations, but rather we consider the possible results over all randomizations and technical errors for each design. The randomized block design includes very many more plans than the Latin square design. For instance, for an experiment involving four replications for each of four treatments, there are 576 different 4×4 Latin square plans and $(4!)^4$ or 576^2 different randomized block plans. The question is whether one should use a random *one* of the 576 different Latin square plans or a random *one* for the 576^2 different randomized block plans each with the analysis appropriate to the design. A partial answer is given by considering the expectations of mean squares in the two cases. These are given in Table 4 for the case of rows as blocks in the randomized block design. The randomized block error mean square is denoted by I^*.

The expectations of Table 4 are in terms of the Σ's as defined for the Latin square. (They would have a rather different but very simple structure in terms

TABLE 4

EXPECTATIONS OF MEAN SQUARES WITH RANDOMIZED BLOCKS

Mean Square	Expectation
B^*	$\Sigma_0 + \frac{(C-t)}{C}\Sigma_c - \frac{1}{C}\Sigma_{ct} + t\Sigma_r$
T^*	$\Sigma_0 + \Sigma_c + \frac{(t-1)}{C}\Sigma_{ct} + t\Sigma_t$
I^*	$\Sigma_0 + \Sigma_c - \frac{1}{C}\Sigma_{ct}$

of Σ's appropriate to a randomized block design.) We see that $1/t(T^*-I^*)$ is an unbiassed estimate of $\Sigma_t+1/C\Sigma_{ct}$ which is equal to $\sigma_t^2-1/R\sigma_{rt}^2$. The bias in the randomized block analysis is $-1/R\sigma_{rt}^2$. If C is small and column-treatment interactions are important then the bias in estimating σ_t^2 from a randomized block experiment may be appreciably less than if a Latin square had been used. On the other hand, if column main effects are important, the variance of comparisons of treatment effects will be much increased, as can be seen from the excess of the expectation of the treatment mean square with randomized blocks over that with the Latin square, namely

$$\begin{aligned}\Sigma_c+\frac{(t-1)}{C}\Sigma_{ct} &= \sigma_c^2-\frac{1}{T}\sigma_{ct}^2-\frac{1}{R}\sigma_{rc}^2+\frac{1}{RT}\sigma_{rct}^2\\ &\quad+\frac{t-1}{C}\sigma_{ct}^2-\frac{t-1}{RC}\sigma_{rct}^2\\ &= \sigma_c^2-\left(\frac{1}{T}-\frac{t-1}{C}\right)\sigma_{ct}^2-\frac{1}{R}\sigma_{rc}^2\\ &\quad+\frac{1}{R}\left(\frac{1}{T}-\frac{t-1}{C}\right)\sigma_{rct}^2.\end{aligned}$$

If R is large compared to r, a situation commonly referred to as "rows are random," the comparison is given in Table 5.

TABLE 5

COMPARISON OF RANDOMIZED BLOCKS AND LATIN SQUARE (ROWS RANDOM)

Expectation of mean squares

	Latin Square	Randomized Blocks
Treatments	$\Sigma_0+t\left(\sigma_t^2-\frac{1}{C}\sigma_{ct}^2\right)$	$\Sigma_0+\sigma_c^2-\left(\frac{1}{T}+\frac{1}{C}\right)\sigma_{ct}^2+t\sigma_t^2$
Error	Σ_0	$\Sigma_0+\sigma_c^2-\left(\frac{1}{T}+\frac{1}{C}\right)\sigma_{ct}^2$

From Table 5 we may guess that even though in this case randomized blocks allows unbiassed estimation of σ_t^2, the Latin square will have in general both a lower real and a lower apparent error. This serves to underline the fact that the Latin square may well give more accurate estimates of treatment comparisons even though their errors cannot be estimated unbiassedly.

As a hypothetical example let $C=T=t$, $\sigma_{rct}^2=0$, $\sigma_{rc}^2=\sigma_{ct}^2=\sigma_{rt}^2=\sigma_0^2$, $\sigma^2=0$, $\sigma_c^2=2\sigma_0^2$. Then if we define the non-centrality factor (NCF.) to be

$$\frac{E\text{ (treatment mean square)}-E\text{ (error mean square)}}{2E\text{ (error mean square)}}$$

we have

$$NCF(L \cdot S) - NCF(RB) = \frac{2t(t-1)}{6(5t-2)} \frac{\sigma_t^2}{\sigma_0^2} - \frac{1}{6}$$

which is positive whenever σ_t^2/σ_0^2 is larger than

$$\frac{1}{2t} + \frac{2}{t-1} - \frac{1}{2t(t-1)}.$$

With t equal to 4 this is $\frac{3}{4}$, so that if $\sigma_t^2 > \frac{3}{4}\sigma_0^2$, $NCF(LS) > NCF(RB)$.

In the case when columns are "random," i.e. C is large compared to t, the situation is as given in Table 6. In this case we note that the bias in estimation of σ_t^2 is the same for both cases. The Latin square will however usually be better for estimation of treatment differences and will usually have a larger non-centrality factor.

TABLE 6

COMPARISON OF RANDOMIZED BLOCKS AND LATIN SQUARE (COLUMNS RANDOM)

Expectation of mean squares

	Latin Square	Randomized Blocks
Treatments	$\Sigma_0 + t\left(\sigma_t^2 - \frac{1}{R}\sigma_{rt}^2\right)$	$\Sigma_0 + \left(\sigma_c^2 - \frac{1}{R}\sigma_{rc}^2 - \frac{1}{T}\sigma_{ct}^2 + \frac{1}{RT}\sigma_{rct}^2\right) + t\left(\sigma_t^2 - \frac{1}{R}\sigma_{rt}^2\right)$
Error	Σ_0	$\Sigma_0 + \left(\sigma_c^2 - \frac{1}{R}\sigma_{rc}^2 - \frac{1}{T}\sigma_{ct}^2 + \frac{1}{RT}\sigma_{rct}^2\right)$

The general conclusion from the results given above is that in general the Latin square design is likely to be better than the randomized block design though it is obvious that, if the column classification of experimental units should be useless in the sense of not accounting for excess variability, they will be essentially equivalent apart from the loss of degrees of freedom for error resulting from the use of column classification in the Latin square.

A GENERALIZATION

There are a number of sampling situations, in addition to the one considered above, within which a Latin square design might be embedded. For example we may have S sources of RC units classified as above; select and use t^2 units from each of s of them, and examine t selected treatments in s Latin squares, each $t \times t$. Our previous situation is a special case with $S = s = 1$. The results for expectations of mean squares for this case are given in Table 7.

The appropriate population model for this situation has the form

$$y_{gijk} = \mu + s_g + r_{gi} + c_{gj} + t_k + (st)_{gk} + (rt)_{gik} + (ct)_{gjk} + (rct)_{gijk} + \epsilon_{gijk}$$

TABLE 7

GENERALIZATION OF LATIN SQUARE DESIGN SITUATION

Mean Square	*df*	Expectation
S^*	$(s-1)$	$\Sigma_0 + t\Sigma_r + t\Sigma_c + t\Sigma_{st} + t^2\Sigma_s$
R^*	$s(t-1)$	$\Sigma_0 + t\Sigma_r$
C^*	$s(t-1)$	$\Sigma_0 + t\Sigma_c$
T^*	$(t-1)$	$\Sigma_0 + t\Sigma_{st} + st\Sigma_t$
I^*_{ST}	$(s-1)(t-1)$	$\Sigma_0 + t\Sigma_{st}$
D^*	$s(t-1)(t-2)$	Σ_0

Definitions of Σ's

$$\Sigma_s = \sigma_s^2 - \frac{1}{T}\sigma_{st}^2 - \frac{1}{R}\sigma_r^2 - \frac{1}{C}\sigma_c^2 + \frac{1}{RT}\sigma_{rt}^2 + \frac{1}{CT}\sigma_{ct}^2 + \frac{1}{RC}\sigma_{rc}^2 - \frac{1}{RCT}\sigma_{rct}^2$$

$$\Sigma_t = \sigma_t^2 - \frac{1}{S}\sigma_{st}^2$$

$$\Sigma_r = \sigma_r^2 - \frac{1}{T}\sigma_{rt}^2 - \frac{1}{C}\sigma_{rt}^2 + \frac{1}{CT}\sigma_{rct}^2$$

$$\Sigma_c = \sigma_c^2 - \frac{1}{T}\sigma_{ct}^2 - \frac{1}{R}\sigma_{rc}^2 + \frac{1}{RCT}\sigma_{rct}^2$$

$$\Sigma_{rc} = \sigma_{rc}^2 - \frac{1}{T}\sigma_{rct}^2$$

$$_{rt} = \sigma_{rt}^2 - \frac{1}{C}\sigma_{rct}^2$$

$$\Sigma_{ct} = \sigma_{ct}^2 - \frac{1}{T}\sigma_{rct}^2$$

$$\Sigma_{rct} = \sigma_{rct}^2$$

$$\Sigma_0 = \sigma^2 + \Sigma_{rct} + \Sigma_{rc} + \Sigma_{rt} + \Sigma_{ct},$$

where $g=1, 2, \cdots, S$; $i=1, 2 \cdots, R$ (for each g); $j=1, 2, \cdots, R$ (for each g); $k=1, 2, \cdots, T$.

The definition of the components of variation is then as exemplified by

$$\sigma_s^2 = \frac{1}{S-1}\sum_g s_g^2; \qquad \sigma_r^2 = \frac{1}{S(r-1)}\sum_{gi} r_{gi}^2;$$

$$\sigma_{rc}^2 = \frac{1}{S(R-1)(C-1)}\sum_{gij}(rc)_{gij}^2; \text{ etc.}$$

The inverse relationship between the σ^2's and the Σ's is as indicated by

$$\sigma_t^2 = \Sigma_t + \frac{1}{S}\left(\Sigma_{st} + \frac{1}{R}\Sigma_{rt} + \frac{1}{C}\Sigma_{ct} + \frac{1}{RC}\Sigma_{rct}\right);$$

$$\sigma_s^2 = \Sigma_s + \frac{1}{T}\Sigma_{st} + \frac{1}{R}\Sigma_r + \frac{1}{C}\Sigma_c + \frac{1}{RT}\Sigma_{rt} + \frac{1}{CT}\Sigma_{ct} + \frac{1}{RCT}\Sigma_{rct}$$

From Table 6 it can be seen that if S is large, then $1/st(T^* - I_{st}^*)$ is essentially unbiassed for σ_t^2, and I_{st}^* would be an appropriate error term for evaluation of observed treatment comparisons, and the assessment or estimation of σ_t^2.

MACHINERY: THE STATISTICAL MODEL

We have found it of considerable value in many respects to formalize and summarize the implications of the randomization procedures and the various assumptions underlying the population model in an explicit statistical model for the observations. This will be described rather briefly for the main situation described above.

We define various sets of "dummy" random variables which reflect and define the essential features in the sampling and randomization procedure. They are as follows:

(1) $\alpha_i^{i^*} =$ 1 if the i^*th selected row is the ith row in the population
$= 0$ otherwise

(2) $\beta_j^{j^*} =$ 1 if the j^*th selected column is the jth column in the population
$= 0$ otherwise

(3) $\gamma_k^{k^*} =$ 1 if the k^*th selected treatment is the kth treatment in the population
$= 0$ otherwise

(4) $\delta_{i^*j^*}^{k^*f} = 1$ if the fth replicate of the k^*th selected treatment occurs on selected i^* and selected column j^*
$= 0$ otherwise

and finally

$\rho_{i^*j^*}^{k^*} = \sum_f \delta_{i^*j^*}^{k^*f}$ which is unity if the k^*th selected treatment occurs on the i^*th selected row and the j^*th selected column.

The α's, β's, γ's and δ's are groupwise independent and all of their distributional properties can in principle be written down or worked out. In the case of the δ's (and the derived ρ's) this may be difficult or, at present, impossible because it would require enumeration of all $t \times t$ Latin squares. It may be recalled that Welch [18] had to use a complex enumeration scheme to obtain some of the less elementary properties for 5×5 and 6×6 squares. In the present case, where we are only concerned with first and second moments the necessary properties of the δ's or ρ's are readily obtained. (They are given with different notation by Kempthorne [10, Section 10.5].)

Using these random variables, explicit models for the observations are

$$\begin{aligned} z_{i^*j^*} &= \sum_{i,j,k,k^*,f} \alpha_i^{i^*}\beta_j^{j^*}\gamma_k^{k^*}\rho_{i^*j^*}^{k^*}y_{ijk} \\ &= \mu + \sum_i \alpha_i^{i^*}r_i + \sum_j \beta_j^{j^*}c_j + \sum_k \gamma_k^{k^*}\rho_{i^*j^*}^{k^*}t_k \\ &\quad + \sum_{ij} \alpha_i^{i^*}\beta_j^{j^*}(rc)_{ij} + \sum_{ikk^*} \alpha_i^{i^*}\gamma_k^{k^*}\rho_{i^*j^*}^{k^*}(rt)_{ik} \\ &\quad + \sum_{jkk^*} \beta_j^{j^*}\gamma_k^{k^*}\rho_{i^*j^*}^{k^*}(ct)_{jk} + \sum_{ijkk^*} \alpha_i^{i^*}\beta_j^{j^*}\gamma_k^{k^*}\rho_{i^*j^*}^{k^*}[(rct)_{ijk} + \epsilon_{ijk}] \end{aligned}$$

and

$$x_{k^*f} = \sum_{i^*j^*} \delta_{i^*j^*}^{k^*f} z_{i^*j^*}$$

An explicit expression for x_{k^*f} can be written down from the relations above. The case when the whole of a population is represented in the sample, is included formally in a simple way: for example, if all rows are included we let $\alpha_i^{i^*}=1$ for i^* equal to i and zero otherwise.

The models given above appear to have a rather complex and formidable structure. One might ask if they are really necessary. Our own development of such models was part of a general investigation [21, 23] into the meaning of the linear models which are commonly used to present the analyses of experimental designs. Such apparently complex models arose quite naturally and we have found them to be easy to handle. They can be used to study many statistical properties of interest, such as expectations of linear estimates, variance of estimates, expectations of mean squares, and estimates of error, by formal algebraic expectation procedures for any experimental design. There is no reason in principle why the models could not be used to find variances and covariances of mean squares. It may be expected however that more purely combinatorial arguments will be developed to handle these problems [Hooke, 9].

Quite apart from the fact that the over-all method has been useful in studying the standard designs, its most important attributes are:

(1) it lays bare the dependence of the meanings of effects and interactions on the populations involved, e.g. of treatment effects, on the populations of rows, columns, etc. in the general experimental situation under study.

(2) it exhibits explicitly the dependence of the meanings of effects and interactions on the scale of observation.

(3) it brings into the open the necessary assumptions or considerations for meaningful interpretation of the analysis of variance. (Here "meaningful" is in contrast with writing down an aribtrary, albeit customary, linear model.)

(4) it helps to indicate which assumptions are very important and which are unimportant.

(5) it helps to bring to the fore one of the basic problems of quantitative experimentation, namely that of finding the functional structure relationship of the variables underlying an experimental situation.

LINEAR ESTIMATES AND ERRORS

We conclude with a short section on estimation of effects. For simplicity we consider the case where treatments are fixed, i.e. $T=t$, so that we can write an observation as x_{kf}. Then as in all cases of randomized designs that we know of, unbiassed estimates of linear population parameters are given by the same linear functions of appropriate sample means. In the present case, regardless of interactions and relative values of R, C, and t an unbiassed estimate of $\sum_k g_k t_k$ is given by $\sum_k g_k(x_{k.}-x_{..})$. Difficulties arise in the estimation of errors of estimates. We may note the following cases:

(1) with interactions of treatments with rows and columns negligible,

$$\text{var}\,(x_{k.} - x_{kl}) = \frac{2}{t}\,(\sigma^2 + \sigma_{rc}^2)$$

which is estimated unbiassedly using D^*. This is of course well known and due to Fisher [6].

(2) as a guide under general conditions, the average variance of estimated treatment differences is, of course, directly connected with the expectation of the treatment mean square and is

$$\frac{2}{t}\left[\sigma^2 + \sigma_{rc}^2 + \left(R + \frac{t}{RC}\right)\sigma_{rct}^2 + \frac{(R-t)}{R}\sigma_{rt}^2 + \frac{(C-t)}{C}\sigma_{ct}^2\right] = \frac{2}{t} \quad (\pounds)$$

and this can be seriously overestimated by $2/tD^*$ if non-additivities are important or R and C are not large. For randomized blocks with rows as blocks the average variance is

$$\frac{2}{t}\left[(\pounds) + \left(\frac{t-1}{C} - \frac{1}{T}\right)\left(\sigma_{ct}^2 - \frac{1}{R}\sigma_{rct}^2\right) + \left(\sigma_c^2 - \frac{1}{R}\sigma_{rc}^2\right)\right]$$

and it is easy to visualize situations in which this average variance will be considerably larger than with the Latin square design. On the other hand the estimate of variance from the randomized block design may be better and less biassed. Where either design is reasonable, the Latin square is usually to be recommended but with some caution to the effect that its error term will usually be too large.

REFERENCES

[1] Bennett, C. A., and Franklin, H. L., *Statistical Analysis in Chemistry and the Chemical Industry*. New York: John Wiley and Sons, 1954.

[2] Cornfield, J., "On samples from finite populations," *Journal of the American Statistical Association*, 39 (1944), 236–9.

[3] Cornfield, J., "The two-way classification in the analysis of variance from the point of view of urn sampling," Office of Biometry (1953). Mimeographed report.

[4] Cornfield, Jerome, and Tukey, John W., "Average values of mean squares in factorials," *Annals of Mathematical Statistics*, 27 (1956), 907–49.

[5] Fisher, R. A., *Statistical Methods for Research Workers*. First Edition. Edinburgh: Oliver and Boyd, 1925.

[6] Fisher, R. A., "The arrangement of field experiments," *Journal of the Ministry of Agriculture*, 33 (1926), 503–13.

[7] Fisher, R. A., *The Design of Experiments*. London: Oliver and Boyd, 1935.

[8] Fisher, R. A., and Yates, F., "The 6×6 Latin squares," *Proceedings of the Cambridge Philosophical Society*, 30 (1934), 492–507.

[9] Hooke, R., "Some applications of Bipolykays to the estimation of variance components and their moments," *Annals of Mathematical Statistics*, 27 (1956), 80–98.

[10] Kempthorne, Oscar, *The Design and Analysis of Experiments*. New York: John Wiley and Sons, 1952.

[11] Kempthorne, Oscar, "The randomization theory of experimental inference," *Journal of the American Statistical Association*, 50 (1955), 946–67.

[12] Neyman, J., with the cooperation of Iwaskiewicz, K. and St. Kolodzieczyk, "Statistical problems in agricultural experimentation," Supplement to the *Journal of the Royal Statistical Society*, 2 (1935), 107–54.

[13] Norton, H. W., "The 7×7 Latin squares," *Annals of Eugenics*, 9 (1939), 269–307.

[14] Pitman, E. J. G., "Significance tests which can be applied to samples from any populations, III, the analysis of variance test," *Biometrika*, 29 (1937) 322–35.

[15] Scheffé, H., "A 'mixed model' for the analysis of variance," *Annals of Mathematical Statistics*, 27 (1956), 23–36.

[16] Smith, H. F., "Variance components, finite populations and statistical inference," North Carolina Institute of Statistics Mimeo Series 135 (1955).

[17] Tukey, J. W., "Interaction in a row by column design," Princeton University: Memorandum Report 18 (1949).

[18] Welch, B. L., "On the z-test in randomized blocks and Latin squares," *Biometrika*, 29 (1937), 21–32.

[19] Wilk, M. B., "The randomization analysis of a generalized randomized block design," Iowa State College: M.S. thesis, 1953.

[20] Wilk, M. B., "The randomization analysis of a generalized randomized block design," *Biometrika*, 42 (1955), 70–9.

[21] Wilk, M. B., "Linear models and randomized experiments," Iowa State College: Ph.D. thesis, 1955.

[22] Wilk, M. B., and Kempthorne, Oscar, "Some aspects of the analysis of factorial experiments in a completely randomized design," *Annals of Mathematical Statistics*, 27 (1956), 950–84.

[23] Wilk, M. B., "The logical derivation of linear models for experimental situations and their use in selecting the appropriate error term in the analysis of variance. I. AV-3; II. On the analysis of general balanced randomized block type design. AV-4; III. (with O. Kempthorne) The expectations of mean squares for a general three-way classification. AV-7; IV (with O. Kempthorne) Fixed, mixed and random models. AV-8; V. The expectation of mean squares for a general three-way classification (Part 2). AV-9; VI. Interactions with experimental units. AV-10. Analysis of variance project reports, Statistical Laboratory, Iowa State College (1953–1954).

[24] Yates, F., "The formation of Latin squares for use in field experiments," *Empire Journal of Experimental Agriculture*, 1 (1933), 235–44.

[25] Yates, F., "Complex experiments," Supplement to the *Journal of the Royal Statistical Society*, 2 (1933), 181–247.

[26] Yates, F., "Incomplete Latin squares," *Journal of Agricultural Sciences*, 26 (1936), 301–15.