

AVERAGE VALUES OF MEAN SQUARES IN FACTORIALS¹

BY JEROME CORNFIELD AND JOHN W. TUKEY

National Institutes of Health and Princeton University

1. Summary. The assumptions appropriate to the application of analysis of variance to specific examples, and the effects of these assumptions on the resulting interpretations, are today a matter of very active discussion. Formulas for average values of mean squares play a central role in this problem, as do assumptions about interactions. This paper presents formulas for crossed (and, incidentally, for nested and for non-interacting completely randomized) classifications, based on a model of sufficient generality and flexibility that the necessary assumptions concern only the selection of the levels of the factors and not the behavior of what is being experimented upon. (This means, in particular, that the average response is an arbitrary function of the factors.) These formulas are not very complex, and specialize to the classical results for crossed and nested classifications, when appropriate restrictions are made.

Complete randomization is only discussed for the elementary case of "no interactions with experimental units" and randomized blocks are not discussed. In discussion and proof, we give most space to the two-way classification with replication, basing our direct proof more closely on the proof independently obtained by Cornfield [17], than on the earlier proof by Tukey [20]. We also treat the three-way classification in detail. Results for the general factorial are also stated and proved.

The relation of this paper to other recent work, published and unpublished, is discussed in Section 4 (average values of mean squares) and in Section 11 (various types of linear models).

INITIAL DISCUSSION

2. Introduction. During the last years of the last decade it was relatively easy to believe that the analysis of variance was well understood. Eisenhart's summary article of 1947 [5], when combined with the work of Pitman [13] and Welch [15] on the randomization approach (work published in 1937-1938, which ever since has been far too much neglected), seemed to provide a simple, easily understandable account of the foundations. But as the years have passed, both statisticians and users of analysis of variance have gradually become aware of a number of areas in which we needed to deepen our understanding. One of these is the relation of formulas for average values of mean squares to assumptions. These are of central importance, since the choice of an "error term" as a basis for *either*

Received May 31, 1955; revised May 24, 1956.

¹ Based in part on the Statistical Research Group, Princeton University, Memorandum Report 18, "Interaction in a row-by-column design," prepared in connection with research sponsored by the Office of Naval Research.

significance tests or confidence statements *must at least* take into account average values of mean squares. (It would be very desirable to consider, also, variances, covariances, and distributions of mean squares, particularly if one is concerned with the detailed validity of an F -test or a multiple-comparisons procedure. Average values, however, seem absolutely essential.) These average values should apply when there are real effects and should not be confined to so called "null" hypotheses. To derive them we must make assumptions, but they should be as weak as possible.

The present paper deals with one aspect of the problem of average mean squares. After introducing this aspect, and describing the situation which we are to treat, we shall relate both this aspect to other aspects of current interest and our work to the work of others.

Difficulties in applying the analysis of variance to even a two-way model have arisen in the so-called mixed model where rows are sampled and columns are completely enumerated. Of two recent standard textbooks, one concludes that the average value of both row and column mean squares includes a component of variance due to interaction [11]; the other finds such a component in column sum of squares, but not in rows [1]. Both texts assume that observed values are linear combinations of certain fixed and random variables, but differ in the nature of the restrictions that are imposed upon these variables. Granting either text its assumptions, its conclusions necessarily follow.

In analyzing the given body of data, the choice among such assumptions can lead to quite different answers. The choice is thus an important one. But it is not a simple choice. At one stage in the development of his ideas about two-way classifications, one of the authors, who leans toward linear models, had 512 alternative sets of assumptions. If they had led to 512 sorts of analysis, the situation would obviously be quite impractical. The only effective way out of such a difficulty was to obtain a single flexible model which could be specialized to any one of the 512 special possibilities. This was done in [20] (and independently in [17] and in [21]).

The question of what assumptions to make seems, at first glance, to be a purely empirical question, one that should be referred to the subject-matter knowledge of the experimenter, who is the expert on such matters. Sometimes this is helpful and sometimes not. But closer study shows that the choice of assumptions depends on more than empirical questions about the behavior of the experimental material. It depends on the nature of the sampling and randomization involved in obtaining the data (as has been recognized by many statisticians, and recently emphasized by Kempthorne and by Wilk). Moreover, it often depends on the purpose of the analysis, as expressed by the situations or populations to which one wishes to make *statistical* inference. These dependences imply diversity, and adequate treatment of diversity requires flexibility of assumption.

After a general initial discussion (Sections 2-6), various descriptions of the situations treated are given (Sections 7-12). The results (Sections 13-19) come next, and are followed by the proofs (Sections 20-28).

3. A pigeonhole model. These requirements of flexibility are met by the system of assumptions which we are about to describe, which specify an object appropriately called a pigeonhole model. Given the values of each of the factors (variables), we are directed to a pigeonhole containing a finite or infinite population. This population represents the possible results of experimentation with these values of the factors. So far we have made no assumptions about how the factors combine to produce the typical effect of each combination. (This typical effect will often be taken as the arithmetic mean of the population of possible values in the corresponding pigeonhole.) We have made the assumption that we can recognize the different levels of each factor, and we shall shortly make assumptions as to how we sample this array of pigeonholes. Neither of these is an assumption about "how the world behaves"; both are assumptions about the experimenter's behavior (and are consequently much easier to check).

We have emphasized the generality of our assumptions; we must also emphasize their limitations. We are concerned with situations where the variability of experimental units (plots, reactors, epochs, mice, etc.) does not play such a predominant role as to require special attention in the design of the experiment. We shall not attempt to treat such well-known situations as randomized blocks, which have been recently studied by Kempthorne [8] and Wilk [16]. (See the next section for a discussion of mutual relations and distinctions.)

Let us be specific about the case of the two-way classification. Let there be RC pigeonholes arranged in R rows and C columns. Let there be at least n elements in the population in each pigeonhole. Let a sample of r rows be drawn from the R potential rows. Let a sample of c columns be drawn from the C potential columns. The rc intersections of a selected row with a selected column specify the rc pigeonholes which become the cells of the actual experiment. In each of these rc cells, let a sample of n elements be drawn. The values of the rcn elements thus obtained are the numbers which are to be analyzed. *Assume that all the samplings—of rows, of columns, and within pigeonholes—are at random and independent of one another.* This is the only assumption we shall make. Note that it is an assumption about the set-up of the experiment and not about the behavior of those things on which the experiment is performed.

While the generality of this model is quite apparent, its flexibility may not be completely evident. If we choose $R = r$ and $C = c$, then rows and columns become fixed, and we have a fixed model which generalizes Model I of Eisenhart [5], since (i) neither normality or constant variance is assumed for the cells, and (ii) no assumption is made about interactions.

If, at the other extreme, we take R and C both infinite, we have a "random" model which generalizes Model II of Eisenhart, since (i) neither normality or constant variance is assumed for the cells, (ii) normality is not assumed for the row and column populations, and (iii) no assumption is made about interactions.

If we take $C = c$ and R infinite, then we obtain a generalization of the conventional mixed model.

Thus our model is flexible enough to cover all the classical cases, and many others besides.

Our model can be described from two apparently quite different points of view, namely, as an example of urn sampling or in terms of a very general linear model. (By urn sampling we shall mean what is usually referred to as sampling without replacements from a finite population.) We shall give both descriptions, striving to take as different points of view as we can about them. An understanding of either description should suffice as a basis for the interpretation of the results.

4. Relation to other problems and workers. Historically, the understanding of the average mean squares, and of the formulas relating them to the underlying models, has developed separately for the different relations which classifications have to one another in the customary situations. The order of development has usually been the same. First, average mean squares are obtained under simplifying assumptions, and then new formulas are obtained as these assumptions are relaxed, usually successively. The initial assumptions, whose removal is often very important, are typically of the following sorts:

- (1) some classification does not interact with one or more others;
- (2) the "errors" are solely (or practically solely) due to the experimental units used;
- (3a) the levels of a certain variable are *fixed*;
- (3b) the levels of a certain variable are a random sample from an *infinite* population.

When we remove these assumptions, we remove (1) or (2) entirely, and replace (3a) or (3b) by

- (3) the levels of a certain variable are a random sample from a population of arbitrary size.

We shall describe assumption (2) as "no free errors" and assumptions (3a) and (3b) as "fixed—" and "infinite sampled—" respectively.

The foremost distinction between the different relations between classifications is the minimum number of classifications which can enter this relation. Thus two (or more) classifications can be crossed or nested. Three (or more) classifications are involved in randomized blocks, in a simple fractional factorial, or in a Latin square. At least four classifications appear in a lattice, and so on.

First, let us discuss the relations which need involve only two classifications:

(1) *Nested* (or as some say, *hierarchical*) *classifications*. Here the general situation was clarified first, and general formulas have been available for some time. (The early clarification of this case may have been due to its intimate relation to the precision of multistage sampling.)

(2) *Crossed classifications*. The case of random interactions was treated early (cf. [5], [8], etc.), and in varying degrees of generality, the most general being in [14].

For the important case of arbitrary interactions, we know of nothing which antedates Memorandum Report 18 [20], which stated a general rule for crossed and nested cases. (However, informed understanding of the limiting cases of "fixed" and "infinitely sampled" cases seems to have been present among some users before this time.) The rule for the limiting cases was stated by Kempthorne

([8], p. 574). The general results for the two-way classification were obtained by Cornfield [17] and by Wilk [21], independently of each other and of Memorandum Report 18. All three approaches to this result were from somewhat different points of view. Kempthorne ([19], pp. 204–220) has discussed the general two-way situation, basing his treatment on [21]. Bennett and Franklin [2] have sketched a proof for the two-way classification (without replication) (pp. 474–477), and have stated the result for the three-way case (p. 394).

Wilk and Kempthorne [24] have treated the case of the general two-way classification with general sampling of the two factors, with replication, and with complete randomization of experimental units, but without free errors or interactions with experimental units. They have also [23] dealt with the three-way classification, with general proportional numbers in a cell, general sampling of all factors, complete randomization of experimental units, free errors only restricted to be of the same variance, but with no interaction with experimental units. Wilk [25] has continued the analysis of the three-way classification, treating both interaction with experimental units, and, separately, the analysis of cell means for general disproportionate cell numbers. Like the earlier work at Ames, this work was carried out independently of ours, in ignorance of the existence of [20] and [17], and before the appearance of [2].

A short proof, using more special techniques, has been found for the two-way classification by Hooke [18], who has been able to obtain variances and covariances as well as average values of mean squares.

(3) *Free random allotment (complete randomization)*. The case of no interaction with experimental units was treated in [14] (p. 73) and is discussed briefly in Section 10 of the present paper. (Other cases appear in [21], [23], [24], and [25].)

The important case of arbitrary interaction was first discussed by Wilk [16], who has treated a more complex case in [25].

Next come the relations involving a minimum of three classifications. The fact that three classifications are necessarily involved does not always appear in the corresponding analysis of variance. Thus in a randomized block experiment, *treatments*, *blocks*, and *plots* must all appear, though there is no trace of plots in the analysis of variance. The recent development of formulas for average mean squares for such relations between classifications has been in the hands of Kempthorne and Wilk, as the detailed summary will now show. The present authors have done no work in this area.

(4) *Randomized blocks*. The case of no interaction with experimental units (with plots) is classical, and formulas are well known.

The important case of interaction with experimental units was first treated, under certain restrictions, by Neyman (with cooperation of Iwaskiewicz and Kolodzieczyk) [12]. Some particular cases were followed up by McCarthy [10]. The case of arbitrary interaction (not distinguishing technical errors) was first discussed in the book of Kempthorne ([8], Sec. 8.4). This treatment was extended to the case where each treatment appears $p > 1$ times per block by Wilk [16].

(5) *Randomized fractionation (as in the classical Latin square where rows and*

columns refer to experimental units). The cases where *all* interactions are assumed zero are classical.

The case of a somewhat restricted interaction of treatments with experimental units was approached by Neyman (with cooperation of Iwaskiewicz and Kolodzieczyk) [12]. We understand that this problem is treated in [23], which we have not seen, and in "complete generality" in further unpublished work of Wilk and Kempthorne.

(6) *Crossed fractionation (as in a simple fractional factorial without blocks)*. Here the results for no interactions are classical. Nothing else seems available.

(7) *Mixed fractionation (as in a "Latin square" with two factors and one family of blocks)*. As for (6).

The work on this class of relations has been restricted to randomized blocks and Latin squares. Thus its importance depends very greatly on the field of experimentation considered.

In much of agriculture, and in many related fields, the thought of *not* having blocks within which to randomize never occurs to the statistician. The errors he faces are large, even in small blocks, and variability from experimental unit to experimental unit may dominate all other sources of variation. Blocks are all important, and confounding with blocks is common. He is almost, but not quite, justified in refusing the adjective "experimental" to a situation without blocks and in calling it "sampling" instead. But there are areas of inquiry, in parts of modern industrial technology, in the study of many measurement processes, and in other areas (usually far from biology), where errors are relatively small (compared to high-order interactions), blocks can be very large, and complete randomization is the order of the day. In these latter areas, treatment of nesting, crossing, and random allotment suffices. In the former areas, treatment of randomized blocks comes first, though it needs to be supplemented with treatments of the simpler relations. The problems selected for initial attack by the independent groups working on average values of mean squares reflect their backgrounds.

We should come now to relations involving a minimum of four classifications (for example, simple lattices). But no work seems to have been done beyond the initial classical results for no interaction.

The present paper is concerned with the crossed classifications and, in the statement of rules, with combinations of crossed, nested, and noninteracting completely randomized classifications. In its general presentation and discussion it undoubtedly makes use not only of ideas from the references and research reports cited, but also of the valuable personal discussions which its authors have had with almost all the persons mentioned above, with H. Fairfield Smith, Franklin E. Satterthwaite, the late Charles P. Winsor, and others. It would be impossible for us now to assign specific credit relating to specific ideas to specific persons. We owe a particular debt to Kempthorne and Wilk for illuminating discussions.

5. The two spans of the bridge of inference. In almost any practical situation where analytical statistics is applied, the inference from the observations to the

real conclusion has two parts, only the first of which is statistical. A genetic experiment on *Drosophila* will usually involve flies of a certain race of a certain species. The statistically based conclusions cannot extend beyond this race, yet the geneticist will usually, and often wisely, extend the conclusion to (a) the whole species, (b) all *Drosophila*, or (c) a larger group of insects. This wider extension may be implicit or explicit, but it is almost always present. If we take the simile of the bridge crossing a river by way of an island, there is a statistical span from the near bank to the island, and a subject-matter span from the island to the far bank. Both are important.

By modifying the observation program and the corresponding analysis of the data, the island may be moved nearer to or farther from the distant bank, and the statistical span may be made stronger or weaker. In doing this it is easy to forget the second span, which usually can only be strengthened by improving the science or art on which it depends. Yet a balanced understanding of, and choice among, the statistical possibilities requires constant attention to the second span. It may often be worth while to move the island nearer to the distant bank, at the cost of weakening the statistical span—particularly when the subject-matter span is weak.

In an experiment where a population of C columns was specified, and a sample of c columns was randomly selected, it is clearly possible to make analyses where

- (1) the c columns are regarded as a sample of c out of C , or
- (2) the c columns are regarded as fixed.

The question about these analyses is not their validity but their wisdom. Both analyses will have the same mean, and will estimate the effects of rows identically. Both analyses will have the same mean squares, but will estimate the accuracy of their estimated effects differently. The analyses will differ in the length of their inferences; both will be equally strong statistically. Usually it will be best to make analysis (1) where the inference is more general. Only if this analysis is entirely unrevealing on one or more points of interest are we likely to be wise in making analysis (2), whose limited inferences may be somewhat revealing.

But what if it is unreasonable to regard c columns as any sort of a fair sample from a population of C columns with $C > c$. We can (at least formally and numerically) carry out an analysis with, say, $C = \infty$. What is the logical position of such an analysis? It would seem to be much as follows: We cannot point to a specific population from which the c columns were a random sample, yet the final conclusion is certainly not to just these c columns. We are likely to be better off to move the island to the far side by introducing an unspecified population of columns "like those observed" and making the inference to the mean of this population. This will lengthen the statistical span at the price of leaving the location of the far end vague. Unless there is a known, fixed, number of reasonably possible columns, this lengthening and blurring is likely to be worth while.

This discussion follows the line of the classical discussion of "selecting the right error term," as developed by Fisher and expounded by many statisticians,

with two considerations rarely faced except in careful discussions of groups or series of experiments (cf. Chapter 28 of [8] for references):

- (1) We admit that more than a single analysis of given data may have "correctness."
- (2) We have tried to state the uncertainties of the *post-facto* $C = \infty$ choice a little more specifically than usual.

In any case, however, this discussion illustrates one way in which the nature of the appropriate analyses of variance depends on the purposes of the analysis.

6. The varied roles of randomization. Emphasis on randomization of arrangement entered modern statistics with the analysis of variance—in the early work of R. A. Fisher. The year 1935, in which Neyman (with cooperation of Iwaskiewicz and Kolodzieczyk) [12] discussed the problem of interaction with experimental units, was marked by the appearance of *The Design of Experiments*, in which Fisher stressed both the role of randomization as a guarantor of the validity of an experiment and the close correspondence, in certain examples, of tests of significance based on randomization (assuming no interaction with experimental units) with those based on an assumption of normality of distribution.

Two years later, in 1937, the first papers of Pitman's series on randomization appeared. Pitman was seeking tests of significance which would be independent of the underlying distribution, naturally tried for randomization tests, and was much surprised when the natural approximation to these tests turned out to be the classical normal theory tests. This series of papers culminated in his *Biometrika* paper [13] which dealt with randomized blocks. In the meantime, Welch [15], had applied similar methods to both randomized blocks and Latin squares. In all of these papers, the assumption was made that there were no interactions with experimental units. As clearly stated, the motivation of these papers was to obtain tests which would apply to any distribution of errors, and randomization was used to mediate this independence of distribution.

The treatment of Pitman and of Welch went far beyond the subject of the present paper, the average values of mean squares. They dealt with a function of the ratio of mean squares, and obtained a number of moments. For the cases they treated, their results go farther than our knowledge for any other situations. They treated randomized blocks and Latin squares explicitly: Implicitly their results cover any less restrictive randomization; in particular, their results also cover complete randomization. Their work was carried out for the case of one classification of treatments. Implicitly, it applies to cases of two or more treatment classifications, but only if these treatment classifications do not interact.

We have today no comparable basis for the analysis of factorial experiments where interactions may be present and where, hence, the main effects will usually be compared with interaction. Our knowledge is limited to average values of mean squares, except for the work of Hooke [18] on second moments in the two-way case. To reach a situation comparable to the case of a single classification of treatments will require an extension of Hooke's work, both to higher moments and to more-way classifications. The essential difficulty is the probable existence

and possible perversity of interactions. No assumptions about randomization will allow us to avoid facing them. For the present, presumably, we shall continue to base our inferences on average values of mean squares, and try to comfort ourselves with distant and tenuous analogies with the single treatment classification case where, if interactions with experimental units be absent, Fisher, Pitman, and Welch have shown us that the situation is rather pleasant.

If one is dealing with situations where the contributions of the experimental units to variability is large or even dominant, it is easy to alter the role of randomization somewhat. Instead of thinking of it as a mediator which assures the validity of significance tests for any shape of error distribution, as Pitman did, one can think of it in itself. (If the effects of experimental units dominate all other sources of variation, it is not only easy but necessary.) This point of view was vigorously taken up by Kempthorne [8], [9], and has strongly motivated the work of Wilk and Kempthorne [21], [22], [23], [24], [25], [26].

While we hold rather definite views, we do not feel that the issues involved have been finally settled. We do feel, however, that a knowledge of what the issues are is essential in understanding just how far average values of mean squares take us and how our work is related to that of others.

DESCRIPTION OF SITUATIONS TREATED

7. A description from the point of view of urn sampling. We consider a finite or infinite number of elements (possible measurement results on animals, plots, batches, samples, etc.). These elements are classified into R rows (litters, days, blocks, pressures, temperatures, times, etc.) and C columns (doses, operators, treatments, temperatures, times, catalysts, etc.). Each of the RC pigeonholes thus formed contains N elements. In some circumstances it is natural to consider the RCN elements as a population, in others it would be unnatural.

A sample of r rows is taken, each row having probability r/R of being selected. Given that a particular row has been selected, the conditional probability that any other given row will have been selected is $(r - 1)/(R - 1)$.

Similarly a sample of c columns is taken, each column having probability c/C of being selected. Given that a particular column is in the sample, the conditional probability that any other given column will be in the sample is $(c - 1)/(C - 1)$.

Every pigeonhole located in a sampled row and a sampled column is thus a cell included in the sample. The pattern of cells so obtained might well be called a bisample, since it is defined by two samples, one of rows and one of columns. In each such cell n elements are sampled, each element in the sampled cell having probability n/N of being selected. Given that a particular element has been selected from this cell, the conditional probability that any other element in the cell has also been selected is $(n - 1)/(N - 1)$.

These are the only assumptions we shall need (for the situation with which we are concerned). From the point of view of urn sampling, we have only to define the variance components corresponding to such an array of RCN elements.

Notice that our definitions of variance components apply to fixed models and mixed models as well as to random ones. There has been a tendency to refer to random models as variance components models, presumably because the analysis associated with a fully random model is most likely to be directed toward the estimation of variance components. This choice of words has already given rise to confusion in connection with mixed models, and its continued use will undoubtedly cause other complications. We recommend that it be discarded.

Since we have to deal with two sets of rows, columns, and cells, one set in the original array of *RCN* elements and one in the array of *rcn* elements we actually have at hand, it is unusually important to make a clear distinction in our notation. We shall do this by using capital letters for all that has to do with the underlying array and lower case for all that has to do with the observed array. Thus

x_{ijk} = value of *k*th element in *j*th column in the *i*th row in the observed array,
 X_{IJK} = value of *K*th element in the *J*th column in the *I*th row in the underlying array.

(This convention will be altered in Sections 23ff). We shall indicate an unweighted mean over observed values by a dot in place of the subscript averaged over, and over underlying values by a dash in place of the subscript averaged over. Thus

$x_{ij.}$ = mean value of all observed elements in the *i*th row and *j*th column,
 X_{IJ-} = average value of all underlying elements in the pigeonhole in the *I*th row and *J*th column.

We can now define the variance components of the underlying array by

$$\sigma_R^2 = \frac{1}{R-1} \sum_{I=1}^R (X_{I--} - X_{---})^2,$$

$$\sigma_C^2 = \frac{1}{C-1} \sum_{J=1}^C (X_{-J-} - X_{---})^2,$$

$$\sigma_{RC}^2 = \frac{1}{(R-1)(C-1)} \sum_{I=1}^R \sum_{J=1}^C (X_{IJ-} - X_{I--} - X_{-J-} + X_{---})^2,$$

$$\sigma_B^2 = \frac{1}{RC} \sum \sum \sigma_{IJ}^2,$$

where

$$\sigma_{IJ}^2 = \frac{1}{N-1} \sum_{K=1}^N (X_{IJK} - X_{IJ-})^2.$$

These formulas reduce to the standard ones in all familiar special cases.

We observe that the variance components thus defined are exactly the mean squares we would obtain if all the values in the entire pigeonhole model were subjected to analysis of variance, except for factors of *N*, *NR* or *NC*.

The description we have just given is both less and more general than the model of Section 3. It is less general because we assumed that all populations

were of the same size. It is more general because we did not require that the sampling be purely at random, but only that any pair of rows, columns, or elements was as likely to enter the sampling as any other. We shall now see that these differences are nonessential.

We assumed a constant population size for the purely expository reason that we had not introduced the subscripts when we mentioned population size. We can now alter three sentences or clauses to read: "The IJ th pigeonhole contains N_{IJ} elements," "each element in the sampled cell having probability n/N_{IJ} of being selected," "the conditional probability \dots is $(n - 1)/(N_{IJ} - 1)$." After this alteration the urn sampling model is at least as general as our initial model (see Section 3).

In the initial model we assumed that all samplings were "at random." Since we are only concerned with averages of quadratic functions of the sampled elements, it would be merely an application of a general principle to conclude that if the results hold for sampling "at random," they must also hold for sampling "in which *individuals* and *pairs* have equal chances of being selected." We may, if we wish, consider that the initial model has been so generalized. The initial model and the urn sampling model then became equivalent.

8. Linear models with "tied" interaction. For convenience in exposition, we shall take $n = 1$ and drop the index k for most of this section. Under these conditions, the conventional linear model of most texts would appear like this:

$$x_{ij} = \theta + \xi_i + \eta_j + \omega_{ij},$$

where the ξ_i are the row contributions (sometimes called "effects"), the η_j are the column contributions, and the ω_{ij} are error or discrepancy contributions. Various normalizing conditions may or may not be applied. The ξ 's, the η 's, and the ω 's will be variously assumed to be fixed or random samples from infinite (or perhaps finite) populations. But the key assumption will be that the variation of the ω 's is *independent* of what the ξ 's and η 's may be. It is this assumption of independence which has made the use of such linear models so special and dangerous.

In particular, this model cannot accommodate the following situation easily described in terms of four pigeonholes:

$N(0, 1)$	$N(2, 1)$
$N(2, 1)$	$N(0, 1)$

where $N(\mu, \sigma^2)$ stands for a normally distributed infinite population with average value μ and variance σ^2 . The independence assumption is an assumption about the behavior of the world, and not just about how we do experiments.

Although some have expressed doubts, there are real advantages to linear models.

It has therefore been worth while to learn how to generalize linear models to apply without assumption.

Let us go, therefore, to a situation with R potential rows and C potential columns, where I designates a potential row and J designates a potential column. If $f(I, J)$ is an arbitrary function of I and J , we may choose to define

$$\begin{aligned}\theta &= f(-, -), \\ \xi_I &= f(I, -) - f(-, -), \\ \eta_J &= f(-, J) - f(-, -), \\ \lambda_{IJ} &= f(I, J) - f(I, -) - f(-, J) + f(-, -),\end{aligned}$$

where replacement of an " I " by a " $-$ " implies averaging over all R potential rows and replacement of a " J " by a " $-$ " implies averaging over all C potential columns. (Note that other definitions are possible.)

We shall then have

$$f(I, J) = \theta + \xi_I + \eta_J + \lambda_{IJ},$$

where

$$\lambda_{I-} = \text{a constant} = \lambda_{-J}$$

(and where indeed this constant is zero, and $\sum \xi_I = 0$ and $\sum \eta_J = 0$, although we shall *not require* any of these to vanish when we use this model).

Now if we pick r rows out of R and c columns out of C , we may designate actual rows and columns with i 's and j 's, and write $I(i)$ for the potential row corresponding to actual row i , and $J(j)$ for the potential column corresponding to the actual column j . If x_{ij} is the value of $f[I(i), J(j)]$, we have

$$x_{ij} = \theta + \xi_i + \eta_j + \lambda_{ij},$$

where we have written ξ_i for $\xi_{I(i)}$, η_j for $\eta_{J(j)}$, and λ_{ij} for $\lambda_{I(i)J(j)}$. We have here an additive model for a bisample of rows and columns drawn from an arbitrary $R \times C$ array of constants. How does this differ from the "independent" model in which ω_{ij} is assumed independent of ξ_i and η_j ? How is it to be described?

It differs in that the λ_{ij} are "tied" to the corresponding ξ_i and η_j . They are "tied" in a very specific way, however. It need not be true that λ_{ij} is a well-defined function of the values of ξ_i and η_j , for there may be, for example, various pairs of values of I and J for which the corresponding ξ_I and η_J have the same values but the corresponding λ_{IJ} are not equal. Thus λ_{ij} is tied to ξ_i and η_j through the values of $I(i)$ and $J(j)$ rather than through the values of ξ_i and η_j .

But what if the pigeonholes contain populations, and we sample one element

from each actual cell. Can this be treated similarly? Quite easily. Let

$$f(I, J) = X_{IJ-} = \frac{1}{N_{IJ}} \sum_K X_{IJK},$$

$$\omega_{IJK} = x_{IJK} - f(I, J).$$

Then, if the sampling in each cell is at random and independent, ω_{ij} is a random sample of 1 from a population $D_{I(i),J(j)}$ whose average value is zero, and

$$x_{ij} = X_{I(i),J(j),K} = f(I, J) + \omega_{ij}$$

$$= \theta + \xi_i + \eta_j + \lambda_{ij} + \omega_{ij},$$

where, although λ_{ij} and ω_{ij} have the same indices, they represent quite different sorts of quantities. The λ 's are interactions tied to the ξ 's and the η 's, while the ω 's are independent fluctuations.

9. Main effects and main contributions. All those who use the analysis of variance are familiar with the words "main effect," but far fewer have any really clear understanding of what they mean. Yet all specific analysis of variances procedures imply very specific interpretations of what it is that concerns us. Any practically satisfactory structure for the analysis of variance must bring the essential definitions into the limelight.

Because there is likelihood of confusing the quantity in the model and its estimate derived from the observations, we call the quantity in the model a main *contribution*. This is its name, but what is it? Usually only a relative definition makes sense. We talk of main contributions, but we work with their differences. What then does the difference $\xi_i - \xi_k$ mean in a situation modeled by

$$x_{ij} = \theta + \xi_i + \eta_j + \lambda_{ij} + \omega_{ij}$$

with the two conditions

$$\lambda_{I-} = \text{a constant} = \lambda_{-J},$$

$$\text{ave } \{\omega_{IJ}\} = 0.$$

Let, again,

$$f(I, J) = \text{average response in the } IJ \text{ pigeonhole};$$

then

$$\theta + \xi_I + \eta_J + \lambda_{IJ} = f(I, J),$$

and we easily find that

$$\xi_I - \xi_K = f(I, -) - f(K, -) = \frac{1}{C} \sum [f(I, J) - f(K, J)].$$

This states that the difference in main contributions between the I th and K th rows is the average over all C potential columns of the average effect of changing

from the I th to the K th row. The average is over *all* potential columns, and for the present definition is *equally* weighted. What is most important in defining row main contributions is our choice of what are the potential columns.

This is the heart of the problem of main contributions—it is our attitude toward the *other* factors which affects the definition of a main contribution. A change from considering *rows* as fixed to considering rows as a sample from a large population need have no effect at all on the definition of *row* main contributions, but it will substantially alter the definitions of the *column* main contributions. It is only when we have relatively explicit definition that we can force ourselves to recognize this fact.

We stated that an alteration in the set of potential rows *need* not alter the definition of the row contributions. This is so because we have *not* required that $\sum \xi_I = 0$. We have not made this requirement because it serves no useful purpose. (Its imposition would make all the ξ_i estimable from an experiment, but there are two reasons why this would not be useful: (i) Because we are interested in estimating only the differences $\xi_i - \xi_j$ anyway. (ii) Because, usually, only some of the ξ_I appear as $\xi_i = \xi_{I(i)}$, and since the others can surely not be estimated, what do we care about one more unestimable parameter? By making one more parameter unestimable, we have gained inestimable freedom.)

It is in presenting explicit quantities which represent main contributions that this linear model—the generalized, nonindependent linear model—has a significant and specific role. It contains exactly the same definition of main contributions, but how many readers recognized this as they read this section? It is concealed in such urn formulas as

$$\sigma_R^2 = \frac{1}{R-1} \sum_{I=1}^R (X_{I--} - X_{---})^2,$$

which implicitly state that what we shall ask the “row” line of the analysis of variance to inform us about, are the X_{I--} . The difference $X_{I--} - X_{K--}$ is identical with $\xi_I - \xi_K$, but the linear model brings the situation out with greater force and clarity.

10. Description by linear models. We now set out the general linear model in the form

$$x_{ijk} = \theta + \xi_i + \eta_j + \lambda_{ij} + \omega_{ijk},$$

where $1 \leq i \leq c$, $1 \leq j \leq r$, $1 \leq k \leq n$ and the assumptions are as below, where in dealing with simple finite populations, we shall always use the modern definition of variance, dividing by *one less* than the number of elements involved. Thus, for example, the variance of the population of potential columns is

$$\sigma_\eta^2 = \frac{1}{C-1} \sum_{j=1}^c (\eta_j - \eta_-)^2,$$

where η_- is the average of η_J over the population of potential columns

$$\eta_- = \frac{1}{C} \sum \eta_J.$$

In general, we shall use dashes for averages of population quantities, and dots for means of sampled quantities.

With these preliminaries we can state the assumptions about θ , ξ_i , and the η_j which we wish to apply:

- (1) There is a population of general contributions, of variance σ_θ^2 , and θ is a random sample of 1 from this population.
- (2) There is a population of row contributions of size R and variance σ_R^2 and $\xi_1, \xi_2, \dots, \xi_r$ are a random sample of size r from this population.
- (3) There is a population of column contributions of size C and variance σ_C^2 , and $\eta_1; \eta_2, \dots, \eta_c$ are a random sample of size c from this population.
- (4) There is a two-way array of interactions λ_{IJ} , one for each row contribution and column contribution in the corresponding populations, the averages (over J) λ_{I-} are independent of I , the averages (over I) λ_{-J} are independent of J , and the interaction actually occurring in the expression for x_{ij} is $\lambda_{I(i),J(j)}$; that is to say, it is the interaction which corresponds to the column contribution and row contribution which occur in the expression for x_{ij} .
- (5) The sampling in (1), (2), and (3) takes place independently.

We define, for reference,

$$\begin{aligned} \sigma_{RC}^2 &= \frac{1}{(R-1)(C-1)} \sum \sum (\lambda_{IJ} - \lambda_{-})^2 \\ &= \frac{1}{(R-1)(C-1)} \sum \sum (\lambda_{IJ} - \lambda_{I-} - \lambda_{-J} + \lambda_{-})^2. \end{aligned}$$

(Note that we do *not* use $(RC - 1)$ as a divisor.)

There remain the assumptions about the ω_{ijk} , where we still have much choice, so long as we require that

$$\text{ave } \{\omega_{IJK}\} = \text{a constant independent of } I \text{ and } J.$$

The assumption which most exactly corresponds to the pigeonhole model is the following:

- (6') For each of the RC combinations of a population row with a population column there is a population of size N_{IJ} , average value μ (the same for all IJ), and variance σ_{IJ}^2 . The ω_{ijk} are a sample of n from the $I(i)J(j)$ th population. Sampling is at random in each cell and independent, both between cells and of the samplings in (1), (2), and (3).

With this choice, the generalized linear model corresponds exactly to the pigeonhole model, provided we place $\sigma_\theta^2 = 0$ and thus keep the general contribution constant.

The variance component for repetition (or "error") σ_E^2 , and the effective population size N , are to be found from

$$\sigma_E^2 = \frac{1}{RC} \sum_I \sum_J \sigma_{IJ}^2.$$

$$\left(1 - \frac{n}{N}\right) \sigma_E^2 = \frac{1}{RC} \sum_I \sum_J \left(1 - \frac{n}{N_{IJ}}\right) \sigma_{IJ}^2.$$

We remarked that (6') was not the only choice for a linear model. There are many. We shall give *detailed* consideration to only one other, namely:

(6'') The cell contributions are

$$\omega_{ijk} = \omega'_{ijk} + \omega''_{ij},$$

where the ω'_{ijk} satisfy the conditions of (6') while the ω''_{ij} are the result of independently randomizing a set of rc values of variance σ^2 among the rc actual cells.

We shall now need to set

$$\sigma_E^2 = \frac{1}{RC} \sum_I \sum_J \sigma_{IJ}^2$$

$$\left(1 - \frac{n}{N}\right) \sigma_E^2 = \sigma^2 + \frac{1}{RC} \sum_I \sum_J \left(1 - \frac{n}{N_{IJ}}\right) \sigma_{IJ}^2$$

(the last of which may lead to negative N 's).

Clearly (6'') is more complicated than (6'), which matched the pigeonhole model. Why then do we wish to consider this added complexity? Because experiments are often more complicated than the simple pigeonhole model makes allowance for. For example, let us suppose that we wish to study a chemical reaction at all combinations of 6 specific pressures and 7 specific temperatures. The pigeonhole model would naturally have 6 rows and 7 columns, and in each pigeonhole we would put an infinite population. The pattern of the averages of these populations would represent the response of this process to pressure and temperature. The variations within each of these populations would represent fluctuations in process behavior and measurement. But we would be ill advised to stop here. Every well-trained statistician would insist, especially if every combination of pressure were to be tried once only, that the order in which the experiments were performed should be randomized. He would do this in fear of systematic errors somehow associated with time. In other words, in fear that the ω''_{ij} of (6'') were not all zero. We need very great flexibility in our models to deal with real situations such as this. (Notice that we do not attempt to discuss the very real and important cases where randomized blocks, and related uses of randomization are involved. We are here concerned with the flexibility required for the simple case of two crossed classifications, where experimental units are not so important as to make complete randomization inadequate.)

The flexibility of the linear model is not limited to (6''). It can easily accommodate any number of terms of each or all of the following kinds:

- (a) Individual distributions of ω'_{IJK} for each IJ .
- (b) Randomization of rc values ω'' over the rc cells.
- (c) Randomization of a sample of rc values ω''' drawn randomly from a population of size $M > rc$ over the rc cells.
- (d) Randomization of RC values ω'''' over the RC pigeonholes.
- (e) Randomization of RC values ω''''' drawn randomly from a population of size $M > RC$ over the RC pigeonholes.

The ease with which such complete randomized contributions can be added to linear models depends on a simple remark, which was apparently first made by McCarthy ([10], p. 358) in the case where all is normal and the variances are constant and was later exploited ([14], p. 107) in the general case; namely, that equal, completely symmetric correlations do not affect the average mean squares, except for the mean square for the general mean (to be discussed in Section 14). That this is true can be easily seen by the following argument. If the correlations are equal and negative, add a varying general contribution with variance equal to the correlations. This will exactly annul the correlations. If the correlations are equal and positive, observe that they are exactly the correlations which would result from such a fluctuating main effect. In either case, the desired result follows immediately.

It may not be immediately appreciated why we need more than one term of a given kind. However, if, in the chemical reaction experiment just considered, we plan to use a separate piece of equipment and a separate operator for each run, we will wish to randomize all three variables—epoch, equipment, and operator. Our model will require three terms of the ω'' kind. There will be three different sorts of experimental units!

If we put down all the variables considered in a really complex experiment, even the linear model begins to look complicated.

11. The development of the linear model. The classical linear model, as described at the beginning of the previous section, had the following disadvantages from our present point of view.

- (1) It made assumptions about the way in which the response depended on the two (or more) factors.
- (2) It assumed that at most one term should appear for a given set of indices.
- (3) It assumed special normalizations (like $\lambda_{I-} = 0$) rather than more general ones (like λ_{I-} independent of I).
- (4) It assumed constant variance "within pigeonholes."

(These have been listed in general order of importance.)

Models which avoided the first disadvantage were introduced (for randomized blocks and Latin squares) by Neyman (with the cooperation of Iwaskiewicz and Kolodzieczyk) [12] and were used by McCarthy [10], but, perhaps because of the generally negative flavor of the papers, seem to have lain forgotten for

many years. The results (for crossed classifications) of Memorandum Report 18 [21] were expounded in the terms used in the last section. Similarly models were independently and extensively used (for various situations) in the book of Kempthorne [8] and have been prominent in the recent work of Wilk and Kempthorne. For all of this, the relationship of $\lambda_{i,j}$ to ξ_i and ξ_j when it is tied through the values of $I(i)$ and $J(j)$ is not yet as widely familiar among statisticians, as it seems to us that it should be.

The introduction of two or more terms with the same set of subscripts was also due to Neyman (with the co-operation of Iwaskiewicz and Kolodzieczyk) [12] who introduced first a single term and then separated it into two parts, corresponding to "plot error" and "technical error." In [14], the approach was to start from the parts and then combine, rather than the reverse.

Both this difference, and the introduction of the weaker normalization in [14], were related to two desires: a desire to weaken assumptions wherever possible, and a desire to treat contributions more as things with independent existence rather than as differences between certain averages. The general philosophy ran along the lines that "if the effect is real and substantial, it should appear in the model whether or not it can be estimated from the data."

The assumption of constant variance "within pigeonholes" is a natural assumption, and is important in connection with both higher moments of mean squares and with the variances of contrasts and other linear combinations. As pointed out in [14], however, it may be dropped without any effect on average mean squares.

A more recent development in the case of linear models has been the introduction of a pair of closely related linear models, called "the population model" and "the statistical model" by Wilk [22] and their further use by Wilk and Kempthorne [23], [24]. By introducing this distinction, the assumptions about the response to the various classifications can be presented first, while the assumption about the randomization involved in setting up the experiment can be added later. This development formalizes and makes definite distinctions hinted at in Kempthorne's book [8].

12. The next consideration. The next step in generality, when we have completely randomized some contribution, to which we will attach the name "experimental units," is to consider the possibility that these units can interact with the other classifications. (Since the results are a function of both classifications jointly, the behavior resembles that of crossed classifications; yet, in any particular experiment, one classification is nested in the other. For these reasons, the term "cross nested" has been used informally to describe the situation.)

It is easiest to make the situation clear when a one-way classification, say temperature, is involved, and epochs are randomized. If we admit that the effect of an epoch can be different for different temperatures, then we are led to a two-way set of pigeonholes—pigeonholes labelled by temperature and epoch. If each temperature is used once only, the actual experiment (if fully randomized) could be described as picking a set of pigeonholes, one in each row and column,

and picking a value from each of these selected pigeonholes. It is as if we had the results of that part of a Latin square occupied by a single letter. Even though we cannot estimate the temperature by epoch interaction variance component, it is clear that it will enter into the average values of the mean squares which we do obtain.

Complete randomization is a binary relation between classifications, but, especially when without replication, it is closely related to the ternary relation of the Latin square. Again independent work has led to related results. Kempthorne treated the Latin square with arbitrary interactions between rows and columns, but with no interactions with treatments in his book ([8], pp. 190-191). In particular he obtained the variance of a treatment mean. This last particular result is also the variance of the general mean in the unreplicated case of complete randomization with arbitrary interaction (with experimental units) which was obtained independently by Cornfield and Evans, and reported, with a modified proof, in Hansen, Hurwitz, and Madow ([6], pp. 262-265). So far as complete randomization is concerned, more general results were obtained by Wilk [16, 25].

It is clear, however, that there are a number of stages of sophistication, care, or cynicism (as you please) about our treatment of such a factor as epoch. Some of these are the following:

- (1) In both design and analysis we neglect it altogether.
- (2) We neglect it in design, and in analysis we use the additive model to show ourselves that we should have randomized it after all.
- (3) We randomize it in design, and in analysis we use the additive model to show ourselves how well off we are.
- (4) We randomize it in design, and in analysis we use the arbitrary-interaction model to show ourselves that the situation is not quite perfect.
- (5) We take it into the design of the experiment as a full-fledged factor.

Each of these attitudes is appropriate in its place. In every experiment there are many variables which *could* enter, and one of the great skills of the experimenter lies in leaving out only inessential ones. What we have just observed is that there are gradations between variables which are entirely out and those which are entirely in.

It is generally understood that in the design of an experiment, there are three classes of variables:

- (i) those treated as factors,
- (ii) those randomized, and
- (iii) those neglected.

The additive model helps us to think about the choice between neglect and randomization. The arbitrary-interaction model helps us to think about the choice between randomization and recognition as a factor.

RESULTS

13. Results for the two-way classification. The average mean squares for the pigeonhole model are set forth for the general case and for three special cases

generalizing the usual fixed, random, and mixed models in Table 1. The important thing to notice in this table is the appearance of the factors

$$1 - \frac{c}{C}, \quad 1 - \frac{r}{R}, \quad 1 - \frac{n}{N},$$

which serve to suppress certain terms completely in the fixed situation and to reduce their effects on the average values of the mean squares when the populations are finite but larger than the samples.

TABLE 1
General results and special cases

Line in the analysis of variance	Average value of mean squares			
	General case	Special cases		
		(Fixed) $c = C, r = R$ N infinite	(Random) c, r finite $N, C, R,$ infinite	(Mixed) $c = C, r$ finite N, R infinite
Rows	$\left(1 - \frac{n}{N}\right) \sigma_E^2$ $+ \left(1 - \frac{c}{C}\right) n\sigma_{RC}^2$ $+ nnc\sigma_R^2$	$\sigma_E^2 + nnc\sigma_R^2$	$\sigma_E^2 + n\sigma_{RC}^2$ $+ nnc\sigma_R^2$	$\sigma_E^2 + nnc\sigma_R^2$
Columns	$\left(1 - \frac{n}{N}\right) \sigma_E^2$ $+ \left(1 - \frac{r}{R}\right) n\sigma_{RC}^2$ $+ nr\sigma_C^2$	$\sigma_E^2 + nr\sigma_C^2$	$\sigma_E^2 + n\sigma_{RC}^2 + nr\sigma_C^2$	$\sigma_E^2 + n\sigma_{RC}^2 + nr\sigma_C^2$
Interaction	$\left(1 - \frac{n}{N}\right) \sigma_E^2 + n\sigma_{RC}^2$	$\sigma_E^2 + n\sigma_{RC}^2$	$\sigma_E^2 + n\sigma_{RC}^2$	$\sigma_E^2 + n\sigma_{RC}^2$
Error	σ_E^2	σ_E^2	σ_E^2	σ_E^2

Notice in particular how, when $R = C = N = \text{infinity}$, σ_E^2 follows σ_{RC}^2 everywhere, just as the lamb followed Mary. This is the usual result for a random model, and appears here in a situation of far greater generality.

The disappearance of σ_{RC}^2 from the average values of row and column mean squares in the fixed case is also familiar.

In the mixed model the average mean square for rows does not involve the interaction variance component (because all columns were observed!), while the average mean square for column involves it with unit weight (because only a small sample of rows were observed!). In each case the behavior of the *other variable* (or, in general, the *other variables*) determines whether the interaction appears or not.

That the result for the mixed model is a consequence solely of the urn sampling approach, and not of the special definitions of interaction and row components of variance used, may be seen from the following simple example. Consider a population composed of three rows and two columns, with one element per cell, having the following numerical values:

0	100
100	0
50	50

Each of the three possible samples of two rows and two columns will yield a row mean square of zero and an interaction mean square greater than zero. Hence for this population the interaction mean square exceeds the row mean square for each sample. Testing the row mean square against interaction would be obviously incorrect.

These results are very general. We assumed only a symmetrical placing of *RC* sets of numbers in *RC* pigeonholes and a well-defined method of extracting the *rcn* numbers entering the actual analysis of variance. Otherwise, the numbers can be as arbitrary as we please.

14. The mean square for the general mean. While the central facts have been pointed out from time to time (e.g., Cochran [3], Hendricks [7]), and while a number of mathematical statisticians have made use of the fact in their lectures, it does not seem to be widely recognized that it is almost always feasible to complete the analysis of variance table by adding a line for the general mean, and that when this has been done, both classical analysis of variance and sampling from finite populations become special cases of a unified situation. We feel that wider recognition and use of this fact would be valuable.

In the present situation we have only to introduce a reference origin *M* (which is freely at our disposal) and define both the sum of squares for the general mean and the mean square for the general mean (there is but one degree of freedom) as

$$rcn(x_{...} - M)^2$$

(This is what is frequently known as the "correction term," and is appropriate to the use of *M* as a computing origin.) It is easy to show that we have the following average value of the mean square for the mean

$$rcn\sigma_{\theta}^2 + cn\left(1 - \frac{r}{R}\right)\sigma_R^2 + rn\left(1 - \frac{c}{C}\right)\sigma_C^2 + n\left(1 - \frac{c}{C}\right)\left(\frac{1-r}{R}\right)\sigma_{RC}^2 + \left(1 - \frac{n}{N}\right)\sigma_E^2 + rcn(\text{ave } \{x_{...}\} - M)^2,$$

which like all average mean squares in a balanced analysis of variance decompose into $(F)[(G) + (H)]$ where

(*F*) = number of units involved in each corresponding mean, (*rcn* in the example),

(*G*) = variance (so far as differences are concerned) of the corresponding mean,

$$\sigma_G^2 + \left(\frac{1}{r} + \frac{1}{R}\right) \sigma_R^2 + \left(\frac{1}{c} - \frac{1}{C}\right) \sigma_C^2 + \left(\frac{1}{c} - \frac{1}{C}\right) \left(\frac{1}{r} - \frac{1}{R}\right) \sigma_{RC} \\ + \frac{1}{rc} \left(\frac{1}{n} - \frac{1}{N}\right) \sigma_E^2 \text{ in the example,}$$

(*H*) = variance of corresponding contributions (usually measured among themselves, but in this case necessarily measured from the arbitrary origin, *M*, since there is only one general mean and hence = (ave {*x*...} - *M*)²).

15. Results for the three-way classification. The pigeonhole model for the three-way classification is the natural generalization of the two-way pigeonhole model. There are *RCS* pigeonholes, one in each combination of *R* rows, *C* columns, and *S* slices. In each pigeonhole there is a population. Random samples of *r* rows from *R*, *c* columns from *C*, and *s* slices from *S* are independently drawn. The intersections of the drawn rows, columns, and slices determine *rcs* cells. In each of these *rcs* cells *n* elements are drawn at random. All drawings, whether of rows, columns, slices, or elements are independent.

The variance components are defined in a way similar to that for the two-way. They differ from the (hypothetical) mean squares we would get, if we analyzed the entire model, by simple factors such as *r*, *c*, *s*, etc.

The average values of the mean squares are given for the general case in Table 2. Here, with obvious extensions of notation,

$$\sigma_E^2 = \frac{1}{RCS} \sum_I \sum_J \sum_K \sigma_{IJK}^2, \\ \left(1 - \frac{n}{N}\right) \sigma_E^2 = \frac{1}{RCS} \sum_I \sum_J \sum_K \left(1 - \frac{n}{N_{IJK}}\right) \sigma_{IJK}^2,$$

and

$$\sigma_G^2 = \text{variance of the general contribution.}$$

(Except for the top line, these results will be found in Bennett and Franklin ([2], p. 394), while generalizations have been given by Kempthorne and Wilk [23] and by Wilk [25]).

While the results for the three-way are not too complex, and are obviously systematic and orderly, they do take up considerable space. If we are to set out the corresponding results for factorial arrangements with more factors, there will be a premium on more compactness.

16. Unreplicated factorials in general. We can obtain this compactness by setting up some rules which will give the coefficient of any variance component

TABLE 2

Average values of mean squares in the general (replicated) three-way classification

Item	DF	Average value of mean square
General mean	1	$\begin{aligned} & \left(1 - \frac{n}{N}\right)\sigma_E^2 + n\left(1 - \frac{r}{R}\right)\left(1 - \frac{c}{C}\right)\left(1 - \frac{s}{S}\right)\sigma_{RCS}^2 \\ & + nr\left(1 - \frac{c}{C}\right)\left(1 - \frac{s}{S}\right)\sigma_{CS}^2 + nc\left(1 - \frac{r}{R}\right) \\ & \cdot \left(1 - \frac{s}{S}\right)\sigma_{RS}^2 + ns\left(1 - \frac{r}{R}\right)\left(1 - \frac{c}{C}\right)\sigma_{RC}^2 \\ & + ncr\left(1 - \frac{s}{S}\right)\sigma_S^2 + ncs\left(1 - \frac{r}{R}\right)\sigma_R^2 \\ & + nrs\left(1 - \frac{c}{C}\right)\sigma_C^2 + ncrs\sigma_G^2 \\ & + ncrs(\text{ave } \{x\} - M) \end{aligned}$
Rows (<i>R</i>)	<i>r</i> - 1	$\begin{aligned} & \left(1 - \frac{n}{N}\right)\sigma_E^2 + n\left(1 - \frac{c}{C}\right)\left(1 - \frac{s}{S}\right)\sigma_{RCS}^2 \\ & + nc\left(1 - \frac{s}{S}\right)\sigma_{RS}^2 + ns\left(1 - \frac{c}{C}\right)\sigma_{CR}^2 + ncs\sigma_R^2 \end{aligned}$
Columns (<i>C</i>)	<i>c</i> - 1	$\begin{aligned} & \left(1 - \frac{n}{N}\right)\sigma_E^2 + n\left(1 - \frac{r}{R}\right)\left(1 - \frac{s}{S}\right)\sigma_{RCS}^2 \\ & + nr\left(1 - \frac{s}{S}\right)\sigma_{CS}^2 + ns\left(1 - \frac{r}{R}\right)\sigma_{CR}^2 + nrs\sigma_C^2 \end{aligned}$
Slices (<i>S</i>)	<i>s</i> - 1	$\begin{aligned} & \left(1 - \frac{n}{N}\right)\sigma_E^2 + n\left(1 - \frac{r}{R}\right)\left(1 - \frac{c}{C}\right)\sigma_{RCS}^2 \\ & + nr\left(1 - \frac{c}{C}\right)\sigma_{CS}^2 + nc\left(1 - \frac{r}{R}\right)\sigma_{RS}^2 + ncr\sigma_S^2 \end{aligned}$
<i>RC</i>	(<i>r</i> - 1)(<i>c</i> - 1)	$\left(1 - \frac{n}{N}\right)\sigma_E^2 + n\left(1 - \frac{s}{S}\right)\sigma_{RCS}^2 + ns\sigma_{RC}^2$
<i>RS</i>	(<i>r</i> - 1)(<i>s</i> - 1)	$\left(1 - \frac{n}{N}\right)\sigma_E^2 + n\left(1 - \frac{c}{C}\right)\sigma_{RCS}^2 + nc\sigma_{RS}^2$
<i>CS</i>	(<i>c</i> - 1)(<i>s</i> - 1)	$\left(1 - \frac{n}{N}\right)\sigma_E^2 + n\left(1 - \frac{r}{R}\right)\sigma_{RCS}^2 + nr\sigma_{CS}^2$
<i>RCS</i>	(<i>r</i> - 1)(<i>c</i> - 1)(<i>s</i> - 1)	$\left(1 - \frac{n}{N}\right)\sigma_E^2 + n\sigma_{RCS}^2$
Repetition	<i>rcs</i> (<i>n</i> - 1)	σ_E^2

in the average value of any mean square

We begin with the case without replication and the almost obvious

Rule 1. Unless the indices of the variance component include all those of the mean square, the coefficient vanishes.

As a result of this rule, we can confine our attention to cases where the indices fall into three categories:

- (1) those appearing in both mean square and variance component,
- (2) those appearing in variance component but not in mean square,
- (3) those appearing in neither.

In these terms we can now state

Rule 2. Any coefficient which does not vanish because of Rule 1 is the product of a small letter for each index which fails to appear in the variance component and of a factor

$$1 - \frac{(\text{small letter})}{(\text{capital letter})}$$

for each index in the variance component which does not appear in the mean square.

Let us discuss some examples. Consider the coefficient of σ_{CS}^2 in the mean square for columns in an unreplicated three-way. Here the behavior of the indices is as follows:

- (1) C appears in both,
- (2) S appears only in the variance component,
- (3) R appears in neither.

The coefficient is, therefore,

$$r \left(1 - \frac{s}{S} \right)$$

and, when we recall that $n = 1$, we see that this checks the entry in Table 2.

Consider a seven-way classification with indices R, C, D, E, F, G, S and the average value of the mean square for the RDG -interaction. What are the coefficients of σ_{RCDDEF}^2 and $\sigma_{RCDDEGS}^2$? Since G is not an index of σ_{RCDDEF}^2 , its coefficient is zero by rule 1. For the second variance component, the indices behave as follows:

- (1) R, D , and G appear in both,
- (2) C, E and S appear only in the variance component,
- (3) F appears in neither.

By rule 2, the coefficient is

$$f \left(1 - \frac{c}{C} \right) \left(1 - \frac{e}{E} \right) \left(1 - \frac{s}{S} \right).$$

The rules are relatively easy to apply.

We can summarize the action of both rules in a 2×2 table to be applied once for each subscript. This is done for a general subscript Q in Table 3.

17. More general designs. Two remarks will conspire to lead us to more general expressions of the rules. First, we note a remarkable similarity between these coefficients, and those which arise when one classification is nested within another. Second, we notice that there is a useful sense in which a replicated factorial is not purely factorial in structure—in the pigeonhole model we allowed a

TABLE 3

2 × 2 table giving the factor in the coefficient due to any one subscript, and thus expressing both rules for factorials

Subscripts of mean square	Subscripts of variance component	
	<i>Q</i> present	<i>Q</i> absent
<i>Q</i> present.....	1	0
<i>Q</i> absent.....	$1 - \frac{q}{Q}$	q

population to nest in each pigeonhole. We are impelled to seek some general rules which apply when classifications are crossed (as in a factorial) or nested in any way. (We shall find that noninteracting completely randomized contributions are more or less automatically included.)

These relations are not the only important relations between categories. We have mentioned randomized blocks above. There are also, for example, the relations involved in balanced and unbalanced incomplete blocks, to which our rules will not apply. But crossing and nesting are the simplest, and a treatment for any combination of them will be well worthwhile.

The scope of such a treatment will be broader than just the arrangements so far mentioned. It is quite possible, for example, to have a one-way array of small pigeonholes nested in each of a two-way array of large pigeonholes and a population nested in each of the small pigeonholes. This will occur in the example of a chemical reaction where temperature and pressure define the large pigeonholes, for instance, if we repeat the reaction several times for each pressure and temperature combination and analyze several portions of each result. Here "sampling and analysis" requires a population nested in "batch," while a one-way array of "batches" is nested in every "temperature-pressure" pigeonhole.

The general rules are based on a system of indices such that a mean square or variance component referring to something of smaller scope must have all the indices of the quantity of larger scope, and one or more in addition. A possible set of indices for the modified chemical reaction example just described might be

T = temperature,

P = pressure,

PT = their interactions,

BPT = batches (within PT combinations).

$EBPT$ = sampling and analysis (within batches).

In this example it makes no sense to define a B main contribution across temperature or pressure, and hence we should not in this notation try to define a B mean square or a B variance component. Indeed we should also not try to use BT - or

BP- quantities of any sort. In order to inquire, for example, how many batches to use to obtain a given accuracy, we should be concerned with "batches within a temperature and pressure combination" and in this notation this is a *BPT* effect. (It would perhaps be clearer to use $B(PT)$ or $B \subset PT$, but we shall not do this here.)

Any set of indices has certain indices (there may be none) toward which it behaves like an interaction. This means that if we sum the corresponding contributions over such an index (covering the pigeonhole model) the total (and hence the mean) is a constant. In the example just discussed

Set of indices	Interactionlike indices
<i>T</i>	<i>T</i>
<i>P</i>	<i>P</i>
<i>PT</i>	<i>T, P</i>
<i>BPT</i>	<i>B</i>
<i>EBPT</i>	<i>E</i>

The appearance of *B* and *E* as interactionlike indices merely means that the average contributions for "batches" and "sampling and analysis" are defined to be zero. The absence of *T* in the last two cases reflects the fact that summing over one batch at each temperature, or summing over one sample and analysis at each temperature, will not ensure the disappearance of the mean batch contribution or the mean sampling and analysis contribution.

We continue to use capital letters for the dimensions of the pigeonhole model, and small letters for the corresponding dimensions of the experiment. We can now state a single unified rule (covering crossed and nested relations, and classifications completely randomized upon them) as follows:

Rule: Divide the indices into five groups as follows:

- (1) those which appear in the mean square but not in the variance component;
- (2) those which appear in the variance component and not in the mean square and are interactionlike for the set of indices defining the variance component;
- (3) those which appear in the variance component but not in the mean square, and are *not* interactionlike;
- (4) those which appear in both;
- (5) those which appear in neither.

The coefficient with which the variance component appears in the average value of the mean square is the product of a factor for each index, the factors being as follows:

- (a) for group (1) each factor is zero,
- (b) for group (2) each factor is

$$\left(1 - \frac{\text{(small letter)}}{\text{(capital letter)}}\right),$$

- (c) for groups (3) and (4) each factor is unity,
 - (d) for group (5) each factor is the corresponding small letter.
- This rule applies so long as all categories are related by crossing or nesting.

In this form, this rule includes the rules of the previous section, and can also be expressed in a simple table, as in Table 4.

TABLE 4

Factor due to any subscript, Q, in the coefficient of any variance component in any average mean square for any combination of crossing and nesting

Subscripts of mean square	Subscripts of variance component		
	Q present and interactionlike	Q present, not interactionlike	Q absent
Q present	1	1	0
Q absent	$1 - \frac{q}{Q}$	1	q

18. Reasons for the general rules. As indicated by the discussion of Section 23, below, the basic reasons for the general rules are two:

- (1) Mean squares are ordinarily expressed on a per-element basis, so the number of elements associated with specific values of the indices of a mean square arises as a factor.
- (2) Where an index is interactionlike we are dealing with variances of sample means, so that factors of

$$\frac{1}{(\text{small letter})} \text{ minus } \frac{1}{(\text{capital letter})}$$

arise from the formula for the variance of a mean from a finite population. These reasons are equally compelling for crossing and nesting.

19. Non-interacting, completely randomized terms in linear models. We discussed briefly, at the end of Section 10, some five kinds of added terms which might be added to the linear model for the two-way classification. We summarize in Table 5 the coefficients with which the corresponding variances or mean variances will appear. These results are easily obtained by the usual way of dealing with independent linear models. Observe that if we square the model, the average value of all cross-terms will cancel out (except in the mean square for the general mean) and hence that average mean squares can be evaluated separately for the original contributions and those of these kinds.

It was emphasized (at the end of Section 10) that the easy addition of such terms was a consequence of a general principle. Several such terms can be added. If the modified chemical reaction example discussed in the last section involved two raw materials, we might have completely randomized packages of one over batches, and have made up one solution of the other for each pressure-temperature combination, randomizing the allotment of solutions. If, in addition, some

TABLE 5

Kind	Average mean of square value for rows, columns, interaction	General Mean
(a)	$1 - \frac{n}{N}$	$1 - \frac{n}{N}$
(b)	1	0
(c)	1	$1 - \frac{rc}{M}$
(d)	1	$1 - \frac{rc}{RC}$
(e)	1	$1 - \frac{rc}{M}$

other relevant item were completely randomized over individual samples, we should have a model made up of at least four parts:

- (1) dependence of yield on pressure and temperature combined with routine fluctuations and errors,
- (2) effects of first raw material,
- (3) effects of second raw material,
- (4) effects of other relevant factors.

If we can properly assume an absence of interaction between these four parts, then we will obtain four sets of (partial) variance components as follows

- (1) appears in *EBPT*, *BPT*, *PT*, *P* and *T*,
- (2) appears in *BPT*, *PT*, *P* and *T*,
- (3) appears in *PT*, *P* and *T*,
- (4) appears in *EBPT*, *BPT*, *PT*, *P* and *T*.

Because of the no-interaction-across-parts assumption, the four sets will behave entirely independently, and the rules of the last section will apply to each separately.

It will only be necessary to remember that (so long as we are not concerned with the general mean) *no* index is interactionlike for contributions which are completely randomized. This same principle will apply to other examples where absence of interaction between completely randomized parts is appropriately assumed.

PROOFS

20. The nature of the various proofs. A number of apparently quite different ways have been developed to carry out the proofs of the formulas for average values of mean squares (see Section 4 for references). They fall quite neatly into two categories:

- (1) Proofs using special machinery or indirect methods (e.g., symmetry arguments and equating of coefficients for special assumptions as in [18] and [20]).

- (2) Proofs using relatively straightforward algebra (e.g., as in [17], [16], and [21]).

While we feel that the first sort of proof offers the real hope for dealing with the more difficult problems which lie ahead, we recognize the usefulness of having examples of the straightforward proofs on record, and have endeavored to keep our proofs direct and to hold the use of special techniques to a minimum.

In the two-way pigeonhole model three samplings take place independently, a sampling of rows, a sampling of columns, and a sampling within cells. The separateness and independence of these samplings is very important in reaching a moderately simple direct proof, as is the possibility of considering the separate samplings as occurring in any order, and then using quantities defined by some array intermediate between the underlying array and the observed array. It is only by combining such a choice with a well-chosen notation and order of procedure that we can keep the algebra from becoming quite heavy.

The original proof [20], was carried through explicitly for the two-way case without replication in the cell ($n = 1$, $N = \infty$), made explicit use of linear models, and depended on two comparisons of three situations:

- (1) rows fixed, columns fixed, within cells sampled,
- (2) rows fixed, columns sampled, within cells sampled,
- (3) rows sampled, columns sampled, within cells sampled.

The next proof [17], found without knowledge of the first result, made explicit use of urn sampling and was carried through explicitly for the two-way case with $\sigma_{IJ}^2 = \text{constant}$, and rested most conveniently on the intermediate situation where the within-cell sampling had been completed in each of the RC cells, but the r rows and c columns of the observed array had not yet been fixed. (Note that this requires thinking about, and calculating with, the cell means for the $RC - rc$ cells which will not be observed.) The material in Section 18 is modelled after this proof.

To obtain the average value of the error mean square it is most convenient to think of the sampling as occurring in exactly the reverse order. Here it is most convenient to rest on the intermediate situation where the r rows and c columns have been fixed, but the n individuals to be selected from H in each of these rc cells are still unspecified.

21. The error mean square. The error mean square can be written in various forms, in particular as

$$\frac{1}{rc(n-1)} \sum_{i=1}^r \sum_{j=1}^c \sum_{k=1}^n (x_{ijk} - x_{ij.})^2 = \frac{1}{rc} \sum_i \sum_j \left[\frac{1}{n-1} \sum_k (x_{ijk} - x_{ij.})^2 \right].$$

So long as we think of the r rows and c columns as fixed, we have the mean of rc terms of the form

$$\frac{1}{n-1} \sum_k (x_{ijk} - x_{ij.})^2,$$

and the average value of this term is well known from the theory of finite sampling to be $\sigma_{I(i)J(j)}^2$. Thus the average value of the error mean square is

$$\frac{1}{RC} \sum_{I=1}^R \sum_{J=1}^C \sigma_{IJ}^2 = \sigma_E^2.$$

Even though this is clear, we shall give a formal proof as an introduction to the use of a significant technique.

22. Indicator variables. The direct evaluation of the average values of mean squares is greatly facilitated by the use of a simple device [4]. We introduce a set of indicator variables a_1, a_2, \dots, a_R , one for each row in the underlying array. The value of these variables depends on the particular sample of rows which has been selected for the actual array, and is given by:

$$a_I = \begin{cases} 1, & \text{if the } I\text{th row is in the sample of rows,} \\ 0, & \text{otherwise.} \end{cases}$$

From our assumptions about the sampling,

$$\text{ave} \{a_I\} = \text{ave} \{a_I^2\} = \frac{r}{R},$$

$$\text{ave} \{a_I a_J\} = \frac{r}{R} \cdot \frac{r-1}{R-1} = \frac{r(r-1)}{R(R-1)},$$

where we write "ave" for the average over all possible samples (we could have written "E" for expectation, but we preferred the more perspicuous notation).

Similarly, we introduce indicator variables b_1, b_2, \dots, b_C for columns, by

$$b_J = \begin{cases} 1, & \text{if the } J\text{th column is in the sample of columns,} \\ 0, & \text{otherwise.} \end{cases}$$

As an illustration of the use of these variables, consider the averaging over all samplings of rows and columns of

$$\frac{1}{rc} \sum_i \sum_j \sigma_{I(i),J(j)}^2 = \frac{1}{rc} \sum_I \sum_J a_I b_J \sigma_{IJ}^2$$

with which we closed the last section. The average value is to be found by replacing a_I by r/R and b_J by c/C , and we have the result announced above.

23. Fundamentals of bisampling. We now consider (i) an arbitrary population of R -by- C arrays $\{y_{IJ}; I = 1, 2, \dots, R; J = 1, 2, \dots, C\}$; (ii) the operation of randomly sampling r rows and c columns; (iii) the resulting r -by- c arrays; and (iv) the row, column, and grand means for these r -by- c arrays. It is our purpose to calculate averages of certain symmetric quadratic expressions in the y_{IJ} , both certain symmetric combinations of variances and covariances of these elements and means, and the averages of certain differences of squares. In the next section knowledge of these combinations will immediately lead us to the formulas for average values of mean squares in a (replicated) two-way classification.

We write, and thus change our convention about capital letters,

$$\begin{aligned} Y_{IJ} &= \text{ave } \{y_{IJ}\}, \\ Y_I &= \frac{1}{C} \sum_J Y_{IJ}, \\ y_I &= \frac{1}{c} \sum_J b_J y_{IJ}, \\ Z_J &= \frac{1}{R} \sum_I Y_{IJ}, \\ z_J &= \frac{1}{r} \sum_I a_I y_{IJ}, \\ Y &= \frac{1}{C} \sum_J Z_J = \frac{1}{R} \sum_I Y_I, \\ y &= \frac{1}{c} \sum_J b_J z_J = \frac{1}{r} \sum_I a_I y_I, \end{aligned}$$

and note that

$$\text{ave } \{y_I\} = Y_I, \quad \text{ave } \{z_J\} = Z_J, \quad \text{ave } \{y\} = Y.$$

We shall find these notations convenient, although we could have written Y_{I-} for Y_I , Y_{-J} for Z_J and Y_{--} for Y . If we had done this, there would have been no convenient parallel notation for the quantities denoted by small letters, since y_I is defined, and will be used, whether or not row I appears in a particular sample of rows, and hence in a particular r -by- c array.

We shall want our results expressed simply. They will have to involve both the average values, Y_{IJ} , and the variances and covariances, of the y_{IJ} . By intuition, or by working out the answer, we can see that they will involve only a few rather symmetric combinations of these moments, namely the three variance components corresponding to the Y_{IJ} :

$$\begin{aligned} \sigma_{\text{row}}^2 &= \frac{1}{R-1} [\sum_I Y_I^2 - RY^2] = \frac{1}{R-1} \sum_I (Y_{I-} - Y_{--})^2, \\ \sigma_{\text{col}}^2 &= \frac{1}{C-1} [\sum_J Z_J^2 - CY^2] = \frac{1}{C-1} \sum_J (Y_{-J} - Y_{--})^2, \\ \sigma_{\text{int}}^2 &= \frac{1}{(R-1)(C-1)} [\sum_I \sum_J Y_{IJ}^2 - C \sum_I Y_I^2 - R \sum_J Z_J^2 + CRY^2] \\ &= \frac{1}{(R-1)(C-1)} \sum_I \sum_J (Y_{IJ} - Y_{I-} - Y_{-J} + Y_{--})^2, \end{aligned}$$

and certain mean variances and covariances, namely,

$$\rho_1 = \frac{1}{RC} \sum_I \sum_J \text{var } \{y_{IJ}\},$$

$$\begin{aligned} \rho_2 &= \frac{1}{RC(C-1)} \sum_I \sum_{J \neq L} \text{cov} \{y_{IJ}, y_{IL}\}, \\ \rho_3 &= \frac{1}{R(R-1)C} \sum_{I \neq K} \sum_J \text{cov} \{y_{IJ}, y_{KJ}\}, \\ \rho_4 &= \frac{1}{R(R-1)C(C-1)} \sum_{I \neq K} \sum_{J \neq L} \text{cov} \{y_{IJ}, y_{KL}\}. \end{aligned}$$

We begin by writing $y_I y_K$ out as

$$y_I y_K = \frac{1}{c^2} \left[\sum_J b_J^2 y_{IJ} y_{KJ} + \sum_{J \neq L} b_J b_L y_{IJ} y_{KL} \right]$$

and using the independence of the b 's and the y 's in

$$\text{ave} \{b_J b_L y_{IJ} y_{KL}\} = \frac{c(c-1)}{C(C-1)} [Y_{IJ} Y_{KL} + \text{cov} \{y_{IJ}, y_{KL}\}]$$

to find

$$\begin{aligned} \text{ave} \{y_I y_K\} &= \frac{1}{cC} \left[\sum_J Y_{IJ} Y_{KJ} + \frac{c-1}{C-1} \sum_{J \neq L} Y_{IJ} Y_{KL} \right] \\ &+ \frac{1}{cC} \left[\sum_J \text{cov} \{y_{IJ}, y_{KJ}\} + \frac{c-1}{C-1} \sum_{J \neq L} \text{cov} \{y_{IJ}, y_{KL}\} \right]. \end{aligned}$$

Now if we use

$$Y_I Y_K = \frac{1}{C^2} \sum_J Y_{IJ} Y_{KJ} + \frac{1}{C^2} \sum_{J \neq L} Y_{IJ} Y_{KL},$$

and $\text{ave} \{y_I\} = Y_I$, $\text{ave} \{y_K\} = Y_K$, and reduce, we have

$$\begin{aligned} \text{cov} \{y_I, y_K\} &= \left(\frac{1}{c} - \frac{1}{C} \right) \frac{1}{C-1} \left[\sum_J Y_{IJ} Y_{KJ} - C Y_I Y_K \right] \\ (1) \quad &+ \frac{1}{cC} \left[\sum_J \text{cov} \{y_{IJ}, y_{KJ}\} + \frac{c-1}{C-1} \sum_{J \neq L} \text{cov} \{y_{IJ}, y_{KL}\} \right]. \end{aligned}$$

This is the key result.

If we sum this over I and K with $I \neq K$ and reduce, using

$$\sum_{I \neq K} \sum Y_{IJ} Y_{KJ} + \sum_I Y_{II}^2 = R^2 Z_J^2,$$

we find

$$\begin{aligned} &\sum_{I \neq K} \text{cov} \{y_I, y_K\} \\ &= -\left(\frac{1}{c} - \frac{1}{C} \right) \frac{1}{C-1} \left[\sum \sum Y_{IJ}^2 - C \sum Y_i^2 - R^2 \sum Z_j^2 + CR^2 Y^2 \right] \end{aligned}$$

$$\begin{aligned}
 (2) \quad & + \frac{1}{cC} \left[\sum_{I \neq K} \sum_J \text{cov} \{y_{IJ}, y_{KJ}\} + \frac{c-1}{C-1} \sum_{I \neq K} \sum_{J \neq L} \text{cov} \{y_{IJ}, y_{KL}\} \right] \\
 & = -\left(\frac{1}{c} - \frac{1}{C}\right) [(R-1)\sigma_{\text{int}}^2 - R(R-1)\sigma_{\text{col}}^2] \\
 & \quad + R(R-1) \frac{\rho_3 + (c-1)\rho_4}{c}.
 \end{aligned}$$

If we go back to (1), and put $K = I$, we have

$$\begin{aligned}
 (3) \quad \text{var} \{y_I\} & = \left(\frac{1}{c} - \frac{1}{C}\right) \frac{1}{C-1} \left[\sum_J Y_{IJ}^2 - CY_I^2 \right] \\
 & \quad + \frac{1}{cC} \left[\sum_J \text{var} \{y_{IJ}\} + \frac{c-1}{C-1} \sum_{J \neq L} \text{cov} \{y_{IJ}, y_{IL}\} \right],
 \end{aligned}$$

and summing over I , we find

$$\begin{aligned}
 (4) \quad \sum_I \text{var} \{y_I\} & = \left(\frac{1}{c} - \frac{1}{C}\right) \frac{1}{C-1} \left[\sum_I \sum_J Y_{IJ}^2 - C \sum_I Y_I^2 \right] \\
 & \quad + \frac{1}{cC} \left[\sum_I \sum_J \text{var} \{y_{IJ}\} + \frac{c-1}{C-1} \sum_I \sum_{J \neq L} \text{cov} \{y_{IJ}, y_{IL}\} \right] \\
 & = \left(\frac{1}{c} - \frac{1}{C}\right) [(R-1)\sigma_{\text{int}}^2 + R\sigma_{\text{col}}^2] + R \frac{\rho_1 + (c-1)\rho_2}{c} \\
 & = R \left[\left(\frac{1}{c} - \frac{1}{C}\right) \left(1 - \frac{1}{R}\right) \sigma_{\text{int}}^2 + \left(\frac{1}{c} - \frac{1}{C}\right) \sigma_{\text{col}}^2 + \frac{1}{c} (\rho_1 + (c-1)\rho_2) \right].
 \end{aligned}$$

When we recall that the relation of y to y_I and Y_I is entirely analogous to that of y_I to y_{IJ} and Y_{IJ} , we see that we can paraphrase (3) to give

$$\begin{aligned}
 (5) \quad \text{var} \{y\} & = \left(\frac{1}{r} - \frac{1}{R}\right) \frac{1}{R-1} \left[\sum_I Y_I^2 - RY^2 \right] \\
 & \quad + \frac{1}{rR} \left[\sum_I \text{var} \{y_I\} + \frac{r-1}{R-1} \sum_{I \neq K} \text{cov} \{y_I, y_K\} \right] \\
 & = \left(\frac{1}{r} - \frac{1}{R}\right) \sigma_{\text{row}}^2 + \frac{1}{rR} \left[\left(\frac{1}{c} - \frac{1}{C}\right) [(R-1)\sigma_{\text{int}}^2 + R\sigma_{\text{col}}^2] \right. \\
 & \quad \left. + R \frac{\rho_1 + (c-1)\rho_2}{c} \right. \\
 & \quad \left. - \frac{r-1}{R-1} \left(\frac{1}{c} - \frac{1}{C}\right) [(R-1)\sigma_{\text{int}}^2 - R(R-1)\sigma_{\text{col}}^2] \right. \\
 & \quad \left. + \frac{r-1}{R-1} R(R-1) \frac{\rho_3 + (c-1)\rho_4}{c} \right]
 \end{aligned}$$

$$= \left(\frac{1}{r} - \frac{1}{R}\right) \sigma_{\text{row}}^2 + \left(\frac{1}{c} - \frac{1}{C}\right) \sigma_{\text{col}}^2 + \left(\frac{1}{r} - \frac{1}{R}\right) \left(\frac{1}{c} - \frac{1}{C}\right) \sigma_{\text{int}}^2 + \frac{1}{rc} \left[\rho_1 + (c - 1)\rho_2 + (r - 1)\rho_3 + (r - 1)(c - 1)\rho_4 \right].$$

We shall also be interested in the combination,

$$(6) \quad \frac{1}{R} \sum_I \text{var} \{y_I\} - \text{var} \{y\} = - \left(\frac{1}{r} - \frac{1}{R}\right) \sigma_{\text{row}}^2 + \frac{r - 1}{r} \left(\frac{1}{c} - \frac{1}{C}\right) \sigma_{\text{int}}^2 + \frac{r - 1}{rc} [\rho_1 - \rho_3 + (c - 1)(\rho_2 - \rho_4)],$$

which becomes, on adding

$$\frac{1}{R} \sum_I (\text{ave} \{y_I\})^2 - (\text{ave} \{y\})^2 = \frac{1}{R} [\sum Y_I^2 - RY^2] = \frac{R - 1}{R} \sigma_{\text{row}}^2$$

on corresponding sides,

$$(7) \quad \text{ave} \left\{ \frac{1}{R} \sum y_I^2 - y^2 \right\} = \left(1 - \frac{1}{r}\right) \left[\sigma_{\text{row}}^2 + \left(\frac{1}{c} - \frac{1}{C}\right) \sigma_{\text{int}}^2 + \frac{1}{c} [\rho_1 - \rho_3 + (c - 1)(\rho_2 - \rho_4)] \right].$$

When we observe that the expression for $\sum_J \text{var} \{z_J\}$ follows from that for $\sum_I \text{var} \{y_I\}$ by symmetry (interchanging c with r , C with R , and "rows" with "col"), we see that we can easily evaluate another combination

$$(8) \quad \sum_I \sum_J \text{var} \{y_{IJ}\} - C \sum_I \text{var} \{y_I\} - R \sum_J \text{var} \{z_J\} + RC \text{var} \{y\} = RC \left[\left[\left(1 - \frac{1}{r}\right) \left(1 - \frac{1}{c}\right) - \left(1 - \frac{1}{R}\right) \left(1 - \frac{1}{C}\right) \right] \sigma_{\text{int}}^2 + \left(1 - \frac{1}{r}\right) \left(1 - \frac{1}{c}\right) [\rho_1 - \rho_2 - \rho_3 + \rho_4] \right],$$

which becomes, on adding

$$\sum_I \sum_J (\text{ave} \{y_{IJ}\})^2 - C \sum_I (\text{ave} \{y_I\})^2 - R \sum_J (\text{ave} \{z_J\})^2 + RC(\text{ave} \{y\})^2 = \sum_I \sum_J Y_{IJ}^2 - C \sum_I Y_I^2 - R \sum_J Z_J^2 + RCY^2 = (R - 1)(C - 1)\sigma_{\text{int}}^2$$

to corresponding sides,

$$(9) \quad \text{ave} \left\{ \sum_I \sum_J y_{IJ}^2 - C \sum_I y_I^2 - R \sum_J z_J^2 + RCy^2 \right\} = RC \left(1 - \frac{1}{r}\right) \left(1 - \frac{1}{c}\right) [\sigma_{\text{int}}^2 + \rho_1 - \rho_2 - \rho_3 + \rho_4].$$

These formulas will provide us with the desired results. For our limited purposes, they are the fundamentals of bisampling.

24. The replicated two-way classification. In order to develop the formulas for the remaining average mean squares, we shall find it convenient to think of the sampling of n from N as taking place first, and taking place within each of the RC cells of the underlying array, and of the sampling of r rows from R and c columns from C as taking place later. Taking this attitude, we can work with the means determined for each of the RC underlying cells after the sampling within cells and before the sampling of rows and columns. We take Y_{IJ} as the mean for the IJ th cell. We then have

$$\begin{aligned} \text{ave } \{y_{IJ}\} &= Y_{IJ} = X_{IJ-}, \\ \text{var } \{y_{IJ}\} &= \left(\frac{1}{n} - \frac{1}{N}\right) \sigma_{IJ}^2, \\ \text{cov } \{y_{IJ}, y_{KL}\} &= 0, \text{ if } (I, J) \neq (K, L). \end{aligned}$$

and

$$\begin{aligned} \sigma_{\text{row}}^2 &= \frac{1}{R-1} \sum_I (X_{I--} - X_{---})^2 = \sigma_R^2, \\ \sigma_{\text{col}}^2 &= \frac{1}{C-1} \sum_J (X_{-J-} - X_{---})^2 = \sigma_C^2, \\ \sigma_{\text{int}}^2 &= \frac{1}{(R-1)(C-1)} \sum_I \sum_J (X_{IJ-} - X_{I--} - X_{-J-} + X_{---})^2 = \sigma_I^2, \\ \rho_1 &= \frac{1}{RC} \left(\frac{1}{n} - \frac{1}{N}\right) \sum_I \sum_J \sigma_{IJ}^2 = \left(\frac{1}{n} - \frac{1}{N}\right) \sigma_E^2 \\ \rho_2 &= \rho_3 = \rho_4 = 0. \end{aligned}$$

The row mean square can be written in various forms, including

$$\begin{aligned} \frac{nc}{r-1} \sum_{i=1}^r (x_{i--} - x_{---})^2 &= \frac{nc}{r-1} \sum_i x_{i--}^2 - \frac{ncr}{r-1} x_{---}^2 \\ &= \frac{nc}{r-1} [\sum_I a_I y_I^2 - ry^2]. \end{aligned}$$

If we use first the independence of a_I from y_I and then (7) of the last section, the average value of this last form becomes

$$\begin{aligned} \frac{nc}{r-1} \text{ave} \left\{ \frac{r}{R} \sum_I y_I^2 - ry^2 \right\} &= \frac{ncr}{r-1} \left(1 - \frac{1}{r}\right) \left[\sigma_R^2 + \left(\frac{1}{c} - \frac{1}{C}\right) \sigma_I^2 \right. \\ &\quad \left. + \frac{1}{c} \left(\frac{1}{n} - \frac{1}{N}\right) \sigma_E^2 \right] = nc\sigma_R^2 + n \left(1 - \frac{c}{C}\right) \sigma_I^2 + \left(1 - \frac{n}{N}\right) \sigma_E^2 \end{aligned}$$

as we wished to show. The average value of the columns mean square follows by symmetry.

The interaction mean square can also be written in many forms, some of which are

$$\begin{aligned} & \frac{n}{(r-1)(c-1)} \sum_i \sum_j (x_{ij} - x_{i..} - x_{.j} + x_{\dots})^2 \\ &= \frac{n}{(r-1)(c-1)} [\sum_i \sum_j x_{ij}^2 - c \sum_i x_{i..}^2 - r \sum_j x_{.j}^2 + rcx_{\dots}^2] \\ &= \frac{n}{(r-1)(c-1)} [\sum_I \sum_J a_I b_J y_{IJ}^2 - c \sum_I a_I y_I^2 - r \sum_J b_J z_J^2 + rcy^2]. \end{aligned}$$

If again we use the independence of a_I and b_J from each other and the other quantities, and then (9) of the last section, the average value of the last form is seen to be

$$\begin{aligned} & \frac{n}{(r-1)(c-1)} \left[\frac{rc}{RC} \sum_I \sum_J Y_{IJ}^2 - \frac{rc}{R} \sum_I Y_I^2 - \frac{rc}{C} \sum_J Z_J^2 + rcY^2 \right] \\ &= \frac{n}{(r-1)(c-1)} RC \frac{rc}{RC} \left(1 - \frac{1}{r}\right) \left(1 - \frac{1}{c}\right) \left[\sigma_I^2 + \left(\frac{1}{n} - \frac{1}{N}\right) \sigma_B^2 \right] \\ &= n\sigma_I^2 + \left(1 - \frac{n}{N}\right) \sigma_B^2 \end{aligned}$$

as we wished to show. Thus we have the average values of the mean squares for the two-way pigeonhole model with replication.

25. The unreplicated three-way classification. To deal with the three-way pigeonhole model, where r rows from R , c columns from C , and s slices from S are independently sampled, we start out to calculate the average mean squares for rows, columns, and their interaction just as for the replicated two-way. Differences will first appear when we come to calculate ρ_1 , ρ_2 , ρ_3 , and ρ_4 , which is most simply done in an indirect way.

We remark that, now, we have

$$\begin{aligned} y_{IJ} &= \frac{1}{s} \sum_{k=1}^s x_{IJK(k)} = \frac{1}{S} \sum_{K=1}^S c_K x_{IJK}, \\ \text{ave } \{y_{IJ}\} &= Y_{IJ} = \frac{1}{S} \sum_K x_{IJK}, \end{aligned}$$

where $\{c_K\}$ are a new set of indicator variables which specify the sampling of s slices from S . Now

$$\text{var } \{y_{IJ}\} = \left(\frac{1}{s} - \frac{1}{S}\right) \frac{1}{S-1} \left[\sum_K x_{IJK}^2 - Sx_{IJ.}^2 \right]$$

since we have a sample mean of s from S , so that

$$RC\rho_1 = \sum_I \sum_J \text{var } \{y_{IJ}\} = \left(\frac{1}{s} - \frac{1}{S}\right) \frac{1}{S-1} \left[\sum_I \sum_J \sum_K x_{IJK}^2 - S \sum_I \sum_J x_{IJ-}^2\right].$$

Moreover

$$\begin{aligned} \text{var } \left\{ \sum_J y_{IJ} \right\} &= \left(\frac{1}{s} - \frac{1}{S}\right) \frac{1}{S-1} \left[\sum_K (\sum_J x_{IJK})^2 - S(\sum_J x_{IJ-})^2 \right] \\ &= \left(\frac{1}{s} - \frac{1}{S}\right) \frac{1}{S-1} \left[C^2 \sum_K x_{I-K}^2 - C^2 S x_{I--}^2 \right]. \end{aligned}$$

Since $\sum_J y_{IJ}$ is the mean of a sample of s from the S values $\sum_J x_{IJK}$, and hence

$$\begin{aligned} RC\rho_1 + RC(C-1)\rho_2 &= \sum_I \text{var } \left\{ \sum_J y_{IJ} \right\} \\ &= \left(\frac{1}{s} - \frac{1}{S}\right) \frac{C^2}{S-1} \left[\sum_I \sum_K x_{I-K}^2 - S \sum_I x_{I--}^2 \right]. \end{aligned}$$

By symmetry, then

$$RC\rho_1 + R(R-1)C\rho_3 = \left(\frac{1}{s} - \frac{1}{S}\right) \frac{R^2}{S-1} \left[\sum_J \sum_K x_{-JK}^2 - S \sum_J x_{-J-}^2 \right]$$

and, indeed,

$$\begin{aligned} RC\rho_1 + RC(C-1)\rho_2 + R(R-1)C\rho_3 + R(R-1)C(C-1)\rho_4 \\ = \text{var } \left\{ \sum_I \sum_J y_{IJ} \right\} &= \left(\frac{1}{s} - \frac{1}{S}\right) \frac{1}{S-1} \left[\sum_K (\sum_I \sum_J x_{IJK})^2 - S(\sum_I \sum_J x_{IJ-})^2 \right] \\ &= \left(\frac{1}{s} - \frac{1}{S}\right) \frac{C^2 R^2}{S-1} \left[\sum_K x_{--K}^2 - S X_{--}^2 \right]. \end{aligned}$$

We now have the basis for evaluating the ρ 's.

We introduce the additional variance components for the three-way classification by

$$\begin{aligned} \sigma_s^2 &= \frac{1}{S-1} \left[\sum_K x_{--K}^2 - S x_{--}^2 \right], \\ \sigma_{rs}^2 &= \frac{1}{(R-1)(S-1)} \left[\sum_I \sum_K x_{I-K}^2 - R \sum_K x_{--K}^2 - S \sum_I x_{I--}^2 + R S x_{--}^2 \right], \\ \sigma_{cs}^2 &= \frac{1}{(C-1)(S-1)} \left[\sum_J \sum_K x_{-JK}^2 - C \sum_J x_{--K}^2 - S \sum_J x_{-J-}^2 + C S x_{--}^2 \right], \end{aligned}$$

and finally,

$$\begin{aligned} \sigma_{rcs}^2 &= \frac{1}{(R-1)(C-1)(S-1)} \left[\sum_I \sum_J \sum_K x_{IJK}^2 - R \sum_J \sum_K x_{-JK}^2 - C \sum_I \sum_K x_{I-K}^2 \right. \\ &\quad \left. - S \sum_I \sum_J x_{IJ-}^2 + RC \sum_K x_{--K}^2 + RS \sum_J x_{-J-}^2 + CS \sum_I x_{I--}^2 - RCS x_{--}^2 \right]. \end{aligned}$$

In terms of these quantities we can now calculate the values of the ρ 's. We can conveniently use

$$\begin{aligned} \sum_I \sum_K x_{I-K}^2 - S \sum_I x_{I--}^2 &= (R - 1)(S - 1)\sigma_{RS} + R[\sum_K x_{--K}^2 - Sx_{---}^2] \\ &= (R - 1)(S - 1)\sigma_{RS}^2 + R(S - 1)\sigma_S^2, \end{aligned}$$

and, by symmetry,

$$\sum_J \sum_K x_{-JK}^2 - S \sum_J x_{-J-}^2 = (C - 1)(S - 1)\sigma_{CS}^2 + C(S - 1)\sigma_S^2,$$

and, by a similar though longer calculation,

$$\begin{aligned} \sum_I \sum_J \sum_K x_{IJK}^2 - S \sum_I \sum_J x_{IJ-}^2 &= (R - 1)(C - 1)(S - 1)\sigma_{RCS}^2 \\ &\quad + R(C - 1)(S - 1)\sigma_{CS}^2 + (R - 1)C(S - 1)\sigma_{RS}^2 + RC(S - 1)\sigma_S^2. \end{aligned}$$

Substituting these into the earlier formulas, and removing factors where convenient, yields

$$\begin{aligned} \rho_1 &= \left(\frac{1}{S} - \frac{1}{S}\right) \left[\left(1 - \frac{1}{R}\right) \left(1 - \frac{1}{C}\right) \sigma_{RCS}^2 + \left(1 - \frac{1}{C}\right) \sigma_{CS}^2 + \left(1 - \frac{1}{R}\right) \sigma_{RS}^2 + \sigma_S^2 \right], \\ \rho_1 + (C - 1)\rho_2 &= \left(\frac{1}{S} - \frac{1}{S}\right) C \left[\left(1 - \frac{1}{R}\right) \sigma_{RS}^2 + \sigma_S^2 \right], \\ \rho_1 + (R - 1)\rho_3 &= \left(\frac{1}{S} - \frac{1}{S}\right) R \left[\left(1 - \frac{1}{C}\right) \sigma_{CS}^2 + \sigma_S^2 \right], \\ \rho_1 + (C - 1)\rho_2 + (R - 1)\rho_3 + (C - 1)(R - 1)\rho_4 &= \left(\frac{1}{S} - \frac{1}{S}\right) RC\sigma_S^2. \end{aligned}$$

From these we could compute all the ρ 's, but it is simpler to find the combinations which will be of most use to us. These are

$$\begin{aligned} \rho_1 - \rho_3 &= \left(\frac{1}{S} - \frac{1}{S}\right) \left[\left(1 - \frac{1}{C}\right) \sigma_{RCS}^2 + \sigma_{RS}^2 \right], \\ \rho_2 - \rho_4 &= \left(\frac{1}{S} - \frac{1}{S}\right) \left[\left(1 - \frac{1}{C}\right) \sigma_{RCS}^2 + \sigma_{RS}^2 \right], \end{aligned}$$

whence

$$\begin{aligned} \rho_1 - \rho_3 + (C - 1)(\rho_2 - \rho_4) &= \left(\frac{1}{S} - \frac{1}{S}\right) \left[\left(1 - \frac{C}{C}\right) \sigma_{RCS}^2 + \sigma_{RS}^2 \right], \\ \rho_1 - \rho_2 - \rho_3 + \rho_4 &= \left(\frac{1}{S} - \frac{1}{S}\right) [\sigma_{RCS}^2]. \end{aligned}$$

We are now ready to return to the evaluations of Section 23, where we found the average mean squares for rows to be

$$nc\sigma_R^2 + n \left(1 - \frac{C}{C}\right) \sigma_I^2 + n[\rho_1 - \rho_3 + (C - 1)(\rho_2 - \rho_4)].$$

We have now to write s in place of n , σ_{RC}^2 in place of σ_I^2 , and to substitute in the value we have just found for quantity in brackets. We obtain, then, for the average value of a row mean square in a pigeonhole three-way

$$cs\sigma_R^2 + s\left(1 - \frac{c}{C}\right)\sigma_{RC}^2 + c\left(1 - \frac{s}{S}\right)\sigma_{RS}^2 + \left(1 - \frac{c}{C}\right)\left(1 - \frac{s}{S}\right)\sigma_{RCS}^2.$$

The average values of column and slice mean squares now follow by symmetry.

We had obtained

$$n\sigma_I^2 + n[\rho_1 - \rho_2 - \rho_3 - \rho_4]$$

for the average value of the interaction mean square. Making the necessary changes, the average value of the RC -interaction mean square for the three-way pigeonhole becomes

$$s\sigma_{RC}^2 + \left(1 - \frac{s}{S}\right)\sigma_{RCS}^2,$$

and the average values for RS and CS interactions follow by symmetry. There remains the triple interaction.

26. The triple interaction. To get a simple hold on the triple interaction, we will find it convenient to introduce some differencing operators with the following definitions:

$$\delta_{ii'}f(i, j, k) = \frac{1}{\sqrt{2}}(f(i, j, k) - f(i', j, k)),$$

$$\Delta_{II'}f(I, J, K) = \frac{1}{\sqrt{2}}(f(I, J, K) - f(I', J, K)).$$

and similarly for other indices. The usefulness of these operators stems from the following chain of representations for interaction sums of squares:

$$\sum_i x_{i..}^2 - rx^2_{..} \equiv \frac{1}{r} \sum_i \sum_{i'} (\delta_{ii'} x_{i..})^2,$$

$$\sum_i \sum_j x_{ij.}^2 - r \sum_j x_{.j.}^2 - c \sum_i x_{i..}^2 + rcx^2_{..} \equiv \frac{1}{rc} \sum_i \sum_{i'} \sum_j \sum_{j'} (\delta_{ii'} \delta_{jj'} x_{ij.})^2.$$

$$\sum_i \sum_j \sum_k x_{ijk}^2 - r \sum_j \sum_k x_{.jk}^2 - c \sum_i \sum_k x_{i.k}^2 - s \sum_i \sum_j x_{ij.}^2 + rc \sum_k x^2_{..k}$$

$$+ rs \sum_j x^2_{.j.} - cs \sum_i x^2_{i..} - rcs^2_{..} = \frac{1}{rcs} \sum_i \sum_{i'} \sum_j \sum_{j'} \sum_k \sum_{k'} (\delta_{ii'} \delta_{jj'} \delta_{kk'} x_{ijk})^2.$$

and so on.

To establish these representations we have only to prove (as we may by direct expansion), that

$$\sum_{h=1}^m u_h^2 - mu^2 \equiv \frac{1}{m} \sum_h \sum_{h'} (\delta_{hh'} u_h)^2$$

and apply induction, writing, for example, the second interaction in the forms

$$\begin{aligned} \sum_j [\sum_i x_{ij}^2 - rx_{.j}^2] - c[\sum_i x_{i..}^2 - rx_{...}^2] \\ = \sum_j \left[\frac{1}{r} \sum_i \sum_{i'} (\delta_{ii'} x_{ij.})^2 \right] - c \left[\frac{1}{r} \sum_i \sum_{i'} (\delta_{ii'} x_{i..})^2 \right] \\ = \sum_i \sum_{i'} \frac{1}{r} \left[\sum_j (\delta_{ii'} x_{ij.})^2 - c(\delta_{ii'} x_{i..})^2 \right] \\ = \sum_i \sum_{i'} \frac{1}{r} \left[\sum_j \sum_{j'} \frac{1}{c} (\delta_{jj'} \delta_{ii'} x_{ij.})^2 \right], \end{aligned}$$

where we have used the identity twice and the fact that the mean value of $\delta_{ii'} x_{ij.}$ over j is $\delta_{ii'} x_{i..}$ itself.

These representations lead at once to average value formulas. Thus, in the one-way case

$$\begin{aligned} \text{ave} \left\{ \sum_i x_i^2 - rx^2 \right\} &= \text{ave} \left\{ \frac{1}{r} \sum_I \sum_{I'} (\delta_{II'} x_I)^2 \right\} = \text{ave} \left\{ \frac{1}{r} \sum_I \sum_{I'} a_I a_{I'} (\Delta_{II'} x_I)^2 \right\} \\ &= \frac{1}{r} \sum_I \sum_{I'} \frac{r(r-1)}{R(R-1)} (\Delta_{II'} x_I)^2 = \left(\frac{r-1}{R-1} \right) \frac{1}{R} \sum_I \sum_{I'} (\Delta_{II'} x_I)^2 \\ &= \frac{r-1}{R-1} [\sum_I x_I^2 - Rx^2], \end{aligned}$$

where we have used the fact that $a_I a_{I'}$ equals $r(r-1)/R(R-1)$ except when $I = I'$, and the fact that we can neglect $I = I'$ because $\delta_{II'} x_I$ vanishes for $I = I'$. Similarly, in a two-way case

$$\begin{aligned} \text{ave} \left\{ \sum_i \sum_j x_{ij}^2 - r \sum_j x_{.j}^2 - c \sum_i x_{i..}^2 + rcx_{...}^2 \right\} \\ = \text{ave} \left\{ \frac{1}{rc} \sum_i \sum_{i'} \sum_j \sum_{j'} (\delta_{ii'} \delta_{jj'} x_{ij.})^2 \right\} \\ = \frac{1}{rc} \text{ave} \left\{ \sum_I \sum_{I'} \sum_J \sum_{J'} a_I a_{I'} b_J b_{J'} (\Delta_{II'} \Delta_{JJ'} x_{IJ})^2 \right\} \\ = \frac{1}{rc} \frac{r(r-1)}{R(R-1)} \frac{c(c-1)}{C(C-1)} \sum_I \sum_{I'} \sum_J \sum_{J'} (\Delta_{II'} \Delta_{JJ'} x_{IJ})^2 \\ = \frac{(r-1)(c-1)}{(R-1)(C-1)} [\sum_I \sum_J x_{IJ}^2 - R \sum_J x_{.J}^2 - C \sum_I x_{i..}^2 + RCx_{...}^2] \end{aligned}$$

and so on. The next case, for the three-way case, shows us that the average value of the RCS -interaction is exactly σ_{RCS}^2 as we wished to prove.

Clearly the extensions of this argument to the four-way, five-way, etc., classifications will always give a similar answer for the lowest interaction (the one involving all the indices!) for any factorial.

27. Further generalization. It is just a matter of moderately stiff algebra to carry on an inductive proof of similar results for more classifications. If $\text{ave } \{y_{IJK}\} = Y_{IJK}$, and if $q(\{y_{ijk}\})$ is a quadratic in the observed y_{ijk} , then

$$\text{ave } \{q(\{y_{ijk}\})\} = \text{ave } \{q(\{Y_{ijk}\})\}$$

+ (a linear form in variances and covariances of the (y_{ijk}))

Now the capital Y 's correspond to an analysis without replication, so that, for the $q(\{Y_{ijk}\})$ we are interested in, the averages on the right can be calculated from the results of the last section. Because of symmetry, the linear form in variances and covariances will be a linear combination of 8 quantities, the mean variance, and seven mean covariances. Working these expressions out, and then calculating the values of these mean variance and covariances in special cases, we can obtain the average values of mean squares for the replicated three-way and unreplicated four-way designs.

After this we are ready for another step of the induction, and so on.

Clearly systematic algebra can take us deep into the forest of notation. But the detailed manipulation will, sooner or later, blot out any understanding we may have started with. If there is a way of seeing some aspects of the final result more directly, then it will be worth while to seize it.

There are a number of such ways involving special tools or devices of varying complexity. Since we have tried to keep the approach of this paper reasonably pedestrian (although indicator variables and the $\delta_{ii'}$ and $\Delta_{II'}$ may be regarded as the equivalent of roller skates!), we shall try to use the least special way that we know.

28. More direct insight. Let us ask about the coefficient with which

$$\sigma_{RCDEFGS}^2$$

appears in the average value of the mean square of the $RCDE$ -interaction in an 11-way design with factors labelled $R, C, D, E, F, G, H, J, K, L, S$. More specifically, let us ask how the coefficient depends on the values of s and S . It will appear that we can answer this rather directly.

There will exist some formulas which make up the fundamentals of decim-sampling. Apply them to the 10-way classification involving $R, C, D, E, F, G, H, J, K, L$ —involving all the classifications but S . They will give the average value of the $RCDE$ -interaction in terms of σ^2 's whose indices do not include S and of a mean variance and $2^{10} - 1 = 1023$ mean covariances. The latter 1024 quantities will be expressible in terms of the σ^2 's which involve S by a process entirely similar (though more complicated in detail) than that used in Section 15, each of the 1024 formulas for a linear combination of mean variance and covariances will involve a factor

$$\left(\frac{1}{s} - \frac{1}{S}\right),$$

since, *in every case*, we shall be sampling s out of S . Hence, when all the algebraic

dust has quieted down, the coefficient of $\sigma_{RCDEFGS}^2$ will include the same factor, and will depend in no other way on S . It will depend on s in a further way, since finding the row, column, D -variable, and E -variable will still leave us with $fg hjkl$ s different cells. Thus a factor of this value will appear in the mean square for $RCDE$ when that mean square is written out in terms of means (not totals). Thus the complete dependence on s and S will be

$$s \left(\frac{1}{s} - \frac{1}{S} \right) = 1 - \frac{s}{S}.$$

With this result, and the result, obtained incidentally, that h and H will enter only through a factor h , we are essentially finished.

By symmetry, we see that the term in the mean-value of the $RCDE$ -interaction with which we are concerned is

$$(\text{constant}) \left(1 - \frac{f}{F} \right) \left(1 - \frac{g}{G} \right) h j k l \left(1 - \frac{s}{S} \right) \sigma_{RCDEFGS}^2$$

If we simplify matters by choosing $h = H = j = J = k = K = l = L = 1$, the whole 11-way classification condenses to a 7-way classification and if, moreover, we revert to Model II (everything independent and normal), we obtain

$$(\text{constant}) \sigma_{RCDEFGS}^2$$

as a term in the average value of the $RCDE$ -mean square, and it is well known that the constant must now be unity.

This provides a not-too-indirect proof for the rules set forth in Section 11.

REFERENCES

- [1] R. L. ANDERSON AND T. A. BANCROFT, *Statistical Theory in Research*, McGraw-Hill, New York, 1952.
- [2] C. A. BENNETT AND N. L. FRANKLIN, *Statistical Analysis in Chemistry and the Chemical Industry*, John Wiley, New York, 1954.
- [3] W. G. COCHRAN, "The use of analysis of variance in enumeration by sampling," *J. Amer. Statist. Assoc.*, Vol. 34 (1939), pp. 492-510.
- [4] J. CORNFIELD, "On samples from finite populations," *J. Amer. Statist. Assoc.*, Vol. 3 (1944), pp. 236.
- [5] C. EISENHART, "The assumptions underlying the analysis of variance," *Biometrics*, Vol. 3 (1947), pp. 1-21.
- [6] M. H. HANSEN, W. N. HURWITZ AND W. G. MADOW, *Sample Survey Methods and Theory*, Vol. 2, *Theory*, John Wiley, New York, 1953.
- [7] W. A. HENDRICKS, "Variance components as a tool for the analysis of sample data," *Biometrics*, Vol. 7 (1951), pp. 97-101.
- [8] O. KEMPTHORNE, *Design and Analysis of Experiments*, John Wiley, New York, 1952.
- [9] O. KEMPTHORNE, "The randomization theory of experimental inference," *J. Amer. Statist. Assoc.*, Vol. 50 (1955), pp. 946-967.
- [10] M. D. MCCARTHY, "On the application of the z -test to randomized blocks," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 337-359.
- [11] A. M. MOOD, *Introduction to the Theory of Statistics*, McGraw-Hill, 1950.
- [12] J. NEYMAN WITH THE COOPERATION OF K. IWASKIEWICZ AND ST. KOŁODZIECZYK, "Statistical problems in agricultural experimentation," *J. Roy. Stat. Soc., Suppl.*, Vol. 2 (1935), pp. 107-180.

- [13] E. J. G. PITMAN, "Significance tests which may be applied to samples from any population, III. The analysis of variance test," *Biometrika*, Vol. 29 (1938), pp. 322-335.
- [14] J. W. TUKEY, "Dyadic anova, an analysis of variance for vectors," *Human Biology*, Vol. 21 (1949), pp. 65-110.
- [15] B. L. WELCH, "On the z-test in randomized blocks and Latin squares," *Biometrika*, Vol. 29 (1937), pp. 21-52.
- [16] M. B. WILK, *The Randomization Analysis of a Generalized Randomized Block Design*, M.S. Thesis, Iowa State College 1953.

RESEARCH REPORTS

- [17] J. CORNFIELD, "The two-way classification in the analysis of variance from the point of view of urn sampling," mimeo from Office of Biometry, National Institutes of Health (ca. 1953).
- [18] R. HOOKE, "Symmetric functions of a two-way array," *Ann. Math. Stat.* Vol. 27 (1956), pp. 55-79.
- [19] O. KEMPTHORNE, dittoed early draft of book on *Genetic Statistics*, used in class about November, 1953.
- [20] J. W. TUKEY, "Interaction in a row-by-column design," Memorandum Report 18, Statistical Research Group, Princeton University, July, 1949.
- [21] M. B. WILK, "The logical derivation of linear models for experimental situations and their use in selecting the appropriate error term in the analysis of variance," Report AV-3 from the Analysis of Variance Project, Statistical Laboratory, Iowa State College, December, 1953.
- [22] M. B. WILK, "The logical derivation of linear models for experimental situations and their use in selecting the appropriate error term in the analysis of variance, II. On the analysis of a general balanced randomized block type design," Report AV-4 from the Analysis of Variance Project, Statistical Laboratory, Iowa State College, March, 1954.
- [23] M. B. WILK AND O. KEMPTHORNE, "The logical derivation of linear models for experimental situations and their use in selecting the appropriate error term in the analysis of variance, III. The expectation of mean square for a general three way classification," Report AV-7 from the Analysis of Variance Project, Statistical Laboratory, Iowa State College, June, 1954.
- [24] M. B. WILK AND O. KEMPTHORNE, "Fixed, mixed, and random models," *J. Amer. Statist. Assoc.*, Vol. 50 (1955), pp. 1144-1167. "The logical derivation of linear models for experimental situations and their use in selecting the appropriate error term in the analysis of variance, IV. Fixed, mixed, and random models," Report AV-8 from the Analysis of Variance Project, Statistical Laboratory, Iowa State College, August, 1954. This report covered by a paper submitted to the *J. Amer. Stat. Assn.*
- [25] M. B. WILK, "The logical derivation of linear models for experimental situations and their use in selecting the appropriate error term in the analysis of variance, V. The expectation of mean squares for a general three-way classification (Part 2)," Report AV-9 from the Analysis of Variance Project, Statistical Laboratory, Iowa State College, August, 1954.
- [26] M. B. WILK AND O. KEMPTHORNE, "Derived linear models and their use in the analysis of randomized experiments," (Final report, covering [21] through [25], Analysis of Variance Project, Statistical Laboratory, Iowa State College, April, 1955.