A Structure Diagram Symbolization for Analysis of Variance
Author(s): Warren H. Taylor, Jr. and H. Gill Hilton
Source: *The American Statistician*, Vol. 35, No. 2 (May, 1981), pp. 85-93
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/2683146
Accessed: 31-07-2017 22:37 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://about.jstor.org/terms

American Statistical Association, Taylor & Francis, Ltd. are collaborating with JSTOR to digitize, preserve and extend access to *The American Statistician*

This content downloaded from 200.17.213.205 on Mon, 31 Jul 2017 22:37:57 UTC
All use subject to http://about.jstor.org/terms

A Structure Diagram Symbolization for Analysis of Variance

WARREN H. TAYLOR, JR. AND H. GILL HILTON*

A structure diagram symbolization for balanced complete experimental designs of the type encountered in standard balanced analysis of variance (ANOVA) calculations is discussed. From the symbolization, rules of formula derivation for obtaining entries in the traditional ANOVA table are introduced. The method requires recognition of characteristic sets of factors which are combined using the concept of formal interaction. Expected mean squares, F-test ratios, and variance component estimates can be quickly derived using the symbolic method.

KEY WORDS: Analysis of variance; Balanced designs; Structure diagram; Experimental design symbolization; F tests; Hasse diagram.

1. INTRODUCTION

The usual method of performing the analysis of variance (ANOVA) calculations relies upon examination of the subscripted terms that appear in the linear statistical model. This article suggests a graphic symbolic approach to ANOVA. In the symbolic approach a structure diagram serves as the basic reference object in deriving statistical tests and computational formulas. This visual representation for analysis provides a better understanding of how the ANOVA calculations proceed by consideration of the relationships that exist among the experimental factors.

Several major benefits of the structure diagram method are attributable to a powerful symbolization that provides a simple, complete picture of an experimental design. This picture describes the fixedness or randomness of the factors, their actual names written out longhand if desired, the number of levels for each factor, and all nesting and crossing relationships among factors. Few if any pictorial methods, simple or otherwise, published in the generally available literature enjoy this descriptive ability. The method works particularly well for complex multifactor designs where it is especially difficult to understand and depict relationships among factors when there are multiple levels of nesting and crossing.

Identification of the correct linear model from a verbal description of the experiment often proves difficult for the expert as well as the novice. By providing a clear picture of the experimental design, the structure diagram forms a visual link connecting the verbal description to the linear model. This visual dimension helps to simplify and clarify the model identification process.

Another frequent source of difficulty for practitioners is the rote memorization and application of mechanical rules for performing ANOVA calculations from the subscripted linear model. This article presents visual rules of thumb that can serve as a valuable independent method to verify the correctness of computational formulas. In particular, these visual rules do not use a subscripted linear model. The visual rules can, therefore, provide increased self-confidence and accuracy for obtaining and checking the ANOVA formulas. An additional bonus is an automatic method to construct correct F-test ratios without the intermediate time-consuming step of writing down expected mean squares.

There are four key elements to successful application of the structure diagram method. The first key is understanding how the diagram portrays the relationships among experimental factors and shows what design effects (linear model terms) appear in the statistical analysis. The second key is learning how any given design effect uniquely partitions the factors in the experiment into disjoint characteristic sets of factors (factor sets), which can be named and identified by reference to the structure diagram. The third key is understanding the concept of formal interaction among effects, and how to obtain these interactions by visual reference to the structure diagram. And the fourth key consists of learning the rules of thumb for calculating statistical quantities by using the preceding concepts. The reader familiar with the traditional subscripted model approach should find that the pictorial method presented here adds refreshing insight into the mechanical rules customarily employed in completing the ANOVA table.

To assist in continuity and clarity of explanation, a rather complex experimental design structure with seven factors accompanies each section. This multifactor example serves to illustrate the general applicability of the method and is useful in all sections for illustrating basic concepts.

2. STRUCTURE DIAGRAMS

It is possible to simply and concisely represent any experimental design, no matter how complex, by drawing a small picture called a structure diagram, provided that the design falls into the large class of balanced complete designs. The picture, more correctly called a Hasse diagram, is a result of the

* Warren H. Taylor, Jr. is with the Systems Performance Department of the Defense & Space Systems Group of TRW Inc., Redondo Beach, CA 90278. H. Gill Hilton is Professor of Statistics at Brigham Young University, Provo, UT 84602. This work was completed while the senior author was a graduate student at Brigham Young University.

This content downloaded from 200.17.213.205 on Mon, 31 Jul 2017 22:37:57 UTC
All use subject to http://about.jstor.org/terms

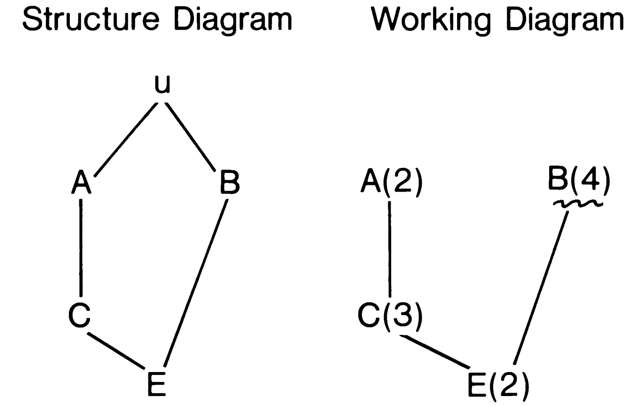


Figure 1. Structure and Working Diagrams for the Three-Factor Experiment of Equation (2.1)

mathematical correspondence between partially ordered lattice structures and experimental design structures, as discussed by Throckmorton (1961), Kempthorne et al. (1961), Zyskind et al. (1964), and Kempthorne and Folks (1971).

The relationships of nesting and crossing are fundamental to explaining the concept of experimental structure, or what is commonly called the design of an experiment. Nesting of factor B in factor A occurs if each level of B appears with one, and only one, level of A. Balanced nesting requires an equal number of B levels to appear with each A level. Complete crossing of factor B and factor A occurs if every level of B appears with every level of A. The class of balanced complete designs may be imprecisely but sufficiently well defined as those designs incorporating the principles of balance and completeness in nesting and crossing relationships among experimental factors. Typical introductory text ANOVA designs, excluding the Latin square type, fall into this class.

To represent an experimental design in diagrammatic form, a capital letter label is first assigned to each factor. The nesting or crossing of factors is then indicated by the presence or absence of connecting upward links between the factor labels. Any factor connected by upward links to other factors is considered to be nested in those factors above it. A factor that is not connected to another factor by upward links is crossed with that factor. The structure diagram then portrays in concise form the relationships of nesting and crossing among all design factors.

For illustrative purposes Figure 1 is an example of a structure diagram with three factors. In this design u represents the experimental mean; A, B, and C are three factors; and E represents the experimental error from basic experimental units treated exactly alike. Factor A is crossed with B, nests C and E, and is nested in u. Factor B is crossed with A and C, nests E, and is nested in u. Factor C is crossed with B, nested in u and A, and nests E. The mean u nests all factors, and the error E is nested in all factors.

The structure diagram representation of the design conveys not only the relationships among the experimental factors, but also implies the linear statistical model corresponding to the structure. The linear model consists of each factor written as an effect, plus one interaction term for each combination of factors that are crossed with each other. Using this rule, the model corresponding to Figure 1 is then

$$Y = u + A + B + AB + C + BC + E \qquad (2.1)$$

The correct subscripts for each model term may be written by assigning a small letter as a subscript to each main effect symbol, putting in parentheses the subscripts of all nesting factors that are linked from above. To remember which subscript goes with which effect, each main effect is assigned exactly the same small letter subscript corresponding to the name of the effect itself. By doing this the correct linear model for the structure of Figure 1 is

$$Y_{abce} = u + A_a + B_b + AB_{ab} + C_{(a)c} + BC_{(a)bc} + E_{(abc)e} \qquad (2.2)$$

From the foregoing discussion it appears that the full linear model for any balanced complete design can be written by inspecting its structural representation. This also suggests, conversely, that the information present in the linear statistical model is present in the structure diagram. It then must follow that all analysis of variance computational formulas and algorithms, as traditionally expressed in terms of the subscripted linear model, could be equivalently expressed in terms of the structure diagram. A change of reference object from the linear model to the structure diagram can provide a new conceptual outlook on the traditional analysis of variance techniques. The concepts of factor sets, formal interaction, and simple visual rules for derivation of sums of squares, expected mean squares, F-test ratios, and components of variance estimates are the results stemming from such a symbolic perspective.

A fixed factor will be designated in the structure diagram by a squiggly underline. The absence of such a designation will imply that the factor is random. The range of a factor subscript is equal to the number of levels at which the factor appears in the design. The factor ranges will be indicated in the diagram by a number in parentheses following the factor symbol. A structure diagram labeled with its fixed and random factors and their ranges can be termed a 'working diagram' because it is possible to work the ANOVA calculations from such a diagram. For example, in Figure 1, if A is random at two levels, B is fixed at four levels, C is random at three levels, and E is random at two levels, the structure diagram becomes the working diagram. The mean is omitted from the working diagram, since it is implicitly understood to be present at one level in all experiments.

It is advantageous to represent the range of a given factor, say C, by the same small letter, c, corresponding to it. The small letters, however, have already been used as subscript variables. It will be understood,

This content downloaded from 200.17.213.205 on Mon, 31 Jul 2017 22:37:57 UTC
All use subject to http://about.jstor.org/terms

therefore, as a matter of convention that a small letter used in a subscript context is a variable quantity, but when used otherwise, a small letter is a fixed quantity representing the factor range. This is the convention used by Lee (1966, 1975) in his design symbolization work. No confusion results in practice and considerable simplification is obtained in multifactor designs. In Figure 1, the factor ranges are $a = 2$, $b = 4$, $c = 3$, and $e = 2$.

To illustrate the conventions discussed, the following hypothetical experiment with seven factors is described and diagramed.

London Lathe Limited (Triple L) is a major supplier of tools and tool rests for the lathes in machine shops around the world. An industrial experiment was conducted where three factories (F) were chosen at random to conduct a study on two kinds (K) of tool rests (R) and three angles (A) of edges for lathe tools (T). Each factory received four tool rests, two of each kind. In each factory three machinists (M) were randomly chosen to use each of the tool rests on each of four randomly selected lathes (L). Each factory was assigned 96 tools, with one third of the tools being of each angle. To control wear and assure uniformity of tool usage, each tool was used only with a specific lathe and a specific tool rest.

To assist the reader in drawing analogies for this example with the factors as stated, the labeled structure diagram may be drawn as shown in Figure 2. Using the first letter of each factor to symbolize the name of the factor, the working diagram of the structure is given in Figure 3. The reader at this point should take special note of the power of the structure diagram to convey simply the names and relationships of factors in this multifactor design. Without a diagram, this design would be very difficult to visualize or to explain; but with the diagram, the complexity is reduced to visible and manageable proportions.

A few statements will be made for purposes of example to illustrate the relationships among factors shown in this design. The factor F is crossed with the factors K and A, and nests M, L, R, T, and E. Factor R is crossed with M, L, and A; is nested in F and K; and nests T and E. Factor T is nested in all factors except M, with which it is crossed, and E which it nests. The factor E is nested in all factors, is inestimable since it appears at only one level, but is retained because it may be needed in an F test later. The number of factors and variety of relationships among factors in this example are useful for illustrating the generality of the principles and methods of the following sections.

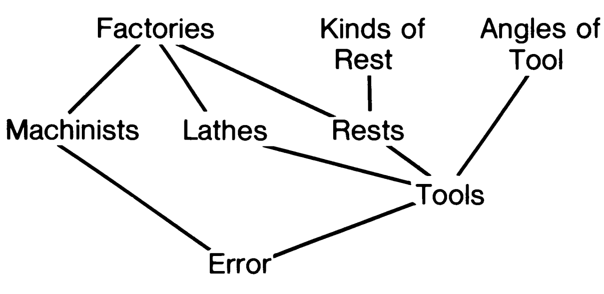


Figure 2. Structure Diagram for the Seven-Factor Triple L Design

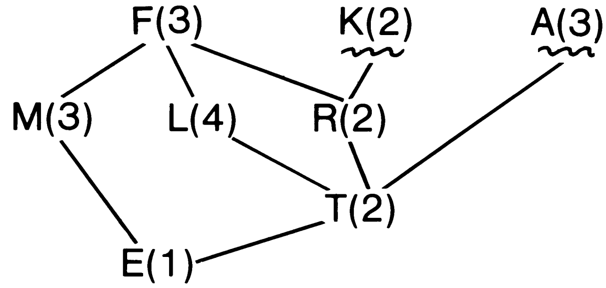


Figure 3. Working Diagram for the Seven-Factor Triple L Design

3. FACTOR SETS

The set of all factors in an experimental design is called the design set. For each effect in any model the design set can be partitioned into smaller functional groupings called factor sets. The basic principle underlying the concept of factor sets is that any effect (model term) uniquely partitions the design set into disjoint characteristic sets that are necessary in performing the details of the analysis of variance computations for that effect. The terminology used in naming factor sets follows Scheffé (1959), Lee (1975), and Collings (1977).

For any effect (linear model term) Scheffé (1959) uses the terminology of live, dead, and absent factors in obtaining statistical results, such as expected mean squares. Factors appearing in the effect name corresponding to the nonbracketed subscripts are termed live factors. Factors that nest the live factors corresponding to subscripts within brackets are termed dead factors. The remaining factors, which neither appear as a factor in the effect name nor nest a factor in the name, are termed absent factors. These classifications of live, dead, and absent partition the design set into three disjoint factor sets. Another recognized manner of partitioning the factors is by noting that a factor may be either fixed or random. Collings (1977) shows that determination of correct numerator and denominator expected mean squared terms for exact or approximate F-test ratios is facilitated by a partition of the random absent factors into two disjoint sets called simple and nonsimple. Simple factors are those random absent factors that are not nested in any other random absent factors.

To facilitate the ANOVA calculations it is convenient to group all the live and dead factors into a group called the symbolic set, as they are referred to in the text by Lee (1975). The absent factor set may also be redesignated the complement set because it is the complement of the symbolic set. With these

This content downloaded from 200.17.213.205 on Mon, 31 Jul 2017 22:37:57 UTC
All use subject to http://about.jstor.org/terms

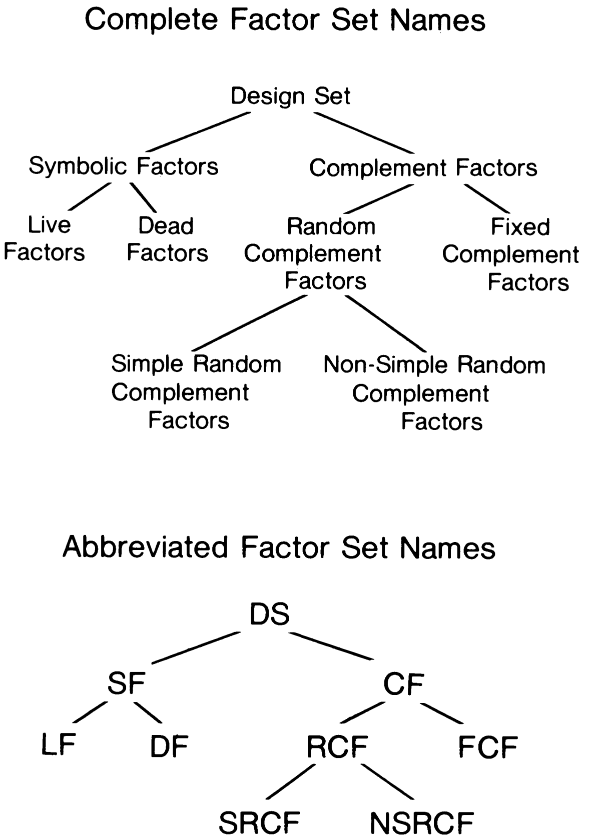


Figure 4. Complete and Abbreviated Factor Set Names Used for Partitioning the Design Set in an Experiment

definitions, a partition of the design set into factor sets useful for ANOVA calculations is effected. Figure 4 represents the partition of the design set into its composite factor sets. Abbreviation of the factor set names, as shown in Figure 4, avoids writing out the entire name for each set.

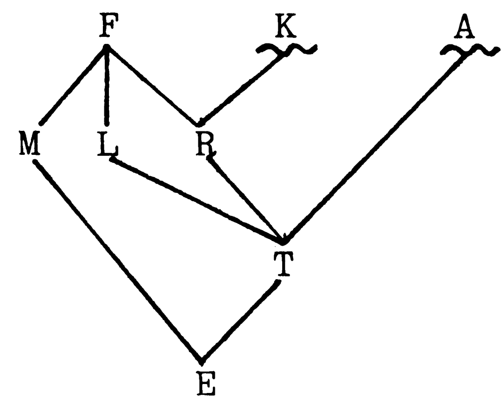
The identification of these factor sets and the rules of Section 5 for combining the factors into statistical results, using the structure diagram as a visual aid,

Table 1. Abbreviations and Identification of Factor Sets Necessary for ANOVA Calculations

Factor Set	Abbr	Identification
Live factors	LF	Factors that appear in the effect name
Dead factors	DF	Factors connected to live factors by upward links (factors nesting the live factors)
Symbolic factors	SF	The live and dead factors
Complement factors	CF	All factors other than SF
Random complement factors	RCF	CF's that are random (random = no squiggle)
Simple random complement factors	SRCF	RCF's not connected to other RCF's by upward links (RCF's not nested in other RCF's)

Table 2. Complete Listings of Factor Sets Necessary for ANOVA Calculations for Selected Effects From the Triple L Design

Structure Diagram



Examples of Factor Sets for Selected Effects

Effect	LF	DF	SF	CF	RCF	SRCF
K	K	None	K	F, A, M, L, R, T, E	F, M, L, R, T, E	F
R	R	F, K	R, F, K	A, M, L, T, E	M, L, T, E	M, L
T	T	F, K, A, L, R	T, F, K, A, L, R	M, E	M, E	M
KL	K, L	F	K, L, F	A, M, R, T, E	M, R, T, E	M, R
AR	A, R	F, K	A, R, F, K	M, L, T, E	M, L, T, E	M, L
AML	A, M, L	F	A, M, L, F	K, R, T, E	R, T, E	R
ALR	A, L, R	F, K	A, L, R, F, K	M, T, E	M, T, E	M, T

constitute the basic features of the structure diagram methodology. The manner of identifying the factor sets for a given effect by referring to the structure diagram is outlined in chart summary form in Table 1. The factor sets given in Table 1 are the only ones necessary for application in future results. Table 2 illustrates the method of identification from an actual structure using examples from the Triple L Design of Figure 3. The visual identification of these factor sets from the structure diagram is necessary for understanding and applying the results of Section 5. Careful note should be taken, therefore, to understand how the factor sets are obtained from viewing the structure diagram.

In addition to and associated with the factor sets named there are two fundamental numerical constants that appear repeatedly in the ANOVA results. The design set of all factors in any experimental design is uniquely partitioned for each effect into the two basic groupings of the symbolic set and the complement set. To each grouping is attributed an algebraic constant. With the symbolic set, for an effect Q, is associated a symbolic product, written df(Q), and traditionally referred to as the degrees of freedom for

This content downloaded from 200.17.213.205 on Mon, 31 Jul 2017 22:37:57 UTC
All use subject to http://about.jstor.org/terms

the effect Q. With the complement set, for an effect Q, is associated a complement product, written k(Q), and traditionally-referred to as the expected mean squared (EMS) coefficient for the effect Q. Both of these characteristic ANOVA quantities are formed by using the ranges of the levels of the factors in their respective sets. For brevity let the term *diminished range* of a factor refer to the number of levels of the factor minus one. The symbolic product df(Q) is then formed by taking the product of all live factor diminished ranges with all dead factor ranges for factors in the symbolic set. The complement product k(Q) is formed by taking the product of factor ranges for all factors in the complement set. These results are summarized and illustrated by example in Table 3. The use of symbolic and complement products will be discussed in the formula derivation rules of Section 5.

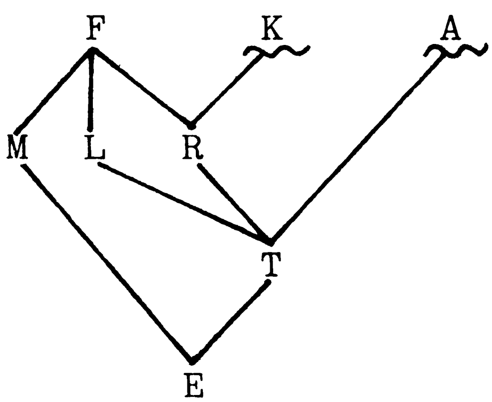
4. FORMAL INTERACTION

The concept of formal interaction is a key ingredient that allows the use of structure diagrams with factor sets to synthesize the visual rules for ANOVA formula and F-test derivations. Formal interaction, as discussed here, is dependent upon a concept with the same name in Collings's (1977) paper that has been suitably modified for application to the structure diagram.

Table 3. Symbols and Formation Rules for Symbolic and Complementary Products With Examples for Selected Effects From the Triple L Design

Product	ANOVA Name	Symbol	Formation Rule
Symbolic product	Degrees of freedom	df(Q)	Product LF diminished ranges times DF ranges
Complementary product	EMS coefficient	k(Q)	Product CF ranges

Structure Diagram



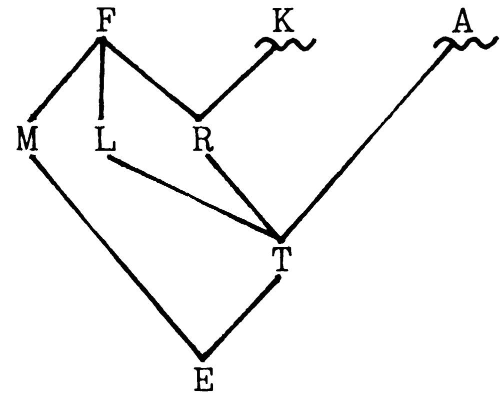
Examples

Effect	Symbolic Product	Complementary Product
K	*k* − 1	*famlrte*
R	(*r* − 1)*fk*	*amlte*
T	(*t* − 1)*fkalr*	*me*
KL	(*k* − 1)(*l* − 1)*f*	*amrte*
AR	(*a* − 1)(*r* − 1)*fk*	*mlte*
AML	(*a* − 1)(*m* − 1)(*l* − 1)*f*	*krte*
ALR	(*a* − 1)(*l* − 1)(*r* − 1)*fk*	*mte*

Table 4. Visual Rule for Formal Interaction With Selected Examples From the Triple L Design

[*Q*1, *Q*2, . . . , *Qn*] = juxtaposition of the factor letters in effects *Q*1, . . . , *Qn*, except for duplicates and any upper factor linked to a lower factor of *Q*1, . . . , *Qn*

Structure Diagram



Examples of Formal Interaction

Formal Interaction	Result
[*F*, *A*]	*FA*
[*FK*, *L*]	*KL*
[*FK*, *R*]	*R*
[*KAM*, *FA*, *MR*]	*AMR*
[*FA*, *KL*]	*KAL*
[*KL*, *AR*, *E*]	*E*
[*L*, *T*]	*T*
[*KAM*, *T*]	*MT*

Defining $Q1, Q2, \ldots, Qn$ to be effect names, the formal interaction among these effects is indicated by square brackets $[Q1, Q2, \ldots, Qn]$ and defined in the following way:

$[Q1, Q2, \ldots, Qn]$ = juxtaposition of all factor letters in effects $Q1, \ldots, Qn$, except for duplicates and any factor nesting another factor of $Q1, \ldots, Qn$.

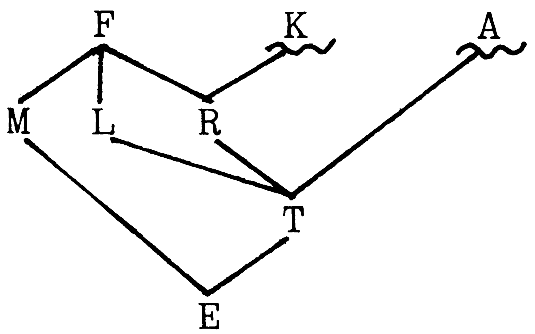
The formal interaction among any number n of effects is itself also an effect. The equivalent definition for visual use with the structure diagram is given in Table 4. The rule, concisely stated, says to discard duplicates and nesting factors within the set $Q1, \ldots, Qn$. The examples of formal interaction of Table 4 may be verified using this rule. In each example of the table, the duplicates and nesting factors within the sets have been omitted. With a small amount of practice, interaction of several sets of factors may be accomplished rapidly and simply while viewing the diagram. Several results of the following section are dependent on the use of formal interaction.

This content downloaded from 200.17.213.205 on Mon, 31 Jul 2017 22:37:57 UTC
All use subject to http://about.jstor.org/terms

Table 5. A Complete Listing of All Effects for the Triple L Design With Sample Calculations of Degrees of Freedom and Sums of Squares for Three Selected Effects

0-way	*1-way*	*2-way*	*3-way*	*4-way*
u	*F L*	*FK KL ML*	*FKA AML*	*KAML*
	K R	*FA AM MR*	*KAM AMR*	*AMLR*
	A T	*KA AL MT*	*KAL ALR*	
	M E	*KM AR LR*	*KML MLR*	

Structure Diagram



Calculation of Sums of Squares for Selected Effects Q

Example 1: $Q = F$

$$\mathrm{df}(F) = \Sigma t = f - 1$$

$$\mathrm{SS}(F) = \frac{1}{\mathrm{k}(f)} \Sigma Y_{f.}^2 - \frac{1}{\mathrm{k}(1)} Y_{.}^2,$$

where

$$\mathrm{k}(f) = kamlrte \quad \text{and} \quad \mathrm{k}(1) = fkamlrte$$

Example 2: $Q = R$

$$\mathrm{df}(R) = \Sigma t = (r - 1)kf = rkf - kf$$

$$\mathrm{SS}(R) = \frac{1}{\mathrm{k}(rkf)} \Sigma Y_{rkf.}^2 - \frac{1}{\mathrm{k}(kf)} \Sigma Y_{kf.}^2,$$

where

$$\mathrm{k}(rkf) = amlte \quad \text{and} \quad \mathrm{k}(kf) = amlrte$$

Example 3: $Q = ML$

$$\mathrm{df}(ML) = \Sigma t = (m - 1)(l - 1)f = mlf - lf - mf + f$$

$$\mathrm{SS}(ML) = \frac{1}{\mathrm{k}(mlf)} \Sigma Y_{mlf.}^2 - \frac{1}{\mathrm{k}(lf)} \Sigma Y_{lf.}^2 - \frac{1}{\mathrm{k}(mf)} \Sigma Y_{mf.}^2 + \frac{1}{\mathrm{k}(f)} \Sigma Y_{f.}^2$$

5. FORMULA DERIVATIONS

ANOVA formula derivation results are traditionally obtained by analysis of the subscripted terms in the linear model. These subscript analysis rules can be adapted to fit the context of the structure diagram when applied in conjunction with the concepts of factor sets and formal interaction. When this is done the traditional rules take on a convenient form suitable for deriving ANOVA results by visual inspection of the structure diagram. The validity of these visual rules is founded on the parallelism that exists with the subscript rules of the linear model representation as discussed in basic references such as those of Bennett and Franklin (1954), Schultz (1955), and Scheffé (1959).

From the verbal description of the experiment, the structure diagram is constructed, with each factor represented by a capital letter designation, and squiggly underline if it is a fixed factor. The allowable effects, which are equivalent to the linear model terms, are the main effects corresponding to the factors themselves, and the higher way effects corresponding to the interaction of factors that are mutually crossed. The mutual crossing of a set of factors in the diagram is manifested by the absence of upward links connecting one factor to another factor in the set. The rule for determination of allowable effects from the structure diagram may be summarized briefly.

Rule 1. Allowable Effects

Allowable Effects = All combinations of factor letters, where no factor in a combination is connected to another factor in the combination by upward links.

A listing of the effects, identical to that obtained from the linear model and written in an ANOVA table under "source," can be formed using Rule 1. Application of Rule 1 is facilitated by the method of successive inclusion of one factor at a time into the structure, while writing down allowable interactions with all previously included factors. An alternative approach is to write the effects in sequence according to the order of the interaction, starting with the mean and all main effects; and then proceeding to all two-way effects, all three-way effects, and so forth, until all terms have been written. The source terms that would appear in an ANOVA table for the Triple L design are listed in Table 5.

The degrees of freedom for any effect are found by calculating the symbolic product for that effect. From the results of Section 3, the following rule may be stated for helping to complete the ANOVA table.

Rule 2. Degrees of Freedom: df(Q)

df(Q) = Symbolic Product
= Product of live factor diminished ranges times dead factor ranges.

The degrees of freedom for any effect are obtained from the working structure diagram by noting which are the live and dead factors for the effect of interest, and then applying the rule. Examples of the application of Rule 2 are found in Table 3.

The sum of squares formula for an effect of interest Q, denoted SS(Q), is obtained by multiplicatively expanding the algebraic formula for df(Q) into a sum of terms, $\Sigma\, t$. Each term, t, consists of a set of letter subscripts along with its positive or negative sign, denoted sign(t), which results from the expansion. Letting the dot notation $Y_{t.}$ indicate a marginal total of observations over absent subscripts, the formula for SS(Q) is written out symbolically in Rule 3.

This content downloaded from 200.17.213.205 on Mon, 31 Jul 2017 22:37:57 UTC
All use subject to http://about.jstor.org/terms

Rule 3. Sum of Squares: SS(Q)

$$\text{df}(Q) = \Sigma t$$

$$\text{SS}(Q) = \sum_{(2)} \left[\frac{\text{sign}(t)}{\text{k}(t)} \sum_{(1)} Y_t^2. \right]$$

(1) indicates sum over all combinations of subscripts in t,

(2) indicates sum over all t, and

k(t) is the complement product of factors whose subscripts appear in t (the product of factor ranges for all factors whose subscripts do not appear in t).

The rule is perhaps most clearly illustrated by the examples of Table 5. For an effect of order n, where $n = 0$ for the mean, $n = 1$ for a main effect, $n = 2$ for a second order interaction, and so forth, the number of terms that appear in the symbolic product expansion is 2^n. This rule is evident from the tabular examples.

Expected mean squares are calculated by using the concepts of factor sets and formal interaction. The two-part rule stated here is an adaptation of the algorithm given by Collings (1977), modified for use with the structure diagram.

Rule 4. Expected Mean Squares: EMS(Q)

Part 1: Form $S_m = \{s \mid s$ is a formal interaction of LF with an m-way interaction of factors in RCF, where $m = 0, 1, 2, \ldots\}$.

Part 2: $\text{EMS}(Q) = \Sigma \text{k}(s)\, \sigma_s^2$, where σ_s^2 is a variance component for random factors and a mean squared deviation from treatment means for fixed factors; and k(s) is the complement product of the effect s.

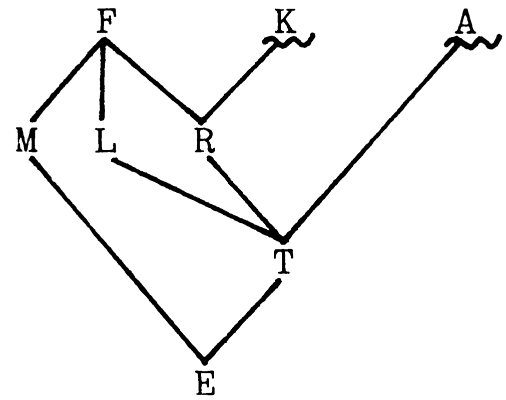
The rule for expected mean squares is illustrated in Table 6, where the abbreviation I/A is used for formal interaction. The interpretation of Rule 4 is that the EMS terms arise from the interactions of the random complement factors (RCF) with the live factor (LF) set.

Because of the way the rule is written, one can make note of the set RCF, and by mentally taking m-way products of RCF with LF off the structure diagram, immediately write down the set S_m that must appear in the expected mean square. It is thus unnecessary to write out the intermediate m-way results, and only the final answer need be written. This manner of intuitively (and rigorously) writing down the expected mean square terms can be accomplished rapidly and quite simply. It is performed by selecting the factor sets RCF and LF and, while viewing the diagram, mentally creating the interactions.

One of the most useful results in Collings's (1977) paper is an automatic method of determining correct numerator and denominator ratios for making exact or approximate F tests of hypotheses on components of variance of the form H_0: $\sigma_Q^2 = 0$. The rule is formulated in terms of the structure diagram.

Table 6. Sample Calculation of Expected Mean Squares for Two Selected Effects From the Triple L Design

Structure Diagram



Example 1 Q = A

RCF = F, M, L, R, T, E; LF = A

0-way RCF = 1	I/A with LF = A
1-way RCF = F, M, L, R, T, E	I/A with LF = FA, AM, AL, AR, T, E
2-way RCF = ML, MR, LR, MT	I/A with LF = AML, AMR, ALR, MT
3-way RCF = MLR	I/A with LF = $AMLR$
(no higher way exists)	

$S_m = \{A, FA, AM, AL, AR, T, E, AML, AMR, ALR, MT, AMLR\}$

Therefore, $\text{EMS}(A) = \Sigma \text{k}(s)\sigma_s^2$

Example 2: Q = KM

RCF = L, R, T, E; LF = K, M

0-way RCF = 1	I/A with LF = KM
1-way RCF = L, R, T, E	I/A with LF = KML, MR, MT, E
2-way RCF = LR	I/A with LF = MLR
(no higher way exists)	

$S_m = \{KM, KML, MR, MT, E, MLR\}$

Therefore, $\text{EMS}(KM) = \Sigma \text{k}(s)\sigma_s^2$

Rule 5. F Tests

Part 1: Form $S_e = \{e \mid e$ is a formal interaction of LF with an even-way $(0, 2, 4, \ldots)$ interaction of factors in SRCF$\}$.
Form $S_o = \{o \mid o$ is a formal interaction of LF with an odd-way $(1, 3, 5, \ldots)$ interaction of factors in SRCF$\}$.

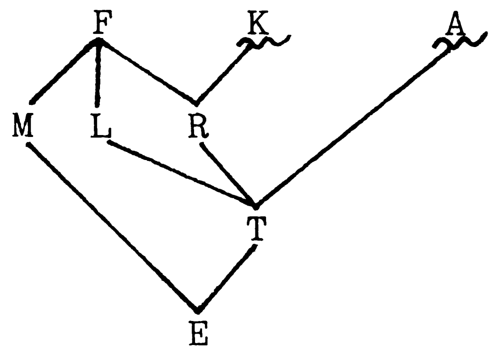
Part 2: $$F_Q = \frac{N_Q}{D_Q} = \frac{\Sigma \text{MS}(e)}{\Sigma \text{MS}(o)}.$$

The ratios formed for F_Q will always have an equal number of terms in the numerator N_Q and denominator D_Q. If there is only one factor in SRCF, there will be only one term each in N_Q and D_Q. If there are two or more, say m terms in SRCF, there will be 2^{m-1} terms each in N_Q and D_Q. In these cases there will be no exact F test and an approximate test of the Satterthwaite type may be used to test the hypothesis. Rule 5 then tells whether there is an exact F test available and generates the correct numerator and denomi-

This content downloaded from 200.17.213.205 on Mon, 31 Jul 2017 22:37:57 UTC
All use subject to http://about.jstor.org/terms

Table 7. Calculation of F Tests for Two Selected Effects From the Triple L Design

Structure Diagram



Example 1: $Q = A$

SRCF = F; LF = A

0-way SRCF = 1 — I/A with LF = A
1-way SRCF = F — I/A with LF = FA
(no higher way exists)

$$S_e = \{A\}$$

$$S_o = \{FA\}$$

$$F_A = \frac{N_A}{D_A} = \frac{\Sigma MS(e)}{\Sigma MS(o)} = \frac{MS(A)}{MS(FA)}$$

Example 2: $Q = KM$

SRCF = L, R; LF = M, K

0-way SRCF = 1 — I/A with LF = KM
1-way SRCF = L, R — I/A with LF = KML, MR
2-way SRCF = LR — I/A with LF = MLR
(no higher way exists)

$$S_e = \{KM, MLR\}$$

$$S_o = \{KML, MR\}$$

$$F_{KM} = \frac{N_{KM}}{D_{KM}} = \frac{\Sigma MS(e)}{\Sigma MS(o)} = \frac{MS(KM) + MS(MLR)}{MS(KML) + MS(MR)}$$

nator ratios for either the exact or approximate test. The F-test rule is illustrated by example in Table 7. As seen from these examples, Rule 5 shows that the F test can be viewed as the quotient of the even-way interactions of the simple random complement factors (SRCF) with the live factor set (LF), divided by the similar odd-way interactions.

One of the benefits of the F test is the provision of a rule to avoid the trial and error construction of two sums of expected mean squares for N_Q and D_Q, whose difference is $k(Q)\sigma_Q^2$, where $k(Q)$ is the complement product. Rule 5 automatically provides that

$$E(N_Q) - E(D_Q) = k(Q)\sigma_Q^2. \qquad (4.1)$$

Because of this property to which every F-test ratio conforms, Rule 6 for variance component estimation may be written:

Rule 6. Variance Component (σ_Q^2) Estimation

$$\hat{\sigma}_Q^2 = \frac{N_Q - D_Q}{k(Q)} = \frac{\Sigma MS(e) - \Sigma MS(o)}{k(Q)}$$

By use of Rules 1 through 6, all information and formulas needed for completing the traditional ANOVA table can now be obtained by visual reference to the structure diagram. Because these rules can be applied rapidly without reference to the linear model, they form an independent cross-checking technique to verify the correctness of formulas otherwise obtainable only from the subscripted model. When using the basic factor sets with the rules for producing ANOVA results, the designer of an experiment can quickly judge the impact of how the addition or deletion of a factor in its relationship to other factors can affect the design of the experiment. The basic capability of drawing the structural picture of a design, coupled with the structural analysis rules, is a potent concept that can deepen understanding of basic ANOVA techniques.

6. CONCLUSIONS

The visual representation of factors in the structure diagram gives an appealing graphical method of picturing experimental designs. This concise picture fully describes fixed and random factors with their names, number of levels, and nesting and crossing relationships. The diagram is a powerful descriptive technique, forming a valuable visual link connecting the verbal description of an experiment to the derivation of the correct linear model. The picture becomes increasingly useful in visualizing complex multifactor designs where students and workers might experience greatest difficulty.

The method can serve as a truly independent cross-check to verify formulas obtained from examination of the linear model, because the visual rules are performed independent of the subscripted model. The visual rules give a particularly clear understanding of how EMS and F-test terms originate in an experimental design from the structural relationship that factors bear relative to one another. The method also incorporates a useful procedure for automatic construction of F-test ratios without intermediate computation of expected mean squares.

A basic difference in perspective between the structure diagram method and the linear model approach is one of generation versus extraction of results. The structure diagram method generates results by using the mechanism of interaction on selected factor sets, while the linear model approach extracts results by searching the written effects of the subscripted linear model. The structure diagram therefore provides a valuable complementary perspective for viewing basic ANOVA design and analysis techniques.

[*Received June 1978. Revised September 1980.*]

This content downloaded from 200.17.213.205 on Mon, 31 Jul 2017 22:37:57 UTC
All use subject to http://about.jstor.org/terms

REFERENCES

BENNETT, CARL A., and FRANKLIN, NORMAN L. (1954), *Statistical Analysis in Chemistry and the Chemical Industry*, New York: John Wiley.

COLLINGS, BRUCE J. (1977), "Determining *F*-Ratios in a Balanced Design," Paper No. SD-016-R, Dept. of Statistics Report Series, Brigham Young University.

KEMPTHORNE, O., and FOLKS, L. (1971), *Probability, Statistics, and Data Analysis*, Ames, Iowa: Iowa State University Press.

KEMPTHORNE, O., ZYSKIND, G., ADDELMAN, S., THROCKMORTON, T.N., WHITE, R.F. (1961), "Analysis of Variance Procedures," *Aeronautical Research Laboratory Technical Report 149*, Office of Technical Services, United States Department of Commerce.

LEE, WAYNE (1966), "Experimental Design Symbolization and Model Derivation," *Psychometrika*, 31, 397–412.

——— (1975), *Experimental Design and Analysis*, San Francisco: W.H. Freeman.

SCHEFFÉ, HENRY (1959), *The Analysis of Variance*, New York: John Wiley.

SCHULTZ, E.F., JR. (1955), "Rules of Thumb for Determining Expectations of Mean Squares in Analysis of Variance," *Biometrics*, 11, 123–135.

THROCKMORTON, THOMAS N. (1961), "Structures of Classification Data," unpublished Ph.D. dissertation, Iowa State University, Dept. of Statistics.

ZYSKIND, G., KEMPTHORNE, O., WHITE, R.F., DAYHOFF, E.E., DOERFLER, T.E. (1964), "Research on Analysis of Variance and Related Topics," *Aerospace Research Laboratory Technical Report 64–193*, Office of Technical Services, United States Department of Commerce.

This content downloaded from 200.17.213.205 on Mon, 31 Jul 2017 22:37:57 UTC
All use subject to http://about.jstor.org/terms