

Federer

EXTENSIONS OF COMPUTER PROGRAMS FOR
ANALYSES OF VARIANCE
OF FACTORIAL EXPERIMENTS

S. R. Searle
Biometrics Unit, Cornell University

BU-209-M

January, 1966

ABSTRACT

Well established rules of thumb for setting up the analysis of variance table for any factorial and/or heierarchical (nested) design are summarized and adapted to the requirements of a computer program.

Biometrics Unit, Plant Breeding Department, Cornell University.

EXTENSIONS OF COMPUTER PROGRAMS FOR
ANALYSES OF VARIANCE
OF FACTORIAL EXPERIMENTS

S. R. Searle
Biometrics Unit, Cornell University

BU-209-M

January, 1966

Rules of thumb for establishing analysis of variance tables for factorial designs, nested designs and combinations thereof are clearly set out by Henderson (1959) in a somewhat obscure and relatively inaccessible publication. Millman and Glass (1965), relying heavily on Henderson (1959), present a similar set of rules, including a neat procedure for symbolizing uncorrected sums of squares. Application of these rules to computer programs is now given. When computer output consists of sums of squares for strictly factorial designs the rules lead very simply to establishing analysis of variance tables for any sort of factorial and/or hierarchical (nested) design.

The discussion is confined entirely to situations of balanced (equal subclass numbers) data, to the complete exclusion of unbalanced data and of partially balanced data such as Latin squares, lattices, balanced incomplete blocks and so on. Unfortunately, notation is sometimes confusing because the meaning of certain symbols differs according to context.

FACTORS AND LEVELS

The analysis of variance procedure is essentially one of apportioning variation found among data according to the various classifications (and their interactions) into which the data can be grouped. We choose to refer to these classifications as factors and to the individual classes of a classification as

Biometrics Unit, Plant Breeding Department, Cornell University.

levels of a factor. Factors will be called A, B, C, ..., the number of levels in them being n_a, n_b, n_c, \dots . If one factor is nested within another (students within a class, for example), the notation will be C:B for factor C within factor B, C:BA for C with AB subclasses and so on. A letter on the left of the colon represents the nested factor and those on the right of the colon represent the factors within which the nested factor is found. With nested factors, C for example, n_c is the number of levels of factor C within each of the factors in which it is nested. Factors that are not nested, namely those forming cross classifications, will be called crossed factors.

Within the smallest subclassification of any data there may be one observation or there may be more than one. In either case these observations can, as Millman and Glass (1965) point out, be referred to as replications within all other subclasses. Following Henderson (1959) we refer to these as the "within" factor, using the notation W:ABC..., the number of levels being n_w (i.e. n_w replicates), $n_w = 1$ or $n_w > 1$. The total number of observations is then

$$N = n_a n_b n_c \dots n_w .$$

RULES OF THUMB

Simple rules of thumb for writing down the analysis of variance table for any factorial and/or nested design can be grouped under a series of headings. What follows is based almost entirely on Henderson (1959), with brevity of expression uppermost in mind.

Lines in the analysis of variance table

- Rule 1. One line for each factor (crossed or nested).
One line for each interaction.

Interactions

Interactions are obtained symbolically as products of factors, both factorial and nested. All products of 2, 3, 4, factors are considered. For

the sake of generality all factorial factors are considered as being expressed with a colon to the right of the symbol; e.g. A:, B:, ... and so on.

Rule 2. Every interaction is of the form

$$ABC...:XYZ...$$

where ABC... is the product on the left of the colon of the factors being combined

and XYZ... is the product on the right of the colon of the factors so associated with A, B and C... .

Rule 3. Repeated letters on the right of the colon are replaced by one of their kind.

Rule 4. If any letter occurs on both sides of a colon that interaction does not exist.

Examples

<u>Factors</u>	<u>Interaction</u>	
A and B	AB	(Rule 2)
A and C:B	AC:B	(Rule 2)
A:B and C:B	AC:BB \equiv AC:B	(Rule 3)
A:B and B:DE	AB:BDE, non-existent	(Rule 4)

Note that by using W:ABC... as the symbolic form for replicates, Rule 4 results in there being no interactions involving W.

Degrees of freedom

Each line in the analysis of variance table is one or another of three forms:

<u>Lines</u>	<u>Examples</u>
Factorial factors:	A: and B:
Nested factors:	C:B and D:CB
Interactions:	AC:B and AD:CB

Any line can therefore be typified by the general expression given for an interaction in Rule 2, namely

$$ABC\dots:XYZ\dots$$

Rule 5. Degrees of freedom for the line denoted by

$$ABC:XYZ \text{ are } (n_a - 1)(n_b - 1)(n_c - 1)n_x n_y n_z .$$

The rule is simple. Degrees of freedom are the product of terms like $(n_a - 1)$ for every letter A on the left of the colon, and of terms like n_x for every letter X on the right of the colon.

Rule 6. The sum of all degrees of freedom is $N - 1$.

Sums of squares

The symbols that specify a line in the analysis of variance are used to establish the corresponding sum of squares. The basic elements are taken to be the uncorrected sums of squares, and the notation is as follows.

l	≡	CF	≡	correction factor for the grand mean			
a	≡	uncorrected sum of squares for factor A					
ab	≡	"	"	"	"	sub-classes AB	
abc	≡	"	"	"	"	"	ABC

We shall refer to this subsequently as the Henderson notation.

Rule 7. Sum of squares for the line denoted by

$$AB:XY \text{ is } (a-b)(b-1)xy = abxy - axy - bxy + xy .$$

Again the rule is simple. The symbolic form for the sum of squares is the product of terms like $(a - 1)$ for every letter A on the left of the colon, and of terms like x for every letter X on the right of the colon. After expanding the product of lower case letters interpretation is in terms of the Henderson notation for uncorrected sums of squares just described. Note that all these sums of squares are expressed essentially in terms of crossed factors. Even when a factor is nested, sums of squares are expressed in terms of uncorrected sums of squares calculated as if the nested factor were a crossed factor. For

example, the sum of squares for A:B (A within B) is $(a-1)b = ab-b$, where ab is the uncorrected sum of squares of the A-B subclasses.

Rule 8. The total of all sums of squares is $\Sigma x^2 - CF$, where Σx^2 represents the sum of squares of the individual observations, $wabc\dots$ in the above notation, and where CF is the correction factor.

Example. Table 1 shows a simple example of using Rules 1 through 8.

The uncorrected sums of squares denoted by lower case letters, a , ab , and so on, have been defined so far solely in words; for example, ab is the uncorrected sum of squares for the A-B subclasses. Henderson (1959) gives no formal, algebraic definition of these terms - and in some sense it is unnecessary so to do, since "everyone knows" what is meant by this: the uncorrected sum of squares for the A-B subclasses is the sum over all such subclasses of the square of each subclass total, the sum being divided by the number of observations in such a subclass (the same number in each). However, Millman and Glass (1965) give a neat procedure for formalizing this. It starts from an expression for the total of all the observations. We state the rule using as an example the uncorrected sum of squares ac in a situation where x_{hijk} is the observation in levels h , i , j and k of factors A, B, C and W respectively.

Rule 9. (a) Write down the total of all observations:

$$\begin{array}{cccc} n_a & n_b & n_c & n_w \\ \Sigma & \Sigma & \Sigma & \Sigma x_{hijk} \\ h=1 & i=1 & j=1 & k=1 \end{array} .$$

(b) Re-order the summation signs so that those pertaining to the letters in the symbolic form of the uncorrected sum of squares of interest (ac , in this case) come first, enclosing the remainder of the sum in parentheses:

$$\begin{array}{cccc} n_a & n_c & n_b & n_w \\ \Sigma & \Sigma & \left(\Sigma & \Sigma x_{hijk} \right) \\ h=1 & j=1 & i=1 & k=1 \end{array} .$$

(c) Square the parenthesis and divide by the product of the n 's therein. The result is the required sum of squares:

$$\text{e.g., } ac = \frac{n_a n_c}{n_b n_w} \left(\Sigma_{h=1} \Sigma_{j=1} \left(\Sigma_{i=1} \Sigma_{k=1} x_{hijk} \right)^2 \right) / n_b n_w .$$

Again, as a workable rule, this is patently very simple.

A convenient notation for sums of squares suggested by Millman and Glass (1965) is to use a, b, c, ... in place of n_a , n_b , n_c , ... for the number of levels of the factors. The sums of squares are then the expanded forms of the expressions used for degrees of freedom. Evidence of this notation is readily provided in Table 1.

Expected values of mean squares, E(m.s.)

Mean squares are sums of squares divided by degrees of freedom. Expected values of mean squares, to be denoted generally by E(m.s.), are obtained by the following rules.

(a) Completely random model

Rule 10. Denote variances by σ^2 with appropriate subscripts. There will be as many σ^2 terms as there are lines in the analysis of variance table, with corresponding subscripts.

Example. When there is an AC:B interaction, $\sigma_{ac:b}^2$ will be involved in the E(m.s.) expressions.

Note that when $n_w = 1$ there will be no W-line in the analysis of variance, although it may be appropriate to envisage σ_w^2 as existing.

Rule 11. Whenever a σ^2 appears in any E(m.s.) expression its coefficient is the product of all n's whose subscripts do not occur in the subscript of that σ^2 .

Example. The coefficient of $\sigma_{ac:b}^2$ is n_w when A, B, C:B and W:ABC are the factors.

This rule implies that the coefficient of $\sigma_{w:abc\dots}^2$ is always unity.

Rule 12. Each E(m.s.) expression contains only those σ^2 's (with coefficients) whose subscripts include all letters pertaining to the m.s.

Example. The E(m.s.) for the AC:B line is $n_w \sigma_{ac:b}^2 + \sigma_{w:abc}^2$.

According to this rule $\sigma^2_{w:abc\dots}$ occurs in every E(m.s.) expression.

The above examples of Rules 10-12 are part of a complete analysis shown in Table 2.

(b) Fixed effects model, or mixed model

Rule 13. Treat the model as completely random, except for deleting certain σ^2 terms from the E(m.s.) expressions. When a σ^2 contains on the left of the colon in its subscript a letter pertaining to a fixed effects factor, that σ^2 is deleted from all E(m.s.) expressions not containing the same letter in their description.

Example. If C:B is considered a fixed effects factor $\sigma^2_{c:b}$ will be deleted from the E(m.s.) for the B line but $\sigma^2_{c:b}$ remains in E(m.s.) of the C:B line.

Rule 14. The σ^2 term corresponding to a fixed effects factor will remain in the E(m.s.) expression pertaining to that factor, but is to be interpreted as a quadratic function of the unknown, fixed effects.

Example. If C:B is a fixed effects factor $\sigma^2_{c:b}$ in the C:B line is interpreted as a quadratic function of the fixed effects relating to C.

These rules are further exemplified in Table 2a.

APPLICATIONS TO COMPUTER PROGRAMS

Many computer programs exist for carrying out the calculations involved in analyzing factorial experiments. A widely used reference for designing such programs is Ralston and Wilf (1956) and accordingly one would expect that many of them have much in common with respect to the kind of output that they yield. What is said in the ensuing pages about the program available at Cornell, FANOV, is therefore likely to apply in large measure to many other factorial analysis of variance programs.

The library of statistics programs at the Cornell Computing Center is known as CUSTAT. Description of it is given elsewhere, Searle (1965). One of its programs is FANOV (see Searle, 1963) which can handle data from factorial experiments having up to 8 factors and a total of $\prod_{i=1}^8 n_i$ levels where $\prod_{i=1}^8 (n_i + 1) \leq 20,000$ for n_i being the number of levels of the i 'th factor. As with many programs of this nature there is no provision for missing observations other than the facility of including in the calculations the number of observations which were missing and for which guessed, estimated or "guesstimated" values have been added to the data; this number is then subtracted from the degrees of freedom for the "within" mean square.

Program FANOV yields as part of its output sums of squares denoted by the labels R, A, B, C, ... corresponding to main effects, and all associated interactions. Thus for factors R, A and B, the sums of squares are R, A, RA, B, RB, AB and RAB. These are not uncorrected sums of squares, but are the sums of squares pertaining to the analysis of variance table for, in this case, cross-classification (factorial) factors R, A and B. Just as Henderson's notation is in terms of crossed factors, so too is the notation of FANOV. The difference is that the Henderson notation is in terms of uncorrected sums of squares whereas the FANOV notation is in terms of analysis of variance sums of squares. But the relationship is simple, as witness the following table.

Sums of squares for 3 crossed
factors R, A and B

<u>Notation in FANOV program</u>	<u>Henderson notation</u>
R	(r-1)
A	(a-1)
RA	(r-1)(a-1)
B	(b-1)
RB	(r-1)(b-1)
AB	(a-1)(b-1)
RAB	(r-1)(a-1)(b-1)

The analogy is obvious: replace r, a and b in the Henderson notation by R+1, A+1 and B+1 to get the FANOV notation.

Rule 15. Convert Henderson's notation for sums of squares to FANOV notation by replacing lower case letters by their upper case counterparts with unity added; e.g. a becomes A+1.

Little insight to the value of this rule is given by the simple example just outlined. But in more complicated situations the analysis of variance tables can readily be established from Rules 1-14 in terms of Henderson's notation, and the calculations derived from a FANOV output by converting to the FANOV notation according to Rule 15.

Henderson's notation shows how uncorrected sums of squares have to be combined in order to derive the sums of squares for a particular analysis. Conversion to FANOV notation indicates the way in which sums of squares that are output from FANOV are to be combined to yield the same required sums of squares.

Example. The sum of squares for C:B in Henderson's notation is $(c-1)b = bc - b$. In FANOV notation it is $C(B+1) = BC + C$.

Additional examples are shown in Table 4.

Preparation of data for FANOV

Data preparation for FANOV is discussed in detail in Searle (1963). There are two important points to be remembered. The first is that the levels of factors must be coded as if all factors were crossed, even if some are nested. Thus cows within herd 1 are given cow numbers 1, 2, 3, ... as are cows within herd 2, even though they are not the same cows. This procedure applies for replicates too, since they are considered simply as a nested factor. The second point to remember is that although FANOV uses the letters R, A, B, C ... for factors, R does not necessarily nor automatically denote replicates, nor need it be used for this purpose. It can be used for any factor of the experiment at all, as can the other letters of FANOV, namely A, B, C, D, E, F or G. The process by which factors of the experiment are allocated to the factors of

FANOV is determined entirely by the order in which the codes for the factor levels are punched in the data cards. If, for an experiment of yearly yields within cows within herds, the year is punched first, followed by the cow code and the herd code, then in FANOV the factor R will represent years, A will represent cows and B will represent herds. Details concerning the mode of coding are described in Searle (1963).

One-degree-of-freedom terms

The analysis of variance for a randomized complete blocks (RCB) experiment of b blocks and k treatments is as follows.

Analysis of variance for RCB experiment

Term	d.f.	Sums of squares
Blocks	b-1	$\sum_{j=1}^b x_{.j}^2/k - x_{..}^2/bk$
Treatments	k-1	$\sum_{i=1}^k x_{i.}^2/b - x_{..}^2/bk$
Residual	(b-1)(k-1)	$\sum_{i=1}^k \sum_{j=1}^b x_{ij}^2 - \sum_{j=1}^b x_{.j}^2/k - \sum_{i=1}^k x_{i.}^2/b + x_{..}^2/bk$
Total	bk-1	$\sum_{i=1}^k \sum_{j=1}^b x_{ij}^2 - x_{..}^2/bk$

Customary summation notation is used in this table: x_{ij} is the observation on the i'th treatment in the j'th block, $x_{i.} = \sum_{j=1}^b x_{ij}$ and so on.

Now suppose there are only three treatments, t_1 , t_2 and t_3 . Then the sum of squares for treatments has 2 degrees of freedom and it can be divided into two 1-degree-of-freedom terms:

<u>1-d.f.-terms</u>	<u>Sums of squares</u>
$t_1 \text{ v. } t_2$	$(x_{1.} - x_{2.})^2/2b$
$t_3 \text{ v. } \frac{1}{2}(t_1 + t_2)$	$(x_{1.} + x_{2.} - 2x_{3.})^2/6b$
<u>Total (2 d.f.)</u>	<u>$(x_{1.}^2 + x_{2.}^2 + x_{3.}^2)/b - x_{..}^2/3b$</u>

Such one-degree-of-freedom terms cannot be obtained directly from FANOV although, of course, the program does yield as output the necessary totals $x_{i.}$ that enable one to readily calculate the sums of squares from the expressions given above. In some situations, however, the volume of data to be analyzed may be such that those calculations become quite tedious. It is then possible, at least in some cases, to do a second analysis of the data that yields the required one-degree-of-freedom terms, using a special coding system for the treatments designed for this purpose.

Consider first, a case of 4 treatments t_1, t_2, t_3 and t_4 . If the comparisons of interest are

$$\begin{aligned} c_1: & t_1 + t_2 - t_3 - t_4 \\ c_2: & t_1 - t_2 + t_3 + t_4 \\ c_3: & t_1 - t_2 - t_3 + t_4, \end{aligned}$$

the four treatments instead of being coded 1, 2, 3 and 4 can be re-coded in terms of two new factors A and B say, each having two levels:

<u>Treatment</u>	<u>New code</u>	
	<u>Factor A</u>	<u>Factor B</u>
1	1	1
2	1	2
3	2	1
4	2	2

Using FANOV to analyze the data as a $b \times 2 \times 2$ design the resulting sums of squares for A, B and AB will be the sums of squares for comparisons c_1 , c_2 and c_3 respectively. Clearly this procedure can be extended to any case where the number of levels of a factor is 2^n for integer n. It yields, of course, only comparisons that are similar to those inherent in the analysis of a 2^n factorial.

To use a scheme as just suggested it is advisable to analyze the data twice, once in the regular manner and once using the 2^n coding system. A check on the computations (and hence on the coding also) is then available. Note, though, that this procedure demands two distinctly different coding systems for the treatments, one as 1, 2, 3 and 4 and one as 11, 12, 21 and 22. It is advisable to put both coding systems in the data cards, duplicate them, and run two separate analyses through FANOV. The cards must be duplicated, for data cannot be read into FANOV under the control of two separate codings. Actually, it is possible to run any analysis through FANOV without any identification codes in the cards - relying solely on having the cards in the sequence demanded by FANOV for whatever coding system is desired. This practice is not to be recommended.

Extensions of this 2^n coding system can be made to cases where there are not 2^n levels of the factor concerned. Consider the case of 3 treatments again, where the one-degree-of-freedom comparisons are t_1 v. t_2 and t_3 v. $\frac{1}{2}(t_1+t_2)$. Suppose we add duplicates of the data on treatment 3 to the data we already have, call them treatment 4 and code the four treatments by the 2^n system. When these data are analyzed by FANOV the sums of squares will be

$$\begin{aligned} A: (x_{1.} + x_{2.} - x_{3.} - x_{4.})^2/4b &= (x_{1.} + x_{2.} - 2x_{3.})^2/4b \\ B: (x_{1.} - x_{2.} + x_{3.} - x_{4.})^2/4b &= (x_{1.} - x_{2.})^2/4b \\ AB: (x_{1.} - x_{2.} - x_{3.} + x_{4.})^2/4b &= (x_{1.} - x_{2.})^2/4b . \end{aligned}$$

Hence two-thirds of the A sum of squares is that for the t_3 v. $\frac{1}{2}(t_1+t_2)$ comparison and twice that for B (or AB) is the sum of squares for the t_1 v. t_2 comparison.

Further work

It would seem worthwhile to investigate further coding systems for extending the use of FANOV (or similar computer programs). It seems reasonable to believe that other adaptations of the 2^n and other systems can be found for deriving the sums of squares for one-degree-of-freedom comparisons between different levels of main effects and of interactions, both when there are 2^n levels and otherwise.

ACKNOWLEDGEMENT

Thanks go to Professor W. T. Federer for suggesting the 2^n coding system when there are 2^n levels of a factor.

References

- Henderson, C. R. (1959). Design and analysis of animal husbandry experiments. Chapter 1 of "Techniques and Procedures in Animal Production Research", American Society of Animal Production Monograph.
- Millman, J., and Glass, G. V. (1965). Rules of thumb for writing the anova table. Mimeograph, Cornell University.
- Ralston, Anthony, and Wilf, Herbert S. (1956). Editors, Mathematical Methods for Digital Computers, J. Wiley and Sons, New York.
- Searle, S. R. (1963). Program documentation: factorial analyses of variance. Mimeo, Cornell Computing Center.
- Searle, S. R. (1965). CUSTAT, the Cornell University Computer programs in Statistics. Paper No. BU-206-M in the Mimeograph Series of the Biometrics Unit, Cornell University.

Analysis of variance example for factors A, B, C:B and W:ABC

Table 1. Analysis of Variance based on Rules 1-8

Line (Rules 1-4)	Degrees of freedom (Rule 5)	Sums of squares (Rule 7)
A	$n_a - 1$	$(a-1) = a-1$
B	$n_b - 1$	$(b-1) = b-1$
C:B	$(n_c - 1)n_b$	$(c-1)b = bc-b$
AB	$(n_a - 1)(n_b - 1)$	$(a-1)(b-1) = ab-a-b+1$
AC:B	$(n_a - 1)(n_c - 1)n_b$	$(a-1)(c-1)b = abc-ab-bc+b$
W:ABC	$(n_w - 1)n_a n_b n_c$	$(w-1)abc = wabc-abc$
Total	$N-1$ (Rule 6)	$\Sigma x^2 - CF \equiv wabc-1$ (Rule 8)

- Notes. 1. Interaction BC:B does not exist (Rule 4).
 2. Total degrees of freedom = $N-1 = n_w n_a n_b n_c - 1$ (Rule 6).

Table 2. Expected Values of Mean Squares: Random Model (Rules 10-12)

Line	Variances (Rule 10) and Coefficients (Rule 11)					
	$n_b n_c n_w \sigma_a^2$	$n_a n_c n_w \sigma_b^2$	$n_a n_w \sigma_{c:b}^2$	$n_c n_w \sigma_{ab}^2$	$n_w \sigma_{ac:b}^2$	$\sigma_{w:abc}^2$
	<u>Terms included (Rule 12)</u>					
A	*			*	*	*
B		*	*	*	*	*
C:B			*		*	*
AB				*	*	*
AC:B					*	*
W:ABC						*

Table 2a. Expected Values of Mean Squares: Mixed Model (Rules 13 and 14)
(Factor C:B considered fixed effects)

Line	$n_b n_c n_w \sigma_a^2$	$n_a n_c n_w \sigma_b^2$	$n_a n_w \sigma_{c:b}^2$	$n_c n_w \sigma_{ab}^2$	$n_w \sigma_{ac:b}^2$	$\sigma_{w:abc}^2$
	<u>Changes to Table 2 (see footnotes)</u>					
A	*			*	[*]	*
B		*	[*]	*	[*]	*
C:B			(*)		*	*
AB				*	[*]	*
AC:B					*	*
W:ABC						*

[*] denotes a σ^2 that is deleted (Rule 13).

(*) denotes a σ^2 interpreted as a quadratic function (Rule 14).

Table 3. Calculating Sums of Squares from FANOV

Sum of Squares	Henderson's Notation (Table 1)	FANOV notation (Rule 15)
A	(a-1)	A
B	(b-1)	B
C:B	(c-1)b	C(B+1) = BC+C
AB	(a-1)(b-1)	AB
AC:B	(a-1)(c-1)b	AC(B+1) = ABC+AC
W:ABC	(w-1)abc	W(A+1)(B+1)(C+1) = WABC+WAB+WAC+WBC +WA+WB+WC+W

Summary of Rules

Notation:	<u>Examples</u>
Crossed factors:	A, B, C, ...
Nested factors:	D:A, E:BC, ...
Replicates "within":	W:ABCD...
Levels:	$n_a, n_b, n_c, \dots, n_w$
No. of observations:	$N = n_a n_b n_c \dots n_w$
Uncorrected sums of squares:	a, b, ab
Correction factor:	$1 \equiv CF$
Total sum of squares:	Σx^2

Lines in the analysis of variance table

- Rule 1. One line for each factor (crossed or nested).
One line for each interaction.

Interactions

- Rule 2. Every interaction is of the form

$$ABC\dots:XYZ\dots$$

where ABC... is the product on the left of the colon of the factors being combined and XYZ... is the product on the right of the colon of the factors so associated with A, B and C

- Rule 3. Repeated letters on the right of the colon are replaced by one of their kind.

- Rule 4. If any letter occurs on both sides of a colon that interaction does not exist.

Degrees of freedom

- Rule 5. Degrees of freedom for the line denoted by

$$ABC:XYZ \text{ are } (n_a - 1)(n_b - 1)(n_c - 1)n_x n_y n_z.$$

Rule 6. The sum of all degrees of freedom is $N-1$.

Sums of squares

Rule 7. The sum of squares for the line denoted by
AB:XY is $(a-1)(b-1)xy = abxy - axy - bxy + xy$.

Rule 8. The total of all sums of squares is $\Sigma x^2 - CF$.

Rule 9. (a) Write down the total of all observations using summation signs.

(b) Re-order the summation signs so that those pertaining to the letters in the symbolic form of the uncorrected sum of squares of interest come first, enclosing the remainder in parenthesis.

(c) Square the parenthesis and divide by the product of the n 's therein. The result is the required uncorrected sum of squares.

Expected values of mean squares E(m.s.)

(a) Completely random model

Rule 10. Denote variances by σ^2 with appropriate subscripts. There will be as many σ^2 terms as there are lines in the analysis of variance table, with corresponding subscripts.

Rule 11. Whenever a σ^2 appears in any E(m.s.) expression its coefficient is the product of all n 's whose subscripts do not occur in the subscript of that σ^2 .

Rule 12. Each E(m.s.) expression contains only those σ^2 's (with coefficients) whose subscripts include all letters pertaining to the m.s.

(b) Fixed effects model, or mixed model

Rule 13. Treat the model as completely random except for deleting certain σ^2 terms from the E(m.s.) expressions. When a σ^2 contains on the left of the colon in its subscript a letter pertaining to a fixed effects factor, that σ^2 is deleted from all E(m.s.) expressions not containing the same letter in their description.

Rule 14. The σ^2 term corresponding to a fixed effects factor will remain in the E(m.s.) expression pertaining to that factor, but is to be interpreted as a quadratic function of the unknown fixed effects.

Computer output (FANOV)

Rule 15. Convert Henderson's notation for sums of squares to FANOV notation by replacing lower case letters by their upper case counterparts with unity added: a becomes A+1.