

Topic 10

Supplement: Rules for determining Expected Mean Squares (EMS)

To perform an ANOVA, one must first determine the sum of squares for each component in the model as well as the number of degrees of freedom associated with each sum of squares. Then, to construct appropriate test statistics (see box below), the expected mean squares must be determined. In complex design situations, particularly those involving random or mixed models, it is frequently helpful to have a formal procedure for this process.

The appropriate test statistic (F) is a ratio of mean squares that is chosen such that the expected value of the *numerator* differs from the expected value of the *denominator* only by the specific variance component or fixed factor being tested.

The following set of rules is appropriate for the manual calculation of the expected mean squares for any **balanced factorial, nested, or nested factorial experiment**. (Note that partially balanced arrangements, such as Latin squares and incomplete block designs, are specifically excluded.) To illustrate the application of these rules, a two-factor factorial model is used.

RULE 1 In addition to the error term, the model contains all the main effects and any interactions that the experimenter assumes exist. In other words, the model is the linear model; it contains all the effects that would appear in the `lm()` statement in R.

Furthermore, the error term in the model, $\varepsilon_{ij\dots m}$, is written as $\varepsilon_{(ij\dots)m}$, where the subscript m denotes the replication subscript. For the two-factor model, this rule implies that ε_{ijk} becomes $\varepsilon_{(ij)k}$. Subsamples, if they appear in the model, are similarly "nested" within their higher-level classification variables.

RULE 2 For each term in the model, the subscripts fall into three classes:

Main: those subscripts present in the term and not in parentheses

Determinant: those subscripts present in the term and are in parentheses (usually higher-level classification subscripts for replications and subsamples/nested factors)

Absent: those subscripts present in the model but not in that particular term

For example, in $(\alpha\beta)_{ij}$, i and j are main subscripts and k is absent; and in $\varepsilon_{(ij)k}$, k is a main subscript and both i and j are determinant.

Note that the number of **degrees of freedom** for any term in the model is the product of the number of levels associated with each determinant subscript and the number of levels minus 1 associated with each main subscript.

For example: $df (\alpha\beta)_{ij} = (a-1)(b-1)$; $df \varepsilon_{(ij)k} = ab(r-1)$

RULE 3 Each effect has either a variance component (random effect) or a fixed factor effect associated with it. If an interaction contains at least one random effect, the entire interaction is considered to be a random effect. For example, in a two-factor mixed model with factor A fixed and factor B random, the variance component for B is σ_B^2 and the variance component for AB is $\sigma_{\alpha\beta}^2$. A fixed effect is always represented by the sum of squares of the model components associated with that factor, divided by its degrees of freedom. In our example, the fixed effect of A is:

$$\frac{1}{a-1} \sum_{i=1}^a \alpha_i^2$$

THE METHOD To obtain the **expected mean squares**, prepare a table with a row for each model component and a column for each subscript. Over each subscript, write the number of levels of the factor associated with that subscript and whether the factor is fixed (F) or random (R). Blocks, replicates, and subsamples are always considered to be random.

(a) In each row, write **1** if a determinant subscript matches the subscript in the column:

Fixed or Random	F	R	R
Number of levels	a	b	r
Subscript	i	j	k
α_i			
β_j			
$(\alpha\beta)_{ij}$			
$\varepsilon_{(ij)k}$	1	1	

(b) In each row, if any of the subscripts match the subscript in the column, write **0** if the column is a fixed factor and **1** if the column is a random factor. For interactions involving at least one random factor, write 1 in the columns that match the row subscript, independent of the nature (fixed or random) of that column:

Fixed or Random	F	R	R
Number of levels	a	b	r
Subscript	i	j	k
α_i	0		
β_j		1	
$(\alpha\beta)_{ij}$	1	1	
$\varepsilon_{(ij)k}$	1	1	1

Note the inclusion of a **1** in the column for the **interaction** term $(\alpha\beta)_{ij}$. This will **determine the presence of $r\sigma_{\alpha\beta}^2$ in the EMS** of the random Factor B.

(c) In the remaining empty cells, write the number of levels shown above the column heading:

Fixed or Random	F	R	R
Number of levels	a	b	r
Subscript	i	j	k
α_i	0	b	r
β_j	a	1	r
$(\alpha\beta)_{ij}$	1	1	r
$\varepsilon_{(ij)k}$	1	1	1

(d) To obtain the EMS for any model component (i.e. row), cover all columns headed by main subscripts associated with that component. Then in each row that contains at least the same subscripts as those of the component being considered, take the product of the visible numbers and multiply by the appropriate fixed or random factor from Rule 5. The sum of these quantities is the EMS of the model component being considered.

To find EMS_A , for example, cover column i (i is the index associated with factor A). The product of the visible numbers in the rows that contain at least subscript i are br (row 1), r (row 3), and 1 (row 4). Note that i is missing in row 2.

The expected mean squares derived using these rules for the two-way mixed model are:

Fixed or Random	F	R	R	Expected Mean Squares
Number of levels	a	b	r	
Subscript	i	j	k	
α_i	0	b	r	$\sigma_\varepsilon^2 + r\sigma_{\alpha\beta}^2 + br\Sigma\alpha^2/(a-1)$
β_j	a	1	r	$\sigma_\varepsilon^2 + r\sigma_{\alpha\beta}^2 + ar\sigma_\beta^2$
$(\alpha\beta)_{ij}$	1	1	r	$\sigma_\varepsilon^2 + r\sigma_{\alpha\beta}^2$
$\varepsilon_{(ij)k}$	1	1	1	σ_ε^2