10.2 SINGLE-FACTOR MODELS

In the case of a single-factor experiment the factor may be referred to as a *treatment effect*, as in Chapter 3; and if the design is completely randomized, the model is

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij} \tag{10.1}$$

Whether the treatment levels are fixed or random, it is assumed in this model that μ is a fixed constant and the errors are normally and independently distributed with a zero mean and the same variance, that is, ε_{ij} are NID $(0, \sigma_\varepsilon^2)$. The decision as to whether the levels of the treatments are fixed or random will affect the assumptions about the treatment term τ_j. The different assumptions and other differences will be compared in parallel columns.

Fixed Model

1. Assumptions: τ_j's are fixed constants.

$$\sum_{j=1}^{k} \tau_j = \sum_{j=1}^{k} (\mu_{.j} - \mu) = 0$$

(These add to zero as they are the only treatment means being considered.)

Random Model

1. Assumptions: τ_j's are random variables and are

$$\text{NID } (0, \sigma_\tau^2)$$

(Here σ_τ^2 represents the variance among the τ_j's or among the true treatment means $\mu_{.j}$. The τ_j average to zero when averaged over all possible levels, but for the k levels of the experiment they usually will not average 0.)

Figure 10.1(a) shows three fixed means whose average is μ as these are the only means of concern and $\sum_j \tau_j = \sum_j (\mu_j - \mu) = 0$. Figure 10.1(b) shows three random means whose average is obviously not μ as these are but three means chosen at random from many possible means. These means and their corresponding τ_j's are assumed to form a normal distribution with a standard deviation of σ_τ.

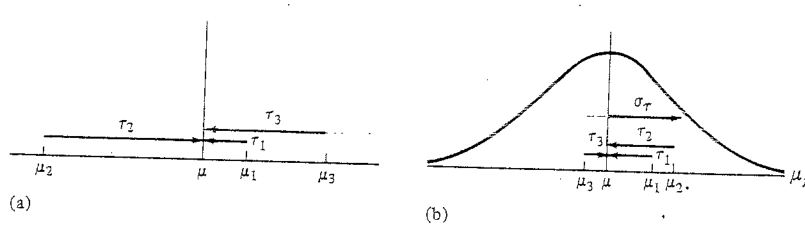

Figure 10.1 Assumed means in (a) fixed and (b) random models.

Fixed Model

2. Analysis: Procedures as given in Chapter 3 for computing SS.

Random Model

2. Analysis: Same as for fixed model.

3. EMS:

Source	df	EMS
τ_j	$k - 1$	$\sigma_\varepsilon^2 + n\phi_\tau$
ε_{ij}	$k(n - 1)$	σ_ε^2

4. Hypothesis tested:

$$H_0: \tau_j = 0 \quad \text{(for all } j\text{)}$$

3. EMS:

Source	df	EMS
τ_j	$k - 1$	$\sigma_\varepsilon^2 + n\sigma_\tau^2$
ε_{ij}	$k(n - 1)$	σ_ε^2

4. Hypothesis tested:

$$H_0: \sigma_\tau^2 = 0$$

The expected mean square (EMS) column turns out to be extremely important in more complex experiments as an aid in deciding how to set up an F test for significance. The EMS for any term in the model is the long-range average of the calculated mean square when the Y_{ij} from the model is substituted in algebraic form into the mean square computation. The derivation of these EMS values is often complicated, but those for the single-factor model were derived in Chapter 3 and those for two-factor models are derived in a later section of this chapter.

For the fixed model, if the hypothesis is true that $\tau_j = 0$ for all j, that is, all the k fixed treatment means are equal, then $\sum_j \tau_j^2 = 0$ and the EMS for τ_j and ε_{ij} are both σ_ε^2. Hence the observed mean squares for treatments and error mean square are both estimates of the error variance, and they can be compared by means of an F test. If this F test shows a significantly high value, it must mean that $n \sum_j \tau_j^2/(k - 1) = n\phi_\tau$ is not zero and the hypothesis is to be rejected.

For the random model, if the hypothesis is true that $\sigma_\tau^2 = 0$, that is, the variance among all treatment means is zero, then again each mean square is an estimate of the error variance. Again, an F test between the two mean squares is appropriate.

From the two tables in step 3 above, it is seen that for a single-factor experiment there is no difference in the test to be made after the analysis, and the only difference is in the generality of the conclusions. If H_0 is rejected, there is probably a difference between the k fixed treatment means for the fixed model; for the random model there is a difference between all treatments of which the k examined are but a random sample.

10.3 TWO-FACTOR MODELS

For two factors A and B the model in the general case is

$$Y_{ijk} = \mu + A_i + B_j + AB_{ij} + \varepsilon_{k(ij)}$$

with

$$i = 1, 2, \ldots, a \qquad j = 1, 2, \ldots, b \qquad k = 1, 2, \ldots, n$$

provided the design is completely randomized. In this model, it is again assumed that μ is a fixed constant and $\varepsilon_{k(ij)}$'s are NID $(0, \sigma_\varepsilon^2)$. If both A and B are at fixed levels, the model is a fixed model. If both are at random levels, the model is a random model, and if one is at fixed levels and the other at random levels, the model is a mixed model. Comparing each of these models gives:

Fixed

1. Assumptions:
   A_i's are fixed constants and
   $$\sum_{i=1}^{a} A_i = 0$$
   B_j's are fixed constants and
   $$\sum_{j=1}^{b} B_j = 0$$
   AB_{ij}'s are fixed constants and
   $$\sum_{i} AB_{ij} = 0$$
   $$\sum_{j} AB_{ij} = 0$$
2. Analysis: Procedures of Chapter 5 for sums of squares
3. EMS:

Random

1. Assumptions:
   A_i's are NID $(0, \sigma_A^2)$
   B_j's are NID $(0, \sigma_B^2)$
   AB_{ij}'s are NID $(0, \sigma_{AB}^2)$
2. Analysis: Same
3. EMS:

Mixed

1. Assumptions:
   A_i's are fixed and
   $$\sum_{i}^{a} A_i = 0$$
   B_j's are NID $(0, \sigma_B^2)$
   AB_{ij}'s are NID $(0, \sigma_{AB}^2)$ but
   $$\sum_{i}^{a} AB_{ij} = 0$$
   $$\sum_{j}^{b} AB_{ij} \neq 0$$
   (for A fixed, B random)
2. Analysis: Same
3. EMS:

Source	df	EMS (Fixed)	EMS (Random)	EMS (Mixed)
A_i	$a-1$	$\sigma_\varepsilon^2 + nb\phi_A$	$\sigma_\varepsilon^2 + n\sigma_{AB}^2 + nb\sigma_A^2$	$\sigma_\varepsilon^2 + n\sigma_{AB}^2 + nb\phi_A$
B_j	$b-1$	$\sigma_\varepsilon^2 + na\phi_B$	$\sigma_\varepsilon^2 + n\sigma_{AB}^2 + na\sigma_B^2$	$\sigma_\varepsilon^2 + na\sigma_B^2$
AB_{ij}	$(a-1)(b-1)$	$\sigma_\varepsilon^2 + n\phi_{AB}$	$\sigma_\varepsilon^2 + n\sigma_{AB}^2$	$\sigma_\varepsilon^2 + n\sigma_{AB}^2$
$\varepsilon_{k(ij)}$	$ab(n-1)$	σ_ε^2	σ_ε^2	σ_ε^2

4. Hypotheses tested:

H_1: $A_i = 0$ for all i
H_2: $B_j = 0$ for all j
H_3: $AB_{ij} = 0$ for all i and j

4. Hypotheses tested:

H_1: $\sigma_A^2 = 0$
H_2: $\sigma_B^2 = 0$
H_3: $\sigma_{AB}^2 = 0$

4. Hypotheses tested:

H_1: $A_i = 0$ for all i
H_2: $\sigma_B^2 = 0$
H_3: $\sigma_{AB}^2 = 0$

In the assumptions for the mixed model the fact that summing the interaction term over the fixed factor ($\sum_i$) is zero but summing it over the random factor ($\sum_j$) is not zero affects the expected mean squares, as seen in item 3 on page 213.

For the fixed model the mean squares for A, B, and AB are each compared with the error mean square to test the respective hypotheses, as should be clear from an examination of the EMS column when the hypotheses are true. For the random model the third hypothesis of no interaction is tested by comparing the mean square for interaction with the mean square for error, but the first and second hypotheses are each tested by comparing the mean square for the main effect (A_i or B_j) with the mean square for the interaction as seen by their expected mean square values. For a mixed model the interaction hypothesis is tested by comparing the interaction mean square with the error mean square. The random effect B_j is also tested by comparing its mean square with the error mean square. The fixed effect (A_i), however, is tested by comparing its mean square with the interaction mean square.

From these observations on a two-factor experiment, the importance of the EMS column is evident, as this column can be used to see how the tests of hypotheses should be run. It is also important to note that these EMS expressions can be determined prior to the running of the experiment. This will indicate whether or not a good test of a hypothesis exists. In some cases the proper test indicated by the EMS column will have insufficient degrees of freedom to be sufficiently sensitive, in which case the investigator might wish to change the experiment. This would involve such changes as a choice of more levels of some factors, or changing from random to fixed levels of some factors.

10.4 EMS RULES

The two examples above have shown the importance of the EMS column in determining what tests of significance are to be run after the analysis is completed. Because of the importance of this EMS column in these and more complex models, it is often useful to have some simple method of determining these values from the model for the given experiment. A set of rules can be stated that will determine the EMS column very rapidly, without recourse to their derivation. The rules will be illustrated on the two-factor mixed model of Section 10.3. To determine the EMS column for any model:

1. Write the variable terms in the model as row headings in a two-way table.

A_i
B_j
AB_{ij}
$\varepsilon_{k(ij)}$

2. Write the subscripts in the model as column headings; over each subscript write F if the factor levels are fixed, R if random. Also write the number of observations each subscript is to cover.

	a	b	n
	F	R	R
	i	j	k
A_i			
B_j			
AB_{ij}			
$\varepsilon_{k(ij)}$			

3. For each row (each term in the model) copy the number of observations under each subscript, providing the subscript does not appear in the row heading.

	a	b	n
	F	R	R
	i	j	k
A_i		b	n
B_j	a		n
AB_{ij}			n
$\varepsilon_{k(ij)}$			

4. For any bracketed subscripts in the model, place a 1 under those subscripts that are inside the brackets.

	a	b	n
	F	R	R
	i	j	k
A_i		b	n
B_j	a		n
AB_{ij}			n
$\varepsilon_{k(ij)}$	1	1	

5. Fill the remaining cells with a 0 or a 1, depending upon whether the subscript represents a fixed F or a random R factor.

	a	b	n
	F	R	R
	i	j	k
A_i	0	b	n
B_j	a	1	n
AB_{ij}	0	1	n
$\varepsilon_{k(ij)}$	1	1	1

6. To find the expected mean square for any term in the model:
   a. Cover the entries in the column (or columns) that contain non-bracketed subscript letters in this term in the model (for example, for A_i, cover column i; for $\varepsilon_{k(ij)}$, cover column k).
   b. Multiply the remaining numbers in each row. Each of these products is the coefficient for its corresponding term in the model, provided the subscript on the term is also a subscript on the term whose expected mean square is being determined. The sum of these coefficients multiplied by the variance of their corresponding terms (ϕ_τ or σ_τ^2) is the EMS of the term being considered (for example, for A_i, cover column i). The products of the remaining coefficients are bn, n, n, and 1, but the first n is not used, as there is no i in its term (B_j). The resulting EMS is then $bn\phi_A + n\sigma_{AB}^2 + 1 \cdot \sigma_\varepsilon^2$. For all terms, these rules give:

	a	b	n	
	F	R	R	
	i	j	k	EMS
A_i	0	b	n	$\sigma_\varepsilon^2 + n\sigma_{AB}^2 + nb\phi_A$
B_j	a	1	n	$\sigma_\varepsilon^2 + na\sigma_B^2$
AB_{ij}	0	1	n	$\sigma_\varepsilon^2 + n\sigma_{AB}^2$
$\varepsilon_{k(ij)}$	1	1	1	σ_ε^2

These results are seen to be in agreement with the EMS values for the mixed model in Section 10.3. Here ϕ_A is, of course, a fixed type of variance

$$\phi_A = \frac{\sum_i A_i^2}{a - 1}$$

Although the rules seem rather involved, they become very easy to use with a bit of practice. Two examples will illustrate the concept.

Example 10.1 The viscosity of a slurry is to be determined by four randomly selected laboratory technicians. Material from each of five mixing machines is bottled and divided in such a way as to provide two samples for each technician to test for viscosity. These are the only mixing machines of interest and the samples can be presented to the technicians in a completely randomized order.

The model here assumes four random technicians and five fixed mixing machines, and each technician measures samples of each machine twice. The model is shown as the vertical column of Table 10.1 and the remainder of the table shows how the EMS column is determined.

Table 10.1 EMS for Example 10.1

Source	df	4 R i	5 F j	2 R k	EMS
T_i	3	1	5	2	$\sigma_\varepsilon^2 + 10\sigma_T^2$
M_j	4	4	0	2	$\sigma_\varepsilon^2 + 2\sigma_{TM}^2 + 8\phi_M$
TM_{ij}	12	1	0	2	$\sigma_\varepsilon^2 + 2\sigma_{TM}^2$
$\varepsilon_{k(ij)}$	20	1	1	1	σ_ε^2

The proper F tests are quite obvious from Table 10.1 and all tests have adequate degrees of freedom for a reasonable test.