Nested ANOVA

Sometimes, constraints prevent us from crossing every level of one factor with every level of the other factor. In these cases we are forced into what is known as a *nested* layout. We say we have a nested layout when fewer than all levels of one factor occur within each level of the other factor. Example, two independent measurements of the left wings of each of 4 female mosquitoes reared in each of 3 cages.

Left wing lengths of mosquitoes (mm).

Cage 1				Cage 2				Cage 3			
1	2	3	4	1	2	3	4	1	2	3	4
58.5	77.8	84.0	70.1	69.8	56.0	50.7	63.8	56.6	77.8	69.9	62.1
59.5	80.9	83.6	68.3	69.8	54.5	49.3	65.8	57.5	79.2	69.2	64.5

I. Models for nested ANOVA

a. A crucial requirement of this type of analysis is that groups representing a subordinate level of classification be randomly chosen. Hence, the subordinate level of a nested ANOVA is always Model II.

b. The highest level of classification in a nested ANOVA may be Model I or Model II.

c. If the highest level is Model II, we call it a **pure Model II nested ANOVA**.

i. $Y_{ijk} = \mu + A_i + B_{ij} + \varepsilon_{ijk}$, where Y_{ijk} is the k^{th} observation in the j^{th} subgroup of the i^{th} group, μ is the parametric mean of the population, A_i is the random contribution for the i^{th} group of the higher level A, B_{ij} is the random contribution for the j^{th} subgroup (level B) of the i^{th} group, and ε_{ijk} is the error term of the k^{th} item in the j^{th} subgroup of the i^{th} group.

d. If the highest level is Model I, we call it a **mixed model nested ANOVA**.

i. $Y_{ijk} = \mu + \alpha_i + B_{ij} + \varepsilon_{ijk}$, where Y_{ijk} is the k^{th} observation in the j^{th} subgroup of the i^{th} group, μ is the parametric mean of the population, α_i is a fixed treatment effect for the i^{th} group of the higher level A, B_{ij} is the random contribution for the j^{th} subgroup (level B) of the i^{th} group, and ε_{ijk} is the error term of the k^{th} item in the j^{th} subgroup of the i^{th} group.

II. ANOVA table for two-level nested ANOVA

Two-level nested ANOVA

Source of variation	Sum of squares (SS)	Degrees of freedom (df)	Mean Square (MS)	F Test Statistic (F)	Expected MS
Among groups	SS_{AG}	df_{AG}	MS_{AG}	F_{AG}	EMS_{AG}
Among subgroups within groups	SS_{ASG}	df_{ASG}	MS_{ASG}	F_{ASG}	EMS_{ASG}
Within subgroups	SS_{WSG}	df_{WSG}	MS_{WSG}		EMS_{WSG}
Total	SS_T	df_T			

III. Formulas for nested ANOVA (equal sample sizes)

a. $SS_{AG} = nb\sum_{i=1}^{a}(\overline{Y}_i - \overline{\overline{Y}})^2$

b. $SS_{ASG} = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{Y}_{ij} - \overline{Y}_i)^2$

c. $SS_{WSG} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \overline{Y}_{ij})^2$

d. $df_{AG} = a - 1$

e. $df_{ASG} = a(b-1)$

f. $df_{ASG} = ab(n-1)$

g. $df_T = abn - 1$

h. $MS_{AG} = SS_{AG} / df_{AG}$

i. $MS_{ASG} = SS_{ASG} / df_{ASG}$

j. $MS_{WSG} = SS_{WSG} / df_{WSG}$

k. Model II

i. $EMS_{AG} = \sigma^2 + n\sigma^2_{B\subset A} + nb\sigma^2_A$

ii. $EMS_{ASG} = \sigma^2 + n\sigma^2_{B\subset A}$

iii. $EMS_{WSG} = \sigma^2$

l. Mixed model

i. $EMS_{AG} = \sigma^2 + n\sigma^2_{B\subset A} + \frac{nb\sum\alpha^2}{a-1}$

ii. $EMS_{ASG} = \sigma^2 + n\sigma^2_{B\subset A}$

iii. $EMS_{WSG} = \sigma^2$

IV. Three-level nested ANOVA

Glycogen content of liver in arbitrary units. Duplicate readings on each of 3 preparations of rat livers from each of 2 rats for each of 3 treatments.

Treatments	Control						Compound 217						Compound 217 plus sugar					
Rats	1			2			1			2			1			2		
Preparations	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
	131	131	136	150	140	160	157	154	147	151	147	162	134	138	135	138	139	134
	130	125	142	148	143	150	145	142	153	155	147	152	125	138	136	140	138	127

Two-way ANOVA

The two-way ANOVA is an extension of the one-way ANOVA, except there are two independent factors (hence the name two-way). If groups are formed along more than one dimension, differences among means are attributable to more than one source. Consider an example with six groups, three of women and three of men in which the DV is scores on a final examination in statistics. One source of variation in means is due to gender (SS_G). If the three groups within each gender are exposed to three different methods of teaching statistics, a second source of differences among means is teaching method (SS_T). The final source of known differences among means is the interaction between gender and teaching methods (SS_{GT}), which assess whether effectiveness of teaching methods varies with gender.

V. Factorial designs
   a. Factorial designs allow us to measure two different sorts of factor effects.
      i. The main effect of each factor is the effect of each factor independent of the other factors.
      ii. The interaction between factors is a measure of how the effects of one factor depend on the level of one or more additional factors.

VI. Null hypotheses
   a. There is no main effect for factor A.
   b. There is no main effect for factor B.
   c. There is no interaction between the two factors.

VII. Models for two-way ANOVA
   a. Model I
   b. Model II
   c. Mixed model

VIII. Two-way ANOVA table

Two-way ANOVA with replication

Source of variation	Sum of squares (SS)	Degrees of freedom (df)	Mean Square (MS)	F Test Statistic (F)	Expected MS
A (columns)	SS_A	d_{fA}	MS_A	F_A	EMS_A
B (rows)	SS_B	df_B	MS_B	F_B	EMS_B
A x B (interaction)	SS_I	df_I	MS_I	F_I	EMS_I
Within	SS_W	df_W	MS_W		EMS_W
Total	SS_T	df_T			

IX. Formulas for two-way ANOVA

a. $SS_A = nb\sum_{i=1}^{a}(\overline{Y}_i - \overline{\overline{Y}})^2$

b. $SS_B = na\sum_{j=1}^{b}(\overline{Y}_j - \overline{\overline{Y}})^2$

c. $SS_I = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{Y}_{ij} - \overline{Y}_i - \overline{Y}_j - \overline{\overline{Y}})^2$

d. $SS_W = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \overline{Y}_{ij})^2$

e. $df_A = a-1$

f. $df_B = b-1$

g. $df_I = (a-1)(b-1)$

h. $df_W = ab(n-1)$

i. $df_T = abn-1$

j. $MS_A = {SS_A}/{df_A}$

k. $MS_B = {SS_B}/{df_B}$

l. $MS_I = {SS_I}/{df_I}$

m. $MS_W = {SS_W}/{df_W}$

n. Model I

i. $EMS_A = \sigma^2 + \frac{nb}{a-1}\sum\alpha^2$

ii. $EMS_B = \sigma^2 + \frac{na}{b-1}\sum\beta^2$

iii. $EMS_I = \sigma^2 + \frac{n}{(a-1)(b-1)}\sum(\alpha\beta)^2$

iv. $EMS_W = \sigma^2$

o. Model II

i. $EMS_A = \sigma^2 + n\sigma^2_{AB} + nb\sigma^2_A$

ii. $EMS_B = \sigma^2 + na\sigma^2_B + na\sigma^2_B$

iii. $EMS_I = \sigma^2 + n\sigma^2_{AB}$

iv. $EMS_W = \sigma^2$

p. Mixed model

i. $EMS_A = \sigma^2 + n\sigma^2_{AB} + \frac{nb}{a-1}\sum\alpha^2$

ii. $EMS_B = \sigma^2 + na\sigma^2_B$

iii. $EMS_I = \sigma^2 + n\sigma_{AB}^2$
iv. $EMS_W = \sigma^2$

X. Two-way ANOVA without replication

a. When there is only a single observation for each combination of factors, it is impossible to test the null hypothesis of no interaction. However, we can test the null hypotheses regarding main effects as long as we assume no interaction.
b. The sum of squares is calculated for each of the two main effects, and a total sum of squares is calculated by considering all of the observations as a single group. After we subtract the sum of squares for columns (factor A) and for rows (factor B) from the total sum of squares, we are left with only a single sum of square.
c. Because we assume no interaction, the row and column mean squares are tested over the error mean square.