17 Nested and Higher Order Designs

17.1 Two-Way Analysis of Variance

Consider an experiment in which the treatments are combinations of two or more influences on the response. The individual influences will be called factors, and the possible values for the factors will be called the levels of the factor.

The important concepts can be best illustrated in the case of two factors in a fixed effects model: In an experiment with two factors A and B, a specific treatment combination consists of factor A at level i and factor B at level j. Assume that there are a levels of factor A and b levels of factor B under investigation. Thus the experiment consists of $a \times b$ treatment combinations. Assume the linear model

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk},$$

with $i = 1, \ldots, a;\ j = 1, \ldots, b;\ k = 1, \ldots, n_{ij};$ and $\varepsilon_{ijk} \sim N(0, \sigma^2)$ independent.

17.1 Theorem: The least-squares estimates are $\hat{\mu}_{ij} = \bar{Y}_{ij\cdot} = \sum_{k=1}^{n_{ij}} Y_{ijk}/n_{ij}$.

17.2 Note: This model is often called the cell means model. This model has $ab - 1$ degrees of freedom and the error degrees of freedom are $\sum_i \sum_j (n_{ij} - 1)$. However, such an analysis is not of much interest since we have not included any structure on the way in which the factors influence the response. One useful structure postulates the existence of an additive structure. Write

$$\mu_{ij} = \bar{\mu}_{\cdot\cdot} + (\bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot\cdot}) + (\bar{\mu}_{\cdot j} - \bar{\mu}_{\cdot\cdot}) + (\mu_{ij} - \bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot j} + \bar{\mu}_{\cdot\cdot}) = \mu + \alpha_i + \beta_j + \gamma_{ij},$$

where

μ	$=$	$\bar{\mu}_{\cdot\cdot}$	is the overall mean,
α_i	$=$	$\bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot\cdot}$	is the effect of the ith level of factor A,
β_j	$=$	$\bar{\mu}_{\cdot j} - \bar{\mu}_{\cdot\cdot}$	is the effect of the jth level of factor B,
γ_{ij}	$=$	$\mu_{ij} - \bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot j} + \bar{\mu}_{\cdot\cdot}$	is the effect of an interaction between levels i and j of factors A and B.

17.3 Note: We use the identifiability constraints

$$\sum_{i=1}^{a} \alpha_i = 0, \quad \sum_{j=1}^{b} \beta_j = 0, \quad \sum_{i=1}^{a} \gamma_{ij} = \sum_{j=1}^{b} \gamma_{ij} = 0.$$

17.4 Note: Hypothesis of interest are: $H_{AB} : \gamma_{ij} = 0\ \forall\, i, j;\quad H_A :\ \alpha_i = 0\ \forall\, i;\quad H_B :\ \beta_j = 0\ \forall\, j.$

17.5 Note: No interaction means that the difference between two groups i and l defined by one factor is independent of the level j of the other factor:

$$\begin{aligned} & \mu_{ij} - \bar{\mu}_{i\cdot} &=& \ \bar{\mu}_{\cdot j} - \bar{\mu}_{\cdot\cdot} & \forall i, j \\ \Longleftrightarrow\ & \mu_{ij} - \bar{\mu}_{i\cdot} &=& \ \mu_{i'j} - \bar{\mu}_{i'\cdot} & \forall i, i', j \\ \Longleftrightarrow\ & \mu_{ij} - \mu_{i'j} &=& \ \bar{\mu}_{i\cdot} - \bar{\mu}_{i'\cdot} & \forall i, i', j \end{aligned}$$

17.6 Note: H_A is equivalent to $\bar{\mu}_{i\cdot} - \bar{\mu}_{i'\cdot} = 0 \;\forall\; i$, i. e. averaged across levels of factor B, the average mean is constant across levels of factor A.

17.7 Note: If $H_{AB} : \gamma_{ij} = 0$ is true, then H_A is equivalent to $\mu_{ij} - \mu_{i'j} = 0 \;\forall\; i, i', j$, i. e. the mean is constant across levels of factor A within each level of factor B. Thus, hypotheses on main effects make most sense if there is no interaction.

17.8 Theorem: In a balanced design (i. e. $n_{ij} = r \;\; \forall i, j$), the least-squares estimates are

$$\hat{\mu} = \bar{Y}_{\cdots}, \quad \hat{\alpha}_i = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots}, \quad \hat{\beta}_j = \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdots}, \quad \hat{\gamma}_{ij} = \bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots}.$$

The ANOVA decomposition is given by

$$\text{SS}_{\text{TOTAL}} = \text{SS}_A + \text{SS}_B + \text{SS}_{AB} + \text{SS}_E$$

where

$$\begin{aligned}
\text{SS}_{\text{TOTAL}} &= \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{\cdots})^2, \\
\text{SS}_A &= \sum_i \sum_j \sum_k (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots})^2 = rb \sum_i (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots})^2 = rb \sum_i \hat{\alpha}_i^2 = RSS_{H_A} - RSS, \\
\text{SS}_B &= \sum_i \sum_j \sum_k (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdots})^2 = ra \sum_j (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdots})^2 = ra \sum_j \hat{\beta}_j^2 = RSS_{H_B} - RSS, \\
\text{SS}_{AB} &= \sum_i \sum_j \sum_k (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots})^2 = r \sum_i \sum_j \hat{\gamma}_{ij}^2 = RSS_{H_{AB}} - RSS, \\
\text{SS}_E &= \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij\cdot})^2 = RSS.
\end{aligned}$$

17.9 Theorem: For the fixed effects model we have:

(a) $E[\text{MS}_A] = E[\text{SS}_A/(a-1)] = \sigma^2 + \frac{rb}{a-1} \sum_i \alpha_i^2$.

(b) $E[\text{MS}_B] = E[\text{SS}_B/(b-1)] = \sigma^2 + \frac{ra}{b-1} \sum_i \beta_i^2$.

(c) $E[\text{MS}_{AB}] = E[\text{SS}_{AB}/((a-1)(b-1))] = \sigma^2 + \frac{r}{(a-1)(b-1)} \sum_i \sum_j \gamma_{ij}^2$.

(d) $E[\text{MS}_E] = E\left[\text{SS}_E/(ab(r-1))\right] = \sigma^2$.

17.10 Theorem: For the fixed effects model we have:

(a) $\text{MS}_A/\text{MS}_E \sim F_{a-1, ab(r-1)}$ if H_A is true.

(b) $\text{MS}_B/\text{MS}_E \sim F_{b-1, ab(r-1)}$ if H_B is true.

(c) $\text{MS}_{AB}/\text{MS}_E \sim F_{(a-1)(b-1), ab(r-1)}$ if H_{AB} is true.

17.11 Note: The above is usually summarized in an ANOVA table:

source	sum of squares	df	mean squares	test statistic
Factor A	SS_A	$a-1$	$MS_A = SS_A/(a-1)$	MS_A/MS_E
Factor B	SS_B	$b-1$	$MS_B = SS_B/(b-1)$	MS_B/MS_E
Interaction	SS_{AB}	$(a-1)(b-1)$	$MS_{AB} = SS_{AB}/((a-1)(b-1))$	MS_{AB}/MS_E
Error	SS_E	$ab(r-1)$	$MS_E = SS_E/(ab(r-1))$	
Total	SS_{TOTAL}	$abr-1$		

17.12 Note: The above ANOVA table is for a fixed effects model. To calculate the correct F statistics for the hypothesis tests in 2-way ANOVA random effects and mixed effects models, we have to consider the expected mean squares for the different factors. The expected mean squares for fixed, random, and mixed (A fixed and B random) effects ANOVAs are shown in the table below:

source	fixed effects	random effects	mixed effects
Factor A	$\sigma^2 + \frac{rb}{a-1}\sum_i \alpha_i^2$	$\sigma^2 + r\sigma_{AB}^2 + rb\sigma_A^2$	$\sigma^2 + r\sigma_{AB}^2 + \frac{rb}{a-1}\sum_i \alpha_i^2$
Factor B	$\sigma^2 + \frac{ra}{b-1}\sum_i \beta_i^2$	$\sigma^2 + r\sigma_{AB}^2 + ra\sigma_B^2$	$\sigma^2 + ra\sigma_B^2$
Interaction	$\sigma^2 + \frac{r}{(a-1)(b-1)}\sum_i\sum_j \gamma_{ij}^2$	$\sigma^2 + r\sigma_{AB}^2$	$\sigma^2 + r\sigma_{AB}^2$
Error	σ^2	σ^2	σ^2

17.2 Nested Designs

In the two-way classification the factors were crossed, i. e. each level i of factor A occurred with each level j of factor B. In contrast, B is called nested within A if the levels of B are sampled within each of the levels of A, and so different levels of A have different levels of B.

17.13 Example: Assume factor A represents different laboratories. Within each lab, 3 technicians are sampled from all the technicians in that lab, and each technician performs some experiment.

Laboratory		1			2	
	↙	↓	↘	↙	↓	↘
Technician	$(1,1)$	$(1,2)$	$(1,3)$	$(2,1)$	$(2,2)$	$(2,3)$
	μ_{11}	μ_{12}	μ_{13}	μ_{21}	μ_{22}	μ_{23}

If instead the same 3 technicians performed the experiment in all the labs, the factors would be crossed:

	Technician		
Laboratory	1	2	3
1	μ_{11}	μ_{12}	μ_{13}
2	μ_{21}	μ_{22}	μ_{23}

17.14 Note: The model for the nested design is $\mu_{ij} = \mu + \alpha_i + \beta_{ij}$, while the model for the crossed design is $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$. The nested design can be seen as an incomplete 2-way layout:

	Technician					
Laboratory	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)
1	μ_{11}	μ_{12}	μ_{13}			
2				μ_{21}	μ_{22}	μ_{23}

17.15 Example: Assume we have a hospitals, b technicians per hospital, and each technician performs r replications of the experiment. The ANOVA model is

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

with $i = 1, \dots, a$; $j = 1, \dots, b$; $k = 1, \dots, r$; and $\varepsilon_{ijk} \sim N(0, \sigma^2)$ independent. We write

$$\mu_{ij} = \bar{\mu}_{\cdot\cdot} + (\bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot\cdot}) + (\mu_{ij} - \bar{\mu}_{i\cdot}) = \mu + \alpha_i + \beta_{ij},$$

where α_i is the Hospital effect, and β_{ij} is the effect of jth technician in ith hospital. In a random effects model, we assume $\alpha_i \sim N(0, \sigma_A^2)$ and $\beta_{ij} \sim N(0, \sigma_{B|A}^2)$.

17.16 Note: Hypothesis of interest are: $H_A:\ \sigma_A^2 = 0$ and $H_B:\ \sigma_{B|A}^2 = 0$.

17.17 Note: H_B implies that $\mu_{ij} = \bar{\mu}_{i\cdot}\ \forall\, i, j \iff \mu_{ij} - \mu_{ij'} = 0\ \forall\, i, j, j'$, i. e. there are no differences between levels of B within each level of A.

17.18 Theorem: The ANOVA decomposition in a balanced design is given by

$$\text{SS}_{\text{TOTAL}} = \text{SS}_A + \text{SS}_{B|A} + \text{SS}_E$$

where

$$\begin{aligned}
\text{SS}_{\text{TOTAL}} &= \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{\cdot\cdot\cdot})^2, \\
\text{SS}_A &= \sum_i \sum_j \sum_k (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2 = rb \sum_i (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2, \\
\text{SS}_{B|A} &= \sum_i \sum_j \sum_k (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot})^2 = r \sum_i \sum_j (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot})^2, \\
\text{SS}_E &= \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij\cdot})^2.
\end{aligned}$$

17.19 Theorem: For the random effects model we have:

(a) $E[\text{MS}_A] = E[\text{SS}_A/(a-1)] = \sigma^2 + r\sigma^2_{B|A} + rb\sigma^2_A$.

(b) $E[\text{MS}_{B|A}] = E[\text{SS}_{AB}/(a(b-1))] = \sigma^2 + r\sigma^2_{B|A}$.

(c) $E[\text{MS}_E] = E\left[\text{SS}_E/(ab(r-1))\right] = \sigma^2$.

17.20 Theorem: For the random effects model we have:

(a) $\text{MS}_A/\text{MS}_{B|A} \sim F_{a-1,a(b-1)}$ if H_A is true.

(b) $\text{MS}_{B|A}/\text{MS}_E \sim F_{a(b-1),ab(r-1)}$ if $H_{B|A}$ is true.

17.21 Note: The above is usually summarized in an ANOVA table:

source	sum of squares	df	mean squares	test statistic
Factor A	SS_A	$a-1$	$\text{MS}_A = \text{SS}_A/(a-1)$	$\text{MS}_A/\text{MS}_{B\|A}$
Factor $B\|A$	$\text{SS}_{B\|A}$	$a(b-1)$	$\text{MS}_{B\|A} = \text{SS}_{B\|A}/(a(b-1))$	$\text{MS}_{B\|A}/\text{MS}_E$
Error	SS_E	$ab(r-1)$	$\text{MS}_E = \text{SS}_E/(ab(r-1))$	
Total	SS_{TOTAL}	$abr-1$		

17.22 Note: The above ANOVA table is for a random effects model. In general, by the definition of a nested model, the nested factor is usually considered random. However, the expected mean squares can be calculated for a fixed effects model too. In the table below we summarize the expected mean squares for random, mixed (A fixed and B random), and fixed effects nested ANOVA models.

source	random effects	mixed effects	fixed effects
Factor A	$\sigma^2 + r\sigma^2_{B\|A} + rb\sigma^2_A$	$\sigma^2 + r\sigma^2_{B\|A} + \frac{rb}{a-1}\sum_i \alpha_i^2$	$\sigma^2 + \frac{rb}{a-1}\sum_i \alpha_i^2$
Factor $B\|A$	$\sigma^2 + r\sigma^2_{B\|A}$	$\sigma^2 + r\sigma^2_{B\|A}$	$\sigma^2 + \frac{r}{a(b-1)}\sum_i\sum_j \beta_{ij}^2$
Error	σ^2	σ^2	σ^2