4.2 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**
- 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 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 p levels of factor A and q levels of factor B under investigation.
  - Thus the experiment consists of pq treatment combinations.

4.2.1 Structure 1: Fixed Effects Models

- If the response to treatment combination i, j is observed on n_{ij} experimental units we say that the i, j treatment combination is **replicated** n_{ij} times.

- Denoting the response for the kth replication by y_{ijk} the model assumes that the y_{ijk} are realized values of random variables Y_{ijk} having the following wide sense model

$$E(Y_{ijk}) = \eta_{ij} \;\text{ and }\; \text{cov}\,(Y_{ijk}, Y_{i'j'k'}) = \begin{cases} \sigma^2 & (i', j', k') = (i, j, k) \\ 0 & \text{elsewhere} \end{cases}$$

  for $k = 1, 2, \ldots, n_{ij}$, $j = 1, 2, \ldots, q$ and $i = 1, 2, \ldots, p$.

- The sum of squares for this model i.e. for testing $H_0 : \eta_{ij} = \eta$ is

$$\sum_{i=1}^{p} \sum_{j=1}^{q} \frac{T_{ij}^2}{n_{ij}} - G^2/n \;\text{ where }\; T_{ij} = \sum_{k=1}^{n_{ij}} y_{ijk} \;,\; G = \sum_{ij} T_{ij}$$

- The error sum of squares is

$$\sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{n_{ij}} y_{ijk}^2 - \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{T_{ij}^2}{n_{ij}} = \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij+})^2$$

  where $\bar{y}_{ij+} = \frac{T_{ij}}{n_{ij}}$.

- This model has $pq - 1$ degrees of freedom and the error degrees of freedom is
$$\sum_{i=1}^{p} \sum_{j=1}^{q} (n_{ij} - 1)$$

- It is often called the **cell means model**.

- 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.

- Define the ith row mean, jth column mean and overall mean by

$$\begin{aligned}
i\text{th row mean} &= \bar{\eta}_{i+} \\
&= \frac{\eta_{i1} + \eta_{i2} + \cdots + \eta_{iq}}{q} \\
j\text{th column mean} &= \bar{\eta}_{+j} \\
&= \frac{\eta_{1j} + \eta_{2j} + \cdots + \eta_{pj}}{p} \\
\text{overall mean} &= \bar{\eta}_{++} \\
&= \frac{\sum_{i=1}^{p} \sum_{j=1}^{q} \eta_{ij}}{pq}
\end{aligned}$$

- Then $\rho_i = \bar{\eta}_{i+} - \bar{\eta}_{++}$ provides a measure of the expected influence of the ith level of A relative to the overall mean.
  - ρ_i is called the **main effect** of the ith level of A or the ith row effect.
- Similarly $\gamma_j = \bar{\eta}_{+j} - \bar{\eta}_{++}$ provides a measure of the expected influence of the jth level of B relative to the overall mean.
  - γ_j is called the main effect of the jth level of B or the jth column effect.
- Note that $\sum_{i=1}^{p} \rho_i = 0$ and $\sum_{j=1}^{q} \gamma_j = 0$.

Since we can write

$$\eta_{ij} = \bar{\eta}_{++} + (\bar{\eta}_{i+} - \bar{\eta}_{++}) + (\bar{\eta}_{+j} - \bar{\eta}_{++}) + (\eta_{ij} - \bar{\eta}_{i+} - \bar{\eta}_{+j} + \bar{\eta}_{++})$$

we have a decomposition of the expected response to the i, j treatment combination into four components:

- a general or overall mean; $\mu = \bar{\eta}_{++}$
- an effect due to level i of factor A; $\rho_i = \bar{\eta}_{i+} - \bar{\eta}_{++}$
- an effect due to level j of factor B; $\gamma_j = \bar{\eta}_{+j} - \bar{\eta}_{++}$
- an effect of the form; $\lambda_{ij} = \bar{\eta}_{ij} - \bar{\eta}_{i+} - \bar{\eta}_{+j} + \bar{\eta}_{++}$ called the **interaction** of the ith level of A with the jth level of B.

One useful way to interpret interaction is to note that the expected response of the i, j treatment combination relative to the jth level of B is

$$\eta_{ij} - \bar{\eta}_{+j}$$

- Since the main effect of the ith level of A is $\bar{\eta}_{i+} - \bar{\eta}_{++}$, comparing the effect of the ith level of A (relative to the jth level of B) to the main effect of the ith level of A yields

$$(\eta_{ij} - \bar{\eta}_{+j}) - (\bar{\eta}_{i+} - \bar{\eta}_{++}) = \lambda_{ij}$$

We obtain the same result when we compare the effect of the jth level of B (relative to the ith level of A) to the main effect of the jth level of B i.e.

$$(\eta_{ij} - \bar{\eta}_{i+}) - (\bar{\eta}_{+j} - \bar{\eta}_{++}) = \lambda_{ij}$$

- Thus interaction may be interpreted as the failure of the effect of a factor to remain constant over different levels of the other factor.
  - Note that $\sum_{i=1}^{p} \lambda_{ij} = 0$ for $j = 1, 2, \ldots, q$ and $\sum_{j=1}^{q} \lambda_{ij} = 0$ for $i = 1, 2, \ldots, p$.
- If all of the λ_{ij} are zero we have a simple additive structure for the way in which the factors influence the expected response since in this case

$$\eta_{ij} = \mu + \rho_i + \gamma_j$$

- When the λ_{ij} are not zero no such simple structure exists even though we can define main effects for the factors.
- Put another way: if the λ_{ij} are not zero we need to know the expected response at the i, j combination of factors A and B; knowing the expected response to the ith level of A and the expected response to the jth level of B is not enough.
- Thus interaction effects measure departure from a simple additive structure involving only main effects.
- Simply put: In the presence of interaction we must resort to the full model in order to obtain a satisfactory representation of the influence of the factors on the expected response.

In order to obtain the sum of squares in an analysis of variance for the model in which

$$E(Y_{ijk}) = \mu + \rho_i + \gamma_j + \lambda_{ij}$$

- We first obtain the analysis for the reduced model

$$E(Y_{ijk}) = \mu + \rho_i + \gamma_j$$

- Then since the error sum of squares for the full model is

$$\sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{n_{ij}} y_{ijk}^2 - \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{T_{ij}^2}{n_{ij}}$$

  and the full model sum of squares is

$$\sum_{i=1}^{p} \sum_{j=1}^{q} \frac{T_{ij}^2}{n_{ij}}$$

  we can obtain the sum of squares for interaction adjusted for main effects by subtraction.

  - (This saves us from solving a set of least squares equations involving $1 + p + q + pq$ estimates).

The design matrix for the model

$$E(Y_{ijk}) = \mu + \rho_i + \gamma_j$$

consists of $n = \sum_{i=1}^{p} \sum_{j=1}^{q} n_{ij}$ rows and $1 + p + q$ columns.

- The rows corresponding to the ith level of A and the jth level of B are of the form

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}$$

  i.e. 1 in the first column, 1 in the $(i+1)$ column and 1 in the $(p+1+j)$ column.

- The least squares equations $\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y}$ are thus

$$\begin{bmatrix} n & r_1 & r_2 & \cdots & r_p & k_1 & k_2 & \cdots & k_q \\ r_1 & r_1 & 0 & \cdots & 0 & n_{11} & n_{12} & \cdots & n_{1q} \\ r_2 & 0 & r_2 & \cdots & 0 & n_{21} & n_{22} & \cdots & n_{2q} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ r_p & 0 & 0 & \cdots & r_p & n_{p1} & n_{p2} & \cdots & n_{pq} \\ k_1 & n_{11} & n_{21} & \cdots & n_{p1} & k_1 & 0 & \cdots & 0 \\ k_2 & n_{12} & n_{22} & \cdots & n_{p2} & 0 & k_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ k_q & n_{1q} & n_{2q} & \cdots & n_{pq} & 0 & 0 & \cdots & k_q \end{bmatrix} \begin{bmatrix} m \\ a_1 \\ a_2 \\ \vdots \\ a_p \\ d_1 \\ d_2 \\ \vdots \\ d_q \end{bmatrix} = \begin{bmatrix} G \\ A_1 \\ A_2 \\ \vdots \\ A_p \\ B_1 \\ B_2 \\ \vdots \\ B_q \end{bmatrix}$$

where

- $n = \sum_{i=1}^{p} \sum_{j=1}^{q} n_{ij}$
- $r_i = n_{i+} = \sum_{j=1}^{q} n_{ij}$
- $k_j = n_{+j} = \sum_{i=1}^{p} n_{ij}$
- $G = \sum_{i=1}^{p} \sum_{j=1}^{q} y_{ijk}$
- $A_i = T_{i+} = \sum_{j=1}^{q} T_{ij}$
- $B_j = T_{+j} = \sum_{i=1}^{p} T_{ij}$
- $T_{ij} = \sum_{k=1}^{n_{ij}} y_{ijk}$
- m is the estimate of μ
- a_i is the estimate of ρ_1
- d_j is the estimate of γ_j

- We may write these equations in matrix notation as

$$\begin{bmatrix} n & \mathbf{r}^T & \mathbf{k}^T \\ \mathbf{r} & \mathbf{R} & \mathbf{N} \\ \mathbf{k} & \mathbf{N}^T & \mathbf{K} \end{bmatrix} \begin{bmatrix} m \\ \mathbf{a} \\ \mathbf{d} \end{bmatrix} \begin{bmatrix} m \\ \mathbf{A} \\ \mathbf{D} \end{bmatrix}$$

where

$$\begin{aligned} \mathbf{r}^T &= (r_1, r_2, \ldots, r_p) \\ \mathbf{k}^T &= (k_1, k_2, \ldots, k_q) \\ \mathbf{R} &= \text{diag}\,(r_1, r_2, \ldots, r_p) \\ \mathbf{K} &= \text{diag}\,(k_1, k_2, \ldots, k_q) \\ \mathbf{N} &= (n_{ij})_{p \times q} \end{aligned}$$

Multiplication of both sides of the least squares equations by the non singular matrix

$$\begin{bmatrix} 1 & \mathbf{0}^T & \mathbf{0}^T \\ -\frac{1}{n}\mathbf{r} & \mathbf{I} & \mathbf{0} \\ -\frac{1}{n}\mathbf{k} & \mathbf{0}^T & \mathbf{I} \end{bmatrix}$$

yields the equivalent equations

$$\begin{bmatrix} n & \mathbf{r}^T & \mathbf{k}^T \\ \mathbf{0} & \mathbf{R} - \frac{1}{n}\mathbf{r}\mathbf{r}^T & \mathbf{N} - \frac{1}{n}\mathbf{r}\mathbf{k}^T \\ \mathbf{0} & \mathbf{N}^T - \frac{1}{n}\mathbf{k}\mathbf{r}^T & \mathbf{K} - \frac{1}{n}\mathbf{k}\mathbf{k}^T \end{bmatrix} \begin{bmatrix} m \\ \mathbf{a} \\ \mathbf{d} \end{bmatrix} \begin{bmatrix} G \\ \mathbf{A} - \frac{1}{n}\mathbf{r}G \\ \mathbf{D} - \frac{1}{n}\mathbf{k}G \end{bmatrix}$$

If all r_i are positive then R^{-1} exists and if we multiply both sides by the non singular matrix

$$\begin{bmatrix} 1 & \mathbf{0}^T & \mathbf{0}^T \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\left(\mathbf{N}^T - \frac{1}{n}\mathbf{k}\mathbf{r}^T\right)\mathbf{R}^{-1} & \mathbf{I} \end{bmatrix}$$

we obtain the equivalent equations

$$\begin{bmatrix} n & \mathbf{r}^T & \mathbf{k}^T \\ \mathbf{0} & \mathbf{R} - \frac{1}{n}\mathbf{r}\mathbf{r}^T & \mathbf{N} - \frac{1}{n}\mathbf{r}\mathbf{k}^T \\ \mathbf{0} & \mathbf{0} & \mathbf{C} \end{bmatrix} \begin{bmatrix} m \\ \mathbf{a} \\ \mathbf{d} \end{bmatrix} \begin{bmatrix} G \\ \mathbf{A} - \frac{1}{n}\mathbf{r}G \\ \mathbf{D}^* \end{bmatrix}$$

where

$$\begin{aligned} \mathbf{C} &= \mathbf{K} - \frac{1}{n}\mathbf{k}\mathbf{k}^T \left[\mathbf{N}^T - \frac{1}{n}\mathbf{k}\mathbf{r}^T\right] \mathbf{R}^{-1} \left[\mathbf{N} - \frac{1}{n}\mathbf{r}\mathbf{k}^T\right] \\ &= \mathbf{K} - \mathbf{N}^T\mathbf{R}^{-1}\mathbf{N} \\ \mathbf{D}^* &= \mathbf{D} - \frac{1}{n}\mathbf{k}G - \left[\mathbf{N}^T - \frac{1}{n}\mathbf{k}\mathbf{r}^T\right] \mathbf{R}^{-1} \left[\mathbf{A} - \frac{1}{n}G\right] \\ &= \mathbf{D} - \mathbf{N}^T\mathbf{R}^{-1}\mathbf{A} \end{aligned}$$

- Thus $\mathbf{d}$ is any solution to $\mathbf{C}\mathbf{d} = \mathbf{D}^*$ and the sum of squares for factor B adjusted for factor A is given by

$$\mathbf{d}^T\mathbf{D}^* = (\mathbf{D}^*)^T\mathbf{C}^-\mathbf{D}^*$$

- To find the sum of squares for factor A ignoring factor B we need to solve the equations

$$\begin{bmatrix} n & \mathbf{r}^T \\ \mathbf{r} & \mathbf{R} \end{bmatrix} \begin{bmatrix} m \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} G \\ \mathbf{A} \end{bmatrix}$$

- Multiplication of both sides by the non-singular matrix

$$\begin{bmatrix} 1 & \mathbf{0}^T \\ -\frac{1}{n}\mathbf{r} & \mathbf{I} \end{bmatrix}$$

yields the equivalent equations

$$\begin{bmatrix} n & \mathbf{r}^T \\ \mathbf{0} & \mathbf{R} - \frac{1}{n}\mathbf{r}\mathbf{r}^T \end{bmatrix} \begin{bmatrix} m \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} G \\ \mathbf{A} - \frac{1}{n}\mathbf{r}G \end{bmatrix}$$

- Since $\mathbf{R}^{-1}$ is a generalized inverse of $\mathbf{R} - \frac{1}{n}\mathbf{r}\mathbf{r}^T$ it follows that

$$\begin{aligned} \mathbf{a} &= \mathbf{R}^{-1}\left[\mathbf{A} - \frac{1}{n}\mathbf{r}G\right] \\ &= \mathbf{R}^{-1}\mathbf{A} - \frac{1}{n}\mathbf{1}G \end{aligned}$$

- Thus the sum of squares for factor A adjusted for the overall mean is

$$\begin{aligned} \left[\mathbf{A} - \frac{1}{n}\mathbf{r}G\right]^T \mathbf{a} &= \left[\mathbf{A} - \frac{1}{n}\mathbf{r}G\right]^T \mathbf{a}\left[\mathbf{R}^{-1}\mathbf{A} - \frac{1}{n}\mathbf{1}G\right] \\ &= \sum_{i=1}^{p} \frac{A_i^2}{r_i} - \frac{G^2}{n} \end{aligned}$$

- The sum of squares for the mean is clearly $\frac{G^2}{n}$.
- In terms of the treatment totals T_i we thus have

$$
\begin{aligned}
\text{SSMean} &= \frac{G^2}{n} \\
&= \frac{(\sum_{i=1}^{p}\sum_{j=1}^{q} T_{ij})^2}{n} \\
\text{SSA} &= \sum_{i=1}^{p}\frac{A_i^2}{r_i} - \frac{G^2}{n} \\
&= \sum_{i=1}^{p}\frac{T_{i+}^2}{r_i} - \frac{G^2}{n} \\
\text{SSB adj for A} &= \mathbf{D}^{*T}\mathbf{C}\mathbf{D}^{*} \\
\text{A} \times \text{B} &= \sum_{i=1}^{p}\sum_{j=1}^{q}\frac{T_{ij}^2}{n_{ij}} - \sum_{i=1}^{p}\frac{T_{i+}^2}{r_i} - \mathbf{D}^{*T}\mathbf{C}\mathbf{D}^{*} \\
\text{SSE} &= \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n_{ij}} y_{ijk}^2 - \sum_{i=1}^{p}\sum_{j=1}^{q}\frac{T_{ij}^2}{n_{ij}}
\end{aligned}
$$

- The degrees of freedom are $1, p-1, \text{rank}\,(\mathbf{C}), pq-p-\text{rank}\,(\mathbf{C})$ and $\sum_{i=1}^{p}\sum_{j=1}^{q}(n_{ij}-1)$ respectively.

We now consider the important special case of balance where $n_{ij} = r \geq 1$ i.e. there are r replicates for each treatment combination.

- In this case

$$\begin{aligned} r_i &= r + r + \cdots + r \\ &= rq \\ k_j &= r + r + \cdots + r \\ &= rp \\ n &= pqr \end{aligned}$$

- Thus

$$\text{SSA} = \sum_{i=1}^{p} \frac{T_{i+}^2}{rq} - \frac{G^2}{rpq}$$

- Note that the j, j' element of $\mathbf{C} = \mathbf{K} - \mathbf{N}^T\mathbf{R}^{-1}\mathbf{N}$ is given by

$$c_{jj'} = \begin{cases} k_j - \sum_{i=1}^{p} \frac{n_{ij}^2}{r_i} & j' = j \\ -\sum_{i=1}^{p} \frac{n_{ij}n_{ij'}}{r_i} & j' \neq j \end{cases}$$

Thus in the balanced case $\mathbf{C} = \mathbf{K} - \mathbf{N}^T\mathbf{R}^{-1}\mathbf{N}$ has j, j' element given by

$$c_{jj'} = \begin{cases} rp - \frac{rp}{q} & j' = j \\ -\frac{rp}{q} & j' \neq j \end{cases}$$

so that

$$\mathbf{C} = rp\mathbf{I}_q - \frac{rp}{q}\mathbf{1}\mathbf{1}^T$$

- The equation

$$\left[rp\mathbf{I}_q - \frac{rp}{q}\mathbf{1}_q\mathbf{1}_q^T\right]\frac{1}{rp}\mathbf{I}_q\left[rp\mathbf{I}_q - \frac{rp}{q}\mathbf{1}_q\mathbf{1}_q^T\right] = \left[rp\mathbf{I}_q - \frac{rp}{q}\mathbf{1}_q\mathbf{1}_q^T\right]$$

shows that a generalized inverse of $\mathbf{C}$ is $\frac{1}{rp}\mathbf{I}_q$

- Also in the balanced case $\mathbf{D}^* = \mathbf{D} - \mathbf{N}^T\mathbf{R}^{-1}\mathbf{A}$ has jth element given by

$$\begin{aligned} D_j - \sum_{i=1}^{p}\frac{n_{ij}A_i}{r_i} &= D_j - \sum_{i=1}^{p}\frac{rA_i}{rq} \\ &= D_j - \frac{1}{q}G \end{aligned}$$

so that $\mathbf{D}^* = \mathbf{D} - \frac{1}{q}\mathbf{1}_qG$.

- It follows that

$$\begin{aligned}\text{SS(B adj A)} &= \left[\mathbf{D} - \frac{1}{q}\mathbf{1}_q G\right]^T \frac{1}{rp}\mathbf{I}_q \left[\mathbf{D} - \frac{1}{q}\mathbf{1}_q G\right] \\ &= \sum_{j=1}^{p} \frac{D_j^2}{rp} - \frac{G^2}{rp}\end{aligned}$$

- We note that in this special case of equal replication, the sum of squares for factor B adjusted for factor A is identical to that which we would have obtained if we had found the sum of squares for B ignoring A.

  - The analysis is thus an orthogonal analysis.

- An explicit expression for the interaction sum of squares in the case of equal replication is given by

$$\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{T_{ij}^2}{r} - \sum_{i=1}^{p}\frac{T_{i+}^2}{rq} - \sum_{j=1}^{q}\frac{T_{+j}^2}{rp} + \frac{G^2}{rpq}$$

  with $pq - p - q + 1 = (p-1)(q-1)$ degrees of freedom.

If we define

$$\bar{y}_{+++} = \frac{G}{rpq}; \quad \bar{y}_{i++} = \frac{T_{i+}}{rq}; \quad \bar{y}_{+j+} = \frac{T_{+j}}{rp}; \quad \bar{y}_{ij+} = \frac{T_{ij}}{r}$$

the expressions for the various sums of squares may be written as

$$\begin{aligned}
\text{SSMean} &= rpq\bar{y}_{+++} \\
\text{SSA} &= rq \sum_{i=1}^{p} (\bar{y}_{i++} - \bar{y}_{+++})^2 \\
\text{SSB} &= rp \sum_{j=1}^{q} (\bar{y}_{+j+} - \bar{y}_{+++})^2 \\
\text{SSA} \times \text{B} &= r \sum_{i=1}^{p} \sum_{j=1}^{q} (\bar{y}_{ij+} - \bar{y}_{i++} - \bar{y}_{+j+} + \bar{y}_{+++})^2 \\
\text{SSE} &= \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{r} (y_{ijk} - \bar{y}_{ij+})^2
\end{aligned}$$

In order to investigate the expectations of the sums of squares for the model we assume that we are operating under the model:

$$E(Y_{ijk}) = \eta_{ij} \quad , \quad \text{cov}\ (Y_{ijk}, Y_{i'j'k'}) = \begin{cases} \sigma^2 & (i',j',k') = (i,j,k) \\ 0 & \text{otherwise} \end{cases}$$

Then we have

$$\begin{aligned}
E(Y_{ijk}^2) &= \text{var}\ (Y_{ijk}) + [E(Y_{ijk})]^2 \\
&= \sigma^2 + \eta_{ij}^2 \\
E\left(\frac{T_{ij}^2}{r}\right) &= \frac{\text{var}\ (\sum_{k=1}^r Y_{ijk}) + [E(\sum_{k=1}^r Y_{ijk})]^2}{r} \\
&= \sigma^2 + r\eta_{ij}^2 \\
E\left(\frac{T_{i+}^2}{rq}\right) &= \frac{\text{var}\ \left(\sum_{j=1}^q \sum_{k=1}^r Y_{ijk}\right) + \left[E(\sum_{j=1}^q \sum_{k=1}^r Y_{ijk})\right]^2}{rq} \\
&= \sigma^2 + rq\bar{\eta}_{i+}^2 \\
E\left(\frac{T_{+j}^2}{rp}\right) &= \frac{\text{var}\ (\sum_{i=1}^p \sum_{k=1}^r Y_{ijk}) + [E(\sum_{i=1}^p \sum_{k=1}^r Y_{ijk})]^2}{rp} \\
&= \sigma^2 + rp\bar{\eta}_{+j}^2 \\
E\left(\frac{G^2}{rpq}\right) &= \frac{\text{var}\ \left(\sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r Y_{ijk}\right) + \left[E(\sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r Y_{ijk})\right]^2}{rpq} \\
&= \sigma^2 + rpq\bar{\eta}_{++}^2
\end{aligned}$$

Putting these together yields

$$
\begin{aligned}
E(\mathrm{SSE}) &= pq(r-1)\sigma^2 \\
E(\mathrm{SS}(A \times B)) &= (p-1)(q-1)\sigma^2 + r\sum_{i=1}^{p}\sum_{j=1}^{q}\eta_{ij}^2 - rq\sum_{i=1}^{p}\bar{\eta}_{i+}^2 - rp\sum_{j=1}^{q}\bar{\eta}_{+j}^2 + rpq\bar{\eta}_{++}^2 \\
&= (p-1)(q-1)\sigma^2 + r\sum_{i=1}^{p}\sum_{j=1}^{q}\lambda_{ij}^2 \\
E(\mathrm{SS}(B)) &= (q-1)\sigma^2 + rp\sum_{j=1}^{q}\bar{\eta}_{+j}^2 - rpq\bar{\eta}_{++}^2 \\
&= (q-1)\sigma^2 + rp\sum_{j=1}^{q}\gamma_{+j}^2 \\
E(\mathrm{SS}(A)) &= (p-1)\sigma^2 + rq\sum_{i=1}^{p}\bar{\eta}_{i+}^2 - rpq\bar{\eta}_{++}^2 \\
&= (p-1)\sigma^2 + rq\sum_{i=1}^{p}\rho_{+j}^2
\end{aligned}
$$

where

$$
\begin{aligned}
\lambda_{ij} &= \eta_{ij} - \bar{\eta}_{i+} - \bar{\eta}_{+j} + \bar{\eta}_{++} \\
\gamma_j &= \bar{\eta}_{+j} - \bar{\eta}_{++} \\
\rho_i &= \bar{\eta}_{i+} - \bar{\eta}_{++}
\end{aligned}
$$

We thus have the following table of expected mean squares:

Source	d.f.	SS	Expectation
Mean	1	$rpq\bar{y}_{+++}^2$	$\sigma^2 + rpq\bar{\eta}_{++}^2$
A	$p-1$	$rq\sum_{i=1}^{p}(\bar{y}_{i++} - \bar{y}_{+++})^2$	$\sigma^2 + rq\sum_{i=1}^{p}\rho_i^2/(p-1)$
B	$q-1$	$rp\sum_{j=1}^{q}(\bar{y}_{+j+} - \bar{y}_{+++})^2$	$\sigma^2 + rp\sum_{j=1}^{q}\gamma_j^2/(q-1)$
A × B	$(p-1)(q-1)$	$r\sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{y}_{ij+} - \bar{y}_{i++} - \bar{y}_{+j+} + \bar{y}_{+++})^2$	$\sigma^2 + r\sum_{i=1}^{p}\sum_{j=1}^{q}\lambda_{ij}^2/(p-1)(q-1)$
Error	$pq(r-1)$	$\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}(y_{ijk} - \bar{y}_{ij+})^2$	σ^2

4.2.2 Structure 2: Random and Mixed Models

When we consider the process by which treatments are selected in the case where the treatments are combinations of two or more factors, we face a more complicated situation than in the one-way situation.

- The reason for the complexity is that some factors may be fixed while others are random.
- In order to consider all possibilities we concentrate on the case of two factors A and B.
- Following previous work on the one way model we assume that the y_{ijk} are realized values of Y_{ijk} with

$$E(Y_{ijk}|S) = \eta_{ij} \;\; ; \;\; \text{cov}\,(Y_{ijk}, Y_{i'j'k'}|S) = \begin{cases} \sigma^2 & (i',j',k') = (i,j,k) \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, \ldots, p; j = 1, 2, \ldots, q$ and $k = 1, 2, \ldots, r$.

- We concentrate on the balanced case because simple closed expressions for all the relevant sums of squares do not exist for the general case.

- We will, however, treat the case of two factors at two levels in the unbalanced case.

- From the ANOVA table of the preceeding section we see that if both factors are fixed i.e. no other levels are of interest other than those used in the experiment, the previous expectations are appropriate.

- Note that one can test the hypotheis of no main effects in the presence of interaction if $r > 1$ (but not if $r = 1$) but that such a test is of doubtful value if interactions are present since reporting of the responses at each treatment combination is the appropriate analysis.

- Later we present a test of a particular form of interaction when $r = 1$.

Both factors random

One useful model in this case assumes that

$$E(\eta_{ij}) = \mu \;\; ; \;\; \text{cov}\,(\eta_{ij}, \eta_{i'j'}) = \begin{cases} \sigma_1^2 + \sigma_2^2 + \sigma_3^2 & (i', j') = (i, j) \\ \sigma_1^2 & i' = i, j' \neq j \\ \sigma_2^2 & i' \neq i, j') = j \\ 0 & \text{otherwise} \end{cases}$$

- Under this model

$$\begin{aligned} E(\eta_{ij}^2) &= \text{var}\,(\eta_{ij}) + [E(\eta_{ij})]^2 \\ &= \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \mu^2 \end{aligned}$$

- It follows that

$$\begin{aligned} E\left(r\sum_{i=1}^{p}\sum_{j=1}^{q}\eta_{ij}^2\right) &= rpq\mu^2 + rpq\sigma_1^2 + rpq\sigma_2^2 + rpq\sigma_3^2 \\ E\left(rq\sum_{i=1}^{p}\bar{\eta}_{i+}^2\right) &= rpq\mu^2 + rpq\sigma_1^2 + rp\sigma_2^2 + rp\sigma_3^2 \\ E\left(rp\sum_{j=1}^{q}\bar{\eta}_{+j}^2\right) &= rpq\mu^2 + rq\sigma_1^2 + rpq\sigma_2^2 + rq\sigma_3^2 \\ E\left(rpq\bar{\eta}_{++}^2\right) &= rpq\mu^2 + rq\sigma_1^2 + rp\sigma_2^2 + r\sigma_3^2 \end{aligned}$$

The table of expected mean squares is thus given by

Source	d.f.	Expectation
Mean	1	$\sigma^2 + rpq\mu^2 + rpq\sigma^2 + rq\sigma_1^2 + rp\sigma_2^2 + r\sigma_3^2$
A	$p-1$	$\sigma^2 + \sigma_3^2 + rq\sigma_1^2$
B	$q-1$	$\sigma^2 + \sigma_3^2 + rp\sigma_2^2$
A × B	$(p-1)(q-1)$	$\sigma^2 + \sigma_3^2$
Error	$pq(r-1)$	σ^2

- The sums of squares being calculated exactly as before.
- Note that in this case we can test main effects even if $r = 1$ when interaction is present.
- But if $r = 1$ we have no test for interaction.
- The same cautions about interpreting main effects still apply although interpretations now depend on the relative importance of the variance components σ_1^2, σ_2^2 and σ_3^2

One factor fixed, one factor random

One useful model in this situation assumes that

$$E(\eta_{ij} = \mu_i) \;\; ; \;\; \text{cov}\,(\eta_{ij}, \eta_{i'j'}) = \begin{cases} \sigma_2^2 + \sigma_3^2 & (i', j') = (i, j) \\ \sigma_2^2 & i' \neq i, j') = j \\ 0 & \text{otherwise} \end{cases}$$

- Under this model

$$\begin{aligned} E(\eta_{ij}^2) &= \text{var}\,(\eta_{ij}) + [E(\eta_{ij})]^2 \\ &= \mu_i^2 + \sigma_2^2 + \sigma_3^2 \end{aligned}$$

- It follows that

$$\begin{aligned} E\left(r\sum_{i=1}^{p}\sum_{j=1}^{q}\eta_{ij}^2\right) &= rq\sum_{i=1}^{p}\mu_i^2 + rpq\sigma_2^2 + rpq\sigma_3^2 \\ E\left(rq\sum_{i=1}^{p}\bar{\eta}_{i+}^2\right) &= rq\sum_{i=1}^{p}\mu_i^2 + rp\sigma_2^2 + rp\sigma_3^2 \\ E\left(rp\sum_{j=1}^{q}\bar{\eta}_{+j}^2\right) &= \frac{rq}{p}\left(\sum_{i=1}^{p}\mu_i\right)^2 + rpq\sigma_2^2 + rq\sigma_3^2 \\ E\left(rpq\bar{\eta}_{++}^2\right) &= \frac{rq}{p}\left(\sum_{i=1}^{p}\mu_i\right)^2 + rpq\sigma_2^2 + r\sigma_3^2 \end{aligned}$$

The table of expected mean squares is thus given by

Source	d.f.	Expectation
A	$p-1$	$\sigma^2 + r\sigma_3^2 + rq\left(\sum_{i=1}^p \mu_i^2 - \frac{(\sum_{i=1}^p \mu_i)^2}{p}\right)/(p-1)$
B	$q-1$	$\sigma^2 + r\sigma_3^2 + rp\sigma_2^2$
A $\times$ B	$(p-1)(q-1)$	$\sigma^2 + r\sigma_3^2$
Error	$pq(r-1)$	σ^2

4.2.3 Scheffe's Approach

Scheffe, in his book, The Analysis of Variance, provides a careful treatment of the random and mixed model. Scheffe's approach is based on treating the cell means as random variables with distribution determined by the method of selection of the levels of the factors. He then defines effects as linear combinations of these random variables. The approach is very informative in situations where it is not obvious as to which effect is fixed and which is random.

4.2.4 Two way analysis of variance two factors each at two levels

We now investigate the unbalanced situation where we have two factors each at two levels.

- Thus we assume y_{ijk} are realized values of Y_{ijk} with

$$E(Y_{ijk}|S) = \eta_{ij} = \mu + \rho_i + \gamma_j + \lambda_{ij}$$

$$\text{cov}\,(Y_{ijk}, Y_{i'j'k'}|S) = \begin{cases} \sigma^2 & (i', j', k') = (i, j, k) \\ 0 & \text{otherwise} \end{cases}$$

- We know that

$$\text{SSE} = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n_{ij}} y_{ijk}^2 - \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{T_{ij}^2}{n_{ij}}$$

- If we find the sum of squares for the model in which $\lambda_{ij} = 0$ we can find the sum of squares for interaction by subtraction (this avoids a set of least squares equations with $1 + 2 + 2 + 4 = 9$ unknowns).

- For the model in which $\lambda_{ij} = 0$ we have the design matrix $\mathbf{X}$ given by

$$\mathbf{X} = \begin{bmatrix} \mathbf{1}_{n_{11}} & \mathbf{1}_{n_{11}} & \mathbf{0}_{n_{11}} & \mathbf{1}_{n_{11}} & \mathbf{0}_{n_{11}} \\ \mathbf{1}_{n_{12}} & \mathbf{1}_{n_{12}} & \mathbf{0}_{n_{12}} & \mathbf{0}_{n_{12}} & \mathbf{1}_{n_{12}} \\ \mathbf{1}_{n_{21}} & \mathbf{0}_{n_{21}} & \mathbf{1}_{n_{21}} & \mathbf{1}_{n_{21}} & \mathbf{0}_{n_{21}} \\ \mathbf{1}_{n_{22}} & \mathbf{0}_{n_{22}} & \mathbf{1}_{n_{22}} & \mathbf{0}_{n_{22}} & \mathbf{1}_{n_{22}} \end{bmatrix}$$

- The least squares equations for this model are thus

$$\begin{bmatrix} n & n_{1+} & n_{2+} & n_{+1} & n_{+2} \\ n_{1+} & n_{1+} & 0 & n_{11} & n_{12} \\ n_{2+} & 0 & n_{2+} & n_{21} & n_{22} \\ n_{+1} & n_{11} & n_{21} & n_{+1} & 0 \\ n_{+2} & n_{12} & n_{22} & 0 & n_{+2} \end{bmatrix} \begin{bmatrix} m \\ a_1 \\ a_2 \\ d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} G \\ A_1 \\ A_2 \\ D_1 \\ D_2 \end{bmatrix}$$

where

$$\begin{aligned} n_{1+} &= n_{11} + n_{12} \\ n_{2+} &= n_{21} + n_{22} \\ n_{+1} &= n_{11} + n_{21} \\ n_{+2} &= n_{12} + n_{22} \\ n &= n_{11} + n_{12} + n_{21} + n_{22} \\ A_1 &= T_{11} + T_{12} \\ A_2 &= T_{21} + T_{22} \\ D_1 &= T_{12} + T_{21} \\ D_2 &= T_{21} + T_{22} \\ G &= T_{11} + T_{12} + T_{21} + T_{22} \end{aligned}$$

and m, a_1, a_2, d_1 and d_2 are estimates of $\mu, \rho_1, \rho_2, \gamma_1$ and γ_2 respectively.

- Incredibly tedious algebra shows that

$$\begin{aligned}
\text{SSMean} &= \frac{G^2}{n} \\
\text{SSA} &= \frac{A^2}{n n_{1+} n_{2+}} \\
\text{SSB adj. A} &= \frac{n(D^*)^2}{a n_{1+} n_{2+}} \\
\text{SS A} \times \text{B adj, A and B} &= \frac{n n_{11} n_{12} n_{21} n_{22} (I^*)^2}{a}
\end{aligned}$$

where

$$\begin{aligned}
G &= T_{11} + T_{12} + T_{21} + T_{22} \\
A &= n_{2+}T_{11} + n_{2+}T_{12} - n_{1+}T_{21} - n_{1+}T_{22} \\
D^* &= n_{2+}n_{12}T_{11} - n_{2+}n_{11}T_{12} + n_{1+}n_{22}T_{21} - n_{1+}n_{21}T_{22} \\
I^* &= \frac{1}{n_{11}}T_{11} - \frac{1}{n_{12}}T_{12} - \frac{1}{n_{21}}T_{21} + \frac{1}{n_{22}}T_{22} \\
a &= n_{1+}n_{+1}n_{2+}n_{+2} - \left(n_{11}n_{22} - n_{12}n_{21}\right)^2
\end{aligned}$$

- Under the model

$$E(Y_{ijk}|S) = \eta_{ij} = \mu + \rho_i + \gamma_j + \lambda_{ij}$$

$$\text{cov}\,(Y_{ijk}, Y_{i'j'k'}|S) = \begin{cases} \sigma^2 & (i',j',k') = (i,j,k) \\ 0 & \text{otherwise} \end{cases}$$

  we have that

$$\begin{aligned}
E(G|S) &= n_{11}\eta_{11} + n_{12}\eta_{12} + n_{21}\eta_{21} + n_{22}\eta_{22} \\
\text{var}\,(G|S) &= n\sigma^2 \\
E(A|S) &= n_{2+}n_{11}\eta_{11} + n_{2+}n_{12}\eta_{12} + n_{1+}n_{21}\eta_{21} + n_{1+}n_{22}\eta_{22} \\
\text{var}\,(A|S) &= n_{1+}n_{2+}n\sigma^2 \\
E(D^*|S) &= n_{1+}n_{2+}n_{11}\eta_{11} - n_{2+}n_{11}n_{12}\eta_{12} + n_{1+}n_{22}n_{21}\eta_{21} - n_{1+}n_{21}n_{22}\eta_{22} \\
\text{var}\,(D^*|S) &= n_{1+}n_{2+}(n_{11}n_{12}n_{2+} + n_{1+}n_{21}n_{22})\sigma^2 \\
E(I^*|S) &= \eta_{11} - \eta_{12} - \eta_{21} + \eta_{22} \\
\text{var}\,(I^*|S) &= \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}\right)\sigma^2
\end{aligned}$$

- We can show that all covariances between G, A, D^* and I^* are zero so that the sums of squares under S and normality assumptions are independent.

- The table of expected values for the mean squares is, under S:

Source	d.f.	Expected Mean Square
A	1	$\sigma^2 + \frac{[n_{2+}(n_{11}\eta_{11}+n_{12}n_{12}\eta_{12})-n_{1+}(n_{21}\eta_{21}+n_{21}n_{22}\eta_{22})]^2}{nn_{1+}n_{2+}}$
B	1	$\sigma^2 + \frac{[n_{2+}n_{12}n_{11}(\eta_{11}-\eta_{12})-n_{1+}n_{21}n_{22}(\eta_{22}-\eta_{21})]^2}{an_{1+}n_{2+}}$
A × B	1	$\sigma^2 + \frac{nn_{11}n_{12}n_{21}n_{22}[\eta_{11}-\eta_{12}-\eta_{21}+\eta_{22}]^2}{a}$
Error	1	$\sigma^2\left(\sum_{i=1}^2\sum_{j=1}^2(n_{ij}-1)\right)$

If we now assume that the S structure for η_{ij} is

$$E(\eta_{ij}) = \mu$$

$$\text{cov}\,(\eta_{ij}, \eta_{i'j'}) = \begin{cases} \sigma_1^2 + \sigma_2^2 + \sigma_3^2 & (i', j') = (i, j) \\ \sigma_1^2 & i' = i, j') \neq j \\ \sigma_2^2 & i' \neq i, j') = j \\ 0 & \text{otherwise} \end{cases}$$

i.e. both factors are random then

$$\begin{aligned} \text{cov}\,(\eta_{11}, \eta_{12}) &= \sigma_1^2 \\ \text{cov}\,(\eta_{11}, \eta_{21}) &= \sigma_2^2 \\ \text{cov}\,(\eta_{11}, \eta_{22}) &= 0 \\ \text{cov}\,(\eta_{12}, \eta_{21}) &= 0 \\ \text{cov}\,(\eta_{21}, \eta_{22}) &= \sigma_1^2 \end{aligned}$$

and

$$\text{var}\,(\eta_{11}) = \text{var}\,(\eta_{12}) = \text{var}\,(\eta_{21}) = \text{var}\,(\eta_{22}) = \sigma_1^2 + \sigma_2^2 + \sigma_3^2$$

- Using these results we can calculate the expectations of the sums of squares. The formulas are complicated and will be determined by an easier method later on.

- What is important is that the covariances between A and I^*, A and D^* and between D^* and I^* are not zero unless

$$n_{11} = n_{12} = n_{21} = n_{22}$$

- Thus the sums of squares are not independent even though they are orthogonal.

- This is one of the major problems in unbalanced situations with random effects models.

- Similar results hold when the model is a mixed model i.e. when

$$E(\eta_{ij}) = \mu_i$$

$$\text{cov}\,(\eta_{ij}, \eta_{i'j'}) = \begin{cases} \sigma_2^2 + \sigma_3^2 & (i', j') = (i, j) \\ \sigma_2^2 & i' \neq i, j' = j \\ 0 & \text{otherwise} \end{cases}$$

4.3 Nested models

Consider an experiment designed to investigate inter-laboratory variability in which we have p laboratories, n_i technicians at the ith laboratory and n_{ij} readings by the jth technician at the ith laboratory.

- One model assumes that the observed responses y_{ijk} are realized values of Y_{ijk} which have structure S defined by

$$E(Y_{ijk}|S) = \eta_{ij} \text{ and } \operatorname{cov}(Y_{ijk}, Y_{i'j'k'}) = \begin{cases} \sigma^2 & (i', j', k') = (i, j, k) \\ 0 & \text{otherwise} \end{cases}$$

- Note that technician j in laboratory i has no relation to technician j in laboratory i'.
- That is, if we consider laboratories as a factor and technicians as a factor there is no logical meaning to any comparisons $\eta_{ij} - \eta_{i'j}$.
- In such cases we say that factor B (technicians) is **nested** within factor A (laboratories).
- Note also that readings are nested within technicians. (In fact the error term is always nested in the other treatment combinations.)
- This fact forces us to define a model for the η_{ij} in a different way from the other two factor models.
- For ease of exposition we consider p levels for factor A, q levels for factor B within each level of factor A and r replications within each level of B.

4.3.1 Case 1: A and B both fixed

A reasonable model for η_{ij} is provided by defining the overall effect as

$$\mu = \frac{1}{pq} \sum_{i=1}^{p} \sum_{j=1}^{q} \eta_{ij} = \bar{\eta}_{++}$$

- Then the main effect of the ith level of A is defined as

$$\rho_i = \frac{1}{q} \sum_{j=1}^{q} \eta_{ij} - \mu = \bar{\eta}_{i+} - \mu$$

- The main effect for the jth level of B **within** the ith level A is defined as

$$\gamma_{ij} = \eta_{ij} - \frac{1}{q} \sum_{j'}^{q} \eta_{ij'}$$

- Thus

$$\eta_{ij} = \mu + \rho_i + \gamma_{ij}$$

- Note that there is no need to introduce an interaction of A with B in this model.

- The sum of squares for the model is clearly

$$\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{T_{ij}^2}{r}$$

with pq degrees of freedom.

- Since the sum of squares for the model in which $\eta_{ij} = \mu + \rho_i$ is

$$\sum_{i=1}^{p}\frac{T_{i+}^2}{rq}$$

with p degrees of freedom we have that the sum of squares for factor B within A is

$$\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{T_{ij}^2}{r} - \sum_{i=1}^{p}\frac{T_{i+}^2}{rq}$$

with $pq - p = p(q-1)$ degrees of freedom.

- The relevant sums of squares are thus

$$\begin{aligned}
\text{SSMean} &= \frac{G^2}{rpq} \\
\text{SSA} &= \sum_{i=1}^{p}\frac{T_{i+}^2}{rq} - \frac{G^2}{rpq} \\
\text{SSB within A} &= \sum_{i=1}^{p}\sum_{j=1}^{q}\frac{T_{ij}^2}{r} - \sum_{i=1}^{p}\frac{T_{i+}^2}{rq} \\
\text{SS Error} &= \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} y_{ijk}^2 - \sum_{i=1}^{p}\sum_{j=1}^{q}\frac{T_{ij}^2}{r}
\end{aligned}$$

- Under the model S we have

$$
\begin{aligned}
E(Y_{ijk}^2) &= \eta_{ij}^2 + \sigma^2 \\
E\left(\frac{T_{ij}^2}{r}\right) &= \frac{1}{r}\left\{[E(T_{ij})]^2 + \text{var}\,(T_{ij})\right\} \\
&= r\eta_{ij}^2 + \sigma^2 \\
E\left(\frac{T_{i+}^2}{rq}\right) &= \frac{1}{r}\left\{[E(T_{i+})]^2 + \text{var}\,(T_{i+})\right\} \\
&= \frac{r}{q}\left(\sum_{j=1}^{q}\eta_{ij}\right)^2 + \sigma^2 \\
E\left(\frac{G^2}{rpq}\right) &= \frac{1}{rpq}\left\{\left(r\sum_{i=1}^{p}\sum_{j=1}^{q}\eta_{ij}\right)^2 + \text{var}\,(G)\right\} \\
&= \frac{r}{pq}\left(\sum_{i=1}^{p}\sum_{j=1}^{q}\eta_{ij}\right)^2 + \sigma^2
\end{aligned}
$$

- Thus we have

$$\begin{aligned}
E(\text{SSE}) &= (r-1)\sigma^2 \\
E(\text{SSB within A}) &= r\sum_{i=1}^{p}\sum_{j=1}^{q}\eta_{ij}^2 + pq\sigma^2 - \frac{r}{q}\sum_{i=1}^{p}\left(\sum_{j=1}^{q}\eta_{ij}\right)^2 + p\sigma^2 \\
&= p(q-1)\sigma^2 + r\sum_{i=1}^{p}\sum_{j=1}^{q}(\eta_{ij}-\bar{\eta}_{i+})^2 \\
E(\text{SSA}) &= \frac{r}{q}\sum_{i=1}^{p}\left(\sum_{j=1}^{q}\eta_{ij}\right)^2 + p\sigma^2 - \frac{r}{pq}\left(\sum_{i=1}^{p}\sum_{j=1}^{q}\eta_{ij}\right)^2 + \sigma^2 \\
&= (p-1)\sigma^2 + rq\sum_{i=1}^{p}(\bar{\eta}_{i+}-\bar{\eta}_{++})^2
\end{aligned}$$

- Hence we have the ANOVA table:

Source	d.f.	Expected Mean Square
Mean	1	$\sigma^2 + rpq\bar{\eta}_{++}^2$
A	$p-1$	$\sigma^2 + rq\sum_{i=1}^{p}(\bar{\eta}_{i+}-\bar{\eta}_{++})^2/(p-1)$
B within A	$p(q-1)$	$\sigma^2 + r\sum_{i=1}^{p}\sum_{j=1}^{q}(\eta_{ij}-\bar{\eta}_{i+})^2/p(q-1)$
Error	$(r-1)pq$	σ^2

4.3.2 Case 2: A and B both random

In this case we assume that

$$E(\eta_{ij}|S) = \mu \ \text{ and } \ \text{cov}\,(\eta_{ij}, \eta_{i'j'}) = \begin{cases} \sigma_1^2 + \sigma_2^2 & (i', j') = (i, j) \\ \sigma_1^2 & i' = i,\ j') \neq j \\ 0 & \text{otherwise} \end{cases}$$

and we have

$$\begin{aligned} E(\eta_{ij}^2) &= \mu^2 + \sigma_1^2 + \sigma_2^2 \\ E(\bar{\eta}_{i+}^2) &= \frac{1}{q^2} E\left(\sum_{j=1}^{q} \eta_{ij}\right)^2 \\ &= \mu^2 + \sigma_1^2 + \frac{\sigma_2^2}{q} \\ E(\bar{\eta}_{++}^2) &= \frac{1}{(pq)^2} E\left(\sum_{i=1}^{p} \sum_{j=1}^{q} \eta_{ij}\right)^2 \\ &= \mu^2 + \frac{1}{p}\sigma_1^2 + \frac{\sigma_2^2}{pq} \end{aligned}$$

It follows that

$$
\begin{aligned}
E(\text{SSB within A}) &= p(q-1)\sigma^2 + rp(q-1)\sigma_2^2 \\
E(\text{SSA}) &= (p-1)\sigma^2 + r(p-1)\sigma_2^2 + rq(p-1)\sigma_1^2
\end{aligned}
$$

which leads to the analysis of variance table:

Source	d.f.	Expected Mean Square
Mean	1	$\sigma^2 + rq\sigma_1^2 + r\sigma_2^2 + rpq\mu^2$
A	$p-1$	$\sigma^2 + r\sigma_2^2 + rq\sigma_1^2$
B within A	$p(q-1)$	$\sigma^2 + r\sigma_2^2$
Error	$pq(r-1)$	σ^2

4.3.3 Case 3: A fixed and B random

Here we assume that

$$E(\eta_{ij}|S) = \mu_i \text{ and } \operatorname{cov}(\eta_{ij}, \eta_{i'j'}) = \begin{cases} \sigma_2^2 & (i', j') = (i, j) \\ 0 & \text{otherwise} \end{cases}$$

In this case

$$\begin{aligned}
E(\eta_{ij}^2) &= \mu_i^2 + \sigma_2^2 \\
E(\bar{\eta}_{i+}^2) &= \frac{1}{q^2} E\left(\sum_{j=1}^{q} \eta_{ij}\right)^2 \\
&= \mu_i^2 + \frac{\sigma_2^2}{q} \\
E(\bar{\eta}_{++}^2) &= \frac{1}{(pq)^2} E\left(\sum_{i=1}^{p}\sum_{j=1}^{q} \eta_{ij}\right)^2 \\
&= \frac{1}{p^2}\left(\sum_{i=1}^{p} \mu_i\right)^2 + \frac{\sigma_2^2}{pq}
\end{aligned}$$

It follows that

$$\begin{aligned} E(\text{SSB within A}) &= p(q-1)\sigma^2 + rp(q-1)\sigma_2^2 \\ E(\text{SSA}) &= (p-1)\sigma^2 + r(p-1)\sigma_2^2 + rq\sum_{i=1}^{p}(\mu_i - \bar{\mu})^2 \end{aligned}$$

Thus the analysis of variance table is

Source	d.f.	Expected Mean Square
Mean	1	$\sigma^2 + rq\sigma_1^2 + r\sigma_2^2 + \frac{rq}{p}\left(\sum_{i=1}^{p}\mu_i\right)^2$
A	$p-1$	$\sigma^2 + r\sigma_2^2 + rq\sum_{i=1}^{p}(\mu_i - \bar{\mu})^2(p-1)$
B within A	$p(q-1)$	$\sigma^2 + r\sigma_2^2$
Error	$pq(r-1)$	σ^2

4.4 Non Additivity

In a two way analysis of variance with interaction present and all $n_{ij} = 1$ it is clear that the error sum of squares is 0 and the presence of interaction cannot be detected.

- In Tukey (Biometrics, 1949a - One Degree of Freedom for Non-Additivity) a test for the presence of interaction of a certain type was developed.
- Tukey assumed the model

  $$E(Y_{ij}) = \mu + \rho_i + \gamma_j + \lambda\rho_i\gamma_j$$

  which is a non-linear model.
- A test for $\lambda = 0$ tests whether interaction of the form $\lambda\rho_i\gamma_j$ is present.
- If μ, ρ_i and γ_j are known the least squares estimate of λ is found by minimizing

  $$\sum_{i=1}^{p}\sum_{j=1}^{q}(y_{ij} - \mu - \rho_i - \gamma_j - \lambda\rho_i\gamma_j)^2$$
- Differentiating with respect to λ yields

  $$\sum_{i=1}^{p}\sum_{j=1}^{q}(y_{ij} - \mu - \rho_i - \gamma_j - \lambda\rho_i\gamma_j)(-2\rho_i\gamma_j)$$

- Equating to zero and solving for λ yields

$$\widehat{\lambda} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{q}(y_{ij} - \mu - \rho_i - \gamma_j - \lambda\rho_i\gamma_j)\rho_i\gamma_j}{\sum_{i=1}^{p}\sum_{j=1}^{q}\rho_i^2\gamma_j^2}$$

- If we replace μ by $m = \bar{y}_{++}$, ρ_i by $a_i = \bar{y}_{i+} - \bar{y}_{++}$ and γ_j by $c_j = \bar{y}_{+j} - \bar{y}_{++}$ we find that an estimate of λ is

$$\begin{aligned}\widehat{\lambda} &= \frac{\sum_{i=1}^{p}\sum_{j=1}^{q}(y_{ij} - \bar{y}_{i+} - \bar{y}_{+j} + \bar{y}_{++})a_i c_j}{\sum_{i=1}^{p}\sum_{j=1}^{q} a_i^2 c_j^2} \\ &= \frac{\sum_{i=1}^{p}\sum_{j=1}^{q} a_i c_j y_{ij}}{\sum_{i=1}^{p}\sum_{j=1}^{q} a_i^2 c_j^2}\end{aligned}$$

- If we define $k_{ij} = a_i c_j$ then

$$\widehat{\lambda} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{q} k_{ij} y_{ij}}{\sum_{i=1}^{p}\sum_{j=1}^{q} k_{ij}^2}$$

- Note that $\sum_{i=1}^{p}\sum_{j=1}^{q} k_{ij} = 0$ so that $\sum_{i=1}^{p}\sum_{j=1}^{q} k_{ij} y_{ij}$ is a treatment contrast (but with coefficients that depend on the data!)

- The sum of squares for such a contrast is formally defined as

$$\text{SSNA} = \frac{\left(\sum_{i=1}^{p}\sum_{j=1}^{q} k_{ij} y_{ij}\right)^2}{\sum_{i=1}^{p}\sum_{j=1}^{q} k_{ij}^2}$$

and is called the non-additivity component of the error sum of squares.

- The non-additivity model is tested by an F test of the form

$$\frac{\text{SSNA}/1}{(\text{SSE-SSNA})/[(p-1)(q-1)-1]}$$

  which under $H_0 : \lambda = 0$ has an F distribution with 1 and $(p-1)(q-1)-1$ degrees of freedom.

- The distributional properties are outlined by Graybill; Linear Statistical Models; McGraw Hill 1961).

- Since the original paper many extensions have been developed.

- It has also been shown that the test is a score test of $H_0 : \lambda = 0$ under normality assumptions.

- Note that the test for non-additivity is not a test of the presence of interaction; only of a particular type of interaction.

- Consider the following data set:

	B1	B2	B3	Total
A1	25	10	0	35
A2	5	30	50	85
Total	30	40	50	120

- Then

$$\begin{aligned} a_1 &= -\frac{25}{3} \\ a_2 &= +\frac{25}{3} \\ b_1 &= -5 \\ b_2 &= 0 \\ b_3 &= +5 \end{aligned}$$

- The "contrast" for non-additivity is given by

$$\begin{aligned} &\left(-\frac{25}{3}\right)(-5)(25) + \left(-\frac{25}{3}\right)(0)(10) + \left(-\frac{25}{3}\right)(+5)(0) \quad + \\ &\left(+\frac{25}{3}\right)(-5)(5) + \left(+\frac{25}{3}\right)(0)(30) + \left(+\frac{25}{3}\right)(+5)(50) \quad = \quad \frac{625 \times 14}{3} \end{aligned}$$

- Thus the sum of squares for non-additivity is

$$\frac{\left(\frac{625\times14}{3}\right)^2}{4 \times \left(\frac{125}{3}\right)^2} = 25 \times 49$$

- The interaction sum of squares is

$$\text{SSInt} = \frac{10100}{6}$$

- Thus

$$\text{SSInt} - \text{SSNA} = \frac{10100}{6} - 25 \times 49 = \frac{2750}{6}$$

and the F statistic is

$$\frac{25 \times 49 \times 6}{2750} = 2.67$$

- Since $F_{.95}(1,1) = 161$ we do not reject $H_o : \lambda = 0$ even though there is clear evidence of interaction.

4.5 Higher Way Models

With current software models involving 3, 4 and more factors can be easily fitted to observed data.

- What is difficult is the interpretation of interaction of three or more factors.
- A three factor interaction $A \times B \times C$ means that the two factor interaction, $A \times B$ changes as the levels of the third factor C change.
  - Thus if A represents levels of a drug, B represents disease severity and C represents gender, a three factor interaction means that the effect of drug dosage changes with disease severity ($A \times B$ interaction) and that this interaction is different for males than for females.
- In general a w order interaction, $w \geq 3$, means that the $w - 1$ order interaction changes as the level of the remaining factor changes.
- Interpreting these interactions requires a great understanding of the subject matter under investigation.

4.6 Generalizations

4.6.1 Conditional Error

It is easy to generalize the principle of conditional error - suitably extended it underlies all linear model procedures. The basic ideas can be illustrated for a model of the form

$$E(\mathbf{Y}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{X}_3\boldsymbol{\beta}_3 = \mathbf{X}\boldsymbol{\beta} \text{ and } \operatorname{var}(\mathbf{Y}) = \sigma^2\mathbf{I}$$

where

$$\begin{aligned} \mathbf{X} &= [\mathbf{X}_1\ \mathbf{X}_2\ \mathbf{X}_3] \text{ is of rank r} \\ \mathbf{X}_1 & \quad \text{is } n \times p_1 \\ \mathbf{X}_2 & \quad \text{is } n \times p_2 \\ \mathbf{X}_3 & \quad \text{is } n \times p_3 \end{aligned}$$

The least squares equations in partitioned form are

$$\begin{bmatrix} \mathbf{X}_1^T\mathbf{X}_1 & \mathbf{X}_1^T\mathbf{X}_2 & \mathbf{X}_1^T\mathbf{X}_3 \\ \mathbf{X}_2^T\mathbf{X}_1 & \mathbf{X}_2^T\mathbf{X}_2 & \mathbf{X}_2^T\mathbf{X}_3 \\ \mathbf{X}_3^T\mathbf{X}_1 & \mathbf{X}_3^T\mathbf{X}_2 & \mathbf{X}_3^T\mathbf{X}_3 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1^T\mathbf{y} \\ \mathbf{X}_2^T\mathbf{y} \\ \mathbf{X}_3^T\mathbf{y} \end{bmatrix}$$

Multiplying both sides by the non-singular matrix

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ -\mathbf{X}_2^T\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^- & \mathbf{I} & \mathbf{0} \\ -\mathbf{X}_3^T\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^- & \mathbf{0} & \mathbf{I} \end{bmatrix}$$

yields the equivalent set of equations.

$$\begin{bmatrix} \mathbf{X}_1^T\mathbf{X}_1 & \mathbf{X}_1^T\mathbf{X}_2 & \mathbf{X}_1^T\mathbf{X}_3 \\ \mathbf{0} & \mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2 & \mathbf{X}_2^T\mathbf{D}_2\mathbf{X}_3 \\ \mathbf{0} & \mathbf{X}_3^T\mathbf{D}_1\mathbf{X}_2 & \mathbf{X}_3^T\mathbf{D}_1\mathbf{X}_3 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1^T\mathbf{y} \\ \mathbf{X}_2^T\mathbf{D}_1\mathbf{y} \\ \mathbf{X}_3^T\mathbf{D}_1\mathbf{y} \end{bmatrix}$$

where $\mathbf{D}_1 = \mathbf{I} - \mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^-\mathbf{X}_1^T$

Multiplying both sides by the non-singular matrix

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{X}_3^T\mathbf{D}_1\mathbf{X}_2(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^- & \mathbf{I} \end{bmatrix}$$

yields the equivalent set of equations:

$$\begin{bmatrix} \mathbf{X}_1^T\mathbf{X}_1 & \mathbf{X}_1^T\mathbf{X}_2 & \mathbf{X}_1^T\mathbf{X}_3 \\ \mathbf{0} & \mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2 & \mathbf{X}_2^T\mathbf{D}_2\mathbf{X}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1^T\mathbf{y} \\ \mathbf{X}_2^T\mathbf{D}_1\mathbf{y} \\ \mathbf{X}_3^T\mathbf{D}_{12}\mathbf{y} \end{bmatrix}$$

where $\mathbf{D}_{12} = \mathbf{D}_1 - \mathbf{D}_1\mathbf{X}_2(\mathbf{X}^T\mathbf{D}_1\mathbf{X})^{-}\mathbf{X}^T\mathbf{D}_1$.

- Note that $\mathbf{D}_1$ and $\mathbf{D}_{12}$ are unique, symmetric, idempotent and that

$$\mathbf{D}_1\mathbf{D}_{12} = \mathbf{D}_{12}\mathbf{D}_1 = \mathbf{D}_{12}$$

- Clearly

$$\mathbf{b}_3 = (\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3)^{-}\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{y} + \left[\mathbf{I} - (\mathbf{X}^T\mathbf{D}_{12}\mathbf{X}_3)^{-}\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3\right]\mathbf{z}_3$$

and thus

$$\begin{aligned} (\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)\mathbf{b}_2 &= \mathbf{X}_2^T\mathbf{D}_1\mathbf{y} - \mathbf{X}_3^T\mathbf{D}_1\mathbf{b}_3 \\ (\mathbf{X}_1^T\mathbf{X}_1)\mathbf{b}_1 &= \mathbf{X}_1^T\mathbf{y} - \mathbf{X}_2^T\mathbf{b}_2 - \mathbf{X}_3^T\mathbf{b}_3 \end{aligned}$$

- It follows that

$$\begin{aligned} \mathbf{b}_2 &= (\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^{-}\left[\mathbf{X}_2^T\mathbf{D}_1\mathbf{y} - \mathbf{X}_3^T\mathbf{D}_1\mathbf{b}_3\right] + \left[\mathbf{I} - (\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^{-}\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2\right]\mathbf{z}_2 \\ \mathbf{b}_1 &= (\mathbf{X}_1^T\mathbf{X}_1)^{-}\left[\mathbf{X}_1^T\mathbf{y} - \mathbf{X}_1^T\mathbf{X}_2\mathbf{b}_2 - \mathbf{X}_1^T\mathbf{X}_3\mathbf{b}_3\right] + \left[\mathbf{I} - (\mathbf{X}_1^T\mathbf{X}_1)^{-}\mathbf{X}_1^T\mathbf{X}_1\right]\mathbf{z}_1 \end{aligned}$$

- We note that

$$\begin{aligned}
\text{SSE} &= \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\mathbf{b} \\
&= \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}_1\mathbf{b}_1 - \mathbf{y}^T\mathbf{X}_2\mathbf{b}_2 - \mathbf{y}^T\mathbf{X}_3\mathbf{b}_3 \\
&= \mathbf{y}^T\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^- \left[\mathbf{X}_1^T\mathbf{y} - \mathbf{X}_1^T\mathbf{X}_2\mathbf{b}_2 - -\mathbf{X}_1^T\mathbf{X}_3\mathbf{b}_3\right] - \mathbf{y}^T\mathbf{X}_2\mathbf{b}_2 - \mathbf{y}^T\mathbf{X}_3\mathbf{b}_3 \\
&= \mathbf{y}^T\mathbf{D}_1\mathbf{y} - \mathbf{y}^T\mathbf{D}_1\mathbf{X}_2\mathbf{b}_2 - \mathbf{D}_1\mathbf{X}_3\mathbf{b}_3 \\
&= \mathbf{y}^T\mathbf{D}_1\mathbf{y} - \mathbf{y}^T\mathbf{D}_1\mathbf{X}_2(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^- \left[\mathbf{X}_2^T\mathbf{D}_1\mathbf{y} - \mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_3\mathbf{b}_3\right] - \mathbf{D}_1\mathbf{X}_3\mathbf{b}_3 \\
&= \mathbf{y}^T\mathbf{D}_1\mathbf{y} - \mathbf{y}^T\mathbf{D}_1\mathbf{X}_2(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^-\mathbf{X}_2^T\mathbf{D}_1\mathbf{y} \\
&\quad -\mathbf{y}^T\mathbf{D}_{12}\mathbf{X}_3(\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3)^-\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{y}
\end{aligned}$$

Suppose now that $\boldsymbol{\beta}_3 = \mathbf{0}$ is equivalent to a hypothesis about a set of linearly estimable functions.

- Setting $\mathbf{b}_3 = \mathbf{0}$ yields

$$\text{SSCE} = \mathbf{y}^T\mathbf{D}_1\mathbf{y} - \mathbf{y}^T\mathbf{D}_1\mathbf{X}_2(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^-\mathbf{X}_2^T\mathbf{D}_1\mathbf{y}$$

- The difference between these two error sum of squares is thus

$$\mathbf{y}^T\mathbf{D}_{12}\mathbf{X}_3(\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3)^-\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{y} = \mathbf{b}_3^T\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{y}$$

  which we call the sum of squares due to $\mathbf{X}_3$ adjusted for $\mathbf{X}_1$ and $\mathbf{X}_2$ and write as

$$\text{SS}\,(\mathbf{X}_3|\mathbf{X}_1, \mathbf{X}_2)$$

- We also note that

$$\begin{aligned}
\text{SS}\ (\mathbf{X}_2|\mathbf{X}_1) &= \mathbf{y}^T\mathbf{D}_1\mathbf{X}_2(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^-\mathbf{X}_2^T\mathbf{D}_1\mathbf{y} \\
&= \mathbf{b}_2^T\mathbf{X}_2^T\mathbf{D}_1\mathbf{y} \\
\text{SS}\ (\mathbf{X}_1) &= \mathbf{y}^T\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^-\mathbf{X}_1^T\mathbf{y} \\
&= \mathbf{b}_1^T\mathbf{X}_1^T\mathbf{y}
\end{aligned}$$

- We thus have the ANOVA table:

Source	d.f.	Sum of Squares (Deviance)
SS $(\mathbf{X}_1)$	rank $(\mathbf{X}_1^T\mathbf{X}_1)$	$\mathbf{b}_1^T\mathbf{X}_1^T\mathbf{y}$
SS $(\mathbf{X}_2\|\mathbf{X}_1)$	rank $(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)$	$\mathbf{b}_2^T\mathbf{X}_2^T\mathbf{D}_1\mathbf{y}$
SS $(\mathbf{X}_3\|\mathbf{X}_1, \mathbf{X}_2)$	rank $(\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3)$	$\mathbf{b}_3^T\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{y}$
SSE	rank $(\mathbf{D}_{123})$	$\mathbf{y}^T\mathbf{D}_{123}\mathbf{y}$
Total	n	$\mathbf{y}^T\mathbf{y}$

where

$$\begin{aligned}
\mathbf{D}_{123} &= \mathbf{D}_1 - \mathbf{D}_1\mathbf{X}_2(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^-\mathbf{b}_2^T\mathbf{X}_2^T\mathbf{D}_1 - \mathbf{D}_{12}\mathbf{X}_3(\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3)^-\mathbf{X}_3^T\mathbf{D}_{12} \\
&= \mathbf{D}_{12} - \mathbf{D}_{12}\mathbf{X}_3(\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3)^-\mathbf{X}_3^T\mathbf{D}_{12}
\end{aligned}$$

It is useful to record the expected values of the sums of squares in the ANOVA table.

- Recall that for any sum of squares we have:

$$\begin{aligned} E(\text{SS}) &= E(\mathbf{Y}^T\mathbf{A}\mathbf{Y}) \\ &= E[\text{tr}\,(\mathbf{A}\mathbf{Y}\mathbf{Y}^T)] \\ &= \text{tr}\,(\mathbf{A}E[\mathbf{Y}\mathbf{Y}^T]) \\ &= \text{tr}\,(\mathbf{A}[\boldsymbol{\mu}\boldsymbol{\mu}^T + \text{var}\,(\mathbf{Y})]) \\ &= \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{A}\mathbf{X}\boldsymbol{\beta} + \text{tr}\,(\mathbf{A})\sigma^2 \end{aligned}$$

- For SSE we thus have

$$\begin{aligned} \mathbf{X}^T\mathbf{A}\mathbf{X} &= \mathbf{X}^T[\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T]\mathbf{X} \\ &= \mathbf{0} \\ \text{tr}\,(\mathbf{A}) &= n - r \end{aligned}$$

- For SS $(\mathbf{X}_3|\mathbf{X}_1, \mathbf{X}_2)$ we have

$$\begin{aligned}
\mathbf{X}^T\mathbf{A}\mathbf{X} &= \mathbf{X}^T\mathbf{D}_{12} - \mathbf{D}_{12}\mathbf{X}_3(\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3)^{-}\mathbf{X}_3^T\mathbf{D}_{12}[\mathbf{X}_1\ \mathbf{X}_2\ \mathbf{X}_3] \\
&= \begin{bmatrix} \mathbf{X}_1^T \\ \mathbf{X}_2^T \\ \mathbf{X}_3^T \end{bmatrix} [\mathbf{0}\ \mathbf{0}\ \mathbf{D}_{12}\mathbf{X}_3] \\
&= \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\text{tr}\,(\mathbf{A}) &= \text{tr}\,(\mathbf{D}_{12}\mathbf{X}_3(\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3)^{-}\mathbf{X}_3^T\mathbf{D}_{12}) \\
&= \text{rank}\,(\mathbf{D}_{12}\mathbf{X}_3) \\
&= \text{rank}\,(\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3)
\end{aligned}$$

- It follows that

$$E(\text{SS}(\mathbf{X}_3|\mathbf{X}_1, \mathbf{X}_2)) = \boldsymbol{\beta}^T\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3\boldsymbol{\beta} + \sigma^2\text{rank}\,(\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3)$$

- For $\text{SS}(\mathbf{X}_2|\mathbf{X}_1)$ we have

$$\begin{aligned}
\mathbf{X}^T\mathbf{A}\mathbf{X} &= \mathbf{X}^T\mathbf{D}_1\mathbf{X}_2(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^-\mathbf{X}_2^T\mathbf{D}_1[\mathbf{X}_1\ \mathbf{X}_2\ \mathbf{X}_3] \\
&= \begin{bmatrix} \mathbf{X}_1^T \\ \mathbf{X}_2^T \\ \mathbf{X}_3^T \end{bmatrix} [\mathbf{0}\ \mathbf{D}_1\mathbf{X}_2\ \mathbf{D}_1\mathbf{X}_2(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^-\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_3] \\
&= \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2 & \mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_3 \\ \mathbf{0} & \mathbf{X}_3^T\mathbf{D}_1\mathbf{X}_2 & \mathbf{X}_3^T\mathbf{D}_1\mathbf{X}_2(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^-\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_3 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\text{tr}\,(\mathbf{D}_1\mathbf{X}_2(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^-\mathbf{X}_2^T\mathbf{D}_1) &= \text{rank}\,(\mathbf{D}_1\mathbf{X}_2) \\
&= \text{rank}\,(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)
\end{aligned}$$

- It follows that

$$\begin{aligned}
E(\text{SS}(\mathbf{X}_2|\mathbf{X}_1)) &= \boldsymbol{\beta}_2^T\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\beta}_3^T\mathbf{X}_3^T\mathbf{D}_1\mathbf{X}_2(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^-\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_3\boldsymbol{\beta}_3 \\
&\quad +2\boldsymbol{\beta}_2^T\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_3\boldsymbol{\beta}_3 + \sigma^2\text{rank}\,(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)
\end{aligned}$$

- For $\mathrm{SS}(\mathbf{X}_1)$ we have

$$\begin{aligned}
\mathbf{X}^T\mathbf{A}\mathbf{X} &= \mathbf{X}^T\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^-\mathbf{X}_1^T[\mathbf{X}_1\ \mathbf{X}_2\ \mathbf{X}_3] \\
&= \begin{bmatrix} \mathbf{X}_1^T \\ \mathbf{X}_2^T \\ \mathbf{X}_3^T \end{bmatrix} [\mathbf{X}_1\ \mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^-\mathbf{X}_1^T\mathbf{X}_2\ \mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^-\mathbf{X}_1^T\mathbf{X}_3 \\
&= \begin{bmatrix} \mathbf{X}_1^T\mathbf{X}_1 & \mathbf{X}_1^T\mathbf{X}_2 & \mathbf{X}_1^T\mathbf{X}_3 \\ \mathbf{X}_2^T\mathbf{X}_1 & \mathbf{X}_2^T\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^-\mathbf{X}_1^T\mathbf{X}_2 & \mathbf{X}_2^T\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^-\mathbf{X}_1^T\mathbf{X}_3 \\ \mathbf{X}_3^T\mathbf{X}_1 & \mathbf{X}_3^T\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^-\mathbf{X}_1^T\mathbf{X}_2 & \mathbf{X}_3^T\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^-\mathbf{X}_1^T\mathbf{X}_3 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\mathrm{tr}\,(\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^-\mathbf{X}_1^T) &= \mathrm{rank}\,(\mathbf{X}_1) \\
&= \mathrm{rank}\,(\mathbf{X}_1^T\mathbf{X}_1)
\end{aligned}$$

- It follows that

$$\begin{aligned}
E(\mathrm{SS}(\mathbf{X}_1)) &= \boldsymbol{\beta}_1^T\mathbf{X}_1^T\mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2^T\mathbf{X}_2^T\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^-\mathbf{X}_1^T\mathbf{X}_2\boldsymbol{\beta}_2 \\
&\quad +\boldsymbol{\beta}_3^T\mathbf{X}_2^T\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^-\mathbf{X}_1^T\mathbf{X}_2\boldsymbol{\beta}_3 + 2\boldsymbol{\beta}_1^T\mathbf{X}_1^T\mathbf{X}_2\boldsymbol{\beta}_2 \\
&\quad +2\boldsymbol{\beta}_1^T\mathbf{X}_1^T\mathbf{X}_3\boldsymbol{\beta}_3 + 2\boldsymbol{\beta}_2^T\mathbf{X}_2^T\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^-\mathbf{X}_1^T\mathbf{X}_3\boldsymbol{\beta}_3 \\
&\quad +\sigma^2\mathrm{rank}\,(\mathbf{X}_1^T\mathbf{X}_1)
\end{aligned}$$

It is clear that we may extend the development to an arbitrary number of sets of covariates. Thus if

$$E(\mathbf{Y}) = \sum_{j=1}^{N} \mathbf{X}_j \boldsymbol{\beta}_j$$

- Define

$$\begin{aligned}
\mathbf{D}_0 &= \mathbf{I} \\
\mathbf{D}_1 &= \mathbf{I} - \mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^-\mathbf{X}_1^T \\
\mathbf{D}_{12} &= \mathbf{D}_1 - \mathbf{D}_1\mathbf{X}_2(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^-\mathbf{X}_2^T\mathbf{D}_1 \\
\mathbf{D}_{123} &= \mathbf{D}_{12} - \mathbf{D}_{12}\mathbf{X}_3(\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3)^-\mathbf{X}_3^T\mathbf{D}_{12} \\
\vdots &= \vdots \\
\mathbf{D}_{\pi(k)} &= \mathbf{D}_{\pi(k-1)} - \mathbf{D}_{\pi(k-1)}\mathbf{X}_k(\mathbf{X}_k^T\mathbf{D}_{\pi(k-1)}\mathbf{X}_k)^-\mathbf{X}_k^T\mathbf{D}_{\pi(k-1)}
\end{aligned}$$

  for $k = 1, 2, \ldots, N$ where $\pi(k) = 12\cdots k$ and $\pi(0) = 0$.

- Now define

$$\begin{aligned}
\mathbf{P}_1 &= \mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^-\mathbf{X}_1^T \\
&= \mathbf{D}_0 - \mathbf{D}_1 \\
\mathbf{P}_2 &= \mathbf{D}_1\mathbf{X}_2(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^-\mathbf{X}_2^T\mathbf{D}_1 \\
&= \mathbf{D}_1 - \mathbf{D}_{12} \\
\vdots &= \vdots \\
\mathbf{P}_k &= \mathbf{D}_{\pi(k-1)}\mathbf{X}_k(\mathbf{X}_k^T\mathbf{D}_{\pi(k-1)}\mathbf{X}_k)^-\mathbf{X}_k^T\mathbf{D}_{\pi(k-1)} \\
&= \mathbf{D}_{\pi(k-1)} - \mathbf{D}_{\pi(k)}
\end{aligned}$$

- We then have

$$\text{SS(Model)} = \sum_{k=1}^{N} \mathbf{y}^T \mathbf{P}_k \mathbf{y}$$

i.e.

$$\text{SS}\left(\mathbf{X}_k | \mathbf{X}_{k-1}, \ldots, \mathbf{X}_0\right) = \mathbf{y}^T \mathbf{P}_k \mathbf{y} \;\; \text{for} \;\; k = 1, 2, \ldots, N$$

where $\mathbf{X}_0 = \mathbf{0}$.

- We note that

$$\mathbf{P}_k \mathbf{X}_j = \mathbf{0}$$

for $j = 0, 1, 2, \ldots, k-1$.

- Since $E(\mathbf{Y}) = \sum_{j=1}^{N} \mathbf{X}_j \boldsymbol{\beta}_j$ it follows that

$$\begin{aligned}
E(\mathbf{Y}^T \mathbf{P}_k \mathbf{Y}) &= E(\text{tr}\,[\mathbf{Y}^T \mathbf{P}_k \mathbf{Y}]) \\
&= E[\text{tr}\,(\mathbf{P}_k \mathbf{Y} \mathbf{Y}^T)] \\
&= \text{tr}\,(\mathbf{P}_k E[\mathbf{Y} \mathbf{Y}^T]) \\
&= \text{tr}\,(\mathbf{P}_k \{\sigma^2 \mathbf{I} + E[\mathbf{Y}] E[\mathbf{Y}^T]\}) \\
&= \text{rank}\,(\mathbf{P}_k) + E[\mathbf{Y}^T] \mathbf{P}_k E[\mathbf{Y}] \\
&= \text{rank}\,(\mathbf{P}_k) + E[\mathbf{P}_k \mathbf{Y}]^T E[\mathbf{P}_k \mathbf{Y}]
\end{aligned}$$

- Now

$$\begin{aligned}\mathbf{P}_k E(\mathbf{Y}) &= \mathbf{P}_k \left(\sum_{j=1}^{N} \mathbf{X}_j \boldsymbol{\beta}_j \right) \\ &= \sum_{j=1}^{N} \mathbf{P}_k \mathbf{X}_j \boldsymbol{\beta}_j \\ &= \sum_{j=k}^{N} \mathbf{P}_k \mathbf{X}_j \boldsymbol{\beta}_j\end{aligned}$$

so that

$$\begin{aligned}E(\mathbf{Y}^T \mathbf{P}_k \mathbf{Y}) &= \sigma^2 \text{rank}\,(\mathbf{P}_k) + \left(\sum_{j=k}^{N} \mathbf{P}_k \mathbf{X}_j \boldsymbol{\beta}_j \right)^T \left(\sum_{j=k}^{N} \mathbf{P}_k \mathbf{X}_j \boldsymbol{\beta}_j \right) \\ &= \sigma^2 \text{rank}\,(\mathbf{P}_k) + \sum_{j=k}^{N} \sum_{j'=k}^{N} \boldsymbol{\beta}_{j'}^T \mathbf{X}_{j'}^T \mathbf{P}_k \mathbf{X}_j \boldsymbol{\beta}_j\end{aligned}$$

Depending on the assumptions about the stochastic structure of the $\boldsymbol{\beta}_j$ we get various types of models. Formally we have:

- **Fixed Effect Models:**

$$E(\boldsymbol{\beta}_j) = \boldsymbol{\beta}_j \ ; \ \text{var}\,(\boldsymbol{\beta}_j) = \mathbf{0}$$

  for $j = 1, 2, \ldots, N$

- **Random Effects Models:**

$$E(\boldsymbol{\beta}_j) = \mathbf{0} \ ; \ \text{var}\,(\boldsymbol{\beta}_j) = \sigma_j^2 \mathbf{I}$$

  for $j = 2, \ldots, N$ and $\mathbf{X}_1 = \mathbf{1}$, $\boldsymbol{\beta}_1 = \mu$.

- **Mixed Models:**
  - $E(\boldsymbol{\beta}_j) = \boldsymbol{\beta}_j \ ; \ \text{var}\,(\boldsymbol{\beta}_j) = \mathbf{0}$ for $j \in F$
  - $E(\boldsymbol{\beta}_j) = \mathbf{0} \ ; \ \text{var}\,(\boldsymbol{\beta}_j) = \sigma_j^2 \mathbf{I}$ for $j \in R$
  - $\text{cov}\,(\boldsymbol{\beta}_{j'}, \boldsymbol{\beta}_j) = \mathbf{0}$ for $j', j \in R; \ j' \neq j$
  - $F \subset \{1, 2, \ldots, N\} \ \ R \subset \{1, 2, \ldots, N\} \ \ F \cap R = \emptyset$

4.7 Balance, Orthogonality and Additivity

Consider, for simplicity, the linear model with three effects, i.e.

$$E(\mathbf{Y}) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{X}_3\boldsymbol{\beta}_3$$

- Of interest is a condition under which the effect of the covariates $\mathbf{X}_3$ is the same whether or not the covariates $\mathbf{X}_2$ are in the model.
- Recall that

  $$\text{SS }(\mathbf{X}_3|\mathbf{X}_1) = \mathbf{y}^T\mathbf{D}_1\mathbf{X}_3(\mathbf{X}_3^T\mathbf{D}_1\mathbf{X}_3)^-\mathbf{X}_3^T\mathbf{D}_1\mathbf{y}$$

  and

  $$\text{SS }(\mathbf{X}_3|\mathbf{X}_1, \mathbf{X}_2) = \mathbf{y}^T\mathbf{D}_{12}\mathbf{X}_3(\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3)^-\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{y}$$

  where

  $$\begin{aligned} \mathbf{D}_1 &= \mathbf{I} - \mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^-\mathbf{X}_1^T \\ \mathbf{D}_{12} &= \mathbf{D}_1 - \mathbf{D}_1\mathbf{X}_2(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^-\mathbf{X}_2^T\mathbf{D}_1 \end{aligned}$$

- Clearly these two sums of squares are identical if and only if

$$\mathbf{D}_1\mathbf{X}_3^T(\mathbf{X}_3\mathbf{D}_1\mathbf{X}_3)^-\mathbf{X}_3^T\mathbf{D}_1 = \mathbf{D}_{12}\mathbf{X}_3(\mathbf{X}_3^T\mathbf{D}_{12}\mathbf{X}_3)^-\mathbf{X}_3^T\mathbf{D}_{12}$$

- This implies that (multiply on the left by $\mathbf{X}_3^T$).

$$\begin{aligned}\mathbf{X}_3^T\mathbf{D}_1 &= \mathbf{X}_3^T\mathbf{D}_{12} \\ &= \mathbf{X}_3^T\mathbf{D}_1 - \mathbf{X}_3^T\mathbf{D}_1\mathbf{X}_2(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^-\mathbf{X}_2^T\mathbf{D}_1\end{aligned}$$

  or

$$\mathbf{X}_3^T\mathbf{D}_1\mathbf{X}_2(\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_2)^-\mathbf{X}_2^T\mathbf{D}_1 = \mathbf{0}$$

- Thus equality of the sum of squares implies (mutiply on the right by $\mathbf{X}_2$)

$$\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_3 = \mathbf{0}$$

- Conversely if $\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_3 = \mathbf{0}$ then

$$\mathbf{D}_{12}\mathbf{X}_3 = \mathbf{D}_1\mathbf{X}_3$$

  and the two sums of squares are equal.

- Thus the two sums of squares are equal if and only if

$$\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_3 = \mathbf{0}$$

- In such a case we say that $\mathbf{X}_2$ and $\mathbf{X}_3$ are **balanced** with respect to $\mathbf{X}_1$

- We consider two examples of this type of balance.

4.7.1 Linear regression - two covariates

Here

$$E(Y_i) = \beta_0 + \beta x_{2i} + \beta_{3i} x_{2i}$$

and

$$\begin{aligned} \mathbf{X}_1^T \mathbf{X}_1 &= n \\ \mathbf{X}_1^T \mathbf{X}_2 &= n\bar{x}_2 \\ \mathbf{X}_1^T \mathbf{X}_3 &= n\bar{x}_3 \\ \mathbf{X}_2^T \mathbf{X}_3 &= \sum_{i=1}^{n} x_{2i} x_{3i} \end{aligned}$$

Thus the condition that $\mathbf{X}_2^T \mathbf{D}_1 \mathbf{X}_3 = \mathbf{0}$ is equivalent to

$$\sum_{i=1}^{n} x_{2i} x_{3i} = \frac{\left(\sum_{i=1}^{n} x_{2i}\right)\left(\sum_{i=1}^{n} x_{3i}\right)}{n}$$

or

$$\sum_{i=1}^{n} (x_{2i} - \bar{x}_2)(x_{3i} - \bar{x}_3)$$

i.e. the sample correlation between x_2 and x_3 is zero.

4.7.2 Two way analysis of variance

The model is

$$E(Y_{ijk}) = \mu + \alpha_i + \beta_j$$

for $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, q$ and $k - 1, 2, \ldots, n_{ij}$. In this case

$$\begin{aligned} \mathbf{X}_1^T\mathbf{X}_1 &= \mathbf{1}^T\mathbf{1} \\ &= n \\ \mathbf{X}_1^T\mathbf{X}_2 &= [n_{1+}\ n_{2+} \ldots n_{p+}] \\ \mathbf{X}_1^T\mathbf{X}_3 &= [n_{+1}\ n_{+2} \ldots n_{+q}] \\ \mathbf{X}_2^T\mathbf{X}_3 &= (n_{ij}) \end{aligned}$$

- Thus the condition that $\mathbf{X}_2^T\mathbf{D}_1\mathbf{X}_3 = \mathbf{0}$ is equivalent to

$$\mathbf{N} = \frac{1}{n}\begin{bmatrix} n_{1+} \\ n_{2+} \\ \vdots \\ n_{p+} \end{bmatrix}[n_{+1}\ n_{+2} \ldots n_{+q}]$$

  i.e.

$$n_{ij} = \frac{n_{i+}n_{+j}}{n}$$

- Note that if $n_{ij} = r$ for all i and j then $n_{i+} = rq$, $n_{+j} = rp$ and $n = rpq$ and the condition is satisfied.

- In general we define a design to be balanced if the number of observations in the finest classification is the same.
- This definition is stronger than the balance condition above but is convenient for complicated designs and is easy to check.
- Orthogonality is defined by the condition that $\mathbf{y}^T\mathbf{Q}_1\mathbf{y}$ and $\mathbf{y}^T\mathbf{Q}2\mathbf{y}$ satisfy $\mathbf{Q}_1\mathbf{Q}_2 = \mathbf{0}$.
  - When the covariance structure of $\mathbf{Y}$ is $\sigma^2\mathbf{I}$ and normality is assumed, orthogonality implies independence of $\mathbf{y}^T\mathbf{Q}_1\mathbf{y}$ and $\mathbf{y}^T\mathbf{Q}_2\mathbf{y}$.
  - If the covariance structure is not $\sigma^2\mathbf{I}$ then orthogonality does not necessarily imply independence (see the unbalanced two factor case).
- Additivity will mean that there is no interaction between two sets of factors so that main effects due to these sets of factors are well defined and interpretable.

4.8 Notation for Models with Linear Predictors

- The quantity $\mathbf{X}\boldsymbol{\beta}$ is called a **linear predictor** and is fundamental in the analysis of the general linear model, logistic regression models and log linear models.
- In this section we introduce notation designed to facilitate description of these models.
- The notation is due to Wilkinson, G.N. and Rogers, C.E. (1973) *Symbolic Description of Factorial Models for Analysis of Variance* Applied Statistician 22, 392-399.

4.8.1 Covariates in Linear Predictors

A **covariate** in a model (also called independent variable, regressor, exogenous variable) is an auxiliarly measurement recorded with each value of the response variable.

- A **continuous** covariate is a covariate whose values are recorded on an interval or ratio scale.
  - Continuous covariates are denoted by X, W, Z and specific values by x, w, z.
- A **factor** is a covariate with nominal or ordinal values. A specific value of a factor is called a **level**.
  - Factors are denoted by the letters $A, B, C, \ldots$ and the levels by $A_1, A_2, \ldots, A_{p_A}$. Typically the levels of a factor are taken to be $1, 2, \ldots, p_A$.
  - A factor with p_A levels in a linear predictor is equivalent to fitting $p_A - 1$ indicator (dummy) variables with the jth indicator having value equal to 1 if A has level j and 0 otherwise.
  - By convention the lowest level of the factor does not have an associated indicator variable.

- The **interaction** of two factors is denoted by $A.B$. Fitting interaction terms in the model is equivalent to fitting $(p_A - 1)(p_B - 1)$ indicator variables with values given by the products of the indicator variables for A and B.
  - In general a kth order interaction is denoted by $A.B.C \ldots K$ and is equivalent to fitting $(p_A - 1)(p_B - 1) \cdots (p_K - 1)$ indicator variables with values given by the product of the indicators of all k factors
  - Equivalently a kth order interaction fits the product of the $(k - 1)st$ interaction indicators and the kth factor indicators.
- The interaction of a factor and a continuous covariate is denoted by a term of the form $A.X$ and is equivalent to fitting a linear predictor in which the continuous covariate is allowed to have a different coefficient for each level of the factor.
  - Similarly the term $(A.B).X$ would indicate a predictor in which the continuous covariate X is allowed to have a different coefficient for each value of the term $A.B$.

4.8.2 Interpretation of Estimates

- The coefficient of a continuous covariate in a linear predictor represents the effect on the linear predictor of a unit change in the covariate.
  - More precisely: If b is the coefficient, b represents the change in the linear predictor associated with a unit change in the covariate, assuming that all other terms in the linear predictor are unchanged.
- The coefficient of an indicator for a factor A, say $a(j)$ represents the effect on the linear predictor of a change in the factor level fron level 1 to level j.
  - More precisely: $a(j)$ represents the difference in the linear predictor at level j of A and the linear predictor at level 1 of A, assuming that all other terms in the linear predictor are unchanged. If there are interaction terms involving A in the model then $a(j)$ repesents the difference in the linear predictor at level j of A and the linear predictor at level 1 of A at level 1 of those factors with which A has an interaction.
- The coefficient of an indicator for an interaction term, say $a(j).b(k)$ represents the difference between
  - the effect on the linear predictor of a change from level 1 of A to level j of A at level k of B.
  - the effect on the linear predictor of a change from level 1 of A to level j of A at level 1 of B.

- Interpreting interaction terms of two factors involves consideration of two by two tables of the form

	Factor B	
Factor A	Level B_1	Level B_j
A_1	$LP(1,1)$	$LP(1,j)$
A_i	$LP(i,1)$	$LP(i,j)$

  where $LP(i_1, i_2)$ denotes the linear predictor at level i_1 of A and level i_2 of B.

- Interpreting interaction terms of a factor and a continuous covariate involves consideration of a table of the form

Factor A	Model
A_1	$LP(1,x) = \beta_0 + \beta_1 x$
A_i	$LP(i,x) = \beta_0 + (A.\beta)(i)x$

  where $LP(i_1, i_2)$ denotes the linear predictor at level i_1 of A and level i_2 of B. Thus $(A.\beta)(i)$ represents the difference between the slope of x at level i of A and the slope of x at level 1 of A.

4.8.3 Combination of Model Terms in Linear Predictors

Model terms may be combined according to a variety of operators, the notation is designed to allow great flexibility in fitting models.

- **dot**
  - $A.B$ denotes the interaction between two factors
  - $A.A = A$
  - $X.X \neq X$ and $X.X \neq X^2$. For continuous covariates create a new covariate equal to the product to fit terms such as X^2.
- **addition**
  - $A + B$ stands for the model intercept plus factor A plus factor B.
  - $A + B + A.B$ stands for the model intercept plus factor A plus factor B plus interaction between A and B.
  - $A + X$ stands for the model intercept plus factor A plus X, a common slope for each level of A
  - $A + X + A.X$ stands for the model intercept plus factor A plus a different slope on X for each level of A

- **crossing**
  - $A * B$ stands for the model intercept plus factor A plus factor B plus interaction between A and B. i.e. $A * B = A + B + A.B$.
- **nesting**
  - A/B stands for the model $A + A.B$
  - Nesting occurs when a linear predictor is such that differences between levels of A are meaningful but differences between levels of B are only meaningful when specific to a level of A. e.g. if A represents different facilities and B represents different technicians then comparing technician 2 and technician 3 in facility 1 makes sense. Comparing technician 1 and in facility 1 with technician 1 in facility 2 makes no sense since the assignment of levels to technicians within a facility is arbitrary.