Chapter 4

Analysis of Variance

In this chapter we discuss analysis of variance models.

- Many of the concepts of linear models are best illustrated in the analysis of variance context.
  - In addition, extensions to generalized linear models are most easily illustrated using analysis of variance models.
- Many think that since use of dummy variables reduces analysis of variance models to regression models there is no need to study analysis of variance models in any detail.
  - However the concepts of interaction, contrasts, confounding and balance are fundamental to the interpretation of analyses and experimental design.

4.1 One Way Analysis of Variance

4.1.1 Treatment structure I: Fixed Effects Models

Let $y_1, y_2, \ldots, y_n$ be realized values of $Y_1, Y_2, \ldots, Y_n$ such that

$$E(Y_i) = \mu \ \text{ and } \ \text{cov}\,(Y_i, Y_j) = \begin{cases} \sigma^2 & i = j \\ 0 & \text{elsewhere} \end{cases}$$

- Thus the $\mathbf{X}$ matrix is $n \times 1$ and equal to $\mathbf{1}$, a column of ones.
- The least squares equations are given by

  $$\mathbf{1}^T\mathbf{1}b = \mathbf{1}^T\mathbf{y}$$

  so that $b = \bar{y}$ is the least squares estimate of μ.
- The associated error sum of squares is given by

  $$\begin{aligned} \mathbf{y}^T\mathbf{y} - (\mathbf{1}^T\mathbf{y})^T b &= \sum_{i=1}^n y_i^2 - n\bar{y}^2 \\ &= \sum_{i=1}^n (y_i - \bar{y})^2 \end{aligned}$$

  with $n - 1$ d.f.

A one way analysis of variance model is defined by the following model: the observations y_{ij} are assumed to be realized values of Y_{ij} where

$$E(Y_{ij}) = \mu + \tau_i \text{ and } \operatorname{cov}(Y_{ij}, Y_{i',j'}) = \begin{cases} \sigma^2 & i' = i,\ j' = j \\ 0 & \text{elsewhere} \end{cases}$$

and

$$i = 1, 2, \ldots, p\ ,\ j = 1, 2, \ldots, n_i$$

- In this case the $\mathbf{X}$ matrix is given by:

$$E(\mathbf{Y}) = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_{n_p} & \mathbf{0}_{n_p} & \mathbf{0}_{n_p} & \cdots & \mathbf{1}_{n_p} \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_p \end{bmatrix}$$

  where $\mathbf{1}_{n_i}$ is an $n_i \times 1$ vector with each element equal to 1.

Interest focuses on assessing differences among the groups.

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

$$\begin{bmatrix} n & n_1 & n_2 & \cdots & n_p \\ n_1 & n_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_p & 0 & 0 & \cdots & n_p \end{bmatrix} \begin{bmatrix} m \\ t_1 \\ t_2 \\ \vdots \\ t_p \end{bmatrix} = \begin{bmatrix} G \\ T_1 \\ T_2 \\ \vdots \\ T_p \end{bmatrix}$$

  where

  - m is the estimate of μ
  - $t_i, i = 1, 2, \ldots, p$ is the estimate of τ_i
  - G is the sum of all observations i.e. $G = \sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij}$ 1
  - T_i is the sum of the observations in the ith group
  - $n = n_1 + n_2 + \cdots + n_p$.

- Note that the sum of the last p rows equals the first row on both sides of the equations. Thus the least squares equations, while consistent, will not have a unique solution since the rank of the $\mathbf{X}^T\mathbf{X}$ matrix is not $p + 1$ but p.

- It is easily verified, however, that a solution is

$$\begin{aligned} m &= \frac{G}{n} \\ t_i &= \frac{T_i}{n_i} - \frac{G}{n} \end{aligned}$$

- The error sum of squares is thus

$$\text{SSE} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{p} \frac{T_i^2}{n_i}$$

  with $n - p$ degrees of freedom.

- We know that a linearly estimable function $\boldsymbol{\ell}^T\boldsymbol{\beta}$ has best linear unbiased estimator $\boldsymbol{\ell}^T\mathbf{b}$ where

$$\mathbf{b}^T = \left[\frac{G}{n}, \frac{T_1}{n_1} - \frac{G}{n}, \frac{T_2}{n_2} - \frac{G}{n}, \ldots, \frac{T_p}{n_p} - \frac{G}{n}\right]$$

  - Since

$$E(T_i) = E\left[\sum_{j=1}^{n_i} Y_{ij}\right] = \sum_{j=1}^{n_i} (\mu + \tau_i) = n_i(\mu + \tau_i)$$

- It follows that

$$E\left(\frac{G}{n}\right) = \mu + \bar{\tau} \quad \text{and} \quad E\left(\frac{T_i}{n_i}\right) = \mu + \tau_i$$

where $\bar{\tau} = \frac{1}{n}\sum_{i=1}^{p} n_i\tau_i$ so that

$$E\left(\frac{T_i}{n_i} - \frac{G}{n}\right) = \tau_i - \bar{\tau}$$

- If $\boldsymbol{\ell}^T\boldsymbol{\beta}$ is to be estimable $\ell_0, \ell_1, \ell_2, \ldots, \ell_p$ must satisfy

$$\ell_0\mu + \sum_{i=1}^{p} \ell_i\tau_i = \ell_0(\mu + \bar{\tau}) + \sum_{i=1}^{p} \ell_i(\tau_i - \bar{\tau})$$

or

$$\ell_0 - \sum_{i=1}^{p} \ell_i = 0 \quad \text{i.e.} \quad \sum_{i=1}^{p} \ell_i = \ell_0$$

- In particular, the following is a set of p linearly independent estimable functions

$$\mathbf{L}^T\boldsymbol{\beta} = \begin{bmatrix} p & +1 & +1 & +1 & \cdots & +1 & +1 \\ 0 & 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & 0 & 0 & \cdots & 0 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_p \end{bmatrix}$$

- The hypothesis that all groups are the same i.e.

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_p$$

is thus equivalent to the hypothesis:

$$\begin{aligned} \tau_1 - \tau_2 &= 0 \\ \tau_1 - \tau_3 &= 0 \\ \cdots &= \cdots \\ \tau_1 - \tau_p &= 0 \end{aligned}$$

which constitutes a set of $p-1$ linearly independent estimable functions.

- Setting

$$\tau_1 = \tau_2 = \cdots = \tau_p = \tau$$

results in a reduced or conditional model having expected values given by

$$E(Y_{ij}) = \mu + \tau = \mu^*$$

for which we have the least squares estimate $m^* = \frac{G}{n}$ and conditional error sum of squares given by

$$\text{SSCE} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{G^2}{n}$$

with $n-1$ d.f.

- It follows that the sum of squares (deviance) due to the hypothesis that all groups are equal is

$$\begin{aligned} \text{SSCE} - \text{SSE} &= \left[\sum_{i=1}^{p}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{G^2}{n}\right] - \left[\sum_{i=1}^{p}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{p}\frac{T_i^2}{n_i}\right] \\ &= \sum_{i=1}^{p}\frac{T_i^2}{n_i} - \frac{G^2}{n} \\ &= \sum_{i=1}^{p} n_i(\bar{y}_{i+} - \bar{y}_{++})^2 \end{aligned}$$

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

  - In this expression $\bar{y}_{i+} = \frac{T_1}{n_i}$ is the mean of the ith group and $\bar{y}_{++} = \frac{G}{n}$ is the mean of all observations.

We may summarize the above analysis in an ANOVA table as

Source	d.f.	Sum of Squares (Deviance)
Groups	$p-1$	$\sum_{i=1}^{p} n_i(\bar{y}_{i+} - \bar{y}_{++})^2$
Error	$n-p$	$\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i+})^2$
Total	$n-1$	$\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{++})^2$

At this point a natural question to ask is: why is the model formulated as $E(Y_{ij}) = \mu + \tau_i$ rather than the more natural formulation $E(Y_{ij}) = \mu_i$?

- In such an alternative formulation we have

$$E(\mathbf{Y}) = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \cdots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_p} & \mathbf{0}_{n_p} & \mathbf{0}_{n_p} & \cdots & \mathbf{1}_{n_p} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

- so that the least squares equations $\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y}$ are given by

$$\begin{bmatrix} n_1 & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n_p \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_p \end{bmatrix} = \begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_p \end{bmatrix}$$

  where m_i denotes the estimate of μ_i and T_i is the sum of all observations in the ith group.

- The least squares estimates are clearly

$$m_i = \frac{T_i}{n_i}$$

  and

$$\begin{aligned} \text{SSE} &= \sum_{i=1}^{p}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{p} \frac{T_i^2}{n_i} \\ &= \sum_{i=1}^{p}\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i+})^2 \end{aligned}$$

  with $n - p$ d.f. since the $\mathbf{X}^T\mathbf{X}$ matrix is of full rank equal to p.

- The hypothesis that all groups are equal is

$$\mu_1 = \mu_2 = \cdots = \mu_p = \mu$$

for which the reduced model is $E(Y_{ij}) = \mu$ which has

$$\text{SSCE} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{G^2}{n}$$

- Thus the sum of squares due to the hypothesis that the groups are equal is

$$\begin{aligned} \text{SSCE} - \text{SSE} &= \left[\sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{G^2}{n} \right] - \left[\sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{p} \frac{T_i^2}{n_i} \right] \\ &= \sum_{i=1}^{p} \frac{T_i^2}{n_i} - \frac{G^2}{n} \\ &= \sum_{i=1}^{p} n_i (\bar{y}_{i+} - \bar{y}_{++})^2 \end{aligned}$$

with $(n-1) - (n-p) = p - 1$ degrees of freedom exactly as in the previous formulation.

- In order to see the reason for the first formulation
  - Define a mean or overall effect by
  $$\mu = \frac{\mu_1 + \mu_2 + \cdots + \mu_p}{p}$$
  - and a main effect of the ith group by
  $$\tau_i = \mu_i - \mu$$
  - then we can write
  $$\mu_i = \mu + \tau_i$$
  where the τ_i obey the natural restriction $\sum_{i=1}^{p} \tau_i = 0$.
  - Thus in the original formulation we view the τ_i as the differential effect of the ith group relative to an overall expected response.
  - For many experiments this is a convenient formulation, i.e.
  $$Y_{ij} = (\text{overall}) + (\text{group effect}) + (\text{error})$$
- Still another parametrization is useful in analysis of variance models.
  - Consider a one way ANOVA in which there are p groups with n_i observations in the ith group.
  - The response variable and the covariate (dummy variable) values are then given by

Response	Dummy Variable			
	D_2	D_3	$\cdots$	D_p
y_{11}	0	0	$\cdots$	0
y_{12}	0	0	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_{1n_1}	0	0	$\cdots$	0
y_{21}	1	0	$\cdots$	0
y_{22}	1	0	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_{2n_2}	1	0	$\cdots$	0
y_{31}	0	1	$\cdots$	0
y_{32}	0	1	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_{3n_3}	0	1	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_{p1}	0	0	$\cdots$	1
y_{p2}	0	0	$\cdots$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_{pn_p}	0	0	$\cdots$	1

- The equations for the regression coefficients are thus given by

$$\begin{aligned} nb_0 + n_2 b_2 + \cdots + n_p b_p &= n\bar{y} \\ n_2 b_0 + n_2 b_2 &= n_2 \bar{y}_{2+} \\ n_3 b_0 + n_3 b_3 &= n_3 \bar{y}_{3+} \\ \cdots &= \cdots \\ n_p b_0 + n_p b_p &= n_p \bar{y}_{p+} \end{aligned}$$

- Adding the last $p-1$ equations together yields the equation

$$(n - n_1)b_0 + n_2 b_2 + \cdots + n_p b_p = n_2 \bar{y}_{2+} + \cdots n_p \bar{y}_{p+} = n\bar{y} - n_1 \bar{y}_{1+}$$

- Subtracting this from the first equation yields

$$n_1 b_0 = n_1 \bar{y}_{1+}$$

- Thus $b_0 = \bar{y}_{1+}$ and substituting this into the remaining equations yields

$$b_i = \bar{y}_{i+} - \bar{y}_{1+} \quad \text{for} \quad i = 2, 3, \ldots, p$$

- This solution corresponds to the solution obtained in a one way analysis of variance model used in many statistical packages (STATA, S-PLUS, SAS).

4.1.2 Contrasts in one way analysis of variance

Consider an experiment designed to investigate the response to varying levels of a quantitative treatment such as amount of a drug, pesticide, etc.

- In such cases there may be a "dose response curve" relating response to dose.
- We may then be interested in some of the features of this function i.e. is it linear, quadratic etc.?
- It is possible to obtain separate measures of such tendencies using a decomposition of the treatment sum of squares using contrasts.

Definition: A linear combination of treatment effects $\sum_{i=1}^{p} \ell_i \tau_i$ is called a **treatment contrast** if $\sum_{i=1}^{p} \ell_i = 0$.

- ○ A treatment contrast is called a **canonical treatment contrast** if it is of the form $\tau_i - \tau_j$.

- The treatment or group sum of squares can be calculated by selecting a set of $p-1$ linearly independent estimable treatment contrasts, finding the coresponding estimates $\mathbf{L}^T\mathbf{b}$ where $\mathbf{b}$ is any solution to the least squares equations and then calculating

$$(\mathbf{L}^T\mathbf{b})^T[\mathbf{L}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{L}]^{-1}(\mathbf{L}^T\mathbf{b})$$

- Equivalently, one may find a matrix $\mathbf{C}^T$ such that

$$\mathbf{C}^T\mathbf{y} = \mathbf{L}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y})$$

- Then the treatment sum of squares is given by

$$(\mathbf{C}^T\mathbf{y})^T(\mathbf{C}^T\mathbf{C})^{-1}(\mathbf{C}^T\mathbf{y})$$

  where $\mathbf{C}^T$ is $(p-1)\times n$ of rank $p-1$.

- If we choose $\mathbf{C}$ so that $\mathbf{C}^T\mathbf{C}$ is a diagonal matrix i.e. the rows of C^T are orthogonal then the treatment sum of squares is given by

$$\sum_{i=1}^{p-1}\frac{(\mathbf{c}_i\mathbf{y})^2}{\mathbf{c}_i^T\mathbf{c}_i}$$

  where $\mathbf{c}_i^T$ is the ith row of $\mathbf{C}^T$.

- If we choose the rows of $\mathbf{C}^T$ appropriately

$$\frac{(\mathbf{c}_i^T\mathbf{y})^2}{\mathbf{c}_i^T\mathbf{c}_i}$$

can be used as a measure of linearity, quadratic tendency etc.

- A general approach to finding orthogonal contrasts uses the Gram Schmidt process. Thus if the dose levels are $x_1, x_2, \ldots, x_p$ and we want $r(< p)$ orthogonal polynomials giving linear, quadratic, etc. components of the responses we simply write

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^r \\ 1 & x_2 & x_2^2 & \cdots & x_2^r \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_p & x_p^2 & \cdots & x_p^r \end{bmatrix}$$

and use the Gram Schmidt process on the successive columns of $\mathbf{X}$.

4.1.3 Treatment structure II: Random Effect Models

In the one way analysis of variance the sums of squares in the analysis of variance are

$$\begin{aligned} \text{SS(Mean)} &= \frac{G^2}{n} \\ \text{SS(Groups)} &= \sum_{i=1}^{p} \frac{T_i^2}{n_i} - \frac{G^2}{n} \\ \text{SS(Error)} &= \sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{p} \frac{T_i^2}{n_i} \end{aligned}$$

with d.f. equal to 1, $p-1$ and $n-p$ respectively, where $n = \sum_{i=1}^{p} n_i$.

- If we view the model as conditional on a structure S, the model we have used thus far is defined by the structure

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

- Under S we thus have

$$
\begin{aligned}
E(Y_{ij}^2|S) &= \text{var}\,(Y_{ij}|S) + [E(Y_{ij}|S)]^2 \\
&= \sigma^2 + \mu_i^2 \\
E(T_i^2|S) &= \text{var}\,(T_i|S) + [E(T_i|S)]^2 \\
&= n_i\sigma^2 + n_i^2\mu_i^2 \\
E(G^2|S) &= \text{var}\,(G|S) + [E(G|S)]^2 \\
&= n\sigma^2 + \left(\sum_{i=1} n_i\mu_i\right)^2
\end{aligned}
$$

- It follows that

$$
\begin{aligned}
E(\text{SS(Mean)}|S) &= \sigma^2 + \frac{\left(\sum_{i=1}^{p} n_i\mu_i\right)^2}{n} \\
E(\text{SS(Treatment)}|S) &= (p-1)\sigma^2 + \sum_{i=1}^{p} n_i\mu_i^2 - n\bar{\mu}^2 \\
E(\text{SS(Error)}|S) &= (n-p)\sigma^2
\end{aligned}
$$

where $\bar{\mu} = \frac{\sum_{i=1}^{p} n_i\mu_i}{n}$

We thus have, conditional on S, the following table for the expected values of the mean squares

Source	d.f.	Sum of Squares	Expected Mean Square
Mean	1	$\frac{G^2}{n}$	$\sigma^2 + \frac{\left(\sum_{i=1}^p n_i\mu_i\right)^2}{n}$
Groups	$p-1$	$\sum_{i=1}^p \frac{T_i^2}{n_i} - \frac{G^2}{n}$	$\sigma^2 + \sum_{i=1}^p n_i\mu_i^2 - n\bar{\mu}^2/(p-1)$
Error	$n-p$	$\sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^p \frac{T_i^2}{n_i}$	σ^2

- The expected mean square column is $E(\text{SS})/\text{d.f.}$
- If the structure S is appropriate the above expectations are correct.
- If, however, we entertain a model which specifies that the μ_i are a sample from a population of possible groups which could have been selected for inclusion in the experiment we have an entirely different situation.
- We must specify the stochastic structure of the μ_i and then take expectations with respect to this structure to obtain the expected values of the sums of squares.

One useful model is the so called **random effects model** in which we assume that

$$E(\mu_i) = \mu \;\text{ and }\; \text{cov}\,(\mu_i, \mu_{i'}) = \begin{cases} \sigma_1^2 & i' = i \\ 0 & \text{elsewhere} \end{cases}$$

- Note that this model implies that the Y_{ij} are correlated since

$$\text{cov}\,(Y_{ij}, Y_{i'j'}) = \begin{cases} \sigma^2 + \sigma_1^2 & (i', j') = (i, j) \\ \sigma_1^2 & i' = i, j' \neq j \\ 0 & \text{elsewhere} \end{cases}$$

- Also note that we repeatedly use the result

$$\text{cov}\,(Y_{ij}, Y_{i'j'}) = E\left[\text{cov}\,(Y_{ij}, Y_{i',j'}|S)\right] + \text{cov}\,\left(E(Y_{ij}|S, E(Y_{i'j'}|S)\right)$$

- Also note that the correlation between any two observations in the same group is given by

$$\text{cov}\,(Y_{ij}, Y_{ij'}) = \frac{\sigma_1^2}{\sigma^2 + \sigma_1^2}$$

  which is called the **intra-class correlation**.

We can easily find the expectations of the various sums of squares under this model by noting that

$$E(\mu_i^2) = \mu^2 + \sigma_1^2$$

- Thus

$$\begin{aligned} E\left[\frac{\left(\sum_{i=1}^p n_i\mu_i\right)^2}{n}\right] &= \frac{\text{var}\ \left[\sum_{i=1}^p n_i\mu_i\right] + \left[E\left(\sum_{i=1}^p n_i\mu_i\right)\right]^2}{n} \\ &= \frac{\left(\sum_{i=1}^p n_i^2\right)\sigma_1^2 + n^2\mu^2}{n} \end{aligned}$$

- and hence

$$\begin{aligned} E(\text{SS(Mean)}) &= \sigma^2 + \left(\sum_{i=1}^p \frac{n_i^2}{n}\right)\sigma_1^2 + n\mu^2 \\ E(\text{SS(Groups)}) &= (p-1)\sigma^2 + \left(n - \sum_{i=1}^p \frac{n_i^2}{n}\right)\sigma_1^2 \\ E(\text{SS(Error)}) &= (n-p)\sigma^2 \end{aligned}$$

- In the important special case where $n_1 = n_2 = \cdots = n_p = r$, called **balance**, we find that

$$\begin{aligned} E(\text{SS(Mean)}) &= \sigma^2 + r\sigma_1^2 + rp\mu^2 \\ E(\text{SS(Groups)}) &= (p-1)\sigma^2 + r(p-1)\sigma_1^2 \\ E(\text{SS(Error)}) &= (n-p)\sigma^2 \end{aligned}$$

- Because of these expectations natural estimators for σ^2 and σ_1^2 are are given by

$$\begin{aligned} \widehat{\sigma^2} &= \frac{\text{SSE}}{p(r-1)} \\ \widehat{\sigma_1^2} &= \frac{\text{SST}}{r(p-1)} - \frac{\text{SSE}}{pr(r-1)} \end{aligned}$$

  - These estimators are unbiased but there is nothing in their derivation which suggests any optimality properties.
  - We shall return to this question later.
  - σ^2 and σ_1^2 are called **variance components** since the variance of an individual response is $\sigma^2 + \sigma_1^2$, a sum of the two components σ^2 and σ_1^2.

Estimation of μ

Consider now the estimation of μ in the random effects model.

- Since $E(\bar{Y}) = \mu$, $\overline{Y}$ is an unbiased estimator of μ.
- However, the presence of correlation between the Y_{ij} does not imply that $\bar{Y}$ has any optimality properties.
- To investigate this let $\sum_{i=1}^{p}\sum_{j=1}^{n_i} a_{ij}Y_{ij}$ be any linear unbiased estimator of μ.
- Then for unbiasedness we must have

$$E\left(\sum_{i=1}^{p}\sum_{j=1}^{n_i} a_{ij}Y_{ij}\right) = \mu \Longrightarrow \sum_{i=1}^{p}\sum_{j=1}^{n_i} a_{ij} = 1$$

- Since

$$\begin{aligned}
\text{var}\left(\sum_{i=1}^{p}\sum_{j=1}^{n_i} a_{ij}Y_{ij}\right) &= E\left[\text{var}\left(\sum_{i=1}^{p}\sum_{j=1}^{n_i} a_{ij}Y_{ij}|S\right)\right] + \text{var}\left[E\left(\sum_{i=1}^{p}\sum_{j=1}^{n_i} a_{ij}Y_{ij}|S\right)\right] \\
&= \left[\sum_{i=1}^{p}\sum_{j=1}^{n_i} a_{ij}^2\right]\sigma^2 + \text{var}\left[\sum_{i=1}^{p}\sum_{j=1}^{n_i} a_{ij}\mu_i\right] \\
&= \left[\sum_{i=1}^{p}\sum_{j=1}^{n_i} a_{ij}^2\right]\sigma^2 + \left[\sum_{i=1}^{p}\left(\sum_{j=1}^{n_i} a_{ij}\right)^2\right]\sigma_1^2
\end{aligned}$$

  we need to find the vales of the a_{ij} which minimize this expression subject to the unbiasedness condition.

- Symmetry requires that $a_{ij} = a_i$ for $j = 1, 2, \ldots, p$. Thus we need to choose $a_1, a_2, \ldots, a_p$ to minimize

$$\left[\sum_{i=1}^{p} n_i a_i^2\right] \sigma^2 + \left[\sum_{i=1}^{p} n_i^2 a_i^2\right] \sigma_1^2$$

subject to the condition for unbiasedness i.e.

$$\sum_{i=1}^{p} n_i a_i = 1$$

- Using the method of Lagrange Multipliers we choose the a_i to minimize

$$h(\mathbf{a}, \lambda) = \left[\sum_{i=1}^{p} n_i a_i^2\right] \sigma^2 + \left[\sum_{i=1}^{p} n_i^2 a_i^2\right] \sigma_1^2 - 2\lambda \left(\sum_{i=1}^{p} n_i a_i - 1\right)$$

- Hence

$$\begin{aligned}
\frac{\partial h}{\partial a_i} &= 2a_i n_i \sigma^2 + 2a_i n_i^2 \sigma_1^2 - 2\lambda n_i \\
\frac{\partial h}{\partial \lambda} &= -2\left(\sum_{i=1}^{p} n_i a_i - 1\right)
\end{aligned}$$

- Equating to 0 results in the following equations

$$\begin{aligned} a_i(\sigma^2 + n_i\sigma_1^2) &= \lambda \\ \sum_{i=1}^{p} n_i a_i &= 1 \end{aligned}$$

- Thus

$$\begin{aligned} a_i &= \frac{\lambda}{\sigma^2 + n_i\sigma_1^2} \\ \frac{1}{\lambda} &= \sum_{i=1}^{p} \left(\frac{n_i}{\sigma^2 + n_i\sigma_1^2} \right) \end{aligned}$$

- It follows that

$$a_i = \frac{1}{\sum_{j=1}^{p}\left(\frac{n_i}{\sigma^2+n_j\sigma_1^2}\right)}\frac{1}{\sigma^2+n_i\sigma_1^2}$$

and the BLUE of μ is thus given by

$$\begin{aligned}\widehat{\mu} &= \left\{\frac{1}{\sum_{j=1}^{p}\left(\frac{n_i}{\sigma^2+n_j\sigma_1^2}\right)}\right\}\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(\frac{1}{\sigma^2+n_j\sigma_1^2}\right)Y_{ij} \\ &= \left\{\frac{1}{\sum_{j=1}^{p}\left(\frac{n_i}{\sigma^2+n_j\sigma_1^2}\right)}\right\}\left(\sum_{i=1}^{p}\frac{n_i\bar{Y}_{i+}}{\sigma^2+n_i\sigma_1^2}\right)\end{aligned}$$

- If $n_1 = n_2 = \cdots = n_p = r$ $\widehat{\mu}$ reduces to $\bar{Y}_{++}$ so that the condition of balance leads to $\bar{Y}_{++}$ as the BLUE even if correlation is present. Note that the BLUE of μ in the general case depends on σ^2 and σ_1^2.

Maximum Likelihood Estimation

If we assume normality and the covariance structure in the preceeding section the likelihood is

$$\begin{aligned}\ell(\mu,\sigma^2,\sigma_1^2;\mathbf{y}) &= (2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n-p}{2}}(\prod_{i=1}^{p}\sigma_1^2+n_i\sigma_1^2)^{-\frac{1}{2}} \\ &\quad \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i+})^2-\frac{1}{2}\sum_{i=1}^{p}\frac{n_i(\bar{y}_{i+}-\mu)^2}{\sigma^2+n_i\sigma_1^2}\right\}\end{aligned}$$

- Clearly the statistics $\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i+})^2$, and $\bar{y}_{1+},\bar{y}_{2+},\ldots,\bar{y}_{p+}$ are sufficient for μ,σ^2 and σ_1^2.
- Note that the dimensionality of the sufficient statistic is $p+1$ whereas the dimensionality of the parameter space is 3.
- If $n_1=n_2=\cdots=n_p=r$ we have

$$\begin{aligned}\sum_{i=1}^{p}\frac{n_i(\bar{y}_{i+}-\mu)^2}{\sigma^2+n_i\sigma_1^2} &= \frac{r\sum_{i=1}^{p}(\bar{y}_{i+}-\mu)^2}{\sigma^2+r\sigma_1^2} \\ &= \frac{r\sum_{i=1}^{p}(\bar{y}_{i+}-\bar{y}_{++})^2+rp(\bar{y}_{++}-\mu)^2}{\sigma^2+r\sigma_1^2}\end{aligned}$$

so that the sufficient statistics are, in the balanced case,

$$\bar{y}_{++},\ \sum_{i=1}^{p}(\bar{y}_{i+}-\bar{y}_{++})^2 \ \text{ and } \ \sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i+})^2$$

and the dimensionality of the sufficient statistics and the dimensionality of the parameter space are the same in the balanced case.

We now investigate properties of the variance component estimators and μ in the one way analysis of variance model under normality assumptions.

- For the ith treatment or group assume that we have realized values y_{ij} of random variables Y_{ij} which have the following stochastic structure

$$E(Y_{ij}) = \mu \text{ and } \operatorname{cov}(Y_{ij}, Y_{i'j'}) = \begin{cases} \sigma^2 + \sigma_1^2 & j = j' \\ \sigma_1^2 & j \neq j' \end{cases}$$

  for $j = 1, 2, \ldots, n_i$ and that the Y_{ij} are jointly normal.

- Thus the distribution for $\mathbf{Y}_i$ is MVN $(\mathbf{1}_{n_i}\mu, \mathbf{V}_i)$ where

$$\mathbf{V}_i = \sigma^2 \mathbf{I}_{n_i} + \sigma_1^2 \mathbf{J}_{n_i}$$

  ($\mathbf{J}_{n_i}$ is an $n_i \times n_i$ matrix with each element equal to 1).

- Using standard matrix results we have

$$\begin{aligned} \mathbf{V}_i^{-1} &= \frac{1}{\sigma^2}\mathbf{I}_{n_i} - \frac{\sigma_1^2}{\sigma^2(\sigma^2 + n_i\sigma_1^2)}\mathbf{J}_{n_i} \\ \det(\mathbf{V}_i) &= (\sigma^2)^{n_i-1}(\sigma^2 + n_i\sigma_1^2) \end{aligned}$$

- Considerable algebra shows that

$$(\mathbf{y}_i - \mu)^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mu) = \frac{\text{SSE}_i}{\sigma^2} + \frac{n_i(\bar{y}_{i+} - \mu)^2}{\sigma^2 + n_i\sigma_1^2}$$

  where $\text{SSE} = \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i+})^2$.

- It follows that the likelihood for the ith group is given by

$$\text{lik}_i(\mu, \sigma^2, \sigma_1^2; \mathbf{y}) = \frac{\exp\left\{-\frac{1}{2}\frac{\text{SSE}_i}{\sigma^2} - \frac{1}{2}\frac{n_i(\bar{y}_{i+}-\mu)^2}{\sigma_i^2+n_i\sigma_1^2}\right\}}{(2\pi)^{-\frac{n_i}{2}}(\sigma^2)^{-\frac{n_i-1}{2}}(\sigma^2+n_i\sigma_1^2)^{-\frac{1}{2}}}$$

- If we assume the groups are independent the overall likelihood becomes

$$\prod_{i=1}^{p} \text{lik}_i(\mu, \sigma^2, \sigma_1^2; \mathbf{y})$$

so that the log likelihood is given by

$$\begin{aligned}\ell(\mu, \sigma^2, \sigma_1^2; \mathbf{y}) = & -\frac{1}{2\sigma^2}\sum_{i=1}^{p}\frac{\text{SSE}_i}{\sigma^2} - \frac{1}{2}\sum_{i=1}^{p}\frac{n_i(\bar{y}_{i+}-\mu)^2}{\sigma^2+n_i\sigma_1^2} \\ & -\frac{n}{2}\log(2\pi) - \sum_{i=1}^{p}\left(\frac{n_i-1}{2}\right)\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{p}\log(\sigma^2+n_i\sigma_1^2)\end{aligned}$$

- The partial derivatives with respect to μ, σ^2 and σ_1^2 are given by

$$
\begin{aligned}
\frac{\partial \ell}{\partial \mu} &= -\frac{1}{2}\sum_{i=1}^{p}\frac{(-2)n_i(\bar{y}_{i+}-\mu)^2}{\sigma^2+n_i\sigma_1^2} \\
\frac{\partial \ell}{\partial \sigma^2} &= \frac{1}{2}\sum_{i=1}^{p}\frac{\text{SSE}_i}{\sigma^4}+\frac{1}{2}\sum_{i=1}^{p}\frac{n_i(\bar{y}_{i+}-\mu)^2}{(\sigma^2+n_i\sigma_1^2)^2} \\
&\quad -\left(\frac{n-p}{2\sigma^2}\right)-\frac{1}{2}\sum_{i=1}^{p}\frac{1}{\sigma^2+n_i\sigma_1^2} \\
\frac{\partial \ell}{\partial \sigma_t^2} &= \frac{1}{2}\sum_{i=1}^{p}\frac{n_i^2(\bar{y}_{i+}-\mu)^2}{(\sigma^2+n_i\sigma_1^2)^2}-\frac{1}{2}\sum_{i=1}^{p}\frac{n_i}{(\sigma^2+n_i\sigma_1^2)}
\end{aligned}
$$

- These maximum likelihood equations have no closed form solution. Large sample properties may be investigated, however by calculating the information matrix. Thus

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \mu^2} &= -\sum_{i=1}^{p} \frac{n_i}{\sigma^2 + n_i\sigma_1^2} \\
\frac{\partial^2 \ell}{\partial (\sigma^2)^2} &= -\sum_{i=1}^{p} \frac{\mathrm{SSE}_i}{(\sigma^2)^3} - \sum_{i=1}^{p} \frac{n_i(\bar{y}_{i+} - \mu)^2}{(\sigma^2 + n_i\sigma_1^2)^3} \\
&\quad + \left(\frac{n-p}{2\sigma^4}\right) + \frac{1}{2}\sum_{i=1}^{p} \frac{1}{(\sigma^2 + n_i\sigma_1^2)^2} \\
\frac{\partial^2 \ell}{\partial (\sigma_1^2)^2} &= -\sum_{i=1}^{p} \frac{n_i^3(\bar{y}_{i+} - \mu)^2}{(\sigma^2 + n_i\sigma_1^2)^3} + \frac{1}{2}\sum_{i=1}^{p} \frac{n_i^2}{(\sigma^2 + n_i\sigma_1^2)^2} \\
\frac{\partial^2 \ell}{\partial \sigma^2 \partial \mu} &= -\sum_{i=1}^{p} \frac{n_i(\bar{y}_{i+} - \mu)}{(\sigma^2 + n_i\sigma_1^2)^2} \\
\frac{\partial^2 \ell}{\partial \sigma_t^2 \partial \mu} &= -\sum_{i=1}^{p} \frac{n_i^2(\bar{y}_{i+} - \mu)}{(\sigma^2 + n_i\sigma_1^2)^2} \\
\frac{\partial^2 \ell}{\partial \sigma^2 \partial \sigma_t^2} &= -\sum_{i=1}^{p} \frac{n_i^2(\bar{y}_{i+} - \mu)^2}{(\sigma^2 + n_i\sigma_1^2)^3} + \frac{1}{2}\sum_{i=1}^{p} \frac{n_i}{(\sigma^2 + n_i\sigma_1^2)^2}
\end{aligned}
$$

- Taking expectations and simplifying yields Fisher's information matrix:

$$\mathbf{I}\left(\mu,\sigma_1^2,\sigma^2\right)=\left[\begin{array}{ccc}\sum_{i=1}^p w_i & 0 & 0\\ 0 & \frac{1}{2}\sum_{i=1}^p w_i^2 & \frac{1}{2}\sum_{i=1}^p \frac{w_i^2}{n_i}\\ 0 & \frac{1}{2}\sum_{i=1}^p \frac{w_i^2}{n_i} & \frac{n-p}{\sigma^4}+\frac{1}{2}\sum_{i=1}^p \frac{w_i^2}{n_i^2}\end{array}\right]$$

  where

$$w_i=\frac{n_i}{\sigma^2+n_i\sigma_1^2}=\frac{1}{\operatorname{var}\left(\bar{Y}_{i+}\right)}$$

- In the special case of balance where $n_1=n_2=\ldots=n_p=r$ the likelihood equation for μ becomes

$$-\frac{1}{2}\sum_{i=1}^p \frac{r\left(\bar{y}_{i+}-\mu\right)}{\sigma^2+r\sigma_1^2}=0$$

  so that $\widehat{\mu}=\bar{y}_{++}$.

- Also in the balanced case

$$\sum_{i=1}^{p}(\bar{y}_{i+}-\mu)^2=\sum_{i=1}^{p}(\bar{y}_{i+}-\bar{y})^2+p(\bar{y}_{++}-\mu)^2$$

so that the likelihood equation for σ_1^2 is

$$\frac{1}{2}\frac{r\sum_{i=1}^{p}(\bar{y}_{i+}-\bar{y})^2+r^2p(\bar{y}_{++}-\mu)^2}{\sigma^2+r\sigma_1^2}-\frac{1}{2}\frac{rp}{\sigma^2+r\sigma_1^2}=0$$

- Thus

$$\widehat{\sigma^2}+r\widehat{\sigma_1^2}=\frac{r}{p}\sum_{i=1}^{p}(\bar{y}_{i+}-\bar{y})^2$$

- The likelihood equation in the balanced case for σ^2 is

$$\frac{1}{2}\sum_{i=1}^{p}\frac{\text{SSE}_i}{\sigma^4}+\frac{1}{2}\sum_{i=1}^{p}\frac{p(\widehat{\sigma^2}+r\widehat{\sigma_1^2})}{(\widehat{\sigma^2}+r\widehat{\sigma_1^2})^2}-\left(\frac{n-p}{2\widehat{\sigma^4}}\right)-\frac{1}{2}\sum_{i=1}^{p}\frac{1}{\widehat{\sigma^2}+r\widehat{\sigma_1^2}}=0$$

so that

$$\widehat{\sigma^2}=\frac{\sum_{i=1}^{p}\text{SSE}_i}{n-p}=\frac{\sum_{i=1}^{p}\sum_{j=1}^{r}(y_{ij}-\bar{y}_{i+})^2}{n-p}$$

- It follows that the maximum likelihood estimate of σ_1^2 is

$$\widehat{\sigma_1^2}=\frac{1}{r}\left[r\frac{\sum_{i=1}^{p}(\bar{y}_{i+}-\bar{y}_{i+})^2}{p}-\frac{\sum_{i=1}^{p}\sum_{j=1}^{r}(y_{ij}-\bar{y}_{i+})^2}{n-p}\right]$$

or 0 if this expression is negative.

- The Fisher information matrix for this special case is given by

$$\mathbf{I}\left(\mu, \sigma_1^2, \sigma^2\right) = \begin{bmatrix} \frac{pr}{\sigma^2 + r\sigma_1^2} & 0 & 0 \\ 0 & \frac{1}{2}\frac{pr^2}{(\sigma^2+r\sigma_1^2)^2} & \frac{1}{2}\frac{pr}{(\sigma^2+r\sigma_1^2)^2} \\ 0 & \frac{1}{2}\frac{pr}{(\sigma^2+r\sigma_1^2)^2} & \frac{1}{2}\frac{pr^2}{(\sigma^2+r\sigma_1^2)^2} \end{bmatrix}$$

which has inverse given by

$$\mathbf{I}^{-1}\left(\mu, \sigma_1^2, \sigma^2\right) = \begin{bmatrix} \frac{\sigma^2}{pr} + \frac{\sigma_1^2}{p} & 0 & 0 \\ 0 & 2\frac{(\sigma^2+r\sigma_1^2)^2}{pr^2} + 2\frac{\sigma^4}{r^2(n-p)} & -2\frac{\sigma^4}{r(n-p)} \\ 0 & -2\frac{\sigma^4}{r(n-p)} & 2\frac{\sigma^4}{n-p} \end{bmatrix}$$

- Since

$$\operatorname{var}\left(\bar{Y}_{++}\right) = \frac{\sigma^2}{rp} + \frac{\sigma_1^2}{p}$$

it follows that $\bar{Y}_{++}$ is the minumum variance unbiased estimator of μ.

- We note that in the general case

$$\begin{aligned} \sum_{i=1}^{p}\sum_{j=1}^{r}(y_{ij} - \bar{y}_{i+})^2 &= \sum_{i=1}^{p}\sum_{j=1}^{r} y_{ij}^2 - \sum_{i=1}^{p} n_i \bar{y}_{i+}^2 \\ &= \sum_{i=1}^{p}\left[\mathbf{y}_i^T\mathbf{y}_i - \frac{1}{n_i}\mathbf{y}_i^T\mathbf{1}\mathbf{1}^T\mathbf{y}_i\right] \\ &= \sum_{i=1}^{p}\mathbf{y}_i^T\mathbf{B}_i\mathbf{y}_i \end{aligned}$$

where $\mathbf{B}_i = \mathbf{I}_{n_i} - \frac{1}{n_i}\mathbf{1}\mathbf{1}^T$.

- Since $\mathbf{V}_i = \sigma^2\mathbf{I}_{n_i} + \sigma_1^2\mathbf{1}\mathbf{1}^T$ we have

$$\begin{aligned} \mathbf{B}_i\mathbf{V}_i &= \sigma^2\mathbf{I}_{n_i} + \sigma_1^2\mathbf{1}\mathbf{1}^T - \frac{\sigma^2}{n_i}\mathbf{1}\mathbf{1}^T\sigma_t^2\mathbf{1}\mathbf{1}^T \\ &= \sigma^2\left[\mathbf{I}_{n_i} - \frac{1}{n_i}\mathbf{1}\mathbf{1}^T\right] \end{aligned}$$

- It follows that $\frac{1}{\sigma^2}\mathbf{B}_i\mathbf{V}_i$ is idempotent. and since

$$\frac{1}{\sigma^2}\mathbf{B}_i\mathbf{1}\mu = \frac{1}{\sigma^2}\left[\mathbf{I}_{n_i} - \frac{1}{n_i}\mathbf{1}\mathbf{1}^T\right]\mathbf{1}\mu = \mathbf{0}$$

  the distribution of $\frac{\mathbf{Y}_i^T\mathbf{B}_i\mathbf{Y}_i}{\sigma^2}$ is chi square with $n_i - 1$ degrees of freedom.

- Thus

$$\sum_{i=1}^{p}\frac{\mathbf{Y}_i^T\mathbf{B}_i\mathbf{Y}_i}{\sigma^2} = \frac{\text{SSE}}{\sigma^2}$$

  is chi-square with $n - p$ degrees of freedom and

$$\begin{aligned} E\left(\sum_{i=1}^{p}\mathbf{Y}_i^T\mathbf{B}_i\mathbf{Y}_i\right) &= (n-p)\sigma^2 \\ \text{var}\left(\sum_{i=1}^{p}\mathbf{Y}_i^T\mathbf{B}_i\mathbf{Y}_i\right) &= 2(n-p)\sigma^4 \end{aligned}$$

- It follows that

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{p}\mathbf{Y}_i^T\mathbf{B}_i\mathbf{Y}_i}{n-p}$$

  is the minimum variance unbiased estimator of σ^2

Recall that

$$\text{SST} = \sum_{i=1}^{p} \frac{T_i^2}{n_i} - \frac{G^2}{n}$$

has expectation

$$(p-1)\sigma^2 + \left(n - \sum_{i=1}^{p} \frac{n_i^2}{n}\right)\sigma_1^2 = (p-1)\sigma^2 + f_1\sigma_1^2$$

so that an unbiased estimator for σ_1^2 is

$$\frac{\text{SST} - (p-1)\widehat{\sigma^2}}{f_1}$$

- Note that we can write

$$\begin{aligned}
\sum_{i=1}^{p} \frac{T_i^2}{n_i} &= \sum_{i=1}^{p} \frac{\mathbf{y}_i^T \mathbf{1}\mathbf{1}^T \mathbf{y}}{n_i} \\
&= \mathbf{y}^T \mathbf{B} \mathbf{y} \\
\frac{G^2}{n} &= \frac{1}{n} \mathbf{y}^T \mathbf{1}\mathbf{1}^T \mathbf{y} \\
&= \mathbf{y}^T \mathbf{C} \mathbf{y}
\end{aligned}$$

  where

$$\begin{aligned}
\mathbf{B} &= \text{diag}\left(\frac{1}{n_i} \mathbf{1}_{n_i} \mathbf{1}_{n_i}^T\right) \\
\mathbf{C} &= \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T
\end{aligned}$$

- It follows that

$$\text{SST} = \mathbf{y}^T(\mathbf{B} - \mathbf{C})\mathbf{y}$$

and

$$(\mathbf{B} - \mathbf{C})\boldsymbol{\mu} = (\mathbf{B} - \mathbf{C})\mathbf{1}\mu = \mathbf{0}$$

- However, there is clearly no divisor of $\mathbf{B} - \mathbf{C}$ which will make $(\mathbf{B} - \mathbf{C})\mathbf{V}$ idempotent so that in the general case

$$\frac{\text{SST}}{E(\text{MST})}$$

is not distributed as a chi square.

- If $n_1 = n_2 = \cdots = n_p = r$ we get quite a different result since in this case

$$\begin{aligned}(\mathbf{B} - \mathbf{C})\mathbf{V} &= \left[\text{diag}\,\frac{1}{r}\mathbf{1}_r\mathbf{1}_r^T - \frac{1}{rp}\mathbf{1}_{rp}\mathbf{1}_{rp}\right]\left[\sigma^2\mathbf{I}_{rp} + \text{diag}\,(\sigma_1^2\mathbf{1}_r\mathbf{1}_r^T)\right] \\ &= (\sigma^2 + r\sigma_1^2)\left[\text{diag}\,\frac{1}{r}\mathbf{1}_r\mathbf{1}_r^T - \frac{1}{rp}\mathbf{1}_{rp}\mathbf{1}_{rp}\right]\text{diag}\,\frac{1}{r}\mathbf{1}_r\mathbf{1}_r^T\end{aligned}$$

which is idempotent.

- Thus
$$\frac{\text{SST}}{\sigma^2 + r\sigma_1^2} = \mathbf{Y}^T \left(\frac{\mathbf{B} - \mathbf{C}}{\sigma^2 + r\sigma_1^2} \right) \mathbf{Y}$$
is chi square with $p - 1$ degrees of freedom.

- **Conclusion:**
  - Balance in one way analysis of variance models with random effects leads to simple sufficient statistics and strong optimality properties for the estimators.
  - Lack of balance results in estimators which have no closed form expressions and very few small sample optimality properties.

4.1.4 Treatment Structure III: Analysis of Covariance

Consider p groups or treatments which are to be compared.

- If there are n_i observations on the ith group a one way analysis of variance gives the variability within groups as

$$\sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i+})^2$$

  which is compared with the variability among or between groups

$$\sum_{i=1}^{p} n_i (\bar{y}_{i+} - \bar{y}_{++})^2$$

- Differences between groups are estimated by $\bar{y}_{i+} - \bar{y}_{i'+}$ with estimated variance

$$\left(\frac{1}{n_i} + \frac{1}{n_i'}\right) \frac{\sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i+})^2}{n - p}$$

Suppose now that we have available another measurement x_{ij} for each response y_{ij} which we expect to be linearly predictive of the response in the sense that

$$E(Y_{ij}) = \alpha_i + \beta x_{ij}$$

- Thus we have a model of the form

$$E(Y_{ij}) = \alpha_i + \beta x_{ij} \text{ and } \operatorname{cov}(Y_{ij}, Y_{i'j'}) = \begin{cases} \sigma^2 & (i'j') = (i,j) \\ 0 & \text{elsewhere} \end{cases}$$

  which is called a one way analysis of covariance model.

- In order to compare the groups it is necessary to take into account the effect of x i.e. we need to "adjust for the effect of the covariate".

- The hypothesis for equality of the groups is now

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p$$

We thus set up the model as:

$$E(\mathbf{Y}) = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} & \mathbf{x}_1 \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \cdots & \mathbf{0}_{n_2} & \mathbf{x}_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{n_p} & \mathbf{0}_{n_p} & \mathbf{0}_{n_p} & \cdots & \mathbf{1}_{n_p} & \mathbf{x}_p \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \\ \beta \end{bmatrix}$$

- The least squares equations are then

$$\begin{bmatrix} n_1 & 0 & 0 & \cdots & 0 & n_1\bar{x}_{1+} \\ 0 & n_2 & 0 & \cdots & 0 & n_2\bar{x}_{2+} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & n_p & n_p\bar{x}_{p+} \\ n_1\bar{x}_{1+} & n_2\bar{x}_{2+} & n_3\bar{x}_{3+} & \cdots & n_p\bar{x}_{p+} & \sum_{i=1}^{p}\sum_{j=1}^{n_i} x_{ij}^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \\ b \end{bmatrix} = \begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_p \\ \sum_{i=1}^{p}\sum_{j=1}^{n_i} x_{ij}y_{ij} \end{bmatrix}$$

- A solution to the least squares equations is

$$b = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})(y_{ij} - \bar{y}_{i+})}{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2} \quad \text{and} \quad a_i = \bar{y}_{i+} - b\bar{x}_{i+}$$

- The corresponding error sum of squares is

$$\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i+})^2 - b^2\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i+})^2$$

  with $n-p-1$ degrees of freedom.

- If

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p = \mu$$

  is true then the model is given by

$$E(Y_{ij}) = \mu + \beta x_{ij} \text{ and } \operatorname{cov}(Y_{ij}, Y_{i'j'}) = \begin{cases} \sigma^2 & (i'j') = (i,j) \\ 0 & \text{elsewhere} \end{cases}$$

- In this case we have estimates

$$m = \bar{y}_{++} \text{ and } b_0 = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{++})(y_{ij}-\bar{y}_{++})}{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{++})^2}$$

- The error sum of squares is

$$\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{++})^2 - b_0^2\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{++})^2$$

  with $n-2$ degrees of freedom.

- Thus the sum of squares for testing H_0 is

$$\left\{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{++})^2 - b_0^2\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{++})^2\right\}$$
$$-\left\{\sum_{i=1}^{p}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i+})^2 - b^2\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i+})^2\right\}$$

  which reduces to

$$\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\bar{y}_{i+}-\bar{y}_{++})^2 + b^2\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i+})^2 - b_0^2\sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{++})^2$$

  with $(n-2)-(n-p-1)=p-1$ degrees of freedom.

- Under this model, it can be shown that the expected value of the sum of squares for groups in the presence of the covariate x is

$$\sum_{i=1}^{p} n_i(\alpha_i-\bar{\alpha})^2 - \frac{\left[\sum_{i=1}^{p} n_i(\bar{x}_{i+}-\bar{x}_{++})\alpha_i\right]^2}{\sum_{i=1}^{p}(x_{ij}-\bar{x}_{++})^2} + (p-1)\sigma^2$$

  where $\bar{\alpha}=\frac{1}{n}\sum_{i=1}^{p} n_i\alpha_i$

- Thus under $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p$ we have the expected value equal to $(p-1)\sigma^2$.

- One important assumption in the above models is that β is constant from group to group.

- This should be tested **before** performing the above covariance analysis.

If the parameter β is not constant the appropriate model is

$$E(Y_{ij}) = \alpha_i + \beta_i x_{ij} \text{ and } \operatorname{cov}(Y_{ij}, Y_{i'j'}) = \begin{cases} \sigma^2 & (i'j') = (i,j) \\ 0 & \text{elsewhere} \end{cases}$$

- The estimates for α_i and β_i are in this case

$$a_i = \bar{y}_{i+} - b_i \bar{x}_{i+} \text{ and } b_i = \frac{\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})(y_{ij} - \bar{y}_{i+})}{\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2}$$

- The error sum of squares for this model is

$$\sum_{i=1}^{p}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{p} n_i \bar{y}_{i+}^2 - \sum_{i=1}^{p}\left(b_i^2 \sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2\right)$$

  with $n - 2p$ degrees of freedom.

- Thus the appropriate sum of squares for testing $H_0 : \beta_1 = \beta_2 = \cdots = \beta_p$ is

$$\sum_{i=1}^{p}\left(b_i^2 \sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2\right) - b^2 \sum_{i=1}^{p}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+})^2$$

  with $(n - p - 1) - (n - 2p) = p - 1$ degrees of freedom.

If we asume the covariance model is appropriate, we note that to compare two groups i and i' at the value x leads to

$$
\begin{aligned}
\widehat{E(Y_{ij})} &= a_i + bx \\
&= \bar{y}_{i+} - b\bar{x}_{i+} + bx \\
&= \bar{y}_{i+} + b(x - \bar{x}_{i+}) \\
\widehat{E(Y_{i'j})} &= a_{i'} + bx \\
&= \bar{y}_{i'+} - b\bar{x}_{i'+} + bx \\
&= \bar{y}_{i'+} + b(x - \bar{x}_{i'+})
\end{aligned}
$$

- The difference between these two estimates is the appropriate estimate of the difference between the two groups and is

$$(\bar{y}_{i+} - \bar{y}_{i'+}) - b(\bar{x}_{i+} - \bar{x}_{i'+})$$

  - a_i and $a_{i'}$ are called "adjusted means" since in the absence of the covariate the estimate would be

$$\bar{y}_{i+} - \bar{y}_{i'+}$$

- Note that the comparison is the same for any value of x and depends only on the difference between the two means **and** the difference between the covariate means for the groups being compared.

- If the parameter β is not constant from group to group the comparison would be

$$\begin{aligned}(a_i + b_i x) - (a_{i'} + b_{i'} x) &= (\bar{y}_{i+} + b_i(x - \bar{x}_{i+})) - (\bar{y}_{i'+} + b_{i'}(x - \bar{x}_{i'+})) \\ &= (\bar{y}_{i+} - \bar{y}_{i'+}) + (b_i - b_{i'})x - (b_i \bar{x}_{i+} - b_{i'} \bar{x}_{i'+})\end{aligned}$$

  and depends on the value of x even if the groups are balanced with respect to the covariate i.e. $\bar{x}_{i+} = \bar{x}_{i'+}$.

4.1.5 Treatment Structure 4: Finite Population of Treatments

In some contexts we may assume that the groups are selected at random from a finite population of possible groups.

- Thus the model is

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

  where S is defined by

$$\begin{aligned} P(\mu_i = v) &= \frac{1}{P} \\ P(\mu_i = v, \mu_{i'} = w) &= \frac{1}{P(P-1)} \end{aligned}$$

  where $v, w \in \mathcal{P}$, $v \neq w$ and

$$\mathcal{P} = \{\mu_1, \mu_2, \ldots, \mu_P\}$$

- We note that under this model for the group means

$$\begin{aligned} E(\mu_i) &= \sum_{v \in \mathcal{P}} \frac{v}{P} \\ &= \mu \\ \operatorname{var}(\mu_i) &= \frac{P-1}{P}\sigma_1^2 \\ \operatorname{cov}(\mu_i) &= -\frac{\sigma_1^2}{P} \end{aligned}$$

where

$$\sigma_1^2 = \frac{\sum_{j=1}^{P}(\mu_j - \mu)^2}{P-1}$$

- Under this struture we can show that

$$\begin{aligned} E(\text{SSE}) &= (n-p)\sigma^2 \\ E(\text{SST}) &= (p-1)\sigma^2 + \left(n - \sum_{i=1}^{p} \frac{n_i^2}{n}\right)\sigma_1^2 \\ E\left(\frac{G^2}{n}\right) &= \sigma^2 + n\mu^2 + \left(\sum_{i=1}^{p} \frac{n_i^2}{n} - \frac{n}{P}\right)\sigma_1^2 \end{aligned}$$

- In the case of balance $n_1 = n_2 = \cdots = n_p$ the ANOVA table becomes

Source	d.f.	Deviance	Expected Mean Square
Mean	1	$\frac{G^2}{n}$	$\sigma^2 + rp\mu^2 + r\left(1 - \frac{p}{P}\right)\sigma_1^2$
Treatments	$p-1$	$\sum_{i=1}^{p} \frac{T_i^2}{n_i} - \frac{G^2}{n}$	$\sigma^2 + r\sigma_1^2$
Error	$p(r-1)$	$\sum_{i=1}^{p} \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{p} \frac{T_i^2}{n_i}$	σ^2

- The term $1 - \frac{p}{P}$ is called a **finite population correction factor**.