STATISTICAL MODELLING

XI. Estimation of linear model parameters

XI.A Linear models for designed experiments

In the analysis of experiments, the models that we have considered are all of the form

$$\psi = E[\mathbf{Y}] = \mathbf{X}\theta \text{ and } \operatorname{var}[\mathbf{Y}] = \mathbf{V} = \sum \sigma_i^2 \mathbf{S}_i$$

where ψ is an $n \times 1$ vector of expected values for the observations,
$\mathbf{Y}$ is an $n \times 1$ vector of random variables for the observations,
$\mathbf{X}$ is an $n \times q$ matrix with $n \geq q$,
θ is a $q \times 1$ vector of unknown parameters,
$\mathbf{V}$ is an $n \times n$ matrix,
σ_i^2 is the variance component for the ith term in the variance model and
$\mathbf{S}_i$ is the summation matrix for the ith term in the variance model.

Note that generally the $\mathbf{X}$ matrices contain indicator variables, each indicating the observations that received a particular level of a (generalized) factor. Also, the summation matrix $\mathbf{S}_i$ is the matrix of ones and zeroes that forms the sums for the generalized factor corresponding to the ith term. Now, $\mathbf{M}_i = \frac{1}{n/f_i}\mathbf{S}_i$ where $\mathbf{M}_i$ is the mean operator for the ith term in the variance model, n is the number of observational units and f_i is the number of levels of the generalized factor for this term. Note that for equally-replicated, generalized factors $n/f_i = g_i$ where g_i is the number of repeats of each of the levels of the generalized factor over the observational units. Rearranging $\mathbf{M}_i = \frac{1}{n/f_i}\mathbf{S}_i$, we have that $\mathbf{S}_i = \frac{n}{f_i}\mathbf{M}_i$. Consequently, the **S**s could be replaced with **M**s.

For example the model with Blocks random for the randomized complete block design is

$$E[\mathbf{Y}] = \mathbf{X}_\mathrm{T}\boldsymbol{\tau} \text{ and } \mathbf{V} = \sigma^2_\mathrm{BU}\mathbf{I}_n + \sigma^2_\mathrm{B}\left(\mathbf{I}_b \otimes \mathbf{J}_t\right) = \sigma^2_\mathrm{BU}\mathbf{M}_\mathrm{BU} + t\sigma^2_\mathrm{B}\mathbf{M}_\mathrm{B} = \sigma^2_\mathrm{BU}\mathbf{S}_\mathrm{BU} + \sigma^2_\mathrm{B}\mathbf{S}_\mathrm{B}$$

since $n/b = bt/b = t$, $\mathbf{M}_\mathrm{BU} = \mathbf{I}_n$ and $\mathbf{M}_\mathrm{B} = \frac{1}{t}\left(\mathbf{I}_b \otimes \mathbf{J}_t\right)$. (see example below)

Now various classes of linear models can be identified. Firstly, linear models split into those for which the expectation model is of **full rank** compared to those that are of **less than full rank**. The rank of an expectation model is given by the following definitions, the second of which is equivalent to that given in chapter I.

Definition XI.1: The rank of an expectation model is equal to the rank of its **X** matrix.■

Definition XI.2: The **rank** of an $n \times q$ matrix **X**, with $n \geq q$, is the number of linearly independent columns in **X**. Such a matrix is of full rank when its rank equal *q* and is of less than full rank when its rank is less than *q*. It will be less than full rank if the linear combination of some columns is equal to the same linear combination as some other columns. ■

Secondly, linear models split into those for which the variation model is **simple** in that it is of the form $\sigma^2\mathbf{I}_n$ as compared to those that are of the more **general** form that involves more than one term.

In the case of the general linear model, it turns out that the **S** matrices in the variation model can be written as the direct products of **I** and **J** matrices. When all the variation terms correspond to only unrandomized, and not randomized, generalized factors, the following rule can be used to derive these expressions:

Definition XI.3: The direct-product operator is denoted $\otimes$ and, if $\mathbf{A}_r$ and $\mathbf{B}_c$ are square matrices of order *r* and *c*, respectively, then the direct product is given by

$$\mathbf{A}_r \otimes \mathbf{B}_c = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1r}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{r1}\mathbf{B} & \cdots & a_{rr}\mathbf{B} \end{bmatrix}.$$ ■

Rule XI.1: The $\mathbf{S}_i$ for the *i*th generalized factor from the unrandomized structure that involves *s* factors, is the direct product of *s* matrices, provided the *s* factors are arranged in standard order. Taking the *s* factors in the sequence specified by the standard order, for each factor in the *i*th generalized factor include an **I** matrix in the direct product and **J** matrices for those that are not. The order of a matrix in the direct product is equal to the number of levels of the corresponding factor. ■

Example IV.1 Penicillin yield (continued)

This experiment was a randomized complete block design and its experimental structure is

Structure	Formula
unrandomized	*5* Blends/*4* Flasks
randomized	*4* Treatments

There are two possible models for this experiment depending on whether Blends is to be considered random or fixed.

For Blends fixed, and taking the observations to be ordered in standard order for Blends then Flasks, the only random term is Blends$\wedge$Flasks and the maximal expectation and variation models are

$$E[\mathbf{Y}] = \mathbf{X}_{\mathrm{B}}\boldsymbol{\beta} + \mathbf{X}_{\mathrm{T}}\boldsymbol{\tau} = \begin{bmatrix}\mathbf{X}_{\mathrm{B}} & \mathbf{X}_{\mathrm{T}}\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta} \\ \boldsymbol{\tau}\end{bmatrix} \text{ and } \mathbf{V} = \sigma^2_{\mathrm{BF}}\mathbf{S}_{\mathrm{BF}} = \sigma^2_{\mathrm{BF}}\mathbf{I}_5 \otimes \mathbf{I}_4 .$$

Suppose that the observations are in standard order for Blend then Treatment and that Flask has been renumbered within Blend to correspond to Treatment, as in the prerandomization layout — the analysis for prerandomization and randomized layouts must be the same as the prerandomization layout is one of the possible randomized arrangements. Then the vectors and matrices in the expectation model are:

$$\mathbf{X}_{\mathrm{B}} = \begin{bmatrix} 1&0&0&0&0\\ 1&0&0&0&0\\ 1&0&0&0&0\\ 1&0&0&0&0\\ 0&1&0&0&0\\ 0&1&0&0&0\\ 0&1&0&0&0\\ 0&1&0&0&0\\ 0&0&1&0&0\\ 0&0&1&0&0\\ 0&0&1&0&0\\ 0&0&1&0&0\\ 0&0&0&1&0\\ 0&0&0&1&0\\ 0&0&0&1&0\\ 0&0&0&1&0\\ 0&0&0&0&1\\ 0&0&0&0&1\\ 0&0&0&0&1\\ 0&0&0&0&1 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\\ \beta_4\\ \beta_5\end{bmatrix}, \quad \mathbf{X}_{\mathrm{T}} = \begin{bmatrix} 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1\\ 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1\\ 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1\\ 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1\\ 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&1 \end{bmatrix}, \quad \boldsymbol{\tau} = \begin{bmatrix}\tau_1\\ \tau_2\\ \tau_3\\ \tau_4\end{bmatrix}.$$

Not that we can write the **X**s as direct products: $\mathbf{X}_{\mathrm{B}} = \mathbf{I}_5 \otimes \mathbf{1}_4$ and $\mathbf{X}_{\mathrm{T}} = \mathbf{1}_5 \otimes \mathbf{I}_4$. Clearly, the sum of the indicator variables (columns) in $\mathbf{X}_{\mathrm{B}}$ and the sum of those in $\mathbf{X}_{\mathrm{T}}$ both equal $\mathbf{1}_{20}$ so that the expectation model involving the **X** matrix $\begin{bmatrix}\mathbf{X}_{\mathrm{B}} & \mathbf{X}_{\mathrm{T}}\end{bmatrix}$ is of less than full rank — its rank is $b+t-1$.

The linear model is a simple linear model because the variance matrix is of the form $\sigma^2\mathbf{I}_n$. That is, the observations are assumed to have the same variance and to be uncorrelated.

On the other hand, for Blends random, the random terms are Blends and Blends$\wedge$Flasks and the maximal expectation and variation models are

$$E[\mathbf{Y}] = \mathbf{X}_{\mathrm{T}}\tau \text{ and } \mathbf{V} = \sigma^2_{\mathrm{BF}}\mathbf{S}_{\mathrm{BF}} + \sigma^2_{\mathrm{B}}\mathbf{S}_{\mathrm{B}} = \sigma^2_{\mathrm{BF}}\mathbf{I}_5 \otimes \mathbf{I}_4 + \sigma^2_{\mathrm{B}}\mathbf{I}_5 \otimes \mathbf{J}_4 .$$

In this case the expectation is of full rank as $\mathbf{X}_{\mathrm{T}}$ consists of t linearly independent columns. However, it is a general, not a simple, linear model as the variance matrix is not of the form $\sigma^2\mathbf{I}_n$. The form is illustrated for b=3, t=4 below, the extension to b=5 being obvious. Clearly, σ^2_{BF} occurs down the diagonal and σ^2_{B} occurs in diagonal 4×4 blocks.

	Blend											
	I				II				III			
Flask	1	2	3	4	1	2	3	4	1	2	3	4
1	$\sigma^2_{\mathrm{BF}}+\sigma^2_{\mathrm{B}}$	σ^2_{B}	σ^2_{B}	σ^2_{B}	0	0	0	0	0	0	0	0
2	σ^2_{B}	$\sigma^2_{\mathrm{BF}}+\sigma^2_{\mathrm{B}}$	σ^2_{B}	σ^2_{B}	0	0	0	0	0	0	0	0
3	σ^2_{B}	σ^2_{B}	$\sigma^2_{\mathrm{BF}}+\sigma^2_{\mathrm{B}}$	σ^2_{B}	0	0	0	0	0	0	0	0
4	σ^2_{B}	σ^2_{B}	σ^2_{B}	$\sigma^2_{\mathrm{BF}}+\sigma^2_{\mathrm{B}}$	0	0	0	0	0	0	0	0
1	0	0	0	0	$\sigma^2_{\mathrm{BF}}+\sigma^2_{\mathrm{B}}$	σ^2_{B}	σ^2_{B}	σ^2_{B}	0	0	0	0
2	0	0	0	0	σ^2_{B}	$\sigma^2_{\mathrm{BF}}+\sigma^2_{\mathrm{B}}$	σ^2_{B}	σ^2_{B}	0	0	0	0
3	0	0	0	0	σ^2_{B}	σ^2_{B}	$\sigma^2_{\mathrm{BF}}+\sigma^2_{\mathrm{B}}$	σ^2_{B}	0	0	0	0
4	0	0	0	0	σ^2_{B}	σ^2_{B}	σ^2_{B}	$\sigma^2_{\mathrm{BF}}+\sigma^2_{\mathrm{B}}$	0	0	0	0
1	0	0	0	0	0	0	0	0	$\sigma^2_{\mathrm{BF}}+\sigma^2_{\mathrm{B}}$	σ^2_{B}	σ^2_{B}	σ^2_{B}
2	0	0	0	0	0	0	0	0	σ^2_{B}	$\sigma^2_{\mathrm{BF}}+\sigma^2_{\mathrm{B}}$	σ^2_{B}	σ^2_{B}
3	0	0	0	0	0	0	0	0	σ^2_{B}	σ^2_{B}	$\sigma^2_{\mathrm{BF}}+\sigma^2_{\mathrm{B}}$	σ^2_{B}
4	0	0	0	0	0	0	0	0	σ^2_{B}	σ^2_{B}	σ^2_{B}	$\sigma^2_{\mathrm{BF}}+\sigma^2_{\mathrm{B}}$

Cleary, this model allows for covariance, and hence correlation, between flasks from the same blend.

The direct-product expressions for both **V**s conform to rule XI.1. In particular, the direct product for σ^2_{BF} is $\mathbf{I}_5 \otimes \mathbf{I}_4$ and it involves two **I** matrices because both unrandomized factors, Blends and Flasks, occur in Blends$\wedge$Flasks; the direct product for σ_{B}^2 is $\mathbf{I}_5 \otimes \mathbf{J}_4$ and it involves an **I** and a **J** because the factor Blends is in this term but Flasks is not. ■

Example IX.1 Production rate experiment (revisited)

The experimental structure for this experiment was:

Structure	Formula
unrandomized	*4* Factories/*3* Areas/*3* Parts
randomized	*3* Methods**3* Sources

Suppose that all the unrandomized factor are random and so the maximal expectation and variation models are, symbolically

ψ = E[Y] = Methods$\wedge$Sources
and var[Y] = Factories + Factories$\wedge$Areas + Factories$\wedge$Areas$\wedge$Parts.

That is

$$\begin{aligned}
\psi = E[\mathbf{Y}] &= \mathbf{X}_{\mathbf{MS}}(\boldsymbol{\alpha\beta}) \\
\text{var}[\mathbf{Y}] &= \mathbf{V} \\
&= \sigma_{\text{FAP}}^2\mathbf{S}_{\text{FAP}} + \sigma_{\text{FA}}^2\mathbf{S}_{\text{FA}} + \sigma_{\text{F}}^2\mathbf{S}_{\text{F}} \\
&= \sigma_{\text{FAP}}^2\mathbf{M}_{\text{FAP}} + \tfrac{36}{4\times3}\sigma_{\text{FA}}^2\mathbf{M}_{\text{FA}} + \tfrac{36}{4}\sigma_{\text{F}}^2\mathbf{M}_{\text{F}} \\
&= \sigma_{\text{FAP}}^2\mathbf{M}_{\text{FAP}} + 3\sigma_{\text{FA}}^2\mathbf{M}_{\text{FA}} + 9\sigma_{\text{F}}^2\mathbf{M}_{\text{F}}.
\end{aligned}$$

Now if the observations are arranged in standard order with respect to the factors Factories, Areas and then Parts, we can use rule XI.1 to obtain the following direct-product expression for **V**.

$$\begin{aligned}
\mathbf{V} &= \sigma_{\text{FAP}}^2\mathbf{S}_{\text{FAP}} + \sigma_{\text{FA}}^2\mathbf{S}_{\text{FA}} + \sigma_{\text{F}}^2\mathbf{S}_{\text{F}} \\
&= \sigma_{\text{FAP}}^2\mathbf{I}_4 \otimes \mathbf{I}_3 \otimes \mathbf{I}_3 + \sigma_{\text{FA}}^2\mathbf{I}_4 \otimes \mathbf{I}_3 \otimes \mathbf{J}_3 + \sigma_{\text{F}}^2\mathbf{I}_4 \otimes \mathbf{J}_3 \otimes \mathbf{J}_3.
\end{aligned}$$

In this matrix σ_{FAP}^2 will occur down the diagonal, σ_{FA}^2 will occur in 3×3 blocks down the diagonal and σ_{F}^2 will occur in 9×9 blocks down the diagonal. So this model allows for covariance between parts from the same area and a different covariance between areas from the same factory.

Actually, in Chapter IX Split-plot experiments, Factors was taken as fixed. However, the rules just given still apply, because the terms in V still only involve unrandomized factors. The expectation and variation models in this case would be

$$\begin{aligned}
\psi = E[\mathbf{Y}] &= \mathbf{X}_{\text{F}}\boldsymbol{\delta} + \mathbf{X}_{\mathbf{MS}}(\boldsymbol{\alpha\beta}) \\
\mathbf{V} &= \sigma_{\text{FAP}}^2\mathbf{S}_{\text{FAP}} + \sigma_{\text{FA}}^2\mathbf{S}_{\text{FA}} \\
&= \sigma_{\text{FAP}}^2\mathbf{I}_4 \otimes \mathbf{I}_3 \otimes \mathbf{I}_3 + \sigma_{\text{FA}}^2\mathbf{I}_4 \otimes \mathbf{I}_3 \otimes \mathbf{J}_3.
\end{aligned}$$

■

XI.B Least squares estimation of the expectation parameters in simple linear models

We want to estimate the parameters θ in the expectation model of a simple linear model by establishing their estimators. There are several different methods for doing this. A common method is the method of least squares, as in the following definition originally given in chapter I.

Definition XI.4: Let $\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ where **X** is an $n \times q$ matrix of rank $m \leq q$ and $n \geq q$, $\boldsymbol{\theta}$ is a $q \times 1$ vector of unknown parameters, $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of errors with mean **0** and variance $\sigma^2\mathbf{I}_n$. The **ordinary least squares (OLS) estimator** of $\boldsymbol{\theta}$ is the value of $\boldsymbol{\theta}$ that minimizes $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = \sum_{i=1}^{n}\varepsilon_i^2$, the sum of squares of the "errors". ■

Note that $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$ is a scalar.

a) Ordinary least squares estimators for full-rank expectation models

Lemma XI.1: Let **X** be an $n \times f$ matrix with $n \geq f$. Then $rank(\mathbf{X}) = rank(\mathbf{X}'\mathbf{X})$.

Proof: not given ■

Theorem XI.1: Let $\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ where **X** is an $n \times q$ matrix of full rank and $n \geq q$, $\boldsymbol{\theta}$ is a $q \times 1$ vector of unknown parameters, $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of errors with mean **0** and variance $\sigma^2\mathbf{I}_n$. The ordinary least squares estimator of $\boldsymbol{\theta}$ is denoted by $\hat{\boldsymbol{\theta}}$ and is given by

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

Proof: In chapter I it was shown that

$$\begin{aligned}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} &= (\mathbf{Y}-\mathbf{X}\boldsymbol{\theta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\theta})\\ &= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\boldsymbol{\theta} - \boldsymbol{\theta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\theta}'\mathbf{X}'\mathbf{X}\boldsymbol{\theta}\end{aligned}$$

and that

$$\begin{aligned}\frac{\partial\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\partial\boldsymbol{\theta}} &= \frac{\partial\left(\mathbf{Y}'\mathbf{Y} - 2(\mathbf{Y}'\mathbf{X})\boldsymbol{\theta} + \boldsymbol{\theta}'(\mathbf{X}'\mathbf{X})\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\\ &= -2(\mathbf{X}'\mathbf{Y}) + 2(\mathbf{X}'\mathbf{X})\boldsymbol{\theta}\end{aligned}$$

so that the solution is obtained by equating this expression to zero and substituting $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$. This yields the normal equations

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\theta}} = \mathbf{X}'\mathbf{Y}.$$

Then the ordinary least squares estimator of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

as claimed. ■

Also, using definition I.7, an estimator of the expected values is $\hat{\boldsymbol{\psi}} = \mathbf{X}\hat{\boldsymbol{\theta}} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$.

Now for experiments the **X** matrix consists of indicator variables and the maximal expectation model is generally the sum of terms each of which is an **X** matrix for a (generalized) factor and an associated parameter vector. In cases where the maximal expectation model involves a single term corresponding to just one generalized factor, the model is necessarily of full rank and the OLS estimator of its parameters is particularly simple. We first prove that it is of full rank and then derive the OLS estimator. But, as an example to aid in following these proofs, consider the maximal model for a two-factor factorial experiment such as was considered in chapter 7, *Factorial experiments*.

Example VII.4 2×2 Factorial experiment

Suppose A and B have two levels each and that each combination of A and B is replicated 3 times. Hence, $a = b = 2$, $r = 3$ and $n = 12$. We assume that the data is ordered for A then B then replicates — that is

$$\mathbf{Y}' = \left(Y_{111} \quad Y_{112} \quad Y_{113} \quad Y_{121} \quad Y_{122} \quad Y_{123} \quad Y_{211} \quad Y_{212} \quad Y_{213} \quad Y_{221} \quad Y_{222} \quad Y_{223}\right).$$

The maximal model involves a single term corresponding to the generalized factor A∧B — it is

$$\psi_{\mathrm{AB}} = E\left[\mathbf{Y}\right] = \mathbf{X}_{\mathrm{AB}}\left(\alpha\beta\right) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} (\alpha\beta)_{11} \\ (\alpha\beta)_{12} \\ (\alpha\beta)_{21} \\ (\alpha\beta)_{22} \end{bmatrix} = \begin{bmatrix} (\alpha\beta)_{11} \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{12} \\ (\alpha\beta)_{12} \\ (\alpha\beta)_{12} \\ (\alpha\beta)_{21} \\ (\alpha\beta)_{21} \\ (\alpha\beta)_{21} \\ (\alpha\beta)_{22} \\ (\alpha\beta)_{22} \\ (\alpha\beta)_{22} \end{bmatrix}.$$

That is, each column of $\mathbf{X}_{\mathrm{AB}}$ indicates which observations received one of the combinations of the levels of A and B.

Now, from theorem XI.1, the estimator of the expectation parameters is given by

$$\widehat{\left(\alpha\beta\right)} = \left(\mathbf{X}'_{\mathrm{AB}}\mathbf{X}_{\mathrm{AB}}\right)^{-1}\mathbf{X}'_{\mathrm{AB}}\mathbf{Y}$$

So we require $\left(\mathbf{X}'_{\mathrm{AB}}\mathbf{X}_{\mathrm{AB}}\right)^{-1}$ and $\mathbf{X}'_{\mathrm{AB}}\mathbf{Y}$.

Firstly,

$$\mathbf{X}'_{\text{AB}}\mathbf{X}_{\text{AB}} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

$$= 3\mathbf{I}_4.$$

Notice in the above that the rows of $\mathbf{X}'_{\text{AB}}$ are the columns of $\mathbf{X}_{\text{AB}}$ and that each element in the product is of the form

$$\mathbf{x}'_i\mathbf{x}_j = \sum_{k=1}^{n} x_{ki}x_{kj}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the n-vectors for the ith and jth columns of $\mathbf{X}$, respectively. That is, the sum of the pairwise products of the elements of the ith and jth columns of $\mathbf{X}$.

Clearly, $\left(\mathbf{X}'_{\text{AB}}\mathbf{X}_{\text{AB}}\right)^{-1} = \frac{1}{3}\mathbf{I}_4$.

Secondly,

$$\mathbf{X}'_{\text{AB}}\mathbf{Y} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{113} \\ Y_{121} \\ Y_{122} \\ Y_{123} \\ Y_{211} \\ Y_{212} \\ Y_{213} \\ Y_{221} \\ Y_{222} \\ Y_{223} \end{bmatrix}$$

$$= \begin{bmatrix} Y_{111} + Y_{112} + Y_{113} \\ Y_{121} + Y_{122} + Y_{123} \\ Y_{211} + Y_{212} + Y_{213} \\ Y_{221} + Y_{222} + Y_{223} \end{bmatrix} = \begin{bmatrix} Y_{11.} \\ Y_{12.} \\ Y_{21.} \\ Y_{22.} \end{bmatrix}$$

where $Y_{ij.}$ is the sum over the replicates for the ijth combination of A and B.

Finally,

$$\widehat{(\alpha\beta)} = \left(\mathbf{X}'_{\text{AB}}\mathbf{X}_{\text{AB}}\right)^{-1}\mathbf{X}'_{\text{AB}}\mathbf{Y}$$

$$= \begin{bmatrix} \frac{1}{3} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} Y_{11.} \\ Y_{12.} \\ Y_{21.} \\ Y_{22.} \end{bmatrix}$$

$$= \begin{bmatrix} \overline{Y}_{11.} \\ \overline{Y}_{12.} \\ \overline{Y}_{21.} \\ \overline{Y}_{22.} \end{bmatrix}$$

where $\overline{Y}_{ij.}$ is the mean over the replicates for the *ij*th combination of A and B.
Also, the estimator of the expected values is:

$$\hat{\boldsymbol{\psi}} = \mathbf{X}_{\text{AB}}\widehat{(\alpha\beta)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \overline{Y}_{11.} \\ \overline{Y}_{12.} \\ \overline{Y}_{21.} \\ \overline{Y}_{22.} \end{bmatrix} = \begin{bmatrix} \overline{Y}_{11.} \\ \overline{Y}_{11.} \\ \overline{Y}_{11.} \\ \overline{Y}_{12.} \\ \overline{Y}_{12.} \\ \overline{Y}_{12.} \\ \overline{Y}_{21.} \\ \overline{Y}_{21.} \\ \overline{Y}_{21.} \\ \overline{Y}_{22.} \\ \overline{Y}_{22.} \\ \overline{Y}_{22.} \end{bmatrix} = \left(\overline{\mathbf{A} \wedge \mathbf{B}}\right) = \mathbf{M}_{\text{AB}}\mathbf{Y}$$

where $\mathbf{M}_{\text{AB}} = \mathbf{X}_{\text{AB}}\left(\mathbf{X}'_{\text{AB}}\mathbf{X}_{\text{AB}}\right)^{-1}\mathbf{X}'_{\text{AB}}$ is the mean operator producing the *n*-vector of means for the combinations of A and B. ■

Now to the proofs.

Lemma XI.2: Let $\mathbf{Y} = \mathbf{X\theta} + \boldsymbol{\varepsilon}$ where **X** is an $n \times f$ matrix corresponding to a generalized factor, F say, with *f* levels and $n \geq f$.

Then, the rank of **X** is *f* so that **X** is of full rank.

Proof: Lemma XI.1 states that $rank(\mathbf{X}) = rank(\mathbf{X}'\mathbf{X})$ and it is easier to derive the rank of $\mathbf{X}'\mathbf{X}$. To do this we need to establish a general expression for $\mathbf{X}'\mathbf{X}$ which we do by considering the form of **X**. The *n* rows of **X** correspond to the observational units in the experiment and the *f* columns to the levels of the generalized factor. All the elements of **X** are either zero or one with a one occurring in the *i*th column for those units with the *i*th level of the generalized factor so that, if the *i*th level of the generalized factor is replicated g_i times, there will be g_i ones in the *i*th column. In each row of **X** there will be a single one as only one level of the generalized factor can be observed with each unit — if it is the *i*th level of the generalized factor that occurs on the unit, then the one will occur in the *i*th column of **X**.

Considering the $f \times f$ matrix $\mathbf{X}'\mathbf{X}$, its *ij*th element is the product of the *i*th column, as a row vector, with the *j*th column as given in the following expression (see example above):

$$\mathbf{x}_i'\mathbf{x}_j = \sum_{k=1}^{n} x_{ki}x_{kj}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the *n*-vectors for the *i*th and *j*th columns of **X**, respectively

That is, the result is the sum of the pairwise products of the elements of the *i*th and *j*th columns of **X**. When $i = j$, the two columns are the same and the products are either of two ones or two zeros and so that the product is one for the units that have the *i*th level of the generalized factor and the sum will be g_i. When $i \neq j$, we have the product of the vector for the *i*th column, as a row vector, with that for the *j*th column. Now as each unit can receive only one level for the generalized factor, the products are either two zeros or a one and a zero. Consequently, the sum is zero.

It is concluded that $\mathbf{X}'\mathbf{X}$ is a $f \times f$ diagonal matrix with diagonal elements g_i. Being a diagonal matrix with all nonzero, diagonal elements, its rank is equal to its order *f*.

Hence, in this case, the $rank(\mathbf{X}) = rank(\mathbf{X}'\mathbf{X}) = f$ and so **X** is of full rank. ■

Theorem XI.2: Let $\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ where **X** is an $n \times f$ matrix corresponding to a generalized factor, F say, with *f* levels and $n \geq f$, $\boldsymbol{\theta}$ is a $f \times 1$ vector of unknown parameters, $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of errors with mean **0** and variance $\sigma^2\mathbf{I}_n$.

The ordinary least squares estimator of $\boldsymbol{\theta}$ is denoted by $\hat{\boldsymbol{\theta}}$ and is given by

$$\hat{\boldsymbol{\theta}} = \ddot{\overline{\mathbf{F}}}_{f\times 1}$$

where $\ddot{\overline{\mathbf{F}}}_{f\times 1}$ is the *f*-vector of means for the levels of the generalized factor F.

Notation: the double over-dots are used to indicate that the vector is of length *f*, the number of levels, rather than *n*, the number of units. An *n*-vector has no over-dots.

Proof: From theorem XI.1 we have that the OLS estimator is $\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Hence, we need to establish expressions for $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{Y}$. In the proof of lemma XI.2 it was shown that $\mathbf{X}'\mathbf{X}$ is a $f \times f$ diagonal matrix with diagonal elements g_i. Consequently, $(\mathbf{X}'\mathbf{X})^{-1}$ is a $f \times f$ diagonal matrix with diagonal elements $\frac{1}{g_i}$.

Now $\mathbf{X}'\mathbf{Y}$ is an *f*-vector whose *i*th element is the total for the *i*th level of the generalized factor of the elements of **Y**. This can be seen when it is realized that the *i*th element of $\mathbf{X}'\mathbf{Y}$ is the product of the *i*th column of **X** with **Y**. Clearly, the result will be the sum of the elements of **Y** corresponding to elements of the *i*th column of **X** that are one — those units with the *i*th level of the generalized factor.

Finally, taking the product of $(\mathbf{X}'\mathbf{X})^{-1}$ with $\mathbf{X}'\mathbf{Y}$ divides each total by its replication forming the *f*-vector of means as stated. ■

Corollary XI.1: With the model for **Y** as in theorem XI.2, the estimator of the expected values is given by

$$\hat{\boldsymbol{\psi}} = \mathbf{X}\hat{\boldsymbol{\theta}} = \overline{\mathbf{F}}_{n\times 1} = \mathbf{M}_{\mathrm{F}}\mathbf{Y}$$

where $\mathbf{F}_{n\times 1}$ is the n-vector an element of which is the mean for the level of the generalized factor F for the corresponding unit and $\mathbf{M}_{\mathrm{F}}$ is the mean operator that replaces each observation in **Y** with the mean for the corresponding level of the generalized factor F.

Proof: Now $\hat{\boldsymbol{\psi}} = \mathbf{X}\hat{\boldsymbol{\theta}}$ and it was pointed out in the proof of lemma XI.2 that each row of **X** contains a single one which, if it is the *i*th level of the generalized factor that occurs on the unit, then the one will occur in the *i*th column of **X**. So the corresponding element of $\hat{\boldsymbol{\psi}} = \mathbf{X}\hat{\boldsymbol{\theta}}$ will be the *i*th element of $\hat{\boldsymbol{\theta}} = \ddot{\overline{\mathbf{F}}}_{f\times 1}$ which is the mean for the *i*th level of the generalized factor F and $\hat{\boldsymbol{\psi}} = \overline{\mathbf{F}}_{n\times 1}$.

To show that $\hat{\boldsymbol{\psi}} = \mathbf{M}_{\mathrm{F}}\mathbf{Y}$, we have from theorem XI.1 that $\hat{\boldsymbol{\psi}} = \mathbf{X}\hat{\boldsymbol{\theta}} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$. Letting $\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' = \mathbf{M}_{\mathrm{F}}$ yields $\hat{\boldsymbol{\psi}} = \mathbf{M}_{\mathrm{F}}\mathbf{Y}$. Now, given $\hat{\boldsymbol{\psi}} = \overline{\mathbf{F}}_{n\times 1}$, $\mathbf{M}_{\mathrm{F}}$ must be the mean operator that replaces each observation in **Y** with the mean for the corresponding level of the generalized factor F. ■

Example XI.2 Rat experiment

In example II.1, we had $\mathbf{Y} = \mathbf{X}_{\mathrm{T}}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$ with

$$\mathbf{X}_{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix}$$

so that the model in this case involves a single, original factor whose levels are unequally replicated.

Now,

$$\mathbf{X}_{\mathrm{T}}'\mathbf{X}_{\mathrm{T}} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{X}_{\mathrm{T}}'\mathbf{Y} = \begin{bmatrix} Y_1 + Y_2 + Y_3 \\ Y_4 + Y_5 \\ Y_6 \end{bmatrix}$$

and

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{X}_T'\mathbf{X}_T\right)^{-1}\mathbf{X}_T'\mathbf{Y}$$
$$= \begin{bmatrix} \tfrac{1}{3} & 0 & 0 \\ 0 & \tfrac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} Y_1 + Y_2 + Y_3 \\ Y_4 + Y_5 \\ Y_6 \end{bmatrix}$$
$$= \begin{bmatrix} \dfrac{Y_1 + Y_2 + Y_3}{3} \\ \dfrac{Y_4 + Y_5}{2} \\ \dfrac{Y_6}{1} \end{bmatrix}$$
$$= \begin{bmatrix} \bar{Y}_A \\ \bar{Y}_B \\ \bar{Y}_C \end{bmatrix}.$$

Finally,

$$\hat{\boldsymbol{\psi}} = \mathbf{X}_T\hat{\boldsymbol{\alpha}} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \bar{Y}_A \\ \bar{Y}_B \\ \bar{Y}_C \end{bmatrix} = \begin{bmatrix} \bar{Y}_A \\ \bar{Y}_A \\ \bar{Y}_A \\ \bar{Y}_B \\ \bar{Y}_B \\ \bar{Y}_C \end{bmatrix} = \bar{\mathbf{T}}.$$

This confirms that, as asserted in chapter **II**, *Completely Randomized Design*, the estimator of the expected values are the Diet means, the vector of which is denoted $\bar{\mathbf{T}}$. ■

b) Ordinary least squares estimators for less-than-full-rank expectation models

The model for the expectation is still of the form $E[\mathbf{Y}] = \mathbf{X}\theta$ but in this case the rank of **X** is less than the number of columns. Now $rank(\mathbf{X}) = rank(\mathbf{X}'\mathbf{X}) = m < q$ and so there is no inverse of $\mathbf{X}'\mathbf{X}$ to use in solving the normal equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\theta}} = \mathbf{X}'\mathbf{Y}$. We will demonstrate that in spite of this complication, and the difficulties that ensure, the estimators of the expected values remain simple function of means. For example, the maximal, less-than-full-rank, expectation model for an RCBD is:

$$\psi_{B+T} = E[\mathbf{Y}] = \mathbf{X}_B\beta + \mathbf{X}_T\tau.$$

For this case we will show that $\hat{\boldsymbol{\psi}} = \bar{\mathbf{B}} + \bar{\mathbf{T}} - \bar{\mathbf{G}}$. Mind you it will take some effort.

The differences less-than-full-rank and full-rank models are as follows:

1. In the full rank model it is assumed that the parameters specified in the model are unique. That is there exists exactly one set of real numbers $\{\theta_1, \theta_2, \ldots, \theta_q\}$ that describes the system.
   In the less-than-full-rank model there are infinitely many sets of real numbers that describe the system. For our example there are infinitely many choices for $\{\beta_1, \ldots, \beta_b, \tau_1, \ldots, \tau_t\}$. The model parameters are said to be **nonidentifiable**.

2. In the full rank model $\mathbf{X'X}$ is nonsingular and so $(\mathbf{X'X})^{-1}$ exists and the normal equations have the one solution $\hat{\theta} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$.
   In a less-than-full-rank model there are infinitely many solutions to the normal equations.
3. In the full rank model all linear functions of $\{\theta_1, \theta_2, \ldots, \theta_q\}$ can be estimated unbiasedly.
   In the less-than-full-rank model this is not the case: some functions can while others cannot. It will be our task to identify those functions whose estimated values are the same regardless of the estimators used to estimate them.

Example XI.3 The RCBD

The simplest model that is less than full rank is the maximal expectation model for the RCBD, E[Y] = Blocks + Treats. Generally, the RCBD involves b blocks in each of which t treatments are observed so that there are $n = b \times t$ observations in all. The maximal model used for an RCBD, in matrix terms, is:

$$\psi_{\mathrm{B+T}} = E[\mathbf{Y}] = \mathbf{X}_{\mathrm{B}}\beta + \mathbf{X}_{T}\tau \text{ and } \mathrm{var}[\mathbf{Y}] = \sigma^2\mathbf{I}_n.$$

where $\mathbf{Y}$ is the n-vector of response variables,

β is the b-vector of parameters specifying a different mean response for each block,

$\mathbf{X}_{\mathrm{B}}$ is the $n \times b$ matrix indicating the block from which an observation came,

τ is the t-vector of parameters specifying a different mean response for each treatment,

$\mathbf{X}_{\mathrm{T}}$ is the $n \times t$ matrix indicating the observations that received each of the treatments.

However, as previously mentioned, the model is not of full rank because the sums of the columns of both $\mathbf{X}_{\mathrm{B}}$ and $\mathbf{X}_{\mathrm{T}}$ are equal to $\mathbf{1}_n$. Consequently, for $\mathbf{X} = [\mathbf{X}_{\mathrm{B}} \quad \mathbf{X}_{\mathrm{T}}]$, $rank(\mathbf{X}) = rank(\mathbf{X'X}) = b + t - 1$.

To illustrate that the parameters are nonidentifiable, suppose an experiment with 2 blocks and 2 treatments is conducted and the following information is known about the parameters:

$$\beta_1 + \tau_1 = 10$$
$$\beta_1 + \tau_2 = 15$$
$$\beta_2 + \tau_1 = 12$$
$$\beta_2 + \tau_2 = 17.$$

Then the parameters are not identifiable for

1. if $\beta_1 = 5$, then $\tau_1 = 5$, $\tau_2 = 10$ and $\beta_2 = 7$;
2. if $\beta_1 = 6$, then $\tau_1 = 4$, $\tau_2 = 9$ and $\beta_2 = 8$.

Clearly we can pick a value for any one of the parameters and then find values of the other parameters that satisfy the above equations and so there are infinitely many possible values for the parameters. However, no matter what values are taken for β_1, β_2, τ_1 and τ_2, the value of $\beta_2 - \beta_1 = 2$ and of $\tau_2 - \tau_1 = 5$; these functions are invariant. ■

To estimate the parameters of a less than full rank model we need to extend the estimation theory presented in section a), *Ordinary least squares estimators for full-rank expectation models*. This will involve obtaining solutions to the normal equations using generalized inverses.

Introduction to generalized inverses

Suppose that we have a system of n linear equations in q unknowns such as $\mathbf{Ax} = \mathbf{y}$ where $\mathbf{A}$ is an $n \times q$ matrix of real numbers, $\mathbf{x}$ is a q-vector of unknowns and $\mathbf{y}$ is a n-vector of real numbers. There are 3 possibilities:

1. the system is inconsistent and has no solution;
2. the system is consistent and has exactly one solution;
3. the system is consistent and has many solutions.

Consider the following equations:

$$x_1 + 2x_2 = 7$$
$$3x_1 + 6x_2 = 21.$$

They are consistent in that the solution of one will also satisfy the second because the second equation is just 3 times the first. However, the following equations are inconsistent in that it is impossible to satisfy both at the same time:

$$x_1 + 2x_2 = 7$$
$$3x_1 + 6x_2 = 24.$$

Generally, to be consistent, any linear relations on the left of the equations must also be satisfied by the righthand side of the equation.

Theorem XI.3: Let $E[\mathbf{Y}] = \mathbf{X}\theta$ be a linear model for the expectation. Then the system of normal equations

$$(\mathbf{X}'\mathbf{X})\hat{\theta} = \mathbf{X}'\mathbf{Y}$$

is consistent.

Proof: not given. ■

Definition XI.5: Let $\mathbf{A}$ be an $n \times q$ matrix. A $q \times n$ matrix $\mathbf{A}^-$ such that

$$\mathbf{AA}^-\mathbf{A} = \mathbf{A}$$

is called a **generalised inverse** (g-inverse for short) for $\mathbf{A}$. ■

Any matrix $\mathbf{A}$ has a generalized inverse but it is not unique unless $\mathbf{A}$ is nonsingular, in which case $\mathbf{A}^- = \mathbf{A}^{-1}$.

Example XI.4 Generalized inverse of a 2×2 matrix

Take $\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$, then $\mathbf{A}^{-} = \begin{bmatrix} 1/2 & 0 \\ 0 & 0 \end{bmatrix}$

since $\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$.

It is easy to see that the following matrices are also generalized inverse for **A**:

$$\begin{bmatrix} 1/2 & 0 \\ 1 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1/2 & 0 \\ 4 & -3 \end{bmatrix}.$$

This illustrates that generalized inverses are not necessarily unique.

An algorithm for finding generalized inverses

To find a generalized inverse $\mathbf{A}^{-}$ for an $n \times q$ matrix **A** of rank *m*:

1. Find any $m \times m$ minor **H** of **A** i.e. the matrix obtained by selecting any *m* rows and any *m* columns of **A**.
2. Find $\mathbf{H}^{-1}$.
3. Replace **H** in **A** with $\left(\mathbf{H}^{-1}\right)'$.
4. Replace all other entries in **A** with zeros.
5. Transpose the resulting matrix.

Properties of generalized inverses

Let **A** be an $n \times q$ matrix of rank *m* with $n \geq q \geq m$. Then

1. $\mathbf{A}^{-}\mathbf{A}$ and $\mathbf{A}\mathbf{A}^{-}$ are idempotent.
2. $rank\left(\mathbf{A}^{-}\mathbf{A}\right) = rank\left(\mathbf{A}\mathbf{A}^{-}\right) = m$.
3. If $\mathbf{A}^{-}$ is a generalized inverse of **A**, then $\left(\mathbf{A}^{-}\right)'$ is a generalized inverse of $\mathbf{A}'$; that is, $\left(\mathbf{A}^{-}\right)' = \left(\mathbf{A}'\right)^{-}$.
4. $\left(\mathbf{A}'\mathbf{A}\right)^{-}\mathbf{A}'$ is a generalized inverse of **A** so that $\mathbf{A} = \mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-}\left(\mathbf{A}'\mathbf{A}\right)$ and $\mathbf{A}' = \left(\mathbf{A}'\mathbf{A}\right)\left(\mathbf{A}'\mathbf{A}\right)^{-}\mathbf{A}'$.
5. $\mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-}\mathbf{A}'$ is unique, symmetric and idempotent as it is invariant to the choice of a generalized inverse. Furthermore $rank\left(\mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-}\mathbf{A}'\right) = m$.

Generalized inverses and the normal equations

Lemma XI.3: Let $\mathbf{Ax} = \mathbf{y}$ be consistent. Then $\mathbf{x} = \mathbf{A}^{-}\mathbf{y}$ is a solution to the system where $\mathbf{A}^{-}$ is any generalized inverse for **A**.

Proof: not given. ■

Theorem XI.4: Let
$$\mathbf{Y} = \mathbf{X\theta} + \boldsymbol{\varepsilon}$$
where **X** is an $n \times q$ matrix of rank $m \le q$ with $n \ge q$, $\boldsymbol{\theta}$ is a $q \times 1$ vector of unknown parameters, $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of errors with mean **0** and variance $\sigma^2\mathbf{I}_n$.

The ordinary least squares estimator of $\boldsymbol{\theta}$ is denoted by $\hat{\boldsymbol{\theta}}$ and is given by
$$\hat{\boldsymbol{\theta}} = (\mathbf{X'X})^{-}\mathbf{X'Y}.$$

Proof: In our proof for the full-rank case, we showed that the ordinary least squares estimator is the solution of the normal equations $\mathbf{X'X}\hat{\boldsymbol{\theta}} = \mathbf{X'Y}$. Nothing in our derivation of this result depended on the rank of **X**. So the ordinary least squares estimator in the less-than-full-rank case is still a solution to the normal equations.

Applying theorem XI.3 and lemma XI.3 to the normal equations $(\mathbf{X'X})\hat{\theta} = \mathbf{X'Y}$, it is easy to see that
$$\hat{\theta} = (\mathbf{X'X})^{-}\mathbf{X'Y}$$

is a solution of the normal equations as claimed. ■

So any generalized inverse of $\mathbf{X'X}$ will generate a solution, however, different generalized inverses result in different solutions.

Theorem XI.5: Let $\mathbf{Ax} = \mathbf{y}$ be consistent and let $\mathbf{A}^{-}$ be any generalized inverse for **A**. Then
$$\mathbf{x}_0 = \mathbf{A}^{-}\mathbf{y} + \left(\mathbf{I} - \mathbf{A}^{-}\mathbf{A}\right)\mathbf{z}$$
is a solution to the system where **z** is an arbitrary *q*-vector.

Proof: not given. ■

In the context of linear models this theorem implies that every vector of the form

$$\hat{\theta} = (\mathbf{X'X})^{-}\mathbf{X'Y} + \left\{\mathbf{I}_q - (\mathbf{X'X})^{-}(\mathbf{X'X})\right\}\mathbf{z}$$

where $(\mathbf{X'X})^{-}$ is a generalized inverse for $\mathbf{X'X}$ and **z** is arbitrary, is a solution to the normal equations.

If $\mathbf{z}=\mathbf{0}$, we get the solution obtained via lemma XI.3 and if $\mathbf{X'X}$ is of full rank then we obtain the usual least squares estimators. It can also be proved that any solution can be written in this form.

Our next theorem derives the ordinary least squares estimators for the less-than-full-rank model for the randomized complete block design. But, first a useful lemma.

Lemma XI.4: The inverse of $a\left(\mathbf{I}_m-\frac{1}{m}\mathbf{J}_m\right)+b\left(\frac{1}{m}\mathbf{J}_m\right)$ is $\dfrac{1}{a}\left(\mathbf{I}_m-\frac{1}{m}\mathbf{J}_m\right)+\dfrac{1}{b}\left(\frac{1}{m}\mathbf{J}_m\right)$ where a and b are scalars.

Proof: First show that $\left(\frac{1}{m}\mathbf{J}_m\right)^2=\frac{1}{m}\mathbf{J}_m$, $\left(\mathbf{I}_m-\frac{1}{m}\mathbf{J}_m\right)^2=\mathbf{I}_m-\frac{1}{m}\mathbf{J}_m$ and $\left(\mathbf{I}_m-\frac{1}{m}\mathbf{J}_m\right)\left(\frac{1}{m}\mathbf{J}_m\right)=\mathbf{0}$. Consequently,

$$\left\{a\left(\mathbf{I}_m-\tfrac{1}{m}\mathbf{J}_m\right)+b\left(\tfrac{1}{m}\mathbf{J}_m\right)\right\}\left\{\frac{1}{a}\left(\mathbf{I}_m-\tfrac{1}{m}\mathbf{J}_m\right)+\frac{1}{b}\left(\tfrac{1}{m}\mathbf{J}_m\right)\right\}=\left(\mathbf{I}_m-\tfrac{1}{m}\mathbf{J}_m\right)+\left(\tfrac{1}{m}\mathbf{J}_m\right)$$
$$=\mathbf{I}_m$$

so that $\dfrac{1}{a}\left(\mathbf{I}_m-\frac{1}{m}\mathbf{J}_m\right)+\dfrac{1}{b}\left(\frac{1}{m}\mathbf{J}_m\right)$ is the inverse of $a\left(\mathbf{I}_m-\frac{1}{m}\mathbf{J}_m\right)+b\left(\frac{1}{m}\mathbf{J}_m\right)$. ■

You should think of **J** matrices as “sum-everything” matrices.

Theorem XI.6: Let **Y** be a n-vector of jointly-distributed random variables with

$$\psi=E\left[\mathbf{Y}\right]=\mathbf{X}_{\mathrm{B}}\beta+\mathbf{X}_{\mathrm{T}}\tau \text{ and } \operatorname{var}\left[\mathbf{y}\right]=\sigma^2\mathbf{I}$$

where β is the b-vector of parameters specifying a different mean response for each block,

$\mathbf{X}_{\mathrm{B}}$ is the $n\times b$ matrix indicating the block from which an observation came,

τ is the t-vector of parameters specifying a different mean response for each treatment,

$\mathbf{X}_{\mathrm{T}}$ is the $n\times t$ matrix indicating the observations that received each of the treatments.

Then a nonunique estimator of θ, obtained by deleting the last row and column of **X** in computing the g-inverse, is

$$\hat{\theta}=\begin{bmatrix}\ddot{\mathbf{B}}_{b\times 1}+\bar{T}_t\mathbf{1}_b-\bar{Y}\mathbf{1}_b\\ \dddot{\mathbf{T}}_{(t-1)\times 1}-\bar{T}_t\mathbf{1}_{(t-1)}\\ 0\end{bmatrix}$$

where $\ddot{\mathbf{B}}_{b\times 1}$ is the b-vector of block means of **Y**, $\dddot{\mathbf{T}}_{(t-1)\times 1}$ is the $(t-1)$-vector of the first t–1 treatment means of **Y,** $\bar{T}_t$ is the mean of **Y** for treatment t and $\bar{Y}$ is the grand mean of **Y**.

Notation: The triple over-dots are used to indicate a modified, double over-dots vector. In this case, whereas the double over-dots vector $\ddot{\bar{\mathbf{T}}}_{t\times 1}$ has t elements, corresponding to the levels of Treatments, the triple over-dots vector $\dddot{\bar{\mathbf{T}}}_{(t-1)\times 1}$ has the last element of $\ddot{\bar{\mathbf{T}}}_{t\times 1}$ removed.

Proof: According to theorem XI.4,

$$\hat{\theta} = \left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{Y}$$

and so we require a generalized inverse of $\mathbf{X}'\mathbf{X}$ where

$$\mathbf{X} = \begin{bmatrix}\mathbf{X}_{\mathrm{B}} & \mathbf{X}_{\mathrm{T}}\end{bmatrix}\text{, so that } \mathbf{X}'\mathbf{X} = \begin{bmatrix}\mathbf{X}'_{\mathrm{B}} \\ \mathbf{X}'_{\mathrm{T}}\end{bmatrix}\begin{bmatrix}\mathbf{X}_{\mathrm{B}} & \mathbf{X}_{\mathrm{T}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}'_{\mathrm{B}}\mathbf{X}_{\mathrm{B}} & \mathbf{X}'_{\mathrm{B}}\mathbf{X}_{\mathrm{T}} \\ \mathbf{X}'_{\mathrm{T}}\mathbf{X}_{\mathrm{B}} & \mathbf{X}'_{\mathrm{T}}\mathbf{X}_{\mathrm{T}}\end{bmatrix} = \begin{bmatrix} t\mathbf{I}_b & \mathbf{J}_{b\times t} \\ \mathbf{J}_{t\times b} & b\mathbf{I}_t\end{bmatrix}.$$

Note that $\mathbf{J}_{b\times t}$ reflects the number of times that each treatment occurs in each block.

Our algorithm tells us that we must take a minor of order $b+t-1$ and obtain its inverse. So we delete any row and column and find its inverse. Suppose we omit the last row and column. To remove the last row and column of $\mathbf{X}'\mathbf{X}$, let $\mathbf{Z} = \begin{bmatrix}\mathbf{X}_{\mathrm{B}} & \mathbf{X}_{\mathrm{T-1}}\end{bmatrix}$ where $\mathbf{X}_{\mathrm{T}-1}$ is the $n\times(t-1)$ matrix formed by taking the first $t-1$ columns of $\mathbf{X}_{\mathrm{T}}$. Then,

$$\mathbf{Z}'\mathbf{Z} = \begin{bmatrix}\mathbf{X}'_{\mathrm{B}}\mathbf{X}_{\mathrm{B}} & \mathbf{X}'_{\mathrm{B}}\mathbf{X}_{\mathrm{T}-1} \\ \mathbf{X}'_{\mathrm{T}-1}\mathbf{X}_{\mathrm{B}} & \mathbf{X}'_{\mathrm{T}-1}\mathbf{X}_{\mathrm{T}-1}\end{bmatrix} = \begin{bmatrix} t\mathbf{I}_b & \mathbf{J}_{b\times(t-1)} \\ \mathbf{J}_{(t-1)\times b} & b\mathbf{I}_{(t-1)}\end{bmatrix}$$

is $\mathbf{X}'\mathbf{X}$ with its last row and column removed. Now $\mathbf{Z}'\mathbf{Z}$ is clearly symmetric so that its inverse will be also. So our generalized inverse will be $\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}$ with a row and column of zeros added. Now, to find the inverse of $\mathbf{Z}'\mathbf{Z}$, we note that for a partitioned, symmetric matrix $\mathbf{H}$ where

$$\mathbf{H} = \begin{bmatrix}\mathbf{A} & \mathbf{C} \\ \mathbf{C}' & \mathbf{D}\end{bmatrix} \text{ then } \mathbf{H}^{-1} = \left(\mathbf{H}^{-1}\right)' = \begin{bmatrix}\mathbf{U} & \mathbf{V} \\ \mathbf{V}' & \mathbf{W}\end{bmatrix} = \begin{bmatrix}\mathbf{A}^{-1} - \mathbf{V}\mathbf{C}'\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{C}\mathbf{W} \\ -\mathbf{W}\mathbf{C}'\mathbf{A}^{-1} & \left(\mathbf{D} - \mathbf{C}'\mathbf{A}^{-1}\mathbf{C}\right)^{-1}\end{bmatrix}.$$

So we first need to find **W**, then **V** and finally **U**. In our case,

$$\begin{aligned}
\mathbf{W} &= \left(\mathbf{D} - \mathbf{C}'\mathbf{A}^{-1}\mathbf{C}\right)^{-1} \\
&= \left(b\mathbf{I}_{(t-1)} - \mathbf{J}_{(t-1)\times b}\tfrac{1}{t}\mathbf{I}_b\mathbf{J}_{b\times(t-1)}\right)^{-1} = \left(b\mathbf{I}_{(t-1)} - \tfrac{b}{t}\mathbf{J}_{(t-1)}\right)^{-1} = \tfrac{1}{b}\left(\mathbf{I}_{(t-1)} - \left(\tfrac{t-1}{t}\right)\tfrac{1}{t-1}\mathbf{J}_{(t-1)}\right)^{-1} \\
&= \tfrac{1}{b}\left(\mathbf{I}_{(t-1)} - \left(1-\tfrac{1}{t}\right)\tfrac{1}{t-1}\mathbf{J}_{(t-1)}\right)^{-1} = \tfrac{1}{b}\left(\left\{\mathbf{I}_{(t-1)} - \tfrac{1}{t-1}\mathbf{J}_{(t-1)}\right\} + \tfrac{1}{t}\left(\tfrac{1}{t-1}\mathbf{J}_{(t-1)}\right)\right)^{-1} \\
&= \tfrac{1}{b}\left(\left\{\mathbf{I}_{(t-1)} - \tfrac{1}{t-1}\mathbf{J}_{(t-1)}\right\} + t\left(\tfrac{1}{t-1}\mathbf{J}_{(t-1)}\right)\right) = \tfrac{1}{b}\left(\mathbf{I}_{(t-1)} + \tfrac{t-1}{t-1}\mathbf{J}_{(t-1)}\right) = \tfrac{1}{b}\left(\mathbf{I}_{(t-1)} + \mathbf{J}_{(t-1)}\right).
\end{aligned}$$

Note that lemma XI.4 is used to obtain the inverse at the end of the third line in the above equation.

$$\mathbf{V} = -\mathbf{A}^{-1}\mathbf{C}\mathbf{W} = -\tfrac{1}{t}\mathbf{I}_b\mathbf{J}_{b\times(t-1)}\tfrac{1}{b}\left(\mathbf{I}_{(t-1)} + \mathbf{J}_{(t-1)}\right) = -\tfrac{1}{bt}\left(\mathbf{J}_{b\times(t-1)} + (t-1)\mathbf{J}_{b\times(t-1)}\right) = -\tfrac{1}{b}\mathbf{J}_{b\times(t-1)}$$

$$\mathbf{U} = \mathbf{A}^{-1} - \mathbf{V}\mathbf{C}'\mathbf{A}^{-1} = \tfrac{1}{t}\mathbf{I}_b + \tfrac{1}{b}\mathbf{J}_{b\times(t-1)}\mathbf{J}_{(t-1)\times b}\tfrac{1}{t}\mathbf{I}_b = \tfrac{1}{t}\mathbf{I}_b + \tfrac{t-1}{bt}\mathbf{J}_b.$$

Consequently,

$$\left(\mathbf{Z}'\mathbf{Z}\right)^{-1} = \begin{bmatrix} \frac{1}{t}\mathbf{I}_b + \frac{t-1}{bt}\mathbf{J}_b & -\frac{1}{b}\mathbf{J}_{b\times(t-1)} \\ -\frac{1}{b}\mathbf{J}_{(t-1)\times b} & \frac{1}{b}\left(\mathbf{I}_{(t-1)} + \mathbf{J}_{(t-1)}\right) \end{bmatrix}$$

and

$$\left(\mathbf{X}'\mathbf{X}\right)^{-} = \begin{bmatrix} \frac{1}{t}\mathbf{I}_b + \frac{(t-1)}{bt}\mathbf{J}_b & -\frac{1}{b}\mathbf{J}_{b\times(t-1)} & \mathbf{0}_{b\times1} \\ -\frac{1}{b}\mathbf{J}_{(t-1)\times b} & \frac{1}{b}\left(\mathbf{I}_{(t-1)} + \mathbf{J}_{(t-1)}\right) & \mathbf{0}_{(t-1)\times1} \\ \mathbf{0}_{1\times b} & \mathbf{0}_{1\times(t-1)} & 0 \end{bmatrix}.$$

Also,

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \mathbf{X}'_{\mathrm{B}} \\ \mathbf{X}'_{\mathrm{T}} \end{bmatrix}\mathbf{Y} = \begin{bmatrix} \mathbf{X}'_{\mathrm{B}} \\ \mathbf{X}'_{\mathrm{T}-1} \\ \mathbf{x}'_t \end{bmatrix}\mathbf{Y} = \begin{bmatrix} \mathbf{X}'_{\mathrm{B}}\mathbf{Y} \\ \mathbf{X}'_{\mathrm{T}-1}\mathbf{Y} \\ \mathbf{x}'_t\mathbf{Y} \end{bmatrix} = \begin{bmatrix} t\ddot{\bar{\mathbf{B}}}_{b\times1} \\ b\dddot{\bar{\mathbf{T}}}_{(t-1)\times1} \\ b\bar{T}_t \end{bmatrix}.$$

That is, $\mathbf{X}'\mathbf{Y}$ is a vector of totals — we express it in terms of means for notational economy, a sum being equal to mean multiplied by the number of observations used in forming the mean.

Hence the nonunique estimators of θ are

$$\begin{aligned} \hat{\theta} &= \left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{Y} \\ &= \begin{bmatrix} \frac{1}{t}\mathbf{I}_b + \frac{(t-1)}{bt}\mathbf{J}_b & -\frac{1}{b}\mathbf{J}_{b\times(t-1)} & \mathbf{0}_{b\times1} \\ -\frac{1}{b}\mathbf{J}_{(t-1)\times b} & \frac{1}{b}\left(\mathbf{I}_{(t-1)} + \mathbf{J}_{(t-1)}\right) & \mathbf{0}_{(t-1)\times1} \\ \mathbf{0}_{1\times b} & \mathbf{0}_{1\times(t-1)} & 0 \end{bmatrix}\begin{bmatrix} t\ddot{\bar{\mathbf{B}}}_{b\times1} \\ b\dddot{\bar{\mathbf{T}}}_{(t-1)\times1} \\ b\bar{T}_t \end{bmatrix} \\ &= \begin{bmatrix} \left(\mathbf{I}_b + \frac{(t-1)}{b}\mathbf{J}_b\right)\ddot{\bar{\mathbf{B}}}_{b\times1} - \mathbf{J}_{b\times(t-1)}\dddot{\bar{\mathbf{T}}}_{(t-1)\times1} \\ -\frac{t}{b}\mathbf{J}_{(t-1)\times b}\ddot{\bar{\mathbf{B}}}_{b\times1} + \left(\mathbf{I}_{(t-1)} + \mathbf{J}_{(t-1)}\right)\dddot{\bar{\mathbf{T}}}_{(t-1)\times1} \\ 0 \end{bmatrix} \end{aligned}$$

Now to simplify this expression further we look to simplify $\mathbf{J}_b\ddot{\bar{\mathbf{B}}}_{b\times1}$, $\mathbf{J}_{(t-1)\times b}\ddot{\bar{\mathbf{B}}}_{b\times1}$, $\mathbf{J}_{(t-1)}\dddot{\bar{\mathbf{T}}}_{(t-1)\times1}$ and $\mathbf{J}_{b\times(t-1)}\dddot{\bar{\mathbf{T}}}_{(t-1)\times1}$.

Firstly, each element of $\mathbf{J}_b\ddot{\bar{\mathbf{B}}}_{b\times1}$ is the sum of the b Block means and $\sum_{j=1}^{b}\bar{B}_j = b\left(\sum_{j=1}^{b}\bar{B}_j \big/ b\right) = b\bar{Y}$ where $\bar{Y}$ is the grand mean of all the observations. Hence, $\mathbf{J}_b\ddot{\bar{\mathbf{B}}}_{b\times1} = b\bar{Y}\mathbf{1}_b$,
Similarly, $\mathbf{J}_{(t-1)\times b}\ddot{\bar{\mathbf{B}}}_{b\times1} = b\bar{Y}\mathbf{1}_{(t-1)}$,

Next, each element of $\mathbf{J}_{(t-1)}\dddot{\bar{\mathbf{T}}}_{(t-1)\times1}$ is the sum of the first $t-1$ Treatment means and $\sum_{i=1}^{t-1}\bar{T}_i = \left(\sum_{i=1}^{t}\bar{T}_i\right) - \bar{T}_t = t\bar{Y} - \bar{T}_t$. Hence, $\mathbf{J}_{(t-1)}\dddot{\bar{\mathbf{T}}}_{(t-1)\times1} = \left(t\bar{Y} - \bar{T}_t\right)\mathbf{1}_{(t-1)}$.

Similarly, $\mathbf{J}_{b\times(t-1)}\ddot{\ddot{\mathbf{T}}}_{(t-1)\times 1} = \left(t\bar{Y} - \bar{T}_t\right)\mathbf{1}_b$.

Therefore

$$\hat{\theta} = \begin{bmatrix} \left(\mathbf{I}_b + \frac{(t-1)}{b}\mathbf{J}_b\right)\ddot{\mathbf{B}}_{b\times 1} - \mathbf{J}_{b\times(t-1)}\dddot{\mathbf{T}}_{(t-1)\times 1} \\ -\frac{t}{b}\mathbf{J}_{(t-1)\times b}\ddot{\mathbf{B}}_{b\times 1} + \left(\mathbf{I}_{(t-1)} + \mathbf{J}_{(t-1)}\right)\dddot{\mathbf{T}}_{(t-1)\times 1} \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} \ddot{\mathbf{B}}_{b\times 1} + (t-1)\bar{Y}\mathbf{1}_b - \left(t\bar{Y} - \bar{T}_t\right)\mathbf{1}_b \\ -t\bar{Y}\mathbf{1}_{(t-1)} + \dddot{\mathbf{T}}_{(t-1)\times 1} + \left(t\bar{Y} - \bar{T}_t\right)\mathbf{1}_{(t-1)} \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} \ddot{\mathbf{B}}_{b\times 1} + \bar{T}_t\mathbf{1}_b - \bar{Y}\mathbf{1}_b \\ \dddot{\mathbf{T}}_{(t-1)\times 1} - \bar{T}_t\mathbf{1}_{(t-1)} \\ 0 \end{bmatrix}.$$

■

So the estimator is a function of block, treatment and grand means, albeit a rather messy one.

c) Estimable functions

In the less-than-full-rank model the problem we face is that θ cannot be estimated uniquely as it changes for the many different choice of $\left(\mathbf{X}'\mathbf{X}\right)^{-}$.

To overcome this problem we consider those linear functions of θ that are invariant to the choice of $\left(\mathbf{X}'\mathbf{X}\right)^{-}$ and hence to the particular $\hat{\theta}$ obtained. That is, we consider functions $\ell'\boldsymbol{\theta}$ whose estimators $\ell'\hat{\boldsymbol{\theta}}$ are invariant to different $\hat{\theta}$s in that their values remain the same regardless of which solution to the normal equations is used. This property is true for all estimable functions.

Definition XI.6: Let

$$E\left[\mathbf{Y}\right] = \mathbf{X}\theta \text{ and } \operatorname{var}\left[\mathbf{Y}\right] = \sigma^2\mathbf{I}_n$$

where $\mathbf{X}$ is $n\times q$ of rank $m \le q$.

A function $\ell'\boldsymbol{\theta}$ is said to be estimable if there exists a vector $\mathbf{c}$ such that $E\left[\mathbf{c}'\mathbf{Y}\right] = \ell'\boldsymbol{\theta}$ for all possible values of θ. ■

This definition is saying that, to be estimable, there must be some linear combination of the random variables $Y_1, Y_2, \ldots, Y_n$ that is a linear unbiased estimator of $\ell'\boldsymbol{\theta}$.

In the case of several linear functions, we can extend the above definition as follows: If **L** is a $k \times q$ matrix of constants, then $\mathbf{L\theta}$ is a set of *k* estimable functions if and only if there is a matrix **C** of constants so that $E[\mathbf{CY}] = \mathbf{L}\theta$ for all possible values of θ.

However, there is the problem of how to identify estimable functions. The following theorems give us some important cases that will cover the quantities of most interest.

But first a lemma that gives us some useful results about the expected values.

Lemma XI.5: Let

a be a $k \times 1$ vector of constants,
A an $s \times k$ matrix of constants and
Y a $k \times 1$ vector of random variables or random vector.

1. $E[\mathbf{a}] = \mathbf{a}$ and $\text{var}[\mathbf{a}] = \mathbf{0}$;
2. $E[\mathbf{a}'\mathbf{Y}] = \mathbf{a}'E[\mathbf{Y}]$ and $\text{var}[\mathbf{a}'\mathbf{Y}] = \mathbf{a}'\,\text{var}[\mathbf{Y}]\mathbf{a}$;
3. $E[\mathbf{AY}] = \mathbf{A}E[\mathbf{Y}]$ and $\text{var}[\mathbf{AY}] = \mathbf{A}\,\text{var}[\mathbf{Y}]\mathbf{A}'$. ■

Theorem XI.7: Let

$$E[\mathbf{Y}] = \mathbf{X}\theta \text{ and } \text{var}[\mathbf{Y}] = \sigma^2\mathbf{I}_n$$

where **X** is $n \times q$ of rank $m \le q$.

Then $\ell'\boldsymbol{\theta}$ is an estimable function if and only if there is a vector of constants **c** such that $\ell' = \mathbf{c}'\mathbf{X}$.

(Note that $\ell' = \mathbf{c}'\mathbf{X}$ implies that ℓ is a linear combination of the rows of **X**.)

Moreover, for a $k \times q$ matrix **L**, $\mathbf{L\theta}$ is a set of *k* estimable functions if and only if there is a matrix **C** of constants such that $\mathbf{L} = \mathbf{CX}$.

Proof: First suppose that $\ell'\boldsymbol{\theta}$ is estimable.

Then, by definition XI.6, there is a vector of constants **c** so that $E[\mathbf{c}'\mathbf{Y}] = \ell'\boldsymbol{\theta}$.

Now, using lemma XI.5, $E[\mathbf{c}'\mathbf{Y}] = \mathbf{c}'E[\mathbf{Y}] = \mathbf{c}'\mathbf{X}\boldsymbol{\theta}$, but $E[\mathbf{c}'\mathbf{Y}] = \ell'\boldsymbol{\theta}$ and so $\mathbf{c}'\mathbf{X}\boldsymbol{\theta} = \ell'\boldsymbol{\theta}$.

Since this last equation is true for all θ, it follows that $\mathbf{c}'\mathbf{X} = \ell'$.

Second suppose that $\ell' = \mathbf{c}'\mathbf{X}$. Then $\ell'\boldsymbol{\theta} = \mathbf{c}'\mathbf{X}\boldsymbol{\theta} = E[\mathbf{c}'\mathbf{Y}]$ and, by definition XI.6, $\ell'\boldsymbol{\theta}$ is estimable.

Similarly, if $E[\mathbf{CY}] = \mathbf{L\theta}$, $\mathbf{CX\theta} = \mathbf{L\theta}$ for all θ and thus $\mathbf{CX} = \mathbf{L}$.

Likewise, if $\mathbf{L} = \mathbf{CX}$, then $\mathbf{L\theta} = \mathbf{CX\theta} = E[\mathbf{CY}]$ and hence $\mathbf{L\theta}$ is estimable. ■

Theorem XI.8: Let

$$\boldsymbol{\psi} = E[\mathbf{Y}] = \mathbf{X}\theta \text{ and } \operatorname{var}[\mathbf{Y}] = \sigma^2\mathbf{I}_n$$

where $\mathbf{X}$ is $n \times q$ of rank $m \leq q$.

Each of the elements of $\boldsymbol{\psi} = \mathbf{X}\theta$ is estimable.

Proof: From theorem XI.7, $\mathbf{L\theta}$ is a set of estimable functions if $\mathbf{L} = \mathbf{CX}$.

In our case of wanting to estimate $\mathbf{X}\theta$, $\mathbf{L} = \mathbf{X}$ and so to fulfil the condition $\mathbf{L} = \mathbf{CX}$ for this case we need to find a $\mathbf{C}$ such that $\mathbf{X} = \mathbf{CX}$. Clearly, setting $\mathbf{C} = \mathbf{I}_n$ gives $\mathbf{CX} = \mathbf{X} = \mathbf{L}$ and as a consequence $\boldsymbol{\psi} = \mathbf{X}\theta$ is estimable. ■

That is, the expected value of an observation, which is a linear function $\mathbf{x}_i'\theta$ where $\mathbf{x}_i'$ is the q-vector that is the ith row of $\mathbf{X}$, is estimable.

Theorem XI.9: Let $\ell_1'\theta, \ell_2'\theta, \ldots, \ell_k'\theta$ be a collection of k estimable functions.

Let $z = a_1\ell_1'\theta + a_2\ell_2'\theta + \ldots + a_k\ell_k'\theta$ be a linear combination of these functions.

Then z is estimable.

Proof: If each of $\ell_i'\boldsymbol{\theta}, i = 1, \ldots, k,$ is estimable then, by definition XI.6, there exists a vector of constants $\mathbf{c}_i$ such that $E[\mathbf{c}_i'\mathbf{Y}] = \ell_i'\boldsymbol{\theta}$.

Then $z = \sum_{i=1}^{k} a_i\ell_i'\boldsymbol{\theta} = \sum_{i=1}^{k} a_i E[\mathbf{c}_i'\mathbf{Y}] = E\left[\sum_{i=1}^{k} a_i\mathbf{c}_i'\mathbf{Y}\right] = E[\mathbf{d}'\mathbf{Y}]$ where $\mathbf{d}' = \sum_{i=1}^{k} a_i\mathbf{c}_i'$ and so z is estimable. ■

That is, a linear combination of estimable functions is itself estimable.

Using the results in this section, we can provide a simple procedure for finding all the estimable functions. First find $\boldsymbol{\psi} = \mathbf{X}\theta$, all of whose rows are estimable functions. It can be shown that all estimable functions can be expressed as a linear combination of the entries in $\boldsymbol{\psi} = \mathbf{X}\theta$. So find all linear combinations of $\boldsymbol{\psi} = \mathbf{X}\theta$ and you have all the estimable functions.

Example XI.3 The RCBD (continued)

For the RCBD, the maximal model is $\boldsymbol{\psi} = E[\mathbf{Y}] = \mathbf{X}_{\mathrm{B}}\beta + \mathbf{X}_{\mathrm{T}}\tau$.

So for $b = 5$, $t = 4$,

$$\boldsymbol{\psi} = \begin{bmatrix} \beta_1 + \tau_1 \\ \beta_1 + \tau_2 \\ \beta_1 + \tau_3 \\ \beta_1 + \tau_4 \\ \beta_2 + \tau_1 \\ \beta_2 + \tau_2 \\ \beta_2 + \tau_3 \\ \beta_2 + \tau_4 \\ \beta_3 + \tau_1 \\ \beta_3 + \tau_2 \\ \beta_3 + \tau_3 \\ \beta_3 + \tau_4 \\ \beta_4 + \tau_1 \\ \beta_4 + \tau_2 \\ \beta_4 + \tau_3 \\ \beta_4 + \tau_4 \\ \beta_5 + \tau_1 \\ \beta_5 + \tau_2 \\ \beta_5 + \tau_3 \\ \beta_5 + \tau_4 \end{bmatrix}.$$

From this we identify the 20 basic estimable functions, $\beta_i + \tau_k$. Any linear combination of these 20 is estimable and any estimable function can be expressed as a linear combination of these. Hence $\beta_i - \beta_{i'}$ and $\tau_k - \tau_{k'}$ are estimable. On the other hand, neither β_i nor $\tau_k + \tau_{k'}$ are estimable. ■

d) Properties of estimable functions

So we have seen what estimable functions are and how to identify them, but why are they important? It turns out that they can be unbiasedly and uniquely estimated using *any* solution to the normal equations and that they are, in some sense, best. It is these properties that we establish in this section. But first a definition.

Definition XI.7: Let $\ell'\boldsymbol{\theta}$ be an estimable function and let $\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$ be a least squares estimator of θ.

Then $\ell'\hat{\boldsymbol{\theta}}$ is called the least squares estimator of $\ell'\boldsymbol{\theta}$. ■

Now the implication of this definition is that $\ell'\hat{\boldsymbol{\theta}}$ is the unique estimator of $\ell'\boldsymbol{\theta}$. This is not at all obvious because there are infinitely many solutions $\hat{\boldsymbol{\theta}}$ to the normal equations when they are less than full rank.

The next theorem establishes that the least squares estimator of an estimable function is unique.

Theorem XI.10: Let $\boldsymbol{\ell}'\boldsymbol{\theta}$ be an estimable function.

Then least squares estimator $\boldsymbol{\ell}'\hat{\boldsymbol{\theta}}$ is invariant to the least squares estimator $\hat{\boldsymbol{\theta}}$.

That is, the expression for $\boldsymbol{\ell}'\hat{\boldsymbol{\theta}}$ is the same for every expression for $\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$.

Proof: From theorem XI.7, we have $\boldsymbol{\ell}' = \mathbf{c}'\mathbf{X}$ and, from theorem XI.4, $\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$.

Using these results we get that

$$\boldsymbol{\ell}'\hat{\boldsymbol{\theta}} = \mathbf{c}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}.$$

Since, from property 5 of generalized inverses, $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ is invariant to choice of $(\mathbf{X}'\mathbf{X})^{-}$, it follows that $\mathbf{c}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$ has the same expression for any choice of $(\mathbf{X}'\mathbf{X})^{-}$.

Hence $\boldsymbol{\ell}'\hat{\boldsymbol{\theta}}$ is invariant to $(\mathbf{X}'\mathbf{X})^{-}$ and hence $\hat{\boldsymbol{\theta}}$. ■

Next we show that the least squares estimator of an estimable function is unbiased and obtain its variance.

Theorem XI.11: Let

$$E[\mathbf{Y}] = \mathbf{X}\theta \text{ and } \operatorname{var}[\mathbf{Y}] = \sigma^2\mathbf{I}_n$$

where **X** is $n \times q$ of rank $m \leq q$.

Also let $\boldsymbol{\ell}'\boldsymbol{\theta}$ be an estimable function and $\boldsymbol{\ell}'\hat{\boldsymbol{\theta}}$ be its least square estimator.

Then $\boldsymbol{\ell}'\hat{\boldsymbol{\theta}}$ is unbiased for $\boldsymbol{\ell}'\boldsymbol{\theta}$.

Moreover, $\operatorname{var}\left[\boldsymbol{\ell}'\hat{\boldsymbol{\theta}}\right] = \sigma^2\boldsymbol{\ell}'(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{\ell}$, which is invariant to choice of generalized inverse.

Proof: Since $\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$ and $\boldsymbol{\ell}' = \mathbf{c}'\mathbf{X}$ we have

$$\begin{aligned}
E\left[\boldsymbol{\ell}'\hat{\boldsymbol{\theta}}\right] &= \boldsymbol{\ell}'E\left[\hat{\boldsymbol{\theta}}\right] \\
&= \boldsymbol{\ell}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'E[\mathbf{Y}] \quad \left(\text{since } (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' \text{ is a matrix of constants}\right) \\
&= \boldsymbol{\ell}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\theta} \\
&= \mathbf{c}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\theta} \\
&= \mathbf{c}'\mathbf{X}\boldsymbol{\theta} \qquad \left(\text{from property 4 of generalized inverses}\right) \\
&= \boldsymbol{\ell}'\boldsymbol{\theta}.
\end{aligned}$$

This verifies that $\ell'\hat{\boldsymbol{\theta}}$ is unbiased for $\ell'\boldsymbol{\theta}$.

Next,

$$\begin{aligned}
\text{var}\left[\ell'\hat{\boldsymbol{\theta}}\right] &= \text{var}\left(\ell'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}\right) \\
&= \ell'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\sigma^2\mathbf{I}_n\mathbf{X}\left\{(\mathbf{X}'\mathbf{X})^{-}\right\}'\ell \quad \left(\text{since var}\left[\mathbf{a}'\mathbf{Y}\right] = \mathbf{a}'\text{var}\left[\mathbf{Y}\right]\mathbf{a} \ \& \ \text{var}[\mathbf{Y}] = \sigma^2\mathbf{I}_n\right) \\
&= \sigma^2\mathbf{c}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\left\{(\mathbf{X}'\mathbf{X})^{-}\right\}'\mathbf{X}'\mathbf{c} \\
&= \sigma^2\mathbf{c}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{c} \quad \left(\text{using } \mathbf{X}'\mathbf{X} = (\mathbf{X}'\mathbf{X})' \ \& \text{ property 3 of g-inverses}\right) \\
&= \sigma^2\mathbf{c}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{c} \quad (\text{from property 4 of generalized inverses}) \\
&= \sigma^2\ell'(\mathbf{X}'\mathbf{X})^{-}\ell \quad (\text{as } \ell' = \mathbf{c}'\mathbf{X}).
\end{aligned}$$

■

The above theorem is most important.

This follows from that fact that $\ell'\hat{\boldsymbol{\theta}}$ is not unbiased for any $\ell'\boldsymbol{\theta}$. In general,

$$E\left[\ell'\hat{\boldsymbol{\theta}}\right] = \ell'E\left[\hat{\boldsymbol{\theta}}\right] = \ell'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\theta}$$

which is not necessarily equal to $\ell'\boldsymbol{\theta}$. Thus the case when $\ell'\boldsymbol{\theta}$ is estimable, and hence $\ell'\hat{\boldsymbol{\theta}}$ is an unbiased estimator of it, is a very special case.

Definition XI.8: **Linear estimators** are estimators of the form $\mathbf{AY}$, where $\mathbf{A}$ is a matrix of real numbers. ■

That is, each element of the estimator is a linear combination of the elements of $\mathbf{Y}$. The unbiased least squares estimator developed here is an example of a linear estimator with $\mathbf{A} = \ell'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$.

Unfortunately, unbiasedness does not guarantee uniqueness. There might be more than one set of unbiased estimators of $\boldsymbol{\theta}$. Another desirable property of an estimator is that it be minimum variance.

Theorem XI.12, the Gauss-Markoff theorem, guarantees that, among the class of linear unbiased estimators of $\ell'\theta$, the least squares estimator is the best in the sense that the variances of the estimator, $\text{var}\left[\ell'\hat{\theta}\right]$, is minimized.

For this reason the least squares estimator is called the BLUE (**b**est **l**inear **u**nbiased **e**stimator).

The following lemma establishes a result that will be required in the proof of the Gauss-Markoff theorem.

Lemma XI.6: Let *U* and *V* be random variables and *a* and *b* be constants. Then

$$\operatorname{var}(aU+bV)=a^2\operatorname{var}(U)+b^2\operatorname{var}(V)+2ab\operatorname{cov}(U,V).$$

Proof:

$$\begin{aligned}\operatorname{var}(aU+bV)&=E\left[\left\{aU+bV-E[aU+bV]\right\}^2\right] \quad \text{(using definition I.5 of the variance)}\\ &=E\left[\left(\left\{aU-E[aU]\right\}+\left\{bV-E[bV]\right\}\right)^2\right]\\ &=E\left[\left\{aU-E[aU]\right\}^2+\left\{bV-E[bV]\right\}^2+2\left\{aU-E[aU]\right\}\left\{bV-E[bV]\right\}\right]\\ &=a^2E\left[\left\{U-E[U]\right\}^2\right]+b^2E\left[\left\{V-E[V]\right\}^2\right]\\ &\qquad\qquad+2abE\left[\left\{U-E[U]\right\}\left\{V-E[V]\right\}\right]\\ &=a^2\operatorname{var}(U)+b^2\operatorname{var}(V)+2ab\operatorname{cov}(U,V)\end{aligned}$$

(using definitions I.5 and 1.6 of the variance and covariance).

■

Theorem XI.12 [Gauss-Markoff]: Let

$$E[\mathbf{Y}]=\mathbf{X}\theta \text{ and } \operatorname{var}[\mathbf{Y}]=\sigma^2\mathbf{I}_n$$

where **X** is $n\times q$ of rank $m\le q$.

Let $\ell'\theta$ be estimable.

Then the best linear unbiased estimator of $\ell'\theta$ is $\ell'\hat{\theta}$ where $\hat{\theta}$ is any solution to the normal equations.

This estimator is invariant to the choice of $\hat{\theta}$.

Proof: Suppose that $\mathbf{a}'\mathbf{Y}$ is some other linear unbiased estimator of $\ell'\theta$ so that $E[\mathbf{a}'\mathbf{Y}]=\ell'\theta$.

The proof proceeds as follows:

1. Establish a result that will be used subsequently.
2. Prove that $\ell'\hat{\theta}$ is the minimum variance estimator by proving that $\operatorname{var}\left[\ell'\hat{\theta}\right]\le\operatorname{var}[\mathbf{a}'\mathbf{Y}]$.
3. Prove that $\ell'\hat{\theta}$ is the unique minimum variance estimator.

Firstly, $E[\mathbf{a}'\mathbf{Y}] = \mathbf{a}'E[\mathbf{Y}] = \mathbf{a}'\mathbf{X}\boldsymbol{\theta}$ and, since also $E[\mathbf{a}'\mathbf{Y}] = \ell'\theta$, $\mathbf{a}'\mathbf{X}\boldsymbol{\theta} = \ell'\theta$ for all θ.

Consequently, $\mathbf{a}'\mathbf{X} = \ell'$ and this is the result that will be used subsequently.

Secondly, we proceed to show that $\text{var}\left[\ell'\hat{\theta}\right] \le \text{var}[\mathbf{a}'\mathbf{Y}]$. Now, from lemma XI.6, as $\ell'\hat{\theta}$ and $\mathbf{a}'\mathbf{Y}$ are scalar random variables and $\ell'\hat{\theta} - \mathbf{a}'\mathbf{Y}$ is their linear combination with coefficients 1 and –1, we have

$$\text{var}\left[\ell'\hat{\theta} - \mathbf{a}'\mathbf{Y}\right] = \text{var}\left[\ell'\hat{\theta}\right] + \text{var}[\mathbf{a}'\mathbf{Y}] - 2\,\text{cov}\left[\ell'\hat{\theta}, \mathbf{a}'\mathbf{Y}\right].$$

But,

$$\begin{aligned}
\text{cov}\left[\ell'\hat{\theta}, \mathbf{a}'\mathbf{Y}\right] &= \text{cov}\left[\ell'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}, \mathbf{a}'\mathbf{Y}\right] \\
&= E\left[\left\{\ell'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y} - E\left[\ell'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}\right]\right\}\{\mathbf{a}'\mathbf{Y} - E[\mathbf{a}'\mathbf{Y}]\}\right] && \left(\begin{array}{l}\text{using definition I.6}\\ \text{of the covariance}\end{array}\right) \\
&= E\left[\left\{\ell'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y} - \ell'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\theta}\right\}\{\mathbf{a}'\mathbf{Y} - \mathbf{a}'\mathbf{X}\boldsymbol{\theta}\}'\right] && \left(\begin{array}{l}\text{as } E[\mathbf{Y}] = \mathbf{X}\theta \text{ and}\\ \{\mathbf{a}'\mathbf{Y} - \mathbf{a}'\mathbf{X}\boldsymbol{\theta}\} \text{ is scalar}\end{array}\right) \\
&= E\left[\ell'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\{\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}\}\left(\mathbf{a}'\{\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}\}\right)'\right] \\
&= E\left[\ell'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\{\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}\}\{\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}\}'\mathbf{a}\right] \\
&= \ell'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'E\left[(\mathbf{Y} - \mathbf{X}\boldsymbol{\theta})(\mathbf{Y} - \mathbf{X}\boldsymbol{\theta})'\right]\mathbf{a} \\
&= \ell'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\,\text{var}[\mathbf{Y}]\mathbf{a} && (\text{using definition I.4 for } \mathbf{V}) \\
&= \sigma^2\ell'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{a} && \left(\text{as var}[\mathbf{Y}] = \sigma^2\mathbf{I}_n\right) \\
&= \sigma^2\ell'(\mathbf{X}'\mathbf{X})^{-}\ell && (\text{as } \mathbf{a}'\mathbf{X} = \ell') \\
&= \text{var}\left[\ell'\hat{\boldsymbol{\theta}}\right].
\end{aligned}$$

(The last step uses theorem XI.11.)

So,

$$\begin{aligned}
\text{var}\left[\ell'\hat{\theta} - \mathbf{a}'\mathbf{Y}\right] &= \text{var}\left[\ell'\hat{\theta}\right] + \text{var}[\mathbf{a}'\mathbf{Y}] - 2\,\text{cov}\left[\ell'\hat{\theta}, \mathbf{a}'\mathbf{Y}\right] \\
&= \text{var}\left[\ell'\hat{\theta}\right] + \text{var}[\mathbf{a}'\mathbf{Y}] - 2\,\text{var}\left[\ell'\hat{\theta}\right] \\
&= \text{var}[\mathbf{a}'\mathbf{Y}] - \text{var}\left[\ell'\hat{\theta}\right]
\end{aligned}$$

Since $\text{var}\left[\ell'\hat{\theta} - \mathbf{a}'\mathbf{Y}\right] \ge 0$, $\text{var}[\mathbf{a}'\mathbf{Y}] \ge \text{var}\left[\ell'\hat{\theta}\right]$. Hence, $\ell'\hat{\theta}$ has minimum variance amongst all linear unbiased estimators.

Finally, to demonstrate that $\ell'\hat{\theta}$ is the only linear unbiased estimator with minimum variance, we suppose that $\text{var}[\mathbf{a}'\mathbf{Y}] = \text{var}\left[\ell'\hat{\theta}\right]$.

Then $\text{var}\left[\ell'\hat{\theta} - \mathbf{a}'\mathbf{Y}\right] = 0$ and hence $\ell'\hat{\theta} - \mathbf{a}'\mathbf{Y}$ is a constant, say *c*.

Since $E\left[\ell'\hat{\theta}\right] = E[\mathbf{a}'\mathbf{Y}]$, $E\left[\ell'\hat{\theta}\right] - E[\mathbf{a}'\mathbf{Y}] = 0$ so that $E\left[\ell'\hat{\theta} - \mathbf{a}'\mathbf{Y}\right] = 0$. But, $\ell'\hat{\theta} - \mathbf{a}'\mathbf{Y} = c$ so that $E\left[\ell'\hat{\theta} - \mathbf{a}'\mathbf{Y}\right] = E[c] = c$. Now, we have $c = 0$ and so $\ell'\hat{\theta} = \mathbf{a}'\mathbf{Y}$.

Thus any linear unbiased estimator of $\ell'\theta$ which has minimum variance must equal $\ell'\hat{\theta}$. ■

So, provided we restrict our attention to estimable functions, our estimators will have the nice properties of being unique and BLU.

e) Properties of the estimators in the full rank case

The above results apply generally to models either of less than full rank or of full rank. However, in the full rank case, some additional results apply.

In particular, $E\left[\hat{\boldsymbol{\theta}}\right] = E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\theta} = \boldsymbol{\theta}$ so that in this particular case θ itself is estimable and so $\hat{\theta}$ are BLUE for θ.

Further, since $\ell'\theta$ is a linear combination of θ, it follows that $\ell'\theta$ is estimable for all constant vectors ℓ.

f) Estimation for the maximal model for the RCBD

In the next theorem, because each of the parameters in $\theta' = \begin{bmatrix}\beta' & \tau'\end{bmatrix}$ is not estimable, we consider the estimators of ψ:

- theorem XI.8 tells us that these are estimable,
- theorem XI.10 that they are invariant to the solution to the normal equation chosen
- theorem XI.12 that they are BLUE.

Theorem XI.13: Let **Y** be a *n*-vector of jointly-distributed random variables with

$$\psi = E[\mathbf{Y}] = \mathbf{X}_{\mathrm{B}}\beta + \mathbf{X}_{\mathrm{T}}\tau \text{ and } \operatorname{var}[\mathbf{y}] = \sigma^2\mathbf{I}_n$$

where β is the *b*-vector of parameters specifying a different mean response for each block,

$\mathbf{X}_{\mathrm{B}}$ is the $n \times b$ matrix indicating the block from which an observation came,

τ is the *t*-vector of parameters specifying a different mean response for each treatment,

$\mathbf{X}_{\mathrm{T}}$ is the $n \times t$ matrix indicating the observations that received each of the treatments.

Then $\hat{\boldsymbol{\psi}} = \bar{\mathbf{B}} + \bar{\mathbf{T}} - \bar{\mathbf{G}}$ where $\bar{\mathbf{B}}$, $\bar{\mathbf{T}}$ and $\bar{\mathbf{G}}$ are the *n*-vectors of block, treatment and grand means, respectively.

Proof: Using the estimator of θ, established in theorem XI.6, the estimator of the expected values is given by

$$\begin{aligned}
\hat{\boldsymbol{\psi}} &= \mathbf{X}\hat{\theta} \\
&= \begin{bmatrix}\mathbf{X}_{\mathrm{B}} & \mathbf{X}_{\mathrm{T}-1} & \mathbf{x}_t\end{bmatrix}\begin{bmatrix}\ddot{\bar{\mathbf{B}}}_{b\times 1} + \bar{T}_t\mathbf{1}_b - \bar{Y}\mathbf{1}_b \\ \dddot{\bar{\mathbf{T}}}_{(t-1)\times 1} - \bar{T}_t\mathbf{1}_{(t-1)} \\ 0\end{bmatrix} \\
&= \mathbf{X}_{\mathrm{B}}\left(\ddot{\bar{\mathbf{B}}}_{b\times 1} + \bar{T}_t\mathbf{1}_b - \bar{Y}\mathbf{1}_b\right) + \mathbf{X}_{\mathrm{T}-1}\left(\dddot{\bar{\mathbf{T}}}_{(t-1)\times 1} - \bar{T}_t\mathbf{1}_{(t-1)}\right).
\end{aligned}$$

1. Now each row of $\begin{bmatrix}\mathbf{X}_{\mathrm{B}} & \mathbf{X}_{\mathrm{T}-1}\end{bmatrix}$ corresponds to an observation, say the observation in the *i*th block that received the *j*th treatment.
2. In the row for this observation the element in the *i*th column of $\mathbf{X}_{\mathrm{B}}$ will be 1, as will the element in the *j*th column of $\mathbf{X}_{\mathrm{T-1}}$; all other elements in the row will be zero.
3. Note that if the treatment is treatment *t*, the only nonzero element will be the *i*th column of $\mathbf{X}_{\mathrm{B}}$.
4. So the estimator of the expected value for the observation in the *i*th block that received the *j*th treatment is the sum of *i*th element of $\ddot{\bar{\mathbf{B}}}_{b\times 1} + \bar{T}_t\mathbf{1}_b - \bar{Y}\mathbf{1}_b$ and the *j*th element of $\dddot{\bar{\mathbf{T}}}_{(t-1)\times 1} - \bar{T}_t\mathbf{1}_{(t-1)}$, unless the treatment is treatment *t* when it will be the *i*th element of $\ddot{\bar{\mathbf{B}}}_{b\times 1} + \bar{T}_t\mathbf{1}_b - \bar{Y}\mathbf{1}_b$.
5. The estimator of the expected value for the observation with
   - treatment *t* in the *i*th block is $\bar{B}_i + \bar{T}_t - \bar{Y}$ and
   - with the *j*th treatment (*j* ≠ *t*) in the *i*th block is $\left(\bar{B}_i + \bar{T}_t - \bar{Y}\right) + \left(\bar{T}_j - \bar{T}_t\right) = \bar{B}_i + \bar{T}_j - \bar{Y}$.

That is, $\psi_{ij} = \bar{B}_i + \bar{T}_j - \bar{Y}$ for all i and j, and clearly $\hat{\boldsymbol{\psi}} = \bar{\mathbf{B}} + \bar{\mathbf{T}} - \bar{\mathbf{G}}$. ■

Now, $\hat{\boldsymbol{\psi}} = \bar{\mathbf{B}} + \bar{\mathbf{T}} - \bar{\mathbf{G}} = \bar{\mathbf{G}} + \left(\bar{\mathbf{B}} - \bar{\mathbf{G}}\right) + \left(\bar{\mathbf{T}} - \bar{\mathbf{G}}\right)$. That is the estimator of the expected values can be written as the sum of the grand mean plus main effects.

Note that the estimator is a function of means with

$$\bar{\mathbf{B}} = \mathbf{M}_{\mathrm{B}}\mathbf{Y},\ \bar{\mathbf{T}} = \mathbf{M}_{\mathrm{T}}\mathbf{Y},\ \bar{\mathbf{G}} = \mathbf{M}_{\mathrm{G}}\mathbf{Y} \text{ and } \mathbf{M}_i = \mathbf{X}_i\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\mathbf{X}_i'$$

where $\mathbf{X}_{\mathrm{i}} = \mathbf{X}_{\mathrm{B}}$, $\mathbf{X}_{\mathrm{T}}$ or $\mathbf{X}_{\mathrm{G}}$.

That is, $\mathbf{M}_{\mathrm{B}}$, $\mathbf{M}_{\mathrm{T}}$ and $\mathbf{M}_{\mathrm{G}}$ are the block, treatment and grand mean operators, respectively.

Further, if the data in the vector **Y** has been arranged in prerandomized order with all the observations for a block placed together, the operators are:

$$\begin{aligned}
\mathbf{M}_{\mathrm{G}} &= n^{-1}\mathbf{J}_b \otimes \mathbf{J}_t = n^{-1}\mathbf{J}_n \\
\mathbf{M}_{\mathrm{B}} &= t^{-1}\mathbf{I}_b \otimes \mathbf{J}_t \\
\mathbf{M}_{\mathrm{T}} &= b^{-1}\mathbf{J}_b \otimes \mathbf{I}_t.
\end{aligned}$$

Example XI.3 The RCBD (continued)

For the estimable function $\psi_{23} = \beta_2 + \tau_3$,

$$\hat{\psi}_{23} = \widehat{\beta_2 + \tau_3} = \bar{Y}_{2.} + \bar{Y}_{.3} - \bar{Y}_{..} = \bar{Y}_{..} + \left(\bar{Y}_{2.} - \bar{Y}_{..}\right) + \left(\bar{Y}_{.3} - \bar{Y}_{..}\right).$$

In this case, the mean operators are as follows:

$$\mathbf{M}_{\mathrm{G}} = 20^{-1}\mathbf{J}_5 \otimes \mathbf{J}_4 = \frac{1}{20}\begin{bmatrix}
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1
\end{bmatrix}$$

$$\mathbf{M}_{\mathrm{B}} = 4^{-1}\mathbf{I}_5 \otimes \mathbf{J}_4 = \frac{1}{4}\begin{bmatrix}
1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1
\end{bmatrix}$$

$$\mathbf{M}_{\mathrm{T}} = 5^{-1}\mathbf{J}_5 \otimes \mathbf{I}_4 = \frac{1}{5}\begin{bmatrix}
1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0\\
0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0\\
0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0\\
0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1\\
1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0\\
0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0\\
0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0\\
0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1\\
1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0\\
0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0\\
0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0\\
0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1\\
1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0\\
0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0\\
0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0\\
0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1\\
1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0\\
0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0\\
0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0\\
0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1
\end{bmatrix}.$$

XI.C Generalized least squares (GLS) estimation of the expectation parameters in general linear models

First we define a generalized least squares estimator.

Definition XI.9: Let **Y** be a random vector with

$$\psi = E[\mathbf{Y}] = \mathbf{X}\theta \text{ and } \mathrm{var}[\mathbf{Y}] = \mathbf{V}$$

where **X** is an $n \times q$ matrix of rank $m \le q$, θ is a $q \times 1$ vector of unknown parameters, **V** is an $n \times n$ positive-definite matrix and $n \ge q$.

Then the generalized least squares estimator of θ is the estimator that minimizes the "sum of squares"

$$\left(\mathbf{y}-\mathbf{X}\theta\right)'\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\theta\right).$$

■

Theorem XI.14: Let **Y** be a random vector with

$$\psi = E\left[\mathbf{Y}\right] = \mathbf{X}\theta \text{ and } \operatorname{var}\left[\mathbf{Y}\right] = \mathbf{V}$$

where **X** is an $n \times q$ matrix of rank $m \leq q$, θ is a $q \times 1$ vector of unknown parameters, **V** is a known, $n \times n$, positive-definite matrix and $n \geq q$.

Then the generalized least squares (GLS) estimator of θ is denoted by $\hat{\theta}$ and is given by

$$\hat{\theta} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}.$$

Proof: similar to proof for theorem XI.1. ■

Of course, for the simple linear model $\mathbf{V} = \sigma^2\mathbf{I}_n$ and the GLS estimator reduces to

$$\begin{aligned}\hat{\theta} &= \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \\ &= \left(\mathbf{X}'\sigma^{-2}\mathbf{I}_n\mathbf{X}\right)^{-}\mathbf{X}'\sigma^{-2}\mathbf{I}_n\mathbf{Y} \\ &= \sigma^2\sigma^{-2}\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{Y} \quad \text{as } \left(c\mathbf{A}\right)^{-} = \tfrac{1}{c}\mathbf{A}^{-} \text{ for } c \text{ a constant} \\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{Y},\end{aligned}$$

the OLS estimator.

Unsurprisingly, the properties of GLS estimators parallel those for OLS estimators. In particular, they are BLUE as the next theorem states.

Theorem XI.15 [Gauss-Markoff]: Let **Y** be a random vector with

$$\psi = E\left[\mathbf{Y}\right] = \mathbf{X}\theta \text{ and } \operatorname{var}\left[\mathbf{Y}\right] = \mathbf{V}$$

where **X** is an $n \times q$ matrix of rank $m \leq q$, θ is a $q \times 1$ vector of unknown parameters, **V** is a known, $n \times n$, positive-definite matrix and $n \geq q$.

Let $\ell'\boldsymbol{\theta}$ be an estimable function.

Then the generalized least squares estimator

$$\ell'\hat{\theta} = \ell'\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$$

is the best linear unbiased estimator of $\ell'\theta$, with $\operatorname{var}\left[\ell'\hat{\theta}\right] = \ell'\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\ell$.

Proof: not given ■

Estimation of fixed effects for the RCBD when Blocks are random

We now derive the estimator of τ in the model for the RCBD where Blocks are random. Symbolically, the model is

$$\text{E[Y]} = \text{Treats and var[Y]} = \text{Blocks + Blocks}\wedge\text{Units.}$$

One effect of making Blocks random is that the maximal model for the experiment is of full rank and so the estimator of τ will be BLUE.

Before deriving estimators under this model, we establish the properties of a complete set of mutually-orthogonal idempotents as they will be useful in proving the results.

Definition XI.10: A set of idempotents $\{\mathbf{Q}_\text{F};\text{F}\in\mathcal{F}\}$ forms a **complete set of mutually-orthogonal idempotents** (CSMOI) if $\mathbf{Q}_\text{F}\mathbf{Q}_{\text{F}'}=\mathbf{Q}_{\text{F}'}\mathbf{Q}_\text{F}=\delta_{\text{FF}'}\mathbf{Q}_\text{F}$ for all $\text{F},\text{F}'\in\mathcal{F}$ and $\sum_{\text{F}\in\mathcal{F}}\mathbf{Q}_\text{F}=\mathbf{I}_n$ where $\delta_{\text{FF}'}=1$ if $\text{F}=\text{F}'$ and $\delta_{\text{FF}'}=0$ otherwise. ■

Lemma XI.7: For a matrix that is the linear combination of a complete set of mutually-orthogonal matrices, such as $\mathbf{V}=\sum_{\text{F}\in\mathcal{F}}\lambda_\text{F}\mathbf{Q}_\text{F}$, its inverse is $\mathbf{V}^{-1}=\sum_{\text{F}\in\mathcal{F}}\lambda_\text{F}^{-1}\mathbf{Q}_\text{F}$.

Proof:

$$\begin{aligned}
\mathbf{V}\mathbf{V}^{-1} &= \left(\sum_{\text{F}\in\mathcal{F}}\lambda_\text{F}\mathbf{Q}_\text{F}\right)\left(\sum_{\text{F}\in\mathcal{F}}\lambda_\text{F}^{-1}\mathbf{Q}_\text{F}\right)\\
&= \sum_{\text{F}\in\mathcal{F}}\ \sum_{\text{F}'\in\mathcal{F}}\lambda_\text{F}\mathbf{Q}_\text{F}\lambda_{\text{F}'}^{-1}\mathbf{Q}_{\text{F}'}\\
&= \sum_{\text{F}\in\mathcal{F}}\lambda_\text{F}\mathbf{Q}_\text{F}\lambda_\text{F}^{-1}\mathbf{Q}_\text{F}\\
&= \sum_{\text{F}\in\mathcal{F}}\mathbf{Q}_\text{F}\\
&= \mathbf{I}_n.
\end{aligned}$$

■

Of particular interest is that the set of **Q** matrices derived from a structure formula based on crossing and nesting relationships forms a complete set of mutually-orthogonal idempotents.

Also, it can be shown that $\mathbf{I}_t-\frac{1}{t}\mathbf{J}_t$ and $\frac{1}{t}\mathbf{J}_t$ are two symmetric idempotent matrices whose sum is **I** and whose product is zero so that they form a complete set of mutually-orthogonal idempotents.

In the following theorem we show that the GLS estimator for the expectation model when only Treatments is fixed is the same as the OLS estimator for a model involving only fixed Treatments, such as would occur with a completely randomized design.

However, before the theorem, a lemma that provides some useful properties of direct products (see definition XI.3).

Lemma XI.8: Let **A**, **B**, **C** and **D** be square matrices and a, b and c be scalar constants. Then,

$(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = \mathbf{AC}\otimes\mathbf{BD}$ provided **A** and **C**, as well as **B** and **D**, are conformable.

$$(\mathbf{A}\otimes\mathbf{B})' = \mathbf{A}'\otimes\mathbf{B}'$$
$$a\mathbf{A} = a\otimes\mathbf{A}$$
$$a\mathbf{A}\otimes b\mathbf{B} = ab\mathbf{A}\otimes\mathbf{B}$$
$$\mathbf{A}\otimes(b\mathbf{B}+c\mathbf{C}) = b\mathbf{A}\otimes\mathbf{B}+c\mathbf{A}\otimes\mathbf{C}$$
$$trace(\mathbf{A}\otimes\mathbf{B}) = trace(\mathbf{A})trace(\mathbf{B}).$$

■

Theorem XI.16: Let **Y** be an $n\times 1$ random vector for the observations from an RCBD with the Y_{ijk} arranged such that those from the same block occur consecutively in treatment order, as in a prerandomized layout.

Also, let $\psi = E[\mathbf{Y}] = \mathbf{X}_\mathrm{T}\tau = (\mathbf{1}_b\otimes\mathbf{I}_t)\tau$ and $\mathbf{V} = \sigma^2_\mathrm{B}\mathbf{S}_\mathrm{B} + \sigma^2_\mathrm{BU}\mathbf{S}_\mathrm{BU} = \sigma^2_\mathrm{B}(\mathbf{I}_b\otimes\mathbf{J}_t) + \sigma^2_\mathrm{BU}(\mathbf{I}_b\otimes\mathbf{I}_t)$.

Then the generalized least squares estimator of τ is given by
$$\hat{\tau} = \left(\mathbf{X}'_\mathrm{T}\mathbf{V}^{-1}\mathbf{X}_\mathrm{T}\right)^{-1}\mathbf{X}'_\mathrm{T}\mathbf{V}^{-1}\mathbf{Y} = \tfrac{1}{b}\mathbf{X}'_\mathrm{T}\mathbf{Y}$$
and the GLS estimator of the expected values is
$$\bar{\mathbf{T}}_{n\times 1} = \mathbf{M}_\mathrm{T}\mathbf{Y}$$
where $\mathbf{M}_\mathrm{T} = \mathbf{X}_\mathrm{T}\left(\mathbf{X}'_\mathrm{T}\mathbf{X}_\mathrm{T}\right)^{-1}\mathbf{X}'_\mathrm{T}$.

Proof: From theorem XI.14, the generalized least squares estimator of τ is given by
$$\hat{\tau} = \left(\mathbf{X}'_\mathrm{T}\mathbf{V}^{-1}\mathbf{X}_\mathrm{T}\right)^{-1}\mathbf{X}'_\mathrm{T}\mathbf{V}^{-1}\mathbf{Y},$$
the generalized inverse being replaced by an inverse here because $\mathbf{X}_\mathrm{T}$ is of full rank.

To prove the second part of the result, we
i) derive an expressions for $\mathbf{V}^{-1}$,
ii) use this to obtain $\left(\mathbf{X}'_\mathrm{T}\mathbf{V}^{-1}\mathbf{X}_\mathrm{T}\right)$,
iii) invert the last result, and
iv) show that $\left(\mathbf{X}'_\mathrm{T}\mathbf{V}^{-1}\mathbf{X}_\mathrm{T}\right)^{-1}\mathbf{X}'_\mathrm{T}\mathbf{V}^{-1}\mathbf{Y} = \tfrac{1}{b}\mathbf{X}'_\mathrm{T}\mathbf{Y}$.

We start by obtaining an expression for **V** in terms of $\mathbf{Q}_\mathrm{G}$, $\mathbf{Q}_\mathrm{B}$ and $\mathbf{Q}_\mathrm{BU}$ as these form a complete set of mutually orthogonal idempotents. Now,

$$\mathbf{V} = \sigma_{\mathrm{B}}^2\mathbf{S}_{\mathrm{B}} + \sigma_{\mathrm{BU}}^2\mathbf{S}_{\mathrm{BU}} = t\sigma_{\mathrm{B}}^2\mathbf{M}_{\mathrm{B}} + \sigma_{\mathrm{BU}}^2\mathbf{M}_{\mathrm{BU}}.$$

But, from section IV.C, *Hypothesis testing using the ANOVA method for an RCBD*, we know that $\mathbf{Q}_{\mathrm{G}} = \mathbf{M}_{\mathrm{G}}$, $\mathbf{Q}_{\mathrm{B}} = \mathbf{M}_{\mathrm{B}} - \mathbf{M}_{\mathrm{G}}$ and $\mathbf{Q}_{\mathrm{BU}} = \mathbf{M}_{\mathrm{BU}} - \mathbf{M}_{\mathrm{B}}$ so that $\mathbf{M}_{\mathrm{B}} = \mathbf{Q}_{\mathrm{B}} + \mathbf{Q}_{\mathrm{G}}$ and $\mathbf{M}_{\mathrm{BU}} = \mathbf{Q}_{\mathrm{BU}} + \mathbf{Q}_{\mathrm{B}} + \mathbf{Q}_{\mathrm{G}}$. Consequently,

$$\begin{aligned}
\mathbf{V} &= t\sigma_{\mathrm{B}}^2\mathbf{M}_{\mathrm{B}} + \sigma_{\mathrm{BU}}^2\mathbf{M}_{\mathrm{BU}} \\
&= t\sigma_{\mathrm{B}}^2\left(\mathbf{Q}_{\mathrm{B}} + \mathbf{Q}_{\mathrm{G}}\right) + \sigma_{\mathrm{BU}}^2\left(\mathbf{Q}_{\mathrm{BU}} + \mathbf{Q}_{\mathrm{B}} + \mathbf{Q}_{\mathrm{G}}\right) \\
&= \left(t\sigma_{\mathrm{B}}^2 + \sigma_{\mathrm{BU}}^2\right)\left(\mathbf{Q}_{\mathrm{B}} + \mathbf{Q}_{\mathrm{G}}\right) + \sigma_{\mathrm{BU}}^2\mathbf{Q}_{\mathrm{BU}} \\
&= \lambda_{\mathrm{B}}\left(\mathbf{Q}_{\mathrm{B}} + \mathbf{Q}_{\mathrm{G}}\right) + \lambda_{\mathrm{BU}}\mathbf{Q}_{\mathrm{BU}}
\end{aligned}$$

where $\lambda_{\mathrm{B}} = t\sigma_{\mathrm{B}}^2 + \sigma_{\mathrm{BU}}^2$ and $\lambda_{\mathrm{BU}} = \sigma_{\mathrm{BU}}^2$.

Since $\mathbf{Q}_{\mathrm{G}}$, $\mathbf{Q}_{\mathrm{B}}$ and $\mathbf{Q}_{\mathrm{BU}}$ form a complete set of mutually orthogonal idempotents, we use lemma XI.7 to obtain

$$\begin{aligned}
\mathbf{V}^{-1} &= \lambda_{\mathrm{B}}^{-1}\left(\mathbf{Q}_{\mathrm{G}} + \mathbf{Q}_{\mathrm{B}}\right) + \lambda_{\mathrm{BU}}^{-1}\mathbf{Q}_{\mathrm{BU}} \\
&= \lambda_{\mathrm{B}}^{-1}\left(\mathbf{M}_{\mathrm{G}} + \mathbf{M}_{\mathrm{B}} - \mathbf{M}_{\mathrm{G}}\right) + \lambda_{\mathrm{BU}}^{-1}\left(\mathbf{M}_{\mathrm{BU}} - \mathbf{M}_{\mathrm{B}}\right) \\
&= \left(\lambda_{\mathrm{B}}^{-1} - \lambda_{\mathrm{BU}}^{-1}\right)\mathbf{M}_{\mathrm{B}} + \lambda_{\mathrm{BU}}^{-1}\mathbf{M}_{\mathrm{BU}} \\
&= \frac{\lambda_{\mathrm{B}}^{-1} - \lambda_{\mathrm{BU}}^{-1}}{t}\mathbf{I}_b \otimes \mathbf{J}_t + \lambda_{\mathrm{BU}}^{-1}\mathbf{I}_b \otimes \mathbf{I}_t
\end{aligned}$$

and we have our expression for $\mathbf{V}^{-1}$.

Next to derive an expression for $\left(\mathbf{X}_{\mathrm{T}}'\mathbf{V}^{-1}\mathbf{X}_{\mathrm{T}}\right)$, note that $\mathbf{X}_{\mathrm{T}} = \mathbf{1}_b \otimes \mathbf{I}_t$ and $\left(\mathbf{A} \otimes \mathbf{B}\right)' = \mathbf{A}' \otimes \mathbf{B}'$ so that

$$\begin{aligned}
&\left(\mathbf{1}_b \otimes \mathbf{I}_t\right)' \mathbf{J}_b \otimes \mathbf{J}_t = b\mathbf{1}_b' \otimes \mathbf{J}_t, \\
&\left(\mathbf{1}_b \otimes \mathbf{I}_t\right)' \mathbf{I}_b \otimes \mathbf{J}_t = \mathbf{1}_b' \otimes \mathbf{J}_t, \\
&\left(\mathbf{1}_b \otimes \mathbf{I}_t\right)' \mathbf{I}_b \otimes \mathbf{I}_t = \mathbf{1}_b' \otimes \mathbf{I}_t \text{ and} \\
&\mathbf{1}_b'\mathbf{1}_b = b.
\end{aligned}$$

We then see that

$$\begin{aligned}
\mathbf{X}_{\mathrm{T}}'\mathbf{V}^{-1}\mathbf{X}_{\mathrm{T}} &= \left(\mathbf{1}_b \otimes \mathbf{I}_t\right)'\left\{\frac{\lambda_{\mathrm{B}}^{-1} - \lambda_{\mathrm{BU}}^{-1}}{t}\mathbf{I}_b \otimes \mathbf{J}_t + \lambda_{\mathrm{BU}}^{-1}\mathbf{I}_b \otimes \mathbf{I}_t\right\}\left(\mathbf{1}_b \otimes \mathbf{I}_t\right) \\
&= \left\{\frac{\lambda_{\mathrm{B}}^{-1} - \lambda_{\mathrm{BU}}^{-1}}{t}\mathbf{1}_b' \otimes \mathbf{J}_t + \lambda_{\mathrm{BU}}^{-1}\mathbf{1}_b' \otimes \mathbf{I}_t\right\}\left(\mathbf{1}_b \otimes \mathbf{I}_t\right) \\
&= \frac{\lambda_{\mathrm{B}}^{-1} - \lambda_{\mathrm{BU}}^{-1}}{t}b\mathbf{J}_t + \lambda_{\mathrm{BU}}^{-1}b\mathbf{I}_t \\
&= b\left\{\left(\lambda_{\mathrm{B}}^{-1} - \lambda_{\mathrm{BU}}^{-1}\right)\tfrac{1}{t}\mathbf{J}_t + \lambda_{\mathrm{BU}}^{-1}\mathbf{I}_t\right\}.
\end{aligned}$$

Thirdly, to find the inverse of this note that $\mathbf{I}_t - \frac{1}{t}\mathbf{J}_t$ and $\frac{1}{t}\mathbf{J}_t$ are two symmetric idempotent matrices whose sum is **I** and whose product is zero so that they form a complete set of mutually orthogonal idempotents (CSMOI) — see also lemma XI.4.

Now

$$\begin{aligned}\left(\mathbf{X}_\mathrm{T}'\mathbf{V}^{-1}\mathbf{X}_\mathrm{T}\right)^{-1} &= \frac{1}{b}\left\{\left(\lambda_\mathrm{B}^{-1} - \lambda_\mathrm{BU}^{-1}\right)\tfrac{1}{t}\mathbf{J}_t + \lambda_\mathrm{BU}^{-1}\mathbf{I}_t\right\}^{-1} \\ &= \frac{1}{b}\left\{\lambda_\mathrm{B}^{-1}\tfrac{1}{t}\mathbf{J}_t + \lambda_\mathrm{BU}^{-1}\left(\mathbf{I}_t - \tfrac{1}{t}\mathbf{J}_t\right)\right\}^{-1}. \\ &= \frac{1}{b}\left\{\lambda_\mathrm{B}\tfrac{1}{t}\mathbf{J}_t + \lambda_\mathrm{BU}\left(\mathbf{I}_t - \tfrac{1}{t}\mathbf{J}_t\right)\right\}.\end{aligned}$$

Finally,

$$\begin{aligned}\hat{\boldsymbol{\tau}} &= \left(\mathbf{X}_\mathrm{T}'\mathbf{V}^{-1}\mathbf{X}_\mathrm{T}\right)^{-1}\mathbf{X}_\mathrm{T}'\mathbf{V}^{-1}\mathbf{Y} \\ &= \frac{1}{b}\left\{\lambda_\mathrm{B}\tfrac{1}{t}\mathbf{J}_t + \lambda_\mathrm{BU}\left(\mathbf{I}_t - \tfrac{1}{t}\mathbf{J}_t\right)\right\}\left\{\left(\lambda_\mathrm{B}^{-1} - \lambda_\mathrm{BU}^{-1}\right)\tfrac{1}{t}\mathbf{1}_b' \otimes \mathbf{J}_t + \lambda_\mathrm{BU}^{-1}\mathbf{1}_b' \otimes \mathbf{I}_t\right\}\mathbf{Y} \\ &= \frac{1}{b}\left\{\begin{matrix}\lambda_\mathrm{B}\left(\lambda_\mathrm{B}^{-1} - \lambda_\mathrm{BU}^{-1}\right)\tfrac{1}{t}\mathbf{1}_b' \otimes \mathbf{J}_t + \lambda_\mathrm{B}\lambda_\mathrm{BU}^{-1}\tfrac{1}{t}\mathbf{1}_b' \otimes \mathbf{J}_t + \mathbf{0} \\ + \lambda_\mathrm{BU}\lambda_\mathrm{BU}^{-1}\mathbf{1}_b' \otimes \left(\mathbf{I}_t - \tfrac{1}{t}\mathbf{J}_t\right)\end{matrix}\right\}\mathbf{Y} \quad \left(\begin{matrix}\text{note } \mathbf{J}_t = 1 \otimes \mathbf{J}_t; \\ \mathbf{0} \text{ using CSMOI}\end{matrix}\right) \\ &= \frac{1}{b}\left\{\lambda_\mathrm{B}\left(\lambda_\mathrm{B}^{-1} - \lambda_\mathrm{BU}^{-1} + \lambda_\mathrm{BU}^{-1}\right)\tfrac{1}{t}\mathbf{1}_b' \otimes \mathbf{J}_t + \left(\mathbf{1}_b' \otimes \mathbf{I}_t - \tfrac{1}{t}\mathbf{1}_b' \otimes \mathbf{J}_t\right)\right\}\mathbf{Y} \\ &= \frac{1}{b}\left\{\tfrac{1}{t}\mathbf{1}_b' \otimes \mathbf{J}_t + \mathbf{1}_b' \otimes \mathbf{I}_t - \tfrac{1}{t}\mathbf{1}_b' \otimes \mathbf{J}_t\right\}\mathbf{Y} \\ &= \frac{1}{b}\left\{\mathbf{1}_b' \otimes \mathbf{I}_t\right\}\mathbf{Y} \\ &= \frac{1}{b}\mathbf{X}_\mathrm{T}'\mathbf{Y}.\end{aligned}$$

The GLS estimator of the expected values is then given by

$$\hat{\boldsymbol{\psi}} = \mathbf{X}_\mathrm{T}\hat{\boldsymbol{\tau}} = \frac{1}{b}\mathbf{X}_\mathrm{T}\mathbf{X}_\mathrm{T}'\mathbf{Y}.$$

But, as

$$\mathbf{X}_\mathrm{T}'\mathbf{X}_\mathrm{T} = \left(\mathbf{1}_b \otimes \mathbf{I}_t\right)'\left(\mathbf{1}_b \otimes \mathbf{I}_t\right) = b\mathbf{I}_t \text{ and,}$$

from corollary XI.1, $\mathbf{M}_\mathrm{T} = \mathbf{X}_\mathrm{T}\left(\mathbf{X}_\mathrm{T}'\mathbf{X}_\mathrm{T}\right)^{-1}\mathbf{X}_\mathrm{T}' = \mathbf{X}_\mathrm{T}\left(b\mathbf{I}_t\right)^{-1}\mathbf{X}_\mathrm{T}' = \frac{1}{b}\mathbf{X}_\mathrm{T}\mathbf{X}_\mathrm{T}'$,

$$\hat{\boldsymbol{\psi}} = \frac{1}{b}\mathbf{X}_\mathrm{T}\mathbf{X}_\mathrm{T}'\mathbf{Y} = \mathbf{M}_\mathrm{T}\mathbf{Y} = \bar{\mathbf{T}}.$$

■

So the estimator for the expected values under the model $\boldsymbol{\psi} = E\left[\mathbf{Y}\right] = \mathbf{X}_\mathrm{T}\boldsymbol{\tau}$ and $\mathbf{V} = \sigma_\mathrm{B}^2\left(\mathbf{I}_b \otimes \mathbf{J}_t\right) + \sigma_\mathrm{BU}^2\left(\mathbf{I}_b \otimes \mathbf{I}_t\right)$ are just the treatment means, the same as for the simple linear model $E\left[\mathbf{Y}\right] = \mathbf{X}_\mathrm{T}\boldsymbol{\tau}$ and $\mathbf{V} = \sigma_\mathrm{BU}^2\left(\mathbf{I}_b \otimes \mathbf{I}_t\right)$. This result is generally true for orthogonal experiments, as stated in the next theorem that we give without proof.

Definition XI.11: An orthogonal experiment is one for which $\mathbf{Q}_F\mathbf{Q}_{F^*}$ equals either $\mathbf{Q}_F$, $\mathbf{Q}_{F^*}$ or $\mathbf{0}$ for all F, F* in the experiment. ■

All the experiments in these notes are orthogonal.

Theorem XI.17: For orthogonal experiments with a variance matrix that can be expressed as a linear combination of a complete set of mutually-orthogonal idempotents, the GLS estimator is the same as the OLS estimator for the same expectation model.
Proof: not given ■

This theorem applies to all orthogonal experiments for which all the variance terms correspond to unrandomized generalized factors.

XI.D Maximum likelihood estimation of the expectation parameters

In previous sections we used least squares estimation to obtain estimators of our parameters. It was mentioned that this is not the only method of estimation. Here we describe an alternative method of estimation, maximum likelihood estimation. It is based on obtaining the values of the parameters that make the data most 'likely'. Compare this with least squares estimation in which the sum of squares of the differences between the data and the expected values is minimized.

Maximum likelihood estimation involves determining the likelihood function for the data and it is this function that is maximized.

Definition XI.12: The **likelihood function** is the joint distribution function, $f(\mathbf{y};\boldsymbol{\xi})$, of n random variables $Y_1, Y_2, \ldots, Y_n$ evaluated at $\mathbf{y} = (y_1, y_2, \ldots, y_n)$. For fixed $\mathbf{y}$ the likelihood function is a function of the k parameters, $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_k)$, and is denoted by $L(\boldsymbol{\xi};\mathbf{y})$. If $Y_1, Y_2, \ldots, Y_n$ represents a random sample from $f(\mathbf{y};\boldsymbol{\xi})$ then

$$L(\boldsymbol{\xi};\mathbf{y}) = f(\mathbf{y};\boldsymbol{\xi}).$$

■

Note that the notation $f(\mathbf{y};\boldsymbol{\xi})$ indicates that $\mathbf{y}$ is the vector of variables for the distribution function and that the values of $\boldsymbol{\xi}$ are fixed for the population under consideration. This is read 'the function f of $\mathbf{y}$ given $\boldsymbol{\xi}$'. The likelihood function reverses the roles in that the variables are $\boldsymbol{\xi}$ and $\mathbf{y}$ is considered fixed so we have the likelihood of $\boldsymbol{\xi}$ given $\mathbf{y}$.

Definition XI.13: Let $L(\boldsymbol{\xi};\mathbf{y}) = f(\mathbf{y};\boldsymbol{\xi})$ be the likelihood function for $Y_1, Y_2, \ldots, Y_n$. For a given set of observations, $\mathbf{y}$, the value $\tilde{\boldsymbol{\xi}}$ that maximizes $L(\boldsymbol{\xi};\mathbf{y})$ is called the

maximum likelihood estimate (MLE) of $\boldsymbol{\xi}$. An expression for this estimate as a function of **y** can be derived. The **maximum likelihood estimators** are defined to be the same function as the estimate, with **Y** substituted for **y**. ■

The procedure for obtaining the maximum likelihood estimators is:

a) Write the likelihood function $L(\boldsymbol{\xi}; \mathbf{y})$ as the joint distribution function for the *n* observations.
b) Find $\ell = \ln\left[L(\boldsymbol{\xi}; \mathbf{y})\right]$.
c) Maximize ℓ with respect to $\boldsymbol{\xi}$ to derive the maximum likelihood estimates.
d) Obtain the maximum likelihood estimators.

The quantity ℓ is referred to as the log likelihood. Maximizing ℓ is equivalent to maximizing $L(\boldsymbol{\xi}; \mathbf{y})$ since *ln* is a monotonic function. It will turn out that the expression for ℓ is often simpler than that for $L(\boldsymbol{\xi}; \mathbf{y})$.

Now the general linear model for a response variable is

$$\boldsymbol{\psi} = E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\theta} \text{ and } \mathrm{var}[\mathbf{Y}] = \mathbf{V}_{\mathbf{Y}} = \sum \sigma_{\mathrm{F}}^2 \mathbf{S}_{\mathrm{F}} = \sum \lambda_{\mathrm{F}} \mathbf{Q}_{\mathrm{F}}.$$

We have not previously specified a probability distribution function to be used as part of the model. However, maximum likelihood estimation requires that we do specify this function (least squares estimation does not). A distribution function commonly used for continuous variables is the multivariate normal distribution.

Definition XI.14: The multivariate normal distribution function for is

$$f(\mathbf{y}; \theta, \mathbf{V}) = (2\pi)^{-n/2} |\mathbf{V}|^{-1/2} \exp\left\{-(\mathbf{y}-\psi)' \mathbf{V}^{-1} (\mathbf{y}-\psi)\right\}$$

where $\psi = E[\mathbf{Y}] = \mathbf{X}\theta$, **X** is an $n \times q$ matrix of rank $m \leq q$ with $n \geq q$, θ is a $q \times 1$ vector of unknown parameters, and $\mathrm{var}[\mathbf{Y}] = \mathbf{V}$. ■

We now derive the maximum likelihood estimators of $\boldsymbol{\theta}$.

Theorem XI.18: Let **Y** be a random vector whose distribution is multivariate normal. Then, the maximum likelihood estimator of θ is denoted by $\tilde{\theta}$ and is given by

$$\tilde{\theta} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-} \mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}.$$

Proof:

The likelihood function $L(\xi;\mathbf{y})$

Given the distribution function for **Y**, the likelihood function $L(\xi;\mathbf{y})$ is

$$L(\xi;\mathbf{y}) = (2\pi)^{-n/2}|\mathbf{V}|^{-1/2}\exp\left\{-(\mathbf{y}-\psi)'\mathbf{V}^{-1}(\mathbf{y}-\psi)\right\}.$$

Find $\ell = \ln\left[L(\boldsymbol{\xi};\mathbf{y})\right]$

$$\begin{aligned}\ell &= \ln\left[L(\xi;\mathbf{y})\right]\\ &= \ln\left[(2\pi)^{-n/2}|\mathbf{V}|^{-1/2}\exp\left\{-(\mathbf{y}-\psi)'\mathbf{V}^{-1}(\mathbf{y}-\psi)\right\}\right]\\ &= -\frac{n}{2}\ln(2\pi)-\frac{1}{2}\ln(|\mathbf{V}|)-(\mathbf{y}-\psi)'\mathbf{V}^{-1}(\mathbf{y}-\psi).\end{aligned}$$

Maximize ℓ with respect to θ

The maximum likelihood estimates of θ are then obtained by maximizing ℓ with respect to θ, that is, by differentiating ℓ with respect to θ and setting the result equal to zero.

Clearly, the maximum likelihood estimates will be the solution of

$$\frac{\partial\left\{(\mathbf{y}-\psi)'\mathbf{V}^{-1}(\mathbf{y}-\psi)\right\}}{\partial\theta} = \frac{\partial\left\{(\mathbf{y}-\mathbf{X}\theta)'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\theta)\right\}}{\partial\theta} = \mathbf{0}.$$

Now,

$$\frac{\partial\left\{(\mathbf{y}-\mathbf{X}\theta)'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\theta)\right\}}{\partial\theta} = \frac{\partial\left\{\mathbf{y}'\mathbf{V}^{-1}\mathbf{y}-2\theta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}+\theta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\theta\right\}}{\partial\theta}$$

and applying the rules for differentiation given in the proof of theorem XI.1, we obtain

$$\begin{aligned}\frac{\partial\left\{\mathbf{y}'\mathbf{V}^{-1}\mathbf{y}-2\theta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}+\theta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\theta\right\}}{\partial\theta} &= -2\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right)+\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)\theta+\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)'\theta\\ &= -2\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right)+2\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)\theta.\end{aligned}$$

Setting this derivative to zero we obtain

$$-2\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right)+2\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)\theta = 0 \text{ or } \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)\theta = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}.$$

Obtain the maximum likelihood estimators

Hence, $\tilde{\theta} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$ as claimed. ■

So for general linear models the maximum likelihood and generalized least squares estimators coincide. It turns out that maximum likelihood estimators have some further desirable properties so that it is an advantage to know that our estimators are maximum likelihood estimators.

XI.E Estimating the variance

a) Estimating the variance for the simple linear model

As we stated in theorem XI.11, $\text{var}\left[\ell'\boldsymbol{\theta}\right] = \sigma^2\ell'\left(\mathbf{X}'\mathbf{X}\right)^{-}\ell$ and so to determine the variance of our OLS estimates we need to know σ^2. Generally, it is unknown and has to be estimated. Now the definition of the variance is $\text{var}\left[Y\right] = E\left[\left(Y - E\left[Y\right]\right)^2\right]$. The logical estimator is one that parallels this definition as much as possible: put in our estimate for $E\left[Y\right]$ and replace the first expectation operator by the mean.

Definition XI.15: The **fitted values** are the estimated expected values for each observation, $\widehat{E\left[Y_i\right]}$. They are obtained by substituting the *x*-values for each observation into the estimated equation. They are given by $\mathbf{X}\hat{\boldsymbol{\theta}}$. The **residuals** are the deviations of the observed values of the response variable from the fitted values. They are denoted by $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\theta}}$ and are the estimates of $\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\theta}$. ■

The logical estimator of σ^2 is:

$$\hat{\sigma}_n^2 = \frac{\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}}\right)'\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}}\right)}{n} = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n}$$

and our estimate would be $\mathbf{e}'\mathbf{e}/n$.

Notice that the numerator of this expression measures the spread of observed *Y*-values from the fitted equation. We expect this to be random variation.

Turns out that this is the maximum likelihood estimator as claimed in the next theorem.

Theorem XI.19: Let $\mathbf{Y}$ be a normally distributed random vector representing a random sample with $E\left[\mathbf{Y}\right] = \mathbf{X}\boldsymbol{\theta}$ and $\text{var}\left[\mathbf{Y}\right] = \mathbf{V}_{\mathbf{Y}} = \sigma^2\mathbf{I}_n$ where $\mathbf{X}$ is an $n \times q$ matrix of rank $m \le q$ with $n \ge q$, and $\boldsymbol{\theta}$ is a $q \times 1$ vector of unknown parameters. The maximum likelihood estimator of σ^2 is denoted by $\tilde{\sigma}^2$ and is given by

$$\tilde{\sigma}_n^2 = \frac{\left(\mathbf{Y}-\mathbf{X}\tilde{\boldsymbol{\theta}}\right)'\left(\mathbf{Y}-\mathbf{X}\tilde{\boldsymbol{\theta}}\right)}{n} = \frac{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}}{n}.$$

Proof: The proof is left as an exercise for you ■

Again we ask the question: is our estimator unbiased?

Theorem XI.20: Let $\hat{\sigma}_n^2 = \left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\theta}}\right)'\left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\theta}}\right)/n$ with $\mathbf{Y}$ the $n \times 1$ random vector of sample random variables, $\mathbf{X}$ is an $n \times q$ matrix of rank $m \leq q$ with $n \geq q$, and $\hat{\boldsymbol{\theta}}$ is a $q \times 1$ vector of estimators of $\boldsymbol{\theta}$ Then $E\left[\hat{\sigma}_n^2\right] = \frac{n-q}{n}\sigma^2$ so that an unbiased estimator of σ^2 is

$$\hat{\sigma}_{n-q}^2 = \frac{\left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\theta}}\right)'\left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\theta}}\right)}{n-q}.$$

Proof: The proof of this theorem is postponed. ■

This latter estimator indicates how we might go more generally, once it is realized that this estimator is one that would be obtained using the Residual mean square from an analysis of variance based on simple linear models.

b) Estimating the variance for the general linear model

Generally, the variance components $\left(\sigma^2\text{s}\right)$ that are the parameters of the variance matrix, **V**, can be estimated using the ANOVA method.

Definition XI.16: The **ANOVA method for estimating the variance components** consists of equating the expected mean squares to the observed values of the mean squares and solving for the variance components. ■

Example IV.1 Penicillin yield

This example involves 4 treatments applied using an RCBD employing 5 Blends as blocks. Suppose it was decided that the Blends are random. The ANOVA table, with expected mean squares, for this case is

Source	df	MSq	E[MSq]	F	Prob
Blends	4	66.0	$\sigma^2_{\mathrm{BF}} + 4\sigma^2_{\mathrm{B}}$	3.50	0.041
Flasks [Blends]	15				
Treatments	3	23.3	$\sigma^2_{\mathrm{BF}} + q_{\mathrm{T}}(\psi)$	1.24	0.339
Residual	12	18.8	σ^2_{BF}		
Total	19				

The estimates of σ^2_{BF} and σ^2_{B} are obtained as follows:

$$\sigma^2_{\mathrm{BF}} + 4\sigma^2_{\mathrm{B}} = 66.0$$
$$\sigma^2_{\mathrm{BF}} = 18.8$$

so that

$$\hat{\sigma}^2_{\mathrm{BF}} = 18.8$$
$$\begin{aligned}\hat{\sigma}^2_{\mathrm{B}} &= \frac{66.0 - \hat{\sigma}^2_{\mathrm{BF}}}{4} \\ &= \frac{66.0 - 18.8}{4} \\ &= 11.8.\end{aligned}$$

That is, the magnitude of the two components of variation is similar.

Now $\mathrm{var}\left[Y_{ijk}\right] = \sigma^2_{\mathrm{BF}} + \sigma^2_{\mathrm{B}}$ (see variance matrix in section XI.A, *Linear models for designed experiments*, above), so that $\widehat{\mathrm{var}\left[Y_{ijk}\right]} = \hat{\sigma}^2_{\mathrm{BF}} + \hat{\sigma}^2_{\mathrm{B}} = 18.8 + 11.8 = 30.6$. Clearly, the two components make similar contributions to the variance of an observation. ■

XI.F Summary

In this chapter we have:

- Distinguished between models for which;
  1. the expectation models are of full rank or are less than full rank;
  2. the variation models are simple or general;
- Investigated the least squares estimation of the expectation model parameters for situations in which the variation models are simple and the expectation models are a) of full rank and b) less than full rank and in which the variation models are general. When the variation model is simple ordinary least squares applies whereas when it is general generalized least squares is required;
- Proven that the expected values for the observations and all linear combinations of them are the set of estimable functions for an experiment;

- Established in the Gauss-Markoff theorem that ordinary and generalized least squares estimators are BLUEs;
- Proven that, for general linear models, the maximum likelihood and generalized least squares estimators of the expectation model parameters coincide;
- Outlined the ANOVA method for estimating the variance components.

XI.G Exercises

XI.1 The analysis for an example involving a 4×4 Latin Square was derived in section VI.B, *The Latin square example*. In particular, the expectation and variation models were written symbolically as

$$\psi = \text{E[Y]} = \text{Additive and}$$
$$\text{var[Y]} = \text{Driver + Car + Driver}\wedge\text{Car}$$

Write down matrix expressions for these models, the expectation model involving indicator-variable (**X**) matrices and the variation model involving summation (**S**) matrices and variance components. Also, give the variance matrix in terms of direct products of **I** and **J** matrices, assuming that the observations are in standard order for Driver then Car.

What will be the estimator of the expected values under this model? Which theorems justify your answer?

XI.2 Suppose that for a particular linear model

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}, \ \mathbf{X}'\mathbf{X} = \begin{bmatrix} 2 & 0 & 0 & 1 & 1 \\ 0 & 2 & 0 & 1 & 1 \\ 0 & 0 & 2 & 1 & 1 \\ 1 & 1 & 1 & 3 & 0 \\ 1 & 1 & 1 & 0 & 3 \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} 19 \\ 29 \\ 14 \\ 24 \\ 12 \\ 22 \end{bmatrix}$$

a) Find a generalized inverse for **X'X** and use it to find a solution to the normal equations by computing $(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$. Obtain the fitted values.
b) Find a second generalized inverse for **X'X** by deleting its first row and column. Use it to find a solution to the normal equations and to obtain the fitted values.
c) How do the solutions and fitted values obtained in the previous parts compare?

XI.3 From theorem XI.8 we know that the expected values for the simple linear model with $\boldsymbol{\psi} = E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\theta}$, where **X** is $n \times q$ of rank $m \leq q$, is estimable. Use theorem XI.11 to show that $\text{var}[\hat{\boldsymbol{\psi}}] = \sigma^2\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}$. (Hint: note that $\psi_i = \mathbf{x}_i'\boldsymbol{\theta}$ where $\mathbf{x}_i$ is the q-vector that is the *i*th row of **X**.)

XI.4 Consider the following expectation models for a two-factor factorial experiment:

$$E\left[Y_{ik\ell}\right]=\left(\alpha\beta\right)_{k\ell},\ E\left[Y_{ik\ell}\right]=\alpha_k+\beta_\ell,\ E\left[Y_{ik\ell}\right]=\alpha_k,\ E\left[Y_{ik\ell}\right]=\beta_\ell \text{ and } E\left[Y_{ik\ell}\right]=\mu.$$

a) What is the estimator of the expected values for the first model?
b) For which models are the individual parameters estimable? Give reasons for your answer.

XI.5 An experiment was conducted to study the effects of temperature on the life (in hours) of a component. An RCBD was employed with five ovens forming the blocks. Four temperatures were randomly assigned to four runs within each oven. The following results were recorded:

		Temperature (degrees)			
		200	300	400	500
	I	340	324	307	274
	II	361	338	312	281
Oven	III	346	328	298	276
	IV	358	332	315	285
	V	343	321	294	269

This data was analysed as part of exercise IV.4 and the following analysis of variance table obtained.

```
> RCBDComponent.aov <- aov(Life ~ Oven + Temperature + Error(Oven/Run),
     RCBDComponent)
> summary(RCBDComponent.aov, split = list(Temperature = list(L = 1, Q = 2)))
Error: Oven
     Df Sum of Sq Mean Sq
Oven  4     845.3 211.325

Error: Run %in% Oven
                Df Sum of Sq  Mean Sq  F Value     Pr(F)
     Temperature  3  14610.60  4870.20  365.493 0.0000000
  Temperature: L  1  14544.36 14544.36 1091.509 0.0000000
  Temperature: Q  1     64.80    64.80    4.863 0.0476871
       Residuals 12    159.90    13.32
```

a) Assume that Ovens and Runs are random factors and that Temperatures is a fixed factor. What are the expected mean squares under the maximal model in these circumstances?
b) Give the expressions for the variance matrix in terms of the variance components and the summation and mean-operator matrices, for this example.
c) What are the estimates of the variance components?

XI.6 Let **Y** be a normally distributed random vector representing a random sample with $E[\mathbf{Y}] = \mathbf{X\theta}$ and $\text{var}[\mathbf{Y}] = \mathbf{V} = \sigma^2\mathbf{I}_n$ where **X** is an $n \times q$ matrix of full rank, $\boldsymbol{\theta}$ is a $q \times 1$ vector of unknown parameters and $n \geq q$. Prove that the maximum likelihood estimator of σ^2 is given by

$$\tilde{\sigma}_n^2 = \frac{\left(\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\theta}}\right)'\left(\mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\theta}}\right)}{n} = \frac{\tilde{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\varepsilon}}}{n}$$

Note that you can use the log likelihood function, ℓ, from Theorem X.17 so that your first step will be to simplify the expression for $\text{var}[\mathbf{Y}] = \mathbf{V} = \sigma^2\mathbf{I}_n$. Then you will need to maximize the log likelihood with respect to σ^2.

Is this estimator unbiased?

XI.7 We have that for the RCBD, provided the observations are ordered on Blocks then Treatments, $\mathbf{M}_{\text{BU}} = \mathbf{I}_n$ and $\mathbf{M}_{\text{T}} = t^{-1}\mathbf{I}_b \otimes \mathbf{J}_t$, $\mathbf{M}_{\text{B}} = b^{-1}\mathbf{J}_b \otimes \mathbf{I}_t$ and $\mathbf{M}_{\text{B}} = b^{-1}\mathbf{J}_b \otimes \mathbf{J}_t$. Also, $\mathbf{Q}_{\text{G}} = \mathbf{M}_{\text{G}}$, $\mathbf{Q}_{\text{B}} = \mathbf{M}_{\text{B}} - \mathbf{M}_{\text{G}}$ and $\mathbf{Q}_{\text{BU}} = \mathbf{M}_{\text{BU}} - \mathbf{M}_{\text{B}}$.

a) Prove that $\mathbf{M}_{\text{T}}\mathbf{M}_{\text{B}} = \mathbf{M}_{\text{B}}\mathbf{M}_{\text{T}} = \mathbf{M}_{\text{G}}$ and that the product of $\mathbf{M}_{\text{G}}$ with any of the other **M** is $\mathbf{M}_{\text{G}}$.
b) Prove that $\mathbf{Q}_{\text{T}}\mathbf{M}_{\text{B}} = \mathbf{M}_{\text{B}}\mathbf{Q}_{\text{T}} = \mathbf{Q}_{\text{T}}\mathbf{M}_{\text{G}} = \mathbf{M}_{\text{G}}\mathbf{Q}_{\text{T}} = \mathbf{0}$.