3. Linear Model

Contents

- Unified approach
  - regression analysis
  - analysis of variance
  - balanced factorial experiments
- Extensions
  - analysis of unbalanced experiments
  - models with correlated responses
    - split plot designs
    - repeated measures
  - models with fixed and random components (mixed models)

Example 3.1: Yield of a chemical process

Yield (%) Y	Temperature (^{o}F) X_1	Time (hr) X_2
77	160	1
82	165	3
84	165	2
89	170	1
94	175	2

Linear regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i, \quad i = 1, 2, 3, 4, 5$$

Matrix formulation:

$$
\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \epsilon_1 \\ \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + \epsilon_2 \\ \beta_0 + \beta_1 X_{13} + \beta_2 X_{23} + \epsilon_3 \\ \beta_0 + \beta_1 X_{14} + \beta_2 X_{24} + \epsilon_4 \\ \beta_0 + \beta_1 X_{15} + \beta_2 X_{25} + \epsilon_5 \end{bmatrix}
$$

$$
= \begin{bmatrix} \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} \\ \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} \\ \beta_0 + \beta_1 X_{13} + \beta_2 X_{23} \\ \beta_0 + \beta_1 X_{14} + \beta_2 X_{24} \\ \beta_0 + \beta_1 X_{15} + \beta_2 X_{25} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}
$$

$$
= \begin{bmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ 1 & X_{13} & X_{23} \\ 1 & X_{14} & X_{24} \\ 1 & X_{15} & X_{25} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}
$$

Analysis of Variance (ANOVA)

Source of Variation	d.f.	Sums of Squares	Mean Squares
Model	2	$\sum_{i=1}^{5}(\hat{Y}_i - \bar{Y}_.)^2$	$\frac{1}{2}\, SS_{\text{model}}$
Error	2	$\sum_{i=1}^{5}(Y_i - \hat{Y}_i)^2$	$\frac{1}{2}\, SS_{\text{error}}$
C. total	4	$\sum_{i=1}^{5}(Y_i - \bar{Y}_.)^2$	

where

$$\bar{Y}_. = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i}$$

$$n = \text{total number of observations}$$

Example 3.2. Blood coagulation times (in seconds) for blood samples from six different rats. Each rat was fed one of three diets.

Diet 1	Diet 2	Diet 3
$Y_{11} = 62$	$Y_{21} = 71$	$Y_{31} = 72$
$Y_{12} = 60$		$Y_{32} = 68$
		$Y_{33} = 67$

Means model

$$Y_{ij} = \mu_i + \epsilon_{ij},$$

where

Y_{ij}: observed time for the j-th rat fed the i-th diet

μ_i: mean time for rats given the i-th diet

ϵ_{ij}: random error with $E(\epsilon_{ij}) = 0$

You can express this model as

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}$$

$$\begin{array}{cccc} \uparrow & \uparrow & \uparrow & \uparrow \\ \mathbf{Y} & X & \boldsymbol{\beta} & \boldsymbol{\epsilon} \end{array}$$

Assuming that $E(\epsilon_{ij}) = 0$ for all (i, j), this is a linear model with

$$E(\mathbf{Y}) = X\boldsymbol{\beta} \text{ and } Var(\mathbf{Y}) = \Sigma$$

An effects model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

This can be expressed as

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}$$

or

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

This is a linear model with

$$E(\mathbf{Y}) = X\boldsymbol{\beta} \text{ and } Var(\mathbf{Y}) = \Sigma$$

You could add the assumptions

- independent errors
- homogeneous variance,

  i.e.,

$$Var(\epsilon_{ij}) = \sigma^2$$

to obtain a linear model

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

with

$$\begin{aligned} E(\mathbf{Y}) &= X\boldsymbol{\beta} \\ Var(\mathbf{Y}) &= Var(\boldsymbol{\epsilon}) = \sigma^2 I \end{aligned}$$

Analysis of Variance (ANOVA)

Source of Variation	d.f.	Sums of Squares	Mean Squares
Diets	3 - 1 = 2	$\sum_{i=1}^{3} n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2$	$\frac{1}{2}\, SS_{\text{diets}}$
Error	$\sum_{i=1}^{3}(n_i - 1) = 3$	$\sum_{i=1}^{3}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i.})^2$	$\frac{1}{3}\, SS_{\text{error}}$
C. total	$\sum_{i=1}^{3}(n_i - 1) = 5$	$\sum_{i=1}^{3}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{..})^2$	

where

$$
\begin{aligned}
n_i &= \text{number of rats fed the i-th diet} \\
\bar{Y}_{i.} &= \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \\
\bar{Y}_{..} &= \frac{1}{n_.} \sum_{i=1}^{3} \sum_{j=1}^{n_i} Y_{ij} \\
n_. &= \sum_{i=1}^{3} n_i \\
&= \text{total number of observations}
\end{aligned}
$$

Example 3.3. A 2×2 factorial experiment

- Experimental units: 8 plots with 5 trees per plot.
- Factor 1: Variety (A or B)
- Factor 2: Fungicide use (new or old)
- Response: Percentage of apples with spots

Percentage of apples with spots	Variety	Fungicide use
$Y_{111} = 4.6$	A	new
$Y_{112} = 7.4$	A	new
$Y_{121} = 18.3$	A	old
$Y_{122} = 15.7$	A	old
$Y_{211} = 9.8$	B	new
$Y_{212} = 14.2$	B	new
$Y_{211} = 21.1$	B	old
$Y_{222} = 18.9$	B	old

Analysis of Variance (ANOVA)

Source of Variation	d.f.	Sums of Squares
Varieties	$2-1=1$	$4\sum_{i=1}^{2}(\bar{Y}_{i..}-\bar{Y}_{...})^2$
Fungicide use	$2-1=1$	$4\sum_{j=1}^{2}(\bar{Y}_{.j.}-\bar{Y}_{...})^2$
Variety $\times$ Fung. use interaction	$(2-1)(2-1)=1$	$2\sum_{i=1}^{2}\sum_{j=1}^{2}(\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}+\bar{Y}_{...})^2$
Error	$4(2-1)=4$	$\sum_{i}\sum_{j}\sum_{k}(Y_{ijk}-\bar{Y}_{ij.})^2$
Corrected total	$8-1=7$	$\sum_{i}\sum_{j}\sum_{k}(Y_{ijk}-\bar{Y}_{...})^2$

Linear model:

Y_{ijk} =	μ +	α_i	+	γ_j	+	δ_{ij}	+	ϵ_{ijk}
$\uparrow$		$\uparrow$		$\uparrow$		$\uparrow$		$\uparrow$
per-		variety		fung.		inter-		random
cent		effects		use		action		error
with		(i=1,2)		(j=1,2)				
spots								

Here we use 9 parameters

$$\boldsymbol{\beta}^T = (\mu\ \alpha_1\ \alpha_2\ \gamma_1\ \gamma_2\ \delta_{11}\ \delta_{12}\ \delta_{21}\ \delta_{22})$$

to represent the 4 response means,

$$E(Y_{ijk}) = \mu_{ij}, \quad i = 1, 2, \text{ and } j = 1, 2,$$

corresponding to the 4 combinations of levels of the two factors.

Write this model in the form

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{224} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \gamma_1 \\ \gamma_2 \\ \delta_{11} \\ \delta_{12} \\ \delta_{21} \\ \delta_{22} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \end{bmatrix}$$

Means model

$$
\begin{aligned}
Y_{ijk} \quad &= \quad \mu_{ij} + \epsilon_{ijk} \\
& \quad \nearrow \\
\mu_{ij} \quad &= E(Y_{ijk}) = \text{ mean percentage of apples with spots}
\end{aligned}
$$

This linear model can be written in the form $\mathbf{Y} = X\beta + \epsilon$, that is,

$$
\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \end{bmatrix}
=
\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix}
+
\begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \end{bmatrix}
$$

General Linear Model

Any linear model can be written as

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{k1} \\ X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & & \vdots \\ X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$\uparrow$ observed responses

$\uparrow$ the elements of X are known (non-random) values

$\uparrow$ random errors are not observed

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \text{ is a random vector with}$$

(1) $E(\mathbf{Y}) = X\boldsymbol{\beta}$

for some known $n \times k$ matrix X of constants and unknown $k \times 1$ parameter vector $\boldsymbol{\beta}$

(2) Complete the model by specifying a probability distribution for the possible values of $\mathbf{Y}$ or ϵ

Sometimes we will only specify the covariance matrix

$$Var(\mathbf{Y}) = \Sigma$$

Since

$$\mathbf{Y} = X\beta + \epsilon$$

we have

$$\epsilon = \mathbf{Y} - X\beta = \mathbf{Y} - E(\mathbf{Y})$$

and

$$\begin{aligned} E(\epsilon) &= \mathbf{0} \\ Var(\epsilon) &= Var(\mathbf{Y}) = \Sigma \end{aligned}$$

Gauss-Markov Model

Defn 3.1: The linear model

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

is a Gauss-Markov model if

$$Var(\mathbf{Y}) = Var(\boldsymbol{\epsilon}) = \sigma^2 I$$

for an unknown constant σ^2.

Notation:

$$\mathbf{Y} \sim \left[E(\mathbf{Y}) = X\boldsymbol{\beta}\,,\; Var(\mathbf{Y}) = \sigma^2 I \right]$$

- The distribution of **Y** is not completely specified.

Normal Theory Gauss-Markov Model

Defn 3.2: A normal theory Gauss-Markov model is a Gauss-Markov model in which $\mathbf{Y}$ (or ϵ) has a multivariate normal distribution.

$$\mathbf{Y} \sim N\,(X\boldsymbol{\beta}\,,\ \sigma^2 I)$$

$$\begin{array}{llll} \nearrow & \uparrow & \nwarrow & \nwarrow \\ \text{distr.} & \text{multivar.} & E(\mathbf{Y}) & Var(\mathbf{Y}) \\ \text{as} & \text{normal} & & \\ & \text{distr.} & & \end{array}$$

The additional assumption of a normal distribution is

(1) not needed for some estimation results

(2) useful in creating

- confidence intervals
- tests of hypotheses

Objectives

(i) Develop estimation procedures

- Estimate β?
- Estimate $E(\mathbf{Y}) = X\beta$
- Estimable functions of β.

(ii) Quantify uncertainty in estimates

- variances, standard deviations
- distributions
- confidence intervals

(iii) Analysis of Variance (ANOVA)

(iv) Tests of hypotheses

- Distributions of quadratic forms
- F-tests
- power

(v) sample size determination

Least Squares Estimation

For the linear model with

$$E(\mathbf{Y}) = X\boldsymbol{\beta} \text{ and } Var(\mathbf{Y}) = \Sigma$$

we have

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{k1} \\ X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & & \vdots \\ X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

and

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \epsilon_i = \mathbf{X}_i^T \boldsymbol{\beta} + \epsilon_i$$

where $\mathbf{X}_i^T = (X_{1i}\, X_{2i} \cdots X_{ki})$ is the i-th row of the model matrix X.

OLS Estimator

Defn 3.3: For a linear model with $E(\mathbf{Y}) = X\beta$, any vector $\mathbf{b}$ that minimizes the sum of squared residuals

$$
\begin{aligned}
Q(\mathbf{b}) &= \sum_{i=1}^{n} (Y_i - \mathbf{X}_i^T \mathbf{b})^2 \\
&= (\mathbf{Y} - X\mathbf{b})^T (\mathbf{Y} - X\mathbf{b})
\end{aligned}
$$

is an ordinary least squares (OLS) estimator for β.

OLS Estimating Equations

For $j = 1, 2, \ldots, k$, solve

$$0 = \frac{\partial Q(\mathbf{b})}{\partial b_j} = 2 \sum_{i=1}^{n} (Y_i - \mathbf{X}_i^T \mathbf{b})\, X_{ij}$$

These equations are expressed in matrix form as

$$\begin{aligned} \mathbf{0} &= X^T(\mathbf{Y} - X\mathbf{b}) \\ &= X^T\mathbf{Y} - X^T X\mathbf{b} \end{aligned}$$

or

$$X^T X\mathbf{b} = X^T \mathbf{Y}$$

These are called the **normal equations.**

If $X_{n\times k}$ has full column rank, i.e., rank(X) = k, then

(i) $X^T X$ is non-singular

(ii) $(X^T X)^{-1}$ exists and is unique

Consequently,

$$(X^T X)^{-1}(X^T X)\mathbf{b} = (X^T X)^{-1} X^T \mathbf{y}$$

and

$$\mathbf{b} = (X^T X)^{-1} X^T \mathbf{Y}$$

is the unique solution to the normal equations.

If $\text{rank}(X) < k$, then

(i) there are infinitely many solutions to the normal equations

(ii) if $(X^T X)^-$ is a generalized inverse of $X^T X$, then

$$\mathbf{b} = (X^T X)^- X^T \mathbf{Y}$$

is a solution of the normal equations.

Generalized Inverse

<u>Defn 3.4:</u> For a given $m \times n$ matrix A, any $n \times m$ matrix G that satisfies

$$AGA = A$$

is a **generalized inverse** of A.

<u>Comments</u>

- We will often use A^- to denote a generalized inverse of A.
- There may be infinitely many generalized inverses.
- If A is an $m \times m$ nonsingular matrix, then $G = A^{-1}$ is the unique generalized inverse for A.

Example 3.5.

$$A = \begin{bmatrix} 16 & -6 & -10 \\ -6 & 21 & -15 \\ -10 & -15 & 25 \end{bmatrix}, \qquad rank(A) = 2.$$

A generalized inverse is

$$G = \begin{bmatrix} \frac{1}{20} & 0 & 0 \\ 0 & \frac{1}{30} & 0 \\ 0 & 0 & \frac{1}{50} \end{bmatrix}, \qquad \text{Note that } AGA = A.$$

Another generalized inverse is

$$G = \begin{bmatrix} \begin{bmatrix} 16 & -6 \\ -6 & 21 \end{bmatrix}^{-1} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & 0 \end{bmatrix} = \begin{bmatrix} \frac{21}{300} & \frac{6}{300} & 0 \\ \frac{6}{300} & \frac{16}{300} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Example 3.2. **Means model**

$$
\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}
$$

For this model

$$
X^T X = \begin{bmatrix} n_1 & 0 & 0 \\ 0 & n_2 & 0 \\ 0 & 0 & n_3 \end{bmatrix}, \quad X^T \mathbf{Y} = \begin{bmatrix} Y_{11} + Y_{12} \\ Y_{21} \\ Y_{31} + Y_{32} + Y_{33} \end{bmatrix}
$$

and the unique OLS estimator for $\beta = (\mu_1\ \mu_2\ \mu_3)^T$ is

$$\begin{aligned} \mathbf{b} &= (X^T X)^{-1} X^T \mathbf{Y} \\ &= \begin{bmatrix} \frac{1}{n_1} & 0 & 0 \\ 0 & \frac{1}{n_2} & 0 \\ 0 & 0 & \frac{1}{n_3} \end{bmatrix} \begin{bmatrix} Y_{1.} \\ Y_{2.} \\ Y_{3.} \end{bmatrix} = \begin{bmatrix} \bar{Y}_{1.} \\ \bar{Y}_{2.} \\ \bar{Y}_{3.} \end{bmatrix} \end{aligned}$$

Example 3.2. **Effects model**

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}$$

Here

$$X^T X = \begin{bmatrix} n_. & n_1 & n_2 & n_3 \\ n_1 & n_1 & 0 & 0 \\ n_2 & 0 & n_2 & 0 \\ n_3 & 0 & 0 & n_3 \end{bmatrix}, \quad X^T \mathbf{Y} = \begin{bmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ Y_{3.} \end{bmatrix}$$

Solution A:

$$(X^T X)^- = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{n_1} & 0 & 0 \\ 0 & 0 & \frac{1}{n_2} & 0 \\ 0 & 0 & 0 & \frac{1}{n_3} \end{bmatrix}$$

and a solution to the normal equations is

$$\mathbf{b} = (X^T X)^- X^T \mathbf{Y}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & n_1^{-1} & 0 & 0 \\ 0 & 0 & n_2^{-1} & 0 \\ 0 & 0 & 0 & n_3^{-1} \end{bmatrix} \begin{bmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ Y_{3.} \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{Y}_{1.} \\ \bar{Y}_{2.} \\ \bar{Y}_{3.} \end{bmatrix}$$

Solution B: Another generalized inverse for X^TX is

$$(X^TX)^- = \begin{bmatrix} A^{-1} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} n_. & n_1 & n_2 \\ n_1 & n_1 & 0 \\ n_2 & 0 & n_2 \end{bmatrix}^{-1} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & \quad 0 & \quad 0 \end{matrix} & 0 \end{bmatrix}.$$

Use result 1.4(ii) to compute

$$A^{-1} = \frac{1}{|A|} \begin{bmatrix} |A_{11}| & -|A_{21}| & |A_{31}| \\ -|A_{12}| & |A_{22}| & -|A_{32}| \\ |A_{13}| & -|A_{23}| & |A_{33}| \end{bmatrix}$$

$$= \frac{1}{n_1 n_2 n_3} \begin{bmatrix} n_1 n_2 & -n_1 n_2 & -n_1 n_2 \\ -n_1 n_2 & n_2(n_1 + n_3) & n_1 n_2 \\ -n_1 n_2 & n_1 n_2 & n_1(n_2 + n_3) \end{bmatrix}$$

Then

$$A^{-1} = \frac{1}{n_3}\begin{bmatrix} 1 & -1 & -1 \\ -1 & \frac{n_1+n_3}{n_1} & 1 \\ -1 & 1 & \frac{n_2+n_3}{n_2} \end{bmatrix}$$

and a solution to the normal equations is

$$\begin{aligned}\mathbf{b} &= (X^T X)^- X^T \mathbf{Y} \\ &= \frac{1}{n_3}\begin{bmatrix} 1 & -1 & -1 & 0 \\ -1 & \frac{n_1+n_3}{n_1} & 1 & 0 \\ -1 & 1 & \frac{n_2+n_3}{n_2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ Y_{3.} \end{bmatrix}\end{aligned}$$

$$
= \frac{1}{n_3}\begin{bmatrix} Y_{..} - Y_{1.} - Y_{2.} \\ -Y_{..} + (\frac{n_1+n_3}{n_1})Y_{1.} + Y_{2.} \\ -Y_{..} + Y_{1.} + (\frac{n_2+n_3}{n_2})Y_{2.} \\ 0 \end{bmatrix} = \begin{bmatrix} \bar{Y}_{3.} \\ \bar{Y}_{1.} - \bar{Y}_{3.} \\ \bar{Y}_{2.} - \bar{Y}_{3.} \\ 0 \end{bmatrix}
$$

This is the OLS estimator for

$$
\beta^T = [\mu, \ \ \alpha_1, \ \ \alpha_2, \ \ \alpha_3]
$$

reported by PROC GLM in the SAS package, but it is not the only possible solution to the normal equations.

Solution C: Another generalized inverse for $X^T X$ is

$$(X^T X)^- = \frac{1}{n_1 n_2 n_3} \begin{bmatrix} n_2 n_3 & 0 & -n_2 n_3 & -n_2 n_3 \\ 0 & 0 & 0 & 0 \\ -n_2 n_3 & 0 & n_3(n_1 + n_2) & n_2 n_3 \\ -n_2 n_3 & 0 & n_2 n_3 & n_2(n_1 + n_3) \end{bmatrix}$$

The corresponding solution to the normal equations is

$$\mathbf{b} = (X^T X)^- X^T \mathbf{Y} = \begin{bmatrix} \bar{Y}_{1.} \\ 0 \\ \bar{Y}_{2.} - \bar{Y}_{1.} \\ \bar{Y}_{3.} - \bar{Y}_{1.} \end{bmatrix}$$

Solution D: Another generalized inverse for $X^T X$ is

$$(X^T X)^- = \begin{bmatrix} \frac{2}{n} & -\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} \\ -\frac{1}{n} & \frac{1}{n_1} & 0 & 0 \\ -\frac{1}{n} & 0 & \frac{1}{n_2} & 0 \\ -\frac{1}{n} & 0 & 0 & \frac{1}{n_3} \end{bmatrix}$$

The corresponding solution to the normal equations is

$$\begin{aligned} \mathbf{b} &= (X^T X)^- X^T \mathbf{Y} \\ &= \begin{bmatrix} \frac{2}{n} & -\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} \\ -\frac{1}{n} & \frac{1}{n_1} & 0 & 0 \\ -\frac{1}{n} & 0 & \frac{1}{n_2} & 0 \\ -\frac{1}{n} & 0 & 0 & \frac{1}{n_3} \end{bmatrix} \begin{bmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ Y_{3.} \end{bmatrix} = \begin{bmatrix} \bar{Y}_{..} \\ \bar{Y}_{1.} - \bar{Y}_{..} \\ \bar{Y}_{2.} - \bar{Y}_{..} \\ \bar{Y}_{3.} - \bar{Y}_{..} \end{bmatrix} \end{aligned}$$

Evaluating Generalized Inverses

Algorithm 3.1:

(i) Find any $r \times r$ nonsingular submatrix of A where r=rank(A). Call this matrix W.

(ii) Invert and transpose W, ie., compute $(W^{-1})^T$.

(iii) Replace each element of W in A with the corresponding element of $(W^{-1})^T$

(iv) Replace all other elements in A with zeros.

(v) Transpose the resulting matrix to obtain G, a generalized inverse for A.

Example 3.6.

$$A = \begin{bmatrix} 4 & 1 & 2 & 0 \\ 1 & 1 & 5 & 15 \\ 3 & 1 & 3 & 5 \end{bmatrix}, \quad \text{Define} \quad W = \begin{bmatrix} 1 & 5 \\ 1 & 3 \end{bmatrix}.$$

Then

$$(W^{-1})^T = \begin{bmatrix} -3/2 & 1/2 \\ 5/2 & -1/2 \end{bmatrix}$$

and thus

$$G = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -\frac{3}{2} & \frac{1}{2} & 0 \\ 0 & \frac{5}{2} & -\frac{1}{2} & 0 \end{bmatrix}^T = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\frac{3}{2} & \frac{5}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}$$

Another solution

$$A = \begin{bmatrix} 4 & 1 & 2 & 0 \\ 1 & 1 & 5 & 15 \\ 3 & 1 & 3 & 5 \end{bmatrix}, \text{Define} \quad W = \begin{bmatrix} 4 & 0 \\ 3 & 5 \end{bmatrix}.$$

Then

$$(W^{-1})^T = \begin{bmatrix} -\frac{5}{20} & -\frac{3}{20} \\ \frac{0}{20} & \frac{4}{20} \end{bmatrix}$$

and thus

$$G = \begin{bmatrix} \frac{5}{20} & 0 & 0 & -\frac{3}{20} \\ 0 & 0 & 0 & 0 \\ \frac{0}{20} & 0 & 0 & \frac{4}{20} \end{bmatrix}^T = \begin{bmatrix} \frac{5}{20} & 0 & \frac{0}{20} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -\frac{3}{20} & 0 & \frac{4}{20} \end{bmatrix}$$

Algorithm 3.2

For any $m \times n$ matrix A with rank(A) = r,

(i) compute a singular value decomposition of A (see result 1.14) to obtain

$$PAQ = \begin{bmatrix} D_{r\times r} & 0 \\ 0 & 0 \end{bmatrix}$$

where

P is an $m \times m$ orthogonal matrix

Q is an $n \times n$ orthogonal matrix

D is an $r \times r$ matrix of singular values

(ii) $G = Q \begin{bmatrix} D^{-1} & F_1 \\ F_2 & F_3 \end{bmatrix} P$ is a generalized inverse for A for any choice of F_1, F_2, F_3.

Proof: Check if $AGA = A$.

$$AGA = P^T \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} Q^T Q \begin{bmatrix} D^{-1} & F_1 \\ F_2 & F_3 \end{bmatrix} P P^T \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} Q^T$$

$$= P^T \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} D^{-1} & F_1 \\ F_2 & F_3 \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} Q^T$$

$$= P^T \begin{bmatrix} I & DF_1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} Q^T$$

$$= P^T \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} Q^T$$

$$= A$$

Moore-Penrose Inverse

Defn 3.5: For any matrix A there is a unique matrix M, called the Moore-Penrose inverse, that satisfies

(i) $AMA = A$

(ii) $MAM = M$

(iii) AM is symmetric

(iv) MA is symmetric

Result 3.1

$$M = Q \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} P$$

is the Moore-Penrose inverse of A, where

$$PAQ = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

is a singular value decomposition of A.

Properties of generalized inverses of $X^T X$

Normal equations: $(X^T X)\mathbf{b} = X^T \mathbf{Y}$

Result 3.3 If G is a generalized inverse of $X^T X$, then

(i) G^T is a generalized inverse of $X^T X$.

(ii) $XGX^T X = X$, i.e., GX^T is a generalized inverse of X.

(iii) XGX^T is invariant with respect to the choice of G.

(iv) XGX^T is symmetric.

Proof:

(i) Since G is a generalized inverse of $(X^T X)$,

$$(X^T X)G(X^T X) = X^T X.$$

Taking the transpose of both sides

$$\begin{aligned} \left[X^T X\right]^T &= \left[(X^T X)G(X^T X)\right]^T \\ &= (X^T X)^T G^T (X^T X)^T \end{aligned}$$

But $(X^T X)^T = X^T (X^T)^T = X^T X$,

hence $(X^T X)G^T (X^T X) = (X^T X)$

(ii) From (i)

$$\underline{(X^T X)G^T}(X^T X) = (X^T X)$$

$\nwarrow$ Call this B

Then

$$\begin{aligned} 0 &= BX^T X - X^T X \\ &= (BX^T X - X^T X)(B^T - I) \\ &= BX^T XB^T - X^T XB^T - BX^T X - X^T X \\ &= (BX^T - X^T)(BX^T - X^T)^T \end{aligned}$$

Hence, $0 = BX^T - X^T$

$$\begin{aligned} \Rightarrow \quad & BX^T = X^T \\ \Rightarrow \quad & X^T X G^T X^T = X^T \end{aligned}$$

Taking the transpose

$$\begin{aligned} X &= (X^T X G^T X^T)^T \\ &= X\underline{GX^T}X \end{aligned}$$

Hence, GX^T is a generalized inverse for X.

(iii) Suppose F and G are generalized inverses for X^TX. Then, from (ii)

$$XGX^TX = X$$

and

$$XFX^TX = X$$

It follows that

$$\begin{aligned} 0 &= X - X \\ &= (XGX^TX - XFX^TX) \\ &= (XGX^TX - XFX^TX)(G^TX^T - F^TX^T) \\ &= (XGX^T - XFX^T)X(G^TX^T - F^TX^T) \\ &= (XGX^T - XFX^T)(XG^TX^T - XF^TX^T) \\ &= (XGX^T - XFX^T)(XGX^T - XFX^T)^T \end{aligned}$$

Since the (i,i) diagonal element of the result of multiplying a matrix by its transpose is the sum of the squared entries in the *i-th* row of the matrix, the diagonal elements of the product are all zero only if all entries are zero in every row of the matrix. Consequently,

$$(XGX^T - XFX^T) = 0$$

(iv) For any generalized inverse G,

$$T = GX^TXG^T$$

is a symmetric generalized inverse. Then

$$XTX^T$$

is symmetric and from (iii),

$$XGX^T = XTX^T.$$

Estimation of the Mean Vector $E(\mathbf{Y}) = X\beta$

For any solution to the normal equations, say

$$\mathbf{b} = (X^T X)^- X^T \mathbf{Y} \ ,$$

the OLS estimator for $E(\mathbf{Y}) = X\beta$ is

$$\hat{\mathbf{Y}} = X\mathbf{b} = X(X^T X)^- X^T \mathbf{Y} = P_X \mathbf{Y}$$

- The matrix $P_X = X(X^T X)^- X^T$ is called an *orthogonal projection matrix*.
- $\hat{\mathbf{Y}} = P_X \mathbf{Y}$ is the projection of $\mathbf{Y}$ onto the space spanned by the columns of X.

Result 3.4 Properties of a projection matrix

$$P_X = X(X^TX)^-X^T$$

(i) P_X is invariant to the choice of $(X^TX)^-$. For any solution

$$\mathbf{b} = (X^TX)^-X^T\mathbf{Y}$$

to the normal equations

$$\hat{\mathbf{Y}} = X\mathbf{b} = P_X\mathbf{Y}$$

is the same. (from Result 3.3 (iii))

(ii) P_X is symmetric (from Result 3.3 (iv))

(iii) P_X is idempotent ($P_X P_X = P_X$)

(iv) $P_X X = X$
(from Result 3.3 (ii))

(v) Partition X as

$$X = [X_1 | X_2 | \cdots | X_k],$$

then $P_X X_j = X_j$

Residuals:

$$e_i = Y_i - \hat{Y}_i \quad i = 1, \ldots, n$$

The vector of residuals is

$$\mathbf{e} \;=\; \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - X\mathbf{b} = \mathbf{Y} - P_X\mathbf{Y} = (I - P_X)\mathbf{Y}$$

Comment:

- $\hat{\mathbf{Y}}$ is in the vector space spanned by the columns of X. It has dimension $n - rank(X)$.
- P_X is a projection matrix that projects $\mathbf{Y}$ onto the space spanned by the columns of X.
- The residual vector $\mathbf{e}$ is in the space orthogonal to the space spanned by the columns of X. It has dimension $n - rank(X)$.
- $I - P_X$ is a projection matrix that projects $\mathbf{Y}$ onto the space orthogonal to the space spanned by the columns of X.

Result 3.5 Properties of $I - P_X$

(i) $I - P_X$ is symmetric
(ii) $I - P_X$ is idempodent

$$(I - P_X)(I - P_X) = I - P_X$$

(iii)

$$(I - P_X)P_X = P_X - P_X P_X = P_X - P_X = 0$$

(iv)

$$(I - P_X)X = X - P_X X = X - X = 0$$

(v) Partition X as $[X_1|X_2|\cdots|X_k]$
then

$$(I - P_X)\mathbf{X}_j = 0$$

(vi) Residuals are invariant with respect to the choice of $(X^T X)^-$, so

$$\mathbf{e} - \mathbf{Y} - X\mathbf{b} = (I - P_X)\mathbf{Y}$$

is the same for any solution

$$\mathbf{b} = (X^T X)^- X^T \mathbf{Y}$$

to the normal equations

Partition of a total sum of squares

Squared length of $\mathbf{Y}$ is

$$\sum_{i=1}^{n} y_i^2 = \mathbf{Y}^T\mathbf{Y}$$

Squared length of the residual vector is

$$\sum_{i=1}^{n} e_i^2 = \mathbf{e}^T\mathbf{e} = [(I - P_X)\mathbf{Y}]^T(I - P_X)\mathbf{Y} = \mathbf{Y}^T(I - P_X)\mathbf{Y}$$

Squared length of $\hat{\mathbf{Y}} = P_X\mathbf{Y}$ is

$$\sum_{i=1}^{n} \hat{Y}_i^2 = \hat{\mathbf{Y}}^T\hat{\mathbf{Y}} = (P_X\mathbf{Y})^T(P_X\mathbf{Y}) = \mathbf{Y}^T P_X P_X \mathbf{Y}$$

$$\Longrightarrow \quad \mathbf{Y}^T\mathbf{Y} = \mathbf{Y}^T(P_X + I - P_X)\mathbf{Y} = \mathbf{Y}^T P_X \mathbf{Y} + \mathbf{Y}^T(I - P_X)\mathbf{Y} .$$

ANOVA

Source of Variation	Degrees of Freedom	Sums of Squares
(uncorrected) model	rank(X)	$\hat{\mathbf{Y}}^T\hat{\mathbf{Y}} = \mathbf{Y}^T P_X \mathbf{Y}$
residuals	n-rank(X)	$\mathbf{e}^T\mathbf{e} = \mathbf{Y}^T(I - P_X)\mathbf{Y}$
(uncorrected) total	n	$\mathbf{Y}^T\mathbf{Y} = \sum_{i=1}^{n} y_i^2$

Properties of $\hat{\mathbf{Y}}$

Result 3.6 For the linear model

$$E(\mathbf{Y}) = X\beta \quad \text{and} \quad Var(Y) = \Sigma,$$

the OLS estimator $\hat{\mathbf{Y}} = X\mathbf{b} = P_X\mathbf{Y}$ for $X\beta$ is

(i) unbiased, i.e., $E(\hat{\mathbf{Y}}) = X\beta$

(ii) a linear function of $\mathbf{Y}$

(iii) has variance-covariance matrix

$$Var(\hat{\mathbf{Y}}) = P_X \Sigma P_X$$

This is true for any solution

$$b = (X^T X)^- X^T \mathbf{Y}$$

to the normal equations.

Proof:

(ii) is trivial, since $\hat{Y} = P_X \mathbf{Y}$

(iii) follows from result 2.1.(ii)

(i)

$$\begin{aligned} E(\hat{\mathbf{Y}}) &= E(P_X \mathbf{Y}) \\ &= P_X E(Y) \text{ from result 2.1.(i)} \\ &= P_X X \beta \\ &= X\beta \text{ since } P_X X = X \end{aligned}$$

Comments:

- $\hat{\mathbf{Y}} = X\mathbf{b} = P_X\mathbf{Y}$ is said to be a linear unbiased estimator for $E(\mathbf{Y}) = X\beta$

- For the Gauss-Markov model, $Var(\mathbf{Y}) = \sigma^2 I$ and

$$
\begin{aligned}
Var(\hat{\mathbf{Y}}) &= P_X(\sigma^2 I)P_X \\
&= \sigma^2 P_X P_X \\
&= \sigma^2 P_X \\
&= \sigma^2 \underline{X(X^T X)^- X^T}
\end{aligned}
$$

$\uparrow$

this is sometimes called the "hat" matrix.

Questions

- Is $\hat{\mathbf{Y}} = X\mathbf{b} = P_X\mathbf{Y}$ the best estimator for $E(\mathbf{Y}) = X\beta$?
- Is $\hat{\mathbf{Y}} = X\mathbf{b} = P_X\mathbf{Y}$ the best estimator for $E(\mathbf{Y}) = X\beta$ in the class of linear, unbiased estimators?
- What other linear functions of β, say

  $$\mathbf{c}^T\beta = c_1\beta_1 + c_2\beta_2 + \cdots + c_k\beta_k,$$

  have OLS estimators that are invariant to the choice of

  $$\mathbf{b} = (X^TX)^-X^T\mathbf{Y}$$

  that solves the normal equations?

Estimable Functions

Some estimates of linear functions of the parameters have the same value, regardless of which solution to the normal equations is used

- These are called estimable functions
- An example is $E(\mathbf{Y}) = X\beta$

Check that $X\mathbf{b}$ has the same value for each solution to the normal equations obtained in Example 3.2, i.e.,

$$X\mathbf{b} = \begin{bmatrix} \bar{Y}_{1.} \\ \bar{Y}_{1.} \\ \bar{Y}_{2.} \\ \bar{Y}_{3.} \\ \bar{Y}_{3.} \\ \bar{Y}_{3.} \end{bmatrix}$$

Defn 3.6: For a linear model

$$E(\mathbf{Y}) = X\boldsymbol{\beta} \ \text{ and } \ Var(\mathbf{Y}) = \Sigma$$

we will say that

$$\mathbf{c}^T\boldsymbol{\beta} = c_1\beta_1 + c_2\beta_2 + \cdots + c_k\beta_k$$

is **estimable** if there exists a linear unbiased estimator $\mathbf{a}^T\mathbf{Y}$ for $\mathbf{c}^T\boldsymbol{\beta}$, i.e., for some non-random vector $\mathbf{a}$, we have $E(\mathbf{a}^T\mathbf{Y}) = \mathbf{c}^T\boldsymbol{\beta}$.

- No dependence on variance structure.

Example 3.2. Blood coagulation times

Diet 1	Diet 2	Diet 3
$Y_{11} = 62$	$Y_{21} = 71$	$Y_{31} = 72$
$Y_{12} = 60$		$Y_{32} = 68$
		$Y_{33} = 67$

The **Effects model** $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ can be written as

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}$$

Examples of estimable functions

$\underline{\mu + \alpha_1}$

Choose $\mathbf{a}^T = (\frac{1}{2}\ \frac{1}{2}\ 0\ 0\ 0\ 0)$. Then,

$$\begin{aligned} E(\mathbf{a}^T\mathbf{Y}) &= E(\frac{1}{2}Y_{11} + \frac{1}{2}Y_{12}) \\ &= \frac{1}{2}E(Y_{11}) + \frac{1}{2}E(Y_{12}) \\ &= \frac{1}{2}(\mu + \alpha_1) + \frac{1}{2}(\mu + \alpha_1) \\ &= \mu + \alpha_1 \end{aligned}$$

Choose $\mathbf{a}^T = (1\ 0\ 0\ 0\ 0\ 0)$ and note that $E(\mathbf{a}^T\mathbf{Y}) = E(Y_{11}) = \mu + \alpha_1$.

$\underline{\mu + \alpha_2}$

Choose $\mathbf{a}^T = (0\ 0\ 1\ 0\ 0\ 0)$. Then,

$$\mathbf{a}^T\mathbf{Y} = Y_{21}$$

and

$$E(\mathbf{a}^T\mathbf{Y}) = E(Y_{21}) = \mu + \alpha_2.$$

$\underline{\mu + \alpha_3}$

Choose $\mathbf{a}^T = (0\ 0\ 0\ 1\ 0\ 0)$. Then,

$$E(\mathbf{a}^T\mathbf{Y}) = E(Y_{31}) = \mu + \alpha_3$$

$\underline{\alpha_1 - \alpha_2}$

Note that

$$\begin{aligned}
\alpha_1 - \alpha_2 &= (\mu + \alpha_1) - (\mu + \alpha_2) \\
&= E(Y_{11}) - E(Y_{21}) \\
&= E(Y_{11} - Y_{21}) \\
&= E(\mathbf{a}^T \mathbf{Y})
\end{aligned}$$

where

$$\mathbf{a}^T = (1\ 0\ \text{-}1\ 0\ 0\ 0)$$

$\underline{2\mu + 3\alpha_1 - \alpha_2}$

Note that

$$\begin{aligned} 2\mu + 3\alpha_1 - \alpha_2 &= 3(\mu + \alpha_1) - (\mu + \alpha_2) \\ &= 3E(Y_{11}) - E(Y_{21}) \\ &= E(3Y_{11} - Y_{21}) \\ &= E(\mathbf{a}^T\mathbf{Y}) \end{aligned}$$

where

$$\mathbf{a}^T = (3\ 0\ \text{-}1\ 0\ 0\ 0)$$

Quantities that are not estimable include

$$\mu,\ \alpha_1,\ \alpha_2,\ \alpha_3,\ 3\alpha_1,\ \alpha_1+\alpha_2$$

To show that a linear function of parameters,

$$c_0\mu + c_1\alpha_1 + c_2\alpha_2 + c_3\alpha_3$$

is not estimable, one must show that there is no non-random vector

$$\mathbf{a}^T = (a_0,\ a_1,\ a_2,\ a_3)$$

for which

$$E(\mathbf{a}^T\mathbf{Y}) = c_0\mu + c_1\alpha_1 + c_2\alpha_2 + c_3\alpha_3$$

For α_1 to be estimable we would need to find an **a** that satisfies

$$\begin{aligned}
\alpha_1 &= E(\mathbf{a}^T\mathbf{Y}) \\
&= a_1 E(Y_{11}) + a_2 E(Y_{12}) + a_3 E(Y_{21}) \\
&\quad + a_4(E(Y_{31}) + a_5 E(Y_{32}) + a_6 E(Y_{33}) \\
&= (a_1 + a_2)(\mu + \alpha_1) + a_3(\mu + \alpha_2) \\
&\quad + (a_4 + a_5 + a_6)(\mu + \alpha_3)
\end{aligned}$$

This implies $0 = a_3 = (a_4 + a_5 + a_6)$.

Then $\alpha_1 = (a_1 + a_2)(\mu + \alpha_1)$ which is impossible.

Example 3.1. Yield of a chemical process

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix} = \begin{bmatrix} 1 & 160 & 1 \\ 1 & 165 & 3 \\ 1 & 165 & 2 \\ 1 & 170 & 1 \\ 1 & 175 & 2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}$$

Since X has full column rank, each element of $\boldsymbol{\beta}$ is estimable.

Consider $\beta_1 = \mathbf{c}^T \boldsymbol{\beta}$ where $c = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

Since X has full column rank, the unique least squares estimator for $\boldsymbol{\beta}$ is

$$\mathbf{b} = (X^T X)^{-1} \mathbf{X}^T \mathbf{Y}$$

and an unbiased linear estimator for $\mathbf{c}^T \boldsymbol{\beta}$ is

$$\mathbf{c}^T \mathbf{b} = \underbrace{\mathbf{c}^T (X^T X)^{-1} X^T}_{\text{call this } \mathbf{a}^T} \mathbf{Y}$$

Result 3.7 For a linear model with

$$E(\mathbf{Y}) = X\beta \text{ and } Var(Y) = \Sigma$$

(i) The expectation of any observation is estimable.

(ii) A linear combination of estimable functions is estimable.

(iii) Each element of β is estimable if and only if rank$(X) = k =$ number of columns.

(iv) Every $\mathbf{c}^T\beta$ is estimable if and only if rank$(X) = k =$ number of columns in X.

Proof:

(i) For $\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ with $E(\mathbf{Y}) = X\beta$, we have

$$Y_i = \mathbf{a}_i^T \mathbf{Y} \text{ where } \mathbf{a}_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow \begin{array}{c} \text{one in} \\ \text{the } i\text{th} \\ \text{position} \end{array}$$

Then

$$E(Y_i) = E(\mathbf{a}_i^T \mathbf{Y}) = \mathbf{a}_i^T E(\mathbf{Y}) = \underline{a_i^T X}\beta = \mathbf{c}_i^T \beta$$

(ii) Suppose $c_i^T\beta$ is estimable. Then, there is an $\mathbf{a}_i$ such that $E(\mathbf{a}_i^T\mathbf{Y}) = \mathbf{c}_i^T\beta$. Now consider a linear combination of estimable functions

$$w_1\mathbf{c}_1^T\beta + w_2\mathbf{c}_2^T\beta + \cdots + w_p\mathbf{c}_p^T\beta$$

Let $\mathbf{a} = w_1\mathbf{a}_1 + w_2\mathbf{a}_2 + \cdots + w_p\mathbf{a}_p$.

Then,

$$\begin{aligned} E(\mathbf{a}^T\mathbf{Y}) &= E(w_1\mathbf{a}_1^T\mathbf{Y} + \cdots + w_p\mathbf{a}_p^T\mathbf{Y}) \\ &= w_1E(\mathbf{a}_1^T\mathbf{Y}) + \cdots + w_pE(\mathbf{a}_p^T\mathbf{Y}) \\ &= w_1\mathbf{c}_1^T\beta + \cdots + w_p\mathbf{c}_p^T\beta \end{aligned}$$

(iii) Previous argument.

(iv) Follows from (ii) and (iii).

Result 3.8. For a linear model with

$$E(\mathbf{Y}) = X\beta \text{ and } Var(\mathbf{Y}) = \Sigma,$$

each of the following is true if and only if $\mathbf{c}^T\beta$ is estimable.

(i) $\mathbf{c}^T = \mathbf{a}^T X$ for some $\mathbf{a}$ i.e., $\mathbf{c}$ is in the space spanned by the rows of X.

(ii) $\mathbf{c}^T\mathbf{a} = 0$ for every $\mathbf{a}$ for which $X\mathbf{a} = \mathbf{0}$.

(iii) $\mathbf{c}^T\mathbf{b}$ is the same for any solution to the normal equations $(X^TX)\mathbf{b} = X^T\mathbf{Y}$, i.e., there is a unique least squares estimator for $\mathbf{c}^T\beta$.

Use Result 3.8. (ii) to show that μ is not estimable in Example 3.2. In that case

$$E(\mathbf{Y}) = X\boldsymbol{\beta} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

and

$$\mu = \mathbf{c}^T\boldsymbol{\beta} = [1\ 0\ 0\ 0]\boldsymbol{\beta}.$$

Let $\mathbf{d}^T = [1\ \ -1\ \ -1\ \ -1]$, then

$$X\mathbf{d} = \mathbf{0}, \quad \text{but} \quad \mathbf{c}^T\mathbf{d} = 1 \neq 0$$

Hence, μ is not estimable.

Part (ii) of Result 3.8 sometimes provides a convenient way to identify all possible estimable functions of β.

In example 3.2, $X\mathbf{d} = \mathbf{0}$ if and only if

$$\mathbf{d} = w \begin{bmatrix} 1 \\ -1 \\ -1 \\ -1 \end{bmatrix} \text{ for some scalar } w.$$

Then, $\mathbf{c}^T\beta$ is estimable if and only if

$$0 = \mathbf{c}^T\mathbf{d} = w(c_1 - c_2 - c_3 - c_4)$$

$$\iff \quad c_1 = c_2 + c_3 + c_4$$

.

Then,

$$(c_2 + c_3 + c_4)\mu + c_2\alpha_1 + c_3\alpha_2 + c_4\alpha_3$$

is estimable for any $(c_2\ c_3\ c_4)$ and these are the only estimable functions of $\mu, \alpha_1, \alpha_2, \alpha_3$.

Some estimable functions are

$$\mu + \frac{1}{3}(\alpha_1 + \alpha_2 + \alpha_3)\ \ (c_2 = c_3 = c_4 = \frac{1}{3})$$

and

$$\mu + \alpha_2 \quad (c_2 = 1\ c_3 = c_4 = 0)$$

but

$$\mu + 2\alpha_2$$

is not estimable.

Defn 3.7: For a linear model with

$$E(\mathbf{Y}) = X\beta \text{ and } Var(\mathbf{Y}) = \Sigma,$$

where X is an $n \times k$ matrix, $C_{r\times k}\beta_{k\times 1}$ is said to be estimable if all of its elements

$$C\beta = \begin{bmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \\ \vdots \\ \mathbf{c}_r^T \end{bmatrix} \beta = \begin{bmatrix} \mathbf{c}_1^T\beta \\ \mathbf{c}_2^T\beta \\ \vdots \\ \mathbf{c}_r^T\beta \end{bmatrix}$$

are estimable.

Result 3.9 For the linear model with $E(\mathbf{Y}) = X\beta$ and $Var(\mathbf{Y}) = \Sigma$, where X is an $n \times k$ matrix, each of the following conditions hold if and only if $C\beta$ is estimable.

(i) $AX = C$ for some matrix A, i.e., each row of C is in the space spanned by the rows of X.

(ii) $C\mathbf{d} = \mathbf{0}$ for any $\mathbf{d}$ for which $X\mathbf{d} = \mathbf{0}$.

(iii) $C\mathbf{b}$ is the same for any solution to the normal equations $(X^T X)\mathbf{b} = X^T\mathbf{y}$.

Summary

For a linear model

$$\mathbf{Y} = X\beta + \epsilon$$

with $E(\mathbf{Y}) = X\beta$ and $Var(\mathbf{Y}) = \Sigma$, we have

- Any estimable function has a unique interpretation
- The OLS estimator for an estimable function $C\beta$ is unique

$$C\mathbf{b} = C\,(X^T X)^{-} X^T \mathbf{Y}$$

- The OLS estimator for an estimable function $C\beta$ is
  - a linear estimator
  - an unbiased estimator

Conjecture on BEST

In the class of linear unbiased estimators for $\mathbf{c}^T\beta$, is the OLS estimator the best?

Here best means smallest expected squared error. Let $t(\mathbf{Y})$ denote an estimator for $\mathbf{c}^T\beta$. Then, the expected squared error is

$$
\begin{aligned}
MSE &= E[t(\mathbf{Y}) - \mathbf{c}^T\beta]^2 \\
&= E[t(\mathbf{Y}) - E(t(\mathbf{Y})) + E(t(\mathbf{Y})) - \mathbf{c}^T\beta]^2 \\
&= E[t(\mathbf{Y}) - E(t(\mathbf{Y}))]^2 + [E(t(\mathbf{Y})) - \mathbf{c}^T\beta]^2 \\
&\qquad + 2[E(t(\mathbf{Y})) - \mathbf{c}^T\beta]E[t(\mathbf{Y}) - E(t(\mathbf{Y}))] \\
&= E[t(\mathbf{Y}) - E(t(\mathbf{Y}))]^2 + [E(t(\mathbf{Y})) - \mathbf{c}^T\beta]^2 \\
&= Var(t(\mathbf{Y})) + [\text{bias}]^2
\end{aligned}
$$

If we restrict our attention to linear unbiased estimators for $\mathbf{c}^T\beta$:

- $E(t(\mathbf{Y})) = \mathbf{c}^T\beta$
- $t(\mathbf{Y}) = \mathbf{a}^T\mathbf{Y}$ for some $\mathbf{a}$

then, $t(\mathbf{Y}) = \mathbf{a}^T\mathbf{Y}$ is the **best linear unbiased estimator (blue)** for $\mathbf{c}^T\beta$ if

$$Var(\mathbf{a}^T\mathbf{Y}) \leq Var(\mathbf{d}^T\mathbf{Y})$$

for any $\mathbf{d}$ and any value of β.

Result 3.10 **Gauss-Markov Theorem**

For the Gauss-Markov model,

$$E(\mathbf{Y}) = X\beta \text{ and } Var(\mathbf{Y}) = \sigma^2 I,$$

the OLS estimator of an estimable function $\mathbf{c}^T\beta$ is the unique best linear unbiased estimator (blue) of $\mathbf{c}^T\beta$.

Proof:

(i) For any solution $\mathbf{b} = (X^TX)^{-}X^T\mathbf{Y}$ to the normal equations, the OLS estimator for $\mathbf{c}^T\beta$ is

$$\mathbf{c}^T\mathbf{b} = \mathbf{c}^T(X^TX)^{-}X^T\mathbf{y}$$

which is a linear function of $\mathbf{Y}$.

(ii) From Result 3.8.(i), there exists a vector **a** such that $\mathbf{c}^T = \mathbf{a}^T X$. Then

$$
\begin{aligned}
E(\mathbf{c}^T\mathbf{b}) &= E(\mathbf{c}^T(X^TX)^- X^T\mathbf{Y}) \\
&= \mathbf{c}^T(X^TX)^- X^T E(\mathbf{Y}) \\
&= \mathbf{c}^T(X^TX)^- X^T X\beta \\
&= \mathbf{a}^T X(X^TX)^- X^T X\beta \\
&= \mathbf{a}^T X\beta \\
&= \mathbf{c}^T\beta
\end{aligned}
$$

Hence, $\mathbf{c}^T\mathbf{b}$ is an unbiased estimator.

(iii) Minimum variance in the class of linear unbiased estimators

Suppose $\mathbf{d}^T\mathbf{Y}$ is any other linear unbiased estimator for $\mathbf{c}^T\beta$. Then

$$E(\mathbf{d}^T\mathbf{Y}) = \mathbf{d}^T E(\mathbf{Y}) = \mathbf{d}^T X\beta = \mathbf{c}^T\beta$$

for every β. Hence, $\mathbf{d}^T X = \mathbf{c}^T$ and $\mathbf{c} = X^T\mathbf{d}$.

We must show that

$$Var(\mathbf{c}^T\mathbf{b}) \leq Var(\mathbf{d}^T\mathbf{Y}).$$

First, note that

$$\begin{aligned} Var(\mathbf{d}^T\mathbf{Y}) &= Var(\mathbf{c}^T\mathbf{b} + [\mathbf{d}^T\mathbf{Y} - \mathbf{c}^T\mathbf{b}]) \\ &= Var(\mathbf{c}^T\mathbf{b}) + Var(\mathbf{d}^T\mathbf{Y} - \mathbf{c}^T\mathbf{b}) \\ &\quad +2Cov(\mathbf{c}^T\mathbf{b}, \mathbf{d}^T\mathbf{Y} - \mathbf{c}^T\mathbf{b}) \end{aligned}$$

Then

$$
\begin{aligned}
Var(\mathbf{d}^T\mathbf{Y}) &\geq Var(\mathbf{c}^T\mathbf{b}) \\
&\quad +2Cov(\mathbf{c}^T\mathbf{b}, \mathbf{d}^T Y - \mathbf{c}^T\mathbf{b}) \\
&= Var(\mathbf{c}^T\mathbf{b})
\end{aligned}
$$

because

$$
Cov(\mathbf{c}^T\mathbf{b}, \mathbf{d}^T Y - \mathbf{c}^T\mathbf{b}) = 0.
$$

To show this first note that

$$
\mathbf{c}^T\mathbf{b} = \mathbf{c}^T(X^TX)^- X^T\mathbf{Y}
$$

is invariant with respect to the choice of $(X^TX)^-$. Consequently, we can use the Moore-Penrose generalized inverse which is symmtric. (Not every generalized inverse of X^TX is symmetric.)

Then,

$$Cov(\mathbf{c}^T\mathbf{b}, \mathbf{d}^T\mathbf{Y} - \mathbf{c}^T\mathbf{b})$$

$$= Cov(\mathbf{c}^T(X^TX)^-\mathbf{X}^T\mathbf{Y}, [\mathbf{d}^T - \mathbf{c}^T(X^TX)^-X^T]\mathbf{Y})$$

$$= (\mathbf{c}^T(X^TX)^-\mathbf{X}^T)\, Var(\mathbf{Y})[\mathbf{d}^T - \mathbf{c}^T(X^TX)^-X^T]^T$$

$$= [\mathbf{c}^T(X^TX)^-\mathbf{X}^T]\sigma^2 I[\mathbf{d} - \mathbf{X}\underline{(\mathbf{X}^TX)^-}\mathbf{c}]$$

$\uparrow$

This is where the symmetry of $(X^TX)^-$ is needed.

$$= \sigma^2[\mathbf{c}^T(X^TX)^-\underline{X^T\mathbf{d}} - \mathbf{c}^T\underline{(X^TX)^-X^TX(X^TX)^-}\mathbf{c}]$$

$\uparrow$

since $X^T\mathbf{d} = \mathbf{c}$

Since $\mathbf{c}^T\mathbf{b}$ is invariant to the choice of $\mathbf{b}$ (result 3.8.(iii)), we were able to use the Moore-Penrose inverse for $(X^TX)^-$ which satisfies

$$(X^TX)^-(X^TX)(X^TX)^- = X^TX$$

by definition. Then,

$$\begin{aligned}
&Cov(\mathbf{c}^T\mathbf{b}, \mathbf{d}^T\mathbf{Y} - \mathbf{c}^T b) \\
=&\sigma^2[\mathbf{c}^T(X^TX)^-\mathbf{c} - \mathbf{c}^T(X^TX)^-\mathbf{c}] \\
=&0
\end{aligned}$$

Consequently,

$$Var(\mathbf{d}^T\mathbf{Y}) \geq Var(\mathbf{c}^T\mathbf{b})$$

and $\mathbf{c}^T\mathbf{b}$ is **blue**.

(iv) To show that the OLS estimator is the unique blue, note that

$$\begin{aligned} Var(\mathbf{d}^T\mathbf{Y}) &= Var(\mathbf{c}^T\mathbf{b} + [\mathbf{d}^T\mathbf{Y} - \mathbf{c}^T\mathbf{b}]) \\ &= Var(\mathbf{c}^T\mathbf{b}) + Var(\mathbf{d}^T\mathbf{Y} - \mathbf{c}^T\mathbf{b}) \end{aligned}$$

because $Cov(\mathbf{c}^T\mathbf{b}, \mathbf{d}^T\mathbf{Y} - \mathbf{c}^T\mathbf{b}) = 0$. Then, $\mathbf{d}^T\mathbf{Y}$ is blue if and only if

$$Var(\mathbf{d}^T\mathbf{Y} - \mathbf{c}^T\mathbf{b}) = 0 \ .$$

This is equivalent to $\mathbf{d}^T\mathbf{Y} - \mathbf{c}^T\mathbf{b} =$ constant.
Since both estimators are unbiased

$$E(\mathbf{d}^T\mathbf{Y} - \mathbf{c}^T\mathbf{b}) = E(\mathbf{d}^T\mathbf{Y}) - E(\mathbf{c}^T\mathbf{b}) = 0.$$

Consequently, $\mathbf{d}^T\mathbf{Y} - \mathbf{c}^T\mathbf{b} = 0$ for all $\mathbf{Y}$ and $\mathbf{c}^T\mathbf{b}$ is the unique blue.

What if you have a linear model that is <u>not</u> a Gauss-Markov model?

$$E(\mathbf{Y}) = X\beta, \quad Var(\mathbf{Y}) = \Sigma \neq \sigma^2 I$$

- Parts (i) and (ii) of the proof of result 3.11 do not require

  $$Var(\mathbf{Y}) = \sigma^2 I \ .$$

  Consequently, the OLS estimator for $\mathbf{c}^T\beta$,

  $$\mathbf{c}^T\mathbf{b} = \mathbf{c}^T(X^TX)^-X^T\mathbf{Y}$$

  is a linear unbiased estimator.

- Result 3.8 does not require

$$Var(\mathbf{Y}) = \sigma^2 I$$

and the OLS estimator for any estimable quantity,

$$\mathbf{c}^T\mathbf{b} = \mathbf{c}^T(X^TX)^{-}X^T\mathbf{Y},$$

is invariant to the choice of $(X^TX)^{-}$.

- The OLS estimator $\mathbf{c}^T\mathbf{b}$ may not be blue. There may be other linear unbiased estimators with smaller variance.

Variance of the OLS estimator of an estimable quantity:

$$\begin{aligned} Var(\mathbf{c}^T\mathbf{b}) =& Var(\mathbf{c}^T(X^TX)^-X^T\mathbf{Y}) \\ =& \mathbf{c}^T(X^TX)^-X^T\Sigma X[(X^TX)^-]^T\mathbf{c} \end{aligned}$$

For the Gauss-Markov model

$$Var(\mathbf{Y}) = \Sigma = \sigma^2 I$$

and

$$\begin{aligned} Var(\mathbf{c}^T\mathbf{b}) &= \sigma^2\mathbf{c}^T(X^TX)^-X^TX[(X^TX)^-]^T\mathbf{c} \\ &= \sigma^2\mathbf{c}^T(X^TX)^-\mathbf{c} \end{aligned}$$

Generalized Least Squares (GLS) Estimation

Defn 3.8: For a linear model with

$$E(\mathbf{Y}) = X\beta \text{ and } Var(\mathbf{Y}) = \Sigma,$$

where Σ is positive definite, a generalized least squares estimator for β minimizes

$$(\mathbf{Y} - X\mathbf{b}_{\text{GLS}})^T \Sigma^{-1} (\mathbf{Y} - X\mathbf{b}_{\text{GLS}})$$

Strategy: Transform **Y** to a random vector **Z** for which the Gauss-Markov model applies.

The spectral decomposition of Σ yields

$$\Sigma = \sum_{j=1}^{n} \lambda_j \mathbf{u}_j \mathbf{u}_j^T \ .$$

Define

$$\Sigma^{-1/2} = \sum_{j=1}^{n} \frac{1}{\sqrt{\lambda_j}} \mathbf{u}_j \mathbf{u}_j^T$$

and create the random vector

$$\mathbf{Z} = \Sigma^{-1/2} \mathbf{Y}.$$

Then

$$Var(\mathbf{Z}) = Var(\Sigma^{-1/2}\mathbf{Y}) = \Sigma^{-1/2}\Sigma\,\Sigma^{-1/2} = I$$

and

$$E(\mathbf{Z}) = E(\Sigma^{-1/2}\mathbf{Y}) = \Sigma^{-1/2}E(\mathbf{Y}) = \Sigma^{-1/2}X\beta = W\beta$$

and we have a Gauss-Markov model for **Z**, where $W = \Sigma^{-1/2}X$ is the model matrix.

Note that

$$
\begin{aligned}
(\mathbf{Z} - W\mathbf{b})^T(\mathbf{Z} - W\mathbf{b}) =& (\Sigma^{-1/2}\mathbf{Y} - \Sigma^{1/2}X\mathbf{b})^T(\Sigma^{-1/2}\mathbf{Y}\Sigma^{-1/2}X\mathbf{b}) \\
=& (Y - X\mathbf{b})^T\Sigma^{-1/2}\Sigma^{-1/2}(Y - X\mathbf{b}) \\
=& (Y - X\mathbf{b})^T\Sigma^{-1}(Y - X\mathbf{b})
\end{aligned}
$$

Hence, any GLS estimator for the **Y** model is an OLS estimator for the **Z** model.

It must be a solution to the normal equations for the $\mathbf{Z}$ model

$$\begin{aligned}
& W^T W\mathbf{b} = W^T\mathbf{Z} \\
& \quad \Leftrightarrow (X^T\Sigma^{-1/2}\Sigma^{-1/2}X)\mathbf{b} = X^T\Sigma^{-1/2}\Sigma^{-1/2}\mathbf{Y} \\
& \quad \Leftrightarrow (X^T\Sigma^{-1}X)\mathbf{b} = X^T\Sigma^{-1}\mathbf{Y}
\end{aligned}$$

These are the generalized least squares estimating equations.

Any solution

$$\begin{aligned}
\mathbf{b}_{\mathrm{GLS}} &= (W^T W)^- W^T\mathbf{Z} \\
&= (X^T\Sigma^{-1}X)^- X^T\Sigma^{-1}\mathbf{Y}
\end{aligned}$$

is called a generalized least squares (GLS) estimator for β.

Result 3.11 For a linear model with $E(\mathbf{Y}) = X\beta$ and $Var(\mathbf{Y}) = \Sigma$, the GLS estimator of an estimable function $\mathbf{c}^T\beta$,

$$\mathbf{c}^T\mathbf{b}_{\text{GLS}} = \mathbf{c}^T(X^T\Sigma^{-1}X)^{-}X^T\Sigma^{-1}\mathbf{Y}\ ,$$

is the unique blue of $\mathbf{c}^T\beta$.

Proof: Since $\mathbf{c}^T\beta$ is estimable, there is an $\mathbf{a}$ such that

$$\begin{aligned}
\mathbf{c}^T\beta &= E(\mathbf{a}^T\mathbf{Y}) \\
&= E(\mathbf{a}^T\Sigma^{1/2}\Sigma^{-1/2}\mathbf{Y}) \\
&= E(\mathbf{a}^T\Sigma^{1/2}\mathbf{Z})
\end{aligned}$$

Consequently, $\mathbf{c}^T\beta$ is estimable for the $\mathbf{Z}$ model. Apply the Gauss-Markov theorem (result 3.10) to the Z model.

Comments

- For the linear model with

$$E(\mathbf{Y}) = X\beta \text{ and } Var(\mathbf{Y}) = \Sigma,$$

both the OLS and GLS estimators for an estimable function $\mathbf{c}^T\beta$ are linear unbiased estimators.

$$Var(\mathbf{c}^T\mathbf{b}_{\mathrm{OLS}}) = \mathbf{c}^T(X^TX)^-X^T\Sigma X[(X^TX)^-]^T\mathbf{c}$$

$$Var(\mathbf{c}^T\mathbf{b}_{\mathrm{GLS}}) = \mathbf{c}^T(X^T\Sigma^{-1}X)^-X^T\Sigma^{-1}X(X^T\Sigma^{-1}X)^-\mathbf{c}$$

and

$$Var(\mathbf{c}^T\mathbf{b}_{\mathrm{OLS}}) \geq Var(\mathbf{c}^T\mathbf{b}_{\mathrm{GLS}})$$

- For the Gauss-Markov model,

$$\mathbf{c}^T b_{\mathrm{GLS}} = \mathbf{c}^T b_{\mathrm{OLS}} \ .$$

- The blue property of $\mathbf{c}^T\mathbf{b}_{\mathrm{GLS}}$ assumes that $Var(\mathbf{Y}) = \Sigma$ is known.
- The same results, including Results 3.12, hold for the Aitken model where $E(\mathbf{Y}) = X\beta$ and $Var(\mathbf{Y}) = \sigma^2 V$ for some known matrix V.

- In practice $Var(\mathbf{Y}) = \Sigma$ is usually unknown. An approximation to

$$\mathbf{b}_{\mathrm{GLS}} = (X^T\Sigma^{-1}X)^{-}X^T\Sigma^{-1}\mathbf{Y}$$

is obtained by substituting a consistent estimator $\hat{\Sigma}$ for Σ.

  - use method of moments or maximum likelihood estimation to obtain $\hat{\Sigma}$
  - the resulting estimator
    - is not a linear estimator
    - is consistent but not necessarily unbiased
    - does not provide a blue for estimable functions
    - may have larger mean squared error than the OLS estimator

Reparameterization, Restrictions, and Avoiding Generalized Inverses

Models that may appear to be different at first sight, may be equivalent in many ways.

Example 3.3 Two-way classification

Consider the **cell mean model**.

$$Y_{ijk} = \mu_{ij} + \epsilon_{ijk} \qquad \begin{aligned} i &= 1,2 \\ j &= 1,2 \\ k &= 1,2 \end{aligned}$$

where $\epsilon_{ijk} \sim NID(0, \sigma^2)$

Matrix notation:

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \end{bmatrix}$$

or

$$\mathbf{Y} = W\gamma + \boldsymbol{\epsilon}$$

The effects model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \quad i = 1,2, \quad j = 1,2, \quad k = 1,2,$$

where $\epsilon_{ijk} \sim NID(0, \sigma^2)$.

Matrix notation:

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{21} \\ \gamma_{22} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \end{bmatrix}$$

or

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

The models are equivalent:

> the space spanned by the columns of W is the same as the spa
> by columns of X.

You can find matrices F and G such that

$$W = X \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = XF$$

and

$$
\begin{aligned}
X &= W \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \\
&= WG
\end{aligned}
$$

Then,

(i) rank(X) = rank (W)

(ii) Estimated mean responses are the same:

$$\hat{\mathbf{Y}} = X(X^TX)^-X^T\mathbf{Y} = W(W^TW)^{-1}W^T\mathbf{Y}$$

or

$$\hat{\mathbf{Y}} = P_X\mathbf{Y} = P_W\mathbf{Y}$$

(iii) Residual vectors are the same

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} \quad = \quad (I - P_X)\mathbf{Y} = (I - P_W)\mathbf{Y}$$

Example 3.1 Regression model for the yield of a chemical process.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

$$\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ \text{yield} & \text{temperature} & \text{time} \end{array}$$

An equivalent model is

$$\begin{aligned} Y_i &= \alpha_0 + \beta_1 (X_{1i} - \bar{X}_{1.}) + \\ &\quad \beta_2 (X_{2i} - \bar{X}_{2.}) + \epsilon_i \end{aligned}$$

For the first model:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ 1 & X_{13} & X_{23} \\ 1 & X_{14} & X_{24} \\ 1 & X_{15} & X_{25} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

For the second model:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix} = \begin{bmatrix} 1 & X_{11} - \bar{X}_1 & X_{21} - \bar{X}_2 \\ 1 & X_{12} - \bar{X}_1 & X_{22} - \bar{X}_2 \\ 1 & X_{13} - \bar{X}_1 & X_{23} - \bar{X}_2 \\ 1 & X_{14} - \bar{X}_1 & X_{24} - \bar{X}_2 \\ 1 & X_{15} - \bar{X}_1 & X_{25} - \bar{X}_2 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix} = W\boldsymbol{\gamma} + \boldsymbol{\epsilon}$$

The space spanned by the columns of X is the same as the space spanned by the columns of W.

$$X = W \begin{bmatrix} 1 & \bar{X}_1 & \bar{X}_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = WG$$

and

$$W = X \begin{bmatrix} 1 & -\bar{X}_1 & -\bar{X}_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = XF$$

and

$$\begin{aligned} \hat{\mathbf{Y}} &= P_X \mathbf{Y} = P_W \mathbf{Y} \\ \mathbf{e} &= \mathbf{Y} - \hat{\mathbf{Y}} = (I - P_X)\mathbf{Y} = (I - P_W)\mathbf{Y} \end{aligned}$$

Defn 3.9: Consider two linear models:

(1) $E(\mathbf{Y}) = X\beta$ and $Var(\mathbf{Y}) = \Sigma$

(2) $E(\mathbf{Y}) = W\gamma$ and $Var(\mathbf{Y}) = \Sigma$

where X is an $n \times k$ model matrix and W is an $n \times q$ model matrix.

We say that one model is a reparameterization of the other if there is a $k \times q$ matrix F and a $q \times k$ matrix G such that

$$W = XF \text{ and } X = WG\,.$$

The previous examples illustrate that if one model is a reparameterization of the other, then

(i) rank(X) = rank(W)

(ii) Least squares estimates of the response means are the same, i.e.,

$$\hat{\mathbf{Y}} = P_X \mathbf{Y} = P_W \mathbf{Y}$$

(iii) Residuals are the same, i.e.,

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = (I - P_X)\mathbf{Y} = (I - P_w)\mathbf{Y}$$

(iv) An unbiased estimator for σ^2 is provided by

$$MSE = SSE/(n - rank(X))$$

where,

$$\begin{aligned} SSE = \mathbf{e}^T\mathbf{e} &= \mathbf{Y}^T(I - P_X)\mathbf{Y} \\ &= \mathbf{Y}^T(I - P_W)\mathbf{Y} \end{aligned}$$

Reasons for reparameterizing models

(i) Reduce the number of parameters
- Obtain a full rank model
- Avoid use of generalized inverses

(ii) Make computations easier
- In the previous examples, $W^T W$ is a diagonal matrix and $(W^T W)^{-1}$ is easy to compute.

(iii) More meaningfull interpretation of parameters.

Result 3.12. Suppose two linear models,

$$(1)\quad E(\mathbf{Y}) = X\beta \quad Var(\mathbf{Y}) = \Sigma$$

and

$$(2)\quad E(\mathbf{Y}) = W\gamma \quad Var(\mathbf{Y}) = \Sigma$$

are reparameterizations of each other, and let F be a matrix such that $W = XF$. Then

(i) If $\mathbf{C}^T\beta$ is estimable for the first model, then $\beta = F\gamma$ and $\mathbf{C}^T F\gamma$ is estimable under Model 2.

(ii) Let $\hat{\beta} = (X^T X)^- X^T \mathbf{Y}$ and $\hat{\gamma} = (W^T W)^- W^T \mathbf{Y}$. If $\mathbf{C}^T \beta$ is estimable, then

$$\mathbf{C}^T \hat{\beta} = \mathbf{C}^T F \hat{\gamma}$$

(iii) If $H_0 : \mathbf{C}^T \beta = \mathbf{d}$ is testable under one model, then $H_0 : \mathbf{C}^T F \gamma = \mathbf{d}$ is testable under the other.

Proof:

(i) If $\mathbf{C}^T\beta$ is estimable for the frist model, then (by Result 3.9 (i))

$$\mathbf{C}^T = \mathbf{a}^T X \text{ for some } \mathbf{a}\,.$$

Hence,

$$\mathbf{C}^T\mathbf{F} = \mathbf{a}^T X\mathbf{F} = \mathbf{a}^T\mathbf{W}$$

which implies that $\mathbf{C}^T F\gamma$ is estimable for the second model.

(ii) Since $\mathbf{C}^T\beta$ is estimable, the unique b.l.u.e. is

$$\begin{aligned}\mathbf{C}^T\hat{\beta} &= \mathbf{C}^T(X^TX)^-X^T\mathbf{Y}\\ &= \mathbf{a}^T X^T(X^TX)^-X^T\mathbf{Y}\\ &= \mathbf{a}^T P_X\mathbf{Y} \text{ for some } \mathbf{a}\end{aligned}$$

Since $\mathbf{C}^T F\gamma$ is also estimable, the unique b.l.u.e. for $\mathbf{C}^T F\gamma$ is

$$\begin{aligned}
\mathbf{C}^T F(W^T W)^- W^T \mathbf{Y} =& \mathbf{a}^T XF(W^T W)^- W^T \mathbf{Y} \\
=& \mathbf{a}^T W(W^T W)^- W^T \mathbf{Y} \\
=& \mathbf{a}^T P_W \mathbf{Y}
\end{aligned}$$

for the same $\mathbf{a}$.

Hence, the estimators are the same if $P_X = P_W$. To show this, note that

$$P_X W = P_X X F = XF = W$$

which implies

$$\begin{aligned} P_X P_W &= P_X W (W^T W)^- W^T \\ &= W (W^T W)^- W^T \\ &= P_W \end{aligned}$$

By a similar argument

$$P_W P_X = P_X$$

Then,

$$\begin{aligned}
P_W &= P + W^T \\
&= (P_X P_X)^T \\
&= P_W^T P_X^T \\
&= P_W P_X \\
&= P_X
\end{aligned}$$

Example 3.2 An effects model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

This model can be expressed as

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}$$

Reparameterize the model as

$$Y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \epsilon_{ij}$$

using "othogonal" polynomial contrasts (for factors with equally spaced levels and balanced designs)

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 1 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 1 & 0 & \frac{-2}{\sqrt{6}} \\ 1 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 1 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 1 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}$$

The unique OLS estimator for $\boldsymbol{\beta} = (\beta_0\ \beta_1\ \beta_2)^T$ is

$$
\begin{aligned}
\mathbf{b} &= (X^T X)^{-1} X^T \mathbf{Y} \\
&= \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 67.000 \\ 5.6568 \\ -4.8989 \end{bmatrix}
\end{aligned}
$$

Note that

$$
\begin{aligned}
\hat{\beta}_0 + \hat{\beta}_1(\frac{-1}{\sqrt{2}}) + \hat{\beta}_2(\frac{1}{\sqrt{6}}) &= 61 = \bar{Y}_{1.} \\
\hat{\beta}_0 + \hat{\beta}_1(0) + \hat{\beta}_2(\frac{-2}{\sqrt{6}}) &= 71 = \bar{Y}_{2.} \\
\hat{\beta}_0 + \hat{\beta}_1(\frac{1}{\sqrt{2}}) + \hat{\beta}_2(\frac{1}{\sqrt{6}}) &= 69 = \bar{Y}_{3.}
\end{aligned}
$$

Reparameterize the model using Helmert contrasts:

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 0 & 2 \\ 1 & 0 & 2 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}$$

Write this model as $\mathbf{Y} = X\beta + \epsilon$ where

$$X = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 0 & 2 \\ 1 & 0 & 2 \\ 1 & 0 & 2 \end{bmatrix} \text{ and } \beta = \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix}$$

Then,

$$X^T X = \begin{bmatrix} n_. & n_2 - n_1 & 2n_3 - n_1 - n_2 \\ n_2 - n_1 & n_1 + n_2 & n_1 - n_2 \\ 2n_3 - n_1 - n_2 & n_1 - n_2 & n_1 + n_2 + 4n_3 \end{bmatrix}$$

and

$$X^T\mathbf{Y} = \begin{bmatrix} Y_{..} \\ Y_{2.} - Y_{1.} \\ 2Y_{3.} - Y_{1.} - Y_{2.} \end{bmatrix}$$

The unique OLS estimator for $\beta = (\gamma_0\ \gamma_1\ \gamma_2)^T$ is

$$\mathbf{b} = (X^TX)^{-1}X^T\mathbf{Y}$$

$$= \begin{bmatrix} \bar{Y}_{..} \\ \frac{1}{2}(\bar{Y}_{2.} - \bar{Y}_{1.}) \\ \frac{1}{3}(\bar{Y}_{3.} - \frac{(\bar{Y}_{1.}+\bar{Y}_{2.})}{2}) \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix} = \begin{bmatrix} 67 \\ 5 \\ 1 \end{bmatrix}$$

Note that

$$\begin{aligned}
\hat{\gamma}_0 + \hat{\gamma}_1(-1) + \hat{\gamma}_2(-1) &= 61 = \bar{Y}_{1.} \\
\hat{\gamma}_0 + \hat{\gamma}_1(1) + \hat{\gamma}_2(-1) &= 71 = \bar{Y}_{2.} \\
\hat{\gamma}_0 + \hat{\gamma}_1(0) + \hat{\gamma}_2(2) &= 69 = \bar{Y}_{3.}
\end{aligned}$$

Restrictions (side conditions)

- Give meaning to individual parameters
- Make individual parameters estimable
- Create a full rank model matrix
- Avoid the use of generalized inverses

Example 3.2 An effects model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

This model can be expressed as

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}$$

Impose the restriction $\alpha_3 = 0$ Then,

$$\begin{aligned} E(Y_{1j}) &= \mu + \alpha_1 \qquad \text{for } j = 1, ..., n_1 \\ E(Y_{2j}) &= \mu + \alpha_2 \qquad \text{for } j = 1, ..., n_2 \\ E(Y_{3j}) &= \mu \qquad\qquad\quad \text{for } j = 1, ..., n_3 \end{aligned}$$

and

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}$$

Write this model as $\mathbf{Y} = X\beta + \epsilon$, where

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \text{ and } \beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix}$$

Then,

$$X^T X = \begin{bmatrix} n_. & n_1 & n_2 \\ n_1 & n_1 & 0 \\ n_2 & 0 & n_2 \end{bmatrix} \text{ and } X^T \mathbf{Y} = \begin{bmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \end{bmatrix}$$

and the unique OLS estimator for $\boldsymbol{\beta} = (\mu\ \alpha_1\ \alpha_2)^T$ is

$$
\begin{aligned}
\mathbf{b} &= (X^T X)^{-1} X^T \mathbf{Y} \\
&= \frac{1}{n_3} \begin{bmatrix} 1 & -1 & -1 \\ -1 & \frac{n_1+n_3}{n_1} & 1 \\ -1 & 1 & \frac{n_2+n_3}{n_2} \end{bmatrix} \begin{bmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \end{bmatrix} \\
&= \begin{bmatrix} \bar{Y}_{3.} \\ \bar{Y}_{1.} - \bar{Y}_{3.} \\ \bar{Y}_{2.} - \bar{Y}_{3.} \end{bmatrix} = \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{bmatrix}
\end{aligned}
$$

Consider the model $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ with the restriction $\alpha_1 + \alpha_2 + \alpha_3 = 0$. Then, $\alpha_3 = -\alpha_1 - \alpha_2$ and

$$\begin{aligned} E(Y_{1j}) &= \mu + \alpha_1 \text{ for } j = 1, ..., n_1 \\ E(Y_{2j}) &= \mu + \alpha_2 \text{ for } j = 1, ..., n_2 \\ E(Y_{3j}) &= \mu + \alpha_3 = \mu - \alpha_1 - \alpha_2 \quad \text{for } j = 1, ..., n_3 \end{aligned}$$

and

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}$$

This model is $\mathbf{Y} = X\beta + \epsilon$ with

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \text{ and } \beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix}$$

The unique OLS estimator for $\boldsymbol{\beta} = (\mu \ \alpha_1 \ \alpha_2)^T$ is

$$
\begin{aligned}
\mathbf{b} &= (X^T X)^{-1} X^T \mathbf{Y} \\
&= \begin{bmatrix} n_{.} & n_1 - n_3 & n_2 - n_3 \\ n_1 - n_3 & n_1 + n_3 & n_3 \\ n_2 - n_3 & n_3 & n_2 + n_3 \end{bmatrix}^{-1} \begin{bmatrix} Y_{..} \\ Y_{1.} - Y_{3.} \\ Y_{2.} - Y_{3.} \end{bmatrix} \\
&= \begin{bmatrix} \bar{Y}_{..} \\ \bar{Y}_{1.} - \bar{Y}_{..} \\ \bar{Y}_{2.} - \bar{Y}_{..} \end{bmatrix} = \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{bmatrix}
\end{aligned}
$$

Consider the model $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ with the restriction $\alpha_1 = 0$. Then,

$$
\begin{aligned}
E(Y_{1j}) &= \mu & \text{for } j = 1, ..., n_1 \\
E(Y_{2j}) &= \mu + \alpha_2 & \text{for } j = 1, ..., n_2 \\
E(Y_{3j}) &= \mu + \alpha_3 & \text{for } j = 1, ..., n_3
\end{aligned}
$$

and

$$
\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}
$$

This model is $\mathbf{Y} = X\beta + \epsilon$, with

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \text{ and } \beta = \begin{bmatrix} \mu \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

The unique OLS estimator for

$\boldsymbol{\beta} = (\mu \ \alpha_1 \ \alpha_2)^T$ is

$$\mathbf{b} = (X^T X)^{-1} X^T \mathbf{Y}$$

$$= \begin{bmatrix} n_. & n_2 & n_3 \\ n_2 & n_2 & 0 \\ n_3 & 0 & n_3 \end{bmatrix}^{-1} \begin{bmatrix} Y_{..} \\ Y_{2.} \\ Y_{3.} \end{bmatrix}$$

$$= \begin{bmatrix} \bar{Y}_{1.} \\ \bar{Y}_{2.} - \bar{Y}_{1.} \\ \bar{Y}_{3.} - \bar{Y}_{1.} \end{bmatrix} = \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \end{bmatrix}$$

The restrictions (i.e. the choice of one particular solution to the normal equations) have no effect on the OLS estimates of estimable quantities. The estimated treatment means are:

$$
\begin{aligned}
E(\hat{Y}_{1j}) &= \hat{\mu} = \bar{Y}_{1.} = 61 \\
E(\hat{Y}_{2j}) &= \hat{\mu} + \hat{\alpha}_2 = \bar{Y}_{2.} = 71 \\
E(\hat{Y}_{3j}) &= \hat{\mu} + \hat{\alpha}_3 = \bar{Y}_{3.} = 69
\end{aligned}
$$