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.

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

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

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

The "effects" model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

where

$$\epsilon_{ijk} \sim NID(0, \sigma^2) \qquad \begin{array}{l} i = 1,2 \\ j = 1,2 \\ k = 1,2 \end{array}$$

Matrix notation:

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

or

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

The models are "equivalent":

the space spanned by the columns of W is the same as the space spanned 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:

$$\begin{aligned} \hat{\mathrm{Y}} &= X(X^T X)^- X^T \mathrm{Y} \\ &= W(W^T W)^{-1} W^T \mathrm{Y} \end{aligned}$$

or

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

(iii) Residual vectors are the same

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

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

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

↑ yield ↑ temperature ↑ time

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:

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

$$= X\beta + \epsilon$$

For the second model:

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

$$= W\gamma + \epsilon$$

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

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

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

<u>Defn 3.9:</u> Consider two linear models:

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

and

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

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 <u>reparameterization</u> 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., $\hat{\mathrm{Y}} = P_X\mathrm{Y} = P_W\mathrm{Y}$

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

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

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

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

where,

$$SSE = \mathrm{e}^T\mathrm{e} \;=\; \mathrm{Y}^T(I - P_X)\mathrm{Y}$$
$$= \mathrm{Y}^T(I - P_W)\mathrm{Y}$$

<u>Reasons for reparameterizing models:</u>

(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^TW is a diagonal matrix and $(W^TW)^{-1}$ is easy to compute.

(iii) More meaningfull interpretation of parameters.

<u>Result 3.12.</u> Suppose two linear models,

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

and

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

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

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

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

$$\mathrm{C}^T\hat{\beta} = \mathrm{C}^TF\hat{\gamma}$$

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

Proof:

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

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

Hence,

$$C^T F = a^T X F = a^T W$$

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

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

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

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

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

for the same a.

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

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

which implies

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

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 $\beta = (\beta_0\ \beta_1\ \beta_2)^T$ is

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

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 $\mathrm{Y} = X\beta + \epsilon$ where

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

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

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

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

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

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

Note that

$$\hat{\gamma}_0 + \hat{\gamma}_1(-1) + \hat{\gamma}_2(-1) = 61 = \bar{Y}_{1.}$$

$$\hat{\gamma}_0 + \hat{\gamma}_1(1) + \hat{\gamma}_2(-1) = 71 = \bar{Y}_{2.}$$

$$\hat{\gamma}_0 + \hat{\gamma}_1(0) + \hat{\gamma}_2(2) = 69 = \bar{Y}_{3.}$$

Restrictions (side conditions)

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

<u>Example 3.2</u> 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,

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

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 $\mathrm{Y} = X\beta + \epsilon$ where

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

Then,

$$X^T X = \begin{bmatrix} n_. & n_1 & n_2 \\ n_1 & n_1 & 0 \\ n_2 & 0 & n_2 \end{bmatrix}$$

and

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

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

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

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

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

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 $\mathrm{Y} = X\beta + \epsilon$ with

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

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

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

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

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

and

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

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

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

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

$$\begin{aligned} \mathrm{b} &= (X^TX)^{-1}X^T\mathrm{Y} \\ &= \begin{bmatrix} n. & n_2 & n_3 \\ n_2 & n_2 & 0 \\ n_3 & 0 & n_3 \end{bmatrix}^{-1} \begin{bmatrix} Y_{..} \\ Y_{2.} \\ Y_{3.} \end{bmatrix} \\ &= \begin{bmatrix} \bar{Y}_{1.} \\ \bar{Y}_{2.} - \bar{Y}_{1.} \\ \bar{Y}_{3.} - \bar{Y}_{1.} \end{bmatrix} = \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \end{bmatrix} \end{aligned}$$

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

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