The LS Estimate of $\boldsymbol{\beta}$

The LS estimate of the regression coefficient vector $\boldsymbol{\beta}$ is defined to be

$$\hat{\boldsymbol{\beta}}_{\mathsf{LS}} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$$

or equivalently

$$\hat{\boldsymbol{\beta}}_{\mathsf{LS}} = \{\mathbf{v} : \mathbf{X}\mathbf{v} = \hat{\mathbf{y}}\}.$$

If the design matrix is *not* of full rank, i.e., rank$(\mathbf{X}) < p$ (here assume $p < n$), then $\hat{\boldsymbol{\beta}}_{\mathsf{LS}}$ is *not* unique: any vector $\mathbf{v}$ satisfying $\mathbf{X}\mathbf{v} = \hat{\mathbf{y}}$ would be a valid LS estimate of $\boldsymbol{\beta}$.

If rank$(\mathbf{X}) = p$, then the LS estimate of $\boldsymbol{\beta}$ is unique:

$$\hat{\boldsymbol{\beta}}_{\mathsf{LS}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\hat{\mathbf{y}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}\mathbf{y}.$$

The Non-uniqueness of $\hat{\beta}_{\mathbf{LS}}$

Consider a one-way ANOVA model with $g = 2$ groups and $n_1 = 2, n_2 = 3$. The element-wise representation of the model is

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1:2,\ j = 1:n_i.$$

The matrix representation of the model is

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{23} \end{pmatrix}_{5\times 1} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}_{5\times 3} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix}_{3\times 1} .$$

Columns of $\mathbf{X}$ are linearly-dependent. Use $\tilde{\mathbf{X}}$ to denote a 5×2 matrix containing the last two columns of $\mathbf{X}$. It is easy to check that *1)* $C(\mathbf{X}) = C(\tilde{\mathbf{X}})$, and *2)* two columns in $\tilde{\mathbf{X}}$ are linearly dependent. So we can compute the projection matrix for this linear model using $\tilde{\mathbf{X}}$:

$$M = \tilde{\mathbf{X}}(\tilde{\mathbf{X}}^t\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^t = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

Using projection matrix M, we can compute $\hat{y}_{ij} = \bar{y}_{i\cdot}$, which is the average of samples in the i-th group.

Since there are just two groups, there are only two different values for $\hat{y}_{ij}$'s. To estimate μ, α_1, α_2, we need to solve:

$$\begin{aligned} \bar{y}_{1\cdot} &= \mu + \alpha_1, \\ \bar{y}_{2\cdot} &= \mu + \alpha_2. \end{aligned}$$

Apparently, the solution of $(\mu, \alpha_1, \alpha_2)$ is not unique (two equations but three unknowns). For example, any of the following would be a valid LS estimate of $(\mu, \alpha_1, \alpha_2)$:

$$\mu = a, \quad \alpha_1 = \bar{y}_{1\cdot} - a, \quad \alpha_2 = \bar{y}_{2\cdot} - a,$$

where a could be any constant.

Next we'll see that although $\boldsymbol{\beta}$ doesn't have a unique LS estimate, some linear-combinations of $\boldsymbol{\beta}$ do. For example, $\alpha_1 - \alpha_2$ for this one-way ANOVA example.

Estimable

Suppose we want to estimate a linear combination of $\boldsymbol{\beta}$, say $\lambda^t\boldsymbol{\beta}$, where λ is a known/given $p \times 1$ vector.

For example, estimating any element of $\boldsymbol{\beta}$ and estimating the mean response at a new value $\mathbf{x}_*$ (i.e., $\mathbf{x}_*^t\boldsymbol{\beta}$) are all special cases of this setup.

Naturally, we can form an estimate of $\lambda^t\boldsymbol{\beta}$ by plugging in the LS estimate, i.e., $\widehat{\lambda^t\boldsymbol{\beta}} = \lambda^t\hat{\boldsymbol{\beta}}_{\mathsf{LS}}$. But if the LS estimate is not unique, we might be wondering: is it possible to estimate $\lambda^t\boldsymbol{\beta}$? For example, for the one-way ANOVA example, is it possible to form a reasonable estimate of μ? The answer is NO.

So we first define what linear combinations are estimable.

A linear combination is estimable if there exists an $n \times 1$ vector $\mathbf{a}$ such that

$$\lambda = \mathbf{X}^t \mathbf{a}. \tag{1}$$

That is, λ is in the row span of $\mathbf{X}$, $\lambda \in C(\mathbf{X}^t)$.

Why (1) implies that $\lambda^t \boldsymbol{\beta}$ is estimable? If (1) holds, we can construct a linear and unbiased estimator of $\lambda^t \boldsymbol{\beta}$: $\mathbf{a}^t \mathbf{y}$, which is unbiased since

$$\mathbb{E}[\mathbf{a}^t \mathbf{y}] = \mathbf{a}^t \mathbb{E}\mathbf{y} = \mathbf{a}^t \mathbf{X}\boldsymbol{\beta} = (\mathbf{X}^t \mathbf{a})^t \boldsymbol{\beta} = \lambda^t \boldsymbol{\beta},$$

where the last equality is due to (1).

Note that such "$\mathbf{a}$" vectors may not be unique. This is because for a given λ, Eq (1) corresponds to p equations, but "$\mathbf{a}$" has n elements, so there must be many "$\mathbf{a}$" vectors satisfying (1) for any given λ.

Back to the one-way ANOVA example on P2. It is easy to check that $\theta = \alpha_1 - \alpha_2$ is estimable. Here are some choices of "$\mathbf{a}$" vectors, which all satisfy $\lambda = \mathbf{X}^t\mathbf{a}$:

$$\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \left[\begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1/2 \\ 1/2 \\ -1/3 \\ -1/3 \\ -1/3 \end{pmatrix} \right]$$