Lecture Notes for Theory of Linear Models

Lecture 19
(Ch 12 in the text)

Theory for Non-Full Rank Linear Models

Longhai Li
Department of Mathematics and Statistics
University of Saskatchewan

Estimation and F-test in Non-full-rank Models

The One-way Model:

Consider the balanced one-way layout model for y_{ij} a response on the j^{th} unit in the i^{th} treatment group. Suppose that there are a treatments and n units in the i^{th} treatment group. The **cell-means** model for this situation is

$$y_{ij} = \mu_i + e_{ij}, \quad i = 1, \ldots, a, j = 1, \ldots, n,$$

where the e_{ij}'s are i.i.d. $N(0, \sigma^2)$.

An alternative, but equivalent, linear model is the **effects model** for the one-way layout:

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, \ldots, a, j = 1, \ldots, n,$$

$$j_N = \sum_{i=1}^{a} x_i$$

with the same assumptions on the errors.

The cell means model can be written in vector notation as

$$\mathbf{y} = \mu_1 \mathbf{x}_1 + \mu_2 \mathbf{x}_2 + \cdots + \mu_a \mathbf{x}_a + \mathbf{e}, \quad \mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}),$$

and the effects model can be written as

$$\mathbf{y} = \mu \mathbf{j}_N + \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \cdots + \alpha_a \mathbf{x}_a + \mathbf{e}, \quad \mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}),$$

where $\mathbf{x}_i$ is an indicator for treatment i, and $N = an$ is the total sample size.

- That is, the effects model has the same model matrix as the cell-means model, but with one extra column, a column of ones, in the first position.

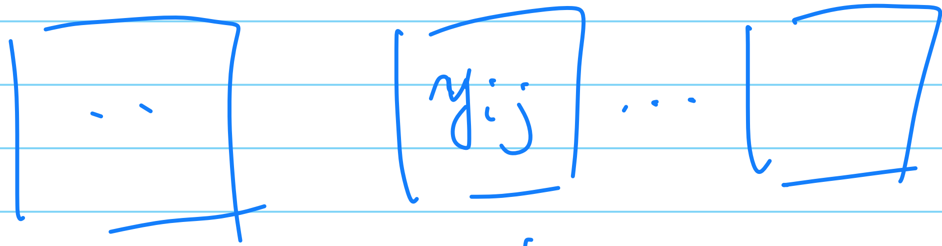
- Notice that $\sum_i \mathbf{x}_i = \mathbf{j}_N$. Therefore, the columns of the model matrix for the effects model are linearly dependent.

Let $\mathbf{X}_1$ denote the model matrix in the cell-means model, $\mathbf{X}_2 = (\mathbf{j}_N, \mathbf{X}_1)$ denote the model matrix in the effects model.

- Note that $C(\mathbf{X}_1) = C(\mathbf{X}_2)$.

In general, two linear models $\mathbf{y} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{e}_1$, $\mathbf{y} = \mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{e}_2$ with the same assumptions on $\mathbf{e}_1$ and $\mathbf{e}_2$ are equivalent linear models if $C(\mathbf{X}_1) = C(\mathbf{X}_2)$.

$$j_6 = x_1 + x_2 + x_3$$



ith

$u_1 \qquad u_2 \quad \cdots \quad u_a$

$$y_{ij} = u + \alpha_i + e_{ij}$$

$$y = \begin{array}{c} \begin{array}{cccc} j_6 & x_1 & x_2 & x_3 \end{array} \\ \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}}_{X} \end{array} \underbrace{\begin{bmatrix} u \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}}_{\beta} + e_{a=3}$$

x_i = indi of group i

$$C([j_6, x_1, x_2, x_3]) = C([x_1, x_2, x_3])$$

$$u_1 = u + \alpha_1$$

$$u_2 = u + \alpha_2$$

$$u_3 = u + \alpha_3$$

4

Illustration of Non-identification

$u_1 = 10$, $u_2 = 15$, $u_3 = 20$

	u	α_1	α_2	α_3
	100	-90	-85	-80
	0	10	15	20
	1M	-1M+10	-1M+15	-1M+20
✓	15	-5	0	5
✓	10	0	5	10
	⋮	⋮	⋮	⋮
	a	$10-a$	$15-a$	$20-a$
		β_1	β_2	β_3

However, **subject to the constraint** $\sum_i \alpha_i = 0$, the parameters of the effects model have the following interpretations:

> μ =grand mean response across all treatments
> α_i =deviation from the grand mean placing μ_i (the i^{th} treatment mean) up or down from the grand mean; i.e., the effect of the i^{th} treatment.

Without the constraint, though, μ is not constrained to fall in the center of the μ_i's. μ is in no sense the grand mean, it is just an arbitrary baseline value.

In addition, adding the constraint $\sum_i \alpha_i = 0$ has essentially the effect of reparameterizing from the overparameterized (non-full rank) effects model to a just-parameterized (full rank) model that is equivalent (in the sense of having the same model space) as the cell means model.

To see this consider the one-way effects model with $a = 3$, $n = 2$. Then $\sum_{i=1}^{a} \alpha_i = 0$ implies $\alpha_1 + \alpha_2 + \alpha_3 = 0$ or $\alpha_3 = -(\alpha_1 + \alpha_2)$. Subject to the constraint, the effects model is

$$\mathbf{y} = \mu \mathbf{j}_N + \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \alpha_3 \mathbf{x}_3 + \mathbf{e}, \quad \text{where } \alpha_3 = -(\alpha_1 + \alpha_2),$$

or

$$\begin{aligned} \mathbf{y} &= \mu \mathbf{j}_N + \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + (-\alpha_1 - \alpha_2)\mathbf{x}_3 + \mathbf{e} \\ &= \mu \mathbf{j}_N + \alpha_1(\mathbf{x}_1 - \mathbf{x}_3) + \alpha_2(\mathbf{x}_2 - \mathbf{x}_3) + \mathbf{e} \\ &= \mu \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \alpha_1 \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ -1 \\ -1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} + \mathbf{e}, \end{aligned}$$

which has the same model space as the cell-means model.

$\boxed{\mu + \alpha_1}$ $\boxed{\mu + \alpha_2}$ $\boxed{\mu + (-\alpha_1 - \alpha_2)}$

Cell 1 Cell 2 Cell 3

6

Another reparametrization

$$y = \underline{\mu_1} x_1 + \mu_2 x_2 + \mu_3 x_3 + e$$

$$= \underbrace{(\mu_2 - \mu_1)}_{=\beta_2} x_2 + \underbrace{(\mu_3 - \mu_1)}_{=\beta_3} x_3 + \underbrace{\mu_1}_{=\beta_0} (x_1 + x_2 + x_3) + e$$

$$= \underbrace{\beta_0}_{\text{mean of cell 1}} + \underbrace{\underline{\beta_2}}_{\mu_2 - \mu_1} x_2 + \underbrace{\underline{\beta_3}}_{\mu_3 - \mu_1} x_3 + e$$

$\boxed{\beta_0}$	$\boxed{\beta_0 + \beta_2}$	$\boxed{\beta_0 + \beta_3}$
cell 1	cell 2	cell 3

Baseline model

7

Thus, when faced with a non-full rank model like the one-way effects model, we have three ways to proceed:

(1) Reparameterize to a full rank model.

- E.g., switch from the effects model to the cell-means model.

(2) Add constraints to the model parameters to remove the overparameterization.

- E.g., add a constraint such as $\sum_{i=1}^{a} \alpha_i = 0$ to the one-way effects model.
- Such constraints are usually called **side-conditions**.
- Adding side conditions essentially accomplishes a reparameterization to a full rank model as in (1).

(3) Analyze the model as a non-full rank model, but limit estimation and inference to those functions of the (overparameterized) parameters that can be uniquely estimated.

- Such functions of the parameters are called **estimable**.
- It is only in this case that we are actually using an overparameterized model, for which some new theory is necessary. (In cases (1) and (2) we remove the overparameterization somehow.)

why over-parametrization?

- Beauty (symmetry) in math.
- what's consequences of using an over-para. model?

8

Least Square Estimation of β

Even if $\mathbf{X}$ is not of full rank, the least-squares criterion is still a reasonable one for estimation, and it still leads to the normal equation:

$$\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{y}. \qquad (\clubsuit)$$

Theorem: For $\mathbf{X}$ and $n \times p$ matrix of rank $k < p \leq n$, ($\clubsuit$) is a consistent system of equations.

$C(X'X) = C(X')$

So ($\clubsuit$) is consistent, and therefore has a *non-unique* (for $\mathbf{X}$ not of full rank) solution given

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y},$$

where $(\mathbf{X}^T\mathbf{X})^{-}$ is some (any) generalized inverse of $\mathbf{X}^T\mathbf{X}$.

GI in Least Square with rank < p.

$\underset{n\times p}{X} = \underset{n\times k}{Q}\,(\underset{k\times k}{R_1}, \underset{k\times(p-k)}{R_2})$, R_1^{-1} exists, $k < p$.

we assume the first k col. of X are LIN.

$$X'X = \begin{bmatrix} R_1' \\ R_2' \end{bmatrix} \cdot \underbrace{Q'Q}_{I}\,(R_1, R_2), \quad X'y = \begin{bmatrix} R_1' \\ R_2' \end{bmatrix} Q'y$$

note

$$X'X\beta = X'y \;\overset{I_k}{\Leftarrow}\; [X'y \in C(X') = C(X'X)]$$

$$\Leftrightarrow \begin{bmatrix} R_1'R_1 & R_1'R_2 \\ R_2'R_1 & R_2'R_2 \end{bmatrix} \beta = \begin{bmatrix} R_1' \\ R_2' \end{bmatrix} Q'y$$

$$\text{let } (X'X)^- = \begin{bmatrix} (R_1'R_1)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{(one version)}$$

$$\hat\beta = (X'X)^- X'y$$

$$= \begin{bmatrix} (R_1'R_1)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} R_1' \\ R_2' \end{bmatrix} Q'y$$

$$= \begin{bmatrix} \underset{k\times 1}{(R_1'R_1)^{-1} R_1' Q'y} \\ 0 \end{bmatrix} = \begin{bmatrix} R_1^{-1}Q'y \\ 0 \end{bmatrix}$$

$$\hat y = X\hat\beta = Q[R_1, R_2]\cdot\hat\beta = Q\cdot R_1(R_1'R_1)^{-1}R_1'Q'y$$
$$= Q\cdot Q'y$$

Thm:

$\hat{\beta} = (X'X)^{-}X'y$ is a solution to

$(X'X)\beta = X'y$

$X\cdot(X'X)^{-1}X'y$

Thm:

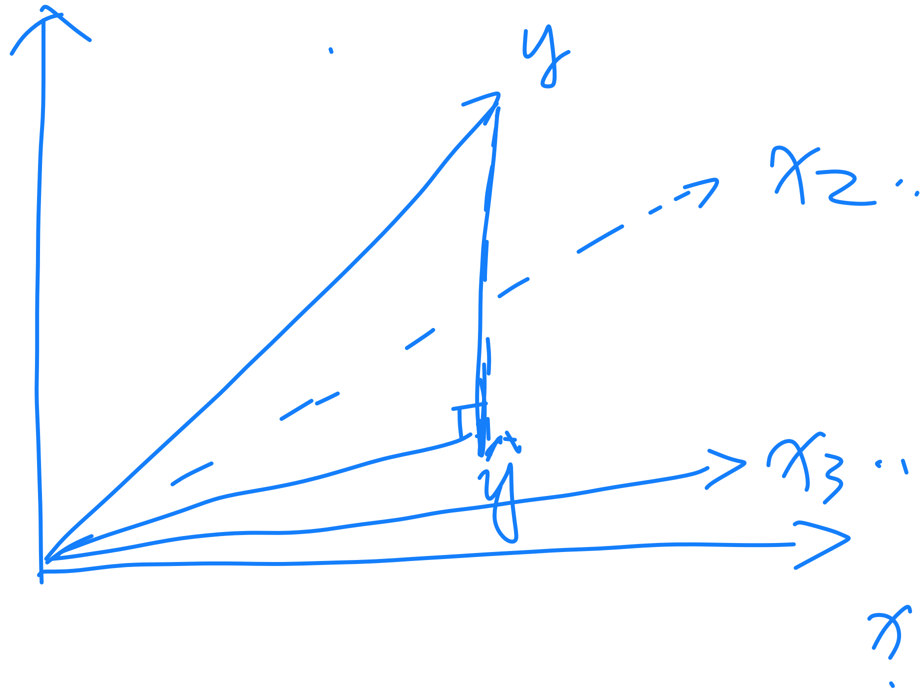
$\hat{y} = X(X'X)^{-}X'y$ is the projection of y onto $C(X)$. ✓

$X'X\beta = X'y$

$\Leftrightarrow X'(y - X\beta) = 0$

Pf: $\hat{\beta} = (X'X)^{-}X'y$ is a solution to the normal equation $X'X\beta = X'y$ [$y - X\hat{\beta} \perp x_i$ for all $i = 1, \cdots, p$]

$\Rightarrow \hat{y} = X\hat{\beta} = X\cdot(X'X)^{-}X'y$ is the proj onto $C(X)$ since the projection is unique.



$x_3 \in C([x_1, x_2])$

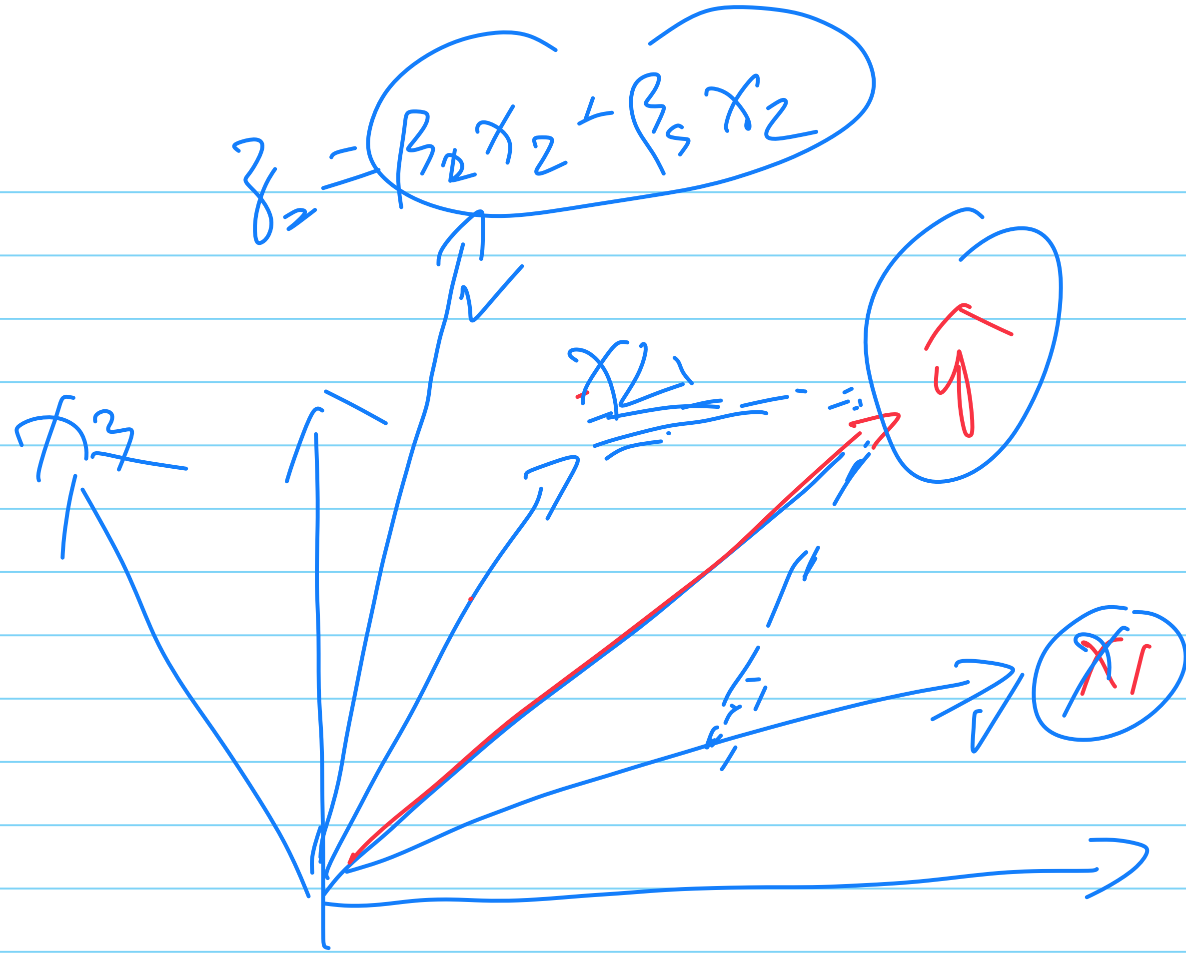
$C([x_1, x_2]) = C([x_1, x_2, x_3])$

$= C([x_2, x_3])$

$\hat{y}$ is unique (well-defined)

$\hat{y} = X \cdot \hat{\beta}$

12


x3
x2
x1

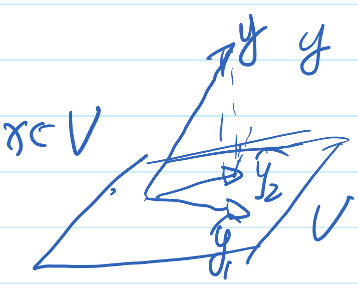
Uniqueness of Projection

Theorem: $\hat{y}_1, \hat{y}_2$ are two projections of y onto V. Then $\hat{y}_1 = \hat{y}_2$.

Pf: $\langle y, x\rangle - \langle \hat{y}_1, x\rangle$

$$\underline{\langle y - \hat{y}_1, x\rangle} = \underline{\langle y - \hat{y}_2, x\rangle} = 0$$

$\forall x \in V$



$$\Rightarrow \langle \hat{y}_1, x\rangle = \langle \hat{y}_2, x\rangle \quad \forall x \in V$$

$$\Rightarrow \langle \hat{y}_1 - \hat{y}_2, x\rangle = 0, \quad \forall x \in V$$

$$\Rightarrow \langle \hat{y}_1 - \hat{y}_2, \hat{y}_1 - \hat{y}_2\rangle = 0$$

$[\langle x+y, z\rangle = \langle x, z\rangle + \langle y, z\rangle]$

$$\Rightarrow \|\hat{y}_1 - \hat{y}_2\|^2 = 0$$

$$\Rightarrow \hat{y}_1 - \hat{y}_2 = 0$$

All the theorems based only on $\hat{y}$ rather than $\hat{\beta}$ are still valid for non-full-rank X, except that the number of columns should be modified to be rank (X)

Distribution of $\hat{\beta}$ and s^2

Theorem: In the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, $\mathrm{E}(\mathbf{e}) = \mathbf{0}$, $\mathrm{var}(\mathbf{e}) = \sigma^2\mathbf{I}$, and where $\mathbf{X}$ is $n \times p$ of rank $k \le p \le n$, we have the following properties of s^2:

(i) (unbiasedness) $\mathrm{E}(s^2) = \sigma^2$.

(ii) (uniqueness) s^2 is invariant to the choice of $\hat{\boldsymbol{\beta}}$ (i.e., to the choice of generalized inverse $(\mathbf{X}^T\mathbf{X})^-$).

$$S^2 = \frac{SSE}{n-k} = \frac{\|y - X\hat{\beta}\|^2}{n-k}$$

where $k = \mathrm{rank}(X)$. $\hat{y} = X\hat{\beta}$

$$y - X\hat{\beta} = P_{C(X)^\perp}\, y$$

Distributions of $\hat{\boldsymbol{\beta}}$ and s^2:

In the normal-errors, not-necessarily full rank model (*), the distribution of $\hat{\boldsymbol{\beta}}$ and s^2 can be obtained. These distributional results are essentially the same as in the full rank case, except for the mean and variance of $\hat{\boldsymbol{\beta}}$:

Theorem: In model (*),

(i) For any given choice of $(\mathbf{X}^T\mathbf{X})^-$,

$$\hat{\boldsymbol{\beta}} \sim N_p[(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X}\{(\mathbf{X}^T\mathbf{X})^-\}^T],$$

(ii) $(n-k)s^2/\sigma^2 \sim \chi^2(n-k)$, and $\frac{SSE}{\sigma^2} \sim \chi^2_{n-k}$

(iii) For any given choice of $(\mathbf{X}^T\mathbf{X})^-$, $\hat{\boldsymbol{\beta}}$ and s^2 are independent.

Thm: Suppose $y \sim N_n(X\beta, \sigma^2 I_n)$ where X is a matrix with rank $k+1$, and $X = [X_1, X_2]$, where $\text{rank}(X_2) = h$. $\hat{y}_0 = P_{C(X_1)} y$, $\hat{y} = P_{C(X)} y$. $\mu_0 = P_{C(X_1)}(X\beta)$. Then,

(i) $\frac{1}{\sigma^2}\|\mathbf{y} - \hat{\mathbf{y}}\|^2 = \frac{1}{\sigma^2}\mathbf{y}^T(\mathbf{I} - \mathbf{P}_{C(\mathbf{X})})\mathbf{y} \sim \chi^2(n-k-1)$;

(ii) $\frac{1}{\sigma^2}\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2 = \frac{1}{\sigma^2}\mathbf{y}^T(\mathbf{P}_{C(\mathbf{X})} - \mathbf{P}_{C(\mathbf{X}_1)})\mathbf{y} \sim \chi^2(h, \lambda_1)$, where

$$\lambda_1 = \frac{1}{2\sigma^2}\|(\mathbf{P}_{C(\mathbf{X})} - \mathbf{P}_{C(\mathbf{X}_1)})\boldsymbol{\mu}\|^2 = \frac{1}{2\sigma^2}\|\boldsymbol{\mu} - \boldsymbol{\mu}_0\|^2;$$

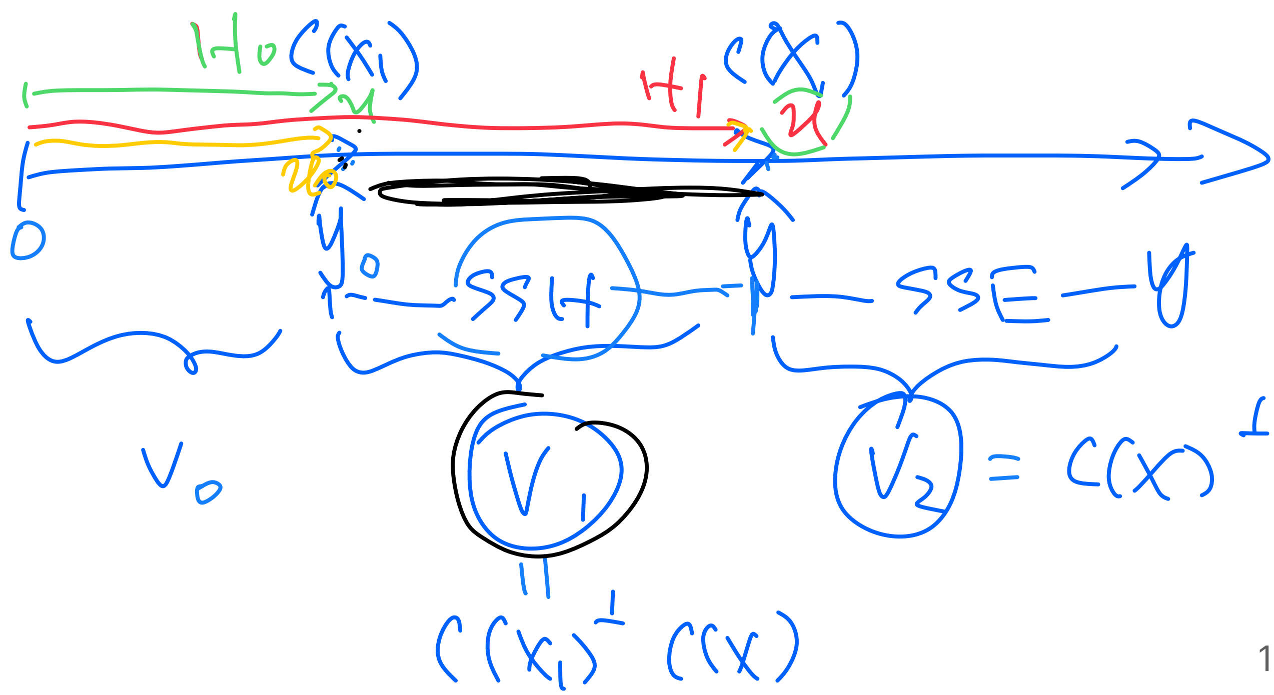
and

(iii) $\frac{1}{\sigma^2}\|\mathbf{y} - \hat{\mathbf{y}}\|^2$ and $\frac{1}{\sigma^2}\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2$ are independent.

Theorem: Under the conditions of the previous theorem,

$$F = \frac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2/h}{s^2} = \frac{\mathbf{y}^T(\mathbf{P}_{C(\mathbf{X})} - \mathbf{P}_{C(\mathbf{X}_1)})\mathbf{y}/h}{\mathbf{y}^T(\mathbf{I} - \mathbf{P}_{C(\mathbf{X})})\mathbf{y}/(n-k-1)}$$
$$\sim \begin{cases} F(h, n-k-1), & \text{under } H_0\text{; and} \\ F(h, n-k-1, \lambda_1), & \text{under } H_1, \end{cases}$$

where λ_1 is as given in the previous theorem.



Example: (one-way ANOVA)

An example of data

$$\underset{\substack{\uparrow\\ y}}{\begin{bmatrix} y_{11}\\ y_{12}\\ y_{21}\\ y_{22}\\ y_{31}\\ y_{32}\end{bmatrix}} \quad \underset{\substack{\uparrow\\ g\\ \text{(group index)}}}{\begin{bmatrix} g_1=1\\ g_2=1\\ g_3=2\\ g_4=2\\ g_5=3\\ g_6=3\end{bmatrix}} \quad \begin{array}{c}\begin{bmatrix} 1 & 0 & 0\\ 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\\ 0 & 0 & 1\end{bmatrix}\\ \begin{array}{ccc}\uparrow & \uparrow & \uparrow\\ x_1 & x_2 & x_3\end{array}\end{array} \quad \underset{j_n}{\begin{bmatrix}1\\1\\1\\1\\1\\1\end{bmatrix}}$$

$j_n \in L(x_1, x_2, x_3)$

$x_i = \mathbb{1}(g=i)$, indicator of group i

$H_0: \quad y_{ij} = \mu + \varepsilon_{ij}$

$H_1: \quad y_{ij} = \mu_i + \varepsilon_i = \mu + \alpha_i + \varepsilon_i$

In matrix,

H_0:

$$y = j_n \cdot \mu + \varepsilon, \quad j_n = (1, 1, \ldots, 1)'$$

H_1:

$$y = [x_1, x_2, x_3] \cdot \begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\end{pmatrix} + \varepsilon$$

$$= [j_6, x_1, x_2, x_3] \begin{pmatrix}\mu\\ \alpha_1\\ \alpha_2\\ \alpha_3\end{pmatrix} + \varepsilon$$

projections:

Under H_0: $\quad proj(y \mid j_n) \equiv P_0 y$

under H_1: $\quad proj(y \mid L(x_1, x_2, x_3)) \equiv P_1 y$

$$L(j_n) \subseteq L(x_1, x_2, x_3)$$

since $j_n = x_1 + x_2 + x_3$

That is, H_0 is a reduced model of H_1.

$$P_0 = \frac{1}{n} j_n j_n'$$

$\bar{y}_{..} j_n$

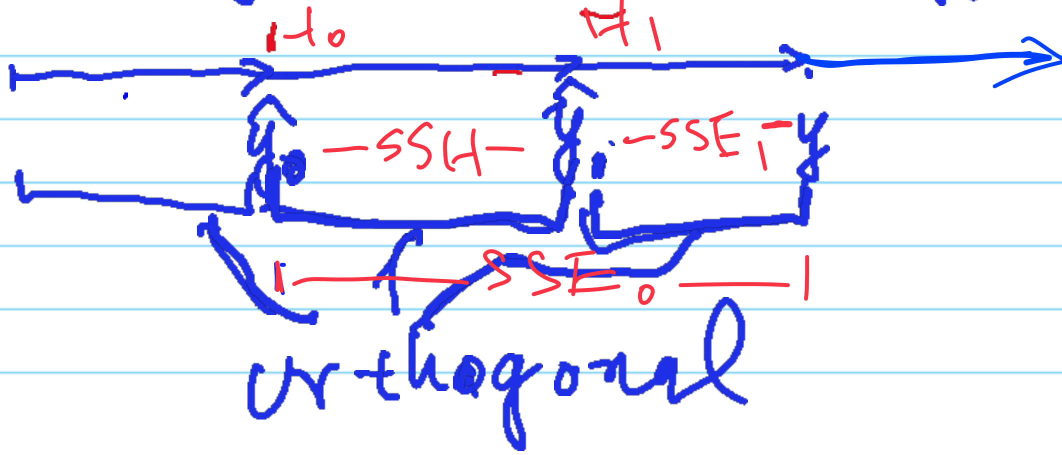
overall mean

$$\hat{y}_0 = P_0 y = (\bar{y}_{..}, \bar{y}_{..}, \cdots, \bar{y}_{..})'$$

group mean

$$\hat{y}_1 = P_1 y = (\bar{y}_{1.}, \bar{y}_{1.}, \bar{y}_{2.}, \bar{y}_{2.}, \bar{y}_{3.}, \bar{y}_{3.})'$$

$$= \bar{y}_{1.} \cdot x_1 + \bar{y}_{2.} \cdot x_2 + \bar{y}_{3.} \cdot x_3$$

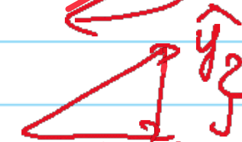


some SS based on $\hat{y}_0$ & $\hat{y}_1$:

$$SSE_0 = \|y - \hat{y}_0\|^2 = \sum_{i,j} (y_{ij} - \bar{y}_{..})^2$$

(SST)

$$= \|y\|^2 - \|\hat{y}_0\|^2$$



$$= \sum_{i,j} y_{ij}^2 - n \cdot \bar{y}_{..}^2$$

$y - \hat{y}_0 \perp \hat{y}_0$

$$= s_y^2$$ sample variance of y

(SSW)

$$SSE_1 = \|y - \hat{y}_1\|^2$$

$$= \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$$

SS within group i

$$= \|y\|^2 - \|\hat{y}_1\|^2$$

$$= \sum_{ij} y_{ij}^2 - \sum_i n_i \bar{y}_{i.}^2$$

$\frac{y_{i.}^2}{n_i}$

(SSB)

$$SSH = SSE_0 - SSE_1$$

$$= \|y - \hat{y}_0\|^2 - \|y - \hat{y}_1\|^2$$

$\frac{y_{..}}{n}$

$$= \|\hat{y}_0 - \hat{y}_1\|^2$$

2

$$= \|\hat{y}_1\|^2 - \|\hat{y}_0\|^2 = \sum_i n_i \bar{y}_{i.}^2 - n \bar{y}_{..}^2$$

$$F = \frac{SSH/2}{SSE/(n-3)}$$

Under H_0:

$$F \sim F_{2,\, n-3}$$

rank(X)

H_0: $y_{ij} = \mu + \varepsilon_{ij}$ 1

H_1: $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$. 3

$a = 3$ $h = 3 - 1 = 2$

Estimation and Testing of Estimable Parameters in Non-full Rank Models

Definition

Estimability: Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)^T$ be a vector of constants. The parameter $\boldsymbol{\lambda}^T\boldsymbol{\beta} = \sum_j \lambda_j \beta_j$ is said to be **estimable** if there exists a vector $\mathbf{a}$ in $\mathcal{R}^n$ such that

$$\mathrm{E}(\mathbf{a}^T\mathbf{y}) = \boldsymbol{\lambda}^T\boldsymbol{\beta}, \quad \text{for all } \boldsymbol{\beta} \in \mathcal{R}^p. \qquad (\dagger)$$

Since ($\dagger$) is equivalent to $\mathbf{a}^T\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\lambda}^T\boldsymbol{\beta}$ for all $\boldsymbol{\beta}$, it follows that $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is estimable if and only if there exists $\mathbf{a}$ such that $\mathbf{X}^T\mathbf{a} = \boldsymbol{\lambda}$ (i.e., iff $\boldsymbol{\lambda}$ lies in the row space of $\mathbf{X}$).

$\lambda' = a'X$

That is $\forall a \in \mathbb{R}^n$, $a'X\beta$ is estimable

because $\widehat{a'X\beta} = a'\hat{y} = a'(X(X'X)^- X'y)$

$E(\hat{y}) = X(X'X)^- X' \cdot X\beta = X\beta$

$E(a'\hat{y}) = a'X\beta$

Note that $\underbrace{X(X'X)^- X'}_{P_{C(X)}} X = X$ $\qquad P_{C(X)} \cdot X = X$

However,

This doesn't mean that

$(X'X)^- X'X = I_p$

useful!

Theorem 12.2b. In the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, the linear function $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is estimable if and only if any one of the following equivalent conditions holds:

(i) $\boldsymbol{\lambda}'$ is a linear combination of the rows of $\mathbf{X}$; that is, there exists a vector $\mathbf{a}$ such that

$\lambda \in C(X')$

$$\mathbf{a}'\mathbf{X} = \boldsymbol{\lambda}'. \tag{12.15}$$

(ii) $\boldsymbol{\lambda}'$ is a linear combination of the rows of $\mathbf{X}'\mathbf{X}$ or $\boldsymbol{\lambda}$ is a linear combination of the columns of $\mathbf{X}'\mathbf{X}$, that is, there exists a vector $\mathbf{r}$ such that

$\lambda \in C(X'X)$

$$\mathbf{r}'\mathbf{X}'\mathbf{X} = \boldsymbol{\lambda}' \quad \text{or} \quad \mathbf{X}'\mathbf{X}\mathbf{r} = \boldsymbol{\lambda}. \tag{12.16}$$

(iii) $\boldsymbol{\lambda}$ or $\boldsymbol{\lambda}'$ is such that

$$\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{\lambda} = \boldsymbol{\lambda} \quad \text{or} \quad \boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \boldsymbol{\lambda}', \tag{12.17}$$

$A A^{-} c = c$

where $(\mathbf{X}'\mathbf{X})^{-}$ is any (symmetric) generalized inverse of $\mathbf{X}'\mathbf{X}$.

Remarks:

An easy way to check whether $\lambda \in C(X'X)$ on computer is let $A = X'X$ and $c = \lambda$ in the following theorem:

Theorem 2.7 The system of equations $\mathbf{A}\mathbf{x} = \mathbf{c}$ has at least one solution vector $\mathbf{x}$ if and only if $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}, \mathbf{c})$.

Given a λ, one can also use condition (iii) to check whether $\lambda \in C(X'X)$

Theorem 2.8f. The system of equations $\mathbf{A}\mathbf{x} = \mathbf{c}$ has a solution if and only if for any generalized inverse $\mathbf{A}^{-}$ of $\mathbf{A}$

$$\mathbf{A}\mathbf{A}^{-}\mathbf{c} = \mathbf{c}.$$

Pf: $A A^{-} \underbrace{A x}_{c} = A x = c$

Theorem: In the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, any estimable function $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ can be obtained by taking a linear combination of the elements of $\mathbf{X}\boldsymbol{\beta}$ or of the elements of $\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}$.

$\lambda'\beta = a'\underline{X\beta}$ for some $a \in \mathbb{R}^{n}$

$\lambda'\beta = r'\underline{X'X\beta}$ for some $r \in \mathbb{R}^{p}$

Example:

$(X'X)^{-1}$ exists, i.e., X is full-rank

$$C(X'X) = C(X') = \mathbb{R}^p$$

$\forall \lambda \in \mathbb{R}^p$, $\lambda'\beta$ is estimable

24

Consider again the effects version of the (balanced) one way layout model:

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, \ldots, a, j = 1, \ldots, n.$$

Suppose that $a = 3$ and $n = 2$. Then, in matrix notation, this model is

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} + \mathbf{e}.$$

$$X\beta = \begin{bmatrix} \mu+\alpha_1 \\ \mu+\alpha_1 \\ \mu+\alpha_2 \\ \mu+\alpha_2 \\ \mu+\alpha_3 \\ \mu+\alpha_3 \end{bmatrix}$$

Let

$\mu_1 = \mu+\alpha_1$

$\mu_2 = \mu+\alpha_2$

$\mu_3 = \mu+\alpha_3$

Any functions of μ_1, μ_2, μ_3

So, any linear combination $\mathbf{a}^T\mathbf{X}\boldsymbol{\beta}$ for some $\mathbf{a}$ is estimable.

μ, $\alpha_1+\alpha_2+\alpha_3$, α_1, α_2, α_3 are non-estimable

$\alpha_i - \alpha_k$ is estimable.

$$X'X = \begin{pmatrix} j_6' \\ x_1' \\ x_2' \\ x_3' \end{pmatrix} \cdot (j_6, x_1, x_2, x_3)$$

$$= \begin{pmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix} = 2 \cdot \begin{pmatrix} 3 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

Estimable λ:

	λ_1	λ_2	λ_3	λ_4	$\lambda_1 - \lambda_2$	
	1	1	1	3	0	
	1	0	0	1	1	
	0	1	0	1	-1	
	0	0	1	1	0	
$\lambda'\beta$	$u+\alpha_1$	$u+\alpha_2$	$u+\alpha_3$	$\Sigma(u+\alpha_i)$	$\alpha_1 - \alpha_2$	$\cdots$

$$\lambda_1 = \begin{array}{c} u \\ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{array} \quad \lambda_2 = \begin{array}{c} \alpha_1 \\ \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \end{array} \notin C(X'X).$$

Example

$$y = \beta_1 x_1 + \beta_2 x_1 + e = (\beta_1 + \beta_2) x_1 + e$$

$$= (x_1, x_1) \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + e$$

$$\lambda' \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = (r_1, r_2) \begin{pmatrix} x_1' \\ x_1' \end{pmatrix} (x_1, x_1) \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

$$= (r_1, r_2) \cdot \begin{pmatrix} x_1' x_1 & x_1' x_1 \\ x_1' x_1 & x_1' x_1 \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

$$= x_1' x_1 \cdot (r_1, r_2) \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

$$= x_1' \cdot x_1 \cdot (r_1, r_2) \cdot \begin{pmatrix} \beta_1 + \beta_2 \\ \beta_1 + \beta_2 \end{pmatrix}$$

$$= x_1' \cdot x_1 \cdot (r_1 + r_2) \cdot (\beta_1 + \beta_2)$$

That is, any function of $\beta_1 + \beta_2$ is estimable!

β_1 & β_2 individually non-estimable.

	λ_1	λ_2	λ_3
	1	1	0
	1	0	1
$\lambda'\beta$	$\beta_1 + \beta_2$	β_1	β_2

Definition

A set of functions $\boldsymbol{\lambda}_1'\boldsymbol{\beta}$, $\boldsymbol{\lambda}_2'\boldsymbol{\beta}, \ldots, \boldsymbol{\lambda}_m'\boldsymbol{\beta}$ is said to be linearly independent if the coefficient vectors $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_m$ are linearly independent [see (2.40)]. The number of linearly independent estimable functions is given in the next theorem.

Theorem 12.2c. In the non-full-rank model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, the number of linearly independent estimable functions of $\boldsymbol{\beta}$ is the rank of $\mathbf{X}$.

Theorem: Let $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ be an estimable function of $\boldsymbol{\beta}$ in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$. Let $\hat{\boldsymbol{\beta}}$ be any solution of the normal equation $\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{y}$. Then the estimator $\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}$ has the following properties:

(i) (unbiasedness) $\mathrm{E}(\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}) = \boldsymbol{\lambda}^T\boldsymbol{\beta}$; and

(ii) (uniqueness) $\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}$ is invariant to the choice of $\hat{\boldsymbol{\beta}}$ (to the choice of generalized inverse $(\mathbf{X}^T\mathbf{X})^-$ in the formula $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y}$.

Proof: Part (i):

$$\mathrm{E}(\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}) = \boldsymbol{\lambda}^T\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\lambda}^T\boldsymbol{\beta}$$

Part (ii): Because $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is estimable, $\boldsymbol{\lambda} = \mathbf{X}^T\mathbf{a}$ for some $\mathbf{a}$. Therefore,

$$\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} = \mathbf{a}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y} = \mathbf{a}^T\mathbf{P}_{C(\mathbf{X})}\mathbf{y}.$$

$$\lambda'(X'X)^- X'X\beta$$
$$= a'\underbrace{X(X'X)^- X'X}_{=X}\beta$$
$$= a'X = \lambda'\beta$$

Note: $\lambda'\hat{\beta} = a'X(X'X)^- X'y$
$$= a'P_{C(X)}y$$
$$= a'\hat{y}$$

$Var(C\hat{\beta}) = \sigma^2 C \cdot (X'X)^- C'$

Theorem: Under the conditions of the previous theorem, and where $\text{var}(\mathbf{e}) = \text{var}(\mathbf{y}) = \sigma^2\mathbf{I}$, the variance of $\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}$ is unique, and is given by

$$\text{var}(\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^-\boldsymbol{\lambda},$$

where $(\mathbf{X}^T\mathbf{X})^-$ is any generalized inverse of $\mathbf{X}^T\mathbf{X}$.

Proof:

$$\begin{aligned}
\text{var}(\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}) &= \boldsymbol{\lambda}^T\text{var}((\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y})\boldsymbol{\lambda} \\
&= \boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{X}\{(\mathbf{X}^T\mathbf{X})^-\}^T\boldsymbol{\lambda} \\
&= \sigma^2\underbrace{\boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{X}}_{=\boldsymbol{\lambda}^T}\{(\mathbf{X}^T\mathbf{X})^-\}^T\boldsymbol{\lambda} \\
&= \sigma^2\boldsymbol{\lambda}^T\{(\mathbf{X}^T\mathbf{X})^-\}^T\boldsymbol{\lambda} \\
&= \sigma^2\mathbf{a}^T\mathbf{X}\{(\mathbf{X}^T\mathbf{X})^-\}^T\mathbf{X}^T\mathbf{a} \quad \text{(for some } \mathbf{a}) \\
&= \sigma^2\mathbf{a}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{a} = \sigma^2\boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^-\boldsymbol{\lambda}.
\end{aligned}$$

Note: $(X^-)' = (X')^-$

unique

Uniqueness: since $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is estimable $\boldsymbol{\lambda} = \mathbf{X}^T\mathbf{a}$ for some $\mathbf{a}$. Therefore,

$$\begin{aligned}
\text{var}(\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}) &= \sigma^2\boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^-\boldsymbol{\lambda} \\
&= \sigma^2\mathbf{a}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{a} = \sigma^2\mathbf{a}^T\mathbf{P}_{C(\mathbf{X})}\mathbf{a}
\end{aligned}$$

Again, the result follows from the fact that projection matrices are unique.

∎

$\lambda' = a'X$ for some $a \in \mathbb{R}^n$

$\lambda'(X'X)^- X'X = a'X \cdot (X'X)^- X'X$

$= a'X = \lambda'$

$[X(X'X)^- X']X = X$

30

Theorem: Let $\boldsymbol{\lambda}_1^T\boldsymbol{\beta}$ and $\boldsymbol{\lambda}_2^T\boldsymbol{\beta}$ be two estimable function in the model considered in the previous theorem (the spherical errors, non-full-rank linear model). Then the covariance of the least-squares estimators of these quantities is

$$\mathrm{cov}(\boldsymbol{\lambda}_1^T\hat{\boldsymbol{\beta}}, \boldsymbol{\lambda}_2^T\hat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{\lambda}_1^T(\mathbf{X}^T\mathbf{X})^-\boldsymbol{\lambda}_2.$$

✓

$$\mathrm{cov}(a'\hat{\beta}, b'\hat{\beta})$$

$$= a' \, \mathrm{cov}(\hat{\beta}) \cdot b$$

Theorem: (Gauss-Markov in the non-full rank case) If $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is estimable in the spherical errors non-full rank linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, then $\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}$ is its BLUE.

Proof: Since $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is estimable, $\boldsymbol{\lambda} = \mathbf{X}^T\mathbf{a}$ for some $\mathbf{a}$. $\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} = \mathbf{a}^T\mathbf{X}\hat{\boldsymbol{\beta}}$ is a linear estimator because it is of the form

$$\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} = \mathbf{a}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y} = \mathbf{a}^T\mathbf{P}_{C(\mathbf{X})}\mathbf{y} = \mathbf{c}^T\mathbf{y}$$

where $\mathbf{c} = \mathbf{P}_{C(\mathbf{X})}\mathbf{a}$. We have already seen that $\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}$ is unbiased. Consider any other linear estimator $\mathbf{d}^T\mathbf{y}$ of $\boldsymbol{\lambda}^T\boldsymbol{\beta}$. For $\mathbf{d}^T\mathbf{y}$ to be unbiased, the mean of $\mathbf{d}^T\mathbf{y}$, which is $\mathrm{E}(\mathbf{d}^T\mathbf{y}) = \mathbf{d}^T\mathbf{X}\boldsymbol{\beta}$, must satisfy $\mathrm{E}(\mathbf{d}^T\mathbf{y}) = \boldsymbol{\lambda}^T\boldsymbol{\beta}$, for all $\boldsymbol{\beta}$, or equivalently, it must satisfy $\mathbf{d}^T\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\lambda}^T\boldsymbol{\beta}$, for all $\boldsymbol{\beta}$, which implies

$$\mathbf{d}^T\mathbf{X} = \boldsymbol{\lambda}^T.$$

The covariance between $\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}$ and $\mathbf{d}^T\mathbf{y}$ is

$$\begin{aligned}\mathrm{cov}(\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}, \mathbf{d}^T\mathbf{y}) &= \mathrm{cov}(\mathbf{c}^T\mathbf{y}, \mathbf{d}^T\mathbf{y}) = \sigma^2\mathbf{c}^T\mathbf{d} \\ &= \sigma^2\boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{d} = \sigma^2\boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^-\boldsymbol{\lambda}.\end{aligned}$$

Now

$$\begin{aligned}0 &\le \mathrm{var}(\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} - \mathbf{d}^T\mathbf{y}) = \mathrm{var}(\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}) + \mathrm{var}(\mathbf{d}^T\mathbf{y}) - 2\mathrm{cov}(\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}, \mathbf{d}^T\mathbf{y}) \\ &= \sigma^2\boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^-\boldsymbol{\lambda} + \mathrm{var}(\mathbf{d}^T\mathbf{y}) - 2\sigma^2\boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^-\boldsymbol{\lambda} \\ &= \mathrm{var}(\mathbf{d}^T\mathbf{y}) - \underbrace{\sigma^2\boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^-\boldsymbol{\lambda}}_{=\mathrm{var}(\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}})}\end{aligned}$$

Therefore,

$$\mathrm{var}(\mathbf{d}^T\mathbf{y}) \ge \mathrm{var}(\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}})$$

with equality holding iff $\mathbf{d}^T\mathbf{y} = \boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}$. I.e., an arbitrary linear unbiased estimator $\mathbf{d}^T\mathbf{y}$ has variance $\ge$ to that of the least squares estimator with equality iff the arbitrary estimator is the least squares estimator. ∎

Definition

A hypothesis such as $H_0: \beta_1 = \beta_2 = \cdots = \beta_q$ is said to be *testable* if there exists a set of linearly independent estimable functions $\boldsymbol{\lambda}_1'\boldsymbol{\beta}, \boldsymbol{\lambda}_2'\boldsymbol{\beta}, \ldots, \boldsymbol{\lambda}_t'\boldsymbol{\beta}$ such that H_0 is true if and only if $\boldsymbol{\lambda}_1'\boldsymbol{\beta} = \boldsymbol{\lambda}_2'\boldsymbol{\beta} = \cdots = \boldsymbol{\lambda}_t'\boldsymbol{\beta} = 0$.

Theorem 12.7b. If $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, if $\mathbf{C}$ is $m \times p$ of rank $m \leq k$ such that $C\boldsymbol{\beta}$ is a set of m linearly independent estimable functions, and if $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y}$, then

(i) $\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{C}'$ is nonsingular.

(ii) $\mathbf{C}\hat{\boldsymbol{\beta}}$ is $N_m[\mathbf{C}\boldsymbol{\beta}, \sigma^2\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{C}']$.

(iii) $\text{SSH}/\sigma^2 = (\mathbf{C}\hat{\boldsymbol{\beta}})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}/\sigma^2$ is $\chi^2(m, \lambda)$, where $\lambda = (\mathbf{C}\boldsymbol{\beta})'[\mathbf{C}(\mathbf{X}'\mathbf{X})^-\mathbf{C}']^{-1}\mathbf{C}\boldsymbol{\beta}/2\sigma^2$.

(iv) $\text{SSE}/\sigma^2 = \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}']\mathbf{y}/\sigma^2$ is $\chi^2(n-k)$.

(v) SSH and SSE are independent.

Summary:

If we restrict ourselves to estimable $\lambda'\beta$, all the previous results for $\lambda'\beta$ can be generalized to non-full-rank models with $(X'X)^{-1}$ being replaced by $(X'X)^-$.

Re-parametrization for Non-full-rank Models

The idea in reparameterization is to transform from the vector of non-estimable parameters $\boldsymbol{\beta}$ in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ where $\mathbf{X}$ is $n \times p$ with rank $k < p \le n$, to a vector of linearly independent estimable functions of $\boldsymbol{\beta}$:

$$\begin{pmatrix} \mathbf{u}_1^T\boldsymbol{\beta} \\ \mathbf{u}_2^T\boldsymbol{\beta} \\ \vdots \\ \mathbf{u}_k^T\boldsymbol{\beta} \end{pmatrix} = \boxed{\mathbf{U}\boldsymbol{\beta} \equiv \boldsymbol{\gamma}.}$$

Here $\mathbf{U}$ is the $k \times p$ matrix with rows $\mathbf{u}_1^T, \ldots, \mathbf{u}_k^T$, so that the elements of $\boldsymbol{\gamma} = \mathbf{U}\boldsymbol{\beta}$ are a "full set" of linearly independent estimable functions of $\boldsymbol{\beta}$.

The new full-rank model is

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}, \qquad (*)$$

where $\mathbf{Z}$ is $n \times k$ of full rank, and $\mathbf{Z}\boldsymbol{\gamma} = \mathbf{X}\boldsymbol{\beta}$ (the mean under the non-full rank model is the same as under the full rank model, we've just changed the parameterization; i.e., switched from $\boldsymbol{\beta}$ to $\boldsymbol{\gamma}$.)

To find the new (full rank) model matrix $\mathbf{Z}$, note that $\mathbf{Z}\boldsymbol{\gamma} = \mathbf{X}\boldsymbol{\beta}$ and $\boldsymbol{\gamma} = \mathbf{U}\boldsymbol{\beta}$ for all $\boldsymbol{\beta}$ imply

$$\begin{aligned} \mathbf{Z}\mathbf{U}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta}, \quad \text{for all } \boldsymbol{\beta}, \quad &\Rightarrow \quad \mathbf{Z}\mathbf{U} = \mathbf{X} \\ &\Rightarrow \quad \mathbf{Z}\mathbf{U}\mathbf{U}^T = \mathbf{X}\mathbf{U}^T \\ &\Rightarrow \quad \boxed{\mathbf{Z} = \mathbf{X}\mathbf{U}^T(\mathbf{U}\mathbf{U}^T)^{-1}.} \end{aligned}$$

- Note that $\mathbf{U}$ is of full rank, so $(\mathbf{U}\mathbf{U}^T)^{-1}$ exists.
- Note also that we have constructed $\mathbf{Z}$ to be of full rank:

$$\text{rank}(\mathbf{Z}) \ge \text{rank}(\mathbf{Z}\mathbf{U}) = \text{rank}(\mathbf{X}) = k$$

  but

$$\text{rank}(\mathbf{Z}) \le k, \quad \text{because } \mathbf{Z} \text{ is } n \times k.$$

  Therefore, $\text{rank}(\mathbf{Z}) = k$.

Example 12.5. We illustrate a reparameterization for the model $y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, 2, \quad j = 1, 2$. In matrix form, the model can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \end{pmatrix}.$$

Since $\mathbf{X}$ has rank 2, there exist two linearly independent estimable functions (see Theorem 12.2c). We can choose these in many ways, one of which is $\mu + \tau_1$ and $\mu + \tau_2$. Thus

$$\boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \mu + \tau_1 \\ \mu + \tau_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}}_{U} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix} = \mathbf{U}\boldsymbol{\beta}.$$

To reparameterize in terms of $\boldsymbol{\gamma}$, we can use

$$\mathbf{Z} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \qquad Z = \underline{X U' (U U')^{-1}}$$

Since $\mathbf{X}$ has rank 2, there exist two linearly independent estimable functions (see Theorem 12.2c). We can choose these in many ways, one of which is $\mu + \tau_1$ and $\mu + \tau_2$. Thus

$$\boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \mu + \tau_1 \\ \mu + \tau_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix} = \mathbf{U}\boldsymbol{\beta}.$$

To reparameterize in terms of $\boldsymbol{\gamma}$, we can use

$$\mathbf{Z} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix},$$

so that $\mathbf{Z}\boldsymbol{\alpha} = \mathbf{X}\boldsymbol{\beta}$:

$$\mathbf{Z}\boldsymbol{\gamma} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_1 \\ \gamma_2 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_2 \\ \mu + \tau_2 \end{pmatrix}.$$

[The matrix $\mathbf{Z}$ can also be obtained directly using (12.31).] It is easy to verify that $\mathbf{ZU} = \mathbf{X}$.

$$\mathbf{ZU} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} = \mathbf{X}.$$

Side Condition

$T\beta = 0$

Theorem 12.6a. If $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ is $n \times p$ of rank $k < p \leq n$, and if $\mathbf{T}$ is a $(p-k) \times p$ matrix of rank $p-k$ such that $\mathbf{T}\boldsymbol{\beta}$ is a set of nonestimable functions, then there is a unique vector $\hat{\boldsymbol{\beta}}$ that satisfies both $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ and $\mathbf{T}\hat{\boldsymbol{\beta}} = \mathbf{0}$.

rank(X) = k, p−k non-est.

Pf: $\begin{array}{c} \overbrace{\quad}^{p} \\ \begin{array}{c} n\{ \\ p-k\{ \end{array} \begin{pmatrix} X \\ T \end{pmatrix} \end{array}$ is a matrix with rank = p.

since each row of T ∉ R(X)

Therefore, $(X', T')\begin{pmatrix} X \\ T \end{pmatrix} = \underbrace{X'X + T'T}_{p \times p}$

is non-singular.

We want to solve these equations:

$$\begin{cases} X'X\beta = X'y \\ T\beta = 0 \Rightarrow T'T\beta = 0 \end{cases}$$

$$\Rightarrow (X'X + T'T)\beta = X'y + 0$$

$$\Rightarrow \hat{\beta} = (X'X + T'T)^{-1} X'y$$

one version of $(X'X)^-$

$y_{ij} = u + \alpha_i + e,$ $\quad \Sigma \alpha_i = 0$

Example 12.6. Consider the model $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $i = 1, 2$, $j = 1, 2$ as in Example 12.5. The function $\tau_1 + \tau_2$ was shown to be nonestimable in Problem 12.5b. The side condition $\tau_1 + \tau_2 = 0$ can be expressed as $(0, 1, 1)\boldsymbol{\beta} = 0$, and $\mathbf{X}'\mathbf{X} + \mathbf{T}'\mathbf{T}$ becomes

$$X = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

$$\underbrace{\begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} (0 \quad 1 \quad 1)}_{X'X + T'T} = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 3 & 1 \\ 2 & 1 & 3 \end{pmatrix}.$$

$$\beta = \begin{pmatrix} u \\ \tau_1 \\ \tau_2 \end{pmatrix}$$

Then

$$(\mathbf{X}'\mathbf{X} + \mathbf{T}'\mathbf{T})^{-1} = \frac{1}{4} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & 0 \\ -1 & 0 & 2 \end{pmatrix}.$$

$T = (0, 1, 1)$

With $\mathbf{X}'\mathbf{y} = (y_{..}, y_{1.}, y_{2.})'$, we obtain, by (12.37)

$T\beta = (0, 1, 1) \cdot \begin{pmatrix} u \\ \tau_1 \\ \tau_2 \end{pmatrix}$

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X} + \mathbf{T}'\mathbf{T})^{-1}\mathbf{X}'\mathbf{y} \\ &= \frac{1}{4} \begin{pmatrix} 2y_{..} - y_{1.} - y_{2.} \\ 2y_{1.} - y_{..} \\ 2y_{2.} - y_{..} \end{pmatrix} = \begin{pmatrix} \bar{y}_{..} \\ \bar{y}_{1.} - \bar{y}_{..} \\ \bar{y}_{2.} - \bar{y}_{..} \end{pmatrix}, \end{aligned} \tag{12.39}$$

$$\tau_1 = u_1 - \frac{u_1 + u_2}{2}$$

since $y_{1.} + y_{2.} = y_{..}$.

We now show that $\hat{\boldsymbol{\beta}}$ in (12.39) is also a solution to the normal equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$:

$$\begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix} \begin{pmatrix} \bar{y}_{..} \\ \bar{y}_{1.} - \bar{y}_{..} \\ \bar{y}_{2.} - \bar{y}_{..} \end{pmatrix} = \begin{pmatrix} y_{..} \\ y_{1.} \\ y_{2.} \end{pmatrix}, \quad \text{or}$$

$$\begin{aligned} 4\bar{y}_{..} + 2(\bar{y}_{1.} - \bar{y}_{..}) + 2(\bar{y}_{2.} - \bar{y}_{..}) &= y_{..} \\ 2\bar{y}_{..} + 2(\bar{y}_{1.} - \bar{y}_{..}) &= y_{1.} \\ 2\bar{y}_{..} + 2(\bar{y}_{2.} - \bar{y}_{..}) &= y_{2.} \end{aligned}$$

These simplify to

$$\begin{aligned} 2\bar{y}_{1.} + 2\bar{y}_{2.} &= y_{..} \\ 2\bar{y}_{1.} &= y_{1.} \\ 2\bar{y}_{2.} &= y_{2.}, \end{aligned}$$

which hold because $\bar{y}_{1.} = y_{1.}/2$, $\bar{y}_{2.} = y_{2.}/2$ and $y_{1.} + y_{2.} = y_{..}$. □