- Make-up tommorow (F, Apr-15, 2:00 - 3:45pm) : E2-194

- ① Uniqueness of $\hat{y} = X\hat{\beta} = \underbrace{X(X^T X)^- X^T}_{=P} y = Py$ (and $\hat{e}$ is also unique) $y = \hat{y} + \hat{e}$

  ② maybe multiple solutions for $\hat{\beta}$

  We can always find one, $\hat{\beta} = (X^T X)^- X^T y$

- Finish reparameterization

  Start Estimability

* g-inverses exist for arbitrary matrices ($\exists G$ s.t $AGA = A$)

* For A symmetric, A^- need not be symmetric

<u>Recall</u> Find a g-inverse of A,

$$A = \left[\begin{array}{c|c} a & b \\ \hline b & b^2/a \end{array}\right]$$

$\Rightarrow$ Our A^- was $A^- = \begin{bmatrix} \frac{1}{a} & 0 \\ 0 & 0 \end{bmatrix}$: symmetric

consider $A^- = \begin{bmatrix} \frac{1}{a} & -1 \\ 1 & 0 \end{bmatrix}$: not symmetric

$$\underset{A}{\begin{bmatrix} a & b \\ b & b^2/a \end{bmatrix}} \underset{A^-}{\begin{bmatrix} \frac{1}{a} & -1 \\ 1 & 0 \end{bmatrix}} \underset{A}{\begin{bmatrix} a & b \\ b & b^2/a \end{bmatrix}} = \begin{bmatrix} 1+b & -a \\ \frac{b}{a} + \frac{b^2}{a} & -b \end{bmatrix} \begin{bmatrix} a & b \\ b & \frac{b^2}{a} \end{bmatrix}$$

$$= \begin{bmatrix} a & b \\ b & \frac{b}{a} \end{bmatrix} = A$$

<u>Bottom Line</u>: For A symmetric, we can find a symmetric A^-

* Later we will discuss spectral decomposition of a symmetric matrix. We will discuss g-inverse more there

$M \quad \mathcal{C}(X)$

$\mathcal{C}(X) = \mathcal{C}(M)$

Two models are equivalent (or reparameterization of each other) if the column spaces of the design matrices are the same.

†: Def 2.1 Two linear models, $\mathbf{y} = \boxed{\mathbf{X}}\beta + \mathbf{e}$ where $\mathbf{X}$ is a $n \times p$ matrix and $\mathbf{y} = \boxed{\mathbf{W}}\gamma + \mathbf{e}$ where $\mathbf{W}$ is a $n \times t$ matrix, are equivalent (or reparameterization of each other) iff $C(\mathbf{X}) = C(\mathbf{W})$.

★ Result 2.8 and Cor 2.4: If $C(\mathbf{X}) = C(\mathbf{W})$,

- $\mathbf{P}_x = \mathbf{P}_w$
- $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \mathbf{W}\hat{\gamma}$
- $\hat{\mathbf{e}} = (I - \mathbf{P}_x)\mathbf{y} = (I - \mathbf{P}_w)\mathbf{y}$

Ex 2 ~~Mode~~

Model 1 $\quad y_{ij} = \beta_0^{(i)} + \beta_1^{(i)} x_{ij} + e_{ij}$ $\qquad i = 1, 2$ two groups, $\quad j = 1, \dots, n$ replicates

Model 2 $\qquad d_{ij} = \begin{cases} 0 & \text{if } i = 1 \text{ (group 1)} \\ 1 & \text{if } i = 2 \text{ (group 2)} \end{cases}$

$$\Rightarrow \quad y_{ij} = \gamma_0 + \gamma_1 x_{ij} + \gamma_2 d_{ij} + \gamma_3 (d_{ij}) x_{ij} + e_{ij}$$

group 1 $\to d_{ij} = 0 \Rightarrow \quad y_{ij} = \underbrace{\gamma_0}_{\beta_0^{(1)}} + \underbrace{\gamma_1}_{\beta_1^{(1)}} x_{ij} + e_{ij}$

group 2 $\to d_{ij} = 1 \Rightarrow \quad y_{ij} = \underbrace{(\gamma_0 + \gamma_2)}_{= \beta_0^{(2)}} + \underbrace{(\gamma_1 + \gamma_3)}_{= \beta_1^{(2)}} x_{ij} + e_{ij}$

$$X = \begin{bmatrix} 1 & x_{11} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & 0 & 0 \\ 0 & 0 & 1 & x_{21} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_{2n} \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0^{(1)} \\ \beta_1^{(1)} \\ \beta_0^{(2)} \\ \beta_1^{(2)} \end{bmatrix}$$

$$W = \begin{array}{c} \left[\begin{array}{cccc} 1 & x_{11} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & 0 & 0 \\ \hline \vdots & \vdots & 1 & x_{21} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{2n} & 1 & x_{2n} \end{array}\right] \end{array} \begin{array}{l} \text{group 1} \\ \\ \\ \text{group 2} \end{array} \qquad \gamma = \left[\begin{array}{c} \gamma_0 \\ \gamma_1 \\ \hline \gamma_2 \\ \gamma_3 \end{array}\right]$$

AMS 256
Chapter 3: Estimability and Least Squares Estimators

Spring 2016

† The form of linear models is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

- $\mathbf{y}$: $n \times 1$ vector of observations (random)
- $\mathbf{X}$: $n \times p$ matrix of known constants (*design* matrix) with $r(\mathbf{X}) = r$
- $\boldsymbol{\beta}$: $p \times 1$ vector of unobservable parameters
- $\mathbf{e}$: $n \times 1$ vector of unobservable random errors

$E(y_i) = x_i^T \beta$

★ Assumptions:

- $\mathrm{E}(\mathbf{e}) = \mathbf{0}$ ($\Leftrightarrow \mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$)
- $\mathrm{Cov}(\mathbf{e}) = \sigma^2 I$ where σ^2 is some unknown parameter ($\sigma^2 > 0$)

$X: n \times p$

- Recall the NEs are $\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y}$.
- $r(\mathbf{X}) = p$ (full rank model)
  $\Rightarrow \hat{\boldsymbol{\beta}}$ is unique.
  $\Rightarrow$ We can estimate any function of $\boldsymbol{\beta}$.
- $r(\mathbf{X}) < p$ (overparameterized)
  $\Rightarrow$ multiple solutions to NEs
  $\Rightarrow$ cannot estimate all function of $\boldsymbol{\beta}$.

† Question: Which function of $\boldsymbol{\beta}$ can and cannot be estimated?

† Def: [Identifiability] The parameterization $\boldsymbol{\beta}$ is *identifiable* if for any $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, $\underbrace{\mathbf{X}\boldsymbol{\beta}_1}_{E(y)} = \mathbf{X}\boldsymbol{\beta}_2$ (two $\boldsymbol{\beta}$ give the same mean for $\mathbf{y}$) implies $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.

In other words,

√ Knowing $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ means knowing $\boldsymbol{\beta}$.

√ $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2 \Rightarrow \mathbf{X}\boldsymbol{\beta}_1 \neq \mathbf{X}\boldsymbol{\beta}_2$.

√ A difference in the parameter values ⇒ difference in the means.

$E(e) = 0$

√ In the linear model, $\mathbf{y}$ depends on $\boldsymbol{\beta}$ only through $\mathbf{X}\boldsymbol{\beta}$. So, if two parameter vectors cannot be distinguished, they lead to the same $\mathbf{X}\boldsymbol{\beta}$ for $\mathbf{y}$.

♣ In short, a parameterization is identifiable if knowing $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ tells us the parameterization vector $\boldsymbol{\beta}$.

♠ Ex2: Consider the one-way ANOVA model.

$$y_{ij} = \mu + \alpha_i + e_{ij}, \qquad i = 1, \cancel{2} \ldots, a; \; j = 1, 2, 3.$$

- Consider two $\boldsymbol{\beta}$

$$\boldsymbol{\beta}_1 = \begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_a \end{bmatrix} \qquad vs. \qquad \boldsymbol{\beta}_2 = \begin{bmatrix} \mu + c \\ \alpha_1 - c \\ \vdots \\ \alpha_a - c \end{bmatrix}$$

for any arbitrary c.

- $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$, BUT $\mathbf{X}\boldsymbol{\beta}_1 = \mathbf{X}\boldsymbol{\beta}_2$.
- $\Rightarrow \boldsymbol{\beta}$ is *not* identifiable!

$\downarrow^{n\times p}$

♠ Ex1: Consider a regression model (that is, model with $r(\mathbf{X}) = p$).

- $(\mathbf{X}^T\mathbf{X})$ is nonsingular $r(X) = r(X^TX)$
- For $\mathbf{X}\boldsymbol{\beta}_1 = \mathbf{X}\boldsymbol{\beta}_2$, then

$$\boldsymbol{\beta}_1 = \underbrace{(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{X})}_{=I}\boldsymbol{\beta}_1 = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{X})\boldsymbol{\beta}_2 = \boldsymbol{\beta}_2$$

- $\Rightarrow$ the parameters are identifiable!

♠ Ex 1 (contd): Consider a regression model in the following cases.

- A person's weight is measured both in pounds and kilos and both variables are entered into the model.
- For each individual, we record the number of years of preuniversity education, the number of years of university education and also the total number of years of education and put all the three variables into the model.

$\Rightarrow$ Columns in $\mathbf{X}$ are linearly dependent

$\Rightarrow$ $\mathbf{X}$ is not of full rank.

† Def: [Identifiability – contd] A vector-valued function $g(\boldsymbol{\beta})$ is *identifiable* if $\mathbf{X}\boldsymbol{\beta}_1 = \mathbf{X}\boldsymbol{\beta}_2$ implies $g(\boldsymbol{\beta}_1) = g(\boldsymbol{\beta}_2)$.

$p \times 1$

♣ *In specific,* focus on linear functions, $g(\boldsymbol{\beta}) = \boldsymbol{\lambda}^T\boldsymbol{\beta}$. Whether a linear function $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is identified depends on whether $\mathbf{X}\boldsymbol{\beta}_1 = \mathbf{X}\boldsymbol{\beta}_2$ implies $\boldsymbol{\lambda}^T\boldsymbol{\beta}_1 = \boldsymbol{\lambda}^T\boldsymbol{\beta}_2$.

$\lambda : p \times 1$

† Th: A function $g(\boldsymbol{\beta})$ is *identifiable* if and only if $g(\boldsymbol{\beta})$ is a function of $\mathbf{X}\boldsymbol{\beta}$.

$= E(y)$

⇒ A linear function $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is identifiable if $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is a function of $\mathbf{X}\boldsymbol{\beta}$.

† Question: Which $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is reasonable to estimate?

Obvious answer: *identifiable* functions!

$$\Leftrightarrow \boldsymbol{\lambda}^T\boldsymbol{\beta} \text{ is a function of } \mathbf{X}\boldsymbol{\beta}$$

† Def 3.1: An estimator $\boxed{t(\mathbf{y})}$ is an unbiased estimator for the scalar $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ iff $\mathrm{E}(t(\mathbf{y})) = \boldsymbol{\lambda}^T\boldsymbol{\beta}$ for all $\boldsymbol{\beta}$.

† Def 3.2: An estimator $t(\mathbf{y})$ is a linear estimator in $\mathbf{y}$ iff $t(\mathbf{y}) = c + \mathbf{a}^T\mathbf{y}$ for constants c, $a_1, \ldots, a_n$.

† Def 3.3: A function $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is linearly estimable iff there exists a linear unbiased estimator for it. If no such estimator exists then the function is called nonestimable.

$t(y) = c + a^T y$: linear

$E(t(y)) = \lambda^T\beta$: unbiased

$E(t(y)) = E(c + a^T y)$

$= \underline{c + a^T X\beta = \lambda^T\beta}$ for any β

$\Leftrightarrow$ $c = 0$ & $a^T X = \lambda^T$

$t(y) = a^T y$

w/ $a^T X = \lambda^T$

† Result 3.1: Under the linear mean model, $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is (linearly) estimable iff there exists a vector $\mathbf{a}$ such that the expectation of the linear combination $\mathbf{a}^T\mathbf{y} = a_1y_1 + \ldots + a_ny_n$ is a linear parametric function $\boldsymbol{\lambda}^T\boldsymbol{\beta}$, that is, for all $\boldsymbol{\beta}$

$$\exists \mathbf{a} \quad \text{s.t} \quad \mathrm{E}(\mathbf{a}^T\mathbf{y}) = \boldsymbol{\lambda}^T\boldsymbol{\beta}.$$

$a^T X = \lambda^T$

$t(y) = a^T y$

$$\Leftrightarrow \qquad \boldsymbol{\lambda}^T = \mathbf{a}^T\mathbf{X} \text{ or } \boldsymbol{\lambda} = \mathbf{X}^T\mathbf{a}.$$

$$\Leftrightarrow \qquad \boldsymbol{\lambda} \in C(\mathbf{X}^T)$$

$\lambda^T\beta$ estimable?

$\Rightarrow$ $\lambda \in C(X^T)$ Y[illegible] $\rightarrow$ est

$\lambda \notin C(X^T)$ [illegible] $\rightarrow$ NOT est

$y = X\beta + e$

$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$ 6x4 $\quad \beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \quad X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

③

basis for $\mathcal{C}(X^T) = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$

basis for $\mathcal{N}(X) = \left\{ \begin{bmatrix} 1 \\ -1 \\ -1 \\ -1 \end{bmatrix} \right\}$

♠ Ex2 (Contd): Consider the one-way ANOVA model.

$$y_{ij} = \mu + \alpha_i + e_{ij},$$

where $i = 1, 2, 3;\ j = 1, \ldots, n_i;\ (n_1, n_2, n_3) = (3, 2, 1)$. Can we estimate the followings?

1. α_1
2. $\mu + \alpha_1$
3. $\alpha_1 - \alpha_3$ $\quad \lambda = [0\ 1\ 0\ -1]^T$
4. $\alpha_1 + \alpha_2 - 2\alpha_3$

$\lambda = [0\ 1\ 1\ -2]^T$

$E(y_{ij})$

$\alpha_1 = \underbrace{[0\ 1\ 0\ 0]}_{\lambda^T} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$ $\quad \lambda \in \mathcal{C}(X^T)$??

$\Leftrightarrow \quad d_1 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + d_2 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + d_3 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$

? $\exists\ (d_1, d_2, d_3)$

No $\Rightarrow$ α_1 is NOT estimable

• Check Section 3.4 of M for more discussion on one-way ANOVA!

#2. $\mu + \alpha_1 = \underbrace{[1\ 1\ 0\ 0]}_{= \lambda^T} \beta$

$\boxed{\exists\, a \text{ s.t. } \lambda = X^T a}$

#3

#3 $\begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}$

$$\begin{cases} \mu+\alpha_1 = E(y_{1j}) \\ \mu+\alpha_2 = E(y_{2j}) \\ \mu+\alpha_3 \end{cases}$$

$\underline{d_1}$

$E(y_{1j}) = \mu + \alpha_1$

$$(\mu+\alpha_1) - (\mu+\alpha_2) = \alpha_1 - \alpha_2$$

$$(\mu+\alpha_1) + (\mu+\alpha_2) - 2(\mu+\alpha_3) = \alpha_1+\alpha_2-2\alpha_3$$

♣ How to know which $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is estimable?

• Method 1 Linear combinations of expected values of observations are estimable. If we can express $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ as a linear combination of $E(y_i)$, then $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is estimable.

is estimable.

• Method 2 If $\boldsymbol{\lambda} \in C(\mathbf{X}^T)$ then $\boldsymbol{\lambda}^T\boldsymbol{\beta}$. So construct a set of basis vectors for $C(\mathbf{X}^T)$, say $\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(r)}\}$, and find constants d_j so that $\boldsymbol{\lambda} = \sum_j d_j \mathbf{v}^{(j)}$.

$\Rightarrow$ In words, if two linear functions are estimable, then any linear combination of them is estimable.

• Method 3 Note that $\boldsymbol{\lambda} \in C(\mathbf{X}^T)$ iff $\boldsymbol{\lambda} \perp \mathcal{N}(\mathbf{X})$. So find a basis for $\mathcal{N}(\mathbf{X})$, say $\{\mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \ldots, \mathbf{c}^{(p-r)}\}$. Then $\boldsymbol{\lambda} \perp \mathbf{c}^{(j)}$ for all $j = 1, \ldots, p-r$, then $\boldsymbol{\lambda} \in C(\mathbf{X}^T)$ and $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is estimable.

$X : n \times p$

$r(X) = r$

♣ How to construct linear unbiased estimators of estimable functions, $\boldsymbol{\lambda}^T\boldsymbol{\beta}$?

• Recall $\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{P}\mathbf{y}$ where $\hat{\boldsymbol{\beta}}$ a LSE of $\boldsymbol{\beta}$ and $\mathbf{P}$ the perpendicular projection operator onto $C(\mathbf{X})$.

† Def 3.4: The least squares estimator of an estimable function $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is $\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}$ where $\hat{\boldsymbol{\beta}}$ is a solution to the NEs.

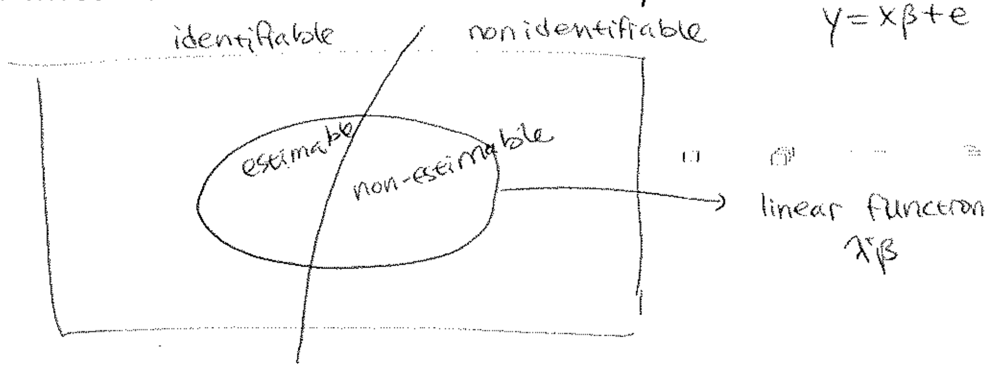
★ Result 3.2: If $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is estimable, then the least squares estimator $\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}$ is the same for all solutions $\hat{\boldsymbol{\beta}}$ to the NEs (i.e. the unique LSE of $\boldsymbol{\lambda}^T\boldsymbol{\beta}$).

$\lambda^T\beta$ is estimable $\Rightarrow$ $\lambda^T\beta$ is identifiable

non-identifiable $\Rightarrow$ non estimable

★ Result: The least squares estimator $\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}$ of an estimable function of $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ is a linear unbiased estimator of $\boldsymbol{\lambda}^T\boldsymbol{\beta}$.

$y = x\beta + e$



$E(y_{ij}) = (\theta_1) \overset{equi}{=} (\mu + \alpha_1)$

$\hat{\theta}_1 = \hat{\mu} + \hat{\alpha}_1$

† Reparameterization

$\hat{y}$ $\lambda^T\beta$

$\hat{e}$

model 1: $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$ where $\mathbf{X}$: $n \times p$
model 2: $\mathbf{y} = \mathbf{W}\gamma + \mathbf{e}$ where $\mathbf{W}$: $n \times t$
where $\mathbf{W} = \mathbf{X}\mathbf{T}$ and $\mathbf{X} = \mathbf{W}\mathbf{S}$.

• Consider two models with different design matrices and different parameters.

• The two models are equivalent if they will give the same least square fit of the data, $\hat{\mathbf{y}}$ ($\Leftrightarrow C(\mathbf{X}) = C(\mathbf{W})$).

One line summary! A function of parameters in one model is estimable $\Rightarrow$ Its equivalent function under the other model is also estimable.

• Check the details in Section 3.7

$r(X) < p$ + conditions $\Rightarrow$ unique solution of β

† Imposing conditions for a unique solution to the NEs.

$(X^TX)\beta = X^Ty$

♠ Ex2 (Contd): Consider the one-way ANOVA model.

$$y_{ij} = \mu + \alpha_i + e_{ij},$$

where $i = 1, 2, 3;\ j = 1, 2.$

$y = X\beta + e$

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ \vdots \\ e_{32} \end{bmatrix}$$

6×4

$r(X) = 3$

$$\underset{4\times4}{X^TX} = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix} \qquad X^Ty = \begin{bmatrix} \sum y_{ij} \\ \sum y_{1j} \\ \sum y_{2j} \\ \sum y_{3j} \end{bmatrix} = \begin{bmatrix} y_{\cdot\cdot} \\ y_{1\cdot} \\ y_{2\cdot} \\ y_{3\cdot} \end{bmatrix}$$

$$\begin{aligned} 6\mu + 2\alpha_1 + 2\alpha_2 + 2\alpha_3 &= y_{\cdot\cdot} \\ 2\mu + 2\alpha_1 &= y_{1\cdot} \\ 2\mu + 2\alpha_2 &= y_{2\cdot} \\ 2\mu + 2\alpha_3 &= y_{3\cdot} \end{aligned}$$

① $\sum n_i \alpha_i = 0$ ($\Rightarrow$ balanced $\sum \alpha_i = 0$)

$0 + \alpha_1 + \alpha_2 + \alpha_3 = 0$

$6\mu = y_{\cdot\cdot}$

$\Rightarrow \hat{\mu} = \bar{y}_{\cdot\cdot} = \frac{y_{\cdot\cdot}}{6}$

$\Rightarrow \hat{\alpha}_i = \frac{y_{i\cdot}}{2} - \hat{\mu}$

$= \bar{y}_{i\cdot} - \hat{\mu}$

② $\alpha_1 = 0$

$0 + \alpha_1 + 0 + 0 = 0$

$\hat{\mu} = \bar{y}_{1\cdot}$

$\hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{y}_{1\cdot}, \quad i = 2, 3$

$= \bar{y}_{i\cdot} - \hat{\mu}$