

# Stat 709: Mathematical Statistics

## Lecture 35

Jun Shao

Department of Statistics  
University of Wisconsin  
Madison, WI 53706, USA

# Lecture 35: The LSE and estimability

## Linear Models

One of the most useful statistical models

$$X_i = \beta^\tau Z_i + \varepsilon_i, \quad i = 1, \dots, n,$$

where  $X_i$  is the  $i$ th observation and is often called the  $i$ th response;  
 $\beta$  is a  $p$ -vector of unknown parameters (main parameters of interest),  
 $p < n$ ;

$Z_i$  is the  $i$ th value of a  $p$ -vector of explanatory variables (or covariates);  
 $\varepsilon_1, \dots, \varepsilon_n$  are random errors (not observed).

Data:  $(X_1, Z_1), \dots, (X_n, Z_n)$ .

$Z_i$ 's are nonrandom or given values of a random  $p$ -vector, in which case our analysis is conditioned on  $Z_1, \dots, Z_n$ .

$X = (X_1, \dots, X_n)$ ,  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$   $Z =$  the  $n \times p$  matrix whose  $i$ th row is the vector  $Z_i$ ,  $i = 1, \dots, n$

A matrix form of the model is

$$X = Z\beta + \varepsilon. \tag{1}$$

# Lecture 35: The LSE and estimability

## Linear Models

One of the most useful statistical models

$$X_i = \beta^\tau Z_i + \varepsilon_i, \quad i = 1, \dots, n,$$

where  $X_i$  is the  $i$ th observation and is often called the  $i$ th response;  $\beta$  is a  $p$ -vector of unknown parameters (main parameters of interest),  $p < n$ ;

$Z_i$  is the  $i$ th value of a  $p$ -vector of explanatory variables (or covariates);  $\varepsilon_1, \dots, \varepsilon_n$  are random errors (not observed).

Data:  $(X_1, Z_1), \dots, (X_n, Z_n)$ .

$Z_i$ 's are nonrandom or given values of a random  $p$ -vector, in which case our analysis is conditioned on  $Z_1, \dots, Z_n$ .

$X = (X_1, \dots, X_n)$ ,  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$   $Z =$  the  $n \times p$  matrix whose  $i$ th row is the vector  $Z_i$ ,  $i = 1, \dots, n$

A matrix form of the model is

$$X = Z\beta + \varepsilon. \quad (1)$$

# Lecture 35: The LSE and estimability

## Linear Models

One of the most useful statistical models

$$X_i = \beta^\tau Z_i + \varepsilon_i, \quad i = 1, \dots, n,$$

where  $X_i$  is the  $i$ th observation and is often called the  $i$ th response;  $\beta$  is a  $p$ -vector of unknown parameters (main parameters of interest),  $p < n$ ;

$Z_i$  is the  $i$ th value of a  $p$ -vector of explanatory variables (or covariates);  $\varepsilon_1, \dots, \varepsilon_n$  are random errors (not observed).

Data:  $(X_1, Z_1), \dots, (X_n, Z_n)$ .

$Z_i$ 's are nonrandom or given values of a random  $p$ -vector, in which case our analysis is conditioned on  $Z_1, \dots, Z_n$ .

$X = (X_1, \dots, X_n)$ ,  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$   $Z =$  the  $n \times p$  matrix whose  $i$ th row is the vector  $Z_i$ ,  $i = 1, \dots, n$

A matrix form of the model is

$$X = Z\beta + \varepsilon. \tag{1}$$

## Definition 3.4

Suppose that the range of  $\beta$  in model (1) is  $B \subset \mathcal{R}^p$ .

A *least squares estimator* (LSE) of  $\beta$  is defined to be any  $\hat{\beta} \in B$  such that

$$\|X - Z\hat{\beta}\|^2 = \min_{b \in B} \|X - Zb\|^2.$$

For any  $l \in \mathcal{R}^p$ ,  $l^\tau \hat{\beta}$  is called an LSE of  $l^\tau \beta$ .

Throughout this book, we consider  $B = \mathcal{R}^p$  unless otherwise stated. Differentiating  $\|X - Zb\|^2$  w.r.t.  $b$ , we obtain that any solution of

$$Z^\tau Zb = Z^\tau X$$

is an LSE of  $\beta$ .

## Full rank $Z$

If the rank of the matrix  $Z$  is  $p$ , in which case  $(Z^\tau Z)^{-1}$  exists and  $Z$  is said to be of full rank, then there is a unique LSE, which is

$$\hat{\beta} = (Z^\tau Z)^{-1} Z^\tau X.$$

### Definition 3.4

Suppose that the range of  $\beta$  in model (1) is  $B \subset \mathcal{R}^p$ .

A *least squares estimator* (LSE) of  $\beta$  is defined to be any  $\hat{\beta} \in B$  such that

$$\|X - Z\hat{\beta}\|^2 = \min_{b \in B} \|X - Zb\|^2.$$

For any  $l \in \mathcal{R}^p$ ,  $l^\tau \hat{\beta}$  is called an LSE of  $l^\tau \beta$ .

Throughout this book, we consider  $B = \mathcal{R}^p$  unless otherwise stated.

Differentiating  $\|X - Zb\|^2$  w.r.t.  $b$ , we obtain that any solution of

$$Z^\tau Zb = Z^\tau X$$

is an LSE of  $\beta$ .

### Full rank $Z$

If the rank of the matrix  $Z$  is  $p$ , in which case  $(Z^\tau Z)^{-1}$  exists and  $Z$  is said to be of full rank, then there is a unique LSE, which is

$$\hat{\beta} = (Z^\tau Z)^{-1} Z^\tau X.$$

### Definition 3.4

Suppose that the range of  $\beta$  in model (1) is  $B \subset \mathcal{R}^p$ .

A *least squares estimator* (LSE) of  $\beta$  is defined to be any  $\hat{\beta} \in B$  such that

$$\|X - Z\hat{\beta}\|^2 = \min_{b \in B} \|X - Zb\|^2.$$

For any  $l \in \mathcal{R}^p$ ,  $l^\tau \hat{\beta}$  is called an LSE of  $l^\tau \beta$ .

Throughout this book, we consider  $B = \mathcal{R}^p$  unless otherwise stated.

Differentiating  $\|X - Zb\|^2$  w.r.t.  $b$ , we obtain that any solution of

$$Z^\tau Zb = Z^\tau X$$

is an LSE of  $\beta$ .

### Full rank $Z$

If the rank of the matrix  $Z$  is  $p$ , in which case  $(Z^\tau Z)^{-1}$  exists and  $Z$  is said to be of full rank, then there is a unique LSE, which is

$$\hat{\beta} = (Z^\tau Z)^{-1} Z^\tau X.$$

## Non full rank $Z$

If  $Z$  is not of full rank, then there are infinitely many LSE's of  $\beta$ .  
Any LSE of  $\beta$  is of the form

$$\hat{\beta} = (Z^T Z)^- Z^T X,$$

where  $(Z^T Z)^-$  is called a *generalized inverse* of  $Z^T Z$  and satisfies

$$Z^T Z (Z^T Z)^- Z^T Z = Z^T Z.$$

Generalized inverse matrices are not unique unless  $Z$  is of full rank, in which case  $(Z^T Z)^- = (Z^T Z)^{-1}$

## Assumptions

To study properties of LSE's of  $\beta$ , we need some assumptions on the distribution of  $X$  or  $\varepsilon$  (conditional on  $Z$  if  $Z$  is random).

A1:  $\varepsilon$  is distributed as  $N_n(0, \sigma^2 I_n)$  with an unknown  $\sigma^2 > 0$ .

A2:  $E(\varepsilon) = 0$  and  $\text{Var}(\varepsilon) = \sigma^2 I_n$  with an unknown  $\sigma^2 > 0$ .

A3:  $E(\varepsilon) = 0$  and  $\text{Var}(\varepsilon)$  is an unknown matrix.

## Non full rank $Z$

If  $Z$  is not of full rank, then there are infinitely many LSE's of  $\beta$ . Any LSE of  $\beta$  is of the form

$$\hat{\beta} = (Z^T Z)^- Z^T X,$$

where  $(Z^T Z)^-$  is called a *generalized inverse* of  $Z^T Z$  and satisfies

$$Z^T Z (Z^T Z)^- Z^T Z = Z^T Z.$$

Generalized inverse matrices are not unique unless  $Z$  is of full rank, in which case  $(Z^T Z)^- = (Z^T Z)^{-1}$

## Assumptions

To study properties of LSE's of  $\beta$ , we need some assumptions on the distribution of  $X$  or  $\varepsilon$  (conditional on  $Z$  if  $Z$  is random).

**A1:**  $\varepsilon$  is distributed as  $N_n(0, \sigma^2 I_n)$  with an unknown  $\sigma^2 > 0$ .

**A2:**  $E(\varepsilon) = 0$  and  $\text{Var}(\varepsilon) = \sigma^2 I_n$  with an unknown  $\sigma^2 > 0$ .

**A3:**  $E(\varepsilon) = 0$  and  $\text{Var}(\varepsilon)$  is an unknown matrix.

## Remarks

- Assumption A1 is the strongest and implies a parametric model.
- We may assume a slightly more general assumption that  $\varepsilon$  has the  $N_n(0, \sigma^2 D)$  distribution with unknown  $\sigma^2$  but a known positive definite matrix  $D$ .

Let  $D^{-1/2}$  be the inverse of the square root matrix of  $D$ .

Then model (1) with assumption A1 holds if we replace  $X$ ,  $Z$ , and  $\varepsilon$  by the transformed variables  $\tilde{X} = D^{-1/2}X$ ,  $\tilde{Z} = D^{-1/2}Z$ , and  $\tilde{\varepsilon} = D^{-1/2}\varepsilon$ , respectively.

- A similar conclusion can be made for assumption A2.
- Under assumption A1, the distribution of  $X$  is  $N_n(Z\beta, \sigma^2 I_n)$ , which is in an exponential family  $\mathcal{P}$  with parameter  $\theta = (\beta, \sigma^2) \in \mathcal{R}^p \times (0, \infty)$ .
- However, if the matrix  $Z$  is not of full rank, then  $\mathcal{P}$  is not identifiable (see §2.1.2), since  $Z\beta_1 = Z\beta_2$  does not imply  $\beta_1 = \beta_2$ .

## Remarks

- Suppose that the rank of  $Z$  is  $r \leq p$ .  
Then there is an  $n \times r$  submatrix  $Z_*$  of  $Z$  such that

$$Z = Z_* Q \quad (2)$$

and  $Z_*$  is of rank  $r$ , where  $Q$  is a fixed  $r \times p$  matrix, and

$$Z\beta = Z_* Q\beta.$$

- $\mathcal{P}$  is identifiable if we consider the reparameterization  $\tilde{\beta} = Q\beta$ .
- The new parameter  $\tilde{\beta}$  is in a subspace of  $\mathcal{R}^p$  with dimension  $r$ .
- In many applications, we are interested in estimating some linear functions of  $\beta$ , i.e.,  $\vartheta = l^\tau \beta$  for some  $l \in \mathcal{R}^p$ .
- From the previous discussion, however, estimation of  $l^\tau \beta$  is meaningless unless  $l = Q^\tau c$  for some  $c \in \mathcal{R}^r$  so that

$$l^\tau \beta = c^\tau Q\beta = c^\tau \tilde{\beta}.$$

The following result shows that  $I^\tau \beta$  is estimable if  $I = Q^\tau c$ , which is also necessary for  $I^\tau \beta$  to be estimable under assumption A1.

### Theorem 3.6

Assume model (1) with assumption A3.

- (i) A necessary and sufficient condition for  $I \in \mathcal{R}^p$  being  $Q^\tau c$  for some  $c \in \mathcal{R}^r$  is  $I \in \mathcal{R}(Z) = \mathcal{R}(Z^\tau Z)$ , where  $Q$  is given by (2) and  $\mathcal{R}(A)$  is the smallest linear subspace containing all rows of  $A$ .
- (ii) If  $I \in \mathcal{R}(Z)$ , then the LSE  $I^\tau \hat{\beta}$  is unique and unbiased for  $I^\tau \beta$ .
- (iii) If  $I \notin \mathcal{R}(Z)$  and assumption A1 holds, then  $I^\tau \beta$  is not estimable.

### Proof

(i) Note that  $a \in \mathcal{R}(A)$  iff  $a = A^\tau b$  for some vector  $b$ .

If  $I = Q^\tau c$ , then

$$I = Q^\tau c = Q^\tau Z_*^\tau Z_* (Z_*^\tau Z_*)^{-1} c = Z_*^\tau [Z_* (Z_*^\tau Z_*)^{-1} c].$$

Hence  $I \in \mathcal{R}(Z)$ .

The following result shows that  $I^\tau \beta$  is estimable if  $I = Q^\tau c$ , which is also necessary for  $I^\tau \beta$  to be estimable under assumption A1.

### Theorem 3.6

Assume model (1) with assumption A3.

- (i) A necessary and sufficient condition for  $I \in \mathcal{R}^p$  being  $Q^\tau c$  for some  $c \in \mathcal{R}^r$  is  $I \in \mathcal{R}(Z) = \mathcal{R}(Z^\tau Z)$ , where  $Q$  is given by (2) and  $\mathcal{R}(A)$  is the smallest linear subspace containing all rows of  $A$ .
- (ii) If  $I \in \mathcal{R}(Z)$ , then the LSE  $I^\tau \hat{\beta}$  is unique and unbiased for  $I^\tau \beta$ .
- (iii) If  $I \notin \mathcal{R}(Z)$  and assumption A1 holds, then  $I^\tau \beta$  is not estimable.

### Proof

(i) Note that  $a \in \mathcal{R}(A)$  iff  $a = A^\tau b$  for some vector  $b$ .

If  $I = Q^\tau c$ , then

$$I = Q^\tau c = Q^\tau Z_*^\tau Z_* (Z_*^\tau Z_*)^{-1} c = Z_*^\tau [Z_* (Z_*^\tau Z_*)^{-1} c].$$

Hence  $I \in \mathcal{R}(Z)$ .

The following result shows that  $I^\tau \beta$  is estimable if  $I = Q^\tau c$ , which is also necessary for  $I^\tau \beta$  to be estimable under assumption A1.

### Theorem 3.6

Assume model (1) with assumption A3.

- (i) A necessary and sufficient condition for  $I \in \mathcal{R}^p$  being  $Q^\tau c$  for some  $c \in \mathcal{R}^r$  is  $I \in \mathcal{R}(Z) = \mathcal{R}(Z^\tau Z)$ , where  $Q$  is given by (2) and  $\mathcal{R}(A)$  is the smallest linear subspace containing all rows of  $A$ .
- (ii) If  $I \in \mathcal{R}(Z)$ , then the LSE  $I^\tau \hat{\beta}$  is unique and unbiased for  $I^\tau \beta$ .
- (iii) If  $I \notin \mathcal{R}(Z)$  and assumption A1 holds, then  $I^\tau \beta$  is not estimable.

### Proof

(i) Note that  $a \in \mathcal{R}(A)$  iff  $a = A^\tau b$  for some vector  $b$ .

If  $I = Q^\tau c$ , then

$$I = Q^\tau c = Q^\tau Z_*^\tau Z_* (Z_*^\tau Z_*)^{-1} c = Z^\tau [Z_* (Z_*^\tau Z_*)^{-1} c].$$

Hence  $I \in \mathcal{R}(Z)$ .

## Proof (continued)

If  $I \in \mathcal{R}(Z)$ , then  $I = Z^\tau \zeta$  for some  $\zeta$  and

$$I = (Z_* Q)^\tau \zeta = Q^\tau c, \quad c = Z_*^\tau \zeta.$$

(ii) If  $I \in \mathcal{R}(Z) = \mathcal{R}(Z^\tau Z)$ , then  $I = Z^\tau Z \zeta$  for some  $\zeta$  and by

$$\widehat{\beta} = (Z^\tau Z)^{-1} Z^\tau X,$$

$$E(I^\tau \widehat{\beta}) = E[I^\tau (Z^\tau Z)^{-1} Z^\tau X] = \zeta^\tau Z^\tau Z (Z^\tau Z)^{-1} Z^\tau Z \beta = \zeta^\tau Z^\tau Z \beta = I^\tau \beta.$$

If  $\bar{\beta}$  is any other LSE of  $\beta$ , then, by  $Z^\tau Z \bar{\beta} = Z^\tau X$ ,

$$I^\tau \widehat{\beta} - I^\tau \bar{\beta} = \zeta^\tau (Z^\tau Z) (\widehat{\beta} - \bar{\beta}) = \zeta^\tau (Z^\tau X - Z^\tau X) = 0.$$

(iii) Under A1, if there is an estimator  $h(X, Z)$  unbiased for  $I^\tau \beta$ , then

$$I^\tau \beta = \int_{\mathcal{R}^n} h(x, Z) (2\pi)^{-n/2} \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \|x - Z\beta\|^2 \right\} dx.$$

Differentiating w.r.t.  $\beta$  and applying Theorem 2.1 lead to

$$I^\tau = Z^\tau \int_{\mathcal{R}^n} h(x, Z) (2\pi)^{-n/2} \sigma^{-n-2} (x - Z\beta) \exp \left\{ -\frac{1}{2\sigma^2} \|x - Z\beta\|^2 \right\} dx,$$

which implies  $I \in \mathcal{R}(Z)$ .

## Proof (continued)

If  $I \in \mathcal{R}(Z)$ , then  $I = Z^\tau \zeta$  for some  $\zeta$  and

$$I = (Z_* Q)^\tau \zeta = Q^\tau c, \quad c = Z_*^\tau \zeta.$$

(ii) If  $I \in \mathcal{R}(Z) = \mathcal{R}(Z^\tau Z)$ , then  $I = Z^\tau Z \zeta$  for some  $\zeta$  and by

$$\widehat{\beta} = (Z^\tau Z)^{-1} Z^\tau X,$$

$$E(I^\tau \widehat{\beta}) = E[I^\tau (Z^\tau Z)^{-1} Z^\tau X] = \zeta^\tau Z^\tau Z (Z^\tau Z)^{-1} Z^\tau Z \beta = \zeta^\tau Z^\tau Z \beta = I^\tau \beta.$$

If  $\bar{\beta}$  is any other LSE of  $\beta$ , then, by  $Z^\tau Z \bar{\beta} = Z^\tau X$ ,

$$I^\tau \widehat{\beta} - I^\tau \bar{\beta} = \zeta^\tau (Z^\tau Z) (\widehat{\beta} - \bar{\beta}) = \zeta^\tau (Z^\tau X - Z^\tau X) = 0.$$

(iii) Under A1, if there is an estimator  $h(X, Z)$  unbiased for  $I^\tau \beta$ , then

$$I^\tau \beta = \int_{\mathcal{R}^n} h(x, Z) (2\pi)^{-n/2} \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \|x - Z\beta\|^2 \right\} dx.$$

Differentiating w.r.t.  $\beta$  and applying Theorem 2.1 lead to

$$I^\tau = Z^\tau \int_{\mathcal{R}^n} h(x, Z) (2\pi)^{-n/2} \sigma^{-n-2} (x - Z\beta) \exp \left\{ -\frac{1}{2\sigma^2} \|x - Z\beta\|^2 \right\} dx,$$

which implies  $I \in \mathcal{R}(Z)$ .

## Proof (continued)

If  $I \in \mathcal{R}(Z)$ , then  $I = Z^\tau \zeta$  for some  $\zeta$  and

$$I = (Z_* Q)^\tau \zeta = Q^\tau c, \quad c = Z_*^\tau \zeta.$$

(ii) If  $I \in \mathcal{R}(Z) = \mathcal{R}(Z^\tau Z)$ , then  $I = Z^\tau Z \zeta$  for some  $\zeta$  and by

$$\widehat{\beta} = (Z^\tau Z)^{-1} Z^\tau X,$$

$$E(I^\tau \widehat{\beta}) = E[I^\tau (Z^\tau Z)^{-1} Z^\tau X] = \zeta^\tau Z^\tau Z (Z^\tau Z)^{-1} Z^\tau Z \beta = \zeta^\tau Z^\tau Z \beta = I^\tau \beta.$$

If  $\bar{\beta}$  is any other LSE of  $\beta$ , then, by  $Z^\tau Z \bar{\beta} = Z^\tau X$ ,

$$I^\tau \widehat{\beta} - I^\tau \bar{\beta} = \zeta^\tau (Z^\tau Z) (\widehat{\beta} - \bar{\beta}) = \zeta^\tau (Z^\tau X - Z^\tau X) = 0.$$

(iii) Under A1, if there is an estimator  $h(X, Z)$  unbiased for  $I^\tau \beta$ , then

$$I^\tau \beta = \int_{\mathcal{R}^n} h(x, Z) (2\pi)^{-n/2} \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \|x - Z\beta\|^2 \right\} dx.$$

Differentiating w.r.t.  $\beta$  and applying Theorem 2.1 lead to

$$I^\tau = Z^\tau \int_{\mathcal{R}^n} h(x, Z) (2\pi)^{-n/2} \sigma^{-n-2} (x - Z\beta) \exp \left\{ -\frac{1}{2\sigma^2} \|x - Z\beta\|^2 \right\} dx,$$

which implies  $I \in \mathcal{R}(Z)$ .

### Example 3.12 (Simple linear regression)

Let  $\beta = (\beta_0, \beta_1) \in \mathcal{R}^2$  and  $Z_i = (1, t_i)$ ,  $t_i \in \mathcal{R}$ ,  $i = 1, \dots, n$ .

Then model (1) is called a *simple linear regression* model.

It turns out that

$$\begin{pmatrix} n & \sum_{i=1}^n t_i \\ \sum_{i=1}^n t_i & \sum_{i=1}^n t_i^2 \end{pmatrix}.$$

This matrix is invertible iff some  $t_i$ 's are different.

Thus, if some  $t_i$ 's are different, then the unique unbiased LSE of  $l^T \beta$  for any  $l \in \mathcal{R}^2$  is  $l^T (Z^T Z)^{-1} Z^T X$ , which has the normal distribution if assumption A1 holds.

The result can be easily extended to the case of *polynomial regression* of order  $p$  in which

$$\beta = (\beta_0, \beta_1, \dots, \beta_{p-1})$$

and

$$Z_i = (1, t_i, \dots, t_i^{p-1}).$$

### Example 3.12 (Simple linear regression)

Let  $\beta = (\beta_0, \beta_1) \in \mathcal{R}^2$  and  $Z_i = (1, t_i)$ ,  $t_i \in \mathcal{R}$ ,  $i = 1, \dots, n$ .

Then model (1) is called a *simple linear regression* model.

It turns out that

$$\begin{pmatrix} n & \sum_{i=1}^n t_i \\ \sum_{i=1}^n t_i & \sum_{i=1}^n t_i^2 \end{pmatrix}.$$

This matrix is invertible iff some  $t_i$ 's are different.

Thus, if some  $t_i$ 's are different, then the unique unbiased LSE of  $l^T \beta$  for any  $l \in \mathcal{R}^2$  is  $l^T (Z^T Z)^{-1} Z^T X$ , which has the normal distribution if assumption A1 holds.

The result can be easily extended to the case of *polynomial regression* of order  $p$  in which

$$\beta = (\beta_0, \beta_1, \dots, \beta_{p-1})$$

and

$$Z_i = (1, t_i, \dots, t_i^{p-1}).$$

### Example 3.13 (One-way ANOVA)

Suppose that  $n = \sum_{j=1}^m n_j$  with  $m$  positive integers  $n_1, \dots, n_m$  and that

$$X_i = \mu_j + \varepsilon_i, \quad i = k_{j-1} + 1, \dots, k_j, \quad j = 1, \dots, m,$$

where  $k_0 = 0$ ,  $k_j = \sum_{l=1}^j n_l$ ,  $j = 1, \dots, m$ , and  $(\mu_1, \dots, \mu_m) = \beta$ .

Let  $J_m$  be the  $m$ -vector of ones.

Then the matrix  $Z$  in this case is a block diagonal matrix with  $J_{n_j}$  as the  $j$ th diagonal column.

Consequently,  $Z^T Z$  is an  $m \times m$  diagonal matrix whose  $j$ th diagonal element is  $n_j$ .

Thus,  $Z^T Z$  is invertible and the unique LSE of  $\beta$  is the  $m$ -vector whose  $j$ th component is

$$\frac{1}{n_j} \sum_{i=k_{j-1}+1}^{k_j} X_i, \quad j = 1, \dots, m.$$

## Example 3.13 (continued)

Sometimes it is more convenient to use the following notation:

$$X_{ij} = X_{k_{i-1}+j}, \quad \varepsilon_{ij} = \varepsilon_{k_{i-1}+j}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, m,$$

and

$$\mu_i = \mu + \alpha_i, \quad i = 1, \dots, m.$$

Then our model becomes

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, m, \quad (3)$$

which is called a *one-way analysis of variance* (ANOVA) model.

Under model (3),  $\beta = (\mu, \alpha_1, \dots, \alpha_m) \in \mathcal{R}^{m+1}$ .

The matrix  $Z$  under model (3) is not of full rank.

An LSE of  $\beta$  under model (3) is

$$\hat{\beta} = (\bar{X}, \bar{X}_1 - \bar{X}, \dots, \bar{X}_m - \bar{X}),$$

where  $\bar{X}$  is still the sample mean of  $X_{ij}$ 's and  $\bar{X}_i$  is the sample mean of the  $i$ th group  $\{X_{ij}, j = 1, \dots, n_i\}$ .

## Example 3.13 (continued)

Sometimes it is more convenient to use the following notation:

$$X_{ij} = X_{k_{i-1}+j}, \quad \varepsilon_{ij} = \varepsilon_{k_{i-1}+j}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, m,$$

and

$$\mu_i = \mu + \alpha_i, \quad i = 1, \dots, m.$$

Then our model becomes

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, m, \quad (3)$$

which is called a *one-way analysis of variance* (ANOVA) model.

Under model (3),  $\beta = (\mu, \alpha_1, \dots, \alpha_m) \in \mathcal{R}^{m+1}$ .

The matrix  $Z$  under model (3) is not of full rank.

An LSE of  $\beta$  under model (3) is

$$\hat{\beta} = (\bar{X}, \bar{X}_1 - \bar{X}, \dots, \bar{X}_m - \bar{X}),$$

where  $\bar{X}$  is still the sample mean of  $X_{ij}$ 's and  $\bar{X}_i$  is the sample mean of the  $i$ th group  $\{X_{ij}, j = 1, \dots, n_i\}$ .

## Example 3.13 (continued)

Sometimes it is more convenient to use the following notation:

$$X_{ij} = X_{k_{i-1}+j}, \quad \varepsilon_{ij} = \varepsilon_{k_{i-1}+j}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, m,$$

and

$$\mu_i = \mu + \alpha_i, \quad i = 1, \dots, m.$$

Then our model becomes

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, m, \quad (3)$$

which is called a *one-way analysis of variance* (ANOVA) model.

Under model (3),  $\beta = (\mu, \alpha_1, \dots, \alpha_m) \in \mathcal{R}^{m+1}$ .

The matrix  $Z$  under model (3) is not of full rank.

An LSE of  $\beta$  under model (3) is

$$\hat{\beta} = (\bar{X}, \bar{X}_{1\cdot} - \bar{X}, \dots, \bar{X}_{m\cdot} - \bar{X}),$$

where  $\bar{X}$  is still the sample mean of  $X_{ij}$ 's and  $\bar{X}_{i\cdot}$  is the sample mean of the  $i$ th group  $\{X_{ij}, j = 1, \dots, n_i\}$ .

The notation used in model (3) allows us to generalize the one-way ANOVA model to any  $s$ -way ANOVA model with a positive integer  $s$  under the so-called factorial experiments.

### Example 3.14 (Two-way balanced ANOVA)

Suppose that

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \quad i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, c, \quad (4)$$

where  $a$ ,  $b$ , and  $c$  are some positive integers.

Model (4) is called a two-way balanced ANOVA model.

If we view model (4) as a special case of model (1), then the parameter vector  $\beta$  is

$$\beta = (\mu, \alpha_1, \dots, \alpha_a, \beta_1, \dots, \beta_b, \gamma_{11}, \dots, \gamma_{1b}, \dots, \gamma_{a1}, \dots, \gamma_{ab}). \quad (5)$$

One can obtain the matrix  $Z$  and show that it is  $n \times p$ , where  $n = abc$  and  $p = 1 + a + b + ab$ , and is of rank  $ab < p$ .

The notation used in model (3) allows us to generalize the one-way ANOVA model to any  $s$ -way ANOVA model with a positive integer  $s$  under the so-called factorial experiments.

### Example 3.14 (Two-way balanced ANOVA)

Suppose that

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \quad i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, c, \quad (4)$$

where  $a$ ,  $b$ , and  $c$  are some positive integers.

Model (4) is called a two-way balanced ANOVA model.

If we view model (4) as a special case of model (1), then the parameter vector  $\beta$  is

$$\beta = (\mu, \alpha_1, \dots, \alpha_a, \beta_1, \dots, \beta_b, \gamma_{11}, \dots, \gamma_{1b}, \dots, \gamma_{a1}, \dots, \gamma_{ab}). \quad (5)$$

One can obtain the matrix  $Z$  and show that it is  $n \times p$ , where  $n = abc$  and  $p = 1 + a + b + ab$ , and is of rank  $ab < p$ .

## Two-way balanced ANOVA

It can also be shown that an LSE of  $\beta$  is given by the right-hand side of (5) with  $\mu$ ,  $\alpha_i$ ,  $\beta_j$ , and  $\gamma_{ij}$  replaced by  $\hat{\mu}$ ,  $\hat{\alpha}_i$ ,  $\hat{\beta}_j$ , and  $\hat{\gamma}_{ij}$ , respectively, where

$$\hat{\mu} = \bar{X}_{...},$$

$$\hat{\alpha}_i = \bar{X}_{i..} - \bar{X}_{...},$$

$$\hat{\beta}_j = \bar{X}_{.j.} - \bar{X}_{...},$$

$$\hat{\gamma}_{ij} = \bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...},$$

and a dot is used to denote averaging over the indicated subscript, e.g.,

$$\bar{X}_{.j.} = \frac{1}{ac} \sum_{i=1}^a \sum_{k=1}^c X_{ijk}$$

with a fixed  $j$ .