Linear Model / Estimation / Least square estimation

Model 1 :

$$Y_{ij} = \mu + \mu_i + \varepsilon_{ij} \qquad 1 \le i \le k, \quad 1 \le j \le n_i$$

- $\mu, \mu_i \in \mathbb{R}$, $1 \le i \le k$
- $\varepsilon_{ij} \sim \text{Normal}(0, \sigma^2)$ & independent $1 \le i \le k$ $1 \le j \le n_i$

Population-1	Population 2	...	Population k
Y_{11}	Y_{21}		Y_{k1}
Y_{12}	⋮		Y_{k2}
⋮			⋮
Y_{1n_1}	Y_{2n_2}		Y_{kn_k}

$n = n_1 + n_2 + n_3 + \cdots + n_k$ - is total sample size.

Questions :-

(i) Can we estimate μ ? If yes then what is the estimate of μ?

(ii) Can we estimate $\mu + \mu_i$? If yes then what is the estimate of $\mu + \mu_i$?

(iii) Can we estimate $\mu_i - \mu_j$? If yes then what is the estimate of $\mu_i - \mu_j$?

$1 \leq i \neq j \leq k$

Within group variation

- Variance of group i can be estimated by

$$i^{th} \text{ Sample group variance} := \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2$$

} unbiased estimate

where $\bar{Y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$

- Sum of squares within:

$$SSW := \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2$$

Between group variation

- Sum of squares between

$$SSB = \sum_{i=1}^{k} n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

where $\bar{Y}_{..} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij}$.

Observations:

(1) $SSW + SSB = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\cdot})^2 + \sum_{i=1}^{k} n_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$

$= \sum_{i=1}^{k} \left[\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\cdot})^2 + (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 \right]$

$= \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{\cdot\cdot})^2$

$\equiv$ Total sum of squares

(2) (Ex.) SSW is independent of SSB.

(3) (Ex.)

$SSW \overset{d}{=} \chi^2_{k-1} \sigma^2$ - Degrees of freedom - $k-1$ for SSW

$SSB \overset{d}{=} \chi^2_{n-k} \sigma^2$ - Degrees of freedom - $n-k$ for SSB

$TSS \overset{d}{=} \chi^2_{n-1} \sigma^2$ - Degrees of freedom - $n-1$ for total

Parameter vector:

$$\beta = (\mu; \mu_1, \mu_2, \dots, \mu_k) \in \mathbb{R}^{k+1}$$

Design Matrix :-

$$X_{n \times k} \; ; \quad E[\underset{\sim}{Y}] = X\beta \quad \text{where}$$

$\underset{\sim}{Y}_{n \times 1}$ - data vector vector

$$= (Y_{11}, \dots, Y_{1n_1}; Y_{21} \dots, Y_{2n_2}; \dots ; Y_{k1} \dots Y_{kn_k})$$

$$\Rightarrow X = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \\ 1 & 0 & 1 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & \\ \vdots & & \vdots & & \end{pmatrix} \begin{matrix} \} n_1 \\ \\ \\ \} n_2 \\ \\ \\ \end{matrix}$$

So

Model

$$\underset{\sim}{Y} = X\beta + \underset{\sim}{\varepsilon}$$

" $Y_{ij} = \mu + \mu_i + \varepsilon_{ij}$ "

$1 \le i \le k$

$1 \le j \le n_i$

where $\underset{\sim}{\varepsilon} = (\varepsilon_{11}, \varepsilon_{12}, \dots \varepsilon_{1n_1}; \varepsilon_{21}, \dots, \varepsilon_{2n_2}; \dots \varepsilon_{kn_k})$

Fact:- . What is the rank of X?

Ex: $\text{rank}(X) = k$.

. How many parameters are there?

A: $k+1$ parameters

Question: Can we estimate all of them?

Answer: No!

Model 1: $Y_{ij} = \mu + \mu_i + \varepsilon_{ij}$ $\quad 1 \le i \le k$, $1 \le j \le n_i$

let. $\tilde{Y}_{ij} = \tilde{\mu} + \tilde{\mu}_i + \varepsilon_{ij}$

Model 2: $\tilde{\mu} = \mu + c$ $\quad 1 \le i \le k$, $1 \le j \le n_i$

$\tilde{\mu}_i = \mu_i - c$

Alas: $\tilde{Y}_{ij} = Y_{ij}$ $\therefore$ Models are same!

Problem: The parameters are not identifiable.

Towards solving / understanding the Problem:

(1) Is there a parameter in the model that is identifiable?

A: Yes! E.g.:- $\gamma_{12} = \mu_1 - \mu_2$

Ans: $\because$ $Y_{11} - Y_{21}$ - is an unbiased estimate.

(as $E[Y_{11} - Y_{12}] = \mu_1 - \mu_2$)

Is d optimal? What is an optimal Estimate?

(2) Suppose we assume $\sum_{i=1}^{k} \mu_i = 0$

$\Rightarrow$ Yay! Then all parameters are identifiable.

What is an optimal Estimate?

Note: In either (1) / (2) there are k - parameters (or their combinations) are identifiable.

Why k? : Because of $\text{rank}(X) = k$?

Theory of linear Models:

- Suppose we have n observations $\{y_1, y_2, \ldots, y_n\}$ which are realizations of random variables.
- let $(\beta_1, \beta_2, \ldots, \beta_m)$ be the parameters on which the distribution of the data depends.

Consider the following model:

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 \cdots + x_{im}\beta_m + \varepsilon_i \qquad 1 \le i \le n$$

where $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m)$ are uncorrelated error random variables. such that $E(\varepsilon_i) = 0 \quad 1 \le i \le n$

$\text{Var}(\varepsilon_i) = \sigma^2$

& (x_{ij}) are known constants.

Such a model is called a Linear Model.

Remark: (i) The model also can be described

$$E[Y_i] = \sum_{j=1}^{n} x_{ij}\beta_j \qquad 1 \le i \le n$$

& $\{Y_i\}_{1 \le i \le n}$ are uncorrelated random variables with variance $= \sigma^2$.

(ii) It is called linear as the expected value is linear in the parameters

Example:

- One way classification

$$Y_{ij} = \mu + \mu_i + \varepsilon_{ij} \qquad \begin{matrix} 1 \le i \le k \\ 1 \le j \le n_i \end{matrix}$$

- Nested classification

$$Y_{ijk} = \mu + \mu_i + \delta_{ij} + \varepsilon_{ijk} \qquad \begin{matrix} 1 \le k \le n_{ij} \\ 1 \le j \le n_i \\ 1 \le i \le k \end{matrix}$$

- Three way classification Model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

$1 \le k \le n_{ij}$

$1 \le j \le J$

$1 \le i \le I$

All models:

$$\underline{Y} = (y_1, y_2, \dots, y_n)^T$$

$$\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_m)^T$$

$$X = (x_{ij})_{\substack{1 \le i \le n \\ 1 \le j \le m}}$$

$$\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$$

Then

$$\underset{Y_{n\times 1}}{Y} = \underset{X_{n\times m}}{X}\ \underset{\beta_{m\times 1}}{\beta} + \underset{\varepsilon_{n\times 1}}{\varepsilon}$$

ε - variance - covariance matrix $\sigma^2 I_{n\times n}$.

Model: $(Y, X\beta, \sigma^2 I_{n\times n})$

- Y: Observation
- $X\beta$: $E[Y]$
- $\sigma^2 I_{n\times n}$: variance - covariance matrix of the data.

The Problem of Estimation of the Parameters

Suppose $(Y, X\beta, \sigma^2 I)$ be a linear model.

The parameters are $\beta_{m\times 1}$ and σ^2

i.e. (H) is $\mathbb{R}^m \times \mathbb{R}_+$

linear functions of the parameters

For any $p \in \mathbb{R}^m$, a linear combination

$$p_1\beta_1 + p_2\beta_2 + \cdots + p_m\beta_m := p^T\beta$$

of the parameter vector $\underline{\beta}$ is called a linear parametric function.

Note: In a Linear model we will be interested only in linear parametric functions.

Definition: Suppose $(Y, X\beta, \sigma^2 I)$ be a linear model and $p \in \mathbb{R}^m$.

The linear parametric function $p^T\beta$ is said to be estimable if

$\exists\ l \in \mathbb{R}^n$ s.t.

$$E[l^T Y] = p^T \beta.$$

(i.e. in other words $\exists\ l \in \mathbb{R}^n$ s.t. $l^T Y$ is unbiased estimate of $p^T \beta$)

Theorem: let $(Y, X\beta, \sigma^2 I_n)$ be a linear model & $p \in \mathbb{R}^n$.

Then $p^T \beta$ is estimable $\iff p \in \mathcal{C}(X^T)$

(where $\mathcal{C}(X^T) = \{z : \exists y \quad X^T y = z\}$)

Proof:

$\underline{p}^T \underline{\beta}$ is estimable

$\iff \exists\ l \in \mathbb{R}^n$ s.t. $E[l^T \underline{Y}] = \underline{p}^T \underline{\beta}. \quad \forall \beta \in \mathbb{R}^n$

$\iff \exists\ l \in \mathbb{R}^n$ s.t. $l^T X \underline{\beta} = \underline{p}^T \underline{\beta} \quad \forall \beta \in \mathbb{R}^n$

(linearity of expectation)

$\iff \exists\ l \in \mathbb{R}^n \quad X^T l = \underline{p}$

(Ex: linear Algebra)

$\iff \underline{p} \in \mathcal{C}(X^T)$ □