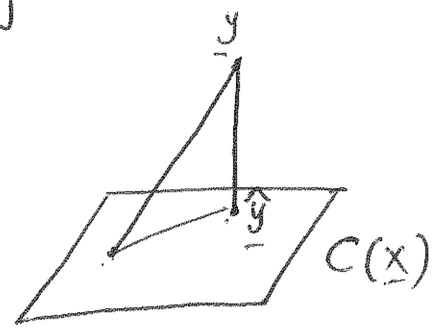
Recap:

Question: $(\underline{X}^T\underline{X})\underline{\beta} = \underline{X}^T\underline{y}$

It can be unique if $(\underline{X}^T\underline{X})$ is invertible and when $\underline{X}^T\underline{X}$ is not invertible, then this equation has infinitely many solution

Projection matrix:



Projection matrix $\underline{P}$ for $C(\underline{X})$ has the property that $\underline{P}\underline{v} = \underline{v}$ if $\underline{v} \in C(\underline{X})$ and $\underline{P}\underline{w} = \underline{0}$ if $\underline{w} \in \mathcal{N}(\underline{X}^T)$:

① $\underline{P}$ is symmetric ② $\underline{P}$ is idempotent i.e. $\underline{P}^2 = \underline{P}$.

If $\underline{P}$ is the projection matrix onto the column space of $\underline{X}$ (i.e. $C(\underline{X})$) then $\underline{I} - \underline{P}$ is the projection matrix onto $\mathcal{N}(\underline{X}^T)$

Thm: $\text{rank}(\underline{P}) + \text{rank}(\underline{I} - \underline{P}) = n$

Pf:- rank of a matrix = number of nonzero eigenvalues of a matrix.

trace of a matrix = sum of the diagonal entries of a matrix

= sum of the eigenvalues of a matrix.

$\underline{P}$ is idempotent $\Rightarrow$ $\underline{P}^2 = \underline{P}$

if λ_i is an eigenvalue with eigenvector $\underline{v}_i$

then $\underline{P}^2 \underline{v}_i = \underline{P}\,\underline{P}\,\underline{v}_i = \underline{P}\lambda_i \underline{v}_i$ $\qquad$ $\underline{P}\underline{v}_i = \lambda_i \underline{v}_i$

$$= \lambda_i P \underline{v}_i$$
$$= \lambda_i \lambda_i v_i = \lambda_i^2 \underline{v}_i$$

$P\underline{v}_i = \lambda_i^2 v_i$

$\Rightarrow$ if λ_i is an eigenvalue then λ_i^2 is also the same eigenvalue

$\Rightarrow \lambda_i = \lambda_i^2 \Rightarrow \lambda_i = 0$ or 1

eigenvalue of an idempotent matrix is either 0 or 1

rank of $\underline{P}$ = number of nonzero eigenvalues

$$= \sum \lambda_i = \text{trace of } \underline{P}$$

rank$(\underline{P})$ = trace$(\underline{P})$.

If $\underline{P}$ is idempotent i.e. $\underline{P}^2 = \underline{P}$ then

$$(\underline{I} - \underline{P})^2 = (\underline{I} - \underline{P})(\underline{I} - \underline{P}) = \underline{I} - 2\underline{P} + \underline{P}^2 = \underline{I} - 2\underline{P} + \underline{P}$$
$$= \underline{I} - \underline{P}$$

$\Rightarrow$ $\underline{I} - \underline{P}$ is also idempotent

hence, rank$(\underline{P})$ + rank$(\underline{I} - \underline{P})$ = trace$(\underline{P})$ + trace$(\underline{I} - \underline{P})$

$$= \text{trace}(\underline{I}) = n.$$

<u>Notice</u>: $\underline{y} = \underline{P}\underline{y} + (\underline{I} - \underline{P})\underline{y}$

where $\underline{P}\underline{y} \in C(\underline{X})$, $(\underline{I} - \underline{P})\underline{y} \in \mathcal{N}(\underline{X}^T)$

$$(\underline{P}\underline{y})^T (\underline{I} - \underline{P})\underline{y} = \underline{y}^T \underline{P}^T (\underline{I} - \underline{P})\underline{y} = \underline{y}^T \underline{P}(\underline{I} - \underline{P})\underline{y}$$
$$= \underline{y}^T(\underline{P} - \underline{P}^2)\underline{y} = \underline{0}$$

We know that there is some matrix $\underline{P}$ which projects any vector onto the $C(\underline{X})$.

What is the formula for $\underline{P}$?

Def:- A generalized inverse of an $m \times n$ matrix $\underline{A}$ is any $n \times m$ matrix $\underline{G}$ satisfying $\underline{A}\,\underline{G}\,\underline{A} = \underline{A}$. The notation $\underline{A}^{-}$ is used to denote the generalized inverse of a matrix.

If $\underline{A}$ is invertible (i.e. $\underline{A}$ has full rank and $\underline{A}$ is a square matrix) then the inverse $\underline{A}^{-1}$ is one of the generalized inverses.

Note that a matrix can have infinitely many generalized inverses.

Result: Let $\underline{A}$ be an $m \times n$ matrix with rank r. If $\underline{A}$ can be partitioned as below,

$$\underline{A} = \begin{bmatrix} \underline{C}_{r \times r} & \underline{D}_{r \times (n-r)} \\ \underline{E}_{(m-r) \times r} & \underline{F}_{(m-r) \times (n-r)} \end{bmatrix}$$

with $\text{rank}(\underline{A}) = \text{rank}(\underline{C}) = r$

then one generalized inverse of $\underline{A}$ is given by

$$\underline{A}^{-} = \begin{bmatrix} \underline{C}^{-1}_{r \times r} & \underline{O}_{r \times (n-r)} \\ \underline{O}_{(m-r) \times r} & \underline{O}_{(m-r) \times (n-r)} \end{bmatrix}$$

Example: $\underline{A} = \begin{bmatrix} 0 & -1 & -2 \\ 1 & 2 & 4 \end{bmatrix}_{2\times 3}$

rank $(\underline{A}) = 2$, as the 3rd column is twice the second column.

$$\underline{A} = \left[\begin{array}{cc|c} 0 & -1 & -2 \\ 1 & 2 & 4 \end{array}\right] \qquad \underline{C} = \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix}$$

$$\underline{D} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$$

Formula says $A^{-} = \begin{bmatrix} \underline{C}^{-1} & : & \underline{0} \end{bmatrix}$

$$\underline{C}^{-1} = \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix}^{-1} = \frac{1}{1} \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}$$

$$\underline{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 2 \end{bmatrix}$$ 1st row = 2nd row + 3rd row

$$\underbrace{\begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix}}_{2\times 2}$$

rank $(\underline{A}) = 2$

$$\underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}}_{\underline{H}_1} \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 2 \end{bmatrix}}_{\underline{A}} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix}}_{\underline{H}_1\underline{A}}$$

$$\underline{H}_1\underline{A} \underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_{\underline{H}_2} = \begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 & 1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\underline{H}_1 \underline{A} \underline{H}_2 = \begin{bmatrix} \underline{C}_{r \times r} & \underline{D}_{r \times (n-r)} \\ \underline{E}_{(m-r)\times r} & \underline{F}_{(m-r)\times(n-r)} \end{bmatrix}$$

Now I can get the generalized inverse of $\underline{A}$ by $A^{-} = \underline{H}_1^{-1} \begin{bmatrix} \underline{C}^{-1} & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix} \underline{H}_2^{-1}$

What are these matrices $\underline{H}_1$ and $\underline{H}_2$.

$$\underline{H}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \qquad \underline{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\underline{I}\,\underline{x} = \underline{x}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[\text{1st and 2nd colum}]{\text{interchange}} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[\text{1st and 3rd colum}]{\text{interchange}} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[\text{1st and 2nd row}]{\text{interchange}} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[\text{2nd and 3rd row}]{\text{interchange}} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

these matrices are called permutation matrices as you permute rows and/or columns of an identity matrix.

Suppose $\underset{n\times p}{\underline{X}}$ is not full column rank.

If $\underline{X}$ is full column rank $\Rightarrow$ $\underline{X}^T\underline{X}$ is invertible

$\Rightarrow$ $\hat{\beta} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{y}$

Suppose $\underline{X}$ is not of full column rank

$\Rightarrow$ $(\underline{X}^T\underline{X})$ is not invertible

~~but~~ let $(\underline{X}^T\underline{X})^-$ be the generalized inverse of $\underline{X}^T\underline{X}$.

<u>Lemma</u>: If $\underline{G}$ and $\underline{H}$ are generalized inverses of $\underline{X}^T\underline{X}$, then

① $\underline{X}\,G\,\underline{X}^T\underline{X} = \underline{X}$ $\underline{H}\,\underline{X}^T\underline{X} = \underline{X}$

② $\underline{X}\,\underline{G}\,\underline{X}^T = \underline{X}\,\underline{H}\,\underline{X}^T$

<u>claim</u>: $(\underline{X}^T\underline{X})^-\underline{X}^T$ is a generalized inverse of $\underline{X}$.

$\underline{X}(\underline{X}^T\underline{X})^-\underline{X}^T\underline{X} = \underline{X}$ $\Rightarrow$ $(\underline{X}^T\underline{X})^-\underline{X}^T$ is the generalized inverse of $\underline{X}$.

<u>Thm</u>: $P = \underline{X}(\underline{X}^T\underline{X})^-\underline{X}^T$ is the projection matrix onto $C(\underline{X})$.

<u>Pf</u>: $\underline{v} \in C(\underline{X})$ $\Rightarrow$ $\underline{v} = \underline{X}\,\underline{a}$ for some $\underline{a}$

$\Rightarrow$ $\underline{P}\,\underline{v} = \underbrace{\underline{X}(\underline{X}^T\underline{X})^-\underline{X}^T\underline{X}}_{\underline{X}}\,\underline{a} = \underline{X}\,\underline{a}$ [by ① of Lemma]

$= \underline{v}$

$\underline{w} \in \mathcal{N}(\underline{X}^T) \Rightarrow \underline{X}^T \underline{w} = \underline{0}$

$\underline{P}\,\underline{w} = \underline{X}(\underline{X}^T\underline{X})^- \underline{X}^T \underline{w} = \underline{0}$

<u>Thm</u>: Let $A\underline{x} = \underline{c}$ be a system of equations and let $\underline{G}$ be the generalized inverse of $\underline{A}$. then $\tilde{\underline{x}}$ is a solution to the system $A\underline{x} = \underline{c}$ if and only if there exists a vector $\underline{z}$ s.t. $\tilde{\underline{x}} = \underline{G}\,\underline{c} + (\underline{I} - \underline{G}\,\underline{A})\underline{z}$

In our case $\underline{A} = (\underline{X}^T\underline{X})$ $\quad \underline{x} = \underline{\beta}$, $\underline{c} = \underline{X}^T\underline{y}$

class of all solutions to the NEs

$$= \left\{ (\underline{X}^T\underline{X})^- \underline{X}^T\underline{y} + (\underline{I} - (\underline{X}^T\underline{X})^- \underline{X}^T\underline{X})\underline{z} \;\middle|\; \text{for all } (\underline{X}^T\underline{X})^- \text{ and for all } \underline{z} \right\}$$

If $\hat{\underline{\beta}}_1$ and $\hat{\underline{\beta}}_2$ are two solutions to the NEs $\Rightarrow$

$$\underline{X}^T\underline{X}\,\hat{\underline{\beta}}_1 = \underline{X}^T\underline{y} \quad \cdots (1)$$

$$\underline{X}^T\underline{X}\,\hat{\underline{\beta}}_2 = \underline{X}^T\underline{y} \quad \cdots (2)$$

subtracting (2) from (1), we get

$$\underline{X}^T\underline{X}(\hat{\underline{\beta}}_1 - \hat{\underline{\beta}}_2) = \underline{0} \Rightarrow (\hat{\underline{\beta}}_1 - \hat{\underline{\beta}}_2) \in \mathcal{N}(\underline{X}^T\underline{X}) = \mathcal{N}(\underline{X})$$

if $\underline{c} \in \mathcal{N}(\underline{X}^T\underline{X}) \Rightarrow \underline{X}^T\underline{X}\,\underline{c} = \underline{0} \Rightarrow \underline{c}^T\underline{X}^T\underline{X}\,\underline{c} = 0$

$\Rightarrow \|\underline{X}\,\underline{c}\|^2 = \underline{0} \Rightarrow \underline{X}\,\underline{c} = 0 \Rightarrow \underline{c} \in \mathcal{N}(\underline{X})$.

Thm: If $C(\underline{W}) \subset C(\underline{X})$ and $\underline{P}_X$ and $\underline{P}_W$ are the projection matrices onto $C(\underline{X})$ and $C(\underline{W})$ respectively, then $\underline{P}_X - \underline{P}_W$ is the projection matrix onto $C((\underline{I} - \underline{P}_W)\underline{X})$

Ex! $y_i = \beta_0 + \beta_1 x_i + e_i$, $i = 1, 2, 3, 4$.

$x_1 = 1$, $x_2 = 2$, $x_3 = 3$, $x_4 = 4$

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}}_{\underline{y}} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}}_{\underline{X}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}}_{\underline{\beta}} + \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}}_{\underline{e}}$$

Model 1

Suppose I do not use the predictor and I fit the response only with an intercept.

$$\underline{y} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}}_{\underline{W}} \gamma + \underbrace{\begin{bmatrix} \tilde{e}_1 \\ \tilde{e}_2 \\ \tilde{e}_3 \\ \tilde{e}_4 \end{bmatrix}}_{\tilde{\underline{e}}}$$

Model 2

$C(\underline{W}) \subset C(\underline{X})$ What $\underline{P}_W = \underline{W}(\underline{W}^T\underline{W})^{-1}\underline{W}^T$

$$= \underline{1}_{4\times1}(\underline{1}_{4\times1}^T \underline{1}_{4\times1})^{-1}\underline{1}_{4\times1}^T$$

$$= \frac{1}{4}\underline{1}_{4\times1}\underline{1}_{4\times1}^T$$

$$= \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} = \frac{1}{4}\underline{J}_{4\times4}$$

Recap: We described what is the projection matrix $\underset{\sim}{P}$. It turns out that the projection matrix onto the $C(\underset{\sim}{X})$ is $\underset{\sim}{P} = \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T$ if $\underset{\sim}{X}^T\underset{\sim}{X}$ has an inverse, o.w. $\underset{\sim}{P} = \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-}\underset{\sim}{X}^T$ where $(\underset{\sim}{X}^T\underset{\sim}{X})^{-}$ is a generalized inverse of $\underset{\sim}{X}^T\underset{\sim}{X}$.

Ex! $y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, 2, 3, 4.$

$x_1 = 1, \; x_2 = 2, \; x_3 = 3, \; x_4 = 4.$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}}_{\underset{\sim}{X}} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \longrightarrow \text{Model 1}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}}_{\underset{\sim}{W}} \gamma + \begin{bmatrix} \tilde{e}_1 \\ \tilde{e}_2 \\ \tilde{e}_3 \\ \tilde{e}_4 \end{bmatrix} \longrightarrow \text{Model 2.}$$

$\underset{\sim}{P}_{\underset{\sim}{W}} = \frac{1}{4}\underset{\sim}{J}_{4\times 4}$. What is the predicted response based on Model 2?

prediction w.r.t. Model 2 is $\underset{\sim}{P}_{W}\underset{\sim}{y}$

$$\underset{\sim}{P}_{W}\underset{\sim}{y} = \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \bar{y} \\ \bar{y} \\ \bar{y} \end{pmatrix}$$

Let us look at Model 1

$$\underline{y} = \underline{X}\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix} + \begin{bmatrix} e_1\\ e_2\\ e_3\\ e_4\end{bmatrix}$$

$$\underline{P}_X = \underline{X}(\underline{X}^T\underline{X})^{-1}\underline{X}^T = \frac{1}{10}\begin{bmatrix} 7 & 4 & 1 & -2\\ 4 & 3 & 2 & 1\\ 1 & 2 & 3 & 4\\ -2 & 1 & 4 & 7\end{bmatrix}$$

the predicted responses based on Model 1 is

$$\underline{P}_X\,\underline{y} = \frac{1}{10}\begin{bmatrix} 7 & 4 & 1 & -2\\ 4 & 3 & 2 & 1\\ 1 & 2 & 3 & 4\\ -2 & 1 & 4 & 7\end{bmatrix}\begin{bmatrix} y_1\\ y_2\\ y_3\\ y_4\end{bmatrix} = \begin{bmatrix} \frac{7y_1+4y_2+y_3-2y_4}{10}\\ \frac{4y_1+3y_2+2y_3+y_4}{10}\\ \frac{y_1+2y_2+3y_3+4y_4}{10}\\ \frac{-2y_1+y_2+4y_3+7y_4}{10}\end{bmatrix}$$

$\underline{y} = \underline{X}\,\underline{\beta} + \underline{e} \longrightarrow$ Model 1

$\underline{y} = \underline{W}\,\underline{\gamma} + \underline{\tilde{e}} \longrightarrow$ Model 2

Qn: when will the predicted responses from two models will be the same?

Ans: $C(\underline{W}) = C(\underline{X})$ then $\underline{P}_X\,\underline{y} = \underline{P}_W\,\underline{y}$

In that case the two models are known to be equivalent and we call one model as the reparametrization of the other model.

Def: Two linear models $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$, where $\underline{X}$ is an $n \times p$ design matrix and $\underline{y} = \underline{W}\underline{\gamma} + \underline{\tilde{e}}$ where $\underline{W}$ is an $n \times t$ matrix are equivalent iff $C(\underline{X}) = C(\underline{W})$. If this holds, then

① $\underline{P}_X = \underline{P}_W$ ② $\underline{\hat{y}} = \underline{P}_X \underline{y} = \underline{P}_W \underline{y}$ ③ $\underline{\hat{e}} = (\underline{I} - \underline{P}_X)\underline{y}$

$$= (\underline{I} - \underline{P}_W)\underline{y}.$$

Example: 1 way ANOVA.

$y_{ij} = \mu + \alpha_i + e_{ij}$, $i = 1, 2, 3$; $j = 1, \dots, n \to$ Model 1

$y_{ij} = \theta_i + \tilde{e}_{ij}$, $i = 1, 2, 3$; $j = 1, \dots, n \to$ Model 2

Model 1: $\underline{y} = \underline{X}\underline{\beta} + \underline{e}$

$$\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \\ y_{21} \\ \vdots \\ y_{2n} \\ y_{31} \\ \vdots \\ y_{3n} \end{pmatrix} = \underbrace{\begin{bmatrix} \underline{1}_n & \underline{1}_n & \underline{0}_n & \underline{0}_n \\ \underline{1}_n & \underline{0}_n & \underline{1}_n & \underline{0}_n \\ \underline{1}_n & \underline{0}_n & \underline{0}_n & \underline{1}_n \end{bmatrix}}_{\underline{X}} \underbrace{\begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}}_{\underline{\beta}} + \underline{e}$$

Model 2:

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{1n} \\ y_{21} \\ \vdots \\ y_{2n} \\ y_{31} \\ \vdots \\ y_{3n} \end{pmatrix} = \underbrace{\begin{bmatrix} \underline{1}_n & \underline{0}_n & \underline{0}_n \\ \underline{0}_n & \underline{1}_n & \underline{0}_n \\ \underline{0}_n & \underline{0}_n & \underline{1}_n \end{bmatrix}}_{\underline{W}} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} + \underline{\tilde{e}}$$

Is $C(\underline{X}) = C(\underline{W})$?

$C(\underline{X}) = C(\underline{W})$ as they have the last three columns identical. The first column in $\underline{X}$ can be written as a sum of the last three columns.

$\Rightarrow$ $\underline{P}_X = \underline{P}_W$ and thus the predicted responses will also be the same.

$\mu + \alpha_i = \theta_i \longrightarrow$ this reparametrization gives Model 2 from Model 1.

<u>Ex!</u> $i = 1, 2;\ j = 1, \dots, n$.

<u>Model 1</u>: $y_{ij} = \beta_0^{(i)} + \beta_1^{(i)} x_{ij} + e_{ij}$

there are two groups of students, $i = 1, 2$.

In each group we are including n students and we observe their heights and weights.

x_{ij} = weight of the jth student in the ith group

y_{ij} = height of the jth student in the ith group.

regression coefficients are specific to the group $i = 1, 2$.

<u>Model 2</u> : $d_{ij} = \begin{cases} 0 & \text{if } i=1 \text{ (for group 1)} \\ 1 & \text{if } i=2 \text{ (group 2)} \end{cases}$.

$$y_{ij} = \gamma_0 + \gamma_1 x_{ij} + \gamma_2 d_{ij} + \gamma_3 d_{ij} x_{ij} + \tilde{e}_{ij}$$

this regression model (Model 2) takes the following forms for groups 1 and 2

For group 1, $y_{ij} = \gamma_0 + \gamma_1 x_{ij} + \tilde{e}_{ij}$

For group 2, $y_{ij} = (\gamma_0 + \gamma_2) + (\gamma_1 + \gamma_3) x_{ij} + \tilde{e}_{ij}$

If I fit Model 1 and Model 2 to the same data, do I get the same predicted responses? Or are they equivalent?

<u>Model 1</u> :

$$\underbrace{\begin{pmatrix} y_{11} \\ \vdots \\ y_{1n} \\ y_{21} \\ \vdots \\ y_{2n} \end{pmatrix}}_{\tilde{y}} = \underbrace{\begin{bmatrix} 1 & x_{11} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & 0 & 0 \\ 0 & 0 & 1 & x_{21} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_{2n} \end{bmatrix}}_{\underline{X}} \begin{pmatrix} \beta_0^{(1)} \\ \beta_1^{(1)} \\ \beta_0^{(2)} \\ \beta_1^{(2)} \end{pmatrix} + \underline{e}$$

<u>Model 2</u> :

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{1n} \\ y_{21} \\ \vdots \\ y_{2n} \end{pmatrix} = \underbrace{\begin{bmatrix} 1 & x_{11} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & 0 & 0 \\ 1 & x_{21} & 1 & x_{21} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{2n} & 1 & x_{2n} \end{bmatrix}}_{\underline{W}} \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} + \tilde{\underline{e}}$$

1st column in $\underline{W}$ = 1st column in $\underline{X}$ + 3rd column in $\underline{X}$

(8) 2nd column in $\underline{W}$ = 2nd column in $\underline{X}$ + 4th column in $\underline{X}$

the last two columns in $\underline{X}$ and $\underline{W}$ are same.

$\Rightarrow \quad C(\underline{X}) = C(\underline{W})$.

these two models are reparametrizations of each other.

Now recall NEs $\quad \underline{X}^T \underline{X} \, \underline{\beta} = \underline{X}^T \underline{y}$

it has a unique soln. when $r(\underline{X}) = p$ [where $\underline{X}$ is $n \times p$ matrix]

when $r(\underline{X}) < p$ it has infinitely many solutions.

~~If an equation~~ If the NEs has multiple solutions then we can't precisely estimate the parameter β from the data in the sense that the parameter is not identifiable.

~~Iden~~ Idenfiability : The parametrization $\underline{\beta}$ is identifiable if for any $\underline{\beta}_1, \underline{\beta}_2$,

$\underline{X} \underline{\beta}_1 = \underline{X} \underline{\beta}_2 \Rightarrow \underline{\beta}_1 = \underline{\beta}_2$.

$E[\underline{y}] = X \underline{\beta}_1$

(1) Knowing $E[\underline{y}] = X \underline{\beta}$ means knowing $\underline{\beta}$.

A difference in the parameter values should imply a difference in the mean.

$y_{ij} = \mu + \alpha_i + e_{ij}$ $\quad E[e_{ij}] = 0$ $\quad i = 1, 2, 3$, $j = 1, \ldots, n$.

$E[y_{ij}] = \mu + \alpha_i = (\mu + c) + (\alpha_i - c)$

$\underline{y} = \underline{X}\,\underline{\beta} + \underline{e}$; $\quad \underline{X} = \begin{bmatrix} \underline{1}_n & \underline{1}_n & \underline{0}_n & \underline{0}_n \\ \underline{1}_n & \underline{0}_n & \underline{1}_n & \underline{0}_n \\ \underline{1}_n & \underline{0}_n & \underline{0}_n & \underline{1}_n \end{bmatrix}$

$\underline{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$

$\underline{\beta}_2 = \begin{pmatrix} \mu + c \\ \alpha_1 - c \\ \alpha_2 - c \\ \alpha_3 - c \end{pmatrix}$ $\quad \underline{\beta}_1 = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$

$X\underline{\beta}_2 = X\underline{\beta}_1$ although $\underline{\beta}_1 \neq \underline{\beta}_2$. This parameter $\underline{\beta}$ is not identifiable.

Ex! Consider a linear model $r(\underline{X}) = p$

If $\underline{X}\,\underline{\beta}_1 = X\underline{\beta}_2$. Premultiply both sides by $(\underline{X}^T\underline{X})^{-1}\underline{X}^T$ $\Rightarrow (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{X}\,\underline{\beta}_1 = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{X}\,\underline{\beta}_2$

$\Rightarrow \underline{\beta}_1 = \underline{\beta}_2$.

when $r(\underline{X}) = p$ then $\underline{\beta}$ becomes identifiable.

Def: A vector valued function $g(\underline{\beta})$ is identifiable if $\underline{X}\,\underline{\beta}_1 = \underline{X}\,\underline{\beta}_2 \Rightarrow g(\underline{\beta}_1) = g(\underline{\beta}_2)$.

Our specific focus lies on the identifiability of linear functions $g(\underline{\beta}) = \underline{\lambda}^T\underline{\beta}$.

thus $\underline{\lambda}^T \underline{\beta}$ is identifiable if

$\underline{X}\,\underline{\beta}_1 = \underline{X}\,\underline{\beta}_2$ implies $\underline{\lambda}^T \underline{\beta}_1 = \underline{\lambda}^T \underline{\beta}_2$.

from any model the functions of the parameters which are reasonable to estimate are the identifiable functions.

Def:- An estimator $t(\underline{y})$ is an unbiased estimator for the function $\underline{\lambda}^T \underline{\beta}$ iff

$E[t(\underline{y})] = \underline{\lambda}^T \underline{\beta}$ for all $\underline{\beta}$.

An estimator $t(\underline{y})$ is a linear estimator in $\underline{y}$ iff $t(\underline{y}) = c + \underline{a}^T \underline{y}$ for some constants

$c, a_1, \ldots, a_n$, $\underline{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$.

A function $\underline{\lambda}^T \underline{\beta}$ is linearly estimable iff there exists a linear unbiased estimator for $\underline{\lambda}^T \underline{\beta}$.

$\underline{\lambda}^T \underline{\beta}$ is linearly estimable

$\Rightarrow \exists$ some $c + \underline{a}^T \underline{y}$ s.t. $E[c + \underline{a}^T \underline{y}] = \lambda^T \underline{\beta}$ --(A)

for all $\underline{\beta}$

think of the linear model

$\underline{y} = X \underline{\beta} + \underline{e}$ and assume $E[\underline{e}] = \underline{0}$

$\Rightarrow E[\underline{y}] = \underline{X}\,\underline{\beta}$

(*) implies $E[c + \underline{a}^T \underline{y}] = \underline{\lambda}^T \underline{\beta} \quad \forall \underline{\beta}$

$\Rightarrow c + \underline{a}^T \underline{X} \underline{\beta} = \underline{\lambda}^T \underline{\beta} \quad \forall \underline{\beta}$

$\Rightarrow c = 0$ and $\underline{a}^T \underline{X} = \underline{\lambda}^T \Rightarrow \underline{\lambda} = \underline{X}^T \underline{a}$

$\Rightarrow \underline{\lambda} \in C(\underline{X}^T)$

A linear function $\underline{\lambda}^T \underline{\beta}$ is linearly estimable iff $\underline{\lambda} \in C(\underline{X}^T)$.

<u>Ex!</u> $i = 1, 2, 3; \quad j = 1, \ldots, n_i \quad , \quad n_1 = 3, \; n_2 = 2, \; n_3 = 1$

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{31} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} + \begin{pmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{31} \end{pmatrix}$$

$y_{ij} = \mu + \alpha_i + e_{ij}$ .. $i = 1, \; n_1 = 3$

$i = 2, \; n_2 = 2$

$i = 3, \; n_3 = 1$

whether the following functions are linearly estimable

1. α_1 2. $\alpha_1 - \alpha_2$ 3. $\mu + \alpha_1$ 4. $\alpha_1 + \alpha_2 - 2\alpha_3$.

$$\underline{X}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

you can check that the basis of $C(\underline{X}^T)$ is $\left\{ \begin{pmatrix}1\\1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\0\\1\end{pmatrix} \right\}$.

$$a_1 \begin{pmatrix}1\\1\\0\\0\end{pmatrix} + a_2 \begin{pmatrix}1\\0\\1\\0\end{pmatrix} + a_3 \begin{pmatrix}1\\0\\0\\1\end{pmatrix} = \begin{pmatrix}0\\0\\0\\0\end{pmatrix}$$

$\Rightarrow a_3 = 0$, $a_2 = 0$ the second equation gives $a_1 = 0$.
indeed they are linearly independent.

$$\alpha_1 = (0 \; 1 \; 0 \; 0) \underbrace{\begin{pmatrix}\mu\\ \alpha_1\\ \alpha_2\\ \alpha_3\end{pmatrix}}_{\underline{\beta}} = \underline{\lambda}^T \underline{\beta}$$

$$\underline{\lambda} = \begin{pmatrix}0\\1\\0\\0\end{pmatrix}$$

Lets check if $\underline{\lambda} = \begin{pmatrix}0\\1\\0\\0\end{pmatrix} \in C(\underline{X}^T)$.

If $\begin{pmatrix}0\\1\\0\\0\end{pmatrix}$ belongs to the $C(\underline{X}^T)$ then it must be written as a linear combination of the basis of $C(\underline{X}^T)$

$$a_1 \begin{pmatrix}1\\1\\0\\0\end{pmatrix} + a_2 \begin{pmatrix}1\\0\\1\\0\end{pmatrix} + a_3 \begin{pmatrix}1\\0\\0\\1\end{pmatrix} = \begin{pmatrix}0\\1\\0\\0\end{pmatrix} \quad \exists\, a_1, a_2, a_3.$$

the last equation $\Rightarrow a_3 = 0$,
second to last equation $\Rightarrow a_2 = 0$
first equation $\Rightarrow a_1 + a_2 + a_3 = 0 \Rightarrow a_1 = 0$
second equation $a_1 = 1$

(back ①)

Hence, $\begin{pmatrix}0\\1\\0\\0\end{pmatrix} \notin C(\underline{X}^T) \Rightarrow \alpha_1$ is not linearly estimable.

2. $\alpha_1 - \alpha_2 = \begin{pmatrix}0 & 1 & -1 & 0\end{pmatrix} \begin{pmatrix}\mu\\ \alpha_1\\ \alpha_2\\ \alpha_3\end{pmatrix} = \underline{\lambda}^T \underline{\beta}$

$\underline{\lambda} = \begin{pmatrix}0\\1\\-1\\0\end{pmatrix}$ my job is to see if

$\underline{\lambda} \in C(\underline{X}^T)$

let $\underline{\lambda} \in C(\underline{X}^T) \Rightarrow \exists\, a_1, a_2, a_3$

$$a_1 \begin{pmatrix}1\\1\\0\\0\end{pmatrix} + a_2 \begin{pmatrix}1\\0\\1\\0\end{pmatrix} + a_3 \begin{pmatrix}1\\0\\0\\1\end{pmatrix} = \begin{pmatrix}0\\1\\-1\\0\end{pmatrix}$$

Last equation $\Rightarrow a_3 = 0$
second to last equation $\Rightarrow a_2 = -1$
second equation $\Rightarrow a_1 = 1$
first eqn. $\Rightarrow a_1 + a_2 + a_3 = 0$ which has been satisfied as $a_1 = 1$ and $a_2 = -1$, $a_3 = 0$
$\Rightarrow \underline{\lambda} \in C(\underline{X}^T) \Rightarrow \alpha_1 - \alpha_2$ is linearly estimable

back 2