$\underline{M} = I_2$: perpendicular ~~matrix~~ projection matrix onto $\mathbb{R}^2$

$\underline{M_0} = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix}$: the perpendicular projection matrix onto

$\mathcal{C}(M_0) = \mathcal{C}(X) \subset \mathcal{C}(M) = \mathbb{R}^2$

$X = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ $\quad \mathcal{C}(M_0) = \mathcal{C}(X) = \left\{ \begin{bmatrix} 2a \\ a \end{bmatrix}, \quad a \in \mathbb{R} \right\}$

$\begin{bmatrix} b \\ -2b \end{bmatrix} \perp \mathcal{C}(X)$ for any $b \in \mathbb{R}$

$$M - M_0 = I_2 - M_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} = \begin{bmatrix} 0.2 & -0.4 \\ -0.4 & 0.8 \end{bmatrix}$$

~~Note)~~ Claim: $\underline{(M - M_0)}$ is the perpendicular projection matrix onto the space of vectors w/ form $\begin{bmatrix} b \\ -2b \end{bmatrix} = N(X^T)$

(i) $V = \begin{bmatrix} b \\ -2b \end{bmatrix} \in N(X^T) \Rightarrow (M - M_0)V = V$?

$$\begin{bmatrix} 0.2 & -0.4 \\ -0.4 & 0.8 \end{bmatrix} \begin{bmatrix} b \\ -2b \end{bmatrix} = \begin{bmatrix} 0.2b + 0.8b \\ -0.4b - 1.6b \end{bmatrix} = \begin{bmatrix} b \\ 2b \end{bmatrix} \checkmark$$

(ii) $W \not\perp N(X^T) \Rightarrow W = \begin{bmatrix} 2a \\ a \end{bmatrix}$

$$(M - M_0) \cdot W = \begin{bmatrix} 0.2 & -0.4 \\ -0.4 & 0.8 \end{bmatrix} \begin{bmatrix} 2a \\ a \end{bmatrix} = \begin{bmatrix} 0.4a - 0.4a \\ -0.8a + 0.8a \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\checkmark$

$$Iy = Py + (I - P)y$$

Ex 1 : $A = \begin{bmatrix} a & b \\ b & \frac{b^2}{a} \end{bmatrix}$ $r(A) = 1$ $A^- = \begin{bmatrix} \frac{1}{a} & 0 \\ 0 & 0 \end{bmatrix}$

$\frac{b}{a}$

Let's check

$AA^-A = A$

$$\begin{bmatrix} a & b \\ b & \frac{b^2}{a} \end{bmatrix}\begin{bmatrix} \frac{1}{a} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ b & \frac{b^2}{a} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{b}{a} & 0 \end{bmatrix}\begin{bmatrix} a & b \\ b & \frac{b^2}{a} \end{bmatrix} = \begin{bmatrix} a & b \\ b & \frac{b^2}{a} \end{bmatrix}$$

† Def A *generalized inverse* of a $m \times n$ matrix $\mathbf{A}$ is any $n \times m$ matrix $\mathbf{G}$ such that $\mathbf{AGA} = \mathbf{A}$. The notation $\mathbf{A}^-$ is used to indicate a generalized inverse of $\mathbf{A}$. $(= A^g)$

★ Result: Let $\mathbf{A}$ be an $m \times n$ matrix with rank r. If $\mathbf{A}$ can be partitioned as below, with $r(\mathbf{A}) = r(\mathbf{C}) = r$,

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

so that $\mathbf{C}$ is nonsingular, then the matrix,

$$G \; \cancel{\mathbf{G}} = \begin{bmatrix} \mathbf{C}^{-1}_{r\times r} & \mathbf{0}_{r\times(m-r)} \\ \mathbf{0}_{(n-r)\times r} & \mathbf{0}_{(n-r)\times(m-r)} \end{bmatrix}$$

is a generalized inverse of $\mathbf{A}$.

Eg2 $A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 2 \end{bmatrix}_{3\times4}$ $r(A) = 2$

$A_1^- = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}_{4\times3}$

★ Result: For a given $m \times n$ matrix $\mathbf{A}$ with rank r, let $\mathbf{P}$ and $\mathbf{Q}$ be permutation matrices such that,

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

where $r(\mathbf{A}) = r(\mathbf{C}) = r$ and $\mathbf{C}$ is nonsingular. Then the matrix $\mathbf{G}$ below is a generalized inverse of $\mathbf{A}$: so that $\mathbf{C}$ is nonsingular, then the matrix,

$$\mathbf{G} = \mathbf{Q} \begin{bmatrix} \mathbf{C}^{-1}_{r\times r} & \mathbf{0}_{r\times(m-r)} \\ \mathbf{0}_{(n-r)\times r} & \mathbf{0}_{(n-r)\times(m-r)} \end{bmatrix} \mathbf{P}.$$

$PAQ = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_{Q}$

$= \begin{bmatrix} 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \boxed{\begin{matrix} 0 & -1 \\ 1 & 2 \end{matrix}} & 0 & 1 \\ & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$

$A_2^- = \underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_{Q} \begin{bmatrix} 2 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & -1 & 0 \end{bmatrix}$

$Ax = c$

$c \in \mathcal{C}(A)$

♣ Th: If $\mathbf{A}$ is nonsingular, the unique generalized inverse of $\mathbf{A}$ is $\mathbf{A}^{-1}$.

♣ Th: For any symmetric matrix $\mathbf{A}$, there exists a generalized inverse of $\mathbf{A}$. (note: For $\mathbf{A}$ symmetric, $\mathbf{A}^{-}$ need not be symmetric)

♠ Observe $\mathbf{X}^T\mathbf{X}$ is symmetric $\rightarrow \exists\ (\mathbf{X}^T\mathbf{X})^{-}$.

★ Lem If $\mathbf{G}$ and $\mathbf{H}$ are generalized inverse of $(\mathbf{X}^T\mathbf{X})$, then

(i) $\mathbf{XGX}^T\mathbf{X} = \mathbf{XHX}^T\mathbf{X} = \mathbf{X}$

(ii) $\mathbf{XGX}^T = \mathbf{XHX}^T$

★ Result: $(\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T$ is a generalized inverse of $\mathbf{X}$.

"Hat"

$\hat{y} = Py = Hy$

$(X^T X)^- X^T$

$= $ g-inverse of X

♣ Th 2.1: $\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T$ is the perpendicular projection operator onto $C(\mathbf{X})$.

Check! (i) $v \in C(X) \Rightarrow Pv = v$

(ii) $w \perp C(X) \Rightarrow Pw = 0$

(i) $v \in C(X) \Rightarrow \exists\, b\, (\neq 0)$ s.t $v = Xb$

$P_v = X\underbrace{(X^T X)^- X^T X}\, b = Xb = v$

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

$P_w = X(X^T X)^- \underbrace{X^T w}_{=0} = 0$

★ Summary! We write

$$\mathbf{y} = \mathbf{P}\mathbf{y} + (I - \mathbf{P})\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{e}}$$

• $\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T$ is the perpendicular projection matrix onto $C(\mathbf{X})$.

• $(I - \mathbf{P})$ is the unique, symmetric perpendicular projection matrix onto the orthogonal complememt of $C(\mathbf{X})$ with respect to $\mathbb{R}^n$.

ye

• $C(\mathbf{X})$ and $\mathcal{N}(\mathbf{X}^T)$ are orthogonal complements to $\mathbb{R}^n$.

• $\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y} = \mathbf{P}\mathbf{y}$: Projecting $\mathbf{y}$ onto $C(\mathbf{X})$.

• $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{P}\mathbf{y} = (I - \mathbf{P})\mathbf{y}$: Projecting $\mathbf{y}$ onto the orthogonal complement of $C(\mathbf{X})$ with respect to $\mathbb{R}^n$.

Connection to $\hat{\boldsymbol{\beta}}$

♣ Th: $\hat{\boldsymbol{\beta}}$ is a least square estimate of $\boldsymbol{\beta}$ if and only if $\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{P}\mathbf{y}$, $= \hat{y}$ where $\mathbf{P}$ is the perpendicular projection operator onto $C(\mathbf{X})$.

♢ Cor: $(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y}$ is a least square estimate of $\boldsymbol{\beta}$.

$$\hat{y} = X\hat{\beta} = Py = \underbrace{X(X^TX)^-X^T}_{P}y$$

$$\hat{\beta} = (X^TX)^-X^Ty$$

$$\underbrace{(X^tX)}_{=A}\beta = \underbrace{X^Ty}_{=c}$$

★ Result A.13: Let $\mathbf{Ax} = \mathbf{c}$ be a consistent system of equations and let $\mathbf{G}$ be a generalized inverse of $\mathbf{A}$; then $\tilde{\mathbf{x}}$ is a solution to the equations $\mathbf{Ax} = \mathbf{c}$ if and only of there exists a vector $\mathbf{z}$ such that $\tilde{x}$ $= \mathbf{Gc} + (I - \mathbf{GA})\mathbf{z}$.

♠ **Implications?**

- By varying $\mathbf{z}$ over $\mathbb{R}^n$, we can obtain all possible solutions to a system of equations.

- The collection of solutions does not depend on the choice of generalized inverse.

- The family of solutions can be decomposed to two parts;
  - Solution to the system of equations $\mathbf{Ax} = \mathbf{c}$: $\mathbf{Gc}$
  - Solution to $\mathbf{Ax} = 0$: $(I - \mathbf{GA})\mathbf{z}$ $(\in \mathcal{N}(\mathbf{A}))$

★ Lem 2.1 $\mathcal{N}(\mathbf{X}) = \mathcal{N}(\mathbf{X}^T\mathbf{X})$

★ Recall: The NEs are

$$\mathbf{X}^T\mathbf{X}\beta = \mathbf{X}^T\mathbf{y}.$$

From the previous slide: The difference between two solutions of the NEs must be a vector in $\mathcal{N}(\mathbf{X})$

◇ Cor 2.3: $\underbrace{\mathbf{X}\hat{\beta}}_{=\hat{y}}$ is invariant to the choice of a solution $\hat{\beta}$ to the NEs.

$\hat{\beta}_1 \quad \hat{\beta}_2$

$X^T X \hat{\beta}_1 = X^T y$

$X^T X \hat{\beta}_2 = X^T y$

$\hat{y}_1 = X\hat{\beta}_1$

$\hat{y}_2 = X\hat{\beta}_2$

$(\hat{y}_1 - \hat{y}_2) = X\hat{\beta}_1 - X\hat{\beta}_2$

$= X(\hat{\beta}_1 - \hat{\beta}_2) = 0$

$(\hat{\beta}_1 - \hat{\beta}_2) \in \mathcal{N}(X)$

♣ Th: **[The Gram-Schmidt Theorem]:** Let $\mathcal{M}$ be a space with basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$. There exists an orthonormal basis for $\mathcal{M}$, say $\{\mathbf{y}_1, \ldots, \mathbf{y}_r\}$, with $\mathbf{y}_s$ in the space spanned by $\{\mathbf{x}_1, \ldots, \mathbf{x}_s\}$, $s = 1, \ldots, r$.

In layman's terms,
A set of linearly independent vectors $\Rightarrow$ a set of mutually orthogonal normalized vectors

★ Th Let $\mathbf{o}_1, \ldots, \mathbf{o}_r$ be an orthonormal basis for $C(\mathbf{X})$, and let $\mathbf{O} = [\mathbf{o}_1, \ldots, \mathbf{o}_r]$. Then $\mathbf{O}\mathbf{O}^T = \sum_{i=1}^r \mathbf{o}_i \mathbf{o}_i^T$ is the perpendicular projection operator onto $C(\mathbf{X})$.

♠ For any $\mathbf{X}$,

1. Get an orthonoraml basis for $C(\mathbf{X})$ by the Gram-Schmidt theorem
2. Obtain the perpendicular projection operator $\Rightarrow P$

   $\hat{y} = Py$, $\hat{\beta}X = Py$

For more details, see section 2.4 (M).

perpendicular proj matrix onto $\mathcal{C}(X)$

" " " $\mathcal{C}(W)$

♣ Th 2.2: If $C(\mathbf{W}) \subset C(\mathbf{X})$, then $\mathbf{P}_x - \mathbf{P}_w$ is the projection onto $C((I - \mathbf{P}_w)\mathbf{X})$.

$\boxed{\text{Ex 2.6}}$ Decomposition for simple linear regression

Model 1 $\quad y_i = \beta_0 + \beta_1 x_i + e_i$

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

① $\mathcal{C}(W) \subset \mathcal{C}(X)$ ✓

② P_X , P_W ✓

③ $(P_X - P_W)$

Model 0 $\quad y_i = \beta_0 + e_i'$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}}_{=W} \beta_0 + \begin{bmatrix} e_1' \\ e_2' \\ e_3' \\ e_4' \end{bmatrix}$$

② let's find P_X & P_W

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

$$P_W = W(W^TW)^{-1}W^T = \frac{1}{4}\underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}}_{=J_{4\times 4}}$$

$$P_W y = \begin{bmatrix} \bar{y} \\ \bar{y} \\ \bar{y} \\ \bar{y} \end{bmatrix}$$

③ $(P_X - P_W) = \frac{1}{100}\begin{bmatrix} 45 & & & \\ 15 & 5 & & \\ -15 & -5 & 5 & \\ -45 & -15 & -15 & 45 \end{bmatrix}$

$= \frac{1}{20}\begin{bmatrix} 3 \\ 1 \\ -1 \\ 3 \end{bmatrix}\begin{bmatrix} 3 & 1 & -1 & 3 \end{bmatrix}$

$\Rightarrow$ orthogonal projection matrix onto $\mathcal{C}((I - P_W)X)$

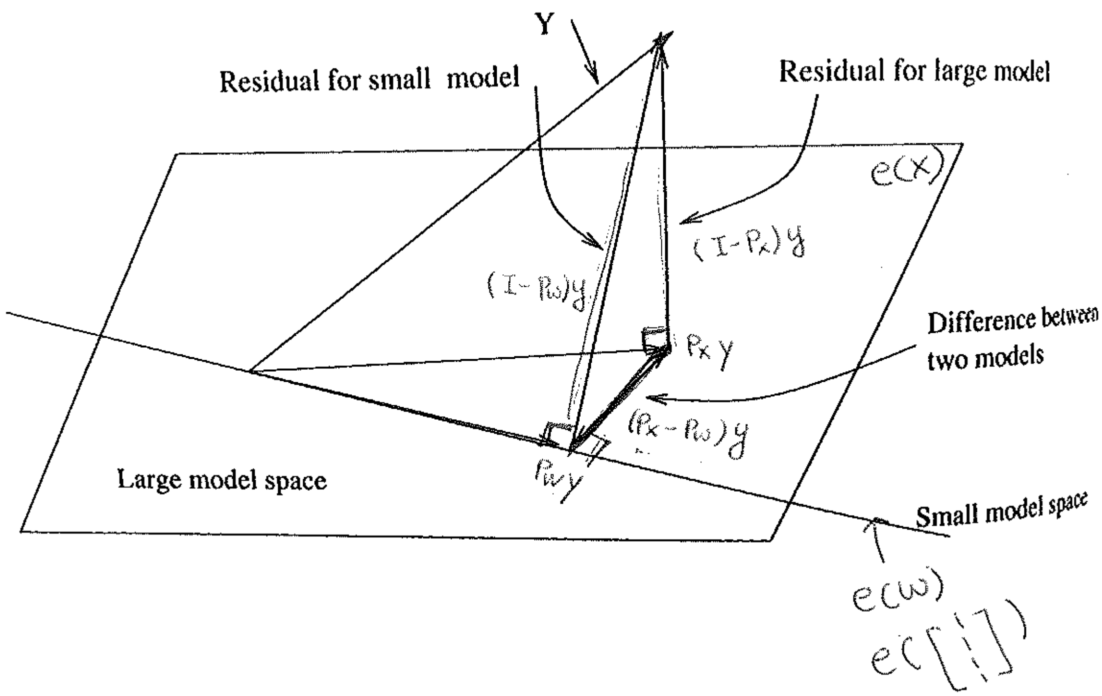
$\Rightarrow$ Very useful for subset regression, decomposition of SS and hypothesis testing

For general notation, let $W = \mathbf{1}$ & $P_W = P_{\mathbf{1}}$

④ $R^2 = \dfrac{\|(P_X - P_{\mathbf{1}})y\|^2}{\underbrace{\|(I - P_{\mathbf{1}})y\|^2}_{\text{Total SS (corrected for mean)}}} = \dfrac{\Sigma(\hat{y}_i - \bar{y})^2}{\Sigma(y_i - \bar{y})^2} = 1 - \dfrac{\Sigma(\hat{y}_i - y_i)^2}{\Sigma(y_i - \bar{y})^2}$

coefficient of determination

★ Figures help!



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, $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$ where $\mathbf{X}$ is a $n \times p$ matrix and $\mathbf{y} = \mathbf{W}\gamma + \mathbf{e}$ 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 1-way ANOVA

Model 1: $y_{ij} = \mu + \alpha_i + e_{ij}$, $i=1,2,3$, $j=1,\dots,n$ $\rightarrow Y = X\beta + e$

Model 2: $y_{ij} = \theta_i + e_{ij}$

$\rightarrow Y = W\gamma + e$

Find X and W $\Rightarrow$ check if $\mathcal{C}(X) = \mathcal{C}(W)$??

$$X = \begin{bmatrix} 1_n & 1_n & 0_n & 0_n \\ 1_n & 0_n & 1_n & 0_n \\ 1_n & 0_n & 0_n & 1_n \end{bmatrix} \qquad \beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

$\mathcal{C}(X) = \mathcal{C}(W)$ ✓

$$W = \begin{bmatrix} 1_n & 0_n & 0_n \\ 0_n & 1_n & 0_n \\ 0_n & 0_n & 1_n \end{bmatrix} \qquad \gamma = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$$