Recap: Linear models: one way anova, two way anova

Date Tuesday 4/10

$\underline{y} = \underline{X}\,\underline{\beta} + \underline{e}$ where $\underline{y}$ is an $n\times 1$ vector

$\underline{X}$: $n\times p$ matrix, $\underline{\beta}$: $p\times 1$ vector and $\underline{e}$: $n\times 1$ vector.

LSE of $\underline{\beta}$ is a solution to the normal equations

$$\underline{X}^T\underline{X}\,\underline{\beta} = \underline{X}^T\underline{y} \quad \text{(NEs)}$$

$$\underline{y} = \underline{X}\,\underline{\beta} + \underline{e}, \qquad \underline{X} = \begin{bmatrix} \underline{1}_n & \underline{1}_n & \underline{0}_n & \cdots & \underline{0}_n \\ \vdots & \underline{0}_n & \underline{1}_n^T & & \vdots \\ \vdots & \vdots & & & \underline{0}_n \\ \underline{1}_n & \underline{0}_n & \cdots & & \underline{1}_n \end{bmatrix}$$

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

$$= \begin{bmatrix} na & n & \cdots & \cdots & n \\ n & n & 0 & \cdots & 0 \\ n & 0 & n & 0 \cdots & 0 \\ \vdots & & & & \\ \vdots & & & & \\ n & 0 & \cdots & 0 & n \end{bmatrix}$$

$$\underline{X}^T\underline{y} = \begin{bmatrix} \underline{1}_n^T & \cdots & \cdots & \underline{1}_n^T \\ \underline{1}_n^T & \underline{0}_n^T & \cdots & \underline{0}_n^T \\ \underline{0}_n^T & \underline{1}_n^T & \cdots & \underline{0}_n^T \\ \vdots & & & \\ \underline{0}_n^T & \cdots & \underline{0}_n^T & \underline{1}_n^T \end{bmatrix} \begin{pmatrix} y_{11} \\ \vdots \\ y_{n1} \\ y_{12} \\ \vdots \\ y_{n2} \\ \vdots \\ y_{1a} \\ \vdots \\ y_{na} \end{pmatrix} = \begin{bmatrix} \sum_{i=1}^{n}\sum_{j=1}^{a} y_{ij} \\ \sum_{i=1}^{n} y_{i1} \\ \sum_{i=1}^{n} y_{i2} \\ \vdots \\ \sum_{i=1}^{n} y_{ia} \end{bmatrix}$$

NES in

$$\underline{X}^T \underline{X} \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_a \end{pmatrix} = \underline{X}^T \underline{y}.$$

<u>Questions</u>: ① $\underline{X}^T \underline{X} \beta = \underline{X}^T \underline{y}$ does it have a solution, and why?

② If there is a solution, is it unique? If yes, when?

③ If it is not unique, can we characterize the class of all possible solutions?

<u>vector space</u>: $S \subset \mathbb{R}^n$ is a set of vectors such that if $\underline{x}, \underline{y} \in S$ then $\alpha \underline{x} + \beta \underline{y} \in S$ for α, β scalars and $\underline{0} \in S$.

<u>Subspace</u>: A subspace of a vector space S is a subset of S that is also a vector space.

<u>Example</u>: $\mathbb{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}$

$(x_1, y_1, z_1) \in \mathbb{R}^3$, $(x_2, y_2, z_2) \in \mathbb{R}^3$

Does $\underbrace{\alpha(x_1, y_1, z_1) + \beta(x_2, y_2, z_2)}_{(\alpha x_1 + \beta x_2,\ \alpha y_1 + \beta y_2,\ \alpha z_1 + \beta z_2) \in \mathbb{R}^3} \in \mathbb{R}^3$ for all $\alpha, \beta \in \mathbb{R}$?

$\underline{0} = (0, 0, 0) \in \mathbb{R}^3$. $\Rightarrow \mathbb{R}^3$ is a vector space.

$\mathcal{M} = \{(x, y, 0) \mid x, y \in \mathbb{R}\}$. Is it a subspace of $\mathbb{R}^3$?

take any $(x, y, 0) \in \mathbb{R}^3 \Rightarrow \mathcal{M} \subset \mathbb{R}^3$.

$(x_1, y_1, 0)\ ;\ (x_2, y_2, 0) \in \mathcal{M}$

$\alpha(x_1, y_1, 0) + \beta(x_2, y_2, 0) = (\alpha x_1 + \beta x_2, \alpha y_1 + \beta y_2, 0) \in \mathcal{M}$

$\underline{0} = (0, 0, 0) \in \mathcal{M}$

$\mathcal{M}$ is a subspace.

If we take $\underline{x}_1, \ldots, \underline{x}_K \in S$. Define,

$$\mathcal{M} = \left\{ \underline{y} : \underline{y} = \sum_{j=1}^{K} c_j \underline{x}_j\ ,\ c_1, \ldots, c_K \text{ are coefficients} \right\}$$

then $\mathcal{M}$ is called the space spanned by $\underline{x}_1, \ldots, \underline{x}_K$.

<u>check</u>: $\mathcal{M}$ is a subspace of S.

<u>column space of a matrix</u>:

Let $\underline{A}$ be a matrix of dimension $m \times n$. The column space of $\underline{A}$, denoted by $C(\underline{A})$ is the vector space spanned by all columns of $\underline{A}$. Let,

$$\underline{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

$$C(\underline{A}) = \left\{ \underline{c} : \underline{c} = \sum_{j=1}^{n} x_j \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix} \right\}$$

$C(\underline{A})$ consists of all possible linear combinations of the columns of $\underline{A}$.

$$\sum_{j=1}^{n} x_j \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix}$$

$$= \underline{A}\,\underline{x} \qquad \underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$C(\underline{A}) = \{ \underline{c} : \underline{c} = \underline{A}\,\underline{x} \text{ for some } \underline{x} \}.$$

Example: $\underline{A} = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 0 & 0 \end{bmatrix}$ claim: $C(\underline{A}) = \left\{ \begin{pmatrix} a \\ b \\ 0 \end{pmatrix} : a, b \in R \right\}$

take any vector from the $C(\underline{A})$. By definition that vector has to be a linear combination of the two columns of $\underline{A}$. So it must be having the form

$$x_1 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_1 + 2x_2 \\ 0 \end{bmatrix} \in \left\{ \begin{pmatrix} a \\ b \\ 0 \end{pmatrix} : a, b \in R \right\}$$

$$\Rightarrow C(\underline{A}) \subset \left\{ \begin{pmatrix} a \\ b \\ 0 \end{pmatrix} : a, b \in R \right\}.$$

Now we will show $\left\{ \begin{pmatrix} a \\ b \\ 0 \end{pmatrix} : a, b \in R \right\} \subset C(\underline{A})$

take any $\begin{pmatrix} a \\ b \\ 0 \end{pmatrix}$; $a, b \in R$.

if $\begin{pmatrix} a \\ b \\ 0 \end{pmatrix}$ is a linear combination of the columns then it has to be of the form

$$x_1 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$$

thus we solve the system of equation

$$x_1 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ b \\ 0 \end{bmatrix}$$

$\Rightarrow \quad x_1 = a, \quad x_1 + 2x_2 = b \Rightarrow x_2 = \dfrac{b-a}{2}$

thus, $a \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \left(\dfrac{b-a}{2}\right) \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ b \\ 0 \end{bmatrix}$

$\Rightarrow \begin{pmatrix} a \\ b \\ 0 \end{pmatrix} \in C(\underline{A}) \quad \Rightarrow C(\underline{A}) = \left\{ \begin{pmatrix} a \\ b \\ 0 \end{pmatrix} : a, b \in R \right\}.$

Qn: does $\begin{pmatrix} 2 \\ 3 \\ -1 \end{pmatrix} \notin C(\underline{A})$.

$\boxed{\text{Note that } \underline{X}\,\underline{\beta} \in C(\underline{X}).}$

- Qn: when does a system of equations of the form $A\underline{x} = \underline{c}$ has a solution?

when $\underline{c} \in C(\underline{A})$

If a system of equations has a solution then that system is called consistent.

Recall the normal equation

$\underline{X}^T \underline{X} \underline{\beta} = \underline{X}^T \underline{y}$. Is this system consistent?

In other words does $\underline{X}^T \underline{y} \in C(\underline{X}^T \underline{X})$

$\underline{X}^T \underline{y} \in C(\underline{X}^T) = C(\underline{X}^T \underline{X})$ (we will prove it later)

$\Rightarrow \underline{X}^T \underline{y} \in C(\underline{X}^T \underline{X}) \Rightarrow$ this system is consistent.

Linear dependence:

Let $\underline{x}_1, \ldots, \underline{x}_k$ be vectors in S. If there exists scalars $\alpha_1, \ldots, \alpha_k$ not all zero such that $\sum_{i=1}^{k} \alpha_i \underline{x}_i = \underline{0}$, then $\underline{x}_1, \ldots, \underline{x}_k$ are known as linearly dependent.

$\alpha_k x_k = - \sum_{i \neq k} \alpha_i x_i$

Look at $\underbrace{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}}_{\underline{x}_1}, \underbrace{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}}_{\underline{x}_2}, \underbrace{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}}_{\underline{x}_3}$. Are they linearly dependent?

$\sum_{i=1}^{3} \alpha_i \underline{x}_i = \underline{0} \Rightarrow \alpha_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \alpha_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

$$\Rightarrow \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$\Rightarrow$ If there is some $\alpha_1, \alpha_2, \alpha_3$ which satisfies the equations then all of them have to be zero.

these vectors are not linearly dependent. We call them linearly independent.

Ex:

$$\underbrace{\begin{pmatrix}1\\0\\0\end{pmatrix}}_{\underline{x}_1}, \quad \underbrace{\begin{pmatrix}0\\1\\-1\end{pmatrix}}_{\underline{x}_2}, \quad \underbrace{\begin{pmatrix}0\\0\\1\end{pmatrix}}_{\underline{x}_3}$$

$$\sum_{i=1}^{3} \alpha_i \underline{x}_i = \underline{0} \Rightarrow \alpha_1 \begin{pmatrix}1\\0\\0\end{pmatrix} + \alpha_2 \begin{pmatrix}0\\1\\-1\end{pmatrix} + \alpha_3 \begin{pmatrix}0\\0\\1\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}$$

$\Rightarrow \alpha_1 = 0, \quad \alpha_2 = 0, \quad -\alpha_2 + \alpha_3 = 0 \Rightarrow \alpha_3 = 0$

$\Rightarrow \underline{x}_1, \underline{x}_2, \underline{x}_3$ are linearly independent.

Ex! $\underline{x}_1 = \begin{pmatrix}1\\-1\end{pmatrix}, \quad \underline{x}_2 = \begin{pmatrix}1\\2\end{pmatrix}, \quad \underline{x}_3 = \begin{pmatrix}2\\1\end{pmatrix}$

$\underline{x}_1 + \underline{x}_2 = \underline{x}_3 \Rightarrow \underline{x}_1 + \underline{x}_2 - \underline{x}_3 = \underline{0}$

there exists $\alpha_1, \alpha_2, \alpha_3$ $(\alpha_1 = 1, \alpha_2 = 1, \alpha_3 = -1)$

s.t. $\sum_{i=1}^{3} \alpha_i \underline{x}_i = \underline{0}$

$\Rightarrow \underline{x}_1, \underline{x}_2, \underline{x}_3$ are linearly dependent.

Basis: If M is a subspace of S and if $\{\underline{x}_1, \dots, \underline{x}_n\}$ is a set of linearly independent vectors which span M then $\{\underline{x}_1, \dots, \underline{x}_n\}$ is called a basis for M.

Ex: $\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}$. They are linearly independent.

What is the span of $\left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}\right\}$?

$\alpha_1, \alpha_2, \alpha_3 \in R$

$$\alpha_1 \begin{pmatrix}1\\0\\0\end{pmatrix} + \alpha_2 \begin{pmatrix}0\\1\\0\end{pmatrix} + \alpha_3 \begin{pmatrix}0\\0\\1\end{pmatrix} = \begin{pmatrix}\alpha_1\\\alpha_2\\\alpha_3\end{pmatrix} \in R^3$$

$\Rightarrow$ ~~Basis~~ A basis of R^3 is $\left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}\right\}$.

Basis is not unique.

Check: $\left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\-1\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}\right\}$ is also a basis for R^3.

Def: For any $m \times n$ matrix $\underline{A}$, rank of the matrix $\underline{A}$, denoted by $r(\underline{A})$, is the number of linearly independent rows or columns of $\underline{A}$.

If $\underline{A}$ is an $m \times n$ matrix and $r(\underline{A}) = m$, then we say that $\underline{A}$ has the full row rank.

If $r(\underline{A}) = n$ then we say that $\underline{A}$ has full column rank.

If $\underline{A}$ is an $n \times n$ matrix, we call $\underline{A}$ nonsingular if there exists a matrix $\underline{A}^{-1}$ s.t. $\underline{A}\,\underline{A}^{-1} = \underline{A}^{-1}\underline{A} = \underline{I}$.

If no such matrix exists then we call $\underline{A}$ a singular matrix.

$\underline{A}^{-1}$ is called the inverse of $\underline{A}$.

For an $n \times n$ matrix, $\underline{A}$ is nonsingular iff $r(\underline{A}) = n$. If $\underline{A}$ is singular $\Rightarrow r(\underline{A}) < n$.

Ex: $\underline{A} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 5 & 5 \\ -2 & 1 & 3 \end{pmatrix}$

1st column + 3rd column = 2nd column.

$r(\underline{A}) < 3$

$$\alpha_1 \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 5 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow 5\alpha_2 = 0 \Rightarrow \alpha_2 = 0 \Rightarrow \alpha_1 = 0$$

$\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 \\ 5 \\ 3 \end{pmatrix}$ are linearly independent

$\Rightarrow r(\underline{A}) = 2. \Rightarrow \underline{A}$ is singular.

$\underline{A} = \begin{pmatrix} 2 & -2 \\ 3 & 4.5 \end{pmatrix} \Rightarrow r(\underline{A}) = 2 \Rightarrow \underline{A}$ has an inverse

Recall that the linear regression model
$\underline{y} = \underline{X}\,\underline{\beta} + \underline{e}$ which leads to NEs:

$$\underset{p\times n}{\underline{X}^T}\,\underset{n\times p}{\underline{X}}\,\underset{p\times 1}{\underline{\beta}} = \underset{p\times n}{\underline{X}^T}\,\underset{n\times 1}{\underline{y}}$$

$\underline{X}^T\underline{X}$ is a $p\times p$ matrix

$r(\underline{X}^T\underline{X})$ = number of linearly independent columns of $\underline{X}^T\underline{X}$

= number of linearly independent columns of $\underline{X}^T$ $\left(C(\underline{X}^T) = C(\underline{X}^T\underline{X}) \right)$

$= r(\underline{X}^T) = r(\underline{X})$.

Typically in linear regressions $n \gg p$

If all columns of $\underline{X}$ are linearly independent

$\Rightarrow r(\underline{X}) = p$. $\Rightarrow r(\underline{X}^T\underline{X}) = p$

Note that $\underline{X}^T\underline{X}$ is a $p\times p$ matrix.

$\Rightarrow$ $\underline{X}^T\underline{X}$ is a full rank matrix

$\Rightarrow$ $\underline{X}^T\underline{X}$ has an inverse.

in that case $\hat{\beta} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{y}$.

If all columns of $\underline{X}$ are not linearly independent then $r(\underline{X}^T\underline{X}) < p \Rightarrow \underline{X}^T\underline{X}$ is not invertible. In this case there is no unique solution

Example: $y_{ij} = \mu + \alpha_i + e_{ij}$, $i = 1, \ldots, n$, $j = 1, \ldots, a$

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

$$\underline{X} = \begin{bmatrix} \underline{1}_n & \underline{1}_n & \underline{0}_n & \cdots & \underline{0}_n \\ \vdots & \underline{0}_n & \underline{1}_n & & \vdots \\ \vdots & \vdots & \underline{0}_n & & \underline{0}_n \\ \underline{1}_n & \underline{0}_n & \underline{0}_n & \cdots & \underline{1}_n \end{bmatrix}$$

1st colum = sum of other colums

$r(\underline{X}) < a+1$

$$\alpha_1 \begin{pmatrix} \underline{1}_n \\ \underline{0}_n \\ \vdots \\ \underline{0}_n \end{pmatrix} + \alpha_2 \begin{pmatrix} \underline{0}_n \\ \underline{1}_n \\ \underline{0}_n \\ \vdots \\ \underline{0}_n \end{pmatrix} + \cdots + \alpha_a \begin{pmatrix} \underline{0}_n \\ \vdots \\ \underline{0}_n \\ \underline{1}_n \end{pmatrix} = \begin{pmatrix} \underline{0}_n \\ \vdots \\ \underline{0}_n \end{pmatrix}$$

$\Rightarrow \alpha_1 = 0, \alpha_2 = 0, \ldots, \alpha_a = 0$

from 2nd to $(a+1)$ colums are linealy independent.

$\Rightarrow r(\underline{X}) = a$

The Normal equations in the case of one way anova do not have a unique solution

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

① Does NEs have any solution?

② When is the solution unique?

We saw the solution is unique if $r(\underline{X}^T\underline{X}) = p$.

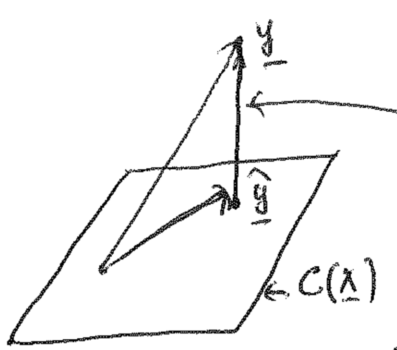
③ When $r(\underline{X}^T\underline{X}) < p$, what are the class of all solutions to NEs.

Instead of finding β, we will start looking at $\hat{\underline{y}} \in C(\underline{X})$. $\hat{\underline{y}}$ will be such that it is closest to $\underline{y}$ in the column space of $\underline{X}$.

Once we find $\hat{\underline{y}}$, we will characterize the class of all $\hat{\beta}$ s.t. $\hat{\underline{y}} = X\hat{\beta}$.

Let $n = 3$, $\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$, let there be two predictors x_1 and x_2

$$\underline{X} = \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ x_{13} & x_{23} \end{bmatrix}$$



$\underline{y} - \hat{\underline{y}} = \hat{\underline{e}}$

We know, that $\underline{X}^T X \hat{\underline{\beta}} = \underline{X}^T \underline{y}$ for any solution $\hat{\beta}$ of the Normal equations.

$\hat{\underline{y}} = \underline{X}\hat{\underline{\beta}}$, $\hat{e} = \underline{y} - \underline{X}\hat{\underline{\beta}}$

$\Rightarrow \underline{X}^T(\underline{y} - \underline{X}\hat{\beta}) = \underline{0}$

$\Rightarrow \underline{X}^T\hat{\underline{e}} = \underline{0}$ - - - - - (*)

Null space of a matrix

If $\underline{A}$ is an $m\times n$ matrix, the null space $\mathcal{N}(\underline{A})$ of $\underline{A}$ is given by $\mathcal{N}(\underline{A}) = \{\underline{y} : \underline{A}\,\underline{y} = \underline{0}\}$

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

$\underline{A}\,\underline{y} = \underline{0} \Rightarrow \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ -1 & 1 & -3 \\ 1 & 2 & 0 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$

$$\Rightarrow \begin{aligned} y_1 + y_2 + y_3 &= 0 \quad \cdots\cdots (1) \\ 2y_1 + 2y_2 + 2y_3 &= 0 \quad \cdots\cdots (2) \\ -y_1 + y_2 - 3y_3 &= 0 \quad \cdots\cdots (3) \\ y_1 + 2y_2 &= 0 \quad \cdots\cdots (4) \end{aligned}$$

From (4) $\Rightarrow$ $y_1 = -2y_2$

use this in (3), to obtain $3y_3 = 3y_2 \Rightarrow y_2 = y_3$

$y_1 = -2y_2, \quad y_3 = y_2$

$\Rightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} -2y_2 \\ y_2 \\ y_2 \end{pmatrix} = y_2 \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$

Any vector $\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \in \mathcal{N}(\underline{A})$ has to have the form

$y_2 \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$, $\Rightarrow \mathcal{N}(\underline{A}) = \left\{ c \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} : c \in \mathbb{R} \right\}$.

$\begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$ is the only basis of $\mathcal{N}(\underline{A})$.

Now go back to (*) which said $\underline{X}^T \hat{\underline{e}} = \underline{0}$.

using definition, $\hat{\underline{e}} \in \mathcal{N}(\underline{X}^T)$.

$\hat{\underline{y}} = \underline{X}\hat{\underline{\beta}} \in C(\underline{X})$.

<u>Def:</u> Two vectors $\underline{x}, \underline{y}$ are said to be orthogonal if $\underline{x}^T \underline{y} = \underline{0}$.

<u>Def:</u> (orthogonal spaces)

Two subspaces $\mathcal{M}_1$ & $\mathcal{M}_2$ are orthogonal spaces if $\underline{x} \in \mathcal{M}_1$ and $\underline{y} \in \mathcal{M}_2$ means $\underline{x}^T \underline{y} = \underline{0}$.

<u>Notation:</u> Let S be a vector space and let $\mathcal{M}$ be a subspace of S. $\mathcal{M}_S^{\perp} = \{ \underline{y} \in S \mid \underline{y}$ is orthogonal to any vector in $\mathcal{M} \}$.

is used to denote the orthogonal space of $\mathcal{M}$.

$\perp$ is used to denote orthogonality.

<u>Thm:</u> For any matrix $\underline{A}_{m \times n}$, $C(\underline{A})$ and $\mathcal{N}(\underline{A}^T)$ are orthogonal spaces.

<u>Pf:</u> $\underline{v} \in C(\underline{A})$ and $\underline{w} \in \mathcal{N}(\underline{A}^T)$

Since $\underline{v} \in C(\underline{A})$, there exists some $\underline{h}$ s.t.

$$\underline{v} = \underline{A}\,\underline{h} \Rightarrow \underline{v}^T \underline{w} = (\underline{A}\,\underline{h})^T \underline{w} = \underline{h}^T \underbrace{\underline{A}^T \underline{w}}_{= \underline{0}}$$

Now $\underline{A}^T \underline{w} = \underline{0}$ as $\underline{w} \in \mathcal{N}(\underline{A}^T)$

Result: $C(\underline{A}) \cap \mathcal{N}(\underline{A}^T) = \{\underline{0}\}$

Pf:- $\underline{v} \in C(\underline{A}) \cap \mathcal{N}(\underline{A}^T)$ $\left[\underline{w} \in C(\underline{A}),\ \underline{z} \in \mathcal{N}(\underline{A}^T) \Rightarrow \underline{w}^T \underline{z} = 0 \right]$

$\underline{v} \in C(\underline{A})$ and $\underline{v} \in \mathcal{N}(\underline{A}^T)$

$\underline{v}^T \underline{v} = \underline{0} \Rightarrow \|\underline{v}\|^2 = 0 \Rightarrow \underline{v} = \underline{0}$

$\Rightarrow$ that $C(\underline{X})$ and $\mathcal{N}(\underline{X}^T)$ are orthogonal spaces and they have no vector in common other than $\underline{0}$.

Thm: Let S be a vector space. Let $\mathcal{M}$ be a subspace of $\underline{x}$. $\mathcal{M}_S^{\perp}$ denotes the orthogonal complement of $\mathcal{M}$ in S.

If $\underline{x} \in S$ then $\underline{x}$ can be written uniquely as $\underline{x} = \underline{x}_0 + \underline{x}_1$ with $\underline{x}_0 \in \mathcal{M}$ and $\underline{x}_1 \in \mathcal{M}_S^{\perp}$.

Ex: $\underline{A} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & 3 \\ 1 & 0 & 3 \end{bmatrix}$ $C(\underline{A})$ and $\mathcal{N}(\underline{A}^T)$.

$C(\underline{A})$: $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} y_1 + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} y_2 + \begin{pmatrix} 2 \\ 3 \\ 3 \end{pmatrix} y_3$

$$= \begin{pmatrix} y_1 + y_2 + 2y_3 \\ y_1 + 3y_3 \\ y_1 + 3y_3 \end{pmatrix}$$

$C(\underline{A}) = \left\{ \begin{pmatrix} a \\ b \\ b \end{pmatrix} : a, b \in \mathbb{R} \right\}$

If $\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \in \mathcal{N}(\underline{A}^T) \Rightarrow \underline{A}^T \underline{y} = \underline{0}$

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 2 & 3 & 3 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \begin{array}{ll} y_1 + y_2 + y_3 = 0 & \cdots (1) \\ y_1 = 0 & \cdots (2) \\ 2y_1 + 3y_2 + 3y_3 = 0 & \cdots (3) \end{array}$$

If we use (2) then both (1) & (3) will give rise to

$y_2 + y_3 = 0 \Rightarrow \cancel{y_3 = 0}\ y_2 = -y_3$

$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \in \mathcal{N}(\underline{A}^T)$ has to have the form $\begin{pmatrix} 0 \\ -y_3 \\ y_3 \end{pmatrix}$

$$\mathcal{N}(\underline{A}^T) = \left\{ \begin{pmatrix} 0 \\ -b \\ b \end{pmatrix} : b \in \mathbb{R} \right\}$$

If you take $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \underbrace{\begin{pmatrix} x \\ \frac{y+z}{2} \\ \frac{y+z}{2} \end{pmatrix}}_{\in C(\underline{A})} \cancel{+ \underbrace{\begin{pmatrix} 0 \\ -\frac{y+z}{2} \\ \frac{y+z}{2} \end{pmatrix}}_{\in \mathcal{N}(\underline{A}^T)}} + \underbrace{\begin{pmatrix} 0 \\ \frac{y-z}{2} \\ \frac{-y+z}{2} \end{pmatrix}}_{\in \mathcal{N}(\underline{A}^T)}$$

$$b = -\frac{(y-z)}{2}$$

Messages so far: Suppose $\hat{\beta}$ is a solution to

NEs $\underline{X}^T \underline{X} \hat{\underline{\beta}} = \underline{X}^T \underline{y}$ and $\underline{y} \in \mathbb{R}^n$, $\underline{X}_{n \times p}$ matrix

(1) $\hat{\underline{y}} = \underline{X} \hat{\underline{\beta}} \in C(\underline{X})$ (2) $\underline{X}^T \hat{\underline{e}} = \underline{0} \Rightarrow \hat{\underline{e}} \in \mathcal{N}(\underline{X}^T)$

$\hat{\underline{e}} = \underline{y} - \hat{\underline{y}}$

(3) $\hat{\underline{y}}$ and $\hat{\underline{e}}$ are orthogonal (4) $\underline{y} = \hat{\underline{y}} + \hat{\underline{e}}$

$\underline{y} = \underline{\hat{y}} + \underline{\hat{e}}$ and $\underline{\hat{y}}^T \underline{\hat{e}} = \underline{0}$

$$\|\underline{y}\|^2 = \|\underline{\hat{y}} + \underline{\hat{e}}\|^2 = (\underline{\hat{y}} + \underline{\hat{e}})^T (\underline{\hat{y}} + \underline{\hat{e}})$$

$$= \underline{\hat{y}}^T \underline{\hat{y}} + \underline{\hat{e}}^T \underline{\hat{e}} + \underbrace{\underline{\hat{y}}^T \underline{\hat{e}}}_{0} + \underbrace{\underline{\hat{e}}^T \underline{\hat{y}}}_{0} = \|\underline{\hat{y}}\|^2 + \|\underline{\hat{e}}\|^2$$

$\|\underline{\hat{y}}\|^2 = \|\underline{X}\underline{\hat{\beta}}\|^2$, $\|\underline{\hat{e}}\|^2 = \|\underline{y} - X\underline{\hat{\beta}}\|^2$

$\|\underline{\hat{y}}\|^2 = \|\underline{X}\underline{\hat{\beta}}\|^2$: the regression sum of squares (SSR)

$\|\underline{\hat{e}}\|^2 = \|\underline{y} - \underline{X}\underline{\hat{\beta}}\|^2$: The error sum of squares (SSE)

Qn: Is there any mechanical way to find $\underline{\hat{y}}$ from $\underline{y}$. In other words, is there a matrix $\underline{M}$ s.t. $\underline{\hat{y}} = \underline{M}\,\underline{y}$

$\underline{M}$ matrix projects any vector $\underline{y} \in \mathbb{R}^n$ onto the column space of $\underline{X}$. These type of matrices are called projection matrices.

Definition:

A square matrix $\underline{M}$ is a perpendicular projection matrix onto $C(\underline{X})$ if and only if

(i) $\underline{v} \in C(\underline{X}) \Rightarrow \underline{M}\,\underline{v} = \underline{v}$.

(ii) $\underline{w} \perp C(\underline{X}) \Rightarrow \underline{M}\,\underline{w} = \underline{0}$.

Thm: $C(\underline{M}) = C(\underline{X})$

Thm: $\underline{M}$ is a projection matrix if and only if

(i) $\underline{M}^2 = \underline{M}.\underline{M} = \underline{M}$ (idempotent) (ii) $\underline{M}^T = \underline{M}$ (symmetric)

Thm: Projection matrix $\underline{M}$ is unique.

Ex: $\underline{X} = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}$, $r(\underline{X}) = 1$

Consider $\underline{M} = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix}$

$$\underline{M}^2 = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix}\begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} = \underline{M}$$

$\underline{M}^T = \underline{M}$

$$C(\underline{X}) = \left\{ a\begin{pmatrix} 2 \\ 1 \end{pmatrix} : a \in R \right\}.$$

any vector $\begin{pmatrix} 2a \\ a \end{pmatrix}$ in the column space of $\underline{X}$

$$\underline{M}\begin{pmatrix} 2a \\ a \end{pmatrix} = \begin{pmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{pmatrix}\begin{pmatrix} 2a \\ a \end{pmatrix} = \begin{pmatrix} 2a \\ a \end{pmatrix}$$

$\mathcal{N}(\underline{X}^T)$: $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathcal{N}(\underline{X}^T) \Rightarrow \underline{X}^T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$\Rightarrow \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{array}{ll} 2x_1 + x_2 = 0 & \cdots (1) \\ 4x_1 + 2x_2 = 0 & \cdots (2) \end{array}$$

$$\Rightarrow x_2 = -2x_1 \qquad \mathcal{N}(\underline{X}^T) = \left\{ \begin{pmatrix} a \\ -2a \end{pmatrix} : a \in R \right\}.$$

$$\underline{M}\begin{pmatrix} a \\ -2a \end{pmatrix} = \begin{pmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{pmatrix}\begin{pmatrix} a \\ -2a \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$\underline{M}$ is clearly the projection matrix onto the column space of $\underline{X}$.

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

$$\underline{M} = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix}$$

$$C(\underline{M}) = \left\{ a\begin{pmatrix} 0.4 \\ 0.2 \end{pmatrix} : a \in \mathbb{R} \right\} = \left\{ c\begin{pmatrix} 2 \\ 1 \end{pmatrix} : c \in \mathbb{R} \right\}$$

$$= C(\underline{X}).$$

$\underline{X}^T \underline{X}\, \hat{\beta} = \underline{X}^T \underline{y}$ $\quad \hat{\underline{y}} \in C(\underline{X}) = X\hat{\underline{\beta}}$

$\hat{\underline{e}} = \underline{y} - X\hat{\underline{\beta}}$ s.t. $\hat{\underline{y}} \perp \hat{\underline{e}}$ and $\underbrace{\underline{y} = \hat{\underline{y}} + \hat{\underline{e}}}_{\text{unique decomposition}}$.

we defined SSR, SSE.

Some matrix $\underline{M}$ s.t. $\hat{\underline{y}} = M\underline{y}$.

1. Some properties of $\underline{M}$.
2. Find the explicit formula for the projection matrix

$\Downarrow$

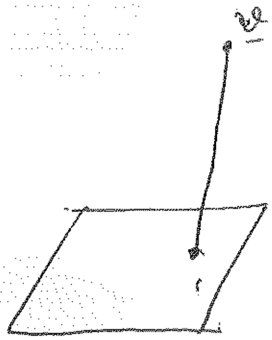
class of all solutions to the NEs.

<u>Thm:</u> If $\underline{M}$ be a projection matrix to a space $\mathcal{M}_1$ and $\underline{M}_0$ be a projection matrix to a space $\mathcal{M}_2$ s.t. $C(\underline{M}_0) \subset C(\underline{M})$

then $\underline{M} - \underline{M}_0$ is also a projection matrix.

$\underline{M}_1$ and $\underline{M}_2$ are the two projection matrices

$(\underline{M}_1 - \underline{M}_2)\underline{v} = \underline{0}$ For all $\underline{v}$.



$$\underline{M}_1 = \begin{bmatrix} m_{11} & \cdots & m_{1n} \\ \vdots & & \vdots \\ m_{n1} & \cdots & m_{nn} \end{bmatrix}$$

$$\underline{M}_2 = \begin{bmatrix} m'_{11} & \cdots & m'_{1n} \\ \vdots & & \vdots \\ m'_{n1} & \cdots & m'_{nn} \end{bmatrix}$$

$$(\underline{M}_1 - \underline{M}_2)\underline{v} = \underline{0} \quad \text{for all } \underline{v}$$

$$\underline{v} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Rightarrow \underline{M}_1 \underline{v} = \begin{pmatrix} m_{11} \\ \vdots \\ m_{n1} \end{pmatrix}$$

$$\underline{M}_2 \underline{v} = \begin{pmatrix} m'_{11} \\ \vdots \\ m'_{n1} \end{pmatrix}$$

$$\underline{M}_1 \underline{v} = \underline{M}_2 \underline{v}$$

$$\Rightarrow \begin{pmatrix} m_{11} \\ \vdots \\ m_{n1} \end{pmatrix} = \begin{pmatrix} m'_{11} \\ \vdots \\ m'_{n1} \end{pmatrix}$$

$$\underline{v} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$