• **Matrix algebra**

$$A = [\,\mathbf{a_1} \quad \mathbf{a_2} \quad \cdots \quad \mathbf{a_n}\,] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = (a_{ij}),\ A' = \begin{bmatrix} \mathbf{a_1}' \\ \mathbf{a_2}' \\ \mathbf{a_3}' \\ \vdots \\ \mathbf{a_n}' \end{bmatrix}$$

①

$$C = A'A = \begin{bmatrix} \mathbf{a_1}' \\ \mathbf{a_2}' \\ \mathbf{a_3}' \\ \vdots \\ \mathbf{a_n}' \end{bmatrix} [\,\mathbf{a_1} \quad \mathbf{a_2} \quad \cdots \quad \mathbf{a_n}\,] = \begin{bmatrix} a_{11} & a_{21} & a_{31} & \cdots & a_{n1} \\ a_{12} & a_{22} & a_{32} & \cdots & a_{n2} \\ a_{13} & a_{23} & a_{33} & \cdots & a_{n3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix} =$$

$$= \begin{bmatrix} \sum a_{i1}^2 & \sum a_{i1}a_{i2} & \sum a_{i1}a_{i3} & \cdots & \sum a_{i1}a_{in} \\ \sum a_{i2}a_{i1} & \sum a_{i2}^2 & \sum a_{i2}a_{i3} & \cdots & \sum a_{i2}a_{in} \\ \sum a_{i3}a_{i1} & \sum a_{i3}a_{i2} & \sum a_{i3}^2 & \cdots & \sum a_{i3}a_{in} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum a_{in}a_{i1} & \sum a_{in}a_{i2} & \sum a_{in}a_{i3} & \cdots & \sum a_{in}^2 \end{bmatrix} = \begin{bmatrix} \mathbf{a_1}'\mathbf{a_1} & \mathbf{a_1}'\mathbf{a_2} & \mathbf{a_1}'\mathbf{a_3} & \cdots & \mathbf{a_1}'\mathbf{a_n} \\ \mathbf{a_2}'\mathbf{a_1} & \mathbf{a_2}'\mathbf{a_2} & \mathbf{a_2}'\mathbf{a_3} & \cdots & \mathbf{a_2}'\mathbf{a_n} \\ \mathbf{a_3}'\mathbf{a_1} & \mathbf{a_3}'\mathbf{a_2} & \mathbf{a_3}'\mathbf{a_3} & \cdots & \mathbf{a_3}'\mathbf{a_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{a_n}'\mathbf{a_1} & \mathbf{a_n}'\mathbf{a_2} & \mathbf{a_n}'\mathbf{a_3} & \cdots & \mathbf{a_n}'\mathbf{a_n} \end{bmatrix} \therefore (c_{ij}) = \mathbf{a_i}'\mathbf{a_j}$$

②

$$C = AB = [\,\mathbf{a_1} \quad \cdots \quad \mathbf{a_n}\,][\,\mathbf{b_1} \quad \cdots \quad \mathbf{b_n}\,] = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1j} & \cdots & b_{1n} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ b_{i1} & \cdots & b_{ij} & \cdots & b_{in} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ b_{n1} & \cdots & b_{nj} & \cdots & b_{nn} \end{bmatrix}$$

$$\text{Let} \begin{bmatrix} a_{i1} \\ \vdots \\ a_{ij} \\ \vdots \\ a_{in} \end{bmatrix} = \mathbf{a_{i\bullet}} \rightarrow \begin{bmatrix} \mathbf{a_{1\bullet}}' \\ \vdots \\ \mathbf{a_{i\bullet}}' \\ \vdots \\ \mathbf{a_{n\bullet}}' \end{bmatrix} [\,\mathbf{b_1} \quad \cdots \quad \mathbf{bj} \quad \cdots \quad \mathbf{b_n}\,] = \begin{bmatrix} \mathbf{a_{1\bullet}}'\mathbf{b_1} & \cdots & \mathbf{a_{i\bullet}}'\mathbf{b_j} & \cdots & \mathbf{a_{1\bullet}}'\mathbf{b_n} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ \mathbf{a_{i\bullet}}'\mathbf{b_1} & \cdots & \mathbf{a_{i\bullet}}'\mathbf{b_j} & \cdots & \mathbf{a_{i\bullet}}'\mathbf{b_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{a_{n\bullet}}'\mathbf{b_1} & \cdots & \mathbf{a_{n\bullet}}'\mathbf{b_j} & \cdots & \mathbf{a_{n\bullet}}'\mathbf{b_n} \end{bmatrix} \therefore (c_{ij}) = \mathbf{a_{i\bullet}}'\mathbf{b_j}$$

• **Simple linear regression with matrix expression**

$$\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} = [\,\mathbf{1} \quad \mathbf{x}\,] \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \boldsymbol{\varepsilon} = \beta_0\mathbf{1} + \beta_1\mathbf{x} + \boldsymbol{\varepsilon} = \leftrightarrow \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix},\ \begin{pmatrix} \boldsymbol{\varepsilon} \sim N(\mathbf{0},\ \mathbf{I}_n\sigma^2) \\ \mathbf{y} \sim N(\mathbf{X\beta},\ \mathbf{I}_n\sigma^2) \end{pmatrix}$$

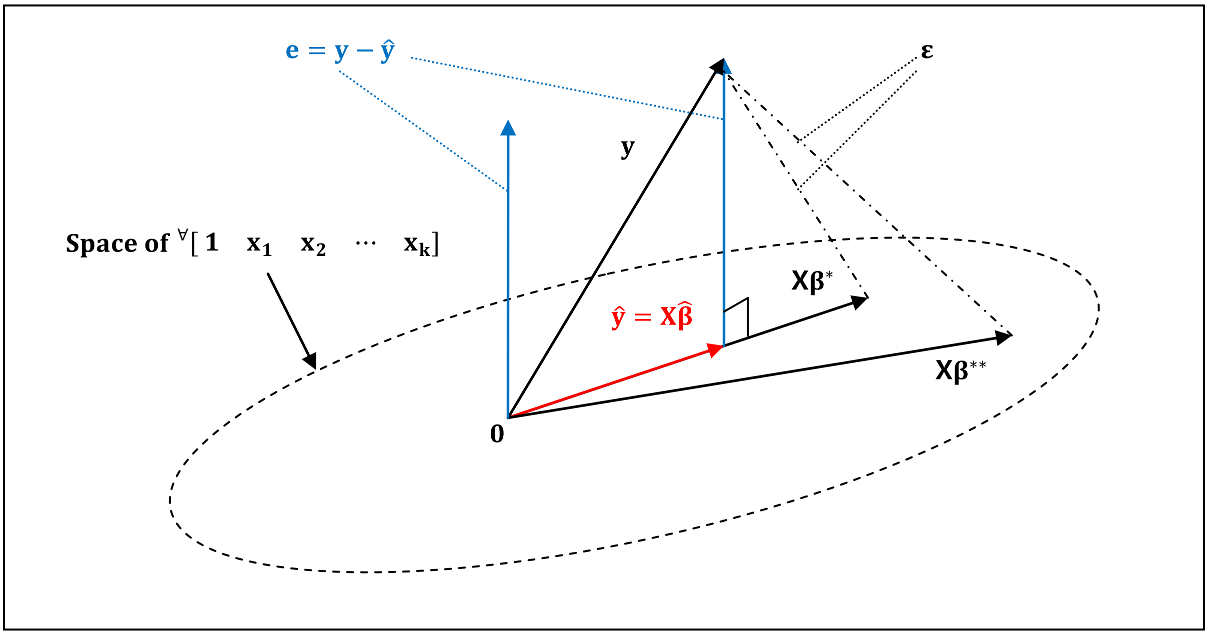
- **Multiple linear regression with matrix expression**

$$\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} = [\,\mathbf{1} \quad \mathbf{x_1} \quad \mathbf{x_2} \quad \cdots \quad \mathbf{x_k}\,]\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \boldsymbol{\varepsilon} = \beta_0\mathbf{1} + \beta_1\mathbf{x_1} + \cdots + \beta_k\mathbf{x_k} + \boldsymbol{\varepsilon}\ ,\ \begin{pmatrix} \boldsymbol{\varepsilon}\sim N(\mathbf{0}\,,\,\mathbf{I}_n\sigma^2) \\ \mathbf{y}\sim N(\mathbf{X\beta}\,,\,\mathbf{I}_n\sigma^2) \end{pmatrix}$$

$$\leftrightarrow \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ 1 & x_{31} & x_{32} & \cdots & x_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}\ ,\ \mathrm{Var}(\boldsymbol{\varepsilon}) = \begin{bmatrix} \boldsymbol{\sigma}^2 & 0 & 0 & \cdots & 0 \\ 0 & \boldsymbol{\sigma}^2 & 0 & \cdots & 0 \\ 0 & 0 & \boldsymbol{\sigma}^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \boldsymbol{\sigma}^2 \end{bmatrix}$$

- **LSE Development**

① Geometric method



$\hat{\mathbf{y}}(=\mathbf{X}\hat{\boldsymbol{\beta}}) \perp \mathbf{e}(=\mathbf{y}-\hat{\mathbf{y}}) \leftrightarrow \hat{\mathbf{y}}'\cdot\mathbf{e} = \mathbf{0}$ **- Inner product between two orthogonal vectors is zero.**

From above result, we can draw followings

$\hat{\mathbf{y}}'\cdot\mathbf{e} = (\mathbf{X}\hat{\boldsymbol{\beta}})'\cdot(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\cdot(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0} \leftrightarrow (\hat{\boldsymbol{\beta}}')\mathbf{X}'\mathbf{y} = (\hat{\boldsymbol{\beta}}')\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \leftrightarrow \mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \leftrightarrow \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = (\,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,)\mathbf{y} = \mathbf{H}\mathbf{y}$, let $\mathbf{H}$: Hat matrix or projection matrix

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

② Differentiation method

Let $\mathbf{S}(\boldsymbol{\beta}) = (\mathbf{y}-\mathbf{X\beta})'\cdot(\mathbf{y}-\mathbf{X\beta}) = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X\beta} = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X\beta}$

$$\left.\frac{\partial \mathbf{S}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \mathbf{0} \leftrightarrow \left.\frac{\partial(\mathbf{y}'\mathbf{y} - 2(\mathbf{y}'\mathbf{X})\boldsymbol{\beta} + \boldsymbol{\beta}'(\mathbf{X}'\mathbf{X})\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \mathbf{0} - 2\mathbf{X}'\mathbf{y} + [(\mathbf{X}'\mathbf{X}) + (\mathbf{X}'\mathbf{X})']\boldsymbol{\beta}|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \mathbf{0}$$

$$\leftrightarrow \mathbf{0} - 2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0} \leftrightarrow \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \leftrightarrow \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

• **Centered simple linear regression**

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i = (\beta_0 + \beta_1 \bar{x}) + \beta_1(x_i - \bar{x}) + \varepsilon_i = \alpha_0 + \beta_1(x_i - \bar{x}) + \varepsilon_i$

$$\mathbf{y} = \alpha_0 \mathbf{1} + \beta_1(\mathbf{x} - \bar{x}\cdot\mathbf{1}) + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = [\,\mathbf{1} \quad \mathbf{x} - \bar{x}\cdot\mathbf{1}\,]\begin{bmatrix}\alpha_0\\ \beta_1\end{bmatrix} + \boldsymbol{\varepsilon} = \begin{bmatrix}1 & x_1 - \bar{x}\\ 1 & x_2 - \bar{x}\\ 1 & x_3 - \bar{x}\\ \vdots & \vdots\\ 1 & x_n - \bar{x}\end{bmatrix}\begin{bmatrix}\alpha_0\\ \beta_1\end{bmatrix} + \begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\\ \vdots\\ \varepsilon_n\end{bmatrix}$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix}1 & \cdots & 1\\ x_1 - \bar{x} & \cdots & x_n - \bar{x}\end{bmatrix}\begin{bmatrix}1 & x_1 - \bar{x}\\ \vdots & \vdots\\ 1 & x_n - \bar{x}\end{bmatrix} = \begin{bmatrix}n & \sum(x_i - \bar{x})\\ \sum(x_i - \bar{x}) & \sum(x_i - \bar{x})^2\end{bmatrix} = \begin{bmatrix}n & 0\\ 0 & s_{xx}\end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{\mathbf{1}}{\mathbf{ns_{xx}}}\begin{bmatrix}s_{xx} & 0\\ 0 & n\end{bmatrix} = \begin{bmatrix}1/n & 0\\ 0 & 1/s_{xx}\end{bmatrix},\ \mathbf{X}'\mathbf{y} = \begin{bmatrix}1 & \cdots & 1\\ x_1 - \bar{x} & \cdots & x_n - \bar{x}\end{bmatrix}\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix} = \begin{bmatrix}\sum y_i\\ \sum(x_1 - \bar{x})y_i\end{bmatrix} = \begin{bmatrix}n\bar{y}\\ s_{xy}\end{bmatrix}$$

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix}1/n & 0\\ 0 & 1/s_{xx}\end{bmatrix}\begin{bmatrix}n\bar{y}\\ s_{xy}\end{bmatrix} = \begin{bmatrix}\bar{y}\\ s_{xy}/s_{xx}\end{bmatrix} = \begin{bmatrix}\hat{\alpha}_0\\ \hat{\beta}_1\end{bmatrix} \to\ \hat{\alpha}_0 = \hat{\beta}_0 + \hat{\beta}_1\bar{x} \leftrightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

• **Only intercept simple linear regression**

$$y_i = \beta_0 + \varepsilon_i \ \to\ \mathbf{y} = \beta_0\mathbf{1} + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = [\,\mathbf{1}\,] + \boldsymbol{\varepsilon} = \begin{bmatrix}1\\ 1\\ 1\\ \vdots\\ 1\end{bmatrix}[\,\beta_0\,] + \begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\\ \vdots\\ \varepsilon_n\end{bmatrix} \to \hat{\beta}_0 = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{y} = \frac{1}{n}\sum y_i$$

• **Properties of Least Squares Estimators**

$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \sim \mathbf{N}(\boldsymbol{\beta}, (\mathbf{X}'\mathbf{X})^{-1}\sigma^2)$

Let $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{A}$ then, $\mathbf{Ay} \sim \mathrm{N}(\mathbf{AX\boldsymbol{\beta}},\ \mathbf{AI}_n\mathbf{A}'\sigma^2)$ $\ast\ \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_n\sigma^2) \ \to \mathbf{y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n\sigma^2)$

① $\mathbf{AX}\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}\ \therefore \mathrm{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$

② $\mathbf{AI}_n\mathbf{A}'\boldsymbol{\sigma}^2 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\cdot\mathbf{I}_n\cdot\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\sigma^2 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\sigma^2 = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2\ \therefore \mathrm{Var}(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$

• **Sequential Sum of Squares**

①

i)	ii)
$y_i = \beta_0 + \varepsilon_i \quad \to$ Base model − I $y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i \quad \to \mathrm{SSR}(\beta_1 \mid \beta_0)$ $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \ \to \mathrm{SSR}(\beta_1, \beta_2 \mid \beta_0)$	$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i \quad \to$ Base model − II $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \ \to \mathrm{SSR}(\beta_2 \mid \beta_0, \beta_1)$

i), ii)에서 $\mathrm{SSR}(\beta_1, \beta_2 \mid \beta_0) = \mathrm{SSR}(\beta_1 \mid \beta_0) + \mathrm{SSR}(\beta_2 \mid \beta_0, \beta_1)$

② $\mathrm{SSR}(\beta_1, \beta_2, \beta_3 \mid \beta_0) = \mathrm{SSR}(\beta_1 \mid \beta_0) + \mathrm{SSR}(\beta_2 \mid \beta_0, \beta_1) + \mathrm{SSR}(\beta_3 \mid \beta_0, \beta_1, \beta_2)$

df 3 1 1 1

- **Partial Sum of Squares**

$SSR(\beta_1|\beta_0,\beta_2,\beta_3) \neq SSR(\beta_2|\beta_0,\beta_1,\beta_3) \neq SSR(\beta_3|\beta_0,\beta_1,\beta_2)$
$SSR(\beta_1|\beta_0,\beta_2,\beta_3) + SSR(\beta_2|\beta_0,\beta_1,\beta_3) + SSR(\beta_3|\beta_0,\beta_1,\beta_2) \neq SSR(\beta_1,\beta_2,\beta_3|\beta_0)$

$$\begin{aligned} &SSR(\beta_4,\beta_5|\beta_0,\beta_1,\beta_2,\beta_3) = SSR(\beta_1,\beta_2,\beta_3,\beta_4,\beta_5|\beta_0) - SSR(\beta_1,\beta_2,\beta_3|\beta_0) \\ &= \big(SSR(\beta_1|\beta_0) + SSR(\beta_2|\beta_0,\beta_1) + SSR(\beta_3|\beta_0,\beta_1,\beta_2) + SSR(\beta_4|\beta_0,\beta_1,\beta_2,\beta_3) + SSR(\beta_5|\beta_0,\beta_1,\beta_2,\beta_3,\beta_4)\big) \\ &- \big(SSR(\beta_1|\beta_0) + SSR(\beta_2|\beta_0,\beta_1) + SSR(\beta_3|\beta_0,\beta_1,\beta_2)\big) = SSR(\beta_4|\beta_0,\beta_1,\beta_2,\beta_3) + SSR(\beta_5|\beta_0,\beta_1,\beta_2,\beta_3,\beta_4) \end{aligned}$$

- **SSE in multiple linear regression**

$$\begin{aligned} SSE &= \mathbf{e'e} = \sum e_i{}^2 = \sum (y_i - \hat{y}_i)^2 = (\mathbf{y}-\hat{\mathbf{y}})'(\mathbf{y}-\hat{\mathbf{y}}) = \big(\mathbf{y}-\mathbf{X}\hat{\boldsymbol\beta}\big)'\big(\mathbf{y}-\mathbf{X}\hat{\boldsymbol\beta}\big) = \mathbf{y'y} - 2\mathbf{y'X}\hat{\boldsymbol\beta} + \hat{\boldsymbol\beta}'\mathbf{X'X}\hat{\boldsymbol\beta} \\ &= \mathbf{y'y} - 2\mathbf{y'X}\hat{\boldsymbol\beta} + ((\mathbf{X'X})^{-1}\mathbf{X'y})' \cdot \mathbf{X'X}\hat{\boldsymbol\beta} = \mathbf{y'y} - 2\mathbf{y'X}\hat{\boldsymbol\beta} + \mathbf{y'X}(\mathbf{X'X})^{-1}(\mathbf{X'X})\hat{\boldsymbol\beta} = \mathbf{y'y} - 2\mathbf{y'X}\hat{\boldsymbol\beta} + \mathbf{y'X}\hat{\boldsymbol\beta} \\ &= \mathbf{y'y} - \mathbf{y'X}\hat{\boldsymbol\beta} \end{aligned}$$

- **Estimation of σ^2 in a multiple linear regression**

$$\begin{aligned} &SSE = \mathbf{e'e} = \sum e_i{}^2 = \sum (y_i - \hat{y}_i)^2 = (\mathbf{y}-\hat{\mathbf{y}})'(\mathbf{y}-\hat{\mathbf{y}}) = \big((\mathbf{I}-\mathbf{H})\mathbf{y}\big)'\big((\mathbf{I}-\mathbf{H})\mathbf{y}\big) = \mathbf{y}'(\mathbf{I}-\mathbf{H})'(\mathbf{I}-\mathbf{H})\mathbf{y} \\ &\quad \leftrightarrow\ \mathbf{y}'(\mathbf{I}-\mathbf{H})\mathbf{y} = (\mathbf{X}\boldsymbol\beta+\boldsymbol\varepsilon)'(\mathbf{I}-\mathbf{H})(\mathbf{X}\boldsymbol\beta+\boldsymbol\varepsilon) = \boldsymbol\varepsilon'(\mathbf{I}-\mathbf{H})\boldsymbol\varepsilon \end{aligned}$$

$$\frac{\mathbf{e'e}}{\sigma^2} = \frac{\boldsymbol\varepsilon'(\mathbf{I}-\mathbf{H})\boldsymbol\varepsilon}{\sigma^2} \sim \chi^2\big(\mathrm{Rank}(\mathbf{I}-\mathbf{H})\big) = \chi^2(n-k-1)$$

$$\begin{aligned} * \ \mathrm{Rank}(\mathbf{I}-\mathbf{H}) &= \mathrm{trace}(\mathbf{I}-\mathbf{H}) = \mathrm{trace}\big(\mathbf{I}_{(n)x(n)}\big) - \mathrm{trace}(\mathbf{H}) = n - \mathrm{trace}(\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X}') \\ &= n - \mathrm{trace}((\mathbf{X'X})^{-1}\mathbf{X'X}) = n - \mathrm{trace}\big(\mathbf{I}_{(k+1)x(k+1)}\big) = n-(k+1) \end{aligned}$$

$$\mathbf{E}\left[\frac{\mathbf{e'e}}{\sigma^2}\right] = n-k-1 \leftrightarrow \mathbf{E}\left[\frac{\mathbf{e'e}}{n-k-1}\right] = \sigma^2 \ \therefore \hat{\sigma}^2(=s^2) = \frac{\sum(y_i-\hat{y}_i)^2}{n-k-1} = \text{MSE is unbiased estimator of } \sigma^2.$$

Therefore, we use $\left(MSE = \dfrac{e'e}{n-k-1}\right)$ as a representative estimator of parameter σ^2

- **Tests for regression coefficients**

① Test of only one regression coeffiecient

$H_0: \beta_j = 0$ vs $H_1: \beta_j \neq 0$

$$\hat{\boldsymbol\beta} = (\mathbf{X'X})^{-1}\mathbf{X'y} \sim \mathbf{N}(\boldsymbol\beta\,, (\mathbf{X'X})^{-1}\sigma^2) \rightarrow \hat{\beta}_j \sim N(\beta_j\,, c_{jj}\sigma^2)$$

$$C_{(k+1)x(k+1)} = \begin{bmatrix} c_{00} & c_{01} & c_{02} & \cdots & c_{0k} \\ c_{10} & c_{11} & c_{12} & \cdots & c_{1k} \\ c_{20} & c_{21} & c_{22} & \cdots & c_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{k0} & c_{k1} & c_{k2} & \cdots & c_{kk} \end{bmatrix} \leftarrow \begin{array}{c} c_{ij} = \mathrm{Cov}\big(\beta_i,\beta_j\big) \text{ and } i,j = 0,1,\cdots,k \\ (\text{For example }, \mathrm{Cov}(\beta_0,\beta_1) = c_{01}\sigma^2) \end{array}$$

$$\text{test statistics } = \frac{\hat{\beta}_j - \beta_j}{s \cdot \sqrt{c_{jj}}} \sim t(n-k-1)$$

② Test of overall regression coefficients

$H_0: \mathbf{y} = \beta_0\mathbf{1} + \boldsymbol{\varepsilon}_\mathbf{A}$ ($\boldsymbol{\beta} = \mathbf{0}$ except for β_0) vs $H_1: \mathbf{y} = \beta_0\mathbf{1} + \beta_1\mathbf{x_1} + \cdots + \beta_k\mathbf{x_k} + \boldsymbol{\varepsilon}$

In this case, we can use SSR and simply compute followings.

$$\begin{aligned} \mathrm{SSR} &= \sum(\hat{y}_i - \bar{y})^2 = (\hat{\mathbf{y}} - \bar{y}\cdot\mathbf{1})'(\hat{\mathbf{y}} - \bar{y}\cdot\mathbf{1}) \ \rightarrow \bar{y} = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{y} \leftrightarrow \bar{y}\cdot\mathbf{1} = [\,\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\,]\mathbf{y} = \mathbf{H_A y} \\ &= (\mathbf{Hy} - \mathbf{H_A y})'(\mathbf{Hy} - \mathbf{H_A y}) = \mathbf{y}'(\mathbf{H} - \mathbf{H_A})'(\mathbf{H} - \mathbf{H_A})\mathbf{y} = \mathbf{y}'(\mathbf{H} - \mathbf{H_A})\mathbf{y} = (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})'(\mathbf{H} - \mathbf{H_A})(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{\varepsilon}'(\mathbf{H} - \mathbf{H_A})\boldsymbol{\varepsilon} + (\mathbf{X}\boldsymbol{\beta})'(\mathbf{H} - \mathbf{H_A})(\mathbf{X}\boldsymbol{\beta}) \leftarrow \textbf{Let } (\mathbf{X}\boldsymbol{\beta})'(\mathbf{H} - \mathbf{H_A})(\mathbf{X}\boldsymbol{\beta}) = \boldsymbol{\delta} \end{aligned}$$

$$f = \frac{\frac{\mathrm{SSR}}{\sigma^2}\cdot\frac{1}{k}}{\left(\frac{\mathrm{SSE}}{\sigma^2}\right)\cdot\frac{1}{(n-k-1)}} = \frac{\left(\frac{\boldsymbol{\varepsilon}'(\mathbf{H}-\mathbf{H_A})\boldsymbol{\varepsilon}+\boldsymbol{\delta}}{\sigma^2}\right)\cdot\frac{1}{k}}{\left(\frac{\boldsymbol{\varepsilon}'(\mathbf{I}-\mathbf{H})\boldsymbol{\varepsilon}}{\sigma^2}\right)\cdot\frac{1}{(n-k-1)}} \overset{H_0}{\Rightarrow} \frac{\left(\frac{\boldsymbol{\varepsilon}'(\mathbf{H}-\mathbf{H_A})\boldsymbol{\varepsilon}}{\sigma^2}\right)\cdot\frac{1}{k}}{\left(\frac{\boldsymbol{\varepsilon}'(\mathbf{I}-\mathbf{H})\boldsymbol{\varepsilon}}{\sigma^2}\right)\cdot\frac{1}{(n-k-1)}} = \frac{\left(\frac{\chi^2(k)}{(k)}\right)}{\left(\frac{\chi^2(n-k-1)}{(n-k-1)}\right)} \sim F(k, n-k-1)$$

$$* \ \frac{\mathrm{SSR}}{\sigma^2} = \frac{\boldsymbol{\varepsilon}'(\mathbf{H}-\mathbf{H_A})\boldsymbol{\varepsilon}+\boldsymbol{\delta}}{\sigma^2} \sim \chi^2\left(\mathrm{Rank}(\mathbf{H}-\mathbf{H_A}), \frac{\boldsymbol{\delta}}{\sigma^2}\right) \overset{H_0}{\rightarrow} \frac{\boldsymbol{\varepsilon}'(\mathbf{H}-\mathbf{H_A})\boldsymbol{\varepsilon}}{\sigma^2} \sim \chi^2(\mathrm{Rank}(\mathbf{H}-\mathbf{H_A}))$$

$*$ Under H_0, $\boldsymbol{\delta} = (\mathbf{X}\boldsymbol{\beta})'(\mathbf{H} - \mathbf{H_A})(\mathbf{X}\boldsymbol{\beta}) = (\beta_0\mathbf{1})'(\mathbf{H} - \mathbf{H_A})(\beta_0\mathbf{1}) = \mathbf{0}$.
$*$ $(\mathbf{H} - \mathbf{H_A})(\mathbf{H} - \mathbf{H_A}) = \mathbf{H} - \mathbf{H_A}$: idempotent matrix
$*$ $\mathrm{Rank}(\mathbf{H} - \mathbf{H_A}) = \mathrm{trace}(\mathbf{H} - \mathbf{H_A}) = \mathrm{trace}(\mathbf{H}) - \mathrm{trace}(\mathbf{H_A}) = (k+1) - \mathrm{trace}(\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}') = k + 1 - 1 = k$
$*$ $(\mathbf{H} - \mathbf{H_A}) \perp (\mathbf{I} - \mathbf{H}) \leftrightarrow (\mathbf{H} - \mathbf{H_A})(\mathbf{I} - \mathbf{H}) = \mathbf{0}$, two chisquares are independent. Thus F dist is satisfied.

③ Test of some regression coefficients (<u>**more general than ②**</u>) **– Extra sum of squares principle**

$H_0: \mathbf{y} = \mathbf{X_A}\boldsymbol{\beta}_\mathbf{A} + \boldsymbol{\varepsilon}_\mathbf{A}$ ($\boldsymbol{\beta}_\mathbf{B} = \mathbf{0}$) vs $H_1: \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = [\,\mathbf{X_A} \quad \mathbf{X_B}\,]\begin{bmatrix}\boldsymbol{\beta}_\mathbf{A} \\ \boldsymbol{\beta}_\mathbf{B}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\varepsilon}_\mathbf{A} \\ \boldsymbol{\varepsilon}_\mathbf{B}\end{bmatrix} \left(* \ \mathbf{X} = \begin{bmatrix}\mathbf{X_A} \\ \mathbf{X_B}\end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix}\boldsymbol{\beta}_\mathbf{A} \\ \boldsymbol{\beta}_\mathbf{B}\end{bmatrix}\right)$

In this case, we should use '**SSE(RM)-SSR(FM)**'. If we use SSR, computation would be complex.

$$\begin{aligned} \mathrm{SSE(RM)} &= \mathbf{y}'(\mathbf{I} - \mathbf{H_A})\mathbf{y} = (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})'(\mathbf{I} - \mathbf{H_A})(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = (\mathbf{X_A}\boldsymbol{\beta}_\mathbf{A} + \mathbf{X_B}\boldsymbol{\beta}_\mathbf{B} + \boldsymbol{\varepsilon})'(\mathbf{I} - \mathbf{H_A})(\mathbf{X_A}\boldsymbol{\beta}_\mathbf{A} + \mathbf{X_B}\boldsymbol{\beta}_\mathbf{B} + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{\varepsilon}'(\mathbf{I} - \mathbf{H_A})\boldsymbol{\varepsilon} + (\mathbf{X_B}\boldsymbol{\beta}_\mathbf{B})'(\mathbf{I} - \mathbf{H_A})(\mathbf{X_B}\boldsymbol{\beta}_\mathbf{B}) \ \leftarrow \textbf{Let } (\mathbf{X_B}\boldsymbol{\beta}_\mathbf{B})'(\mathbf{I} - \mathbf{H_A})(\mathbf{X_B}\boldsymbol{\beta}_\mathbf{B}) = \boldsymbol{\delta} \end{aligned}$$

$$\begin{aligned} \mathrm{SSE(FM)} &= \mathbf{e}'\mathbf{e} = \sum {e_i}^2 = \sum (y_i - \hat{y}_i)^2 = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) = \left((\mathbf{I} - \mathbf{H})\mathbf{y}\right)'\left((\mathbf{I} - \mathbf{H})\mathbf{y}\right) = \mathbf{y}'(\mathbf{I} - \mathbf{H})'(\mathbf{I} - \mathbf{H})\mathbf{y} \\ &\leftrightarrow \ \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} = (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})'(\mathbf{I} - \mathbf{H})(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\varepsilon}'(\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon} \end{aligned}$$

$$\mathrm{SSE(RM)} - \mathrm{SSE(FM)} = \boldsymbol{\varepsilon}'(\mathbf{I} - \mathbf{H_A})\boldsymbol{\varepsilon} + \boldsymbol{\delta} - \boldsymbol{\varepsilon}'(\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}'(\mathbf{H} - \mathbf{H_A})\boldsymbol{\varepsilon} + \boldsymbol{\delta}$$

$$f = \frac{\left(\frac{\mathrm{SSE(RM)} - \mathrm{SSE(FM)}}{\sigma^2}\right)\cdot\frac{1}{r}}{\left(\frac{\mathrm{SSE}}{\sigma^2}\right)\cdot\frac{1}{(n-k-1)}} = \frac{\left(\frac{\boldsymbol{\varepsilon}'(\mathbf{H}-\mathbf{H_A})\boldsymbol{\varepsilon}+\boldsymbol{\delta}}{\sigma^2}\right)\cdot\frac{1}{r}}{\left(\frac{\boldsymbol{\varepsilon}'(\mathbf{I}-\mathbf{H})\boldsymbol{\varepsilon}}{\sigma^2}\right)\cdot\frac{1}{(n-k-1)}} \overset{H_0}{\Rightarrow} \frac{\left(\frac{\boldsymbol{\varepsilon}'(\mathbf{H}-\mathbf{H_A})\boldsymbol{\varepsilon}}{\sigma^2}\right)\cdot\frac{1}{r}}{\left(\frac{\boldsymbol{\varepsilon}'(\mathbf{I}-\mathbf{H})\boldsymbol{\varepsilon}}{\sigma^2}\right)\cdot\frac{1}{(n-k-1)}} = \frac{\left(\frac{\chi^2(r)}{(r)}\right)}{\left(\frac{\chi^2(n-k-1)}{(n-k-1)}\right)} \sim F(r, n-k-1)$$

$$* \ \frac{\mathrm{SSE(RM)} - \mathrm{SSE(FM)}}{\sigma^2} = \frac{\boldsymbol{\varepsilon}'(\mathbf{H}-\mathbf{H_A})\boldsymbol{\varepsilon}+\boldsymbol{\delta}}{\sigma^2} \sim \chi^2\left(\mathrm{Rank}(\mathbf{H}-\mathbf{H_A}), \frac{\boldsymbol{\delta}}{\sigma^2}\right) \overset{H_0}{\rightarrow} \frac{\boldsymbol{\varepsilon}'(\mathbf{H}-\mathbf{H_A})\boldsymbol{\varepsilon}}{\sigma^2} \sim \chi^2(\mathrm{Rank}(\mathbf{H}-\mathbf{H_A}))$$

$*$ Under H_0, $\boldsymbol{\delta} = (\mathbf{X_B}\boldsymbol{\beta}_\mathbf{B})'(\mathbf{I} - \mathbf{H_A})(\mathbf{X_B}\boldsymbol{\beta}_\mathbf{B}) = \mathbf{0}$.
$*$ $(\mathbf{H} - \mathbf{H_A})(\mathbf{H} - \mathbf{H_A}) = \mathbf{H} - \mathbf{H_A}$: idempotent matrix
$*$ $\mathrm{Rank}(\mathbf{H} - \mathbf{H_A}) = \mathrm{trace}(\mathbf{H} - \mathbf{H_A}) = \mathrm{trace}(\mathbf{H}) - \mathrm{trace}(\mathbf{H_A}) = (k+1) - (k_A + 1) = k - k_A = r$
$*$ $(\mathbf{H} - \mathbf{H_A}) \perp (\mathbf{I} - \mathbf{H}) \leftrightarrow (\mathbf{H} - \mathbf{H_A})(\mathbf{I} - \mathbf{H}) = \mathbf{0}$, two chisquares are independent. Thus F dist is satisfied.

• **Confidence intervals and prediction intervals in multiple regression**

First, remind this matrix structure.

$$\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} = \beta_0\mathbf{1} + \beta_1\mathbf{x_1} + \cdots + \beta_k\mathbf{x_k} + \boldsymbol{\varepsilon} \leftrightarrow \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ 1 & x_{31} & x_{32} & \cdots & x_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

If we want to know the mean response of a set of observations (low vector) like below,

$$\begin{bmatrix} y_1 \\ \vdots \\ E|_{\mathbf{x=x_0}} \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ 1 & x_{01} & x_{02} & \cdots & x_{0k} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon|_{\mathbf{x=x_0}} \\ \vdots \\ \varepsilon_n \end{bmatrix} \leftarrow \text{Let an observation vector } \mathbf{x_0} = \begin{bmatrix} 1 \\ x_{01} \\ x_{02} \\ \vdots \\ x_{0k} \end{bmatrix}$$

The mean response will be expressed as follows.

$$E(\mathbf{y}|_{\substack{\mathbf{x_1=x_{01}} \\ \mathbf{x_2=x_{02}} \\ \vdots \\ \mathbf{x_k=x_{0k}}}}) = E(\beta_0 1 + \beta_1 x_{01} + \beta_2 x_{02} \cdots + \beta_k x_{0k} + \varepsilon) \leftrightarrow E\left(\mathbf{y}|_{\mathbf{x=x_0}}\right) = \mathbf{x_0}'\boldsymbol{\beta} \leftarrow E(\varepsilon) = 0$$

① Disitribution : $\widehat{E(\mathbf{y}|\mathbf{x} = \mathbf{x_0})} = \mathbf{x_0}'\widehat{\boldsymbol{\beta}} \sim N(\mathbf{x_0}'\boldsymbol{\beta}, (\mathbf{x_0}'(\mathbf{X'X})^{-1}\mathbf{x_0})\sigma^2) \rightarrow \widehat{Var}\left(E(\mathbf{y}|\mathbf{x} = \mathbf{x_0})\right) = (\mathbf{x_0}'(\mathbf{X'X})^{-1}\mathbf{x_0})s^2$

② $(1-\alpha)100\%$ Confidence Intervals : $\mathbf{x_0}'\widehat{\boldsymbol{\beta}} \pm t_{\frac{\alpha}{2}}(n-k-1) \cdot s \cdot \sqrt{\mathbf{x_0}'(\mathbf{X'X})^{-1}\mathbf{x_0}}$

Similarly, the prediction of a set of observations would be interesting.

$$\begin{bmatrix} y_1 \\ \vdots \\ y|_{\mathbf{x=x_0}} \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ 1 & x_{01} & x_{02} & \cdots & x_{0k} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon|_{\mathbf{x=x_0}} \\ \vdots \\ \varepsilon_n \end{bmatrix} \leftarrow \text{Let an observation vector } \mathbf{x_0} = \begin{bmatrix} 1 \\ x_{01} \\ x_{02} \\ \vdots \\ x_{0k} \end{bmatrix}$$

The prediction will be expressed as follows.

$$\mathbf{y}|_{\substack{\mathbf{x_1=x_{01}} \\ \mathbf{x_2=x_{02}} \\ \vdots \\ \mathbf{x_k=x_{0k}}}} = \beta_0 1 + \beta_1 x_{01} + \beta_2 x_{02} \cdots + \beta_k x_{0k} + \varepsilon \leftrightarrow \mathbf{y}|_{\mathbf{x=x_0}} = \mathbf{x_0}'\boldsymbol{\beta} + \varepsilon$$

① Disitribution : $\widehat{\mathbf{y}|\mathbf{x} = \mathbf{x_0}} = \mathbf{x_0}'\widehat{\boldsymbol{\beta}} + \hat{\varepsilon} \sim N\left(\mathbf{x_0}'\boldsymbol{\beta}, \left(1 + \mathbf{x_0}'(\mathbf{X'X})^{-1}\mathbf{x_0}\right)\sigma^2\right) \rightarrow \widehat{Var}\left(\widehat{\mathbf{y}|\mathbf{x} = \mathbf{x_0}}\right) = \left(1 + \mathbf{x_0}'(\mathbf{X'X})^{-1}\mathbf{x_0}\right)s^2$

② $(1-\alpha)100\%$ Prediction Intervals : $\mathbf{x_0}'\widehat{\boldsymbol{\beta}} \pm t_{\frac{\alpha}{2}}(n-k-1) \cdot s \cdot \sqrt{1 + \mathbf{x_0}'(\mathbf{X'X})^{-1}\mathbf{x_0}}$

Pay attention to the difference between case of mean response and that of prediction.

Only '1' is added to the variance in the prediction.

- **Test of general linear hypothesis**

$H_0: \mathbf{C\beta} = \mathbf{d}$ vs $H_1: \mathbf{C\beta} \neq \mathbf{d}$

$\mathbf{C\hat{\beta}} \sim \mathbf{N}(\mathbf{C\beta}, \mathbf{C}(\mathbf{X'X})^{-1}\mathbf{C'}\sigma^2) \leftarrow \hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y} \sim \mathbf{N}(\boldsymbol{\beta}, (\mathbf{X'X})^{-1}\sigma^2)$

$\mathbf{C\hat{\beta}} \sim \mathbf{N}(\mathbf{C\beta}, \mathbf{C}(\mathbf{X'X})^{-1}\mathbf{C'}\sigma^2) \xrightarrow{\text{standardized}} \dfrac{(\mathbf{C\hat{\beta}} - \mathbf{C\beta})'(\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{C'})^{-1}(\mathbf{C\hat{\beta}} - \mathbf{C\beta})}{\sigma^2} \sim \chi^2(\mathrm{r})$

$* \dfrac{(\mathbf{C\hat{\beta}} - \mathbf{C\beta})'(\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{C'})^{-1}(\mathbf{C\hat{\beta}} - \mathbf{C\beta})}{\sigma^2} = \dfrac{(\mathbf{A'\varepsilon})'(\mathbf{A'A})^{-1}(\mathbf{A'\varepsilon})}{\sigma^2} = \dfrac{\boldsymbol{\varepsilon}'(\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'})\boldsymbol{\varepsilon}}{\sigma^2} \sim \chi^2(\mathrm{Rank}(\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'}) = \mathrm{r})$

$* \mathbf{C\hat{\beta}} = \mathbf{C}(\mathbf{X'X})^{-1}\mathbf{X'y} = \mathbf{C}(\mathbf{X'X})^{-1}\mathbf{X'}(\mathbf{X\beta} + \boldsymbol{\varepsilon}) = \mathbf{C}(\mathbf{X'X})^{-1}(\mathbf{X'X})\boldsymbol{\beta} + [\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{X'}]\boldsymbol{\varepsilon} \leftarrow \textbf{Let } \mathbf{A'} = \mathbf{C}(\mathbf{X'X})^{-1}\mathbf{X'}$

$= \mathbf{C\beta} + \mathbf{A'\varepsilon} \leftrightarrow \mathbf{C\hat{\beta}} - \mathbf{C\beta} = \mathbf{A'\varepsilon}$

$* \mathbf{A'A} = [\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{X'}][\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{C'}] = \mathbf{C}(\mathbf{X'X})^{-1}(\mathbf{X'X})(\mathbf{X'X})^{-1}\mathbf{C'} = \mathbf{C}(\mathbf{X'X})^{-1}\mathbf{C'}$

$* (\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'})(\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'}) = \mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'}$: idempotent mattrix

$* \mathrm{Rank}(\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'}) = \mathrm{trace}((\mathbf{A'A})^{-1}\mathbf{A'A}) = \mathrm{trace}(\mathbf{I}_{(r)x(r)}) = \mathrm{r} \leftarrow \because \mathbf{C} = (\mathrm{r})\mathrm{x}(\mathrm{p})$ matrix

$$f = \frac{\mathbf{f_0}(H_0)}{\mathbf{f_1}(H_1)} \xrightarrow{H_0:\mathbf{C\beta}=\mathbf{d}} \frac{\left(\dfrac{(\mathbf{C\hat{\beta}} - \mathbf{d})'(\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{C'})^{-1}(\mathbf{C\hat{\beta}} - \mathbf{d})}{\sigma^2}\right)\cdot\dfrac{1}{r}}{\left(\dfrac{SSE}{\sigma^2}\right)\cdot\dfrac{1}{(n-k-1)}} = \frac{(\mathbf{C\hat{\beta}} - \mathbf{d})'(\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{C'})^{-1}(\mathbf{C\hat{\beta}} - \mathbf{d})}{r\cdot s^2}$$

$$= \frac{\left(\dfrac{\boldsymbol{\varepsilon}'(\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'})\boldsymbol{\varepsilon}}{\sigma^2}\right)\cdot\dfrac{1}{r}}{\left(\dfrac{\boldsymbol{\varepsilon}'(\mathbf{I}-\mathbf{H})\boldsymbol{\varepsilon}}{\sigma^2}\right)\cdot\dfrac{1}{(n-k-1)}} = \frac{\left(\dfrac{\chi^2(r)}{(r)}\right)}{\left(\dfrac{\chi^2(n-k-1)}{(n-k-1)}\right)} \sim F(r, n-k-1)$$

$* (\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'}) \perp (\mathbf{I}-\mathbf{H}) \leftrightarrow (\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'})(\mathbf{I}-\mathbf{H}) = \mathbf{0}$. Orthogonality is satisfied. Thus F dist is satisfied.

$\because (\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'})(\mathbf{I}-\mathbf{H}) = \mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'} - \mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'H} = \mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'} - \mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'} = \mathbf{0}$

- **Simultaneous test**

$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta_0}$ vs $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta_0}$

$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y} \sim \mathbf{N}(\boldsymbol{\beta}, (\mathbf{X'X})^{-1}\sigma^2) \xrightarrow{\text{standardized}} \dfrac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{X'X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sigma^2} \sim \chi^2(\mathrm{p})$

$* \dfrac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{X'X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sigma^2} = \dfrac{(\mathbf{A'\varepsilon})'(\mathbf{A'A})^{-1}(\mathbf{A'\varepsilon})}{\sigma^2} = \dfrac{\boldsymbol{\varepsilon}'(\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'})\boldsymbol{\varepsilon}}{\sigma^2} \sim \chi^2(\mathrm{Rank}(\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'}) = \mathrm{p})$

$* \hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y} = (\mathbf{X'X})^{-1}\mathbf{X'}(\mathbf{X\beta} + \boldsymbol{\varepsilon}) = (\mathbf{X'X})^{-1}(\mathbf{X'X})\boldsymbol{\beta} + [(\mathbf{X'X})^{-1}\mathbf{X'}]\boldsymbol{\varepsilon} \leftarrow \textbf{Let } \mathbf{A'} = (\mathbf{X'X})^{-1}\mathbf{X'}$

$= \boldsymbol{\beta} + \mathbf{A'\varepsilon} \leftrightarrow \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \mathbf{A'\varepsilon}$

$* \mathbf{A'A} = [(\mathbf{X'X})^{-1}\mathbf{X'}][\mathbf{X}(\mathbf{X'X})^{-1}] = (\mathbf{X'X})^{-1}(\mathbf{X'X})(\mathbf{X'X})^{-1} = (\mathbf{X'X})^{-1} \leftrightarrow (\mathbf{X'X}) = (\mathbf{A'A})^{-1}$

$* (\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'})(\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'}) = \mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'}$: idempotent mattrix

$* \mathrm{Rank}(\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'}) = \mathrm{trace}((\mathbf{A'A})^{-1}\mathbf{A'A}) = \mathrm{trace}(\mathbf{I}_{(p)x(p)}) = \mathrm{p} \leftarrow \because \mathbf{X} = (\mathrm{n})\mathrm{x}(\mathrm{p})$ matrix

$$f = \frac{\mathbf{f_0}(H_0)}{\mathbf{f_1}(H_1)} \xrightarrow{H_0:\boldsymbol{\beta}=\boldsymbol{\beta_0}} = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{X'X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{p\cdot s^2} = \frac{\left(\dfrac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{X'X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sigma^2}\right)\cdot\dfrac{1}{p}}{\left(\dfrac{SSE}{\sigma^2}\right)\cdot\dfrac{1}{(n-p)}}$$

$$= \frac{\left(\dfrac{\boldsymbol{\varepsilon}'(\mathbf{A}(\mathbf{A'A})^{-1}\mathbf{A'})\boldsymbol{\varepsilon}}{\sigma^2}\right)\cdot\dfrac{1}{p}}{\left(\dfrac{\boldsymbol{\varepsilon}'(\mathbf{I}-\mathbf{H})\boldsymbol{\varepsilon}}{\sigma^2}\right)\cdot\dfrac{1}{(n-p)}} = \frac{\left(\dfrac{\chi^2(p)}{(p)}\right)}{\left(\dfrac{\chi^2(n-p)}{(n-p)}\right)} \sim F(p, n-p) \quad * p = k+1$$

$*$ $(1-\alpha)100\%$ joint confidence region on $\boldsymbol{\beta}$: $\left\{\boldsymbol{\beta}: \dfrac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\mathbf{X'X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{p\cdot s^2} \leq \mathbf{F_\alpha}(p, n-p)\right\}$

- **Test under some constraints**

Let Full model : $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \varepsilon_i$

Then, consider below several constraints.

① $H_0 : \beta_1 = \beta_3$

② $H_0 : \beta_1 = \beta_3 | \beta_2 = \beta_4 = \beta_5 = \beta_6 = 0$

③ $H_0 : \beta_1 = \beta_3 , \beta_2 = \beta_4 = \beta_5 = \beta_6 = 0$

④ $H_0 : \beta_1 + \beta_3 = 1 \mid \beta_2 = \beta_4 = \beta_5 = \beta_6 = 0$

⑤ $H_0 : \beta_1 = \beta_3 ,\ \beta_1 + \beta_3 = 1 \mid \beta_2 = \beta_4 = \beta_5 = \beta_6 = 0$

Solutions)

① $H_0 : \beta_1 = \beta_3$

RM : $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + (\beta_1) x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \varepsilon_i$
$\leftrightarrow y_i = \beta_0 + \beta_1 (x_{i1} + x_{i3}) + \beta_2 x_{i2} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \varepsilon_i \leftarrow$ Let $v_i = (x_{i1} + x_{i3})$
$\leftrightarrow y_i = \beta_0 + \beta_1 (v_i) + \beta_2 x_{i2} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \varepsilon_i$

FM : $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \varepsilon_i$

Test statistics : $f = \dfrac{(SSE(RM) - SSE(FM))/(r)}{SSE/(n-k-1)} = \dfrac{(SSE(RM) - SSE(FM))/(7-6)}{SSE/(n-6-1)} \sim F(1, n-7)$
In this test , $SST(RM) = SST(FM)$ and 'r' means the difference of number of β_j.

② $H_0 : \beta_1 = \beta_3 | \beta_2 = \beta_4 = \beta_5 = \beta_6 = 0$

RM : $y_i = \beta_0 + \beta_1 x_{i1} + (\beta_1) x_{i3} + \varepsilon_i$
$\leftrightarrow y_i = \beta_0 + \beta_1 (x_{i1} + x_{i3}) + \varepsilon_i \leftarrow$ Let $v_i = (x_{i1} + x_{i3})$
$\leftrightarrow y_i = \beta_0 + \beta_1 (v_i) + \varepsilon_i$

FM : $y_i = \beta_0 + \beta_1 x_{i1} + \beta_3 x_{i3} + \varepsilon_i$

Test statistics : $f = \dfrac{\big(SSE(RM) - SSE(FM)\big)/(r)}{SSE/(n-k-1)} = \dfrac{\big(SSE(RM) - SSE(FM)\big)/(3-2)}{SSE/(n-2-1)} \sim F(1, n-3)$
In this test , $SST(RM) = SST(FM)$ and 'r' means the difference of number of β_j.

③ $H_0 : \beta_1 = \beta_3 , \beta_2 = \beta_4 = \beta_5 = \beta_6 = 0$

RM : $y_i = \beta_0 + \beta_1 x_{i1} + (\beta_1) x_{i3} + \varepsilon_i$
$\leftrightarrow y_i = \beta_0 + \beta_1 (x_{i1} + x_{i3}) + \varepsilon_i \leftarrow$ Let $v_i = (x_{i1} + x_{i3})$
$\leftrightarrow y_i = \beta_0 + \beta_1 (v_i) + \varepsilon_i$

FM : $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \varepsilon_i$

Test statistics : $f = \dfrac{\big(SSE(RM) - SSE(FM)\big)/(r)}{SSE/(n-k-1)} = \dfrac{\big(SSE(RM) - SSE(FM)\big)/(7-2)}{SSE/(n-6-1)} \sim F(5, n-7)$
In this test , $SST(RM) = SST(FM)$ and 'r' means the difference of number of β_j.

$*$ As test ② eliminated useless variables before the test , thus test ② is more sensitive to reject H_0 than ③.

④ $H_0: \beta_1 + \beta_3 = 1 \mid \beta_2 = \beta_4 = \beta_5 = \beta_6 = 0$

$RM: y_i = \beta_0 + \beta_1 x_{i1} + (1-\beta_1)x_{i3} + \varepsilon_i$
$\leftrightarrow (y_i - x_{i3}) = \beta_0 + \beta_1(x_{i1} - x_{i3}) + \varepsilon_i \leftarrow \text{Let } y_i' = (y_i - x_{i3}),\ v_i = (x_{i1} - x_{i3})$
$\leftrightarrow y_i' = \beta_0 + \beta_1 v_i + \varepsilon_i$

$FM: y_i = \beta_0 + \beta_1 x_{i1} + \beta_3 x_{i3} + \varepsilon_i$

$$\text{Test statistics}: f = \frac{\big(SSE(RM) - SSE(FM)\big)/(r)}{SSE/(n-k-1)} = \frac{\big(SSE(RM) - SSE(FM)\big)/(3-2)}{SSE/(n-2-1)} \sim F(1, n-3)$$
In this test, $SST(RM) \neq SST(FM)$ $(\because y_i \to y_i')$ and $'r'$ means the difference of number of β_j.

⑤ $H_0: \beta_1 = \beta_3,\ \beta_1 + \beta_3 = 1 \mid \beta_2 = \beta_4 = \beta_5 = \beta_6 = 0$

$RM: y_i = \beta_0 + (0.5)x_{i1} + (0.5)x_{i3} + \varepsilon_i$
$\leftrightarrow (y_i - 0.5x_{i1} - 0.5x_{i3}) = \beta_0 + \varepsilon_i \leftarrow \text{Let } y_i'' = (y_i - 0.5x_{i1} - 0.5x_{i3})$
$\leftrightarrow y_i'' = \beta_0 + \varepsilon_i$

$FM: y_i = \beta_0 + \beta_1 x_{i1} + \beta_3 x_{i3} + \varepsilon_i$

$$\text{Test statistics}: f = \frac{\big(SSE(RM) - SSE(FM)\big)/(r)}{SSE/(n-k-1)} = \frac{\big(SSE(RM) - SSE(FM)\big)/(3-1)}{SSE/(n-2-1)} \sim F(2, n-3)$$
In this test, $SST(RM) \neq SST(FM)$ $(\because y_i \to y_i'')$ and $'r'$ means the difference of number of β_j.

• **Hat matrix**

$$\begin{bmatrix} y_1 \\ \vdots \\ E|_{\mathbf{x}=\mathbf{x_i}} \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ 1 & x_{i1} & x_{i2} & \cdots & x_{ik} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon|_{\mathbf{x}=\mathbf{x_i}} \\ \vdots \\ \varepsilon_n \end{bmatrix} \leftarrow \text{Let an observation vector } \mathbf{x_i} = \begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{ik} \end{bmatrix}$$

$\widehat{\mathbf{y}|_{\mathbf{x}=\mathbf{x_i}}} = \mathbf{x_i}'\widehat{\boldsymbol{\beta}} \sim N(\mathbf{x_i}'\boldsymbol{\beta}, \mathbf{x_i}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x_i}\sigma^2) \rightarrow$ Let $h_{ii} = \mathbf{x_i}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x_i}$ $*$ h_{ii} is a diagonal element of $\mathbf{H}$ matrix

\# **3.4. Q**uestion : $\frac{1}{n} \le h_{ii} \le 1$?

Sol)

① Proof of $(h_{ii} \le 1)$

$$\mathbf{HH} = \mathbf{H} : \textbf{idempotent matrix} \rightarrow \begin{bmatrix} h_{i1} & \cdots & h_{ii} & \cdots & h_{in} \end{bmatrix} \begin{bmatrix} h_{i1} \\ \vdots \\ h_{ii} \\ \vdots \\ h_{in} \end{bmatrix} = \sum_{j=1}^{n} {h_{ij}}^2 = {h_{ii}}^2 + \sum_{j\ne i}^{n} {h_{ij}}^2 = h_{ii}$$

$$\leftrightarrow h_{ii} + \frac{\sum_{j\ne i}^{n} {h_{ij}}^2}{h_{ii}} = 1 \;\; \therefore h_{ii} \le 1$$

② Proof of $\left(\frac{1}{n} \le h_{ii}\right)$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \beta_0\mathbf{1} + \beta_1\mathbf{x_1} + \cdots + \beta_k\mathbf{x_k} + \boldsymbol{\varepsilon} \leftrightarrow \xrightarrow{\text{Centered}} = \alpha\mathbf{1} + \beta_1(\mathbf{x_1} - \overline{x_1}\mathbf{1}) + \cdots + \beta_k(\mathbf{x_k} - \overline{x_k}\mathbf{1}) + \boldsymbol{\varepsilon}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} - \overline{x_1} & \cdots & x_{1k} - \overline{x_k} \\ 1 & x_{21} - \overline{x_1} & \cdots & x_{2k} - \overline{x_k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} - \overline{x_1} & \cdots & x_{nk} - \overline{x_k} \end{bmatrix}, \mathbf{X}'\mathbf{X} = \begin{bmatrix} n & 0 \\ 0 & \mathbf{X}^{*\prime}\mathbf{X}^{*} \end{bmatrix}, (\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1/n & 0 \\ 0 & (\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1} \end{bmatrix}$$

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \begin{bmatrix} \mathbf{1} & \mathbf{x_1} - \overline{x_1}\mathbf{1} & \cdots & \mathbf{x_k} - \overline{x_k}\mathbf{1} \end{bmatrix} \begin{bmatrix} 1/n & [\quad 0 \quad] \\ [0] & [(\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1}] \end{bmatrix} \begin{bmatrix} \mathbf{1}' \\ \mathbf{x_1} - \overline{x_1}\mathbf{1}' \\ \vdots \\ \mathbf{x_k} - \overline{x_k}\mathbf{1}' \end{bmatrix}$$

$$h_{ii} = \mathbf{x_i}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x_i} = \begin{bmatrix} 1 & x_{i1} - \overline{x_1} & \cdots & x_{ik} - \overline{x_k} \end{bmatrix} \begin{bmatrix} 1/n & [\quad 0 \quad] \\ [0] & [(\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1}] \end{bmatrix} \begin{bmatrix} 1 \\ x_{i1} - \overline{x_1} \\ \vdots \\ x_{ik} - \overline{x_k} \end{bmatrix} = \frac{1}{n} + \alpha \ge \frac{1}{n}$$

$*$ $(\mathbf{X}^{*\prime}\mathbf{X}^{*})$은 **posivitve (semi) definite** 행렬이므로, $^{\forall}\mathbf{x_i}$에 대해 $\alpha = \mathbf{x_i}'(\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1}\mathbf{x_i} \ge 0$이 성립한다.

$*$ $\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, (\mathbf{X}'\mathbf{X})^{-1}\sigma^2) \leftarrow {}^{\forall}\{(\mathbf{X}'\mathbf{X})^{-1}\sigma^2\}_{ij} \ge 0 \rightarrow \mathbf{x_i}'\widehat{\boldsymbol{\beta}} \sim N(\mathbf{x_i}'\boldsymbol{\beta}, \mathbf{x_i}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x_i}\sigma^2) \leftarrow {}^{\forall}\{\mathbf{x_i}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x_i}\sigma^2\}_{ij} \ge 0$

$\therefore$ ①, ② 로부터 $\frac{1}{n} \le h_{ii} \le 1$

When predict a data point at given location$(\mathbf{x} = \mathbf{x_i})$,

$$\frac{1}{n} \le h_{ii} \le 1 \rightarrow \frac{\sigma^2}{n} \le h_{ii}\sigma^2 \le \sigma^2 = \mathrm{Var}(\bar{y}) \le \mathrm{Var}\left(E(\widehat{\mathbf{y}|\mathbf{x} = \mathbf{x_i}})\right) \le \mathrm{Var}(\varepsilon_i)$$

$\therefore$ $\frac{1}{n} \approx h_{ii}$ means that location$(\mathbf{x} = \mathbf{x_i})$ is very near the mean of the dataset.

- **Detection methods of Multicollinearity**

① Correlation matrix

$$\mathbf{X}=\begin{bmatrix}1 & x_{11} & x_{12} & \cdots & x_{1k}\\ \vdots & \cdots & \vdots & \cdots & \vdots\\ 1 & x_{i1} & x_{i2} & \cdots & x_{ik}\\ \vdots & \cdots & \vdots & \cdots & \vdots\\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk}\end{bmatrix}\xrightarrow[\textbf{Scaling}]{\textbf{Centering}}\begin{bmatrix}1/\sqrt{n} & x_{11}^{*} & x_{12}^{*} & \cdots & x_{1k}^{*}\\ \vdots & \cdots & \vdots & \cdots & \vdots\\ 1/\sqrt{n} & x_{i1}^{*} & x_{i2}^{*} & \cdots & x_{ik}^{*}\\ \vdots & \cdots & \vdots & \cdots & \vdots\\ 1/\sqrt{n} & x_{n1}^{*} & x_{n2}^{*} & \cdots & x_{nk}^{*}\end{bmatrix}\qquad \leftarrow x_{ij}^{*}=\frac{x_{ij}-\bar{x}_j}{\sqrt{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}}$$

$$y_i=\beta_0+\beta_1x_{i1}+\cdots+\beta_kx_{ik}+\varepsilon_i\xrightarrow{\substack{\textbf{Centering}\\ \textbf{Scaling}}}y_i=\alpha_0+\beta_1x_{i1}^{*}+\cdots+\beta_kx_{ik}^{*}+\varepsilon_i$$

$$\mathbf{X}'\mathbf{X}=\begin{bmatrix}1/\sqrt{n} & \cdots & 1/\sqrt{n} & \cdots & 1/\sqrt{n}\\ x_{11}^{*} & \cdots & x_{i1}^{*} & \cdots & x_{n1}^{*}\\ x_{12}^{*} & \cdots & x_{i2}^{*} & \cdots & x_{n2}^{*}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ x_{1k}^{*} & \cdots & x_{ik}^{*} & \cdots & x_{nk}^{*}\end{bmatrix}\begin{bmatrix}1/\sqrt{n} & x_{11}^{*} & x_{12}^{*} & \cdots & x_{1k}^{*}\\ \vdots & \cdots & \vdots & \cdots & \vdots\\ 1/\sqrt{n} & x_{i1}^{*} & x_{i2}^{*} & \cdots & x_{ik}^{*}\\ \vdots & \cdots & \vdots & \cdots & \vdots\\ 1/\sqrt{n} & x_{n1}^{*} & x_{n2}^{*} & \cdots & x_{nk}^{*}\end{bmatrix}=\begin{bmatrix}1 & 0 & 0 & \cdots & 0\\ 0 & 1 & r_{12} & \cdots & r_{1k}\\ 0 & r_{12} & 1 & \cdots & r_{1k}\\ \vdots & \cdots & \vdots & \cdots & \vdots\\ 0 & r_{1k} & r_{1k} & \cdots & 1\end{bmatrix}$$

$$\leftrightarrow\begin{bmatrix}\frac{1}{\sqrt{n}}\mathbf{1}'\\ \mathbf{X}^{*\prime}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{n}}\mathbf{1} & \mathbf{X}^{*}\end{bmatrix}=\begin{bmatrix}1 & 0\\ 0 & \mathbf{X}^{*\prime}\mathbf{X}^{*}\end{bmatrix}\leftarrow \text{Let } \mathbf{X}=\begin{bmatrix}\frac{1}{\sqrt{n}}\mathbf{1} & \mathbf{X}^{*}\end{bmatrix}\qquad \uparrow r_{jk}^{*}=\frac{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)(x_{ik}-\bar{x}_k)}{\sqrt{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}\sqrt{\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2}}$$

$$\underset{(n)x(k)}{\mathbf{X}^{*}}=\begin{bmatrix}x_{11}^{*} & x_{12}^{*} & \cdots & x_{1k}^{*}\\ \cdots & \vdots & \cdots & \vdots\\ x_{i1}^{*} & x_{i2}^{*} & \cdots & x_{ik}^{*}\\ \cdots & \vdots & \cdots & \vdots\\ x_{n1}^{*} & x_{n2}^{*} & \cdots & x_{nk}^{*}\end{bmatrix}\text{ and }\underset{\text{(Correlation matrix)}}{\mathbf{X}^{*\prime}\mathbf{X}^{*}:\ (k)x(k)}=\begin{bmatrix}1 & r_{12} & \cdots & r_{1k}\\ r_{12} & 1 & \cdots & r_{1k}\\ \cdots & \vdots & \cdots & \vdots\\ r_{1k} & r_{1k} & \cdots & 1\end{bmatrix}$$

$\mathbf{X}^{*\prime}\mathbf{X}^{*}$ shows linear relationship between two regressor.

If there is a strong linearity, then we should doubt the multicollnearity.

Correlation matrix ($\mathbf{X}^{*\prime}\mathbf{X}^{*}$) is very important itself, and its inverse matrix conducts more important works.

$$(\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1}\rightarrow\begin{bmatrix}1 & r_{12} & \cdots & r_{1k}\\ r_{12} & 1 & \cdots & r_{1k}\\ \cdots & \vdots & \cdots & \vdots\\ r_{1k} & r_{1k} & \cdots & 1\end{bmatrix}^{-1}\text{ then, }\left\{(\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1}\right\}_{jj}=\frac{1}{1-R_j^2}=\mathrm{VIF}_j\text{ in }\left(\substack{y_i=\alpha_0+\beta_1x_{i1}^{*}+\cdots+\beta_kx_{ik}^{*}+\varepsilon_i\\ \text{Scaled and Centered model}}\right)$$

VIF_j can show evidence of multicollinearity. So, with correlation matrix, we can expect these problems.

② **Variance Inflation Factors (VIF)**

$$\mathrm{VIF_j} = \frac{1}{1-\mathrm{R_j}^2} = \{(\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1}\}_{jj}$$

$\mathrm{R_j}^2$ is the coefficient of multiple determinant in a below model which explained the other regressors.
$x_j^* = \beta_1 x_1^* + \beta_2 x_2^* + \cdots + \beta_{j-1} x_{j-1}^* + \cdots + \beta_k x_k^* + \varepsilon_i \leftarrow$ intercept isn't existed because of centering.

If there is a high $\mathrm{VIF_j} \geq 10$, then we should doubt the multicollinearity.

 # Example of $\mathrm{VIF_1}$ regarding x_1

$$\mathbf{X}^* = [\,\mathbf{x_1}^* \quad \mathbf{x_2}^* \quad \cdots \quad \mathbf{x_k}^*\,] = [\,\mathbf{x_1}^* \quad \mathbf{X_1}^*\,] \rightarrow \mathbf{X}^{*\prime}\mathbf{X}^* = \begin{bmatrix} (\mathbf{x_1}^*)'(\mathbf{x_1}^*) & (\mathbf{x_1}^*)'(\mathbf{X_1}^*) \\ (\mathbf{X_1}^*)'(\mathbf{x_1}^*) & (\mathbf{X_1}^*)'(\mathbf{X_1}^*) \end{bmatrix}$$

$$\mathrm{Var}(\hat{\beta}_1) = \{(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\}_{11}\cdot\sigma^2 \leftarrow \hat{\beta}_j \text{ are in } \begin{pmatrix} y_i = \alpha_0 + \beta_1 x_{i1}^* + \cdots + \beta_k x_{ik}^* + \varepsilon_i \\ \text{Scaled and Centered model} \end{pmatrix}$$

$$= \left[(\mathbf{x_1}^*)'(\mathbf{x_1}^*) - (\mathbf{x_1}^*)'\left\{(\mathbf{X_1}^*)\left((\mathbf{X_1}^*)'(\mathbf{X_1}^*)\right)^{-1}(\mathbf{X_1}^*)'\right\}(\mathbf{x_1}^*)\right]^{-1}\cdot\sigma^2 \leftarrow \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \left(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{B}'\right)^{-1} & \cdots \\ \cdots & \cdots \end{pmatrix}$$

$$= \frac{\sigma^2}{(\mathbf{x_1}^*)'(\mathbf{x_1}^*) - (\mathbf{x_1}^*)'\left\{(\mathbf{X_1}^*)\left((\mathbf{X_1}^*)'(\mathbf{X_1}^*)\right)^{-1}(\mathbf{X_1}^*)'\right\}(\mathbf{x_1}^*) \leftarrow \text{This is a scalar}}$$

$$= \frac{\sigma^2}{(\mathbf{x_1}^*)'(\mathbf{I}-\mathbf{H}^*)(\mathbf{x_1}^*)} = \frac{\sigma^2}{\frac{(\mathbf{x_1}^*)'(\mathbf{I}-\mathbf{H}^*)(\mathbf{x_1}^*)}{(\mathbf{x_1}^*)'(\mathbf{x_1}^*)}} = \frac{\sigma^2}{\frac{\mathrm{SSE_1}}{\mathrm{SST_1}}} = \frac{\sigma^2}{1-\frac{\mathrm{SSR_1}}{\mathrm{SST_1}}} = \frac{\sigma^2}{1-\mathrm{R_1}^2} = \mathrm{VIF_1}\cdot\sigma^2$$

$$\leftrightarrow \frac{\mathrm{Var}(\hat{\beta}_1)}{\sigma^2} = \mathrm{VIF_1} = \{(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\}_{11}$$

$$\therefore \frac{\mathrm{Var}(\hat{\beta}_j)}{\sigma^2} = \mathrm{VIF_j} = \{(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\}_{jj} : \text{This is satisfied } \textbf{only} \text{ in } \begin{pmatrix} y_i = \alpha_0 + \beta_1 x_{i1}^* + \cdots + \beta_k x_{ik}^* + \varepsilon_i \\ \text{Scaled and Centered model} \end{pmatrix}$$

$$* \ (\mathbf{x_1}^*)'(\mathbf{I}-\mathbf{H}^*)(\mathbf{x_1}^*) = \frac{(\mathbf{x_1}^*)'(\mathbf{I}-\mathbf{H}^*)(\mathbf{x_1}^*)}{(\mathbf{x_1}^*)'(\mathbf{x_1}^*) \leftarrow \text{scalar '1'}} = \frac{\mathrm{SSE_1}}{\mathrm{SST_1}} = 1 - \frac{\mathrm{SSR_1}}{\mathrm{SST_1}} = 1-\mathrm{R_1}^2$$

$* \ (\mathbf{x_1}^*)'(\mathbf{x_1}^*) = 1$ and $\bar{x}_1^* = 0$ by centering

$$* \ (\mathbf{x_1}^*)'(\mathbf{x_1}^*) = \sum (x_{i1}^*)^2 = \sum (x_{i1}^* - 0)^2 = \sum (x_{i1}^* - \bar{x}_1^*)^2 = \mathrm{SST_1}$$

$$* \ (\mathbf{x_1}^*)'(\mathbf{I}-\mathbf{H}^*)(\mathbf{x_1}^*) \xrightarrow{\mathrm{SSE}=\mathbf{y}'(\mathbf{I}-\mathbf{H}^*)\mathbf{y}} \mathrm{SSE_1} \text{ in } \left(x_{i1}^* = \beta_2 x_{i2}^* + \cdots + \beta_{j-1} x_{j-1}^* + \cdots + \beta_k x_k^* + \varepsilon_i\right)$$

$$* \ \mathrm{VIF_j} \text{ from } \begin{pmatrix} y_i = \alpha_0 + \beta_1 x_{i1}^* + \cdots + \beta_k x_{ik}^* + \varepsilon_i \\ \text{Scaled and Centered model} \end{pmatrix} \text{ equals to } \mathrm{VIF_j} \text{ from } \begin{pmatrix} y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i \\ \text{Unadjusted model} \end{pmatrix}$$

 # Example

Let $\mathbf{X}^* = [\,\mathbf{x_1} \quad \mathbf{x_2}\,]$ then

① $\mathbf{X}^{*\prime}\mathbf{X}^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \rightarrow (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \rightarrow \mathrm{Var}(\hat{\beta}_{1(2)}) = \sigma^2$: $\mathbf{x_1}$ and $\mathbf{x_2}$ are orthogonal each other.

② $\mathbf{X}^{*\prime}\mathbf{X}^* = \begin{bmatrix} 1 & 0.99215 \\ 0.99215 & 1 \end{bmatrix} \rightarrow (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} = \begin{bmatrix} 63.94 & -63.44 \\ -63.44 & 63.94 \end{bmatrix} \rightarrow \frac{\mathrm{Var}(\hat{\beta}_{1(2)})}{\sigma^2} = \mathrm{VIF}_{1(2)} = 63.94$

(a strong evidence of MC)

③ Property of Eigen value - I

$$\mathbf{V}'(\mathbf{X}^{*\prime}\mathbf{X}^*)\mathbf{V} \binom{\text{Eigenvalue}}{\text{Decomposition}} = \mathbf{D} = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_k \end{bmatrix} \leftarrow \quad \mathbf{V} = [\,\mathbf{v_1} \quad \mathbf{v_2} \quad \cdots \quad \mathbf{v_k}\,] \begin{pmatrix} \mathbf{V} : \text{orthogonal matrix} \\ \mathbf{v_i} : \text{Eigen vector} \\ \lambda_i : \text{Eigen value} \end{pmatrix}$$

$* \ \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0$ If $^\forall \mathbf{x_i} \perp {}^\forall \mathbf{x_j}$ of X , then $^\forall \lambda_i = 1$

$* \ (\mathbf{X}^{*\prime}\mathbf{X}^*)$: positive semi definite matrix $\rightarrow \ \therefore \ ^\forall \lambda_i \geq 0$

$$\leftrightarrow (\mathbf{X}^{*\prime}\mathbf{X}^*) = \mathbf{V} \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_k \end{bmatrix} \mathbf{V}' \leftrightarrow (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} = \mathbf{V} \begin{bmatrix} 1/\lambda_1 & & & 0 \\ & 1/\lambda_2 & & \\ & & \ddots & \\ 0 & & & 1/\lambda_k \end{bmatrix} \mathbf{V}'$$

If a multicollinearity is presented, at least one $\lambda_i \cong 0$.

Proof)

Let $\lambda_k \cong 0 \leftrightarrow \lambda_k = \mathbf{v_k}'(\mathbf{X}^{*\prime}\mathbf{X}^*)\mathbf{v_k} = (\mathbf{X}^*\mathbf{v_k})'(\mathbf{X}^*\mathbf{v_k}) \cong 0 \ \rightarrow$ Let $\mathbf{a} = (\mathbf{X}^*\mathbf{v_k})$: (n)x(1) vector

$\leftrightarrow \lambda_k = \mathbf{a}'\mathbf{a} = \sum_{j=1}^{n} a_j^2 = (a_1^2 + a_2^2 + \cdots a_n^2) = 0 \ \rightarrow \ ^\forall a_j \cong 0 \ \therefore \ \mathbf{a} = (\mathbf{X}^*\mathbf{v_k}) \cong \mathbf{0}$

$$\mathbf{X}^*\mathbf{v_k} = [\,\mathbf{x_1}^* \quad \mathbf{x_2}^* \quad \cdots \quad \mathbf{x_k}^*\,] \begin{bmatrix} v_{1k} \\ v_{2k} \\ \vdots \\ v_{kk} \end{bmatrix} = v_{1k}\mathbf{x_1}^* + v_{2k}\mathbf{x_2}^* + \cdots + v_{kk}\mathbf{x_k}^* = \sum_{j=1}^{k} v_{jk}\mathbf{x_j}^* \cong \mathbf{0}$$

$\leftrightarrow \therefore \mathbf{x_1}^* = -\dfrac{v_{2k}}{v_{1k}}\mathbf{x_2}^* + \cdots - \dfrac{v_{kk}}{v_{1k}}\mathbf{x_k}^*$: $\mathbf{x_1}^*$ is expressed by a linear combination of the other regressors.

④ Property of Eigen value - II

$$\mathbf{E}(\hat{\boldsymbol\beta}'\hat{\boldsymbol\beta}) - \boldsymbol\beta'\boldsymbol\beta = \mathbf{E}\left((\hat{\boldsymbol\beta} - \boldsymbol\beta)'(\hat{\boldsymbol\beta} - \boldsymbol\beta)\right) = \mathbf{E}\left((\hat\beta_1 - \beta_1)^2 + (\hat\beta_2 - \beta_2)^2 + \cdots + (\hat\beta_k - \beta_k)^2\right)$$

$$= \mathrm{E}(\hat\beta_1 - \beta_1)^2 + \mathrm{E}(\hat\beta_2 - \beta_2)^2 + \cdots + \mathrm{E}(\hat\beta_k - \beta_k)^2 = \sum_{j=1}^{k} \mathrm{Var}(\hat\beta_j) = \mathrm{trace}[(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}] \cdot \sigma^2$$

$$= \mathrm{trace}\left[\mathbf{V} \cdot \mathrm{diag}\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \cdots, \frac{1}{\lambda_k}\right) \cdot \mathbf{V}'\right] \cdot \sigma^2 = \mathrm{trace}\left[\mathbf{V}'\mathbf{V} \cdot \mathrm{diag}\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \cdots, \frac{1}{\lambda_k}\right)\right] \cdot \sigma^2 = \sum_{i=1}^{k} \frac{1}{\lambda_i}\sigma^2$$

$$\leftrightarrow \therefore \mathbf{E}(\hat{\boldsymbol\beta}'\hat{\boldsymbol\beta}) = \boldsymbol\beta'\boldsymbol\beta + \sum_{i=1}^{k} \frac{1}{\lambda_i}\sigma^2 = \sum_{i=1}^{k} \beta_i^2 + \sum_{i=1}^{k} \frac{1}{\lambda_i}\sigma^2 \quad \therefore \text{If one } \lambda_i \cong 0 \text{ then, } \mathbf{E}(\hat{\boldsymbol\beta}'\hat{\boldsymbol\beta}) \rightarrow \infty$$

$*$ $\hat\beta_j$ are in $\begin{pmatrix} y_i = \alpha_0 + \beta_1 x_{i1}^* + \cdots + \beta_k x_{ik}^* + \varepsilon_i \\ \text{Scaled and Centered model} \end{pmatrix}$

$$* \ \mathbf{V}'\mathbf{V} = \begin{bmatrix} \mathbf{v_1}' \\ \mathbf{v_2}' \\ \vdots \\ \mathbf{v_k}' \end{bmatrix} [\,\mathbf{v_1} \quad \mathbf{v_2} \quad \cdots \quad \mathbf{v_k}\,] = \begin{bmatrix} \mathbf{v_1}'\mathbf{v_1} & \mathbf{v_1}'\mathbf{v_2} & \cdots & \mathbf{v_1}'\mathbf{v_k} \\ \mathbf{v_2}'\mathbf{v_1} & \mathbf{v_2}'\mathbf{v_2} & \cdots & \mathbf{v_2}'\mathbf{v_k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v_k}'\mathbf{v_1} & \mathbf{v_k}'\mathbf{v_2} & \cdots & \mathbf{v_k}'\mathbf{v_k} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \mathbf{I}_{(\mathbf{k})\mathbf{x}(\mathbf{k})} \ \leftrightarrow \ \mathbf{V}' = \mathbf{V}^{-1}$$

$*$ $\mathbf{v_j}$ are orthonormal vectors $\mathbf{v_i} \perp \mathbf{v_j}$ $(i \neq j)$ and $\mathbf{v_j}'\mathbf{v_j} = 1 \rightarrow \mathbf{V}'\mathbf{V} = \mathbf{V}\mathbf{V}' = \mathbf{I}$

$*$ **Is a multicollinearity always serious problem**?

① Multicollinearity is related with the stability of coefficients of regression.

② Multicollinearity isn't a big problem when we do prediction of $\mathrm{E}(\hat{y})$ in the range of the dataset.