Multivariate regression with orthogonal $\underset{\sim}{X}$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & & & & \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$\underset{\sim}{X}$ has <u>orthogonal</u> columns means

(*) $\sum_{i=1}^{n} x_{ji} x_{li} = 0$ when $j \neq l$

(**) $\sum_{i=1}^{n} x_{li} = 0$ for $l = 1, \cdots, k$ (column sums are 0 because of first column is $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$)

<u>Estimation of β:</u>

$$Q = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{1i} - \cdots - b_k x_{ki})^2$$

$$\frac{\partial Q}{\partial b_0} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{1i} - \cdots - b_k x_{ki})$$

$$= -2 \left[\sum_{i=1}^{n} y_i - n b_0 \right] \quad \text{by } (**)$$

so $\frac{\partial Q}{\partial b_0} = 0 \Leftrightarrow \sum_{i=1}^{n} y_i = n b_0 \Leftrightarrow b_0 = \frac{1}{n} \sum_{i=1}^{n} y_i = \bar{y}$

$$\frac{\partial Q}{\partial b_1} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{1i} - \cdots - b_k x_{ki}) x_{1i}$$

$$= -2 \left[\sum_{i=1}^{n} x_{1i} y_i - b_1 \sum_{i=1}^{n} x_{1i}^2 \right] \quad \text{by } (*) \text{ and } (**)$$

so $\frac{\partial Q}{\partial b_1} = 0 \Leftrightarrow \sum_{i=1}^{n} x_{1i} y_i = b_1 \sum_{i=1}^{n} x_{1i}^2 \Leftrightarrow b_1 = \frac{\sum_{i=1}^{n} x_{1i} y_i}{\sum_{i=1}^{n} x_{1i}^2}$

<u>Similarly</u>

$$\boxed{b_j = \frac{\sum_{i=1}^{n} x_{ji} y_i}{\sum_{i=1}^{n} x_{ji}^2} \quad \text{for } j = 1, 2, \cdots, k}$$

$$Var(b_j) = \frac{1}{\left(\sum_{i=1}^{n} x_{ji}^2\right)^2} \cdot \sum_{i=1}^{n} x_{ji}^2 \underbrace{Var(y_i)}_{\sigma^2} = \frac{\sigma^2}{\sum_{i=1}^{n} x_{ji}^2}$$

$$\boxed{Var(b_j) = \frac{\sigma^2}{\sum_{i=1}^{n} x_{ji}^2}}$$

Estimation of σ^2 etc.:

$$SST = SSR + SSE$$

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \underbrace{\bar{y}}_{b_0})^2$$

$$= \sum_{i=1}^{n} (b_0 + b_1 x_{1i} + \dots + b_k x_{ki} - b_0)^2$$

$$= \sum_{i=1}^{n} (b_1 x_{1i} + \dots + b_k x_{ki})^2 \quad \leftarrow \text{(product terms cancel by (*))}$$

so
$$\boxed{SSR = b_1^2 \sum_{i=1}^{n} x_{1i}^2 + b_2^2 \sum_{i=1}^{n} x_{2i}^2 + \dots + b_k^2 \sum_{i=1}^{n} x_{ki}^2}$$
by (*)

or
$$\boxed{SSR = R(\beta_1) + R(\beta_2) + \dots + R(\beta_k)}$$

[while in general $= R(\beta_1) + R(\beta_2 | \beta_1) + R(\beta_3 | \beta_1, \beta_2) + \dots + R(\beta_k | \beta_1, \beta_2, \dots, \beta_{k-1})$]

Thus

$$SSE = SST - SSR = \sum_{i=1}^{n} (y_i - \underbrace{\bar{y}}_{b_0})^2 - b_1^2 \sum_{i=1}^{n} x_{1i}^2 - \dots - b_k^2 \sum_{i=1}^{n} x_{ki}^2$$

Thus to test

$H_0: \beta_j = 0$ we use

$$Z_j = \frac{b_j}{\frac{\sigma}{\sqrt{\sum_{i=1}^{n} x_{ji}^2}}} = \frac{\sqrt{\sum_{i=1}^{n} x_{ji}^2} \cdot b_j}{\sigma} \sim N(0,1) \text{ \underline{if} } \beta_j = 0$$

$$\text{so } Z_j^2 = \frac{R(\beta_j)}{\sigma^2} \sim \chi^2_1 \text{ \underline{if} } \beta_j = 0$$

$$\text{and } \frac{SSR}{\sigma^2} \sim \chi^2_k \text{ \underline{if} } \beta_1 = \dots = \beta_k = 0 \text{ (no regression)}$$