1) Write down SSE for this model

model → $\hat{y}_i = b_0 + b_1 z_{i1} + b_2 z_{i2}$

residual $(e_i) = (y_i - \hat{y}_i)^2$

$$SSE(b_0, b_1, b_2) = \sum_{i=1}^{N} (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2})^2$$

2) take partial derivative w/ respect to b_0, b_1, b_2

$e_i = y_i - b_0 - b_1 z_{i1} - b_2 z_{i2}$

$\therefore$ $SSE = \sum_i e_i^2$

derivative rule

∂ → partial derivative

↳ $\frac{d}{dx}(e_i^2) = 2e_i \cdot \frac{de_i}{dx}$

a. derivative with respect to b_0

$$\frac{\partial SSE}{\partial b_0} = \sum_{i=1}^{N} 2(y_i - b_0 - b_1 z_{i1} - b_2 z_{i2})(-1)$$

$$= -2\sum_{i=1}^{N} (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2}) = -2\sum_{i=1}^{N} e_i$$

b. derivative with respect to b_1

$$\frac{\partial SSE}{\partial b_1} = \sum_{i=1}^{N} 2(y_i - b_0 - b_1 z_{i1} - b_2 z_{i2})(-z_{i1})$$

$$= -2\sum_{i=1}^{N} (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2}) z_{i1} = -2\sum_{i=1}^{N} e_i z_{i1}$$

c. derivative with respect to b_2

$$\frac{\partial SSE}{\partial b_2} = \sum_{i=1}^{N} (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2})(-z_{i2})$$

$$= -2\sum_{i=1}^{N} (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2}) z_{i2} = -2\sum_{i=1}^{N} e_i z_{i2}$$

3) Verify that avg error = 0 and $e \cdot z = 0$ @ optimum

a. avg error is zero

at optimum b_0, b_1, b_2

$$\frac{\partial SSE}{\partial b_0} = 0 = -2\sum_i e_i = 0 \Rightarrow \sum_i e_i = 0 \Rightarrow \frac{1}{N}\sum_i e_i = \bar{e} = 0$$

b. $e \cdot z = 0$

$$\frac{\partial SSE}{\partial b_0} = 0 = \sum e_i z_{i1} = 0 \ \& \ \sum e_i z_{i2} = 0$$

$$e \cdot z_j = \sum_i e_i z_{ij} \Rightarrow e \cdot z_1 = 0 \ \& \ e \cdot z_2 = 0$$

4) show that $b^*_0 = \bar{y}$

$$\frac{\partial SSE}{\partial b_0} = -2\sum_{i=1}^{N} e_i \rightarrow @ \min \rightarrow \sum_{i=1}^{N} e_i = 0$$

$$0 = \sum_{i=1}^{N} (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2}) = \sum_{i=1}^{N} y_i - N b_0 - b_1 \sum_{i=1}^{N} z_{i1} - b_2 \sum_{i=1}^{N} z_{i2}$$

$$\sum_i z_{i1} = 0 \ \& \ \sum_i z_{i2} = 0$$

$$0 = \sum_{i=1}^{N} y_i = N b_0 \Rightarrow b_0^* = \bar{y}$$

$$e_i = (y_i - \bar{y}) - b_1 z_{i1} - b_2 z_{i2}$$

$$\sum_{i=1}^{N} e_i z_{i1} = 0 \ \& \ \sum_{i=1}^{N} e_i z_{i2} = 0$$

Substitute e_i Equations:

$$\sum_{i=1}^{N} z_{i1}(y_i - \bar{y}) = b_1 \sum_{i=1}^{N} z_{i1}^2 + b_2 \sum_{i=1}^{N} z_{i1} z_{i2}$$

$$\sum_{i=1}^{N} z_{i2}(y_i - \bar{y}) = b_1 \sum_{i=1}^{N} z_{i1} z_{i2} + b_2 \sum_{i=1}^{N} z_{i2}^2$$

(b_0 is removed)

5) write eq in form "Ab = c" ⇒ normal functions

$$\begin{array}{ccc} A & B & \\ \begin{bmatrix} \sum_i z_{i1}^2 & \sum_i z_{i1}z_{i2} \\ \sum_i z_{i1}z_{i2} & \sum_i z_{i2}^2 \end{bmatrix} & \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} & = \begin{bmatrix} \sum_i z_{i1}(y_i - \bar{y}) \\ \sum_i z_{i2}(y_i - \bar{y}) \end{bmatrix} \end{array}$$

6) Divide by N & substitute $z_{ij} = x_{ij} - m_j$

$$\frac{1}{N}\left(\begin{bmatrix} \sum_i z_{i1}^2 & \sum_i z_{i1}z_{i2} \\ \sum_i z_{i1}z_{i2} & \sum_i z_{i2}^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum_i z_{i1}(y_i - \bar{y}) \\ \sum_i z_{i2}(y_i - \bar{y}) \end{bmatrix}\right)$$

$$\begin{array}{ccc} A & B & C \\ \begin{bmatrix} \frac{1}{N}\sum(x_{i1} - m_1)^2 & \frac{1}{N}\sum(x_{i1} - m_1)(x_{i2} - m_2) \\ \frac{1}{N}\sum(x_{i1} - m_1)(x_{i2} - m_2) & \frac{1}{N}\sum(x_{i2} - m_2)^2 \end{bmatrix} & \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} & = \begin{bmatrix} \frac{1}{N}\sum(x_{i1} - m_1)(y_i - \bar{y}) \\ \frac{1}{N}\sum(x_{i2} - m_2)(y_i - \bar{y}) \end{bmatrix} \end{array}$$

A → covariance matrix of predictors
↳ its avg products of x deviations
tells us how spread out & correlated predictions are

C → covariance between predictor & response
↳ avg product of deviations btwn x_i & y

dividing by N gives averages

means that coeff. proportional to how much x_i covaries w/ y
and coeffi ⇒ $\underline{b = A^{-1} C}$