Q1

1. Write down the SSE for this model.

$\hat{y}_i = b_0 + b_1 z_{i_1} + b_2 z_{i_2}$.

$e_i = y^i - \hat{y}^i = y^i - b_0 - b_1 z_{i_1} - b_2 z_{i_2}$

$$SSE = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (y^i - b_0 - b_1 z_{i_1} - b_2 z_{i_2})^2$$

2. Take partial derivatives w/respect to b_0, b_1, and b_2.

Derivative w/respect to b_0:

$$\frac{\partial SSE}{\partial b_0} = \sum_{i=1}^{N} 2(y^i - b_0 - b_1 z_{i_1} - b_2 z_{i2})(-1)$$

$$= -2 \sum_{i=1}^{N} (y^i - b_0 - b_1 z_{i_1} - b_2 z_{i_2})$$

$$\sum_{i=1}^{N} (y^i - b_0 - b_1 z_{i_1} - b_2 z_{i_2}) = 0$$

Derivative w/respect to b_1:

$$\frac{\partial SSE}{\partial b_1} = \sum_{i=1}^{N} 2(y^i - b_0 - b_1 z_{i_1} - b_2 z_{i2})(-z_{i1})$$

$$= -2 \sum_{i=1}^{N} (y^i - b_0 - b_1 z_{i_1} - b_2 z_{i_2}) z_{i1}$$

$$\sum_{i=1}^{N} (y^i - b_0 - b_1 z_{i1} - b_2 z_{i2}) z_{i_1} = 0$$

Derivative w/respect to b_2:

$$\frac{\partial SSE}{\partial b_2} = \sum_{i=1}^{N} 2(y^i - b_0 - b_1 z_{i_1} - b_2 z_{i_2}) \cdot (-z_{i2})$$

$$= -2 \sum_{i=1}^{N} (y^i - b_0 - b_1 z_{i_1} - b_2 z_{i_2}) z_{i2}$$

$$\sum_{i=1}^{N} (y^i - b_0 - b_1 z_{i_1} - b_2 z_{i_2}) z_{i2} = 0$$

3. Verify that average error is zero and $e \cdot z = 0$ at the optimum, just as in the single linear regression case.

$e_i = y^i - b_0 - b_1 z_{i1} - b_2 z_{i2}$

First order conditions: $\sum_{i=1}^{N} e_i = 0$ $\quad \sum_{i=1}^{N} e_i z_{i1} = 0$ $\quad \sum_{i=1}^{N} e_i z_{i2} = 0$

Average: $\frac{1}{N} \sum_{i=1}^{N} e_i = 0$

Residuals orthogonal to predictors: $e \cdot z_1 = \sum_{i=1}^{N} e_i z_{i1} = 0$

$e \cdot z_2 = \sum_{i=1}^{N} e_i z_{i2} = 0$

4. Show that the optimal intercept is $b_0^* = \bar{y}$. Eliminate b_0^* from the remaining equations, and focus on b_1 and b_2.

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

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

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

$$\sum_{i=1}^{N} z_{i1} = 0, \quad \sum_{i=1}^{N} z_{i2} = 0$$

$$\sum_{i=1}^{N} y^i - N b_0 = 0$$

$$b_0^* = \frac{1}{N} \sum_{i=1}^{N} y^i = \bar{y}$$

5. Write your results as a matrix equation in the form "Ab=C". These are called the normal equations.

$$\begin{bmatrix} \sum z_{i1}^2 & \sum z_{i1} z_{i2} \\ \sum z_{i1} z_{i2} & \sum z_{i2}^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum (y^i - \bar{y}) z_{i1} \\ \sum (y^i - \bar{y}) z_{i2} \end{bmatrix}$$

6. Divide both sides by N and substitute $z_{ij} = x_{ij} - M_j$ back into your normal equations for x_{ij}. What is the matrix A? What is the vector C? Explain the intuition of your discovery.

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

$$\frac{1}{N} \sum_{i=1}^{N} z_{i1}^2 = \frac{1}{N} \sum_{i=1}^{N} (x_{i1} - M_1)^2 = Var(x_1)$$

$$\frac{1}{N} \sum_{i=1}^{N} z_{i2}^2 = Var(x_2), \quad \frac{1}{N} \sum_{i=1}^{N} z_{i1} z_{i2} = Cov(x_1, x_2)$$

$$\frac{1}{N} \sum_{i=1}^{N} (y^i - \bar{y}) z_{i1} = Cov(x_1, y), \quad \frac{1}{N} \sum_{i=1}^{N} (y^i - \bar{y}) z_{i2} = Cov(x_2, y)$$

$$\underbrace{\begin{bmatrix} Var(x_1) & Cov(x_1, x_2) \\ Cov(x_1, x_2) & Var(x_2) \end{bmatrix}}_{A} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \underbrace{\begin{bmatrix} Cov(x_1, y) \\ Cov(x_2, y) \end{bmatrix}}_{C}$$

Matrix A captures how the predictors vary and relate to each other, vector C captures how each predictor relates to the response, and solving Ab=C gives slopes that account for correlations between predictors to minimize residuals.