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

2. $\frac{dSSE}{db_0} = -2 \sum_i y_i - b_0 - b_1 z_{i1} - b_2 z_{i2}$

$\frac{dSSE}{db_1} = -2 \sum_i z_{i1} (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2})$

$\frac{dSSE}{db_2} = -2 \sum_i z_{i2} (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2})$

3. residuals $= e_i = y_i - \hat{y}_i = y_i - b_0 - b_1 z_{i1} - b_2 z_{i2}$

At optimum partial derivative $= 0$

$MSE = \frac{1}{N} \sum_i e_i$

$\frac{dSSE}{db_0} = -2 \sum_i e_i = 0$ at optimum and $\sum_i e_i = 0$

therefore $MSE = \frac{1}{N}(0) = 0$

$\frac{dSSE}{db_1} = -2 \sum_i z_{i1} e_i = 0$ and $\frac{dSSE}{db_2} = \sum_i z_{i2} e_i = 0$

4. At optimum: $\sum_{i=1}^{N} e_i = 0$

$\sum_i (y_i - b_0 - b_1 z_{i1} - b_2 z_{i2}) = 0$

$$\sum_i y_i - N b_0 \underbrace{- b_1 \sum_i z_{i1} - b_2 \sum_i z_{i2}}_{0 \text{ at optimum}} = 0$$

$\sum_i y_i - N b_0 = 0$

$\sum_i y_i = N b_0$

$b_0^* = \sum y_i / N = \bar{y}$

5. $$\begin{pmatrix} N & \cancelto{0}{\sum z_{i1}} & \cancelto{0}{\sum z_{i2}} \\ \cancelto{0}{\sum z_{i1}} & \sum z_{i1}^2 & \sum z_{i1} z_{i2} \\ \cancelto{0}{\sum z_{i2}} & \sum z_{i1} z_{i2} & \sum z_{i2}^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i z_{i1} \\ \sum y_i z_{i2} \end{pmatrix}$$

$$\underbrace{\begin{pmatrix} N & 0 & 0 \\ 0 & \sum z_{i1}^2 & \sum z_{i1} z_{i2} \\ 0 & \sum z_{i1} z_{i2} & \sum z_{i2}^2 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}}_{b} = \underbrace{\begin{pmatrix} \sum y_i \\ \sum y_i z_{i1} \\ \sum y_i z_{i2} \end{pmatrix}}_{C}$$

6. $$\frac{A}{N} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{N}\sum z_{i1}^2 & \frac{1}{N}\sum z_{i1} z_{i2} \\ 0 & \frac{1}{N}\sum z_{i1} z_{i2} & \frac{1}{N}\sum z_{i2}^2 \end{pmatrix}$$

$$\left\{ \begin{array}{l} var(X) = \frac{1}{N} \sum (x_{ij} - m_j)^2 = \frac{1}{N} \sum z_{ij}^2 \\ cov(X_1, X_2) = \frac{1}{N} \sum (x_1 - m_1)(x_2 - m_2) = \frac{1}{N} \sum z_1 z_2 \end{array} \right.$$

$$\frac{A}{N} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & var(x_1) & cov(x_1, x_2) \\ 0 & cov(x_1, x_2) & var(x_2) \end{pmatrix} \qquad \frac{C}{N} = \begin{pmatrix} \frac{1}{N}\sum y_i \\ \frac{1}{N}\sum y_i z_{i1} \\ \frac{1}{N}\sum y_i z_{i2} \end{pmatrix} = \begin{pmatrix} \bar{y} \\ cov(y, x_1) \\ cov(y, x_2) \end{pmatrix}$$

Dividing by N or the sample size takes the average of the product. Variance is the average of squared terms and covariance is the average of cross-product terms. Therefore it makes sense that dividing by N leads to variance and covariance