Q2: Show the following equations -

β_0

a) $\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})(Y_i - \hat{Y}_i) = 0$

$$LHS = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})(Y_i - \hat{Y}_i)$$

$$= \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y}) e_i$$

$$= \sum_{i=1}^{n} \hat{Y}_i e_i - \bar{Y} \sum_{i=1}^{n} e_i$$

$$= \sum_{i=1}^{n} \hat{Y}_i e_i - 0$$

$$= 0$$

Properties of fitted regression line:

1. Sum of residuals is zero:

$$\sum_{i=1}^{n} e_i = 0$$

5. Sum of the weighted residuals is zero, when the residual in the i^{th} trial is weighted by the fitted value of the response variable for the i^{th} trial:

$$\sum_{i=1}^{n} \hat{Y}_i e_i = 0$$

b) $E(b_1) = \beta_1$ and $var(b_1) = \dfrac{\sigma^2}{S_{xx}}$

In Week 1's lecture, it has been shown that —

$$b_1 = \sum_{i=1}^{n} k_i Y_i \rightarrow b_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

where $k_i = \frac{X_i - \bar{X}}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{X_i - \bar{X}}{S_{xx}}$ and these constants k_i have the following properties

$$\sum_{i=1}^{n} k_i = 0, \quad \sum_{i=1}^{n} k_i X_i = 1, \quad \sum_{i=1}^{n} k_i^2 = \frac{1}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{1}{S_{xx}}.$$

Then,

$$E(b_1) = E\left(\sum_{i=1}^{n} k_i Y_i\right)$$

since - $E(a) = a$, a: constant

$$= \sum_{i=1}^{n} k_i E(Y_i)$$

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

$E(Y_i) = \beta_0 + \beta_1 X_i$

$$= \sum_{i=1}^{n} k_i (\beta_0 + \beta_1 X_i)$$

$$= \beta_0 \sum_{i=1}^{n} k_i + \beta_1 \sum_{i=1}^{n} k_i X_i$$

Since - $\sum_{i=1}^{n} k_i = 0$ and $\sum_{i=1}^{n} k_i X_i = 1$

$$= 0 + \beta_1 \times 1$$

$$= \beta_1$$

$\sum_{i=1}^{n} k_i = 0$: $\sum_{i=1}^{n} \frac{x_i - \bar{x}}{S_{xx}} = \frac{1}{S_{xx}} \sum_{i=1}^{n} x_i - \bar{x}$

$$= \frac{1}{S_{xx}} \left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} \right]$$

$$= \frac{1}{S_{xx}} \left[\sum_{i=1}^{n} x_i - n\bar{x} \right]$$

$$= \frac{1}{S_{xx}} \left[\sum_{i=1}^{n} x_i - n\left(\frac{1}{n} \sum_{i=1}^{n} x_i\right) \right]$$

$$= \frac{1}{S_{xx}} \left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i \right]$$

$$= 0$$

$\sum_{i=1}^{n} k_i x_i = 1$

$$\sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{S_{xx}} \right) x_i = \frac{1}{S_{xx}} \sum_{i=1}^{n} (x_i - \bar{x}) x_i$$

$$= \frac{1}{S_{xx}} \left[\sum_{i=1}^{n} x_i^2 - \bar{x} \sum_{i=1}^{n} x_i \right]$$

$$= \frac{1}{S_{xx}} \left[\sum_{i=1}^{n} x_i^2 - \bar{x}^2 \cdot n \right]$$

Then, we can show $S_{xx} = \sum_{i=1}^{n} x_i^2 - \bar{x}^2 \cdot n$:

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \sum_{i=1}^{n} (x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

$$= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x}^2$$

$$= \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2$$

$$= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$, hence shown

$$= \frac{1}{S_{xx}} \cdot S_{xx}$$

$$= 1$$

As responses Y_i are uncorrelated,

$$\text{var}(b_1) = \text{Var}\left(\sum_{i=1}^{n} k_i Y_i\right)$$

$$= \sum_{i=1}^{n} k_i^2 \,\text{var}(Y_i)$$

$$= \sum_{i=1}^{n} k_i^2 \sigma^2$$

$$= \sigma^2 \sum_{i=1}^{n} k_i^2$$

$$= \frac{\sigma^2}{S_{xx}}$$

Since -

$\text{var}(aX) = a^2 \text{Var}(X)$,

a: constant

X : random variable

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$\text{var}(Y_i) = \text{Var}(\beta_0 + \beta_1 X_i + \varepsilon_i)$$

$$= \text{var}(\varepsilon_i)$$

$$= \sigma^2$$

Since -

$$\sum_{i=1}^{n} k_i^2 = \frac{1}{S_{xx}}$$

$$\sum_{i=1}^{n} k_i^2 = \frac{1}{S_{xx}}$$

$$\sum_{i=1}^{n} k_i^2 = \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{S_{xx}}\right)^2$$

$$= \frac{1}{S_{xx}^2} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

$$= \frac{1}{S_{xx}^2} \cdot S_{xx}$$

$$= \frac{1}{S_{xx}}$$

c) $E(b_0) = \beta_0$ and $Var(b_0) = \sigma^2\left[\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}}\right]$

As we know, $b_0 = \bar{Y} - b_1\bar{X}$:

$$E(b_0) = E(\bar{Y} - b_1\bar{X})$$
$$= E(\bar{Y}) - \bar{X}E(b_1)$$
$$= E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) - \bar{X}\beta_1$$
$$= \frac{1}{n}\sum_{i=1}^{n} E(Y_i) - \beta_1\bar{X}$$
$$= \frac{1}{n}\sum_{i=1}^{n}(\beta_0 + \beta_1 X_i + \varepsilon_i) - \beta_1\bar{X}$$
$$= \beta_0 + \cancel{\beta_1\bar{X}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i}_{=0} - \cancel{\beta_1\bar{X}}$$
$$= \beta_0$$

Then, we can rewrite b_0 as:

$$b_0 = \bar{Y} - b_1\bar{X}$$
$$= \frac{1}{n}\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} k_i Y_i \cdot \bar{X}$$
$$= \sum_{i=1}^{n} Y_i\left(\frac{1}{n} - k_i\bar{X}\right)$$
$$= \sum_{i=1}^{n} c_i Y_i$$

where $c_i = \frac{1}{n} - k_i\bar{X}$

Then,

$$\sum_{i=1}^{n} c_i^2 = \sum_{i=1}^{n}\left(\frac{1}{n} - k_i\bar{X}\right)^2$$

$$= \sum_{i=1}^{n}\left(\frac{1}{n^2} - \frac{2}{n}k_i\bar{X} + k_i^2\bar{X}^2\right)$$

$$= \sum_{i=1}^{n}\frac{1}{n^2} - \frac{2\bar{X}}{n}\sum_{i=1}^{n}k_i + \bar{X}^2\sum_{i=1}^{n}k_i^2$$

$$= n\cdot\frac{1}{n^2} - \cancel{\frac{2\bar{X}}{n}\cdot 0} + \frac{\bar{X}^2}{S_{xx}}$$

$$= \frac{1}{n} + \frac{\bar{X}^2}{S_{xx}}$$

Hence,

$$Var(b_0) = Var\left(\sum_{i=1}^{n} c_i Y_i\right)$$

$$= \sum_{i=1}^{n} c_i^2 \cdot Var(Y_i)$$

$$= \sum_{i=1}^{n} c_i^2 \cdot \sigma^2$$

$$= \sigma^2\sum_{i=1}^{n} c_i^2$$

$$= \sigma^2\left[\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}}\right]$$