Linear Regression HW 1 MATH PART

① Prove: $\sum \hat{y_i}\hat{e_i} = 0$

Proof: We know $\sum \hat{e_i} = 0$. In SLR, we minimize

$$Q = \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 x_i))^2$$

Set $$\frac{\partial Q}{\partial \beta_0} = \sum_{i=1}^{n} 2(Y_i - (\beta_0 + \beta_1 x_i))(-1) = 0$$

$$\frac{\partial Q}{\partial \beta_1} = -2\sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 x_i))\, x_i = 0$$

let $\hat{Y_i} = \beta_0 + \beta_1 x_i$, $\hat{e_i} = Y_i - \hat{Y_i}$
and we get

$$\sum_{i=1}^{n} (Y_i - \hat{Y_i}) = 0 \Rightarrow \sum_{i=1}^{n} \hat{e_i} = 0$$

$$\sum_{i=1}^{n} x_i (y_i - \hat{y_i}) = 0. \Rightarrow \sum_{i=1}^{n} x_i \hat{e_i} = 0$$

Since

$$\beta_1 x_i + \beta_0 = \hat{Y},$$

$$\frac{\hat{Y} - \beta_0}{\beta_1} = x_i.$$

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

$$\Rightarrow \sum_{i=1}^{n} \frac{(\hat{y_i} - \beta_0)}{\beta_1} \hat{e_i} = 0$$

$$\Rightarrow \frac{1}{\beta_1} \sum_{i=1}^{n} (\hat{y_i}\hat{e_i} - \beta_0 \hat{e_i}) = 0$$

$$\Rightarrow \frac{1}{\beta_1} \left(\sum_{i=1}^{n} \hat{y_i}\hat{e_i} - \beta_0 \sum_{i=1}^{n} \hat{e_i} \right) = 0 \quad \text{and} \quad \sum_{i=1}^{n} \hat{e_i} = 0$$

So $$\frac{1}{\beta_1} \sum_{i=1}^{n} \hat{y_i}\hat{e_i} = 0$$

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

② $\sum (y_i - \hat{y_i})(\hat{y_i} - \bar{y})$

$= \sum \hat{e_i}(\hat{y_i} - \bar{y})$

$= \sum \hat{e_i}\hat{y_i} - \sum \hat{e_i}\bar{y}$ as proved in ①, $\sum_{i=1}^{n} \hat{e_i}\hat{y_i} = 0$ and $\sum_{i=1}^{n} e_i = 0$

$= 0$

B.)

$$\hat{\beta}_{Y|X} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{cov(x,y)}{var(x)} = r_{xy} \frac{S_Y}{S_X}.$$

$sd(x) = sd(y) \Rightarrow var(x) = var(y)$

$$\Rightarrow \hat{\beta}_{Y|X} = \frac{cov(x,y)}{var(y)} = \hat{\beta}_{X|Y}$$

$$S_Y = S_X \Rightarrow \frac{S_Y}{S_X} = 1 \Rightarrow \hat{\beta}_{Y|X} = \hat{\beta}_{X|Y} = r_{XY}.$$

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$ $\quad \bar{y}$ & $\bar{x}$ are constants.

So $\hat{\alpha}_{Y|X} = \hat{\alpha}_{X|Y}$ as $\hat{\beta}_{Y|X} = \hat{\beta}_{X|Y}$

C) let $\alpha_{X|Y} = \alpha_{Y|X} = \alpha$, $\beta_{X|Y} = \beta_{Y|X} = \beta$.

then we have $y = \beta x + \alpha$ and $x = \beta y + \alpha$

$\Rightarrow \frac{1}{\beta}x - \frac{\alpha}{\beta} = y$ & $\beta x + \alpha = y$

Since $\frac{1}{\beta} = \beta$ and $-\frac{\alpha}{\beta} = \alpha$ can't be realized

they're different.

3. This is a biased sample. The children who had difficulty in reading cannot represent the whole population, because proper randomization is not achieved. Also, they should set up a control group and an experimental group, to make sure that the improvement comes from the treatment. Otherwise, I could propose that

the improvement is due to other reasons such as the natural mental development of kids in a year.