Practice Sheet-5

1) a) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n$

let $L = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$

$$\frac{\partial L}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \beta_0 - \sum_{i=1}^{n} \beta_1 x_i = 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \beta_0 = \frac{\sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i}{n}$$

$$= \bar{y} - \beta_1 \bar{x}$$

$$\frac{\partial L}{\partial \beta_1} = -2 \sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} \beta_0 x_i - \sum_{i=1}^{n} \beta_1 x_i^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i = (\bar{y} - \beta_1 \bar{x}) \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i = (\bar{y} - \beta_1 \bar{x}) n\bar{x} + \beta_1 \sum_{i=1}^{n} x_i^2$$

$$\Rightarrow \beta_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

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

$$\Rightarrow \boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2}}$$

$$\boxed{\hat{\beta}_0 = \bar{y} - \bar{x}\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum x_i^2 - n\bar{x}^2}\right)}$$

b) ε_i is a random error with $E(\varepsilon_i)=0$ and $Var(\varepsilon_i) = \sigma^2$; this is the assumption made on ε_i, $i=1,2,\dots,n$

c) Since $\varepsilon_i \sim N(0,\sigma^2)$

we have $y_i \sim N(\beta_0+\beta_1 x_i, \sigma^2)$

$$L = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\,\sigma}\exp.\left[-\frac{(y_i-\beta_0-\beta_1x_i)^2}{2\sigma^2}\right]$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp.\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)^2\right]$$

$$\Rightarrow \ln L = -n\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)^2$$

If we do $\frac{\partial}{\partial\beta_0}\ln L = 0$ and $\frac{\partial}{\partial\beta_1}\ln L = 0$ and solve the 2 equations, we get

$$\left.\begin{array}{l}\beta_0 = \hat{\beta}_0\\ \beta_1 = \hat{\beta}_1\end{array}\right\} \text{part (a)}$$

$\boxed{\text{So } \hat{\beta}_0 \text{ and } \hat{\beta}_1 \text{ are the MLE of } \beta_0 \text{ and } \beta_1 \text{ resp.}}$

2.) $SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$ $\quad [\hat{\beta}_1 = S_{xy}/S_{xx}]$

$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$

we have,

$$SSE = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

$$= \sum_{i=1}^{n} \left(y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i\right)^2$$

$$= \sum_{i=1}^{n} [(y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})]^2$$

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2 - 2\hat{\beta}_1 \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) + \hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= SST - 2 \frac{S_{xy}}{S_{xx}} \times S_{xy} + \left(\frac{S_{xy}}{S_{xx}}\right)^2 S_{xx}$$

$$= SST - 2 \frac{S_{xy}^2}{S_{xx}} + \frac{S_{xy}^2}{S_{xx}}$$

$$= SST - \frac{S_{xy}^2}{S_{xx}}$$

$$\Rightarrow SST = SSE + \frac{S_{xy}^2}{S_{xx}}$$

Now we have to prove $SSR = \frac{S_{xy}^2}{S_{xx}}$

$$SSR = \sum_{i=1}^{n} (\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y})^2$$

$$= \sum_{i=1}^{n} \left(-\hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i\right)^2$$

$$= \hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \left(\frac{S_{xy}}{S_{xx}}\right)^2 S_{xx} = \frac{S_{xy}^2}{S_{xx}}$$

Hence, proved.

3.) $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$

Applying this formula for each of the 4 datasets in Anscombe's quartet, we get the same value of $\boxed{r = 0.816}$

All the 4 datasets have same r, regression line, R^2, but do not display the same relationship.

4.) a.) $S_{xx} = \sum x_i^2 - \dfrac{(\sum x_i)^2}{n}$

$$= 251970 - \frac{(1950)^2}{18}$$

$$= \boxed{40720}$$

$$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

$$= 5530.92 - \frac{1950 \times 47.92}{18}$$

$$= \boxed{338.87}$$

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$= 130.6074 - \frac{(47.92)^2}{18}$$

$$\approx \boxed{3.034}$$

b.) $$R^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}}$$

$$= \frac{(338.87)^2}{40720 \times 3.034}$$

$$= \boxed{0.9295}$$

c.) $$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{338.87}{40720} = \boxed{0.0083}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{47.92}{18} - 0.0083 \times \frac{1950}{18}$$

$$= \boxed{1.763}$$

5.) a.) $\hat{Y} = XB$ where $\boxed{\beta = \begin{bmatrix} -199.556 \\ 0.21 \\ 3 \end{bmatrix}}$

b.) $$\text{F-statistic} = \frac{MSR}{MSE} = \frac{357.75}{1.12037} = \boxed{319.314}$$

c.) $$R^2 = \frac{SSR}{SST} = \frac{715.5}{722.2222} = \boxed{0.9907}$$

—— x ——