Recitation Hour (03.06.2020)

1) i) Use least-squares regression to fit a straight line to the following data.

| x | 1 | 3 | 5 | 7 | 10 | 12 | 13 | 16 | 18 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| y | 4 | 5 | 6 | 5 | 8 | 7 | 6 | 9 | 12 | 11 |

ii) Calculate the correlation coefficient.

Soln.

$y_i = \alpha + \beta x_i + e_i$

$e_i = y_i - (\alpha + \beta x_i)$

$n$ : # of samples.

$$S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

$$\beta = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{S_{xy}}{S_{xx}}$$

$$\beta = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

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

$S_x$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

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

$$= \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$= \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n}$$

$n = 10$

$\sum x_i = 105$

$\sum y_i = 73$

$\bar{x} = \frac{1}{10} \sum x_i = 10.5$

$\bar{y} = \frac{1}{10} \sum y_i = 7.3$

$\sum x_i^2 = 1477$

$\sum x_i y_i = 906$

$$\beta = \frac{S_{xy}}{S_{xx}} = \frac{906 - \frac{105 \cdot 73}{10}}{1477 - \frac{105^2}{10}} = 0.37725$$

$$\alpha = \bar{y} - \beta \bar{x}$$

$$\alpha = 7.3 - 0.37725 \cdot (10.5) = 3.388$$

$$y_i = 3.39\, x_i - 0.37725$$

ii) $$r = \frac{S_{xy}}{\sigma_x \sigma_y} = \frac{906 - \frac{105 - 73}{10}}{\sqrt{\sum_{i=1}^{10} (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^{10} (x_i - \bar{x})^2}} = 0.9$$

2) Consider fitting the simple linear regression model.

$$\hat{y}_i = \alpha + \beta x_i + e_i$$

to the following bivariate data:

| $i$ | $x_i$ | $y_i$ |
|---|---|---|
| 1 | -5 | -2 |
| 2 | -2 | 0 |
| 3 | 3 | 3 |
| 4 | 4 | 5 |

a) Solve the least squares estimates of the intercept and slope.

b) find the distribution of $\hat{\alpha}$, the slope of the least square line.

soln.

$$\bar{x} = \frac{1}{4}\sum_{i=1}^{4} x_i = 0$$

$$\bar{y} = \frac{1}{4}\sum_{i=1}^{4} y_i = 1.5$$

$$S_{xx} = \sum_{i=1}^{4} (x_i - \bar{x})^2 = 54$$

$$S_{xy} = \sum_{i=1}^{4} (x_i - \bar{x})(y_i - \bar{y})$$

$$= \sum_{i=1}^{4} x_i y_i - \underbrace{\sum_{i=1}^{n} x_i}_{0} \cdot \sum_{i=1}^{n} y_i = \sum_{i=1}^{4} x_i y_i = 39$$

$$\beta = \frac{S_{xy}}{S_{xx}} = \frac{39}{54} = 0.722$$

$$\alpha = \bar{y} - \beta\bar{x} \Rightarrow 1.5 - 0.7222 \cdot (0) = 1.5$$

3

b) $E[\hat{\alpha}] = \alpha$

$$\sigma_{\hat{\alpha}}^2 = \sigma^2 \left[ \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right]$$

$$E[\hat{\alpha}] = \alpha = 1,5$$

$$\sigma^2 = Var[\epsilon_i]$$

$$\epsilon_i = y_i - \hat{y}_i$$

$$\epsilon_i = y_i - (\alpha x_i + \beta) \rightarrow \text{residuals.}$$

for $i=1$
$$\epsilon_1 = -2 - (1,5 \cdot (-5) + 0,722) = 4,78$$

for $i=2$
$$\epsilon_2 = 0 - (1,5 \cdot (-2) + 0,722) = 2,28$$

for $i=3$
$$\epsilon_3 = 3 - (1,5 \cdot (3) + 0,722) = -2,22$$

for $i=4$
$$\epsilon_4 = 5 - (1,5 \cdot (4) + 0,722) = -1,722$$

$$\bar{\epsilon} = \frac{1}{4} \sum_{i=1}^{n} \epsilon_i = \frac{3,12}{4} = 0,8$$

$$\sigma^2 = Var(\epsilon_i) = \sum_{i=1}^{4} (\epsilon_i - \bar{\epsilon}) = 33,51$$

$$\sigma_{\hat{\alpha}}^2 = \sigma^2 \left[ \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right] = 33,51 \left[ \frac{1}{4} + \frac{\cancel{0}^2}{S_{xx}} \right] = 8,3779.$$

4