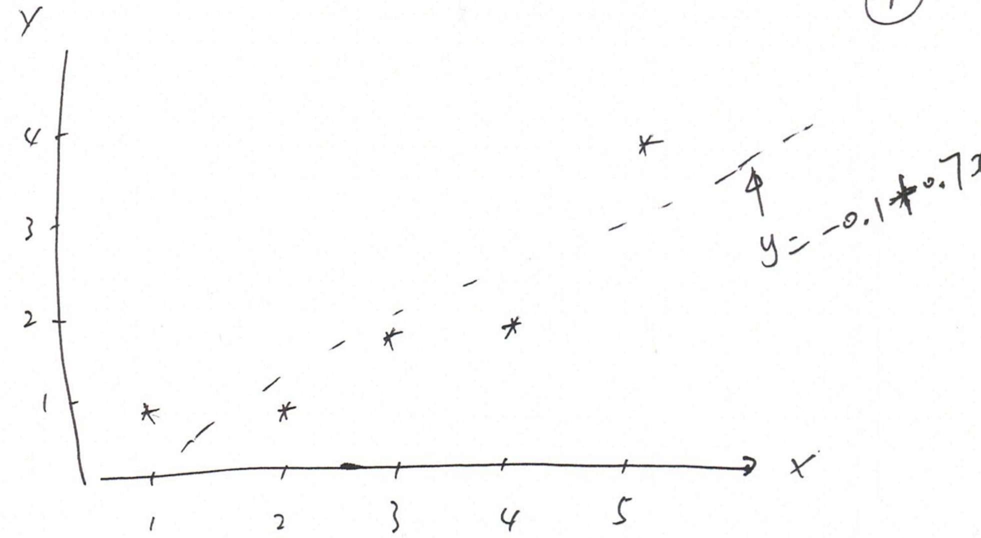
Assignment 1
Solutions.

1. (a)



It shows a rough linear relationship between y and x, but the correlation is quite strong.

(b) A linear regression model can be expressed as

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2).$$

main assumptions are:

(i) the relationship between y_i and x_i is linear

(ii) the variance of y is a constant

(iii) the observations are independent

(iv) (optional) Y is distributed normally.

(c) $\bar{x} = 3 \quad \bar{y} = 2$

$$Sxx = \Sigma(x_i - \bar{x})^2 = 10$$

$$Sxy = \Sigma(x_i - \bar{x})(y_i - \bar{y}) = 7$$

$$Syy = \Sigma(y_i - \bar{y})^2 = 6$$

$$\hat{\beta}_1 = Sxy/Sxx = 7/10 = 0.7$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 2 - 0.7 * 3 = -0.1$$

Thus the least squares line is $y = -0.1 + 0.7x$, which has been added to the scatterplot

(d) $H_0: \beta_1 = 0$ v.s. $H_1: \beta_1 \neq 0$ $n=5$

$$SSE = \Sigma(y_i - \hat{y}_i)^2 = 1.1$$

$$s^2 = \frac{SSE}{n-2} = 0.3667$$

$$t = \frac{\hat{\beta}_1}{(S/S_{xx})^{1/2}} = \frac{0.7}{(0.3667/10)^{1/2}} = 3.6556$$

$$\geq t_{0.025,3} = 3.1824$$

Thus, reject H_0, meaning the Adverting Expenditure has effect of the Sales Revenue.

(e) $t = \frac{\hat{\beta}_1 - \beta_1}{S/S_{xx}^{1/2}} \sim t_3$, the 95% C.I. of β_1 is

$$\hat{\beta}_1 \pm t_{0.025,3} \cdot S/S_{xx}^{1/2}$$

$$= 0.7 \pm 3.1824 * \left(\frac{0.3667}{10}\right)^{1/2}$$

$$= (0.0906, 1.3094)$$

There is 95% chance that β_1 would take values between 0.0906 and 1.3094.

(f) $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{1.1}{6} = 0.8167$

meaning 81.67% percent of the variations of the Sales Revenue can be explained by the model.

(g) $x_h = 4$, $\hat{Y}_h = \hat{\beta}_0 + \hat{\beta}_1 x_h = 2.7$

C.I. of $E(Y_h)$ is

$$\hat{Y}_h \pm t_{0.025,3} \cdot s \cdot \sqrt{\frac{1}{n} + \frac{(x_h - \bar{x})^2}{S_{xx}}}$$

$= (1.6445, 3.7559)$

Predictive interval of Y_h at $x_h = 4$ is

$$\hat{Y}_h \pm t_{0.025,3} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_h - \bar{x})^2}{S_{xx}}}$$

$= (0.5028, 4.8972)$.

Assignment 1.

Data $\begin{pmatrix} x_i \\ y_i \end{pmatrix}$ $i=1,\cdots n$.

2. (a) $y_i = \beta_0 + \beta_1 x_i + e_i$ $\quad e_i \overset{iid}{\sim} N(0, \sigma^2)$.

$$l(\underline{\theta}) \propto \sum_{i=1}^{n} \log p(y_i | \underline{\theta}) \qquad \underline{\theta} = (\beta_0, \beta_1, \sigma^2).$$

$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$. independently for $i=1,\cdots n$.

$$p(y_i | \underline{\theta}) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{ \frac{-(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \right\}.$$

$$\log p(y_i | \underline{\theta}) = -\frac{1}{2}\log(2\pi) - \log\sigma + \frac{-(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}.$$

$$\Rightarrow l(\underline{\theta}) \propto \sum_{i=1}^{n} \left[-\log\sigma + \frac{-(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \right]$$

$$= -n\log\sigma - \sum_{i=1}^{n} \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \qquad \underline{\theta} = (\beta_0, \beta_1, \sigma^2).$$

$$\frac{\partial l(\underline{\theta})}{\partial \beta_1} = -\sum_{i=1}^{n} \frac{2(y_i - \beta_0 - \beta_1 x_i)(-x_i)}{2\sigma^2} = 0 \Rightarrow \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)\, x_i = 0 \quad ①$$

$$\frac{\partial l(\underline{\theta})}{\partial \beta_0} = +\sum_{i=1}^{n} \frac{2(y_i - \beta_0 - \beta_1 x_i)}{2\sigma^2} = 0 \Rightarrow \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0 \quad ②$$

$$② \Rightarrow \sum y_i - n\beta_0 - \beta_1 \sum x_i = 0 \Rightarrow \beta_0 = \frac{\sum y_i - \beta_1 \sum x_i}{n}$$

$$= \frac{n\bar{y} - \beta_1 n\bar{x}}{n} = \bar{y} - \beta_1 \bar{x} \quad \text{代入} ①$$

$$\sum (y_i - \bar{y} + \beta_1 \bar{x} - \beta_1 x_i)\, x_i = 0.$$

$$\sum x_i y_i - \bar{y}\sum x_i + \underline{\underline{\beta_1}}\bar{x}\sum x_i - \underline{\underline{\beta_1}} \sum x_i^2 = 0 \Rightarrow \hat{\beta}_1 = \frac{-n\overline{xy} + n\bar{x}\bar{y}}{n\bar{x}^2 - n\overline{x^2}}$$

$$\Rightarrow \hat{\beta}_1 = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} = \frac{S_{xy}}{S_{xx}}.$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

$$l(\underline{\theta}) \propto -\frac{n}{2}\log\sigma^2 - \sum_{i=1}^{n}\frac{(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma^2}$$

$$\frac{\partial l(\underline{\theta})}{\partial\sigma^2} = \frac{-n}{2\sigma^2} + \sum_{i=1}^{n}\frac{(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma^4} = 0.$$

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

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

(b) ** To obtain LSE, we don't need to have a normal assumption for ϵ_i.

** The estimation of $\hat{\beta}_0$ and $\hat{\beta}_1$ is the same for MLE. and LSE.

** Using LSE, we usually have $\hat{\sigma}^2_{LSE} = \frac{SSE}{n-2}$ which is an unbiased MLE of σ^2. MLE is an asymptotic unbiased estimator of σ^2.