MA 542: Regression Analysis HW-2: Solutions

Q1 $\beta_0 = 100,\quad \beta_1 = 20,\quad \sigma^2 = 25,\quad X = 5$

a) $Y = \underbrace{100 + 20(5)}_{200} + \varepsilon$, where $E(\varepsilon) = 0$ and $\text{var}(\varepsilon) = 25$.

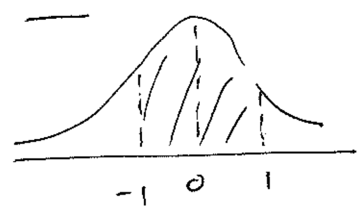
No, Probabilities of Y can not be found, because the distribution of Y is unknown.

b) Yes, now $Y \sim N(200, 25)$.

$$P(195 \le Y \le 205) = P\left(\frac{195-200}{5} \le \frac{Y-200}{5} \le \frac{205-200}{5}\right)$$

$$= P(-1 \le Z \le 1), \text{ where } Z \sim N(0,1).$$

$$= P(Z \le 1) - P(Z \le -1)$$



$$= P(Z \le 1) - [1 - P(Z \le 1)]$$

$$= 2P(Z \le 1) - 1$$

$$= 2[0.8413] - 1 \quad \text{(from Standard normal table)}$$

$$= 1.6826 - 1 = \boxed{0.6826}$$

Q2 You can use the residuals calculated in HW-I, Q # 4.

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} = \frac{(-2.15)^2 + (3.85)^2 + (-5.15)^2 + (-1.15)^2 + (0.57)^2 + (2.57)^2 + (-2.43)^2 + (5.57)^2 + (3.30)^2 + (0.30)^2 + (1.30)^2 + (-3.70)^2 + (0.02)^2 + (-1.98)^2 + (3.02)^2 + (-3.98)^2}{16-2}$$

$$= 10.4587$$

$\therefore$ Estimate of $\sigma = \hat{\sigma} = \sqrt{10.4587} = 3.234$

Q3

a) The likelihood function,

$$L(\beta_1) = \prod_{i=1}^{6} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y_i - \beta_1 X_i)^2}{2\sigma^2}}$$

$$= \left(\frac{1}{2\pi(16)}\right)^3 \text{Exp}\left[\frac{-[(128-7\beta_1)^2 + (213-12\beta_1)^2 + (75-4\beta_1)^2 + (250-14\beta_1)^2 + (446-25)^2 + (540-30\beta_1)^2]}{2(16)}\right]$$

$$= \left(\frac{1}{32\pi}\right)^3 \text{Exp}\left[\frac{-[(128-7\beta_1)^2 + (213-12\beta_1)^2 + (75-4\beta_1)^2 + (250-14\beta_1)^2 + (446-25\beta_1)^2 + (540-30\beta_1)^2]}{32}\right]$$

b) When $\beta_1 = 17$,

$$(128-7(17))^2 + (213-12(17))^2 + (75-4(17))^2 + (250-14(17))^2 + (446-25(17))^2 + (540-30(17))^2$$

$$= 1696$$

$$\therefore L(17) = \left(\frac{1}{32\pi}\right)^3 \text{Exp}(-1696) = 9.4513 \times 10^{-30}$$

When $\beta_1 = 18$

$$(128-7(18))^2 + (213-12(18))^2 + (75-4(18))^2 + (250-14(17))^2 + (446-25(18))^2 + (540-30(18))^2$$

$$= 42$$

$$\therefore L(18) = \left(\frac{1}{32\pi}\right)^3 \text{Exp}(-42) = 2.6490 \times 10^{-7}$$

When $\beta_1 = 19$

$$(128-7(19))^2 + (213-12(19))^2 + (75-4(19))^2 + (250-4(19))^2 + (446-25(19))^2 + (540-30(19))^2$$

$$= 2248$$

$$\therefore L(19) = \left(\frac{1}{32\pi}\right)^3 \text{Exp}(-2248) = 3.0472 \times 10^{-37}$$

So the likelihood function is largest when $\beta_1 = 18$.

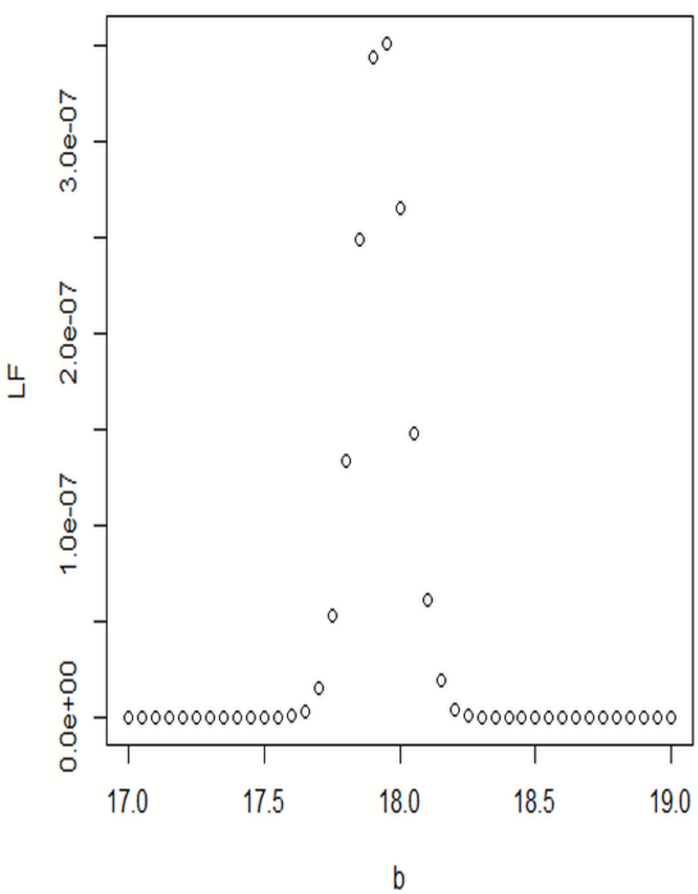
c) $\hat{\beta}_{1_{MLE}} = b_1 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$

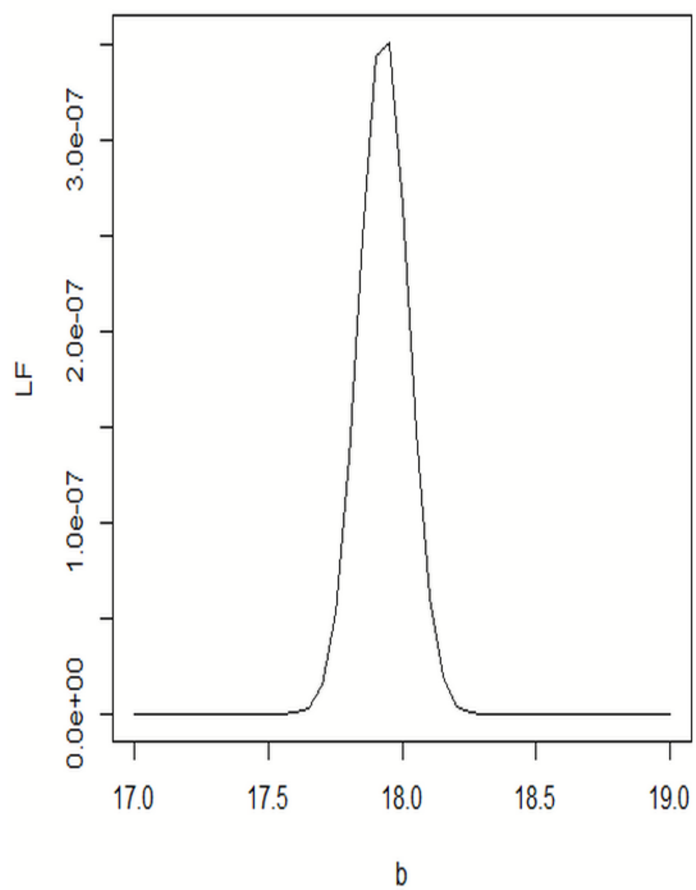
$$= \frac{(7)(128) + (12)(213) + (4)(75) + 14(250) + 25(446) + 30(540)}{7^2 + 12^2 + 4^2 + 14^2 + 25^2 + 30^2}$$

$$= 17.\cancel{5658}\ 9284$$

Yes, $\hat{\beta}_{1_{MLE}}$ is very close to the value in part a).

d)





Yes. Based on the graph likelihood function has it's maximum around 18.

Q4 $b_1 = \sum_{i=1}^{n} k_i Y_i$, where $k_i = \dfrac{X_i - \bar{X}}{\Sigma (X_i - \bar{X})^2}$

a)

$\Sigma k_i = 0$.

Proof:

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

$$= \frac{\Sigma X_i - n\bar{X}}{\Sigma (X_i - \bar{X})^2}$$

$$= \frac{n\cancel{\bar{X}} - n\cancel{\bar{X}}}{\Sigma (X_i - \bar{X})^2} \quad \left(\because \ \bar{X} = \frac{\Sigma X_i}{n} \Rightarrow \Sigma X_i = n\bar{X} \right)$$

$$= 0 \ \square.$$

b) $\sum_{i=1}^{n} k_i X_i = 1$

Proof:

$$\Sigma k_i X_i = \frac{\Sigma (X_i - \bar{X}) X_i}{\Sigma (X_i - \bar{X})^2}$$

$$= \frac{\Sigma X_i^2 - \bar{X} \Sigma X_i}{\Sigma (X_i^2 - 2\bar{X} X_i + \bar{X}^2)}$$

$$= \frac{\Sigma X_i^2 - n\bar{X}^2}{\Sigma X_i^2 - 2\bar{X} \Sigma X_i + n\bar{X}^2}$$

$$= \frac{\Sigma X_i^2 - n\bar{X}^2}{\Sigma X_i^2 - 2n\bar{X}^2 + n\bar{X}^2} = \frac{\Sigma X_i^2 - n\bar{X}^2}{\Sigma X_i^2 - n\bar{X}^2}$$

$$= 1 \ \square.$$

c) $$\sum_{i=1}^{n} K_i^2 = \Sigma \left[\frac{X_i - \bar{X}}{\Sigma (X_i - \bar{X})^2} \right]^2 = \frac{1}{\left[\Sigma (X_i - \bar{X})^2\right]^{\cancel{2}}} \cancel{\Sigma (X_i - \bar{X})^2}$$

$$= \frac{1}{\Sigma (X_i - \bar{X})^2}$$ a.

Q5 $b_0 = \bar{Y} - b_1 \bar{X}$

a) $E(b_0) = E(\bar{Y} - b_1 \bar{X}) = E(\bar{Y}) - E(b_1)\bar{X} \longrightarrow (1)$

But $E(\bar{Y}) = E\left[\frac{\Sigma Y_i}{n}\right] = \frac{1}{n} \Sigma E(Y_i)$

$$= \frac{1}{n} \Sigma E(\beta_0 + \beta_1 X_i + \varepsilon_i)$$

$$= \frac{1}{n} \Sigma (\beta_0 + \beta_1 X_i + 0)$$

$$= \frac{1}{n} (n\beta_0 + \beta_1 \Sigma X_i)$$

$$= \beta_0 + \beta_1 \bar{X}$$

and $E(b_1) = \beta_1$ (proved in the class note)

From (1),

$\therefore$ $E(b_0) = \beta_0 + \cancel{\beta_1 \bar{X}} - \cancel{\beta_1 \bar{X}}$

$= \beta_0$.

ie MLE of $\beta_0 = b_0$ is an unbiased estimator of β_0.

b) $Var(b_0) = var(\bar{Y} - b_1 \bar{X})$

$$= Var\left(\frac{\Sigma Y_i}{n} - \Sigma k_i Y_i \bar{X} \right)$$

$$= Var\left(\Sigma \left(\frac{1}{n} + \bar{X} k_i \right) Y_i \right)$$

$$= \Sigma \left(\frac{1}{n} + \bar{X} k_i \right)^2 \underbrace{var(Y_i)}_{\sigma^2} \quad (\because Y_i\text{s are independent}).$$

$$= \Sigma \left(\frac{1}{n^2} + \frac{2}{n} \bar{X} k_i + \bar{X}^2 k_i^2 \right) \sigma^2$$

$$= \left(\frac{n}{n^2} + \frac{2\bar{X}}{n} \underbrace{\Sigma k_i}_{=0} + \bar{X}^2 \Sigma k_i^2 \right) \sigma^2$$

$$= \left(\frac{1}{n} + \frac{\bar{X}^2}{\Sigma (X_i - \bar{X})^2} \right) \sigma^2.$$

c) Estimated variance of b_0 is obtained when σ^2 is replaced with MSE.

So $$\widehat{Var}(b_0) = \left(\frac{1}{n} + \frac{\bar{X}^2}{\Sigma (X_i - \bar{X})^2} \right) MSE,$$

where $$MSE = \frac{\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2}{n-2}.$$

Q6) $\hat{Y}_n = b_0 + b_1 X_n$

a) $E(\hat{Y}_n) = E(b_0 + b_1 X_n) = E(b_0) + X_n E(b_1) = \beta_0 + \beta_1 X_n \longrightarrow$ 1

$E(Y_n) = E(\beta_0 + \beta_1 X_n + \varepsilon_n) = \beta_0 + \beta_1 X_n + \underbrace{E(\varepsilon_n)}_{=0} = \beta_0 + \beta_1 X_n \rightarrow$ 2

By 1 and 2,

$$E(\hat{Y}_n) = E(Y_n).$$

b) $var(\hat{Y}_n) = var(\bar{Y} - b_1\bar{X} + b_1 X_n)$

$$= var\left(\frac{\sum Y_i}{n} + (X_n - \bar{X})\sum k_i Y_i\right)$$

$$= var\left(\sum \left(\frac{1}{n} + (X_n - \bar{X})k_i\right) Y_i\right)$$

$$= \sum \left(\frac{1}{n} + (X_n - \bar{X})k_i\right)^2 var(Y_i) \quad (\because Y_i\text{s are independent}).$$

$$= \sum \left(\frac{1}{n^2} + \frac{2}{n}(X_n - \bar{X})k_i + (X_n - \bar{X})^2 k_i^2\right)\sigma^2$$

$$= \left(\frac{1}{n} + \frac{2}{n}(X_n - \bar{X})\underbrace{\sum k_i}_{=0} + (X_n - \bar{X})^2 \sum k_i^2\right)\sigma^2$$

$$= \sigma^2\left[\frac{1}{n} + \frac{(X_n - \bar{X})^2}{\sum (X_i - \bar{X})^2}\right].$$

c) Estimated variance of $\hat{Y}_n$:

$$\widehat{var}(\hat{Y}_n) = MSE\left[\frac{1}{n} + \frac{(X_n - \bar{X})^2}{\sum (X_i - \bar{X})^2}\right].$$