# MA 542: Homework 2 Solutions

1. (a) We can't find the exact probability that $Y$ will fall in a certain interval because the distribution of $Y$ is unknown.

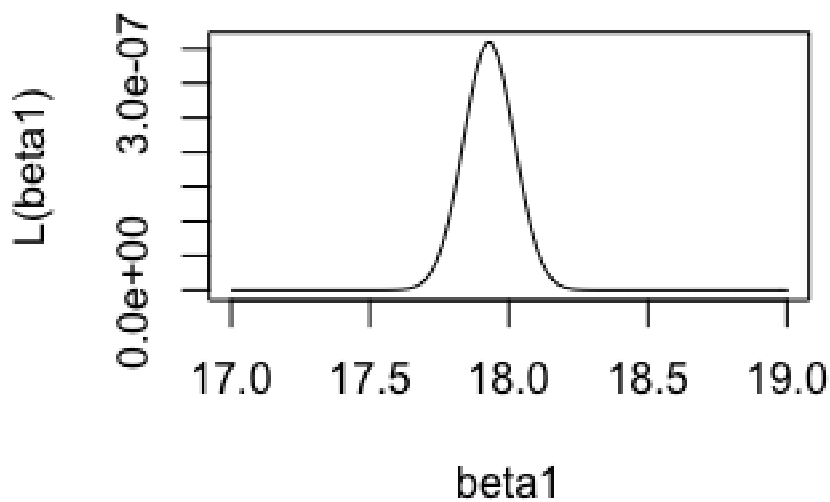
   (b) If the normal error model is applicable, $Y \sim N(\beta_0 + \beta_1 X, \sigma) = N(200, 5)$. Interval $(195, 105) = (200 - 1 \cdot \sigma, 200 + 1 \cdot \sigma)$, so we know the probability that Y will fall between 195 and 205 is 0.682.

2. (a) The likelihood is $L(\beta_1) = (32\pi)^{-3} exp(-\Lambda/32)$, where $\Lambda = (128 - 7\beta_1)^2 + (213 - 12\beta_1)^2 + (75 - 4\beta_1)^2 + (250 - 14\beta_1)^2 + (446 - 25\beta_1)^2 + (540 - 20\beta_1)^2$.

   (b) $L(17) = 9.45133 \times 10^{-30}, L(18) = 2.65 \times 10^{-7}, L(19) = 3.0473 \times 10^{-37}$.

   (c) $b_1 = 17.9285, L(b_1) = 3.61 \times 10^{-7}$. Yes, the results in (b) are consistent with this.

   (d) Yes, the function is maximized at 17.9285.




3. (a) $\sum_{i=1}^n k_i = \frac{\sum_{i=1}^n (X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n X_i - n\bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{0}{\sum_{i=1}^n (X_i - \bar{X})^2} = 0.$

   (b) $\sum_{i=1}^n k_i X_i = \frac{\sum_{i=1}^n (X_i^2 - X_i\bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n X_i^2 - \bar{X}\sum_{i=1}^n X_i}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{\sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2} = 1.$

   (c) $\sum_{i=1}^n k_i^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(\sum_{i=1}^n (X_i - \bar{X})^2)^2} = \frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2}.$

4. (a) $E(b_0) = E(\bar{Y} - b_1\bar{X}) = E(\bar{Y}) - E(b_1\bar{X}) = \frac{1}{n}\sum E(Y_i) - \bar{X}E(b_1) = \frac{1}{n}\sum(\beta_0 + \beta_1 X_i) - \beta_1\bar{X} = \beta_0.$

(b) First, prove $Cov(\bar{Y}, b_1) = 0$. $Cov(\bar{Y}, b_1) = Cov(\frac{\sum Y_i}{n}, \sum k_i Y_i) = \sum \frac{k_i}{n} Cov(Y_i, Y_i) = \frac{\epsilon^2}{n} \sum k_i = 0$. Therefore, $Var(b_0) = Var(\bar{Y} - b_1 \bar{X}) = Var(\bar{Y}) + Var(b_1 \bar{X}) = \frac{Var(Y_i)}{n} + \bar{X}^2 Var(b_1) = \sigma^2(\frac{1}{n} + \frac{\bar{X}}{\sum(X_i - \bar{X})^2})$.

(c) $\hat{Var}(b_0) = MSE(\frac{1}{n} + \frac{\bar{X}}{\sum(X_i - \bar{X})^2})$.

5. (a) $E(\hat{Y}_h) = E(b_0 + b_1 X_h) = \beta_0 + \beta_1 X_h = E(Y_h)$.

(b) We will use $Cov(\bar{Y}, b_1) = 0$ here. $Var(\hat{Y}_h) = Var(b_0 + b_1 X_h) = Var(b_0 + b_1 \bar{X} + b_1(X_h - \bar{X})) = Var(\bar{Y} + b_1(X_h - \bar{X})) = Var(\bar{Y}) + Var(b_1(X_h - \bar{X})) = \sigma^2/n + \frac{(X_h - \bar{X})^2 \sigma^2}{\sum(X_i - \bar{X})^2} = \sigma^2(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2})$.

(c) $\hat{Var}(\hat{Y}_h) = MSE(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2})$

6. (a) 95% confidence interval for $\beta_1$: (2.92,5.08). Interpretation:(1): We estimate the difference in mean response per unit increase in the number of times the carton was transferred from one aircraft to another in between 2.92 and 5.08. (2): In 95% of all experiments with these same X values, the confidence interval so computed will contain the true $\beta_1$.

(b) We test $H_0 : \beta_1 = 0$ versus $H_a$: $\beta_1 \neq 0$. We reject $H_0$ in favor of $H_a$ if the p-value is less than 0.05. Since the p-value is $2.75 \times 10^{-5} < 0.0001$, we reject $H_0$ in favor of $H_a$ and conclude there is a linear association between the number of ampules found to be broken upon arrival(Y) and the number of times the carton was transferred from one aircraft to another(X).

(c) 95% confidence interval for $\beta_0$:(8.67,11.73). Interpretation:(1): We estimate the mean number of ampules broken when no transfers of the shipment are made in between 8.67 and 11.73. (2): In 95% of all experiments with these same data, the confidence interval so computed will contain true $\beta_0$.

(d) We test $H_0 : \beta_0 \leq 9$ versus $H_a$: $\beta_0 > 9$. We reject $H_0$ in favor of $H_a$ if the p-value is less than 0.05. Since the p-value is $0.054 > 0.05$, we fail to reject $H_0$ and conclude that the mean number of broken ampules doesn't significantly exceed 9.0.

(e) 99% confidence interval for mean response when $X = 2, 4$ is (15.974,20.426), (21.223,31.177) respectively. Interpretation: (1): We estimate the mean response when X=2 (X=4) is between 15.974 and 20.426 (21.223 and 31.177). (2): In 99% of all experiments with these same data, the confidence interval so computed will contain true mean response.

7. (a) 90% confidence interval for $\beta_1$: (14.223,15.847). Interpretation:(1): We estimate the difference in mean response per unit increase in the number of copiers serviced in between 14.223 and 15.847. (2): In 90% of all experiments with these same X values, the confidence interval so computed will contain the true $\beta_1$.

(b) We test $H_0 : \beta_1 = 0$ versus $H_a$: $\beta_1 \neq 0$. We reject $H_0$ in favor of $H_a$ if the p-value is less than 0.1. Since the p-value $< 2 \times 10^{-16} < 0.1$, we reject $H_0$ in favor of $H_a$ and conclude there is a linear association between X and Y.

(c) Yes, the results are consistent. The confidence interval computed in part(a) doesn't include 0, which means $\beta_1$ is significant.

(d) We test $H_0 : \beta_1 \leq 14$ versus $H_a$: $\beta_1 > 14$. We reject $H_0$ in favor of $H_a$ if the p-value is less than 0.05. Since the p-value is $0.0189 < 0.05$, we reject $H_0$ and conclude that this standard is not being satisfied.

(e) No, $b_0$ is a negative number here, which does not give any useful information about the "start-up" time.