Solutions to Tutorial 2

1. (a)

$$\begin{array}{ccccc} \hat{Y} & = & 10.2 & + & 4.00X \\ (SE) & & (0.6633) & & (0.4690) \end{array}$$

$$MSE = 2.199289, \quad R^2 = 0.9009, \quad F = 72.73$$

see code **(R code)** for the plot.

Yes, the linear regression function fits the data well.

(b)

$$\hat{Y} = 10.2 + 4.00 * 1 = 14.2$$

(c)

$$b_1 \pm t(0.975, 8) * s(b_1) = 4 \pm 2.306 * 0.469 = [2.918486, 5.081514]$$

The interval does not include 0, indicating that we have (big) confidence that β_1 is different from 0, i.e. β_1 is not zero, and thus the linear association is significant.

(d) We need to test

$$H_0 : \beta_1 = 0 \qquad v.s. H_1 : \beta_1 \neq 0$$

Note that

$$|t| = |\frac{b_1 - 0}{s(b_1)}| = 8.528 > t(1 - \alpha/2, n - 2) = 2.306$$

We reject H_0. In other words, there is significant linear association.

(e)

$$b_0 \pm t(0.975, 8) * s(b_0) = 10.2 \pm 2.306 * 0.6633 = [8.67043, 11.72957]$$

The interval does not include 0, indicating we have (big) confidence that β_0 is different from 0. Even if no shipment, there are still broken ampules (due to the other reasons)

(f) We need to test

$$H_0 : \beta_1 = 2 \qquad v.s. \qquad H_1 : \beta_1 \neq 2$$

Note that

$$|t| = |\frac{b_1 - 2}{s(b_1)}| = 4.264392 > t(1 - \alpha/2, n - 2) = 2.9$$

We reject H_0. In other words, the past experience is not applicable to the data

2. The absolute value of the coefficient -1.4 looks different from 0, but its p-value is very big, indicating that the coefficient is not significantly different from 0. Thus, the correlation between X and Y is not strong and the conclusion is not statistically solid.

3. (a) see **(R code)**. the 99% CI (confidence interval) is

$$b_1 \pm t(1-0.01/2, 118) * s(b_1) = 0.03883 \pm 2.62 * 0.01277 = [0.0053726, 0.0722874]$$

It does not include 0 (or we have 99% confidence to believe it is not 0). If it is not 0, a student's PCA can be predicted by his ACT. In other word, ACT can be an indicator for universities to evaluate a students and to decide whether a student should be admitted.

(b)

$$H_0 : \beta_1 = 0 \quad H_1 : \beta_1 \neq 0$$

From the calculation

$$|t| = |0.03883/0.01277| = |3.040| > t(0.995, 118) = 2.62$$

Thus, we reject H_0, i.e. there is significant linear association between ACT and GPA when $\alpha = 0.01$

(c) Because

$$p - value = 0.00292 < 0.01$$

we also reject the H_0 above

4. (a) The estimator is to minimize

$$Q = \sum_{i=1}^{n} (Y_i - b_1 X_i)^2.$$

By calculus, we have the estimator (minimizer) is

$$b_1 = \sum_{i=1}^{n} X_i Y_i / \sum_{i=1}^{n} X_i^2$$

(b) NO

(c) It follows from the model

$$b_1 = \frac{\sum_{i=1}^{n} X_i(\beta_1 X_i + \varepsilon_i)}{\sum_{i=1}^{n} X_i^2} = \beta_1 + \frac{\sum_{i=1}^{n} X_i \varepsilon_i}{\sum_{i=1}^{n} X_i^2}$$

It is easy to check

$$Eb_1 = \beta_1, \quad Var(b_1) = \frac{\sigma^2}{\sum_{i=1}^n X_i^2}$$

Thus

$$b_1 \sim N(\beta_1, \frac{\sigma^2}{\sum_{i=1}^n X_i^2})$$

To estimate σ^2, we consider the fitted residuals

$$e_i = Y_i - \hat{Y}_i$$

We estimate σ^2 by

$$MSE = \frac{1}{n-1}\sum_{i=1}^n e_i^2$$

(Why n-1?) then the standard error is

$$s(b_1) = \sqrt{MSE/\sum_{i=1}^n X_i^2}$$

(d)

$$H_0 : \beta_1 = 0 \qquad H_1 : \beta_1 \neq 0$$

parallel to the simple linear regresison, we have

$$\frac{\mathbf{b}_1 - \beta_1}{s(b_1)} \sim t(n-1)$$

Thus, under H_0

$$t = \frac{\mathbf{b}_1}{s(b_1)} \sim t(n-1)$$

For a t-value t^*,

If $|t^*| < t(1-\alpha/2, n-1)$, we accept H_0

If $|t^*| \geq t(1-\alpha/2, n-1)$, we reject H_0

5. We have proved from the lecture notes (page 3, part 2 chapter 1) that

$$b_1 = \beta_1 + \frac{\sum_{i=1}^n (X_i - \bar{X})\varepsilon_i}{\sum_{i=1}^n (X_i - \bar{X})^2} = \beta_1 + \sum_{i=1}^n \frac{(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}\varepsilon_i$$

Let $A_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$. By the conditions, A_i and ε_i are independent. Thus

$$E\{A_i\varepsilon_i\} = E\{A_i\}E\{\varepsilon_i\} = 0$$

It follows immediately

$$Eb_1 = \beta_1 + \sum_{i=1}^{n} E\{A_i \varepsilon_i\} = \beta_1$$

Note that

$$\begin{aligned} b_0 &= \bar{Y} - b_1 \bar{X} = \beta_0 + (\beta_1 - b_1)\bar{X} + \bar{\varepsilon} \\ &= \beta_0 - \sum_{i=1}^{n} A_i \varepsilon_i \bar{X} + \bar{\varepsilon} \end{aligned}$$

It is easy to see that $A_i \bar{X}$ is independent of ε_i. Thus

$$\begin{aligned} Eb_0 &= \beta_0 - \sum_{i=1}^{n} E\{A_i \bar{X}\} E\varepsilon_i + E\bar{\varepsilon} \\ &= \beta_0 \end{aligned}$$