Tutorial Notes 3 of MATH3424

1 Summary of course material

1.1 Covariance, Correlation Coefficient

- Covariance of X and Y is defined by

$$\text{Cov}(X,Y) = \frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{n-1} \tag{1}$$

- Correlation and coefficient between X and Y is given by

$$\text{Cor}(X,Y) = \frac{\text{Cov}(Y,X)}{s_y s_x} \tag{2}$$

1.2 Simple Linear Regression Model

- A simple linear model:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Core assumption:

$$\epsilon_1, \cdots, \epsilon_n \overset{i.i.d.}{\sim} N(0,\sigma^2)$$

- Parameter estimation (least square estimates/unbiased/standard error/distribution):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}$$

- Fitted values and residuals

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \quad e_i = y_i - \hat{y}_i$$

- Measuring the quality of fit:

$$\text{SST} = \sum(y_i-\bar{y})^2 \quad \text{SSR} = \sum(y_i-\bar{y})^2 \quad \text{SSE} = \sum(y_i-\bar{y})^2$$

$$R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}$$

- Hypothesis test and confidence intervals:

$$t = \frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)} \qquad \hat{\beta}_1 \pm t_{(n-2,\alpha/2)} \times s.e.(\hat{\beta}_1)$$

- No intercept model: difference

2 Questions

1. Consider a simple linear regression model: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ for $i = 1, \cdots, n$.

   - You are given 5 pairs of (x_i, y_i) where y_4 is missing and the fitted line passes through the point $(3, 1.65)$. Find c and then determine $\sum_{i=1}^{5}(y_i - \bar{y})^2$.

x_i	1	2	3	4	5
y_i	0.25	1.75	1.5	c	2.5

   - Given the following statistics from 25 pairs of (x_i, y_i):

$$\bar{x} = 0, \quad \hat{\sigma}^2 = 100, \quad \hat{\beta}_0 = 3$$

   determine the length of a 98% confidence interval for β_0.

   - Given the following statistics from 10 pairs of (x_i, y_i):

$$\sum_{i=1}^{10}(x_i - \bar{x})^2 = 400, \quad \sum_{i=1}^{10}(y_i - \bar{y})^2 = 425, \quad \sum_{i=1}^{10}(\hat{y}_i - \bar{y})^2 = 225$$

   Calculate the test statistic for testing the hypothesis $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ by t test. Write down your conclusion clearly. Set the significance level at $\alpha = 0.05$

2. Consider a linear model, for $i = 1, \cdots, 3$

$$y_i = \beta_0 + i\beta_1 + \epsilon_i$$

where ϵ_i follows independent normal distribution with mean 0 and variance $i\sigma^2$.

- Find the least squares estimates of β_0 and β_1 in terms of y_i
- Find the $Var(\hat{\beta}_0)$ and $Var(\hat{\beta}_1)$.

3. Consider the simple linear regression model: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ for $i = 1, \cdots, n$. Show that $Cov(\bar{y}, \hat{\beta}_1) = 0$ and $Cov(e_i, \hat{\beta}_1) = 0$.