The Hong Kong University of Science & Technology

MATH3424 - Regression Analysis

Quiz 1

Answer ALL Questions **Date: 9 October 2020**

Full marks: 25 + Bonus: 2 **Time Allowed: 1 hour**

---

1. Consider a simple linear regression model: $y_i = \beta_0 + \beta_1 x_i + e_i$, where e_i is normally distribution with mean 0 and variance σ^2.

   (a) **(2 marks)** You are given 5 pairs of (x_i, y_i) where y_4 is missing

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

   and the fitted line passes through the point (3, 1.65). Find c and then determine $\sum_{i=1}^{5}(y_i - \bar{y})^2$.

   (b) **(3 marks)** 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.

   (c) **(4 marks)** 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. Using the following summary statistics

$n = 20$, $\sum_{i=1}^{20} x_{i1} = 124$, $\sum_{i=1}^{20} x_{i2} = 114$, $\sum_{i=1}^{20} y_i = 138$,

$\sum_{i=1}^{20} x_{i1}^2 = 1012$, $\sum_{i=1}^{20} x_{i1}x_{i2} = 875$, $\sum_{i=1}^{20} x_{i2}^2 = 834$, $\sum_{i=1}^{20} x_{i1}y_i = 962$,

$\sum_{i=1}^{20} x_{i2}y_i = 1002$, $\sum_{i=1}^{20} y_i^2 = 1324$,

$S_{x_1x_1} = 243.2$, $S_{x_1x_2} = 168.2$, $S_{x_2x_2} = 184.2$, $S_{x_1y} = 106.4$,

$S_{x_2y} = 215.4$, $S_{yy} = 371.8$.

and

$$\begin{pmatrix} 243.2 & 168.2 \\ 168.2 & 184.2 \end{pmatrix}^{-1} = \begin{pmatrix} 0.011159 & -0.010190 \\ -0.010190 & 0.014734 \end{pmatrix},$$

to fit a model of y on x_1 and x_2, i.e., do the following regression model,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i, \quad e_i \sim N(0, \sigma^2).$$

From previous calculation, it is known that $\hat{\beta}_0 = 1.237088$, $\hat{\beta}_1 = -1.007608$ and $\hat{\beta}_2 = 2.089488$.

(a) **(3 marks)** Find Residual Sum of Squares and the unbiased estimate of the unknown parameter σ^2. No need to show that it is unbiased.

(b) **(2 marks)** $H_0 : \beta_2 = 2$ against the alternative hypothesis that $H_1 : \beta_2 \neq 2$ at the significance level of $\alpha = 0.05$ by t-test. Write down the test statistic, the critical value and your conclusion clearly.

Assume that $\beta_0 = 2$.

(c) **(2 marks)** Write down $\underset{\sim}{X}^T\underset{\sim}{X}$, $(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}$ and $\underset{\sim}{X}^T\underset{\sim}{Y}$ in terms of values of summary statistics.

(d) **(2 marks)** Find the least squares estimates of the unknown parameters β_1 and β_2. Then, write down the fitted line.

(e) **(3 marks)** Find the Residual Sum of Squares and the unbiased estimate of the unknown parameter σ^2. No need to show that it is unbiased.

3. Consider a linear model

$$y_i = \beta_0 + i\beta_1 + e_i$$
$$\text{for } i = 1, 2, 3$$

where e_i follows independent normal distribution with mean 0 and variance $\boldsymbol{i} \times \boldsymbol{\sigma^2}$.

(a) **(2 marks)** Find the least squares estimates of β_0 and β_1 in terms of y_i.

(b) **(2 marks)** Find the $Var(\hat{\beta}_0)$ and $Var(\hat{\beta}_1)$.

(c) **(Bonus: 2 marks)** Find the expectation vector and covariance matrix of residual $\hat{\boldsymbol{e}}$.

********** END **********