The Hong Kong University of Science & Technology

MATH3424 - Regression Analysis

Quiz 2

Answer ALL Questions **Date: 13 November 2020**

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

---

1. Short questions

   (a) **(3 marks)** A linear trend model with two seasonal indicator components is used to model 8 successive data points representing quarterly sales for the years 2018 and 2019, $t = 1, 2, \ldots, 8$ (where t refers to quarters). The first seasonal indicator, IND_{i1}, corresponds to the first quarter of each year, so that $\text{IND}_{i1} = 1$ if time t is the first quarter of a year and is 0 otherwise. The second seasonal indicator, IND_{i2}, corresponds in a similar way to the third quarter of each year. The model used is

   $$y_i = \beta_0 + \beta_1 t_i + \delta_1 \text{IND}_{i1} + \delta_2 \text{IND}_{i2} + e_i,$$

   Write down the design matrix $\underset{\sim}{X}$ for this model and then the design matrix for the corresponding centered model, $\underset{\sim}{X}_c$. Hence, find $\underset{\sim}{X}_c^T \underset{\sim}{X}_c$.

   (b) **(2 marks)** You are given 20 pairs of values (x_i, y_i) which will be represented by the following model: $y_i = \beta_0 + \beta_1 x_{i1} + e_i$ where e_i is normally distributed with mean 0 and variance σ^2. You have found the following statistics

   $$\sum_{i=1}^{20} x_i = 200; \quad \sum_{i=1}^{20} y_i = 400; \quad \sum_{i=1}^{20} x_i^2 = 2400; \quad \sum_{i=1}^{20} y_i^2 = 8600; \quad \hat{\sigma}^2 = 100$$

   Determine the estimated standard deviation in the prediction of y_k when $x_k = 26$.

   (c) **(Bonus: 2 marks)** The model of

   $$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + e_i,$$

   where e_i is normally distributed with mean 0 and variance σ^2, was fitted by 30 observations and the following statistics were obtained

   $$\sum_{i=1}^{30} (y_i - \bar{y})^2 = 900; \quad \sum_{i=1}^{30} (\hat{y}_i - \bar{y})^2 = 300.$$

   You then decided to modify your model to include a new variable, x_5, with the following properties:

   $$\sum_{i=1}^{30} x_{i5} = 0; \quad \sum_{i=1}^{30} x_{i5} x_{ik} = 0, \quad k \neq 5; \quad \sum_{i=1}^{30} x_{i5}^2 = 100; \quad \sum_{i=1}^{30} x_{i5} y_i = 150$$

   Calculate the absolute change in R^2 resulting from the inclusion of x_5 in the model.

2. Use the following summary statistics

Overall

$n = 28,\qquad \sum_{i=1}^{28} x_i = 72,\qquad \sum_{i=1}^{28} y_i = 1950,$

$\sum_{i=1}^{28} x_i^2 = 218,\qquad \sum_{i=1}^{28} x_i y_i = 5240,\qquad \sum_{i=1}^{28} y_i^2 = 143100$

Repeated measures of y on different values of x

$\boldsymbol{x = 1}$
$n = 6,\quad \sum y_i = 340,\quad \sum y_i^2 = 20600;$

$\boldsymbol{x = 2}$
$n = 7,\quad \sum y_i = 510,\quad \sum y_i^2 = 38100;$

$\boldsymbol{x = 3}$
$n = 8,\quad \sum y_i = 520,\quad \sum y_i^2 = 35600;$

$\boldsymbol{x = 4}$
$n = 7,\quad \sum y_i = 580,\quad \sum y_i^2 = 48800.$

to answer the following questions.

(a) 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.

i. **(2 marks)** Find the unbiased estimate of the unknown parameter σ^2. No need to show that it is unbiased.

ii. **(4 marks)** Conduct a lack of fit test at $\alpha = 0.05$. Write down your test statistic, critical value and your conclusions clearly.

iii. **(1 mark)** Based on the conclusion from the lack of fit test, find the unbiased estimate of σ^2. Explain your choice.

(b) Treat x as a categorical variable with 4 levels. Then, the model becomes

$$y_{ij} = \mu_i + e_{ij} \qquad \text{for } i = 1, 2, 3, 4;\ j = 1, \ldots, n_i$$

i. **(1 mark)** Find the unbiased estimate of the unknown parameter σ^2. No need to show that it is unbiased.

ii. **(4 marks)** Test all population means of y are equal at $\alpha = 0.05$. Write down null hypothesis, test statistic, critical value and your conclusion clearly.

iii. **(4 marks)** Test $H_0 : \mu_1 = \mu_2 = \mu_3$ at $\alpha = 0.05$. Write down null hypothesis, test statistic, critical value and your conclusion clearly. Hint: Write $H_0 : \underset{\sim}{C}\underset{\sim}{\beta} = \underset{\sim}{d}$.

iv. **(4 marks)** Test whether the population mean of y at the 4^{th} level is equal to the average of population means of y at other levels by F test at $\alpha = 0.05$. Write down null hypothesis, contrast sum of squares, test statistic, critical value and your conclusion clearly.