MATH3424 Regression Analysis

Assignment 1

1. Using the following summary statistics

$n = 20,$ $\quad \sum_{i=1}^{20} x_{i1} = 114,$ $\quad \sum_{i=1}^{20} x_{i2} = -136,$ $\quad \sum_{i=1}^{20} y_i = 222,$

$\sum_{i=1}^{20} x_{i1}^2 = 860,$ $\quad \sum_{i=1}^{20} x_{i1}x_{i2} = -1025,$ $\quad \sum_{i=1}^{20} x_{i2}^2 = 1228,$ $\quad \sum_{i=1}^{20} x_{i1}y_i = 1537,$

$\sum_{i=1}^{20} x_{i2}y_i = -1824,$ $\quad \sum_{i=1}^{20} y_i^2 = 2950,$

$S_{x_1x_1} = 210.2,$ $\quad S_{x_1x_2} = -249.8,$ $\quad S_{x_2x_2} = 303.2,$ $\quad S_{x_1y} = 271.6,$

$S_{x_2y} = -314.4,$ $\quad S_{yy} = 485.8.$

and

$$\begin{pmatrix} 210.2 & -249.8 \\ -249.8 & 303.2 \end{pmatrix}^{-1} = \begin{pmatrix} 0.227525 & 0.187453 \\ 0.187453 & 0.157737 \end{pmatrix},$$

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

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

(a) **Assume that $\boldsymbol{\beta_0 = 2}$.**

i. Find the least squares estimates of the unknown parameters β_1 and β_2. Then, write down the fitted line.

ii. Find the Residual Sum of Squares and the unbiased estimate of the unknown parameter σ^2. No need to show that it is unbiased.

(b) **Assume that $2\boldsymbol{\beta_1} = \boldsymbol{\beta_2}$.**

i. Find the least squares estimates of the unknown parameters β_0 and β_1. Then, write down the fitted line..

ii. Find the Residual Sum of Squares and the unbiased estimate of the unknown parameter σ^2. No need to show that it is unbiased.

iii. Test H_0: $\beta_1 = 2$ by t-test at significance level of $\alpha = 0.05$. Write down your test statistic, critical value and your conclusions clearly.

(c) **Assume that $\boldsymbol{\beta_0, \beta_1}$ and $\boldsymbol{\beta_2}$ are unknown.**

i. Find the least squares estimates of the unknown parameters β_0, β_1 and β_2. Then, write down the fitted line.

ii. Find Residual Sum of Squares and the unbiased estimate of the unknown parameter σ^2. No need to show that it is unbiased.

iii. Test the assumption in Part II that $H_0 : 2\beta_1 = \beta_2$ against the alternative hypothesis that $H_1 : 2\beta_1 \neq \beta_2$ at the significant level of $\alpha = 0.05$ by t-test. Write down the test statistic, the critical value and your conclusion clearly.

2. Consider a situation in which the regression data set is divided into two parts as follows.

x	y
x_1	y_1
x_2	y_2
$\vdots$	$\vdots$
x_{n_1}	y_{n_1}
x_{n_1+1}	y_{n_1+1}
$\vdots$	$\vdots$
$x_{n_1+n_2}$	$y_{n_1+n_2}$

The model is given by

$$\begin{aligned} y_i &= \beta_0^{(1)} + \beta_1 x_i + e_i \qquad \text{for } i = 1, \ldots, n_1 \\ &= \beta_0^{(2)} + \beta_1 x_i + e_i \qquad \text{for } i = n_1 + 1, \ldots, n_1 + n_2 \end{aligned}$$

In other words there are two regression lines with common slope. Using the centered model,

$$\begin{aligned} y_i &= \beta_0^{(1)^*} + \beta_1 (x_i - \bar{x}_1) + e_i \qquad \text{for } i = 1, \ldots, n_1 \\ &= \beta_0^{(2)^*} + \beta_1 (x_i - \bar{x}_2) + e_i \qquad \text{for } i = n_1 + 1, \ldots, n_1 + n_2 \end{aligned}$$

where $\bar{x}_1 = \sum_{i=1}^{n_1} \frac{x_i}{n_1}$ and $\bar{x}_2 = \sum_{i=n_1+1}^{n_1+n_2} \frac{x_i}{n_2}$.

Show that the least squares estimate of β_1 is given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n_1} (x_i - \bar{x}_1) y_i + \sum_{i=n_1+1}^{n_1+n_2} (x_i - \bar{x}_2) y_i}{\sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2 + \sum_{i=n_1+1}^{n_1+n_2} (x_i - \bar{x}_2)^2}.$$

Hint: Write the model in matrix form.

3. Consider the simple linear regression model, $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ for $i = 1, 2, \ldots, n$, show that the least squares slope is given by

$$\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} d_i \varepsilon_i \quad \text{where} \quad d_i = \frac{x_i - \bar{x}}{S_{xx}}$$

Also show that

$$\hat{\beta}_0 = \beta_0 + \bar{\varepsilon} - \bar{x} \sum_{i=1}^{n} d_i \varepsilon_i$$

4. For the model of $y_i = \beta_0 + \beta_1 x_{i1} + e_i$ for $i = 1, \ldots, n$, where $e_i \underset{\sim}{iid} N(0, \sigma^2)$, find $E(y_i - \hat{y}_i)$ and $Var(y_i - \hat{y}_i)$.

   Hint: Define

$$\begin{aligned} \hat{e}_i &= y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} \\ &= y_i - \sum_{j=1}^{n} (c_j + d_j x_{i1}) y_j \\ &= \sum_{j=1}^{n} \left(\delta_{ij} - (c_j + d_j x_{i1})\right) y_j \end{aligned}$$

   where $c_j = \frac{1}{n} - \frac{(x_{j1} - \bar{x}_1)\bar{x}_1}{S_{x_1 x_1}}$, $d_j = \frac{x_{j1} - \bar{x}_1}{S_{x_1 x_1}}$ and $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$

5. Fit the model of y on x_1 and x_2, i.e.,

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

   and get its fitted line by the method of least squares

$$\boldsymbol{\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}}$$

   Now fit another model of x_2 on x_1 and $\hat{y}_i$, i.e.,

$$\boldsymbol{x_{i2} = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 \hat{y}_i + \epsilon_i}$$

   Write down $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in terms of $S_{x_1x_1}$, $S_{x_1x_2}$, $S_{x_2x_2}$, S_{x_1y}, S_{x_2y} and S_{yy}. Hence or otherwise, prove that $\hat{\epsilon}_i = 0$ for all i.