MATH3424 Regression Analysis

Assignment 3

1. Using the following data set

x_1	x_2	y	x_1	x_2	y
1	2	12	5	9	19
1	3	13	6	7	25
2	1	6	6	8	16
2	1	3	8	9	30
2	1	5	9	9	30
3	2	13	10	11	33
3	2	8	10	11	28
4	5	14	10	11	40
5	6	16	11	14	40
5	6	14	12	15	45

with summary statistics:

$$n = 20, \qquad \sum_{i=1}^{20} x_{i1} = 115, \qquad \sum_{i=1}^{20} x_{i2} = 133, \qquad \sum_{i=1}^{20} y_i = 410,$$

$$\sum_{i=1}^{20} x_{i1}^2 = 905, \qquad \sum_{i=1}^{20} x_{i1}x_{i2} = 1055, \qquad \sum_{i=1}^{20} x_{i2}^2 = 1261, \qquad \sum_{i=1}^{20} x_{i1}y_i = 3163,$$

$$\sum_{i=1}^{20} x_{i2}y_i = 3729, \qquad \sum_{i=1}^{20} y_i^2 = 11404,$$

$$S_{x_1x_1} = 243.75, \qquad S_{x_1x_2} = 290.25, \qquad S_{x_2x_2} = 376.55, \qquad S_{x_1y} = 805.5,$$

$$S_{x_2y} = 1002.5, \qquad S_{yy} = 2999.$$

and

$$\begin{pmatrix} 243.75 & 290.25 \\ 290.25 & 376.55 \end{pmatrix}^{-1} = \begin{pmatrix} 0.049947 & -0.038450 \\ -0.038450 & 0.032332 \end{pmatrix},$$

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 the previous calculation, it is known that $\hat{\beta}_0 = 1.774373$, $\hat{\beta}_1 = 1.636212$ and $\hat{\beta}_2 = 1.401114$.

(a) Re-write the model to a centered model in matrix form. Define $\underset{\sim}{Y}$ and $\underset{\sim}{X}$ in terms of data. 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.

(b) Find Residual Sum of Squares and Pure Error Sum of Squares.

(c) Fill the following table.

Source	Sum of Squares	D.F.	Mean Squares	F value
Regression				
Residual				–
Lack of fit				
Pure error				–
Total			–	–

    i. Conduct a lack of fit test at $\alpha = 0.05$. Write down your test statistic, critical value and your conclusions clearly.

    ii. Find the unbiased estimate of σ^2 based on the conclusion from the lack of fit test above.

    iii. Test H_0: $\beta_1 = \beta_2 = 0$ at significance level of $\alpha = 0.05$. Write down your test statistic, critical value and your conclusions clearly.

    iv. Calculate the coefficient of determination.

(d) Construct the 95% predication interval for individual value of y at $x_1 = 5$ and $x_2 = 5$.

(e) Test the null hypothesis that $H_0 : \beta_1 = \beta_2$ against the alternative hypothesis that $H_1 : \beta_1 \neq \beta_2$ by F test at the significance level of $\alpha = 0.05$.

    i. Find the Residual Sum of Squares of the model under the null hypothesis.

    ii. Construct the test statistic in terms of "Increase in Regression Sum of Squares". Write down the test statistic, the critical value and your conclusion clearly.

(f) Test $H_0 : \beta_0 = 1$ against the alternative hypothesis $H_1 : \beta_0 \neq 1$ by <u>t-test</u> at the significance level of $\alpha = 0.05$. Write down the test statistic, the critical value and your conclusion clearly.

(g) Test $H_0 : \beta_1 + \beta_2 = 2$ against the alternative hypothesis that $H_1 : \beta_1 + \beta_2 \neq 2$ by F test for the General Linear Hypothesis at the significance level of $\alpha = 0.05$. Write down the test statistic, the critical value and your conclusion clearly.

Hint: Write down the null hypothesis as $H_0 : C\beta = d$.

(h) **Assume that $\beta_0 = 1$.** Estimate the unbiased estimate of the unknown parameter σ^2. No need to show that it is unbiased.

(i) **Assume that $\beta_0 = 1$ and $\beta_1 = \beta_2$.** Find the least squares estimate of the unknown parameter in the model. Then, write down the fitted line. Estimate the unbiased estimate of the unknown parameter σ^2. No need to show that it is unbiased.

2. Three catalysts are used in a chemical process. The following are yield data from the process:

	Catalyst		
	1	2	3
	79.5	81.5	78.1
	82.0	82.3	80.2
	80.6	81.4	81.5
	84.9	79.5	83.0
	81.0	83.0	82.1
Mean	81.6000	81.5400	80.9800
Variance	4.2050	1.7230	3.6270

Given that the overall sample variance is 2.8135.

(a) Write down a regression model (a model in terms of dummy variables) for the analysis of the above data set. Define the variables clearly.

(b) Write down a one-way classification model (a model in terms of population means of catalysts) for the analysis of the above data set. Define the variables clearly.

(c) Estimate the unknown parameters in part 2.

(d) Hence or otherwise, estimate the unknown parameters in part 1.

(e) Test all population means of catalysts are equal at $\alpha = 0.05$. Write down the test statistic, the critical value and your conclusion clearly.