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:

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

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

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

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

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

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 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.

$$\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{e} \text{ where } \underset{\sim}{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix}, \underset{\sim}{X} = \begin{pmatrix} 1 & x'_{11} & x'_{12} \\ 1 & x'_{21} & x'_{22} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x'_{n1} & x'_{n2} \end{pmatrix}, x'_{ij} = x_{ij} - \bar{x}_j, j = 1, 2 \text{ and}$$

$$\underset{\sim}{e}= \begin{pmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{pmatrix}$$

$$\underset{\sim}{X}^T\underset{\sim}{X}= \begin{pmatrix} n & 0 & 0 \\ 0 & S_{x_1x_1} & S_{x_1x_2} \\ 0 & S_{x_1x_2} & S_{x_2x_2} \end{pmatrix} = \begin{pmatrix} 20 & 0 & 0 \\ 0 & 243.75 & 290.25 \\ 0 & 290.25 & 376.55 \end{pmatrix}$$

$$(\underset{\sim}{X}^T\underset{\sim}{X})^{-1} = \begin{pmatrix} 0.05 & 0 & 0 \\ 0 & 0.049947 & -0.038450 \\ 0 & -0.038450 & 0.032332 \end{pmatrix}$$

$$\underset{\sim}{X}^T\underset{\sim}{Y}= \begin{pmatrix} \sum_{i=1}^n y_i \\ S_{x_1y} \\ S_{x_2y} \end{pmatrix} = \begin{pmatrix} 410 \\ 805.5 \\ 1002.5 \end{pmatrix}$$

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

ResS.S $= S_{yy} - \hat{\beta}_1 S_{x_1y} - \hat{\beta}_2 Sx_2y = 276.414449$

Pure Error S.S $= 91.833333$

(c) Fill the following table.

| Source | Sum of Squares | D.F. | Mean Squares | F value |
|---|---|---|---|---|
| Regression | 2722.585551 | 2 | 1361.292776 | 83.722024 |
| Residual | 276.414449 | 17 | 16.259673 | – |
| Lack of fit | 184.581116 | 11 | 16.780101 | 1.096341 |
| Pure error | 91.833333 | 6 | 15.305556 | – |
| Total | 2999 | 19 | – | – |

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

$F_{obs} = 1.096341$

Critical value $= F_{0.05,11,6} > F_{0.05,12,6} = 4.00 > 1.096341$

Hence, we do not reject $H_0$.

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

Since we do not reject $H_0$: There is no lack of fit, $\hat{\sigma}^2 = \frac{276.414449}{17} = 16.259673$

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.

$F_{obs} = 83.722024$

Critical value $= F_{0.05,2,17} = 3.59 < 83.722024$

Hence, we reject $H_0$.

iv. Calculate the coefficient of determination.

$R^2 = \frac{2722.585551}{2999} = 0.907831$

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

We construct the 95% prediction interval by making use of the centered model.

$\hat{y}|_{x_1=x_2=5} = 1.774374 + 5(1.636212 + 1.401114) = 16.961003$

$x_1 = 5, x_2 = 5 \Leftrightarrow x_1' = 5 - \bar{x_1} = -0.75, x_2' = 5 - \bar{x_2} = -1.65$

$$\underset{\sim}{x}_0^T(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{x}_0 = \begin{pmatrix} 1 & -0.75 & -1.65 \end{pmatrix} \begin{pmatrix} 0.05 & 0 & 0 \\ 0 & 0.049947 & -0.038450 \\ 0 & -0.038450 & 0.032332 \end{pmatrix} \begin{pmatrix} 1 \\ -0.75 \\ -1.65 \end{pmatrix}$$
$= 0.070955.$

s.e $= \sqrt{(16.259673)(1 + 0.070955)} = 4.172934$

Hence the required 95% prediction interval
$= [16.961003 - 2.110(4.172934), 16.961003 + 2.110(4.172934)] = [8.156112, 25.765894]$

(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.

Under $H_0$, the model becomes $y_i = \beta_0 + \beta_1(x_{i1} + x_{i2}) + e_i = \beta_0 + \beta_1 x_i' + e_i$

$\hat{\beta}_1 = \frac{S_{x'y}}{S_{x'x'}} = \frac{S_{x_1y} + S_{x_2y}}{S_{x_1x_1} + 2S_{x_1x_2} + S_{x_2x_2}} = 1.505663$

Hence, ResS.S $= S_{yy} - \hat{\beta}_1 S_{x'y} = S_{yy} - \hat{\beta}_1(S_{x_1y} + S_{x_2y}) = 276.761296$

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.

$F_{obs} = \frac{(276.761296-276.414449)/(18-17)}{276.414449/17} = 0.021332 < F_{0.05,1,17} = 4.45$

Hence, we do not reject $H_0$.

(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.

Since $\widehat{Cov}(\hat{\beta}'_0, \hat{\beta}_1) = \widehat{Cov}(\hat{\beta}'_0, \hat{\beta}_2) = 0$

$\widehat{Var}(\hat{\beta}_0) = \widehat{Var}(\hat{\beta}'_0 - \hat{\beta}_1\bar{x}_1 - \hat{\beta}_2\bar{x}_2) = \widehat{Var}(\hat{\beta}'_0) + \bar{x}_1^2\widehat{Var}(\hat{\beta}_1) + \bar{x}_2^2\widehat{Var}(\hat{\beta}_2) + 2\bar{x}_1\bar{x}_2\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)$

s.e of $\hat{\beta}_0$
$= \sqrt{(16.259673)(0.05 + 5.75^2(0.049947) + 6.65^2(0.032332) + 2(5.75)(6.65)(-0.038450))}$
$= 1.760936$

$t_{obs} = |\frac{1.774373-1}{1.760936}| = 0.439751 < t_{0.025,17} = 2.110$

Hence, we do not reject $H_0$.

(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 $H_0 : \underset{\sim}{C}\underset{\sim}{\beta} = \underset{\sim}{d}$.

We make use of the centered model to perform the hypothesis testing:

$$H_0 : \underset{\sim}{C}\underset{\sim}{\beta} = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \beta'_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 2 \end{pmatrix} = \underset{\sim}{d}$$

$\underset{\sim}{C}\hat{\underset{\sim}{\beta}} - \underset{\sim}{d} = \hat{\beta}_1 + \hat{\beta}_2 - 2 = 1.037326$

$\underset{\sim}{C}(\underset{\sim}{X}^T\underset{\sim}{X})\underset{\sim}{C}^T = 0.005379 \Rightarrow (\underset{\sim}{C}(\underset{\sim}{X}^T\underset{\sim}{X})\underset{\sim}{C}^T)^{-1} = 185.908161$

$(\underset{\sim}{C}\hat{\underset{\sim}{\beta}} - \underset{\sim}{d})^T(\underset{\sim}{C}(\underset{\sim}{X}^T\underset{\sim}{X})\underset{\sim}{C}^T)^{-1}(\underset{\sim}{C}\hat{\underset{\sim}{\beta}} - \underset{\sim}{d}) = 200.045590$

$F_{obs} = \frac{(\underset{\sim}{C}\hat{\underset{\sim}{\beta}} - \underset{\sim}{d})^T(\underset{\sim}{C}(\underset{\sim}{X}^T\underset{\sim}{X})\underset{\sim}{C}^T)^{-1}(\underset{\sim}{C}\hat{\underset{\sim}{\beta}} - \underset{\sim}{d})}{r\hat{\sigma}^2} = \frac{200.045590/1}{276.414449/17} = 12.303174 > F_{0.05,1,17} = 4.45$

Hence, we reject $H_0$.

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

The model becomes $y_i - 1 = \beta_1 x_{i1} + \beta_2 x_{i2} + e_i \Rightarrow y_i' = \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$

$$\underset{\sim}{X}^T \underset{\sim}{X} = \begin{pmatrix} 905 & 1055 \\ 1055 & 1261 \end{pmatrix} \Rightarrow (\underset{\sim}{X}^T \underset{\sim}{X})^{-1} = \begin{pmatrix} 0.044748 & -0.037438 \\ -0.037438 & 0.032115 \end{pmatrix}$$

$$\underset{\sim}{X}^T \underset{\sim}{Y} = \begin{pmatrix} \sum_{i=1}^n x_{i1} y_i - \sum_{i=1}^n x_{i1} \\ \sum_{i=1}^n x_{i2} y_i - \sum_{i=1}^n x_{i2} \end{pmatrix} = \begin{pmatrix} 3048 \\ 3596 \end{pmatrix}$$

$$\underset{\sim}{\hat{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X^T X)^{-1} X^T Y = \begin{pmatrix} 1.764856 \\ 1.374516 \end{pmatrix}$$

ResS.S $= \underset{\sim}{Y}^T \underset{\sim}{Y} - \underset{\sim}{\hat{\beta}} \underset{\sim}{X}^T \underset{\sim}{Y} = 281.959376$

Hence, $\hat{\sigma}^2 = \frac{ResS.S}{n-p'} = \frac{281.959376}{20-2} = 15.664410$

(i) **Assume that $\boldsymbol{\beta_0 = 1}$ and $\boldsymbol{\beta_1 = \beta_2}$.** 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 $\sigma^2$. No need to show that it is unbiased.

The model becomes $y_i - 1 = \beta_1 (x_{i1} + x_{i2}) + e_i \Rightarrow y_i' = \beta_1 x_i' + e_i$

$X^T X = \sum_{i=1}^{20} x_i'^2 = 4276$

$X^T Y = \sum_{i=1}^{20} x_i' y_i' = 6644$

$\hat{\beta}_1 = \frac{6644}{4276} = 1.553789$

Hence, the fitted line is $\hat{y} = 1 + 1.553789(x_1 + x_2)$.
$Y^T Y = \sum_{i=1}^{20} y_i'^2 = 10604$

ResS.S $= Y^T Y - \hat{\beta}_1 X^T Y = 280.625884$

Hence, $\hat{\sigma}^2 = \frac{280.625884}{20-1} = 14.769783$

2. (a)

| catalyst | $g_{i1}$ | $g_{i2}$ |
|---|---|---|
| 1 | 1 | 0 |
| 2 | 0 | 1 |
| 3 | 0 | 0 |

$$y_i = \beta_0 + \beta_{g_1} g_{i1} + \beta_{g_2} g_{i2} + e_i$$

(b)

$$y_{ij} = \mu_i + e_{ij}$$

with i=1,2,3 j= 1,2,3,4,5

(c)

$$\hat{\boldsymbol{\mu}} = \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{Y} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_3 \end{pmatrix} = \begin{pmatrix} 81.6 \\ 81.54 \\ 80.98 \end{pmatrix}$$

(d) $\hat{\beta}_0 = \hat{\mu}_3$=80.98
$\hat{\beta}_{g_1} = \hat{\mu}_1 - \hat{\beta}_0 =$ 81.60-80.98=0.62
$\hat{\beta}_{g_2} = \hat{\mu}_3 - \hat{\beta}_0 =$ 81.54-80.98=0.56

(e) Total Sum of Squares $= \sum\sum(y_{ij} - \bar{y}_{..}) =$ 2.81354*14 = 39.3893
Residual Sum of Squares $= \sum_{i=1}^{3} s_i^2(n_i - 1) =$ 4.205*4 + 1.723*4 + 3.627*4 = 38.22
Regression Sum of Squares = Total Sum of Squares - Residual Sum of Squares = 39.3893 - 38.22 = 1.1693
Test statistic =

$$F = \frac{Reg.S.S./(m-1)}{Res.S.S/(n-m)} = \frac{1.169/2}{38.22/12} = 0.1835$$

Critical Value is $F_{0.05,2,12} = 3.89 > F_{obs} = 0.1835$
So we do not reject the null hypothesis