8. In a small-scale experimental study of the relation between degree of brand liking (Y) and moisture content (x_1) and sweetness (x_2) of the product, the following results were obtained:

$$\mathbf{X}^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} 16 & 112 & 48 \\ 112 & 864 & 336 \\ 48 & 336 & 160 \end{bmatrix},$$

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \begin{bmatrix} 1.2375 & -0.0875 & -0.1875 \\ -0.0875 & 0.0125 & 0 \\ -0.1875 & 0 & 0.0625 \end{bmatrix}, \quad \mathbf{X}^{\mathrm{T}}\mathbf{y} = \begin{bmatrix} 1308 \\ 9510 \\ 3994 \end{bmatrix} \quad \text{and} \quad \sum_{i=1}^{16} y_i^2 = 108896,$$

$= \underset{\sim}{y}^T \underset{\sim}{y}$

with $3 < x_1 < 11$ and $3 < x_2 < 5$. Assume that a multiple regression model is appropriate to explaned the expected value of Y.

(a) Fit a multiple regression model to the data. State the estimated regression function. How should $\hat{\beta}_1$ be interpreted here?

(b) Estimate the variance of Y.

(c) Test the significance of the regression, using $\alpha = 0.01$. State the hypotheses, test statistic, decision rule and conclusion.

(d) Obtain the coefficient of multiple determination. Comment the result.

(e) Derive a 99% confidence interval for the expected value of brand liking when $x_1 = 5$ and $x_2 = 4$. It is possible to make intervalar estimation when $x_1 = 12$ and $x_2 = 4$? Why? (Exam 01/02/2017)

a) $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 \underline{x_2}$
$= 37.65 + 4.425\, x_1 + 4.375\, x_2$

if we increase one unit of x_2 (with x_1 fixed) the $\hat{E}[Y|x_0]$ will increase $\hat{\beta}_1$

b) $\hat{Var}(Y) = \hat{Var}(\varepsilon) = \hat{\sigma}^2 = MSE = \dfrac{SSE}{n-p}$

$n = 16$
$p = 3$

$SSE =$

$SST = SSR + SSE \Leftrightarrow SSE = \underline{SST} - \underline{SSR}$

$$SST = \underset{\sim}{y}^T \underset{\sim}{y} - n\bar{y}^2$$

$$SSR = \underbrace{\hat{\underset{\sim}{\beta}}^T \underset{\sim}{X}^T \underset{\sim}{y}} - n\bar{y}^2$$

$$\Rightarrow SSE = \underset{\sim}{y}^T \underset{\sim}{y} - \underbrace{\hat{\beta}^T \underset{\sim}{X}^T y}$$

$$= 94.3$$

Group I

Y - degree of brand liking; x_1 - moisture content $3 < x_1 < 11$; x_2 - sweetness $3 < x_2 < 5$

model: $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$, $i = 1, \dots, 16$

a) $\hat{\beta} = (X^T X)^{-1} X^T y = \begin{bmatrix} 37.650 \\ 4.425 \\ 4.375 \end{bmatrix}$ $\hat{y} = 37.65 + 4.425 x_1 + 4.375 x_2$; $3 < x_1 < 11$; $3 < x_2 < 5$

$\hat{\beta}_1 = 4.425$, so if we have a brand with one more unit in moisture the expected brand liking will increase 4.425.

b) $\hat{var}(Y) = MSE = \frac{SSE}{n-p} = \frac{SST - SSR}{3} = \frac{y^T y - \hat{\beta}^T X^T y}{3} = \frac{108896 - 108801.7}{3} = \frac{94.3}{3} = 7.254$

c) $H_0: \beta_1 = \beta_2 = 0$ vs $H_1: \beta_1 \neq 0 \vee \beta_2 \neq 0$ (Admit that $\varepsilon_i \underset{iid}{\sim} N(0, \sigma^2)$), $\forall i$

under H_0 we have the test statistic $F_0 = \frac{MSR}{MSE} \sim F(p-1, n-p)$

$\alpha = 0.01$, we will reject H_0 if $F_0 > c = F^{-1}_{F(2,13)}(0.99) = 6.70$; $CR =]6.70; +\infty[$

observed value: $f_0 = \frac{936.35}{7.254} = 129.0832$ since $MSR = \frac{SSR}{p-1} = \frac{\hat{\beta}^T X^T y - n\bar{y}^2}{p-1} =$
$= \frac{108801.7 - 16 \times (81.75)^2}{2} = \frac{1872.7}{2} = 936.36$

Decision: Since $f_0 \in CR$, for $\alpha = 0.01$, rej H_0

$\Rightarrow \exists$ evidence that there is a linear association between degree of brand liking and the two explanatory variables x_1 and x_2.

d) $r^2 = 1 - \frac{SSE}{SST} = 1 - \frac{94.3}{1967} = 0.9521$; $SST = SSE + SSR = 94.3 + 1872.7 = 1967$

$\simeq 95.21\%$ of the total variability of y is explained with this regression model

$\Rightarrow$ the model fit well to this data set.

e) $CI_{99\%}(\mu_Y | x_0)$ where $x_0^T = (1, 5, 4)$

Pivotal quantity: $T = \frac{\hat{\mu}_{Y|x_0} - \mu_{Y|x_0}}{\sqrt{\hat{\sigma}^2 x_0^T (X^T X)^{-1} x_0}} \sim t(n-p)$; $F^{-1}_{t(13)}(0.995) = 3.012$

$CI_{99\%}(\mu_Y | x_0) = (\hat{\mu}_{Y|x_0} \pm 3.012 \sqrt{\hat{\sigma}^2 x_0^T (X^T X)^{-1} x_0})$

$\hat{\mu}_{Y|x_0} = 37.65 + 4.425 \times 5 + 4.375 \times 4 = 77.275$; $\hat{\sigma}^2 = 7.254$; $x_0^T (X^T X)^{-1} x_0 = 0.175$

$CI_{99\%}(\mu_Y | x_0) = (73.88142; 80.66858)$

the C.I. for $x_0^T = (1, 12, 4)$ it is not possible since the value for the product is out of the range of the data set used for estimate the model parameters.