Solution

1. Short questions

   (a) The design matrices are given by

$$X = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 3 & 0 & 1 \\ 1 & 4 & 0 & 0 \\ 1 & 5 & 1 & 0 \\ 1 & 6 & 0 & 0 \\ 1 & 7 & 0 & 1 \\ 1 & 8 & 0 & 0 \end{pmatrix}, \; X_c = \begin{pmatrix} 1 & -3.5 & 0.75 & -0.25 \\ 1 & -2.5 & -0.25 & -0.25 \\ 1 & -1.5 & -0.25 & 0.75 \\ 1 & -0.5 & -0.25 & -0.25 \\ 1 & 0.5 & 0.75 & -0.25 \\ 1 & 1.5 & -0.25 & -0.25 \\ 1 & 2.5 & -0.25 & 0.75 \\ 1 & 3.5 & -0.25 & -0.25 \end{pmatrix},$$

$$X_c^\top X_c = \begin{pmatrix} 8 & 0 & 0 & 0 \\ 0 & 42 & -3 & 1 \\ 0 & -3 & 1.5 & -0.5 \\ 0 & 1 & -0.5 & 1.5 \end{pmatrix}.$$

   (b) We have

$$s.e.(\hat{y_k} - y_k) = \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_k - \bar{x})^2}{S_{xx}}} = 10 \times \sqrt{1 + 1/20 + 16^2/400} = 13.$$

   (c) Let X be the original design matrix of x_1 to x_4, and $X_5 \in \mathbb{R}^{n\times 1}$ be the observations of x_5. Then the new design matrix is $X' = (X, X_5)$. Let $\hat{\beta} \in \mathbb{R}^5$ be the LSE of the original model, and $\hat{\beta}' \in \mathbb{R}^6$ be the LSE of the new one. We have

$$X'^\top X' = \begin{pmatrix} X^\top \\ X_5^\top \end{pmatrix} \begin{pmatrix} X & X_5 \end{pmatrix} = \begin{pmatrix} X^\top X & X^\top X_5 \\ X_5^\top X & X_5^\top X_5 \end{pmatrix} = \begin{pmatrix} X^\top X & 0 \\ 0 & S_{x_5 x_5} \end{pmatrix}, \; X'^\top y = \begin{pmatrix} X^\top y \\ X_5^\top y \end{pmatrix},$$

$$(X'^\top X')^{-1} = \begin{pmatrix} (X^\top X)^{-1} & 0 \\ 0 & 1/S_{x_5 x_5} \end{pmatrix},$$

   which implies

$$\hat{\beta}' = (X'^\top X')^{-1} X'^\top y = \begin{pmatrix} (X^\top X)^{-1} X^\top y \\ S_{x_5 y}/S_{x_5 x_5} \end{pmatrix} = \begin{pmatrix} \hat{\beta} \\ \hat{\beta}_5 \end{pmatrix}.$$

   Hence

$$\begin{aligned} \hat{y}'^\top \hat{y}' &= (X\hat{\beta} + X_5\hat{\beta}_5)^\top (X\hat{\beta} + X_5\hat{\beta}_5) = \hat{\beta}^\top X^\top X \hat{\beta} + 2\hat{\beta}^\top X^\top X_5 \hat{\beta} + \hat{\beta}_5^2 S_{x_5 x_5} \\ &= \hat{y}^\top \hat{y} + \hat{\beta}_5^2 S_{x_5 x_5}. \end{aligned}$$

   Then we know $Reg.SS' = \hat{y}'^\top \hat{y}' - n\bar{y}^2 = \hat{y}^\top \hat{y} + \hat{\beta}_5^2 S_{x_5 x_5} - n\bar{y}^2 = Reg.SS + \hat{\beta}_5^2 S_{x_5 x_5}$. Therefore the absolute change in R^2 resulting from the inclusion of x_5 in the model is

$$\frac{\hat{\beta}_5^2 S_{x_5 x_5}}{TSS} = \frac{1.5^2 * 100}{900} = 0.25.$$

2. (a) i. We have $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{5240-72*1950/28}{218-72^2/28} = \frac{225.7143}{32.85714} = 6.869565$. Then

$$\hat{\sigma}^2 = \frac{RSS}{n-2} = \frac{S_{yy} - \hat{\beta}_1^2 S_{xx}}{n-2} = \frac{5745.87}{26} = 220.995.$$

ii. From above we know $Res.SS = 5745.87$. Pure error SS $= (20600 - 340^2/6) + (35600 - 520^2/8) + (38100 - 510^2/7) + (48800 - 580^2/7) = 4819.048$. Then Lack of fit SS $= 5745.87 - 4819.048 = 926.822$.
Then the test statistic is $F = \frac{926.822/2}{4819.048/24} = 2.307896 < F_{0.05}(2, 24) = 3.402826$. Hence we cannot reject the null hypothesis of lack of fit.

iii. Since there is no lack of fit, the unbiased estimate of σ^2 is still 220.995.

(b) i.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{m} \sum_{i=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2}{n-m} = \frac{4819.048}{24} = 200.7937.$$

ii. $H_0 : \mu_1 = \cdots = \mu_m$. We have $TSS = S_{yy} - n\bar{y}^2 = 143100 - 1950^2/28 = 7296.429$, $RSS = 4819.048$ and then $RegSS = TSS - RSS = 2477.381$.
Then the test statistic is

$$F = \frac{RegSS/(m-1)}{\hat{\sigma}^2} = \frac{2477.381/3}{200.7937} = 4.112647.$$

The critical value is $F_\alpha(3, 24) = 3.008787 < F$.
Therefore we reject H_0.

iii. Let $\beta^T = (\mu_1, \mu_2, \mu_3, \mu_4)$ and

$$C = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \end{pmatrix}.$$

Then $H_0 \Leftrightarrow C\beta = 0$. Note that $\text{rank}(C) = 2$. We know $Cov(\hat{\beta}) = \sigma^2 diag(1/n_i)$ and $C\hat{\beta} = (\hat{\mu}_1 - \hat{\mu}_2, \hat{\mu}_2 - \hat{\mu}_3)^\top = (-16.19048, 7.857143)^\top$. Then

$$Cov(C\hat{\beta}) = \sigma^2 C diag(1/n_i) C^\top = \sigma^2 \begin{pmatrix} \frac{1}{n_1} + \frac{1}{n_2} & -\frac{1}{n_2} \\ -\frac{1}{n_2} & \frac{1}{n_2} + \frac{1}{n_3} \end{pmatrix}.$$

Hence the test statistic is given by

$$F = \frac{(C\hat{\beta})^\top [C diag(1/n_i) C^\top]^{-1} C\hat{\beta}/2}{\hat{\sigma}^2} = \frac{847.619/2}{200.7937} = 2.110671.$$

The critical value is $F_\alpha(2, 24) = 3.402826 > F$.
Therefore we cannot reject H_0.

iv. $H_0 : \mu_1 + \mu_2 + \mu_3 - 3\mu_4 = 0$.
We have

$$SSW = \frac{(340/6 + 510/7 + 520/8 - 580/7 * 3)^2}{1/6 + 1/7 + 1/8 + 9/7} = \frac{2921.145}{1.720238} = 1698.105.$$

Test statistic is $F = \frac{SSW}{\hat{\sigma}^2} = \frac{1698.105}{200.7937} = 8.456964$.
The critical value is $F_\alpha(1, 24) = 4.259677 < F$.
Hence we reject H_0.