MA2261 - DLI, Linear Statistical Models, Year 2022-2023

Solutions of exercises for feedback class 8
(Note: the exercise number refers to the workbook)

EXERCISE 4.4 ii)

ii) From the R listing we have SSE (null model) $-SSE(x_1) = 1203.94$
$SSE(x_1) - SSE(x_1, x_2) = 142.92$

Also, SSE (null model) $= SST$. Hence

$SSM(x_1, x_2) = SST - SSE(x_1, x_2) = SST - SSE(x_1) + SSE(x_1) - SSE(x_1, x_2) = 1203.94 + 142.92 = 1346.86$.

Therefore $R^2 = \dfrac{SSM(x_1, x_2)}{SST} = \dfrac{1346.86}{1700.3} = 0.7921$.

Thus 79.21% of the total variation in Y is explained by the complete model.

EXERCISE 4.5

i)

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_{25} \end{pmatrix} \qquad \boldsymbol{X} = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & x_{1,1}x_{2,1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1,25} & x_{2,25} & x_{1,25}x_{2,25} \end{pmatrix} \qquad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{25} \end{pmatrix}$$

The model in matrix form is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \ .$$

ii)

$$
\begin{aligned}
SSE &= \hat{\sigma}^2(25-4) = 16.0986 \times 21 = 338.07 \\
R^2 &= \frac{SSM}{SST} = \frac{SSM}{SSM+SSE}
\end{aligned}
$$

$$
\begin{aligned}
R^2\,SSM + R^2\,SSE &= SSM \\
SSM(1-R^2) &= R^2\,SSE \\
SSM &= \frac{R^2}{1-R^2}SSE = \frac{0.8012}{1-0.8012} \times 338.07 = 1362.483\ . \\
SST &= \frac{SSM}{R^2} = 1700.3 \\
MSM &= SSM/(p-1) = 1362.483/3 = 454.161 \\
MSE &= SSE/(n-p) = 338.07/21 = 16.0986
\end{aligned}
$$

The ANOVA table is therefore

	SS	d.f.	MS	F
SSM	1362.483	p-1=3	454.161	28.21
SSE	338.07	n-p=21	16.0986	
SST	1700.3	n-1=24		

iii)

Let the model with the interaction term be the full model and the model with only x_1 and x_2 be the reduced model. From R listings in Exercise 4.4, we have

$$SSE_{\mathrm{R}} = SST_{\mathrm{R}} - SSM_{\mathrm{R}} = 1700.3 - 1203.94 - 142.92 = 353.44$$

Therefore

$$SS_{extra} = SSE_{\mathrm{R}} - SSE_{\mathrm{F}} = 353.44 - 338.07 = 15.37\ .$$

Hence

$$\frac{SS_{extra}/(p-q)}{SSE_{\mathrm{F}}/(n-p)} = \frac{15.37/1}{338.07/21} = 0.9547 \sim F_{1,21}.$$

The critical region is $(4.325, +\infty)$. We accept $H_0 : \beta_3 = 0$. We conclude that the addition of the interaction term does not give an improvement in fit. Thus the interaction between variables x_1 and x_2 in model (4.1) does not have a statistically significant effect on the mean life expectancy.