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

Solutions of exercises for feedback class 9
(Note: the exercise number refers to the workbook)

EXERCISE 5.2

i) The response variable is the weight of potatoes, the explanatory variable is the fertilizer. The explanatory variable is categorical, with 3 levels A,B,C.

ii) The design matrix is the 11×3 matrix

$$\boldsymbol{X} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

The parameter vector is $\boldsymbol{\beta} = (\mu_1, \mu_2, \mu_3)^T$, the error vector is
$\boldsymbol{\varepsilon} = (\varepsilon_{11}\ \varepsilon_{21}\ \varepsilon_{31}\ \varepsilon_{12}\ \varepsilon_{22}\ \varepsilon_{32}\ \varepsilon_{42}\ \varepsilon_{13}\ \varepsilon_{23}\ \varepsilon_{33}\ \varepsilon_{43})^T$ with $\varepsilon_{ij} \sim N(0, \sigma^2)$ iid.

The response vector is
$\boldsymbol{Y} = (y_{11}\ y_{21}\ y_{31}\ y_{12}\ y_{22}\ y_{32}\ y_{42}\ y_{13}\ y_{23}\ y_{33}\ y_{43})^T$.

The model is

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

iii) The analysis of variance table for the one-way ANOVA gives a decomposition of the total sum of squares SST, which is the quantity measuring the total variation of the observations from the overall mean. This decomposition is given by

$$SST = SSB + SSE\ .$$

Here SSE is the error variation, which is a measure of the random variation of the observations around the respective factor level sample means.

SSB is the between group variation, which is a measure of the extent of differences between factor level sample means, based on the deviation of the factor level sample means $\bar{y}_j$ around the overall mean $\bar{y}$.

For the above data, we calculate

$$SST = \sum_{j=1}^{k} \sum_{i=1}^{n_j} y_{ij}^2 - CF = 43.0925 - 42.2184 = 0.8741$$

$$SSB = \sum_{j=1}^{k} n_j \bar{y}_j^2 - CF = 3 \times 3.4225 + 4 \times 3.0888 + 4 \times 5.0288 - 42.2184 = 0.5195$$

$$SSE = SST - SSB = 0.8741 - 0.5195 = 0.3546$$

The analysis of variance also shows the degree of freedom associated with each component of SST and the mean squares

$$MSB = \frac{SSB}{k-1} \qquad MSE = \frac{SSE}{N-k} \; .$$

For the above data, $k = 3$, $N = 11$, therefore

$$MSB = \frac{0.5195}{3-1} = 0.25975 \qquad MSE = \frac{0.3545}{11-3} = 0.0443$$

$$F = \frac{MSB}{MSE} = 5.863 \sim F_{2,8}$$

The ANOVA table is therefore

	SS	d.f.	MS	F
SSB	0.5195	2	0.25975	5.863
SSE	0.3546	8	0.0443	
SST	0.8741	10		

EXERCISE 5.3

i) If the hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu$ holds, then

$$F = \frac{MSB}{MSE} = 5.863 \sim F_{k-1,N-k} = F_{2,8}$$

The critical region is $(4.459, +\infty)$. We reject H_0. Thus we conclude that there is an evidence of difference between the three fertilizers.

ii) Consider the contrast $\mathcal{C} = \dfrac{\mu_1 + \mu_2}{2} - \mu_3$. This is of the form $\mathcal{C} = c_1\mu_1 + c_2\mu_2 + c_3\mu_3$ with $c_1 = c_2 = \dfrac{1}{2}$, $c_3 = -1$, $c_1 + c_2 + c_3 = 0$.

The point estimate of $\mathcal{C}$ is

$$\hat{\mathcal{C}} = \frac{1}{2}\bar{y}_1 + \frac{1}{2}\bar{y}_2 - \bar{y}_3 = 0.5 \times 1.85 + 0.5 \times 1.7575 - 2.2425 = -0.43875 \ ,$$

while

$$s^2(\hat{\mathcal{C}}) = MSE \sum_{j=1}^{k} \frac{c_j^2}{n_j} = 0.0443(0.25/3 + 0.25/4 + 1/4) = 0.01754 \ .$$

The 95% confidence interval for $\mathcal{C}$ is

$$\hat{\mathcal{C}} \pm t_{0.025,8}\, s(\hat{\mathcal{C}}) = -0.43875 \pm 2.306 \times 0.13244 = (-0.7441,\ -0.1334) \ .$$

iii) The confidence interval in part b) ii) does not contain 0. Hence the hypothesis $\mathcal{C} = 0$ is rejected. Thus there is a significant difference in mean weight of potatoes between using the fertilizer with additive and the one without it.