MLDS 401/IEMS 404-1 (Fall 2023): Lab 4 Solution – 10/17/2023

3.7 (Derivation of the one-way ANOVA F-test using the extra SS method):

a. The ANOVA identity follows immediately from (3.13) by noting that $\widehat{y}_{ij} = \overline{y}_i$ and $e_{ij} = y_{ij} - \widehat{y}_{ij} = y_{ij} - \overline{y}_i$ for $j = 1, \ldots, n_i$ and $i = 1, \ldots, k$. The squared norms of the vectors are then given by the corresponding sums of squares.

b. Under H_0, the one-way ANOVA model becomes $y_{ij} = \mu + \varepsilon_{ij}$, where μ is the common mean of all groups under H_0. It is easy to show that the LS estimate of μ is $\widehat{\mu} = \overline{\overline{y}}$. So $\text{SSE}_0 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \overline{\overline{y}})^2 = \text{SST}$. Therefore $\text{SSH}_0 = \text{SSE}_0 - \text{SSE} = \text{SST} - \text{SSE} = \text{SSG}$ from the ANOVA identity. The hypothesis d.f. is $k-1$ since without H_0 there are k independent parameters, $\mu_1, \ldots, \mu_k$, while under H_0, there is only one, μ; so H_0 imposes $k-1$ linearly independent restrictions.. The error d.f. is, of course, $N-k$. This explains the extra SS F-statistic, which equals the ANOVA F-statistic.

6.2 (Extra sum of squares test in terms of R^2): The extra SS test statistic for comparing the two models is given by

$$F = \frac{(\text{SSE}_q - \text{SSE}_p)/(p-q)}{\text{SSE}_p/[n-(p+1)]}.$$

We have $\text{SSE}_p = \text{SST}(1 - R_p^2)$ and $\text{SSE}_q = \text{SST}(1 - R_q^2)$. Substituting in the extra SS F-statistic we get

$$\begin{aligned} F &= \frac{\text{SST}[(1-R_q^2) - (1-R_p^2)]}{\text{SST}(1-R_p^2)/[n-(p+1)]} \\ &= \frac{(R_p^2 - R_q^2)/(p-q)}{(1-R_p^2)/[n-(p+1)]}. \end{aligned}$$

For the given data

$$F = \frac{(0.90 - 0.80)/(5-3)}{(1-0.90)/(26-6)} = 10 > f_{2,20,.01} = 5.85.$$

So the increase R^2 is significant at $\alpha = .01$.

3.9 (Alternate coding of categorical variables): With the new coding: $x_1 = \pm 1$ and $x_2 = \pm 1$, we get the following equations:

$$\begin{aligned}
\beta_0 - \beta_1 - \beta_2 + \beta_3 &= 40 \\
\beta_0 + \beta_1 - \beta_2 - \beta_3 &= 45 \\
\beta_0 - \beta_1 + \beta_2 - \beta_3 &= 50 \\
\beta_0 + \beta_1 + \beta_2 + \beta_3 &= 65.
\end{aligned}$$

The solution to this square system of equations $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{y}$ is the LS estimate $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ where

$$\boldsymbol{X} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \boldsymbol{y} = \begin{bmatrix} 40 \\ 45 \\ 50 \\ 65 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}.$$

Note that $\boldsymbol{X}$ is an orthogonal matrix and $\boldsymbol{X}'\boldsymbol{X} = 4\boldsymbol{I}$. Hence

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\
&= \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}\begin{bmatrix} 40 \\ 45 \\ 50 \\ 65 \end{bmatrix} \\
&= \frac{1}{4}\begin{bmatrix} 40 + 45 + 50 + 65 \\ -40 + 45 - 50 + 65 \\ -40 - 45 + 50 + 65 \\ 40 - 45 - 50 + 65 \end{bmatrix} = \begin{bmatrix} 50 \\ 5 \\ 7.5 \\ 2.5 \end{bmatrix}.
\end{aligned}$$

The interpretations of these coefficients are as follows.

1. $\widehat{\beta}_0$ is the overall mean of all the y's.
2. $\widehat{\beta}_1$ is (1/4)th times the change in y when x_1 is changed from -1 to $+1$ summed over the two levels of x_2 (called the main effect of x_1).
3. $\widehat{\beta}_2$ is (1/4)th times the change in y when x_2 is changed from -1 to $+1$ summed over the two levels of x_1 (called the main effect of x_2).
4. $\widehat{\beta}_3$ is (1/4)th times the difference in the changes in y when x_1 is changed from -1 to $+1$ between the two levels of x_2 (called the interaction between x_1 and x_2). This is also equal to (1/4)th times the difference in the changes in y when x_2 is changed from -1 to $+1$ between the two levels of x_1.