ANOVA using Linear Models in R

Consider the following data corresponding to a randomized block design with $k = 4$ treatments and $b = 5$ blocks.

	Block 1	Block 2	Block 3	Block 4	Block 5
Treatment 1	-3.0	-5.0	-5.5	-6.4	-5.5
Treatment 2	-2.3	-2.3	-1.8	-2.8	-1.6
Treatment 3	-0.3	1.6	1.5	-1.6	0.3
Treatment 4	6.2	4.6	5.5	5.5	3.7

Modeling as a Linear Model

We can use the model

$$Y_{ij} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \epsilon_{ij},$$

where

$$x_i = \begin{cases} 1, & \text{if the observations is from block } i, \\ 0, & \text{otherwise} \end{cases}$$

for $i = 1, 2, 3, 4$, and

$$x_{i+4} = \begin{cases} 1, & \text{if treatment } i \text{ is applied to the observation,} \\ 0, & \text{otherwise} \end{cases}$$

for $i = 1, 2, 3$. Here, we assume $\epsilon_{ij} \sim N(0, \sigma^2)$.

We can construct the data matrices for the linear model using the following code.

```r
# Store observations as vector.
y <- c(-3.0, -2.3, -0.3, 6.2,
                     -5.0, -2.3, 1.6, 4.6,
                     -5.5, -1.8, 1.5, 5.5,
                     -6.4, -2.8, -1.6, 5.5,
                     -5.5, -1.6, 0.3, 3.7)

# Make first indicator variable by setting first column of
# four by matrix equal to 1.
x1 <- matrix(0, nrow = 4, ncol = 5)
x1[,1] <- 1

# Remaining indicator variables for blocks.
x2 <- matrix(0, nrow = 4, ncol = 5)
x2[,2] <- 1
x3 <- matrix(0, nrow = 4, ncol = 5)
x3[,3] <- 1
x4 <- matrix(0, nrow = 4, ncol = 5)
x4[,4] <- 1
```

```r
# Make indicator variable for first treatment by setting first row of zero matrix
# equal to 1.
x5 <- matrix(0, nrow = 4, ncol = 5)
x5[1,] <- 1

# Repeat for remaining indicator variables for blocks.
x6 <- matrix(0, nrow = 4, ncol = 5)
x6[2,] <- 1
x7 <- matrix(0, nrow = 4, ncol = 5)
x7[3,] <- 1

# Store all indicator variables as matrix.
X <- cbind(rep(x = 1, times = 20), # beta0 column
                 as.vector(x1),
                 as.vector(x2),
                 as.vector(x3),
                 as.vector(x4),
                 as.vector(x5),
                 as.vector(x6),
                 as.vector(x7))

# Make X'X and X'Y.
XX <- t(X) %*% X
XY <- t(X) %*% y
```

This yields

$$\mathbf{Y} = \begin{bmatrix} -3 \\ -2.3 \\ -0.3 \\ 6.2 \\ -5 \\ -2.3 \\ 1.6 \\ 4.6 \\ -5.5 \\ -1.8 \\ 1.5 \\ 5.5 \\ -6.4 \\ -2.8 \\ -1.6 \\ 5.5 \\ -5.5 \\ -1.6 \\ 0.3 \\ 3.7 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{X'X} = \begin{bmatrix} 20 & 4 & 4 & 4 & 4 & 5 & 5 & 5 \\ 4 & 4 & 0 & 0 & 0 & 1 & 1 & 1 \\ 4 & 0 & 4 & 0 & 0 & 1 & 1 & 1 \\ 4 & 0 & 0 & 4 & 0 & 1 & 1 & 1 \\ 4 & 0 & 0 & 0 & 4 & 1 & 1 & 1 \\ 5 & 1 & 1 & 1 & 1 & 5 & 0 & 0 \\ 5 & 1 & 1 & 1 & 1 & 0 & 5 & 0 \\ 5 & 1 & 1 & 1 & 1 & 0 & 0 & 5 \end{bmatrix} \qquad \mathbf{X'Y} = \begin{bmatrix} -9.2 \\ 0.6 \\ -1.1 \\ -0.3 \\ -5.3 \\ -25.4 \\ -10.8 \\ 1.5 \end{bmatrix}$$

We calculate the least squares estimator

$$\hat{\beta} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$$

using the following code.

```r
hBeta <- solve(XX, XY)
```

This gives

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 4.785 \\ 0.925 \\ 0.5 \\ 0.7 \\ -0.55 \\ -10.18 \\ -7.26 \\ -4.8 \end{bmatrix}$$

with sum of squared errors given by

$$\text{SSE}_C = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} = 301.02 - 288.468 = 12.552.$$

Testing Significance of Blocking

Suppose that we want to test if there is a significant difference between block means. The appropriate test uses null hypothesis is $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, and reduced model

$$Y = \beta_0 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \epsilon.$$

The corresponding $\mathbf{X}$ matrix consists of the first, sixth, seventh, and eighth columns of $\mathbf{X}$, which can be formed using the following code.

```r
Xr <- X[, c(1, 6:8)]
```

We can calculate the reduced model using the following code.

```r
hBetaR <- solve(t(Xr) %*% Xr, t(Xr) %*% y)
```

This gives

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{R}} = \begin{bmatrix} 5.1 \\ -10.18 \\ -7.26 \\ -4.8 \end{bmatrix}$$

with sum of squared errors given by

$$\text{SSE}_R = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}_{\boldsymbol{R}}'\mathbf{X}_{\mathbf{R}}'\mathbf{Y} = 301.02 - 282.86 = 18.16.$$

To test H_0 we use the F statistic

$$F = \frac{(\text{SSE}_R - \text{SSE}_C)/(k-g)}{\text{SSE}_C/(n-(k+1))},$$

where $k = 7$, $g = 3$, and n, which we calculate using the following code.

```r
sseR <- t(y) %*% y -  t(hBetaR) %*% t(Xr) %*% y
sseC <- t(y) %*% y -  t(hBeta) %*% t(X) %*% y

fval <- (sseR - sseC)/(7 -3)/(sseC/(20 - 8))
```

This yields

$$F = 1.3403442.$$

Since the critical value of F with $\nu_1 = 4$ and $\nu_2 = 12$ for an $\alpha = 0.05$ level test is $f_{0.05,\nu_1,\nu_2} = 3.2592$, we cannot reject H_0. This implies that there is little evidence that blocking has a significant effect.