

# STA207 Homework 3

Juanjuan Hu, Zhen Zhang

January 28, 2016

## 23.13

```
transform_function <- function(x) {  
  log10(x + 1)  
}  
kidney$Days <- transform_function(as.numeric(as.character(kidney$Days)))  
kidney <- subset(kidney, Duration != 2 | Weight != 1) # empty y_21 cell
```

(a)

The full model is:

$$Y_{ijk} = \mu_{..} + \alpha_1 * X_{ij1} + \beta_1 * X_{ij2} + \beta_2 * X_{ij3} + \epsilon_{ijk}$$

The reduced model for testing for factor A main effects is:

$$Y_{ijk} = \mu_{..} + \beta_1 * X_{ij2} + \beta_2 * X_{ij3} + \epsilon_{ijk}$$

The reduced model for testing for factor B main effects is:

$$Y_{ijk} = \mu_{..} + \alpha_1 * X_{ij1} + \epsilon_{ijk}$$

Where:

$$X_{ij1} = \begin{cases} 1, & \text{if case from level 1 for factor A} \\ -1, & \text{if case from level 2 for factor A} \end{cases}$$

$$X_{ij2} = \begin{cases} 1, & \text{if case from level 1 for factor B} \\ -1, & \text{if case from level 3 for factor B} \\ 0, & \text{otherwise} \end{cases}$$

$$X_{ij3} = \begin{cases} 1, & \text{if case from level 2 for factor B} \\ -1, & \text{if case from level 3 for factor B} \\ 0, & \text{otherwise} \end{cases}$$

(b)

```
kidney$X1 = ifelse(kidney$Duration == 1, 1, -1)  
kidney$X2 = ifelse(kidney$Weight == 1, 1, ifelse(kidney$Weight == 3, -1, 0))  
kidney$X3 = ifelse(kidney$Weight == 2, 1, ifelse(kidney$Weight == 3, -1, 0))  
fullModel = with(kidney, lm(Days~X1+X2+X3))  
fullModel
```

```
##
## Call:
## lm(formula = Days ~ X1 + X2 + X3)
##
## Coefficients:
## (Intercept)          X1          X2          X3
##    0.66939    0.11733   -0.34323    0.02608
```

```
sse.full = anova(fullModel)[4,2]
sse.full
```

```
## [1] 4.489821
```

```
reducedModelTestA = with(kidney, lm(Days~X2+X3))
reducedModelTestA
```

```
##
## Call:
## lm(formula = Days ~ X2 + X3)
##
## Coefficients:
## (Intercept)          X2          X3
##    0.70850   -0.26502   -0.01303
```

```
sse.reduceA = anova(reducedModelTestA)[3,2]
sse.reduceA
```

```
## [1] 5.040447
```

```
reducedModelTestB = with(kidney, lm(Days~X1))
reducedModelTestB
```

```
##
## Call:
## lm(formula = Days ~ X1)
##
## Coefficients:
## (Intercept)          X1
##    0.75520    0.03152
```

```
sse.reduceB = anova(reducedModelTestB)[2,2]
sse.reduceB
```

```
## [1] 7.10425
```

From R output above, we can fit the full model as:

$$\hat{Y} = 0.66939 + 0.11733 * X_{ij1} - 0.34323 * X_{ij2} + 0.02608 * X_{ij3}$$

And the SSE of full model is 4.4898209.

The reduced model for testing A main effects is:

$$\hat{Y} = 0.70850 - 0.26502 * X_{ij2} - 0.01303 * X_{ij3}$$

The corresponding SSE is 5.0404474.

The reduced model for testing B main effects is:

$$\hat{Y} = 0.75520 + 0.03152 * X_{ij1}$$

The corresponding SSE is 7.1042504.

#### TESTING A MAIN EFFECTS:

$$H_0 : \alpha_1 = 0$$

$$H_1 : \alpha_1 \neq 0$$

$$F^* = \frac{(5.0404474 - 4.4898209)/1}{(4.4898209)/46} = 0.5506265/0.0976048 = 5.6414$$

$$F(0.95, 1, 46) = 4.0517$$

$$p - val = 0.02176$$

The decision rule is: if  $F^*$  is greater than 4.0517, then reject  $H_0$ , otherwise, accept  $H_1$ . Here,  $5.6414 \geq 4.0517$ , so we reject  $H_0$ , concluding that factor A main effects are present. The p-value is 0.02176, which is less than 0.05, leading to the same conclusion.

#### TESTING B MAIN EFFECTS:

$$H_0 : \beta_1 = \beta_2 = 0$$

$$H_1 : \text{not all } \beta \text{ equal to } 0$$

$$F^* = \frac{(7.1042504 - 4.4898209)/2}{(4.4898209)/46} = 1.307215/0.0976048 = 13.39294$$

$$F(0.95, 2, 46) = 3.1996$$

$$p - val = 2.608231e - 05$$

The decision rule is: if  $F^*$  is greater than 3.1996, then reject  $H_0$ , otherwise, accept  $H_1$ . Here,  $13.39294 \geq 3.1996$ , so we reject  $H_0$ , concluding that factor B main effects are present. The p-value is almost zero, which is less than 0.05, leading to the same conclusion.

## 23.19

### (a)

The ANOVA model is:

$$Y_{ij} = \mu_{..} + \rho_i + \tau_j + \epsilon_{ij}, \text{ Where } i = 1, 2, \dots, 5, j = 1, 2, 3$$

The corresponding regression model is:

$$Y_{ij} = \mu_{..} + \rho_1 * X_{ij1} + \rho_2 * X_{ij2} + \rho_3 * X_{ij3} + \rho_4 * X_{ij4} + \tau_1 * X_{ij5} + \tau_2 * X_{ij6} + \epsilon_{ij}$$

Where:

$$X_{ij1} = \begin{cases} 1, & \text{if case from block 1} \\ -1, & \text{if case from block 5} \\ 0, & \text{otherwise} \end{cases}$$

$$X_{ij2} = \begin{cases} 1, & \text{if case from block 2} \\ -1, & \text{if case from block 5} \\ 0, & \text{otherwise} \end{cases}$$

$$X_{ij3} = \begin{cases} 1, & \text{if case from block 3} \\ -1, & \text{if case from block 5} \\ 0, & \text{otherwise} \end{cases}$$

$$X_{ij4} = \begin{cases} 1, & \text{if case from block 4} \\ -1, & \text{if case from block 5} \\ 0, & \text{otherwise} \end{cases}$$

$$X_{ij5} = \begin{cases} 1, & \text{if case from treatment 1} \\ -1, & \text{if case from treatment 3} \\ 0, & \text{otherwise} \end{cases}$$

$$X_{ij6} = \begin{cases} 1, & \text{if case from treatment 2} \\ -1, & \text{if case from treatment 3} \\ 0, & \text{otherwise} \end{cases}$$

(b)

The reduced model for testing for differences in the mean reductions in lipid level for treatment is:

$$Y_{ij} = \mu_{..} + \rho_1 * X_{ij1} + \rho_2 * X_{ij2} + \rho_3 * X_{ij3} + \rho_4 * X_{ij4} + \epsilon_{ij}$$

(c)

```
Yij = c(0.73, 0.67, 0.15, 0.86, 0.75, 0.21, 0.94, 0.81, 0.26, 1.4, 1.32, 0.75, 1.62, 1.41, 0.78)
obs = data.frame(matrix(Yij, 5,3,2))
rownames(obs) = c("block1", "block2", "block3", "block4", "block5")
names(obs) = c("treatment1", "treatment2", "treatment3")
obs[1,3] = NA
obs[5,1] = NA
obs
```

```
##          treatment1 treatment2 treatment3
## block1          0.73          0.67          NA
## block2          0.86          0.75          0.21
## block3          0.94          0.81          0.26
## block4          1.40          1.32          0.75
## block5           NA          1.41          0.78
```

```
Y = c(0.73, 0.67, 0.86, 0.75, 0.21, 0.94, 0.81, 0.26, 1.4, 1.32, 0.75, 1.41, 0.78)
X1 = c(1,1,0,0,0,0,0,0,0,0,0,-1,-1)
X2 = c(0,0,1,1,1,0,0,0,0,0,0,-1,-1)
X3 = c(0,0,0,0,0,1,1,1,0,0,0,-1,-1)
X4 = c(0,0,0,0,0,0,0,0,1,1,1,-1,-1)
X5 = c(1,0,1,0,-1,1,0,-1,1,0,-1,0,-1)
X6 = c(0,1,0,1,-1,0,1,-1,0,1,-1,1,-1)
df = cbind(Y, X1, X2, X3, X4, X5,X6)
df = data.frame(df)
```

```
fullModel2 = with(df, lm(Y~X1+X2+X3+X4+X5+X6))
fullModel2
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X3 + X4 + X5 + X6)
##
## Coefficients:
## (Intercept)          X1          X2          X3          X4
##      0.8294      -0.3361      -0.2227      -0.1594       0.3273
##          X5          X6
##      0.2508       0.1626
```

```
sse.full12 = anova(fullModel2)[7,2]
sse.full12
```

```
## [1] 0.00350582
```

```
reducedModel2 = with(df, lm(Y~X1+X2+X3+X4))
reducedModel2
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X3 + X4)
##
## Coefficients:
## (Intercept)          X1          X2          X3          X4
##      0.8457      -0.1457      -0.2390      -0.1757       0.3110
```

```
sse.reduce2 = anova(reducedModel2)[5,2]
sse.reduce2
```

```
## [1] 0.9541833
```

From R output above, we can fit the full model as:

$$\hat{Y} = 0.8294 - 0.3361 * X_1 - 0.2227 * X_2 - 0.1594 * X_3 + 0.3273 * X_4 + 0.2508 * X_5 + 0.1626 * X_6$$

And the SSE of full model is 0.0035058.

The reduced model for testing differences in the treatment is:

$$\hat{Y} = 0.8457 - 0.1457 * X_1 - 0.2390 * X_2 - 0.1757 * X_3 + 0.311 * X_4$$

The corresponding SSE is 0.9541833.

### TESTING TREATMENT EFFECTS:

$$H_0 : \tau_1 = \tau_2 = 0$$

$$H_1 : \text{not all equal to zero}$$

$$F^* = \frac{(0.9541833 - 0.0035058)/2}{(0.0035058)/6} = 0.4753/0.00058 = 819.48$$

$$F(0.95, 2, 6) = 5.1433$$

The decision rule is: if  $F^*$  is greater than 5.1433, then reject  $H_0$ , otherwise, accept  $H_1$ . Here,  $819.48 \geq 4.0517$ , so we reject  $H_0$ , concluding that the mean reductions in lipid level differ for the three diets. The result is the same as obtained in Problem 23.17d.

(d)

```
vcov(fullModel2)
```

```
##           (Intercept)           X1           X2           X3
## (Intercept)  4.760990e-05  1.298452e-05 -8.656346e-06 -8.656346e-06
## X1          1.298452e-05  2.448509e-04 -5.193808e-05 -5.193808e-05
## X2          -8.656346e-06 -5.193808e-05  1.644706e-04 -3.029721e-05
## X3          -8.656346e-06 -5.193808e-05 -3.029721e-05  1.644706e-04
## X4          -8.656346e-06 -5.193808e-05 -3.029721e-05 -3.029721e-05
## X5           4.328173e-06 -3.524369e-05 -4.328173e-06 -4.328173e-06
## X6          -8.656346e-06 -1.298452e-05  8.656346e-06  8.656346e-06
##           X4           X5           X6
## (Intercept) -8.656346e-06  4.328173e-06 -8.656346e-06
## X1          -5.193808e-05 -3.524369e-05 -1.298452e-05
## X2          -3.029721e-05 -4.328173e-06  8.656346e-06
## X3          -3.029721e-05 -4.328173e-06  8.656346e-06
## X4           1.644706e-04 -4.328173e-06  8.656346e-06
## X5          -4.328173e-06  1.051128e-04 -4.328173e-05
## X6           8.656346e-06 -4.328173e-05  8.656346e-05
```

Construct:  $L = \tau_1 - \tau_3 = 2 * \tau_1 + \tau_2$

$$\hat{L} = 2 * \hat{\tau}_1 + \hat{\tau}_2 = 2 * 0.2508 + 0.1626 = 0.6642$$

According to the covariance matrix of model coefficients,  $s^2\{\hat{\tau}_1\} = 1.051128e - 04$ ,  $s^2\{\hat{\tau}_2\} = 8.656346e - 05$ ,  $s\{\hat{\tau}_1, \hat{\tau}_2\} = -4.328173e - 05$ . Therefore:  $s\{\hat{L}\} = \text{sqrt}(4 * 1.051128e - 04 + 8.656346e - 05 + 4 * (-4.328173e - 05)) = 0.0182726$

$$t(0.99, 6) = 3.142668$$

$$\hat{L} + t(0.99, 6) * s\{\hat{L}\} = 0.6642 + 3.142668 * 0.0182726 = 0.7216247 \quad \hat{L} - t(0.99, 6) * s\{\hat{L}\} = 0.6642 - 3.142668 * 0.0182726 = 0.6067753$$

Therefore, the 98% confidence interval for difference in diet1 and diet3 is [0.6068, 0.7216]. We can find that the CI does not include zero, indicating that mean reduction in lipid for diet 1 is significantly larger than the reduction for diet3.