Assignment 1 : Linear Models

Subhrajyoty Roy

September 17, 2018

Question 1

A firm wishes to compare four programs for training workers to perform a certain manual task.Twenty new employees are randomly assigned to the training programs, with 5 in each program. At the end of the training period, a test is conducted to see how quickly trainees can perform the task. The number of times the task is performed per minute is recorded for each trainee, with the following results:

- Program 1: 9, 12, 14, 11, 13
- Program 2: 10, 6, 9, 9, 10
- Program 3: 12, 14, 11, 13, 11
- Program 4: 9, 8, 11, 7, 8

Each one of you select two random numbers from 1 to 4, and again one random number from 1 to 5. Corresponding to the two random numbers selected from 1 to 4, you consider the two programs and delete the observation corresponding to the number selected from 1 to 5, for each of the two selected programs. Thus in the above data, all of you have to work with 5 observations for two programs and 4 observations for the other two programs

```
prog1 <- c(9,12,14,11,13)
prog2 <- c(10,6,9,9,10)
prog3 <- c(12,14,11,13,11)
prog4 <- c(9,8,11,7,8)   #generate the vectors with the data

set.seed(2018)  # a seed to generate the random number so that the work is reproducible
random1 <- sample(1:4,2)
print(random1)  #check the first random numbers
```

```
[1] 2 4
```

```
random2 <- sample(1:5,1)
print(random2) #check the second random number
```

```
[1] 1
```

Since the first two random numbers between 1 and 4 are 2 and 4, and the second random number between 1 and 5 is given as 1, hence, we should delete the 1st elements of *prog2* and *prog4* variable.

```
prog2 <- prog2[-1]  #remove the 1st element and update as necessary
prog4 <- prog4[-1]  #remove the 1st element
```

Now, we should construct the data frame, with one column containing the observations and another column containing the factor, which shows the index of the program it is assigned to.

```
observations <- c(prog1, prog2, prog3, prog4)
programs <- factor(rep(c(1:4), c(5,4,5,4)))
#contains the factor variable of 5 many 1's, followed by 4 many 2's, then 5 many 3's and 4 many
 4's

data <- data.frame(obs = observations, fac = programs)
```

Let us look at the structure of the data just to check whether everything we worked so far is correct.

```
str(data)
```

```
'data.frame':   18 obs. of  2 variables:
 $ obs: num  9 12 14 11 13 6 9 9 10 12 ...
 $ fac: Factor w/ 4 levels "1","2","3","4": 1 1 1 1 1 2 2 2 2 3 ...
```

(a) Construct the ANOVA table.

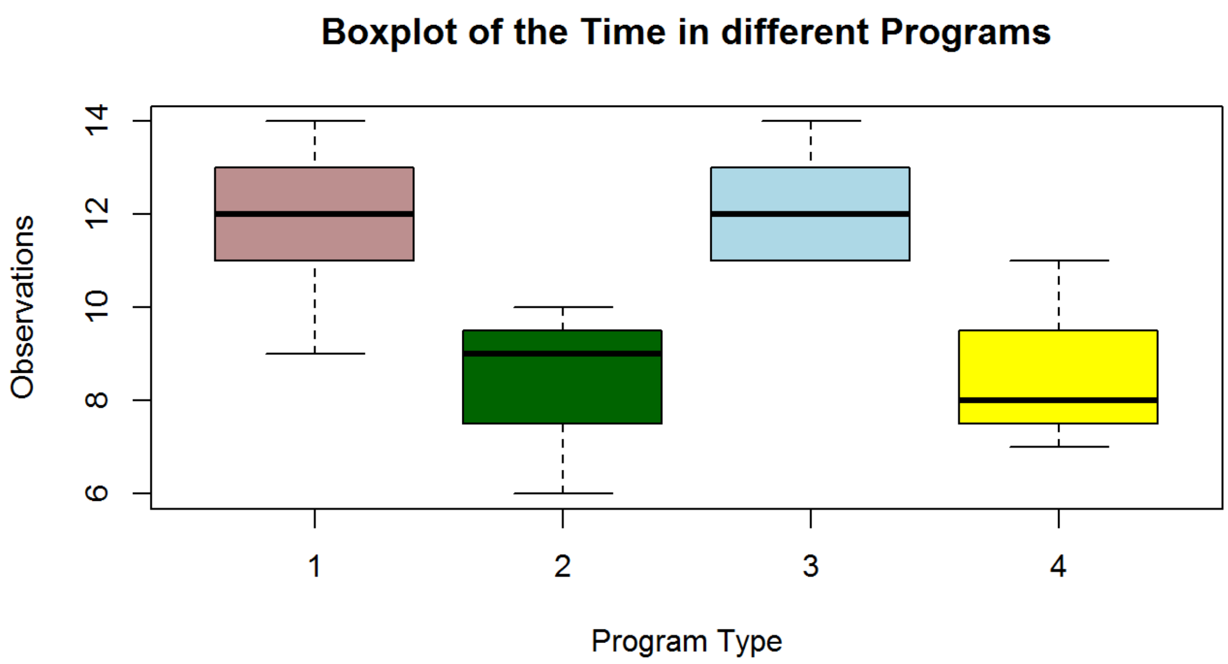
```
model <- aov(obs ~ fac, data = data)
summary(model)
```

```
            Df Sum Sq Mean Sq F value  Pr(>F)
fac          3  54.84  18.281   6.463 0.00569 **
Residuals   14  39.60   2.829
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(b) Using $\alpha = 0.05$, determine whether the treatments differ in their effectiveness.

Before going into testing of hypothesis, let us consider the boxplot to see the variation of effects in an exploratory basis.

```
boxplot(data$obs ~ data$fac, col = c("rosybrown","darkgreen","lightblue","yellow"),
        main = "Boxplot of the Time in different Programs", xlab = "Program Type", ylab = "Obser
vations")
```



Now, observe that, in the above ANOVA table, the p-value is smaller than the given test of significance level 0.05, hence, we reject the null hypothesis and conclude that the data shows evidence that the treatments differ in their effectiveness.

(c) Construct 3 independent meaningful contrasts of the effects of programs such that the S.S due to these contrasts add up to the between group contrasts.

We know that since there are 4 levels of factors i.e there are 4 types of programs, so the Sum of Squares can be expressed as sum of squares due to 3 contrasts whose BLUE's are uncorrelated. Consider the following contrasts.

1. $-\tau_1 + \tau_3$, which means difference in the effect of program 1 and program 3. The BLUE of this is $\bar{Y}_{3.} - \bar{Y}_{1.}$
2. $-\tau_2 + \tau_4$, which means the difference of effect of program 2 from the effect of program 4. The BLUE of this is $\bar{Y}_{4.} - \bar{Y}_{2.}$
3. $\tau_1 - \tau_2 + \tau_3 - \tau_4$, which means the difference of improvement from program 1 to 2 and improvement from program 3 to 4. The BLUE of this is $\bar{Y}_{1.} - \bar{Y}_{2.} + \bar{Y}_{3.} - \bar{Y}_{4.}$.

Observe that, the BLUE of first and second contrasts are uncorrelated. Now, consider, BLUEs of contrast 1 and 3. We have, $((-1 * 1)/5) + 0 + ((1 * 1)/5) + 0 = 0$, which they are uncorrelated. Similarly, considering BLUEs of contrast 2 and 3, we have $0 + ((-1 * -1)/4) + 0 + ((1 * -1)/4) = 0$, they are uncorrelated.

Question 2

Twenty-four missiles were selected from a large production batch. The missiles were randomly split into three groups of size eight. The first group of eight had engine type 1 installed, the second group had engine type 2, and the third group received engine type 3. Each group of eight was randomly divided into four groups of two. The first such group was assigned propellant type 1, the second group was assigned propellant type 2, and so on. Data on burn rate were collected, as follows:

Each one of you select a random sample of size 12 with replacement from {1,2} and select the observation corresponding to your sample from each of this cell. Thus you have a data on two way table with single observation per cell.

```
obs1 <- c(34.0, 32.00, 28.4, 30.1, 30.2, 27.3, 29.8, 28.7, 29.7, 29.0, 27.6, 28.8)
obs2 <- c(32.7, 33.2, 29.3, 32.8, 29.8, 28.9, 26.7, 28.1, 27.3, 28.9, 27.8, 29.1)
observations <- cbind(obs1, obs2)

set.seed(2018)  #set the seed for generating random number
index <- sample(1:2, 12, replace = TRUE)  #generate index to work with
print(index)  #take a look at the random numbers
```

```
 [1] 1 1 1 1 1 1 2 1 2 2 1 2
```

```
obs <- sapply(1:nrow(observations), function(i){observations[i,index[i]]})
#take the corresponding observation specified by index

missiledata <- data.frame(BurnRate = obs, EngineType = factor(rep(1:3, 4)),
                          PropType = factor(rep(1:4, each = 3)))

knitr::kable(missiledata)  #take a look at the data
```

BurnRate	EngineType	PropType
34.0	1	1

BurnRate	EngineType	PropType
32.0	2	1
28.4	3	1
30.1	1	2
30.2	2	2
27.3	3	2
26.7	1	3
28.7	2	3
27.3	3	3
28.9	1	4
27.6	2	4
29.1	3	4

Now complete the ANOVA table assuming a fixed effects two way model without interaction and carry out the appropriate test to determine whether either factor, engine type (factor A) or propellant type (factor B), has a significant effect on burn rate.

```
fit <- aov(BurnRate ~ EngineType + PropType, data = missiledata)
summary(fit)
```

```
            Df Sum Sq Mean Sq F value Pr(>F)
EngineType   2  8.347   4.173   1.507  0.295
PropType     3 24.749   8.250   2.979  0.118
Residuals    6 16.613   2.769
```

We observe that the p-value of F-statistic is not smaller than 0.05. Hence, we may conclude that the data shows no strong evidence for rejecting the null hypothesis i.e. the data shows no strong evidence that Engine Type (factor A) or Propellant Type (factor B) has any significant effect on burn rate.

THANK YOU