

# STAT430 Assignment2

Jiajun Zhang

October 11, 2019

## Question 4.8

Let  $\mu_1$  be the mean of tensile strengths for treatment 1. Similarly, we can define  $\mu_2$ ,  $\mu_3$ , and  $\mu_4$ . Then, we will test:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$H_a$  : at least one  $\mu_i$  is different than the other, where  $i \in \{1, 2, 3, 4\}$

```
chem_data <- data.frame(Chemical=factor(c(rep(1,5), rep(2,5), rep(3,5),
                                         rep(4,5))),
                        Bolt=factor(rep(1:5,4)),
                        Tensile.Strength=c(73, 68, 74, 71, 67, 73, 67,
                                           75, 72, 70, 75, 68, 78, 73,
                                           68, 73, 71, 75, 75, 69))
anova(lm(Tensile.Strength~Chemical+Bolt, data=chem_data))

## Analysis of Variance Table
##
## Response: Tensile.Strength
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Chemical    3  12.95   4.317   2.3761    0.1211
## Bolt        4 157.00  39.250  21.6055 2.059e-05 ***
## Residuals  12  21.80   1.817
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the result, we will not reject  $H_0$  for treatments since  $0.121 > 0.05$ , and conclude that there is no significant difference in means among four chemicals.

## Question 4.9

Let  $\mu_1$  be the mean of bacteria growth for washing solution 1. Similarly, we can define  $\mu_2, \mu_3$ .

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$H_a$  : at least one  $\mu_i$  is different than the other, where  $i \in \{1, 2, 3\}$

```
bacteria_data <- data.frame(Solution=factor(c(rep(1,4), rep(2,4), rep(3,4))),
                           Days=factor(rep(1:4, 3)),
                           Observations=c(13, 22, 18, 39, 16, 24, 17, 44,
                                           5, 4, 1, 22))
anova(lm(Observations~Solution+Days, data=bacteria_data))

## Analysis of Variance Table
##
## Response: Observations
##          Df  Sum Sq Mean Sq F value    Pr(>F)
## Solution   2   703.50   351.75  40.717 0.0003232 ***
## Days       3  1106.92   368.97  42.711 0.0001925 ***
## Residuals   6    51.83     8.64
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the result, we have a p-value of  $0.0003 < 0.05$ . Therefore, we will reject  $H_0$  and conclude that there is significant difference in means among three washing solutions. In order to determine which treatment is significant different than the other, we will run the LSD test.

```
library(agricolae)
LSD.test(bacteria_data$Observations, bacteria_data$Solution, DError=6,
         MSError=8.64, console=T)
```

```
##
## Study: bacteria_data$Observations ~ bacteria_data$Solution
##
## LSD t Test for bacteria_data$Observations
##
## Mean Square Error: 8.64
##
## bacteria_data$Solution, means and individual ( 95 %) CI
##
##  bacteria_data.Observations      std r      LCL      UCL Min Max
## 1                23.00 11.284207 4 19.403789 26.59621 13 39
## 2                25.25 12.996794 4 21.653789 28.84621 16 44
## 3                 8.00  9.486833 4  4.403789 11.59621  1 22
##
## Alpha: 0.05 ; DF Error: 6
## Critical Value of t: 2.446912
##
## least Significant Difference: 5.085811
##
## Treatments with the same letter are not significantly different.
##
##  bacteria_data$Observations groups
## 2                25.25      a
## 1                23.00      a
## 3                 8.00      b
```

From the LSD test, we can see that there is no significant difference between treatments 1 and 2. Alternatively, we can conclude that there is significant difference in means between washing solutions 1 & 3 and solutions 2 & 3.

## Question 4.14

(a) Let  $\mu_1$  be the mean of fuel consumption for treatment(oil) 1. Similarly, we can define  $\mu_2$  and  $\mu_3$ .

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$H_a$  : at least one  $\mu_i$  is different than the other, where  $i \in \{1, 2, 3\}$

```
truck_data <- data.frame(Oil=factor(c(rep(1,5), rep(2,5), rep(3,5))),
                          Truck=factor(rep(1:5, 3)),
                          Consumptions=c(0.500, 0.634, 0.487, 0.329, 0.512,
                                          0.535, 0.675, 0.520, 0.435, 0.540,
                                          0.513, 0.595, 0.488, 0.400, 0.510))
anova(lm(Consumptions~Oil+Truck, data=truck_data))

## Analysis of Variance Table
##
## Response: Consumptions
##           Df    Sum Sq   Mean Sq F value    Pr(>F)
## Oil        2  0.006706  0.0033529   6.3527  0.02229 *
## Truck      4  0.092100  0.0230249  43.6257 1.781e-05 ***
## Residuals  8  0.004222  0.0005278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the result, we can see that the p-value of oil is  $0.022 < 0.05$  which leads us to reject  $H_0$  and conclude that there is a significant difference in means of fuel consumptions among three lubricating oils.

(b)

```

LSD.test(y=truck_data$Consumptions, trt=truck_data$Oil, DError=8,
         MSerror=0.0005278, console=T)

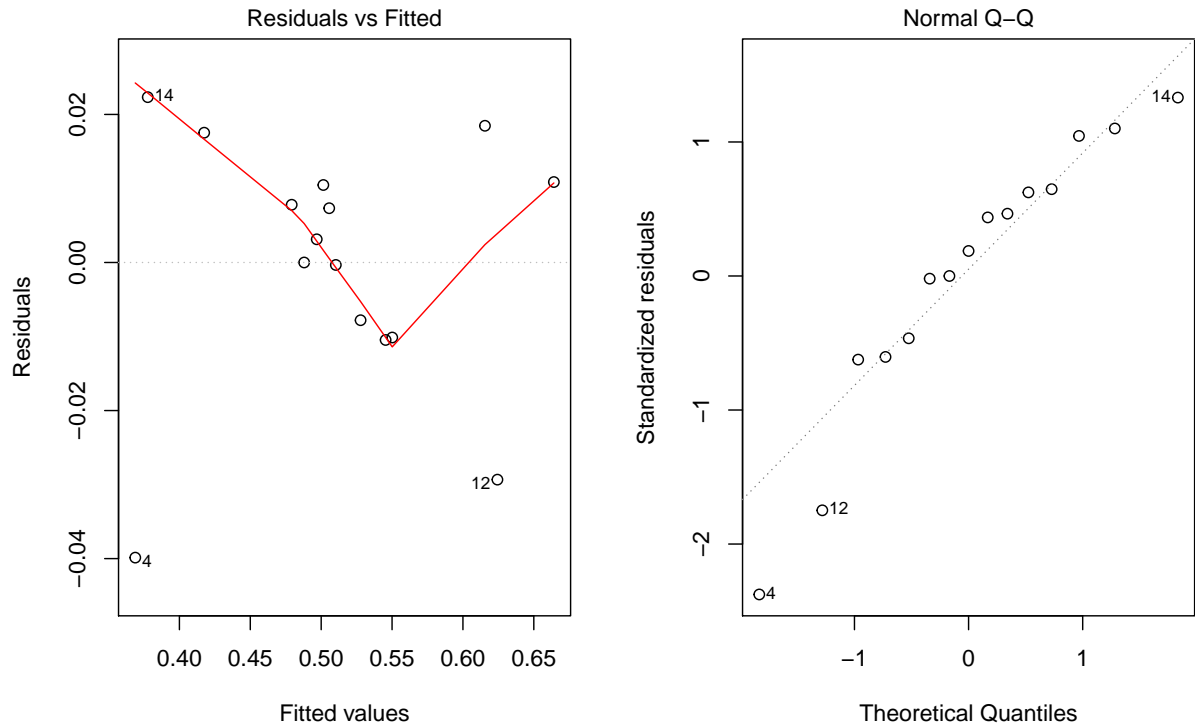
##
## Study: truck_data$Consumptions ~ truck_data$Oil
##
## LSD t Test for truck_data$Consumptions
##
## Mean Square Error: 0.0005278
##
## truck_data$Oil, means and individual ( 95 %) CI
##
##   truck_data.Consumptions      std r      LCL      UCL   Min   Max
## 1           0.4924 0.10865220 5 0.4687076 0.5160924 0.329 0.634
## 2           0.5410 0.08612491 5 0.5173076 0.5646924 0.435 0.675
## 3           0.5012 0.06969720 5 0.4775076 0.5248924 0.400 0.595
##
## Alpha: 0.05 ; DF Error: 8
## Critical Value of t: 2.306004
##
## least Significant Difference: 0.03350617
##
## Treatments with the same letter are not significantly different.
##
##   truck_data$Consumptions groups
## 2           0.5410      a
## 3           0.5012      b
## 1           0.4924      b

```

From the LSD test result, we can conclude that there are significant differences in means between type 2 & 3 oils, and type 2 & 1 oils.

(c)

```
res <- aov(lm(Consumptions~Oil+Truck, data=truck_data))
par(mfrow=c(1,2))
plot(res, 1); plot(res, 2)
```



It is hard to tell whether or not the assumptions of constant variance and normality are violated from the two graphs above. Therefore, we can perform Bartlett test and Shapiro-Wilk's test to investigate. Then, our hypothesis becomes:

$H_0$  : Variances are equal for across the treatments

$H_a$  : At least one variance is different than the other.

Similarly, we will need to test whether data follows a normal distribution under null hypothesis using Shapiro-Wilk's test.

```
#Constant Variance; H0: all variances are equal
bartlett.test(truck_data$Consumptions, truck_data$Oil)
```

```
##
## Bartlett test of homogeneity of variances
##
## data: truck_data$Consumptions and truck_data$Oil
## Bartlett's K-squared = 0.70349, df = 2, p-value = 0.7035

# H0: Data follows a normal distribution
shapiro.test(truck_data$Consumptions)

##
## Shapiro-Wilk normality test
##
## data: truck_data$Consumptions
## W = 0.95496, p-value = 0.6056
```

Since the p-values from both result are greater than 0.05, we do not reject  $H_0$  and conclude that the assumptions of constant variance and normality are not violated.



## Question 4.28

Let  $\mu_A$  be the mean assembly time for treatment A. Similarly, we can define  $\mu_B$ ,  $\mu_C$ , and  $\mu_D$ .

```
engineer_data <- data.frame(Order.Assembly=factor(c(rep(1,4), rep(2,4),
                                                    rep(3,4), rep(4,4))),
                             Operator=factor(rep(1:4, 4)),
                             Methods=factor(c("C", "D", "A", "B", "B", "C",
                                                "D", "A", "A", "B", "C", "D",
                                                "D", "A", "B", "C")),
                             Time=c(10, 14, 7, 8, 7, 18, 11, 8, 5, 10, 11, 9,
                                     10, 10, 12, 14))

anova(lm(Time~Order.Assembly+Operator+Methods, data=engineer_data))

## Analysis of Variance Table
##
## Response: Time
##
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Order.Assembly  3    18.5   6.1667   3.5238 0.088519 .
## Operator        3    51.5  17.1667   9.8095 0.009926 **
## Methods         3    72.5  24.1667  13.8095 0.004213 **
## Residuals       6    10.5   1.7500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since we obtained a p-value of treatments  $0.004 < 0.05$ , we will reject  $H_0$  and conclude that there is a evidence to indicate that the mean assembly time is significantly different among assembly methods. To specify which treatment is different than the other, we can perform a LSD test.

```

LSD.test(engineer_data$Time, engineer_data$Methods, DError=6,
          MSerror=1.75, console=T)

##
## Study: engineer_data$Time ~ engineer_data$Methods
##
## LSD t Test for engineer_data$Time
##
## Mean Square Error: 1.75
##
## engineer_data$Methods, means and individual ( 95 %) CI
##
##   engineer_data.Time      std r      LCL      UCL Min Max
## A           7.50 2.081666 4  5.88152  9.11848   5  10
## B           9.25 2.217356 4  7.63152 10.86848   7  12
## C          13.25 3.593976 4 11.63152 14.86848  10  18
## D          11.00 2.160247 4  9.38152 12.61848   9  14
##
## Alpha: 0.05 ; DF Error: 6
## Critical Value of t: 2.446912
##
## least Significant Difference: 2.288876
##
## Treatments with the same letter are not significantly different.
##
##   engineer_data$Time groups
## C           13.25      a
## D           11.00     ab
## B           9.25     bc
## A           7.50      c

```

Based on the LSD test result, we will conclude that there are significant differences between methods A & C, A & D and B & C.

## Question 4.29

```
bacteria_data <- data.frame(Solution=factor(c(rep(1,4), rep(2,4), rep(3,4))),
                           Days=factor(rep(1:4, 3)),
                           Observations=c(13, 22, 18, 39, 16, 24, 17, 44,
                                           5, 4, 1, 22))

library(lme4)

fit.rand <- lmer(Observations~Solution+(1|Days), data=bacteria_data, REML=T)

summary(fit.rand)

## Linear mixed model fit by REML ['lmerMod']
## Formula: Observations ~ Solution + (1 | Days)
## Data: bacteria_data
##
## REML criterion at convergence: 60.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -0.93800 -0.64423 -0.04722  0.39940  1.44360
##
## Random effects:
## Groups   Name      Variance Std.Dev.
## Days     (Intercept) 120.111  10.960
## Residual                    8.639   2.939
## Number of obs: 12, groups: Days, 4
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    23.000     5.673   4.054
## Solution2       2.250     2.078   1.083
```

```
## Solution3      -15.000      2.078  -7.217
##
## Correlation of Fixed Effects:
##           (Intr) Soltn2
## Solution2 -0.183
## Solution3 -0.183  0.500
```

Now, the blocks(Days) are random, then the estimate block variance component is 120.11 as we obtained from the result.

## Question 4.41

By introducing a fourth factor, we denote them as  $a = \beta$ ,  $b = \gamma$ ,  $c = \delta$ , and  $d = \alpha$ . As previous stated, we are interested in testing the differences among assembly methods (treatments). Let  $\mu_A$  be the mean of assembly time for assembly method A. Similarly, we can define  $\mu_B$ ,  $\mu_C$ , and  $\mu_D$ . Then, we will test:

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D$$

$H_a$  : at least one  $\mu_i$  is different than the other, where  $i \in \{A, B, C, D\}$

```
engineer_data <- data.frame(Order.Assembly=factor(c(rep(1,4), rep(2,4),
                                                    rep(3,4), rep(4,4))),
                           Operator=factor(rep(1:4, 4)),
                           Methods=factor(c("C", "B", "D", "A",
                                             "B", "C", "A", "D",
                                             "A", "D", "B", "C",
                                             "D", "A", "C", "B")),
                           Workplace=factor(c("a", "b", "c", "d",
                                                "d", "c", "b", "a",
                                                "c", "d", "a", "b",
                                                "b", "a", "d", "c")),
                           Time=c(11, 10, 14, 8, 8, 12, 10, 12, 9,
                                   11, 7, 15, 9, 8, 18, 6))

anova(lm(Time~Order.Assembly+Operator+Methods+Workplace, data=engineer_data))

## Analysis of Variance Table
##
## Response: Time
##              Df Sum Sq Mean Sq F value Pr(>F)
## Order.Assembly  3    0.5   0.167   0.0182 0.9960
## Operator        3   19.0   6.333   0.6909 0.6157
## Methods         3   95.5  31.833   3.4727 0.1669
```

## Workplace	3	7.5	2.500	0.2727	0.8429
## Residuals	3	27.5	9.167		

From the result, we have a p-value of 0.167 on treatments  $> 0.05$ . Therefore, we will reject  $H_0$  and conclude that there is no significant difference in means among assembly methods.

## Question 4.45

Let  $\mu_1$  be the mean mileage for gasoline additive type 1. Similarly, we can define  $\mu_2, \mu_3, \mu_4, \mu_5$ .

```
gas_data <- data.frame(Additive=factor(c(rep(1,4), rep(2,4), rep(3,4),
                                         rep(4,4), rep(5,4))),
                      Car=factor(c(2, 3, 4, 5, 1, 2, 4, 5, 1, 3, 4,
                                     5, 1, 2, 3, 4, 1, 2, 3, 5)),
                      Mileage=c(17, 14, 13, 12, 14, 14, 13, 10, 12,
                                13, 12, 9, 13, 11, 11, 12, 11, 12,
                                10, 8))

fit <- lm(Mileage~Additive+Car, data=gas_data)
library(car)
Anova(fit, type=2)

## Anova Table (Type II tests)
##
## Response: Mileage
##           Sum Sq Df F value    Pr(>F)
## Additive  35.733  4  9.8103 0.001247 **
## Car       35.233  4  9.6730 0.001321 **
## Residuals 10.017 11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the result, we have the p-value on treatments is  $0.0012 < 0.05$ . Therefore, we reject  $H_0$  and conclude that there is a significant difference in means among the types of gasoline additives.



```

LSD.test(gas_data$Mileage, gas_data$Additive, DError=11,
         MSerror=10.017/11, console=T)

##
## Study: gas_data$Mileage ~ gas_data$Additive
##
## LSD t Test for gas_data$Mileage
##
## Mean Square Error: 0.9106364
##
## gas_data$Additive, means and individual ( 95 %) CI
##
##   gas_data.Mileage      std r      LCL      UCL Min Max
## 1          14.00 2.1602469 4 12.94983 15.05017 12 17
## 2          12.75 1.8929694 4 11.69983 13.80017 10 14
## 3          11.50 1.7320508 4 10.44983 12.55017 9 13
## 4          11.75 0.9574271 4 10.69983 12.80017 11 13
## 5          10.25 1.7078251 4 9.19983 11.30017 8 12
##
## Alpha: 0.05 ; DF Error: 11
## Critical Value of t: 2.200985
##
## least Significant Difference: 1.485165
##
## Treatments with the same letter are not significantly different.
##
##   gas_data$Mileage groups
## 1          14.00      a
## 2          12.75     ab
## 4          11.75      b

```

## 3	11.50	bc
## 5	10.25	c

Based on the LSD test result, we will conclude that there is significant differences between gas additives type 1 & 3, 1 & 4, 1 & 5, 2 & 5, and 4 & 5.

## Question 4.50

Let  $a$  be the number of treatments,  $b$  be the number of blocks,  $k$  be the number of treatments in each block,  $r$  be the number of times each treatment occurs in the design.

Now, we have  $a=4$ ,  $b=6$ ,  $k=2$ . And based on the formula,

$$\lambda = \frac{r(k-1)}{a-1}$$

Let  $\lambda=1$ , we will have  $r=3$ . That means each treatment occurs exactly three times in the design. Therefore, the design can be constructed as below.

```
matrix(c("x", "x", "x", "", "", "",
        "x", "", "", "x", "x", "",
        "", "x", "", "x", "", "x",
        "", "", "x", "", "x", "x"),
       nrow=4, ncol=6, byrow=T, dimnames=list(c("Treatment1", "Treatment2",
        "Treatment3", "Treatment4"),
        c("Block1", "Block2", "Block3",
        "Block4", "Block5", "Block6")))
```

	Block1	Block2	Block3	Block4	Block5	Block6
Treatment1	"x"	"x"	"x"	" "	" "	" "
Treatment2	"x"	" "	" "	"x"	"x"	" "
Treatment3	" "	"x"	" "	"x"	" "	"x"
Treatment4	" "	" "	"x"	" "	"x"	"x"

## Question 4.54

Recall the formula,

$$\lambda = \frac{r(k-1)}{a-1}$$

Now, we have  $a=8$ ,  $r=8$ ,  $k=4$ . Therefore,

$$\lambda = \frac{8(4-1)}{8-1} = \frac{24}{7}$$

Since  $\lambda$  is not integer, therefore we conclude that the BIBD with provided parameters does not exist.