Homework3

Evangeline Lim

10/16/2019

Question 9

We can include a quadratic trend term because if the plot of residuals against time exhibits a quadratic trend, the error variance of y first increases then decreases as time duration increases. This suggests a time trend or autocorrelation over time.

Question 11

(a) A one-way layout should be used. Let each of four drugs experiment on 5 participants who are selected and distributed randomly. Randomize the order of screening as well. Check if there are differences among drugs in treating the illness.

(b)

```
## [1] 8.983264
```

```
## [1] 4.63657
```

```
## [1] 0.0002976389
```

F-statistic is 8.983. P-value in this test is 0.003. At the 0.01 level, the critical value is equal to 4.637, which is less than the F-statistic. Hence, we reject the null that four treatments have the same bioactivity.

(c)

```
## [1] 3.440076
```

```
## [1] -0.437439
```

```
## [1] -4.453904
```

```
## [1] -2.615993
```

```
## [1] -4.153339
```

```
## [1] -2.27568
```

```
## [1] 1.497311
```

The Tukey critical value at the .01 level is 3.44. The t-statistics for comparisons between A&C and B&C are larger than 3.44. This means that A and D, and, B and C have different bioactivity at 0.01.

(d)

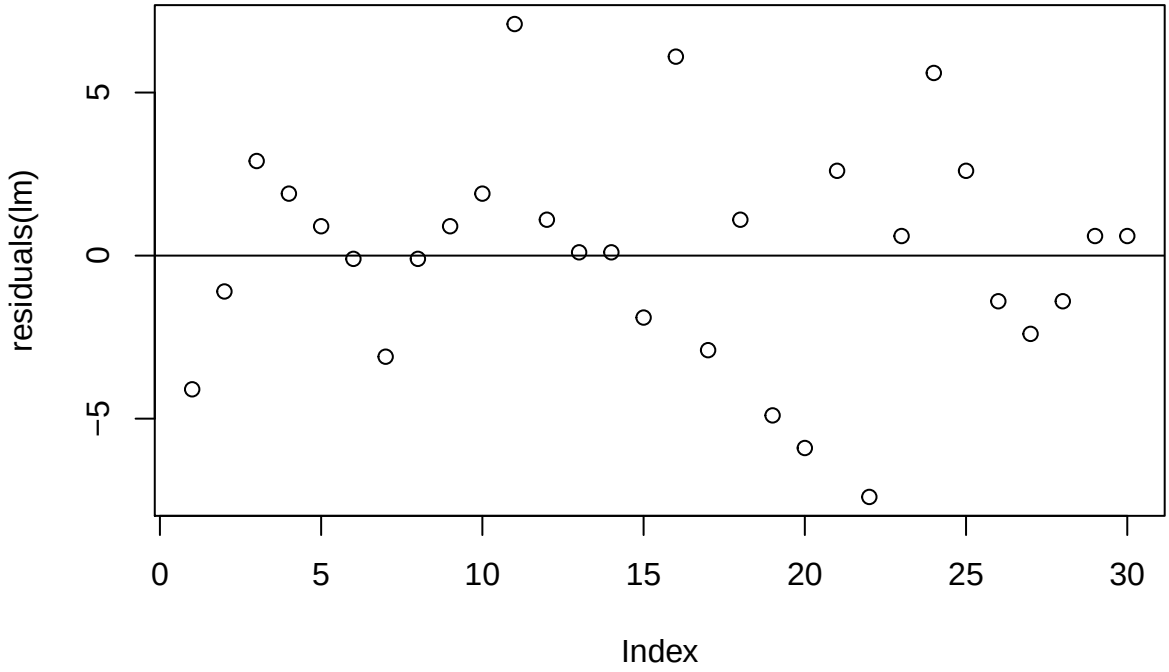
```
## [1] 6.739054
```

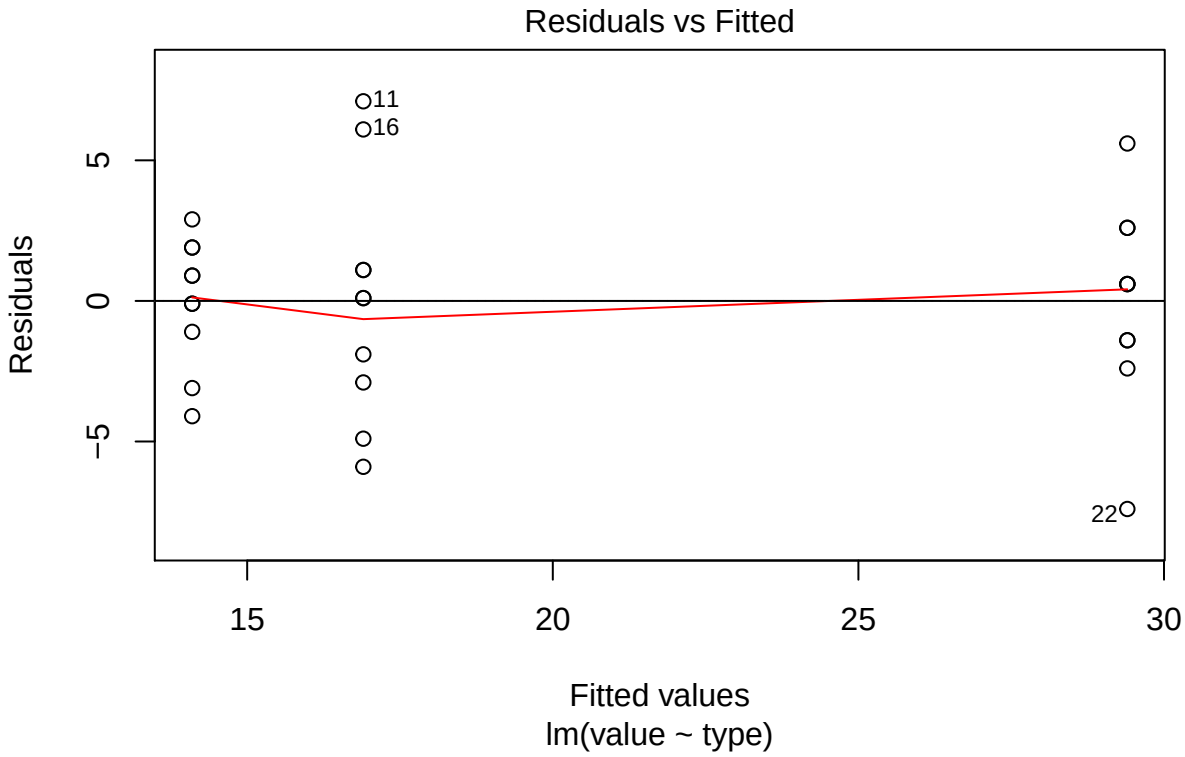
```
## [1] 1.881272e-07
```

The p-value is 1.88e-07, which is significantly small at the 0.01 level. We reject the null hypothesis that the brand-name and generic drugs have same bioactivity.

Question 12

```
## Analysis of Variance Table
##
## Response: value
##           Df Sum Sq Mean Sq F value    Pr(>F)
## type       2 1327.3  663.63  56.667 2.171e-10 ***
## Residuals 27  316.2   11.71
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```






```
## [1] 1.829636
```

```
## [1] 9.997651
```

```
## [1] 8.168016
```

```
## [1] 2.552459
```

```
## [1] 2.479418
```

The Bonferroni method gives a critical value of 2.552. The P&O & HT group and the C&W & HT group are both different at the 0.05 level. Also, the Tukey method gives a critical value of 2.479 and produces the same results as the Bonferroni method.

Question 14

(a)

```
## [1] "0.016" "0.030" "0.044" "0.058"
```

```
## Analysis of Variance Table
##
## Response: Velocity
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Area       3  13515  4505.0  7.0666 0.0003813 ***
## Residuals 60  38250   637.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] 7.066667
```

```
## [1] 0.0003813011
```

```
## [1] 2.728552
## [1] 2.642521
## [1] -0.8905781
## [1] -3.424385
## [1] -3.829066
## [1] -2.533807
## [1] -2.938488
## [1] -0.4046809
```

The one-way anova gives an F-statistic of 7.067 and a p-value of 0.0004, which is significant at the 0.05 level to reject the null that muzzle velocity is the same for these 4 different discharge hole areas and conclude that at least one pair show significant difference.

The critical value given by Bonferroni method is 2.729. According to the Bonferroni method, the 0.016 & 0.044 group, 0.016 & 0.058 group, and 0.03 & 0.058 group are different at level 0.05.

The critical value derived with Tukey method is 2.643. It gives the same results as the Bonferroni method.

Question 16

```
## Analysis of Variance Table
##
## Response: weight
##           Df Sum Sq Mean Sq F value    Pr(>F)
## machines   2 31.652 15.8262  59.508 1.107e-14 ***
## Residuals 57 15.159  0.2659
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] 1.107126e-14
```

According to the ANOVA table, the F-statistic is equal to 59.508. The p-value is 1.107e-14, which is significantly small to reject the null that three machines have the same performance.

(b) From (a) we can conclude that at least one pair has different performance.

(c)

```
## Analysis of Variance Table
##
## Response: weight
##           Df Sum Sq Mean Sq F value    Pr(>F)
## machines   2 31.652 15.8262  59.508 1.107e-14 ***
## Residuals 57 15.159  0.2659
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] 1.107126e-14
## [1] 0.2659
## [1] 0.778015
```

We apply the same ANOVA and F test as in the fixed effects case, which give an F-statistic of 59.508. So, we reject the null that the variance among operators is equal to zero at the 0.05 level.

The variance components are estimated to be 0.2659 and 0.778.

(d)

```
## [1] 47.89172 52.31128
```

The 95% confidence interval is (47.892, 52.311).

Question 17

(a)

```
## Analysis of Variance Table
##
## Response: bp
##           Df Sum Sq Mean Sq F value Pr(>F)
## dev        2    0.0   0.007   1e-04 0.9999
## Residuals 42 3699.4  88.082
```

```
## [1] 1e-04
```

```
## [1] 88.082
```

```
## [1] -5.871667
```

```
## Analysis of Variance Table
##
## Response: BP
##           Df Sum Sq Mean Sq F value    Pr(>F)
## DOC        2 496.33 248.163   139.1 < 2.2e-16 ***
## Residuals 42  74.93   1.784
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] 139.1
```

```
## [1] 1.784
```

```
## [1] 16.42527
```

For devices, the F statistic is equal to 1e-04 with a p-value close to 1, thus we cannot reject the null which states that the variance among operators in the population is equal to zero. The variance components are 88.082 and 0 respectively.

For doctors, the F statistic is equal to 139.1 with a p-value close to 0, thus we reject the null which states that the variance among operators in the population is equal to zero. The variance components are 1.784 and 16.425 respectively.

(b) For devices, we cannot reject the null, thus we conclude that the variance among devices might be 0. For doctors, we reject the null, thus we conclude that the variance among at least on pair of doctors are different.

```
# Devices
bpbar = sum(blood$bp)/45
t = qt(0.025, 2)
MSTr4 = 0.007
n4 = 15
```

```
k = 3
lower_dev = bpbar + t*sqrt(MSTr4/(n4*k))
upper_dev = bpbar - t*sqrt(MSTr4/(n4*k))
(c(lower_dev, upper_dev))

# Doctors
BPbar = sum(blood2$BP)/45
t = qt(0.025, 2)
MSTr5 = 248.163
n5 = 15
k = 3
lower_doc = BPbar + t*sqrt(MSTr5/(n5*k))
upper_doc = BPbar - t*sqrt(MSTr5/(n5*k))
(c(lower_doc, upper_doc))
```

The 95% confidence interval for the mean blood pressure measured by devices is [128.0128, 128.1201];

The 95% confidence interval for the mean blood pressure measured by doctors is [119.9172, 140.1255].

Appendix

```
MSR = 21.47
MSE = 2.39
f = MSR/MSE
f
qf(0.01, 3, 26, lower.tail = F)
pf(f, 3, 26, lower.tail = F)
```

```
qc = qtukey(p = 0.99, nmeans = 4, df = 26)/sqrt(2)
qc
tAB = (65.75-66.10)/(sqrt(MSE)*sqrt((1/8) + (1/7)))
tAB
tAC = (62.63-66.10)/(sqrt(MSE)*sqrt((1/9) + (1/7)))
tAC
tAD = (63.85-66.10)/(sqrt(MSE)*sqrt((1/6) + (1/7)))
tAD
tBC = (62.63-65.75)/(sqrt(MSE)*sqrt((1/9) + (1/8)))
tBC
tBD = (63.85-65.75)/(sqrt(MSE)*sqrt((1/6) + (1/8)))
tBD
tCD = (63.85-62.63)/(sqrt(MSE)*sqrt((1/6) + (1/9)))
tCD
```

```
tcontrast = ((66.1+65.75)/2  -(62.62+63.85)/2)/(sqrt(MSE)*sqrt((1/(7+8)/2+1/(9+6)/2)))
tcontrast #formula??
pt(tcontrast, 26, lower.tail = F)
```

```
data = read.csv("bolt.csv")
data$type = factor(data$type, labels = c("P&O", "C&W", "HT"))
lm = lm(value~type, data = data)
anova(lm)

plot(residuals(lm))
```

```
abline(0,0)
plot(lm, which = c(1,1))
abline(0,0)

onebar = sum(data$value[1:10])/10
twobar = sum(data$value[11:20])/10
threebar = sum(data$value[21:30])/10
s_e = sqrt(11.71)
t12 = (twobar - onebar)/(s_e*sqrt(1/10 + 1/10))
t12
t13 = (threebar - onebar)/(s_e*sqrt(1/10 + 1/10))
t13
t23 = (threebar - twobar)/(s_e*sqrt(1/10 + 1/10))
t23

# Bonferroni
tcr = qt(0.05/(2*3), 27, lower.tail = F) # 3 = 3 choose 2
tcr
# Tukey
qcr = qtukey(p = 0.95, nmeans = 3, df = 27)/sqrt(2)
qcr
```

```
data2 = read.csv("onewaymuzzle.csv")
data2$Area = factor(data2$Area, labels = c("0.016", "0.030", "0.044",  "0.058"))
levels(data2$Area)
lm2 = lm(Velocity ~ Area, data = data2)
anova(lm2)
f2 = 4505/637.5
f2
pf(f2, 3, 60, lower.tail = F)
# Bonferroni
tcrit = qt(0.05/(2*6), 60, lower.tail = F) # 6 = 4 choose 2
tcrit
# Tukey
qcrit = qtukey(p = 0.95, nmeans = 4, df = 60)/sqrt(2)
qcrit
# Multiple Comparisons
a = sum(data2$Velocity[1:16])/16
b = sum(data2$Velocity[17:32])/16
c = sum(data2$Velocity[33:48])/16
d = sum(data2$Velocity[49:64])/16
se2 = sqrt(637.5)
na = nb = nc = nd = 16
tab = (b - a)/(se2*(sqrt(1/na + 1/nb)))
tab
tac = (c - a)/(se2*(sqrt(1/na + 1/nc)))
tac
tad = (d - a)/(se2*(sqrt(1/na + 1/nd)))
tad
tbc = (c - b)/(se2*(sqrt(1/nb + 1/nc)))
tbc
tbd = (d - b)/(se2*(sqrt(1/nb + 1/nd)))
tbd
tcd = (d - c)/(se2*(sqrt(1/nc + 1/nd)))
```

```
tcd
```

```
pack = read.csv("PackingMachine.csv")
pack$machines = factor(pack$machines, labels = c("MC1", "MC2", "MC3"))
lm3 = lm(weight ~ machines, data = pack)
anova(lm3)
pf(59.508, 2, 57, lower.tail = F)
```

```
anova(lm3)
pf(59.508, 2, 57, lower.tail = F)
va_hat = 0.2659
va_hat
va_o_hat = (15.8262 - 0.2659)/20
va_o_hat
```

```
## [1] 47.89172 52.31128
```

```
# Devices
blood = read.csv("BloodPressure.csv")
blood$dev = factor(blood$dev, labels = c("Dev1", "Dev2", "Dev3"))
lm4 = lm(bp~dev, data = blood)
anova(lm4)
(f4 = 1e-04)
(va_hat4 = 88.082)
(va_o_hat4 = (0.007-88.082)/15) #less than 0 -> =0

# Doctors
blood2 = read.csv("BloodPressure2.csv")
blood2$DOC = factor(blood2$DOC, labels = c("Doc1", "Doc2", "Doc3"))
lm5 = lm(BP~DOC, data = blood2)
anova(lm5)
(f5 = 139.1)
(va_hat5 = 1.784)
(va_o_hat5 = (248.163 - 1.784)/15)
```

```
# Devices
bpbar = sum(blood$bp)/45
t = qt(0.025, 2)
MSTr4 = 0.007
n4 = 15
k = 3
lower_dev = bpbar + t*sqrt(MSTr4/(n4*k))
upper_dev = bpbar - t*sqrt(MSTr4/(n4*k))
(c(lower_dev, upper_dev))

# Doctors
BPbar = sum(blood2$BP)/45
t = qt(0.025, 2)
MSTr5 = 248.163
n5 = 15
k = 3
lower_doc = BPbar + t*sqrt(MSTr5/(n5*k))
upper_doc = BPbar - t*sqrt(MSTr5/(n5*k))
(c(lower_doc, upper_doc))
```