

Second Component SS3843 Final Project

Luchang Liu 251107314

2021/12/12

Q1

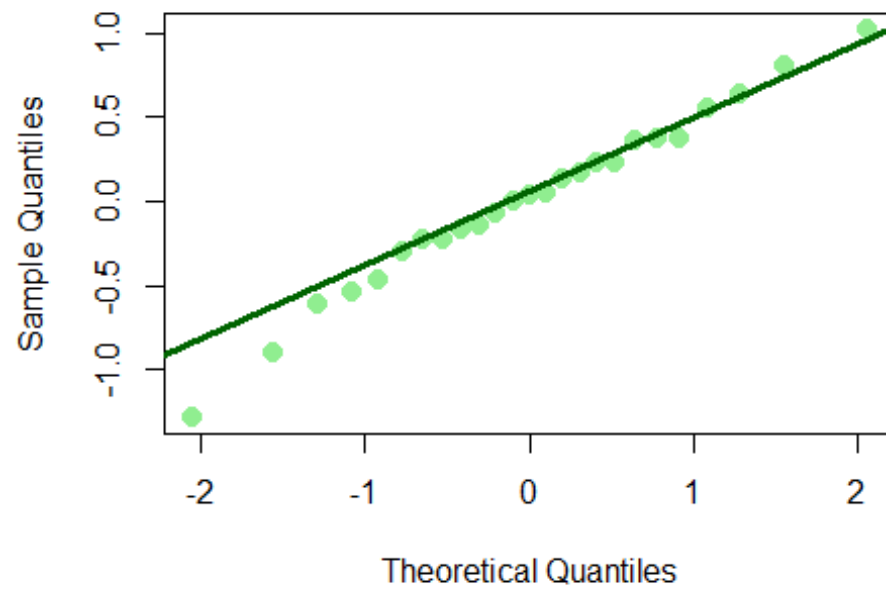
```
# b.
y = c(8.04, 6.61, 11.99, 7.78, 8.40,
      9.58, 6.58, 6.66, 5.34, 7.92,
      7.98, 7.98, 8.98, 7.94, 11.32,
      9.74, 9.46, 9.14, 12.00, 9.32,
      9.66, 11.28, 8.04, 8.12, 6.72)
yogurts = factor(c('M', 'H', 'N', 'Y', 'C',
                   'N', 'Y', 'C', 'M', 'H',
                   'Y', 'M', 'H', 'C', 'N',
                   'H', 'C', 'M', 'N', 'Y',
                   'C', 'N', 'Y', 'H', 'M'))
batches = factor(c(rep(1:5, each = 5)))
positions = factor(c(rep(1:5, times = 5)))

d1 = data.frame(batches, positions, yogurts, y)

q1 = lm(y ~ yogurts + batches + positions)
resi1 = resid(q1)

qqnorm(resi1, col = "lightgreen", pch = 20, cex = 2)
qqline(resi1, col = "darkgreen", lwd = 3)
```

Normal Q-Q Plot



Comment:

From the QQ-plot, majority points are very close to the line, which means the residuals follow a normal distribution. Therefore, normality assumption holds.

Q2

e.

```
cool.zone = rep(c(5,10,15,20,25), times=6)
```

```
preheat = rep(c(5,25), each=15)
```

```
y = c(6.45, 2.51, 4.71, 11.47, 10.69,  
      2.71, 4.42, 4.91, 9.31, 9.25,  
      3.08, 4.20, 8.19, 12.04, 10.00,  
      2.86, 5.43, 5.35, 8.92, 13.13,  
      5.51, 8.37, 4.20, 7.57, 12.01,  
      5.66, 3.54, 6.49, 7.14, 11.71)
```

```
d2 = data.frame(cool.zone, preheat, y)
```

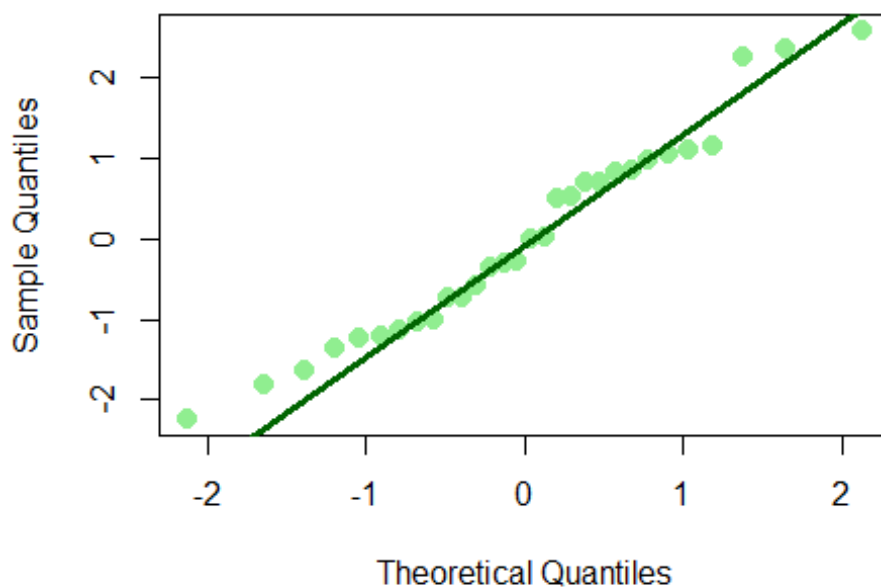
```
q2 = lm(y ~ factor(cool.zone) * factor(preheat))
```

```
resi2 = resid(q2)
```

```
qqnorm(resi2, col = "lightgreen", pch = 20, cex = 2)
```

```
qqline(resi2, col = "darkgreen", lwd = 3)
```

Normal Q-Q Plot



Comment

From the QQ-plot, points in left and right parts of the plot are very far from the line, which means the residuals do not follow a normal distribution. Therefore, normal assumption is violated.

```

# f.
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.1.2

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.1.2

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

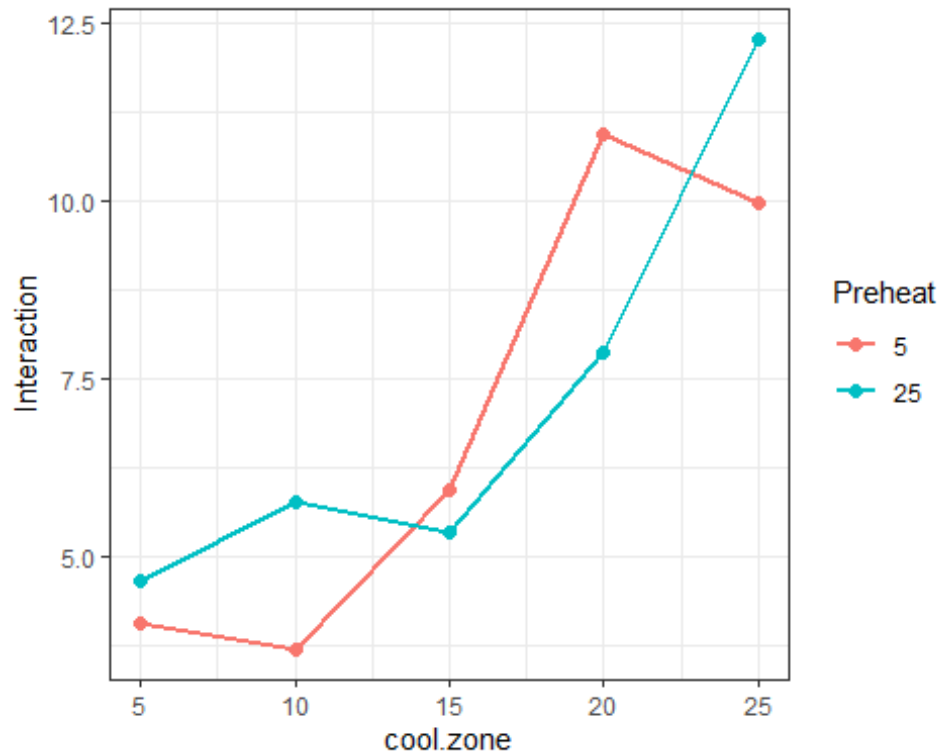
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

means = d2 %>% group_by(cool.zone, preheat) %>% summarise(Mean=mean(y))

## `summarise()` has grouped output by 'cool.zone'. You can override using the `.groups` argument.

ggplot(means, aes(x = cool.zone, y = Mean, colour = factor(preheat))) +
  geom_line(size=1) + geom_point(size=2) +
  labs(colour="Preheat", y="Interaction") +
  theme_bw()

```



Comment

From the plot, we can see that both levels of preheat show an increasing trend with different slopes, which means there is a significant difference between preheat level and cool-zone level. Furthermore, from the plot we also can see that the larger cool-zone level is, the larger interaction will be.

```

# g.
lm.q2 = lm(y ~ factor(cool.zone) + factor(preheat))
anova(lm.q2)

## Analysis of Variance Table
##
## Response: y
##
##           Df Sum Sq Mean Sq F value    Pr(>F)
## factor(cool.zone)  4 219.932   54.983  17.6477 7.086e-07 ***
## factor(preheat)    1   0.520    0.520   0.1669   0.6865
## Residuals         24  74.774    3.116
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Q4

$$4(a) Y_{ijkl} = \mu + T_i + B_j + A_k + Y_l + (TB)_{ij} + (TA)_{ik} + (TY)_{il} + (BA)_{jk} + (BY)_{jl} + (AY)_{kl} + (TBA)_{ijk} + (TBY)_{ijl} + (BAY)_{ikl} + (TBAY)_{ijkl} + E_{ijkl}$$

Y_{ijkl} is the observed hardness of the gold fillings on i th dentist (our blocking variable), i th condensation filling method, k th sintered temperature and l th gold alloy.

T_i is the effect of the i th dentist, $i = 1, \dots, 5$

B_j is the effect of the j th method, $j = 1, \dots, 3$

A_k is the effect of k th temperature, $k = 1, \dots, 3$

Y_l is the effect of l th alloy, $l = 1, 2$

E_{ijkl} is the random error.

$(TB)_{ij}$ is the interaction between the i th dentist and j th method

$(TA)_{ik}$ is the interaction between the i th dentist and k th temperature

$(TY)_{il}$ is the interaction between the i th dentist and l th alloy

$(BA)_{jk}$ is the interaction between the j th method and k th temperature

$(BY)_{jl}$ is the interaction between the j th method and l th alloy

$(AY)_{kl}$ is the interaction between the k th temperature and l th alloy

$(TBA)_{ijk}$ is the interaction between the i th dentist + j th method + k th temperature

$(TBY)_{ijl}$ is the interaction between the i th dentist + j th method + l th alloy

$(BAY)_{ikl}$ is the interaction between the j th method + k th temperature + l th alloy

$(TBAY)_{ijkl}$ is the interaction between the i th dentist + j th method, k th temperature, and l th alloy

$$\sum(T_i) = 0$$

$$\sum(B_j) = 0$$

$$\sum(A_k) = 0$$

$$\sum(Y_l) = 0$$

$$\sum((TB)_{ij}) = 0$$

$$\sum((TA)_{ik}) = 0$$

$$\sum((TY)_{il}) = 0$$

$$\text{sum}((BA)_{jk}) = 0$$

$$\text{sum}((BY)_{jl}) = 0$$

$$\text{sum}((AY)_{kl}) = 0$$

$$\text{sum}((TBA)_{ijk}) = 0$$

$$\text{sum}((TBY)_{ijl}) = 0$$

$$\text{sum}((BAY)_{ikl}) = 0$$

$$\text{sum}((TBAY)_{ijkl}) = 0$$

$$4(b) \text{ df}_{\text{dentist}} = 5-1 = 4$$

$$\text{df}_{\text{method}} = 3-1 = 2$$

$$\text{df}_{\text{alloy}} = 2-1 = 1$$

$$\text{df}_{\text{temperature}} = 3-1 = 2$$

$$\text{df}_{\text{dentist, method}} = 4 \times 2 = 8$$

$$\text{df}_{\text{dentist, temperature}} = 4 \times 2 = 8$$

$$\text{df}_{\text{dentist, alloy}} = 4 \times 1 = 4$$

$$\text{df}_{\text{method, temperature}} = 2 \times 2 = 4$$

$$\text{df}_{\text{method, alloy}} = 2 \times 1 = 2$$

$$\text{df}_{\text{temperature, alloy}} = 2 \times 1 = 2$$

$$\text{df}_{\text{dentist, method, temperature}} = 4 \times 2 \times 2 = 16$$

$$\text{df}_{\text{dentist, method, alloy}} = 4 \times 2 \times 1 = 8$$

$$\text{df}_{\text{dentist, temperature, alloy}} = 4 \times 2 \times 1 = 8$$

$$\text{df}_{\text{method, temperature, alloy}} = 2 \times 2 \times 1 = 4$$

$$\text{df}_{\text{dentist, method, temperature and alloy}} = 4 \times 2 \times 2 \times 1 = 16 \quad \text{df}_{\text{total}} = N-1 = 90-1 = 89$$
$$\text{df}_{\text{error}} = 0$$

Q5

#a

```
y = c(41.2, 41.2, 39.8, 41.5, 41.9, 45.5,
      42.6, 41.4, 40.3, 43.0, 42.7, 44.7,
      135.7, 143.0, 132.4, 134.4, 137.4, 141.1,
      136.8, 143.3, 130.3, 130.0, 135.2, 139.1,
      163.2, 181.4, 173.6, 174.9, 166.6, 175.0,
      163.3, 180.3, 173.9, 175.6, 165.5, 172.0)
con = factor(rep(c(1:3), each=12))
day = factor(rep(c(1:3), each=2, times=6))
run = factor(rep(c(1:6), times=6))

library(lme4)

## Loading required package: Matrix

model = lmer(y ~ con + (1|day) + (1|day:run) + (1|con:day) + (1|con:day:
run))

summary(model)

## Linear mixed model fit by REML ['lmerMod']
## Formula:
## y ~ con + (1 | day) + (1 | day:run) + (1 | con:day) + (1 | con:day:r
un)
##
## REML criterion at convergence: 163.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.97752 -0.33411 -0.02637  0.42400  1.69412
##
## Random effects:
##   Groups             Name             Variance  Std.Dev.
## con:day:run (Intercept) 1.711e+01  4.1366339
## con:day      (Intercept) 7.217e-05  0.0084951
## day:run      (Intercept) 3.657e+00  1.9124384
## day          (Intercept) 1.072e-07  0.0003274
## Residual                    1.436e+00  1.1983739
## Number of obs: 36, groups:  con:day:run, 18; con:day, 9; day:run, 6;
day, 3
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   42.150      1.892    22.27
## con2           94.408      2.438    38.73
## con3          129.958      2.438    53.31
##
## Correlation of Fixed Effects:
##      (Intr) con2
```

```
## con2 -0.644
## con3 -0.644  0.500

#b
#Test 1: concentration effect
fullmodel1 = lmer(y ~ con + (1|day) + (1|day:run) + (1|con:day) + (1|con:day:run), REML = FALSE)

## boundary (singular) fit: see ?isSingular

reducedmodel1 = lmer(y ~ 1 + (1|day) + (1|day:run) + (1|con:day) + (1|con:day:run), REML = FALSE)

#LRT
D1 = as.numeric(2*(logLik(fullmodel1)-logLik(reducedmodel1)))
pvalue1 = pchisq(D1, df=(3-1), lower.tail = FALSE)
pvalue1

## [1] 1.467227e-12
```

Comment

Because the p-value = 1.467227e-12 is smaller than 0.1,0.05,0.01, there is strong evidence to reject H_0 , so the concentration effect is not 0.

```
#Test 2: day effect
fullmodel2 = lmer(y ~ con + (1|day) + (1|day:run) + (1|con:day) + (1|con:day:run))
reducedmodel2 = lmer(y ~ con + (1|day:run) + (1|con:day) + (1|con:day:run))

#LRT
D2 = as.numeric(2*(logLik(fullmodel2)-logLik(reducedmodel2)))
pvalue2 = pchisq(D2, df=(3-1), lower.tail = FALSE)
pvalue2

## [1] 1
```

Comment

Because p_value = 1 is greater than 0.001,0.05,0.1, the data have very weak evidence against H_0 , so the day effect is 0.

```
#Test 3: run effect
fullmodel3 = lmer(y ~ con + (1|day) + (1|day:run) + (1|con:day) + (1|con:day:run))
reducedmodel3 = lmer(y ~ con + (1|day) + (1|con:day) + (1|con:day:run))

## boundary (singular) fit: see ?isSingular

# LRT
D3 = as.numeric(2*(logLik(fullmodel3)-logLik(reducedmodel3)))
pvalue3 = pchisq(D3, df=(2-1), lower.tail = FALSE)
pvalue3
```

```
## [1] 0.5267768
```

Comment

Because $p_value = 0.5267768$ is greater than 0.001, 0.05, 0.1, the data have very weak evidence against H_0 , so run effect is 0.

```
#Test 4: interaction effect between concentration and day
fullmodel4 = lmer(y ~ con + (1|day) + (1|day:run) + (1|con:day) + (1|con:day:run))
reducedmodel4 = lmer(y ~ con + (1|day) + (1|day:run) + (1|con:day:run))

## boundary (singular) fit: see ?isSingular

# LRT
D4 = as.numeric(2*(logLik(fullmodel4)-logLik(reducedmodel4)))
pvalue4 = pchisq(D4, df=(3-1)*(3-1), lower.tail = FALSE)
pvalue4

## [1] 1
```

Comment

Because $p_value = 1$ is greater than 0.001, 0.05, 0.1, the data have very weak evidence against H_0 , so no interaction effect between concentration and day.

```
#Test 5: interaction effect between concentration day run
fullmodel5 = lmer(y ~ con + (1|day) + (1|day:run) + (1|con:day) + (1|con:day:run))
reducedmodel5 = lmer(y ~ con + (1|day) + (1|day:run) + (1|con:day))

## boundary (singular) fit: see ?isSingular

# LRT
D5 <- as.numeric(2*(logLik(fullmodel5)-logLik(reducedmodel5)))
pvalue5 <- pchisq(D5, df=(3-1)*(3-1)*(2-1), lower.tail = FALSE)
pvalue5

## [1] 4.586637e-05
```

Comment

Because $p_value = 4.586637e-05$ is smaller than 0.1, 0.05, 0.01, there is strong evidence to reject H_0 , so the interaction effect is not 0.