

STAT 407 Homework 3

Charles Hwang

10/20/2021

Problem 5.6

(a)

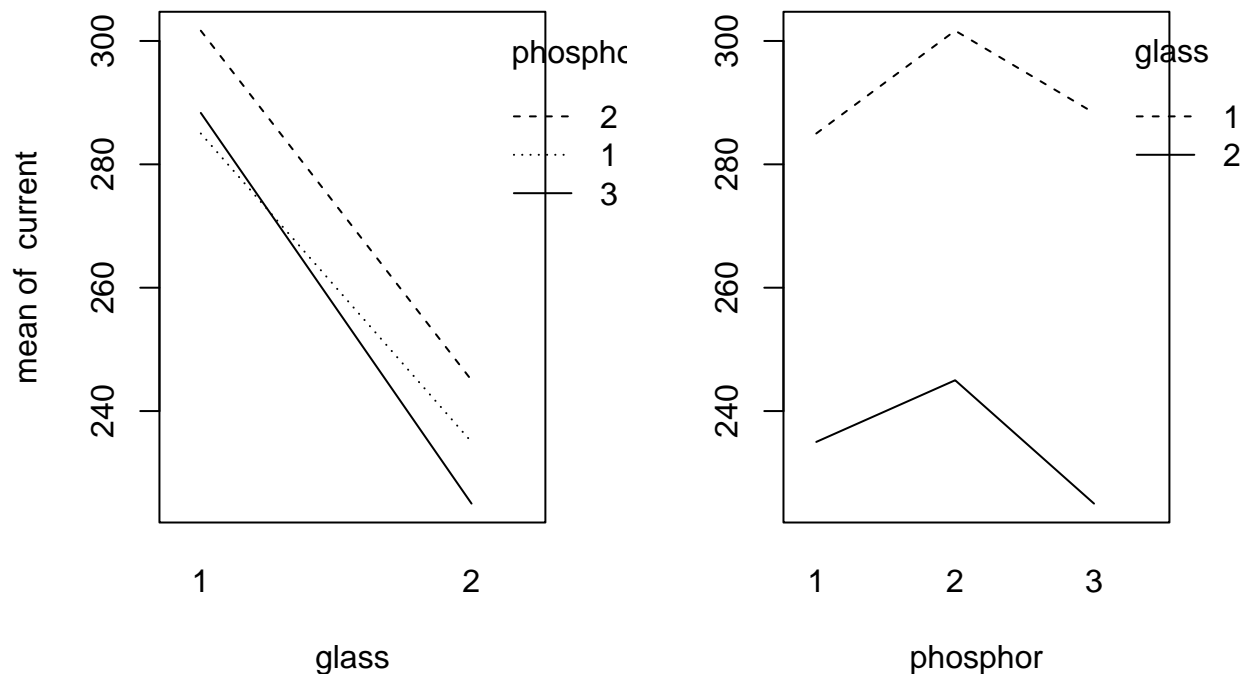
```
rm(list=ls())
current<-c(280,300,290,290,310,285,285,295,290,230,260,220,235,240,225,240,235,230)
glass<-as.factor(rep(1:2,each=9))
phosphor<-as.factor(rep(1:3,times=6))
t<-lm(current~glass*phosphor)
anova(t)
```

```
## Analysis of Variance Table
##
## Response: current
##           Df Sum Sq Mean Sq F value    Pr(>F)
## glass       1 14450.0  14450.0  273.7895 1.259e-09 ***
## phosphor    2   933.3    466.7    8.8421 0.004364 **
## glass:phosphor 2   133.3     66.7    1.2632 0.317801
## Residuals   12   633.3     52.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We reject both null hypotheses at the $\alpha = 0.05$ level. There is sufficient evidence ($p < 0.000001$, $p = 0.004364$) that both factors influence brightness.

(b)

```
par(mfrow=c(1,2))
interaction.plot(glass,phosphor,current)
interaction.plot(phosphor,glass,current,ylab="")
```



There is interaction between phosphor types 1 and 3. There may be slight interaction # between phosphor type 2 and the other two phosphor types.

```
anova(t)["glass:phosphor", "Pr(>F)"]
```

```
## [1] 0.3178005
```

We can see the overall interaction is not significant ($p = 0.3178005$). The two factors do interact, but only at the levels of phosphor types 1 and 3.

(c)

```
shapiro.test(t$residuals)
```

```
##
## Shapiro-Wilk normality test
##
## data:  t$residuals
## W = 0.9525, p-value = 0.4655
```

```
bartlett.test(t$residuals~glass)
```

```
##
## Bartlett test of homogeneity of variances
##
## data:  t$residuals by glass
## Bartlett's K-squared = 1.4693, df = 1, p-value = 0.2255
```

```
bartlett.test(t$residuals~phosphor)
```

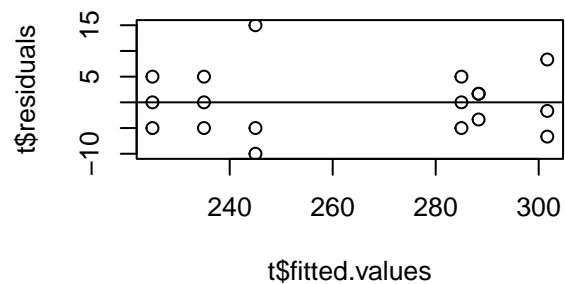
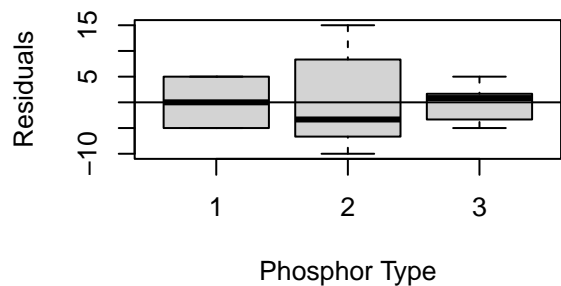
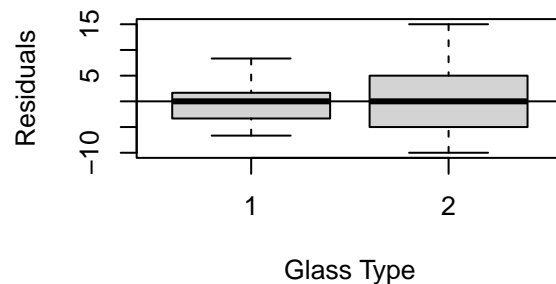
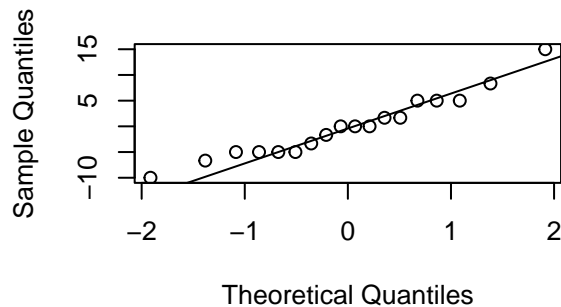
```
##
## Bartlett test of homogeneity of variances
##
## data:  t$residuals by phosphor
## Bartlett's K-squared = 5.0816, df = 2, p-value = 0.0788
```

```
library(car)
leveneTest(lm(t))

## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 5    0.57  0.722
##      12

# None of the null hypotheses were rejected, so the normality and
par(mfrow=c(2,2)) # equal variance assumptions appear to be met.
qqnorm(t$residuals)
qqline(t$residuals)
plot(c(glass),t$residuals,xlab="Glass Type",ylab="Residuals")
abline(h=0)
plot(c(phosphor),t$residuals,xlab="Phosphor Type",ylab="Residuals")
abline(h=0)
plot(t$fitted.values,t$residuals)
abline(h=0)
```

Normal Q-Q Plot

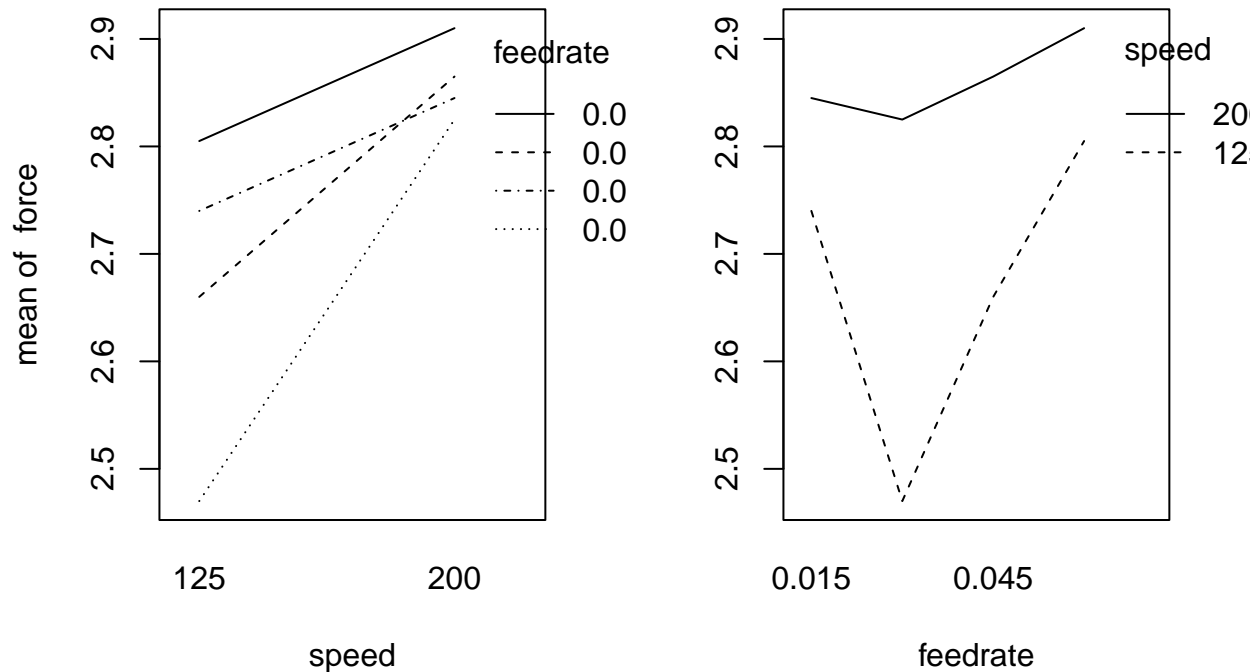


*# There is a slight variation in the Q-Q plot, but there does not appear to be any pattern
in the residual plots.*

Problem 5.9

```
force<-c(2.70,2.45,2.60,2.75,2.78,2.49,2.72,2.86,2.83,2.85,2.86,2.94,2.86,2.80,2.87,2.88)
speed<-as.factor(rep(c(125,200),each=8))
feedrate<-as.factor(rep(c(0.015,0.03,0.045,0.06),times=4))
par(mfrow=c(1,2))
interaction.plot(speed,feedrate,force)
```

```
interaction.plot(feedrate,speed,force,ylab="")
```



We can see that all of the levels have some degree of interaction with each other.

```
d<-lm(force~speed+feedrate+speed*feedrate)
anova(d)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: force
```

```
##          Df    Sum Sq  Mean Sq F value    Pr(>F)
## speed      1  0.148225   0.148225  57.0096 6.605e-05 ***
## feedrate   3  0.092500   0.030833  11.8590  0.002582 **
## speed:feedrate 3  0.041875   0.013958   5.3686  0.025567 *
## Residuals  8  0.020800   0.002600
```

```
## ---
```

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

We reject both null hypotheses at the $\alpha = 0.05$ level. There is sufficient evidence ($p = 0.000066$, $p = 0.002582$) that both factors influence thrust force.

```
shapiro.test(d$residuals)
```

```
##
```

```
##  Shapiro-Wilk normality test
```

```
##
```

```
## data:  d$residuals
```

```
## W = 0.96815, p-value = 0.8078
```

```
bartlett.test(d$residuals~speed)
```

```
##
```

```
##  Bartlett test of homogeneity of variances
```

```
##
```

```
## data:  d$residuals by speed
```

```
## Bartlett's K-squared = 3.7165, df = 1, p-value = 0.05388
```

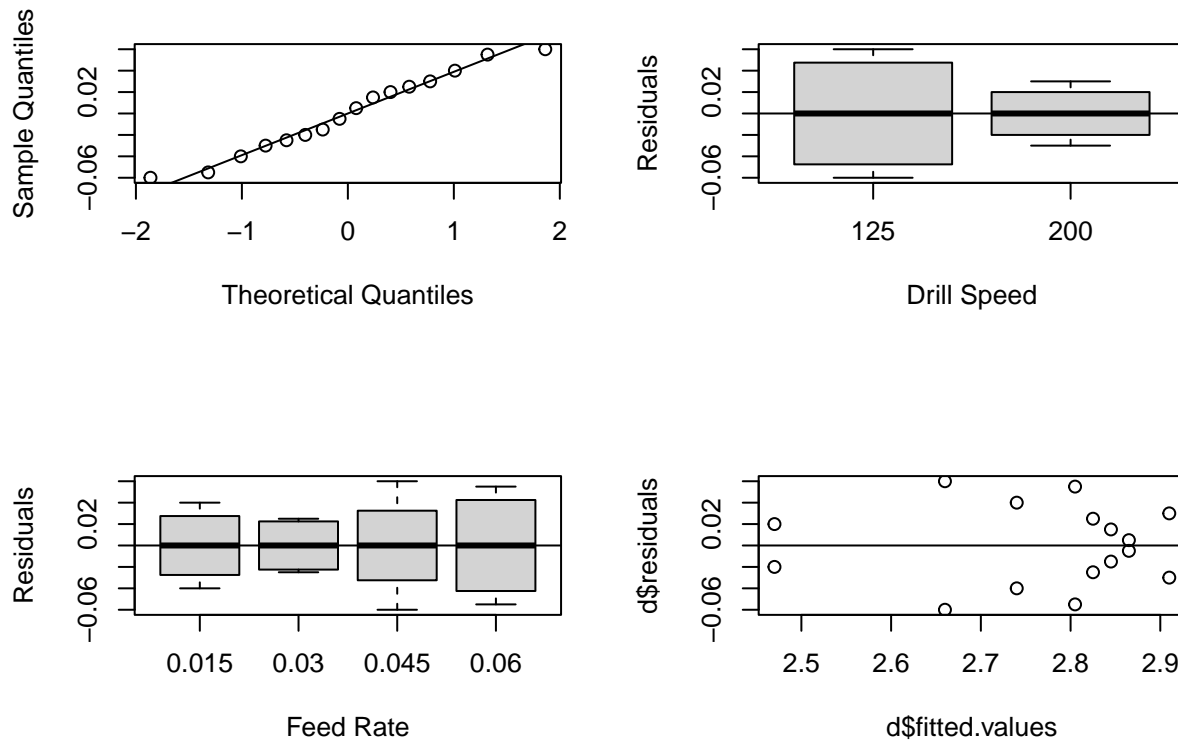
```

bartlett.test(d$residuals~feedrate)

##
## Bartlett test of homogeneity of variances
##
## data: d$residuals by feedrate
## Bartlett's K-squared = 1.4239, df = 3, p-value = 0.6999
# None of the null hypotheses were rejected, so the normality and
par(mfrow=c(2,2)) # equal variance assumptions appear to be met.
qqnorm(d$residuals)
qqline(d$residuals)
plot(c(speed),d$residuals,xlab="Drill Speed",ylab="Residuals")
abline(h=0)
plot(c(feedrate),d$residuals,xlab="Feed Rate",ylab="Residuals")
abline(h=0)
plot(d$fitted.values,d$residuals)
abline(h=0)

```

Normal Q-Q Plot



There is a slight variation in the Q-Q plot and a slight pattern in the residuals vs. fitted values plot.

```

TukeyHSD(aov(force~speed+feedrate+speed*feedrate))$speed # Post-hoc analysis

```

```

##           diff          lwr          upr          p adj
## 200-125 0.1925 0.1337082 0.2512918 6.605525e-05

```

```

TukeyHSD(aov(force~speed+feedrate+speed*feedrate))$feedrate

```

```

##           diff          lwr          upr          p adj
## 0.03-0.015 -0.145 -0.2604624436 -0.02953756 0.016179049
## 0.045-0.015 -0.030 -0.1454624436 0.08546244 0.838049193

```

```
## 0.06-0.015    0.065 -0.0504624436  0.18046244 0.338536209
## 0.045-0.03    0.115 -0.0004624436  0.23046244 0.050910663
## 0.06-0.03     0.210  0.0945375564  0.32546244 0.001766307
## 0.06-0.045    0.095 -0.0204624436  0.21046244 0.111326674
```

```
# The two drill speeds are clearly different. The 0.03 feed rate appears to be different
# from the 0.015 feed rate and the 0.06 feed rate. The 0.03 feed rate may also be
# different from the 0.045 feed rate, but we narrowly failed to reject the null
# hypothesis at the alpha = 0.05 level (p = 0.050910663).
```

Problem 5.12

```
yield<-c(90.4,90.7,90.2,90.2,90.6,90.4,90.1,90.5,89.9,90.3,90.6,90.1,90.5,90.8,90.4,90.7,90.9,90.1)
tempr<-as.factor(rep(c(150,160,170),each=6))
pressure<-as.factor(rep(c(200,215,230),times=6))
TukeyHSD(aov(yield~tempr+pressure+tempr*pressure))$pressure
```

```
##           diff          lwr          upr          p adj
## 215-200  0.3166667  0.1017380  0.53159536 0.0066518030
## 230-200 -0.1833333 -0.3982620  0.03159536 0.0944905140
## 230-215 -0.5000000 -0.7149287 -0.28507131 0.0002951232
```

The 215 psig level appears to be different from both the 200 psig level and the 230 psig level.

Problem 5.13

Model without interaction: $y_{ijk} = \mu = \tau_i + \beta_j + \epsilon_{ijk}$ (no $(\tau\beta)_{ij}$ term)

i = 1, 2; j = 1, 2, 3; k = 1, 2, 3

```
density<-c(570,1063,565,565,1080,510,583,1043,590,528,988,526,547,1026,538,521,1004,532)
position<-as.factor(rep(1:2,each=9))
tempc<-as.factor(rep(c(800,825,850),times=6))
a<-lm(density~position+tempc)
anova(a)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: density
```

```
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## position   1   7160     7160   16.197 0.001254 **
## tempc       2  945342  472671 1069.257 4.924e-16 ***
## Residuals 14   6189       442
```

```
## ---
```

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

```
# We reject both null hypotheses at the alpha = 0.05 level. There is sufficient
# evidence (p = 0.001254, p < 0.000001) that both factors influence baked density.
shapiro.test(a$residuals)
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data:  a$residuals
```

```
## W = 0.91252, p-value = 0.09529
```

```
bartlett.test(a$residuals~position)
```

```
##
## Bartlett test of homogeneity of variances
##
## data: a$residuals by position
## Bartlett's K-squared = 2.0896, df = 1, p-value = 0.1483
```

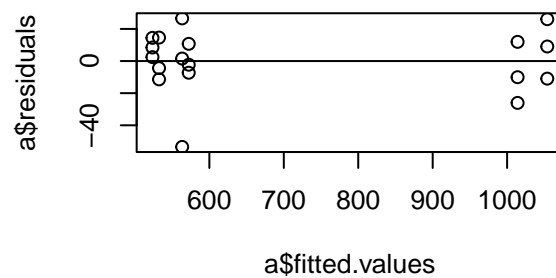
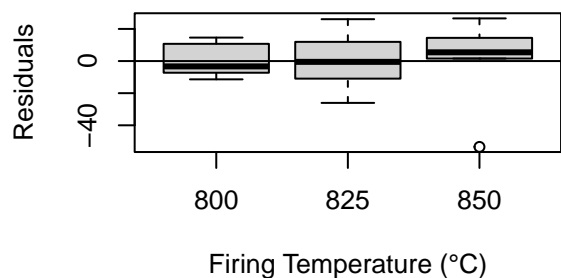
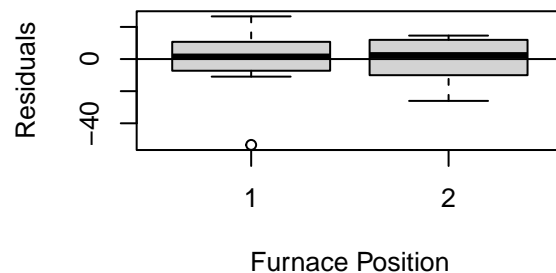
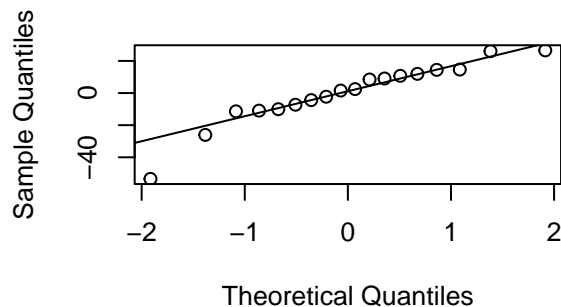
```
bartlett.test(a$residuals~tempc)
```

```
##
## Bartlett test of homogeneity of variances
##
## data: a$residuals by tempc
## Bartlett's K-squared = 3.9449, df = 2, p-value = 0.1391
```

None of the null hypotheses were rejected, but we should exercise slight caution as the # Shapiro-Wilk normality test is close to the alpha = 0.05 significance level (p = 0.09529).

```
par(mfrow=c(2,2))
qqnorm(a$residuals)
qqline(a$residuals)
plot(c(position),a$residuals,xlab="Furnace Position",ylab="Residuals")
abline(h=0)
plot(c(tempc),a$residuals,xlab="Firing Temperature (°C)",ylab="Residuals")
abline(h=0)
plot(a$fitted.values,a$residuals)
abline(h=0)
```

Normal Q-Q Plot



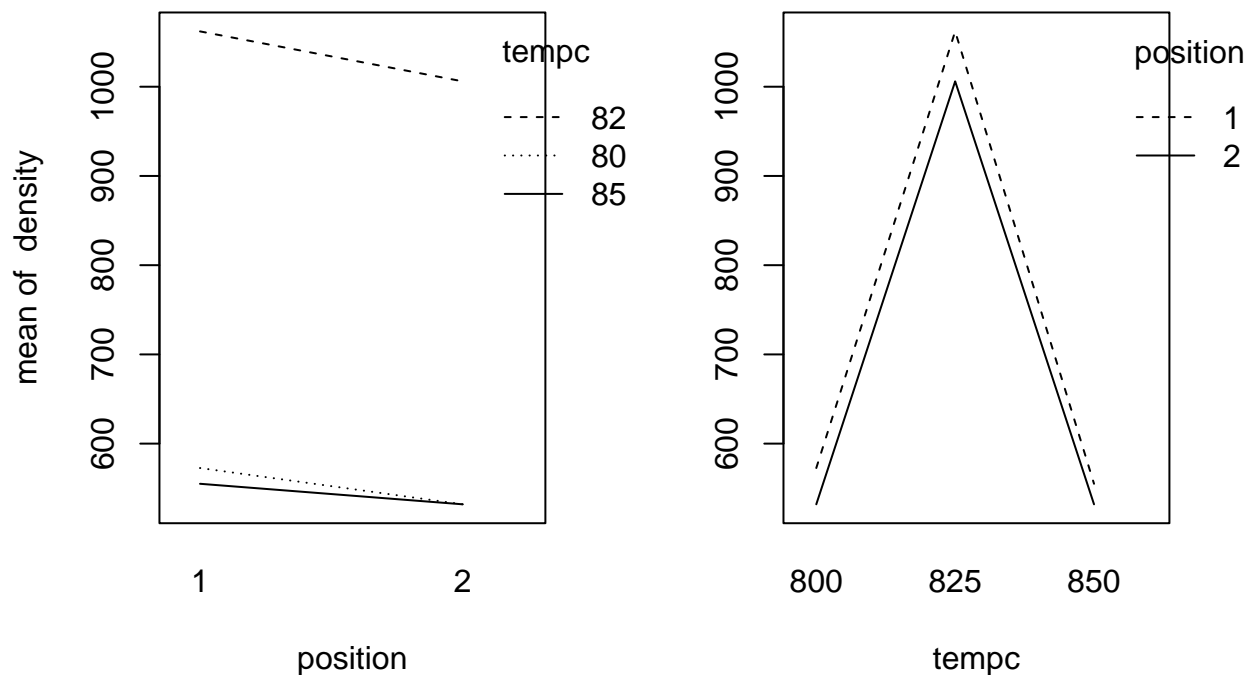
There is a slight variation in the Q-Q plot, but not much pattern in the residual plots. TukeyHSD(aov(density~position+tempc)) # Post-hoc analysis

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = density ~ position + tempc)
##
## $position
##      diff      lwr      upr    p adj
## 2-1 -39.88889 -61.14659 -18.63119 0.0012542
##
## $tempc
##      diff      lwr      upr    p adj
## 825-800 481.666667 449.89587 513.43746 0.0000000
## 850-800 -8.833333 -40.60413 22.93746 0.7515016
## 850-825 -490.500000 -522.27079 -458.72921 0.0000000
```

We can see that the two furnace positions are different. The 825C level is also clearly different from both the 800C level and the 850C level.

“Q”

```
par(mfrow=c(1,2))
interaction.plot(position,tempc,density)
interaction.plot(tempc,position,density,ylab="")
```



```
anova(lm(density~position+tempc+position*tempc))["position:tempc", "Pr(>F)"]
```

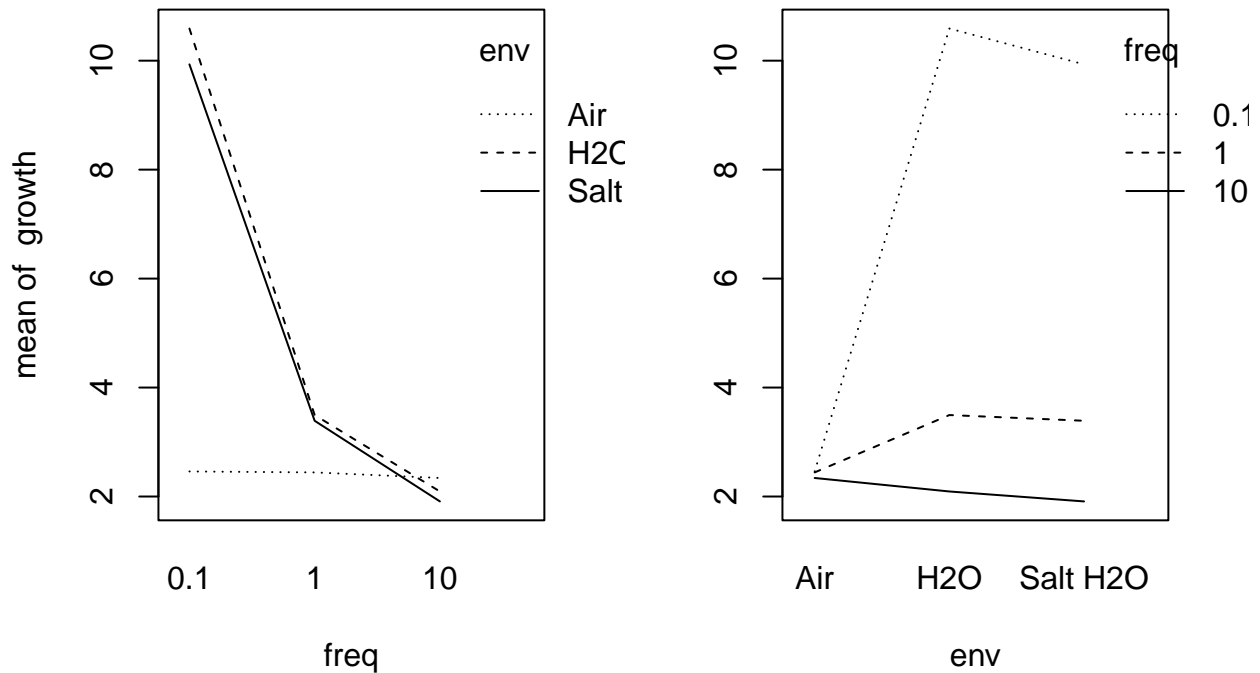
```
## [1] 0.4271101
```

Looking at the interaction plots, there does not appear to be any interaction. The assumption of a lack of interaction appears to have been made in good faith and should be considered valid. As such, the model as written appears to be adequate.

Problem 5.24

(a)

```
growth<-c(2.29,2.06,1.90,2.47,2.05,1.93,2.48,2.23,1.75,2.12,2.03,2.06,2.65,3.20,3.10,2.68,3.18,3.24,2.06)
freq<-as.factor(rep(c(10,1,0.1),each=12))
env<-as.factor(rep(c("Air","H2O","Salt H2O"),times=12))
par(mfrow=c(1,2))
interaction.plot(freq,env,growth)
interaction.plot(env,freq,growth,ylab="")
```



We can see that all of the levels have some degree of interaction with each other,
although the interaction between H2O and salt H2O appears to be slight.

```
c<-lm(growth~freq*env)
anova(c)
```

Analysis of Variance Table

##

```
## Response: growth
```

##	Df	Sum Sq	Mean Sq	F value	Pr(>F)
## freq	2	209.893	104.946	522.40	< 2.2e-16 ***
## env	2	64.252	32.126	159.92	1.076e-15 ***
## freq:env	4	101.966	25.491	126.89	< 2.2e-16 ***
## Residuals	27	5.424	0.201		

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

We reject both null hypotheses at the alpha = 0.05 level. There is sufficient
evidence ($p < 0.000001$, $p < 0.000001$) that both factors influence crack growth rate.

(b)

```
shapiro.test(c$residuals)
```

##

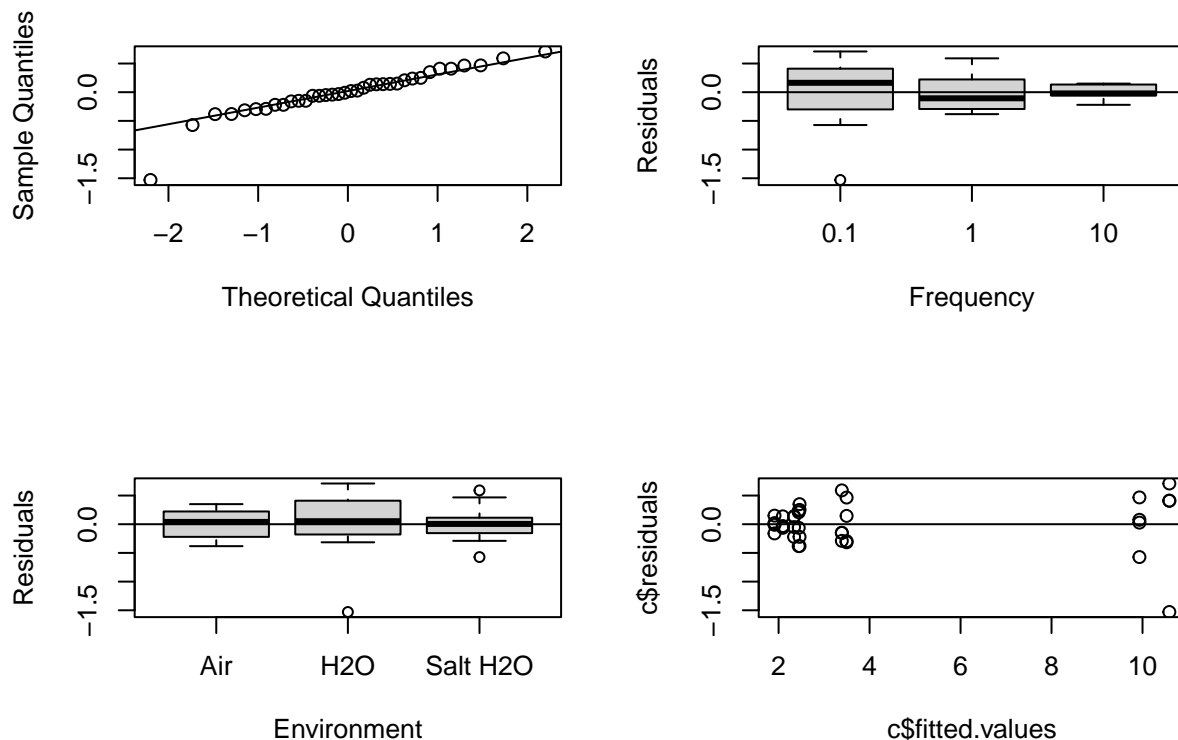
```
## Shapiro-Wilk normality test
```

```
##
## data: c$residuals
## W = 0.8949, p-value = 0.002483
bartlett.test(c$residuals~freq)

##
## Bartlett test of homogeneity of variances
##
## data: c$residuals by freq
## Bartlett's K-squared = 21.62, df = 2, p-value = 2.019e-05
bartlett.test(c$residuals~env)

##
## Bartlett test of homogeneity of variances
##
## data: c$residuals by env
## Bartlett's K-squared = 7.8992, df = 2, p-value = 0.01926
# All of the null hypotheses were rejected. The normality and equal variance assumptions
par(mfrow=c(2,2)) # have clearly been violated and this analysis should not be used.
qqnorm(c$residuals)
qqline(c$residuals)
plot(c(freq),c$residuals,xlab="Frequency",ylab="Residuals")
abline(h=0)
plot(c(env),c$residuals,xlab="Environment",ylab="Residuals")
abline(h=0)
plot(c$fitted.values,c$residuals)
abline(h=0)
```

Normal Q-Q Plot



```
# There is a slight variation in the Q-Q plot and there appears to be a megaphone pattern
# in the residuals vs. fitted values plot. This is further evidence that this analysis
# should not be used.
```

```
(c)
```

```
lngrowth<-log(growth)
lnc<-lm(lngrowth~freq+env+freq*env)
anova(lnc)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: lngrowth
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## freq         2  7.5702   3.7851  404.095 < 2.2e-16 ***
## env          2  2.3576   1.1788  125.849 2.061e-14 ***
## freq:env      4  3.5284   0.8821   94.172 1.885e-15 ***
## Residuals    27  0.2529   0.0094
```

```
## ---
```

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

```
# We reject both null hypotheses at the alpha = 0.05 level. There is sufficient
# evidence (p < 0.000001, p < 0.000001) that both factors influence crack growth rate.
shapiro.test(lnc$residuals)
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: lnc$residuals
```

```
## W = 0.9782, p-value = 0.6842
```

```
bartlett.test(lnc$residuals~freq)
```

```
##
```

```
## Bartlett test of homogeneity of variances
```

```
##
```

```
## data: lnc$residuals by freq
```

```
## Bartlett's K-squared = 3.798, df = 2, p-value = 0.1497
```

```
bartlett.test(lnc$residuals~env)
```

```
##
```

```
## Bartlett test of homogeneity of variances
```

```
##
```

```
## data: lnc$residuals by env
```

```
## Bartlett's K-squared = 1.8086, df = 2, p-value = 0.4048
```

```
leveneTest(lm(c))
```

```
## Levene's Test for Homogeneity of Variance (center = median)
```

```
##           Df F value Pr(>F)
```

```
## group      8  0.7096 0.6808
```

```
##           27
```

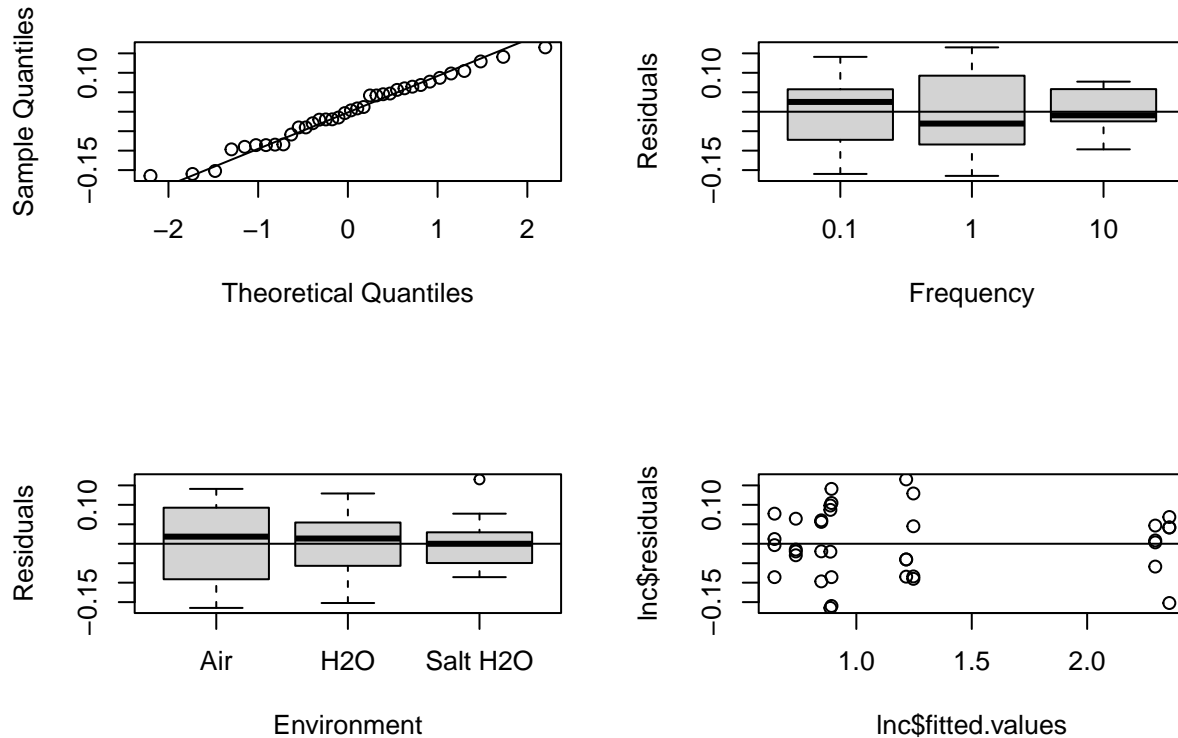
```
# None of the null hypotheses were rejected, so the normality and
par(mfrow=c(2,2)) # equal variance assumptions appear to be met.
```

```
qqnorm(lnc$residuals)
```

```
qqline(lnc$residuals)
```

```
plot(c(freq),lnc$residuals,xlab="Frequency",ylab="Residuals")
abline(h=0)
plot(c(env),lnc$residuals,xlab="Environment",ylab="Residuals")
abline(h=0)
plot(lnc$fitted.values,lnc$residuals)
abline(h=0)
```

Normal Q-Q Plot



There is a slight variation in the Q-Q plot, but the intensity of the megaphone pattern in the residuals vs. fitted values plot appears to have decreased.
 TukeyHSD(aov(lngrowth~freq+env+freq*env))\$freq *# Post-hoc analysis*

```
##           diff          lwr          upr          p adj
## 1-0.1 -0.7309649 -0.8289301 -0.6329997 4.440892e-15
## 10-0.1 -1.1040940 -1.2020593 -1.0061288 4.218847e-15
## 10-1 -0.3731291 -0.4710944 -0.2751639 1.417139e-09
```

```
TukeyHSD(aov(lngrowth~freq+env+freq*env))$env
```

```
##           diff          lwr          upr          p adj
## H20-Air      0.57097507  0.4730098 0.66894032 9.880985e-14
## Salt H20-Air 0.50952788  0.4115626 0.60749313 1.412537e-12
## Salt H20-H20 -0.06144719 -0.1594124 0.03651806 2.820305e-01
```

We can see that all levels of both factors except H2O and salt H2O for the environment factor are clearly different from one another.

We saw that the levels of the frequency are powers of 10 ($0.1 = 10^{-1}$, $1 = 10^0$, and $10 = 10^1$). A logarithmic transformation of the data appears to have created a more suitable model where the normality and equal variance assumptions are met.

Extra Credit

Model: $y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$ ($i = 1, 2$; $j = 1, 2, 3$; $k = 1, 2, 3$)

y_{ijk} : response variable

μ : overall mean

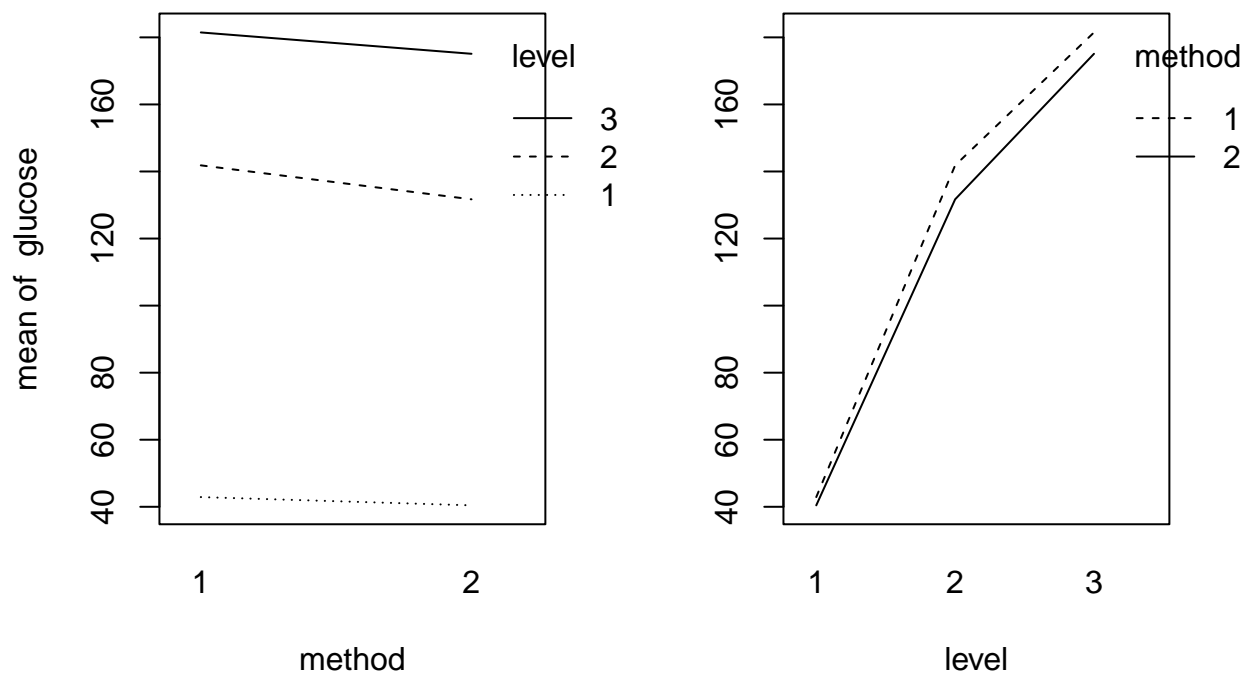
τ_i : level i of factor τ (method)

β_j : level j of factor β (glucose level)

$(\tau\beta)_{ij}$: interaction term

ϵ_{ijk} : error/residual

```
glucose<-c(42.5,138.4,180.9,39.8,132.4,176.8,43.3,144.4,180.5,40.3,132.4,173.6,42.9,142.7,183,41.2,130.5)
method<-as.factor(rep(1:2,times=3,each=3))
level<-as.factor(rep(1:3,times=6))
par(mfrow=c(1,2))
interaction.plot(method,level,glucose)
interaction.plot(level,method,glucose,ylab="")
```



```
# There may be slight interaction between method 1 and method 2.
g<-lm(glucose~method*level)
anova(g)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: glucose
```

```
##      Df Sum Sq Mean Sq    F value    Pr(>F)
## method      1      180      179.9      67.0729 2.955e-06 ***
## level       2     58864     29431.8    10975.1736 < 2.2e-16 ***
## method:level 2         44         22.0         8.2202 0.005642 **
## Residuals   12         32         2.7
```

```
## ---
```

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

```

# We reject both null hypotheses at the alpha = 0.05 level. There is sufficient
# evidence (p = 0.000003, p < 0.000001) that both factors influence concentration of glucose.
shapiro.test(g$residuals)

##
## Shapiro-Wilk normality test
##
## data:  g$residuals
## W = 0.9711, p-value = 0.8181
bartlett.test(g$residuals~method)

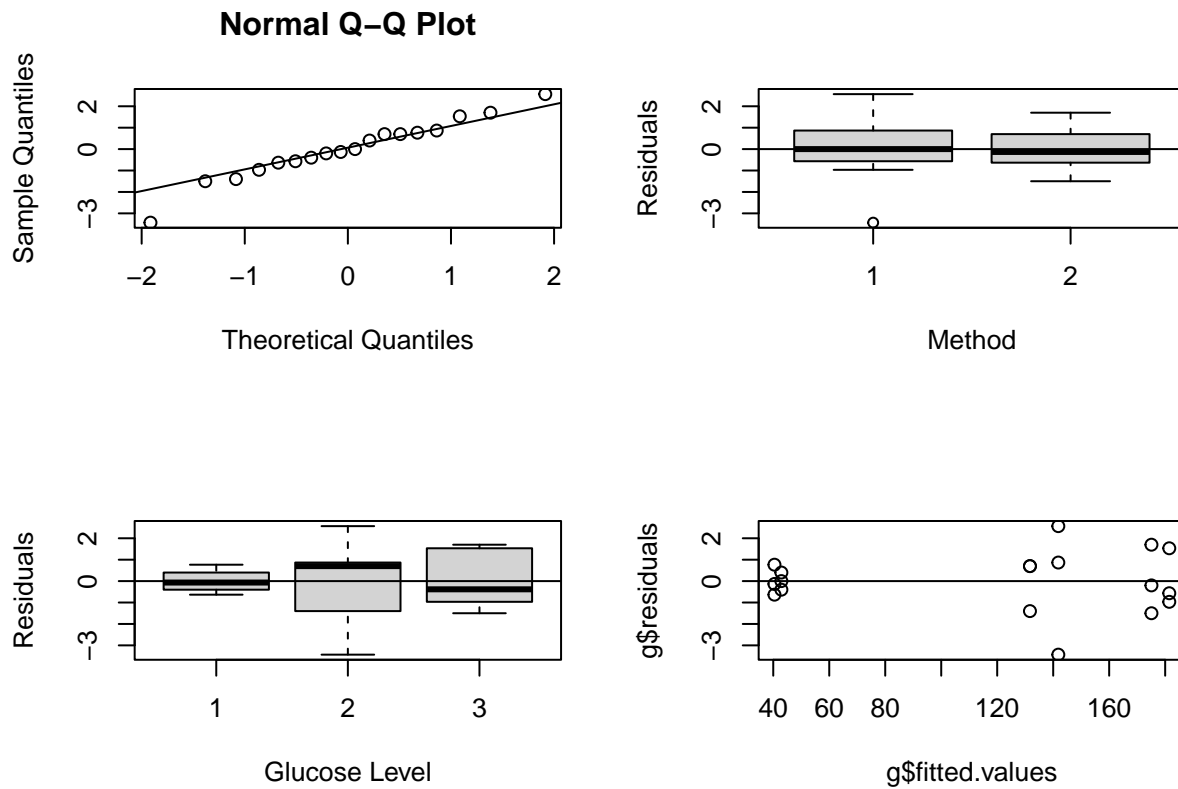
##
## Bartlett test of homogeneity of variances
##
## data:  g$residuals by method
## Bartlett's K-squared = 1.5615, df = 1, p-value = 0.2114
bartlett.test(g$residuals~level)

##
## Bartlett test of homogeneity of variances
##
## data:  g$residuals by level
## Bartlett's K-squared = 7.201, df = 2, p-value = 0.02731
leveneTest(g)

## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group  5  0.7826 0.5812
##      12

# We should exercise caution here as the null hypothesis for Bartlett's test for homoscedasticity
# by level has been rejected at the alpha = 0.05 level (p = 0.02731).
par(mfrow=c(2,2))
qqnorm(g$residuals)
qqline(g$residuals)
plot(c(method),g$residuals,xlab="Method",ylab="Residuals")
abline(h=0)
plot(c(level),g$residuals,xlab="Glucose Level",ylab="Residuals")
abline(h=0)
plot(g$fitted.values,g$residuals)
abline(h=0)

```



There is a slight variation in the Q-Q plot and a megaphone/football pattern in the # residuals vs. fitted values plot.

`TukeyHSD(aov(glucose~method+level+method*level))$method` *# Post-hoc analysis*

```
##          diff          lwr          upr          p adj
## 2-1 -6.322222 -8.004184 -4.640261 2.955106e-06
```

`TukeyHSD(aov(glucose~method+level+method*level))$level`

```
##          diff          lwr          upr          p adj
## 2-1  95.10000  92.57765  97.62235 2.708944e-14
## 3-1 136.61667 134.09432 139.13902 2.708944e-14
## 3-2  41.51667  38.99432  44.03902 1.120215e-13
```

We can see that all levels of both factors are clearly different from one another.