

Stat461 HW7

Enbo Cao

Oct 23, 2020

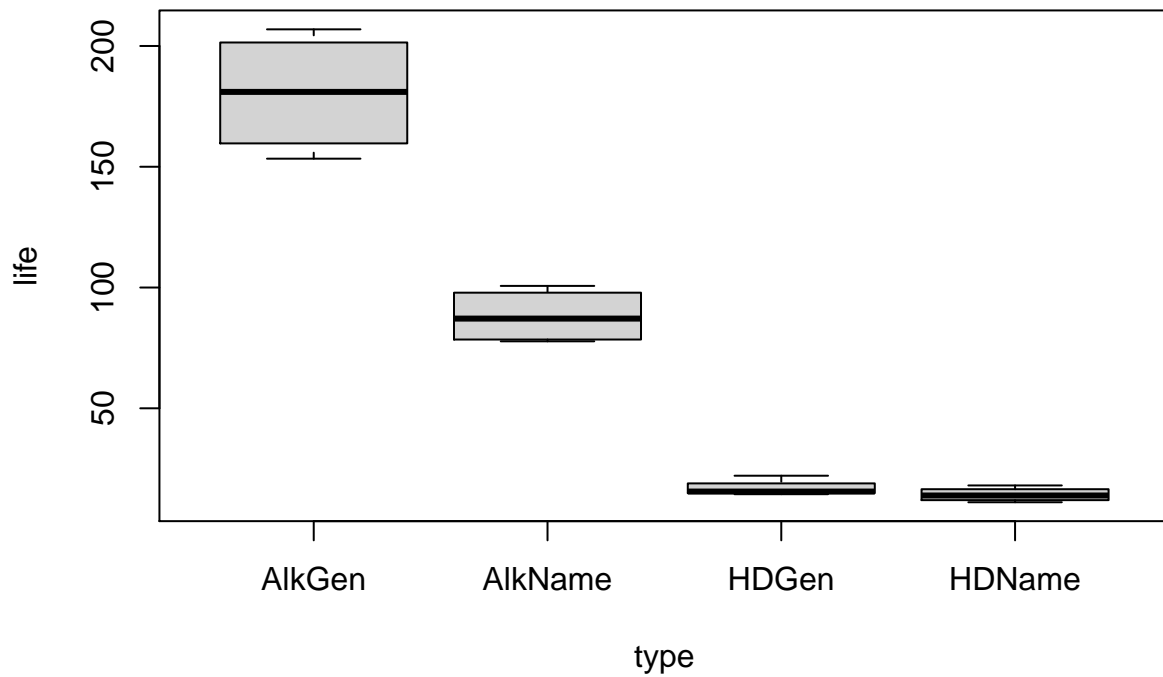
Question 1

1

```
type<-c("AlkName", "AlkName", "AlkName", "AlkName", "AlkGen", "AlkGen", "AlkGen", "AlkGen",  
"HDName", "HDName", "HDName", "HDName", "HDGen", "HDGen", "HDGen", "HDGen")  
life<-c(100.668, 77.734, 79.210, 95.063, 206.880, 153.347, 165.980, 196.000,  
14.951, 18.063, 11.111, 12.840, 15.340, 22.090, 15.734, 14.440)  
batt<-data.frame(type=type, life=life)  
batt
```

```
##      type    life  
## 1 AlkName 100.668  
## 2 AlkName  77.734  
## 3 AlkName  79.210  
## 4 AlkName  95.063  
## 5  AlkGen 206.880  
## 6  AlkGen 153.347  
## 7  AlkGen 165.980  
## 8  AlkGen 196.000  
## 9  HDName  14.951  
## 10 HDName  18.063  
## 11 HDName  11.111  
## 12 HDName  12.840  
## 13  HDGen  15.340  
## 14  HDGen  22.090  
## 15  HDGen  15.734  
## 16  HDGen  14.440
```

```
boxplot(life~type, data = batt)
```



2

(a)

$$Y_{it} = \mu + \tau_i + \epsilon_{it}, \quad i = AN, AG, HN, HG \quad t = 1, 2, 3, 4$$

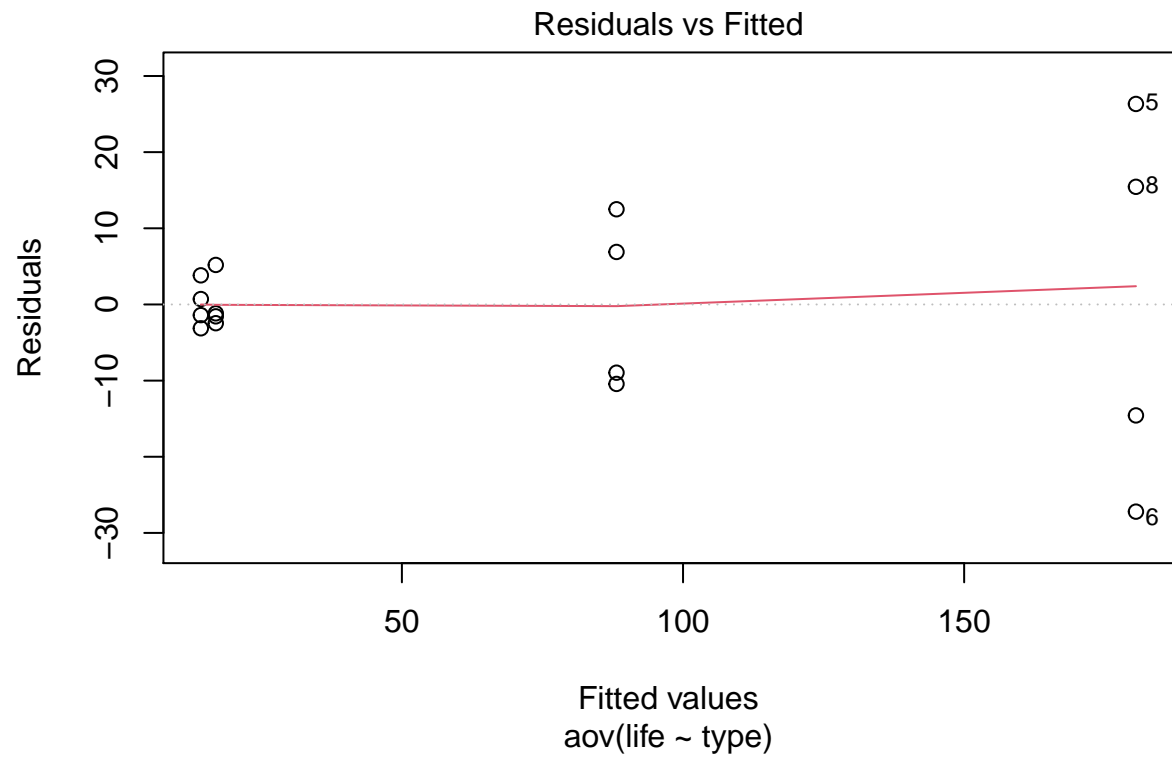
$$\epsilon_{it} \stackrel{iid}{\sim} N(0, \sigma^2)$$

#(b)

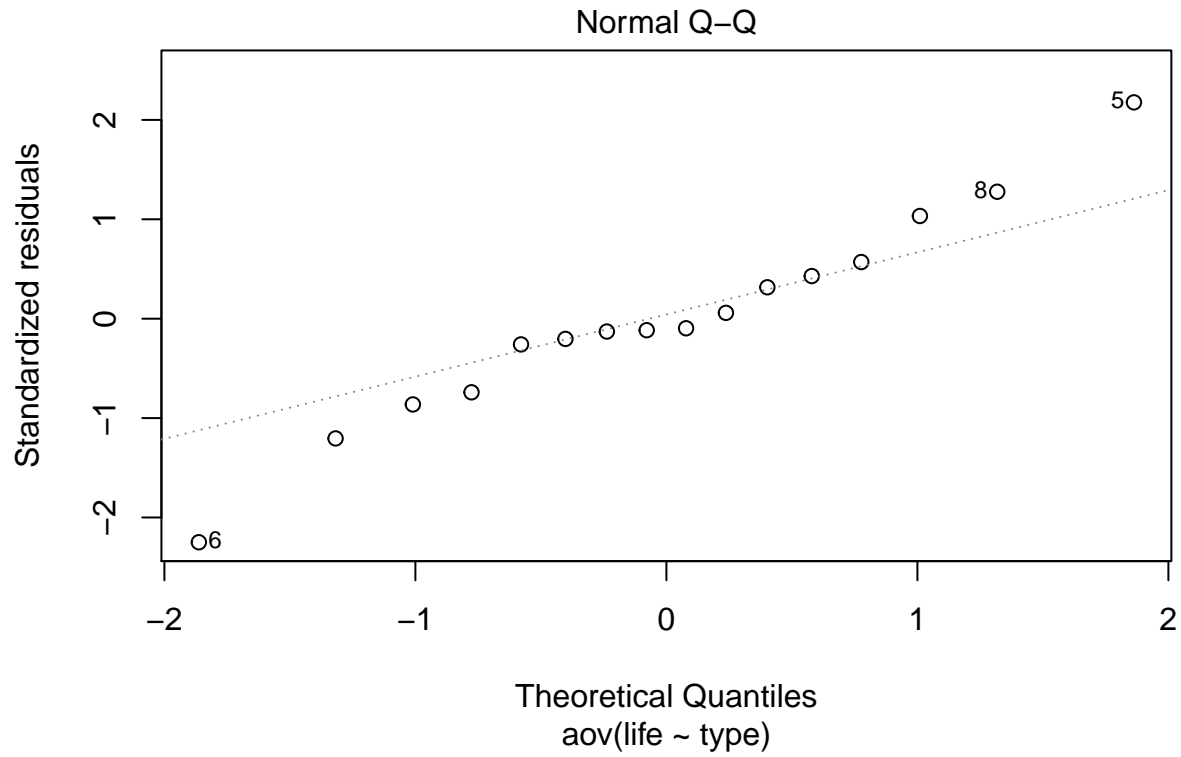
```
batt1=aov(life~type, data = batt)
summary(batt1)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## type         3  73527    24509   125.6 2.5e-09 ***
## Residuals   12   2341     195
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(batt1, which = 1)
```



```
plot(batt1, which = 2)
```



```
# The residuals are approximately normal because the QQ line is a straight line
# which indicate the the residuals are normally distributed.
```

```
##(c)
```

```
v1 = var(life[type=="AlkName"])
```

```
v2 = var(life[type=="AlkGen"])
```

```
v3 = var(life[type=="HDName"])
```

```
v4 = var(life[type=="HDGen"])
```

```
v1
```

```
## [1] 130.9684
```

```
v2
```

```
## [1] 628.0865
```

```
v3
```

```
## [1] 8.957162
```

```
v4
```

```
## [1] 12.26028
```

```
v2/v3
```

```
## [1] 70.12115
```

```
# The assumption of constant error variance among treatments is not justified,
# since the largest variance is more than 3 times larger than smallest variance.
```

3

(a)

$$\sqrt{Y_{it}} = \mu + \tau_i + \epsilon_{it}, \quad i = AN, AG, HN, HG \quad t = 1, 2, 3, 4$$

$$\epsilon_{it} \stackrel{iid}{\sim} N(0, \sigma^2)$$

#3-5 transformations

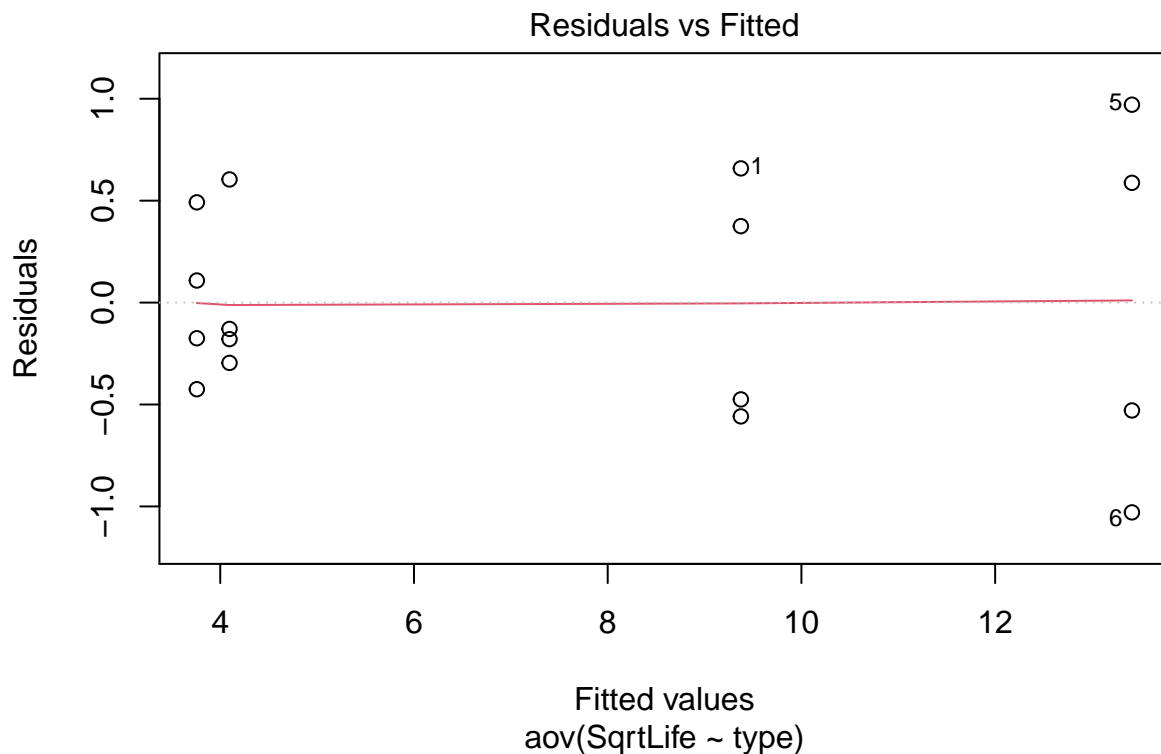
```
batt$SqrtLife = sqrt(life)
batt$LogLife = log(life)
batt$SqrLife = (life)^2
```

#(b)

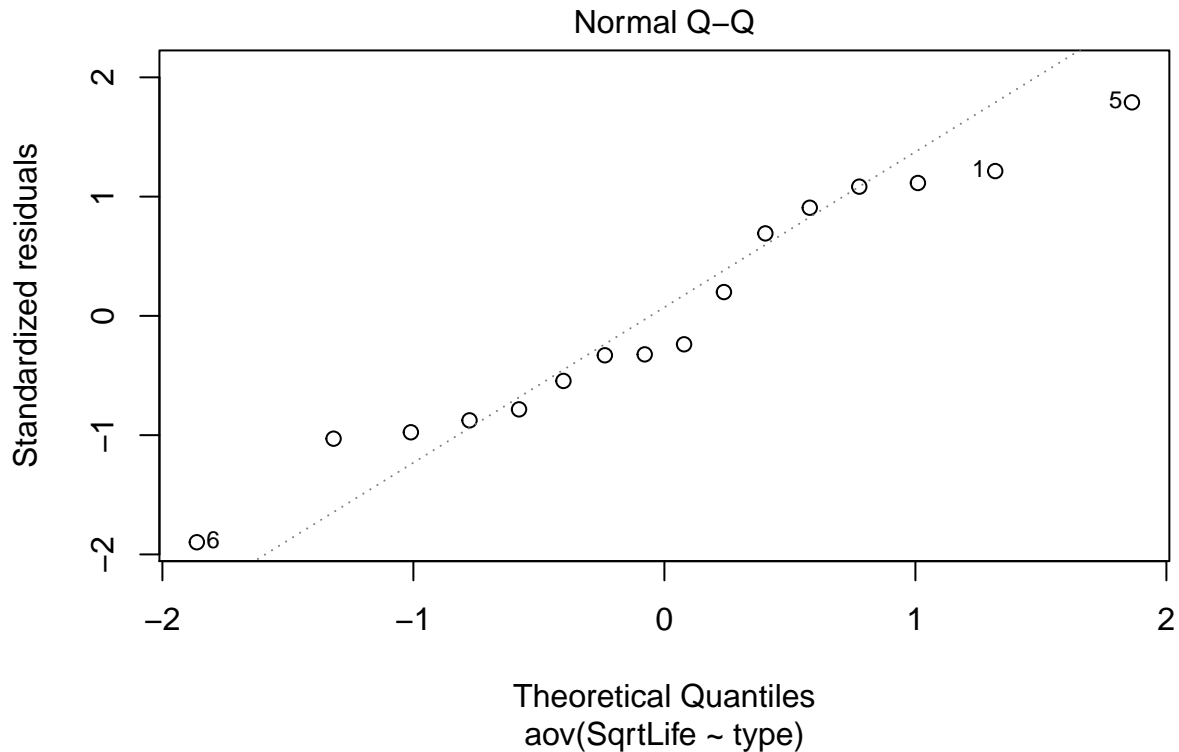
```
sqrtAnova = aov(SqrtLife~type,data = batt)
summary(sqrtAnova)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## type         3  255.8   85.28    217.5 1.01e-10 ***
## Residuals    12    4.7    0.39
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(sqrtAnova, which = 1)
```



```
plot(sqrtAnova, which = 2)
```



```
# The residuals are approximately normal since the dots on QQ graph  
# are shown as a straight line.
```

```
##(c)
```

```
v1 = var(batt$SqrtLife[type=="AlkName"])
```

```
v2 = var(batt$SqrtLife[type=="AlkGen"])
```

```
v3 = var(batt$SqrtLife[type=="HDName"])
```

```
v4 = var(batt$SqrtLife[type=="HDGen"])
```

```
v1
```

```
## [1] 0.370466
```

```
v2
```

```
## [1] 0.875621
```

```
v3
```

```
## [1] 0.1549349
```

```
v4
```

```
## [1] 0.1671155
```

```
v2/v3
```

```
## [1] 5.651542
```

```
# The assumption of constant error variance among treatments is not justified,  
# since the largest variance is more than 3 times larger than smallest variance.
```

#4 (a)

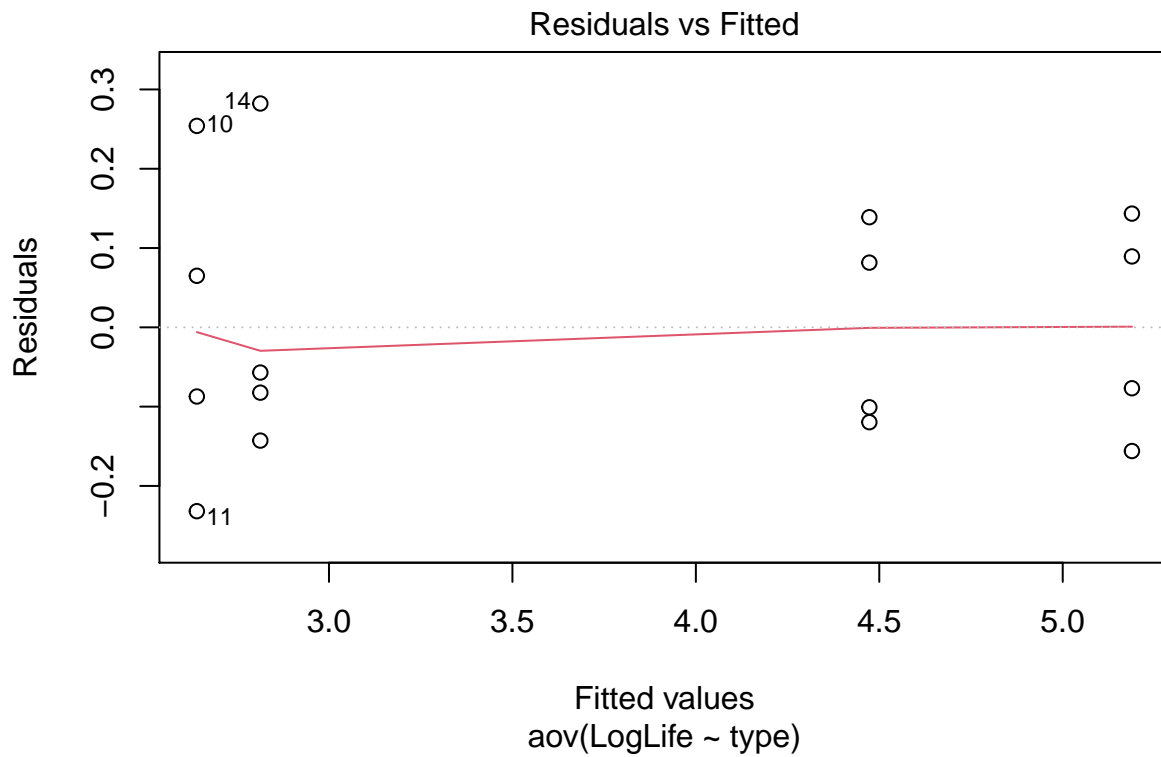
$$\log(Y_{it}) = \mu + \tau_i + \epsilon_{it}, \quad i = AN, AG, HN, HG \quad t = 1, 2, 3, 4$$

$$\epsilon_{it} \stackrel{iid}{\sim} N(0, \sigma^2)$$

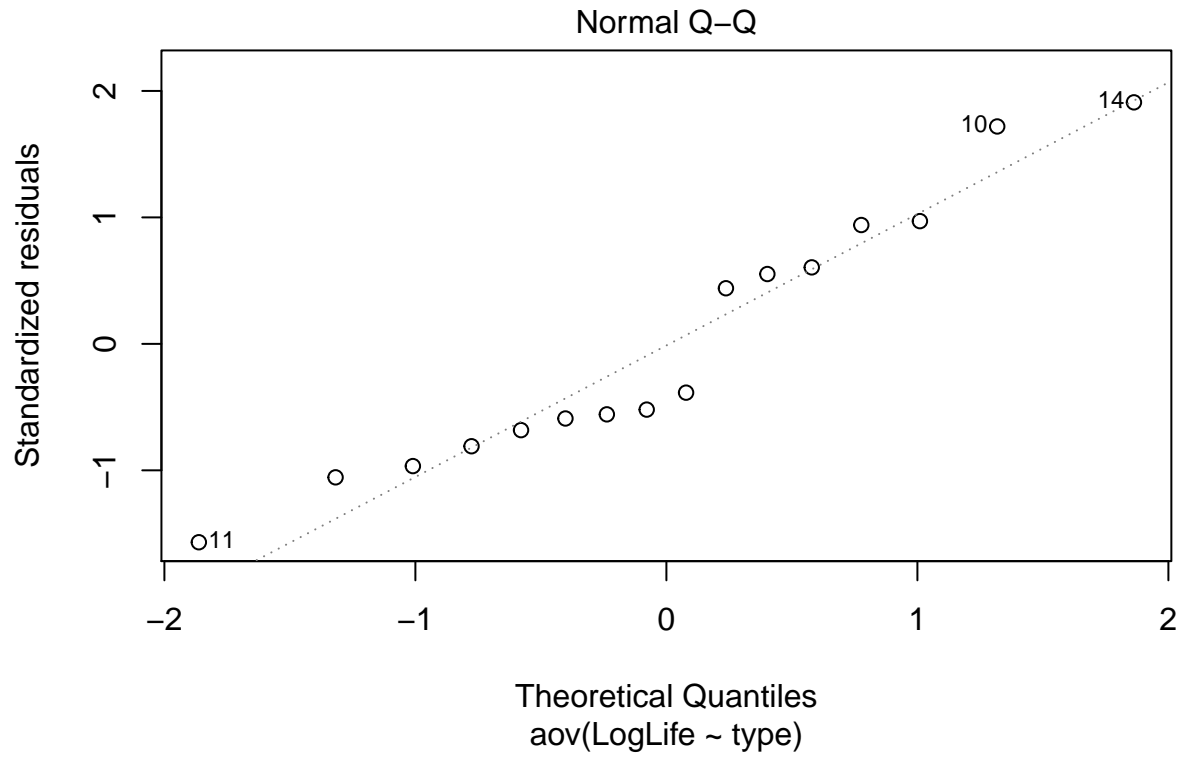
```
logAnova = aov(LogLife ~ type, data = batt)
summary(logAnova)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## type          3  18.80   6.267    215.2 1.08e-10 ***
## Residuals    12   0.35   0.029
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(logAnova, which = 1)
```



```
plot(logAnova, which = 2)
```



```
# The residuals are approximately normal since the dots on QQ graph  
# are shown as a straight line.
```

```
v1 = var(batt$LogLife[type=="AlkName"])  
v2 = var(batt$LogLife[type=="AlkGen"])  
v3 = var(batt$LogLife[type=="HDName"])  
v4 = var(batt$LogLife[type=="HDGen"])  
v1
```

```
## [1] 0.01681069
```

```
v2
```

```
## [1] 0.01960264
```

```
v3
```

```
## [1] 0.04338084
```

```
v4
```

```
## [1] 0.03670805
```

```
v3/v1
```

```
## [1] 2.580551
```

```
# The assumption of constant error variance among treatments is justified,  
# since the largest variance is not more than 3 times larger than smallest variance.
```


#5 (a)

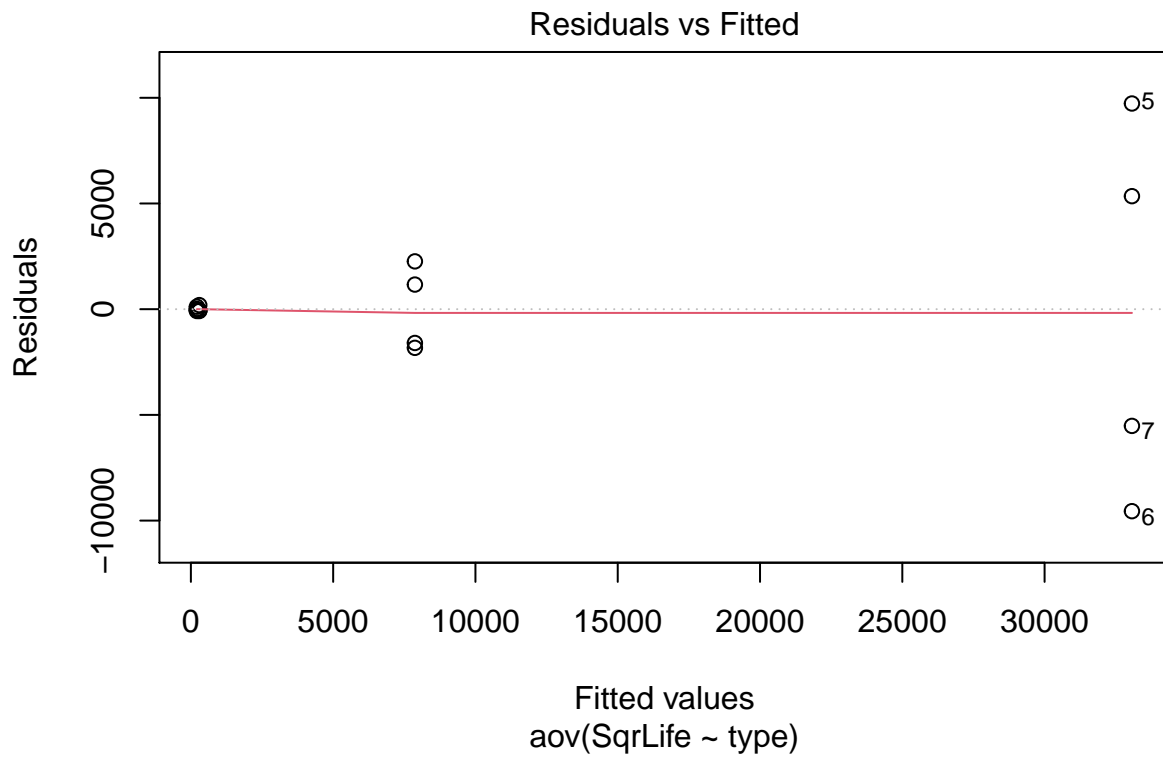
$$(Y_{it})^2 = \mu + \tau_i + \epsilon_{it}, \quad i = AN, AG, HN, HG \quad t = 1, 2, 3, 4$$

$$\epsilon_{it} \stackrel{iid}{\sim} N(0, \sigma^2)$$

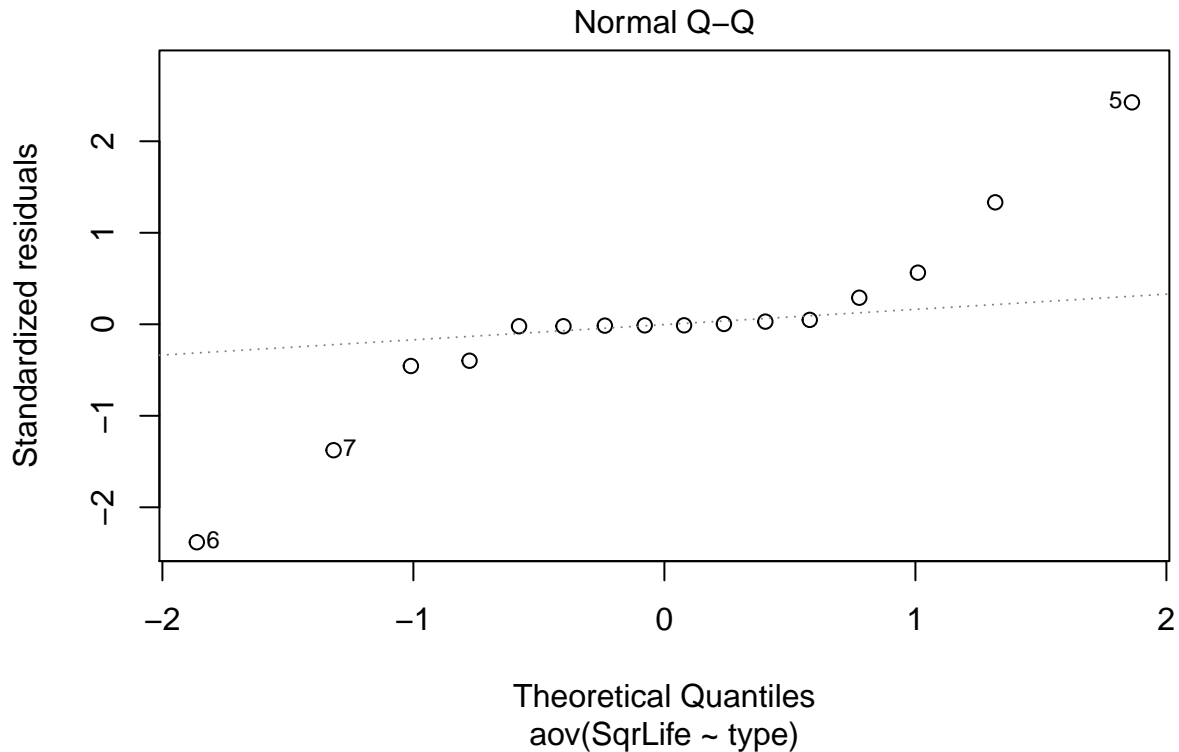
```
sqrAnova = aov(SqrLife~type,data = batt)
summary(sqrAnova)
```

```
##           Df      Sum Sq   Mean Sq F value    Pr(>F)
## type         3 2.905e+09 968365217   45.13 8.23e-07 ***
## Residuals    12 2.575e+08  21454736
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(sqrAnova, which = 1)
```



```
plot(sqrAnova, which = 2)
```



*# The residuals are not approximately normal since the dots on QQ graph
are not shown as a straight line. It is the "trumpet" shape.*

```
v1 = var(batt$SqrLife[type=="AlkName"])
v2 = var(batt$SqrLife[type=="AlkGen"])
v3 = var(batt$SqrLife[type=="HDName"])
v4 = var(batt$SqrLife[type=="HDGen"])
v1
```

```
## [1] 4124567
```

```
v2
```

```
## [1] 81669791
```

```
v3
```

```
## [1] 7742.895
```

```
v4
```

```
## [1] 16843.13
```

```
v2/v3
```

```
## [1] 10547.71
```

*# The assumption of constant error variance among treatments is not justified,
since the largest variance is more than 3 times larger than smallest variance.*

```
#6
```

It is very hard to decide which one is the best model among the four models. I would say the log model is

the best one. On the log model, most points lie on the line and only a part of points are slightly skewed, the beginning and ending points are also more closer to the straight line than other models.

#7

$H_0 : \tau_{AN} - \tau_{AG} = \tau_{AN} - \tau_{HN} = \tau_{AN} - \tau_{HG} = \tau_{AG} - \tau_{HN} = \tau_{AG} - \tau_{HG} = \tau_{HN} - \tau_{HG}$

H_a : At least there is one difference between batteries is different

```
library(lsmeans)
```

```
## Loading required package: emmeans
```

```
## The 'lsmeans' package is now basically a front end for 'emmeans'.
```

```
## Users are encouraged to switch the rest of the way.
```

```
## See help('transition') for more information, including how to
```

```
## convert old 'lsmeans' objects and scripts to work with 'emmeans'.
```

```
lsm.batt=lsmeans(logAnova,"type")
```

```
library(knitr)
```

```
model=logAnova
```

```
summary(contrast(lsm.batt,method="pairwise", adjust="tukey"),
```

```
infer=c(T,T), level=0.95, side="two-sided")
```

```
## contrast estimate SE df lower.CL upper.CL t.ratio p.value
```

```
## AlkGen - AlkName 0.716 0.121 12 0.357 1.074 5.931 0.0003
```

```
## AlkGen - HDGen 2.376 0.121 12 2.018 2.734 19.688 <.0001
```

```
## AlkGen - HDName 2.549 0.121 12 2.191 2.907 21.122 <.0001
```

```
## AlkName - HDGen 1.660 0.121 12 1.302 2.018 13.757 <.0001
```

```
## AlkName - HDName 1.833 0.121 12 1.475 2.191 15.191 <.0001
```

```
## HDGen - HDName 0.173 0.121 12 -0.185 0.531 1.434 0.5034
```

```
##
```

```
## Confidence level used: 0.95
```

```
## Conf-level adjustment: tukey method for comparing a family of 4 estimates
```

```
## P value adjustment: tukey method for comparing a family of 4 estimates
```

```
# T ratio for AG-AN is 5.931, AG-HG is 19.688, AG-HN is 21.122
```

```
# AN-HG is 13.757, AN-HN is 15.191, HG-HN is 1.434.
```

```
#Since the all the P-value is <= 0.05, except HDGen-HDName.
```

```
#We reject the Ho.
```

```
#It means at least there is one difference between batteries is different.
```

Question 2

```
Types=c(rep("Beef",20),rep("Pork",17),rep("Chicken",17))
```

```
Calories=c(186,181,176,149,184,190,158,139,175,148,152,111,141,153,190,157,131,149,
```

```
135,132,173,191,182,190,172,147,146,139,175,136,179,153,107,195,135,140,
```

```
138,129,132,102,106,94,102,87,99,107,113,135,142,86,143,152,146,144)
```

```
hotdogs=data.frame(Types,Calories)
```

```
hotdogs
```

```
## Types Calories
```

```
## 1 Beef 186
```

```
## 2 Beef 181
```

```
## 3 Beef 176
```

```
## 4 Beef 149
```

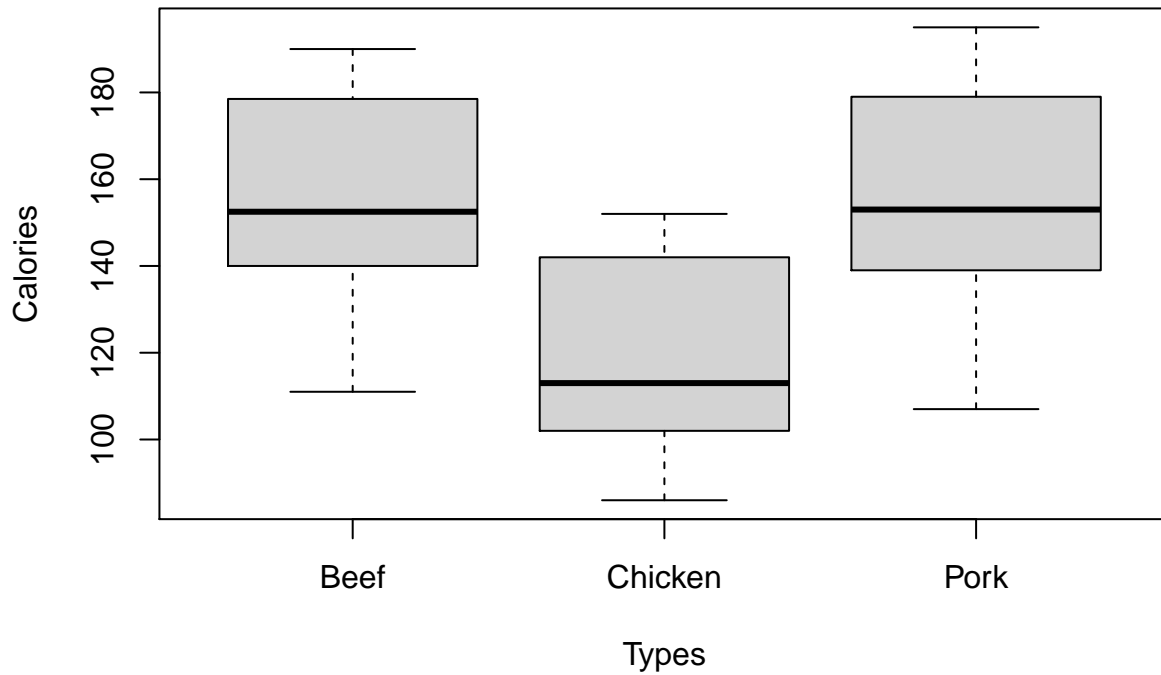
```
## 5 Beef 184
```

```
## 6 Beef 190
```

## 7	Beef	158
## 8	Beef	139
## 9	Beef	175
## 10	Beef	148
## 11	Beef	152
## 12	Beef	111
## 13	Beef	141
## 14	Beef	153
## 15	Beef	190
## 16	Beef	157
## 17	Beef	131
## 18	Beef	149
## 19	Beef	135
## 20	Beef	132
## 21	Pork	173
## 22	Pork	191
## 23	Pork	182
## 24	Pork	190
## 25	Pork	172
## 26	Pork	147
## 27	Pork	146
## 28	Pork	139
## 29	Pork	175
## 30	Pork	136
## 31	Pork	179
## 32	Pork	153
## 33	Pork	107
## 34	Pork	195
## 35	Pork	135
## 36	Pork	140
## 37	Pork	138
## 38	Chicken	129
## 39	Chicken	132
## 40	Chicken	102
## 41	Chicken	106
## 42	Chicken	94
## 43	Chicken	102
## 44	Chicken	87
## 45	Chicken	99
## 46	Chicken	107
## 47	Chicken	113
## 48	Chicken	135
## 49	Chicken	142
## 50	Chicken	86
## 51	Chicken	143
## 52	Chicken	152
## 53	Chicken	146
## 54	Chicken	144

```
boxplot(Calories~Types,main="Calories of different hot dogs")
```

Calories of different hot dogs



$$Y_{it} = \mu + \tau_i + \epsilon_{it}, \quad i = B, P, C \quad t = 1, 2, \dots, r_i \quad r_B = 20 \quad r_P = 17 \quad r_C = 17$$

$$\epsilon_{it} \stackrel{iid}{\sim} N(0, \sigma^2)$$

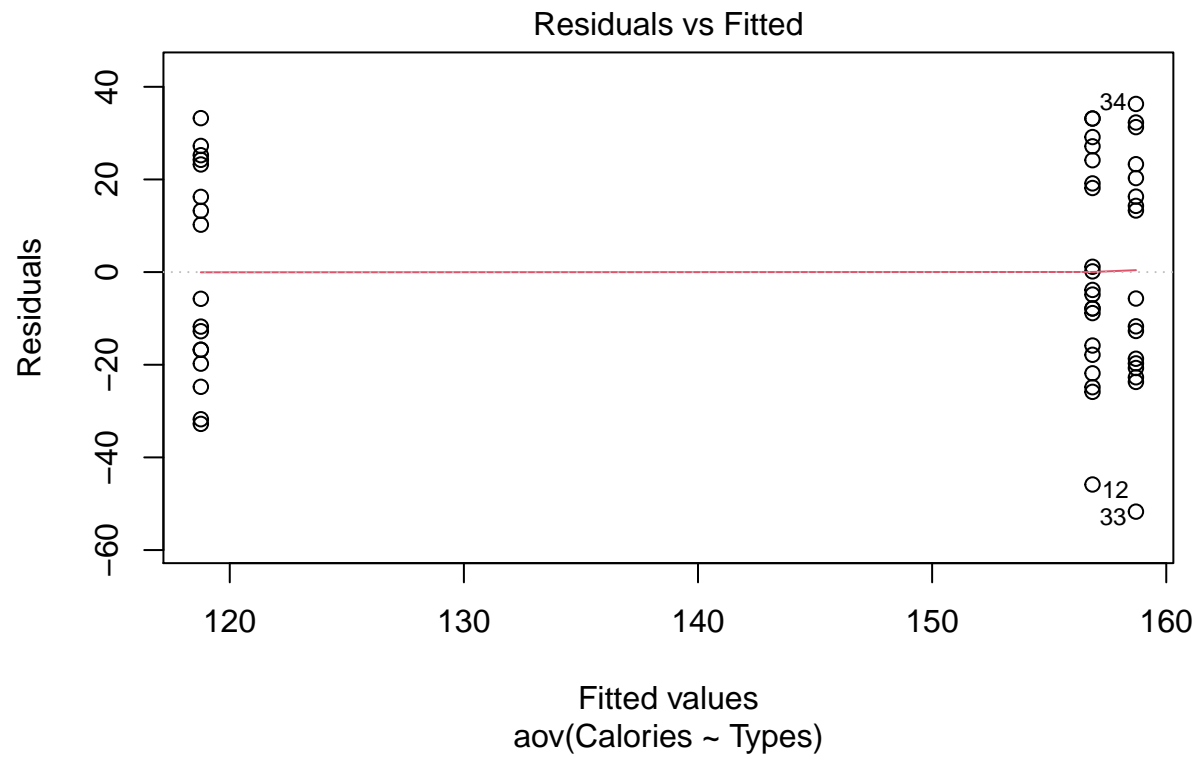
$$H_0 : \tau_B - \tau_P = \tau_B - \tau_C = \tau_P - \tau_C$$

H_a : At least there is one difference between the types is different.

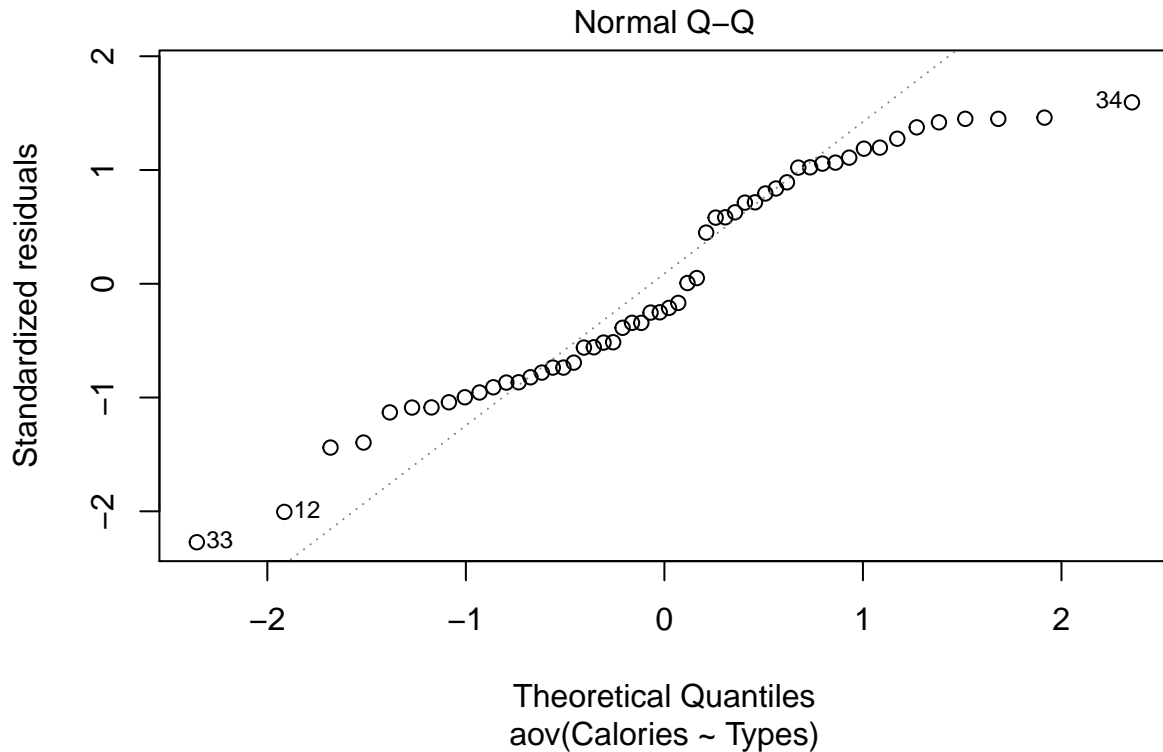
```
hotdog1=aov(Calories~Types, data = hotdogs)
summary(hotdog1)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Types      2  17692    8846   16.07 3.86e-06 ***
## Residuals  51  28067     550
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(hotdog1, which = 1)
```



```
plot(hotdog1, which = 2)
```



```
v1 = var(Calories[Types=="Beef"])
v2 = var(Calories[Types=="Pork"])
v3 = var(Calories[Types=="Chicken"])
v1

## [1] 512.6605

v2

## [1] 636.8456

v3

## [1] 508.5662

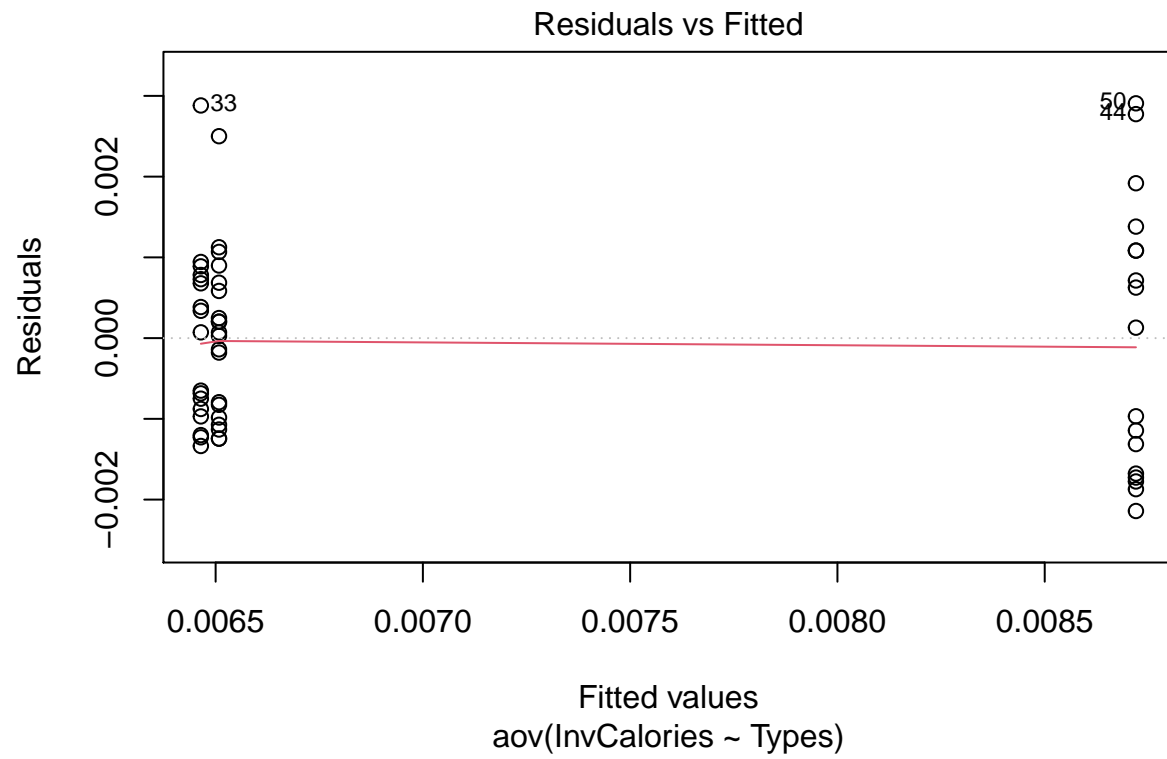
v2/v3

## [1] 1.252237

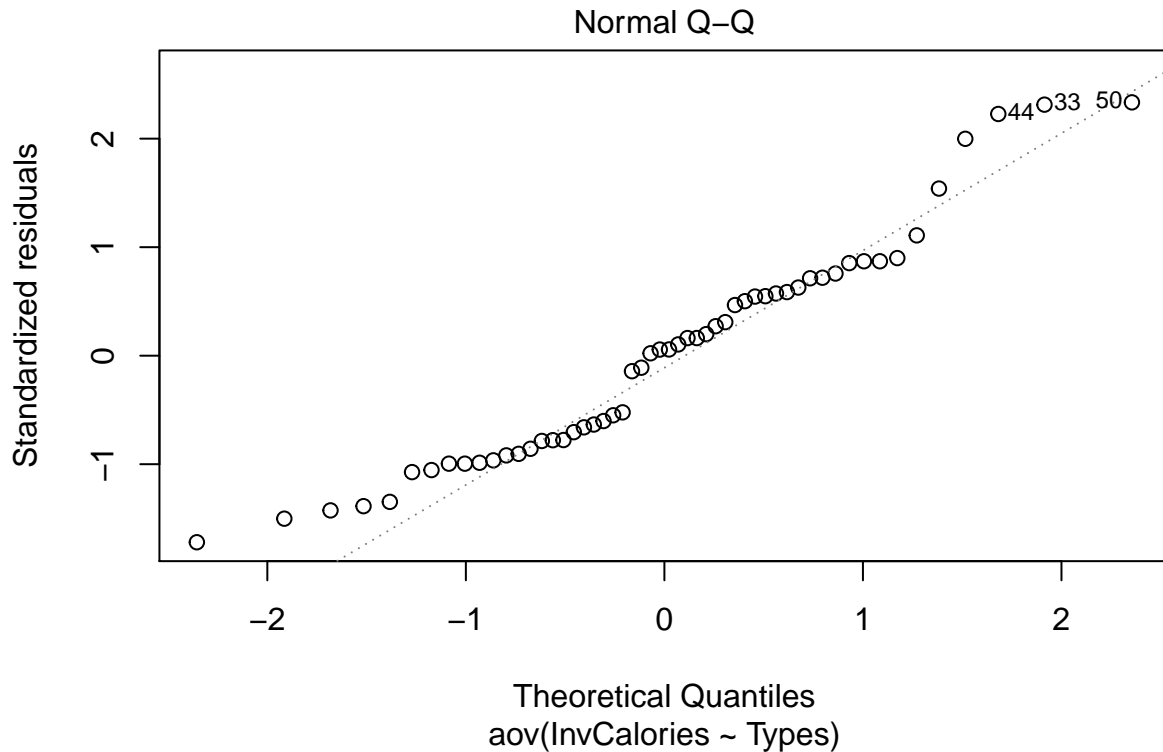
hotdogs$InvCalories = 1/Calories
InvAnova=aov(InvCalories~Types, data = hotdogs)
summary(InvAnova)

##           Df    Sum Sq  Mean Sq F value   Pr(>F)
## Types      2 5.805e-05  2.903e-05   17.61 1.53e-06 ***
## Residuals  51 8.406e-05  1.648e-06
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(InvAnova, which = 1)
```



```
plot(InvAnova, which = 2)
```

```
v1 = var(hotdogs$InvCalories[Types=="Beef"])
v2 = var(hotdogs$InvCalories[Types=="Pork"])
v3 = var(hotdogs$InvCalories[Types=="Chicken"])
v1
```

```
## [1] 9.629329e-07
```

```
v2
```

```
## [1] 1.234792e-06
```

```
v3
```

```
## [1] 2.875638e-06
```

```
v3/v1
```

```
## [1] 2.986333
```

```
library(lsmeans)
lsm.Inv=lsmeans(InvAnova,"Types")
library(knitr)
model=InvAnova
summary(contrast(lsm.Inv,method="pairwise", adjust="tukey"),
infer=c(T,T), level=0.95, side="two-sided")
```

```
## contrast      estimate      SE df lower.CL upper.CL t.ratio p.value
## Beef - Chicken -2.21e-03 0.000424 51 -0.003234 -0.00119 -5.223 <.0001
## Beef - Pork    4.38e-05 0.000424 51 -0.000979 0.00107 0.103 0.9941
## Chicken - Pork 2.26e-03 0.000440 51 0.001193 0.00332 5.122 <.0001
```

```
##  
## Confidence level used: 0.95  
## Conf-level adjustment: tukey method for comparing a family of 3 estimates  
## P value adjustment: tukey method for comparing a family of 3 estimates  
#Since the p-value for beef-chicken and chicken-pork is smaller than 0.05  
#We reject the Ho.  
#Therefore, at least there is one difference between the types is different.
```