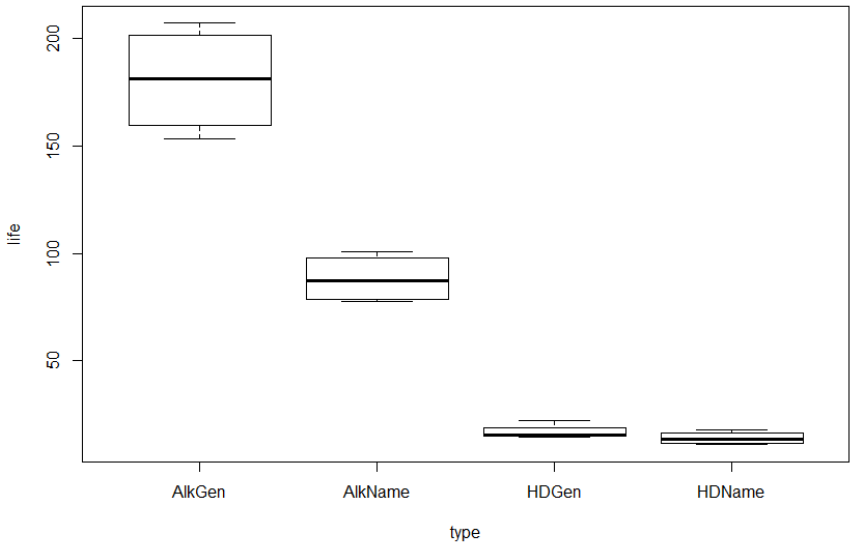
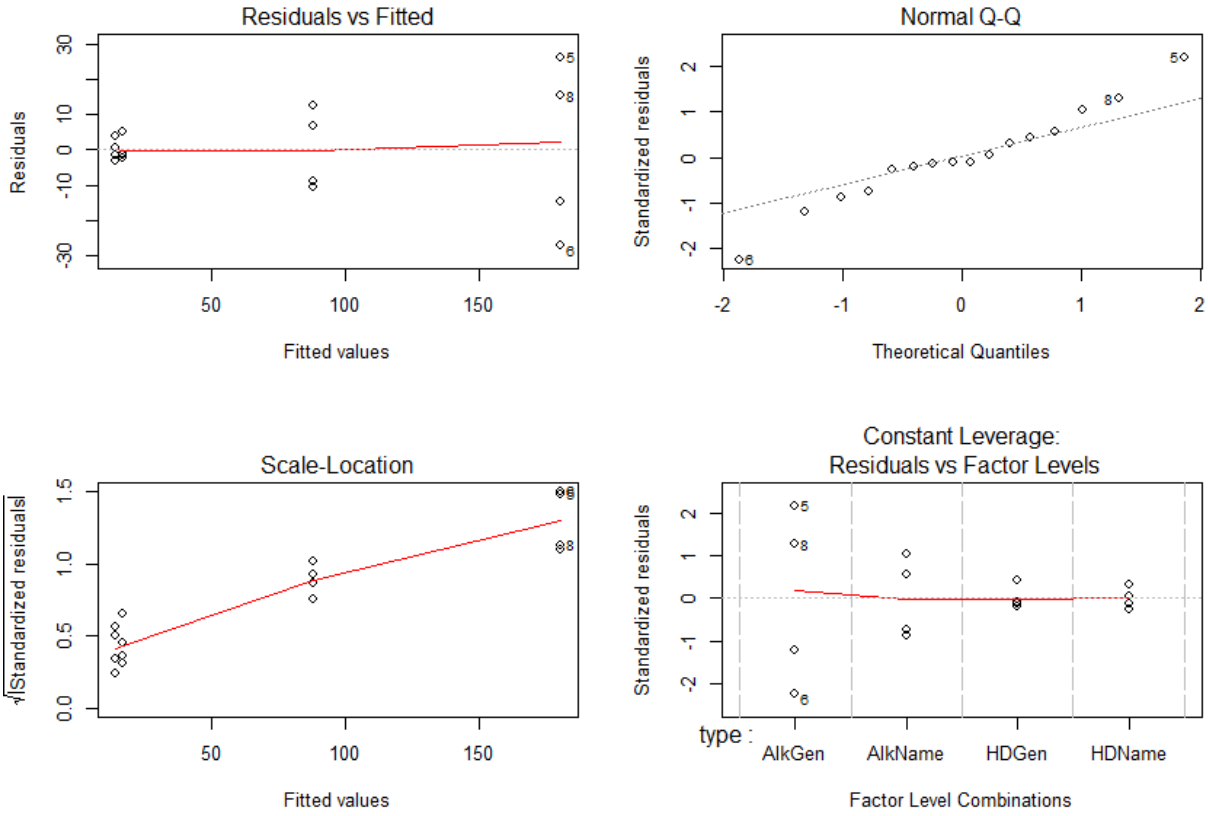
Homework 6

Jiaqi Li

1.



2.



(a)

$Y_{it} = \mu + \tau_i + \epsilon_{it}$, $\epsilon_{it} \sim N(0, \sigma^2)$ with iid.

i = 1, 2, 3, 4

t = 1, 2, 3, 4

(b)

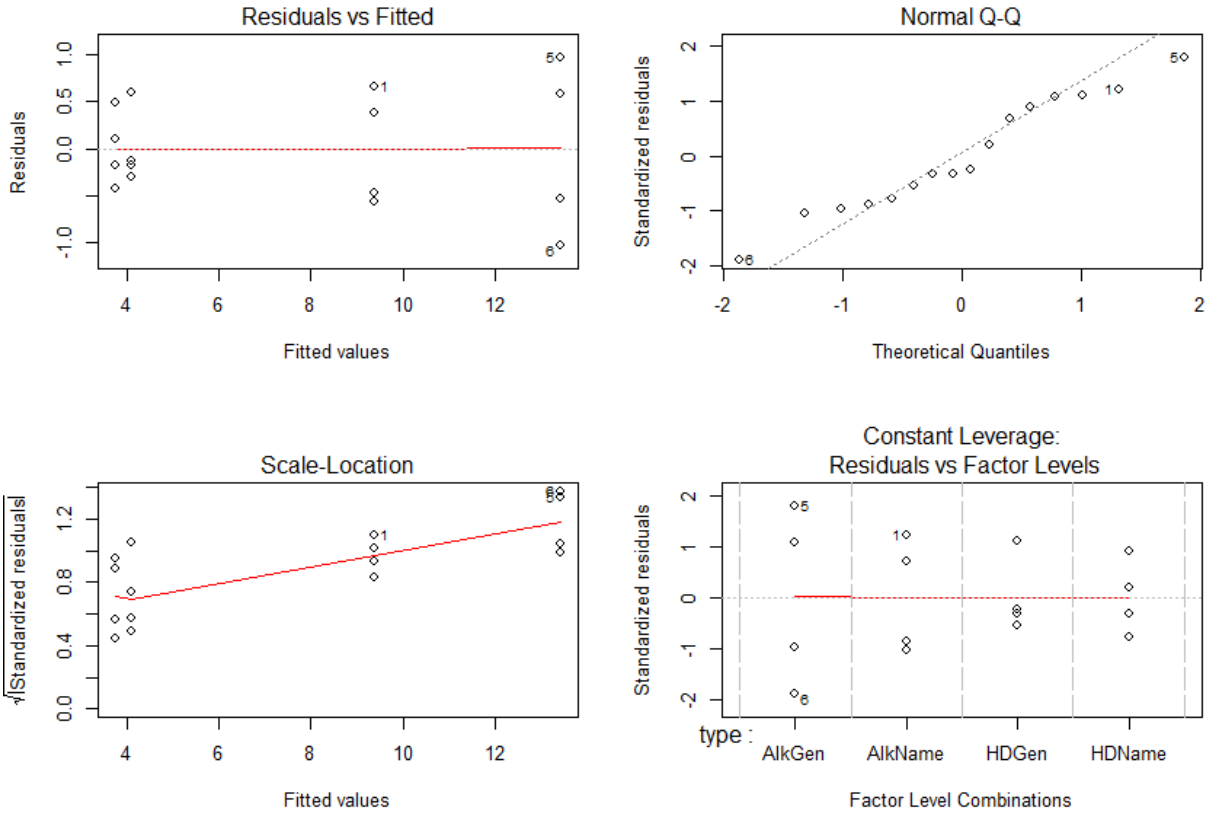
By Q-Q plot, we observe that residuals are not fitting the line well, which indicates that residuals violate normality.

(c)

By the scatter plot, we observe that residuals do not have constant variance since the variance of residuals are getting larger.

Thus, the assumption of constant error variance among treatments is <u>NOT</u> justified.

3.



(a)

$\sqrt{Y_{it}}$ = μ + τ_i + ϵ_{it} , ϵ_{it} ~ N(0, σ^2) with iid.

i = 1, 2, 3, 4

t = 1, 2, 3, 4

(b)

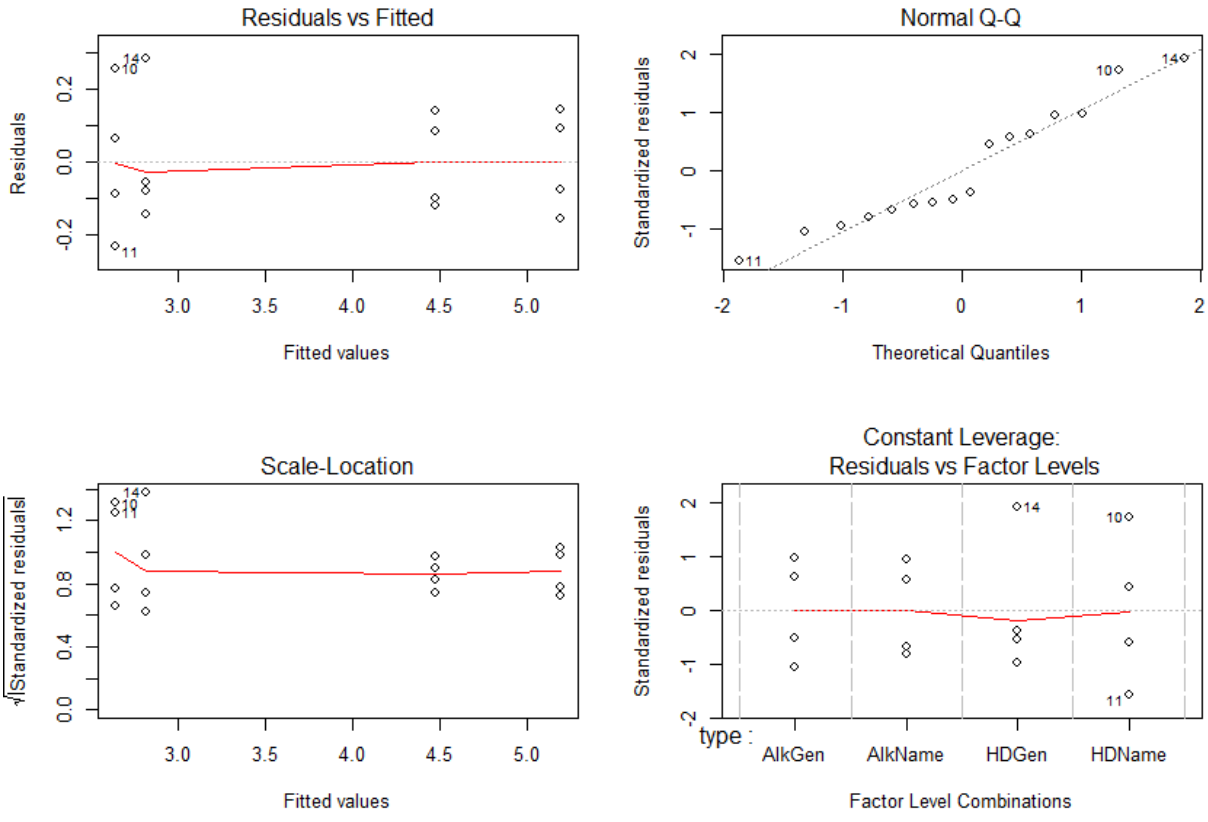
By Q-Q plot, we observe that residuals fit the line well, which indicates that residuals satisfy normality.

(c)

By the scatter plot, we observe that variances of residuals are approximately constant.

Thus, the assumption of constant error variance among treatments is justified.

4.



(a)

$\log Y_{it}$ = μ + τ_i + ϵ_{it} , ϵ_{it} ~ N(0, σ^2) with iid.

$i = 1, 2, 3, 4$

$t = 1, 2, 3, 4$

(b)

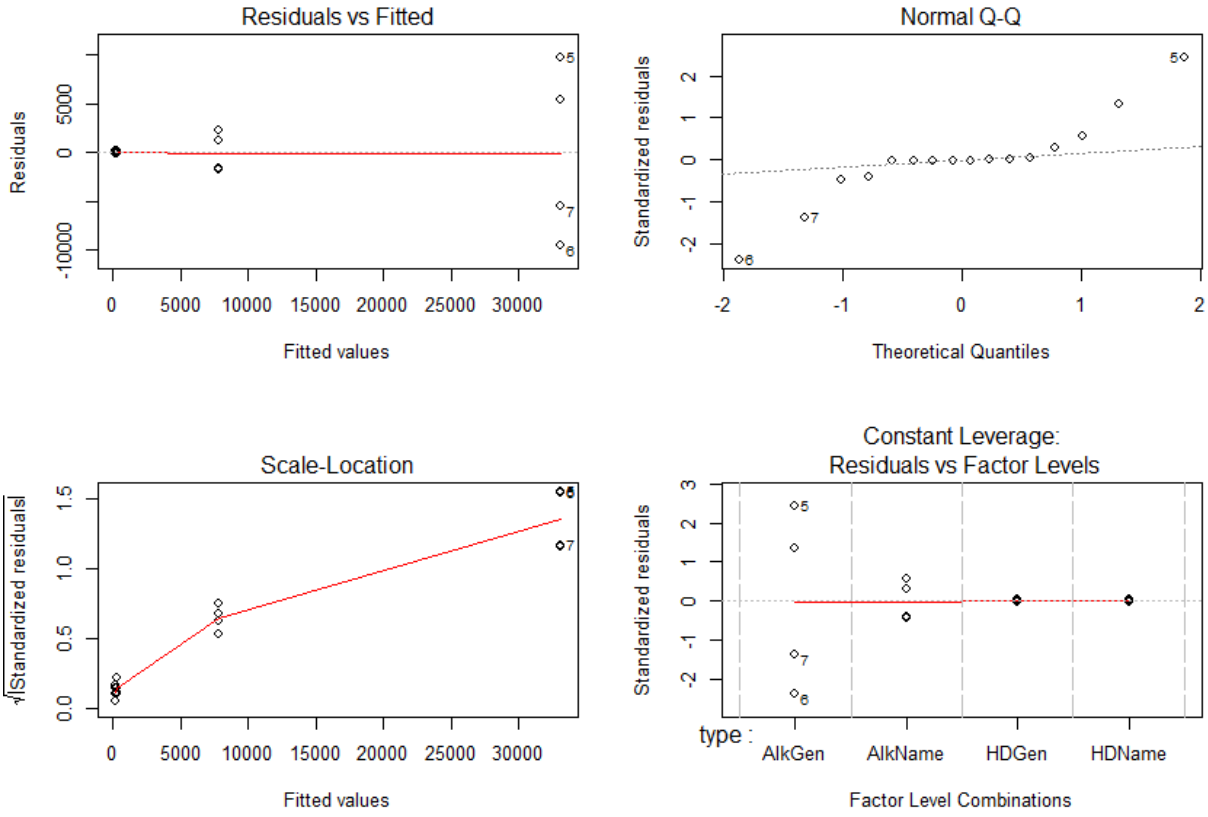
By Q-Q plot, we observe that residuals fit the line well, which indicates that residuals satisfy normality.

(c)

By the scatter plot, we observe that variances of residuals are approximately constant.

Thus, the assumption of constant error variance among treatments is justified.

5.



(a)

$Y_{it}^2 = \mu + \tau_i + \epsilon_{it}$, $\epsilon_{it} \sim N(0, \sigma^2)$ with iid.

$i = 1, 2, 3, 4$

$t_i = 1, 2, 3, 4$

(b)

By Q-Q plot, we observe that residuals are not fitting the line well, which indicates that residuals violate normality.

(c)

By the scatter plot, we observe that residuals do not have constant variance since the variance of residuals are getting larger.

Thus, the assumption of constant error variance among treatments is NOT justified.

6.

Square-root of the battery life and log of the battery life satisfy the assumptions of ANOVA model.

Also, by observing the scatter plot, we see that residuals of log of the battery life have smaller variance.

Thus, the model with the log of the battery life as a response variable best satisfy the assumptions of ANOVA model.

7.

```
> summary(fit3)
            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
```

H_0: $\tau_1=\tau_2=\tau_3=\tau_4$

H_1: At least one of τ_i is different with others.

$T^* = F_{3,12} = 215.2$

p-value = $1.08 \times 10^{-10} < \alpha = 0.05$

Thus, reject null hypothesis and conclude that there is enough evidence to show that at least one of τ_i is different with others for i=1,2,3,4.

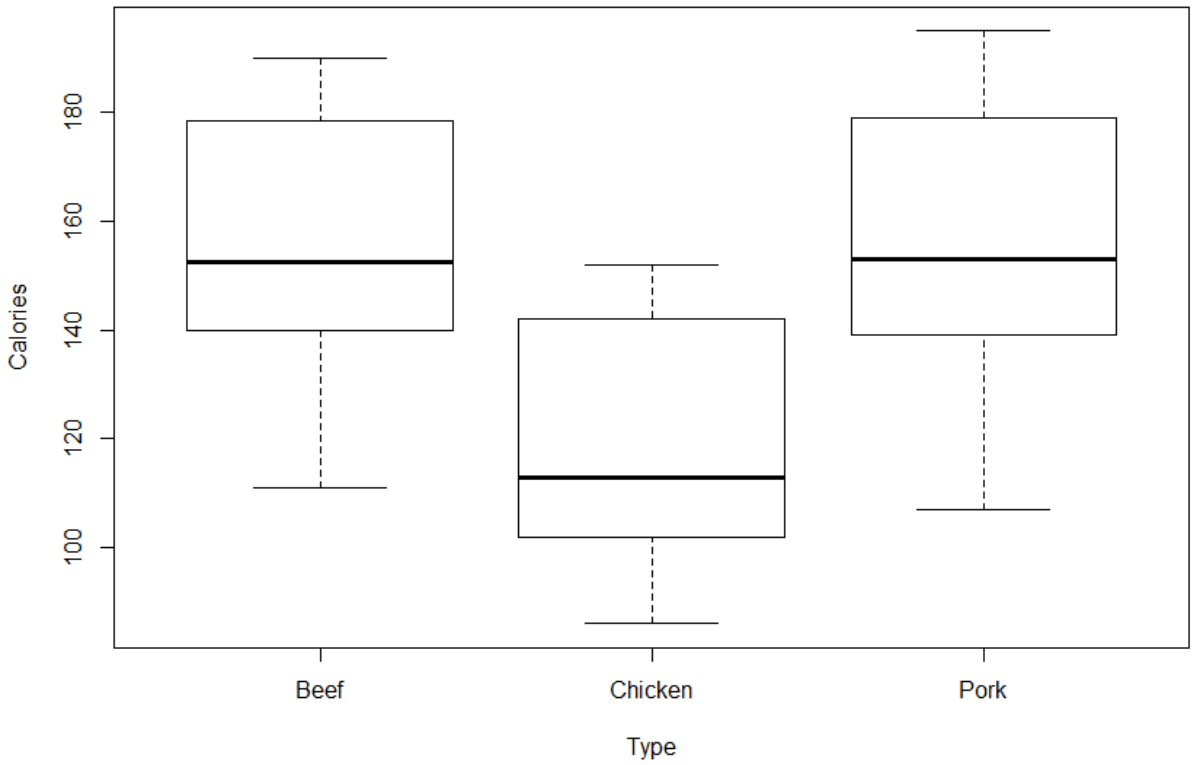
```
> lsm.life=lsmeans(fit3, ~type)
> kable(summary(contrast(lsm.life,method="pairwise",adjust="tukey"),infer=c(T,T),level=0.5,side="
two-sided"))


|contrast         |  estimate|        SE| df|   lower.CL|  upper.CL|   t.ratio|   p.value|
|:----------------|---------:|---------:|--:|----------:|---------:|---------:|---------:|
|AlkGen - AlkName | 0.7157653| 0.1206763| 12|  0.5419570| 0.8895736|  5.931281| 0.0003481|
|AlkGen - HDGen   | 2.3758523| 0.1206763| 12|  2.2020440| 2.5496606| 19.687807| 0.0000000|
|AlkGen - HDName  | 2.5489199| 0.1206763| 12|  2.3751116| 2.7227281| 21.121954| 0.0000000|
|AlkName - HDGen  | 1.6600870| 0.1206763| 12|  1.4862788| 1.8338953| 13.756526| 0.0000001|
```

```
|AlkName - HDName | 1.8331546| 0.1206763| 12|  1.6593463| 2.0069629| 15.190672| 0.0000000|
|HDGen - HDName   | 0.1730675| 0.1206763| 12| -0.0007407| 0.3468758|  1.434147| 0.5034253|
```

From the table above, we see that only the p-value of contrast between HDGen and HDName is larger than α=0.05.

Thus, all contrasts have significant pairwise differences in mean lifetime of different battery types except the contrast between HDGen and HDName.

8.



First, we try this model:

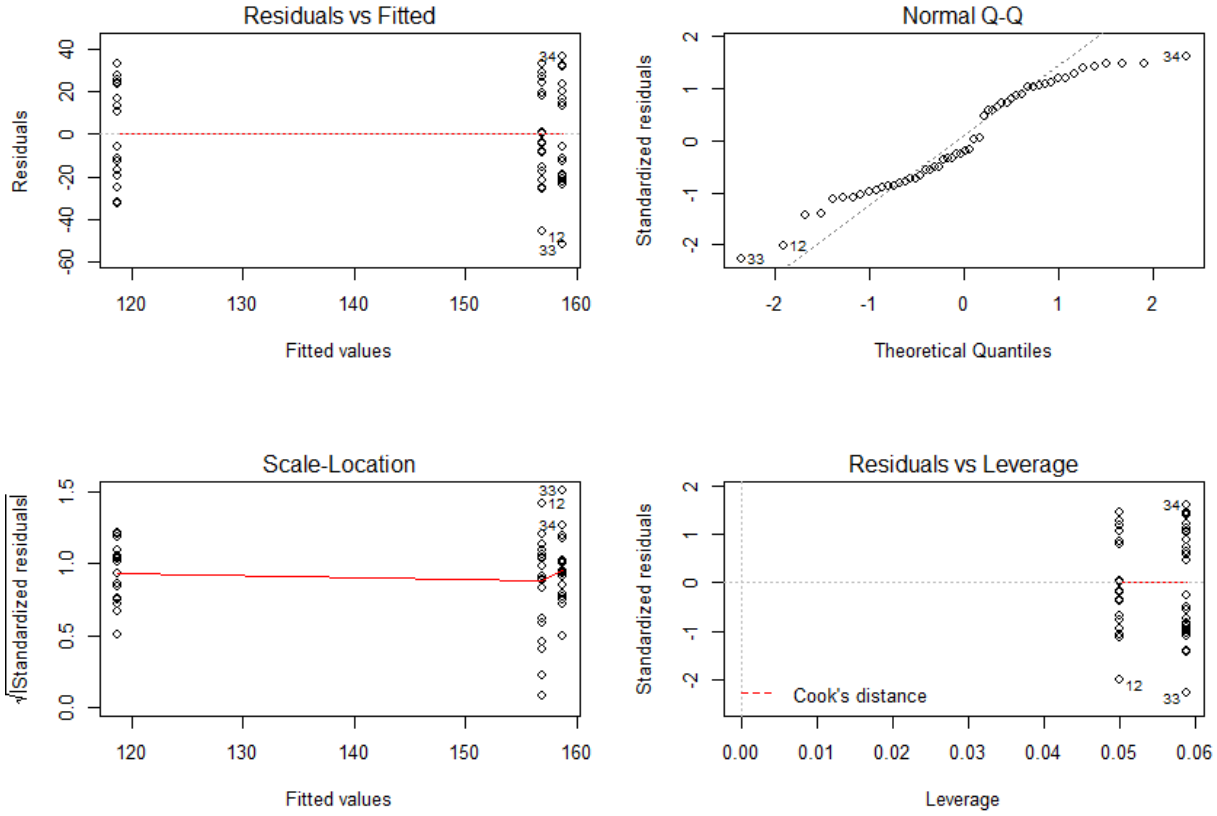
$Y_{it} = \mu + \tau_i + \epsilon_{it}$, $\epsilon_{it} \sim N(0, \sigma^2)$ with iid.

Response is simply the Calories

i = 1, 2, 3

t = 1, …, r_i where $r_1 = 20$, $r_2 = 17$, $r_3 = 17$

Then, we look at the residual plots:



From the Q-Q plot, we can observe that residuals do not fit the line well at the right upper corner, which indicates that residuals in this model may violate the normality assumption.

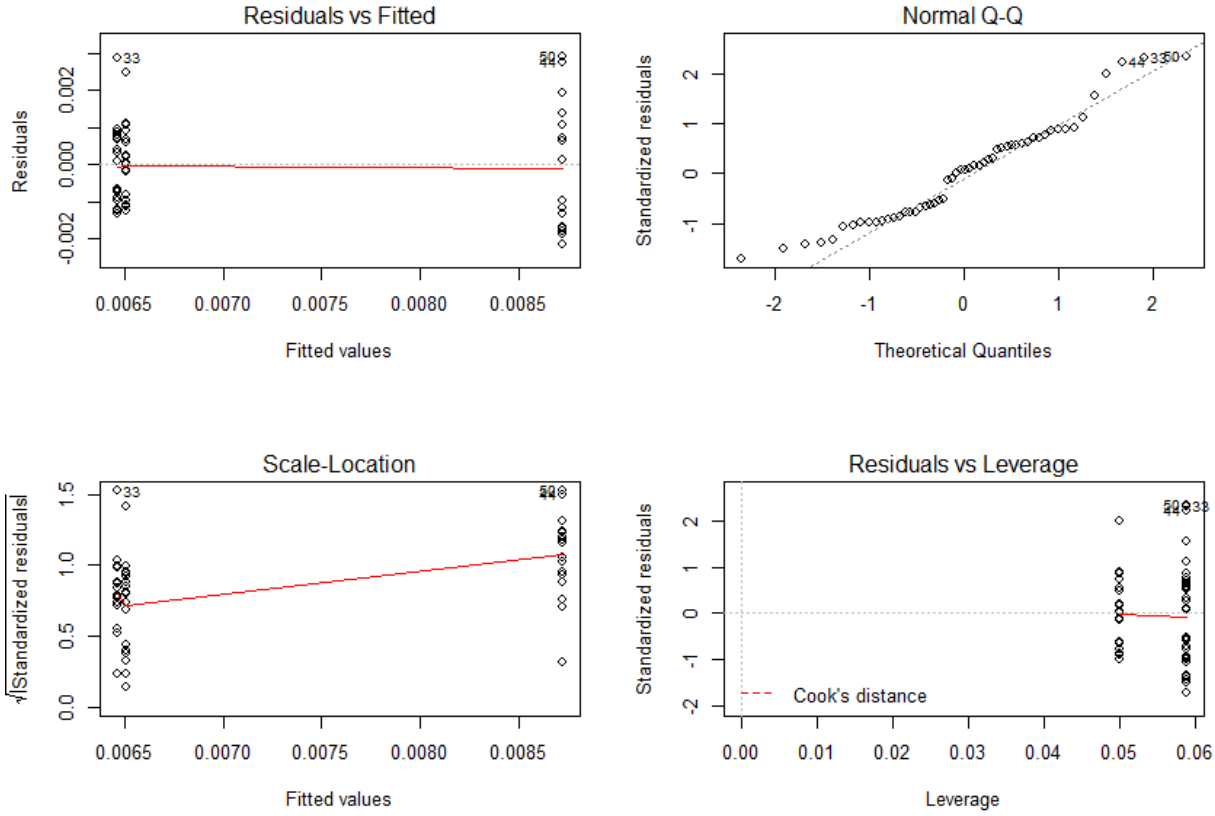
Thus, we use transformations and try other 5 models and finally select the following model (coded as m5 in R):

$1/Y_{it} = \mu + \tau_i + \epsilon_{it}$, $\epsilon_{it} \sim N(0, \sigma^2)$ with iid.

Response is the Inverse of Calories

i = 1, 2, 3

t = 1, …, r_i where $r_1 = 20$, $r_2 = 17$, $r_3 = 17$



By Q-Q plot, we observe that residuals fit the line well, which indicates that residuals satisfy normality.

By the scatter plot, we observe that variances of residuals are approximately constant.

By the scatter plot, we can observe a linear trend among residuals, which satisfies the linearity.

Thus, assumptions of ANOVA model are all justified. Then, we can analyze data by using this model.

```
> summary(m5)
             Df    Sum Sq   Mean Sq F value   Pr(>F)
Type          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
```

H_0: $\tau_1=\tau_2=\tau_3$

H_1: At least one of τ_i is different with others.

$T^* = F_{2,51} = 17.61$

p-value = $1.53 \times 10^{-6} < \alpha = 0.05$

Thus, reject null hypothesis and conclude that there is enough evidence to show that at least one of τ_i is different with others for i=1,2,3.

```
> lsm.hotdog=lsmeans(m5, ~Type)
> kable(summary(contrast(lsm.hotdog,method="pairwise",adjust="tukey"),infer=c(T,T),level=0.5,side
="two-sided"))
```

contrast	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
Beef - Chicken	-0.0022119	0.0004235	51	-0.0026907	-0.0017331	-5.2226554	0.0000097
Beef - Pork	0.0000438	0.0004235	51	-0.0004350	0.0005226	0.1034495	0.9941180
Chicken - Pork	0.0022557	0.0004404	51	0.0017579	0.0027535	5.1224837	0.0000138

From the table above, we see that only the p-value of contrast between Beef and Pork is larger than α=0.05.

Beef hotdog has larger mean hotdog calories than Chicken hotdog.

Pork hotdog has larger mean hotdog calories than Chicken hotdog.

Contrast between Beef and Pork does NOT have significant pairwise differences in mean hotdog calories of different meat types

No other significant pairwise differences in mean hotdog calories of different meat types.

R code:

```
type<-
c("AlkName","AlkName","AlkName","AlkName","AlkGen","AlkGen","AlkGen","AlkGen","HDName","HDN
ame","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)

plot(batt)

fit1<-aov(life~type,data=batt)
par(mfrow=c(2,2))
plot(fit1)

batt$SqrtLife<-sqrt(life)
batt$LogLife<-log(life)
batt$SqLife<-life*life

fit2<-aov(SqrtLife~type,data = batt)
par(mfrow=c(2,2))
plot(fit2)

fit3<-aov(LogLife~type,data = batt)
par(mfrow=c(2,2))
plot(fit3)

fit4<-aov(SqLife~type, data = batt)
par(mfrow=c(2,2))
plot(fit4)

summary(fit3)
install.packages("knitr")
library(knitr)
install.packages("lsmeans")
library(lsmeans)
lsm.life=lsmeans(fit3, ~type)
kable(summary(contrast(lsm.life,method="pairwise",adjust="tukey"),infer=c(T,T),level=0.5,side="two-
sided"))

#8-----------------------------------------------------------------------------------------------------------
hotdog=read.table("hotdogs.txt",header=TRUE)
hotdog
attach(hotdog)
plot(Calories~Type)

m1<-aov(Calories~Type,data=batt)
```

```
par(mfrow=c(2,2))
plot(m1)

hotdog$SqrtC<-sqrt(Calories)
hotdog$LogC<-log(Calories)
hotdog$SqC<-Calories*Calories
hotdog$InvC<-1/Calories
hotdog$TrirtC<-Calories^(1/3)

m2<-aov(SqrtC~Type,data =hotdog)
par(mfrow=c(2,2))
plot(m2)

m3<-aov(LogC~Type,data = hotdog)
par(mfrow=c(2,2))
plot(m3)

m4<-aov(SqC~Type, data = hotdog)
par(mfrow=c(2,2))
plot(m4)

m5<-aov(InvC~Type, data = hotdog)
par(mfrow=c(2,2))
plot(m5)

m6<-aov(TrirtC~Type, data = hotdog)
par(mfrow=c(2,2))
plot(m6)

summary(m5)
lsm.hotdog=lsmeans(m5, ~Type)
kable(summary(contrast(lsm.hotdog,method="pairwise",adjust="tukey"),infer=c(T,T),level=0.5,side="two-sided"))
```