

# HW7

Adhrit Srivastav, eid: Ams22362

11/8/2021

## 6.5 Running dogs

- a) The response is the running speed of the dogs.
- b) Vitamin C diet or not and the dogs.
- c) Experimental for the 3 diets given thus 3 levels. Observational for the 5 greyhounds thus 5 levels.
- d) Yes, complete block design because each of the 5 greyhounds got all 3 diets.

## 6.6 Fat rats

- a) Amount eaten by the rat.
- b) Leptin or insulin shot. Male or female rat.
- c) Experimental for the leptin or insulin shot thus 2 levels. Observational for the male or female rats thus 2 levels.
- d) Potentially not complete block design because the male and female rats were randomly assigned the shots, so there's no guarantee all combinations were applied.

## 6.27 Running dogs: df

- a) With blocks: Diets = 2 ( $I - 1$ ), Blocks = 4 ( $J - 1$ ), Error/Residual = 8 ( $(I - 1)(J - 1)$ ), Total = 14 ( $IJ - 1$ )
- b) No blocks/Completely Randomized: Diets = 2, Error/Residual = 2
- b) The advantage of the No Blocks experimental design is that it removes the possibility of carry-over effects of the diets, but it potentially adds more variability due to there being more test subjects. The advantage of the block design is that it uses fewer dogs, but it'll take longer. It has less degrees of freedom for error/residual.

## 6.28 Migraines

- a) Give all 4 volunteers the 4 medications. The units would be each of the 4 people. I will assign the treatments randomly to each of the units. Each subject is a block, time slot experimental unit. All volunteers will get a different treatment at a different time slot. They will be assigned in a way that each subject gets a different treatment and there are no overlaps.
- b) In the attached image.

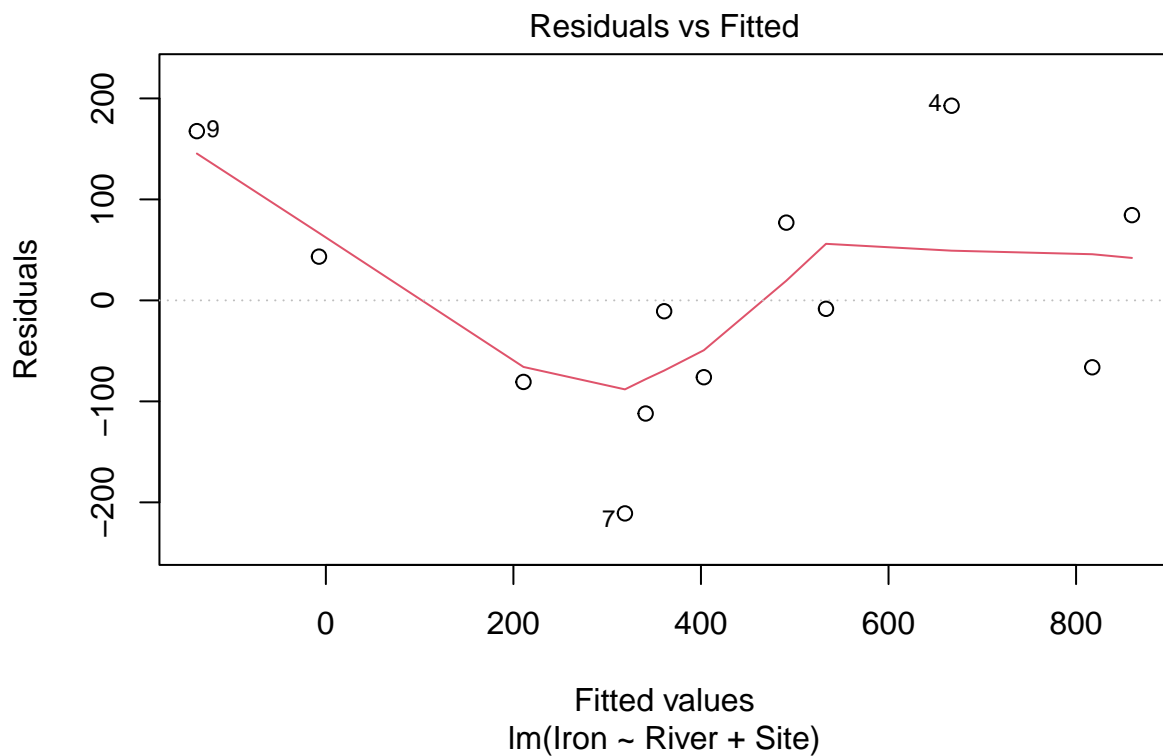
## 6.30 River iron

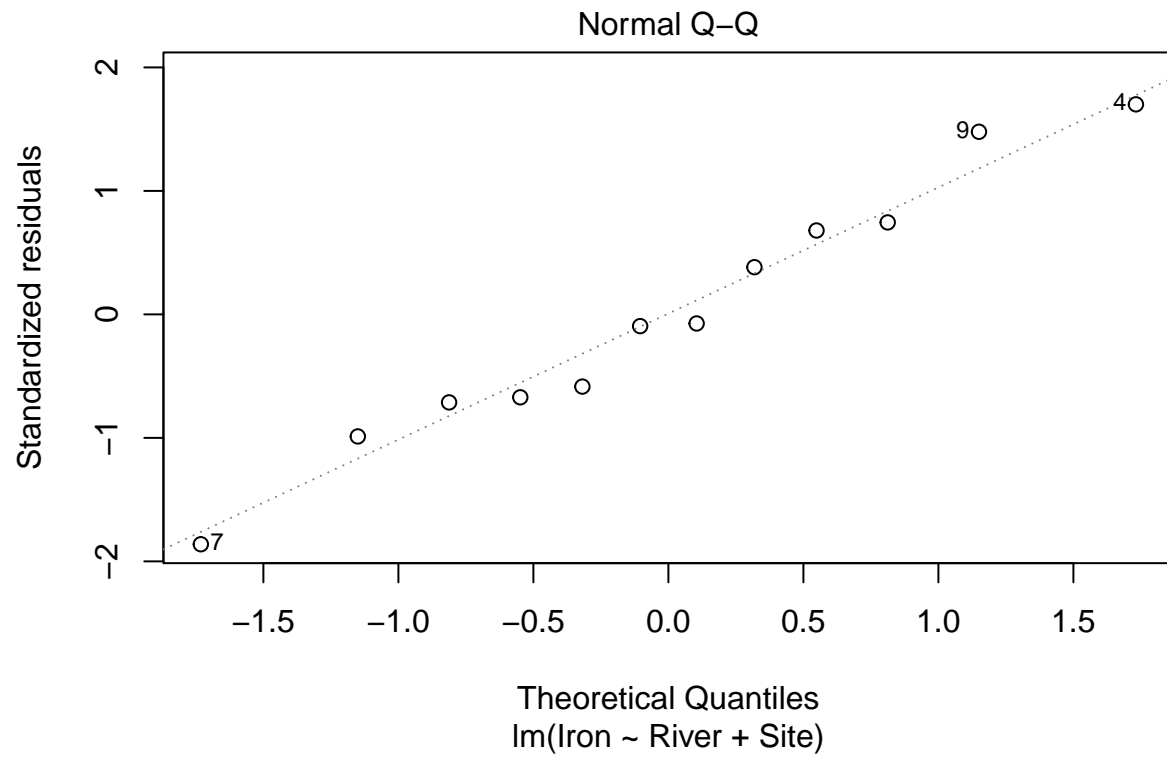
```
iron = read.csv("C:/Users/adhri/OneDrive/Documents/R/App_Reg_and_Time_Series/chpt6/datasets/Riveriron.csv")
attach(iron)
```

```
aov2I = aov(Iron ~ River + Site)
summary(aov2I)
```

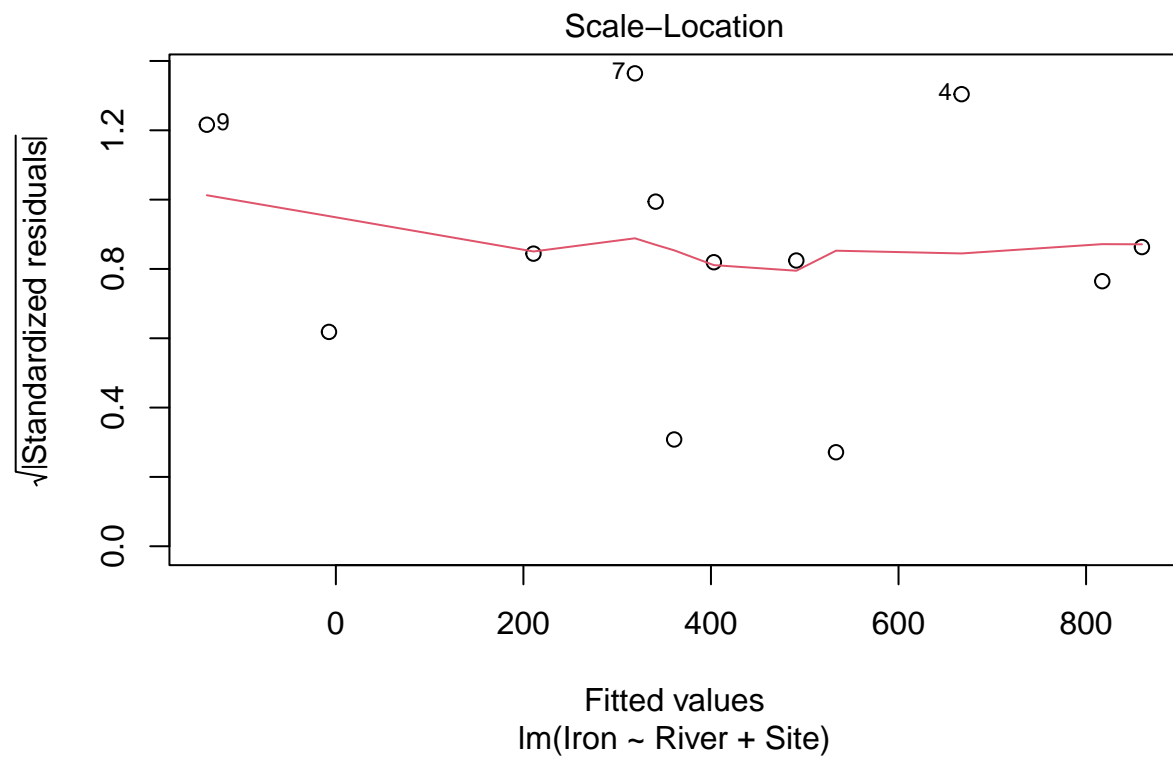
```
##           Df Sum Sq Mean Sq F value Pr(>F)
## River      3 542458  180819   7.041 0.0216 *
## Site       2 442395   221198   8.613 0.0172 *
## Residuals   6 154083    25680
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
modell = lm(Iron ~ River + Site)
plot(modell)
```





```
## hat values (leverages) are all = 0.5
## and there are no factor predictors; no plot no. 5
```



- a) Outputted above.
- b) The Q-Q Norm plot appears straight thus the data appears normal.
- c) There is a curve to the residuals, thus it might not pass the linearity condition. There don't appear to be pronounced clusters. The data might be fanning out which means it might not pass the equal variance condition.
- d) The estimated fitted value of the leftmost point is -175. This one fitted value strongly suggests that the model is not appropriate because it should not be negative.

## 6.36 Alfalfa sprouts

```
library(mosaic)
```

```
## Registered S3 method overwritten by 'mosaic':  
##   method                from  
##   fortify.SpatialPolygonsDataFrame ggplot2
```

```
##  
## The 'mosaic' package masks several functions from core packages in order to add  
## additional features. The original behavior of these functions should not be affected by this.
```

```
##  
## Attaching package: 'mosaic'
```

```
## The following objects are masked from 'package:dplyr':
##
##   count, do, tally
```

```
## The following object is masked from 'package:Matrix':
##
##   mean
```

```
## The following object is masked from 'package:ggplot2':
##
##   stat
```

```
## The following objects are masked from 'package:stats':
##
##   binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,
##   quantile, sd, t.test, var
```

```
## The following objects are masked from 'package:base':
##
##   max, mean, min, prod, range, sample, sum
```

```
alfalfa = read.csv("C:/Users/adhri/OneDrive/Documents/R/App_Reg_and_Time_Series/chpt6/datasets/Alfalfa.
attach(alfalfa)
```

```
# a
favstats(Ht4 ~ factor(Row))
```

```
##   factor(Row)  min    Q1 median    Q3  max mean      sd n missing
## 1          a 1.00 1.015   1.03 1.240 1.45 1.16 0.2515949 3      0
## 2          b 0.70 0.960   1.22 2.005 2.79 1.57 1.0880717 3      0
## 3          c 0.45 0.910   1.37 1.650 1.93 1.25 0.7472617 3      0
## 4          d 1.65 1.990   2.33 2.565 2.80 2.26 0.5781868 3      0
## 5          e 1.07 1.265   1.46 3.155 4.85 2.46 2.0789661 3      0
```

```
favstats(Ht4 ~ factor(Acid))
```

```
##   factor(Acid)  min    Q1 median    Q3  max mean      sd n missing
## 1      1.5HCl 0.70 1.00   1.37 1.46 2.80 1.466 0.8050963 5      0
## 2      3.0HCl 0.45 1.03   1.07 1.22 1.65 1.084 0.4312540 5      0
## 3      water 1.45 1.93   2.33 2.79 4.85 2.670 1.3151426 5      0
```

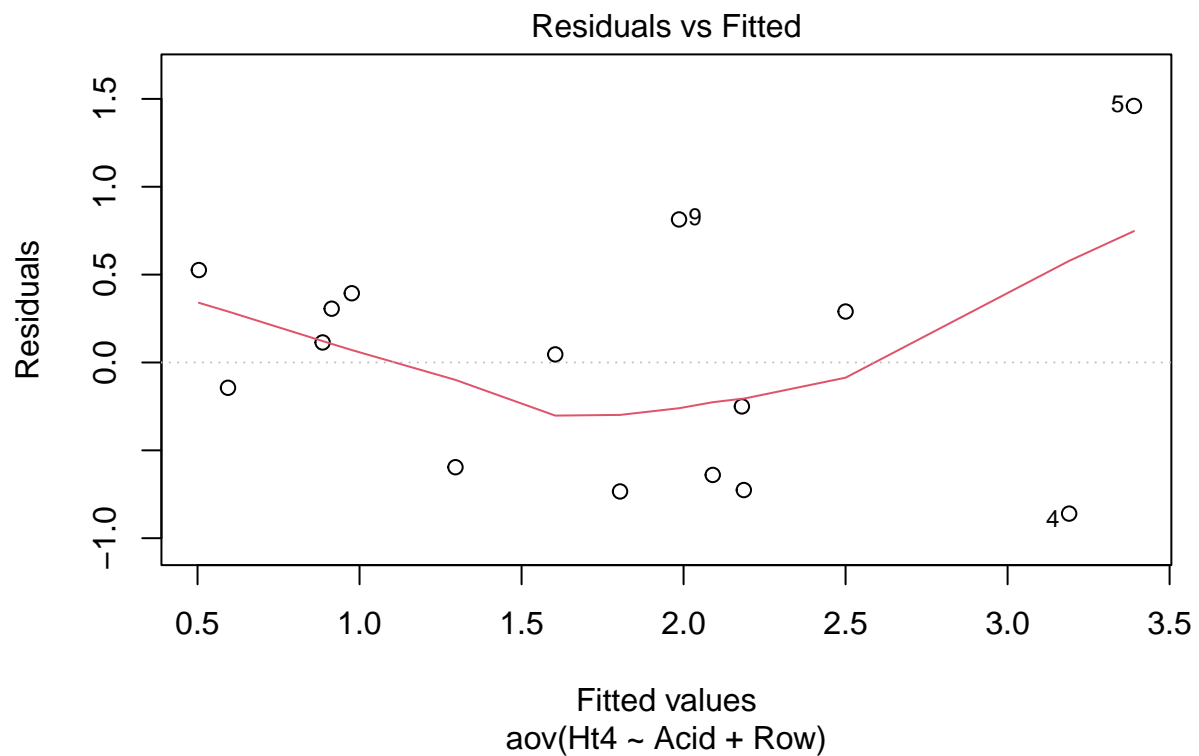
```
favstats(Ht4)
```

```
##   min    Q1 median    Q3  max mean      sd n missing
## 0.45 1.05   1.45 2.13 4.85 1.74 1.105396 15      0
```

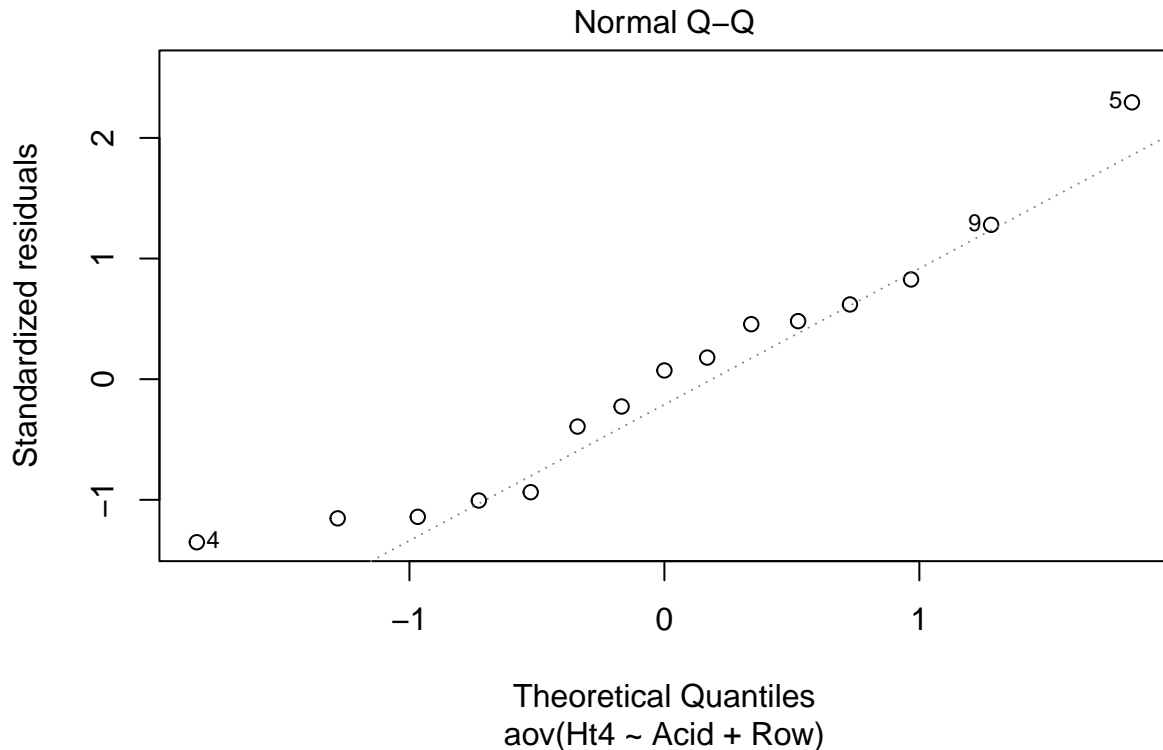
```
# b
aov2A = aov(Ht4 ~ Acid + Row)
summary(aov2A)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Acid       2  6.852   3.426   4.513 0.0487 *
## Row        4  4.183   1.046   1.378 0.3235
## Residuals   8  6.072   0.759
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# c
plot(aov2A, which = 1)
```



```
plot(aov2A, which = 2)
```



- Means for rows of cups and treatments outputted above. For all 15 cups average = 1.74 and sd = 1.105.
- In the above code.
- There is slight curvature in the residual plot but it appears to pass the linearity test. The Q-Q Norm plot has slight deviations from the line so it might not pass the normality condition.
- I would conclude there is significant difference in average growth due to treatments because the p-value of .0487 is less than alpha and is thus significant.
- I would not conclude there is significant difference in average growth due to distance from the window because the p-value of .3235 is greater than alpha and is thus not significant.

### 6.37 Alfalfa sprouts: Fisher's LSD

```
library(mosaic)
library(emmeans)

alfalfa = read.csv("C:/Users/adhri/OneDrive/Documents/R/App_Reg_and_Time_Series/chpt6/datasets/Alfalfa.")
attach(alfalfa)

## The following objects are masked from alfalfa (pos = 4):
##
##   Acid, Ht4, Row
```



```
aovA = aov(Ht4 ~ Acid)
```

```
Fitted.Model<-emmeans(aovA, ~factor(Acid))  
Fitted.Model
```

```
##  Acid    emmean      SE df lower.CL upper.CL  
##  1.5HCl    1.47 0.413 12    0.565    2.37  
##  3.0HCl    1.08 0.413 12    0.183    1.98  
##  water     2.67 0.413 12    1.769    3.57  
##  
## Confidence level used: 0.95
```

```
pairs(Fitted.Model, adjust='none')    # the estimate is the mean difference between groups
```

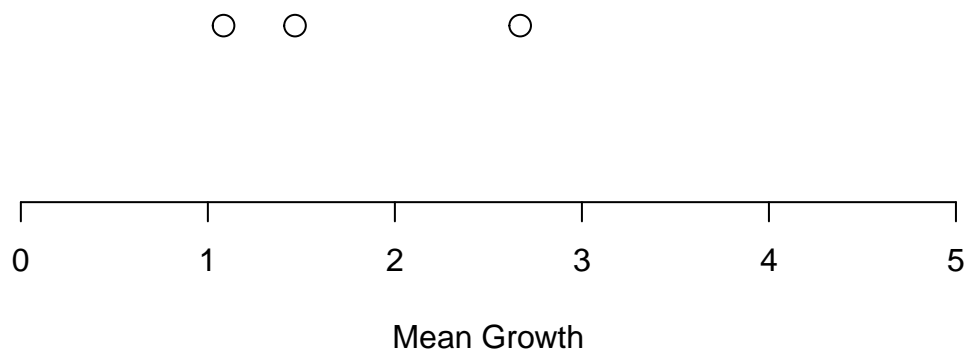
```
##  contrast      estimate      SE df t.ratio p.value  
##  1.5HCl - 3.0HCl    0.382 0.585 12   0.653  0.5258  
##  1.5HCl - water    -1.204 0.585 12  -2.059  0.0618  
##  3.0HCl - water    -1.586 0.585 12  -2.713  0.0189
```

```
FFStats=favstats(Ht4 ~ factor(Acid))
```

```
tstar=qt(0.975,aovA$df.residual)  
MSE=summary(aovA)[[1]]$"Mean Sq"[2]    #gets the MSE from aov summary  
ni=FFStats$n[1]  
LSD=tstar*sqrt(MSE)*sqrt(1/ni+1/ni)  
round(LSD,2)
```

```
## [1] 1.27
```

```
y=rep(2,3)  
plot(FFStats$mean,y, xlim=c(0,5), ylim=c(1,4), yaxt="n",ylab="", bty = "n", xlab="Mean Growth",cex=1.5)
```



```
#rest below is labeling the points and showing LSD
FFMean=FFStats$mean
```

The LSD value was calculated to be 1.2738, so we are looking for estimates greater than that to declare those differences as significant. The only significant difference was between 3.0Hcl - water. The p-value was also .0189 which is less than alpha thus the levels are different.

## 6.38 Unpopped popcorn

```
popcorn = read.csv("C:/Users/adhri/OneDrive/Documents/R/App_Reg_and_Time_Series/chpt6/datasets/Popcorn.csv")
attach(popcorn)

# a
favstats(Unpopped)
```

```
##  min Q1 median   Q3 max    mean      sd  n missing
##    6 14   23.5 35.5  47 25.58333 14.29214 12      0
```

```
favstats(Unpopped ~ factor(Brand))
```

```
##  factor(Brand) min    Q1 median   Q3 max    mean      sd n missing
## 1      Orville   6  9.50  16.0 24.0 35 17.83333 11.07098 6      0
## 2      Seaway  14 24.25  35.5 44.5 47 33.33333 13.51542 6      0
```

```
favstats(Unpopped ~ factor(Acid))
```

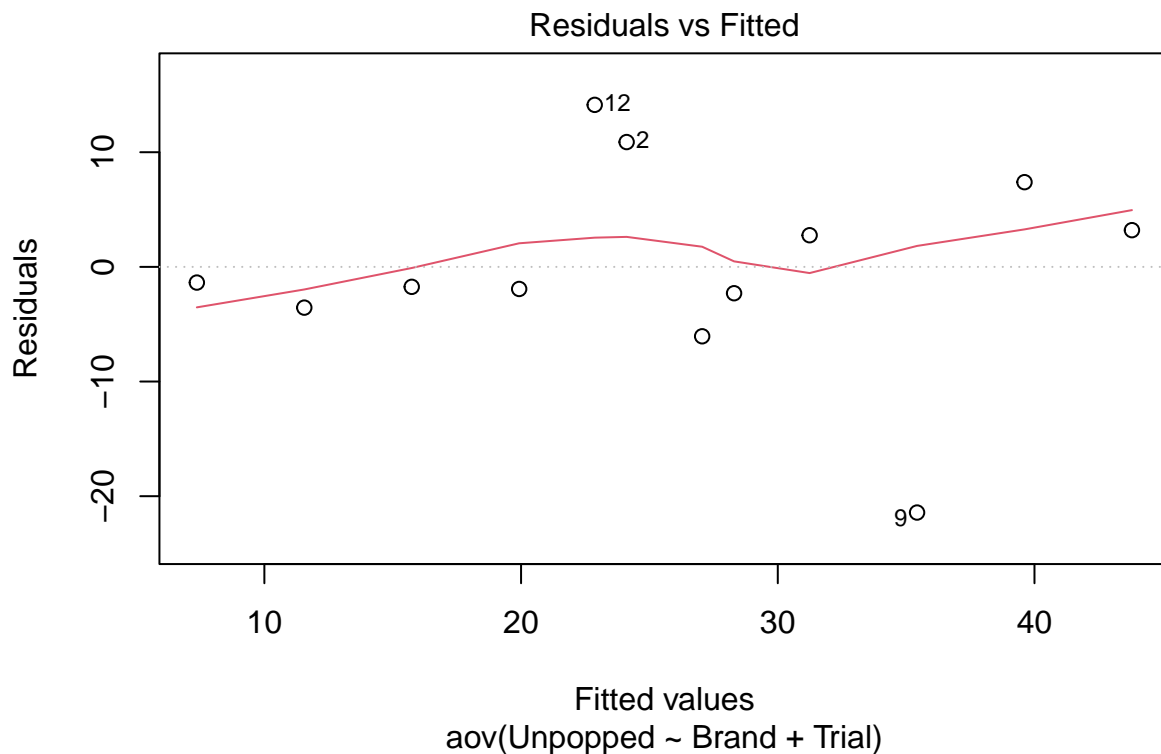
```
## Warning in split.default(evalF$left[, 1], mosaicCore::joinFrames(evalF$right, :  
## data length is not a multiple of split variable
```

```
##   factor(Acid) min Q1 median Q3 max mean      sd n missing  
## 1      1.5HCl   6 14      34 47  47 29.6 18.87591 5      0  
## 2      3.0HCl  21 25      29 33  37 29.0 11.31371 2      0  
## 3      water   8 14      18 26  35 20.2 10.54514 5      0
```

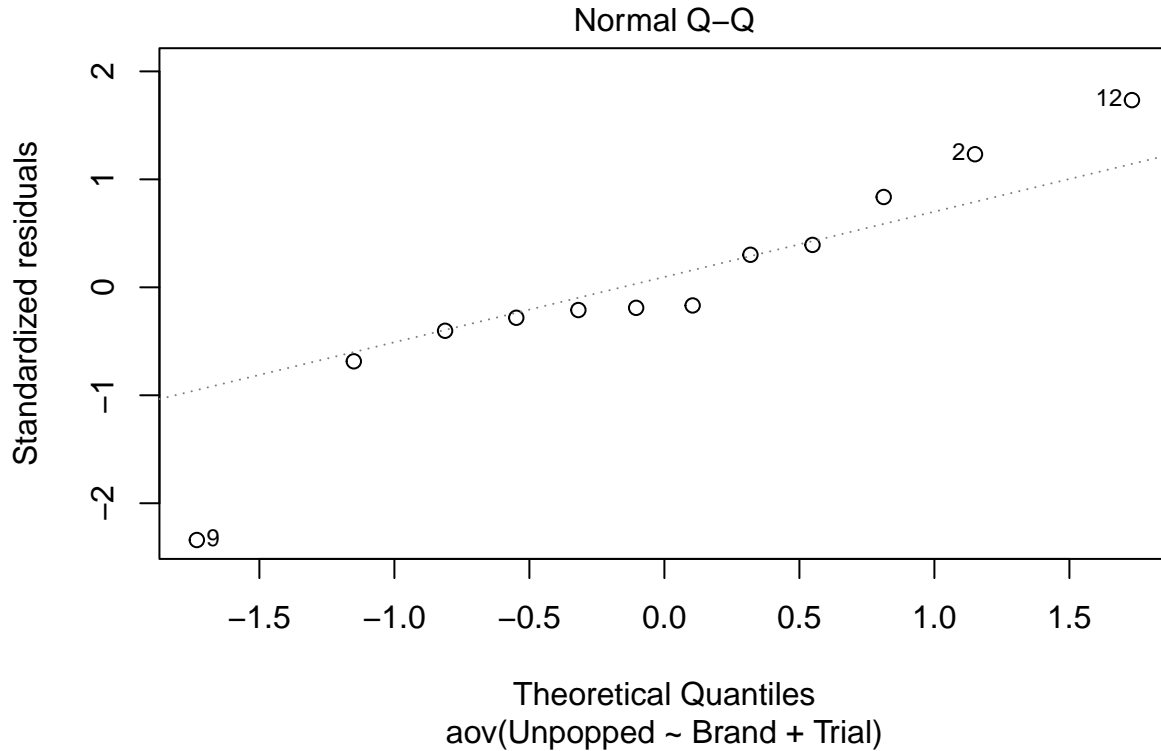
```
# b  
aov2P = aov(Unpopped ~ Brand + Trial)  
summary(aov2P)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)  
## Brand      1  720.7    720.7    7.105 0.0258 *  
## Trial      1   613.2    613.2    6.045 0.0362 *  
## Residuals  9   913.0    101.4  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(aov2P, which = 1)
```



```
plot(aov2P, which = 2)
```

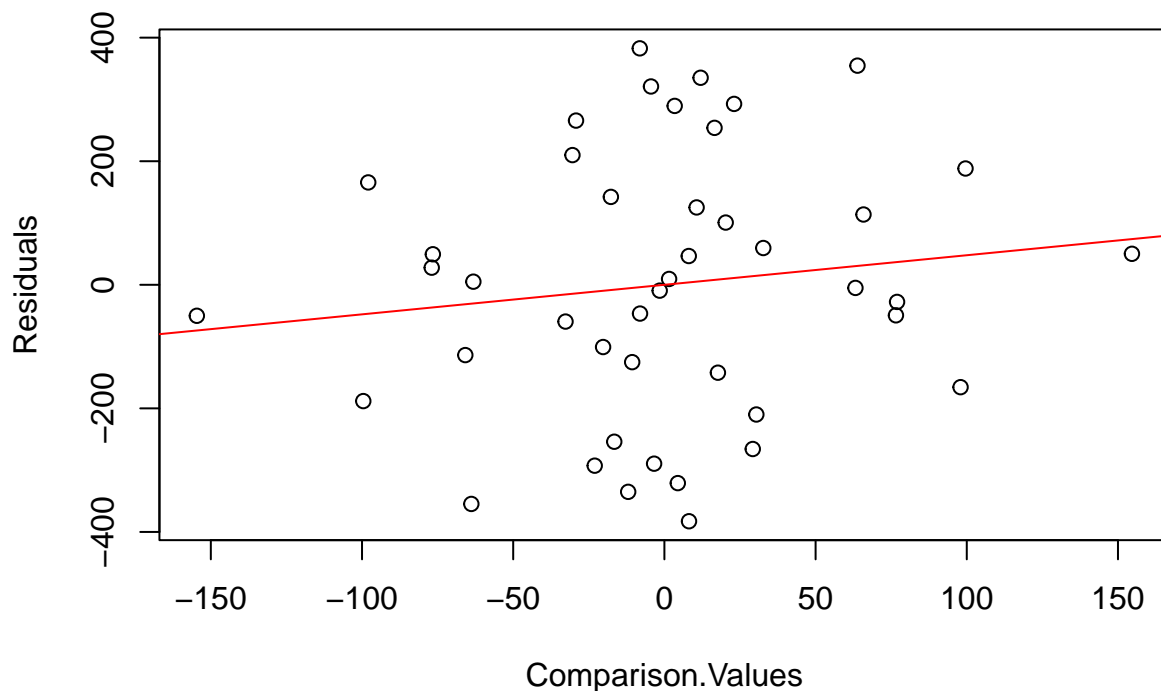


- The mean = 25.5833 for the entire sample. The estimated effects of the brand of the popcorn is that Seaway has more unpopped kernels based on the mean unpopped kernels for each brand.
- Two-way ANOVA is in the output above. The Q-Q Norm plot has points not in the straight line so it might not pass the normality test. There is not too much fanning in the plot so it likely passes equal variance.
- The brand does appear to make a difference in the mean number of unpopped kernels as the p-value is .0258 which is less than alpha and thus significant. The trial does appear to make a difference in the mean number of unpopped kernels as the p-value is .0362 which is less than alpha and thus significant.

#### 6.40 Oral contraceptives: Tukey additivity plot

```
contraceptive = read.csv("C:/Users/adhri/OneDrive/Documents/R/App_Reg_and_Time_Series/chpt6/datasets/Con
attach(contraceptive)

plot(Residuals ~ Comparison.Values)
model0 = lm(Residuals ~ Comparison.Values)
abline(model0, col = "Red")
```



```
summary(model0)
```

```
##
## Call:
## lm(formula = Residuals ~ Comparison.Values)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -386.6 -143.1    0.0  143.1  386.6
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.281e-16  3.212e+01   0.000    1.000
## Comparison.Values  4.784e-01  5.598e-01   0.855    0.398
##
## Residual standard error: 213 on 42 degrees of freedom
## Multiple R-squared:  0.01709,    Adjusted R-squared:  -0.006314
## F-statistic: 0.7302 on 1 and 42 DF,  p-value: 0.3977
```

Once we plot residuals versus comparison values, we can see the data relatively follows the line. This means the additive model will decently fit the transformed data. Although the points were scattered, there were so many data points that they sent a consistent message. Transforming to logs was indicated. The slope of the line is relatively zero, so there is no need to express the additive model.

## 7.7 Heart and soul

H	S	Diff
---	---	------

D	2.5 11 -8.5	R 6 9.5 -3.5	Diff 3.5 1.5
---	-------------	--------------	--------------

Yes, there is interaction present because when we compare the differences of the means of the rows and columns, they are different.

## 7.8 Blood, sweat, and tears

B	S	T	Diff
---	---	---	------

M	7.5 15 15 -22.5	F 17.5 25 25 -32.5	Diff -10 -10 -10
---	-----------------	--------------------	------------------

There is not interaction present because when we compare the differences of the means of the rows and columns, they are roughly equal.

## 7.14 Hypothetical decomposition #1

MSE, MSAB, MSB, MSA

## 7.15 Hypothetical decomposition #2

MSB, MSE, MSA, MSAB