

marks_proj_work

Mark Austin

2022-11-15

```
#using Halid's code to keep data frame consistent
library(tidyverse)
df_eff <- read_csv('effervescence.csv', col_types = 'ffnmm')
```

Exploratory Analysis

Data Preparation

The Effervescent Experiment data contains 48 rows, 6 columns and no missing data points. As a result of having intact data there will not be a need for data remediation.

The data set contains three explanatory variables; Brand, Temp and Stirred. Brand accounts for cold medication that are either name brand vs store brand. Stirred considers observations where the water was agitated or not. Temperature data points record the water temperature at specific predetermined levels—6°C, 23°C, and 40°C. The data set also contains a response variable labeled Time. Here are the first few rows of the data to illustrate the structure of our dataset.

```
## # A tibble: 6 x 6
##   Brand Temp  Stirred Order  Time OrgTime
##   <fct> <fct> <fct>   <dbl> <dbl>   <dbl>
## 1 name   6     yes       8  77.2    75.5
## 2 name  23     yes       3  75.4    68.1
## 3 name  40     yes       7  68.1    44.8
## 4 store 6     yes       1  77.9    78.4
## 5 store 23     yes       2  66.4    40.6
## 6 store 40     yes      18  59.8    27.4
```

Summary Statistics

```
# Time, Generalized
summary(df_eff$Time)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  58.24   66.09   70.92   70.77   76.93   80.92
```

```
df_eff %>%
  group_by(Brand) %>%
  summarize(min = min(Time),
            q1 = quantile(Time, 0.25),
            median = median(Time),
            mean = mean(Time),
            q3 = quantile(Time, 0.75),
            max = max(Time))
```

```
## # A tibble: 2 x 7
##   Brand   min    q1 median  mean    q3   max
##   <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 name   64.5  70.3  75.6  73.4  76.4  79.7
## 2 store  58.2  59.3  66.7  68.1  77.5  80.9
```

```
df_eff %>%
  group_by(Temp) %>%
  summarize(min = min(Time),
            q1 = quantile(Time, 0.25),
            median = median(Time),
            mean = mean(Time),
            q3 = quantile(Time, 0.75),
            max = max(Time))
```

```
## # A tibble: 3 x 7
##   Temp   min    q1 median  mean    q3   max
##   <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 6      75.8  77.0  77.9  78.0  79.2  80.9
## 2 23     65.9  66.9  68.5  70.7  75.6  76.8
## 3 40     58.2  59.1  62.1  63.6  68.6  71.2
```

```
df_eff %>%
  group_by(Stirred) %>%
  summarize(min = min(Time),
            q1 = quantile(Time, 0.25),
            median = median(Time),
            mean = mean(Time),
            q3 = quantile(Time, 0.75),
            max = max(Time))
```

```
## # A tibble: 2 x 7
##   Stirred   min    q1 median  mean    q3   max
##   <fct>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 yes     58.2  65.7  68.6  69.6  75.9  77.9
## 2 no      58.5  67.7  73.6  72.0  78.4  80.9
```

Visualization

< add >

Model Assumptions

Outliers

All of the data falls within two standard deviations of the mean and there does not appear to be any outliers.

```
sd(df_eff$Time)
```

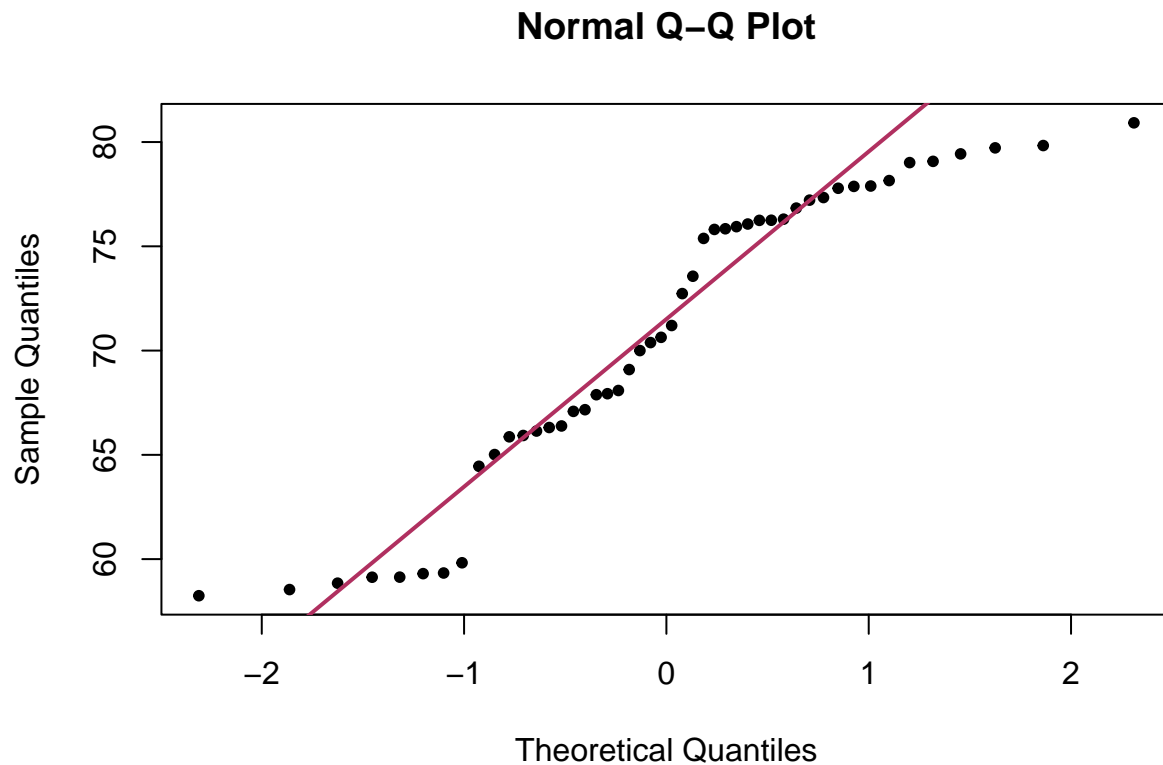
```
## [1] 7.10074
```

<add Cooks D discussion/check here as well.

Normality

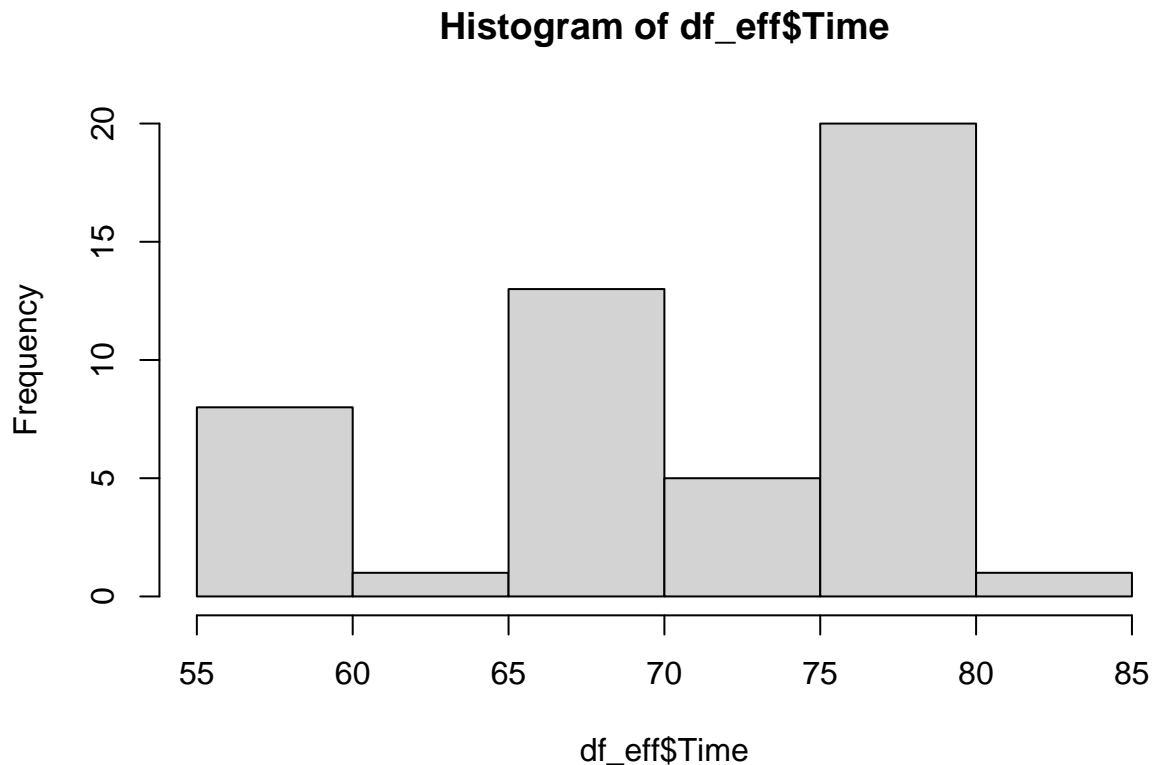
We want to check the residuals of our data for normality since many of the tests performed will rely on this as an assumption. Using a quantile-quantile plot, we can see that our observed residual values deviate from what would be considered normal. Although the plot trends upward diagonally, there are two rapid transitions in our residuals which indicate the possibility of bimodality. From the tail-ends of our Q-Q plot a right-skew and/or outliers may also be present. Before beginning our analysis, a histogram check may reveal the presence of a multimodal distribution and tests such as Shapiro-Wilk can check for normality while a boxplot can check shape of distribution and check for outliers.

```
options(repr.plot.width = 8, repr.plot.height = 6, repr.plot.res = 200)
qqnorm(df_eff$Time, pch = 20)
qqline(df_eff$Time, col = "maroon", lwd = 2)
```



A histogram of the Time response variable further illustrates multi-modality. This would be indicative of a pattern in the Time variable as a result of the explanatory variables. In a thorough analysis, we will test the means of time on these factors to see how they vary.

```
hist(df_eff$Time)
```



Homogeneity of Variance < homoscadascity visual illustration>

Homogeneity of covariances Analogue of homoscadascity for multivariate analysis

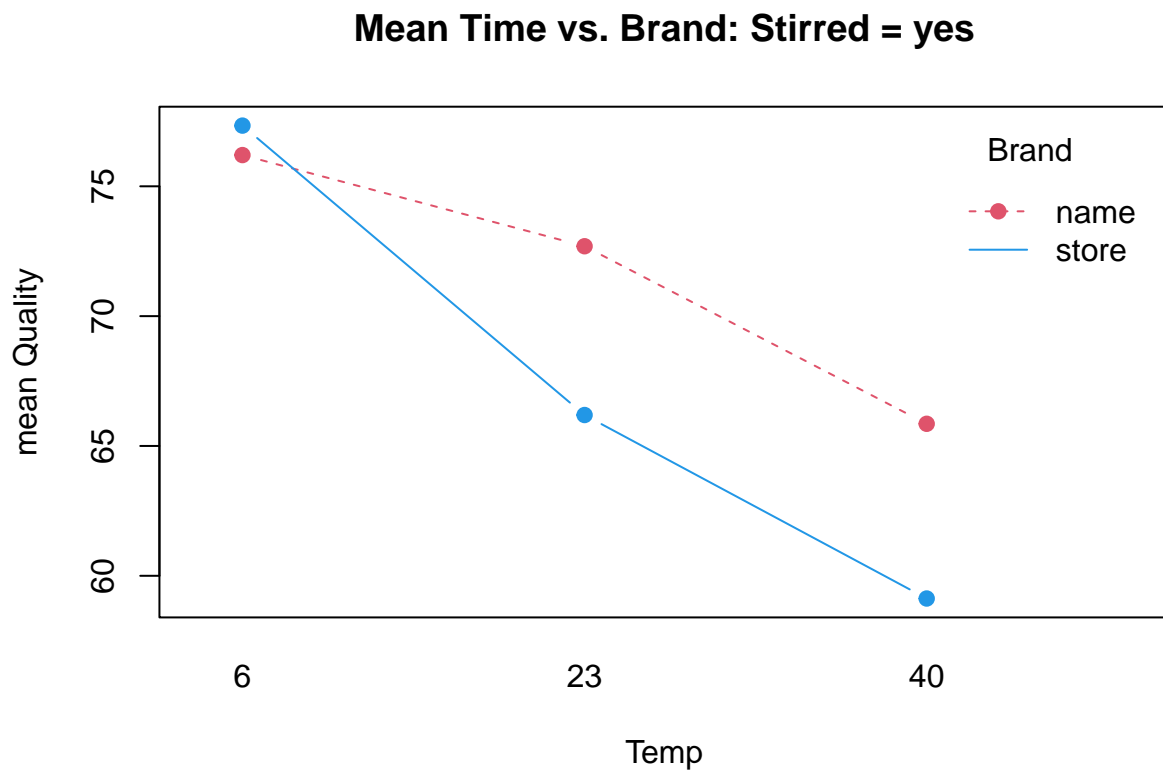
Exploratory Analysis (Original)

```
#using Halid's code to keep data frame consistent  
library(tidyverse)  
df_eff <- read_csv('effervescence.csv', col_types = 'ffnmm')  
df_eff %>% head()
```

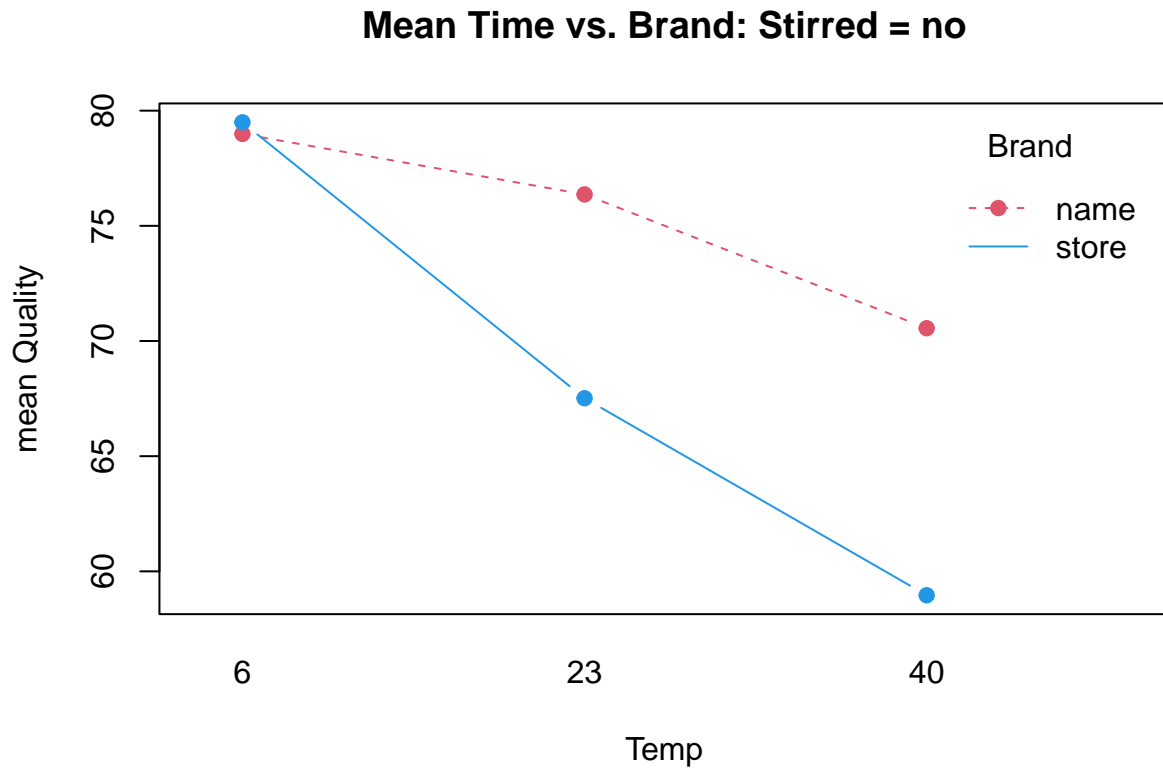
```
## # A tibble: 6 x 6
##   Brand Temp  Stirred Order  Time OrgTime
##   <fct> <fct> <fct>   <dbl> <dbl>   <dbl>
## 1 name   6     yes       8  77.2    75.5
## 2 name  23     yes       3  75.4    68.1
## 3 name  40     yes       7  68.1    44.8
## 4 store  6     yes       1  77.9    78.4
## 5 store 23     yes       2  66.4    40.6
## 6 store 40     yes      18  59.8    27.4
```

##3 factor interaction plot based on HW7 code

```
with(df_eff%>%filter(Stirred=="yes"),interaction.plot(Temp,Brand,Time,
  type="b", pch=19, col=c(2,4), ylab="mean Quality",
  main="Mean Time vs. Brand: Stirred = yes"))
```



```
with(df_eff%>%filter(Stirred=="no"),interaction.plot(Temp,Brand,Time,
  type="b", pch=19, col=c(2,4), ylab="mean Quality",
  main="Mean Time vs. Brand: Stirred = no"))
```



From the three factor interaction plots, does the 3 factor interaction look obvious here? The brand by temp interaction is clear.

Analysis and Results

```
#model with stirred as block effect without interaction
aov_block_eff <- aov(lm_block_eff <- lm(Time ~ Brand * Temp + Stirred, data = df_eff))
summary(lm_block_eff)
```

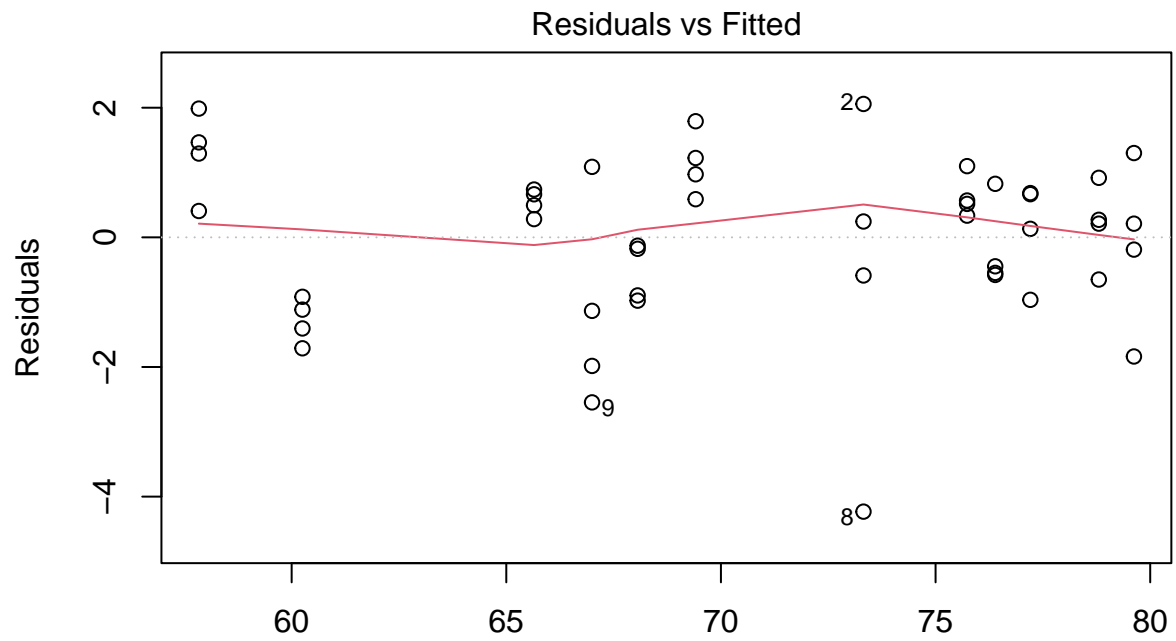
```
##
## Call:
## lm(formula = Time ~ Brand * Temp + Stirred, data = df_eff)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.2315 -0.7120  0.2577  0.7596  2.0577
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    76.3899    0.5035 151.710 < 2e-16 ***
## Brandstore      0.8182    0.6593   1.241   0.222
## Temp23         -3.0690    0.6593  -4.655 3.38e-05 ***
## Temp40         -9.3922    0.6593 -14.246 < 2e-16 ***
## Stirredno       2.4133    0.3806   6.340 1.42e-07 ***
## Brandstore:Temp23 -8.4923    0.9323  -9.108 2.11e-11 ***
## Brandstore:Temp40 -9.9781    0.9323 -10.702 1.93e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.319 on 41 degrees of freedom
## Multiple R-squared:  0.9699, Adjusted R-squared:  0.9655
## F-statistic: 220.3 on 6 and 41 DF,  p-value: < 2.2e-16
```

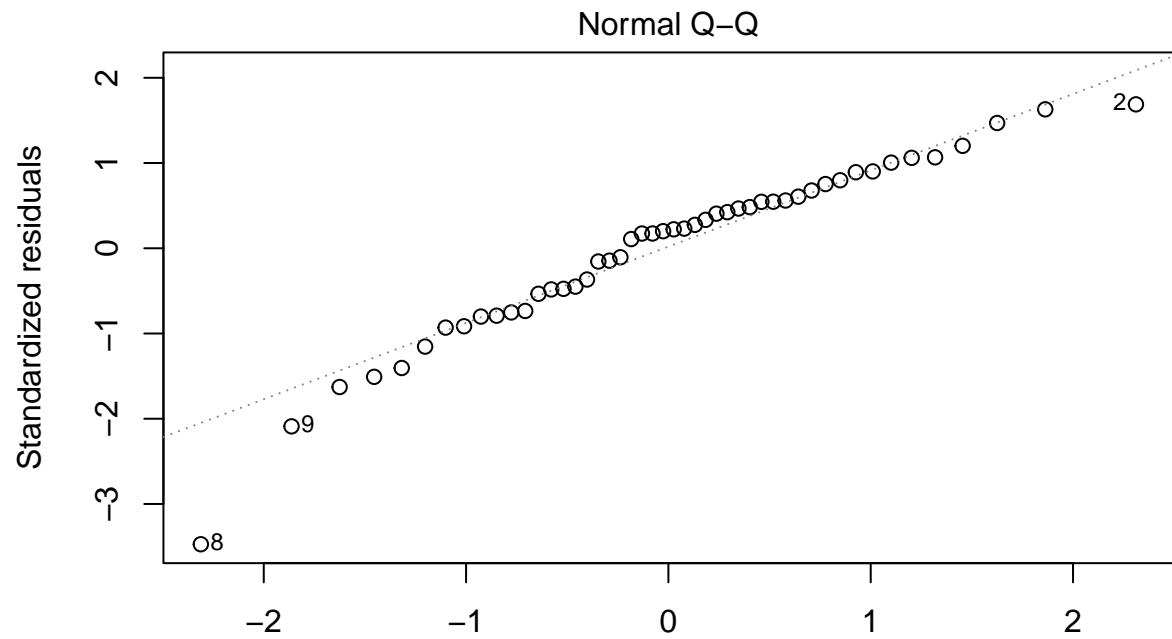
```
summary(aov_block_eff)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Brand         1  342.0   342.0  196.72 < 2e-16 ***
## Temp          2 1654.7   827.4  475.89 < 2e-16 ***
## Stirred       1   69.9    69.9   40.20 1.42e-07 ***
## Brand:Temp    2  231.9   115.9   66.68 1.30e-13 ***
## Residuals    41   71.3     1.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

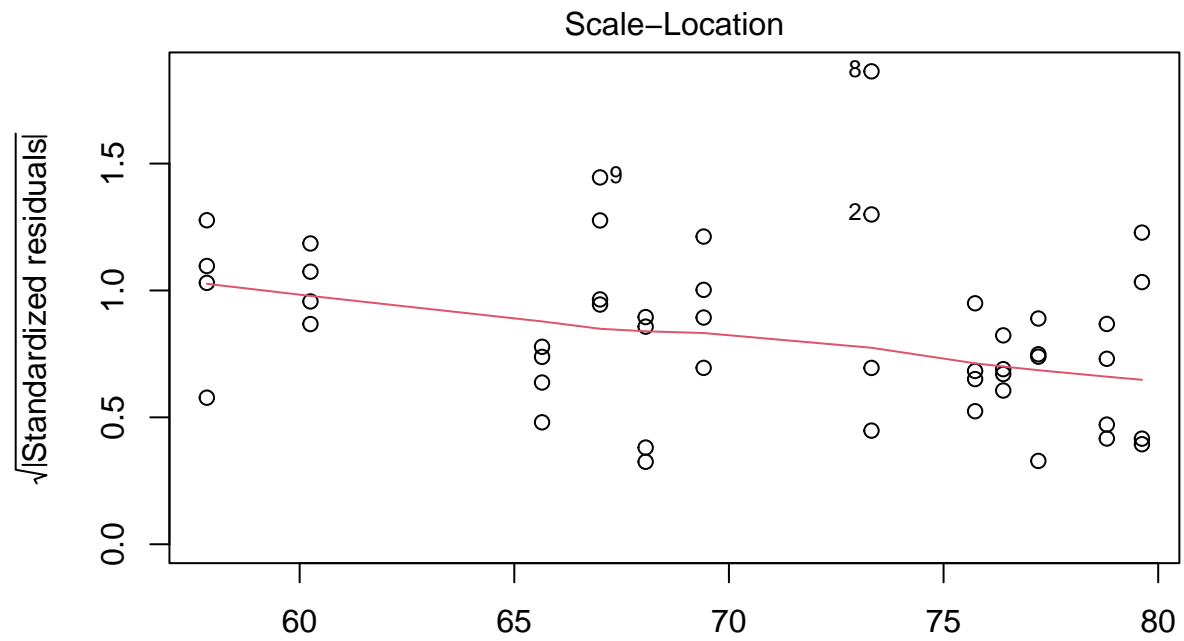
```
plot(aov_block_eff)
```



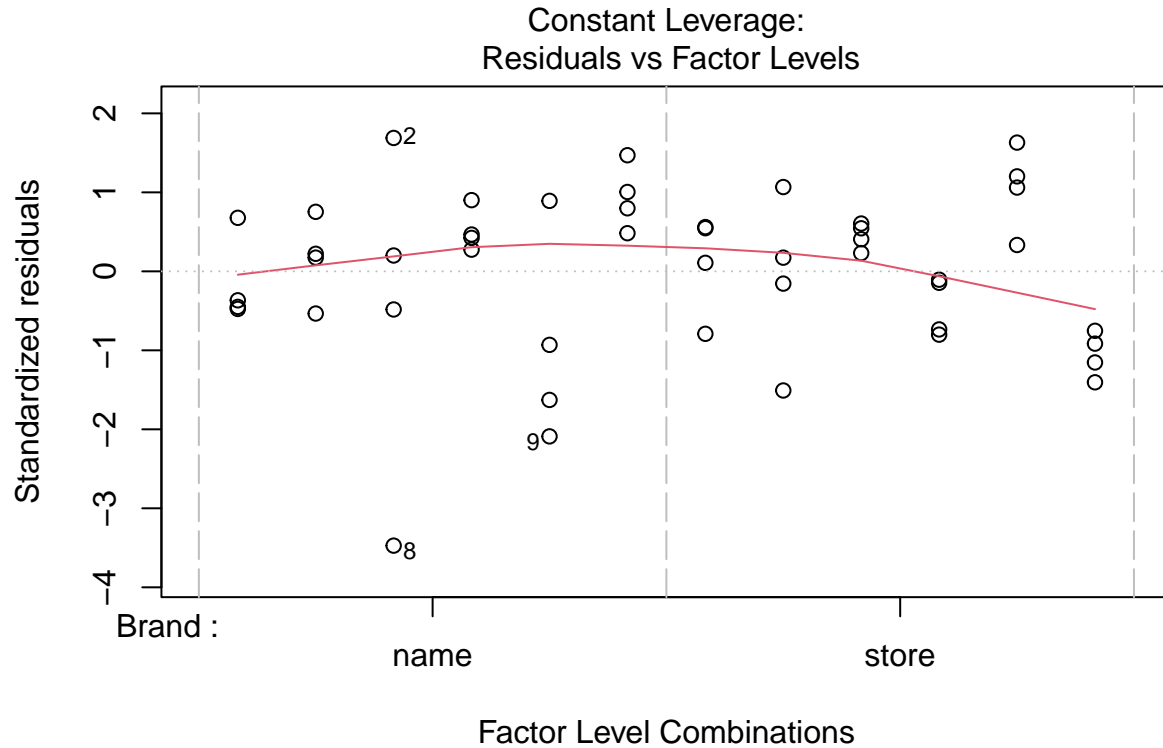
Fitted values
`aov(lm_block_eff <- lm(Time ~ Brand * Temp + Stirred, data = df_eff))`



aov(lm_block_eff <- lm(Time ~ Brand * Temp + Stirred, data = df_eff))



Fitted values
`aov(lm_block_eff <- lm(Time ~ Brand * Temp + Stirred, data = df_eff))`



```
library(olsrr)
```

```
## Warning: package 'olsrr' was built under R version 4.1.3
```

```
##
```

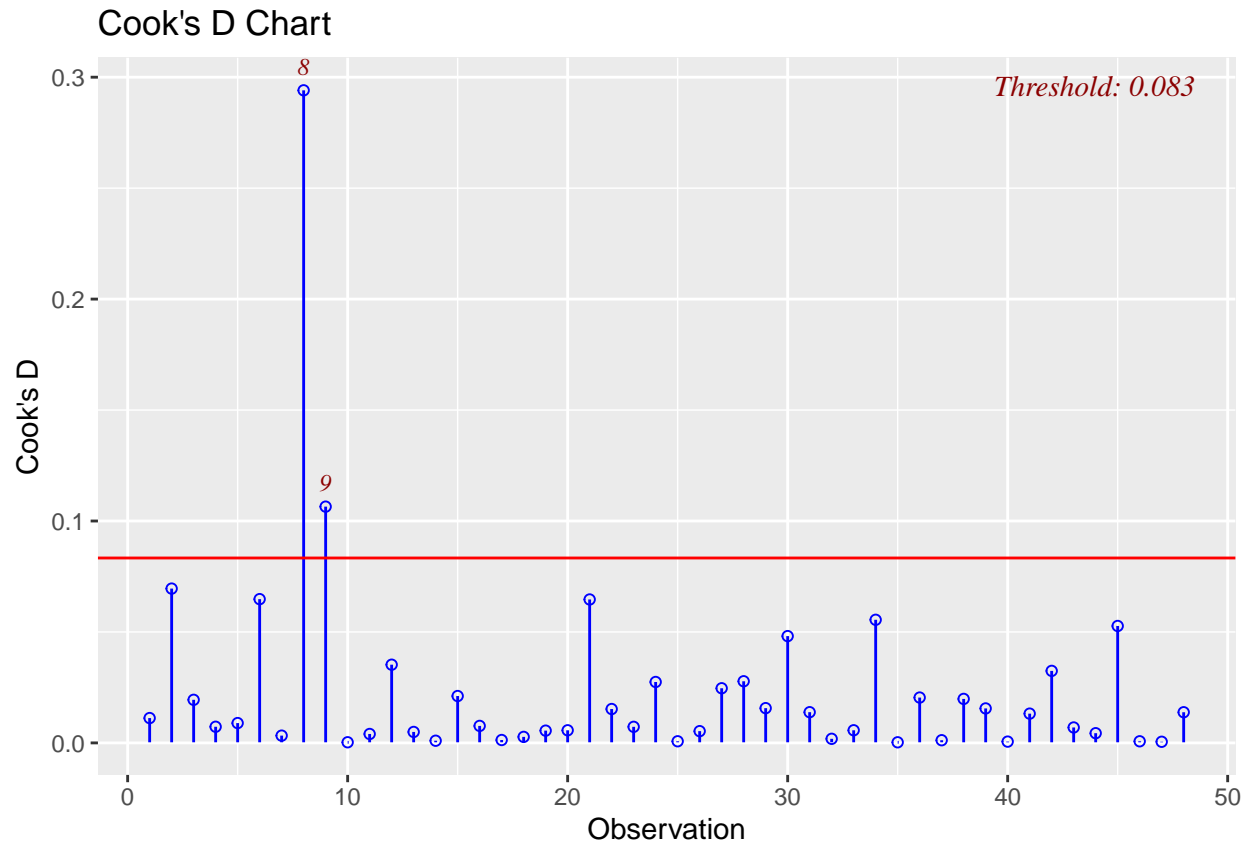
```
## Attaching package: 'olsrr'
```

```
## The following object is masked from 'package:datasets':
```

```
##
```

```
## rivers
```

```
ols_plot_cooks_d_chart(lm_block_eff)
```



```
#added covariate Order model with stirred as block effect without interaction
aov_block_order_eff <- aov(lm_block_order_eff <- lm(Time ~ Brand * Temp + Stirred + Order, data = df_eff))
summary(lm_block_order_eff)
```

```
##
## Call:
## lm(formula = Time ~ Brand * Temp + Stirred + Order, data = df_eff)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.9257 -0.7714  0.0969  0.8613  2.4030
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    75.76840    0.66484  113.965 < 2e-16 ***
## Brandstore      0.95630    0.65881   1.452  0.1544
## Temp23         -2.91120    0.66104  -4.404 7.73e-05 ***
## Temp40         -9.17522    0.66943 -13.706 < 2e-16 ***
## Stirredno       1.46630    0.77003   1.904  0.0641 .
## Order           0.03946    0.02800   1.409  0.1664
## Brandstore:Temp23 -8.64026    0.92729  -9.318 1.42e-11 ***
## Brandstore:Temp40 -10.22471    0.93780 -10.903 1.51e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.303 on 40 degrees of freedom
```

```
## Multiple R-squared:  0.9713, Adjusted R-squared:  0.9663
## F-statistic: 193.7 on 7 and 40 DF,  p-value: < 2.2e-16
```

```
summary(aov_block_order_eff)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Brand      1  342.0   342.0 201.452 < 2e-16 ***
## Temp       2 1654.7   827.4 487.344 < 2e-16 ***
## Stirred    1   69.9    69.9 41.166 1.23e-07 ***
## Order      1    0.9     0.9  0.534   0.469
## Brand:Temp  2  234.3   117.2 69.010 1.08e-13 ***
## Residuals 40   67.9     1.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.1.3
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.1.3
```

```
##
## Attaching package: 'car'
```

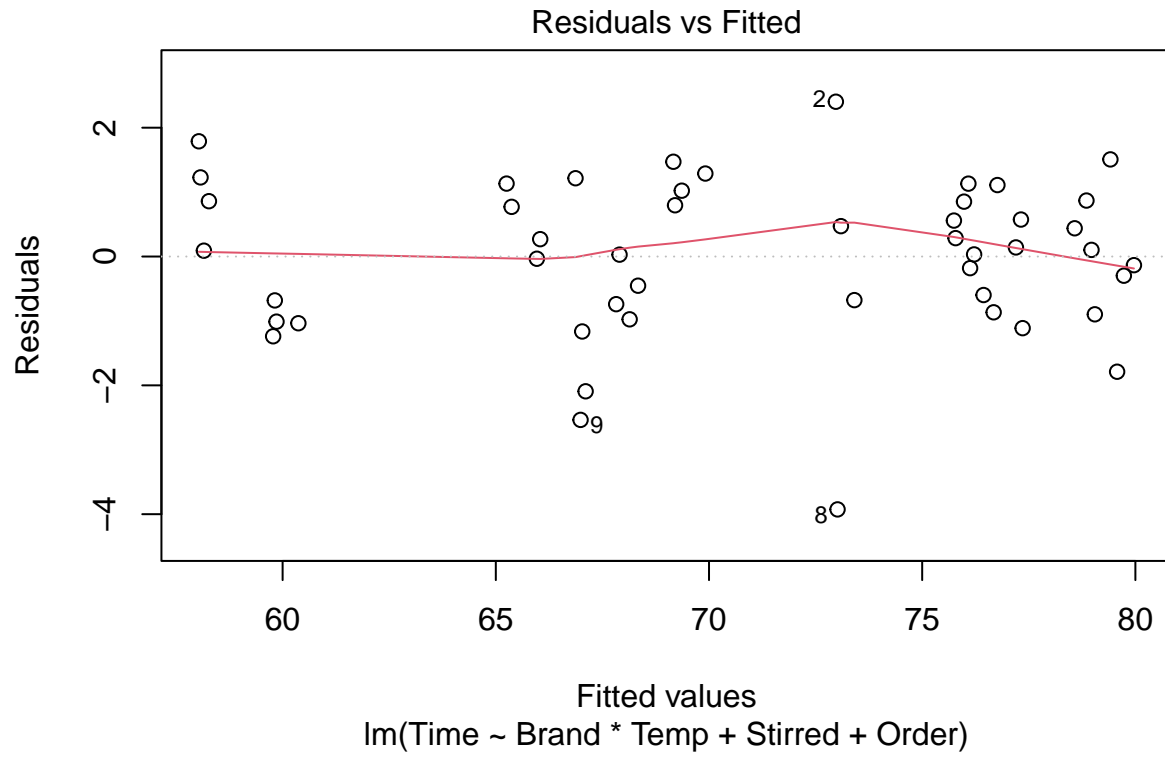
```
## The following object is masked from 'package:dplyr':
##
##      recode
```

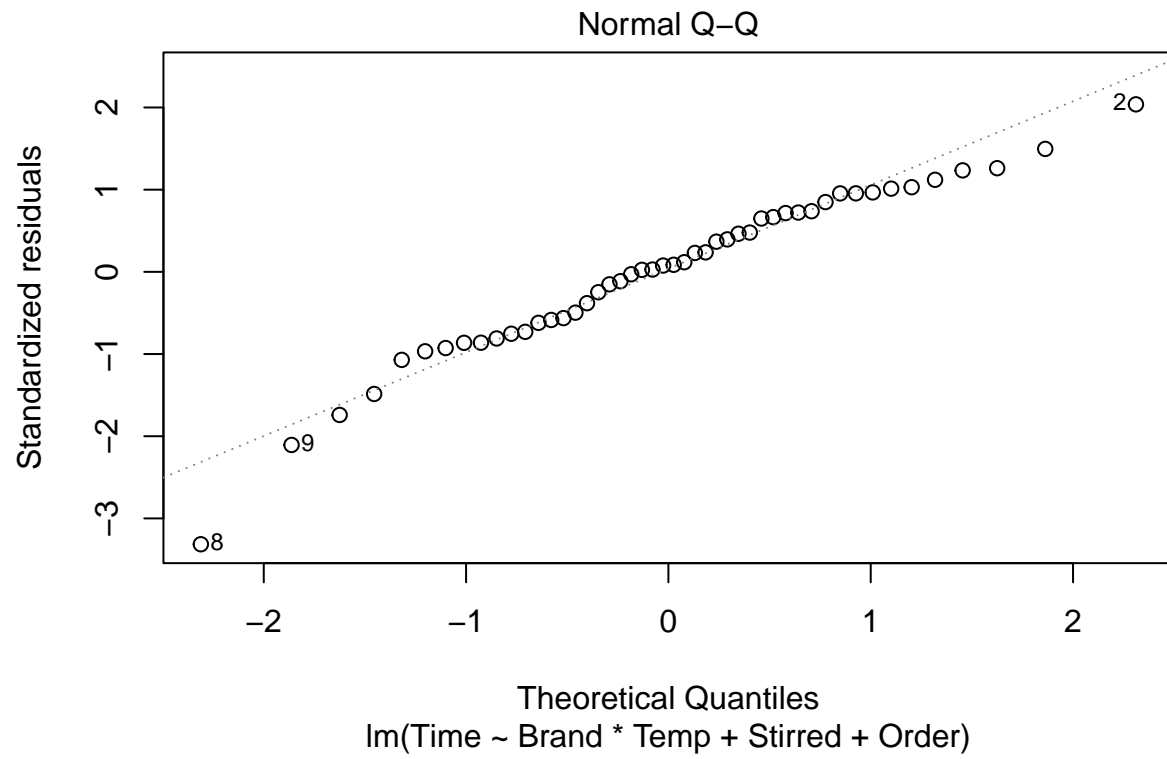
```
## The following object is masked from 'package:purrr':
##
##      some
```

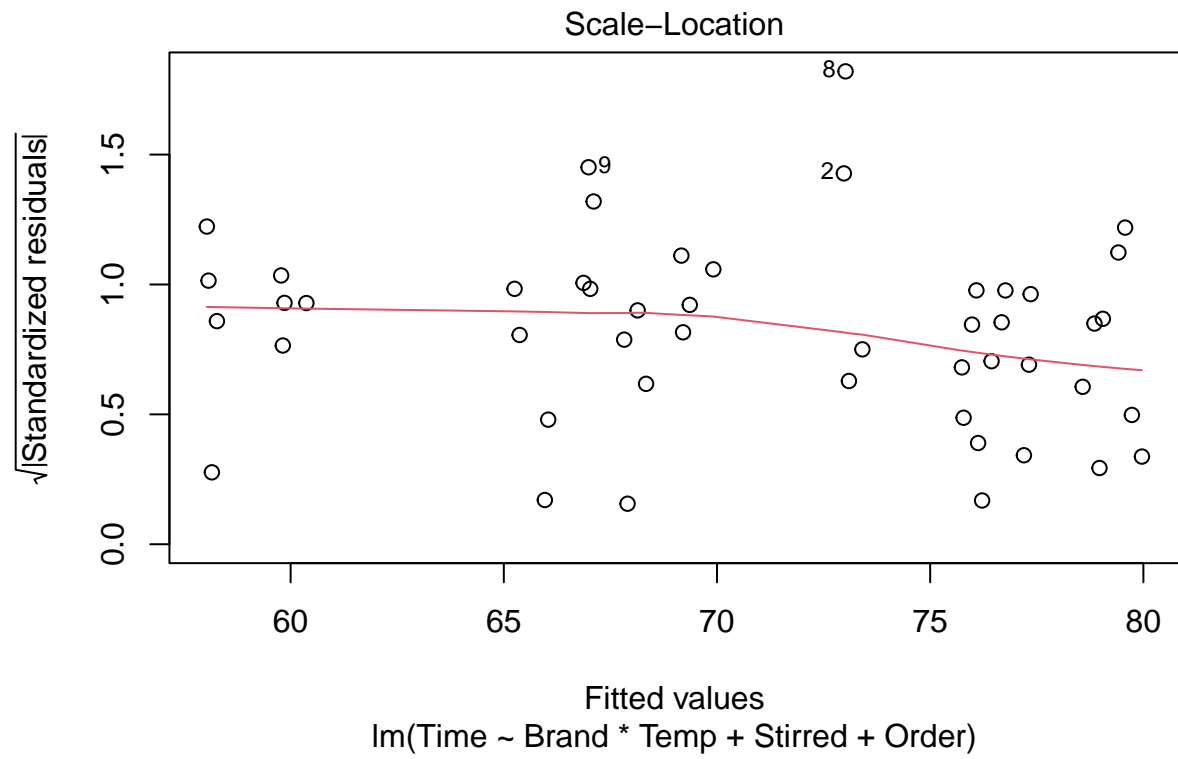
```
Anova(aov_block_order_eff, type=3) # type 3 SS
```

```
## Anova Table (Type III tests)
##
## Response: Time
##           Sum Sq Df    F value    Pr(>F)
## (Intercept) 22049.8 1 12987.9591 < 2.2e-16 ***
## Brand        3.6  1    2.1070   0.15442
## Temp       335.9  2   98.9237 3.275e-16 ***
## Stirred      6.2  1    3.6261   0.06409 .
## Order        3.4  1    1.9864   0.16645
## Brand:Temp   234.3  2   69.0102 1.076e-13 ***
## Residuals    67.9 40
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(lm_block_order_eff)
```

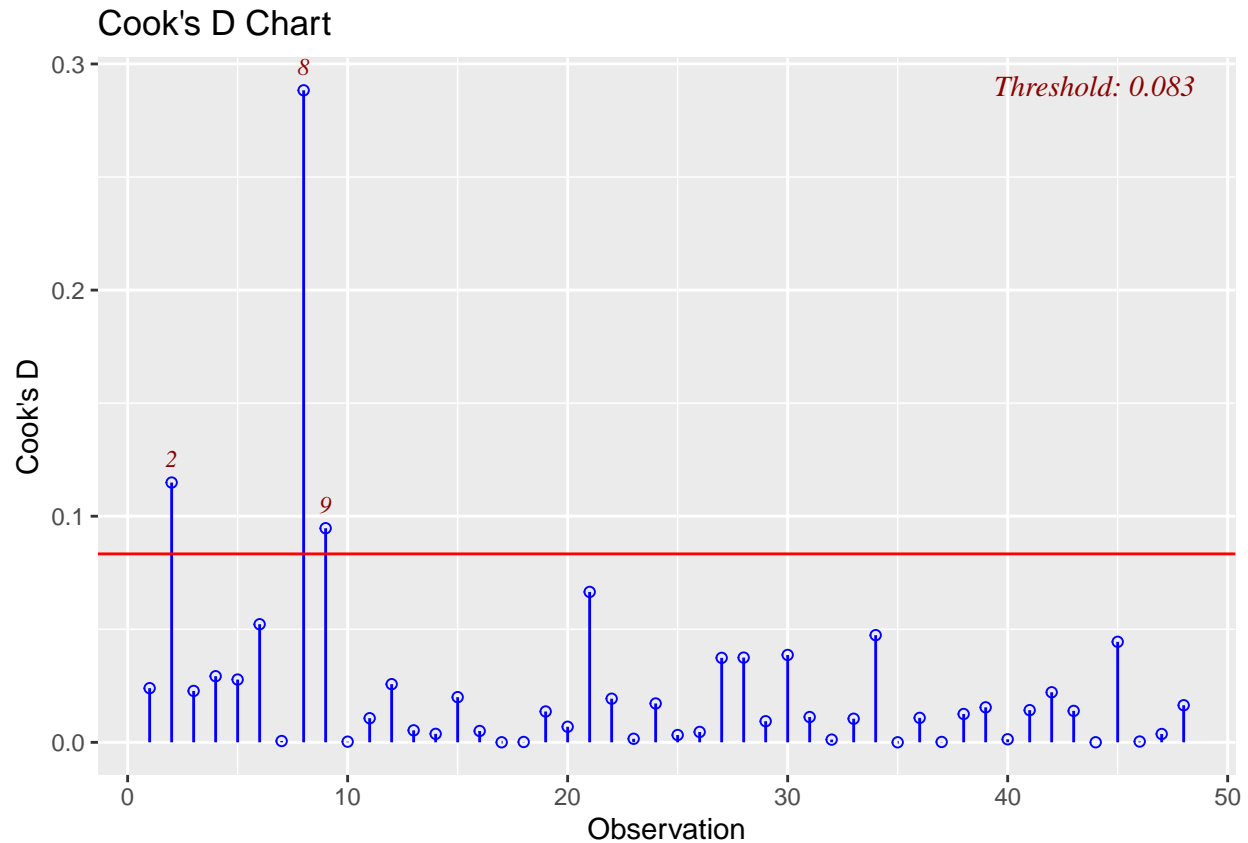








```
ols_plot_cooksd_chart(lm_block_order_eff)
```



#added covariate Order to model with 3 factor interaction

```
aov_three_order_eff <- aov(lm_three_order_eff <- lm(Time ~ Brand * Temp*Stirred + Order, data = df_eff))
summary(lm_three_order_eff)
```

```
##
## Call:
## lm(formula = Time ~ Brand * Temp * Stirred + Order, data = df_eff)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6373 -0.3259  0.0077  0.4420  2.6391
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      76.3849    0.6759  113.009 < 2e-16 ***
## Brandstore         1.0930    0.7740   1.412  0.16673
## Temp23            -3.6070    0.7971  -4.525 6.67e-05 ***
## Temp40           -10.4002    0.7768 -13.389 2.47e-15 ***
## Stirredno          3.1339    1.0819   2.897  0.00646 **
## Order             -0.0128    0.0282  -0.454  0.65270
## Brandstore:Temp23  -7.5228    1.1145  -6.750 8.06e-08 ***
## Brandstore:Temp40  -7.6899    1.1517  -6.677 1.00e-07 ***
## Brandstore:Stirredno -0.6392    1.0870  -0.588  0.56026
## Temp23:Stirredno    0.9735    1.1047   0.881  0.38420
## Temp40:Stirredno    1.8751    1.0902   1.720  0.09427 .
## Brandstore:Temp23:Stirredno -1.8430    1.5628  -1.179  0.24622
```

```
## Brandstore:Temp40:Stirredno -4.4163      1.5906 -2.777  0.00876 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.087 on 35 degrees of freedom
## Multiple R-squared:  0.9826, Adjusted R-squared:  0.9766
## F-statistic: 164.3 on 12 and 35 DF,  p-value: < 2.2e-16
```

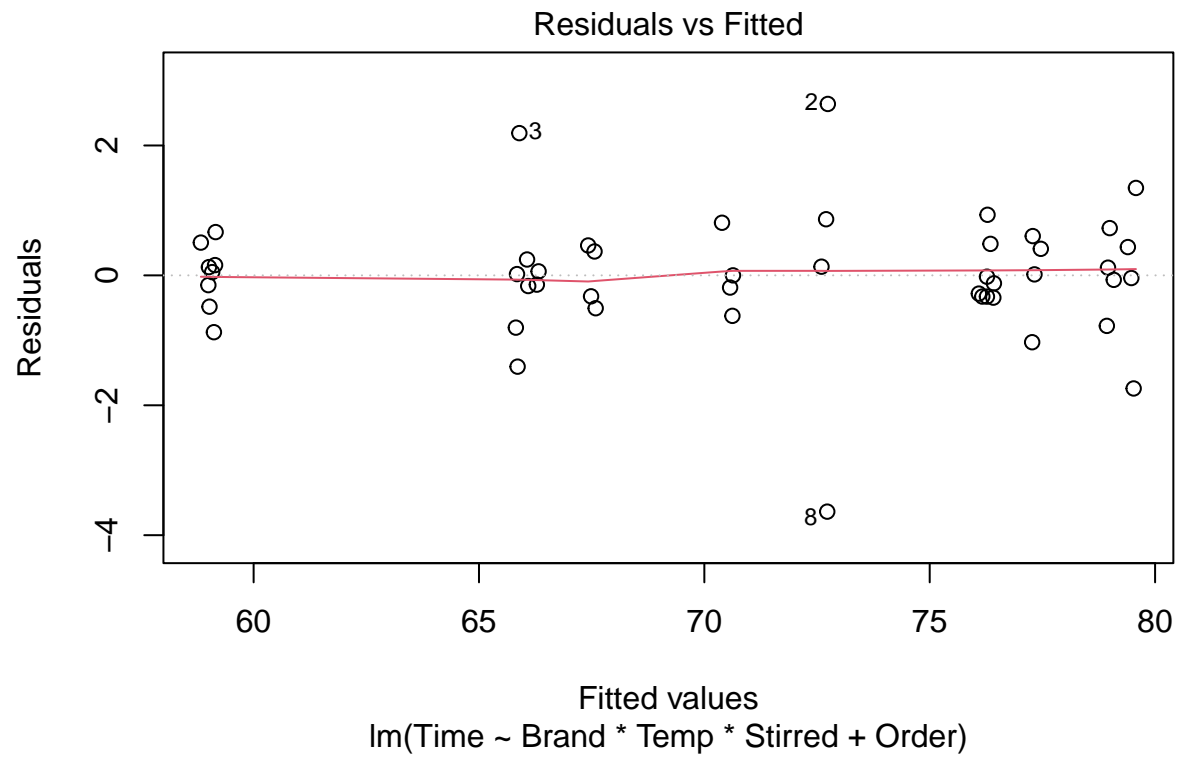
```
summary(aov_three_order_eff)
```

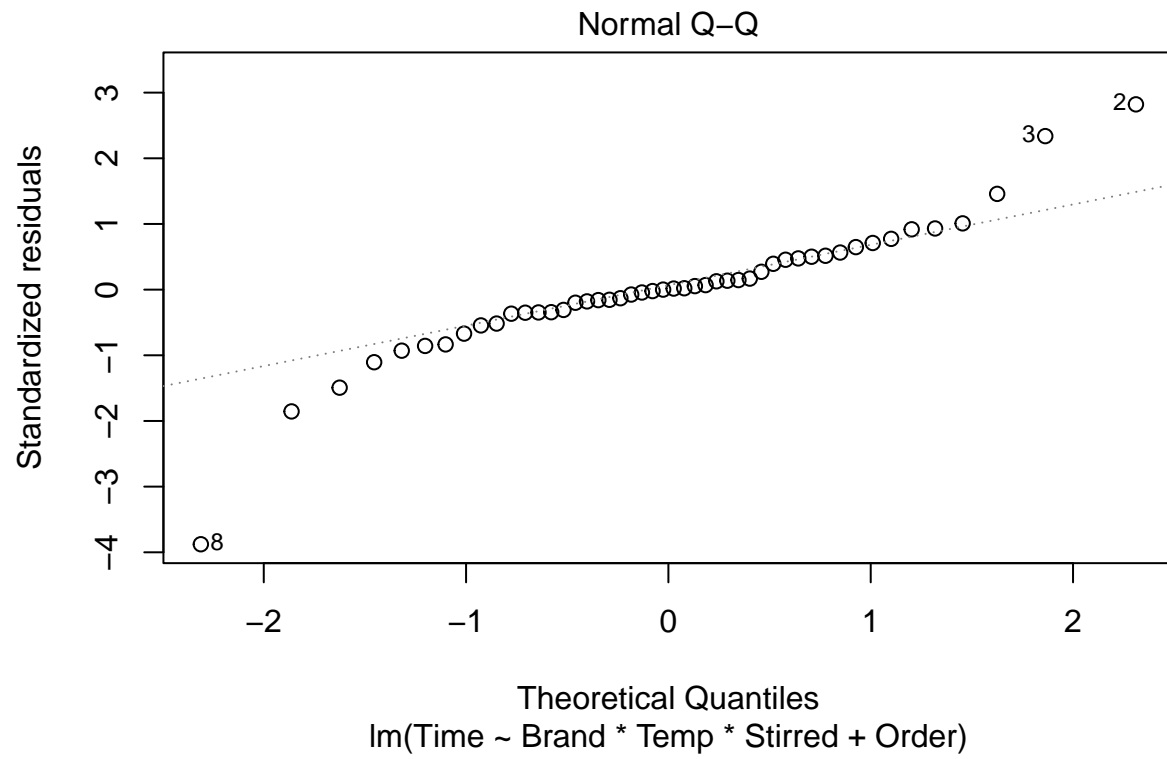
```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Brand          1  342.0   342.0 289.512 < 2e-16 ***
## Temp           2 1654.7   827.4 700.374 < 2e-16 ***
## Stirred        1   69.9    69.9  59.161 5.01e-09 ***
## Order          1    0.9     0.9   0.767 0.387161
## Brand:Temp      2  234.3   117.2  99.176 3.81e-15 ***
## Brand:Stirred   1   17.3    17.3  14.641 0.000514 ***
## Temp:Stirred    2    0.0     0.0   0.018 0.982371
## Brand:Temp:Stirred 2    9.2     4.6   3.904 0.029469 *
## Residuals      35   41.3     1.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

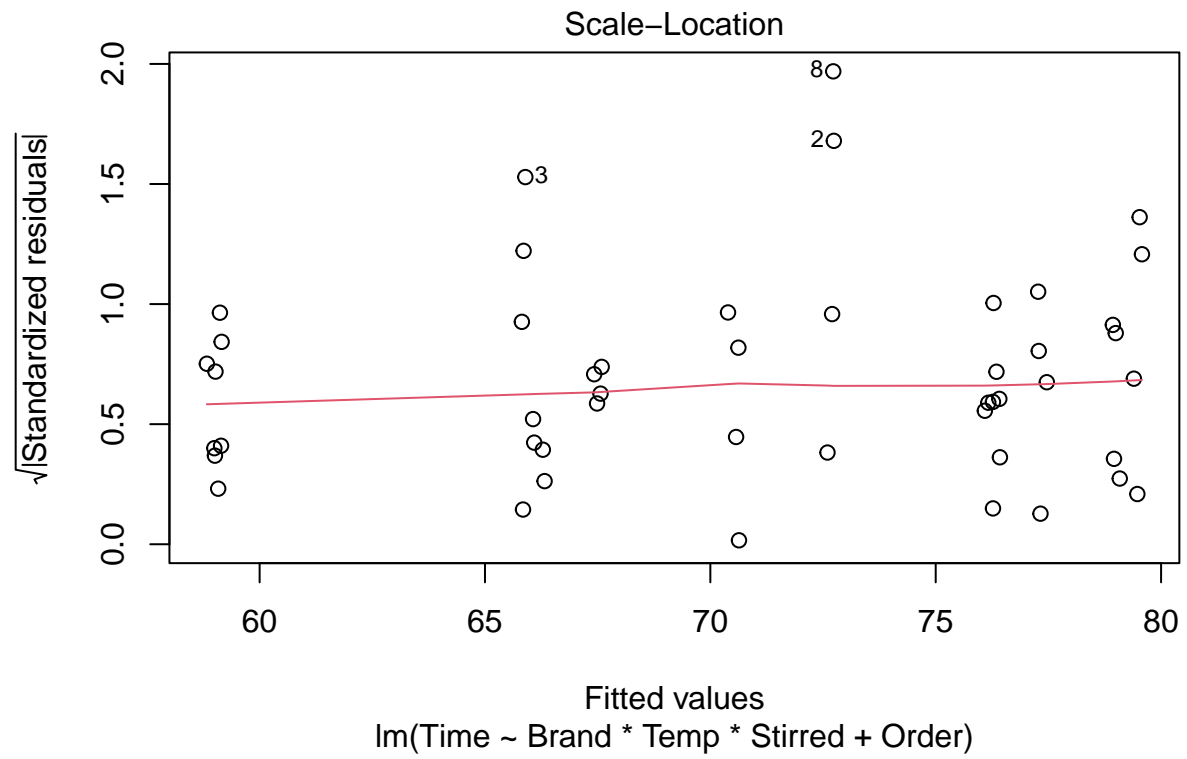
```
library(car)
Anova(aov_three_order_eff, type=3) # type 3 SS
```

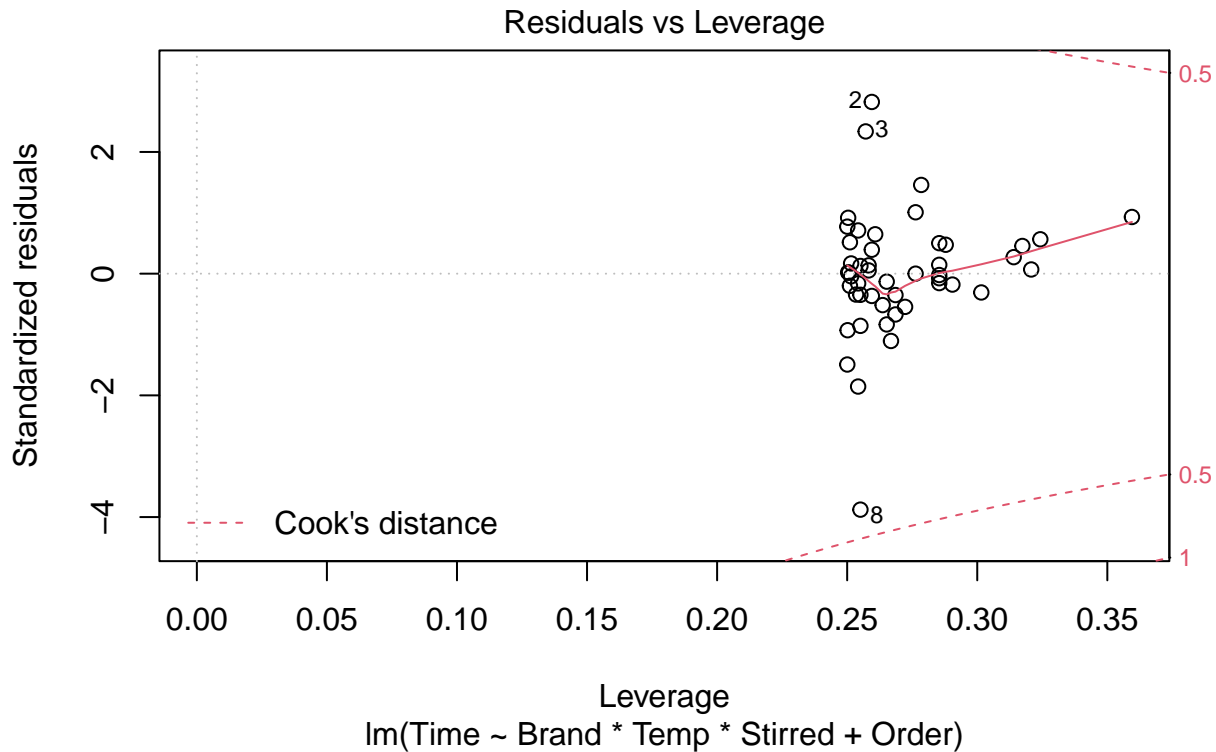
```
## Anova Table (Type III tests)
##
## Response: Time
##              Sum Sq Df    F value    Pr(>F)
## (Intercept) 15086.6  1 12770.9606 < 2.2e-16 ***
## Brand         2.4   1    1.9942  0.166731
## Temp        220.9   2   93.4777 9.153e-15 ***
## Stirred       9.9   1    8.3899 0.006465 **
## Order         0.2   1    0.2060 0.652697
## Brand:Temp    69.5   2   29.3957 3.224e-08 ***
## Brand:Stirred  0.4   1    0.3458 0.560255
## Temp:Stirred   3.5   2    1.4816 0.241180
## Brand:Temp:Stirred 9.2   2    3.9044 0.029469 *
## Residuals    41.3  35
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(lm_three_order_eff)
```









```
#adding Halid's code to see how full 3 factor interaction compares
aov_eff <- aov(lm_eff <- lm(Time ~ Brand * Temp * Stirred, data = df_eff))
summary(lm_eff)
```

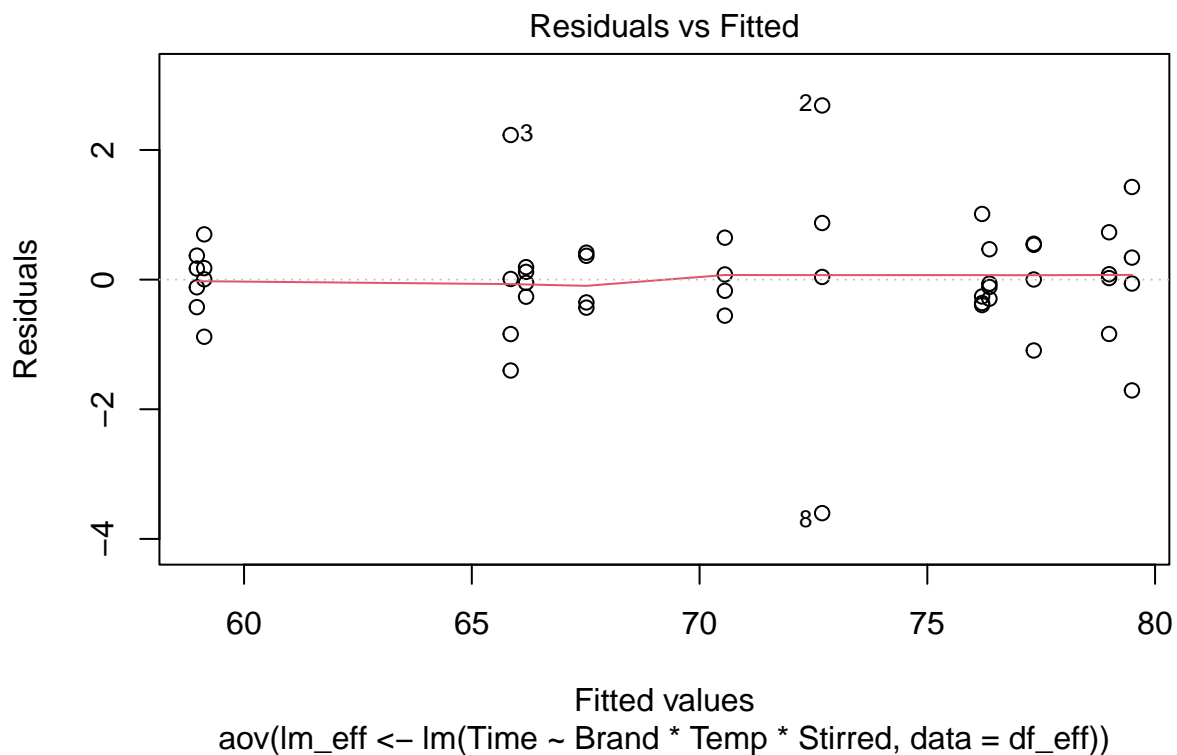
```
##
## Call:
## lm(formula = Time ~ Brand * Temp * Stirred, data = df_eff)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6021 -0.3538  0.0077  0.3816  2.6871
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      76.2024     0.5374 141.794 < 2e-16 ***
## Brandstore         1.1346     0.7600   1.493 0.144185
## Temp23            -3.5110     0.7600  -4.620 4.78e-05 ***
## Temp40           -10.3490     0.7600 -13.617 9.05e-16 ***
## Stirredno         2.7882     0.7600   3.669 0.000783 ***
## Brandstore:Temp23  -7.6348     1.0748  -7.103 2.40e-08 ***
## Brandstore:Temp40  -7.8628     1.0748  -7.315 1.27e-08 ***
## Brandstore:Stirredno -0.6328     1.0748  -0.589 0.559694
## Temp23:Stirredno    0.8839     1.0748   0.822 0.416308
## Temp40:Stirredno    1.9135     1.0748   1.780 0.083476 .
## Brandstore:Temp23:Stirredno -1.7150     1.5200  -1.128 0.266685
```

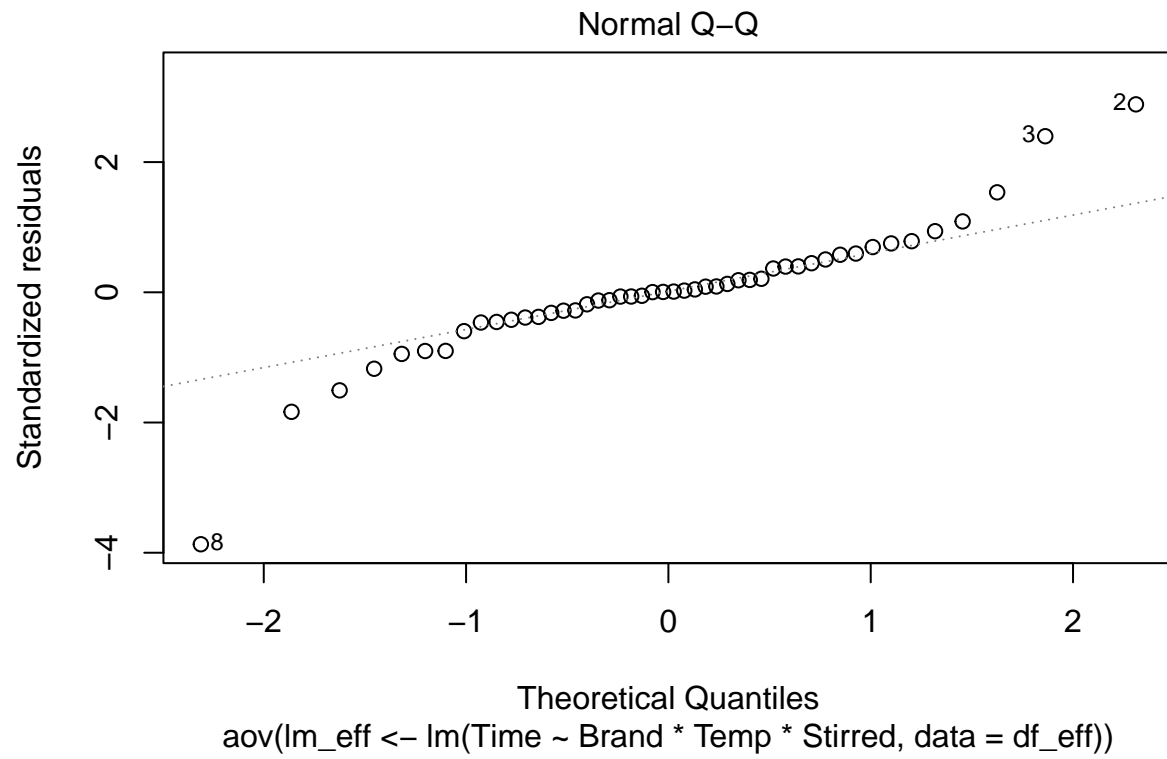
```
## Brandstore:Temp40:Stirredno -4.2307      1.5200 -2.783 0.008519 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.075 on 36 degrees of freedom
## Multiple R-squared:  0.9824, Adjusted R-squared:  0.9771
## F-statistic: 183.2 on 11 and 36 DF,  p-value: < 2.2e-16
```

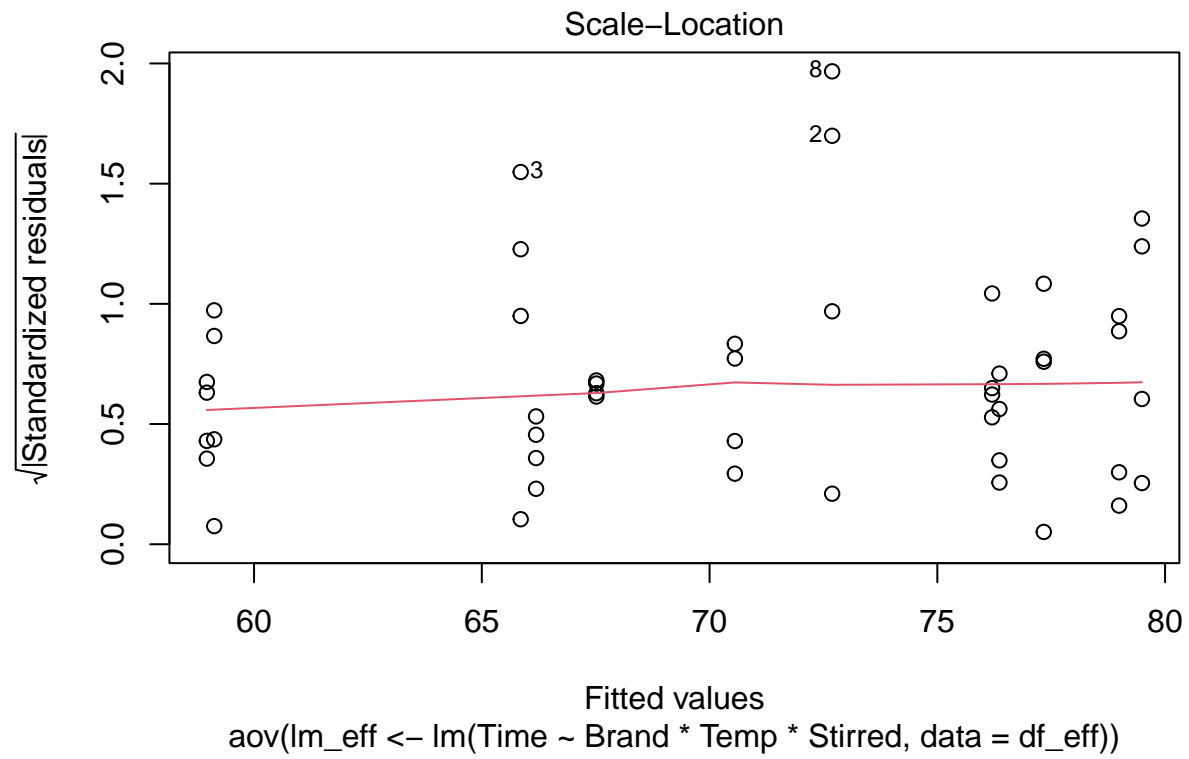
```
summary(aov_eff)
```

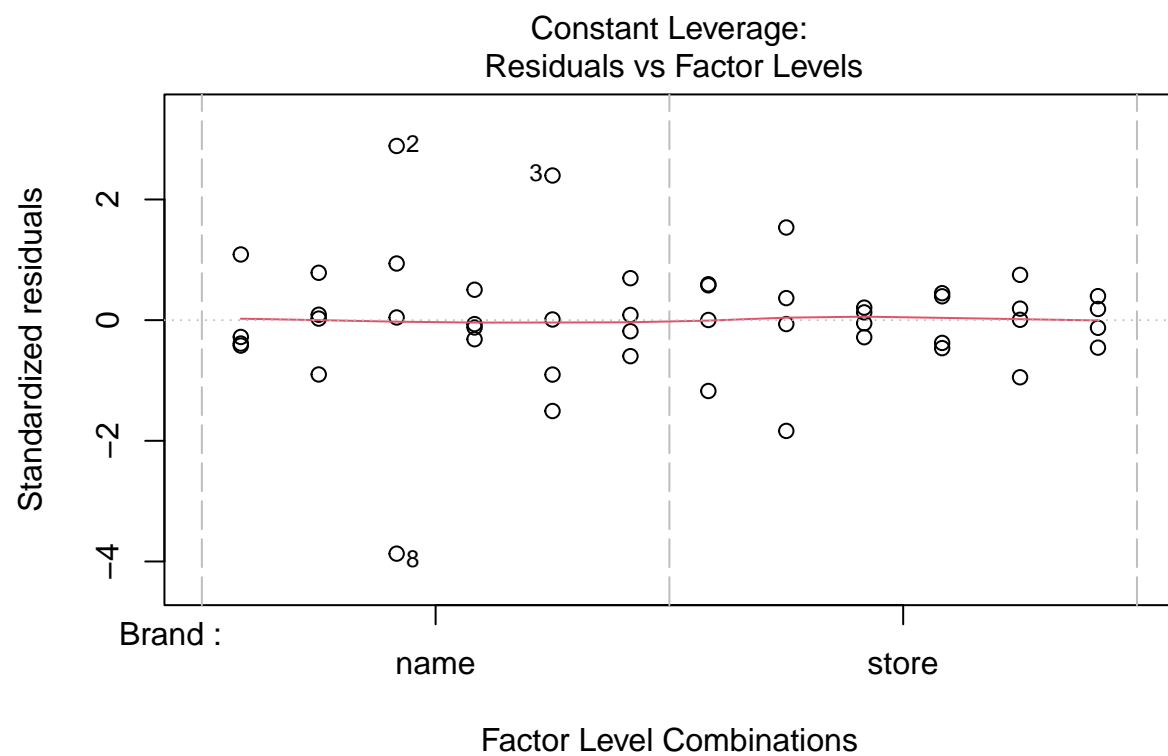
```
##
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Brand          1  342.0   342.0 296.041 < 2e-16 ***
## Temp           2 1654.7   827.4 716.169 < 2e-16 ***
## Stirred        1   69.9    69.9  60.495 3.22e-09 ***
## Brand:Temp      2  231.9   115.9 100.345 1.90e-15 ***
## Brand:Stirred   1   20.5    20.5  17.753 0.000161 ***
## Temp:Stirred    2    0.1     0.1   0.054 0.947535
## Brand:Temp:Stirred 2    9.1     4.5   3.919 0.028838 *
## Residuals      36   41.6     1.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(aov_eff)
```









```
#adding Cook's D here  
ols_plot_cooksd_chart(lm_eff)
```

