STAT 425 Final Project

Michael Garbus mgarbus2

12/13/2021

Contents

Company XX is a manufacturer of several types of protective packaging, including bubble wrap sold in both retail and bulk. The objective of this project is to determine the best operating conditions for the bubble wrap lines to increase production capacity.

The response variable was the production rate measured in lbs/hr. The experiment was replicated 3 times and the randomization order for each replication was also recorded.

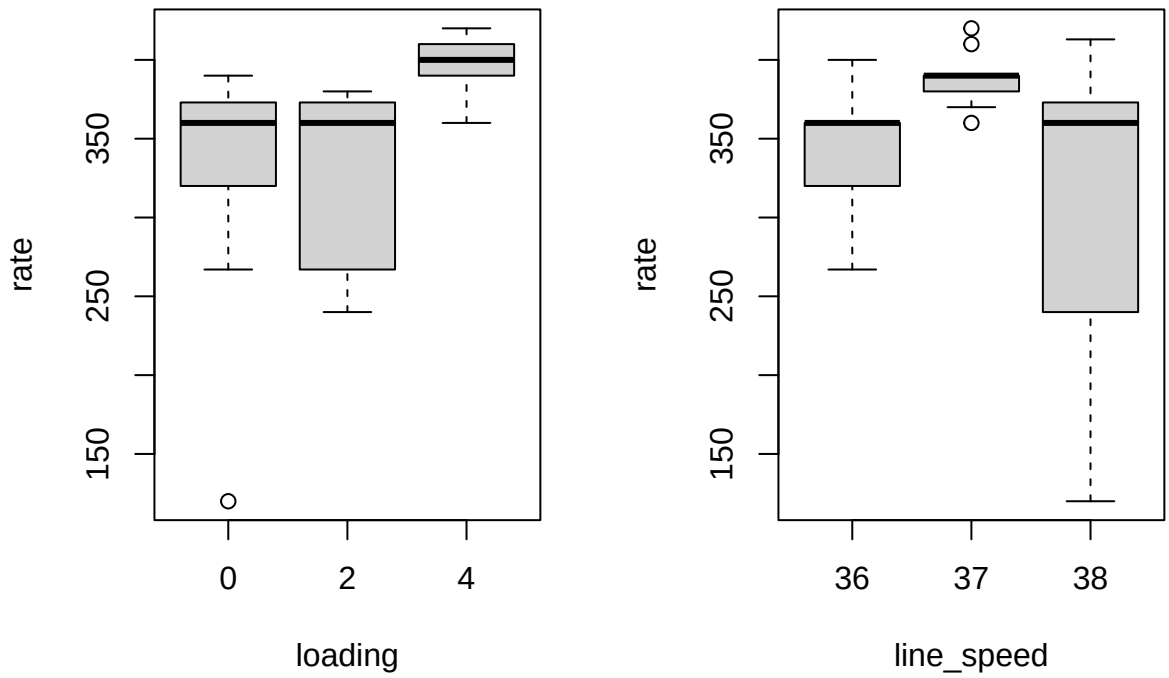
```
bubblewrap <- read.csv('bubblewrap.csv')
```

```
# set factors
bubblewrap$line_speed <- as.factor(bubblewrap$line_speed)
bubblewrap$loading <- as.factor(bubblewrap$loading)
head(bubblewrap)
```

```
##   replication run_order line_speed loading rate
## 1           1         6         38       2  240
## 2           1         8         37       4  390
## 3           1         1         36       0  360
## 4           1         9         38       4  400
## 5           1         3         38       0  320
## 6           1         7         36       4  400
```

```
#Means
aov_model <- aov(rate ~ -1 + loading + line_speed, data = bubblewrap)
group_means <- aov(rate ~ -1 + loading + line_speed, data = bubblewrap)
```

```
par(mfrow = c(1,2))
boxplot(rate ~ loading, data = bubblewrap)
boxplot(rate ~ line_speed, data = bubblewrap)
```



There appears to be a difference in mean loading and line speed with respect to rate. Loading 4 and Line Speed 37 seem to have different means than their counterparts.

```
library(car)
```

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

```
Anova(lm(rate ~ -1 + loading + line_speed, data = bubblewrap), type = "III")
```

```
## Anova Table (Type III tests)
##
## Response: rate
##             Sum Sq Df  F value   Pr(>F)
## loading    1089609  3 115.7807 1.26e-13 ***
## line_speed   23945  2   3.8165  0.03777 *
## Residuals    69014 22
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It seems like both loading and line speed are significant.

We check our model assumptions.

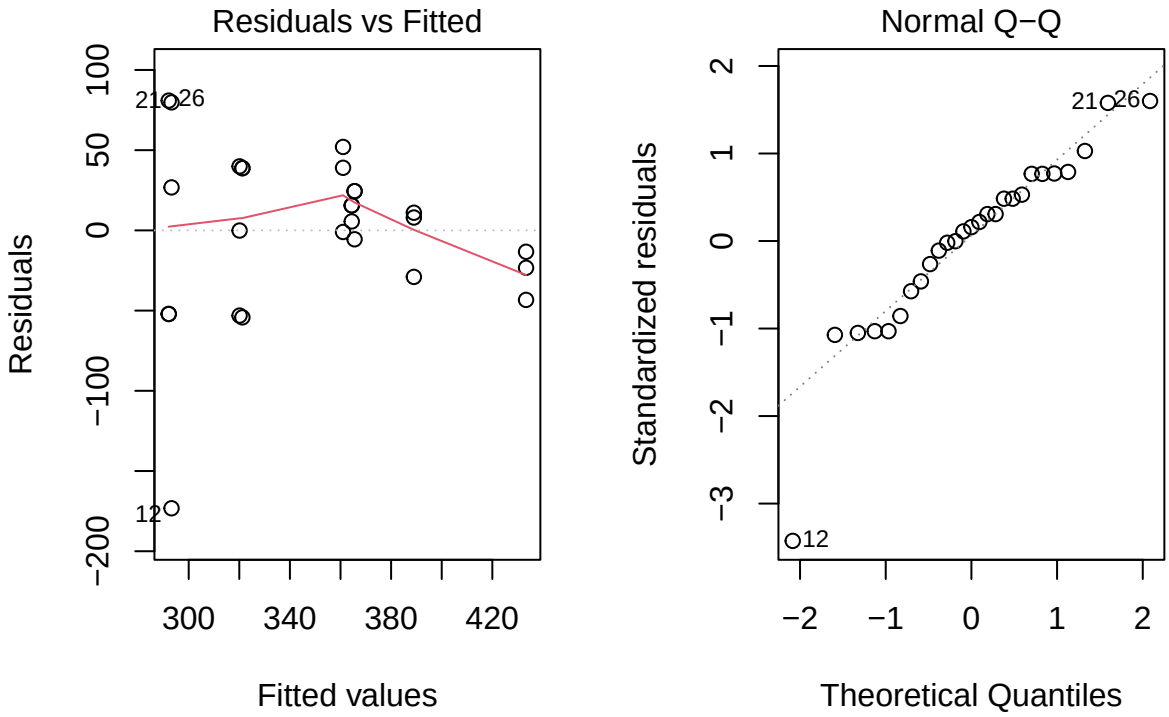
```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
par(mfrow = c(1,2))
add_mod <- lm(rate ~ -1 + loading + line_speed, data = bubblewrap)
plot(add_mod, which = c(1,2))
```




```
shapiro.test(resid(add_mod))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(add_mod)
## W = 0.89977, p-value = 0.01321
```

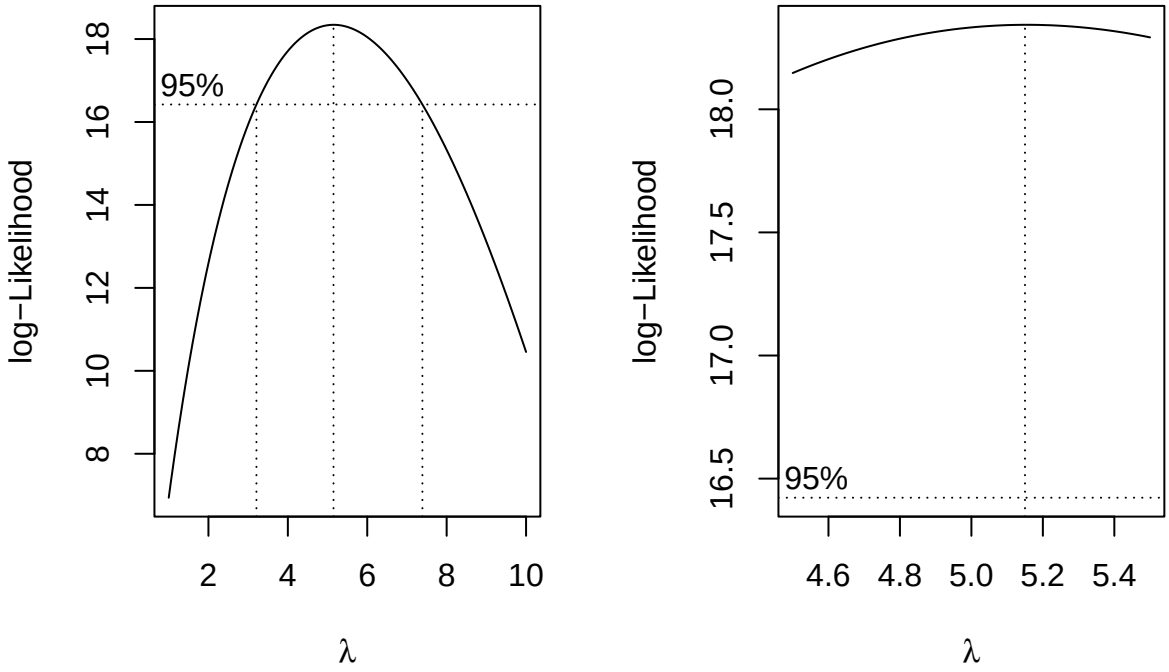
```
bptest(add_mod)
```

```
##
##  studentized Breusch-Pagan test
##
```

```
## data:  add_mod
## BP = 7.3817, df = 4, p-value = 0.117
```

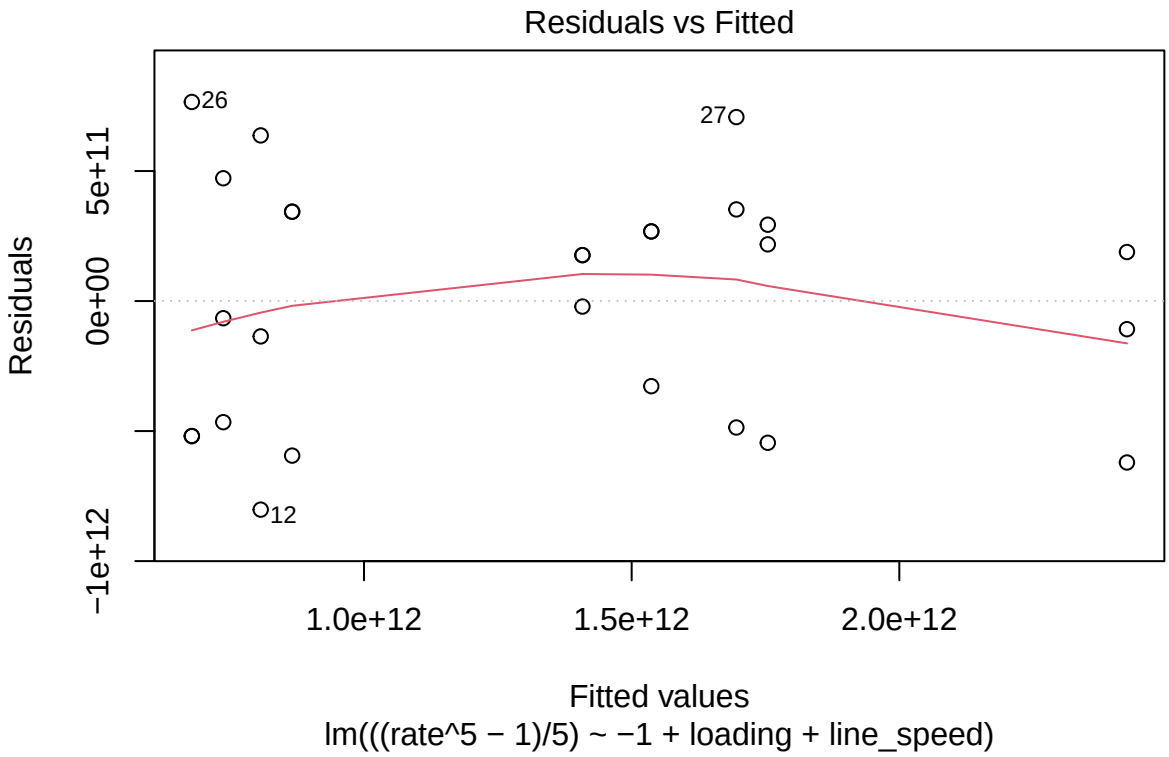
The model appears to be heteroskedastic according to the Breusch-Pagan test, but its residuals do not follow the normal distribution according to the Shapiro-Wilk test. The violation of the residuals assumption can be ignored due to small sample size. We performed a box-cox transformation just to see if we could achieve this assumption regardless.

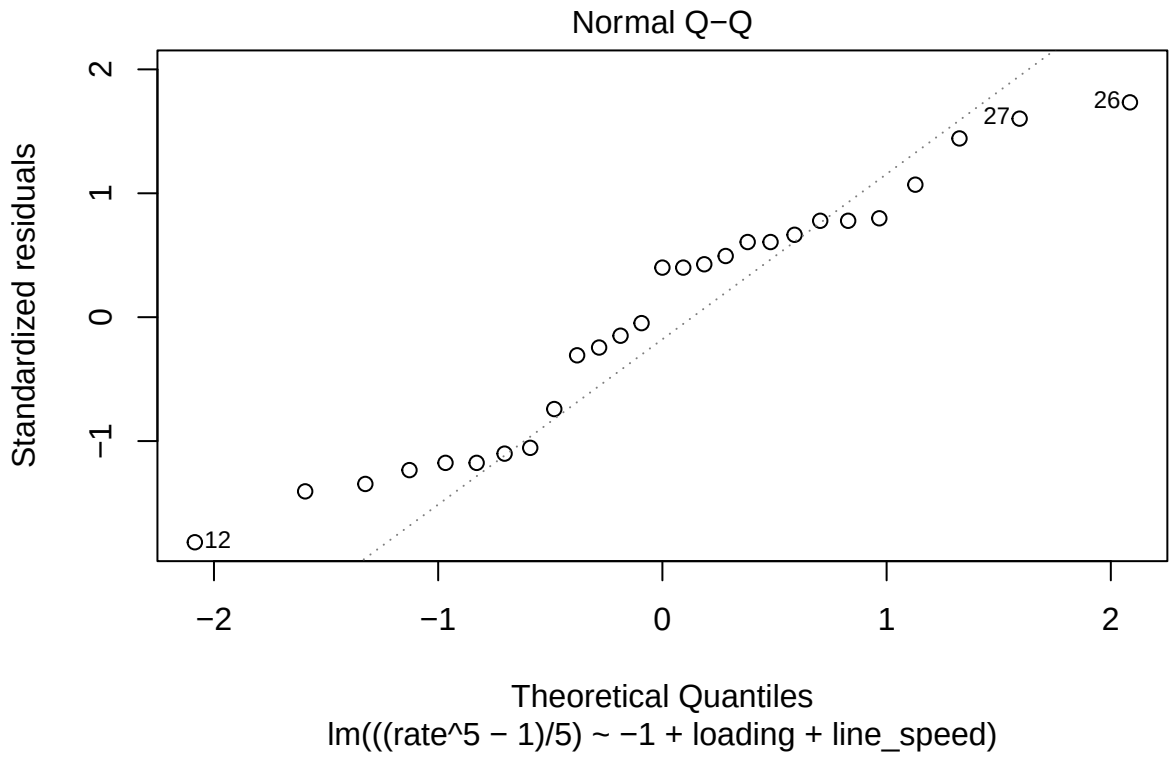
```
library(MASS)
par(mfrow = c(1,2))
boxcox(add_mod, lambda = seq(1,10,0.05))
boxcox(add_mod, lambda = seq(4.5,5.5,0.01))
```




```
#Lambda = 5.18
```

```
#transformed model:
tfmod <- lm(((rate^5 - 1)/5) ~ -1 + loading + line_speed, data = bubblewrap)
plot(tfmod, which = c(1,2))
```


Residuals vs Fitted
Residuals
5e+11
0e+00
−1e+12
26
27
12
1.0e+12
1.5e+12
2.0e+12
Fitted values
lm(((rate^5 − 1)/5) ~ −1 + loading + line_speed)




```
shapiro.test(resid(tfmod))
```

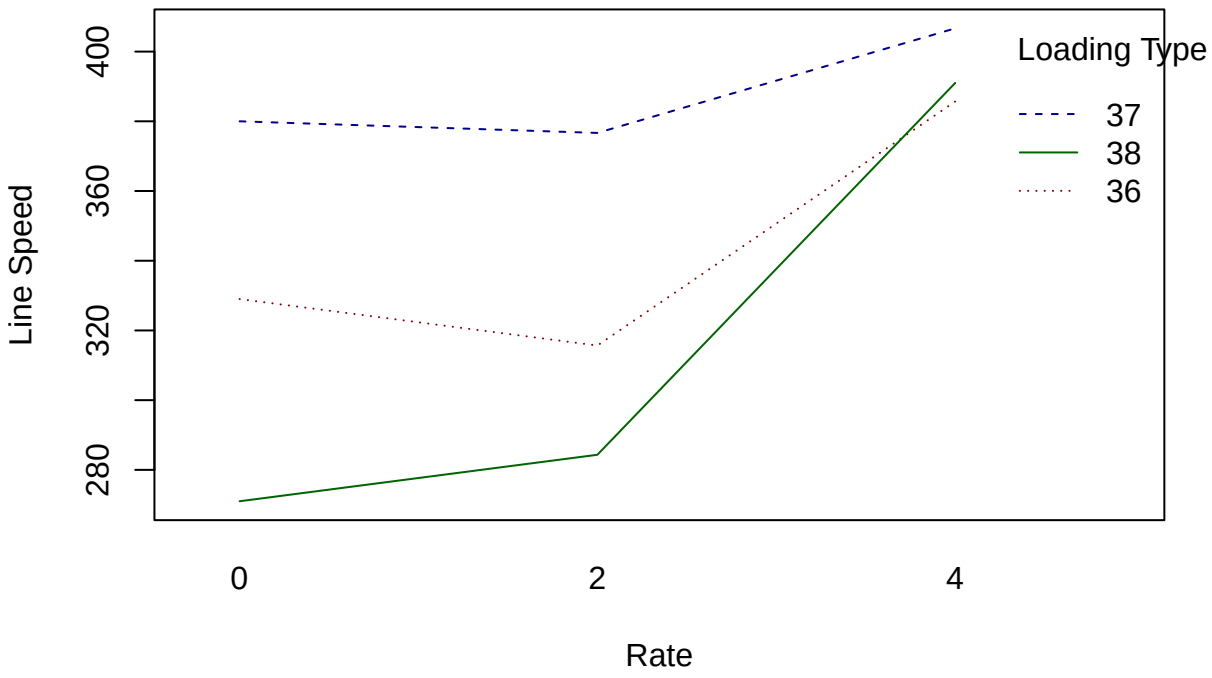
```
##
##  Shapiro-Wilk normality test
##
## data:  resid(tfmod)
## W = 0.94206, p-value = 0.137
```

```
bptest(tfmod)
```

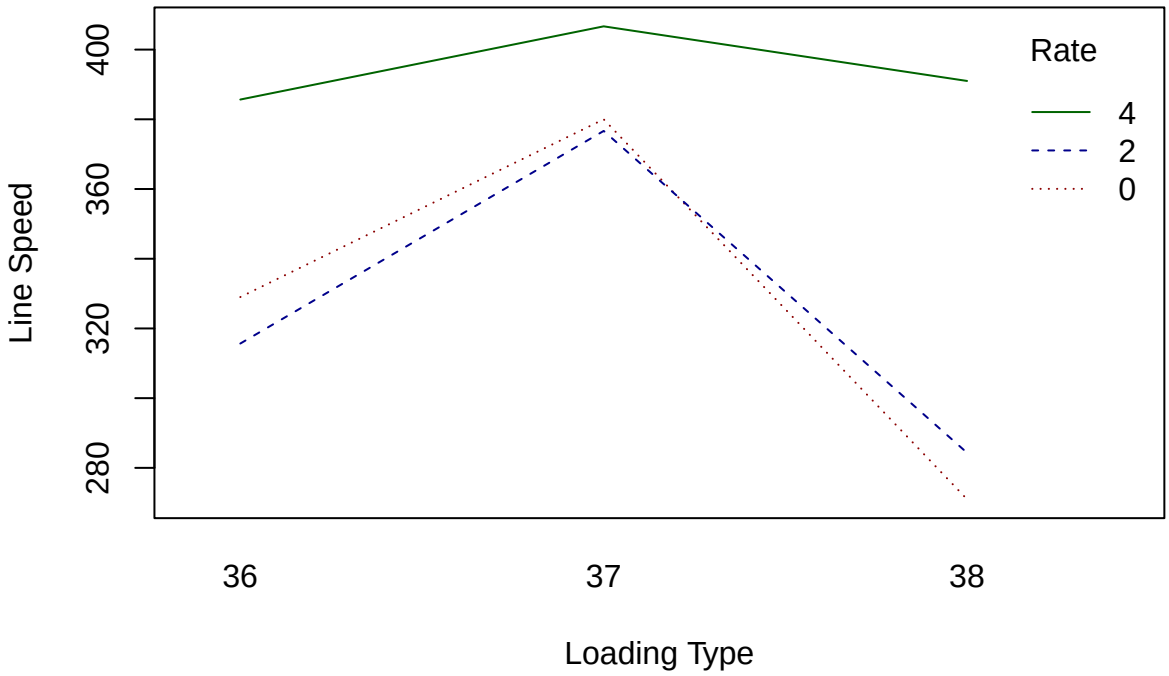
```
##
##  studentized Breusch-Pagan test
##
## data:  tfmod
## BP = 9.5987, df = 4, p-value = 0.04776
```

The box-cox transformation violated more assumptions than the original model. We will cotinue to use the additive model.

```
loading_and_line_plot <- interaction.plot(x.factor = bubblewrap$loading, response = bubblewrap$rate, tra
```




```
line_and_loading_plot <- interaction.plot(trace.factor = bubblewrap$loading, response = bubblewrap$rate
```



Some interactions may be present in the model. We test for their significance.

```
int_mod  <- lm(rate ~ -1 + loading * line_speed, data = bubblewrap)
anova(add_mod, int_mod)
```

```
## Analysis of Variance Table
##
## Model 1: rate ~ -1 + loading + line_speed
## Model 2: rate ~ -1 + loading * line_speed
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1     22 69014
## 2     18 61169  4    7844.4 0.5771 0.6829
```

The interaction term is not significant.

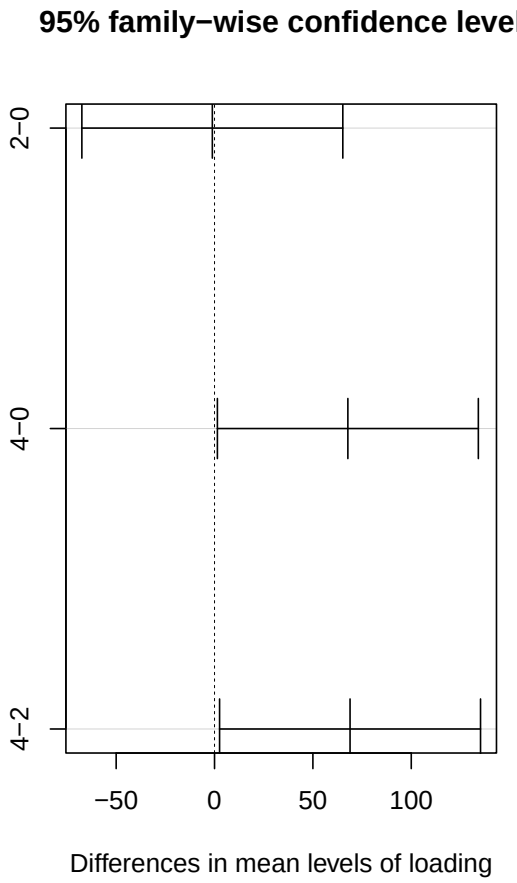
We move on to the pairwise differences.

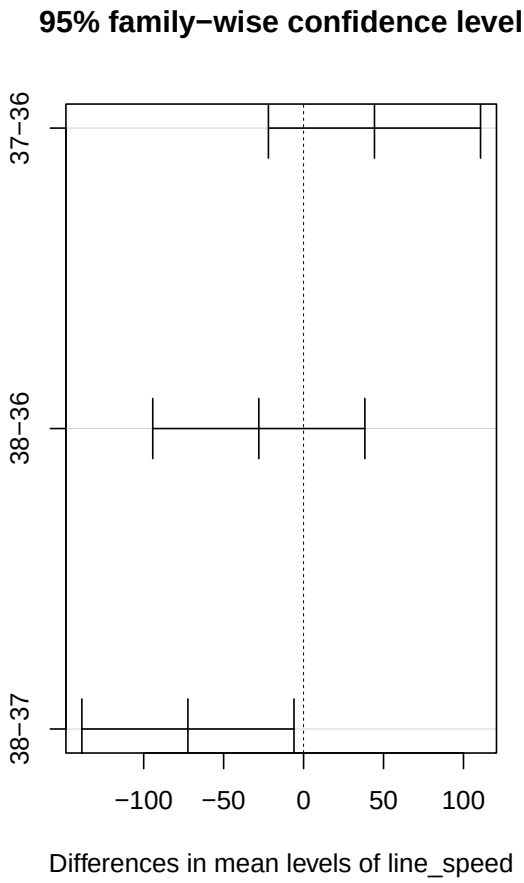
```
par(mfrow = c(1,2))
TukeyHSD(aov_model)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = rate ~ -1 + loading + line_speed, data = bubblewrap)
##
## $loading
```

```
##          diff        lwr       upr     p adj
## 2-0 -1.111111 -67.436661  65.21444 0.9990241
## 4-0 67.777778   1.452228 134.10333 0.0445852
## 4-2 68.888889   2.563339 135.21444 0.0408082
##
## $line_speed
##            diff        lwr        upr     p adj
## 37-36  44.33333  -21.99222 110.658883 0.2353973
## 38-36 -28.00000  -94.32555  38.325550 0.5478326
## 38-37 -72.33333 -138.65888  -6.007784 0.0308834
```

```
plot(TukeyHSD(aov_model))
```





#Modelling