

705604096_stats101b_hw2

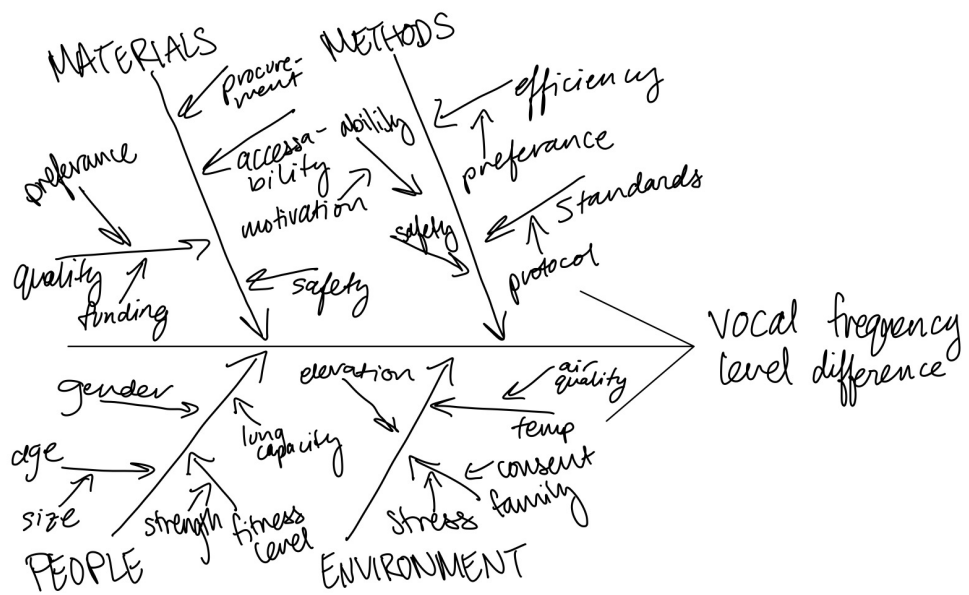
Jade Gregory

2023-08-24

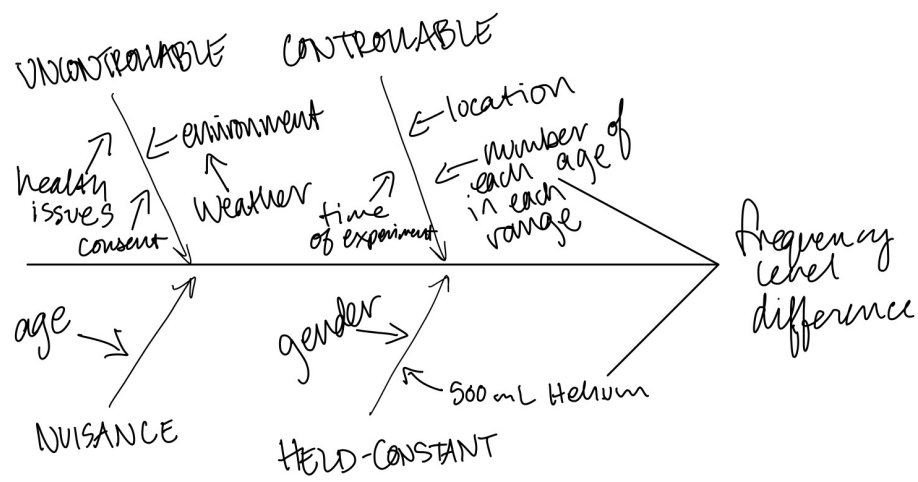
Question 1

```
idat <- read.csv("101bislandhw2.csv", header = TRUE)
idat
```

##	gender	name	age	vocal_frequency_before	vocal_frequency_after
## 1	female	Sarah Bager	92	205	280
## 2	female	Helga Thorn	102	186	254
## 3	female	Anika Olsen	24	183	240
## 4	female	Tindra Landvik	22	195	259
## 5	female	Summer Sato	8	248	346
## 6	female	Krista Mohr	16	205	220
## 7	female	Gabrielle Larsen	68	200	272
## 8	female	Sakura McCarthy	43	174	243
## 9	female	Kirsten Wolff	29	192	262
## 10	female	Vanessa Sorensen	65	195	268
## 11	female	Rosemarie Kramer	10	248	351
## 12	female	Sonja Jensen	27	181	225
## 13	female	Nicole Brown	29	191	262
## 14	female	Tineka Thorn	38	186	257
## 15	female	Pia Kapoor	8	241	320
## 16	female	Kelda Eklund	11	260	318
## 17	female	Ariane Baumann	30	183	242
## 18	female	Charu Datta	7	235	297
## 19	female	Summer Yamada	21	188	255
## 20	female	Kana Morris	120	197	271
## 21	female	Ciara Jones	22	174	259
## 22	female	Viktorija Ibsen	21	187	258
## 23	female	Sacha Ibsen	68	194	256
## 24	female	Elena Solberg	12	231	308



a)



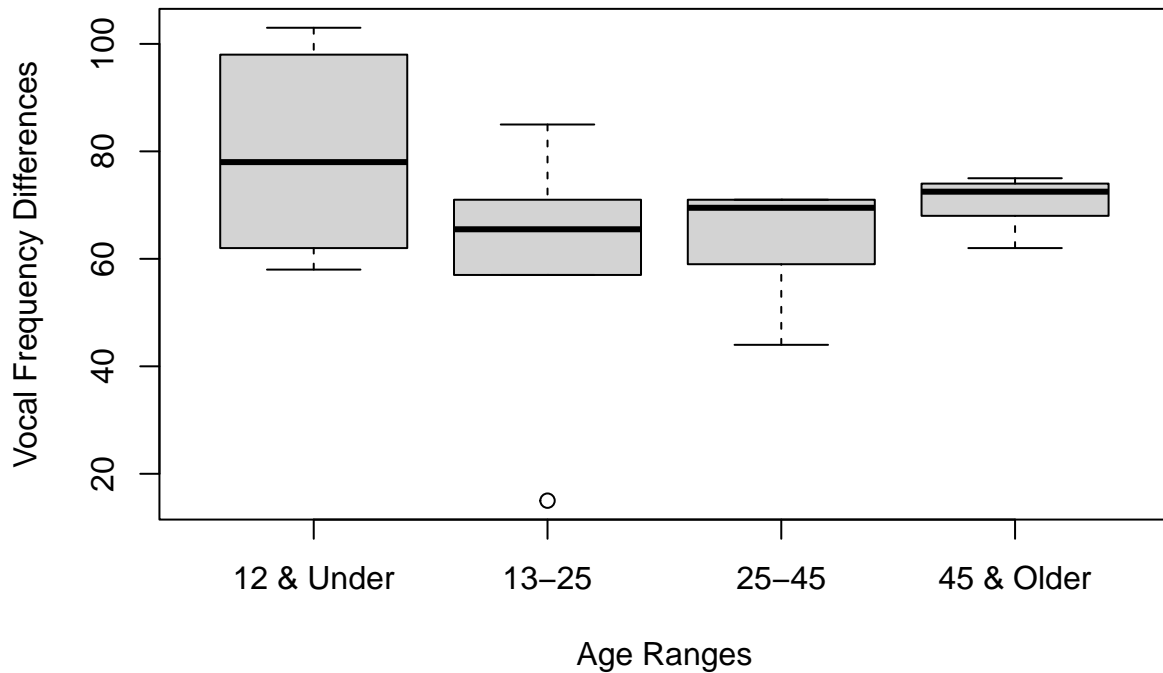
b)

c)

```

idat$difference <- idat$vocal_frequency_after - idat$vocal_frequency_before
idat$agerange <- cut(idat$age, breaks = c(0, 12, 25, 45, Inf), labels = c("12 & Under", "13-25", "25-45"))
boxplot(idat$difference ~ idat$agerange, xlab = "Age Ranges", ylab = "Vocal Frequency Differences")

```



This plot shows us that the means of each respective age range are relatively similar. This could indicate that the age range brackets do not have a significant effect on the vocal frequency difference in this experiment. For our 12 & Under age group, we see that the vocal frequency difference has a larger range of values than our older groups of ages. It appears that the oldest age range, 45 & Older, has the smallest range of values for vocal frequency difference.

d)

```
idataov <- aov(difference ~ agerange, data = idat)
summary(idataov)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## agerange    3   1326    442.1   1.695   0.2
## Residuals   20   5218    260.9
```

Our p-value is 0.2 which is greater than our significance level of 0.05 so we fail to reject the null hypothesis stating that there is no significant difference in the means of the age ranges. Therefore, we can conclude that there is evidence to support the idea that age range does not affect vocal frequency difference in our data.

Question 2

```
df_er <- 19 - 3
ss_er <- 196.04 - 36.15
ms_fac <- 36.15 / 3
ms_er <- ss_er / df_er
f <- ms_fac / ms_er
print(c(df_er, ss_er, ms_fac, ms_er, f))

## [1] 16.000000 159.890000 12.050000 9.993125 1.205829

p <- 0.339533
```

DF Error = 16 SS Error = 159.89 MS Factor = 12.05 MS Error = 9.993125 F = 1.205829 P = 0.339533

Question 3

a)

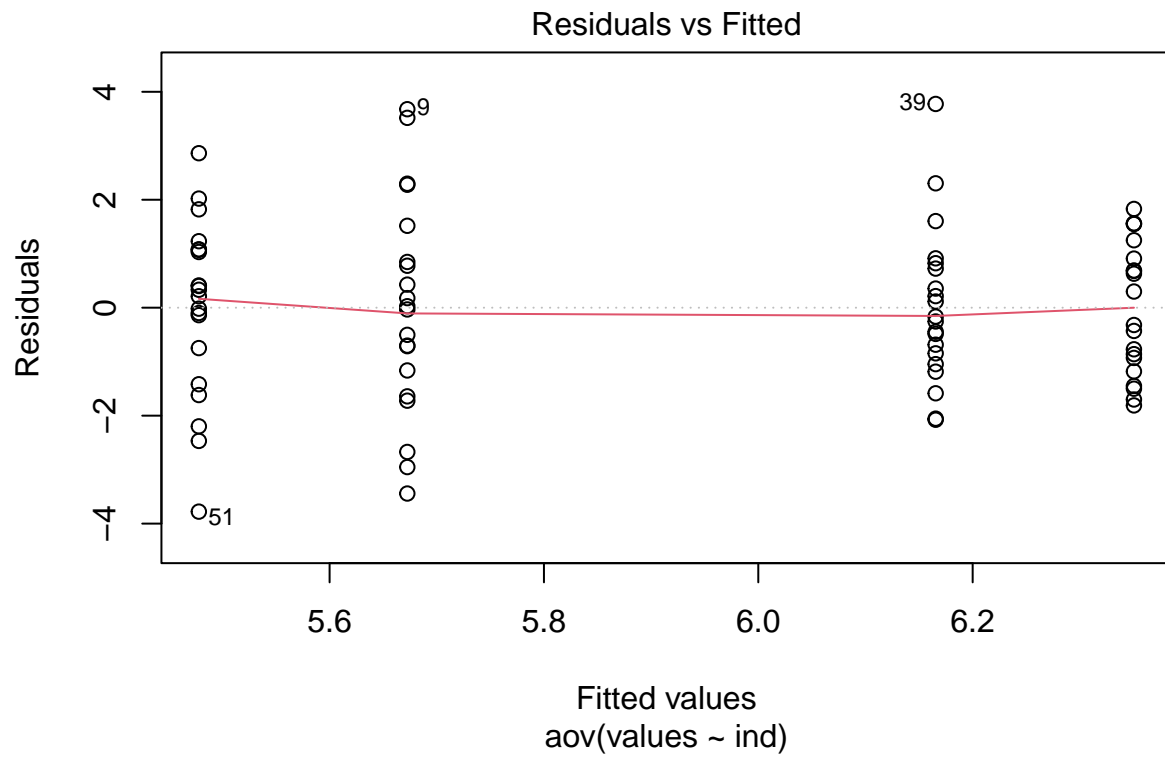
```
q3dat <- read.csv("HW2 Q3 Data Su C 2023.csv", header = TRUE)
q3aov <- aov(values ~ ind, data = stack(q3dat))
summary(q3aov)
```

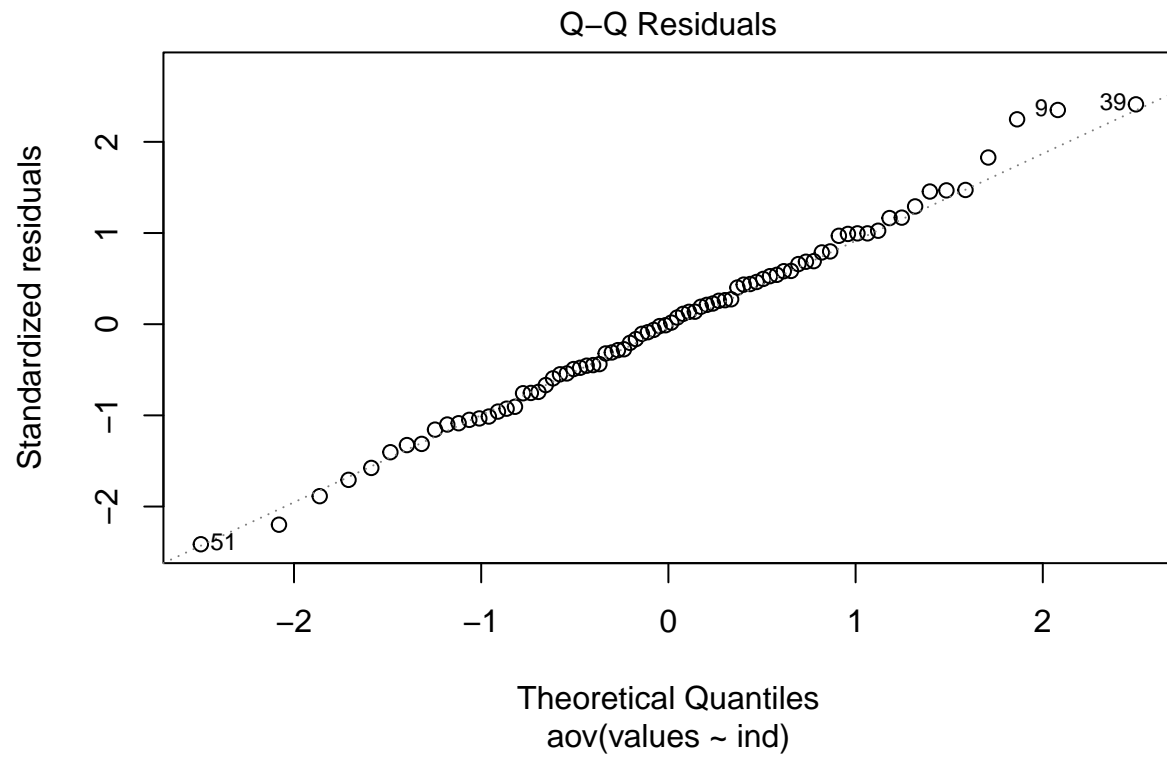
```
##           Df Sum Sq Mean Sq F value Pr(>F)
## ind         3  10.04   3.348   1.298  0.281
## Residuals   76 196.03   2.579
```

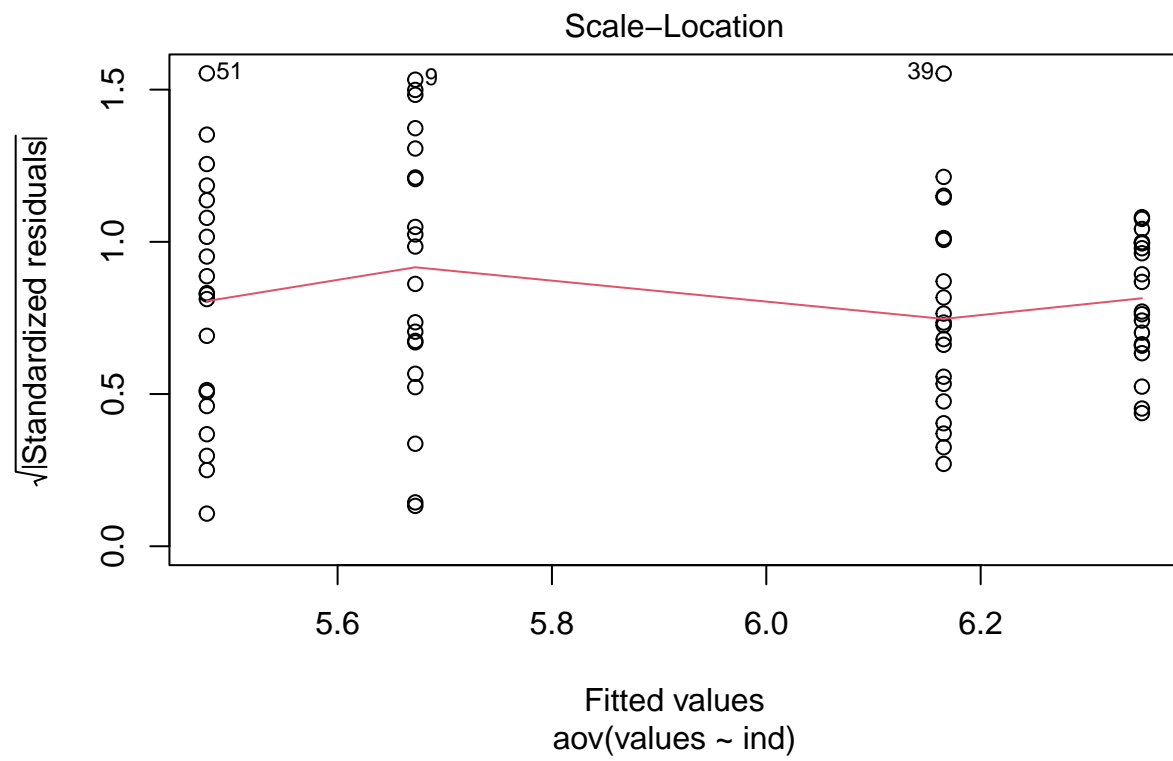
Since our p-value of 0.281 is greater than our significance level of 0.05, we fail to reject the null hypothesis stating there is no difference in the means of the different factor levels. So, we can conclude that there is no evidence to support a claim that PEMF usage affects BMD loss.

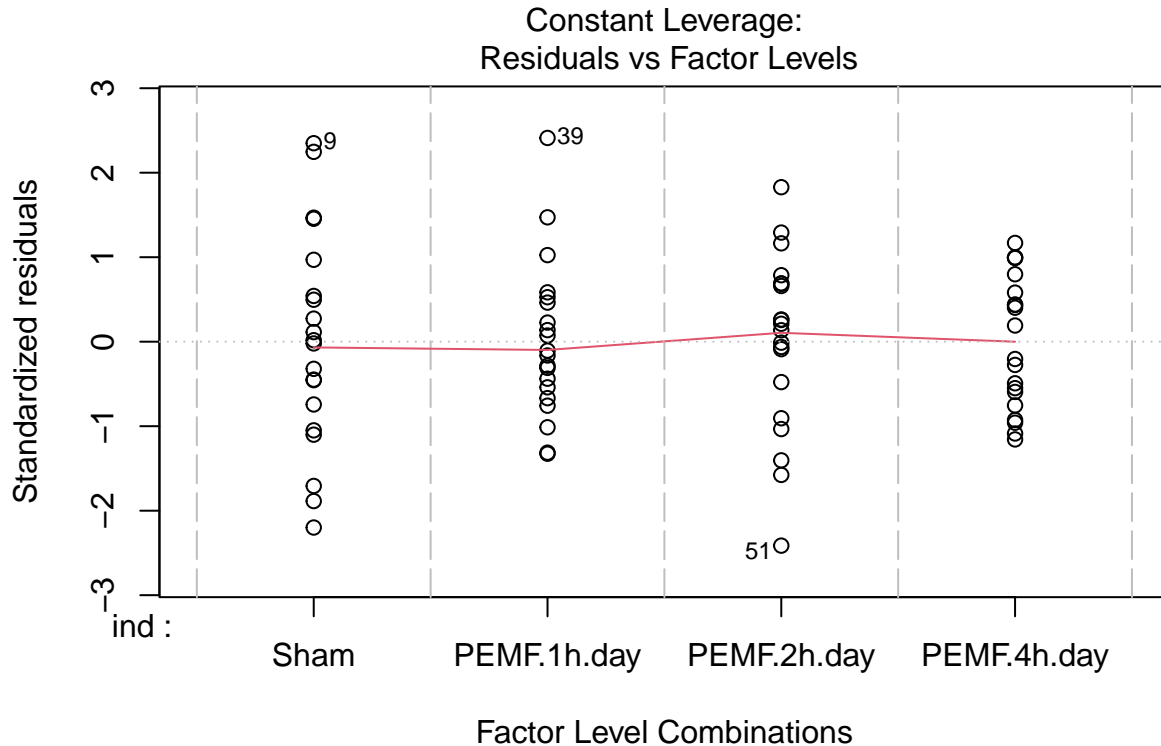
b)

```
plot(q3aov)
```









In our residuals vs fitted plot, we can see our data points are plotted equally and horizontally across the graph with no noticeable trend supporting the idea that the constant variance assumption is held by our model. In our QQ norm plot we observe that the data points tightly follow the dashed line and do not often stray suggesting that our model holds the normality assumption. In our scale location plot our data points are plotted horizontally across the graph with no noticeable pattern further supporting the idea that our constant variance assumption is held by our model. In our residuals vs factor levels plot, we can see that the data points are plotted equally and horizontally across the graph supporting the idea that our model appropriately accounts for the contribution of each factor level in it. Overall, I believe it is safe to conclude that our model is a great fit for our data based off of these graphs.

Question 4

a)

```
cotdat <- read.csv("HW2 Q4 data Su C 2023.csv", header = TRUE)
cotdat$CottonWeight <- as.factor(cotdat$CottonWeight)
cdataov <- aov(Observations ~ CottonWeight, data = cotdat)
summary(cdataov)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## CottonWeight  4  475.8   118.94   14.76 9.13e-06 ***
## Residuals    20   161.2     8.06
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With a p-value of 9.13e-06 which is less than our significance level of 0.05, we reject our null hypothesis and can conclude that there is evidence to support the claim that cotton contents appear to have an effect on the mean tensile strength.

b)

```
#installing packages
install.packages("agricolae")

## Error in contrib.url(repos, "source"): trying to use CRAN without setting a mirror

library(agricolae)

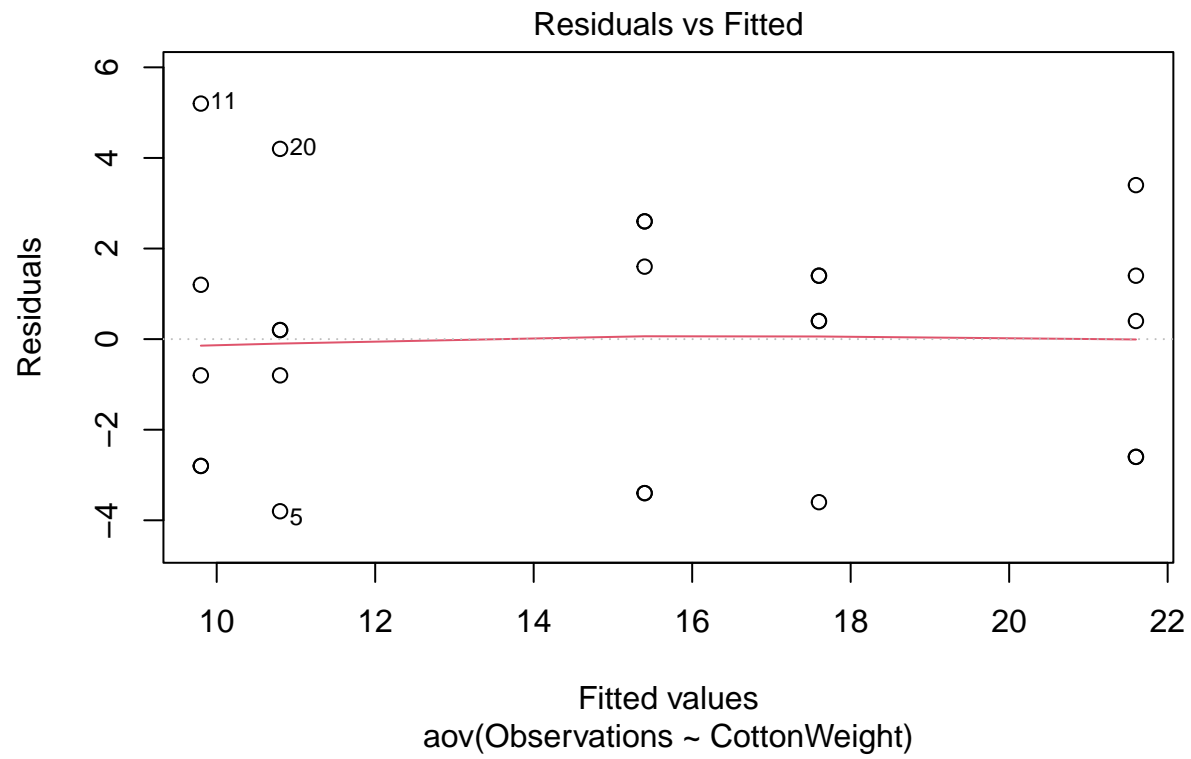
print(LSD.test(cdataov, "CottonWeight"))

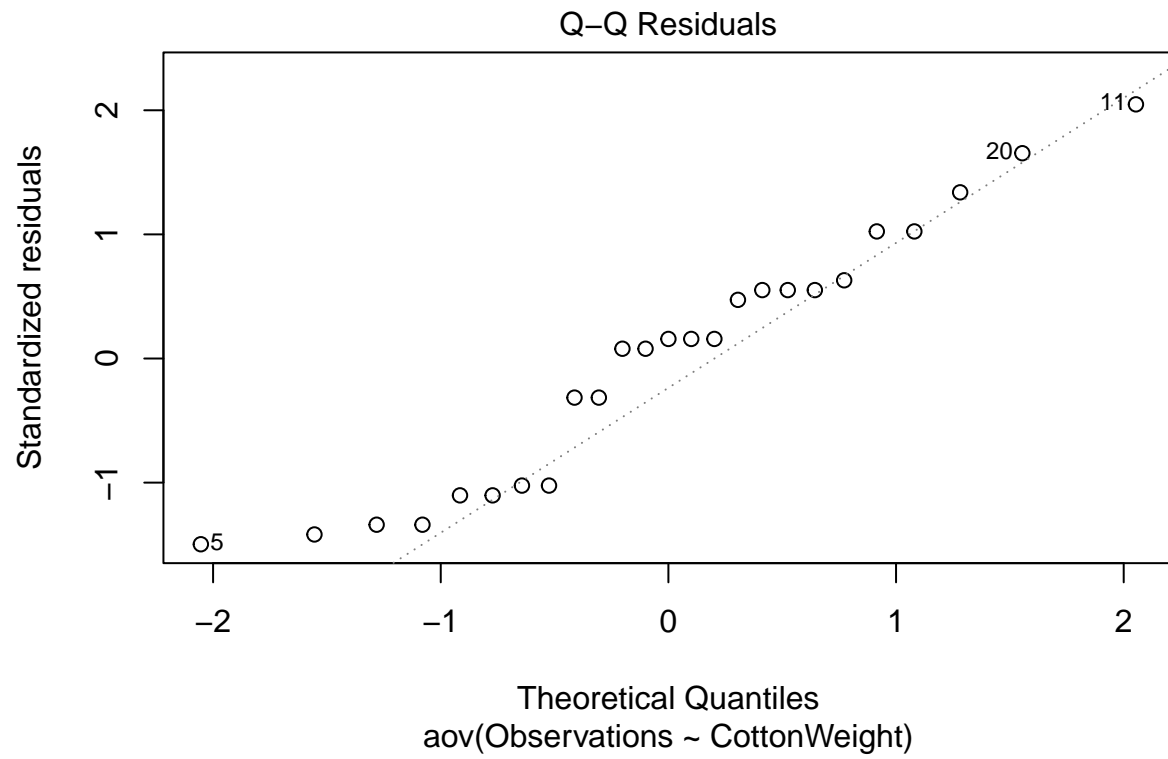
## $statistics
##      MSerror Df  Mean      CV  t.value    LSD
##      8.06 20 15.04 18.87642 2.085963 3.745452
##
## $parameters
##      test p.adjusted      name.t ntr alpha
##  Fisher-LSD      none CottonWeight   5  0.05
##
## $means
##      Observations      std r      se      LCL      UCL Min Max Q25 Q50 Q75
## 15      9.8 3.346640 5 1.269646  7.151566 12.44843   7 15   7   9  11
## 20     15.4 3.130495 5 1.269646 12.751566 18.04843  12 18  12  17  18
## 25     17.6 2.073644 5 1.269646 14.951566 20.24843  14 19  18  18  19
## 30     21.6 2.607681 5 1.269646 18.951566 24.24843  19 25  19  22  23
## 35     10.8 2.863564 5 1.269646  8.151566 13.44843   7 15  10  11  11
##
## $comparison
## NULL
##
## $groups
##      Observations groups
## 30      21.6      a
## 25      17.6      b
## 20      15.4      b
## 35      10.8      c
## 15      9.8      c
##
## attr(,"class")
## [1] "group"
```

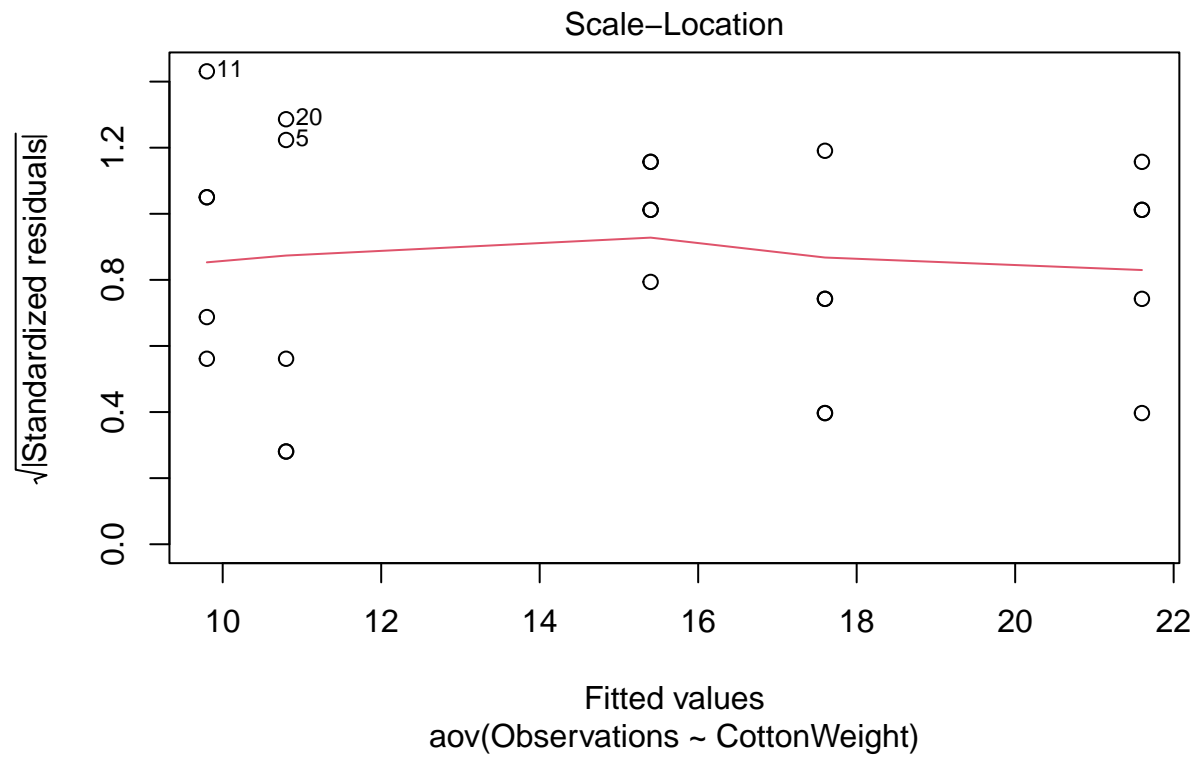
From our Fisher LSD test we can see that 15% cotton and 35% cotton do not have significantly different mean observations. We also see that 20% cotton and 25% cotton do not have significantly different mean observations, while the rest do have significantly different mean observation values.

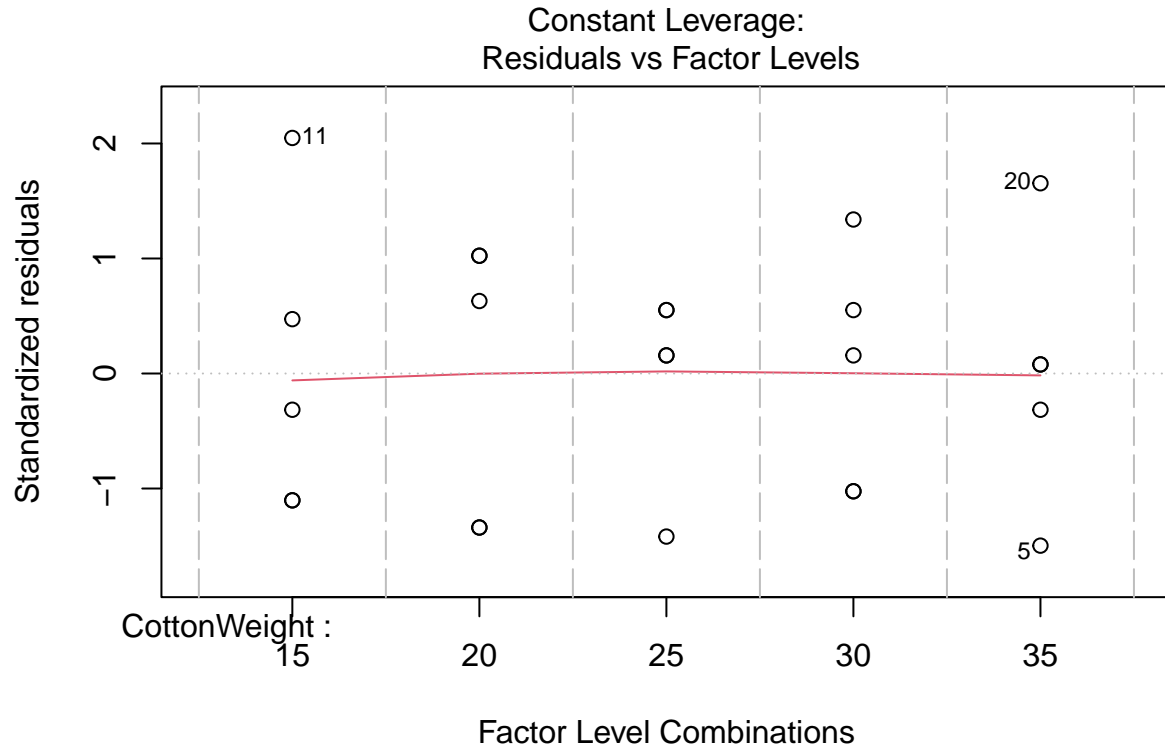
c)

```
plot(cdataov)
```









In our residuals vs fitted plot, we can see that our data points are plotted randomly and horizontally across the graph. This supports the idea that the constant variance assumption is held by our model. There may be a parabola like pattern that appears to some, but more data would be needed to be sure. If this is the case, it could potentially affect the constant variance assumption of our model. Our QQ norm plot displays our data points as following the dashed line, though they are not extremely tight to the dashed line they do not stray far from it therefore indicating that our normality assumption is held in our model. Our scale location plot displays our data points plotted equally and horizontally across the graph indicating that our constant variance assumption is held by our model. The residuals vs factor levels graph has our data points plotted equally across the graph indicating that our model accurately accounts for the contribution of each factor level in our model. Overall, I would support the idea that this model is a good fit for our data.