HW Week 6

Md Ariful Haque Miah

10/8/2022

```r
# Problem 3.23
ft1 <- c(17.6,18.9,16.3,17.4,20.1,21.6)
ft2 <- c(16.9,15.3,18.6,17.1,19.5,20.3)
ft3 <- c(21.4,23.6,19.4,18.5,20.5,22.3)
ft4 <- c(19.3,21.1,16.9,17.5,18.3,19.8)
dat <- data.frame(ft1,ft2,ft3,ft4)
dat
```

```
##    ft1  ft2  ft3  ft4
## 1 17.6 16.9 21.4 19.3
## 2 18.9 15.3 23.6 21.1
## 3 16.3 18.6 19.4 16.9
## 4 17.4 17.1 18.5 17.5
## 5 20.1 19.5 20.5 18.3
## 6 21.6 20.3 22.3 19.8
```

```r
library(tidyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
dat1 <- pivot_longer(dat,c(ft1,ft2,ft3,ft4))
dat1
```

```
## # A tibble: 24 x 2
##    name  value
##    <chr> <dbl>
##  1 ft1    17.6
##  2 ft2    16.9
##  3 ft3    21.4
##  4 ft4    19.3
##  5 ft1    18.9
##  6 ft2    15.3
##  7 ft3    23.6
##  8 ft4    21.1
##  9 ft1    16.3
```

```
## 10 ft2    18.6
## # ... with 14 more rows
```

```r
colnames(dat1)<-c("Fluid Type","Life")
dat1
```

```
## # A tibble: 24 x 2
##    `Fluid Type`  Life
##    <chr>        <dbl>
##  1 ft1           17.6
##  2 ft2           16.9
##  3 ft3           21.4
##  4 ft4           19.3
##  5 ft1           18.9
##  6 ft2           15.3
##  7 ft3           23.6
##  8 ft4           21.1
##  9 ft1           16.3
## 10 ft2           18.6
## # ... with 14 more rows
```

```r
dat1$`Fluid Type` <- as.factor(dat1$`Fluid Type`)
dat1$Life <- as.numeric(dat1$Life)
str(dat1)
```

```
## tibble [24 x 2] (S3: tbl_df/tbl/data.frame)
##  $ Fluid Type: Factor w/ 4 levels "ft1","ft2","ft3",..: 1 2 3 4 1 2 3 4 1 2 ...
##  $ Life      : num [1:24] 17.6 16.9 21.4 19.3 18.9 15.3 23.6 21.1 16.3 18.6 ...
```

```r
model <- aov(dat1$Life~dat1$`Fluid Type`,data=dat1)
summary(model)
```

```
##                   Df Sum Sq Mean Sq F value Pr(>F)
## dat1$`Fluid Type`  3  30.16   10.05   3.047 0.0525 .
## Residuals         20  65.99    3.30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Problem 3.23.(a)
# Hypothesis
# H0: u1=u2=u3=u4
# Ha: at least one of the u(i) differs
# u(i) = mean of the Fluid Type (i=1,2,3,4)
# Since P-value (0.0525) > alpha so we fail to reject H0 at alpha = 0.05.
# There are no differences in fluid type, also the P-value is just slightly above 0.05,
# there is probably a difference in fluid type.
# Problem 3.23.(b)
library(agricolae)
LSD.test(model,"dat1$`Fluid Type`",console = TRUE)
```

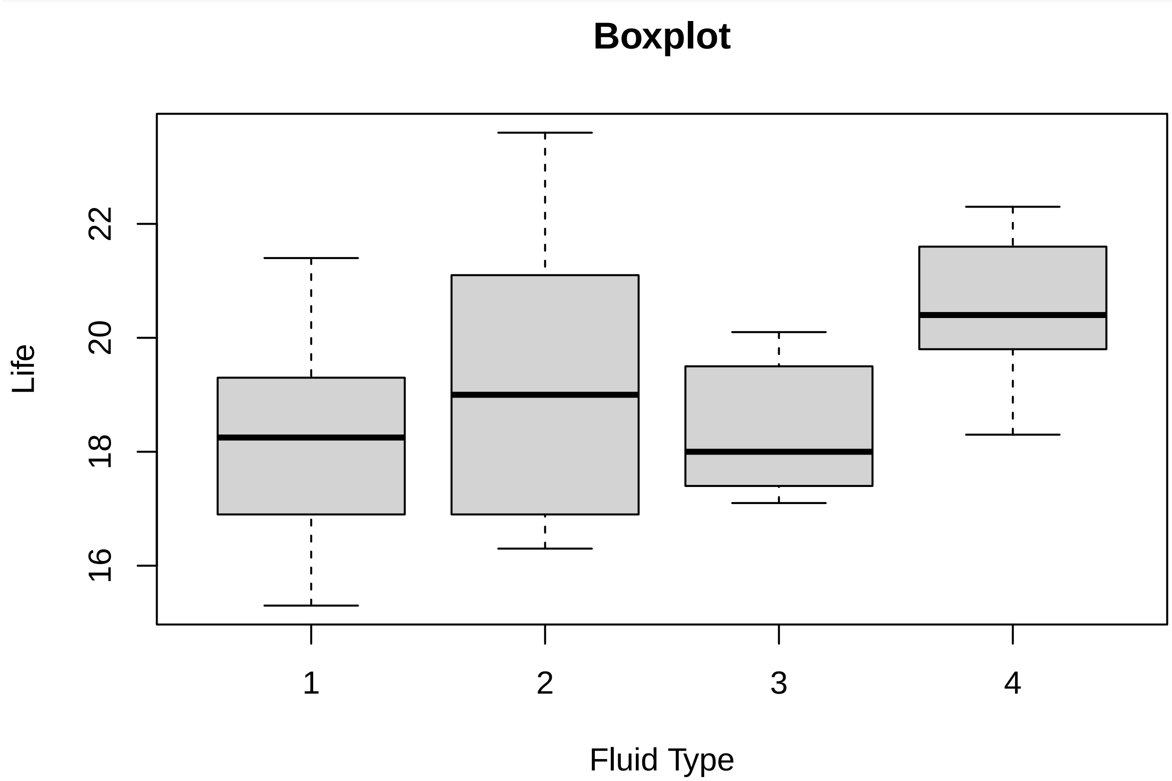
```
##
## Study: model ~ "dat1$`Fluid Type`"
##
## LSD t Test for dat1$Life
##
## Mean Square Error:  3.299667
##
```

```
## dat1$`Fluid Type`,  means and individual ( 95 %) CI
##
##      dat1.Life      std r      LCL      UCL  Min  Max
## ft1  18.65000 1.952178 6 17.10309 20.19691 16.3 21.6
## ft2  17.95000 1.854454 6 16.40309 19.49691 15.3 20.3
## ft3  20.95000 1.879096 6 19.40309 22.49691 18.5 23.6
## ft4  18.81667 1.554885 6 17.26975 20.36358 16.9 21.1
##
## Alpha: 0.05 ; DF Error: 20
## Critical Value of t: 2.085963
##
## least Significant Difference: 2.187666
##
## Treatments with the same letter are not significantly different.
##
##      dat1$Life groups
## ft3  20.95000      a
## ft4  18.81667     ab
## ft1  18.65000      b
## ft2  17.95000      b
```

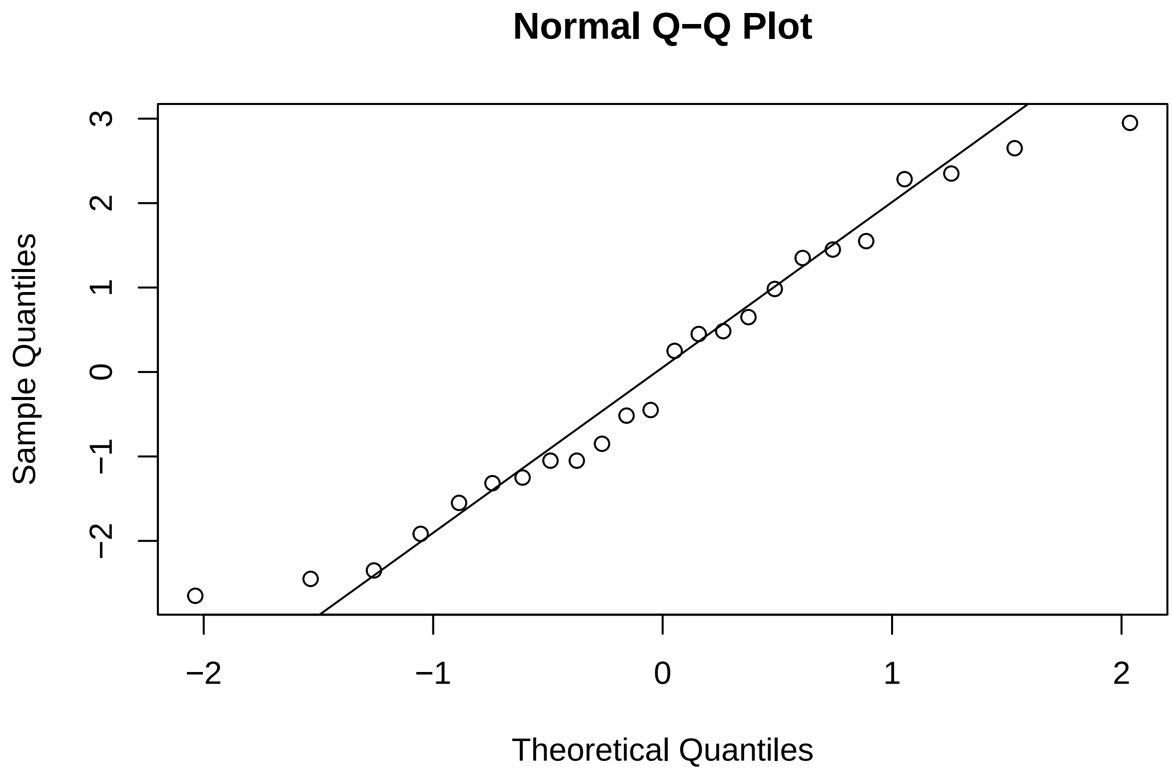
```
# According to the LSD test, I would select fluid type 3 as the fluid type 3 is different from
# the others, and it's mean life also exceeds the mean lives of the other three fluids.
# Problem 3.23.(c)
Life <- c(dat1$Life)
str(Life)
```

```
##  num [1:24] 17.6 16.9 21.4 19.3 18.9 15.3 23.6 21.1 16.3 18.6 ...
```

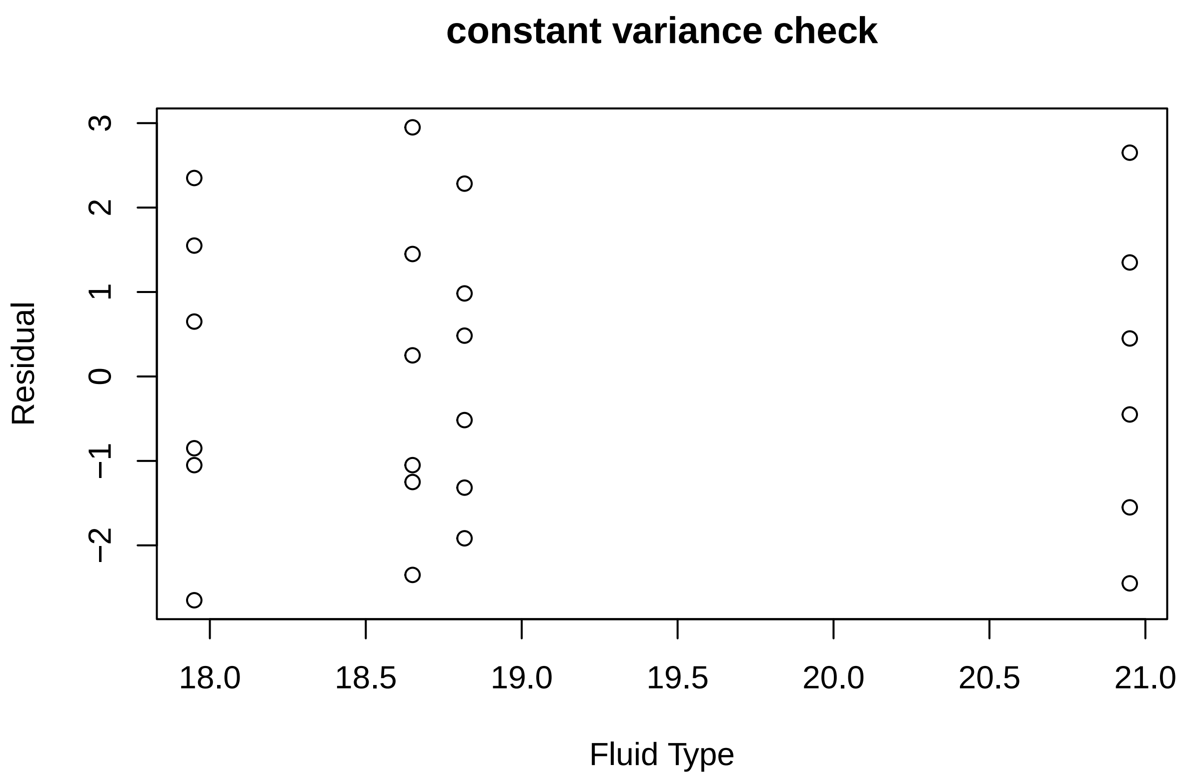
```
x <- c(rep(1,6),rep(2,6),rep(3,6),rep(4,6))
boxplot(Life~x,xlab="Fluid Type",ylab="Life",main="Boxplot")
```



```
meanx<-c(rep(mean(ft1),6),rep(mean(ft2),6),rep(mean(ft3),6),rep(mean(ft4),6))
Type<-c(ft1,ft2,ft3,ft4)
res<-Type-meanx
qqnorm(res)
qqline(res)
```




```
plot(meanx,res,xlab="Fluid Type",ylab="Residual",
     main="constant variance check")
```




```r
# From the basic analysis of variance assumptions, we see that the residuals have
# the same spread/disperse. Hence the variance is constant and there is nothing
# unusual in the residual plots.
# From the qq plot, we see that the data appears to be normally distributed.
# Hence our ANOVA assumptions are satisfied.
# Problem 3.28
m1 <- c(110,157,194,178)
m2 <- c(1,2,4,18)
m3 <- c(880,1256,5276,4355)
m4 <- c(495,7040,5307,10050)
m5 <- c(7,5,29,2)
dat <- data.frame(m1,m2,m3,m4,m5)
dat
```

```
##    m1 m2   m3    m4 m5
## 1 110  1  880   495  7
## 2 157  2 1256  7040  5
## 3 194  4 5276  5307 29
## 4 178 18 4355 10050  2
```

```r
library(tidyr)
library(dplyr)
dat1 <- pivot_longer(dat,c(m1,m2,m3,m4,m5))
dat1
```

```
## # A tibble: 20 x 2
##    name  value
##    <chr> <dbl>
##  1 m1      110
##  2 m2        1
##  3 m3      880
```

```
##   4 m4      495
##   5 m5        7
##   6 m1      157
##   7 m2        2
##   8 m3     1256
##   9 m4     7040
##  10 m5        5
##  11 m1      194
##  12 m2        4
##  13 m3     5276
##  14 m4     5307
##  15 m5       29
##  16 m1      178
##  17 m2       18
##  18 m3     4355
##  19 m4    10050
##  20 m5        2
```

```r
colnames(dat1)<-c("Material","Failure Time")
dat1
```

```
## # A tibble: 20 x 2
##    Material `Failure Time`
##    <chr>             <dbl>
##  1 m1                  110
##  2 m2                    1
##  3 m3                  880
##  4 m4                  495
##  5 m5                    7
##  6 m1                  157
##  7 m2                    2
##  8 m3                 1256
##  9 m4                 7040
## 10 m5                    5
## 11 m1                  194
## 12 m2                    4
## 13 m3                 5276
## 14 m4                 5307
## 15 m5                   29
## 16 m1                  178
## 17 m2                   18
## 18 m3                 4355
## 19 m4                10050
## 20 m5                    2
```

```r
dat1$Material <- as.factor(dat1$Material)
dat1$`Failure Time` <- as.numeric(dat1$`Failure Time`)
str(dat1)
```

```
## tibble [20 x 2] (S3: tbl_df/tbl/data.frame)
##  $ Material    : Factor w/ 5 levels "m1","m2","m3",..: 1 2 3 4 5 1 2 3 4 5 ...
##  $ Failure Time: num [1:20] 110 1 880 495 7 ...
```

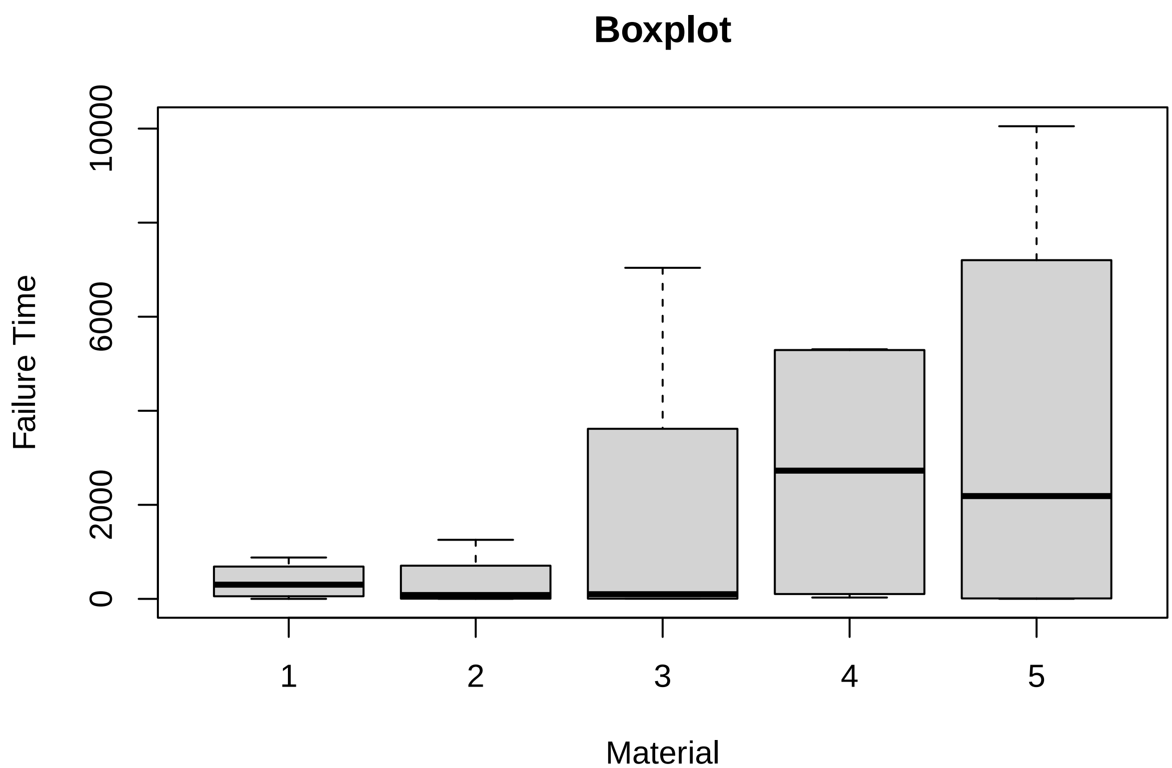
```r
model2 <- aov(dat1$`Failure Time`~dat1$Material,data=dat1)
summary(model2)
```

```
##                Df    Sum Sq  Mean Sq F value  Pr(>F)
## dat1$Material  4 103191489 25797872   6.191 0.00379 **
## Residuals     15  62505657  4167044
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

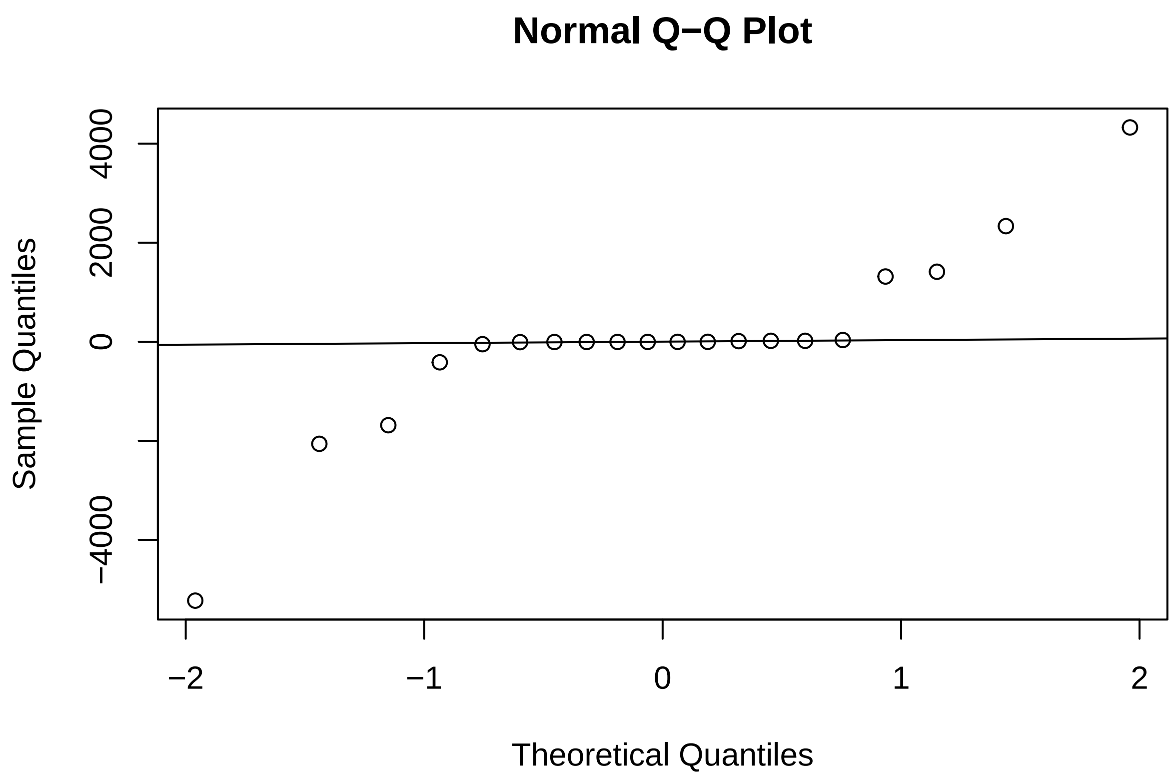
```
# Problem 3.28.(a)
# Hypothesis
# HO: u1=u2=u3=u4=u5
# Ha: at least one of the u(i) differs
# u(i) = mean of the Material (i=1,2,3,4,5)
# Since P-value (0.00379) < alpha so we reject HO at alpha = 0.05.
# No not all the five materials have the same effect on mean failure time.
# at least one material is different.
# Problem 3.28.(b)
ft <- c(dat1$`Failure Time`)
str(ft)
```

```
##  num [1:20] 110 1 880 495 7 ...
```

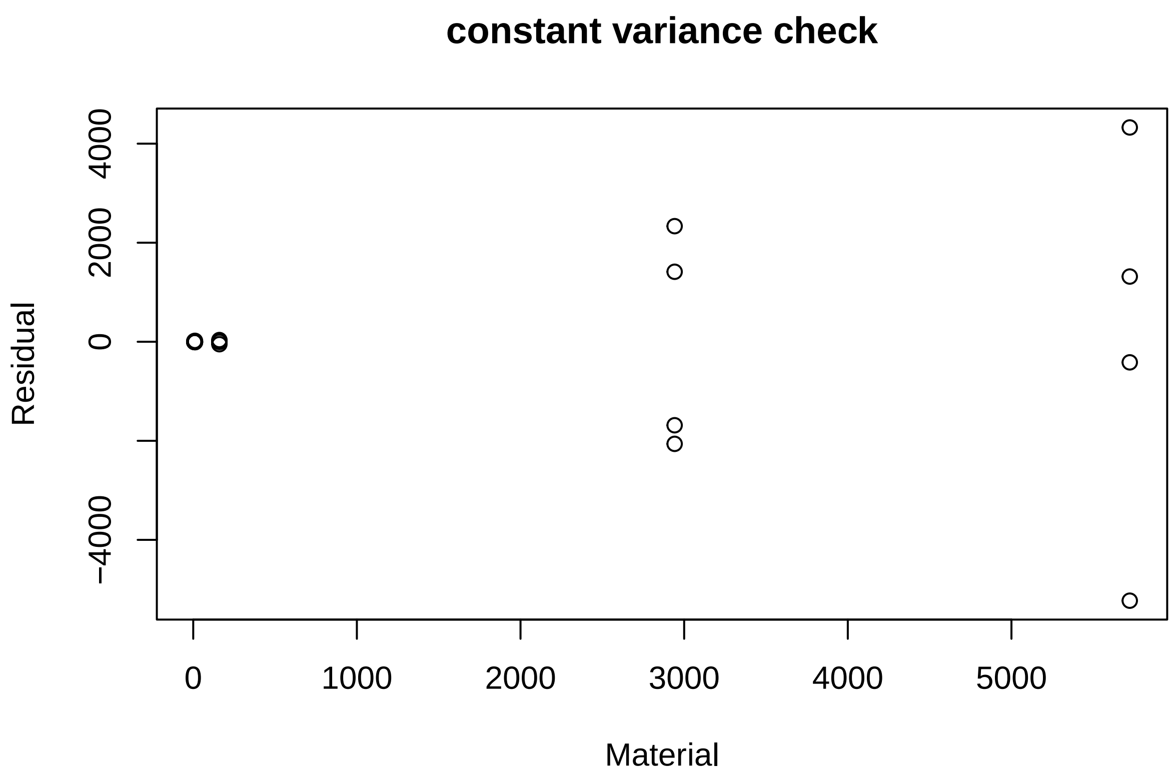
```
x <- c(rep(1,4),rep(2,4),rep(3,4),rep(4,4),rep(5,4))
boxplot(ft~x,xlab="Material",ylab="Failure Time",main="Boxplot")
```




```
meanx<-c(rep(mean(m1),4),rep(mean(m2),4),rep(mean(m3),4),rep(mean(m4),4),rep(mean(m5),4))
Material<-c(m1,m2,m3,m4,m5)
res<-Material-meanx
qqnorm(res)
qqline(res)
```




```
plot(meanx,res,xlab="Material",ylab="Residual",
     main="constant variance check")
```




```
# The plot of residuals versus predicted indicates the variance of the original observations
# is not constant. The normal probability plot also indicates that the normality assumption
# is not valid. Boxcox transformation is needed.
```

```r
# Problem 3.28.(c)
library(MASS)
```

```
##
## Attaching package: 'MASS'

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

```r
m1 <- c(110,157,194,178)
m2 <- c(1,2,4,18)
m3 <- c(880,1256,5276,4355)
m4 <- c(495,7040,5307,10050)
m5 <- c(7,5,29,2)
dat <- data.frame(m1,m2,m3,m4,m5)
dat
```

```
##    m1 m2   m3    m4 m5
## 1 110  1  880   495  7
## 2 157  2 1256  7040  5
## 3 194  4 5276  5307 29
## 4 178 18 4355 10050  2
```

```r
library(tidyr)
library(dplyr)
dat1 <- pivot_longer(dat,c(m1,m2,m3,m4,m5))
dat1
```

```
## # A tibble: 20 x 2
##    name  value
##    <chr> <dbl>
##  1 m1      110
##  2 m2        1
##  3 m3      880
##  4 m4      495
##  5 m5        7
##  6 m1      157
##  7 m2        2
##  8 m3     1256
##  9 m4     7040
## 10 m5        5
## 11 m1      194
## 12 m2        4
## 13 m3     5276
## 14 m4     5307
## 15 m5       29
## 16 m1      178
## 17 m2       18
## 18 m3     4355
## 19 m4    10050
## 20 m5        2
```

```r
colnames(dat1)<-c("Material","Failure Time")
dat1
```

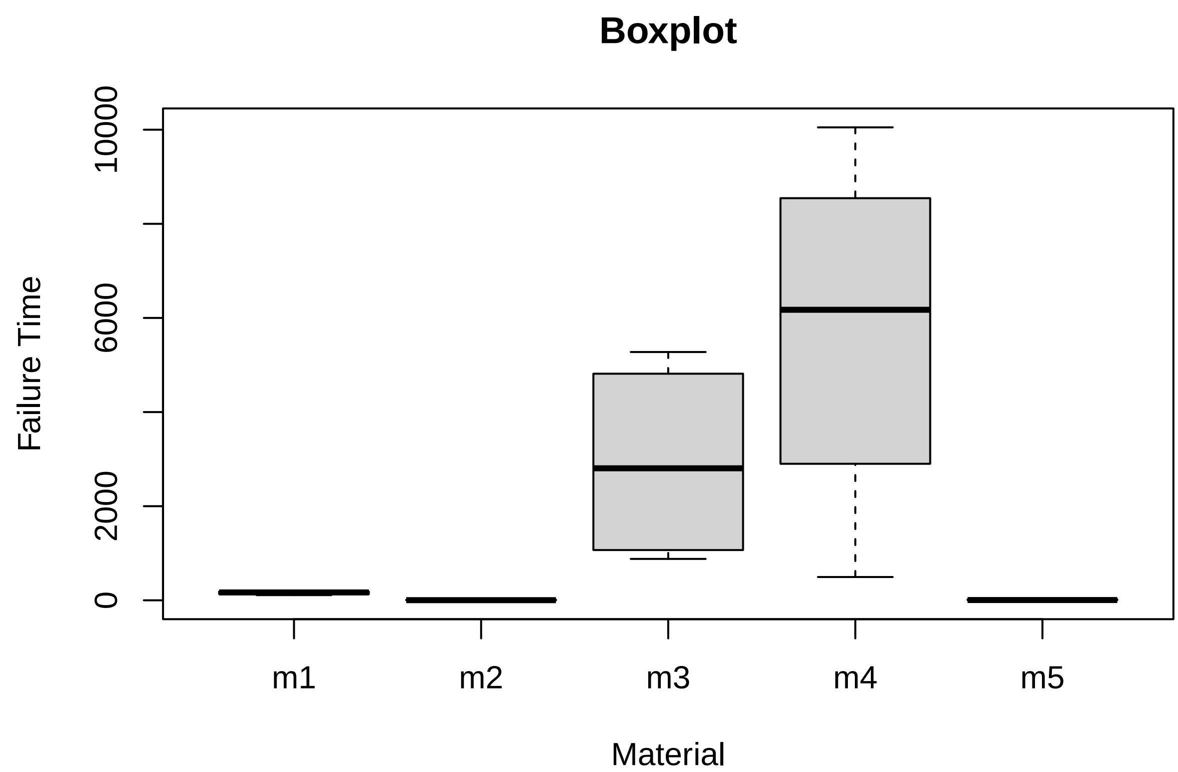
```
## # A tibble: 20 x 2
```

```
##    Material `Failure Time`
##    <chr>               <dbl>
##  1 m1                    110
##  2 m2                      1
##  3 m3                    880
##  4 m4                    495
##  5 m5                      7
##  6 m1                    157
##  7 m2                      2
##  8 m3                   1256
##  9 m4                   7040
## 10 m5                      5
## 11 m1                    194
## 12 m2                      4
## 13 m3                   5276
## 14 m4                   5307
## 15 m5                     29
## 16 m1                    178
## 17 m2                     18
## 18 m3                   4355
## 19 m4                  10050
## 20 m5                      2
```

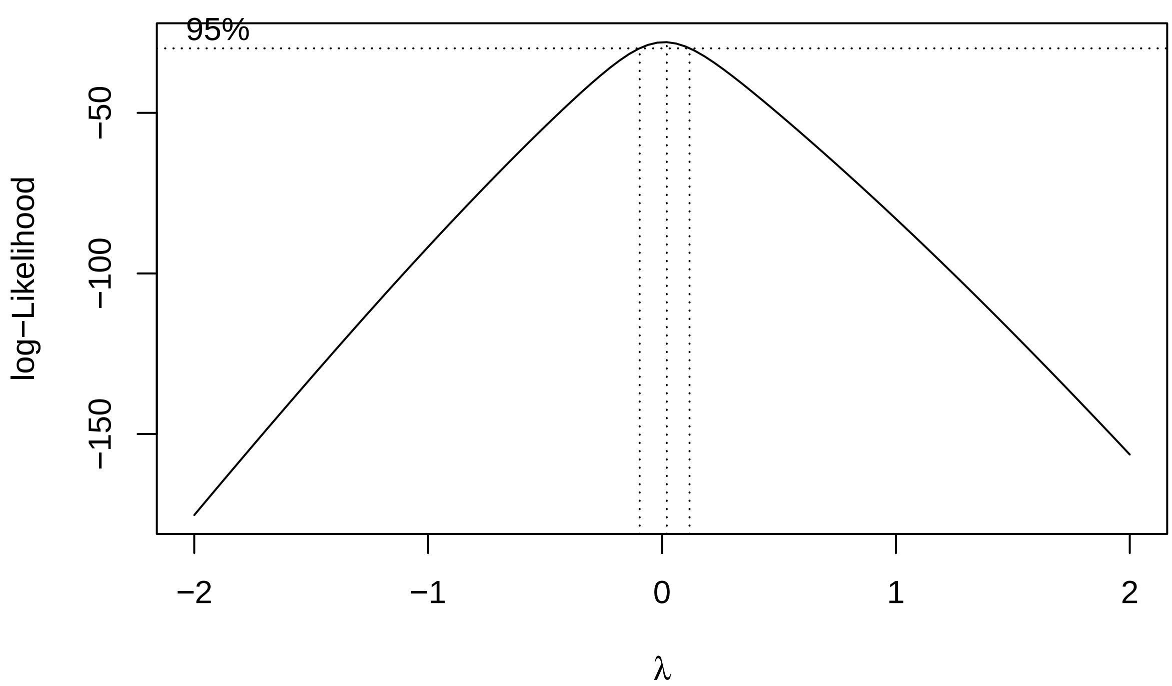
```
dat1$Material <- as.factor(dat1$Material)
dat1$`Failure Time` <- as.numeric(dat1$`Failure Time`)
str(dat1)
```

```
## tibble [20 x 2] (S3: tbl_df/tbl/data.frame)
##  $ Material    : Factor w/ 5 levels "m1","m2","m3",..: 1 2 3 4 5 1 2 3 4 5 ...
##  $ Failure Time: num [1:20] 110 1 880 495 7 ...
```

```
boxplot(dat1$`Failure Time`~dat1$Material,xlab="Material",ylab="Failure Time",main="Boxplot")
```



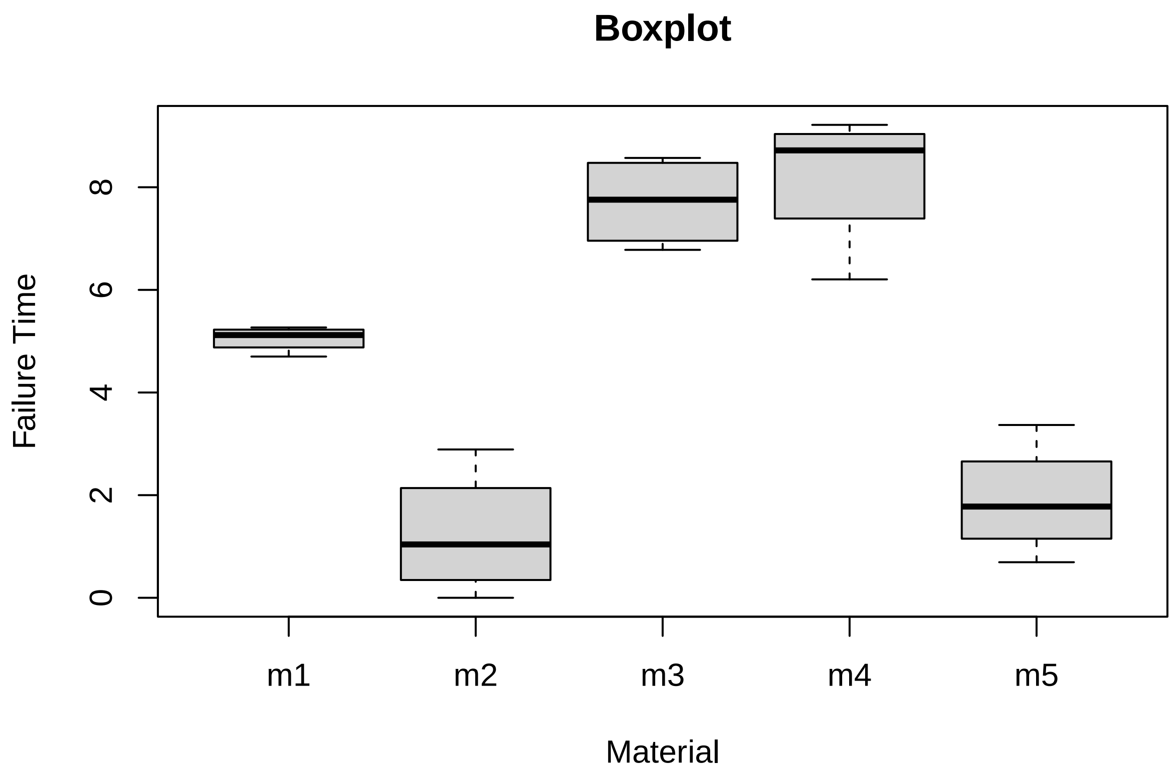

```r
boxcox(dat1$`Failure Time`~dat1$Material,data=dat1)
```




```r
# lambda=1 is not within 95% CI so we can go ahead to do boxcox transformations
y <- log(dat1$`Failure Time`)
head(y)
```

```
## [1] 4.700480 0.000000 6.779922 6.204558 1.945910 5.056246
```

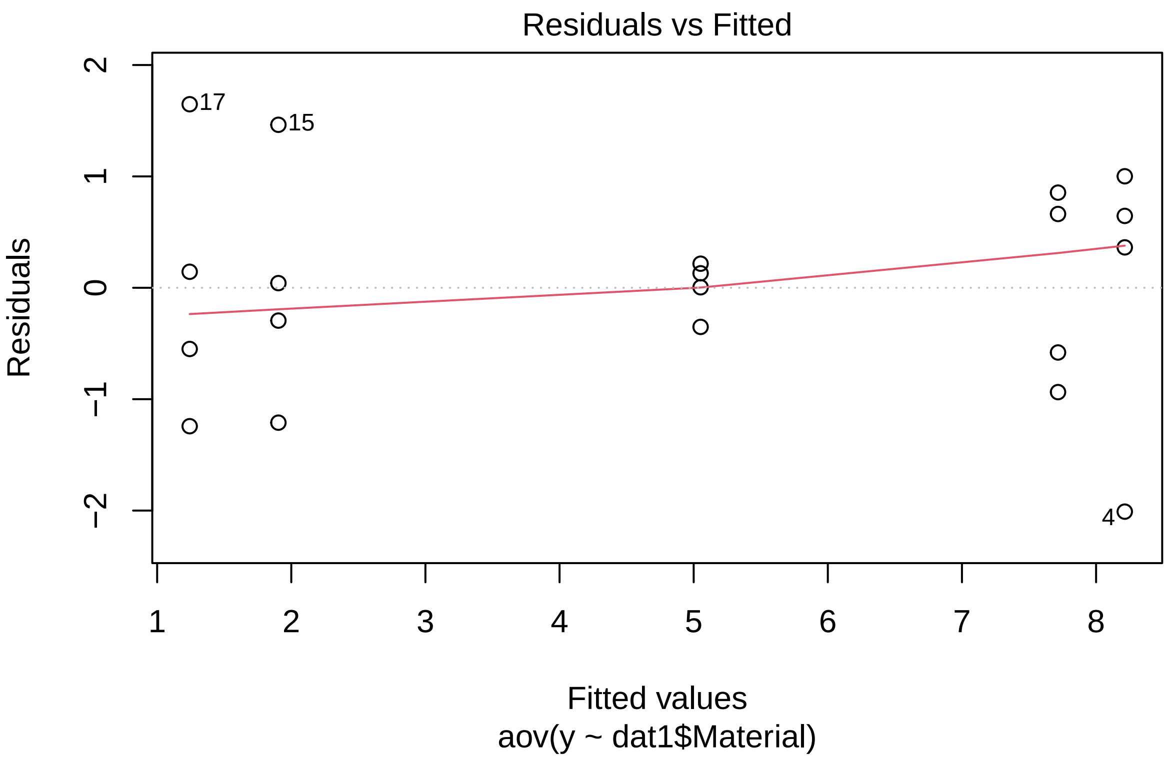
```r
boxplot(y~dat1$Material,xlab="Material",ylab="Failure Time",main="Boxplot")
```

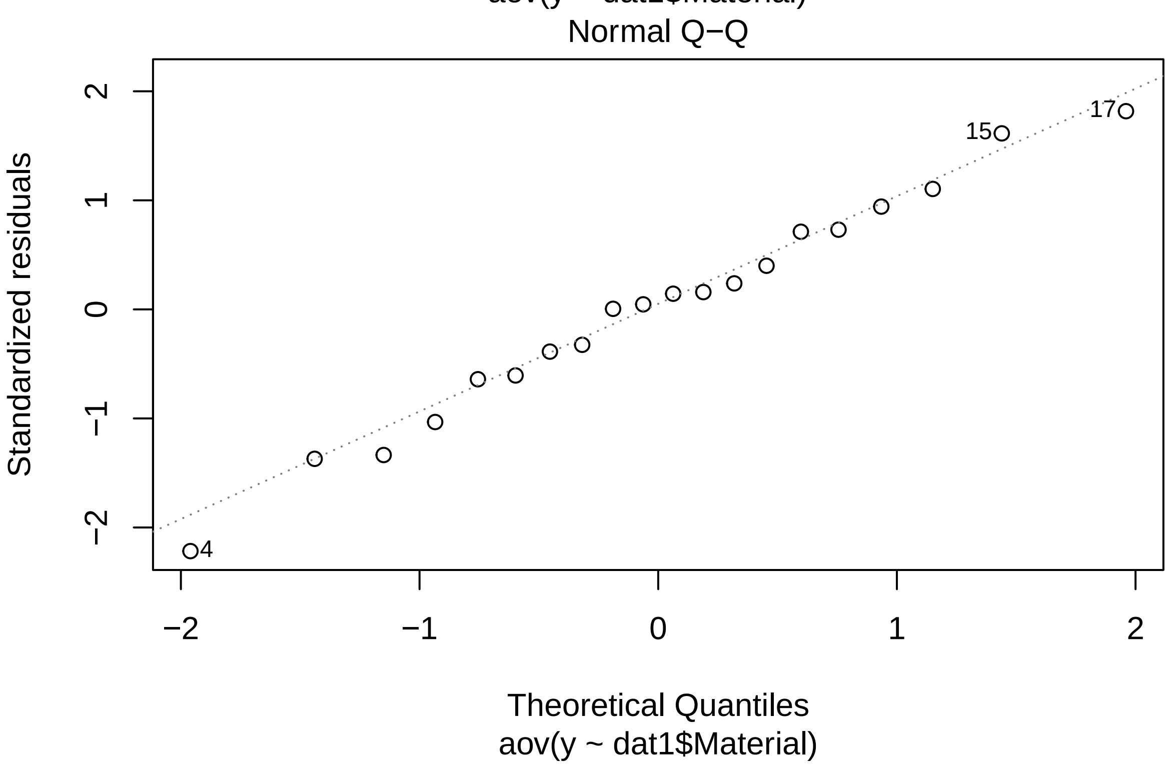


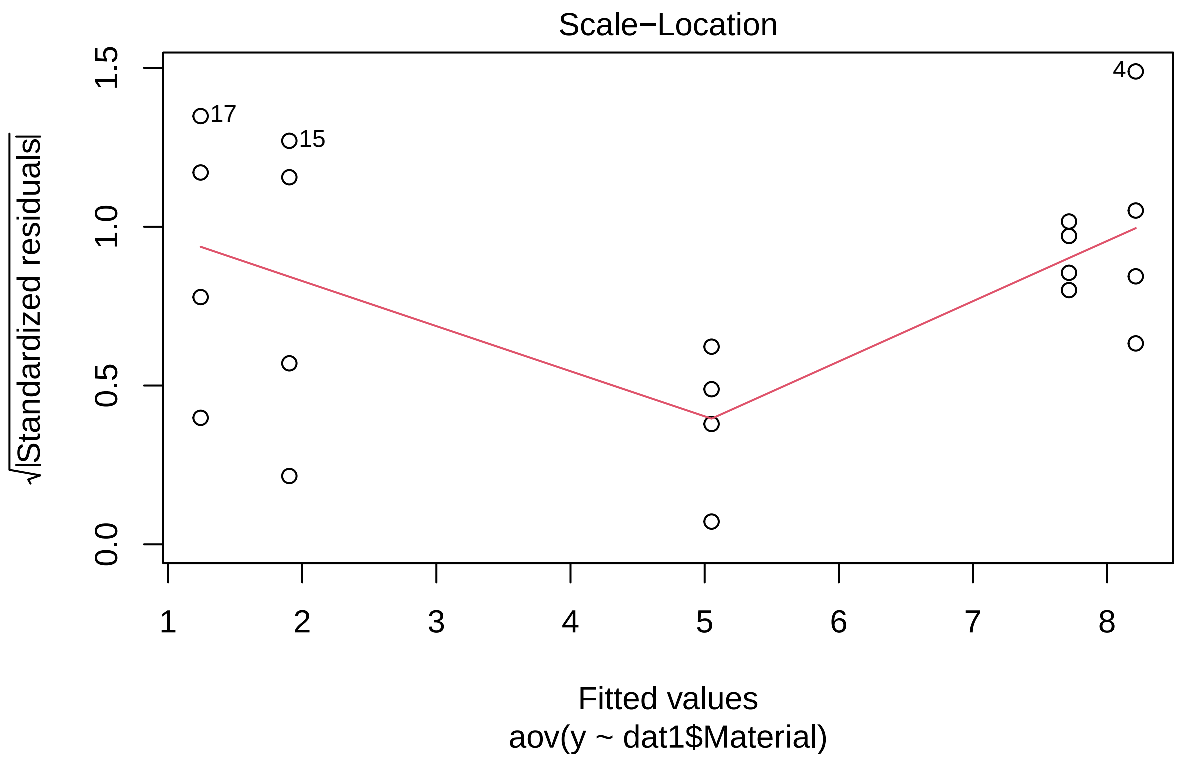

```r
model3 <- aov(y~dat1$Material,data=dat1)
summary(model3)
```

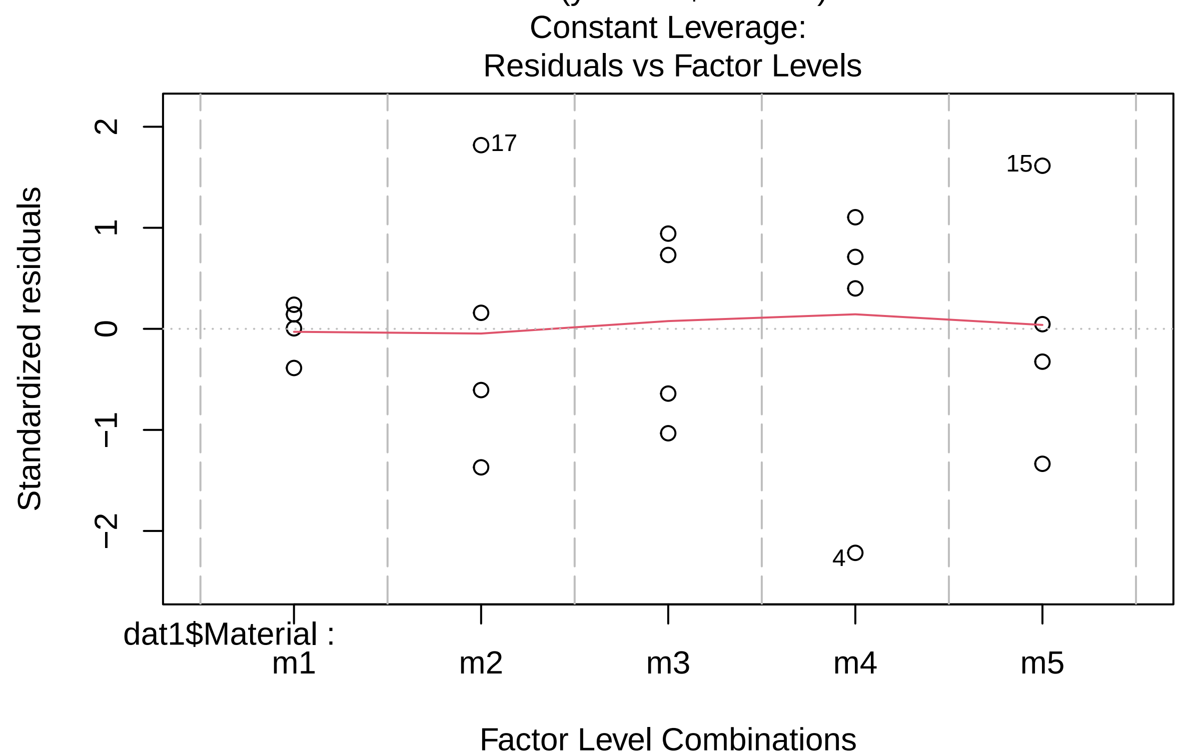
```
##               Df Sum Sq Mean Sq F value   Pr(>F)    
## dat1$Material  4 165.06   41.26   37.66 1.18e-07 ***
## Residuals     15  16.44    1.10                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plot(model3)
```


Residuals vs Fitted
Residuals
2
1
0
−1
−2
17
15
4
1
2
3
4
5
6
7
8
Fitted values
aov(y ~ dat1$Material)


Normal Q−Q
Standardized residuals
2
1
0
−1
−2
15
17
4
−2
−1
0
1
2
Theoretical Quantiles
aov(y ~ dat1$Material)






```
# Since P-value (1.18e-07) < alpha so we reject H0 at alpha = 0.05.
# No not all the five materials have the same effect on mean failure time.
# at least one material is different.
# After doing natural log transformation, we see that there is nothing unusual in the
# NPP and the residual plots so ANOVA assumptions are adequate with normal distribution
# and constant variance.
```

```r
# Problem 3.29
m1 <- c(31,10,21,4,1)
m2 <- c(62,40,24,30,35)
m3 <- c(53,27,120,97,68)
dat <- data.frame(m1,m2,m3)
dat
```

```
##   m1 m2  m3
## 1 31 62  53
## 2 10 40  27
## 3 21 24 120
## 4  4 30  97
## 5  1 35  68
```

```r
library(tidyr)
library(dplyr)
dat1 <- pivot_longer(dat,c(m1,m2,m3))
dat1
```

```
## # A tibble: 15 x 2
##    name  value
##    <chr> <dbl>
##  1 m1       31
##  2 m2       62
##  3 m3       53
##  4 m1       10
##  5 m2       40
##  6 m3       27
##  7 m1       21
##  8 m2       24
##  9 m3      120
## 10 m1        4
## 11 m2       30
## 12 m3       97
## 13 m1        1
## 14 m2       35
## 15 m3       68
```

```r
colnames(dat1)<-c("Method","Count")
dat1
```

```
## # A tibble: 15 x 2
##    Method Count
##    <chr>  <dbl>
##  1 m1        31
##  2 m2        62
##  3 m3        53
##  4 m1        10
##  5 m2        40
##  6 m3        27
##  7 m1        21
##  8 m2        24
##  9 m3       120
## 10 m1         4
## 11 m2        30
## 12 m3        97
```

```
## 13 m1          1
## 14 m2         35
## 15 m3         68
```

```r
dat1$Method <- as.factor(dat1$Method)
dat1$Count <- as.numeric(dat1$Count)
str(dat1)
```

```
## tibble [15 x 2] (S3: tbl_df/tbl/data.frame)
##  $ Method: Factor w/ 3 levels "m1","m2","m3": 1 2 3 1 2 3 1 2 3 1 ...
##  $ Count : num [1:15] 31 62 53 10 40 27 21 24 120 4 ...
```

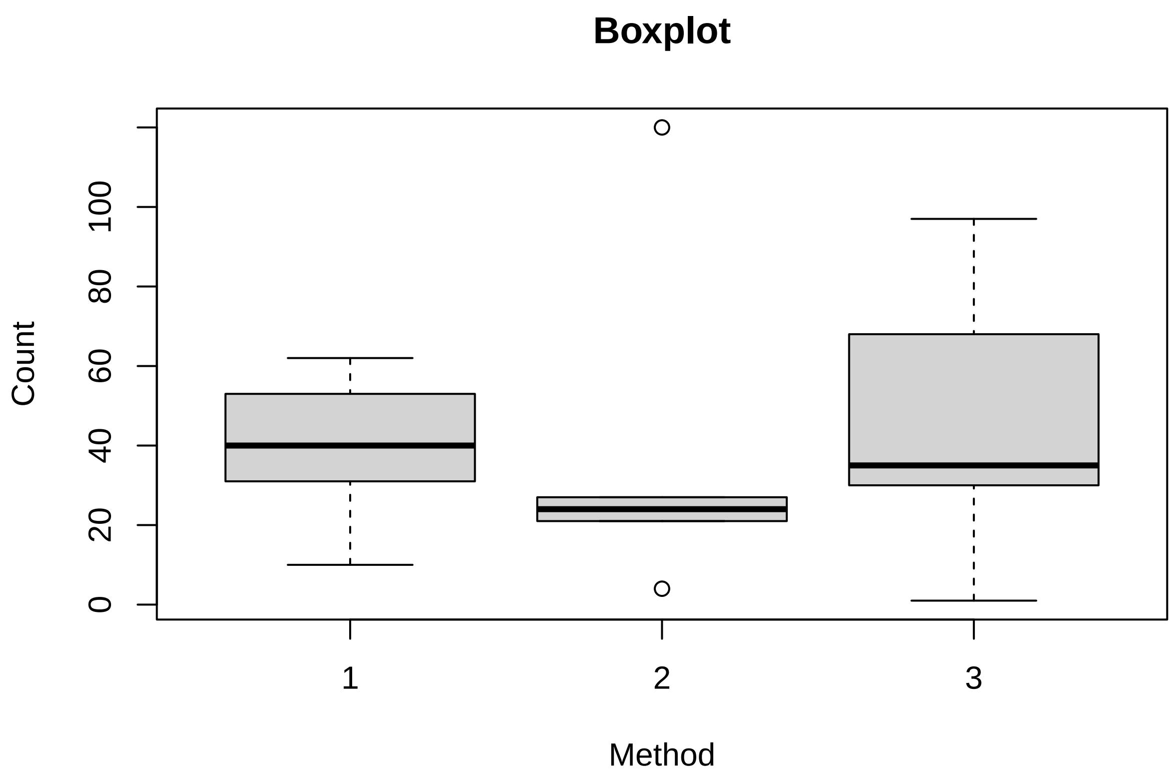
```r
model3 <- aov(dat1$Count~dat1$Method,data=dat1)
summary(model3)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## dat1$Method  2   8964    4482   7.914 0.00643 **
## Residuals   12   6796     566
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

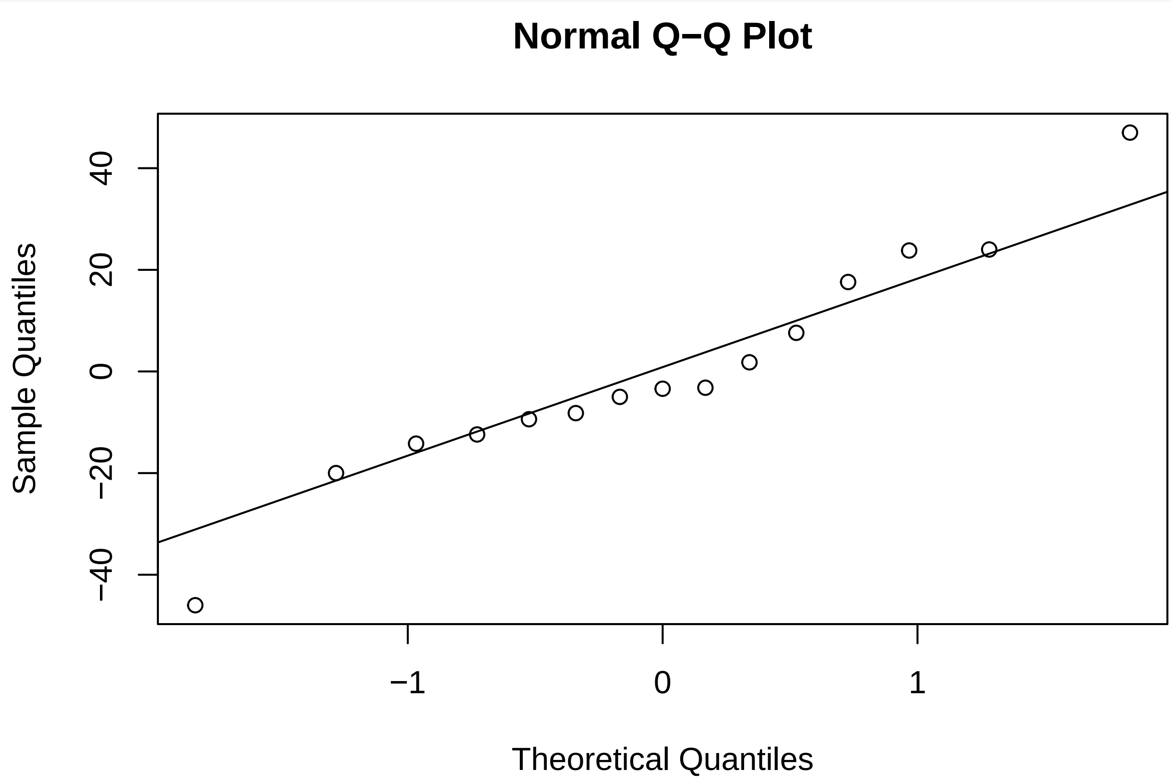
```r
# Problem 3.29.(a)
# Hypothesis
# H0: u1=u2=u3
# Ha: at least one of the u(i) differs
# u(i) = mean of the Method (i=1,2,3)
# Since P-value (0.00643) < alpha so we reject H0 at alpha = 0.05.
# No not all the methods have the same effect on mean particle count.
# at least one method is different.
# Problem 3.29.(b)
count <- c(dat1$Count)
str(count)
```

```
##  num [1:15] 31 62 53 10 40 27 21 24 120 4 ...
```

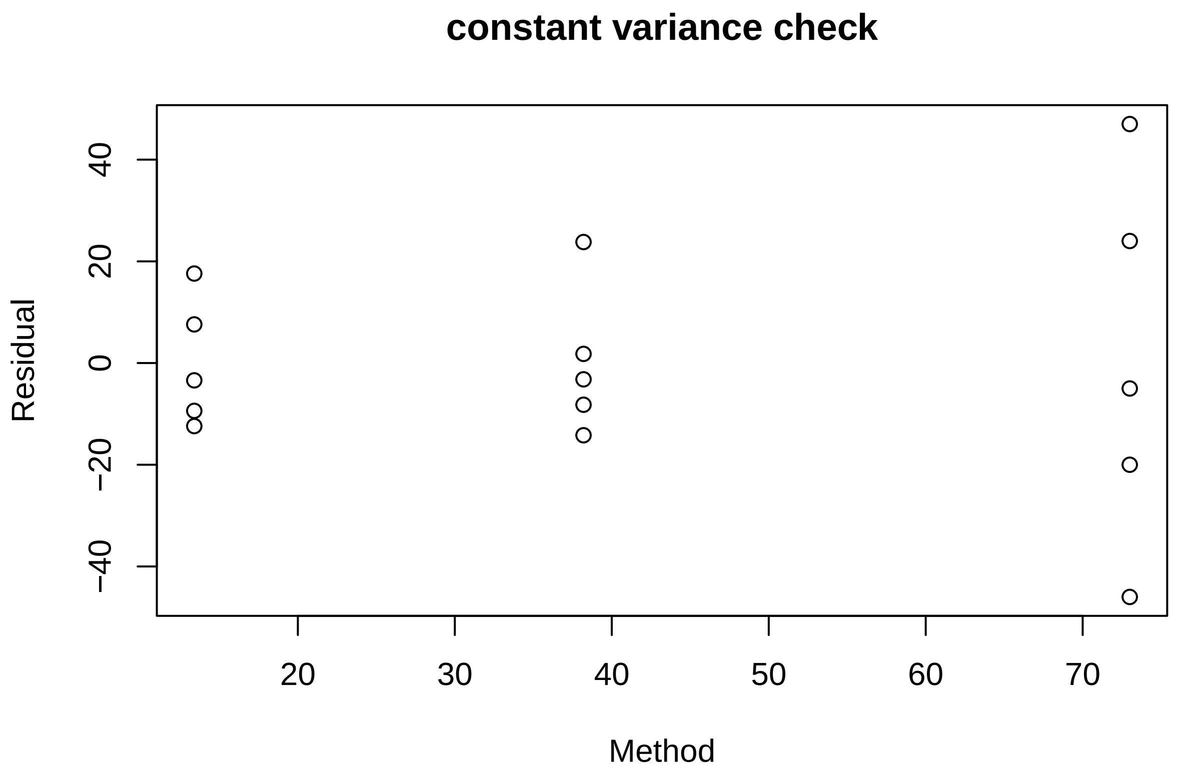
```r
x <- c(rep(1,5),rep(2,5),rep(3,5))
boxplot(count~x,xlab="Method",ylab="Count",main="Boxplot")
```




```
meanx<-c(rep(mean(m1),5),rep(mean(m2),5),rep(mean(m3),5))
Method<-c(m1,m2,m3)
res<-Method-meanx
qqnorm(res)
qqline(res)
```



```r
plot(meanx,res,xlab="Method",ylab="Residual",
     main="constant variance check")
```




```r
# The plot of residuals versus predicted response indicates the variance of the original observations
# is not constant. The normal probability plot indicates that the data is approximately
# normally distributed. So, Boxcox transformation is needed and then testing the
# hypothesis for adequacy of ANOVA.
# Problem 3.29.(c)
library(MASS)
m1 <- c(31,10,21,4,1)
m2 <- c(62,40,24,30,35)
m3 <- c(53,27,120,97,68)
dat <- data.frame(m1,m2,m3)
dat
```

```
##   m1 m2  m3
## 1 31 62  53
## 2 10 40  27
## 3 21 24 120
## 4  4 30  97
## 5  1 35  68
```

```r
library(tidyr)
library(dplyr)
dat1 <- pivot_longer(dat,c(m1,m2,m3))
dat1
```

```
## # A tibble: 15 x 2
##    name  value
##    <chr> <dbl>
##  1 m1       31
```

```
##   2 m2       62
##   3 m3       53
##   4 m1       10
##   5 m2       40
##   6 m3       27
##   7 m1       21
##   8 m2       24
##   9 m3      120
##  10 m1        4
##  11 m2       30
##  12 m3       97
##  13 m1        1
##  14 m2       35
##  15 m3       68
```

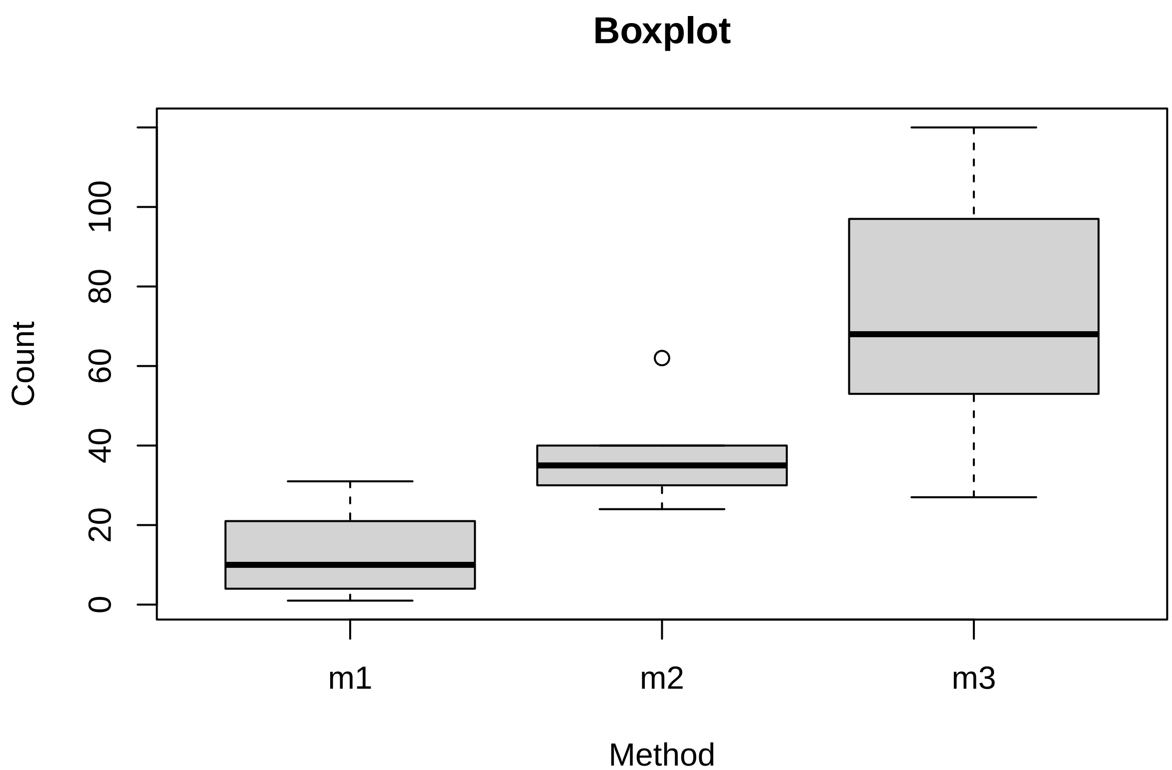
```r
colnames(dat1)<-c("Method","Count")
dat1
```

```
## # A tibble: 15 x 2
##    Method Count
##    <chr>  <dbl>
##  1 m1        31
##  2 m2        62
##  3 m3        53
##  4 m1        10
##  5 m2        40
##  6 m3        27
##  7 m1        21
##  8 m2        24
##  9 m3       120
## 10 m1         4
## 11 m2        30
## 12 m3        97
## 13 m1         1
## 14 m2        35
## 15 m3        68
```

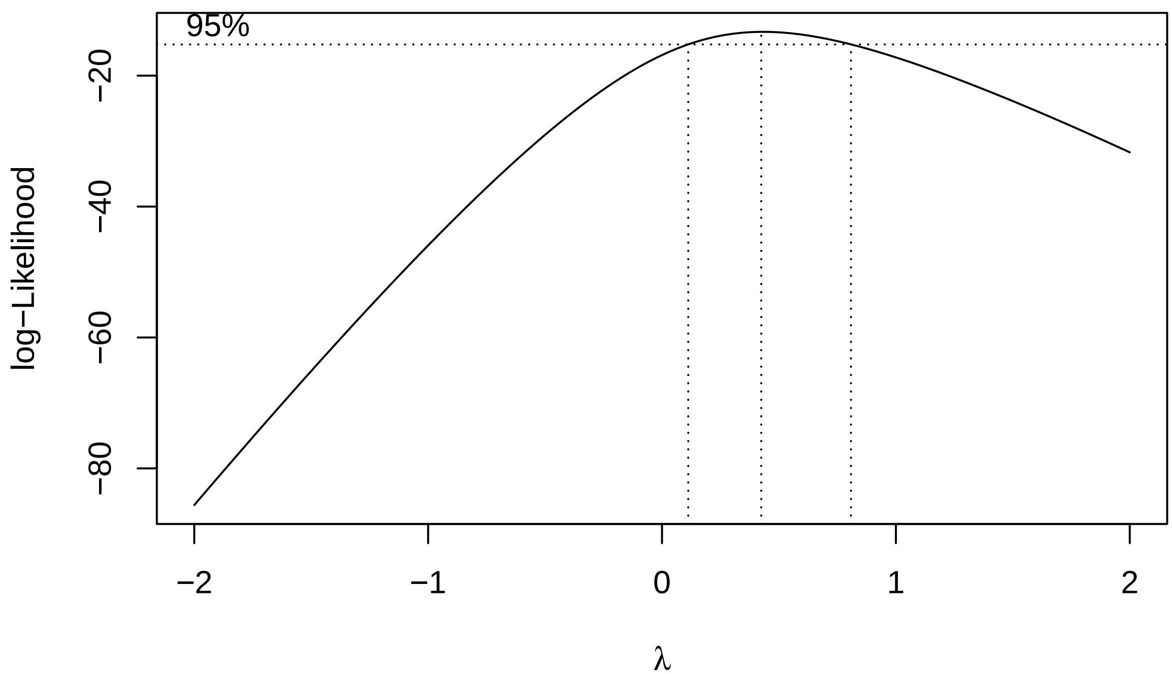
```r
dat1$Method <- as.factor(dat1$Method)
dat1$Count <- as.numeric(dat1$Count)
str(dat1)
```

```
## tibble [15 x 2] (S3: tbl_df/tbl/data.frame)
##  $ Method: Factor w/ 3 levels "m1","m2","m3": 1 2 3 1 2 3 1 2 3 1 ...
##  $ Count : num [1:15] 31 62 53 10 40 27 21 24 120 4 ...
```

```r
boxplot(dat1$Count~dat1$Method,xlab="Method",ylab="Count",main="Boxplot")
```



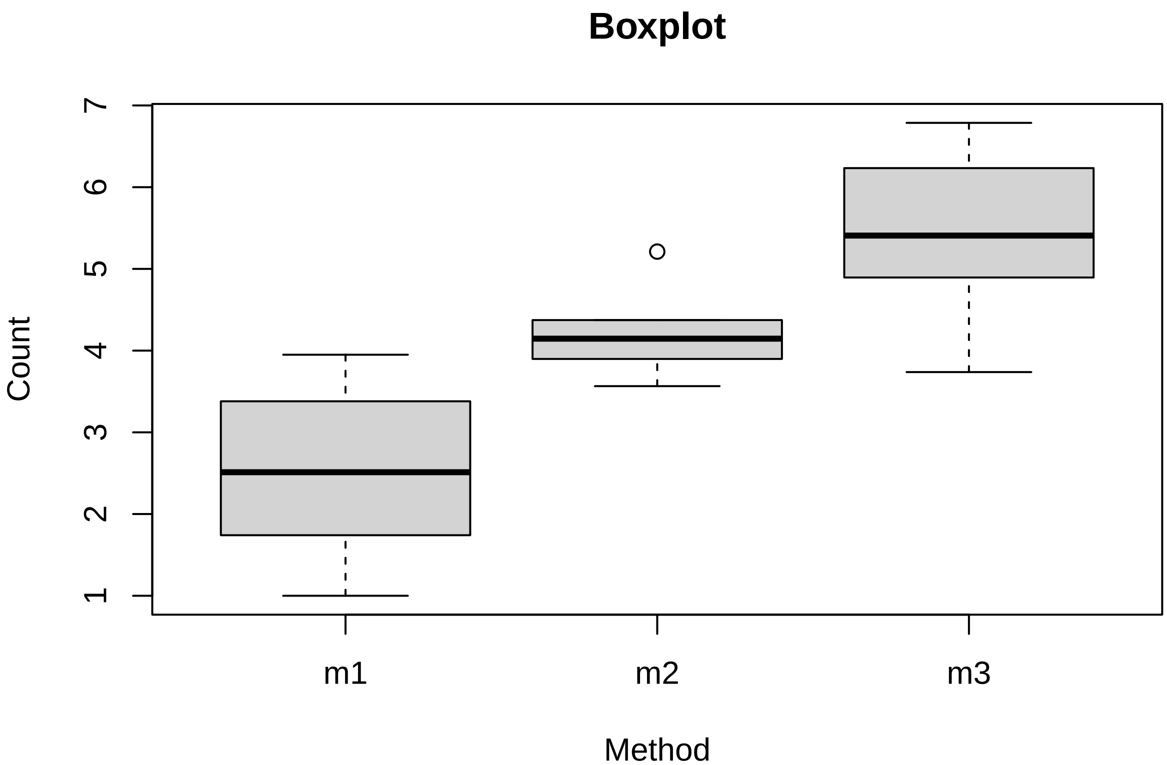

```
boxcox(dat1$Count~dat1$Method,data=dat1)
```




```
# lambda=1 is not within 95% CI so we can go ahead to do boxcox transformations
y <- (dat1$Count)^0.4
head(y)
```

```
## [1] 3.949523 5.211427 4.894523 2.511886 4.373448 3.737193
```

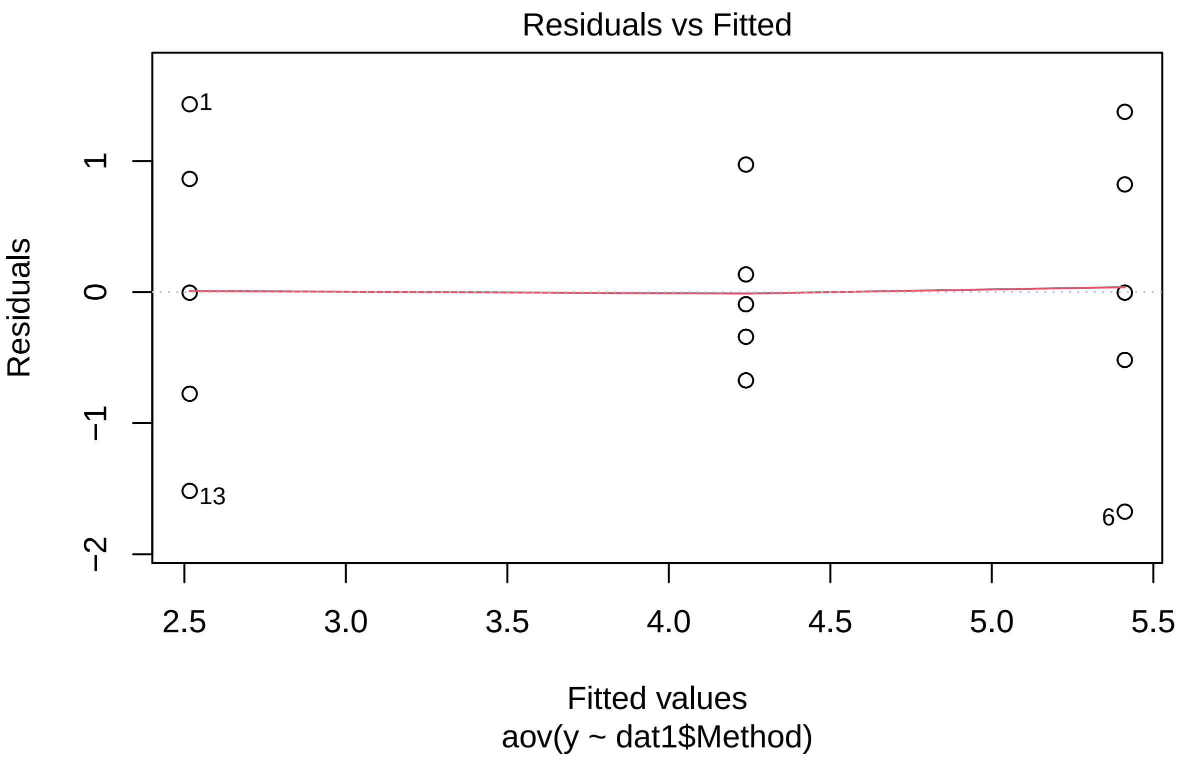
```
boxplot(y~dat1$Method,xlab="Method",ylab="Count",main="Boxplot")
```

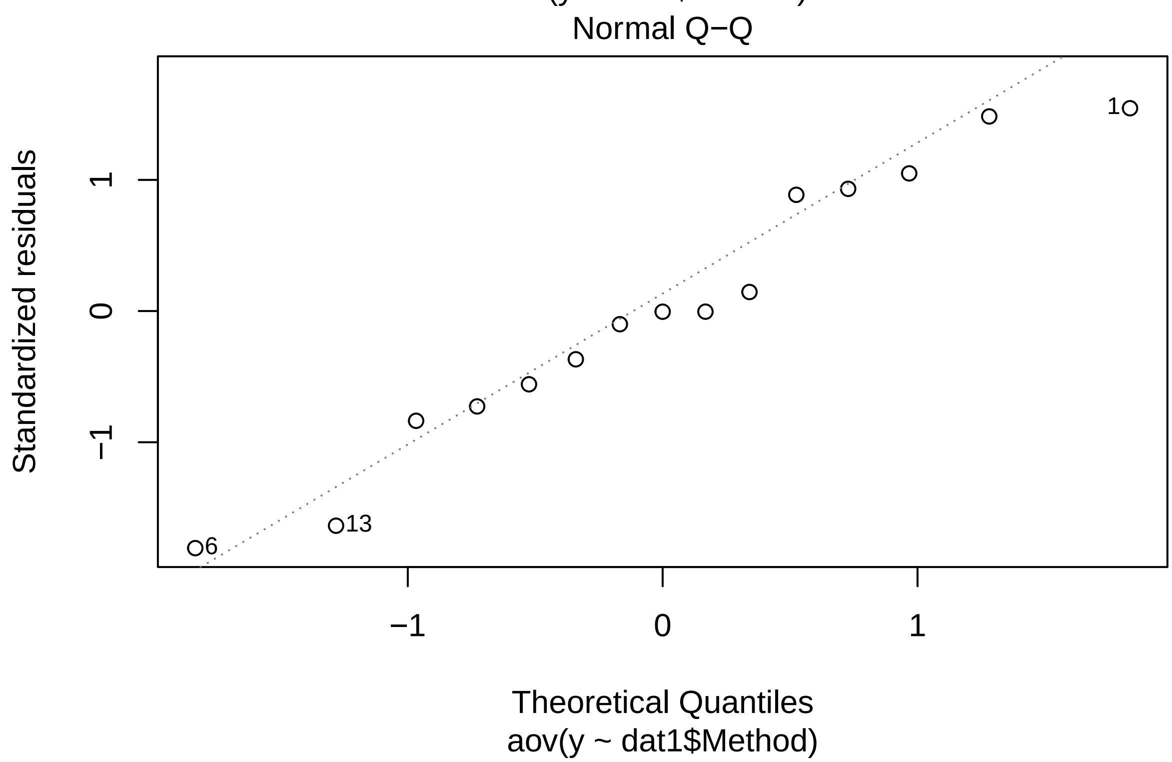


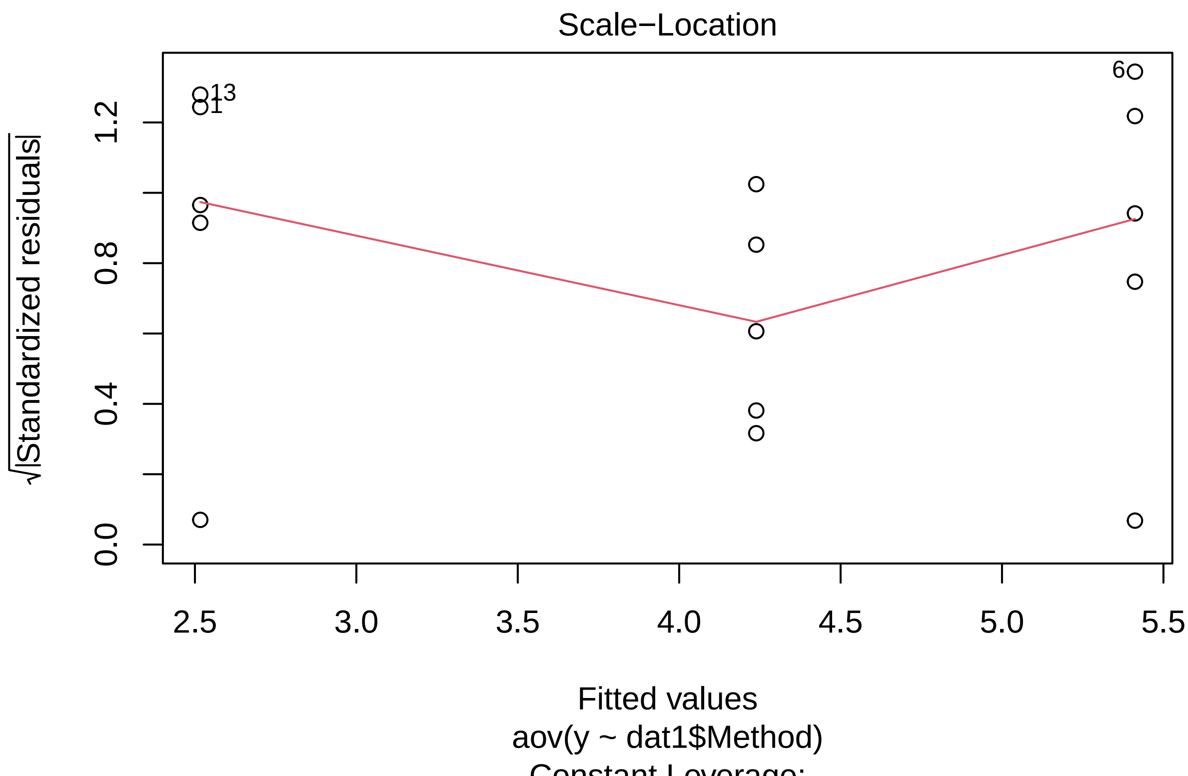

```
model4 <- aov(y~dat1$Method,data=dat1)
summary(model4)
```

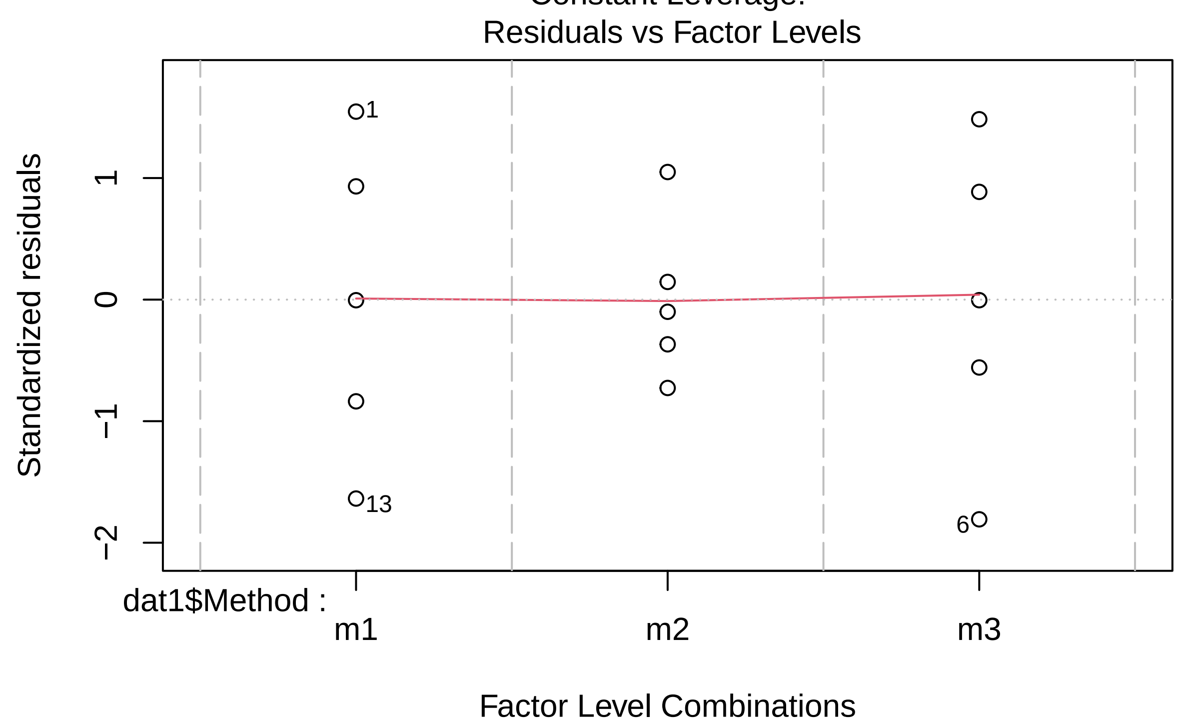
```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## dat1$Method  2  21.21  10.605   9.881 0.00291 **
## Residuals   12  12.88   1.073
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(model4)
```


Residuals vs Fitted
Residuals
1
0
−1
−2
2.5
3.0
3.5
4.0
4.5
5.0
5.5
1
13
6
Fitted values
aov(y ~ dat1$Method)


Normal Q−Q
Standardized residuals
1
0
−1
−1
0
1
1
6
13
Theoretical Quantiles
aov(y ~ dat1$Method)






```
# Since P-value (0.00291) < alpha so we reject H0 at alpha = 0.05.
# No not all the methods have the same effect on mean particle count.
# at least one method is different.
# After doing Boxcox transformation, we see that there is nothing unusual in the
# residual and NPP plots so ANOVA assumptions are adequate with normal distribution
# and constant variance.
```

```r
# Problem 3.51
ft1 <- c(17.6,18.9,16.3,17.4,20.1,21.6)
ft2 <- c(16.9,15.3,18.6,17.1,19.5,20.3)
ft3 <- c(21.4,23.6,19.4,18.5,20.5,22.3)
ft4 <- c(19.3,21.1,16.9,17.5,18.3,19.8)
dat <- data.frame(ft1,ft2,ft3,ft4)
dat
```

```
##    ft1  ft2  ft3  ft4
## 1 17.6 16.9 21.4 19.3
## 2 18.9 15.3 23.6 21.1
## 3 16.3 18.6 19.4 16.9
## 4 17.4 17.1 18.5 17.5
## 5 20.1 19.5 20.5 18.3
## 6 21.6 20.3 22.3 19.8
```

```r
library(tidyr)
library(dplyr)
dat1 <- pivot_longer(dat,c(ft1,ft2,ft3,ft4))
dat1
```

```
## # A tibble: 24 x 2
##    name  value
##    <chr> <dbl>
##  1 ft1    17.6
##  2 ft2    16.9
##  3 ft3    21.4
##  4 ft4    19.3
##  5 ft1    18.9
##  6 ft2    15.3
##  7 ft3    23.6
##  8 ft4    21.1
##  9 ft1    16.3
## 10 ft2    18.6
## # ... with 14 more rows
```

```r
colnames(dat1)<-c("Fluid Type","Life")
dat1
```

```
## # A tibble: 24 x 2
##    `Fluid Type`  Life
##    <chr>        <dbl>
##  1 ft1           17.6
##  2 ft2           16.9
##  3 ft3           21.4
##  4 ft4           19.3
##  5 ft1           18.9
##  6 ft2           15.3
##  7 ft3           23.6
##  8 ft4           21.1
##  9 ft1           16.3
## 10 ft2           18.6
## # ... with 14 more rows
```

```r
dat1$`Fluid Type` <- as.factor(dat1$`Fluid Type`)
dat1$Life <- as.numeric(dat1$Life)
```

```r
str(dat1)
```

```
## tibble [24 x 2] (S3: tbl_df/tbl/data.frame)
##  $ Fluid Type: Factor w/ 4 levels "ft1","ft2","ft3",..: 1 2 3 4 1 2 3 4 1 2 ...
##  $ Life      : num [1:24] 17.6 16.9 21.4 19.3 18.9 15.3 23.6 21.1 16.3 18.6 ...
```

```r
model <- aov(dat1$Life~dat1$`Fluid Type`,data=dat1)
summary(model)
```

```
##                     Df Sum Sq Mean Sq F value Pr(>F)
## dat1$`Fluid Type`   3  30.16   10.05   3.047 0.0525 .
## Residuals          20  65.99    3.30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
kruskal.test(dat1$Life~dat1$`Fluid Type`,data=dat1)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  dat1$Life by dat1$`Fluid Type`
## Kruskal-Wallis chi-squared = 6.2177, df = 3, p-value = 0.1015
```

```r
# Hypothesis
# H0: u1=u2=u3=u4
# Ha: at least one of the u(i) differs
# u(i) = mean of the Fluid Type (i=1,2,3,4)
# In ANOVA, we see that P-value (0.0525) > alpha so we fail to reject H0 at alpha = 0.05.
# In Kruskal-Wallis Test, we again see that P-value (0.1015) > alpha so we fail to
# reject H0 again at alpha = 0.05
# Hence, from both ANOVA and Kruskal-Wallis Test we conclude that
# there are no differences in fluid type. This agrees with the analysis of variance.
# Problem 3.52
cd1 <- c(19,20,19,30,8)
cd2 <- c(80,61,73,56,80)
cd3 <- c(47,26,25,35,50)
cd4 <- c(95,46,83,78,97)
dat <- data.frame(cd1,cd2,cd3,cd4)
dat
```

```
##   cd1 cd2 cd3 cd4
## 1  19  80  47  95
## 2  20  61  26  46
## 3  19  73  25  83
## 4  30  56  35  78
## 5   8  80  50  97
```

```r
library(tidyr)
library(dplyr)
dat1 <- pivot_longer(dat,c(cd1,cd2,cd3,cd4))
dat1
```

```
## # A tibble: 20 x 2
##    name  value
##    <chr> <dbl>
##  1 cd1      19
##  2 cd2      80
```

```
##   3 cd3      47
##   4 cd4      95
##   5 cd1      20
##   6 cd2      61
##   7 cd3      26
##   8 cd4      46
##   9 cd1      19
##  10 cd2      73
##  11 cd3      25
##  12 cd4      83
##  13 cd1      30
##  14 cd2      56
##  15 cd3      35
##  16 cd4      78
##  17 cd1       8
##  18 cd2      80
##  19 cd3      50
##  20 cd4      97
```

```
colnames(dat1)<-c("Circuit Design","Noise Observed")
dat1
```

```
## # A tibble: 20 x 2
##    `Circuit Design` `Noise Observed`
##    <chr>                       <dbl>
##  1 cd1                            19
##  2 cd2                            80
##  3 cd3                            47
##  4 cd4                            95
##  5 cd1                            20
##  6 cd2                            61
##  7 cd3                            26
##  8 cd4                            46
##  9 cd1                            19
## 10 cd2                            73
## 11 cd3                            25
## 12 cd4                            83
## 13 cd1                            30
## 14 cd2                            56
## 15 cd3                            35
## 16 cd4                            78
## 17 cd1                             8
## 18 cd2                            80
## 19 cd3                            50
## 20 cd4                            97
```

```
dat1$`Circuit Design` <- as.factor(dat1$`Circuit Design`)
dat1$`Noise Observed` <- as.numeric(dat1$`Noise Observed`)
str(dat1)
```

```
## tibble [20 x 2] (S3: tbl_df/tbl/data.frame)
##  $ Circuit Design: Factor w/ 4 levels "cd1","cd2","cd3",..: 1 2 3 4 1 2 3 4 1 2 ...
##  $ Noise Observed: num [1:20] 19 80 47 95 20 61 26 46 19 73 ...
```

```
model <- aov(dat1$`Noise Observed`~dat1$`Circuit Design`,data=dat1)
summary(model)
```

```
##                        Df Sum Sq Mean Sq F value  Pr(>F)
## dat1$`Circuit Design`  3  12042    4014   21.78 6.8e-06 ***
## Residuals             16   2949     184
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
kruskal.test(dat1$`Noise Observed`~dat1$`Circuit Design`,data=dat1)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  dat1$`Noise Observed` by dat1$`Circuit Design`
## Kruskal-Wallis chi-squared = 14.931, df = 3, p-value = 0.001877
```

```
# Hypothesis
# H0: u1=u2=u3=u4
# Ha: at least one of the u(i) differs
# u(i) = mean of the circuit design type (i=1,2,3,4)
# In ANOVA, we see that P-value (6.8e-06) < alpha so we reject H0 at alpha = 0.05.
# In Kruskal-Wallis Test, we again see that P-value (0.001877) < alpha so we
# reject H0 again at alpha = 0.05
# Hence, from both ANOVA and Kruskal-Wallis Test we conclude that the amount of noise present
# in all four circuit designs are different. This agrees with the analysis of variance.
```

Source Code

```
ft1 <- c(17.6,18.9,16.3,17.4,20.1,21.6)
ft2 <- c(16.9,15.3,18.6,17.1,19.5,20.3)
ft3 <- c(21.4,23.6,19.4,18.5,20.5,22.3)
ft4 <- c(19.3,21.1,16.9,17.5,18.3,19.8)
dat <- data.frame(ft1,ft2,ft3,ft4)
dat
library(tidyr)
library(dplyr)
dat1 <- pivot_longer(dat,c(ft1,ft2,ft3,ft4))
dat1
colnames(dat1)<-c("Fluid Type","Life")
dat1
dat1$`Fluid Type` <- as.factor(dat1$`Fluid Type`)
dat1$Life <- as.numeric(dat1$Life)
str(dat1)
model <- aov(dat1$Life~dat1$`Fluid Type`,data=dat1)
summary(model)
library(agricolae)
LSD.test(model,"dat1$`Fluid Type`",console = TRUE)
Life <- c(dat1$Life)
str(Life)
x <- c(rep(1,6),rep(2,6),rep(3,6),rep(4,6))
boxplot(Life~x,xlab="Fluid Type",ylab="Life",main="Boxplot")
meanx<-c(rep(mean(ft1),6),rep(mean(ft2),6),rep(mean(ft3),6),rep(mean(ft4),6))
Type<-c(ft1,ft2,ft3,ft4)
res<-Type-meanx
qqnorm(res)
qqline(res)
plot(meanx,res,xlab="Fluid Type",ylab="Residual",
     main="constant variance check")
```

```r
m1 <- c(110,157,194,178)
m2 <- c(1,2,4,18)
m3 <- c(880,1256,5276,4355)
m4 <- c(495,7040,5307,10050)
m5 <- c(7,5,29,2)
dat <- data.frame(m1,m2,m3,m4,m5)
dat
library(tidyr)
library(dplyr)
dat1 <- pivot_longer(dat,c(m1,m2,m3,m4,m5))
dat1
colnames(dat1)<-c("Material","Failure Time")
dat1
dat1$Material <- as.factor(dat1$Material)
dat1$`Failure Time` <- as.numeric(dat1$`Failure Time`)
str(dat1)
model2 <- aov(dat1$`Failure Time`~dat1$Material,data=dat1)
summary(model2)
ft <- c(dat1$`Failure Time`)
str(ft)
x <- c(rep(1,4),rep(2,4),rep(3,4),rep(4,4),rep(5,4))
boxplot(ft~x,xlab="Material",ylab="Failure Time",main="Boxplot")
meanx<-c(rep(mean(m1),4),rep(mean(m2),4),rep(mean(m3),4),rep(mean(m4),4),rep(mean(m5),4))
Material<-c(m1,m2,m3,m4,m5)
res<-Material-meanx
qqnorm(res)
qqline(res)
plot(meanx,res,xlab="Material",ylab="Residual",
     main="constant variance check")
library(MASS)
m1 <- c(110,157,194,178)
m2 <- c(1,2,4,18)
m3 <- c(880,1256,5276,4355)
m4 <- c(495,7040,5307,10050)
m5 <- c(7,5,29,2)
dat <- data.frame(m1,m2,m3,m4,m5)
dat
library(tidyr)
library(dplyr)
dat1 <- pivot_longer(dat,c(m1,m2,m3,m4,m5))
dat1
colnames(dat1)<-c("Material","Failure Time")
dat1
dat1$Material <- as.factor(dat1$Material)
dat1$`Failure Time` <- as.numeric(dat1$`Failure Time`)
str(dat1)
boxplot(dat1$`Failure Time`~dat1$Material,xlab="Material",ylab="Failure Time",main="Boxplot")
boxcox(dat1$`Failure Time`~dat1$Material,data=dat1)
# lambda=1 is not within 95% CI so we can go ahead to do boxcox transformations
y <- log(dat1$`Failure Time`)
head(y)
boxplot(y~dat1$Material,xlab="Material",ylab="Failure Time",main="Boxplot")
model3 <- aov(y~dat1$Material,data=dat1)
```

```r
summary(model3)
plot(model3)
m1 <- c(31,10,21,4,1)
m2 <- c(62,40,24,30,35)
m3 <- c(53,27,120,97,68)
dat <- data.frame(m1,m2,m3)
dat
library(tidyr)
library(dplyr)
dat1 <- pivot_longer(dat,c(m1,m2,m3))
dat1
colnames(dat1)<-c("Method","Count")
dat1
dat1$Method <- as.factor(dat1$Method)
dat1$Count <- as.numeric(dat1$Count)
str(dat1)
model3 <- aov(dat1$Count~dat1$Method,data=dat1)
summary(model3)
count <- c(dat1$Count)
str(count)
x <- c(rep(1,5),rep(2,5),rep(3,5))
boxplot(count~x,xlab="Method",ylab="Count",main="Boxplot")
meanx<-c(rep(mean(m1),5),rep(mean(m2),5),rep(mean(m3),5))
Method<-c(m1,m2,m3)
res<-Method-meanx
qqnorm(res)
qqline(res)
plot(meanx,res,xlab="Method",ylab="Residual",
     main="constant variance check")
library(MASS)
m1 <- c(31,10,21,4,1)
m2 <- c(62,40,24,30,35)
m3 <- c(53,27,120,97,68)
dat <- data.frame(m1,m2,m3)
dat
library(tidyr)
library(dplyr)
dat1 <- pivot_longer(dat,c(m1,m2,m3))
dat1
colnames(dat1)<-c("Method","Count")
dat1
dat1$Method <- as.factor(dat1$Method)
dat1$Count <- as.numeric(dat1$Count)
str(dat1)
boxplot(dat1$Count~dat1$Method,xlab="Method",ylab="Count",main="Boxplot")
boxcox(dat1$Count~dat1$Method,data=dat1)
# lambda=1 is not within 95% CI so we can go ahead to do boxcox transformations
y <- (dat1$Count)^0.4
head(y)
boxplot(y~dat1$Method,xlab="Method",ylab="Count",main="Boxplot")
model4 <- aov(y~dat1$Method,data=dat1)
summary(model4)
plot(model4)
```

```r
ft1 <- c(17.6,18.9,16.3,17.4,20.1,21.6)
ft2 <- c(16.9,15.3,18.6,17.1,19.5,20.3)
ft3 <- c(21.4,23.6,19.4,18.5,20.5,22.3)
ft4 <- c(19.3,21.1,16.9,17.5,18.3,19.8)
dat <- data.frame(ft1,ft2,ft3,ft4)
dat
library(tidyr)
library(dplyr)
dat1 <- pivot_longer(dat,c(ft1,ft2,ft3,ft4))
dat1
colnames(dat1)<-c("Fluid Type","Life")
dat1
dat1$`Fluid Type` <- as.factor(dat1$`Fluid Type`)
dat1$Life <- as.numeric(dat1$Life)
str(dat1)
model <- aov(dat1$Life~dat1$`Fluid Type`,data=dat1)
summary(model)
kruskal.test(dat1$Life~dat1$`Fluid Type`,data=dat1)
cd1 <- c(19,20,19,30,8)
cd2 <- c(80,61,73,56,80)
cd3 <- c(47,26,25,35,50)
cd4 <- c(95,46,83,78,97)
dat <- data.frame(cd1,cd2,cd3,cd4)
dat
library(tidyr)
library(dplyr)
dat1 <- pivot_longer(dat,c(cd1,cd2,cd3,cd4))
dat1
colnames(dat1)<-c("Circuit Design","Noise Observed")
dat1
dat1$`Circuit Design` <- as.factor(dat1$`Circuit Design`)
dat1$`Noise Observed` <- as.numeric(dat1$`Noise Observed`)
str(dat1)
model <- aov(dat1$`Noise Observed`~dat1$`Circuit Design`,data=dat1)
summary(model)
kruskal.test(dat1$`Noise Observed`~dat1$`Circuit Design`,data=dat1)
```