assg-3

Pradeep Paladugula

5/19/2020

```
Data:
Age: Categorical character variable (60-69, 70-79, >80 )
Condition: categorical character variable (task1, task2, task3)
Performance score: numerical variable
```

```r
library('xlsx')
library('ggplot2')
library(gplots)
```

```
##
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':
##
##     lowess
```

```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
```

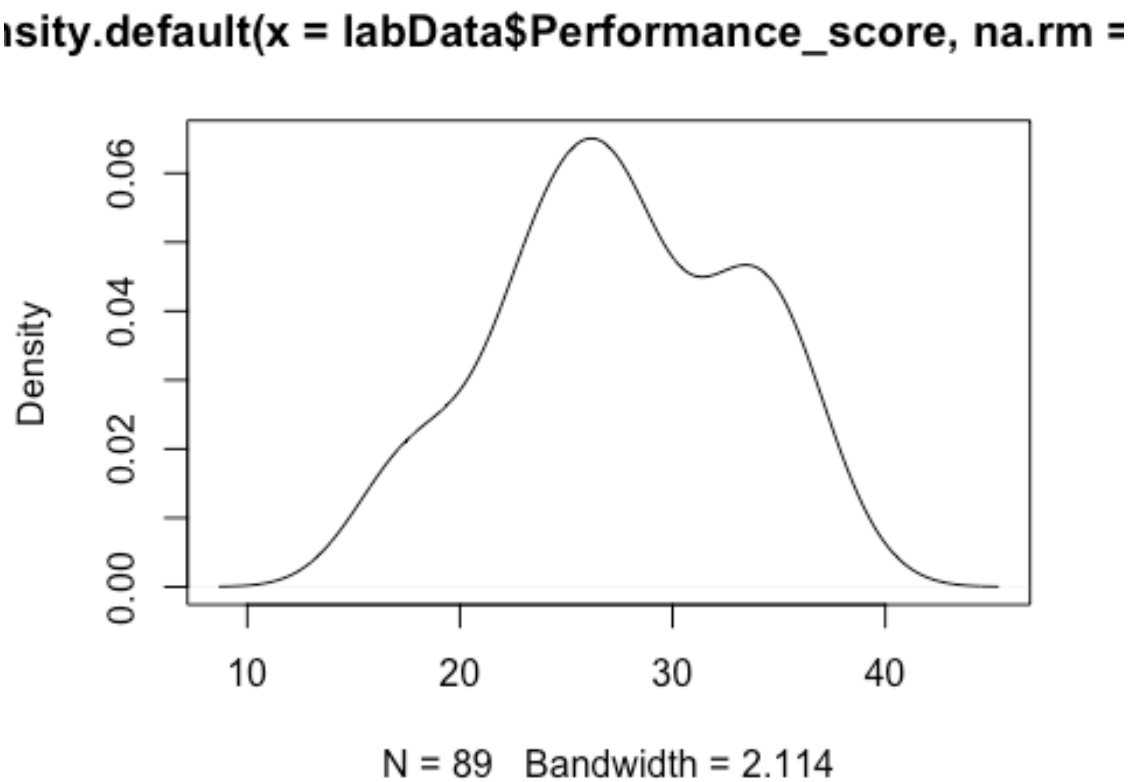
```r
#labData = read.xlsx2('lab3.xlsx', 1)
labData = read.csv('Lab3.csv')
```

1. What assumption must we test to include a variable as a blocking factor?

Solution: - H0: Their is no difference in the performance inbetween age group in any condition - H1: thier is significant difference in the performacne between age group under three different condiontions - performnace density plots of three different age groups and also performance density plot when considered all the the groups together tp check the normality - skewness and also kurtosis to check greater extremity of deviations or outliers - higher the value of leads to greater extremity and the value also depend on p-value - Analysis of variance (ANOVA) and eqality of variance test between the age groups.

```
library('ggplot2')
perfData = labData$Performance_score
qplot(Performance_score, data=labData, geom="density", fill= Age,
alpha=I(.5),
   main="Density plot of performance of skillset between different age
groups", xlab="performance score of skillset",
   ylab="Density", xlim = c(5,50))
```


```
plot(density(x = labData$Performance_score, na.rm = TRUE))
```



2. Recognize the IV, DV, block and create a table for the following research statement.

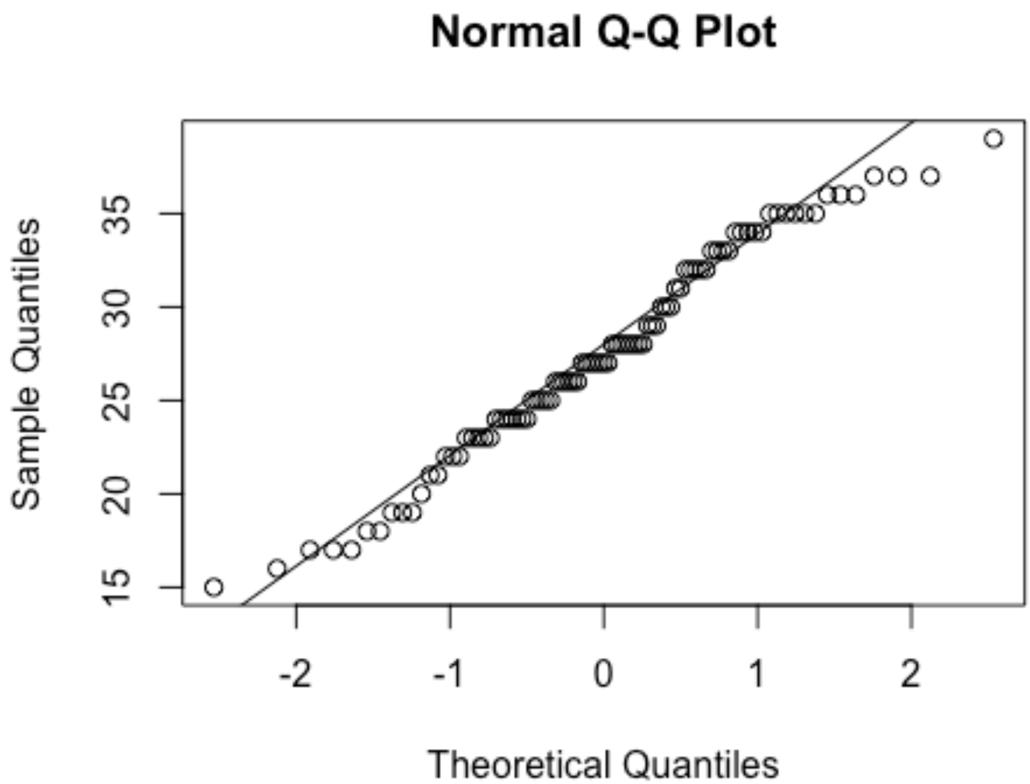
   "A company is planning to investigate the motor skills or elderly population. The company separates the target population into three age categories: 60 – 69, 70 – 79, and above 80 then randomly assign the participants in the study to one of the three task conditions. After individuals have completed the task, their performance will be compared."

```
labData = read.csv('Lab3.csv')
labData$Age <- factor(labData$Age, level = c(1, 2, 3), labels = c('60-69',
'70-79', '>80'))
labData$Condition <- factor(labData$Condition, level = c(1, 2, 3), labels =
c('task1', 'task2', 'task3'))
labDataNew <- table(labData$Condition, labData$Age)
labDataNew

##
##         60-69 70-79 >80
##   task1     9    10   9
##   task2    10    10  10
##   task3    10    11  10
```

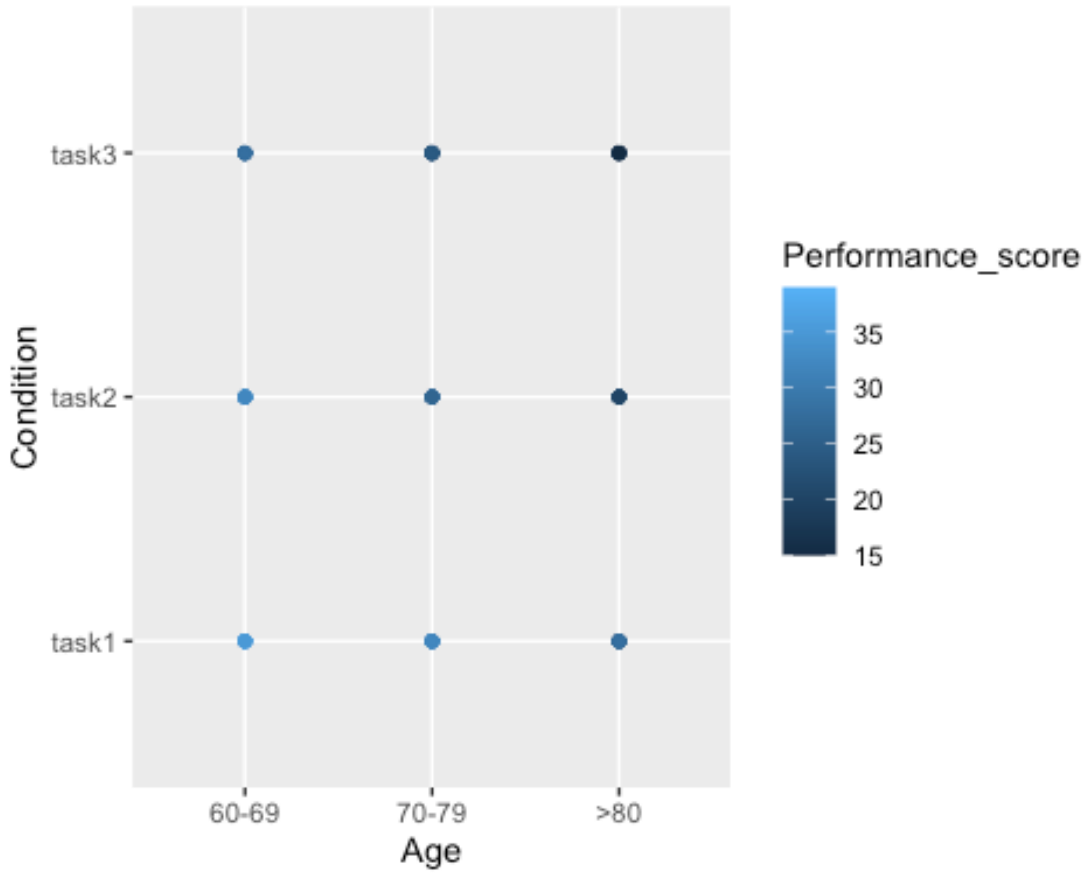
3. Use the data "Lab 3" with the research question to perform a fine report.

*age “1”:60-69, “2”: 70-79 and “3”: above 80.

```
library('ggplot2')
summary(labData)

##     Age      Performance_score  Condition
##  60-69:29   Min.   :15.00     task1:28
##  70-79:31   1st Qu.:24.00     task2:30
##  >80  :29   Median :27.00     task3:31
##             Mean   :27.52
##             3rd Qu.:32.00
##             Max.   :39.00

var(labDataNew)

##           60-69     70-79       >80
## 60-69 0.3333333 0.1666667 0.3333333
## 70-79 0.1666667 0.3333333 0.1666667
## >80   0.3333333 0.1666667 0.3333333

cor(labDataNew)

##       60-69 70-79 >80
## 60-69   1.0   0.5 1.0
## 70-79   0.5   1.0 0.5
## >80     1.0   0.5 1.0

library(moments)
agostino.test(labData$Performance_score)

##
##  D'Agostino skewness test
##
## data:  labData$Performance_score
## skew = -0.11171, z = -0.45976, p-value = 0.6457
## alternative hypothesis: data have a skewness

shapiro.test(labData$Performance_score)

##
##  Shapiro-Wilk normality test
##
## data:  labData$Performance_score
## W = 0.9755, p-value = 0.09018

qqnorm(labData$Performance_score)
qqline(labData$Performance_score)
```

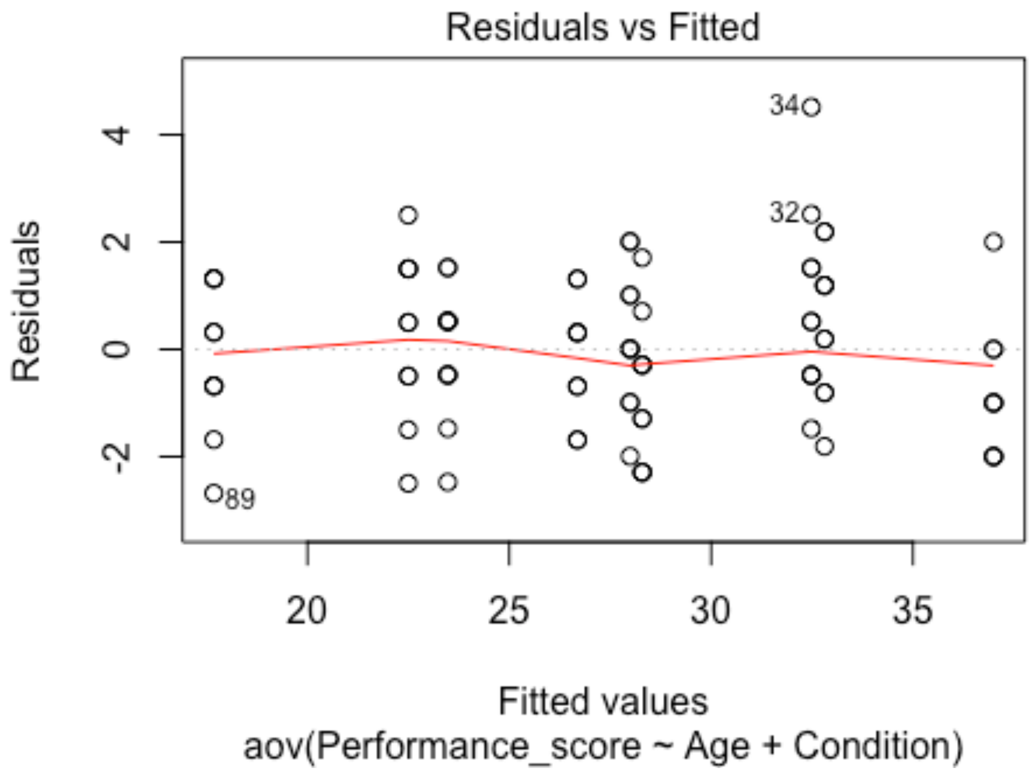


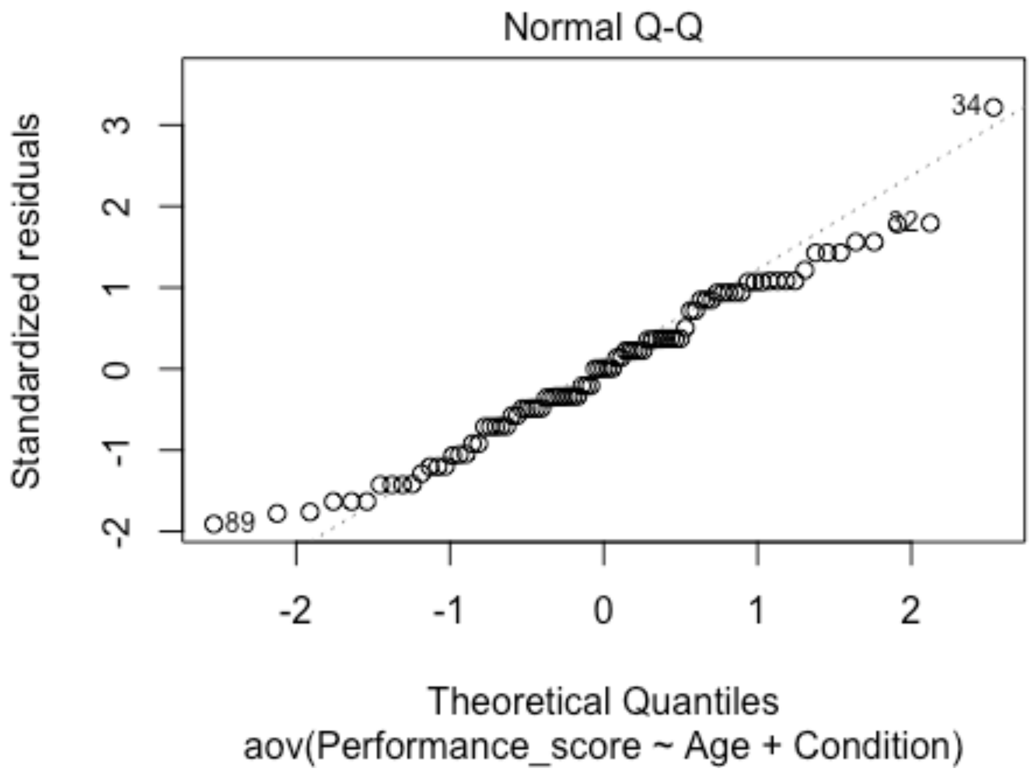

```
library('ggplot2')
ggplot(data = labData, aes(x = Age, y = Condition, color =
Performance_score)) + geom_point()
```

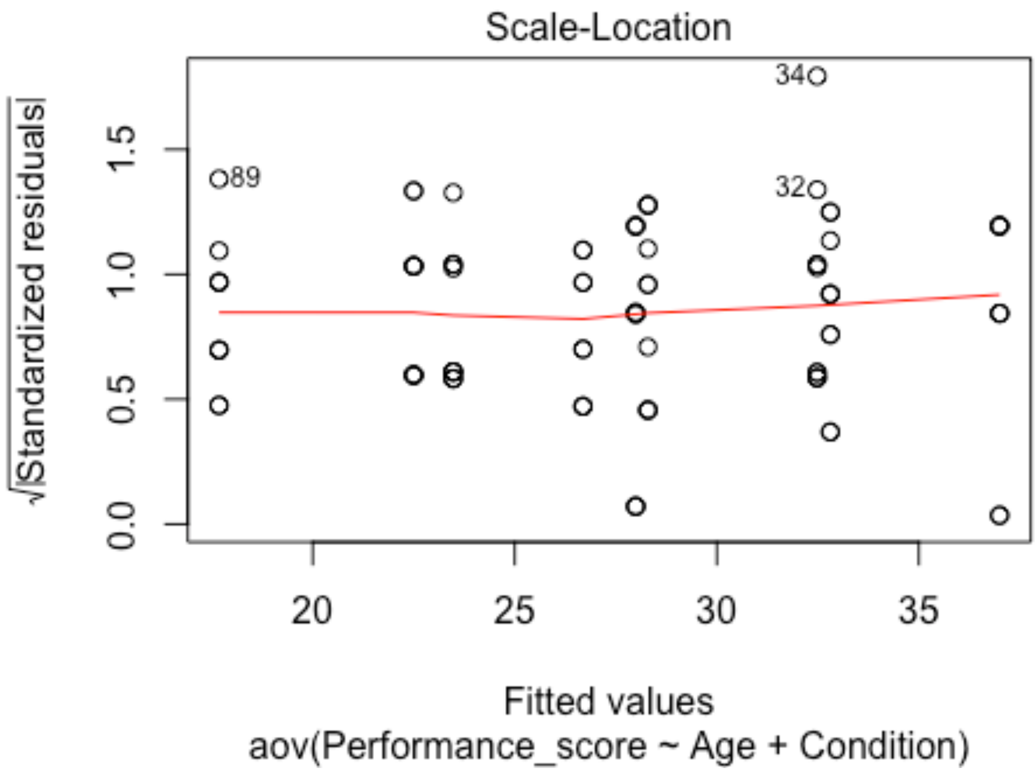


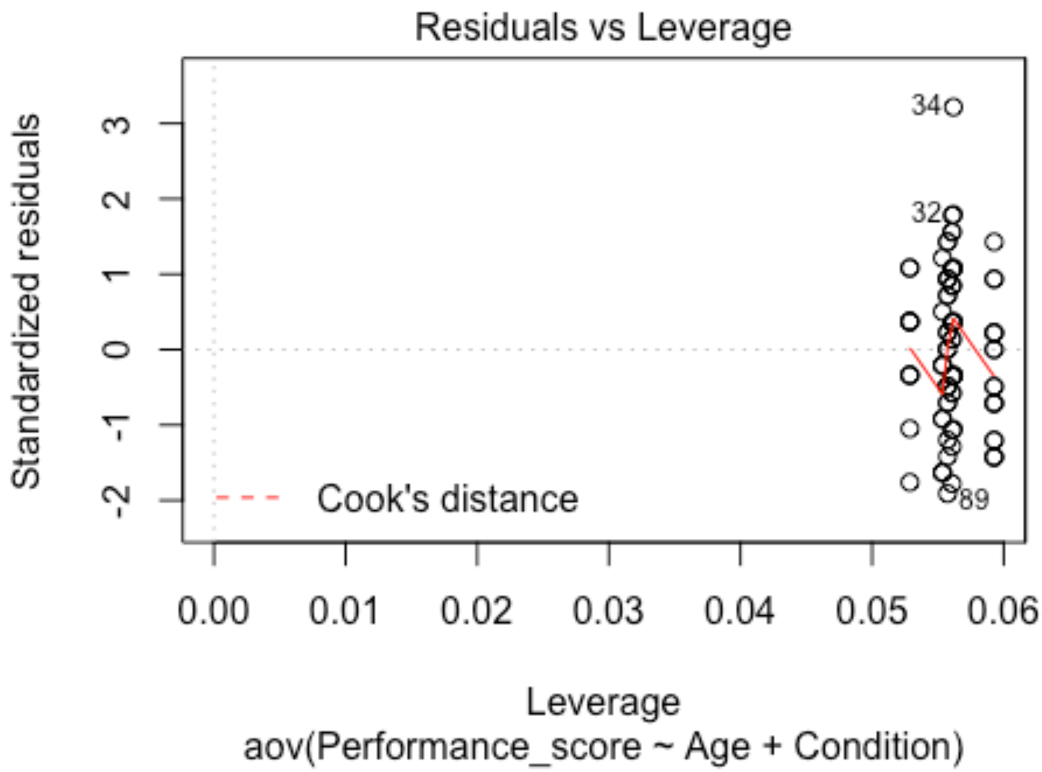

```
bartlett.test(labData$Performance_score, labData$Age)

##
##  Bartlett test of homogeneity of variances
##
## data:  labData$Performance_score and labData$Age
## Bartlett's K-squared = 1.0587, df = 2, p-value = 0.589

tapply(labData$Performance_score, labData$Age, var)

##    60-69    70-79      >80
## 12.89901 18.99570 15.83744

bartlett.test(labData$Performance_score, labData$Condition)

##
##  Bartlett test of homogeneity of variances
##
## data:  labData$Performance_score and labData$Condition
## Bartlett's K-squared = 0.14381, df = 2, p-value = 0.9306

tapply(labData$Performance_score, labData$Condition, var)

##    task1    task2    task3
## 18.36508 20.94713 20.72903
```

```
summary(aov(Performance_score ~ Age + Condition, data = labData))

##             Df Sum Sq Mean Sq F value Pr(>F)    
## Age          2 1549.7   774.9   370.8 <2e-16 ***
## Condition    2 1198.9   599.5   286.9 <2e-16 ***
## Residuals   84  175.5     2.1                   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

plot(aov(Performance_score ~ Age + Condition, data = labData))
```

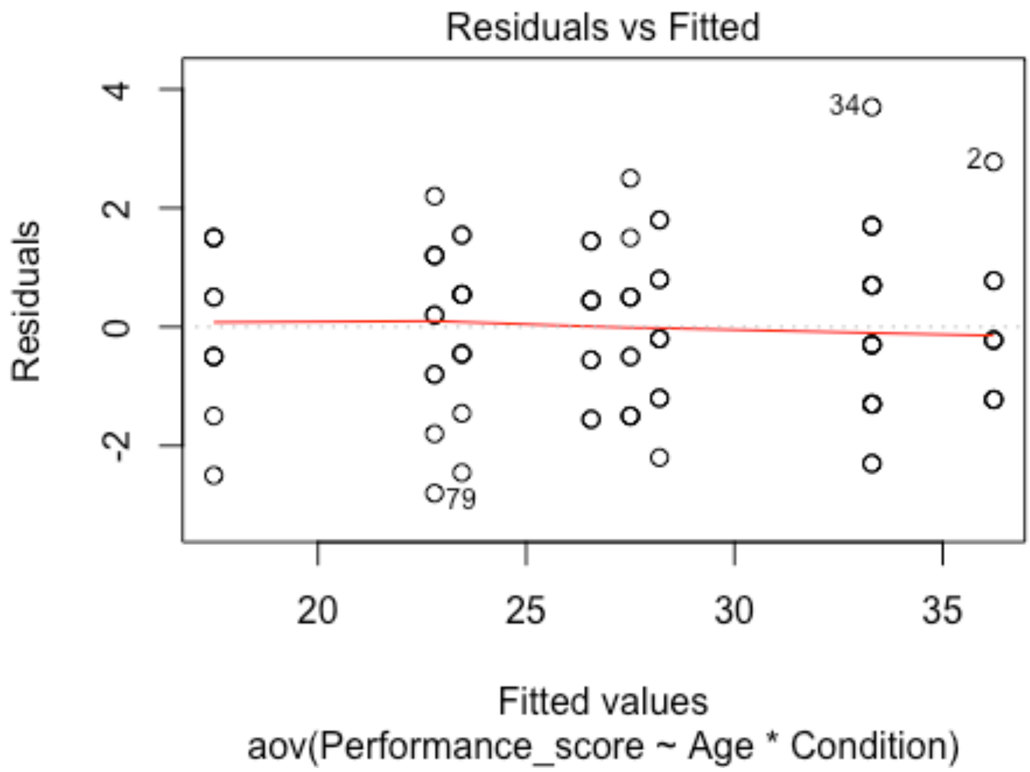

Residuals vs Fitted
Residuals
Fitted values
aov(Performance_score ~ Age + Condition)
4
2
0
-2
20
25
30
35
34
32
89


Normal Q-Q
Standardized residuals
3
2
1
0
-1
-2
34
32
89
-2
-1
0
1
2
Theoretical Quantiles
aov(Performance_score ~ Age + Condition)

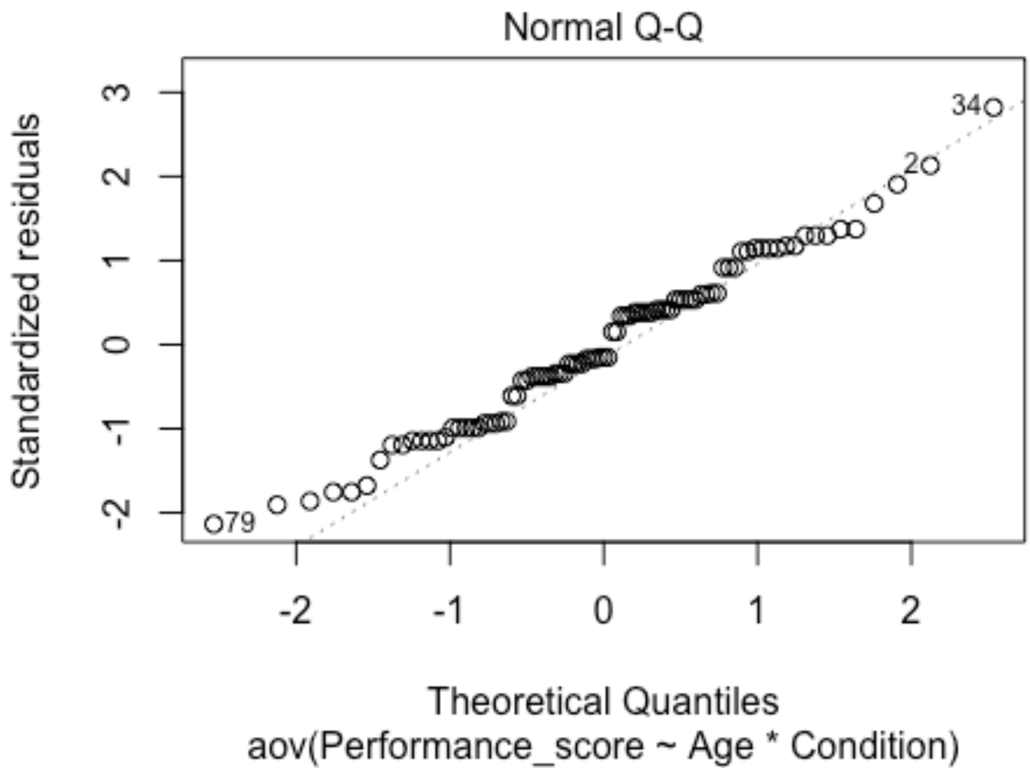

Scale-Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
20
25
30
35
89
34
32
Fitted values
aov(Performance_score ~ Age + Condition)

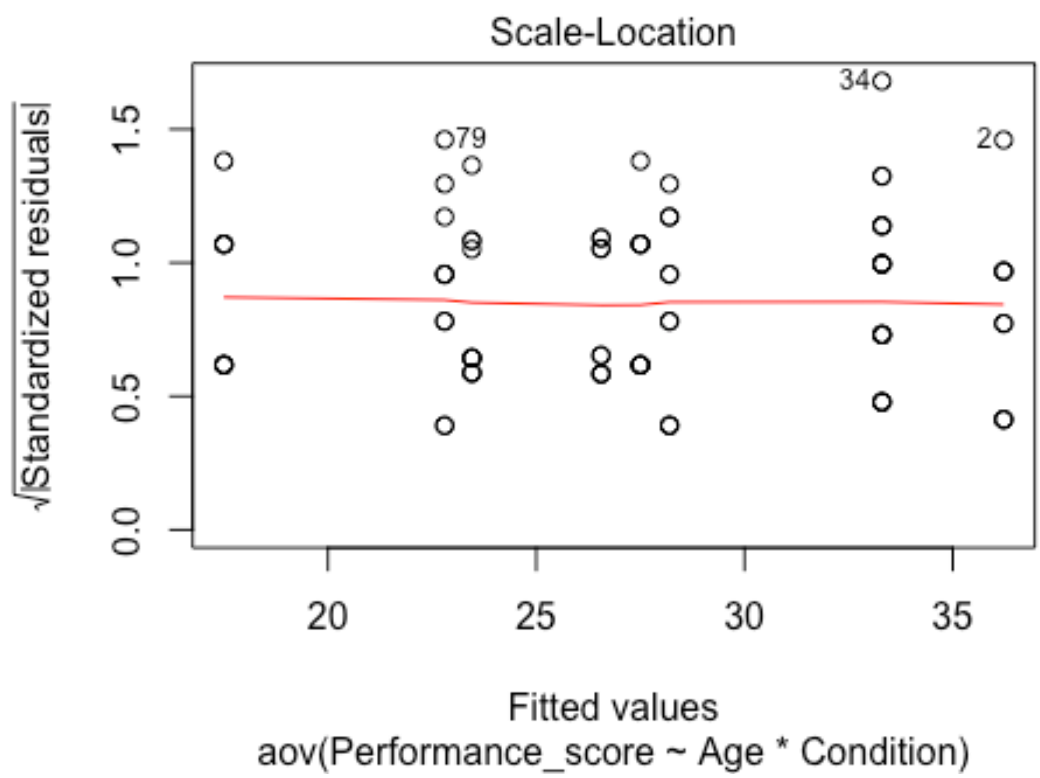


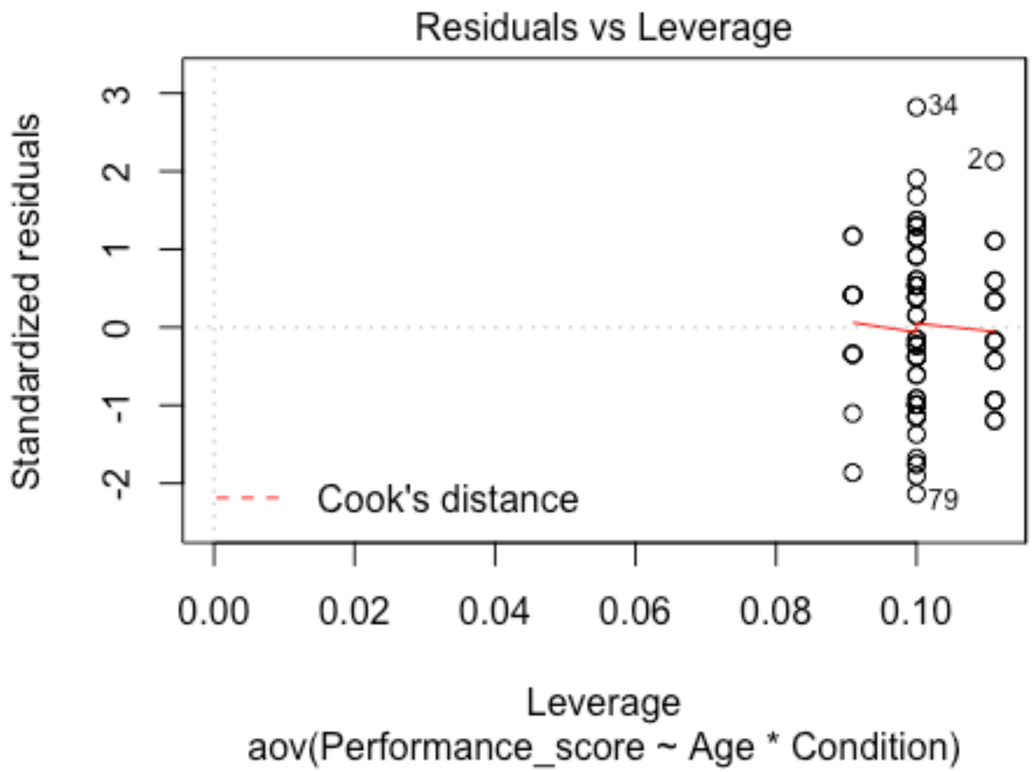

```
anovaInteraction = aov(Performance_score ~ Age * Condition, data = labData)
summary(aov(Performance_score ~ Age * Condition, data = labData))

##                Df Sum Sq Mean Sq F value Pr(>F)
## Age             2 1549.7   774.9 405.411 <2e-16 ***
## Condition       2 1198.9   599.5 313.645 <2e-16 ***
## Age:Condition   4   22.6     5.7   2.961 0.0246 *
## Residuals      80  152.9     1.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

plot(aov(Performance_score ~ Age * Condition, data = labData))
```



Residuals vs Fitted
Residuals
4
2
0
-2
34
2
79
20
25
30
35
Fitted values
aov(Performance_score ~ Age * Condition)


Normal Q-Q
Standardized residuals
3
2
1
0
-1
-2
34
2
79
-2
-1
0
1
2
Theoretical Quantiles
aov(Performance_score ~ Age * Condition)


Scale-Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
20
25
30
35
79
34
2
Fitted values
aov(Performance_score ~ Age * Condition)



Conclusion:

By looking at the desity graphs, the data looks like attaining nearly normality. in further analysis from the agostino skewness test p-value = 0.6457 which is greater than 0.05, skew = -0.11171 even though skew value shows the negative skewness but it very minimal and from the shapiro normality test where p-value = 0.09018, which clearly rejects the alernative hypothesis. While I ran Bartlett's variance equlity check variance between the means between Performance and conditions and Age separately I found the analysis rejecting the alternative hypothesis. where p-values are greater than 0.05.

Bartlett's variance equlity check between Performance Score and Age (p-value = 0.589)
Bartlett's variance equlity check between Performance Score and Condition (p-value = 0.9306)

Then from the anaalysis of variance (ANOVA), the test analysis supporting the alternative hyposthesis. where the p-value is less than the <0.05 and f-values are very high.
Age: f(2,80) = 405.411, pvalue = <2e-16
condition: f(2,80) = 313.645, pvalue = <2e-16
Age:Condition: f(4,80) = 2.961, pvalue = 0.0246

```
when graphs are consider the performance score of three different age group
dcreasing evetually from task to task and viceversa. from the Cooks's
distance plot, it is observed that the outliers are inside the approximate
value. which hleps to clarify that the normal Q_Q graph attains nearly
linearity and also outliers are very close to the normality line.
```