

HW1

```
y1<-c(3129, 3000, 2865, 2890)
y2<-c(3200, 3300, 2975, 3150)
y3<-c(2800, 2900, 2985, 2765)
y4<-c(2600, 2700, 2600, 2765)
y <- c(y1, y2, y3, y4)
n <- rep(4, 4)
group <- rep(1:4, n)
group_df <- data.frame(y, group)
group_df <- transform(group_df, group = factor(group))
```

Problem 1. (a)

```
summary(aov(y ~ group, data = group_df))#MSE= 12380
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## group          3 503865   167955    13.57 0.000367 ***
## Residuals     12 148564    12380
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

p-value 가 0.000367 이므로 유의수준 0.05 에서 Mixing Techniques 이 시멘트의 강도에 영향을 미치지 않는다는 귀무가설을 기각할 수 있다.

Problem 1. (b)

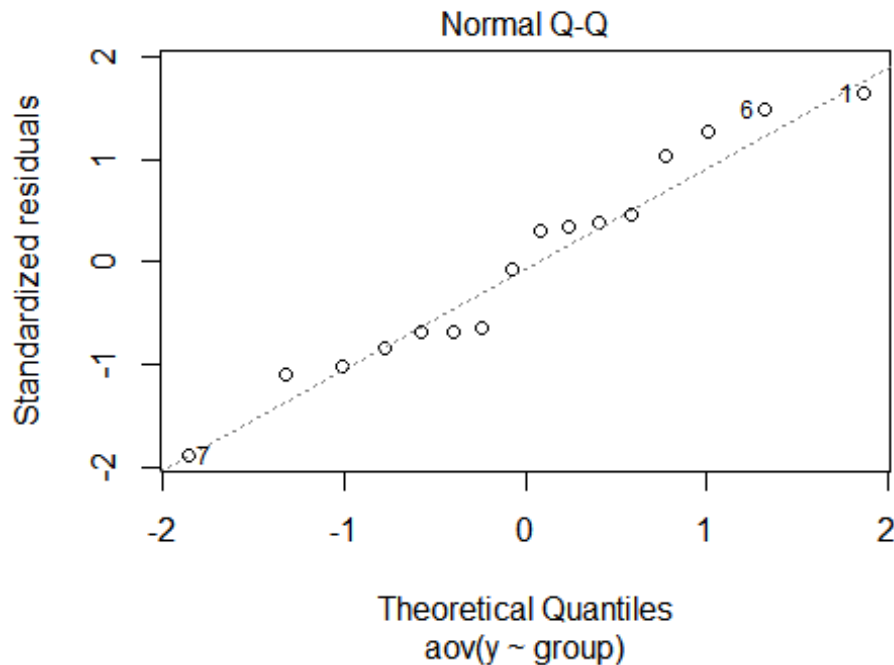
```
lsd<-qt(p=.975, df=12)*sqrt(12380*2/4)
abs(mean(y1)-mean(y2))#185.25
## [1] 185.25
abs(mean(y1)-mean(y3))#108.5
## [1] 108.5
abs(mean(y1)-mean(y4))#304.75
## [1] 304.75
abs(mean(y2)-mean(y3))#293.75
## [1] 293.75
abs(mean(y2)-mean(y4))#490
## [1] 490
abs(mean(y3)-mean(y4))#196.25
```

```
## [1] 196.25
```

유의수준 0.05 에서 LSD 값은 171.4 이고, 108.5 가 나오는 Mixing Techniques 1 과 3 을 제외하고는 나머지 Mixing Techniques 사이에서는 통계량이 LSD 값보다 크므로 Mixing Techniques 1 과 3 사이에서만 두 Mixing Techniques 간 평균이 같다는 귀무가설을 기각할 수 없다.

Problem 1. (c)

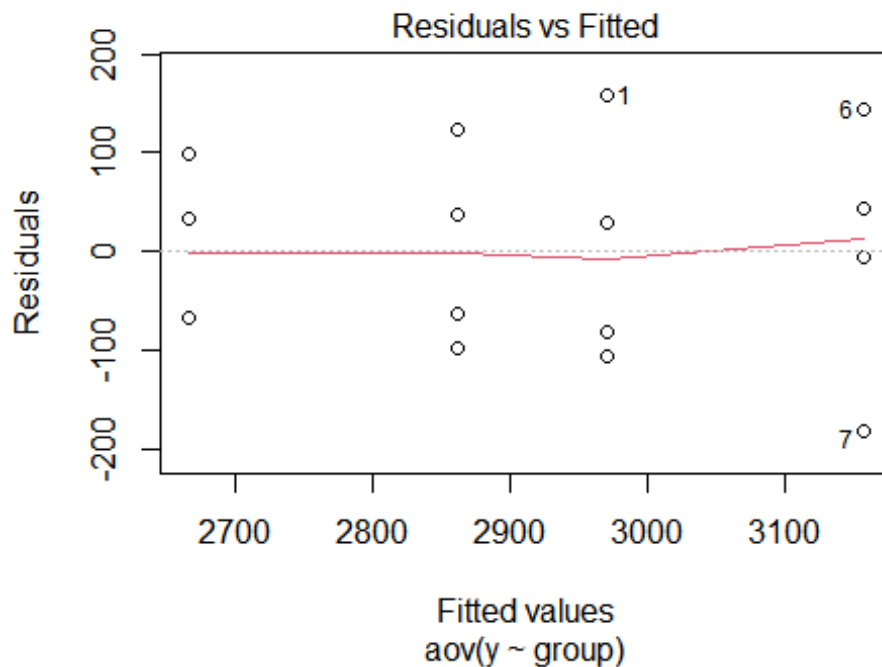
```
plot(aov(y ~ group, group_df),2)
```



Q-Q plot 을 그렸을 때 점들이 직선 주위에 위치하므로 정규성을 가진다고 말할 수 있다.

Problem 1. (d)

```
plot(aov(y ~ group, group_df),1)
```



잔차도를 그렸을 때 Fitted values 와 Residuals 간 관계가 존재하지 않음을 확인할 수 있다. 또한 빨간색 실선도 0 에 수평선 가깝게 나타나므로 등분산성을 가진다고 말할 수 있다.

Problem 2.

```
TukeyHSD(x=aov(y ~ group, data = group_df), conf.level=0.95)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = y ~ group, data = group_df)
##
## $group
##      diff      lwr      upr    p adj
## 2-1  185.25 -48.33644 418.83644 0.1398505
## 3-1 -108.50 -342.08644 125.08644 0.5345349
## 4-1 -304.75 -538.33644 -71.16356 0.0102980
## 3-2 -293.75 -527.33644 -60.16356 0.0131466
## 4-2 -490.00 -723.58644 -256.41356 0.0002228
## 4-3 -196.25 -429.83644  37.33644 0.1114776
```

Fisher LSD method 와 다르게 Tukey's test 에서는 Mixing Techniques 1 과 3 뿐만 아니라 Mixing Techniques 1 과 2, 3 과 4 에서도 p-value 가 0.05 보다 커 두 Mixing Technique 간 평균이 같다는 귀무가설을 기각할 수 없다.

```

Z1<-c(100, 96, 92, 96, 92)
Z2<-c(76, 80, 75, 84, 82)
Z3<-c(108, 100, 96, 98, 100)
Z <- c(Z1, Z2, Z3)
nZ <- rep(5, 3)
groupZ <- rep(1:3, nZ)
groupZ_df <- data.frame(Z, groupZ)
groupZ_df <- transform(groupZ_df, groupZ = factor(groupZ))

```

Problem 3. (a)

```
summary(aov(Z ~ groupZ, data = groupZ_df))
```

```

##              Df Sum Sq Mean Sq F value    Pr(>F)    
## groupZ         2 1196.1    598.1   38.34 6.14e-06 ***
## Residuals     12  187.2     15.6                
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

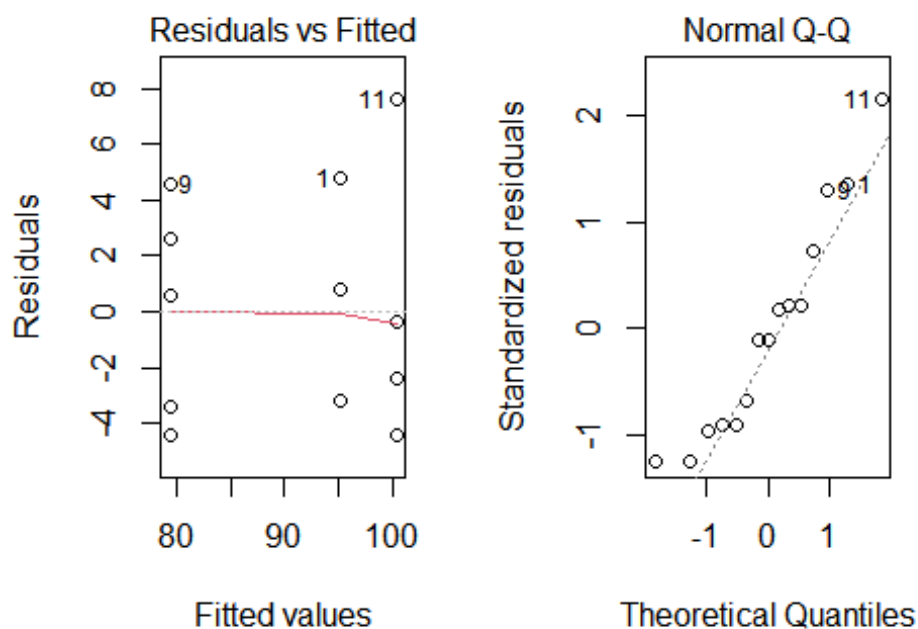
p-value 가 6.14e-06 이므로 0.05 보다 작아 배터리 브랜드들 간 평균의 차이가 없다는 귀무가설을 기각할 수 있다.

Problem 3. (b)

```

par(mfrow=c(1,2))
plot(lm(Z ~ groupZ, data = groupZ_df), which=1:2)

```



Q-Q plot 을 그렸을 때 점들이 직선 주위에 위치하므로 정규성을 가진다고 말할 수 있다.
마찬가지로 잔차도를 그렸을 때 빨간색 실선이 0 에 수평선 가깝게 나타나므로
등분산성을 가진다고 말할 수 있다.

Problem 3. (c)

```
summary(Z2)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      75.0   76.0   80.0   79.4   82.0   84.0

n<-5
xbar<-mean(Z2)
s<-var(Z2)
margin <- qt(0.975,df=n-1)*sqrt(s/n)
xbar-margin

## [1] 74.62322

xbar+margin

## [1] 84.17678

TukeyHSD(aov(Z ~ groupZ, data = groupZ_df), conf.level = 0.99)

##      Tukey multiple comparisons of means
##      99% family-wise confidence level
##
## Fit: aov(formula = Z ~ groupZ, data = groupZ_df)
##
## $groupZ
##      diff      lwr      upr      p adj
## 2-1 -15.8 -24.712897 -6.887103 0.0001044
## 3-1  5.2  -3.712897 14.112897 0.1355226
## 3-2 21.0 12.087103 29.912897 0.0000063
```

mean life of battery brand 2 의 95% 신뢰구간은 one sample t test 을 이용해 구해보면,
[74.62322, 84.17678]이고, mean different between the lives of battery brands 2 and 3 는
problem 3. (a)에서 세 브랜드 간의 평균이 같다는 귀무가설을 기각했으므로, Multiple
Comparisons approaches 중 하나인 Tukey method 를 이용해 구해보면 [12.087103,
29.912897]을 신뢰구간으로 얻을 수 있다.

Problem 4.

```
library(car)

## Warning: 패키지 'car'는 R 버전 4.1.3 에서 작성되었습니다
```

```
## 필요한 패키지를 로딩중입니다: carData

## Warning: 패키지 'carData'는 R 버전 4.1.3 에서 작성되었습니다

leveneTest(aov(Z ~ groupZ, data = groupZ_df), center=median)#등분산성 만족

## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group  2    0.07 0.9328
##      12
```

p-value 가 $0.9328 > 0.05$ 이므로 유의수준 0.05 에서 잔차도를 이용한 Problem 3 과 동일하게 등분산성이 성립한다는 귀무가설을 기각할 수 없다.