

14
22

STAT 425, Ass02

Daniel Girvitz

05/02/2022

4.12

```
responses = c(250,350,219,375,
              400,525,390,580,
              275,340,200,310)
treatment = rep(c("1","2","3"), each=4) #design
block = rep(c("NE", "NW", "SE", "SW"),3) #region

eData=data.frame(responses,treatment,block)
```

4.12.a

```
eData.aov=aov(responses~treatment+block,data=eData)
summary(eData.aov)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## treatment    2   90755    45378   50.15 0.00018 ***
## block        3   49036    16345   18.07 0.00208 **
## Residuals    6    5429     905
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since p-values are far below significance of $\alpha=0.05$, we REJECT the null hypothesis of equality of means for both designs and rejections, and conclude they both may have an influence on response rate.

4.12.b

```
design_means=tapply(responses,treatment,mean)

#Refer to p.87 of the textbook
alpha <- 0.05 #significance level
N <- 12 #total number of observations
a <- 3 #total number of treatment effects ie levels
n <- 4 #observations per treatment effect or observations per level
MSE <- 13656
```

```
LSD <- qt(p=alpha, df=N-a, lower.tail=FALSE) * sqrt(2*MSE/n) #142.7277
```

#There are 3C2=3 pairs

```
LSD_procedure = c(
  abs(design_means[1]-design_means[2]),
  abs(design_means[1]-design_means[3]),
  abs(design_means[2]-design_means[3])
)
LSD_procedure-LSD
```

```
##          1          1          2
## 23.77674 -134.22326  41.02674
```

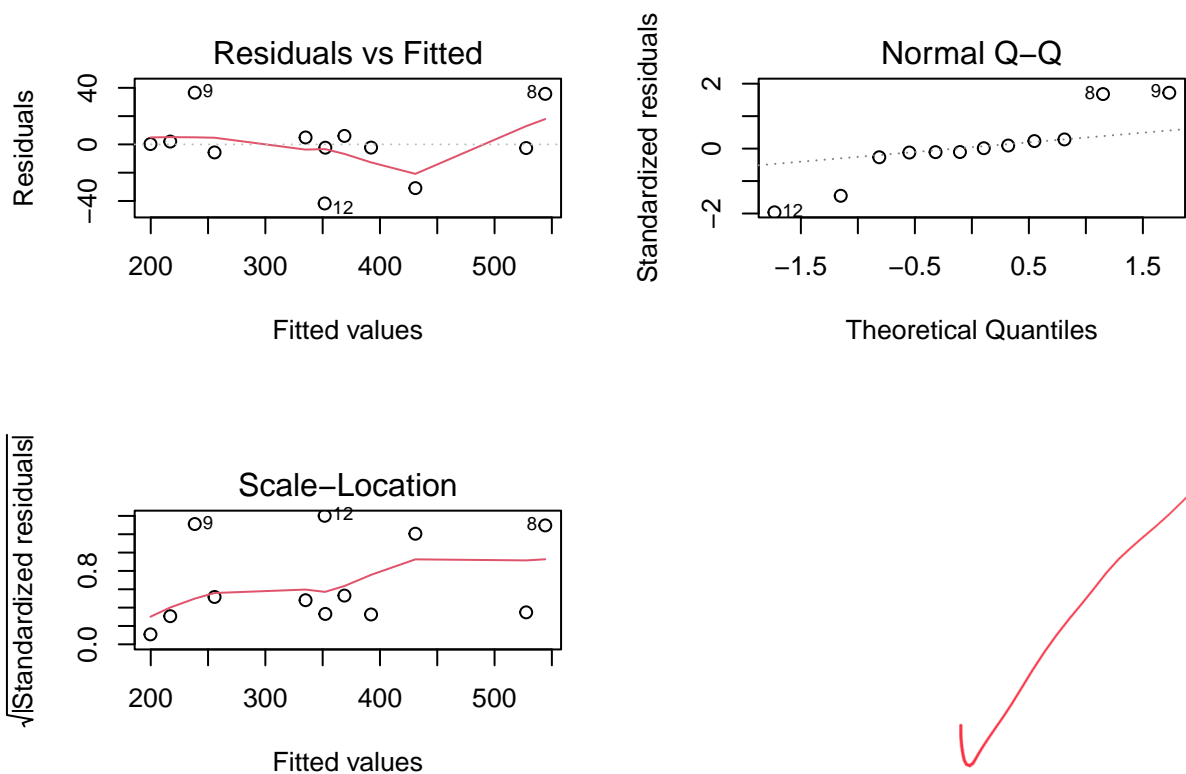
*#Any values above 0 imply the absolute value of difference in means between the
#corresponding pair of averages is greater than the Fisher LSD value.
#Thus only the means of the pair of ybar_1. to ybar_3. do not differ.*

The only significant difference is between designs 1 and 3.

4.12.c

```
par(mfrow=c(2,2))
plot(eData.aov)
```

```
## hat values (leverages) are all = 0.5
## and there are no factor predictors; no plot no. 5
```



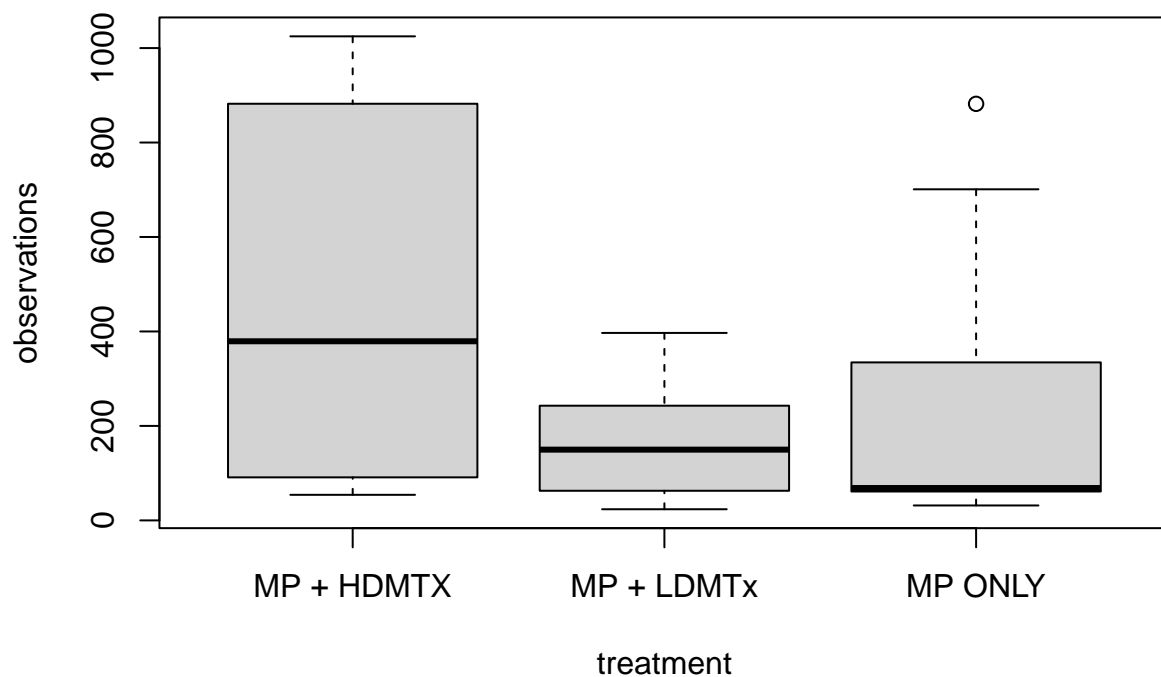
1. Residuals vs fitted: There appears to be a curved relationship, which suggests interaction between blocks and treatments.
2. Normal Q-Q: Points appear to fit diagonal well, so normality can be assumed.

4.15

```
observations = c(334.5,31.6,701,41.2,61.2,69.6,67.5,66.6,120.7,881.9,
                919.4,404.2,1024.8,54.1,62.8,671.6,882.1,354.2,321.9,91.1,
                108.4,26.1,240.8,191.1,69.7,242.8,62.7,396.9,23.6,290.4)
treatment = rep(c("MP ONLY", "MP + HDMTX", "MP + LDMTx"), each=10) #design
eData = data.frame(observations,treatment)
```

4.15.a

```
boxplot(observations~treatment,data=eData)
```



```
eData.aov=aov(observations~treatment,data=eData)
summary(eData.aov)
```

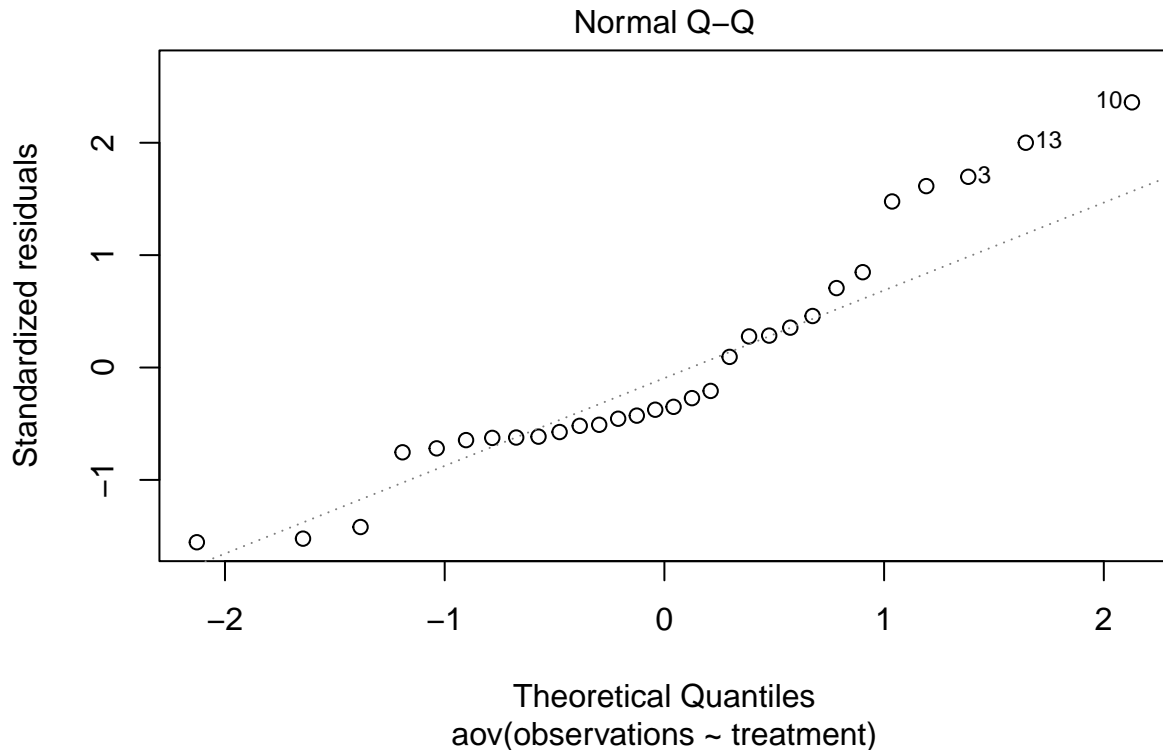
```
##           Df  Sum Sq Mean Sq F value Pr(>F)
## treatment    2   538442   269221    3.249  0.0544 .
## Residuals   27  2236974    82851
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



At an alpha level of 0.05, we do not have evidence to support the claim that the treatment means differ (FTR null hypothesis of equality of means).

4.15.b

```
par(mfrow=c(1,1))
plot(eData.aov,2)
```



Since the standardized residuals follow the diagonal quite well, we can assume normality.

4.15.c

```
eData.log.aov=aov(log(observations)~treatment,data=eData)
summary(eData.log.aov)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## treatment    2   6.30   3.148    2.571 0.0951 .
## Residuals   27  33.07   1.225
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

*→ wrong
but right
conclusion*

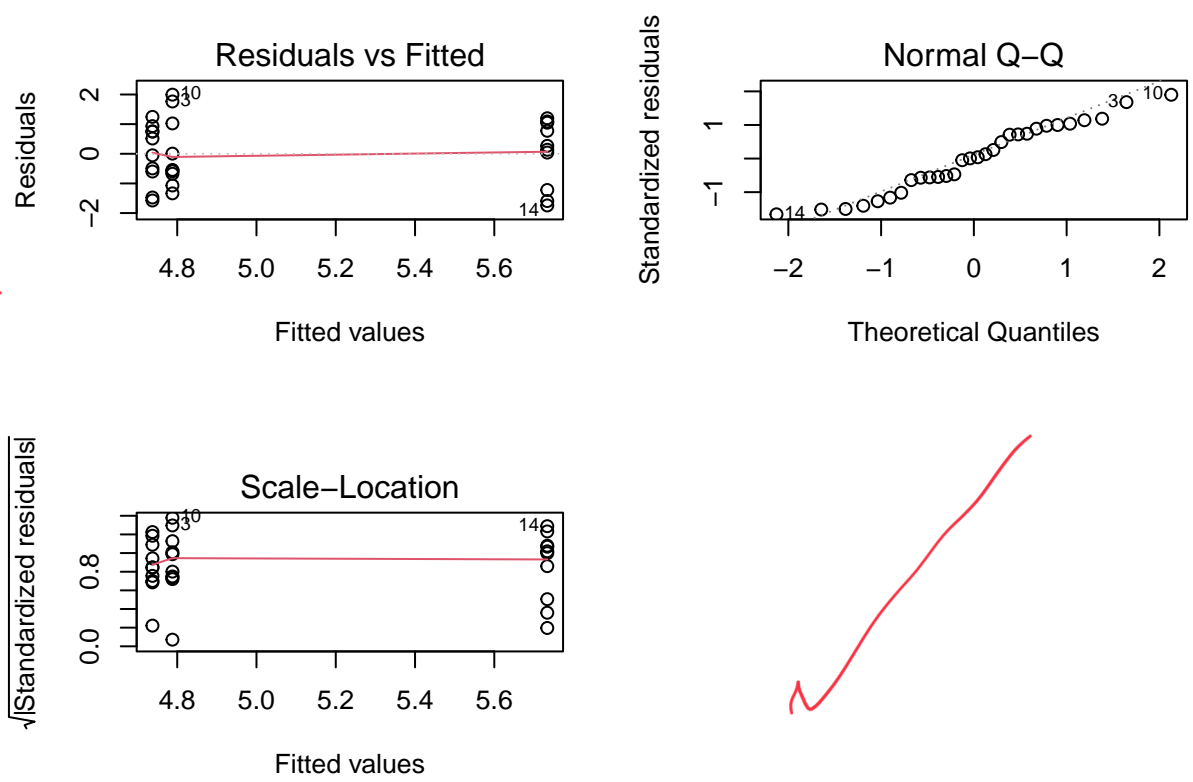
At an alpha level of 0.05, we do not have evidence to support the claim that the treatment means differ (FTR null hypothesis of equality of means).

4.15.d

```
par(mfrow=c(2,2))
plot(eData.log.aov)
```

```
## hat values (leverages) are all = 0.1
## and there are no factor predictors; no plot no. 5
```

2/4



1. Residuals vs Fitted: Residuals appear to be spread evenly along the red line, which is almost perfectly straight, thus, we can assume homogeneity of variance (homoscedasticity).
2. Normal Q-Q: Since the standardized residuals follow the diagonal quite well, we can assume normality.

4.17

```
grain.size = c(8,4,5,6,
               14,5,6,9,
               14,6,9,2,
               17,9,3,6)
treatment = rep(c("5","10","15","20"), each=4) #stirring rate
block = rep(c("1","2","3","4"),4) #furnace

eData = data.frame(grain.size,treatment,block)
```

4.17.a

```
aov.eData = aov(grain.size~treatment+block,data=eData)
summary(aov.eData)
```

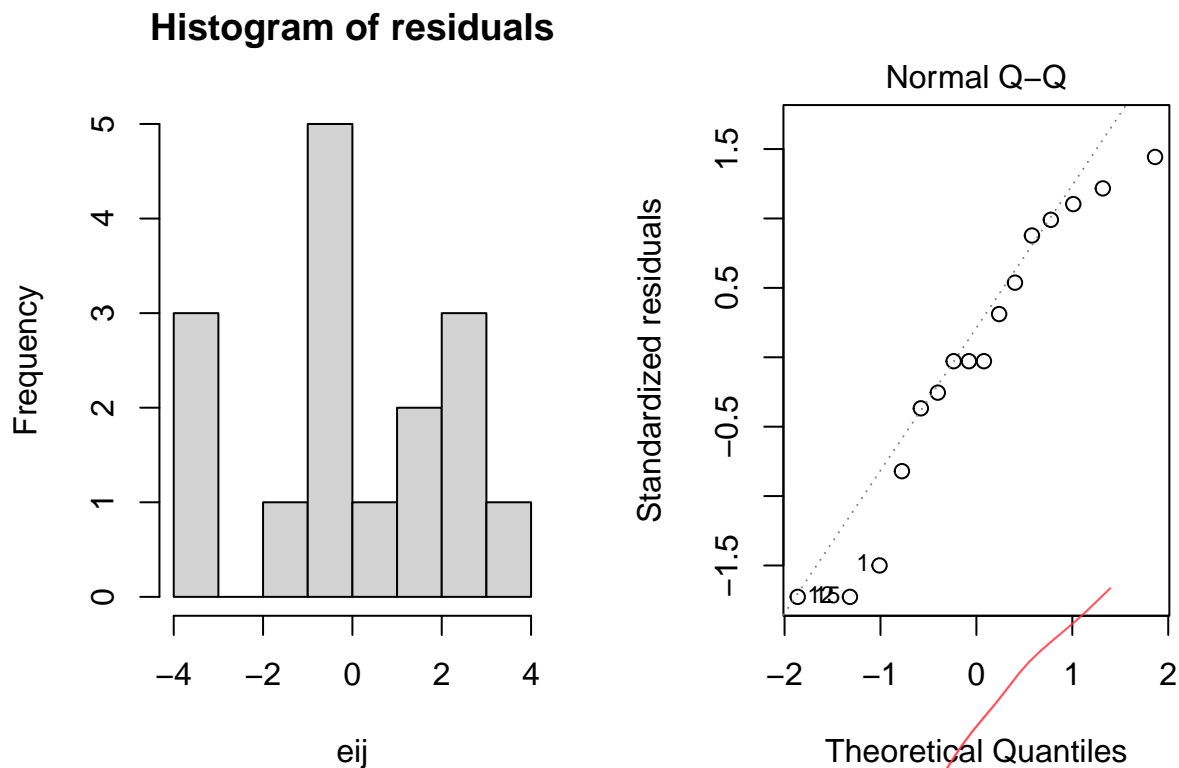
```
##           Df Sum Sq Mean Sq F value Pr(>F)
```

```
## treatment      3  22.19    7.40    0.853 0.4995
## block          3 165.19   55.06    6.348 0.0133 *
## Residuals      9   78.06    8.67
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With a p-value of 0.4995, there is strong evidence to FTR the null hypothesis of equality of all treatments, thus there is little to no evidence that stirring rates affect grain size.

4.17.b

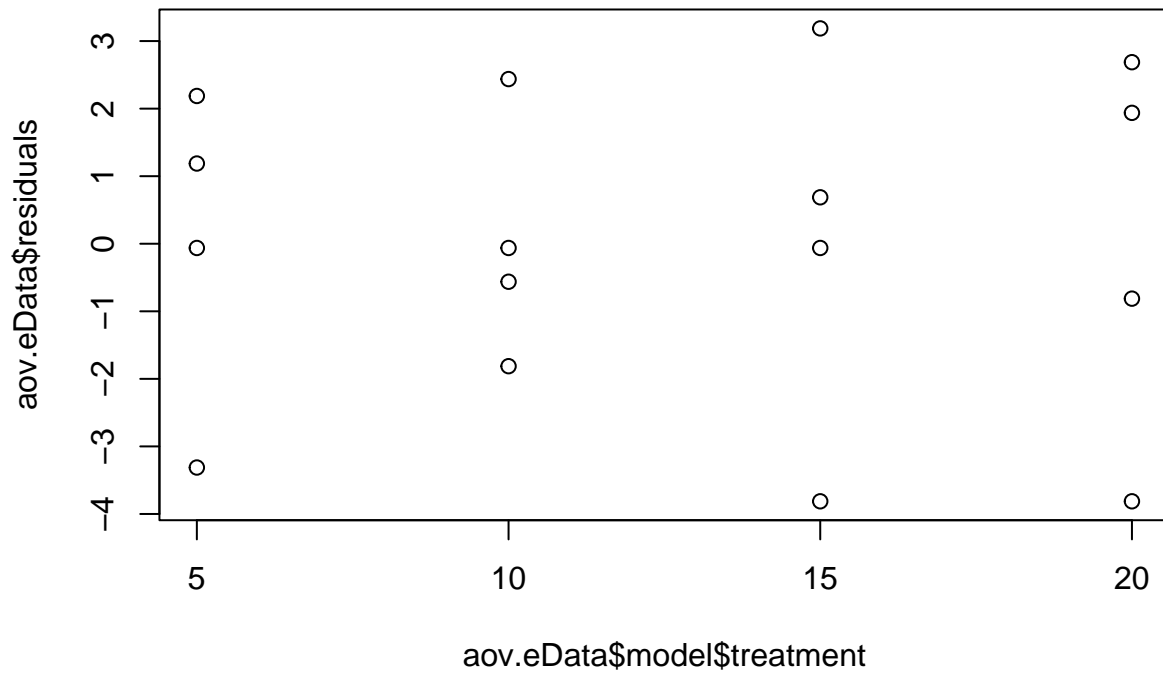
```
par(mfrow=c(1,2))
eij=residuals(aov.eData)
hist(eij,main="Histogram of residuals")
#plot(density(eij),main="Density plot of residuals",ylab="Density",xlab="Residuals")
plot(aov.eData,2)
```



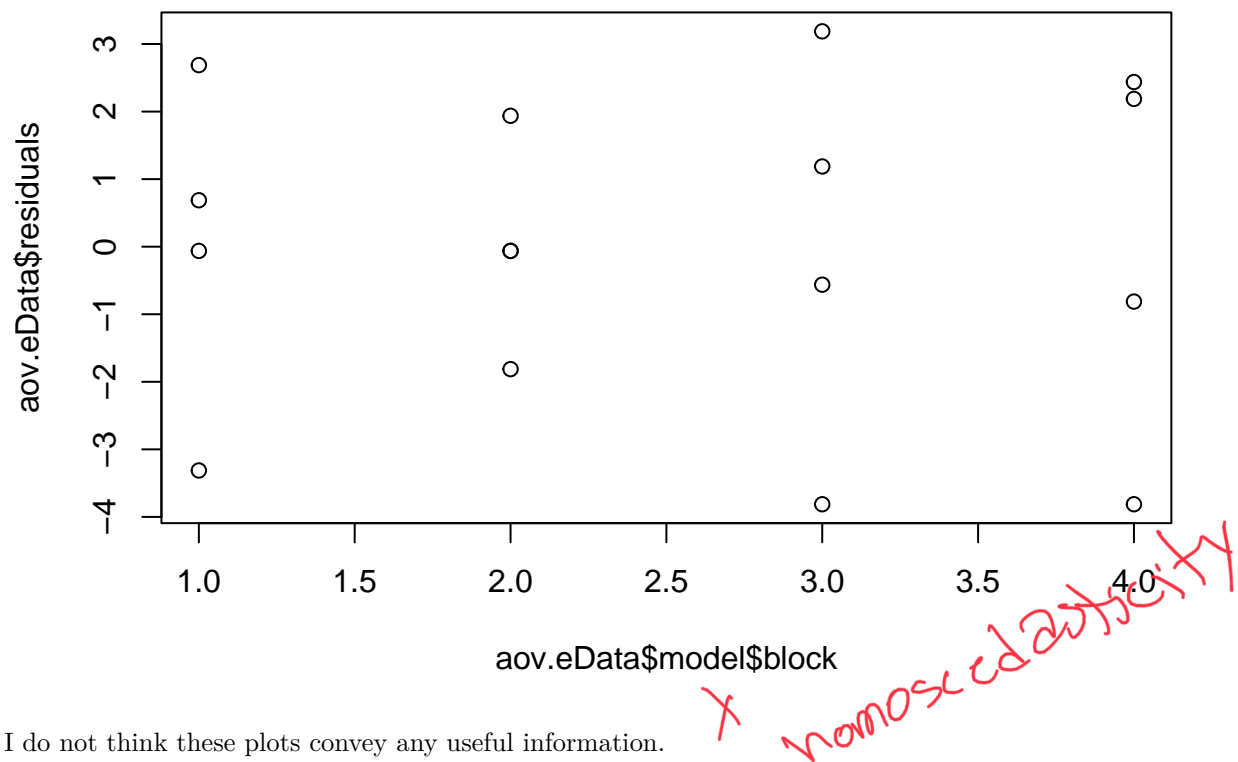
Residuals do not appear to distribute normally on a histogram, and they do not fit the diagonal on the Normal Q-Q plot well, thus normality cannot be assumed.

4.17.c

```
# Residuals vs stirring rate  
plot(aov.eData$model$treatment,aov.eData$residuals)
```



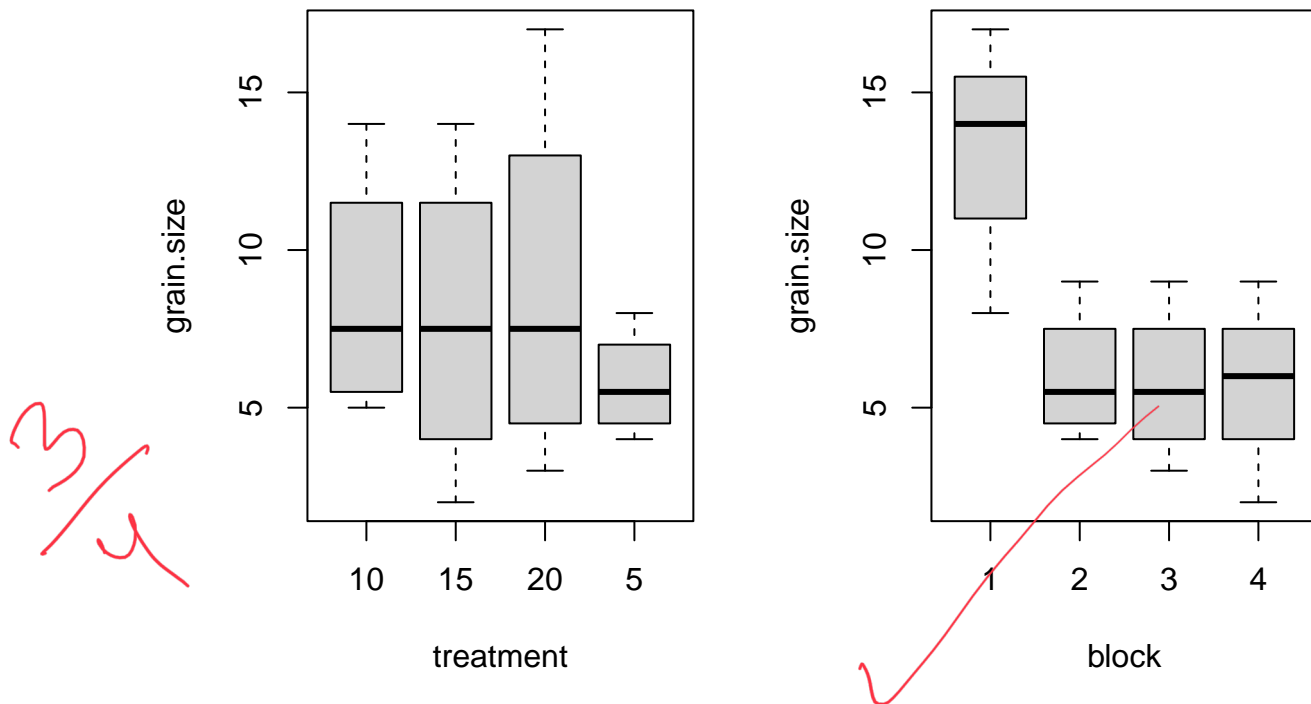
```
# Residuals vs furnace  
plot(aov.eData$model$block,aov.eData$residuals)
```

I do not think these plots convey any useful information.

4.17.d

```
par(mfrow=c(1,2))
boxplot(grain.size~treatment,data=eData) #stirring rate
boxplot(grain.size~block,data=eData) #furnace
```



If smaller grain size is desirable, process engineers should recommend stirring rate 5 and any of furnaces 2,3, or 4.

4.24

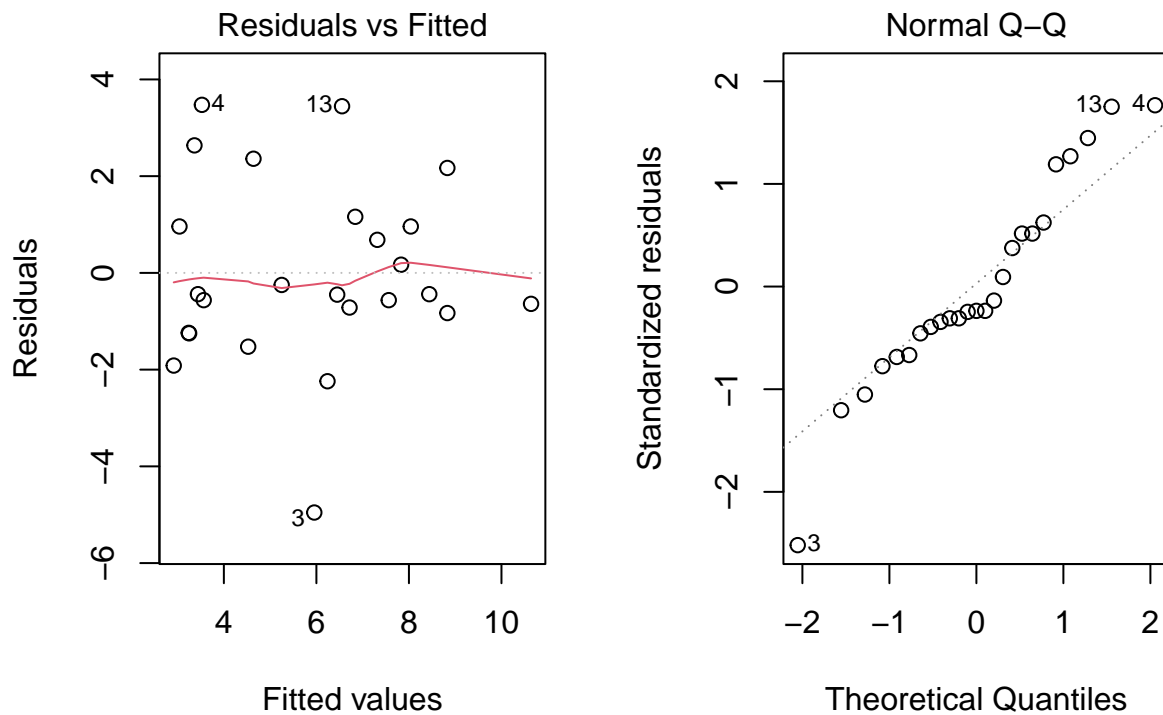
```
treatment = rep(c("1","2","3","4","5"), each=5) #batch
block = rep(c("1", "2", "3", "4","5"),5) #day
formula=c("A","B","C","D","E",
          "C","E","A","D","B",
          "B","A","C","E","D",
          "D","C","E","B","A",
          "E","D","B","A","C") #ingredients
responses = c(8,7,1,7,3,
              11,2,7,3,8,
              4,9,10,1,5,
              6,8,6,6,10,
              4,2,3,8,8)
df = data.frame(treatment,block,formula,responses)
one.way = aov(responses~treatment+block+formula, data=df)
summary(one.way)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## treatment  4  15.44   3.860   0.536  0.712
## block      4  12.24   3.060   0.425  0.788
```

```
## formula      4  92.54  23.136   3.213  0.052 .
## Residuals    12  86.42   7.201
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- (a) At a significance level of $\alpha=0.05$, there is evidence ($p=0.052$) that ingredients may affect reaction time; REJECT null hypothesis.

```
par(mfrow=c(1,2))
plot(one.way,1)
plot(one.way,2)
```



- (b) As the theoretical quantities are well-described by the diagonal, which appears to intersect $(x,y)=(0,0)$ with a slope of $m=1$, one can conclude that the normality assumption holds, although there is a minor dip around the origin, which may argue otherwise.
- (c) As the Residuals vs Fitted plot does not appear increase or decrease from left to right, with residuals randomly scatter throughout, the homogeneity assumption appears to hold.

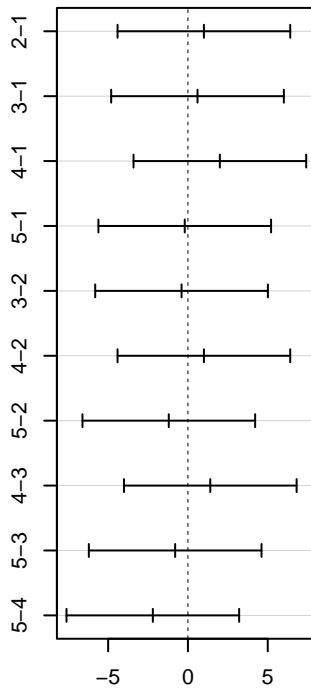
```
TukeyTest=TukeyHSD(one.way, conf.level=0.95)
TukeyTest
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = responses ~ treatment + block + formula, data = df)
```

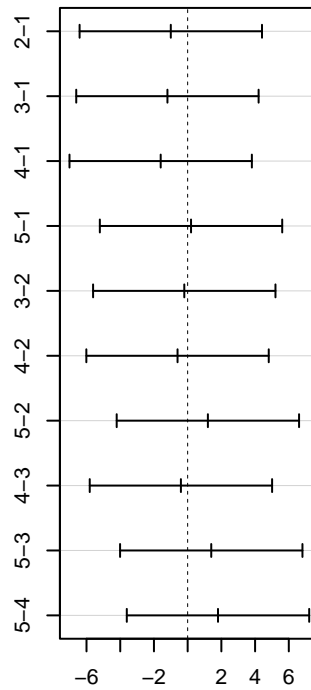
```
##
## $treatment
##      diff      lwr      upr      p adj
## 2-1  1.0 -4.409729 6.409729 0.9741580
## 3-1  0.6 -4.809729 6.009729 0.9961769
## 4-1  2.0 -3.409729 7.409729 0.7630942
## 5-1 -0.2 -5.609729 5.209729 0.9999495
## 3-2 -0.4 -5.809729 5.009729 0.9992120
## 4-2  1.0 -4.409729 6.409729 0.9741580
## 5-2 -1.2 -6.609729 4.209729 0.9509276
## 4-3  1.4 -4.009729 6.809729 0.9178209
## 5-3 -0.8 -6.209729 4.609729 0.9886044
## 5-4 -2.2 -7.609729 3.209729 0.6983860
##
## $block
##      diff      lwr      upr      p adj
## 2-1 -1.0 -6.409729 4.409729 0.9741580
## 3-1 -1.2 -6.609729 4.209729 0.9509276
## 4-1 -1.6 -7.009729 3.809729 0.8747511
## 5-1  0.2 -5.209729 5.609729 0.9999495
## 3-2 -0.2 -5.609729 5.209729 0.9999495
## 4-2 -0.6 -6.009729 4.809729 0.9961769
## 5-2  1.2 -4.209729 6.609729 0.9509276
## 4-3 -0.4 -5.809729 5.009729 0.9992120
## 5-3  1.4 -4.009729 6.809729 0.9178209
## 5-4  1.8 -3.609729 7.209729 0.8226133
##
## $formula
##      diff      lwr      upr      p adj
## B-A -2.80 -8.209729 2.6097289 0.4964575
## C-A -0.88 -6.289729 4.5297289 0.9837723
## D-A -3.72 -9.129729 1.6897289 0.2464383
## E-A -5.20 -10.609729 0.2097289 0.0616300
## C-B  1.92 -3.489729 7.3297289 0.7876550
## D-B -0.92 -6.329729 4.4897289 0.9808987
## E-B -2.40 -7.809729 3.0097289 0.6308789
## D-C -2.84 -8.249729 2.5697289 0.4835104
## E-C -4.32 -9.729729 1.0897289 0.1441919
## E-D -1.48 -6.889729 3.9297289 0.9017590
```

```
par(mfrow=c(1,3))
plot(TukeyTest)
```

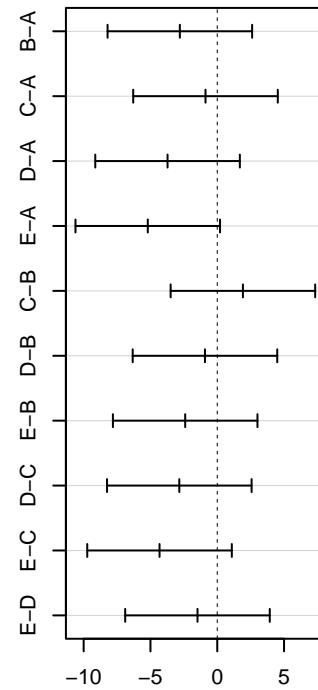
95% family-wise confidence lev 95% family-wise confidence lev 95% family-wise confidence lev



Differences in mean levels of treatment



Differences in mean levels of block



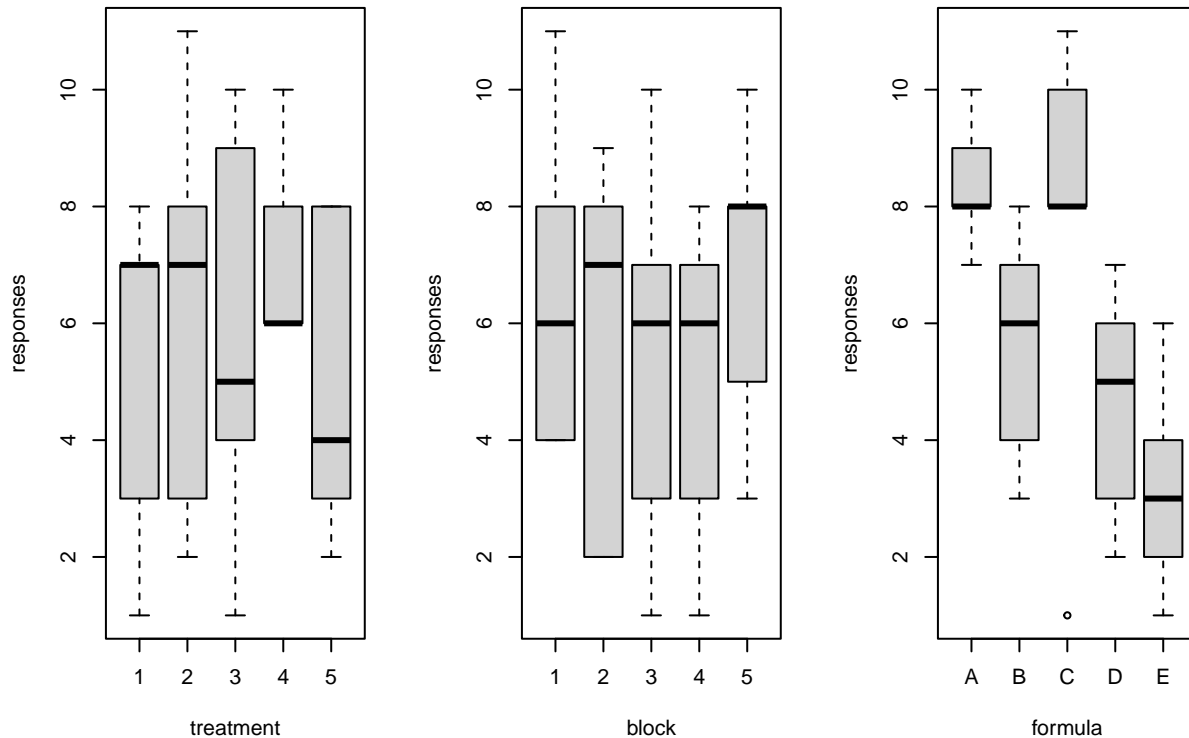
Differences in mean levels of formula

Formally, all confidence intervals, at the 95% level, cross 0, so there are no significantly different response means, for batches, days, and ingredients.

Boxplots

```
par(mfrow=c(1,3))
boxplot(responses~treatment,data=df) #batch
boxplot(responses~block,data=df) #day
boxplot(responses~formula,data=df) #ingredients
```

2/9



Although the boxplots might contest those findings.

4.37

```

treatment = rep(c("1","2","3","4","5"), each=5) #batches of raw material
block = rep(c("1", "2", "3","4","5"),5) #acid concentration
latin = c("A","B","C","D","E",
          "B","C","D","E","A",
          "C","D","E","A","B",
          "D","E","A","B","C",
          "E","A","B","C","D") #standing times
greek = c("a","b","g","d","e",
          "g","d","e","a","b",
          "e","a","b","g","d",
          "b","g","d","e","a",
          "d","e","a","b","g") #catalyst concentrations
yield = c(26, 16, 19, 16, 13,
          18, 21, 18, 11, 21,
          20, 12, 16, 25, 13,
          15, 15, 22, 14, 17,
          10, 24, 17, 17, 14)

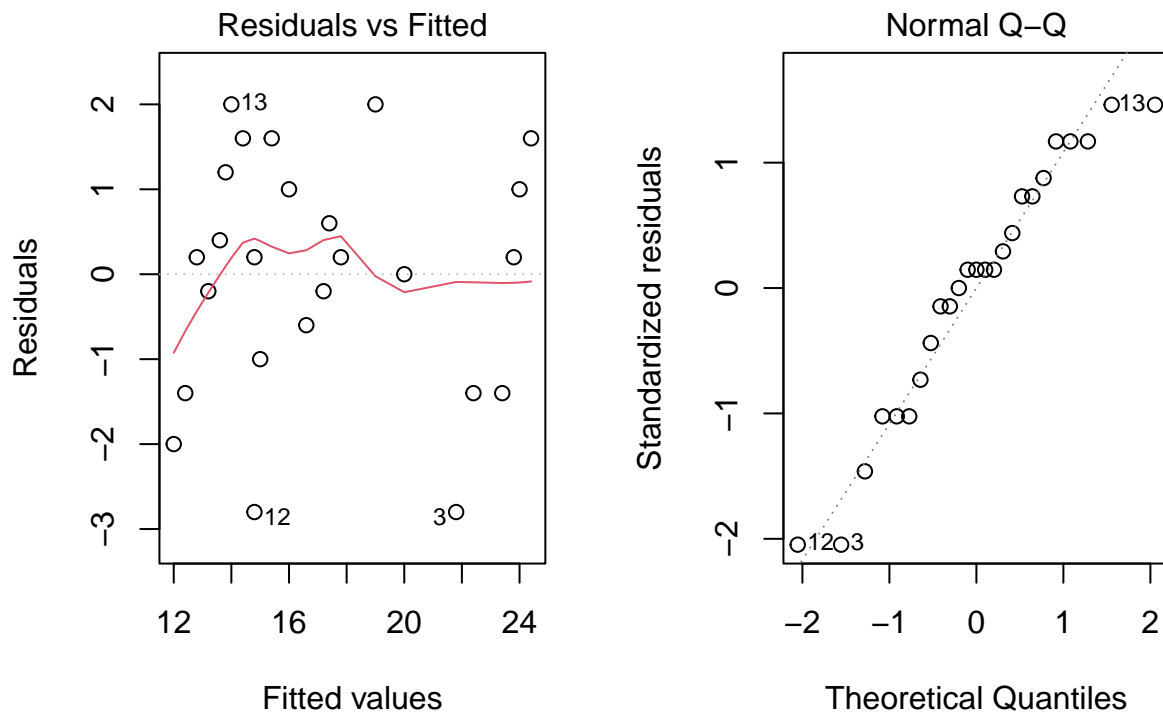
df = data.frame(treatment,block,latin,greek,yield)
one.way = aov(yield~treatment+block+latin+greek, data=df)
summary(one.way)

```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## treatment  4   10.0     2.50   0.427 0.785447
## block      4   24.4     6.10   1.043 0.442543
## latin      4  342.8    85.70  14.650 0.000941 ***
## greek      4   12.0     3.00   0.513 0.728900
## Residuals  8   46.8     5.85
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- (a) At a significance level of $\alpha=0.05$, only the standing times ($p=0.000941$) appear to significantly affect the yield.

```
par(mfrow=c(1,2))
plot(one.way,1)
plot(one.way,2)
```



- (b) As the theoretical quantities are well-described by the diagonal, which appears to intersect $(x,y)=(0,0)$ with a slope of $m=1$, one can conclude that the normality assumption holds, although there is a minor dip around the origin, which may argue otherwise.
- (c) As the Residuals vs Fitted plot does not appear increase or decrease from left to right, with residuals randomly scatter throughout, the homogeneity assumption appears to hold.

```
TukeyTest=TukeyHSD(one.way, conf.level=0.95)
TukeyTest
```

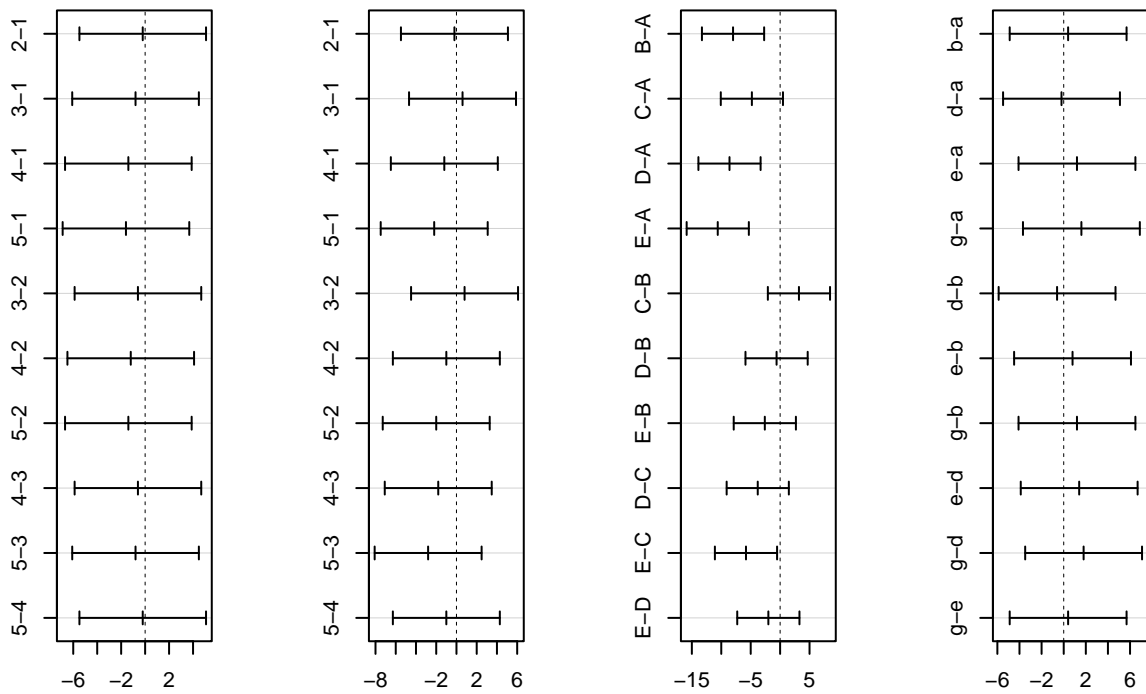
```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = yield ~ treatment + block + latin + greek, data = df)
##
## $treatment
##      diff      lwr      upr      p adj
## 2-1 -0.2 -5.484751 5.084751 0.9999182
## 3-1 -0.8 -6.084751 4.484751 0.9823986
## 4-1 -1.4 -6.684751 3.884751 0.8834072
## 5-1 -1.6 -6.884751 3.684751 0.8279246
## 3-2 -0.6 -5.884751 4.684751 0.9939694
## 4-2 -1.2 -6.484751 4.084751 0.9281909
## 5-2 -1.4 -6.684751 3.884751 0.8834072
## 4-3 -0.6 -5.884751 4.684751 0.9939694
## 5-3 -0.8 -6.084751 4.484751 0.9823986
## 5-4 -0.2 -5.484751 5.084751 0.9999182
##
## $block
##      diff      lwr      upr      p adj
## 2-1 -0.2 -5.484751 5.084751 0.9999182
## 3-1  0.6 -4.684751 5.884751 0.9939694
## 4-1 -1.2 -6.484751 4.084751 0.9281909
## 5-1 -2.2 -7.484751 3.084751 0.6232282
## 3-2  0.8 -4.484751 6.084751 0.9823986
## 4-2 -1.0 -6.284751 4.284751 0.9610846
## 5-2 -2.0 -7.284751 3.284751 0.6948188
## 4-3 -1.8 -7.084751 3.484751 0.7640759
## 5-3 -2.8 -8.084751 2.484751 0.4197369
## 5-4 -1.0 -6.284751 4.284751 0.9610846
##
## $latin
##      diff      lwr      upr      p adj
## B-A -8.0 -13.284751 -2.7152488 0.0051639
## C-A -4.8 -10.084751  0.4847512 0.0770797
## D-A -8.6 -13.884751 -3.3152488 0.0032815
## E-A -10.6 -15.884751 -5.3152488 0.0008219
## C-B  3.2 -2.084751  8.4847512 0.3087034
## D-B -0.6 -5.884751  4.6847512 0.9939694
## E-B -2.6 -7.884751  2.6847512 0.4837165
## D-C -3.8 -9.084751  1.4847512 0.1869031
## E-C -5.8 -11.084751 -0.5152488 0.0317351
## E-D -2.0 -7.284751  3.2847512 0.6948188
##
## $greek
##      diff      lwr      upr      p adj
## b-a  0.4 -4.884751 5.684751 0.9987373
## d-a -0.2 -5.484751 5.084751 0.9999182
## e-a  1.2 -4.084751 6.484751 0.9281909
## g-a  1.6 -3.684751 6.884751 0.8279246
```



```
## d-b -0.6 -5.884751 4.684751 0.9939694
## e-b 0.8 -4.484751 6.084751 0.9823986
## g-b 1.2 -4.084751 6.484751 0.9281909
## e-d 1.4 -3.884751 6.684751 0.8834072
## g-d 1.8 -3.484751 7.084751 0.7640759
## g-e 0.4 -4.884751 5.684751 0.9987373
```

```
par(mfrow=c(1,4))
plot(TukeyTest)
```

% family-wise confidence % family-wise confidence % family-wise confidence % family-wise confidence



Differences in mean levels of trea Differences in mean levels of bl Differences in mean levels of lε Differences in mean levels of gr

Tukey's test confirms that only standing times significantly affect yield quantities, with significantly different pairs: B-A, D-A, E-A, E-C.

O/S