

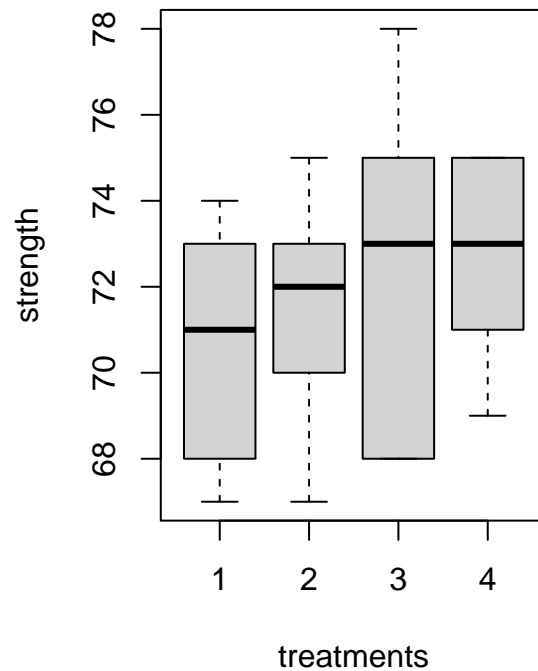
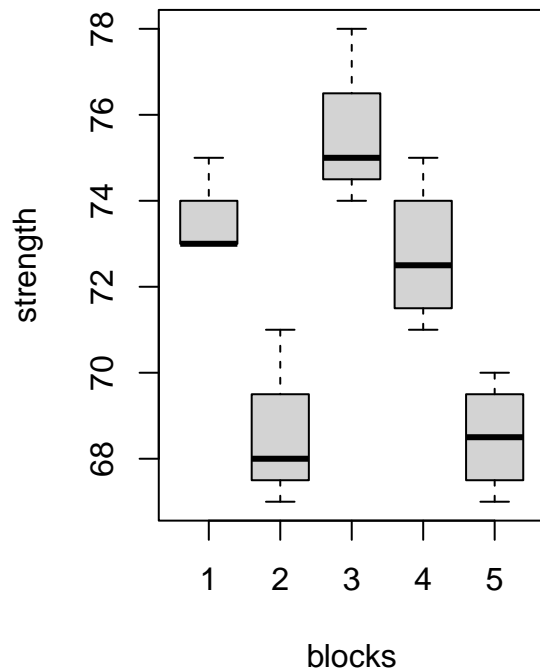
# STAT 407 Homework 2

Charles Hwang

10/8/2021

## Problem 4.3

```
rm(list=ls())
A<-c(73,68,74,71,67)
B<-c(73,67,75,72,70)
C<-c(75,68,78,73,68)
D<-c(73,71,75,75,69)
strength<-c(A,B,C,D)
blbolt<-as.factor(rep(1:5,times=4))
trchem<-as.factor(rep(1:4,each=5))
par(mfrow=c(1,2))
boxplot(strength~blbolt,xlab="blocks")
boxplot(strength~trchem,xlab="treatments")
```



```
t<-lm(strength~trchem+blbolt)
library(car)
anova(t)
```

```
## Analysis of Variance Table
##
## Response: strength
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## trchem      3  12.95   4.317  2.3761    0.1211
## blbolt      4 157.00  39.250 21.6055 2.059e-05 ***
## Residuals  12   21.80   1.817
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# We fail to reject the null hypothesis at the alpha = 0.05 level. There is insufficient
# evidence (p = 0.1211) that the mean tensile strengths of the chemical agents are different.
bartlett.test(strength~trchem) # Testing normality and equal variance assumptions

##
## Bartlett test of homogeneity of variances
##
## data:  strength by trchem
## Bartlett's K-squared = 1.1508, df = 3, p-value = 0.7648
bartlett.test(strength~blbolt)

##
## Bartlett test of homogeneity of variances
##
## data:  strength by blbolt
## Bartlett's K-squared = 1.1175, df = 4, p-value = 0.8915
bartlett.test(t$residuals~trchem)

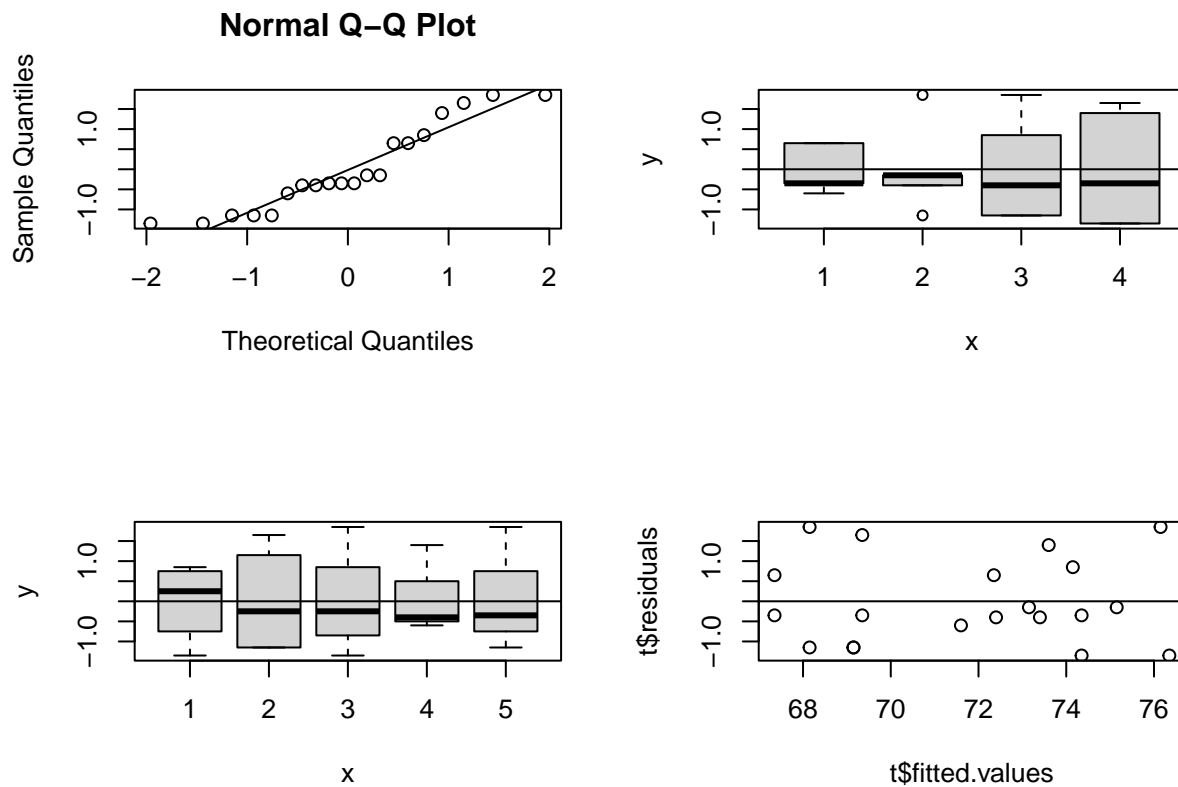
##
## Bartlett test of homogeneity of variances
##
## data:  t$residuals by trchem
## Bartlett's K-squared = 2.6757, df = 3, p-value = 0.4444
leveneTest(strength~trchem)

## Levene's Test for Homogeneity of Variance (center = median)
##           Df F value Pr(>F)
## group      3  0.5815 0.6357
##           16
shapiro.test(t$residual)

##
## Shapiro-Wilk normality test
##
## data:  t$residual
## W = 0.8996, p-value = 0.04054

# We should exercise caution here because the null hypothesis for the Shapiro-Wilk test is
# rejected (p = 0.04054) and the residuals may not be normally distributed.
par(mfrow=c(2,2))
qqnorm(t$residuals)
qqline(t$residuals)
plot(c(trchem),t$residuals)
abline(h=0)
plot(c(blbolt),t$residuals)
abline(h=0)
plot(t$fitted.values,t$residuals)
```

```
abline(h=0)
```

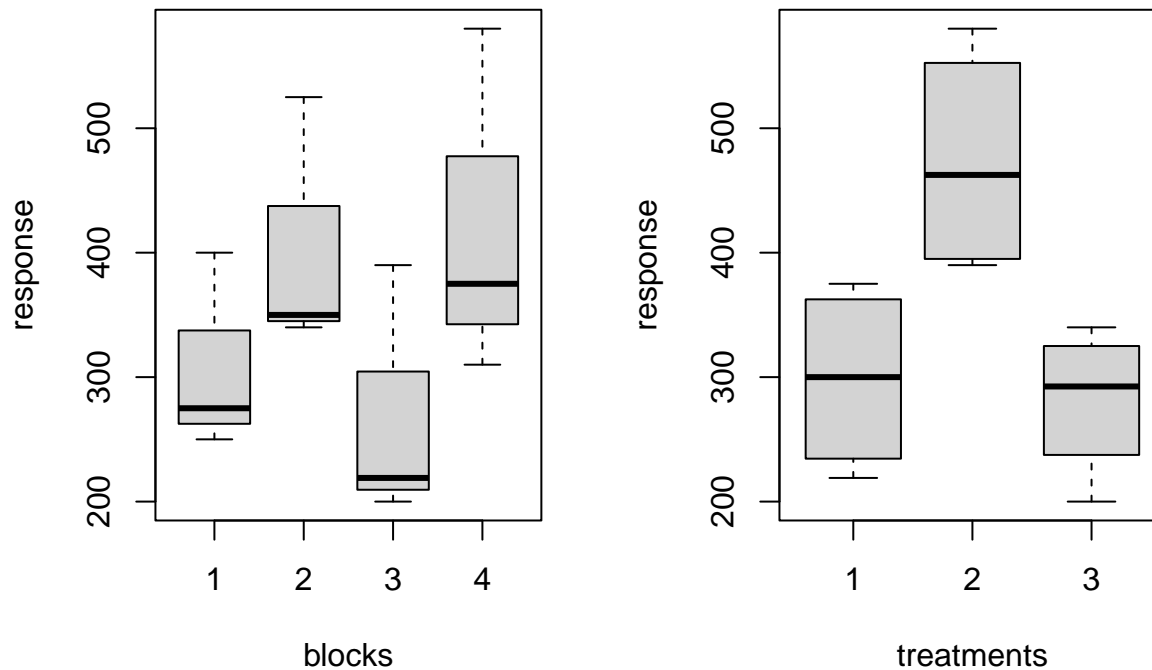


*# There does not appear to be any pattern in the Q-Q plot or residual plots.*

## Problem 4.8

(a)

```
NE<-c(250,400,275)
NW<-c(350,525,340)
SE<-c(219,390,200)
SW<-c(375,580,310)
response<-as.vector(rbind(NE,NW,SE,SW))
blregion<-as.factor(rep(1:4,times=3))
trdesign<-as.factor(rep(1:3,each=4))
par(mfrow=c(1,2))
boxplot(response~blregion,xlab="blocks")
boxplot(response~trdesign,xlab="treatments")
```



```
r<-lm(response~trdesign+blregion)
anova(r)
```

```
## Analysis of Variance Table
##
## Response: response
##          Df Sum Sq Mean Sq F value    Pr(>F)
## trdesign   2  90755   45378   50.152 0.0001798 ***
## blregion   3  49036   16345   18.065 0.0020837 **
## Residuals  6   5429     905
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# We reject the null hypothesis at the alpha = 0.05 level. There is sufficient
# evidence (p = 0.0002) that the mean response rate of at least one mailer is different.
bartlett.test(response~trdesign) # Testing normality and equal variance assumptions
```

```
##
## Bartlett test of homogeneity of variances
##
## data:  response by trdesign
## Bartlett's K-squared = 0.49796, df = 2, p-value = 0.7796
```

```
bartlett.test(response~blregion)
```

```
##
## Bartlett test of homogeneity of variances
##
## data:  response by blregion
## Bartlett's K-squared = 0.52645, df = 3, p-value = 0.913
```

```
leveneTest(response~trdesign)
```

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

```
## group 2 1.6742 0.2409
##      9
```

(b)

```
library(agricolae)
LSDr<-LSD.test(aov(response~trdesign+blregion),"trdesign")
LSDr
```

```
## $statistics
##      MSerror Df      Mean      CV  t.value      LSD
##    904.8056  6 351.1667 8.565729 2.446912 52.04523
##
## $parameters
##      test p.adjusted name.t ntr alpha
## Fisher-LSD      none trdesign  3  0.05
##
## $means
##      response      std r      LCL      UCL Min Max      Q25      Q50      Q75
## 1    298.50 75.66814 4 261.6985 335.3015 219 375 242.25 300.0 356.25
## 2    473.75 93.75278 4 436.9485 510.5515 390 580 397.50 462.5 538.75
## 3    281.25 60.32896 4 244.4485 318.0515 200 340 256.25 292.5 317.50
##
## $comparison
## NULL
##
## $groups
##      response groups
## 2    473.75      a
## 1    298.50      b
## 3    281.25      b
##
## attr(,"class")
## [1] "group"
```

We can see that design 2 clearly has a different mean response rate than designs 1 and 3.

(c)

```
bartlett.test(r$residuals~trdesign)
```

```
##
## Bartlett test of homogeneity of variances
##
## data:  r$residuals by trdesign
## Bartlett's K-squared = 6.255, df = 2, p-value = 0.04383
# We should exercise caution here because the null hypothesis for the Bartlett's test
# is rejected (p = 0.04383) and the residuals may not have equal variance.
shapiro.test(r$residual)
```

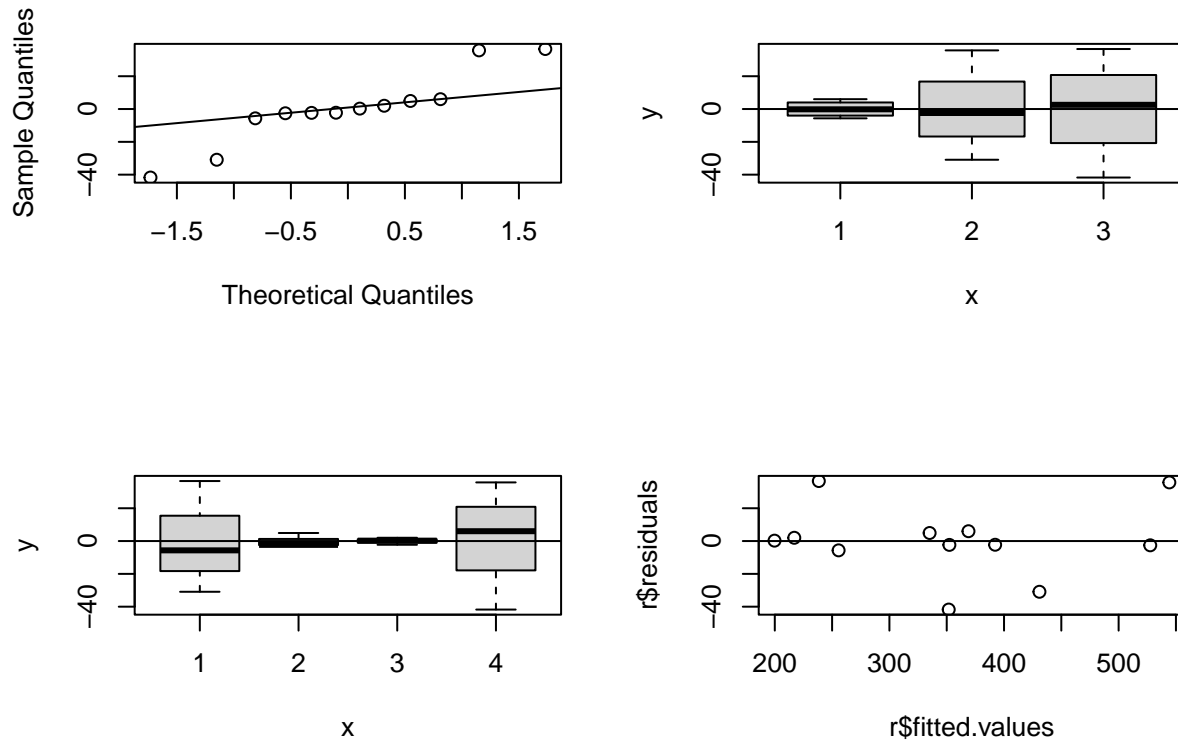
```
##
## Shapiro-Wilk normality test
##
## data:  r$residual
## W = 0.88867, p-value = 0.1133
```

```

par(mfrow=c(2,2))
qqnorm(r$residuals)
qqline(r$residuals)
plot(c(trdesign),r$residuals)
abline(h=0)
plot(c(blregion),r$residuals)
abline(h=0)
plot(r$fitted.values,r$residuals)
abline(h=0)

```

### Normal Q-Q Plot



*# There is some variation in the Q-Q plot, but there does not appear to be any pattern in the residual plots.*

*friedman.test(response,trdesign,blregion) # Additional post-hoc analysis*

```

##
## Friedman rank sum test
##
## data: response, trdesign and blregion
## Friedman chi-squared = 6.5, df = 2, p-value = 0.03877
TukeyHSD(aov(response~trdesign+blregion),conf.level=0.95)$trdesign

```

```

##      diff      lwr      upr      p adj
## 2-1  175.25  109.98853  240.51147 0.0004235806
## 3-1  -17.25  -82.51147   48.01147 0.7104869430
## 3-2 -192.50 -257.76147 -127.23853 0.0002508095

```

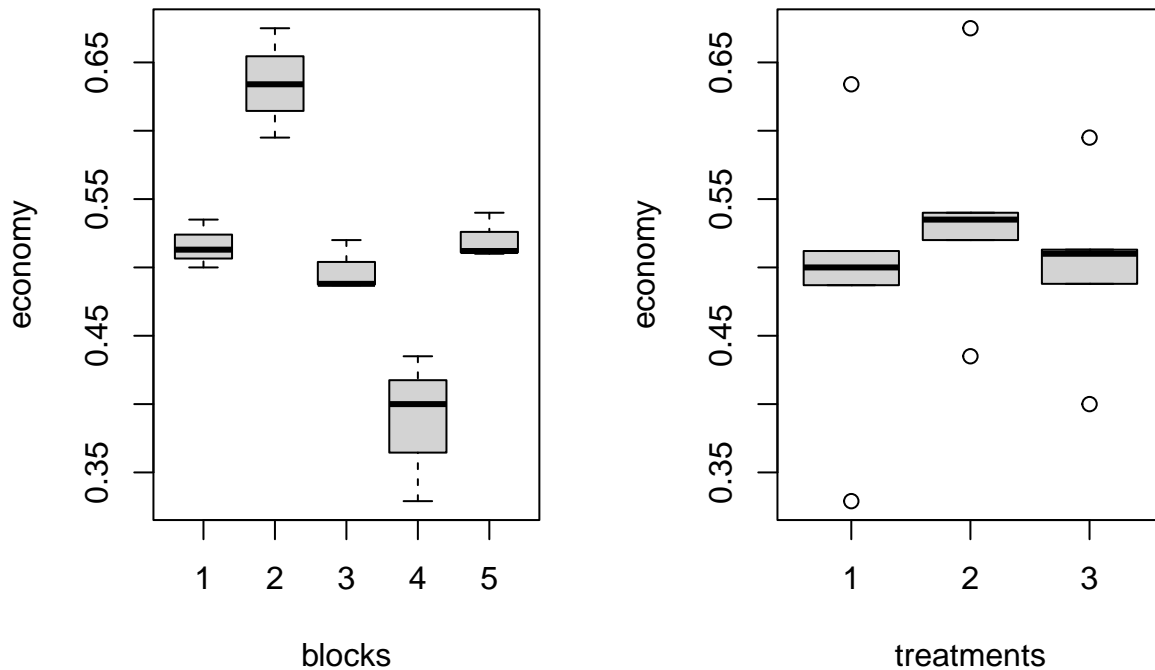
## Problem 4.9

(a)

```

X<-c(0.500,0.634,0.487,0.329,0.512)
Y<-c(0.535,0.675,0.520,0.435,0.540)
Z<-c(0.513,0.595,0.488,0.400,0.510)
economy<-c(X,Y,Z)
bltruck<-as.factor(rep(1:5,times=3))
troil<-as.factor(rep(1:3,each=5))
par(mfrow=c(1,2))
boxplot(economy~bltruck,xlab="blocks")
boxplot(economy~troil,xlab="treatments")

```



```

e<-lm(economy~troil+bltruck)
anova(e)

```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: economy
```

```
##          Df    Sum Sq   Mean Sq F value    Pr(>F)
## troil      2  0.006706  0.0033529   6.3527  0.02229 *
## bltruck    4  0.092100  0.0230249  43.6257 1.781e-05 ***
## Residuals  8  0.004222  0.0005278
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

# We reject the null hypothesis at the alpha = 0.05 level. There is sufficient
# evidence (p = 0.0223) that the mean fuel economy of at least one oil is different.
bartlett.test(economy~troil) # Testing normality and equal variance assumptions

```

```
##
```

```
## Bartlett test of homogeneity of variances
```

```
##
```

```
## data: economy by troil
```

```
## Bartlett's K-squared = 0.70349, df = 2, p-value = 0.7035
```

```
bartlett.test(economy~bltruck)
```

```
##
## Bartlett test of homogeneity of variances
##
## data: economy by bltruck
## Bartlett's K-squared = 3.9598, df = 4, p-value = 0.4115
```

```
leveneTest(economy~troil)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 2  0.1425 0.8686
##      12
```

(b)

```
LSDe<-LSD.test(aov(economy~troil+bltruck),"troil")
LSDe
```

```
## $statistics
##      MSerror Df      Mean      CV  t.value      LSD
## 0.0005277833  8 0.5115333 4.491112 2.306004 0.03350564
##
## $parameters
##      test p.adjusted name.t ntr alpha
## Fisher-LSD      none troil   3  0.05
##
## $means
## economy      std r      LCL      UCL   Min   Max   Q25   Q50   Q75
## 1  0.4924 0.10865220 5 0.4687079 0.5160921 0.329 0.634 0.487 0.500 0.512
## 2  0.5410 0.08612491 5 0.5173079 0.5646921 0.435 0.675 0.520 0.535 0.540
## 3  0.5012 0.06969720 5 0.4775079 0.5248921 0.400 0.595 0.488 0.510 0.513
##
## $comparison
## NULL
##
## $groups
## economy groups
## 2  0.5410      a
## 3  0.5012      b
## 1  0.4924      b
##
## attr("class")
## [1] "group"
```

We can see that oil 2 has a different mean fuel economy than oils 1 and 3.

(c)

```
bartlett.test(e$residuals~troil)
```

```
##
## Bartlett test of homogeneity of variances
##
## data: e$residuals by troil
## Bartlett's K-squared = 1.0484, df = 2, p-value = 0.592
```



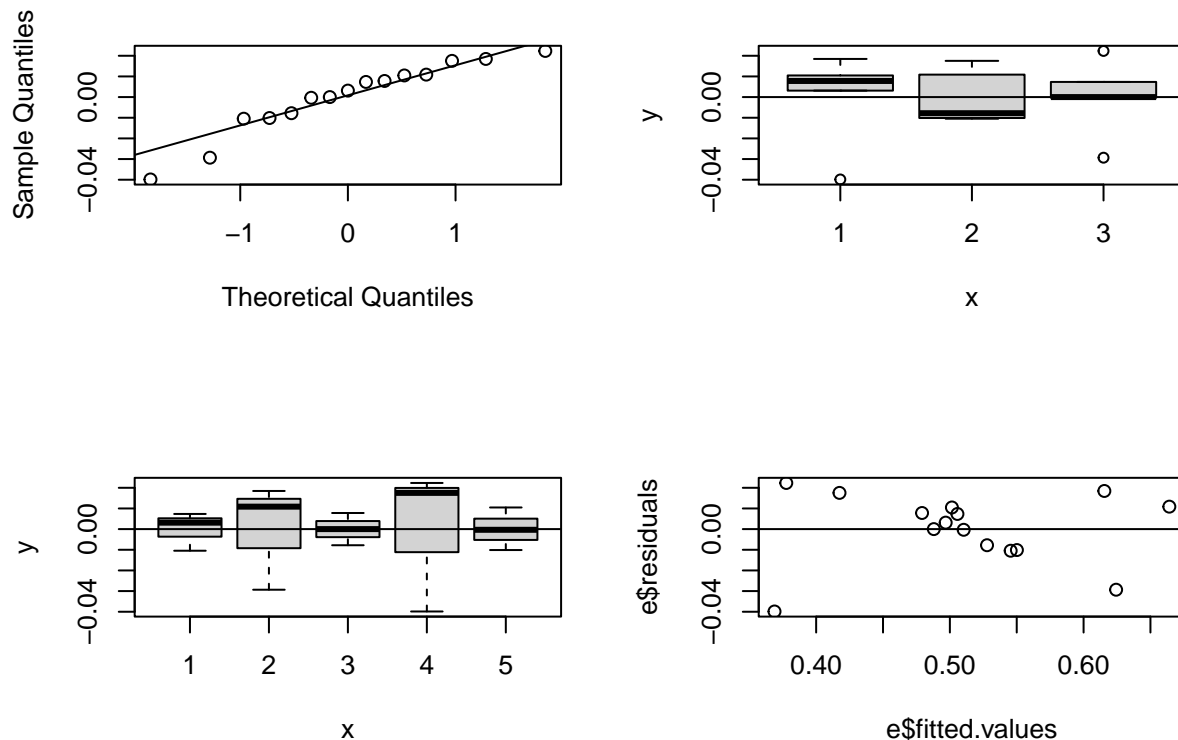
```
shapiro.test(e$residual)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  e$residual
## W = 0.91855, p-value = 0.183
```

```
# None of the null hypotheses were rejected, so the normality and
par(mfrow=c(2,2)) # equal variance assumptions appear to be met.
```

```
qqnorm(e$residuals)
qqline(e$residuals)
plot(c(troil),e$residuals)
abline(h=0)
plot(c(bltruck),e$residuals)
abline(h=0)
plot(e$fitted.values,e$residuals)
abline(h=0)
```

**Normal Q-Q Plot**



```
# There does not appear to be any pattern in the Q-Q plot or residual plots.
friedman.test(economy,troil,bltruck) # Additional post-hoc analysis
```

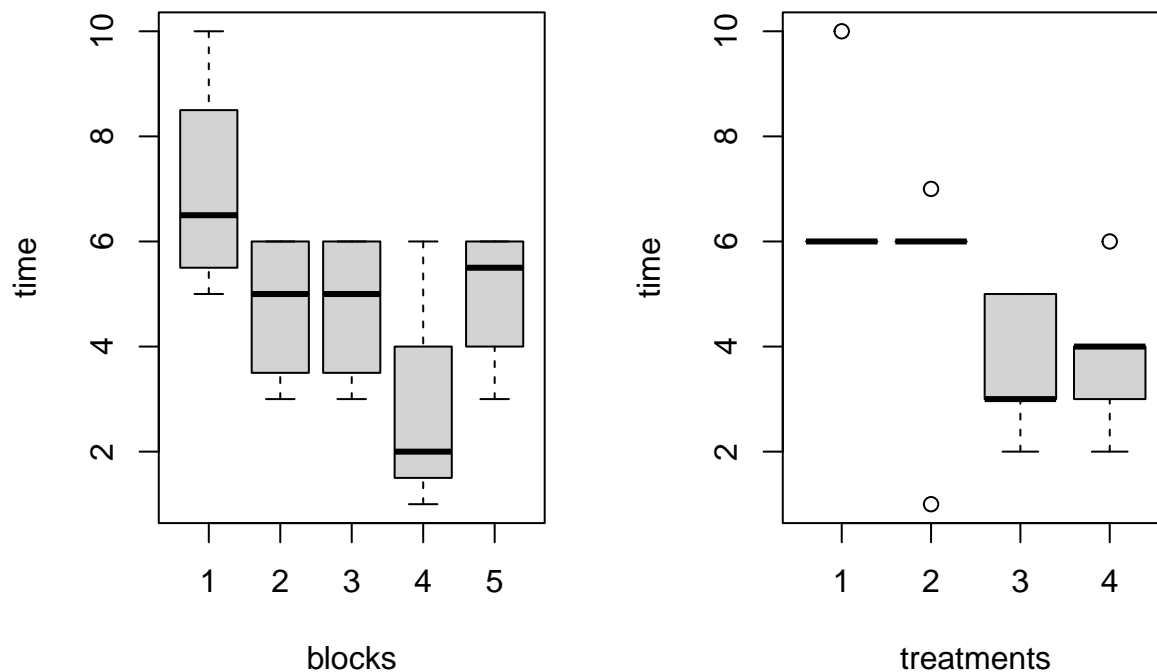
```
##
##  Friedman rank sum test
##
## data:  economy, troil and bltruck
## Friedman chi-squared = 7.6, df = 2, p-value = 0.02237
TukeyHSD(aov(economy~troil+bltruck),conf.level=0.95)$troil
```

```
##          diff          lwr          upr          p adj
```

```
## 2-1  0.0486  0.007082078 0.090117922 0.02458094
## 3-1  0.0088 -0.032717922 0.050317922 0.82109697
## 3-2 -0.0398 -0.081317922 0.001717922 0.05949792
```

## Problem 4.21

```
fou<-c(10,6,6,6,6)
six<-c(7,6,6,1,6)
eig<-c(5,3,3,2,5)
ten<-c(6,4,4,2,3)
time<-c(fou,six,eig,ten)
blsubej<-as.factor(rep(1:5,times=4))
trdist<-as.factor(rep(1:4,each=5))
par(mfrow=c(1,2))
boxplot(time~blsubej,xlab="blocks")
boxplot(time~trdist,xlab="treatments")
```



```
ft<-lm(time~trdist+blsubej)
anova(ft)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: time
```

```
##      Df Sum Sq Mean Sq F value    Pr(>F)
## trdist  3  32.95   10.983    8.6144 0.002543 **
## blsubej  4  36.30    9.075    7.1176 0.003548 **
## Residuals 12  15.30    1.275
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# We reject the null hypothesis at the alpha = 0.05 level. There is sufficient
# evidence (p = 0.0025) that the mean focus time for at least one distance is different.
bartlett.test(time~trdist) # Testing normality and equal variance assumptions
```

```

##
## Bartlett test of homogeneity of variances
##
## data: time by trdist
## Bartlett's K-squared = 1.4631, df = 3, p-value = 0.6908
bartlett.test(time~blsubj)

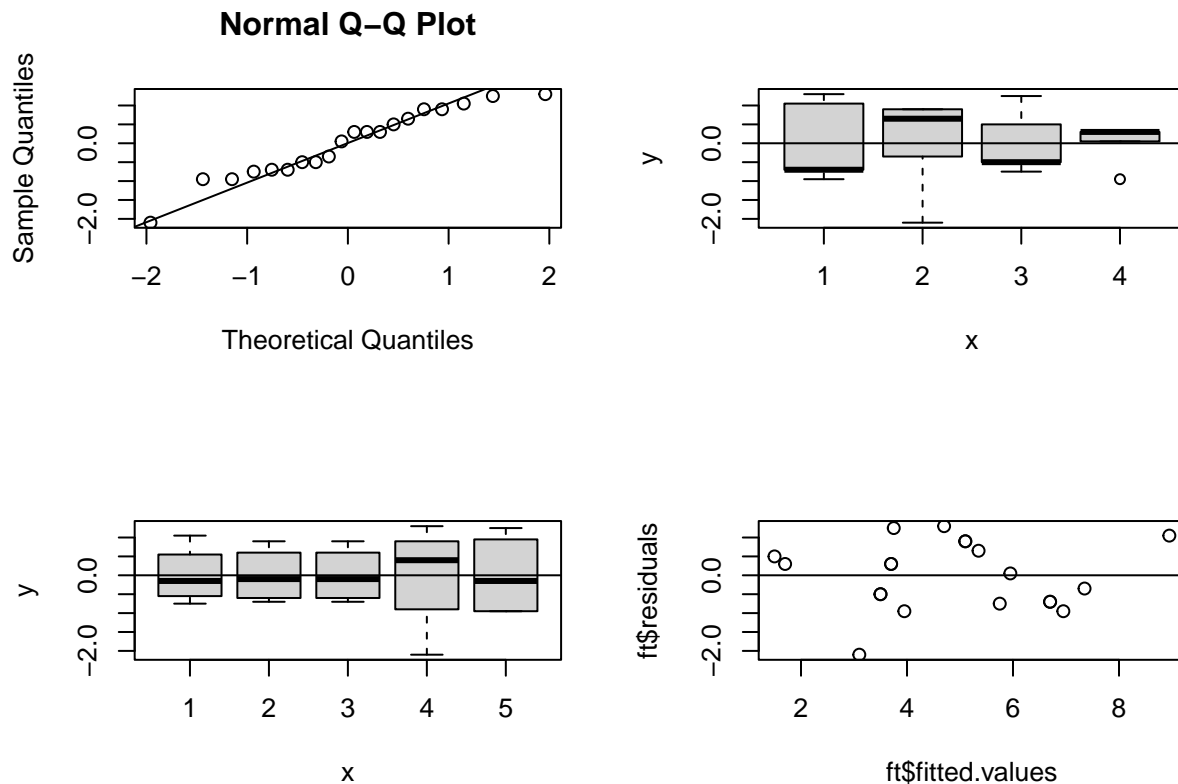
##
## Bartlett test of homogeneity of variances
##
## data: time by blsubj
## Bartlett's K-squared = 1.042, df = 4, p-value = 0.9034
bartlett.test(ft$residuals~trdist)

##
## Bartlett test of homogeneity of variances
##
## data: ft$residuals by trdist
## Bartlett's K-squared = 2.6207, df = 3, p-value = 0.4539
leveneTest(time~trdist)

## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 3  0.0539 0.9829
##      16
shapiro.test(ft$residual)

##
## Shapiro-Wilk normality test
##
## data: ft$residual
## W = 0.94976, p-value = 0.3634
# None of the null hypotheses were rejected, so the normality and
par(mfrow=c(2,2)) # equal variance assumptions appear to be met.
qqnorm(ft$residuals)
qqline(ft$residuals)
plot(c(trdist),ft$residuals)
abline(h=0)
plot(c(blsubj),ft$residuals)
abline(h=0)
plot(ft$fitted.values,ft$residuals)
abline(h=0)

```



*# There does not appear to be any pattern in the Q-Q plot. There appears to be some  
# pattern in the residuals vs. fitted values plot, but remember that the sample sizes  
# for each distance are relatively small.*  
friedman.test(time,trdist,blsubj) *# Post-hoc analysis*

```
##
## Friedman rank sum test
##
## data: time, trdist and blsubj
## Friedman chi-squared = 9.6522, df = 3, p-value = 0.02177
LSDft<-LSD.test(aov(time~trdist+blsubj),"trdist")
LSDft
```

```
## $statistics
##   MSerror Df Mean      CV  t.value      LSD
##   1.275 12 4.85 23.28163 2.178813 1.555984
##
## $parameters
##      test p.adjusted name.t ntr alpha
## Fisher-LSD      none trdist   4 0.05
##
## $means
##   time      std r      LCL      UCL Min Max Q25 Q50 Q75
## 1  6.8 1.788854 5 5.699753 7.900247  6 10  6  6  6
## 2  5.2 2.387467 5 4.099753 6.300247  1  7  6  6  6
## 3  3.6 1.341641 5 2.499753 4.700247  2  5  3  3  5
## 4  3.8 1.483240 5 2.699753 4.900247  2  6  3  4  4
##
##
## $comparison
```

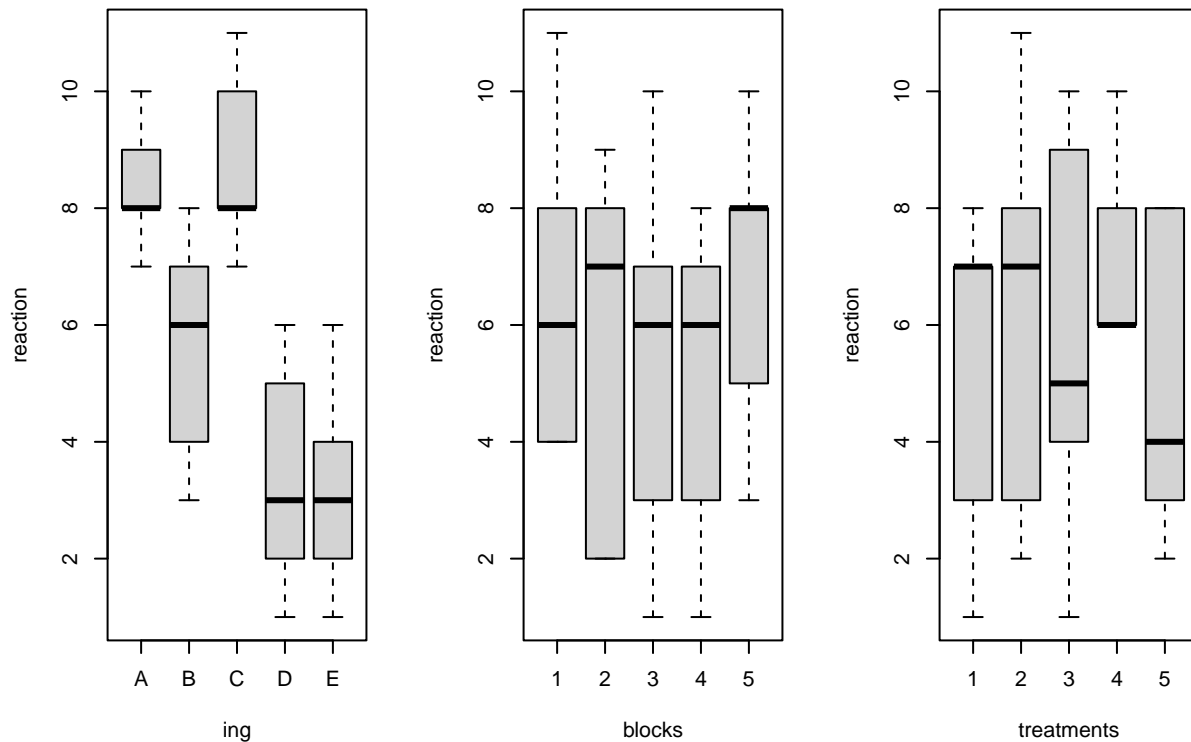
```
## NULL
##
## $groups
##   time groups
## 1  6.8      a
## 2  5.2      b
## 4  3.8     bc
## 3  3.6      c
##
## attr(,"class")
## [1] "group"
```

```
TukeyHSD(aov(time~trdist+blsubj),conf.level=0.95)$trdist
```

```
##      diff      lwr      upr      p adj
## 2-1 -1.6 -3.720219  0.5202195 0.167524230
## 3-1 -3.2 -5.320219 -1.0797805 0.003615573
## 4-1 -3.0 -5.120219 -0.8797805 0.005836959
## 3-2 -1.6 -3.720219  0.5202195 0.167524230
## 4-2 -1.4 -3.520219  0.7202195 0.255437448
## 4-3  0.2 -1.920219  2.3202195 0.991929366
```

## Problem 4.22

```
reaction<-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)
day<-as.factor(rep(1:5,times=5))
batch<-as.factor(rep(1:5,each=5))
ing<-c("A","B","D","C","E","C","E","A","D","B","B","A","C","E","D","D","C","E","B","A","E","D","B","A",
par(mfrow=c(1,3))
boxplot(reaction~ing)
boxplot(reaction~day,xlab="blocks")
boxplot(reaction~batch,xlab="treatments")
```



```
rt<-lm(reaction~ing+day+batch)
anova(rt)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: reaction
```

```
##          Df Sum Sq Mean Sq F value    Pr(>F)
## ing         4 141.44   35.360  11.3092 0.0004877 ***
## day         4  12.24    3.060   0.9787 0.4550143
## batch        4  15.44    3.860   1.2345 0.3476182
## Residuals  12   37.52    3.127
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# We reject the null hypothesis at the alpha = 0.05 level. There is sufficient
# evidence (p = 0.0005) that the mean reaction time for at least one ingredient is different.
```

```
bartlett.test(reaction~ing) # Testing normality and equal variance assumptions
```

```
##
```

```
## Bartlett test of homogeneity of variances
```

```
##
```

```
## data: reaction by ing
```

```
## Bartlett's K-squared = 1.5544, df = 4, p-value = 0.817
```

```
bartlett.test(reaction~batch)
```

```
##
```

```
## Bartlett test of homogeneity of variances
```

```
##
```

```
## data: reaction by batch
```

```
## Bartlett's K-squared = 2.1504, df = 4, p-value = 0.7081
```

```

bartlett.test(reaction~day)

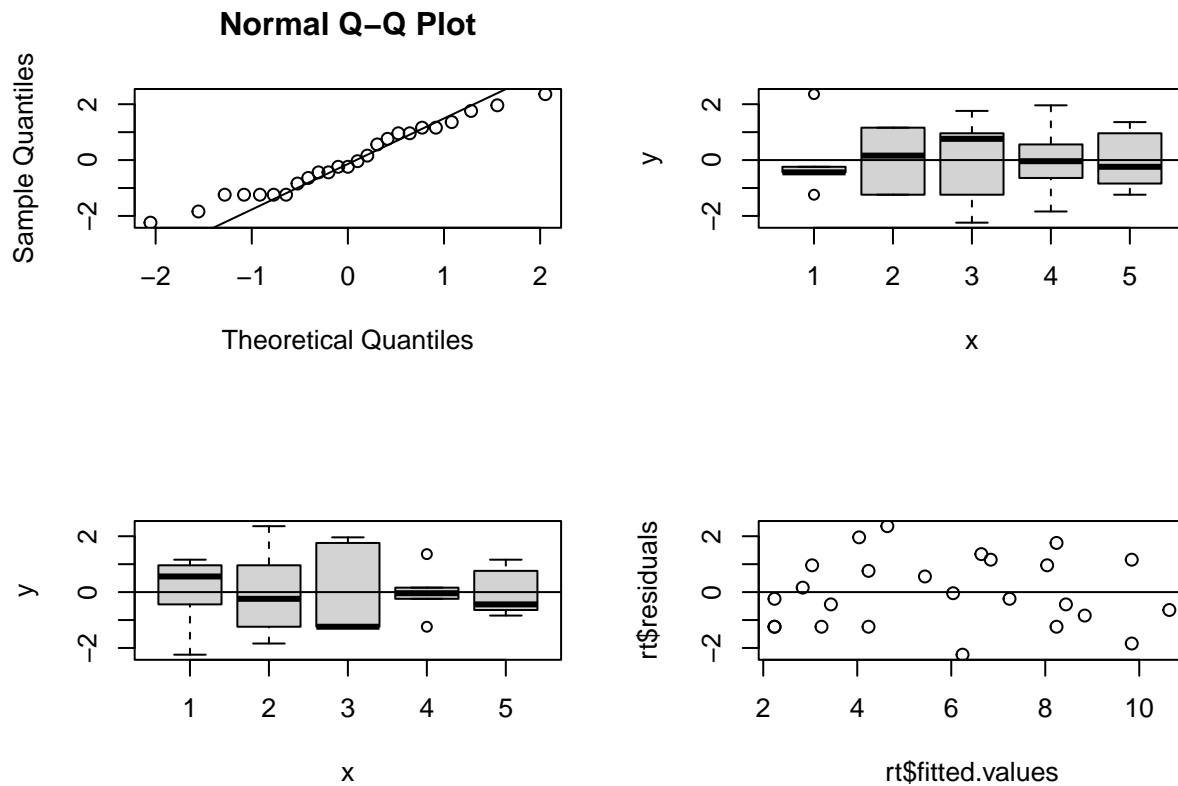
##
## Bartlett test of homogeneity of variances
##
## data: reaction by day
## Bartlett's K-squared = 0.29268, df = 4, p-value = 0.9903
bartlett.test(rt$residuals~batch)

##
## Bartlett test of homogeneity of variances
##
## data: rt$residuals by batch
## Bartlett's K-squared = 0.69132, df = 4, p-value = 0.9524
f<-vector() # Levene's test
for(j in c("A","B","C","D")) f[ing==j]<-abs(reaction[ing==j]-median(reaction[ing==j]))
lt<-lm(f~ing)
anova(lt)

## Analysis of Variance Table
##
## Response: f
##          Df Sum Sq Mean Sq F value Pr(>F)
## ing        3    2.2  0.73333   0.5867 0.6325
## Residuals 16   20.0  1.25000
shapiro.test(rt$residual)

##
## Shapiro-Wilk normality test
##
## data: rt$residual
## W = 0.96606, p-value = 0.5476
# None of the null hypotheses were rejected, so the normality and
par(mfrow=c(2,2)) # equal variance assumptions appear to be met.
qqnorm(rt$residuals)
qqline(rt$residuals)
plot(c(batch),rt$residuals)
abline(h=0)
plot(c(day),rt$residuals)
abline(h=0)
plot(rt$fitted.values,rt$residuals)
abline(h=0)

```



*# There is some variation in the Q-Q plot, but there does not appear to be any pattern  
# in the residual plots.*

```
LSDlt<-LSD.test(aov(reaction~ing+day+batch),"ing") # Post-hoc analysis
LSDlt
```

```
## $statistics
##      MSerror Df Mean      CV  t.value    LSD
##      3.126667 12  5.88 30.07208 2.178813 2.436636
##
## $parameters
##      test p.adjusted name.t ntr alpha
##      Fisher-LSD      none   ing   5  0.05
##
## $means
##      reaction      std r      LCL      UCL Min Max Q25 Q50 Q75
## A      8.4 1.140175 5 6.677038 10.122962  7 10  8  8  9
## B      5.6 2.073644 5 3.877038  7.322962  3  8  4  6  7
## C      8.8 1.643168 5 7.077038 10.522962  7 11  8  8 10
## D      3.4 2.073644 5 1.677038  5.122962  1  6  2  3  5
## E      3.2 1.923538 5 1.477038  4.922962  1  6  2  3  4
##
## $comparison
## NULL
##
## $groups
##      reaction groups
## C      8.8      a
## A      8.4      a
## B      5.6      b
```



```
## D      3.4      b
## E      3.2      b
##
## attr(,"class")
## [1] "group"
TukeyHSD(aov(reaction~ing+day+batch),conf.level=0.95)$ing
```

```
##      diff      lwr      upr      p adj
## B-A -2.8 -6.3646078  0.7646078 0.153943335
## C-A  0.4 -3.1646078  3.9646078 0.996001220
## D-A -5.0 -8.5646078 -1.4353922 0.005586216
## E-A -5.2 -8.7646078 -1.6353922 0.004143094
## C-B  3.2 -0.3646078  6.7646078 0.086435305
## D-B -2.2 -5.7646078  1.3646078 0.336581142
## E-B -2.4 -5.9646078  1.1646078 0.263155088
## D-C -5.4 -8.9646078 -1.8353922 0.003082228
## E-C -5.6 -9.1646078 -2.0353922 0.002300665
## E-D -0.2 -3.7646078  3.3646078 0.999734935
```

## Problem 4.27

```
SLIM<-c(1244,21,82,2221,905,839)
COCOMOA<-c(281,129,396,1306,336,910)
COCOMOR<-c(220,84,458,543,300,794)
COCONOC<-c(225,83,425,552,291,826)
FUNCTIONPOINTS<-c(19,11,-34,121,15,103)
ESTIMALS<-c(-20,35,-53,170,104,199)
accuracy<-c(SLIM,COCOMOA,COCOMOR,COCONOC,FUNCTIONPOINTS,ESTIMALS)
blproj<-as.factor(rep(1:6,times=6))
tralgo<-as.factor(rep(1:6,each=6))
anova(lm(accuracy~tralgo+blproj))
```

```
## Analysis of Variance Table
##
## Response: accuracy
##           Df Sum Sq Mean Sq F value    Pr(>F)
## tralgo      5 2989130   597826   5.3770 0.001720 **
## blproj       5 2287339   457468   4.1146 0.007295 **
## Residuals  25 2779574   111183
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
library(dplyr) # Using Equation 4.16: var_blocks=(MSB-MSE)/a on page 152
(anova(lm(accuracy~tralgo+blproj))$"Mean Sq"[2]-anova(lm(accuracy~tralgo+blproj))$"Mean Sq"[3])/n_disti
```

```
## [1] 57714.16
```