Question 2 - BIOS14 (HT2021)

Read and explore data

```
#Question 2

rm(list=ls())

library(car)

## Loading required package: carData

library(lmtest)

## Warning: package 'lmtest' was built under R version 4.1.2

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

library(multcomp)

## Warning: package 'multcomp' was built under R version 4.1.2

## Loading required package: mvtnorm

## Loading required package: survival

## Warning: package 'survival' was built under R version 4.1.2

## Loading required package: TH.data

## Warning: package 'TH.data' was built under R version 4.1.2

## Loading required package: MASS

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
##     geyser

#read and explore data
pine <- read.csv('pine.csv')
pine$treat <- factor(pine$treat)
```
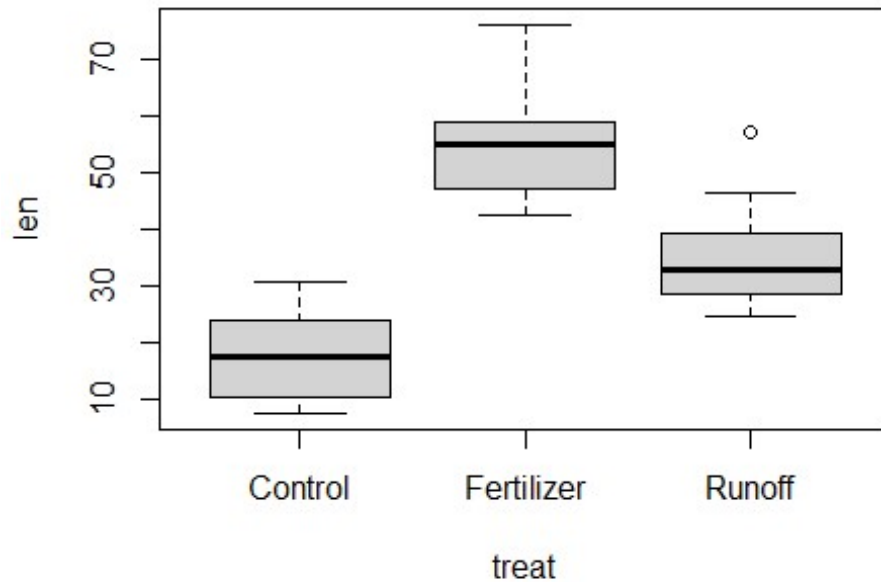
```
plot(len~treat, data=pine)
```


Part (a)

```
#one way ANOVA

#construct model

model.pine <- lm(len~treat-1, data=pine)
#to treat groups as three different individuals rather than as comparison to
control (given as intercept value in the summary)

anova(model.pine)

## Analysis of Variance Table
##
## Response: len
##           Df Sum Sq Mean Sq F value    Pr(>F)
## treat      3  46073 15357.8  175.07 < 2.2e-16 ***
## Residuals 27   2368    87.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(model.pine)

##
## Call:
```

```
## lm(formula = len ~ treat - 1, data = pine)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -12.6330  -7.3797  -0.9502   4.6811  21.8950
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## treatControl      17.821      2.962   6.017 2.02e-06 ***
## treatFertilizer   55.128      2.962  18.613  < 2e-16 ***
## treatRunoff       35.364      2.962  11.940 2.78e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.366 on 27 degrees of freedom
## Multiple R-squared:  0.9511, Adjusted R-squared:  0.9457
## F-statistic: 175.1 on 3 and 27 DF,  p-value: < 2.2e-16
```

```
#posthoc test
ph.pine <- glht(model.pine, linfct=mcp(treat='Tukey'))

summary(ph.pine)
```

```
##
##    Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = len ~ treat - 1, data = pine)
##
## Linear Hypotheses:
##                          Estimate Std. Error t value Pr(>|t|)
## Fertilizer - Control == 0   37.307      4.189   8.907  < 1e-04 ***
## Runoff - Control == 0       17.543      4.189   4.188 0.000753 ***
## Runoff - Fertilizer == 0   -19.764      4.189  -4.719 0.000173 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

```
cld(ph.pine)
```

```
##    Control Fertilizer     Runoff
##        "a"        "c"        "b"
```

```
cld <- c('a','c','b')
#agriculture run-off is significantly different from control - effective
fertilizer
```

Part (b)

A bar plot with the standard errors used here to compare the means of the different groups along with the possible error in computing the means. This is helpful in comparing how group means vary with treatment, which is essentially what we aim at studying in an ANOVA analysis.

```
#illustrate results

means.len <- summary(model.pine)$coefficients[,1]
se.len <- summary(model.pine)$coefficients[,2]
bp <- barplot(means.len, xlab='Treatments', ylab='Seedling length',
main='Barplot illustration of one-way ANOVA', names=c('Control',
'Commercial\n fertilizer', 'Agricultural\n runoff'), col='light blue',
ylim=c(0,65))

arrows(bp, means.len+se.len, bp, means.len-se.len, code=3, angle=90)
text(bp, means.len+se.len+5, cld)
```


Barplot illustration of one-way ANOVA
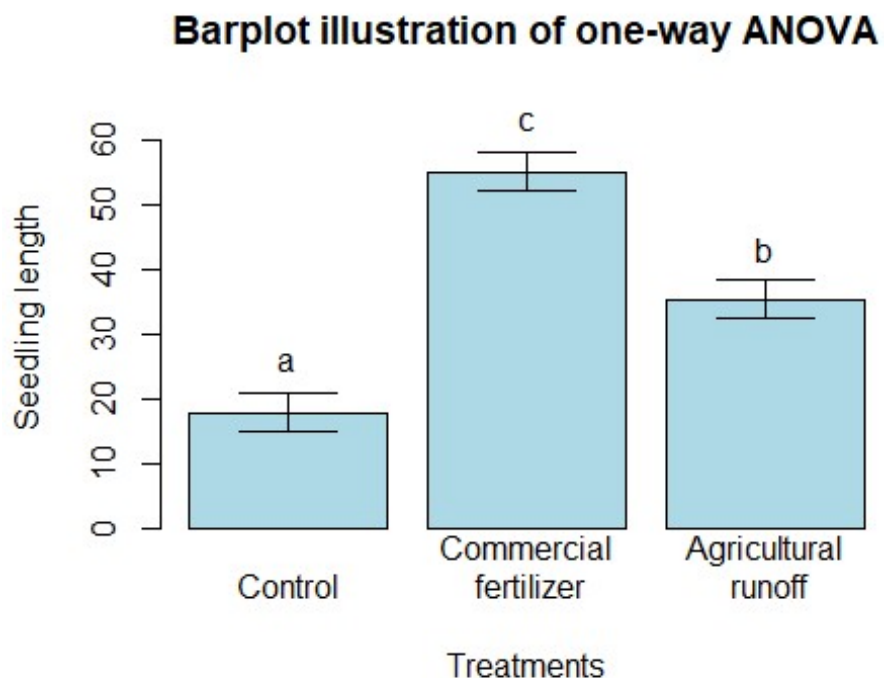
Part (c)

```
#test assumptions

#assumption 1: independent observations

dwtest(model.pine)
```
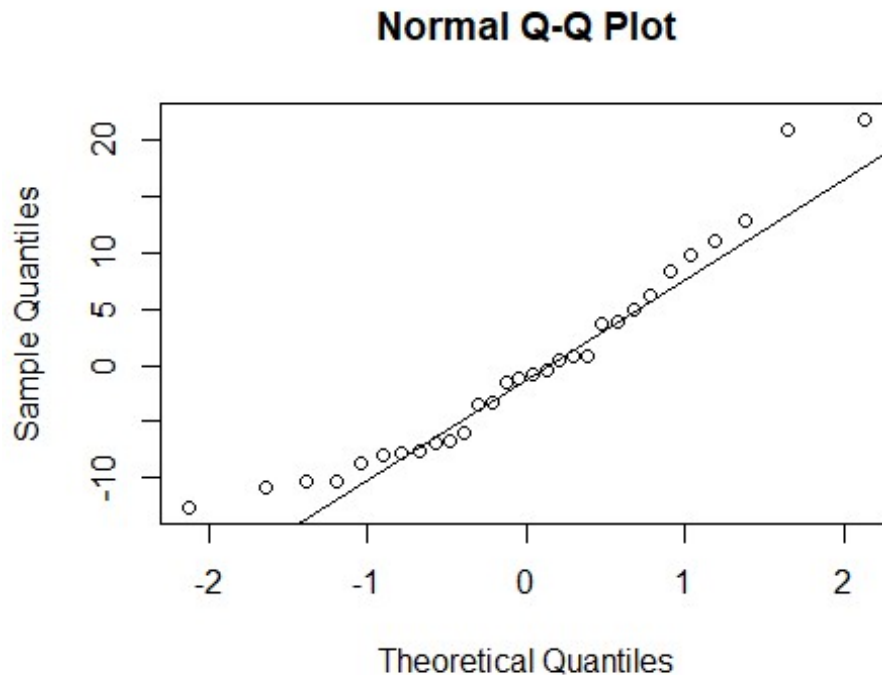
```
## 
##   Durbin-Watson test
## 
## data:  model.pine
## DW = 1.5514, p-value = 0.04921
## alternative hypothesis: true autocorrelation is greater than 0
```

#null hypothesis: true autocorrelation is equal to zero
#not significant - accept null hypothesis - assumption is true

#assumption 2: normality of residuals
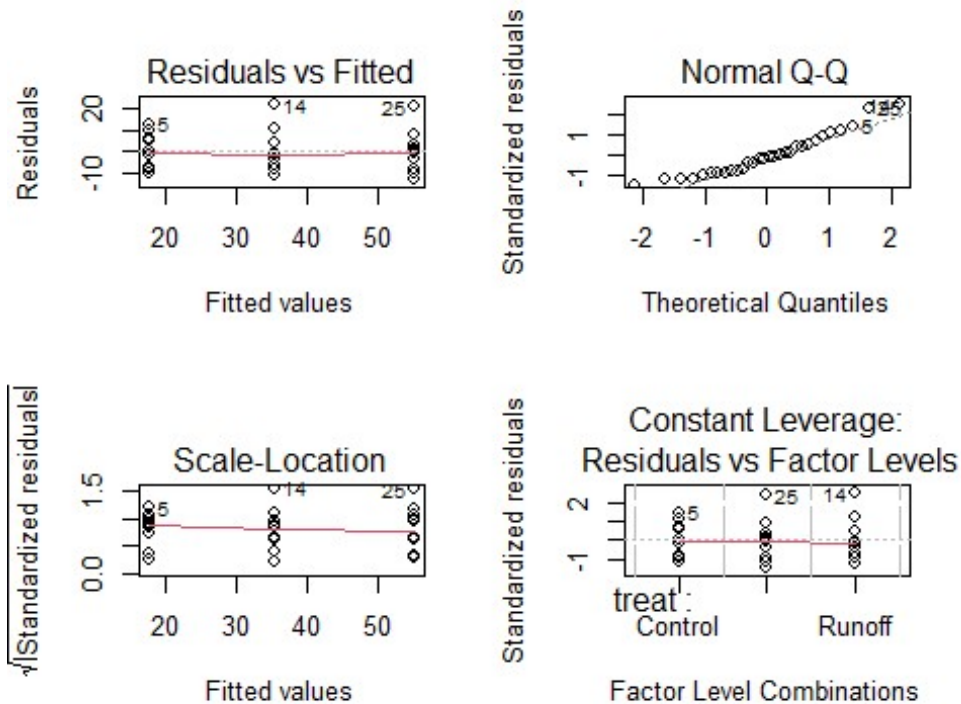
```
qqnorm(resid(model.pine))
qqline(resid(model.pine))
```

Normal Q-Q Plot


```
shapiro.test(resid(model.pine))
```

```
## 
##   Shapiro-Wilk normality test
## 
## data:  resid(model.pine)
## W = 0.93222, p-value = 0.05625
```

#null hypothesis: data is normally distributed
#data not significantly different from normal distribution - accept null
hypothesis - normality of residuals true

```
par(mfrow=c(2,2))
plot(model.pine)
```

Residuals vs Fitted

Normal Q-Q

Scale-Location

Constant Leverage:
Residuals vs Factor Levels

```
par(mfrow=c(1,1))
```

```
#assumption 3: homogeneity of variances

leveneTest(model.pine)

## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   2  0.0013 0.9987
##        27

#null hypothesis: all population variances are equal
#not significant - no significant differences between variances - homogeneity
of variances true
```

Part (d)

One way anova null hypothesis : No difference between the population means of different groups (grouping based on a factor, a categorical predictor variable with multiple levels) When the result is highly significant, we reject null hypothesis

The null hypotheses for the assumptions in part (c) are written with the tests