

```
library(utils)
library(readxl)
library(car)
```

```
## Warning: package 'car' was built under R version 4.0.4
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.0.3
```

```
## Registered S3 methods overwritten by 'tibble':
##   method      from
##   format.tbl  pillar
##   print.tbl   pillar
```

```
library(base)
library(graphics)
library(stats)
library(olsrr)
```

```
## Warning: package 'olsrr' was built under R version 4.0.5
```

```
##
## Attaching package: 'olsrr'
```

```
## The following object is masked from 'package:datasets':
##
##   rivers
```

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:olsrr':
##
##   cement
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v tibble  3.0.3    v dplyr    1.0.7
## v tidyr   1.1.2    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.5.0
## v purrr   0.3.4
```

```
## Warning: package 'dplyr' was built under R version 4.0.5

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
## x dplyr::recode() masks car::recode()
## x dplyr::select() masks MASS::select()
## x purrr::some() masks car::some()
```

```
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 4.0.5
```

```
library(rstatix)
```

```
## Warning: package 'rstatix' was built under R version 4.0.5
```

```
##
## Attaching package: 'rstatix'

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

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

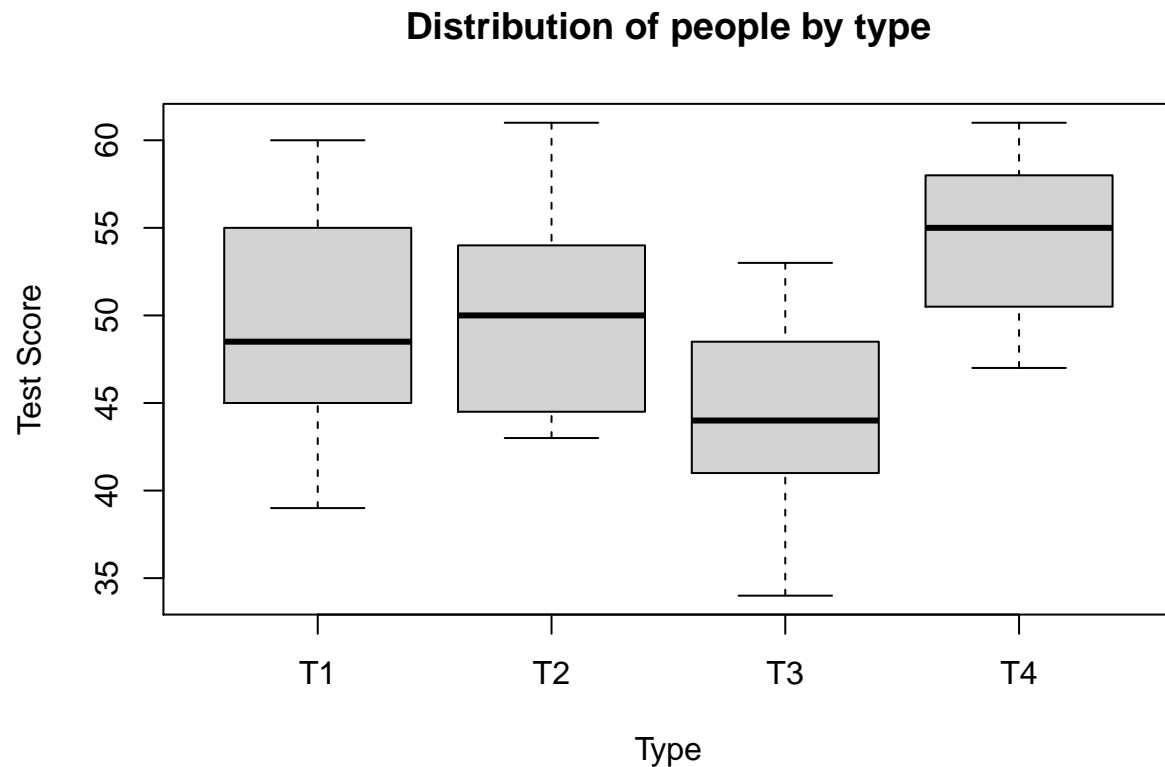
```
# Q1
score<-c(44, 46, 59, 48, 49, 60, 51, 39,
        43, 45, 57, 51, 51, 49, 44, 61,
        51, 34, 53, 45, 41, 46, 41, 43,
        52, 47, 59, 56, 49, 61, 57, 54)
type<-rep(c("T1","T2","T3","T4"), each = 8)
tapply(score, type, mean)
```

```
##      T1      T2      T3      T4
## 49.500 50.125 44.250 54.375
```

```
type.ANOVA<-aov(score~factor(type))
summary(type.ANOVA)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## factor(type)  3  413.6   137.87   3.648 0.0245 *
## Residuals    28 1058.3    37.79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
boxplot(score~factor(type),
        main = "Distribution of people by type",
        xlab = "Type",
        ylab = "Test Score")
```

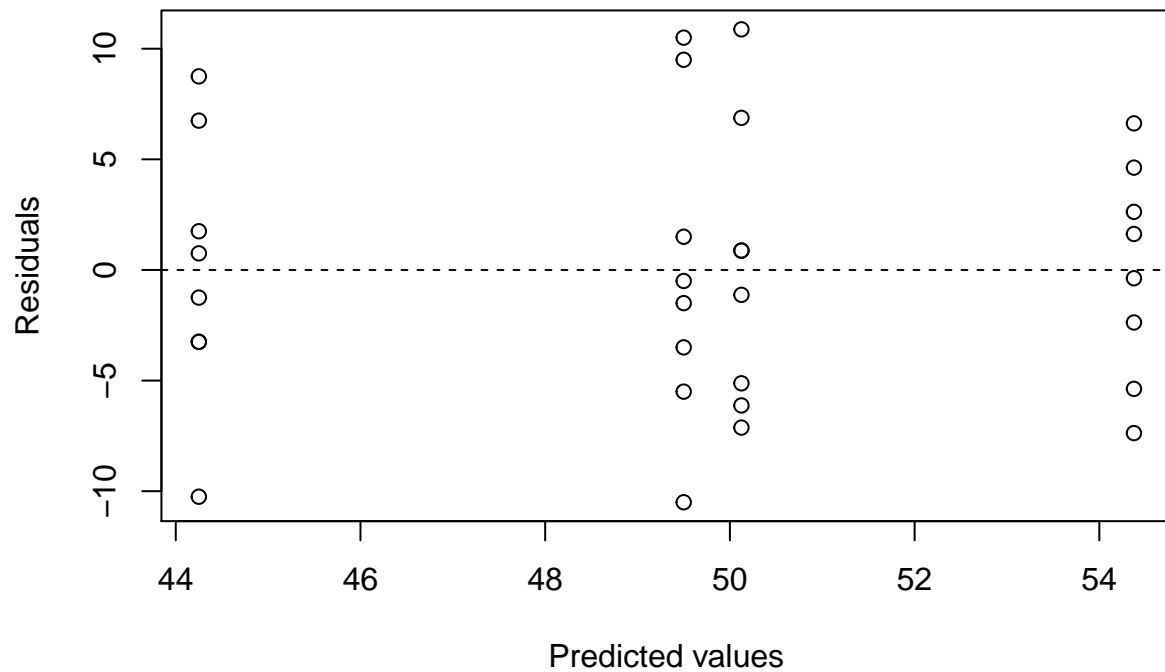


```
leveneTest(score~factor(type), center = "mean")
```

```
## Levene's Test for Homogeneity of Variance (center = "mean")
##      Df F value Pr(>F)
## group 3  0.2523  0.859
##      28
```

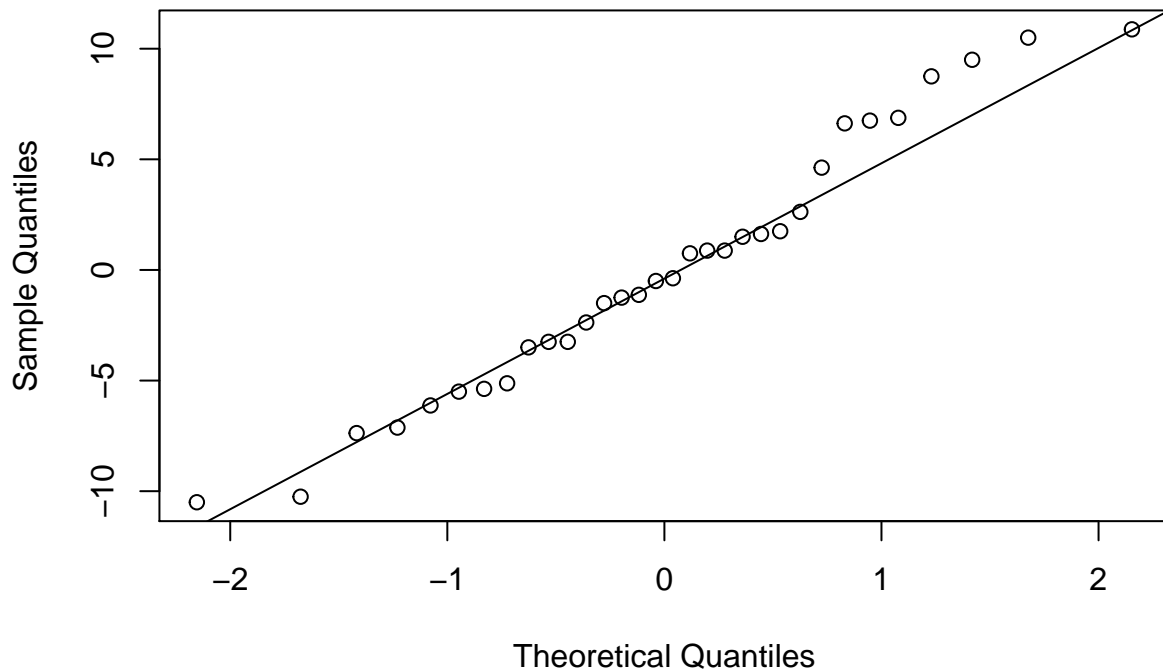
```
plot(x = type.ANOVA$fitted.values, y = type.ANOVA$residuals,
     main = "Residuals vs. fitted values\n ANOVA on type effects data",
     xlab = "Predicted values", ylab = "Residuals")
abline(h = 0, lty = 2)
```

Residuals vs. fitted values ANOVA on type effects data



```
qqnorm(type.ANOVA$residuals,  
        main = "Normal Q-Q plot of residuals\n ANOVA on type effects data")  
qqline(type.ANOVA$residuals)
```

Normal Q-Q plot of residuals ANOVA on type effects data



```
#Q2
zinc<-c( 6340, 4280, 5170, 2880, 4330, 3050,
        3690, 4750, 5100, 2360, 1990, 2140,
        250,  470,  330,  400,  310,  430,
        2850, 2380, 3130, 1070, 960, 1300)
pH<-rep(c("pH 5.5(acid)", "pH7(neutral)"), each = 3, time=4)
plant<-rep(c("Alpine pennycress", "Bladder campion", "Lettuce", "Martin red fescue"), each = 6)

cbind(pH, plant, zinc)
```

```
##      pH      plant      zinc
## [1,] "pH 5.5(acid)" "Alpine pennycress" "6340"
## [2,] "pH 5.5(acid)" "Alpine pennycress" "4280"
## [3,] "pH 5.5(acid)" "Alpine pennycress" "5170"
## [4,] "pH7(neutral)" "Alpine pennycress" "2880"
## [5,] "pH7(neutral)" "Alpine pennycress" "4330"
## [6,] "pH7(neutral)" "Alpine pennycress" "3050"
## [7,] "pH 5.5(acid)" "Bladder campion"   "3690"
## [8,] "pH 5.5(acid)" "Bladder campion"   "4750"
## [9,] "pH 5.5(acid)" "Bladder campion"   "5100"
## [10,] "pH7(neutral)" "Bladder campion"   "2360"
## [11,] "pH7(neutral)" "Bladder campion"   "1990"
## [12,] "pH7(neutral)" "Bladder campion"   "2140"
## [13,] "pH 5.5(acid)" "Lettuce"           "250"
## [14,] "pH 5.5(acid)" "Lettuce"           "470"
## [15,] "pH 5.5(acid)" "Lettuce"           "330"
```

```
## [16,] "pH7(neutral)" "Lettuce"          "400"
## [17,] "pH7(neutral)" "Lettuce"          "310"
## [18,] "pH7(neutral)" "Lettuce"          "430"
## [19,] "pH 5.5(acid)" "Martin red fescue" "2850"
## [20,] "pH 5.5(acid)" "Martin red fescue" "2380"
## [21,] "pH 5.5(acid)" "Martin red fescue" "3130"
## [22,] "pH7(neutral)" "Martin red fescue" "1070"
## [23,] "pH7(neutral)" "Martin red fescue" "960"
## [24,] "pH7(neutral)" "Martin red fescue" "1300"
```

```
##### raw data #####
```

```
zinc.ANOVA<-aov(zinc~factor(pH)*factor(plant))
summary(zinc.ANOVA)
```

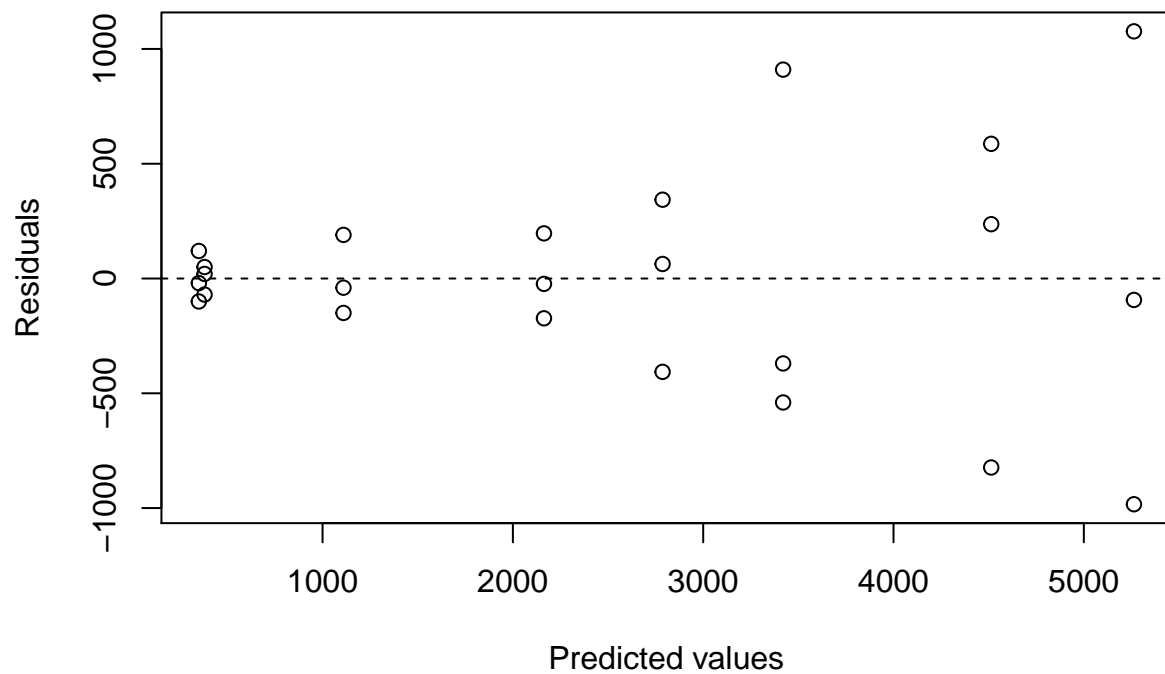
```
##              Df    Sum Sq  Mean Sq F value    Pr(>F)
## factor(pH)      1 12789600 12789600  41.602 8.02e-06 ***
## factor(plant)    3  53742533 17914178  58.271 7.85e-09 ***
## factor(pH):factor(plant) 3  4809133  1603044   5.214  0.0106 *
## Residuals      16  4918867    307429
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#boxplot(zinc~factor(plant):factor(pH), main = "Boxplot of zinc")
```

```
#leveneTest(zinc~factor(plant)*factor(pH), center = "mean")
```

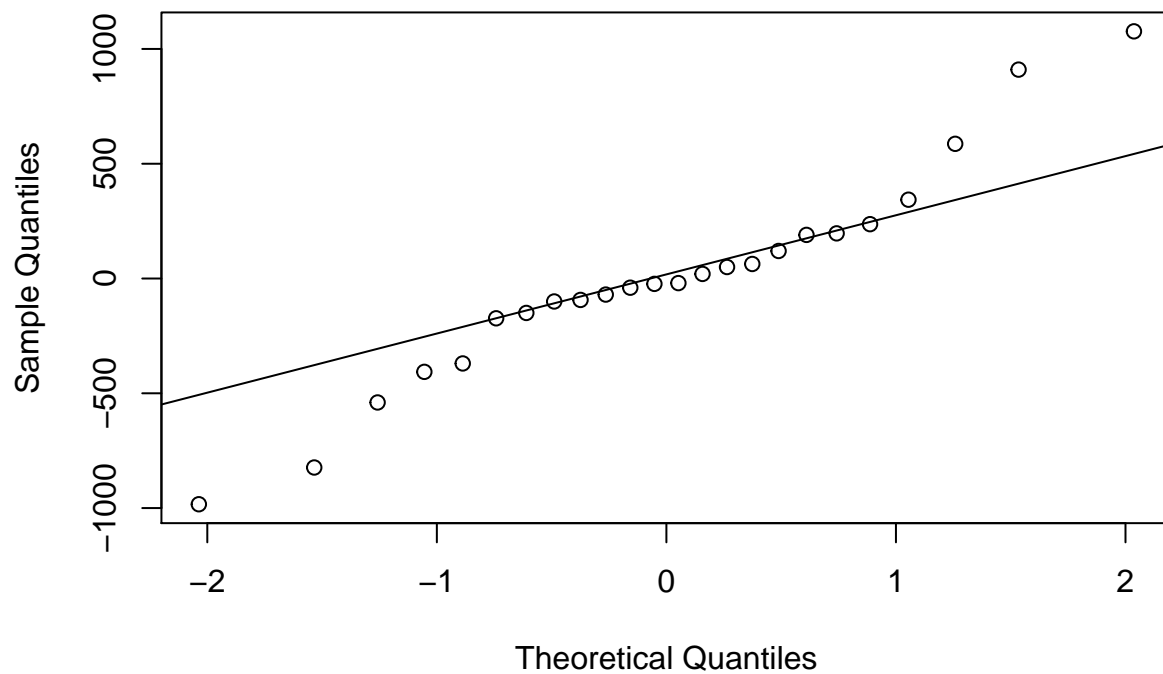
```
plot(x = zinc.ANOVA$fitted.values, y = zinc.ANOVA$residuals,
     main = "Residuals vs. fitted values\n ANOVA on uptake of zinc data",
     xlab = "Predicted values", ylab = "Residuals")
abline(h = 0, lty = 2)
```

Residuals vs. fitted values ANOVA on uptake of zinc data



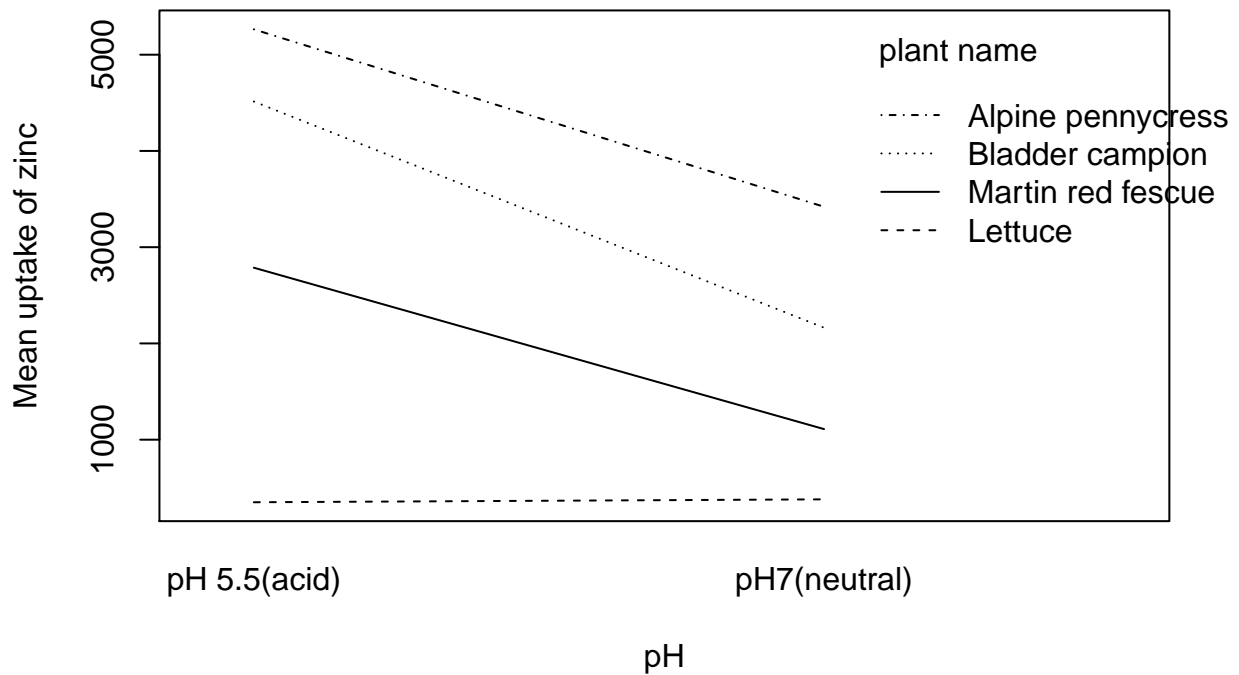
```
qqnorm(zinc.ANOVA$residuals,  
       main = "Normal Q-Q plot of residuals\n ANOVA on uptake of zinc data")  
qqline(zinc.ANOVA$residuals)
```

Normal Q-Q plot of residuals ANOVA on uptake of zinc data



```
interaction.plot(x.factor = pH,  
                trace.factor = plant,  
                response = zinc,  
                fun = mean,  
                xlab = "pH",  
                ylab = "Mean uptake of zinc",  
                main = "Interaction graph for four species of plant",  
                legend = TRUE, xpd=TRUE,  
                trace.label="plant name")
```


Interaction graph for four species of plant



```
#####log -transformed data #####
```

```
zinc.log.ANOVA<-aov(log(zinc)~factor(pH)*factor(plant))
summary(zinc.log.ANOVA)
```

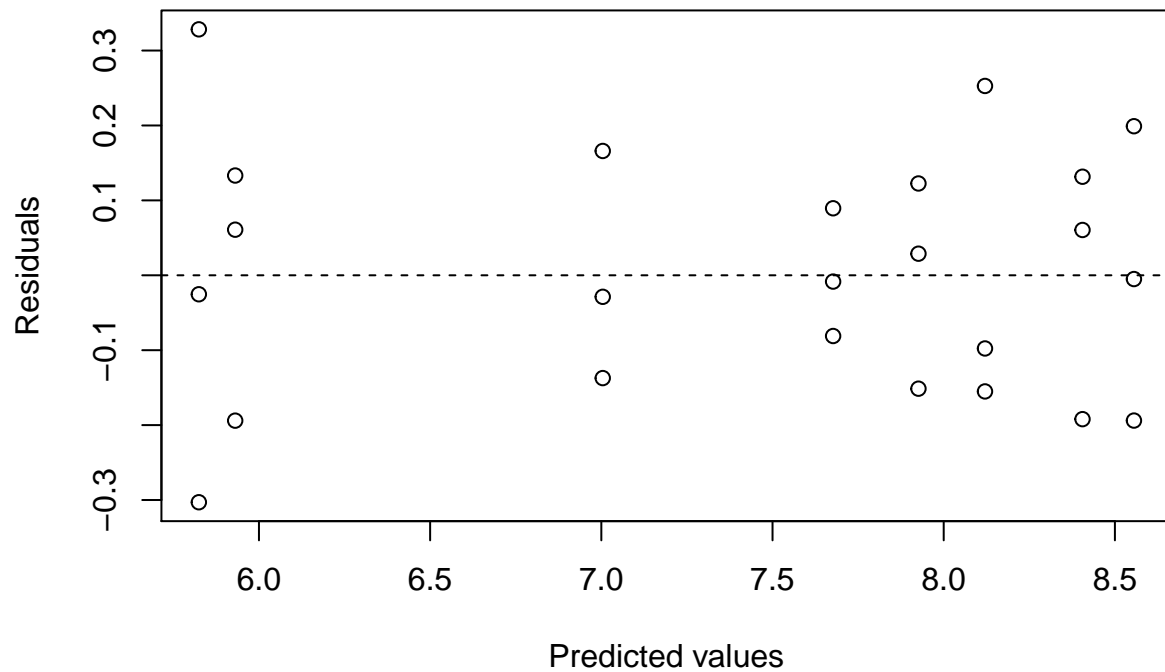
```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## factor(pH)      1  1.469    1.469  39.681 1.06e-05 ***
## factor(plant)    3 21.658    7.219 194.968 8.33e-13 ***
## factor(pH):factor(plant) 3  0.903    0.301   8.128 0.00163 **
## Residuals      16  0.592    0.037
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#boxplot(log(zinc)~factor(plant):factor(pH), main = "Boxplot of logged zinc data")
```

```
#leveneTest(log(zinc)~factor(plant)*factor(pH), center = "mean")
```

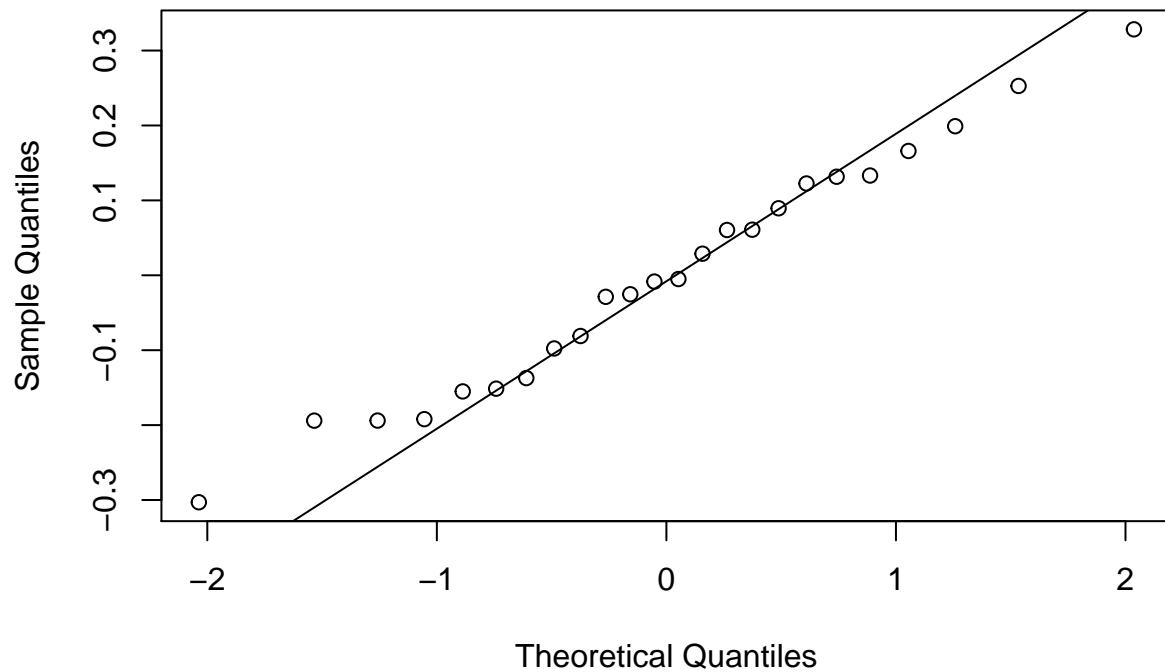
```
plot(x = zinc.log.ANOVA$fitted.values, y = zinc.log.ANOVA$residuals,
     main = "Residuals vs. fitted values\n ANOVA on uptake of logged zinc data",
     xlab = "Predicted values", ylab = "Residuals")
abline(h = 0, lty = 2)
```

Residuals vs. fitted values
ANOVA on uptake of logged zinc data



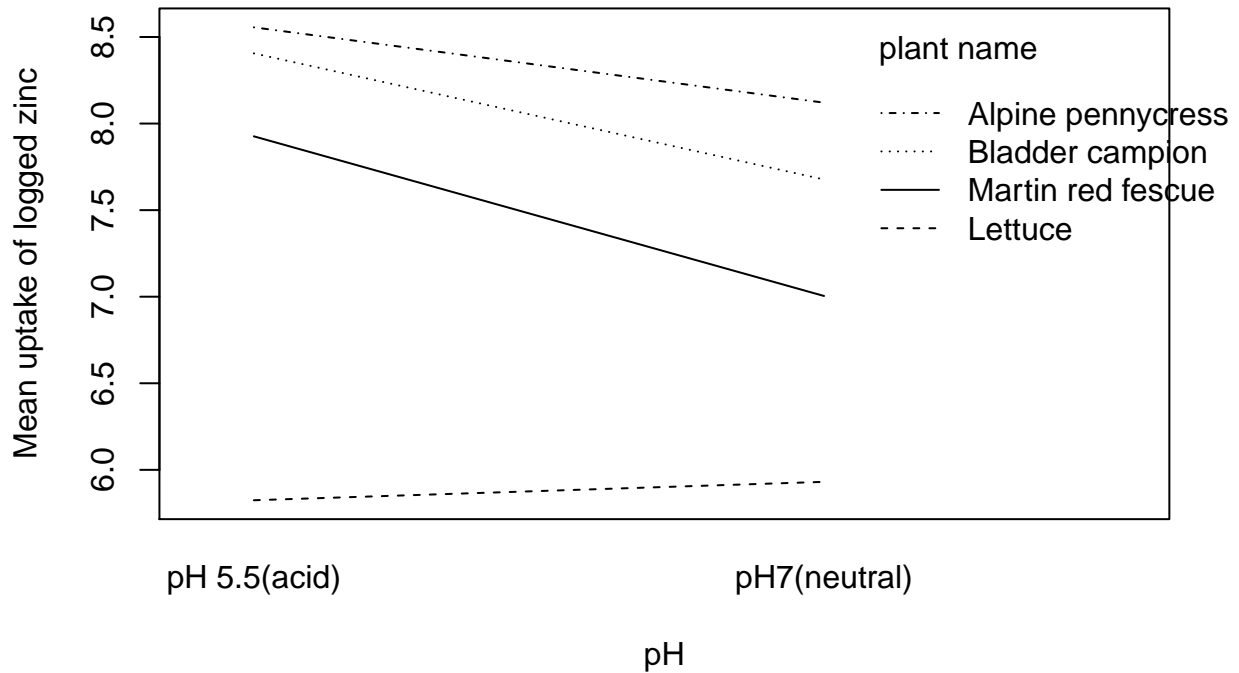
```
qqnorm(zinc.log.ANOVA$residuals,  
       main = "Normal Q-Q plot of residuals\n ANOVA on uptake of logged zinc data")  
qqline(zinc.log.ANOVA$residuals)
```

Normal Q–Q plot of residuals
ANOVA on uptake of logged zinc data



```
interaction.plot(x.factor = pH,  
                trace.factor = plant,  
                response = log(zinc),  
                fun = mean,  
                xlab = "pH",  
                ylab = "Mean uptake of logged zinc",  
                main = "Interaction graph for logged zinc data",  
                legend = TRUE, xpd=TRUE,  
                trace.label="plant name")
```

Interaction graph for logged zinc data



```
Marks <- c(67, 66, 75, 76, 71, 70, 72, 63, 72, 62, 61, 69, 64, 71, 68, 56,
          69, 57, 55, 63, 65, 55, 59, 47, 49,
          30, 47, 39, 33)
Ethnity <- rep(c("E1", "E2", "E3"), c(16, 9, 4))
Sex <- rep(c(rep(c("F", "M"), c(8, 8)),
             rep(c("F", "M"), c(6, 3)),
             rep(c("F", "M"), c(2, 2))))
cbind(Ethnity, Sex, Marks)
```

```
##      Ethnity Sex Marks
## [1,] "E1"    "F"  "67"
## [2,] "E1"    "F"  "66"
## [3,] "E1"    "F"  "75"
## [4,] "E1"    "F"  "76"
## [5,] "E1"    "F"  "71"
## [6,] "E1"    "F"  "70"
## [7,] "E1"    "F"  "72"
## [8,] "E1"    "F"  "63"
## [9,] "E1"    "M"  "72"
## [10,] "E1"   "M"  "62"
## [11,] "E1"   "M"  "61"
## [12,] "E1"   "M"  "69"
## [13,] "E1"   "M"  "64"
## [14,] "E1"   "M"  "71"
## [15,] "E1"   "M"  "68"
```

```
## [16,] "E1"      "M" "56"
## [17,] "E2"      "F" "69"
## [18,] "E2"      "F" "57"
## [19,] "E2"      "F" "55"
## [20,] "E2"      "F" "63"
## [21,] "E2"      "F" "65"
## [22,] "E2"      "F" "55"
## [23,] "E2"      "M" "59"
## [24,] "E2"      "M" "47"
## [25,] "E2"      "M" "49"
## [26,] "E3"      "F" "30"
## [27,] "E3"      "F" "47"
## [28,] "E3"      "M" "39"
## [29,] "E3"      "M" "33"
```

#Q3

```
mark_data<- read_xlsx("question3.xlsx", sheet = "Data")
head(mark_data)
```

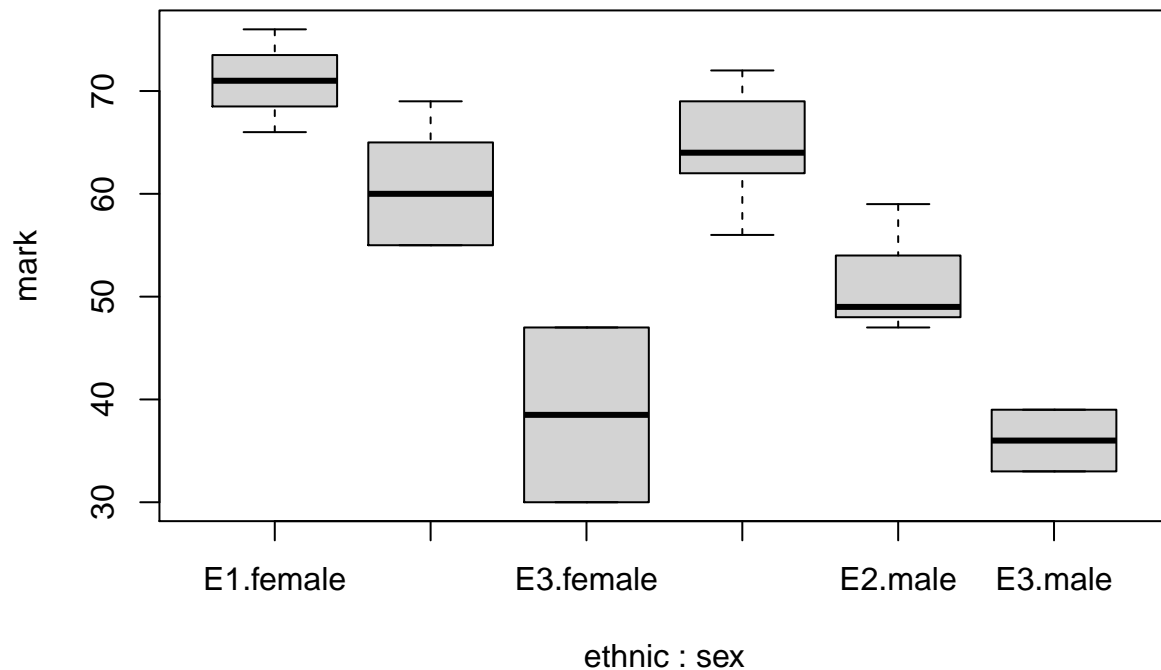
```
## # A tibble: 6 x 3
##   mark sex    ethnic
##   <dbl> <chr>  <chr>
## 1    67 female E1
## 2    66 female E1
## 3    75 female E1
## 4    76 female E1
## 5    71 female E1
## 6    70 female E1
```

```
mark<-mark_data$mark
sex<-as.factor(mark_data$sex)
ethnic<-as.factor(mark_data$ethnic)
```

```
#mark.ANOVA<-aov(mark~sex*ethnic, data =mark_data)
```

```
boxplot(mark~ethnic*sex, main = "Boxplot of test marks")
```

Boxplot of test marks



```
mark.ANOVA<-aov(mark~ethnic*sex, data =mark_data)
summary(mark.ANOVA)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## ethnic      2 3060.6  1530.3   48.791 5.31e-09 ***
## sex         1  275.1   275.1    8.771 0.00699 **
## ethnic:sex   2    29.7    14.8    0.473 0.62892
## Residuals   23  721.4    31.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

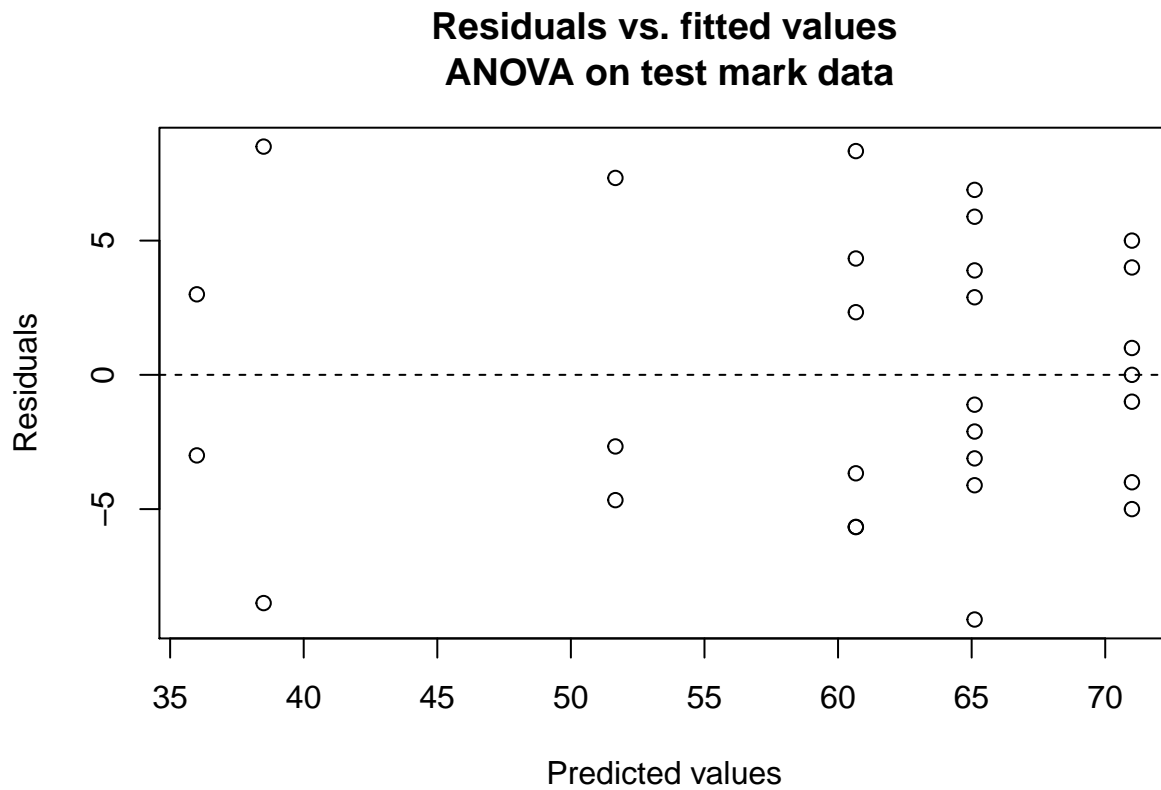
```
TukeyHSD(mark.ANOVA, which = "sex")
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = mark ~ ethnic * sex, data = mark_data)
##
## $sex
##           diff       lwr       upr      p adj
## male-female -6.033036 -10.33829 -1.727782 0.0080933
```

```
leveneTest(mark~ethnic*sex, center = "mean", data =mark_data)
```

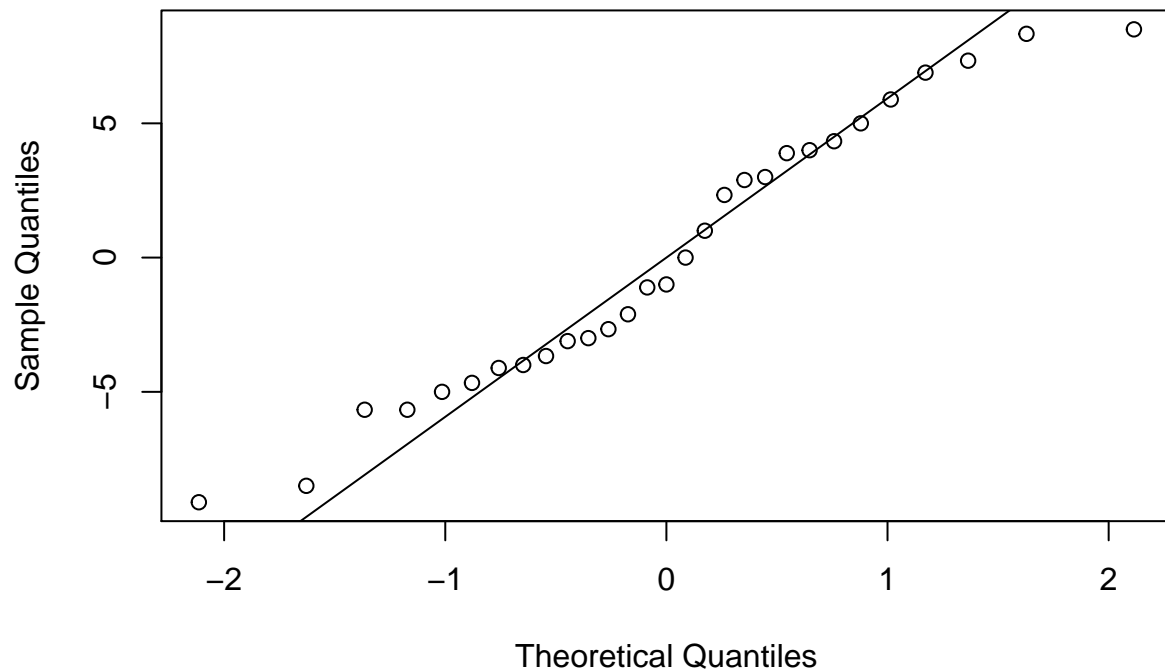
```
## Levene's Test for Homogeneity of Variance (center = "mean")
##      Df F value Pr(>F)
## group 5  2.3854 0.0698 .
##      23
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(x = mark.ANOVA$fitted.values, y = mark.ANOVA$residuals,
     main = "Residuals vs. fitted values\n ANOVA on test mark data",
     xlab = "Predicted values", ylab = "Residuals")
abline(h = 0, lty = 2)
```



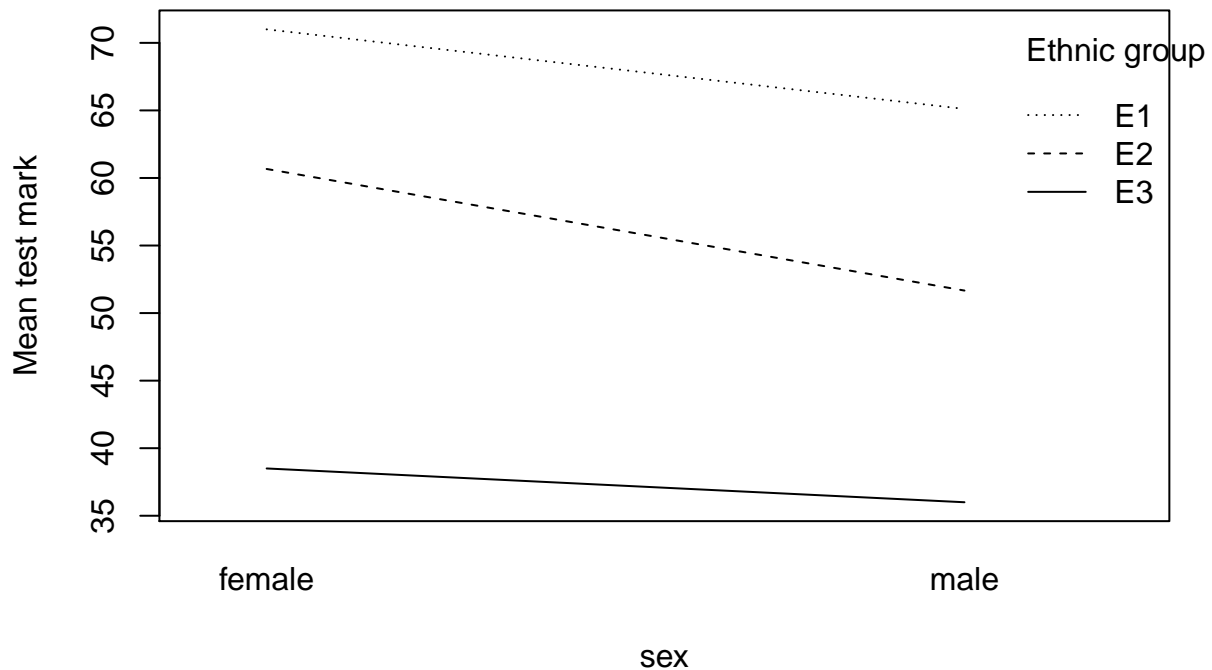
```
qqnorm(mark.ANOVA$residuals,
       main = "Normal Q-Q plot of residuals\n ANOVA on test mark data")
qqline(mark.ANOVA$residuals)
```

Normal Q-Q plot of residuals ANOVA on test mark data



```
interaction.plot(x.factor = sex,  
                 trace.factor = ethnic,  
                 response = mark,  
                 fun = mean,  
                 xlab = "sex",  
                 ylab = "Mean test mark",  
                 main = "Interaction graph for test mark data",  
                 legend = TRUE, xpd=TRUE,  
                 trace.label="Ethnic group")
```


Interaction graph for test mark data



#####Q4

```
area<-c(516,469.06,462.25,938.6,1357.15,1773.66,1686.01,1786.29,3090.07,3980.12,4424.84,4451.68,4982.89,
        4450.86,5490.74,7476.21,7138.82,9149.94,10133.07,9287.69,13729.13,20300.77,24712.72,27144.03,26
species<-c(3,7,6,8,10,9,10,11,16,9,13,14,12,
          14,20,22,15,20,22,21,15,24,25,25,24)
head(cbind(area,species))
```

```
##      area species
## [1,]  516.00      3
## [2,]  469.06      7
## [3,]  462.25      6
## [4,]  938.60      8
## [5,] 1357.15     10
## [6,] 1773.66      9
```

raw data

```
# x = Area, y = number of different species
macro.lm<-lm(species~area)
anova(macro.lm)
```

```
## Analysis of Variance Table
##
## Response: species
##      Df Sum Sq Mean Sq F value    Pr(>F)
## area   1  712.94    712.94   50.444 3.102e-07 ***
```

```
## Residuals 23 325.06 14.13
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

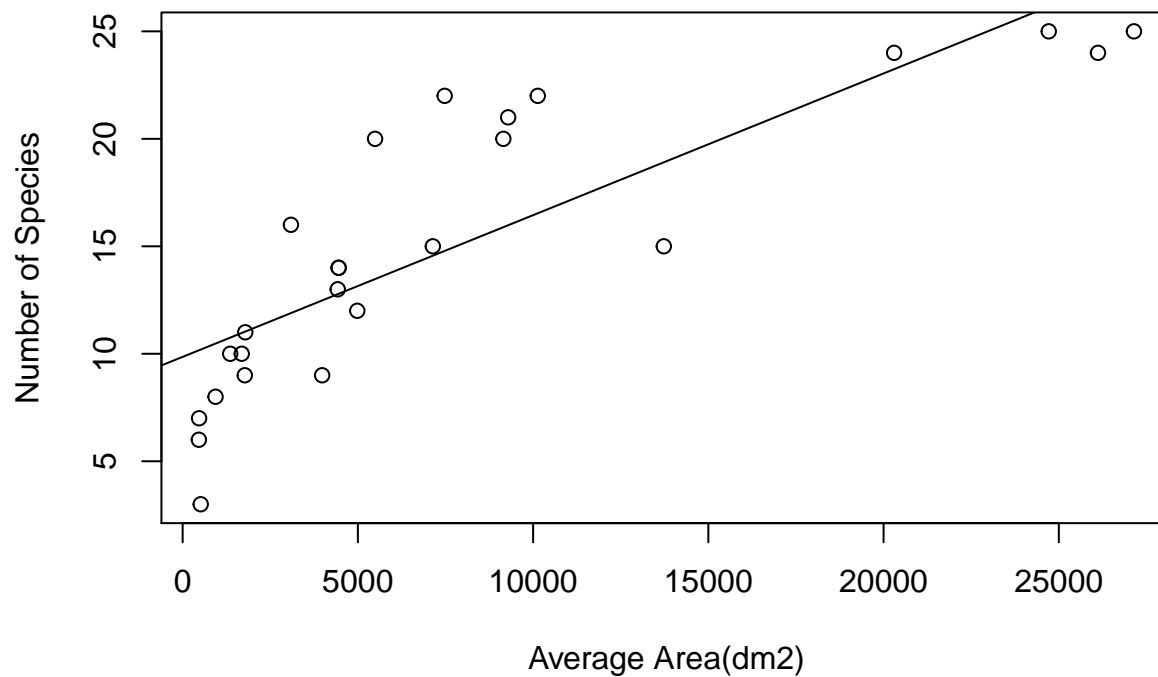
```
summary(macro.lm)
```

```
##
## Call:
## lm(formula = species ~ area)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.1964 -2.7521 -0.7509  1.2094  7.2148
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.856e+00  1.044e+00   9.441 2.24e-09 ***
## area         6.593e-04  9.283e-05   7.102 3.10e-07 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.759 on 23 degrees of freedom
## Multiple R-squared:  0.6868, Adjusted R-squared:  0.6732
## F-statistic: 50.44 on 1 and 23 DF, p-value: 3.102e-07
```

```
# Produce a scatterplot of area vs. species
```

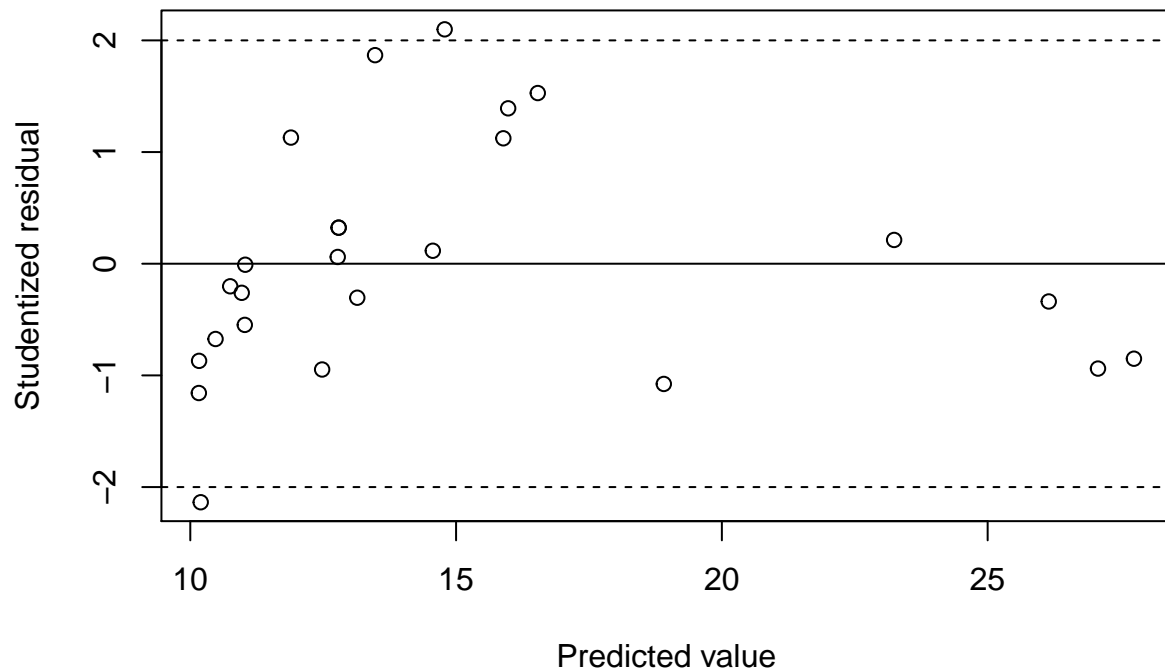
```
plot(x=area, y=species,
     main="Scatterplot of the mussel clumps area vs. species of\n macroinvertebrates found with fitted",
     xlab = "Average Area(dm2)", ylab = "Number of Species")
abline(macro.lm)
```

Scatterplot of the mussel clumps area vs. species of macroinvertebrates found with fitted regression line



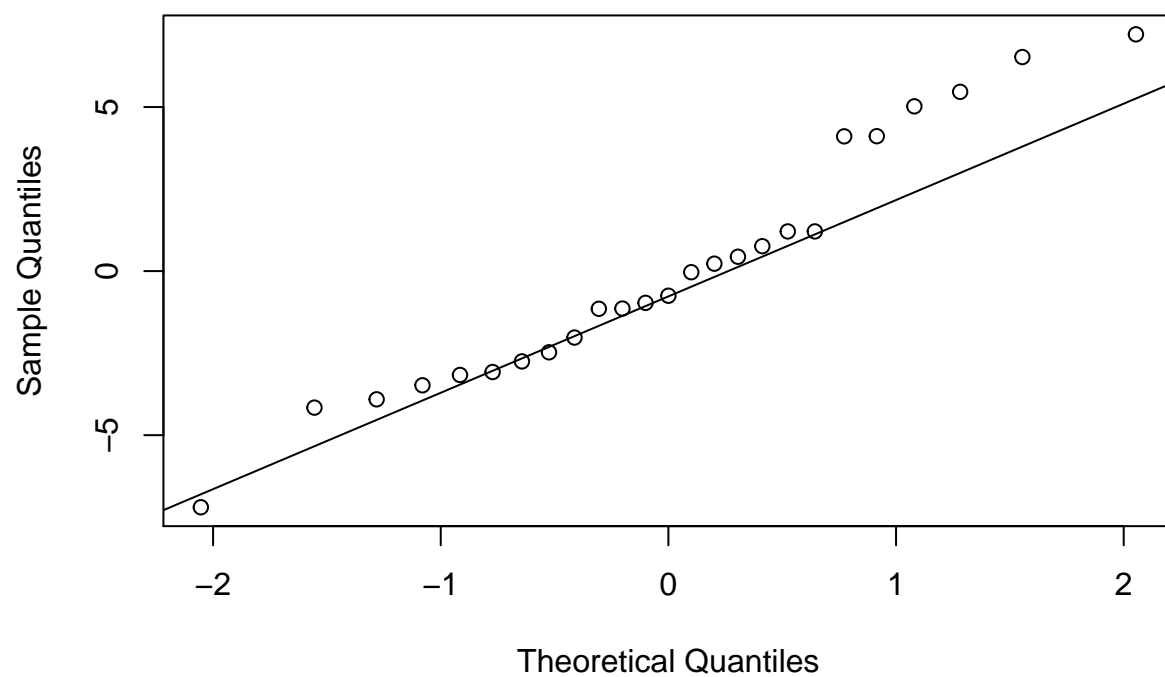
```
# Produce a scatterplot of studentized residuals versus fitted values
plot(x=macro.lm$fitted.values, y = rstudent(macro.lm),
     main="Studentized residuals vs. fitted values\n regression for macroinvertebrates data",
     xlab = "Predicted value", ylab = "Studentized residual")
abline(h = 0)
abline(h = c(-2, 2), lty = 2)
```

Studentized residuals vs. fitted values regression for macroinvertebrates data

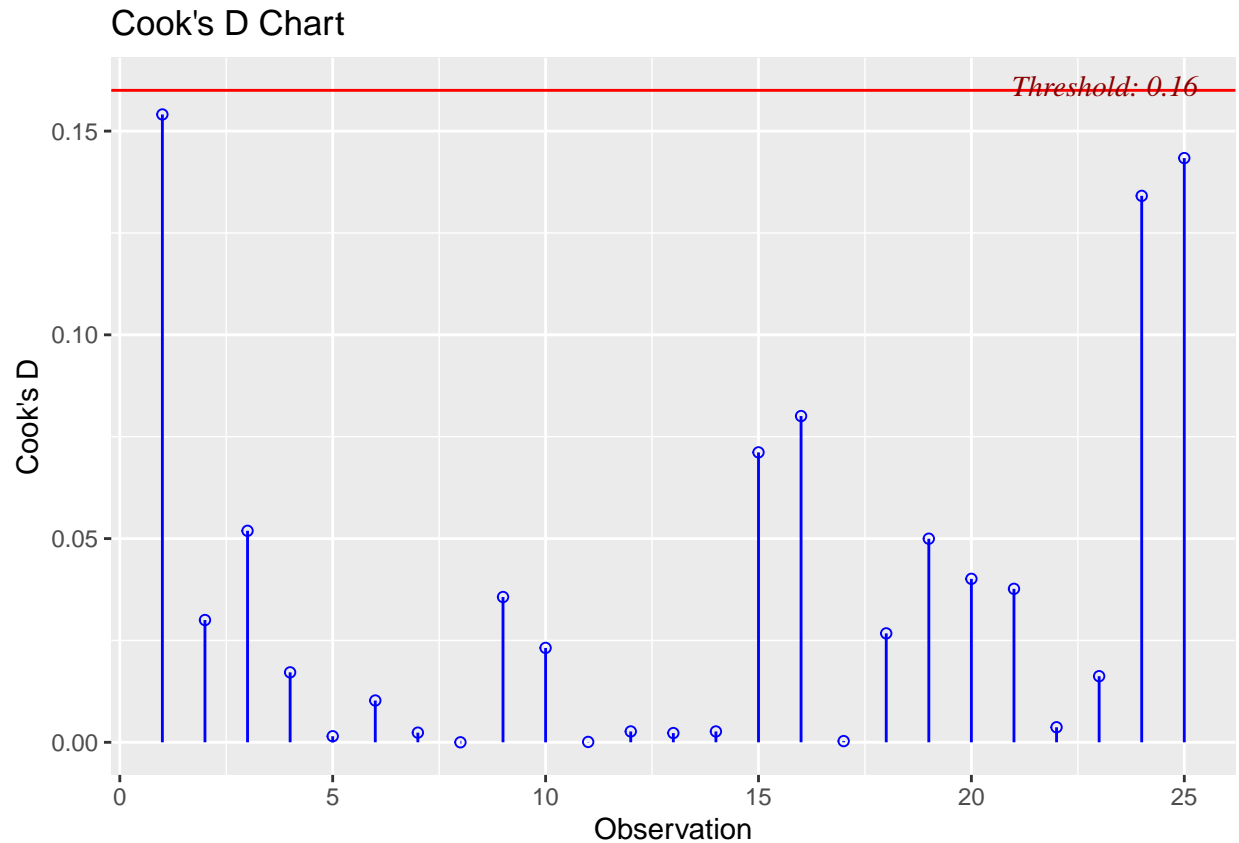


```
# Produce a normal Q-Q plot of residuals  
qqnorm(macro.lm$residuals,  
main = "Normal Q-Q plot of residuals\n regression for macroinvertebrates data")  
qqline(macro.lm$residuals)
```

**Normal Q-Q plot of residuals
regression for macroinvertebrates data**



```
ols_plot_cooksd_chart(macro.lm)
```



```
#####
##### logged data #####
macro.lm2<-lm(species~log(area))
anova(macro.lm2)
```

```
## Analysis of Variance Table
##
## Response: species
##           Df Sum Sq Mean Sq F value    Pr(>F)
## log(area)  1 869.52   869.52   118.7 1.481e-10 ***
## Residuals 23  168.48     7.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

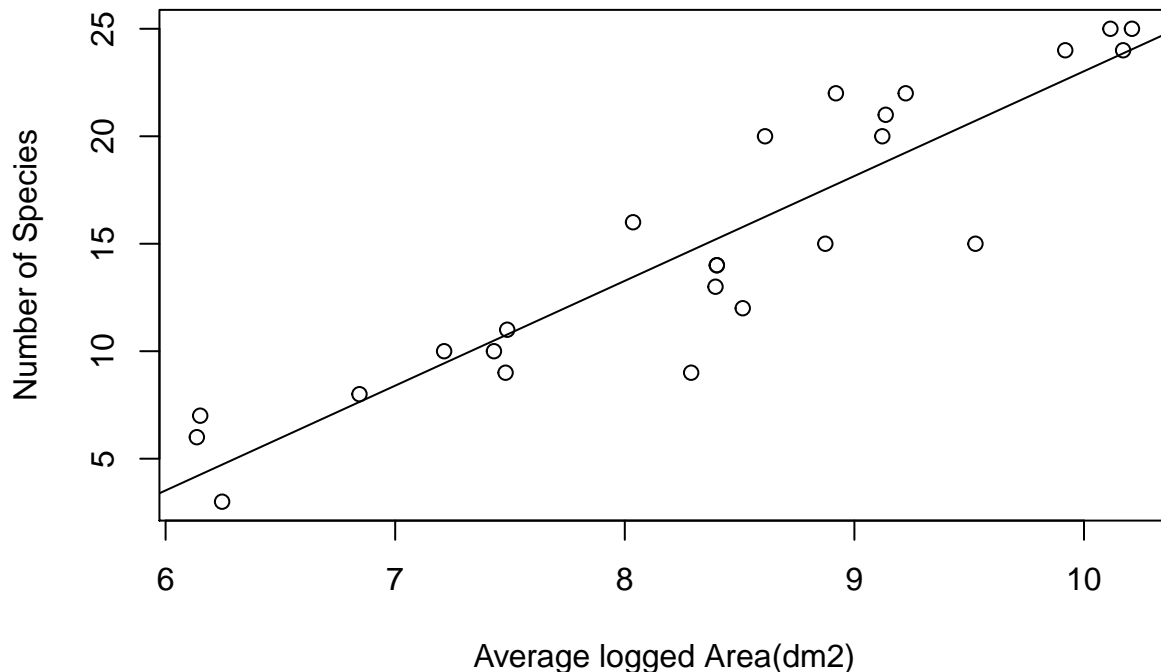
```
summary(macro.lm2)
```

```
##
## Call:
## lm(formula = species ~ log(area))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.7204 -1.7227  0.3603  1.8136  4.2430
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -25.7308      3.7774  -6.812 6.02e-07 ***
## log(area)    4.8756      0.4475  10.895 1.48e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.707 on 23 degrees of freedom
## Multiple R-squared:  0.8377, Adjusted R-squared:  0.8306
## F-statistic: 118.7 on 1 and 23 DF,  p-value: 1.481e-10
```

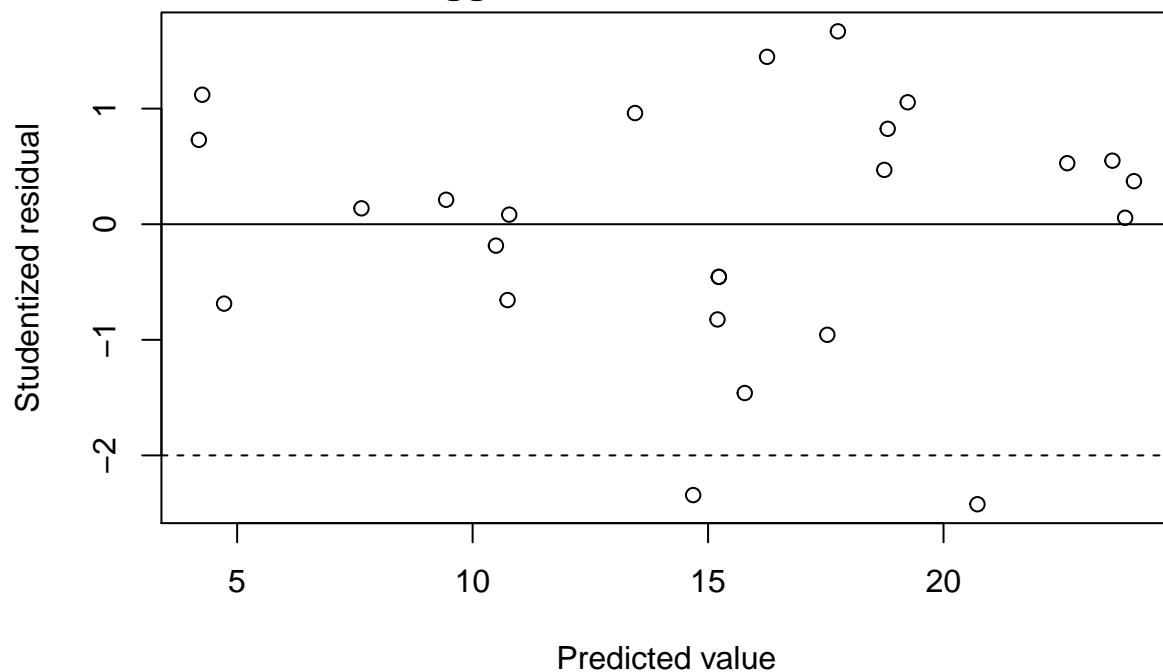
```
# Produce a scatterplot of area vs. species
plot(x=log(area), y=species,
     main="Scatterplot of the mussel clumps area vs. species of\n
          macroinvertebrates found with fitted regression line",
     xlab = "Average logged Area(dm2)", ylab = "Number of Species")
abline(macro.lm2)
```

**Scatterplot of the mussel clumps area vs. species of
macroinvertebrates found with fitted regression line**



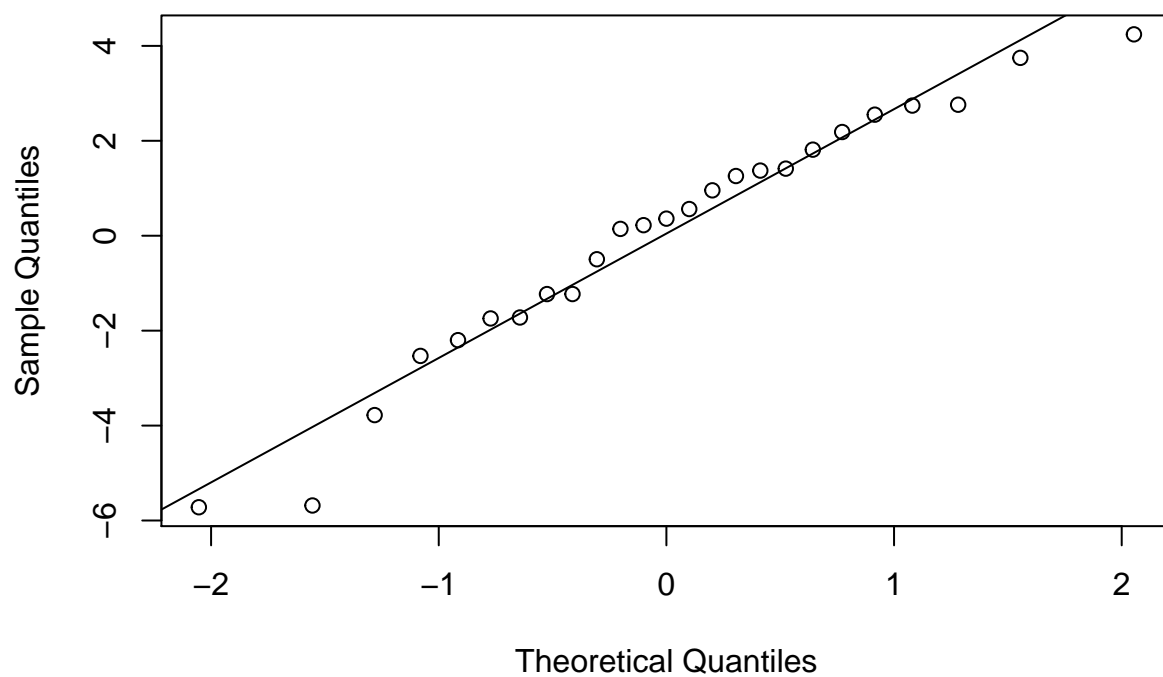
```
# Produce a scatterplot of studentized residuals versus fitted values
plot(x=macro.lm2$fitted.values, y = rstudent(macro.lm2),
     main="Studentized residuals vs. fitted values\n
          regression for\n
          logged macroinvertebrates data",
     xlab = "Predicted value", ylab = "Studentized residual")
abline(h = 0)
abline(h = c(-2, 2), lty = 2)
```

**Studentized residuals vs. fitted values
regression for
logged macroinvertebrates data**

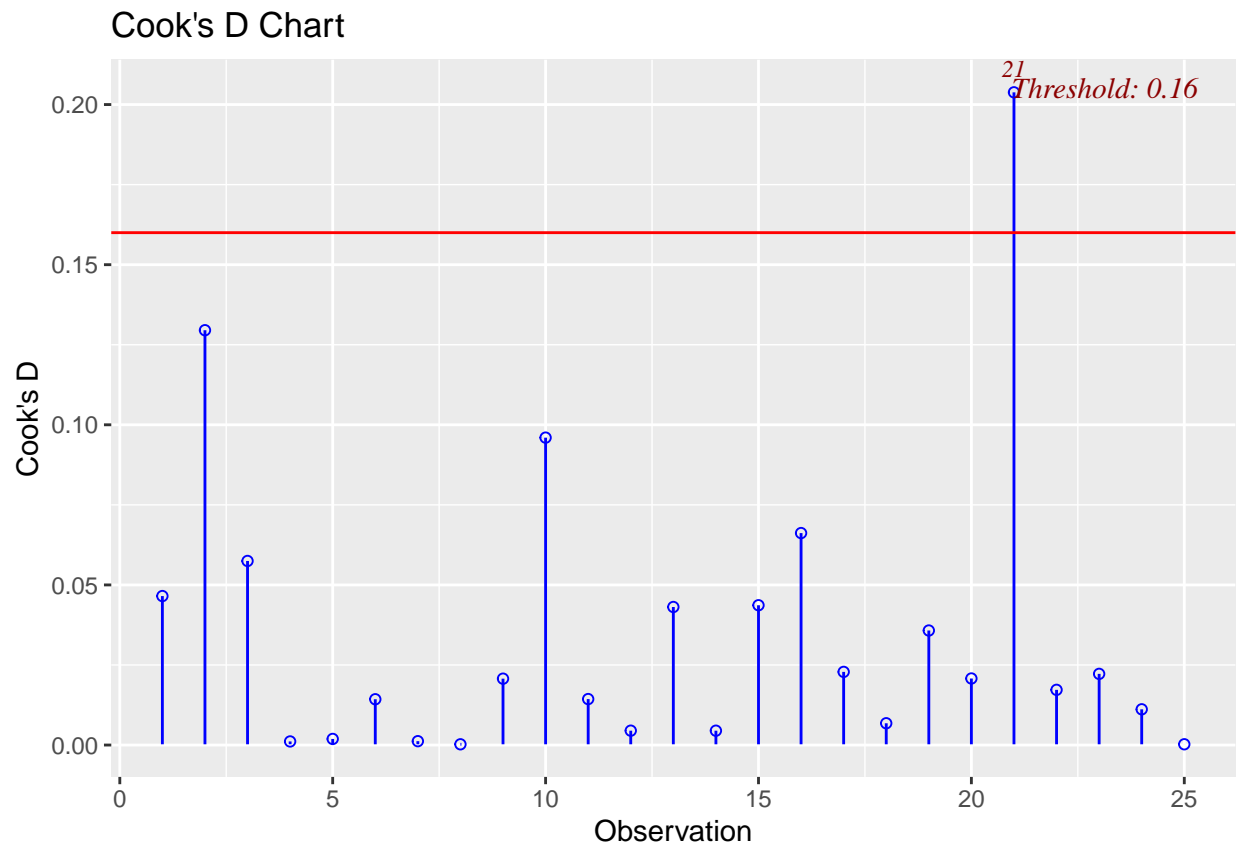


```
# Produce a normal Q-Q plot of residuals
qqnorm(macro.lm2$residuals,
main = "Normal Q-Q plot of residuals\n regression for logged macroinvertebrates data")
qqline(macro.lm2$residuals)
```


**Normal Q–Q plot of residuals
regression for logged macroinvertebrates data**



```
# Plot Cook's distances  
ols_plot_cooksd_chart(macro.lm2)
```



```
# Obtain 95% confidence intervals for beta_0 and beta_1.  
confint.default(macro.lm2)
```

```
##           2.5 %    97.5 %  
## (Intercept) -33.134401 -18.327116  
## log(area)    3.998505  5.752689
```