```r
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
```

```r
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
```

```r
library(MASS)
```

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

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

```r
library(ggplot2)
```

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

```r
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------------------------------------
```

```
## 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 ---------------------------------------------------------------
## 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()
```

```r
library(ggpubr)
```

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

```r
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
```

```r
# unbalanced data entry
Development_time <- c(78,91,97,82,85,77,
                      75,93,78,71,63,76,69,
                      64,72,68,77,56,95,71,
                      55,66,49,64,70,68)
Temperature <- rep(c(12,16,20,24), c(6,7,7,6))
result <- 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)
ethnic <- rep(c("E1","E2","E3"), c(16,9,4))
sex<-rep(c(rep(c("femal","male"),c(7,9)),
          rep(c("femal","male"),c(6,3)),
          rep(c("femal","male"),c(2,2))))
```

ONE-WAY ANOVA and Tukeytest:categorical variables (factors)

```r
# csv
areas1 <- read.csv("areasLess3.csv")
#xlsx
areas.less6<-read_xlsx("areasLess6.xlsx", sheet = "Data")
tapply(areas.less6$Size, areas.less6$Shape, mean)
```

```
##    circle rectangle    square  triangle
##  19.34667  20.64000  16.11452  20.75806
```

```r
boxplot(Size ~ factor(Shape), main = "Distribution of Size: Areas data without 6 'outliers'", xlab = "Sl
```

**Distribution of Size: Areas data without 6 'outliers'**
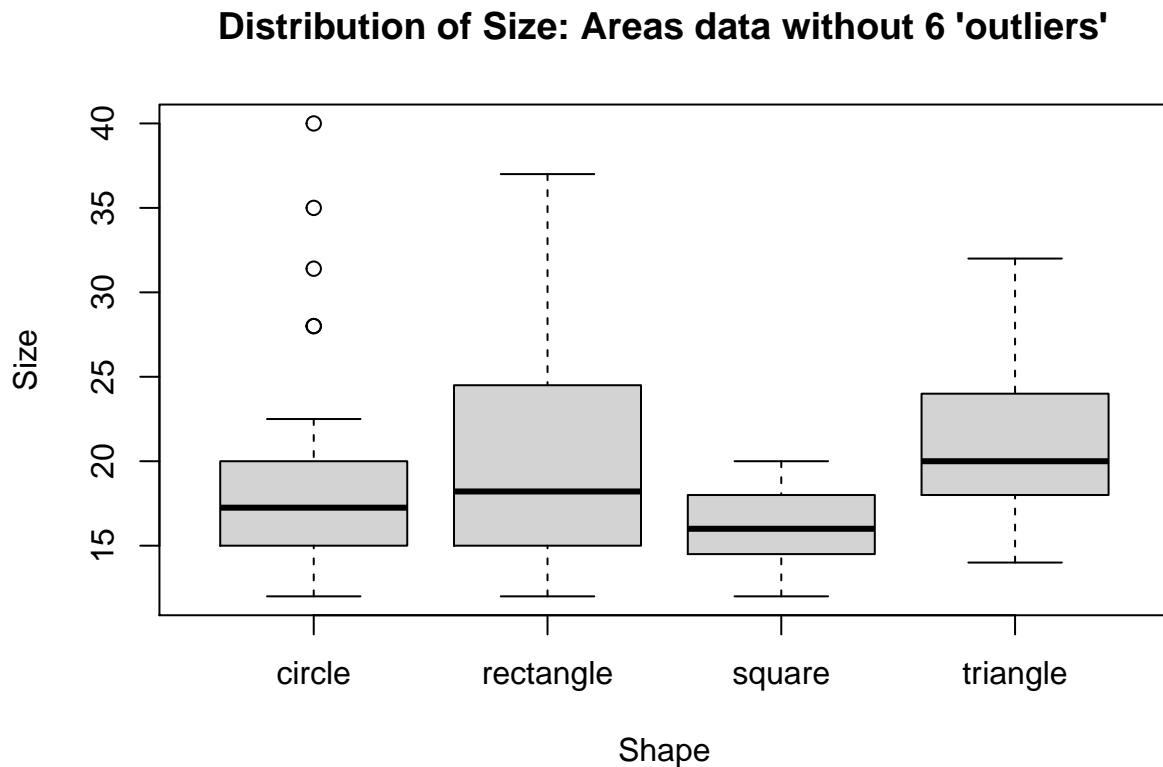

```r
size.less6.ANOVA <- aov(Size ~ factor(Shape), data = areas.less6)
summary(size.less6.ANOVA)
```

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## factor(Shape)  3    429  143.01   4.551 0.00472 **
## Residuals    116   3645   31.42
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
leveneTest(Size ~ factor(Shape), center = "mean", data = areas.less6)
```

```
## Levene's Test for Homogeneity of Variance (center = "mean")
##        Df F value     Pr(>F)
## group   3  7.3576 0.0001477 ***
##       116
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(x = size.less6.ANOVA$fitted.values, y = size.less6.ANOVA$residuals,
     main = "Residuals vs. fitted values\n ANOVA on Areas data without 6 'outliers'",
     xlab = "Predicted values", ylab = "Residuals")
abline(h = 0, lty=2)
```

Residuals vs. fitted values
ANOVA on Areas data without 6 'outliers'
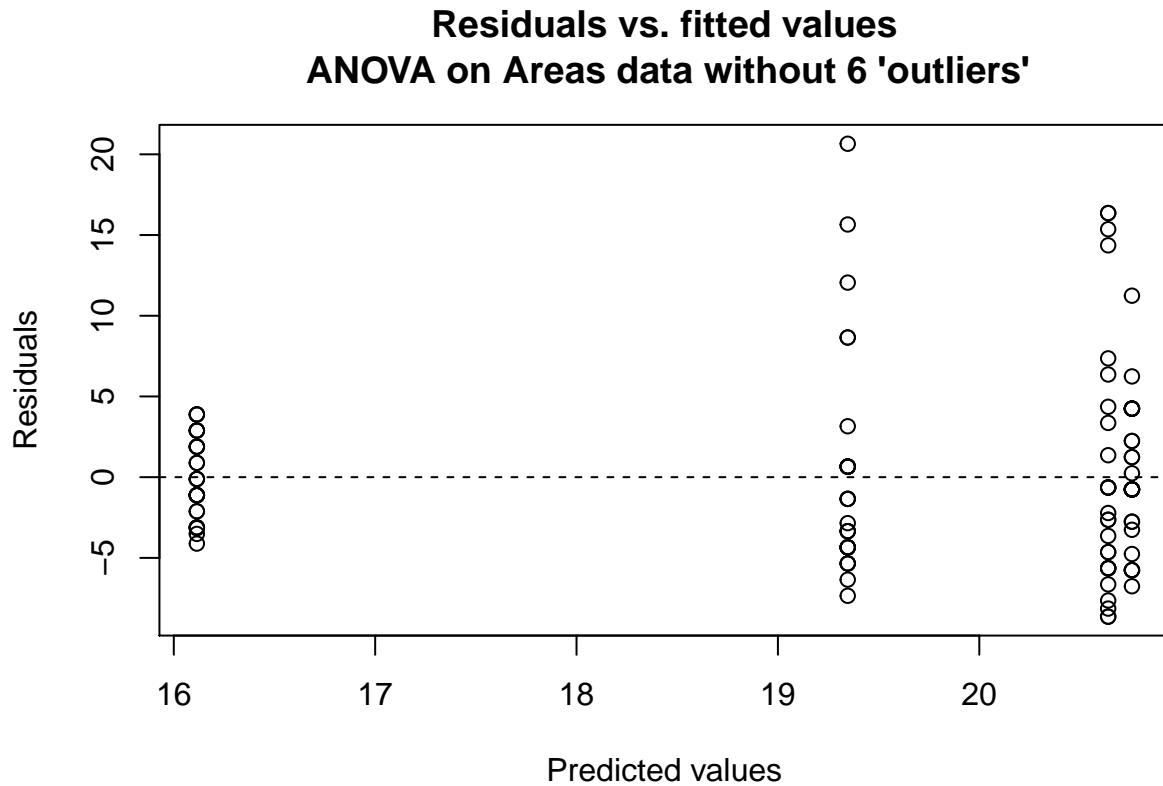

```
qqnorm(size.less6.ANOVA$residuals,
       main = "Normal Q-Q plot of residuals\n ANOVA on Areas data without 6 'outliers'")
qqline(size.less6.ANOVA$residuals)
```

Normal Q–Q plot of residuals
ANOVA on Areas data without 6 'outliers'
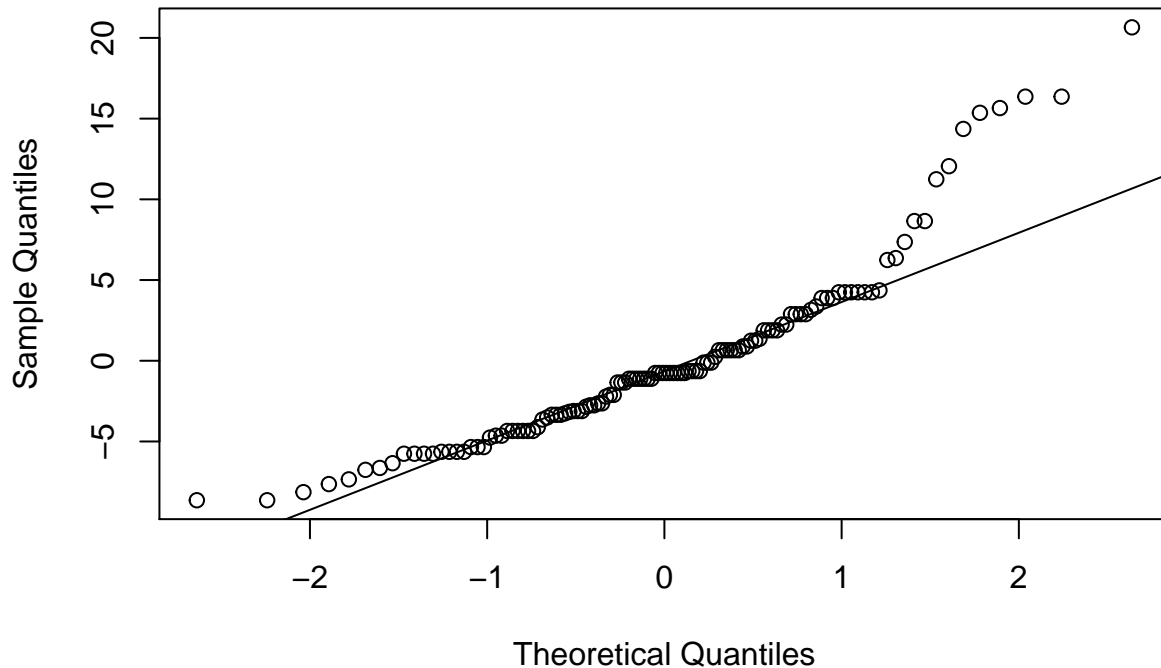

```r
# Non-parametric Kruskal-Wallis test
kruskal.test(Size ~ factor(Shape), data = areas.less6)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Size by factor(Shape)
## Kruskal-Wallis chi-squared = 17.687, df = 3, p-value = 0.0005104
```

```r
# Tukey's HSD
TukeyHSD(size.less6.ANOVA)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Size ~ factor(Shape), data = areas.less6)
##
## $`factor(Shape)`
##                         diff        lwr        upr      p adj
## rectangle-circle   1.2933333 -2.5461151  5.1327818 0.8162121
## square-circle     -3.2321505 -6.9742724  0.5099713 0.1157309
## triangle-circle    1.4113978 -2.3307240  5.1535197 0.7593206
## square-rectangle  -4.5254839 -8.3349194 -0.7160484 0.0129383
## triangle-rectangle 0.1180645 -3.6913710  3.9275000 0.9998099
## triangle-square    4.6435484  0.9322264  8.3548704 0.0078367
```

```
club <- rep(c("A","B","C","D"), each = 10)
times <- c(52.59, 55.55, 57.17, 50.31, 55.86, 56.01, 53.85, 53.91, 53.87, 53.22,
55.05, 54.00, 54.45, 56.13, 57.92, 55.78, 54.98, 54.18, 54.33, 60.83,
54.27, 53.02, 53.12, 54.92, 52.61, 51.10, 55.15, 51.95, 53.97, 52.13,
60.20, 57.05, 56.58, 57.00, 54.74, 55.66, 53.64, 55.32, 57.41, 57.07)
tapply(times, club, mean)
```

```
##      A      B      C      D
## 54.234 55.765 53.224 56.467
```

```
mean(times)
```

```
## [1] 54.9225
```

```
boxplot(times ~ factor(club),
        main = "Distribution of Albatross Wingspan by Colony",
        xlab = "Colony", ylab = "Wingspan (cm)")
```
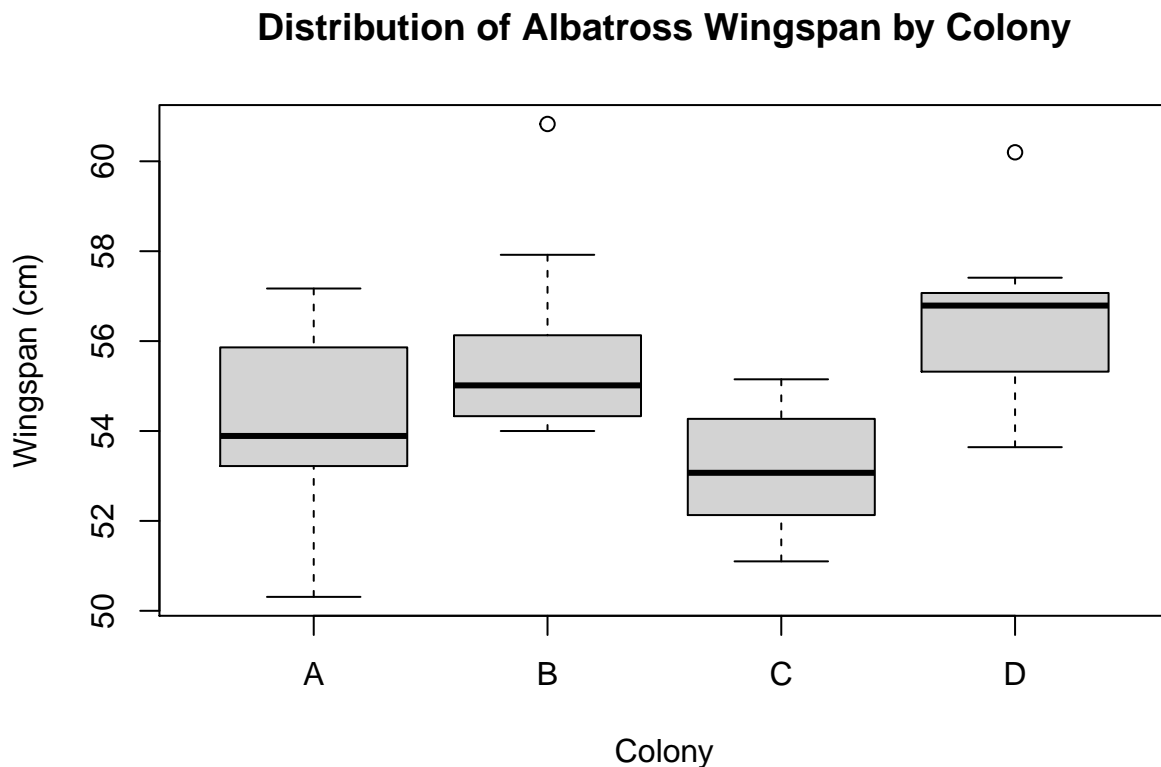


```
times.ANOVA <- aov(times ~ factor(club))
summary(times.ANOVA)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## factor(club)  3  64.54  21.514    6.38 0.0014 **
```

6

```
## Residuals    36 121.40   3.372
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
leveneTest(times ~ factor(club), center = "mean")
```

```
## Levene's Test for Homogeneity of Variance (center = "mean")
##       Df F value Pr(>F)
## group  3  0.3507 0.7889
##       36
```

```r
TukeyHSD(times.ANOVA)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = times ~ factor(club))
##
## $`factor(club)`
##        diff         lwr        upr      p adj
## B-A  1.531 -0.68082063  3.7428206 0.2612210
## C-A -1.010 -3.22182063  1.2018206 0.6123797
## D-A  2.233  0.02117937  4.4448206 0.0471202
## C-B -2.541 -4.75282063 -0.3291794 0.0190435
## D-B  0.702 -1.50982063  2.9138206 0.8278042
## D-C  3.243  1.03117937  5.4548206 0.0019040
```

```r
#enter data
wingspan <- c(160, 183, 192, 168, 177, 160, 169, 185,
              173, 157, 171, 166, 154, 172, 165, 166,
              147, 135, 156, 153, 148, 156, 150, 145,
              173, 159, 162, 169, 173, 178, 156, 163)
colony<-rep(c("A","B","C","D"), each = 8)
cbind(colony, wingspan)
```

```
##       colony wingspan
##  [1,] "A"    "160"
##  [2,] "A"    "183"
##  [3,] "A"    "192"
##  [4,] "A"    "168"
##  [5,] "A"    "177"
##  [6,] "A"    "160"
##  [7,] "A"    "169"
##  [8,] "A"    "185"
##  [9,] "B"    "173"
## [10,] "B"    "157"
## [11,] "B"    "171"
## [12,] "B"    "166"
## [13,] "B"    "154"
## [14,] "B"    "172"
## [15,] "B"    "165"
## [16,] "B"    "166"
```

```
## [17,] "C"      "147"
## [18,] "C"      "135"
## [19,] "C"      "156"
## [20,] "C"      "153"
## [21,] "C"      "148"
## [22,] "C"      "156"
## [23,] "C"      "150"
## [24,] "C"      "145"
## [25,] "D"      "173"
## [26,] "D"      "159"
## [27,] "D"      "162"
## [28,] "D"      "169"
## [29,] "D"      "173"
## [30,] "D"      "178"
## [31,] "D"      "156"
## [32,] "D"      "163"
```
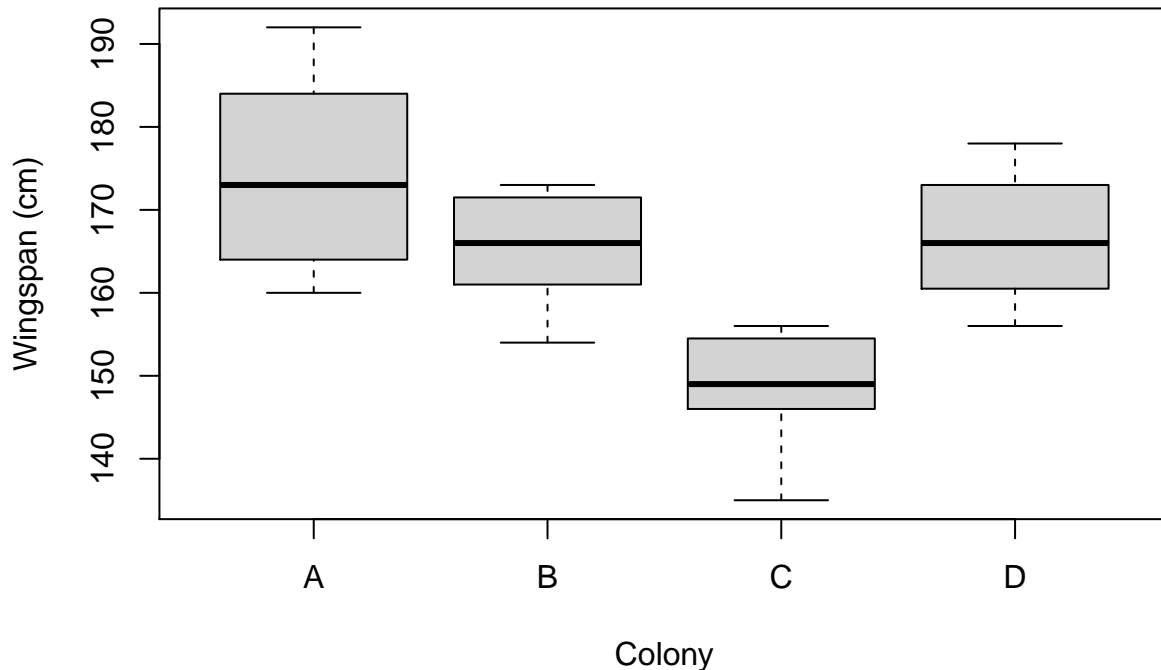
```r
# Display the mean, useful for Tukey underlining diagram
tapply(wingspan, colony, mean)
```

```
##       A       B       C       D
## 174.250 165.500 148.750 166.625
```

```r
#boxplot
boxplot(wingspan ~ factor(colony),
        main = "Distribution of Albatross Wingspan by Colony",
        xlab = "Colony", ylab = "Wingspan (cm)")
```

Distribution of Albatross Wingspan by Colony


```r
# Levene's test
leveneTest(wingspan ~ factor(colony), center = "mean")
```

```
## Levene's Test for Homogeneity of Variance (center = "mean")
##       Df F value  Pr(>F)
## group  3  2.3646 0.09238 .
##       28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Carry out a one-way ANOVA
wingspan.ANOVA <- aov(wingspan ~ factor(colony))
summary(wingspan.ANOVA)
```

```
##                 Df Sum Sq Mean Sq F value   Pr(>F)
## factor(colony)  3   2773   924.2   12.47 2.33e-05 ***
## Residuals      28   2075    74.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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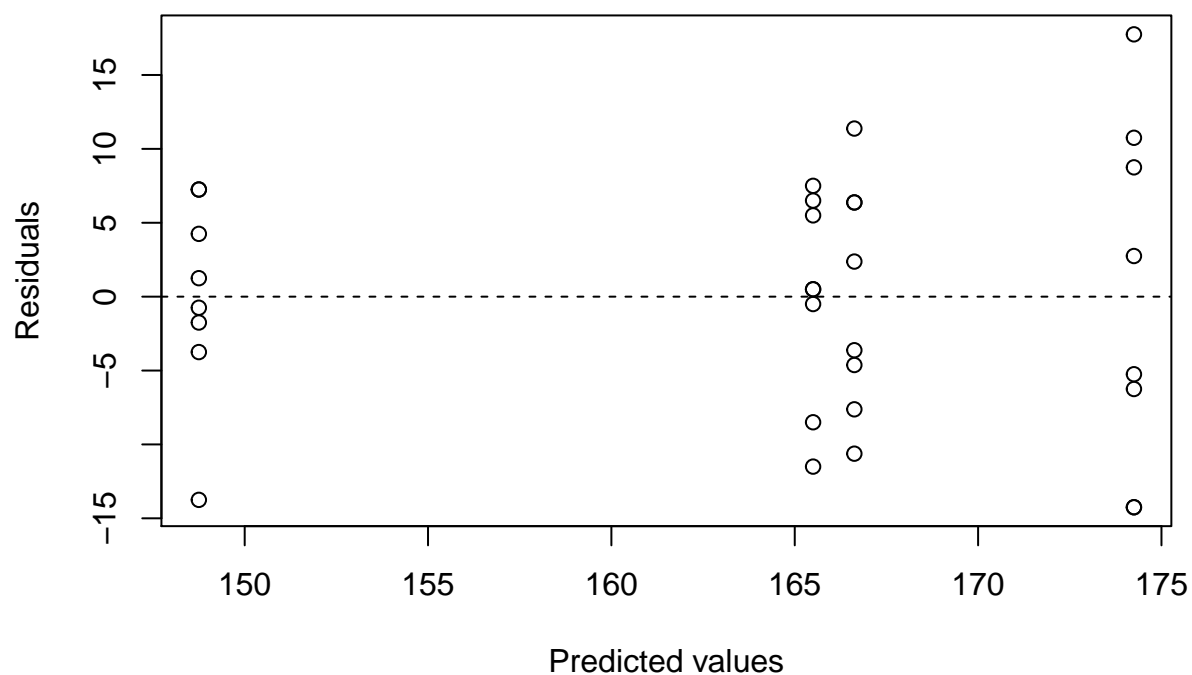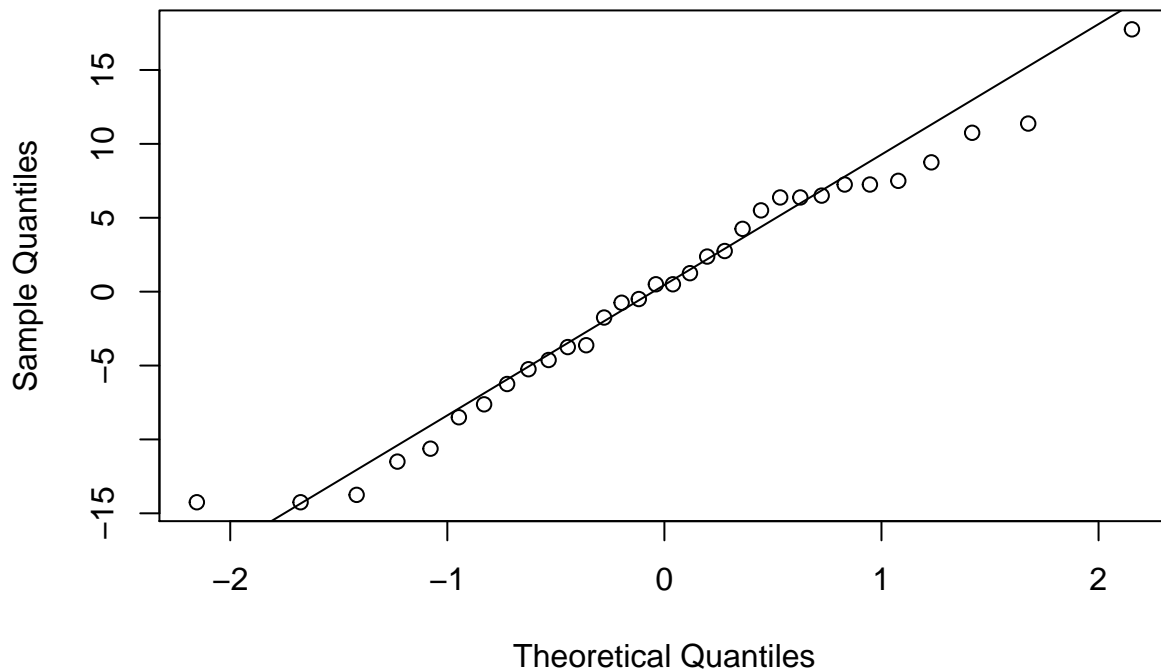
Residuals vs. fitted values
ANOVA on albatross data


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

Normal Q–Q plot of residuals
ANOVA on albatross data


```
# Carry out Tukey's HSD test
TukeyHSD(wingspan.ANOVA)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = wingspan ~ factor(colony))
##
## $`factor(colony)`
##         diff       lwr        upr       p adj
## B-A   -8.750 -20.501659   3.001659 0.2003031
## C-A  -25.500 -37.251659 -13.748341 0.0000128
## D-A   -7.625 -19.376659   4.126659 0.3077431
## C-B  -16.750 -28.501659  -4.998341 0.0029790
## D-B    1.125 -10.626659  12.876659 0.9935997
## D-C   17.875   6.123341  29.626659 0.0015025
```

random effects ONE-WAY ANOVA: same code as means model (categorical variables (factors))

```
score <- c(69, 77, 59, 71, 63, 69, 62, 52, 71, 79,
           52, 58, 47, 38, 63, 58, 47, 49, 58, 49,
           72, 46, 56, 73, 53, 53, 60, 54, 71, 52,
```

```
        53, 84, 76, 81, 61, 74, 59, 53, 74, 70,
        83, 76, 81, 59, 80, 72, 69, 67, 78, 78)
teacher <- rep(c(1,2,3,4,5), each = 10)
# Display the mean, useful for Tukey underlining diagram
tapply(score, teacher, mean)
```

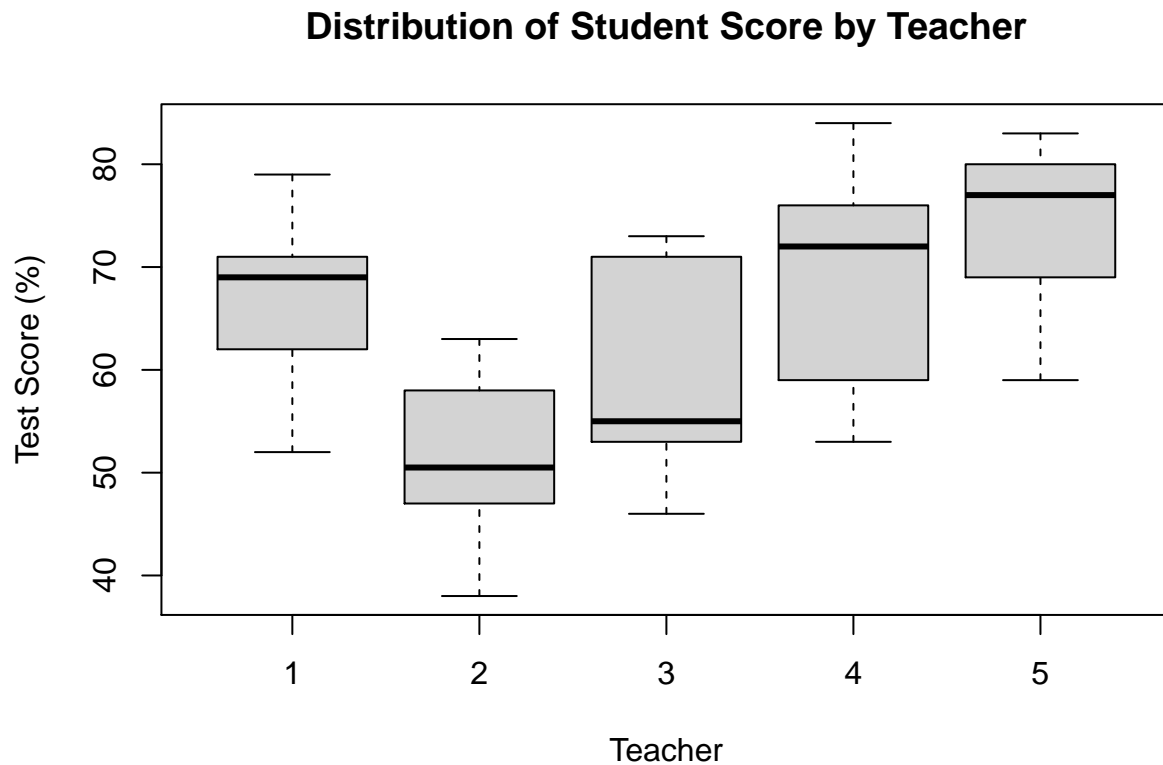```
1   2   3   4   5
67.2 51.9 59.0 68.5 74.3
```

```
#boxplot
boxplot(score ~ factor(teacher),
        main = "Distribution of Student Score by Teacher",
        xlab = "Teacher", ylab = "Test Score (%)")
```

Distribution of Student Score by Teacher


```
# Levene's test
leveneTest(score ~ factor(teacher), center = "mean")
```

```
## Levene's Test for Homogeneity of Variance (center = "mean")
##       Df F value Pr(>F)
## group  4  1.2552 0.3016
##       45
```

```
# Carry out a one-way ANOVA
teacher.ANOVA <- aov(score ~ factor(teacher))
summary(teacher.ANOVA)
```

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)
## factor(teacher)   4   3078   769.6   9.654 1.01e-05 ***
## Residuals        45   3587    79.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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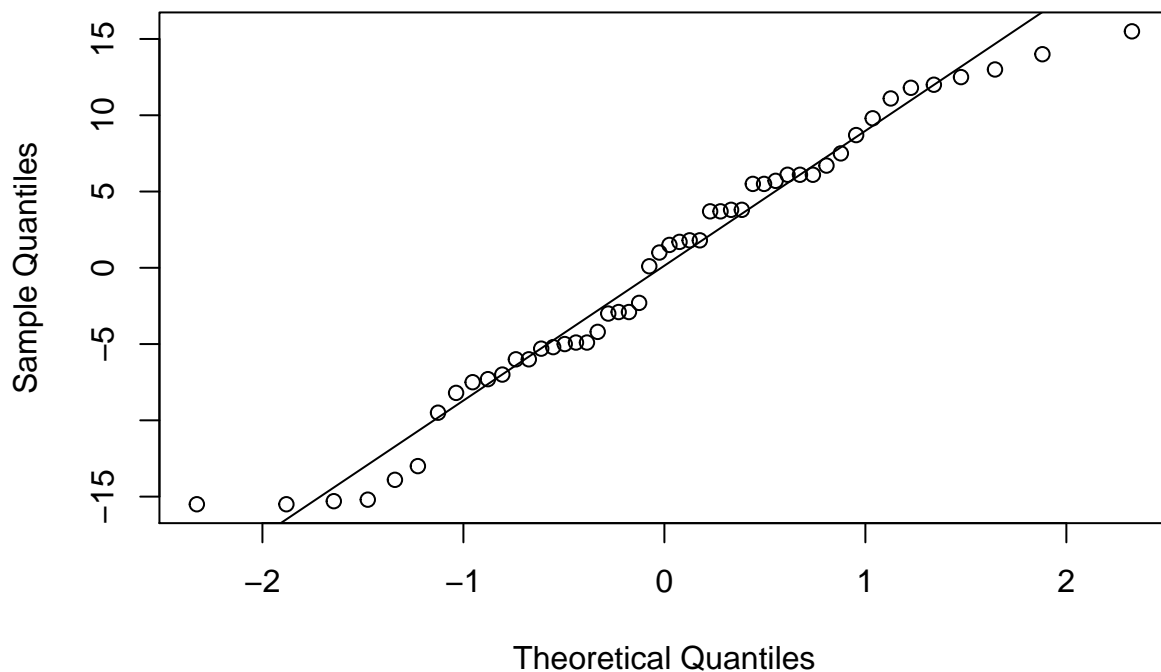
Residuals vs. fitted values
ANOVA on teacher effects data
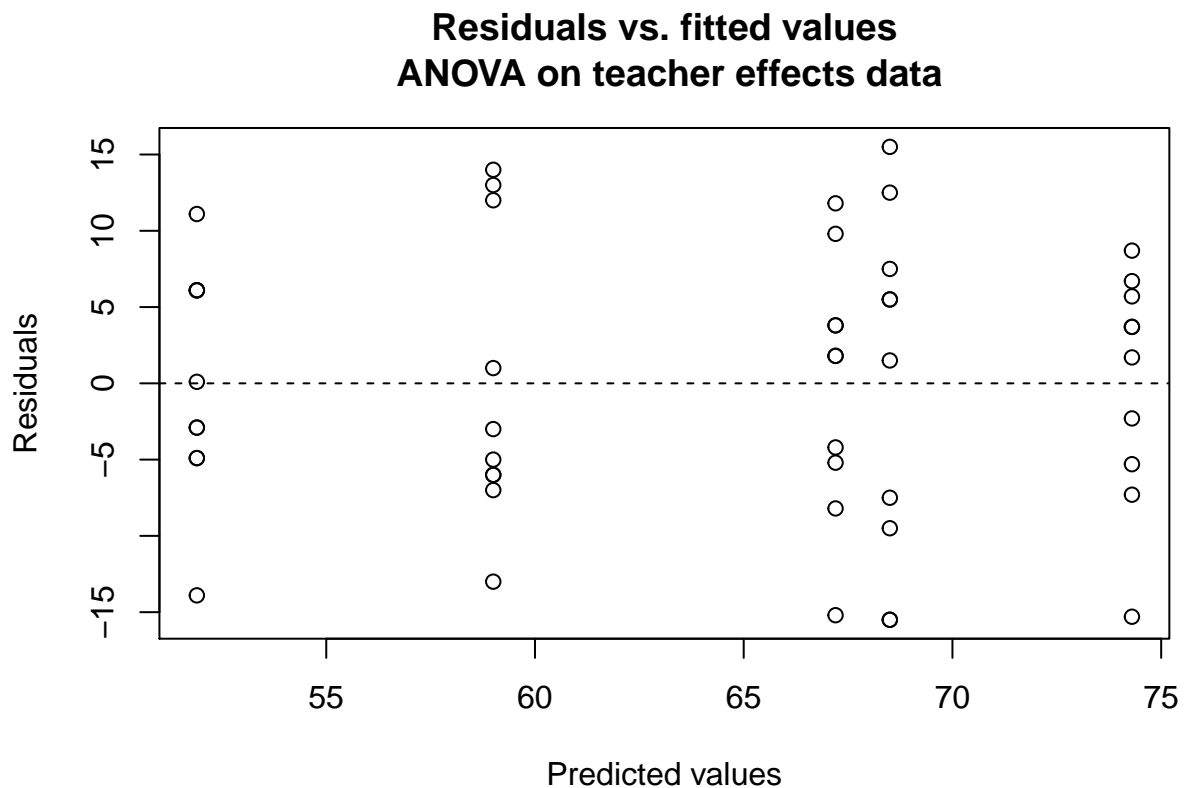

```
# Normal Q-Q plot
qqnorm(teacher.ANOVA$residuals,
       main = "Normal Q-Q plot of residuals\n ANOVA on teacher effects data")
qqline(teacher.ANOVA$residuals)
```

13

Normal Q–Q plot of residuals
ANOVA on teacher effects data

two-way ANOVA:interaction test (categorical variables (factors))

```
conc <- c(12, 14, 19, 17, 21, 24, 11, 14, 25, 23, 32, 37)
dose <- rep(rep(c("1.Low", "2.Medium","3.High"),each =2), times = 2) # times: 2 rows
delivery<- rep(c("Patch", "Capsule"), each=6) #each row has 6 value
# Fit a two-way ANOVA
drug.ANOVA <- aov(conc ~ factor(dose) * factor(delivery))
summary(drug.ANOVA)
```

```
##                            Df Sum Sq Mean Sq F value   Pr(>F)
## factor(dose)                2  496.5  248.25  54.164 0.000145 ***
## factor(delivery)            1  102.1  102.08  22.273 0.003260 **
## factor(dose):factor(delivery) 2   78.2   39.08   8.527 0.017627 *
## Residuals                   6   27.5    4.58
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Levene's test
leveneTest(conc ~ factor(dose) * factor(delivery), center = "mean")
```

```
## Warning in anova.lm(lm(resp ~ group)): ANOVA F-tests on an essentially perfect
## fit are unreliable
```

```
## Levene's Test for Homogeneity of Variance (center = "mean")
##       Df    F value    Pr(>F)
## group  5 1.0258e+30 < 2.2e-16 ***
##        6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Scatterplot of residuals vs. fitted values
plot(x = drug.ANOVA$fitted.values, y = drug.ANOVA$residuals,
     main = "Residuals vs. fitted values\n ANOVA for drug concentration data",
     xlab = "Predicted values", ylab = "Residuals")
abline(h = 0, lty = 2)
```

Residuals vs. fitted values
ANOVA for drug concentration data
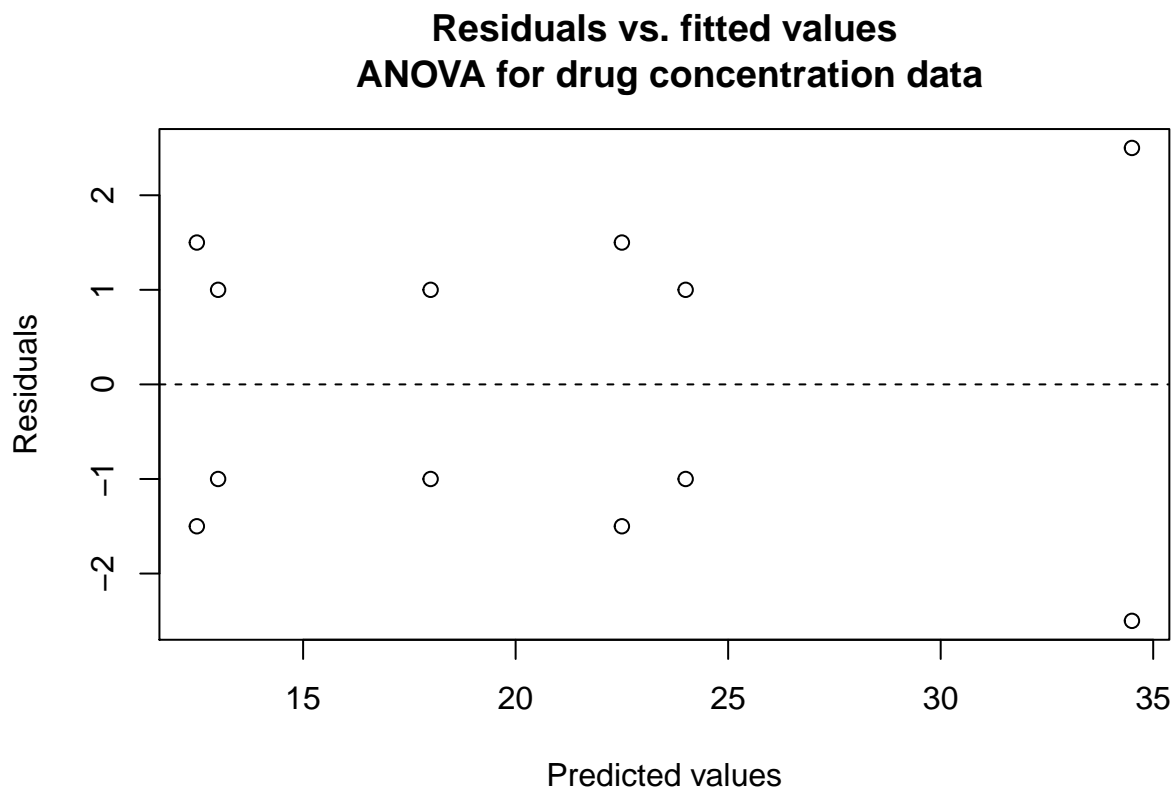

```r
# Normal Q-Q plot
qqnorm(drug.ANOVA$residuals,
       main = "Normal Q-Q plot of residuals\n ANOVA for drug concentration data")
qqline(drug.ANOVA$residuals)
```

Normal Q–Q plot of residuals
ANOVA for drug concentration data
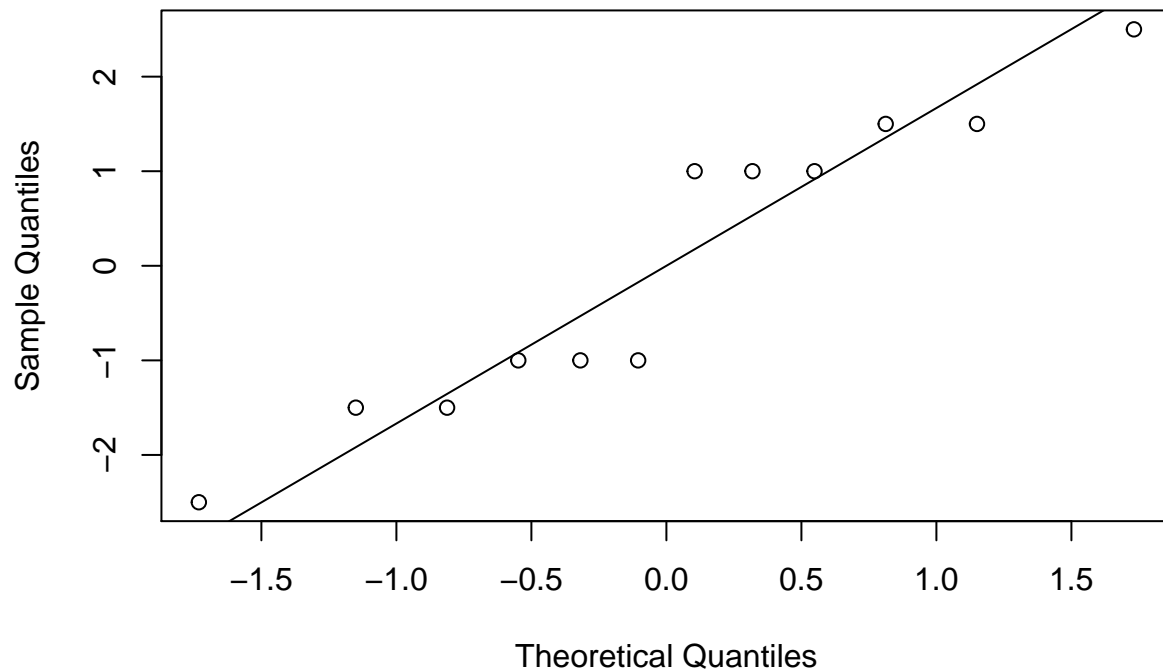

```
# Interaction graph
interaction.plot(x.factor = dose,
                 trace.factor = delivery,
                 response = conc,
                 fun = mean,
                 xlab = "Dose",
                 ylab = "Mean drug concentration",
                 main = "Interaction graph for drug delivery",
                 legend = TRUE, xpd=TRUE,
                 trace.label="Delivery")
```

Interaction graph for drug delivery
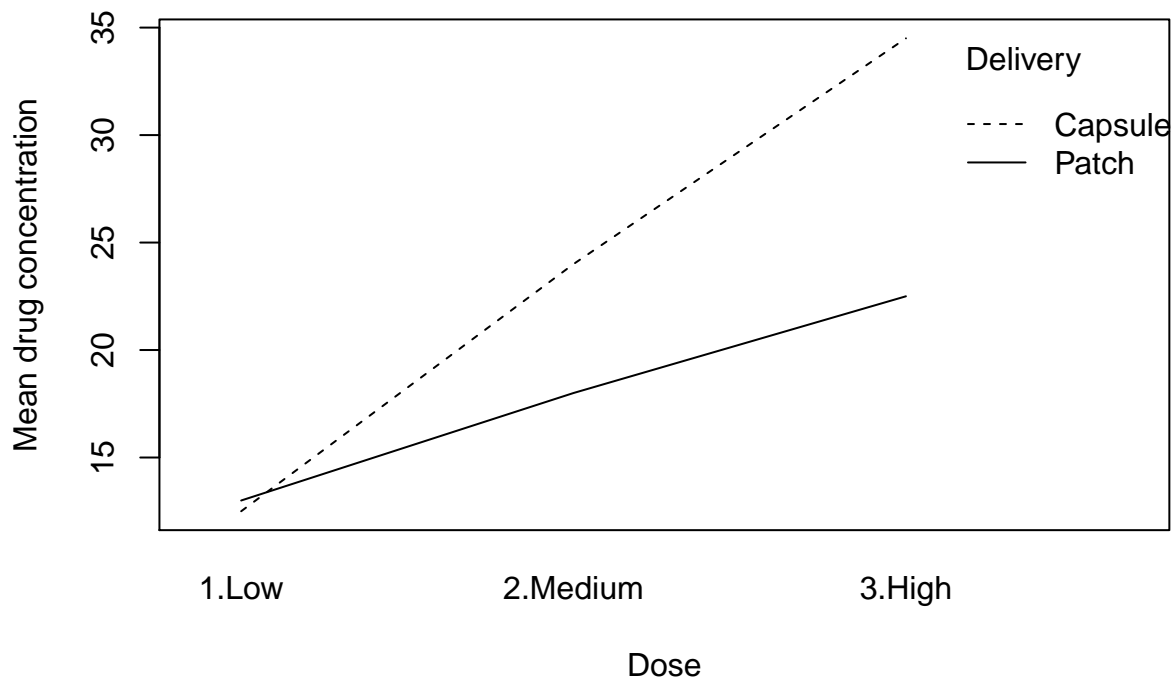
```r
#TukeyHSD(rat.logmistakes.ANOVA, which = "factor(strain)")
```

three way ANOVA: interaction test (categorical variables (factors))

```r
Recovery <- c(6, 3, 7, 4, 5, 4, 7, 7, 8, 5, 6, 7, 4, 6, 8,
              8, 5, 6, 9, 5, 6, 6, 7, 2, 4, 6, 7, 3, 9, 5, 8, 12, 3, 6, 8, 7, 5, 6, 6, 6, 8, 9, 6,
              7, 9, 9, 11, 6, 7, 8, 6, 10, 11, 12, 6, 13, 7, 9, 8)
Age <- factor(rep(c("Y","M","O"), c(15,28,16)), levels = c("Y","M","O")) # each ega group(row) it has 1
Severity <- factor(c(rep(c("L","H"),c(6,9)),
                     rep(c("L","H"),c(13,15)),
                     rep(c("L","H"),c(8,8))),
                   levels = c("L","H")) # three rows with each Low/High
SelfMed <- factor(c(rep(c("Y","N"),c(2,4)),rep(c("Y","N"),c(3,6)),
                    rep(c("Y","N"),c(7,6)),rep(c("Y","N"),c(11,4)),
                    rep(c("Y","N"),c(5,3)),rep(c("Y","N"),c(6,2))),
                  levels = c("Y","N"))
head(cbind(Age, Severity, SelfMed))
```

```
##      Age Severity SelfMed
## [1,]   1        1       1
## [2,]   1        1       1
## [3,]   1        1       2
```

```
## [4,]    1       1       2
## [5,]    1       1       2
## [6,]    1       1       2
```

```r
# Fit a three-way ANOVA
recover.ANOVA <- aov(Recovery ~ Age * Severity + Age * SelfMed + Severity * SelfMed)
summary(recover.ANOVA)
```

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)
## Age               2  77.36   38.68  10.150 0.000205 ***
## Severity          1  30.11   30.11   7.903 0.007078 **
## SelfMed           1   8.25    8.25   2.164 0.147693
## Age:Severity      2   0.86    0.43   0.112 0.893859
## Age:SelfMed       2   1.40    0.70   0.184 0.832266
## Severity:SelfMed  1   0.92    0.92   0.241 0.625898
## Residuals        49 186.73    3.81
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Scatterplot of residuals vs. fitted values
plot(x = recover.ANOVA$fitted.values, y = recover.ANOVA$residuals,
     main = "Residuals vs. fitted values\n ANOVA for flu recovery data",
     xlab = "Predicted values", ylab = "Residuals")
abline(h = 0, lty = 2)
```

Residuals vs. fitted values
ANOVA for flu recovery data
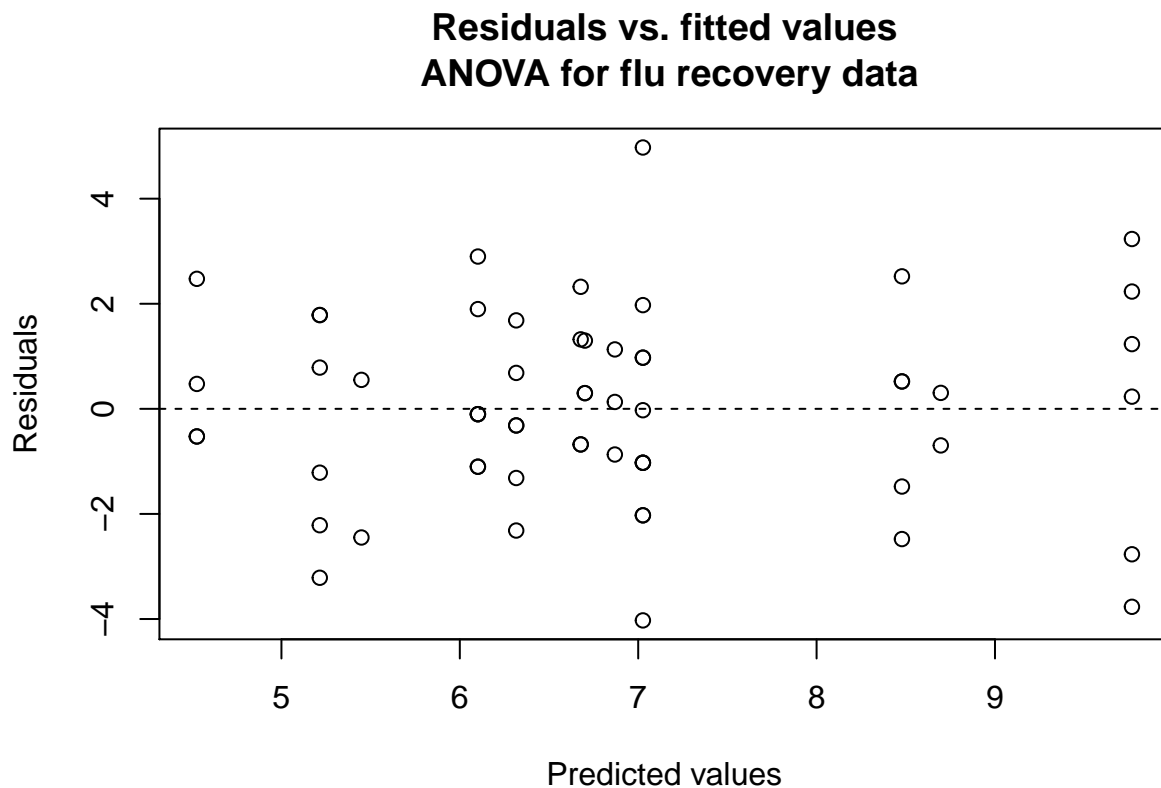
```
# Normal Q-Q plot
qqnorm(recover.ANOVA$residuals,
       main = "Normal Q-Q plot of residuals\n ANOVA for flu recovery data")
qqline(recover.ANOVA$residuals)
```

Normal Q–Q plot of residuals
ANOVA for flu recovery data
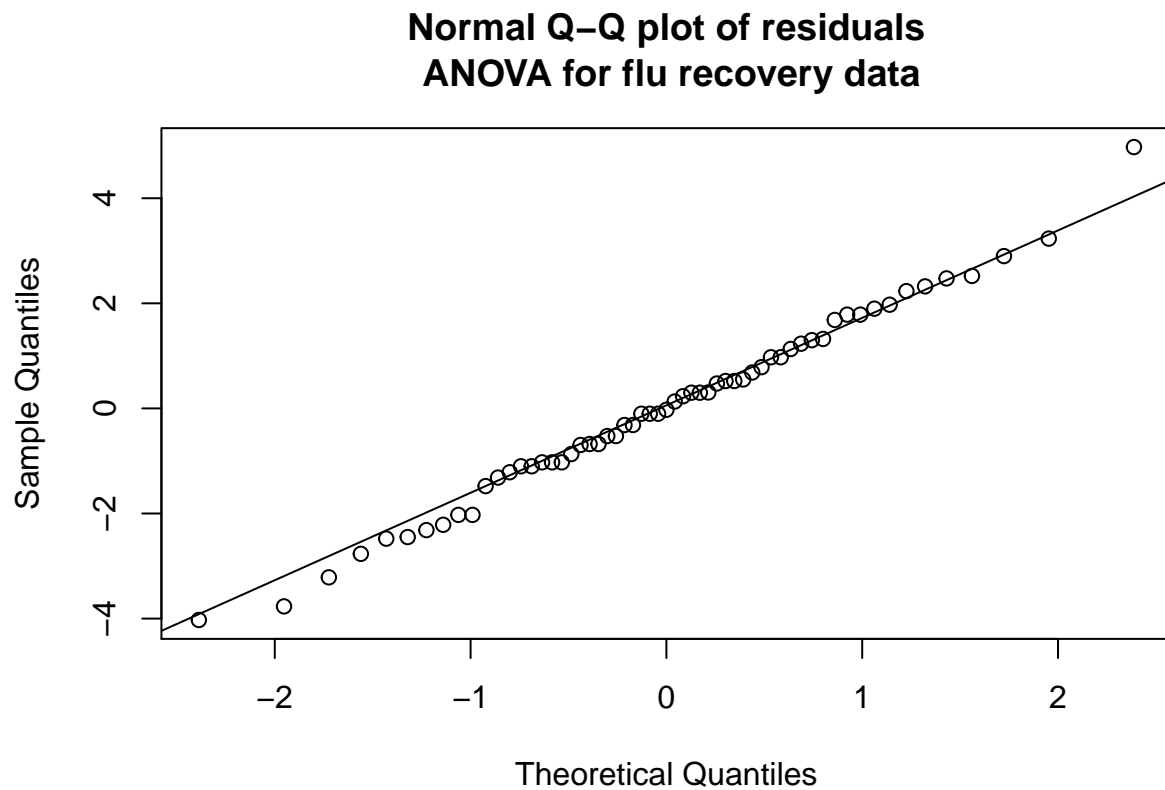

```
# Interaction graph for flu recovery data.
interaction.plot(x.factor = Age,
                 trace.factor = Severity,
                 response = Recovery,
                 fun = mean,
                 ylab = "Mean recovery (days)",
                 main = "Interaction graph for flu recovery",
                 legend = TRUE, xpd=TRUE)
```

Interaction graph for flu recovery
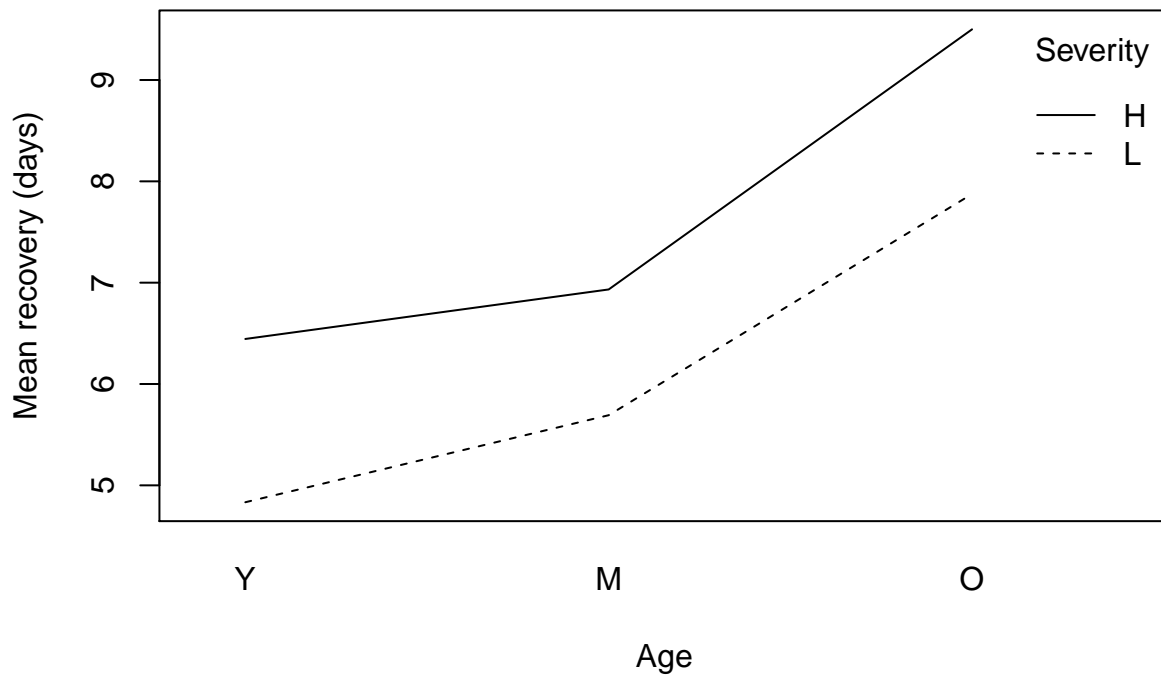

```
#multiple comparisons test
# Fit a one-way ANOVA
# including only the Age factor. Age has more than one comparison
recover2.ANOVA <- aov(Recovery ~ Age)
summary(recover2.ANOVA)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Age           2  77.36   38.68   9.489 0.000282 ***
## Residuals    56 228.27    4.08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(recover2.ANOVA) # Since there is no significant interaction
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Recovery ~ Age)
##
## $Age
##          diff        lwr      upr     p adj
## M-Y 0.5571429 -0.9981564 2.112442 0.6659994
## O-Y 2.8875000  1.1405535 4.634447 0.0005806
## O-M 2.3303571  0.8070351 3.853679 0.0014890
```

```r
TukeyHSD(recover.ANOVA)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Recovery ~ Age * Severity + Age * SelfMed + Severity * SelfMed)
##
## $Age
##           diff        lwr      upr      p adj
## M-Y 0.5571429 -0.9525108 2.066796 0.6478910
## O-Y 2.8875000  1.1918237 4.583176 0.0004269
## O-M 2.3303571  0.8517423 3.808972 0.0011154
##
## $Severity
##           diff        lwr      upr      p adj
## H-L 1.430126 0.4049967 2.455255 0.0072241
##
## $SelfMed
##           diff        lwr      upr      p adj
## N-Y -0.7176088 -1.751146 0.3159279 0.1692188
##
## $`Age:Severity`
##                  diff         lwr      upr      p adj
## M:L-Y:L   0.9487738 -1.90830084 3.805848 0.9206314
## O:L-Y:L   3.1016840 -0.02465080 6.228019 0.0529545
## Y:H-Y:L   1.7128106 -1.33817978 4.763801 0.5609388
## M:H-Y:L   2.1360773 -0.66020164 4.932356 0.2279423
## O:H-Y:L   4.7286887  1.60235386 7.855024 0.0005993
## O:L-M:L   2.1529102 -0.44835760 4.754178 0.1582110
## Y:H-M:L   0.7640368 -1.74617578 3.274249 0.9440076
## M:H-M:L   1.1873034 -1.00627821 3.380885 0.5992318
## O:H-M:L   3.7799149  1.17864705 6.381183 0.0010565
## Y:H-O:L  -1.3888734 -4.20174759 1.424001 0.6880534
## M:H-O:L  -0.9656068 -3.49994976 1.568736 0.8665553
## O:H-O:L   1.6270047 -1.26741898 4.521428 0.5595659
## M:H-Y:H   0.4232666 -2.01752569 2.864059 0.9953758
## O:H-Y:H   3.0158781  0.20300391 5.828752 0.0290187
## O:H-M:H   2.5926115  0.05826848 5.126954 0.0421940
##
## $`Age:SelfMed`
##                  diff         lwr       upr      p adj
## M:Y-Y:Y   0.3833350 -2.5430722  3.3097422 0.9987827
## O:Y-Y:Y   2.9147307 -0.2075414  6.0370027 0.0800426
## Y:N-Y:Y  -0.5809163 -3.7515985  2.5897660 0.9940230
## M:N-Y:Y  -0.2143801 -3.3850623  2.9563022 0.9999523
## O:N-Y:Y   1.5883045 -2.0728840  5.2494929 0.7908717
## O:Y-M:Y   2.5313957  0.3159633  4.7468280 0.0165571
## Y:N-M:Y  -0.9642513 -3.2474034  1.3189009 0.8087175
## M:N-M:Y  -0.5977151 -2.8808672  1.6854371 0.9702587
## O:N-M:Y   1.2049694 -1.7214377  4.1313766 0.8245000
## Y:N-O:Y  -3.4956469 -6.0249764 -0.9663175 0.0020510
## M:N-O:Y  -3.1291107 -5.6584402 -0.5997813 0.0074852
## O:N-O:Y  -1.3264262 -4.4486982  1.7958458 0.8049042
```

21

```
## M:N-Y:N  0.3665362 -2.2223150  2.9553874 0.9982315
## O:N-Y:N  2.1692207 -1.0014615  5.3399030 0.3414512
## O:N-M:N  1.8026845 -1.3679977  4.9733668 0.5473483
##
## $'Severity:SelfMed'
##                diff        lwr        upr      p adj
## H:Y-L:Y  1.1296379 -0.6794335  2.9387093 0.3552274
## L:N-L:Y -1.0255514 -3.0251437  0.9740408 0.5275825
## H:N-L:Y  0.6141812 -1.4281553  2.6565176 0.8541814
## L:N-H:Y -2.1551894 -4.0047371 -0.3056416 0.0164153
## H:N-H:Y -0.5154568 -2.4111350  1.3802215 0.8873931
## H:N-L:N  1.6397326 -0.4385422  3.7180074 0.1680073
```

```r
diversity <- c(2.8, 1.9, 2.6, 2.7, 2.4, 3.4, 3.2, 3.3, 2.1, 2.1,
               3.4, 2.6, 0.4, 3.3, 3.2, 2.0, 3.9, 2.8, 3.3, 2.6,
               0.9, 4.6, 4.0, 3.6, 2.1, 2.5, 3.6, 4.1, 3.6, 2.1,
               2.8, 1.4, 3.2, 3.9, 2.5, 3.7, 2.1, 3.9, 2.8, 1.8,
               4.7, 2.9, 4.6, 2.0, 2.8, 2.4, 3.8, 4.1, 2.7, 1.3,
               0.8, 2.6, 3.1, 0.8, 2.6, 2.7, 4.2, 3.0, 0.9, 3.0)
rainfall <- c(55, 45, 64, 24, 61, 69, 58, 44, 73, 32,
              49, 16, 20, 46, 37, 70, 34, 75, 41, 58,
              83, 40, 52, 40, 72, 74, 44, 20, 71, 78,
              24, 17, 85, 61, 65, 17, 81, 55, 40, 61,
              33, 22, 24, 38, 16, 46, 30, 30, 82, 15,
              76, 79, 81, 19, 41, 78, 61, 25, 67, 63)
temperature <- c(15.1, 14.9, 15.8, 14.8, 16.7, 16.5, 14.8, 15.2, 16.0, 15.5,
                 15.5, 14.5, 14.5, 16.8, 16.5, 16.3, 16.7, 14.4, 15.2, 15.1,
                 14.6, 17.0, 15.7, 14.9, 16.7, 15.7, 14.7, 15.7, 15.7, 15.5,
                 16.3, 14.6, 14.6, 16.4, 15.2, 14.6, 15.8, 15.4, 16.2, 14.5,
                 16.1, 16.5, 16.3, 16.5, 15.6, 14.1, 17.0, 15.2, 14.4, 14.1,
                 14.3, 15.2, 16.6, 14.7, 14.8, 16.2, 14.6, 14.6, 14.1, 16.5)

Model1 <- lm(diversity ~ rainfall + temperature)
summary(Model1)
```

```
##
## Call:
## lm(formula = diversity ~ rainfall + temperature)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.14690 -0.61415  0.02996  0.70570  1.90082
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.366512   2.135966  -2.044 0.045555 *
## rainfall    -0.007229   0.005420  -1.334 0.187575
## temperature  0.486759   0.137496   3.540 0.000804 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9027 on 57 degrees of freedom
## Multiple R-squared:  0.1967, Adjusted R-squared:  0.1685
## F-statistic: 6.977 on 2 and 57 DF,  p-value: 0.001947
```

```
Model2 <- lm(diversity ~ rainfall )
summary(Model2)
```

```
##
## Call:
## lm(formula = diversity ~ rainfall)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.59459 -0.50551  0.00892  0.69499  1.78936
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.123745   0.320688   9.741 8.14e-14 ***
## rainfall    -0.006458   0.005930  -1.089    0.281
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9884 on 58 degrees of freedom
## Multiple R-squared:  0.02004,    Adjusted R-squared:  0.003144
## F-statistic: 1.186 on 1 and 58 DF,  p-value: 0.2806
```

```
Model3 <- lm(diversity ~ temperature )
summary(Model3)
```

```
##
## Call:
## lm(formula = diversity ~ temperature)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.93993 -0.65359 -0.04741  0.63069  1.81213
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.6112     2.1423  -2.152   0.0355 *
## temperature   0.4794     0.1383   3.466   0.0010 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9087 on 58 degrees of freedom
## Multiple R-squared:  0.1716, Adjusted R-squared:  0.1573
## F-statistic: 12.01 on 1 and 58 DF,  p-value: 0.001
```

```
29/(29+47)
```

```
## [1] 0.3815789
```

simple linear regression:numerical variables (independent variables)

```r
# data from csv
marks <- read.csv("marks.csv")
head(marks)
```

```
##   ï..Student Final_Mark Midterm_Test
## 1          1         63         41.0
## 2          2         71         34.5
## 3          3         87         37.5
## 4          4         73         28.0
## 5          5         58         40.0
## 6          6         75         33.0
```

```r
# Fit a linear regression
marks.lm <- lm(Final_Mark ~ Midterm_Test, data = marks)
# ANOVA table
anova(marks.lm)
```

```
## Analysis of Variance Table
##
## Response: Final_Mark
##               Df Sum Sq Mean Sq F value    Pr(>F)
## Midterm_Test   1  13011 13010.6  161.28 < 2.2e-16 ***
## Residuals    166  13391    80.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(marks.lm)
```

```
##
## Call:
## lm(formula = Final_Mark ~ Midterm_Test, data = marks)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -27.2794  -5.2387   0.7115   6.1934  18.9235
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 39.54766    2.29592   17.23   <2e-16 ***
## Midterm_Test 0.92029    0.07247   12.70   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.982 on 166 degrees of freedom
## Multiple R-squared:  0.4928, Adjusted R-squared:  0.4897
## F-statistic: 161.3 on 1 and 166 DF,  p-value: < 2.2e-16
```

```
# Produce a scatterplot
plot(x = marks$Midterm_Test, y = marks$Final_Mark,
     main = "Scatterplot of Assessment Marks with Fitted Regression Line",
     xlab = "Mid-term test mark (/50)", ylab = "Final mark (%)")
abline(marks.lm)
```

Scatterplot of Assessment Marks with Fitted Regression Line
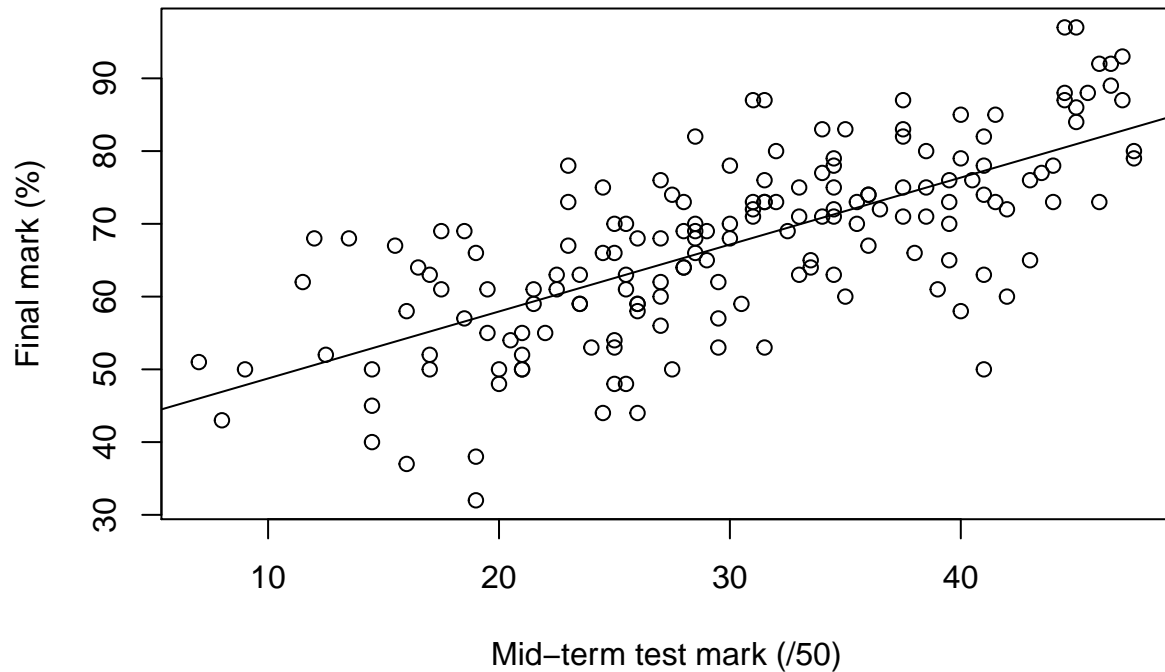


```
# studentized residuals versus fitted values
plot(x = marks.lm$fitted.values, y = rstudent(marks.lm),
     main = "Studentized residuals vs. fitted values\n regression for assessment example",
     xlab = "Predicted value", ylab = "Studentized residual")
abline(h = 0)
abline(h = c(-2, 2), lty = 2)
```

Studentized residuals vs. fitted values
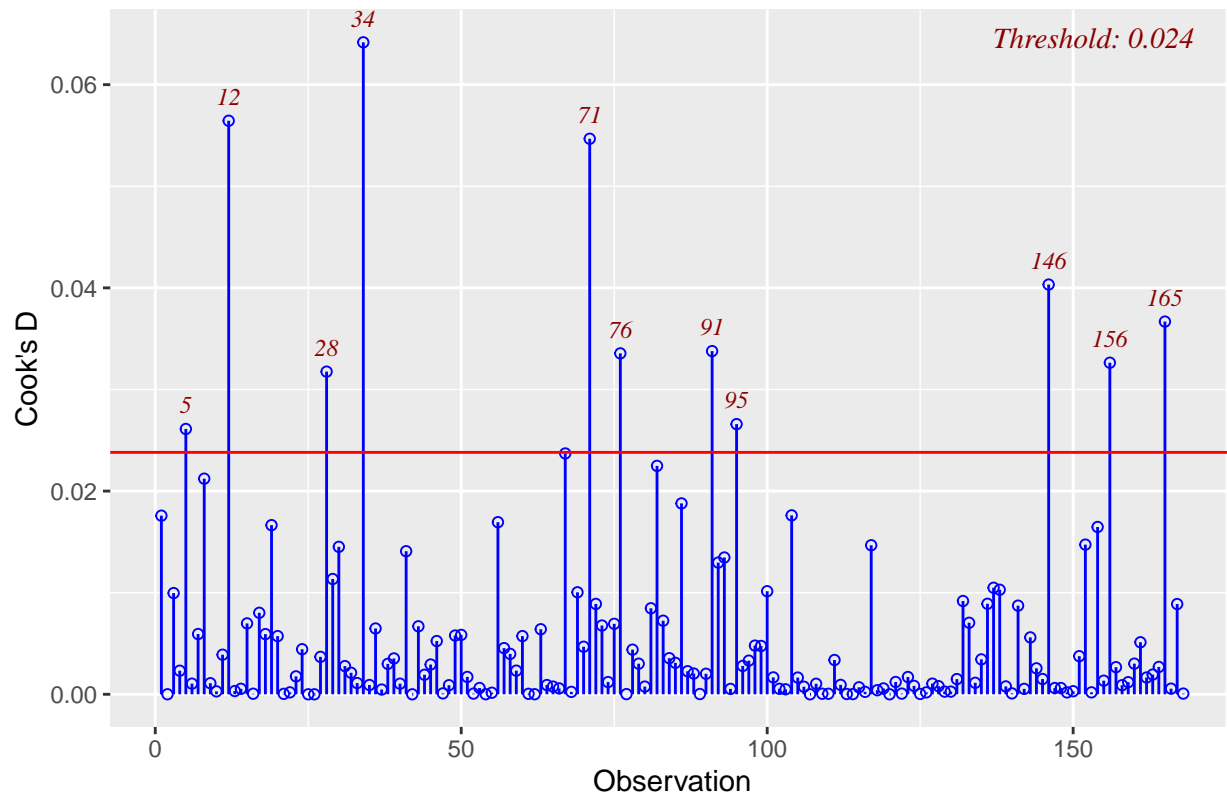## regression for assessment example
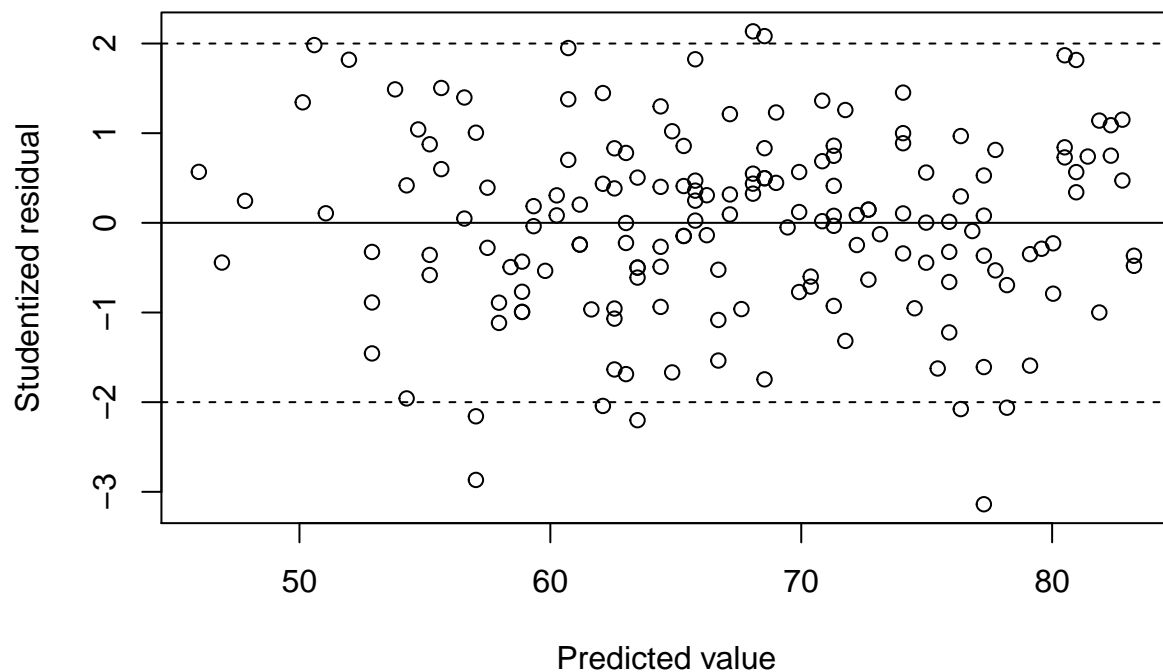

```
# Normal Q-Q plot
qqnorm(marks.lm$residuals,
       main = "Normal Q-Q plot of residuals\n regression for assessment example")
qqline(marks.lm$residuals)
```

Normal Q–Q plot of residuals
regression for assessment example
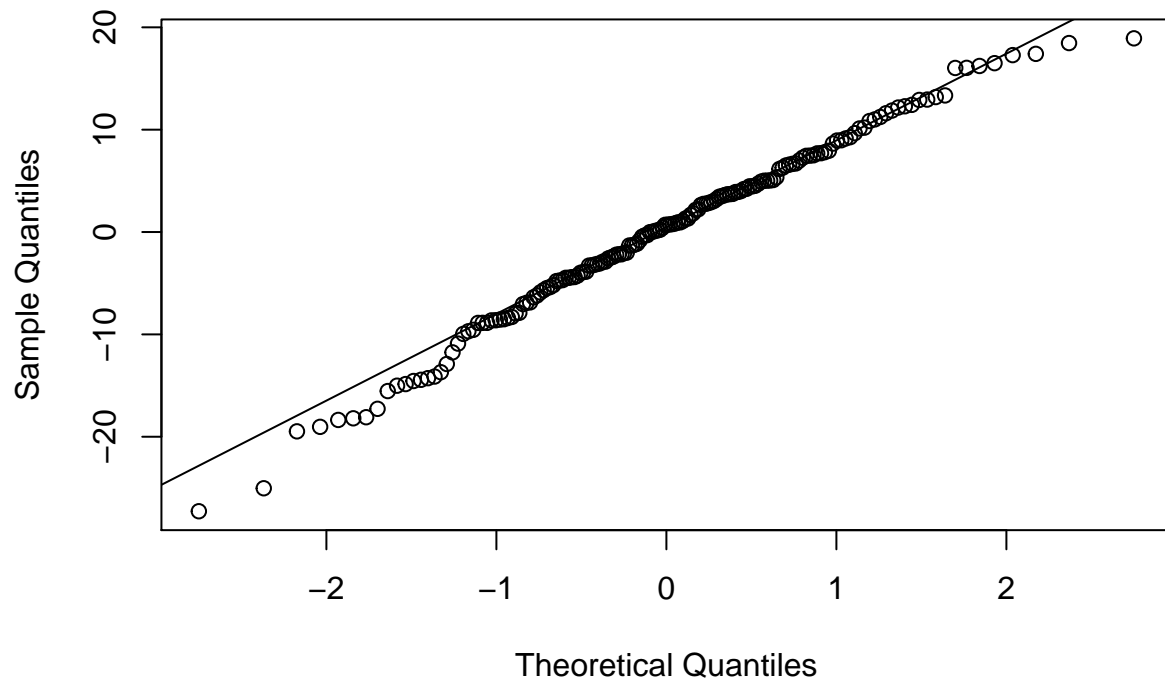

```
# Cook's distances
ols_plot_cooksd_chart(marks.lm)
```

Cook's D Chart

Threshold: 0.024


```r
# Obtain 95% confidence intervals for beta_0 and beta_1.
confint.default(marks.lm) # with original data
```

```
##                  2.5 %    97.5 %
## (Intercept)  35.0477448 44.047582
## Midterm_Test  0.7782563  1.062316
```

```r
# Prediction
predict(marks.lm, newdata = data.frame(Midterm_Test = 20),
        interval = "prediction")
```

```
##        fit      lwr      upr
## 1 57.95339 40.10779 75.79899
```

```r
##########################################################################

# enter data
mercury <- c(1.23, 1.33, 0.04, 0.44, 1.2, 0.27, 0.48, 0.19, 0.83,
        0.81, 0.71, 0.5, 0.49, 1.16, 0.05, 0.15, 0.19, 0.77,
        1.08, 0.98, 0.63, 0.56, 0.41, 0.73, 0.34, 0.59, 0.34,
        0.84, 0.5, 0.34, 0.28, 0.34, 0.87, 0.56, 0.17, 0.18,
        0.19, 0.04, 0.49, 1.1, 0.16, 0.1, 0.48, 0.21, 0.86,
        0.52, 0.65, 0.27, 0.94, 0.4, 0.43, 0.25, 0.27)
pH <- c(6.1, 5.1, 9.1, 6.9, 4.6, 7.3, 5.4, 8.1, 5.8,
```

```
      6.4, 5.4, 7.2, 7.2, 5.8, 7.6, 8.2, 8.7, 7.8,
   5.8, 6.7, 4.4, 6.7, 6.1, 6.9, 5.5, 6.9, 7.3,
   4.5, 4.8, 5.8, 7.8, 7.4, 3.6, 4.4, 7.9, 7.1,
   6.8, 8.4, 7, 7.5, 7, 6.8, 5.9, 8.3, 6.7,
   6.2, 6.2, 8.9, 4.3, 7, 6.9, 5.2, 7.9)
head(cbind(mercury,pH))
```

```
##      mercury  pH
## [1,]    1.23 6.1
## [2,]    1.33 5.1
## [3,]    0.04 9.1
## [4,]    0.44 6.9
## [5,]    1.20 4.6
## [6,]    0.27 7.3
```

```
# Fit a linear regression
bass.lm <- lm(log(mercury) ~ pH)
# ANOVA table
anova(bass.lm)
```

```
## Analysis of Variance Table
##
## Response: log(mercury)
##           Df Sum Sq Mean Sq F value    Pr(>F)
## pH         1 13.961 13.9611  31.388 8.544e-07 ***
## Residuals 51 22.685  0.4448
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(bass.lm)
```

```
##
## Call:
## lm(formula = log(mercury) ~ pH)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.67936 -0.43150  0.09943  0.44216  1.37147
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.73999    0.48187   3.611 0.000696 ***
## pH          -0.40215    0.07178  -5.602 8.54e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6669 on 51 degrees of freedom
## Multiple R-squared:  0.381,  Adjusted R-squared:  0.3688
## F-statistic: 31.39 on 1 and 51 DF,  p-value: 8.544e-07
```

```r
# Produce a scatterplot
plot(x = pH, y = log(mercury),
main = "Scatterplot of mercury in largemouth bass vs. water pH\n with fitted regression line",
xlab = "Water pH", ylab = "Average mercury")
abline(bass.lm)
```

Scatterplot of mercury in largemouth bass vs. water pH with fitted regression line
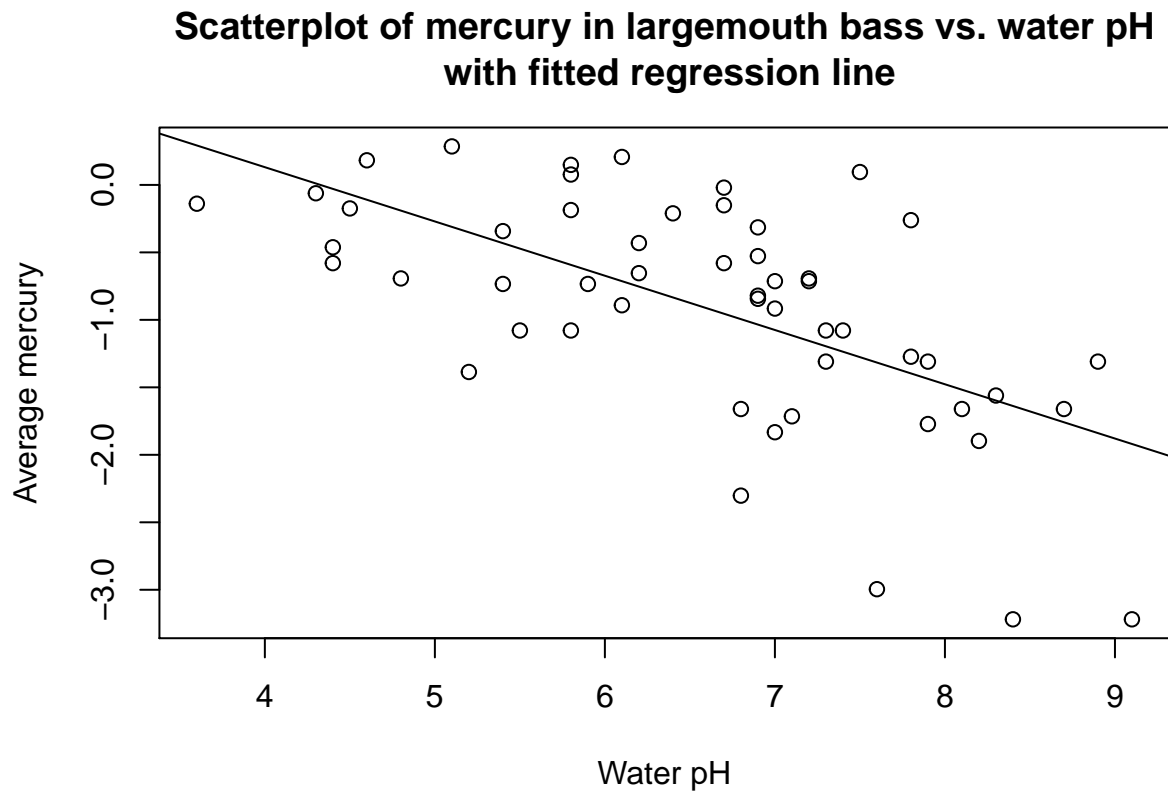

```r
# Studentized residuals versus fitted values
plot(x = bass.lm$fitted.values, y = rstudent(bass.lm),
main = "Studentized residuals vs. fitted values\n regression for largemouth bass data",
xlab = "Predicted value", ylab = "Studentized residual")
abline(h = 0)
abline(h = c(-2, 2), lty = 2)
```

Studentized residuals vs. fitted values
regression for largemouth bass data
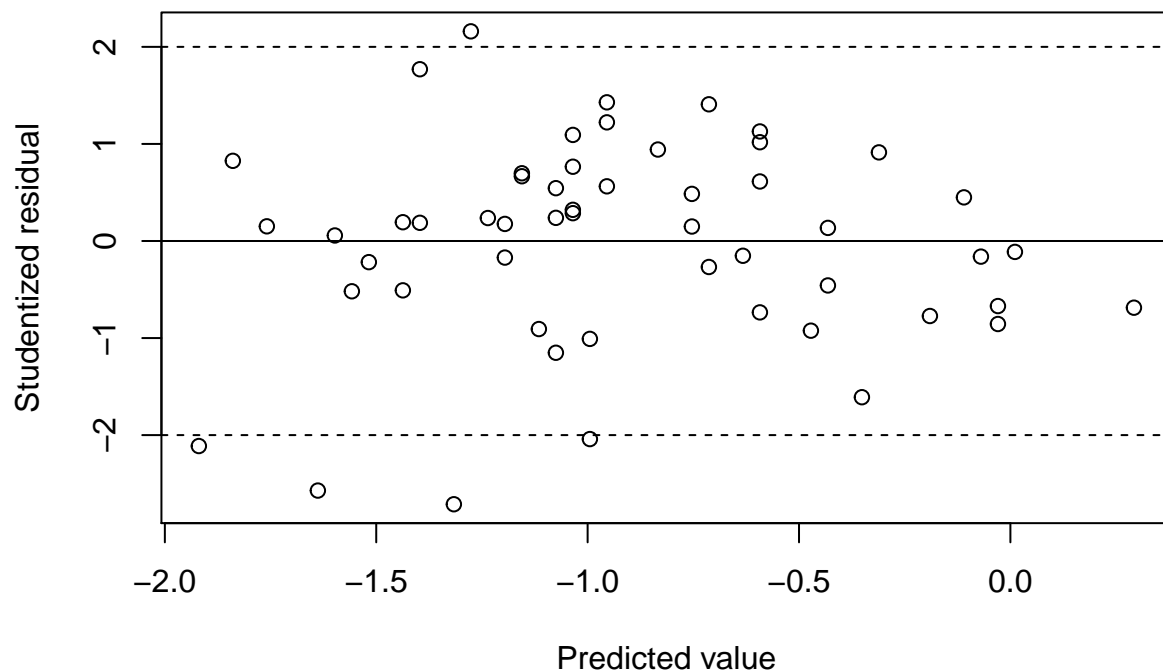
Normal Q-Q plot of residuals
```
# Normal Q-Q plot of residuals
qqnorm(bass.lm$residuals,
main = "Normal Q-Q plot of residuals\n regression for largemouth bass data")
qqline(bass.lm$residuals)
```

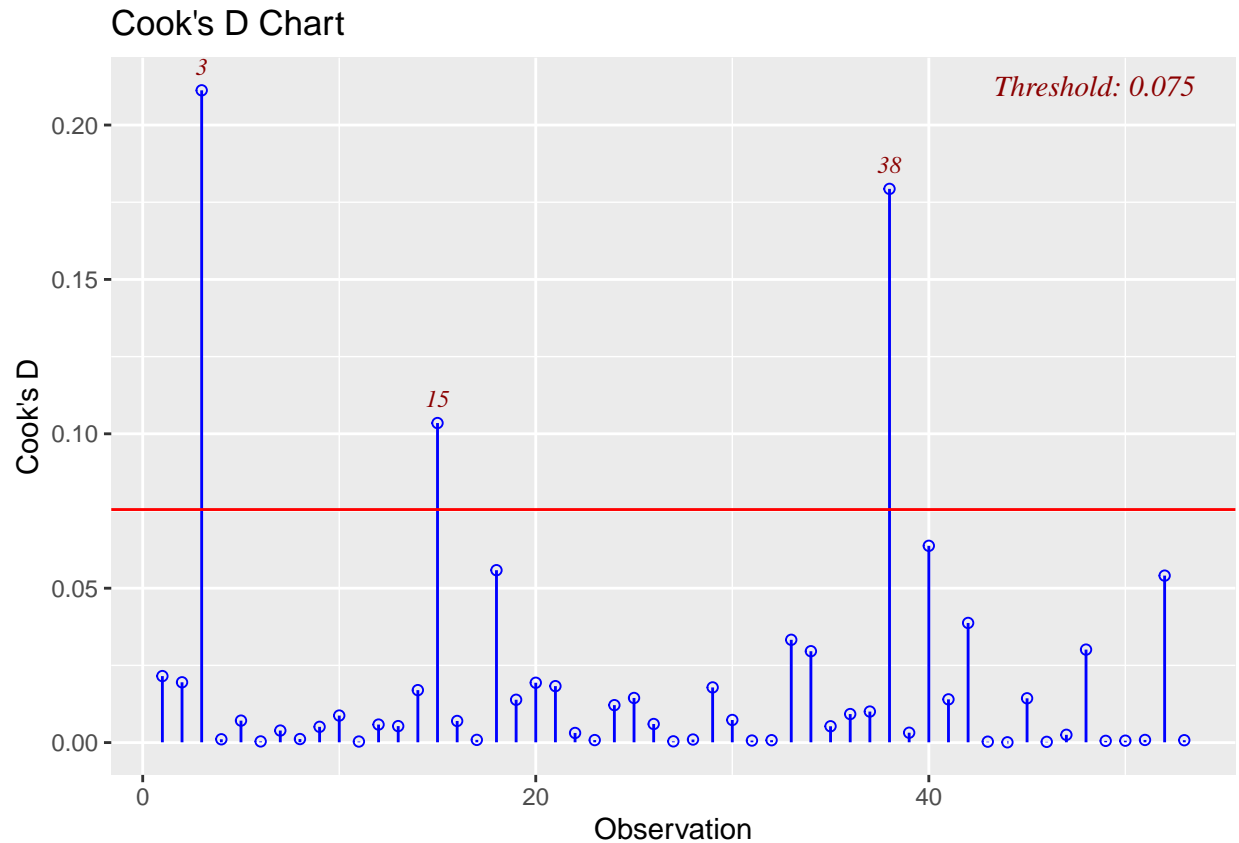**Normal Q–Q plot of residuals
regression for largemouth bass data**
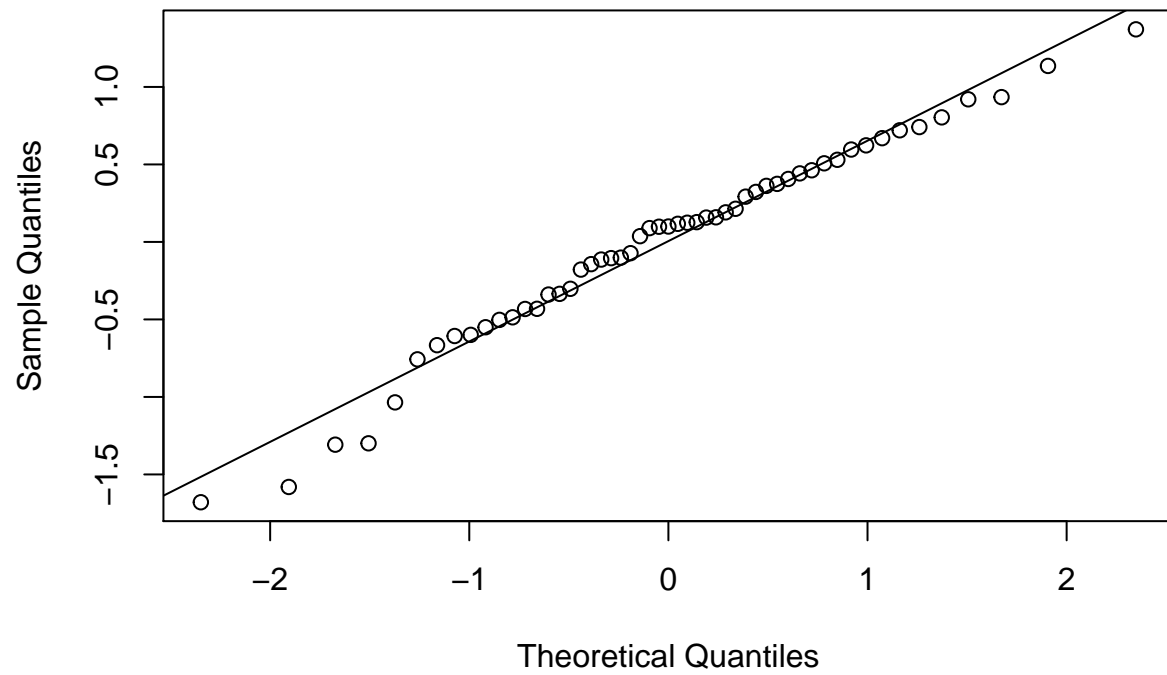

```r
# Plot Cook's distances for the largemouth bass data.
ols_plot_cooksd_chart(bass.lm)
```

Cook's D Chart


```r
# Obtain 95% confidence intervals for beta_0 and beta_1.
confint.default(lm(mercury ~ pH))
```

```
##                  2.5 %      97.5 %
## (Intercept)  1.1320800   1.92975741
## pH          -0.2117138  -0.09288796
```

```r
# Prediction
predict(lm(mercury ~ pH), newdata = data.frame(pH = 7),
        interval = "prediction")
```

```
##         fit        lwr      upr
## 1 0.4648127 -0.1064657 1.036091
```

mutiple linear regression:numerical variables (independent variables)

```r
loyn <- read.csv("loyn.csv")
lyon.lm <- lm(ABUND ~ log(AREA) + log(DIST) + GRAZE, data = loyn)
anova(lyon.lm) #F-test P-value
```

```
## Analysis of Variance Table
##
## Response: ABUND
##            Df Sum Sq Mean Sq F value    Pr(>F)
## log(AREA)   1 3471.0  3471.0 84.0053 1.929e-12 ***
## log(DIST)   1   65.5    65.5  1.5847 0.2137067
## GRAZE       1  652.9   652.9 15.8015 0.0002181 ***
## Residuals  52 2148.6    41.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(lyon.lm)
```

```
##
## Call:
## lm(formula = ABUND ~ log(AREA) + log(DIST) + GRAZE, data = loyn)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.227  -3.769  -0.163   3.037  13.256
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  26.6284     5.4254   4.908 9.49e-06 ***
## log(AREA)     3.1695     0.5806   5.459 1.36e-06 ***
## log(DIST)    -1.0750     0.9548  -1.126 0.265385
## GRAZE        -2.8269     0.7112  -3.975 0.000218 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.428 on 52 degrees of freedom
## Multiple R-squared:  0.661,  Adjusted R-squared:  0.6414
## F-statistic:  33.8 on 3 and 52 DF,  p-value: 2.921e-12
```

```r
# Fit each model in turn
#3 predictors
mod1.lm <- lm(ABUND ~ log(AREA) + log(DIST) + GRAZE, data = loyn)
summary(mod1.lm) #t-test P-value
```

```
##
## Call:
## lm(formula = ABUND ~ log(AREA) + log(DIST) + GRAZE, data = loyn)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.227  -3.769  -0.163   3.037  13.256
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  26.6284     5.4254   4.908 9.49e-06 ***
## log(AREA)     3.1695     0.5806   5.459 1.36e-06 ***
## log(DIST)    -1.0750     0.9548  -1.126 0.265385
## GRAZE        -2.8269     0.7112  -3.975 0.000218 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.428 on 52 degrees of freedom
## Multiple R-squared:  0.661,  Adjusted R-squared:  0.6414
## F-statistic:  33.8 on 3 and 52 DF,  p-value: 2.921e-12
```

```r
# 2 predictors
mod2.lm <- lm(ABUND ~ log(AREA) + log(DIST), data = loyn)
summary(mod2.lm)
```

```
##
## Call:
## lm(formula = ABUND ~ log(AREA) + log(DIST), data = loyn)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.189  -5.172   1.335   4.379  14.944
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  16.1343     5.3609   3.010    0.004 **
## log(AREA)     4.4316     0.5498   8.061 8.94e-11 ***
## log(DIST)    -1.2013     1.0793  -1.113    0.271
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.27 on 53 degrees of freedom
## Multiple R-squared:  0.558,  Adjusted R-squared:  0.5413
## F-statistic: 33.45 on 2 and 53 DF,  p-value: 4.019e-10
```

```r
#2 predictors
mod3.lm <- lm(ABUND ~ log(AREA) + GRAZE, data = loyn)
summary(mod3.lm)
```

```
##
## Call:
## lm(formula = ABUND ~ log(AREA) + GRAZE, data = loyn)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.4296  -4.3186  -0.6323   4.1273  13.0739
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  21.6029     3.0917   6.987 4.73e-09 ***
## log(AREA)     2.9923     0.5603   5.341 1.98e-06 ***
## GRAZE        -2.8535     0.7125  -4.005 0.000195 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.444 on 53 degrees of freedom
## Multiple R-squared:  0.6527, Adjusted R-squared:  0.6396
## F-statistic: 49.81 on 2 and 53 DF,  p-value: 6.723e-13
```

```
#2 predictors
mod4.lm <- lm(ABUND ~ log(AREA) + GRAZE, data = loyn)
summary(mod4.lm)
```

```
##
## Call:
## lm(formula = ABUND ~ log(AREA) + GRAZE, data = loyn)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -13.4296  -4.3186  -0.6323   4.1273  13.0739
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  21.6029     3.0917   6.987 4.73e-09 ***
## log(AREA)     2.9923     0.5603   5.341 1.98e-06 ***
## GRAZE        -2.8535     0.7125  -4.005 0.000195 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.444 on 53 degrees of freedom
## Multiple R-squared:  0.6527, Adjusted R-squared:  0.6396
## F-statistic: 49.81 on 2 and 53 DF,  p-value: 6.723e-13
```

```
# 1 predictor
mod5.lm <- lm(ABUND ~ log(AREA), data = loyn)
summary(mod5.lm)
```

```
##
## Call:
## lm(formula = ABUND ~ log(AREA), data = loyn)
##
## Residuals:
##     Min      1Q  Median      3Q      Max
## -13.380  -6.119   1.372   4.631  14.255
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.4014     1.4894   6.984 4.38e-09 ***
## log(AREA)     4.2467     0.5252   8.086 7.18e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.286 on 54 degrees of freedom
## Multiple R-squared:  0.5477, Adjusted R-squared:  0.5393
## F-statistic: 65.38 on 1 and 54 DF,  p-value: 7.178e-11
```

```
# 1 predictor
mod6.lm <- lm(ABUND ~ log(DIST), data = loyn)
summary(mod6.lm)
```

```
##
```

```
## Call:
## lm(formula = ABUND ~ log(DIST), data = loyn)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.095  -8.104   1.789   8.858  17.537
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   12.229      7.892   1.550    0.127
## log(DIST)      1.428      1.521   0.939    0.352
##
## Residual standard error: 10.75 on 54 degrees of freedom
## Multiple R-squared:  0.01606,    Adjusted R-squared:  -0.002162
## F-statistic: 0.8813 on 1 and 54 DF,  p-value: 0.352
```

```
# 1 predictor
mod7.lm <- lm(ABUND ~ GRAZE, data = loyn)
summary(mod7.lm)
```

```
##
## Call:
## lm(formula = ABUND ~ GRAZE, data = loyn)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.1066  -5.4097   0.0934   4.4856  18.2747
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  34.3692     2.4095  14.264  < 2e-16 ***
## GRAZE        -4.9813     0.7259  -6.862  6.9e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.918 on 54 degrees of freedom
## Multiple R-squared:  0.4658, Adjusted R-squared:  0.4559
## F-statistic: 47.09 on 1 and 54 DF,  p-value: 6.897e-09
```

```
# null model
mod8.lm <- lm(ABUND ~ NULL, data = loyn)
summary(mod8.lm)
```

```
##
## Call:
## lm(formula = ABUND ~ NULL, data = loyn)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.014  -7.114   1.536   8.786  20.086
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)    19.514       1.434     13.6    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.73 on 55 degrees of freedom
```

#ANCOVA: categorical and numerical variables

```r
Height1 <- c(154, 155, 157, 157, 158, 163, 165, 168, 171, 173,
             156, 160, 163, 168, 168, 169, 173, 175, 178, 179)
Height2 <- c(157, 160, 159, 161, 160, 166, 166, 171, 172, 173,
             162, 163, 166, 172, 170, 175, 176, 179, 179, 182)
gain <- Height2-Height1
Sex <- factor(rep(c("Female", "Male"), each = 10))
# scatter of data
plot(x = Height1, y = gain, xlab = "Height1",
     ylab = "gain", col = as.numeric(factor(Sex)),
     pch = c(rep(16,times=10), rep(15,times=10)))
legend("topright", legend = unique(Sex),
       col = unique(as.numeric(factor(Sex))),
        pch = c(16,15))
```
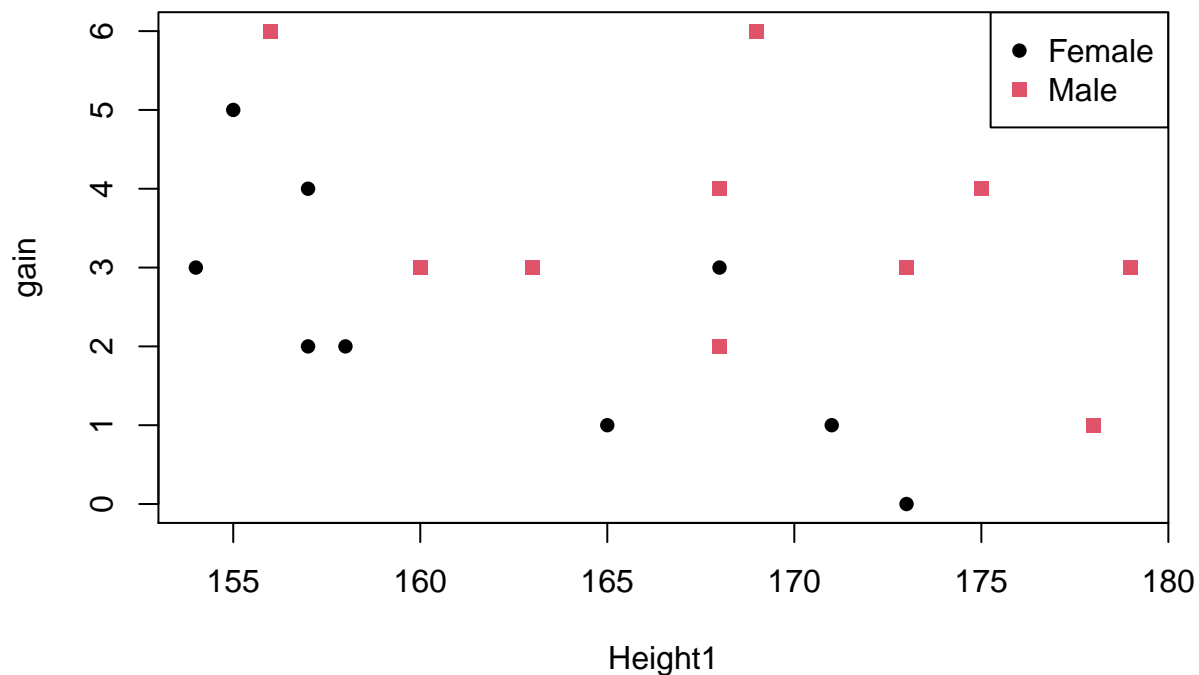


```r
# Enter Height1 (the covariate) first in this model
height.gain.lm1 <- lm(gain ~ Height1*Sex) #carrot.lm <- lm(increase ~ initial * factor(fertiliser))
# ANOVA table output for the ANCOVA.
anova(height.gain.lm1)
```
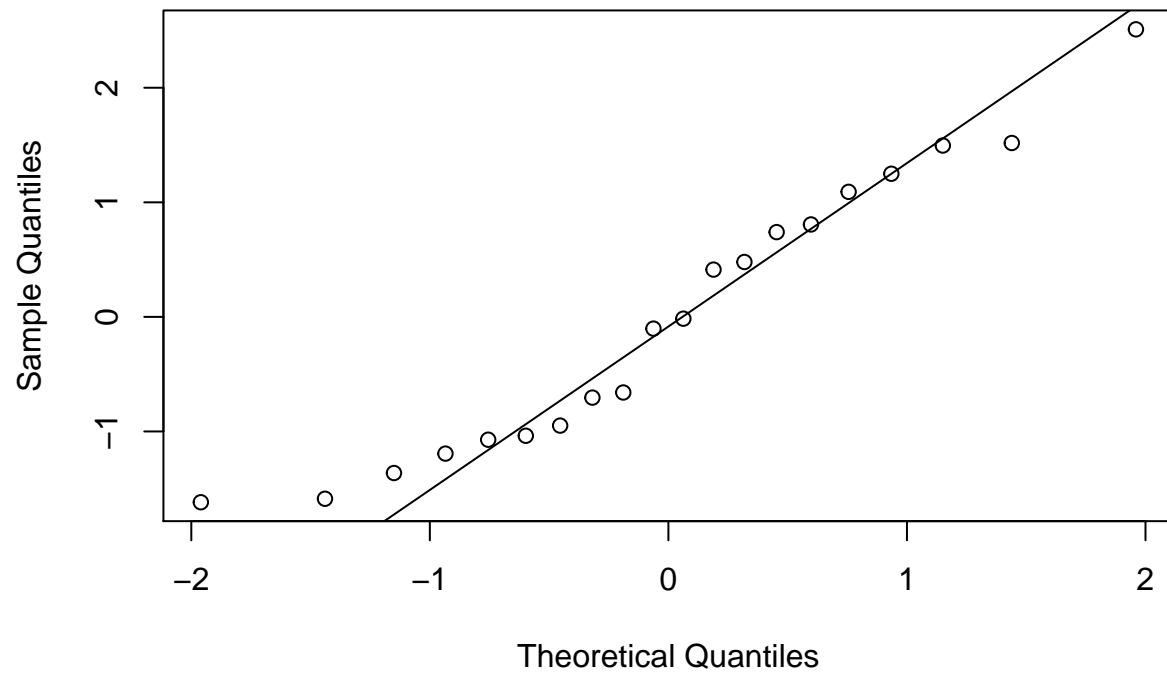
```
## Analysis of Variance Table
##
## Response: gain
##             Df  Sum Sq Mean Sq F value   Pr(>F)
## Height1      1  5.3607  5.3607  3.0991 0.097436 .
## Sex          1 15.1076 15.1076  8.7339 0.009307 **
## Height1:Sex  1  0.8055  0.8055  0.4656 0.504756
## Residuals   16 27.6763  1.7298
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#studentized residuals versus fitted values
plot(x = height.gain.lm1$fitted.values, y = rstudent(height.gain.lm1), main = "Studentized
residuals vs. fitted values\n ANCOVA for height gain data",
xlab = "Predicted value", ylab = "Studentized residual")
abline(h = 0)
abline(h = c(-2, 2), lty = 2)
```

**Studentized
residuals vs. fitted values
ANCOVA for height gain data**
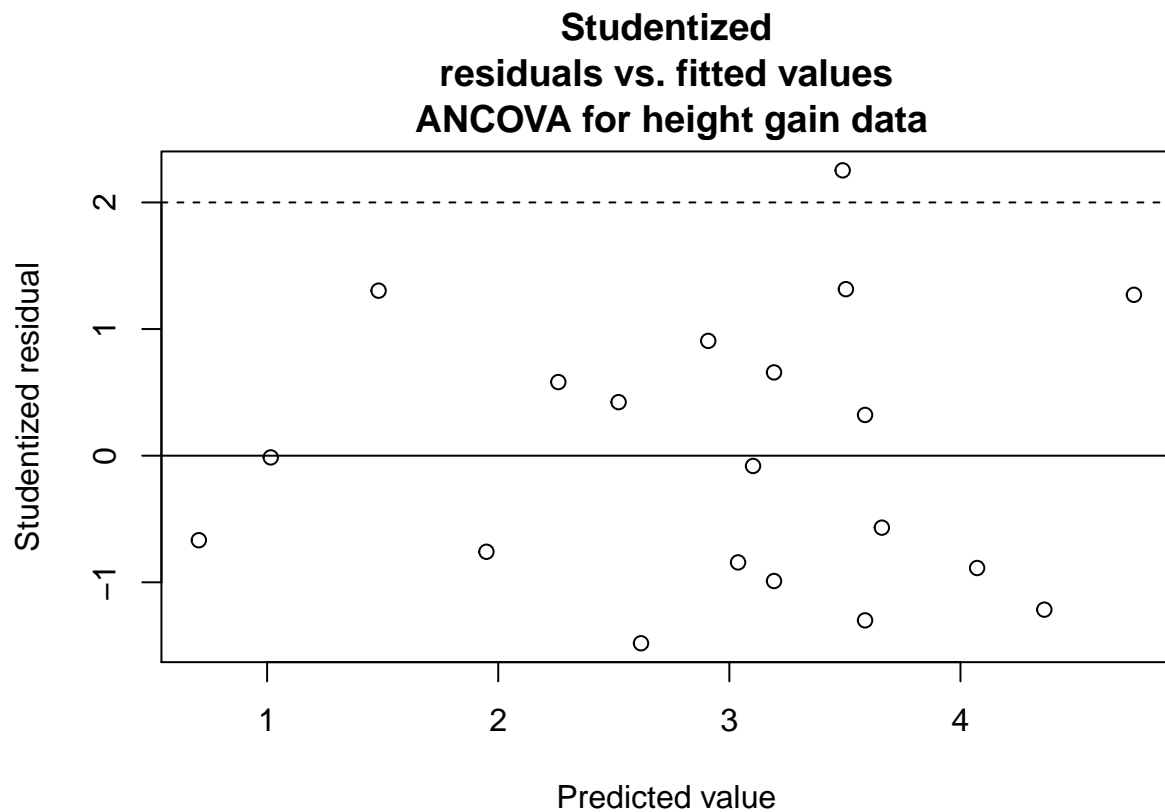

```r
#normal Q-Q
qqnorm(height.gain.lm1$residuals,
main = "Normal Q-Q plot of residuals\n ANCOVA for height gain data")
qqline(height.gain.lm1$residuals)
```
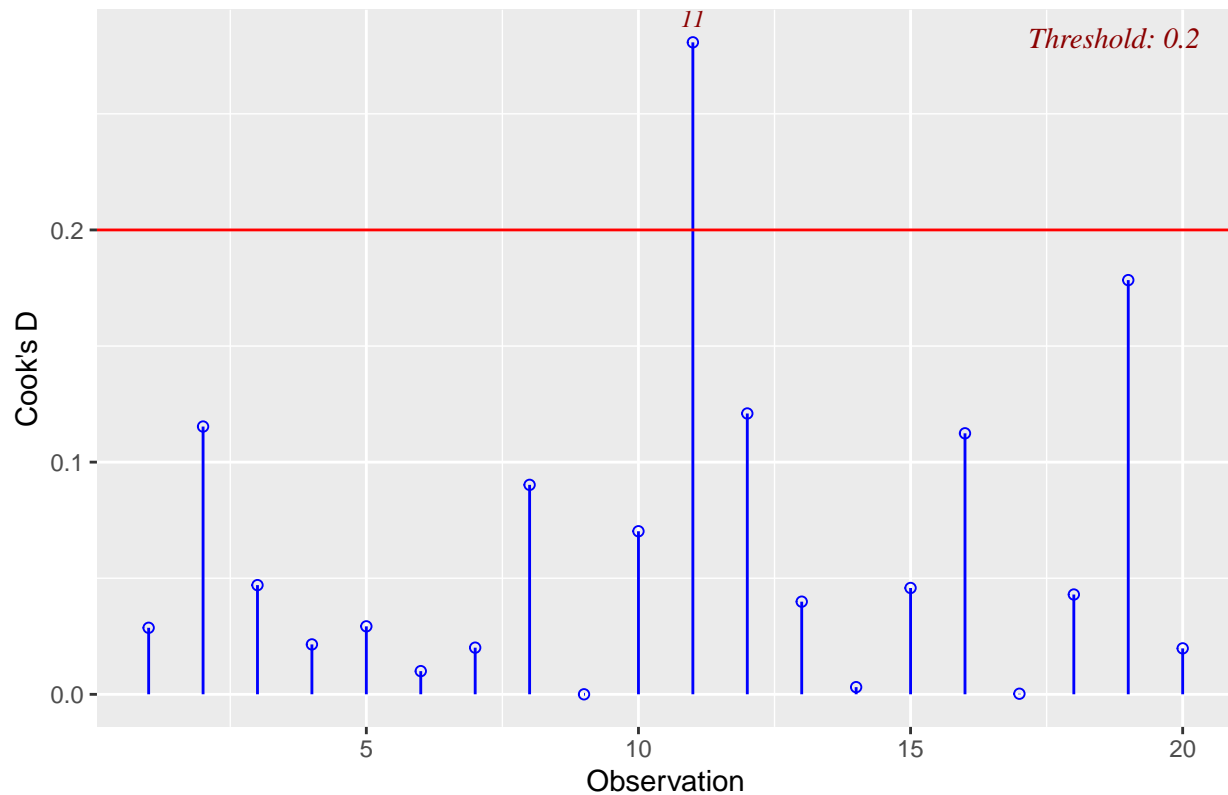
Normal Q–Q plot of residuals
ANCOVA for height gain data


```
#Cook's distances
ols_plot_cooksd_chart(height.gain.lm1)
```

Cook's D Chart


```r
# Enter Height1 (the covariate) second in this model
height.gain.lm2 <- lm(gain ~ Sex*Height1)
# ANOVA table output for the ANCOVA.
anova(height.gain.lm2)
```

```
## Analysis of Variance Table
##
## Response: gain
##              Df  Sum Sq Mean Sq F value  Pr(>F)
## Sex           1  6.0500  6.0500  3.4976 0.07987 .
## Height1       1 14.4182 14.4182  8.3354 0.01072 *
## Sex:Height1   1  0.8055  0.8055  0.4656 0.50476
## Residuals    16 27.6763  1.7298
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANCOVA: Exploratory Analysis - no specific hypotheses

```r
fertiliser <- rep(c("Growmost", "Superior"), each = 10)
initial <- c(40, 43, 43, 45, 48, 51, 52, 57, 59, 65,
             42, 44, 46, 52, 52, 57, 59, 63, 68, 69)
increase <- c(6.2, 8.7, 5.6, 7, 4.5, 3.2, 4.8, 2.2, 1.3, 2,
              7.1, 9.3, 8.2, 8.9, 7.7, 7.2, 7, 5.5, 3.2, 3.9)
#Interactive model (Non-parallel lines)
```

```
carrot.m1 <- lm(increase ~ initial+factor(fertiliser)+initial * factor(fertiliser))
anova(carrot.m1)    # F-test P-value, SSE & MSE
```

```
## Analysis of Variance Table
##
## Response: increase
##                           Df Sum Sq Mean Sq F value    Pr(>F)
## initial                    1 38.542  38.542 26.3359 0.0001004 ***
## factor(fertiliser)         1 49.586  49.586 33.8824 2.604e-05 ***
## initial:factor(fertiliser) 1  2.073   2.073  1.4168 0.2512909
## Residuals                 16 23.416   1.463
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(carrot.m1) # t-test P-value, adj R & R-squared
```

```
##
## Call:
## lm(formula = increase ~ initial + factor(fertiliser) + initial *
##     factor(fertiliser))
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -2.05129 -0.85817 -0.07469  0.75969  2.28359
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       17.41036    2.54243   6.848 3.92e-06 ***
## initial                           -0.25567    0.04997  -5.117 0.000104 ***
## factor(fertiliser)Superior        -0.77769    3.45459  -0.225 0.824737
## initial:factor(fertiliser)Superior 0.07755    0.06515   1.190 0.251291
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.21 on 16 degrees of freedom
## Multiple R-squared:  0.7939, Adjusted R-squared:  0.7553
## F-statistic: 20.55 on 3 and 16 DF,  p-value: 9.819e-06
```

```
#Parallel lines model
carrot.m2 <- lm(increase ~ initial+factor(fertiliser))#1. Initial (the covariate) specified first
anova(carrot.m2)
```

```
## Analysis of Variance Table
##
## Response: increase
##                    Df Sum Sq Mean Sq F value    Pr(>F)
## initial             1 38.542  38.542  25.706 9.466e-05 ***
## factor(fertiliser)  1 49.586  49.586  33.072 2.353e-05 ***
## Residuals          17 25.489   1.499
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(carrot.m2) # for adj R & R-squared
```

```
##
## Call:
## lm(formula = increase ~ initial + factor(fertiliser))
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -2.47268 -0.62767  0.07307  0.64984  2.61663
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               15.11558    1.67764   9.010 6.98e-08 ***
## initial                   -0.21005    0.03245  -6.473 5.74e-06 ***
## factor(fertiliser)Superior  3.27925    0.57023   5.751 2.35e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.224 on 17 degrees of freedom
## Multiple R-squared:  0.7757, Adjusted R-squared:  0.7493
## F-statistic: 29.39 on 2 and 17 DF,  p-value: 3.039e-06
```

```
carrot.m3 <- lm(increase ~ factor(fertiliser)+initial)# 2. Fertiliser (the factor) specified first
anova(carrot.m3)
```

```
## Analysis of Variance Table
##
## Response: increase
##                    Df Sum Sq Mean Sq F value    Pr(>F)
## factor(fertiliser)  1 25.313  25.313  16.882 0.0007327 ***
## initial             1 62.816  62.816  41.895 5.735e-06 ***
## Residuals          17 25.489   1.499
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(carrot.m3) # for adj R & R-squared
```

```
##
## Call:
## lm(formula = increase ~ factor(fertiliser) + initial)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -2.47268 -0.62767  0.07307  0.64984  2.61663
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               15.11558    1.67764   9.010 6.98e-08 ***
## factor(fertiliser)Superior  3.27925    0.57023   5.751 2.35e-05 ***
## initial                   -0.21005    0.03245  -6.473 5.74e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 1.224 on 17 degrees of freedom
## Multiple R-squared:  0.7757, Adjusted R-squared:  0.7493
## F-statistic: 29.39 on 2 and 17 DF,  p-value: 3.039e-06
```

```r
#One-way ANOVA of fertiliser factor
carrot.m4 <- lm(increase ~ factor(fertiliser))
anova(carrot.m4)
```

```
## Analysis of Variance Table
##
## Response: increase
##                    Df Sum Sq Mean Sq F value  Pr(>F)
## factor(fertiliser)  1 25.313 25.3125  5.1597 0.03562 *
## Residuals          18 88.305  4.9058
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(carrot.m4) # for adj R & R-squared
```

```
##
## Call:
## lm(formula = increase ~ factor(fertiliser))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.600 -1.600  0.275  1.462  4.150
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                4.5500     0.7004   6.496 4.15e-06 ***
## factor(fertiliser)Superior  2.2500     0.9905   2.271   0.0356 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.215 on 18 degrees of freedom
## Multiple R-squared:  0.2228, Adjusted R-squared:  0.1796
## F-statistic:  5.16 on 1 and 18 DF,  p-value: 0.03562
```

```r
#The regression model with x = initial yield
carrot.m5 <- lm(increase ~ initial)
anova(carrot.m5)
```

```
## Analysis of Variance Table
##
## Response: increase
##           Df Sum Sq Mean Sq F value   Pr(>F)
## initial    1 38.542  38.542  9.2408 0.007046 **
## Residuals 18 75.075   4.171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(carrot.m5) # for adj R & R-squared
```

```
##
## Call:
## lm(formula = increase ~ initial)
##
## Residuals:
##     Min     1Q  Median     3Q    Max
## -3.3874 -1.6465  0.0175  1.5899  3.1065
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.00993    2.77964    5.04 8.51e-05 ***
## initial     -0.15801    0.05198   -3.04  0.00705 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.042 on 18 degrees of freedom
## Multiple R-squared:  0.3392, Adjusted R-squared:  0.3025
## F-statistic: 9.241 on 1 and 18 DF,  p-value: 0.007046
```

```
#The null model with only an overall mean
carrot.m6 <- lm(increase ~ NULL)
anova(carrot.m6)
```

```
## Analysis of Variance Table
##
## Response: increase
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 19 113.62  5.9799
```

```
summary(carrot.m6) # for adj R & R-squared
```

```
##
## Call:
## lm(formula = increase ~ NULL)
##
## Residuals:
##     Min     1Q Median     3Q    Max
## -4.375 -1.950  0.225  1.650  3.625
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.6750     0.5468   10.38 2.88e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.445 on 19 degrees of freedom
```