UCC_Course_Problems.R

akane

Fri Jan 13 09:39:06 2017

```r
# Exercise 1
height <- c(163, 185, 155, 195, 168, 198, 200, 146, 179, 160, 180, 170, 190)
weight <- c(65, 85, 70, 120, 73, 100, 103, 50, 81, 64, 90, 78, 71)
# mean
mean(height)
```

```
## [1] 176.0769
```

```r
mean (weight)
```

```
## [1] 80.76923
```

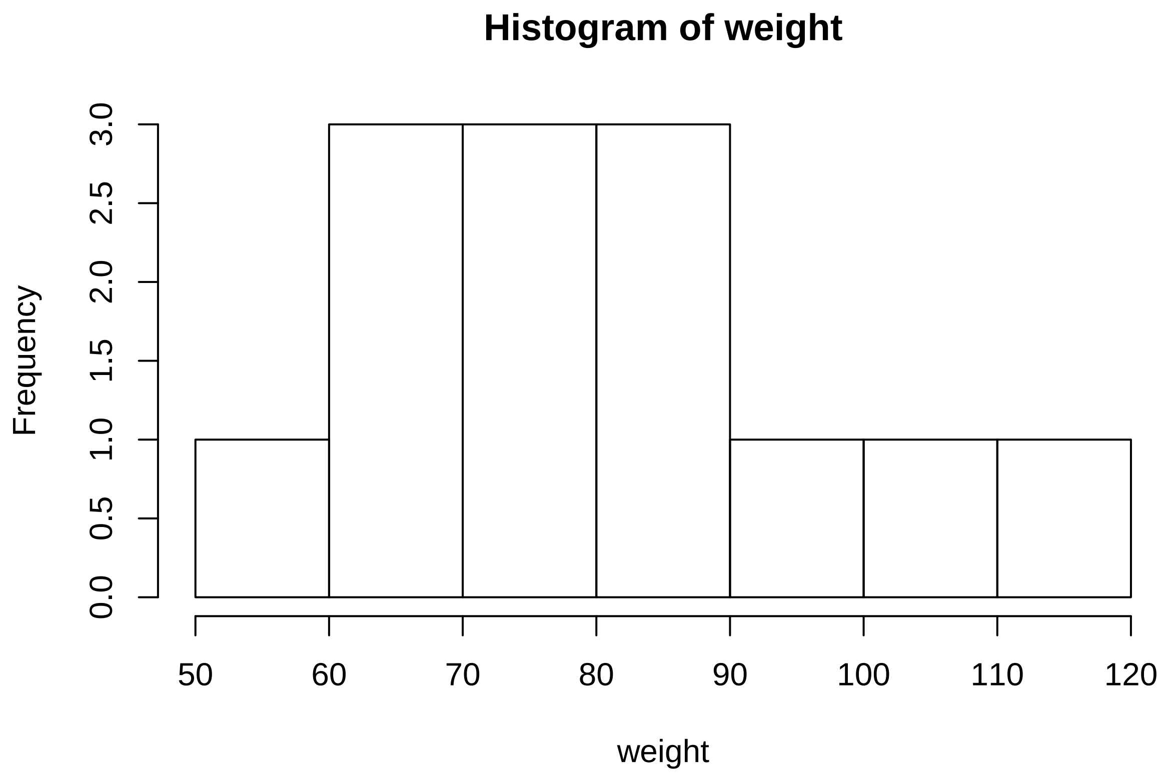
```r
# sd
sd(height)
```

```
## [1] 17.29384
```

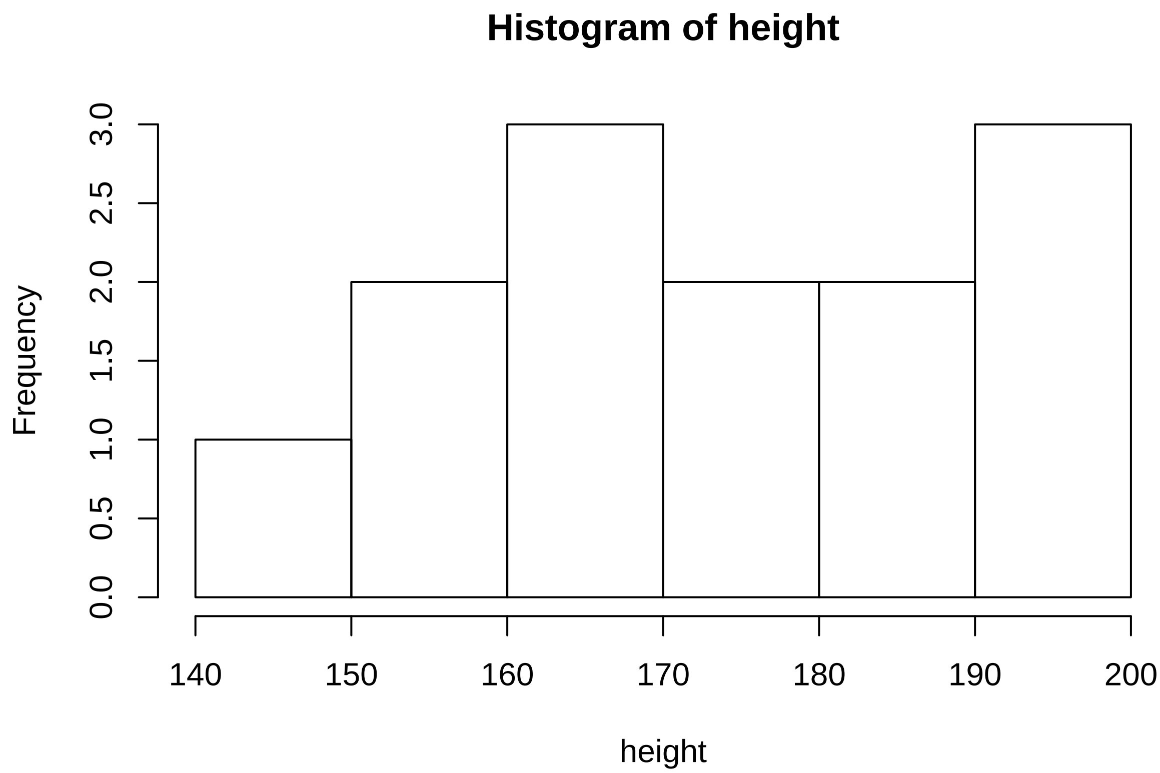
```r
sd(weight)
```

```
## [1] 18.84655
```

```r
# plot - histogram
hist(weight)
```




```r
hist(height)
```

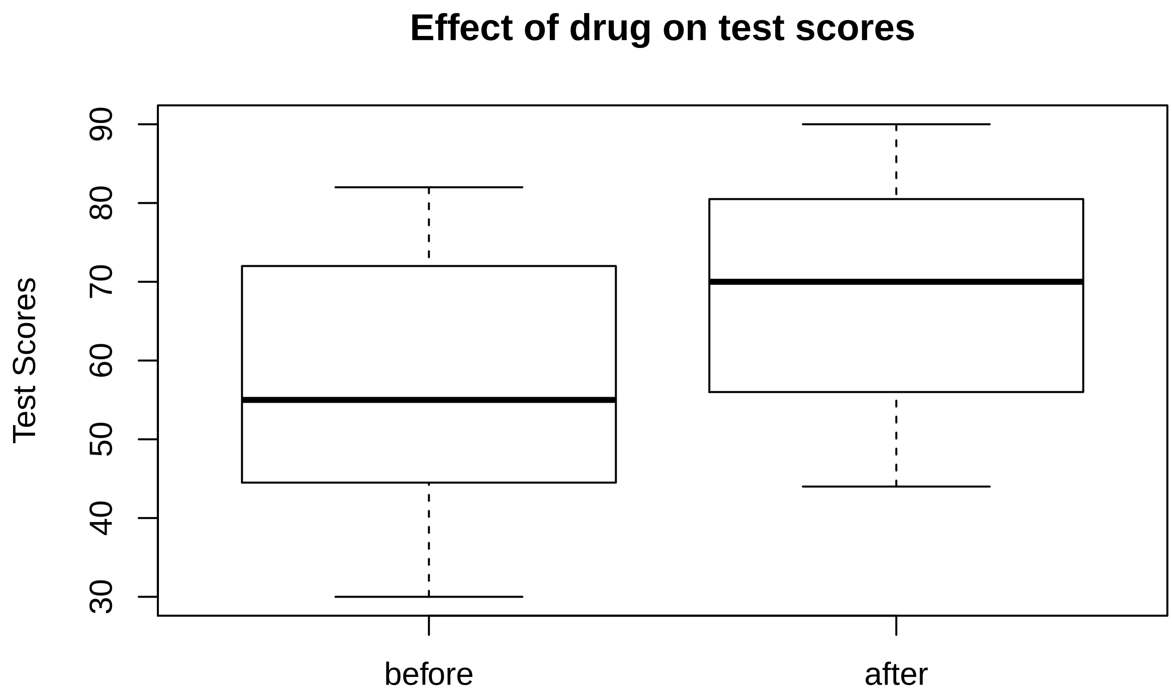



```r
# Exercise 2
#The following distribution has a mean of 5 and a standard deviation of 1.
# Using **pnorm** find the area under the curve from negative infinity to 4.
pnorm(4,mean=5,sd = 1)
```

```
## [1] 0.1586553
```

```r
# How many standard deviations is 4 away from the mean? 1

#Exercise 3 t test
before<-c(30,68,45,60,79,40,55,49,82,44,76)
after<-c(70,68,50,59,81,44,78,53,90,87,80)
boxplot(before, after, ylab="Test Scores",
        names=c("before","after"),
        main="Effect of drug on test scores")
```




```r
mean(before)
```

```
## [1] 57.09091
```

```r
mean(after)
```

```
## [1] 69.09091
```

```r
t.test(before,after, alternative = "less" )
```

```
##
##  Welch Two Sample t-test
##
## data:  before and after
## t = -1.7042, df = 19.803, p-value = 0.05199
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf 0.1501547
## sample estimates:
## mean of x mean of y
##  57.09091  69.09091
```

```r
t.test(before,after, alternative = "less" , paired = T)
```

```
##
##  Paired t-test
##
## data:  before and after
## t = -2.4992, df = 10, p-value = 0.01574
```

```
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -3.297441
## sample estimates:
## mean of the differences
##                     -12
```

```r
#Note that the t.test() function always subtracts first group minus second group. Since the "before" we
mean(before) - mean(after)
```

```
## [1] -12
```

```r
# Exercise 4
#To estimate the mean amount spent by customers in a restaurant, data was collected for 75 customers. W

# 1. At the 95% confidence, what is the margin of error?
# 2. If the sample mean is 20 euro, what is the 95% confidence interval for the population mean (all cu
n<-75
sd<-4
mu<-20
qnorm(0.975)
```

```
## [1] 1.959964
```

```r
margin<-qnorm(0.975)*(sd/sqrt(n));margin
```

```
## [1] 0.9052686
```

```r
mu+margin
```

```
## [1] 20.90527
```

```r
mu-margin
```

```
## [1] 19.09473
```

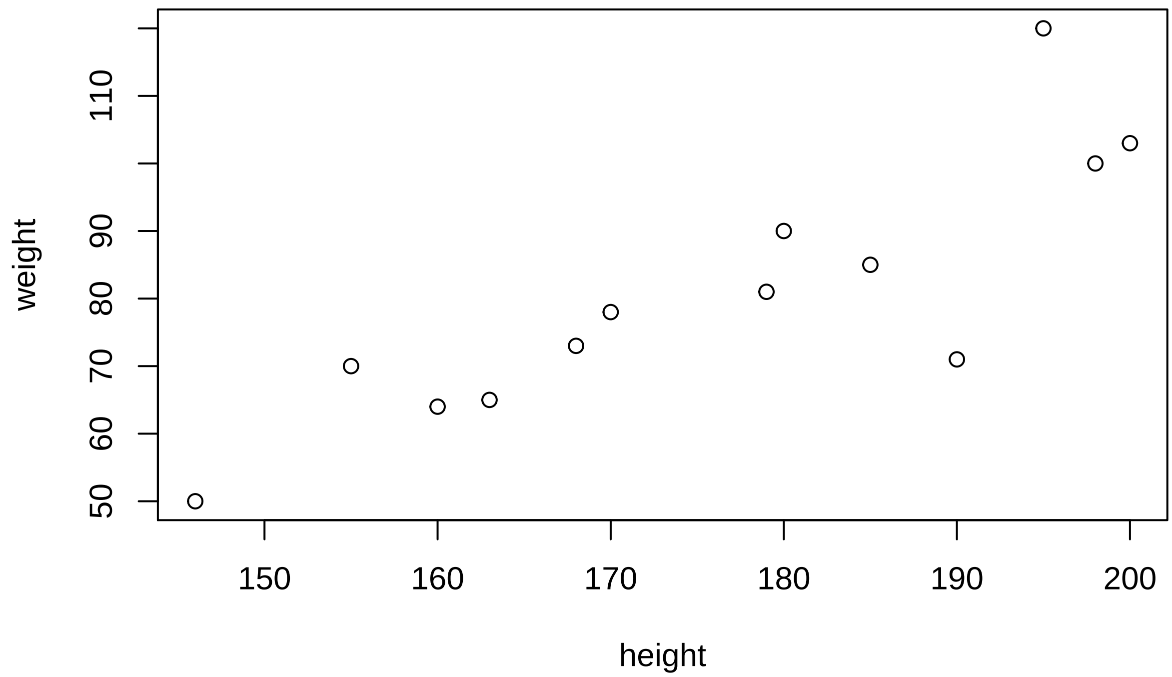
```r
#All samples of n = 75 will have 0.91 as the margin
#of error, assuming the population standard deviation
#is known to be 4.

#95% of all intervals using a sample mean plus or minus
#the margin of error will contain the unknown population
#mean.

# If we take 100 samples of n = 75, and make intervals of their
# sample mean +/- 0.91, 95 of them will contain the population mean

# That's what we mean when we say we're 95% confident

# Exercise 5
# create 2 vectors of numerical data
plot(height, weight)
```



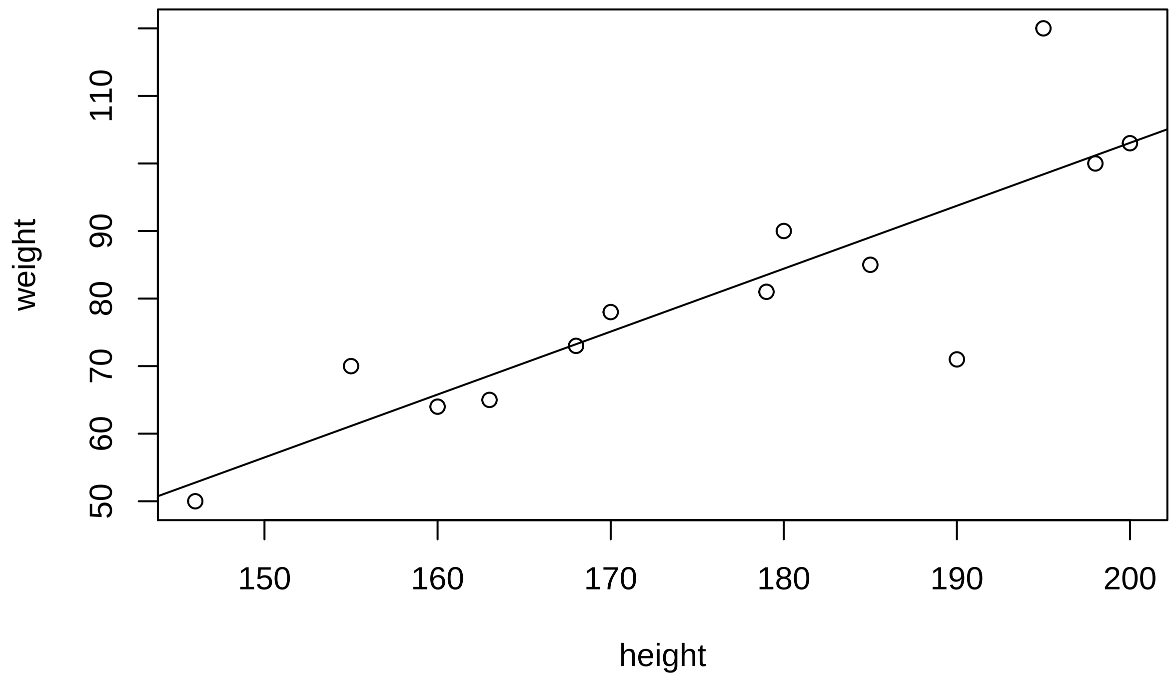

```r
# covariance and correlation of the two variables
cov(height, weight)
```

```
## [1] 278.5192
```

```r
cor.test(height, weight)
```

```
##
##  Pearson's product-moment correlation
##
## data:  height and weight
## t = 5.4568, df = 11, p-value = 0.0001988
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5736507 0.9555940
## sample estimates:
##       cor
## 0.8545391
```

```r
# Exercise 6
# create 2 vectors of numerical data
height <- c(163, 185, 155, 195, 168, 198, 200, 146, 179, 160, 180, 170, 190)
weight <- c(65, 85, 70, 120, 73, 100, 103, 50, 81, 64, 90, 78, 71)
plot(weight~height)
m1<-lm(weight~height)
abline(m1)
```




```r
print(m1)
```

```
##
## Call:
## lm(formula = weight ~ height)
##
## Coefficients:
## (Intercept)       height
##    -83.2047       0.9313
```

```r
y<-coef(m1)["(Intercept)"] + coef(m1)["height"]*height[1];y
```

```
## (Intercept)
##    68.59118
```

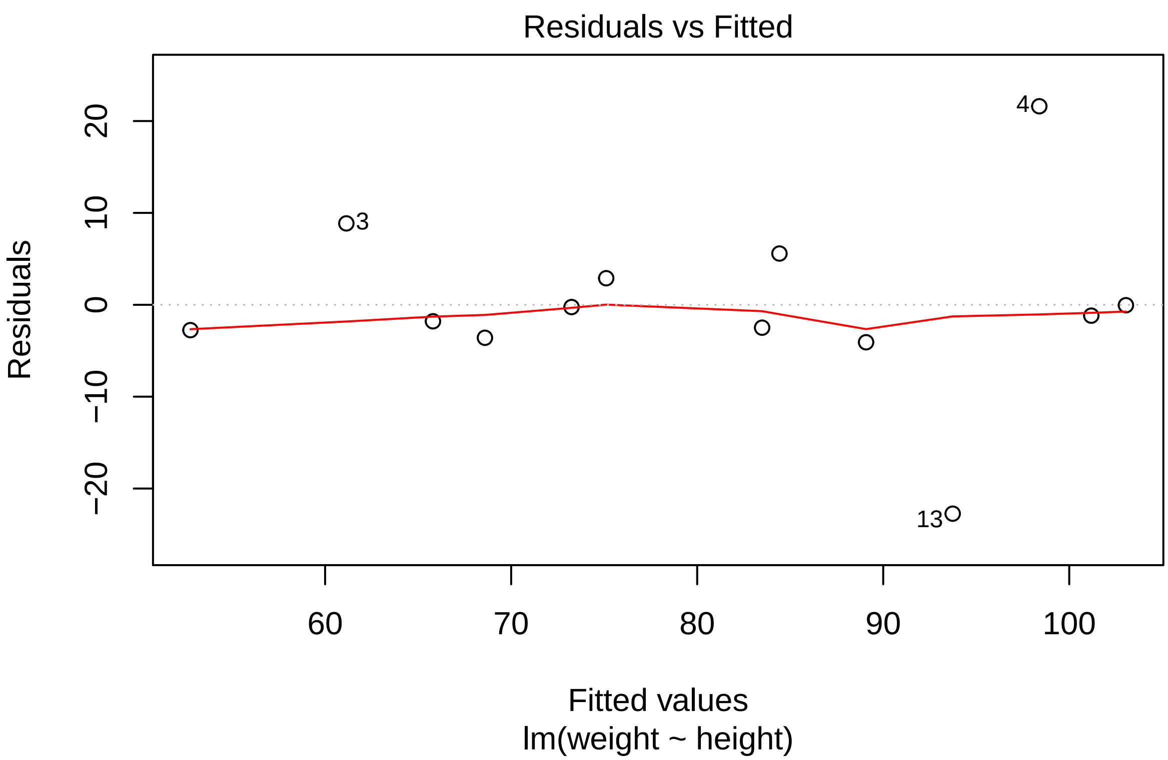
```r
resid(m1)
```

```
##            1            2            3            4            5
##  -3.59117798  -4.07896091   8.85892490  21.60841049  -0.24749228
##            6            7            8            9           10
##  -1.18537809  -0.04790381  -2.75970936  -2.49138374  -1.79738940
##           11           12           13
##   5.57735340   2.88998200 -22.73527521
```

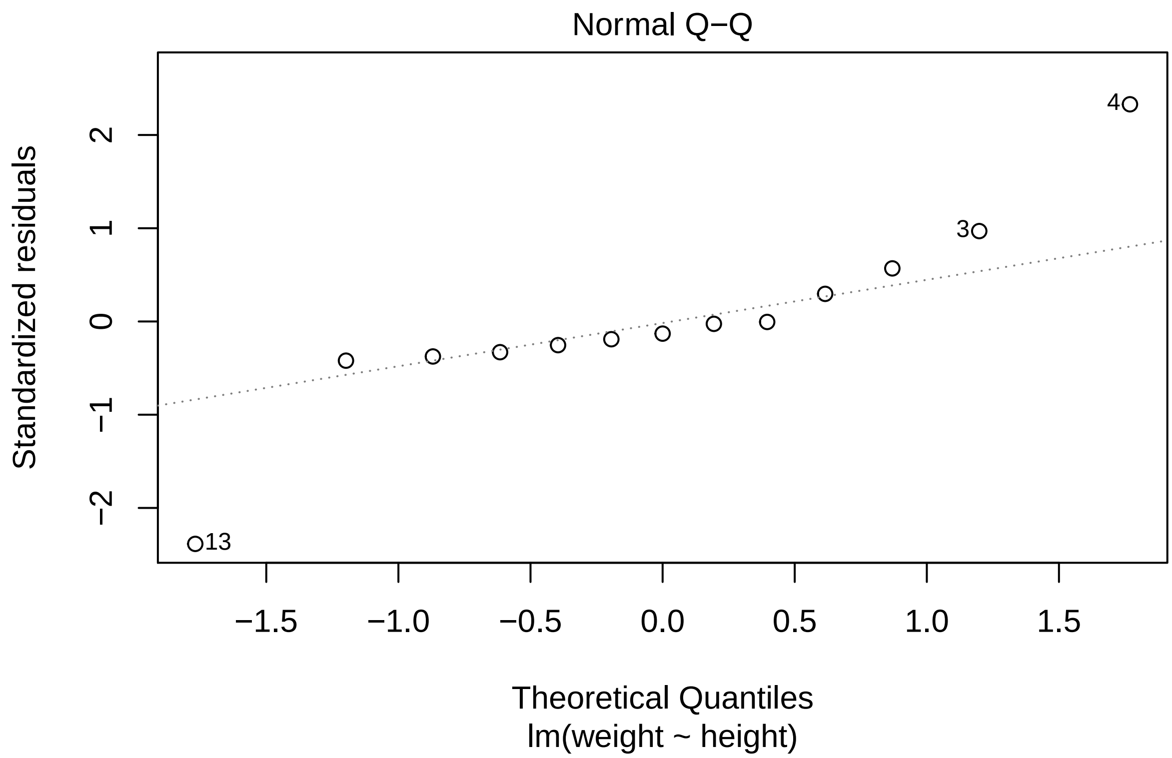
```r
y - 3.59117798
```

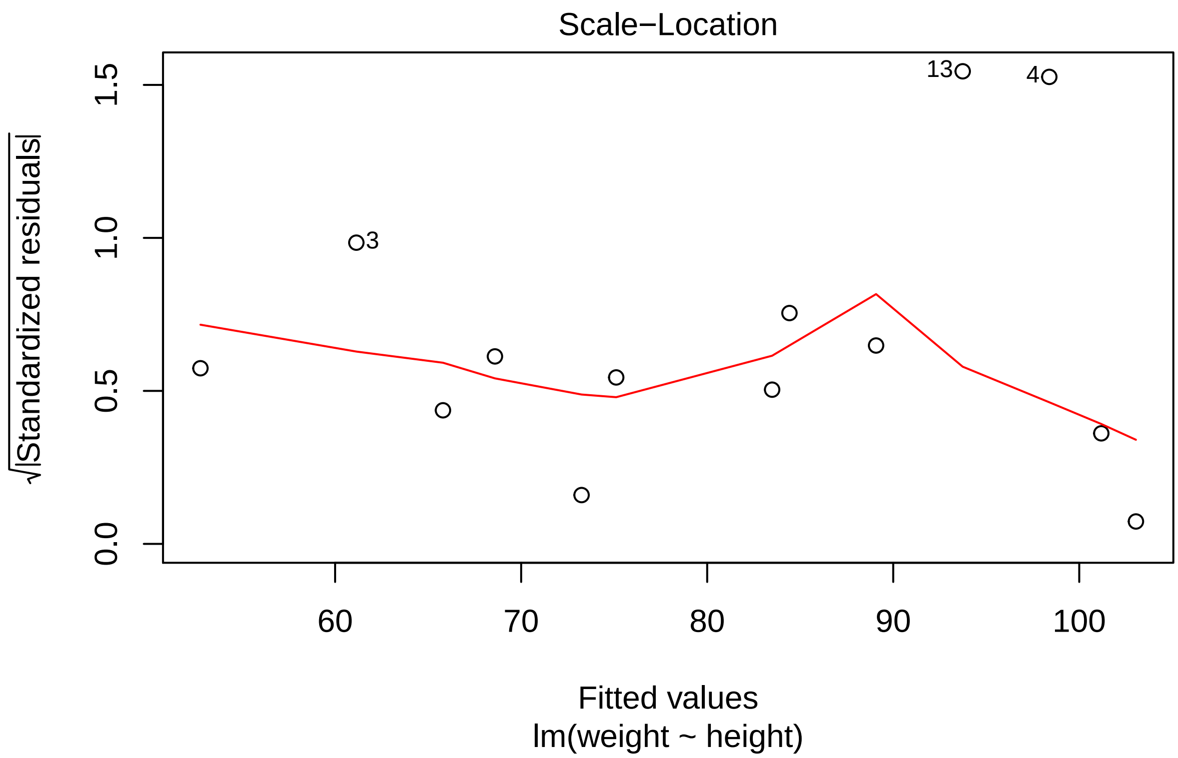
```
## (Intercept)
##          65
```

```r
plot(m1)
```

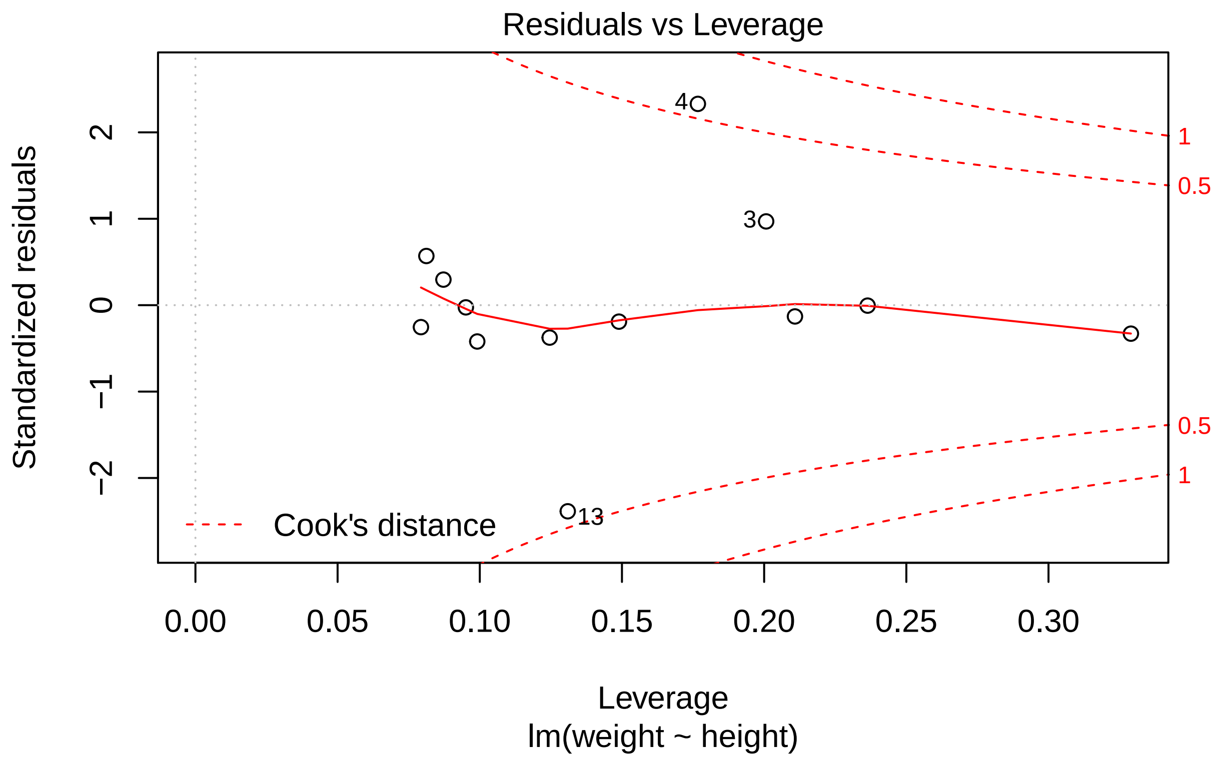


Residuals vs Fitted
Residuals
20
10
0
−10
−20
3
4
13
60
70
80
90
100
Fitted values
lm(weight ~ height)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(weight ~ height)
−2
−1
0
1
2
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
4
3
13



Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
60
70
80
90
100
Fitted values
lm(weight ~ height)
3
13
4




```r
# Exercise 7 Chi-square test of independence
men = c(100, 120, 60)
women = c(350, 200, 90)

Observed <- matrix(c(100,120,60,350,200,90),nrow = 2,ncol = 3, byrow = T)
chisq.test(Observed)
```

```
##
##  Pearson's Chi-squared test
##
## data:  Observed
## X-squared = 28.362, df = 2, p-value = 6.938e-07
```

```r
#It provides strong evidence to suggest that men and women tend to have difference preferences for ice
```