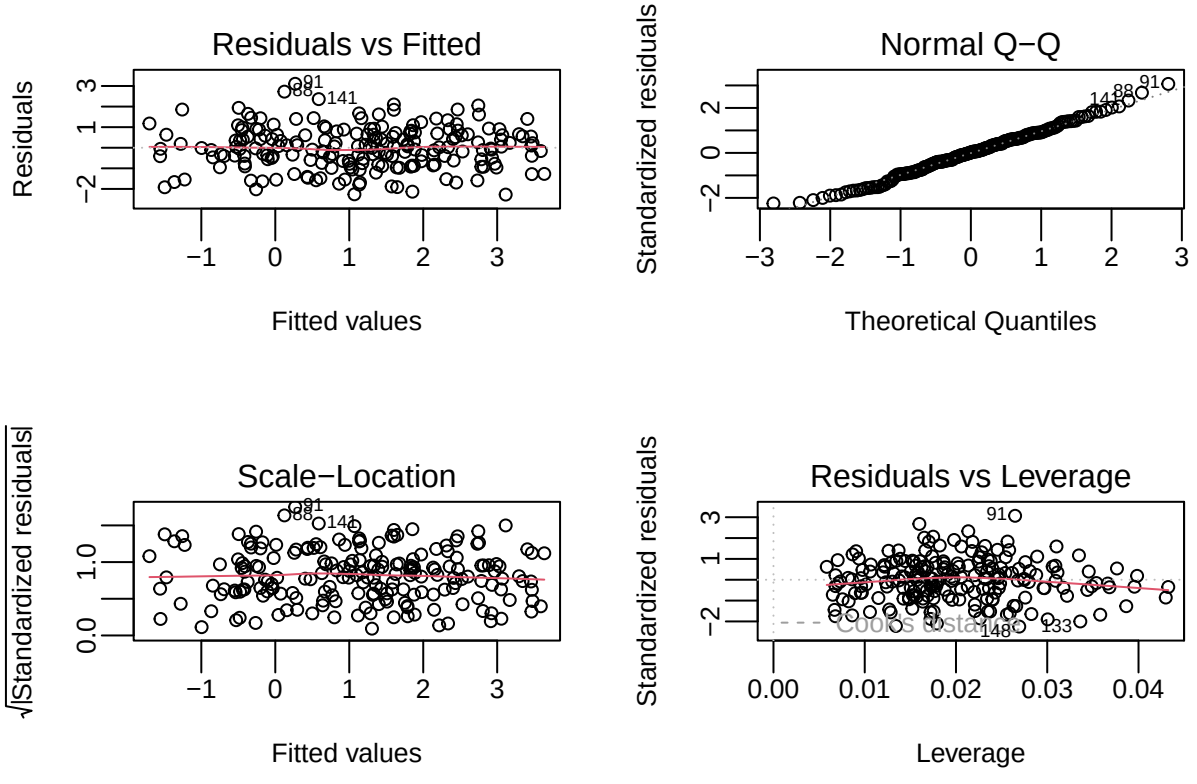
II Diagnostics

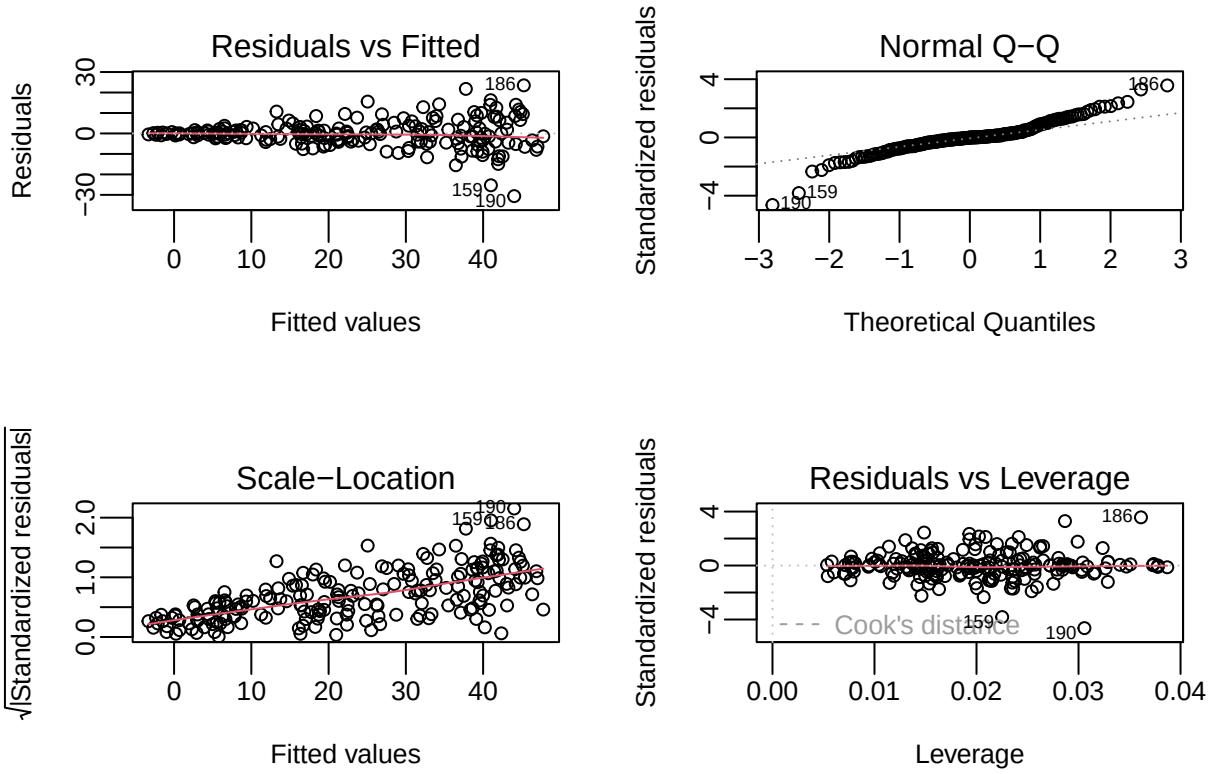
```
set.seed(0)

N <- 200 #number of samples
```

```
par(mfrow = c(2, 2), tcl=-1)

x1 <- runif(N)
x2 <- runif(N)
x3 <- runif(N)

y <- 0.1*x1 - 2*x2 + 4*x3 + rnorm(N) #normal noises

l <- lm(y ~ x1 + x2 + x3)

plot(l)
```

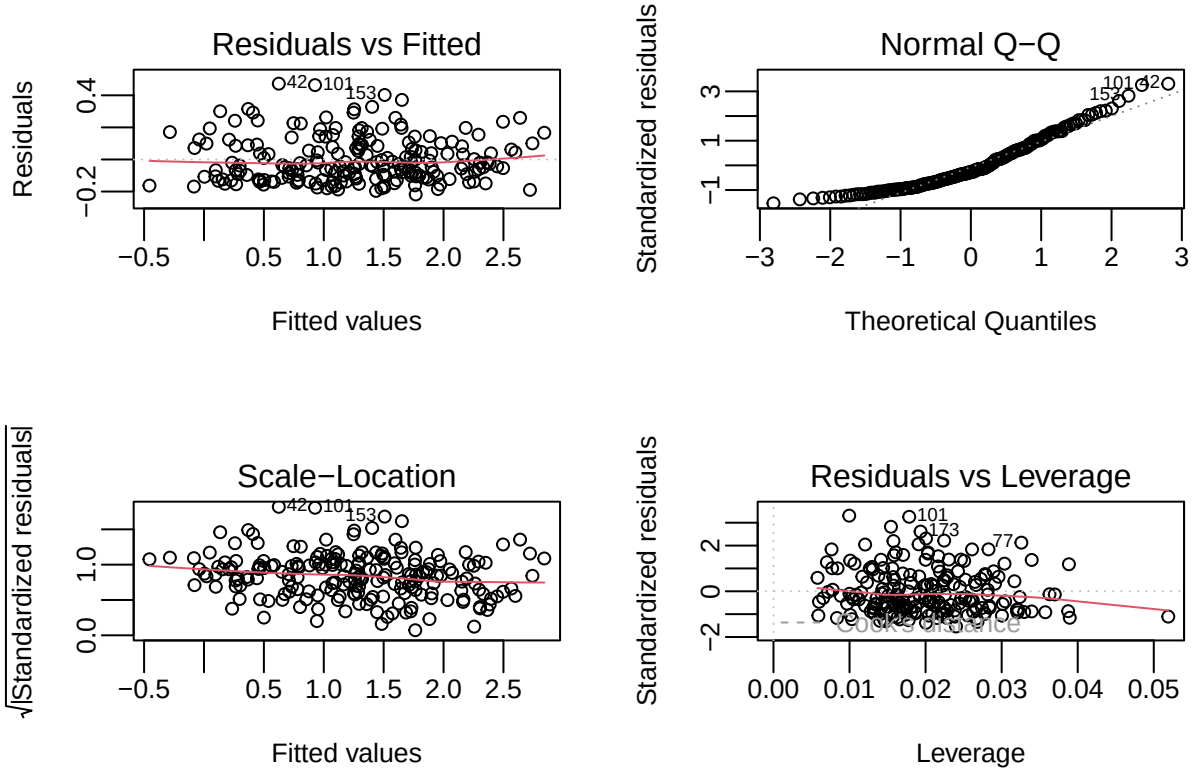



```
par(mfrow = c(2, 2), tcl=-1)

x1 <- runif(N)
```

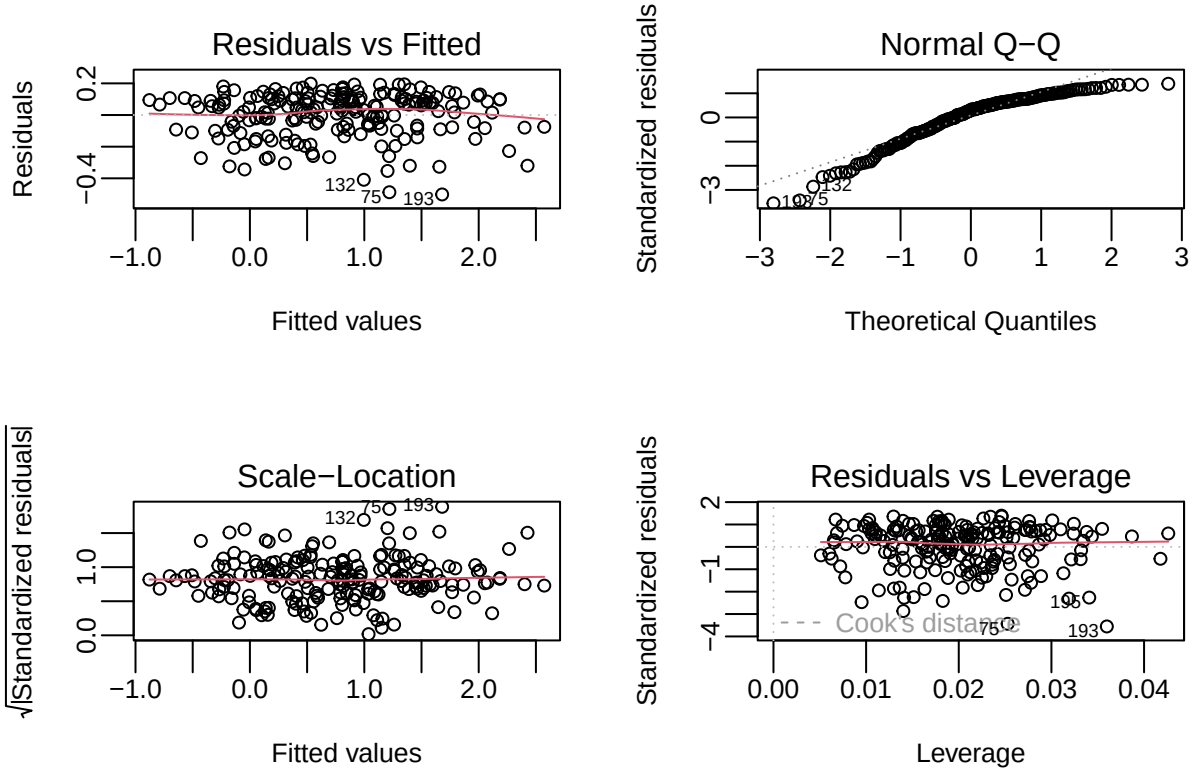
```
x2 <- runif(N)
x3 <- sort(runif(N))

y <- x1 - 5*x2 + 50*x3 + 10 * x3 * rnorm(N) #normal noises with heteroscedasticity

l <- lm(y ~ x1 + x2 + x3)

plot(l)
```



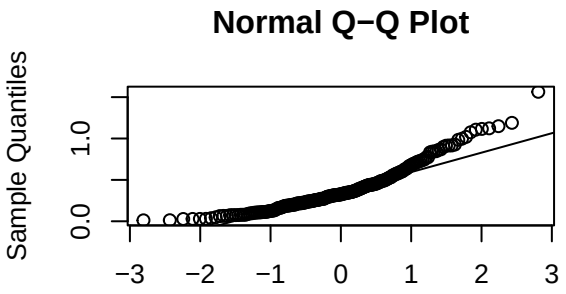

```
par(mfrow = c(2, 2), tcl=-1)

x1 <- runif(N)
x2 <- runif(N)
x3 <- sort(runif(N))

y <- x1 - x2 + 2*x3 + rgamma(N, 2, 10) #right skew

l <- lm(y ~ x1 + x2 + x3)

plot(l)
```



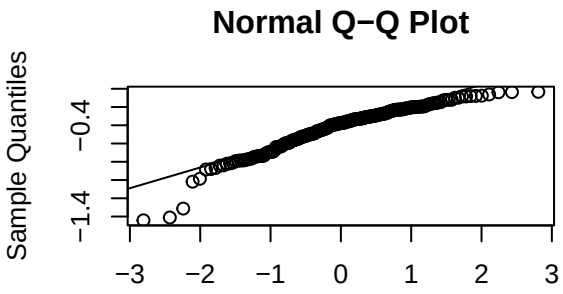

```
par(mfrow = c(2, 2), tcl=-1)

x1 <- runif(N)
x2 <- runif(N)
x3 <- sort(runif(N))

y <- x1 - x2 + 2*x3 - rgamma(N, 2, 10) #left skew

l <- lm(y ~ x1 + x2 + x3)

plot(l)
```



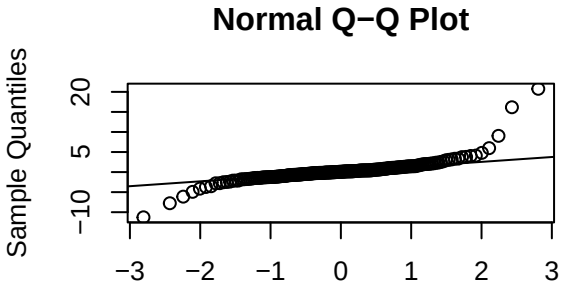

```
par(mfrow = c(2, 2), tcl=-0.5)

y <- rgamma(N, 2, 5)
qqnorm(y)
qqline(y)
title(sub = "right skew", font.sub = 4)

y <- -rgamma(N, 2, 5)
qqnorm(y)
qqline(y)
title(sub = "left skew", font.sub = 4)

y <- rt(N, 2)
qqnorm(y)
qqline(y)
title(sub = "platykurtic", font.sub = 4)

y <- runif(N, -10, 10)
qqnorm(y)
qqline(y)
title(sub = "leptokurtic", font.sub = 4)
```



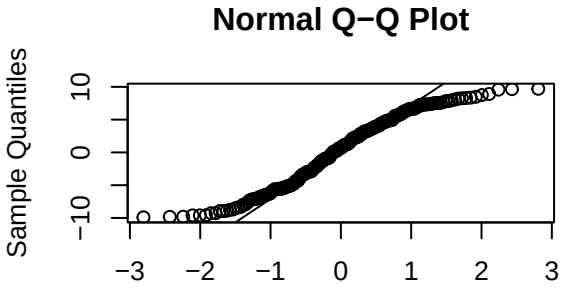
right skew



left skew



platykurtic



leptokurtic

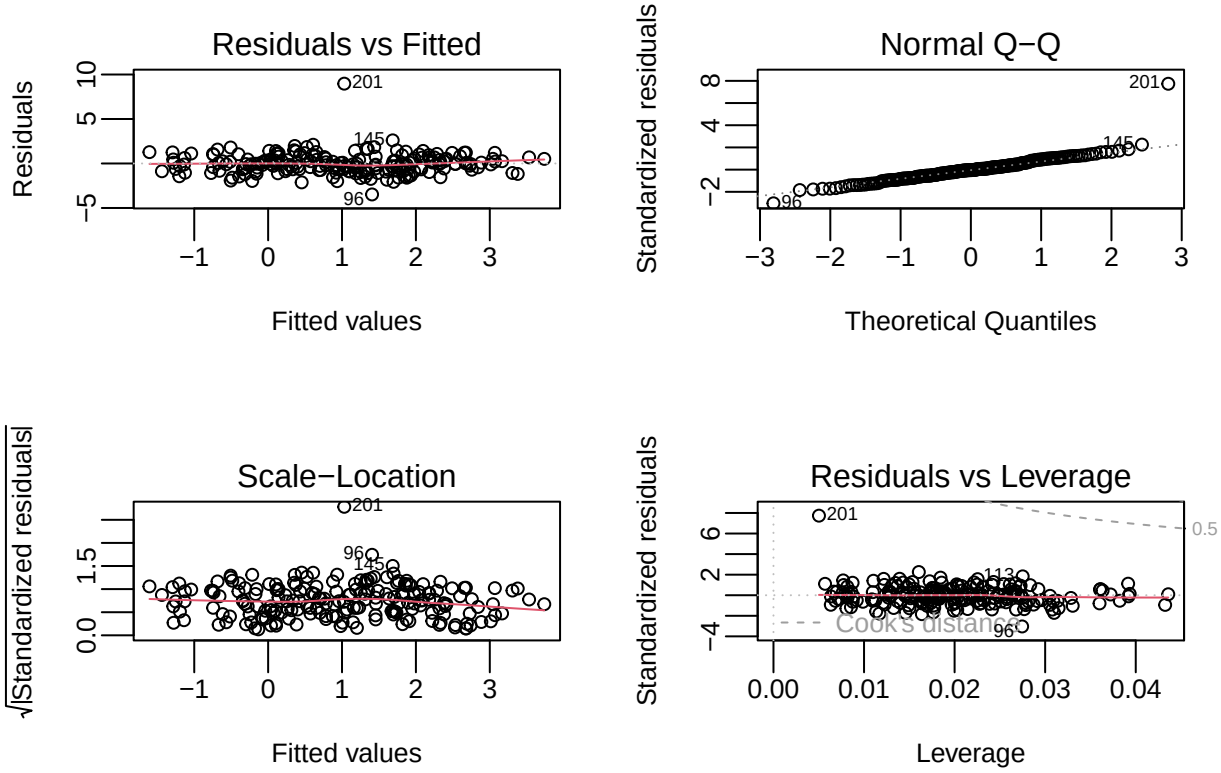
```
par(mfrow = c(2, 2), tcl=-1)

x1 <- runif(N)
x2 <- runif(N)
x3 <- sort(runif(N))

y <- 0.1*x1 - 2*x2 + 4*x3 + rnorm(N) #normal noises
print(l$coefficients)
```

```
## (Intercept)          x1          x2          x3
##  -0.2151462   0.9379543  -0.9468195   2.0155145
```

```
# outlier (not following general trend)
y <- c(y, 10)
x1 <- c(x1, 0.5)
x2 <- c(x2, 0.5)
x3 <- c(x3, 0.5)

l <- lm(y ~ x1 + x2 + x3)
print(l$coefficients)
```

```
## (Intercept)          x1          x2          x3
## -0.15988167  0.06459868 -1.56506896  3.88070528
```

```
plot(l)
```



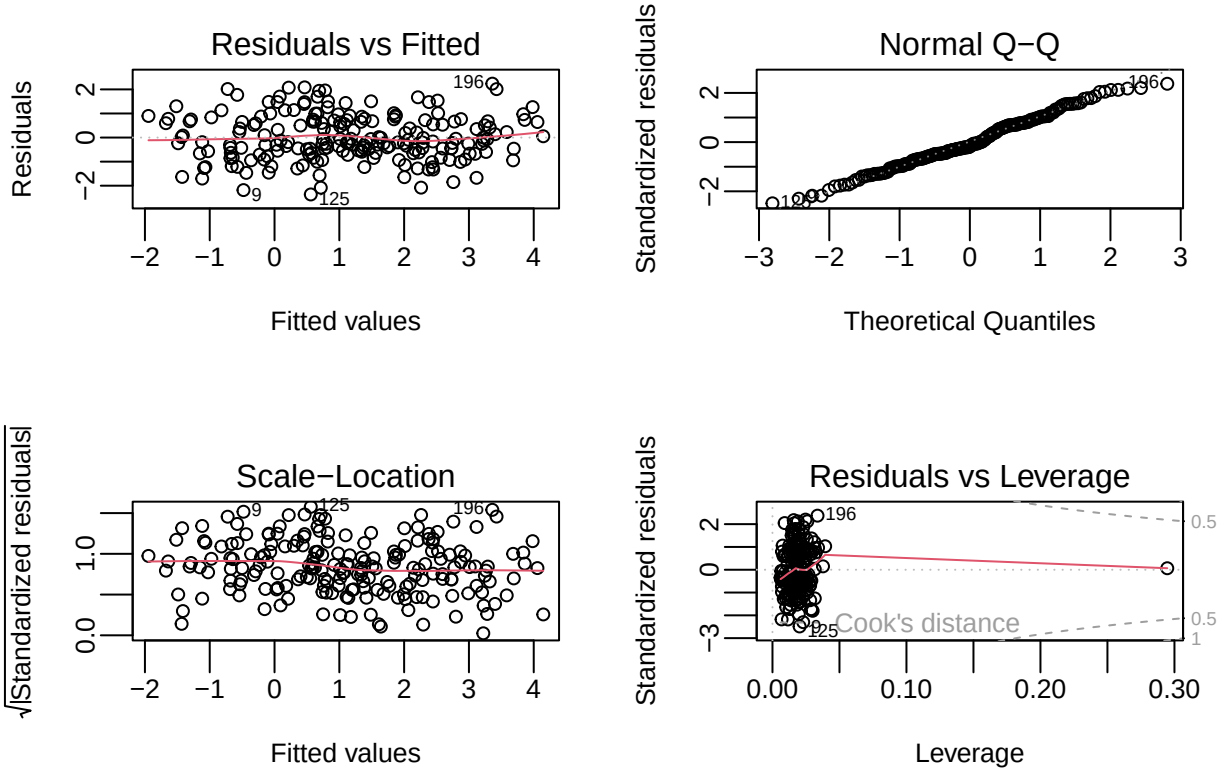

```
par(mfrow = c(2, 2), tcl=-1)

x1 <- runif(N)
x2 <- runif(N)
x3 <- sort(runif(N))

y <- 0.1*x1 - 2*x2 + 4*x3 + rnorm(N) #normal noises
print(l$coefficients)
```

```
## (Intercept)          x1          x2          x3
## -0.15988167  0.06459868 -1.56506896  3.88070528
```

```
# high leverage point (far from explanatory variable distribution cluster)
x1 <- c(x1, 2)
x2 <- c(x2, 2)
x3 <- c(x3, 2)
y <- c(y, 0.1*2 - 2*2 + 4*2)

l <- lm(y ~ x1 + x2 + x3)
print(l$coefficients)
```

```
## (Intercept)          x1          x2          x3
##   0.1143523   0.4422960  -2.3493162   3.9233728
```

```
plot(l)
```



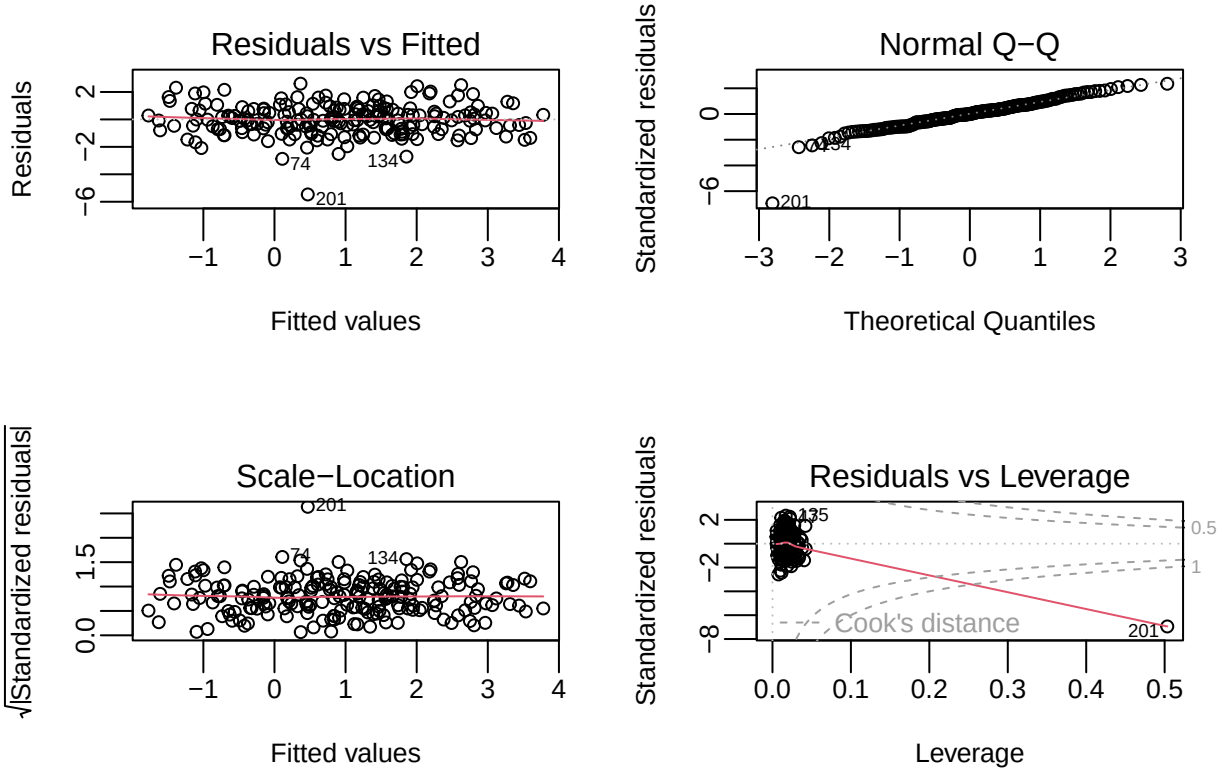

```
par(mfrow = c(2, 2), tcl=-1)

x1 <- runif(N)
x2 <- runif(N)
x3 <- sort(runif(N))

y <- 0.1*x1 - 2*x2 + 4*x3 + rnorm(N) #normal noises

l <- lm(y ~ x1 + x2 + x3)
print(l$coefficients)
```

```
##  (Intercept)           x1           x2           x3
##  0.007011215  0.236232122 -2.133004429  3.898039679
```

```
# high leverage point and outlier
x1 <- c(x1, 3)
x2 <- c(x2, 3)
x3 <- c(x3, 3)
y <- c(y, -5)

l <- lm(y ~ x1 + x2 + x3)
print(l$coefficients)
```

```
## (Intercept)          x1          x2          x3
##   1.1073736  -0.5572203  -2.8625408   3.2068098
```

```
plot(l)
```


Residuals vs Fitted
Residuals
Fitted values
74
134
201
Normal Q−Q
Standardized residuals
Theoretical Quantiles
201
Scale−Location
√|Standardized residuals|
Fitted values
201
74
134
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
0.5
1
201