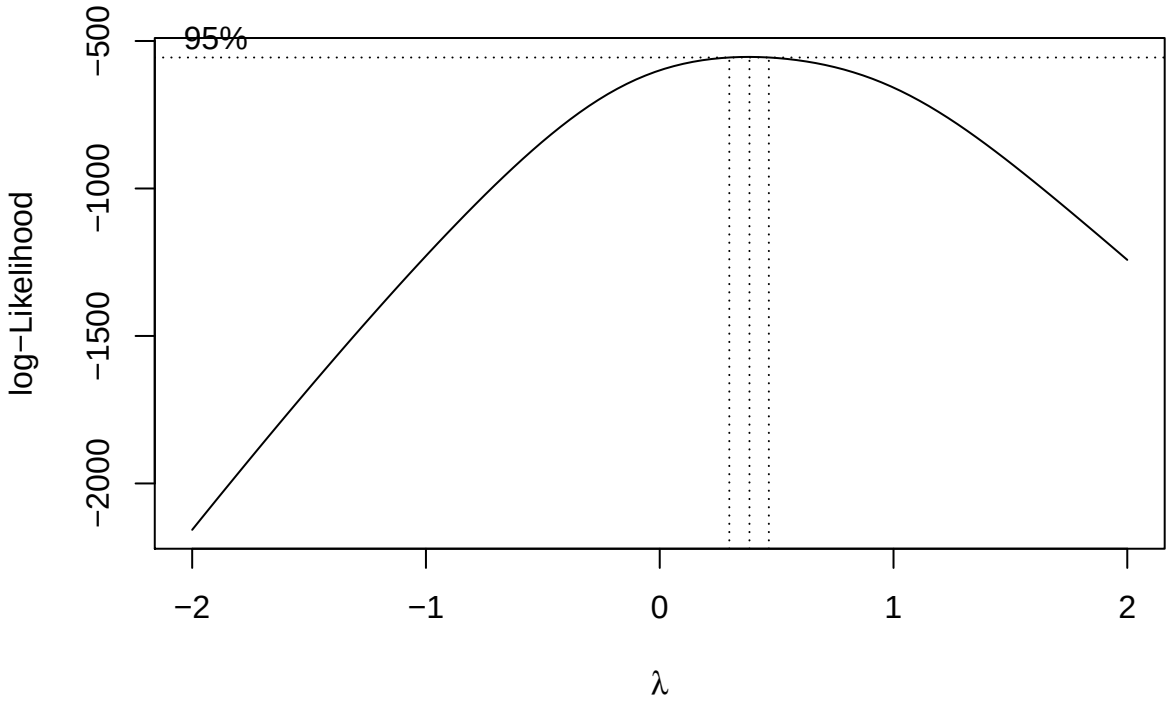
```r
## Effect of individual observations

## Python data revisited
python <- read.csv("csv/FLpython.csv")
python$male <- ifelse(python$sex == 'M', 1, 0) # 1 = M, 0 =F
mpf2 <- lm(fat ~ male + mass + svl, data = python)

# Last time we used a Box-Cox transformation
library(MASS)
bc <- boxcox(mpf2)
```

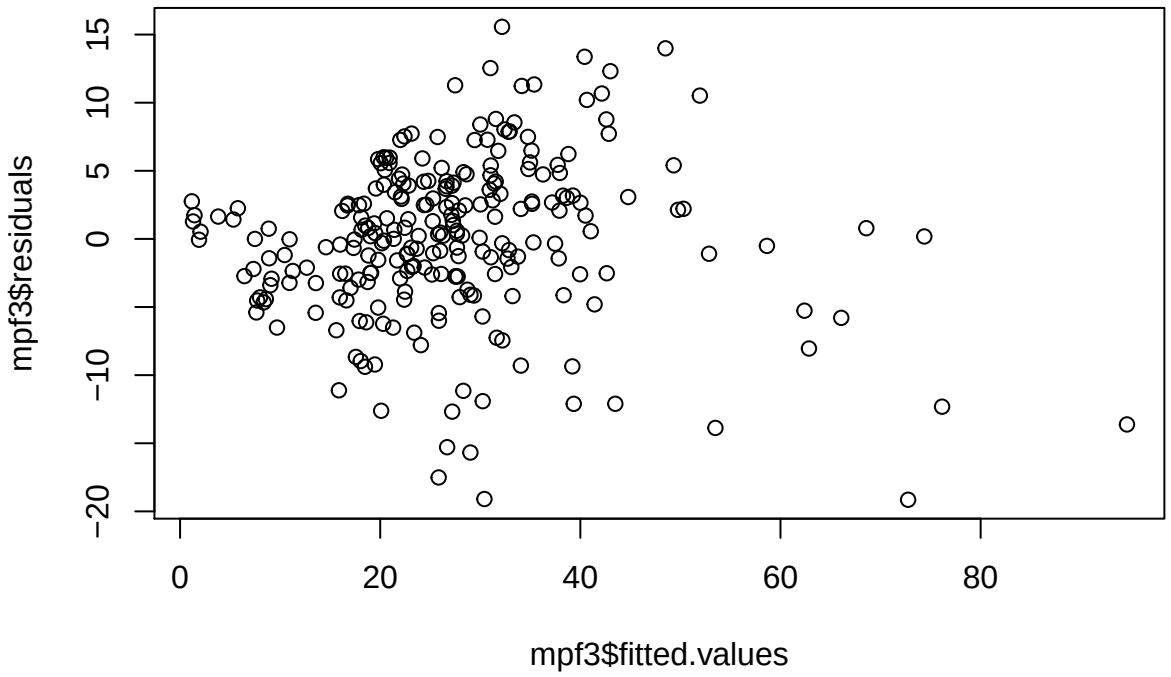



```r
lambda <- bc$x[which.max(bc$y)]
mpf3 <- lm((fat ^ lambda - 1) / lambda ~ male + mass + svl, data = python)
summary(mpf3)
```

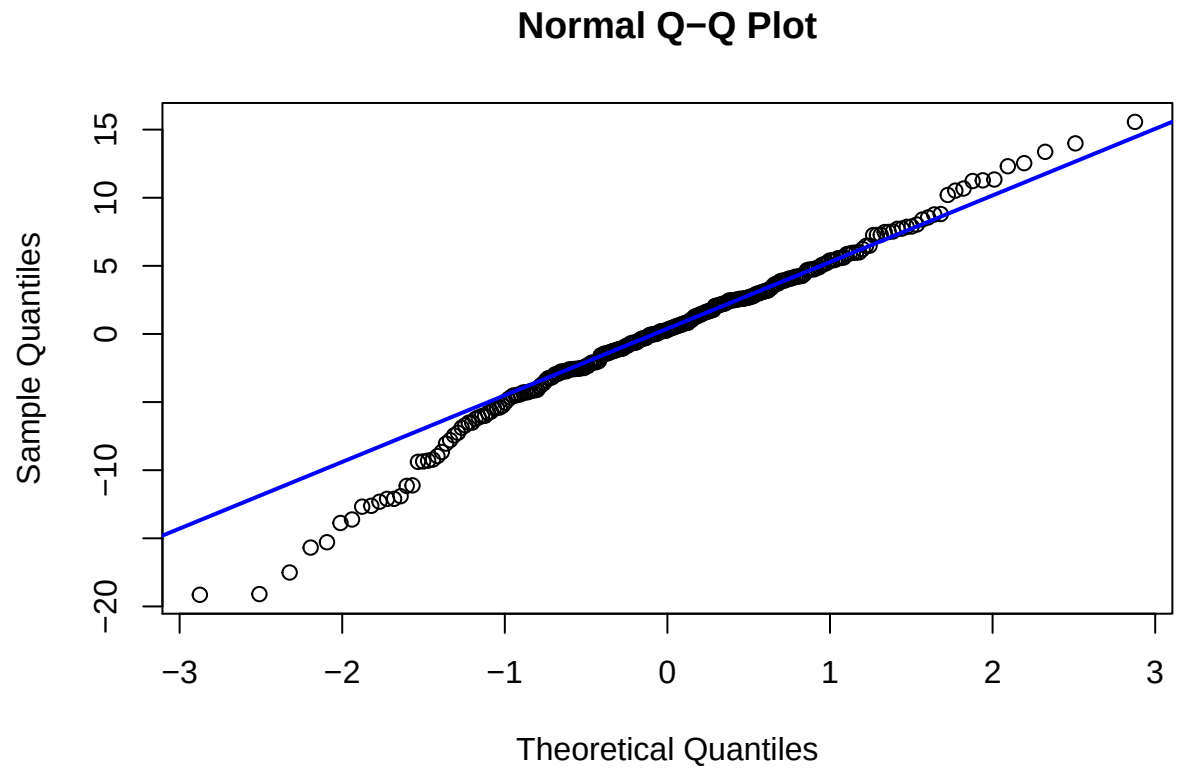
```
##
## Call:
## lm(formula = (fat^lambda - 1)/lambda ~ male + mass + svl, data = python)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.146  -2.910   0.297   3.688  15.568
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.0558134  2.1813183  -3.693 0.000273 ***
## male        -1.7849310  0.7776166  -2.295 0.022560 *
## mass         0.0004461  0.0000864   5.164 5.03e-07 ***
## svl          0.1431492  0.0126019  11.359  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 5.939 on 244 degrees of freedom
## Multiple R-squared:  0.8356, Adjusted R-squared:  0.8336
## F-statistic: 413.5 on 3 and 244 DF,  p-value: < 2.2e-16
```

```r
plot(mpf3$fitted.values, mpf3$residuals)
```

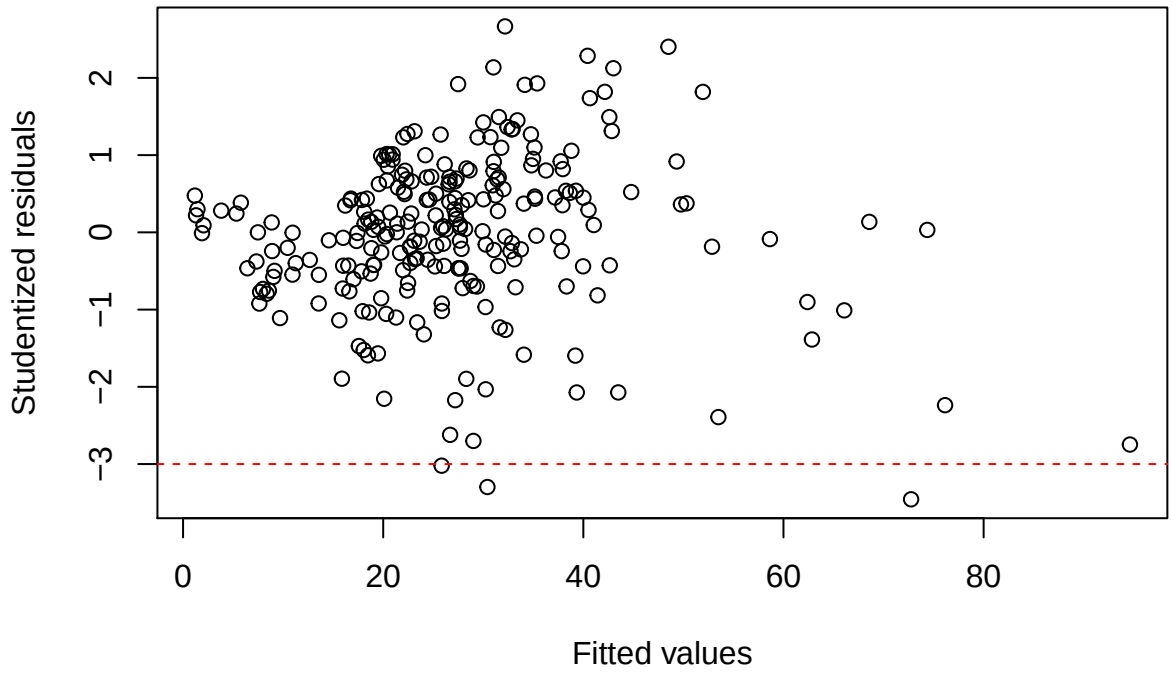



```r
qqnorm(mpf3$residuals)
qqline(mpf3$residuals, col = "blue", lwd = 2)
```



```r
# Quantities for individual observations
studres(mpf3)  # studentized residuals
hatvalues(mpf3) # leverage
cooks.distance(mpf3) # Cook's distance
```

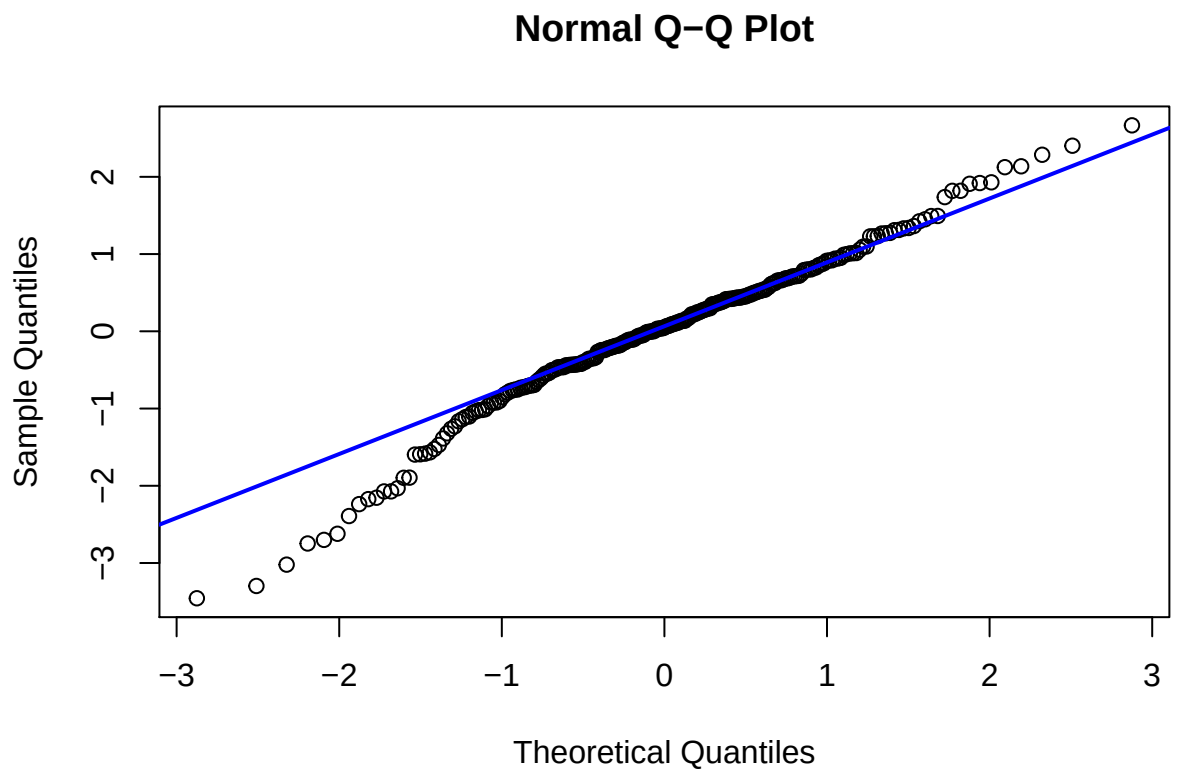
```r
# Residual plots with studentized residuals
plot(mpf3$fitted.values,
     studres(mpf3),
     xlab = "Fitted values",
     ylab = "Studentized residuals")
abline(h = c(3, -3), col = "red", lty = 2)
```



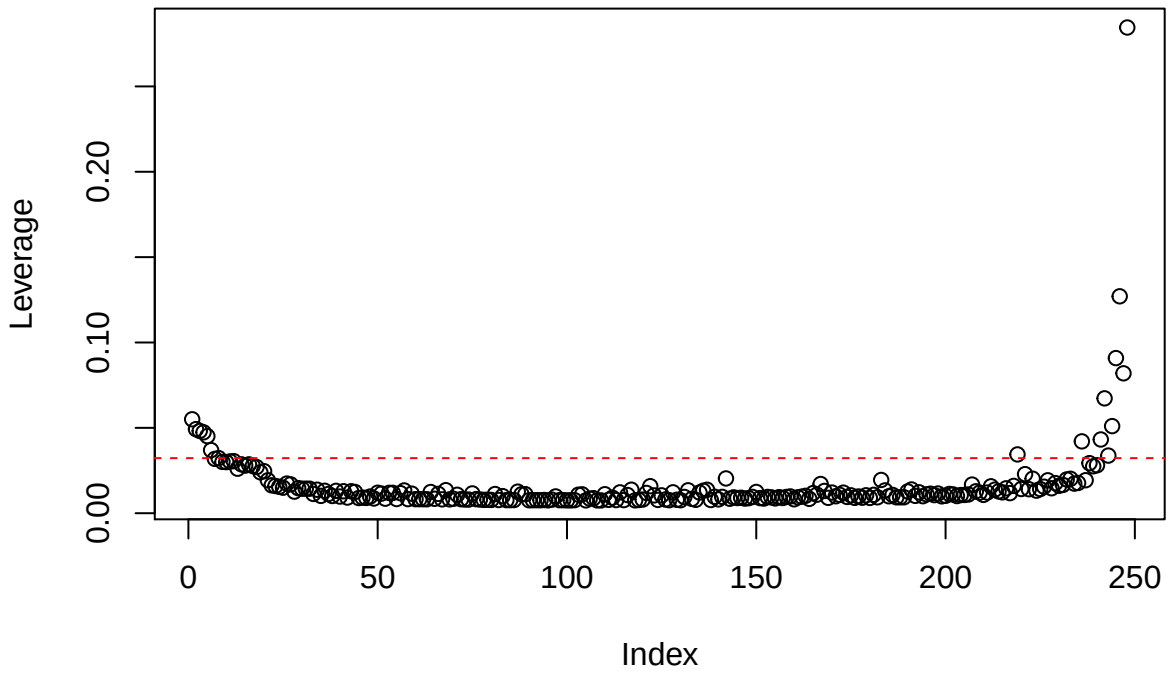

```r
which(abs(studres(mpf3)) > 3)
```

```
## 122 181 245
## 122 181 245
```

```r
qqnorm(studres(mpf3))
qqline(studres(mpf3), col = "blue", lwd = 2)
```




```
# Leverage
plot(hatvalues(mpf3), ylab = "Leverage")
abline(h = 2 * mean(hatvalues(mpf3)),
       col = "red",
       lty = 2)
```




```
which(hatvalues(mpf3) > 2 * mean(hatvalues(mpf3)))
```

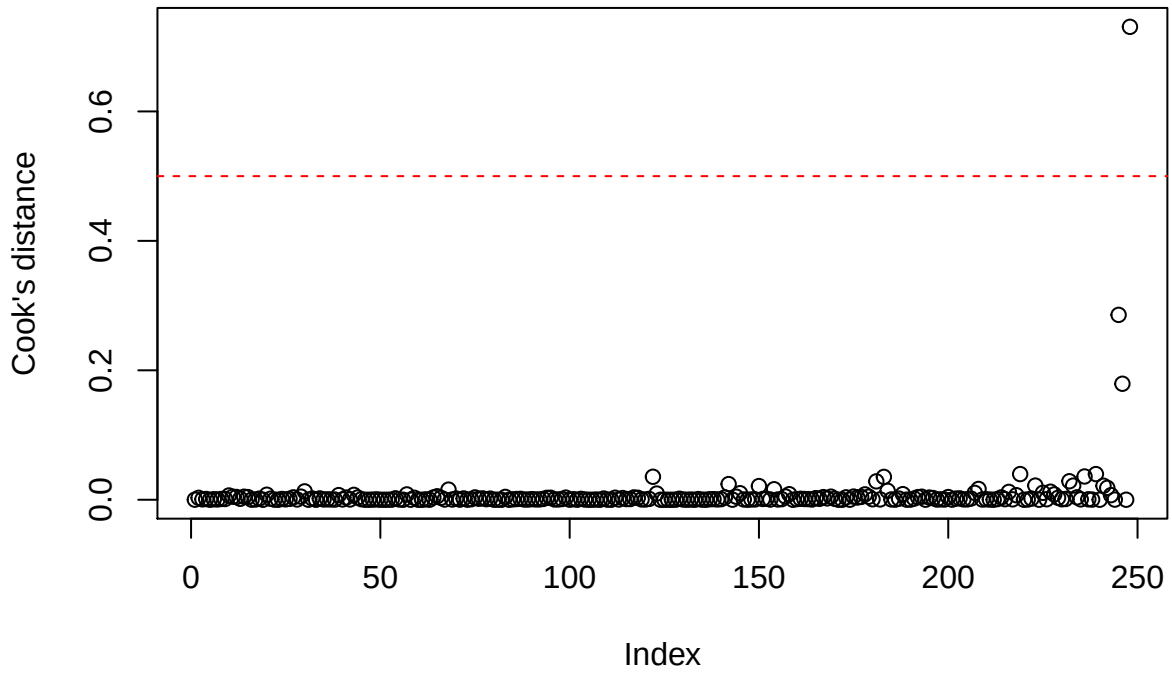
```
##   1   2   3   4   5   6   8 219 236 241 242 243 244 245 246 247 248
```

```
##   1   2   3   4   5   6   8 219 236 241 242 243 244 245 246 247 248
```

```
python[which(hatvalues(mpf3) > 2 * mean(hatvalues(mpf3))), ]
```

```
##     sex   svl  mass length      fat male
## 1     F  70.0   186   77.5    6.000    0
## 2     M  76.0   310   83.8   11.000    1
## 3     M  77.0   260   86.1    6.000    1
## 4     M  78.0   262   87.1    8.000    1
## 5     M  81.0   306   91.1    4.000    1
## 6     M  93.5   605  104.6   18.959    1
## 8     F 105.0   800  117.5   17.000    0
## 219   M 285.0 27000  316.2 3230.000    1
## 236   M 330.0 32600  370.9 4374.000    1
## 241   F 376.0 38280  424.2 3156.000    0
## 242   F 381.0 43910  424.9 4002.000    0
## 243   F 384.5 34540  432.4 3500.000    0
## 244   F 405.5 41660  455.3 5688.000    0
## 245   F 409.0 49900  460.2 2988.000    0
## 246   F 416.0 55260  469.1 4618.000    0
## 247   F 422.0 49350  473.4 6818.000    0
## 248   F 482.0 75500  545.0 8406.000    0
```

```
# Cook's distance
plot(cooks.distance(mpf3), ylab = "Cook's distance")
abline(h = 0.5, col = "red", lty = 2)
```




```
which(cooks.distance(mpf3) > 0.5)
```

```
## 248
## 248
```

```
# Let's look at actual changes in beta estimates
mpf3$coefficients # with all the data
```

```
##   (Intercept)          male          mass           svl
## -8.0558134354 -1.7849310101  0.0004461197  0.1431491887
```

```
# e.g., fit without obs 248
mpf4 <-
  lm((fat ^ lambda - 1) / lambda ~ male + mass + svl, data = python[-248, ])
mpf4$coefficients
```

```
##   (Intercept)          male          mass           svl
## -6.6475616056 -1.6605858218  0.0005743312  0.1313793189
```

```
# e.g., fit without obs 50
mpf5 <-
  lm((fat ^ lambda - 1) / lambda ~ male + mass + svl, data = python[-50, ])
mpf5$coefficients
```

```
##   (Intercept)          male          mass           svl
## -8.0628754675 -1.7805093651  0.0004462354  0.1431573753
```