Notizen - Sitzung 13

Jonas Schulte

05.07.2023

Verwendete Packages

```
library(here)
library(data.table)
library(tidyverse)
library(ggplot2)
library(ggfortify)
library(car)
library(lmtest)
library(sandwich)
```

Datensatz einlesen

```
data_path <- here("data/bip-mortrate.csv")
lifeexp_data <- fread(data_path,
              colClasses = c("character", "integer", rep("double", 5)))
head(lifeexp_data, 3)
```

```
##    country year LIFEEXP GDP_PPPpc MORTRATE GINI_post GINI_pre
## 1:     AFG 2010  61.028  1671.581     88.0      30.2     30.9
## 2:     ALB 2010  76.562  9927.630     13.3      38.4     44.0
## 3:     ARM 2010  73.331  6702.829     18.5      35.9     47.4
```

Regressionen: Lebenserwartung und BIP pro Kopf

Modelle

```
model1 <- lm(LIFEEXP ~ GDP_PPPpc, lifeexp_data)
summary(model1)
```

```
##
## Call:
## lm(formula = LIFEEXP ~ GDP_PPPpc, data = lifeexp_data)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -23.865  -3.826   2.087   4.875   9.498
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.521e+01  7.351e-01   88.70   <2e-16 ***
## GDP_PPPpc   3.148e-04  2.996e-05   10.51   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.663 on 148 degrees of freedom
## Multiple R-squared:  0.4272, Adjusted R-squared:  0.4233
## F-statistic: 110.4 on 1 and 148 DF,  p-value: < 2.2e-16
```

```
model2 <- lm(LIFEEXP ~ log(GDP_PPPpc), lifeexp_data)
summary(model2)
```

```
##
## Call:
## lm(formula = LIFEEXP ~ log(GDP_PPPpc), data = lifeexp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.3420  -1.7187   0.6502   3.0746   9.7919
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     15.1124     3.0494   4.956 1.95e-06 ***
## log(GDP_PPPpc)   6.0748     0.3322  18.287  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.876 on 148 degrees of freedom
## Multiple R-squared:  0.6932, Adjusted R-squared:  0.6911
## F-statistic: 334.4 on 1 and 148 DF,  p-value: < 2.2e-16
```

```
model3 <- lm(log(LIFEEXP) ~ log(GDP_PPPpc), lifeexp_data)
summary(model3)
```

```
##
## Call:
## lm(formula = log(LIFEEXP) ~ log(GDP_PPPpc), data = lifeexp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31957 -0.02348  0.01000  0.04694  0.15263
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.430276   0.048212   71.15   <2e-16 ***
## log(GDP_PPPpc) 0.089613   0.005252   17.06   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07709 on 148 degrees of freedom
## Multiple R-squared:  0.663,  Adjusted R-squared:  0.6607
## F-statistic: 291.1 on 1 and 148 DF,  p-value: < 2.2e-16
```

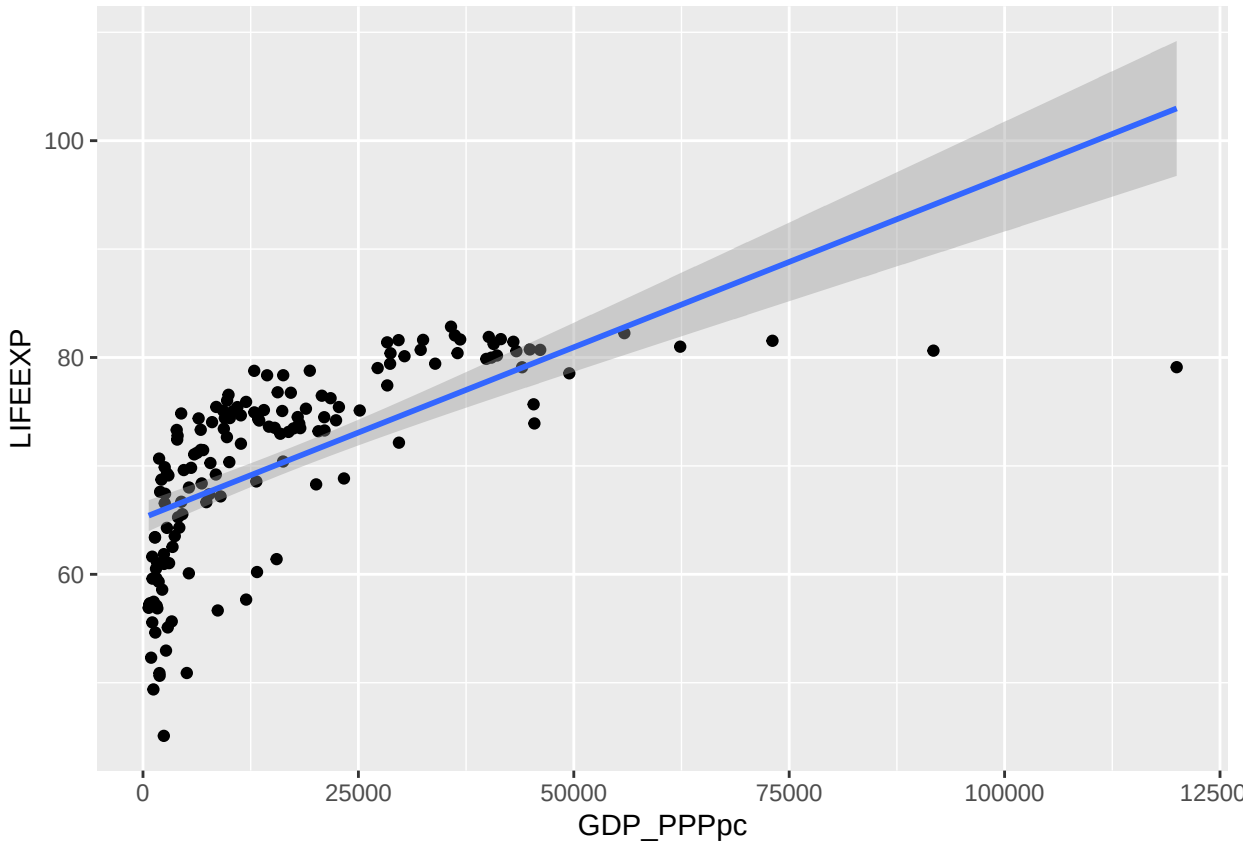
```
model4 <- lm(LIFEEXP ~ log(GDP_PPPpc) + I(log(GDP_PPPpc)**2), lifeexp_data)
summary(model4)
```

```
##
## Call:
## lm(formula = LIFEEXP ~ log(GDP_PPPpc) + I(log(GDP_PPPpc)^2),
##     data = lifeexp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.3626  -1.6936   0.8357   2.9741   9.9662
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -7.8443    21.5501  -0.364   0.7164
## log(GDP_PPPpc)       11.2891     4.8570   2.324   0.0215 *
## I(log(GDP_PPPpc)^2)  -0.2907     0.2702  -1.076   0.2837
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.874 on 147 degrees of freedom
## Multiple R-squared:  0.6956, Adjusted R-squared:  0.6915
## F-statistic:   168 on 2 and 147 DF,  p-value: < 2.2e-16
```

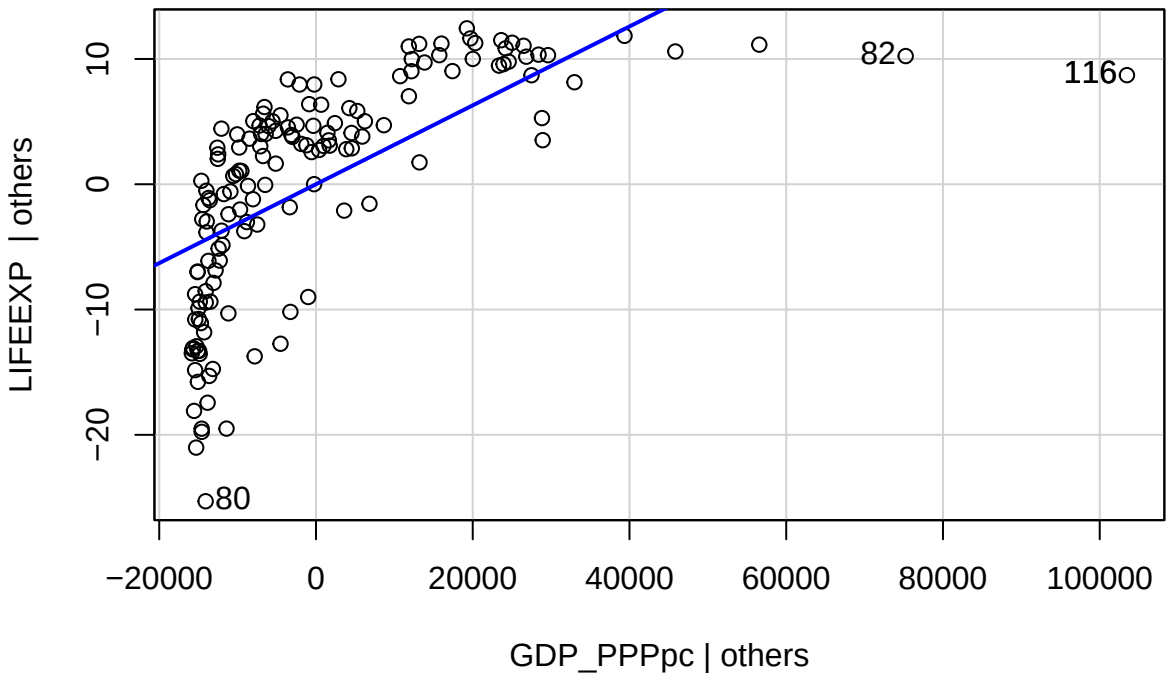
Regressionsdiagnostik

Scatter-Plots

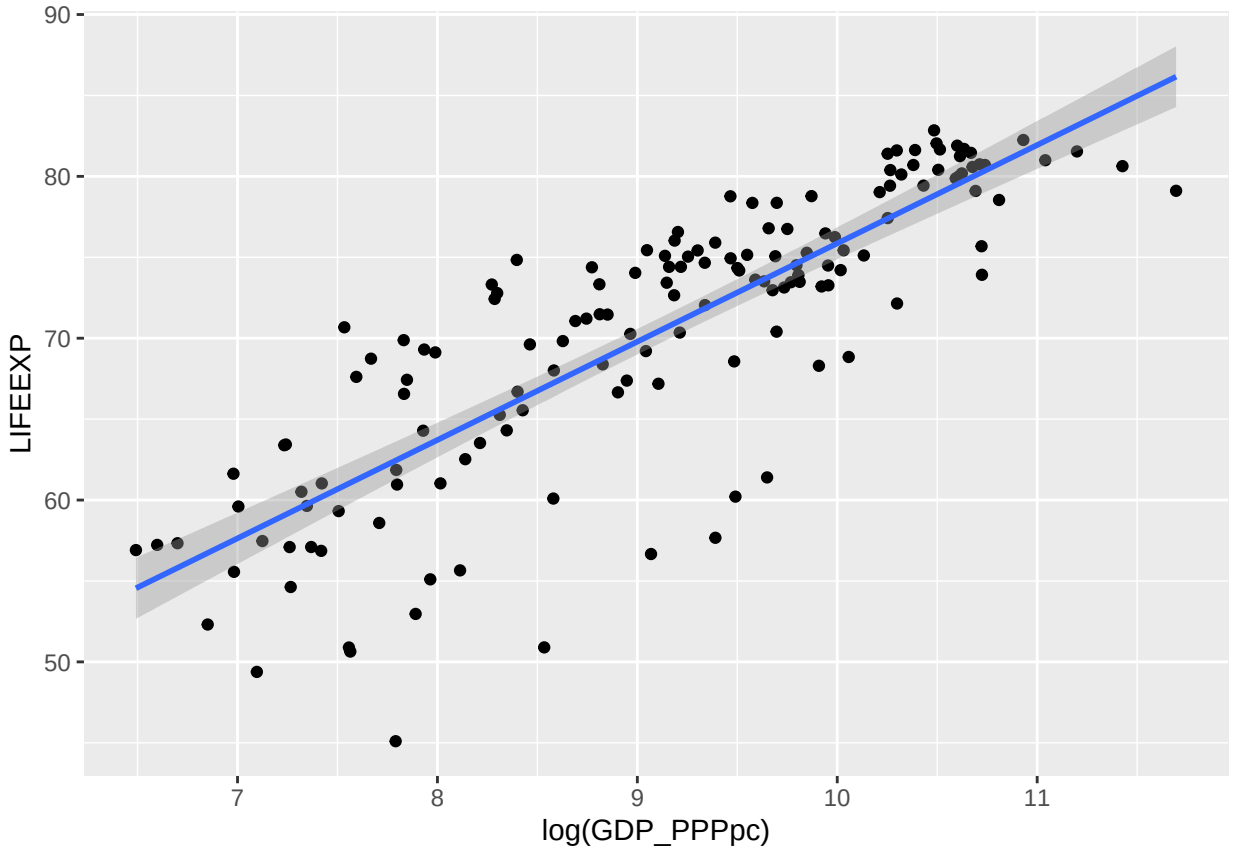
```
# Modell 1
ggplot(lifeexp_data, aes(x = GDP_PPPpc, y = LIFEEXP))+
  geom_point()+
  geom_smooth(method = "lm")
```



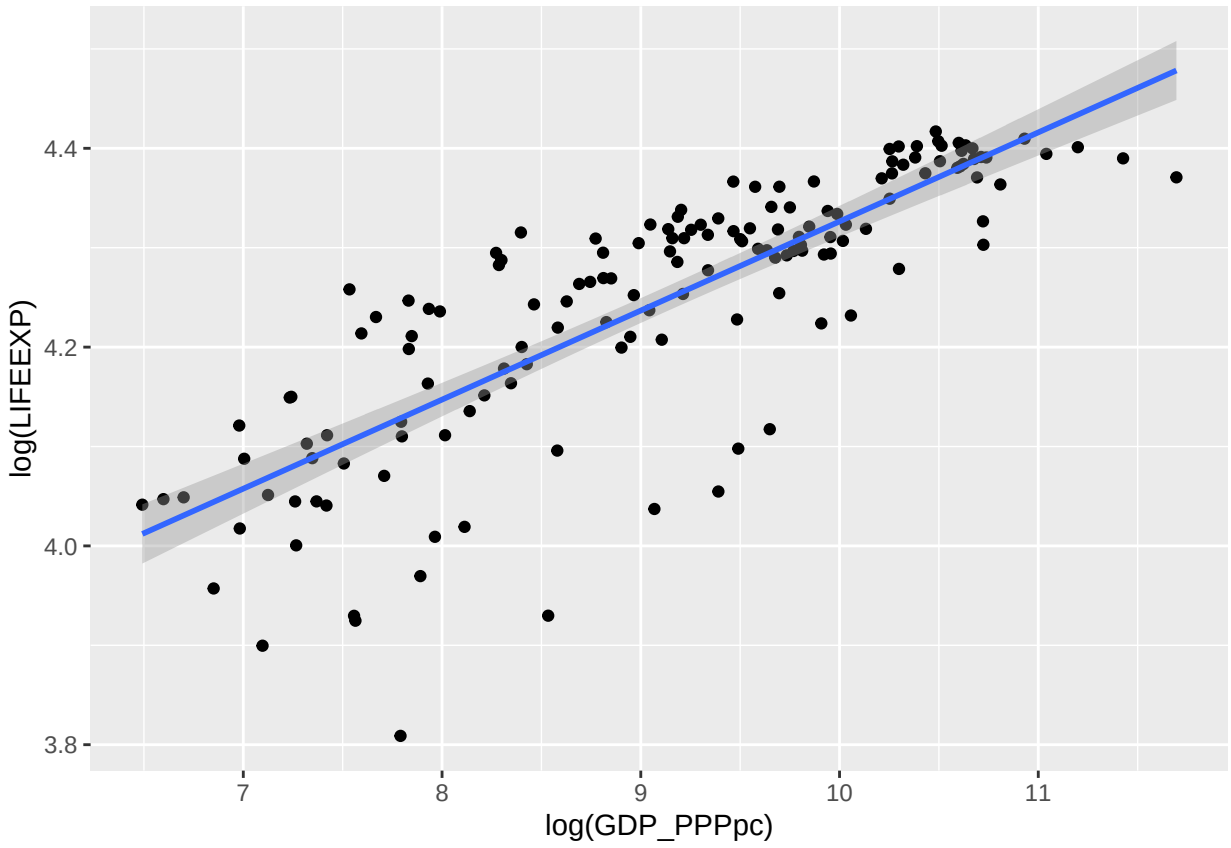

```
avPlots(model1)
```



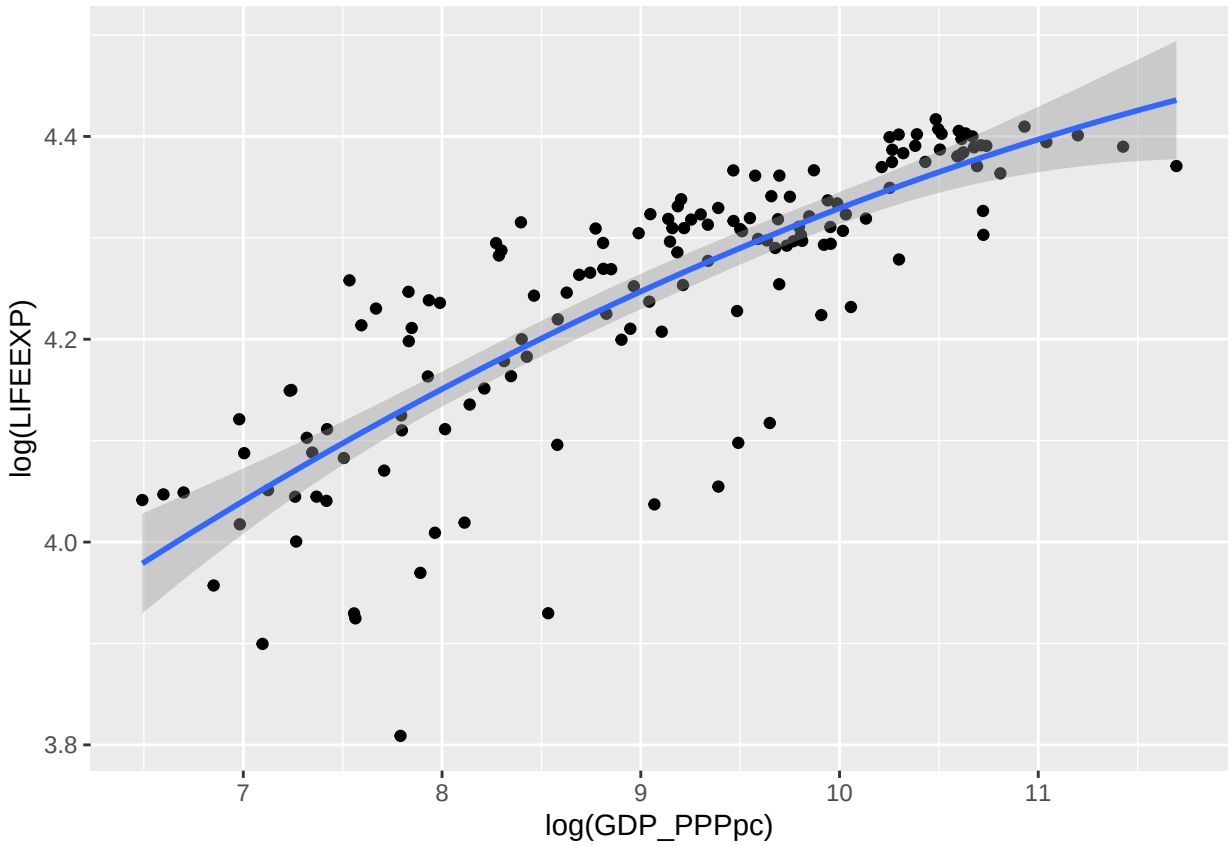

```
# Modell 2
ggplot(lifeexp_data, aes(x = log(GDP_PPPpc), y = LIFEEXP))+
  geom_point()+
  geom_smooth(method = "lm")
```




```
# Modell 3
ggplot(lifeexp_data, aes(x = log(GDP_PPPpc), y = log(LIFEEXP)))+
  geom_point()+
  geom_smooth(method = "lm")
```




```
# Modell 4
ggplot(lifeexp_data, aes(x = log(GDP_PPPpc), y = log(LIFEEXP)))+
  geom_point()+
  geom_smooth(method = "lm", formula = y ~ x + I(x**2))
```



RESET-Test

```
library(lmtest)
resettest(model1) # Mit dem Power Argument, kann man festlegen, für welche Potenzen getestet werden sol
```

```
##
##  RESET test
##
## data:  model1
## RESET = 47.873, df1 = 2, df2 = 146, p-value < 2.2e-16
```

```
resettest(model2)
```

```
##
##  RESET test
##
## data:  model2
## RESET = 1.3622, df1 = 2, df2 = 146, p-value = 0.2593
```

```
resettest(model3)
```

```
##
##  RESET test
```

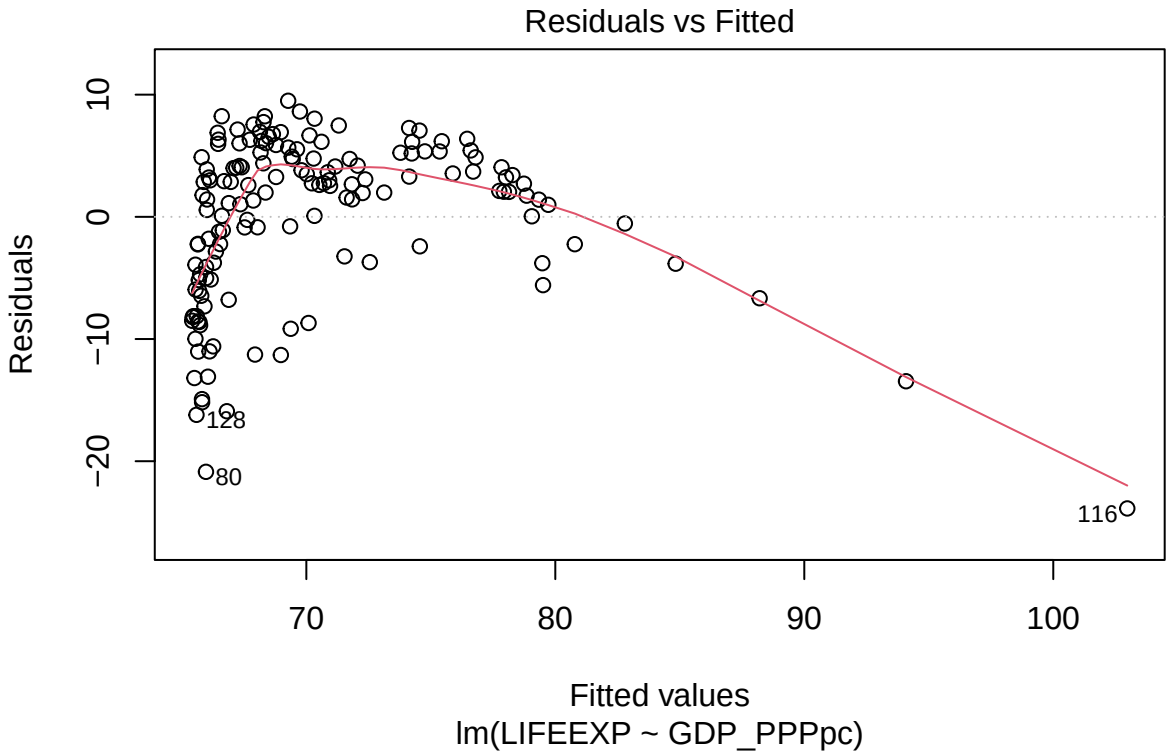
```
##
## data:  model3
## RESET = 2.055, df1 = 2, df2 = 146, p-value = 0.1318
```

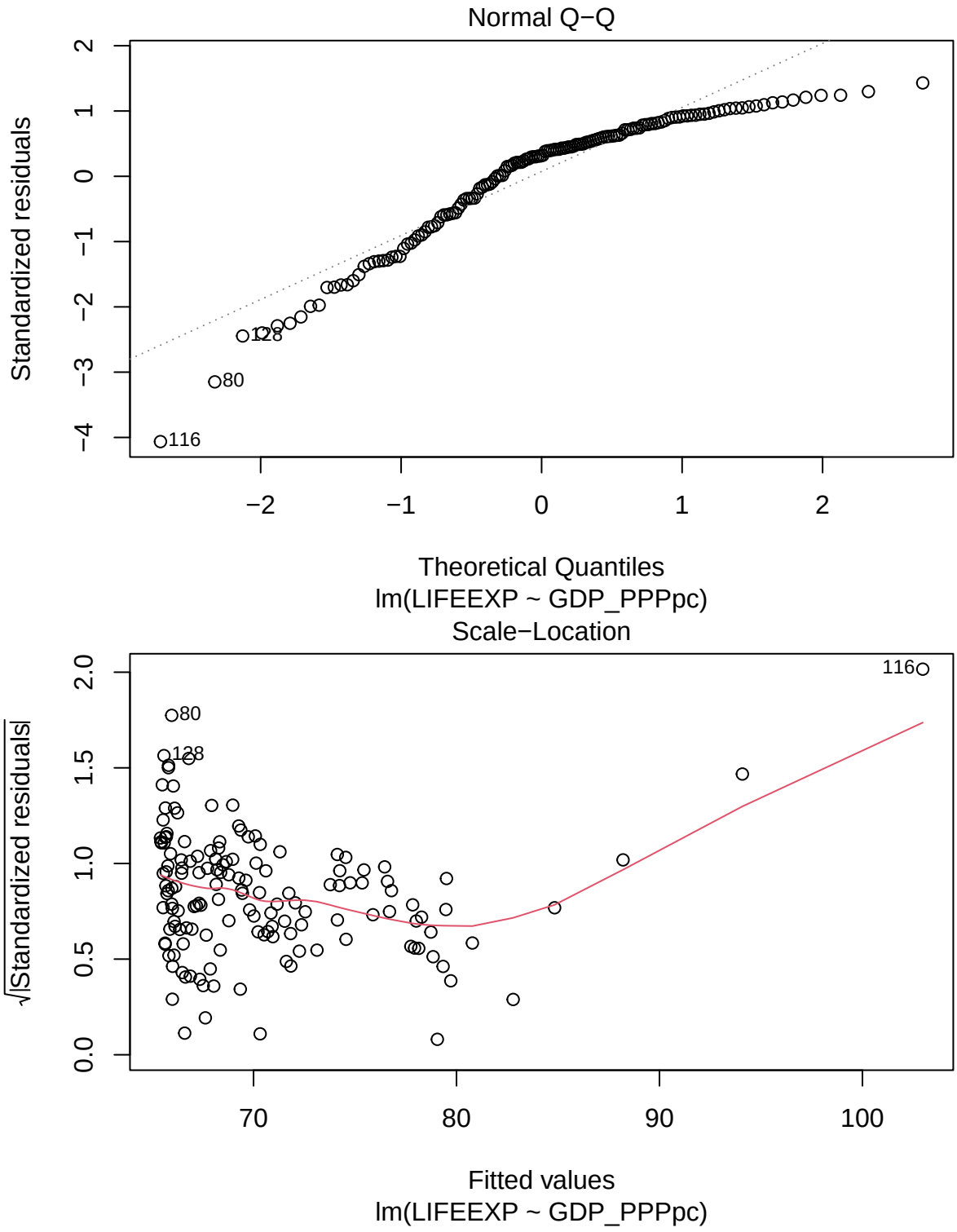
```
resettest(model4, power = 3)
```

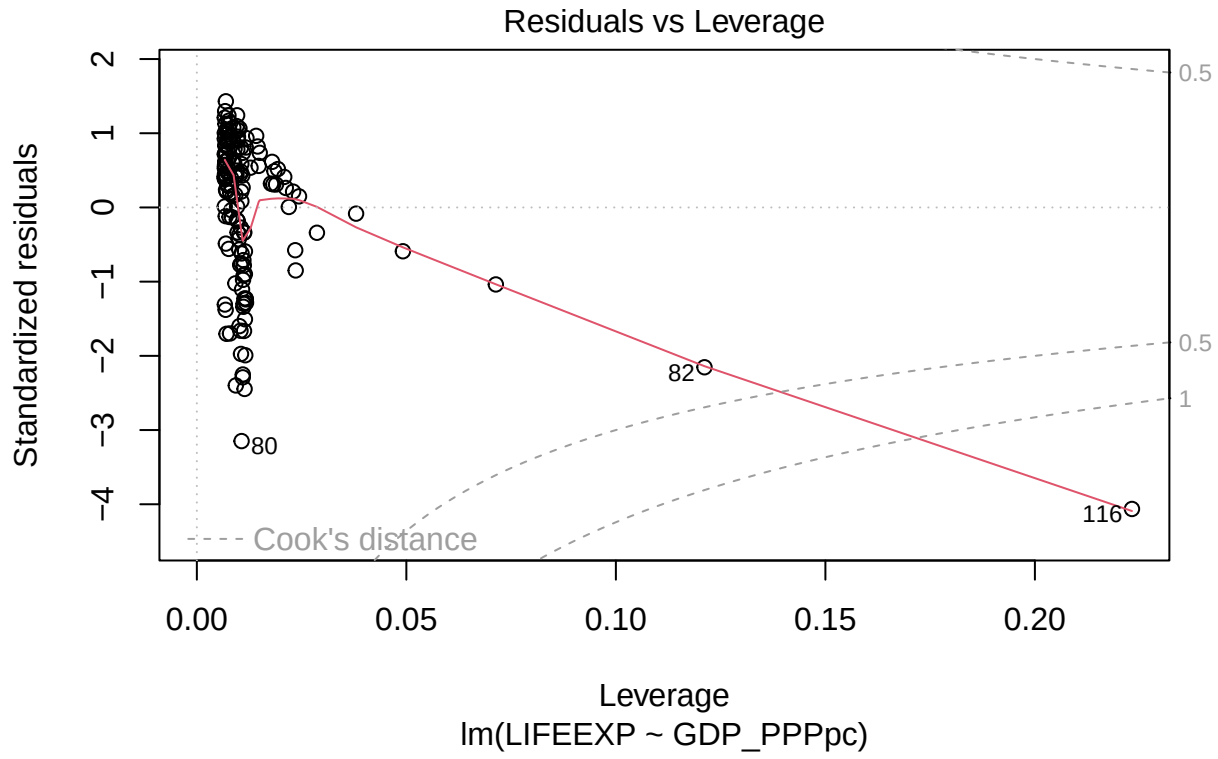
```
##
##  RESET test
##
## data:  model4
## RESET = 1.4865, df1 = 1, df2 = 146, p-value = 0.2247
```

TA-Plot, QQ_plot, etc.

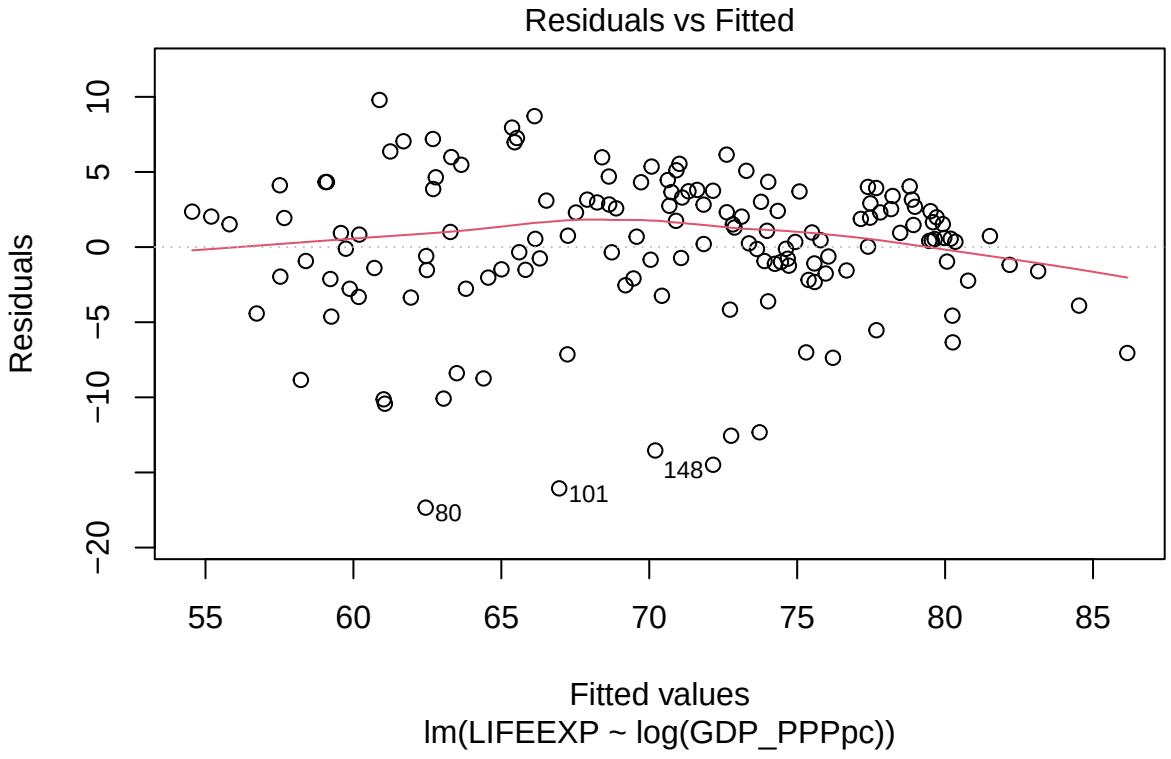
```
plot(model1)
```

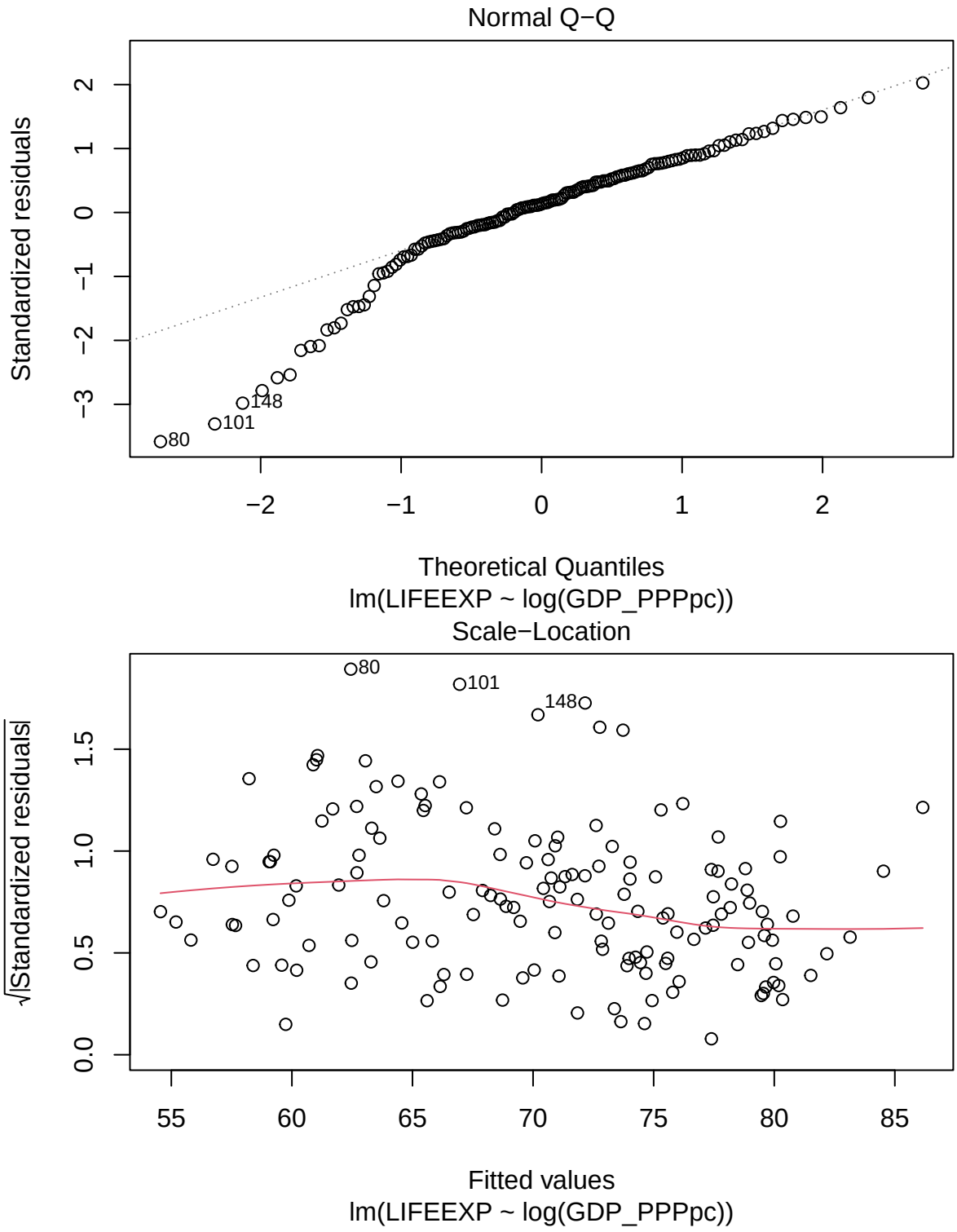


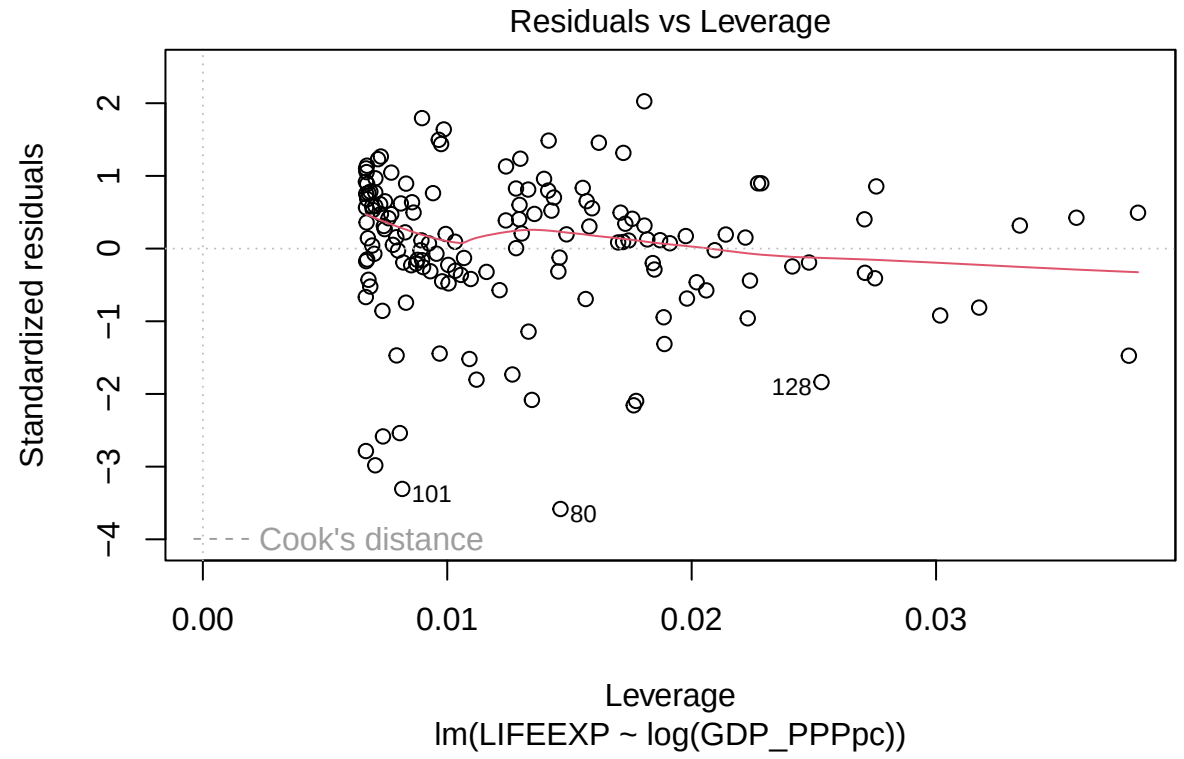

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(LIFEEXP ~ GDP_PPPpc)
Scale−Location
√|Standardized residuals|
Fitted values
lm(LIFEEXP ~ GDP_PPPpc)
116
80
128


Residuals vs Leverage
Standardized residuals
Leverage
lm(LIFEEXP ~ GDP_PPPpc)
Cook's distance
0.5
0.5
1
80
82
116
0.00
0.05
0.10
0.15
0.20
2
1
0
−1
−2
−3
−4

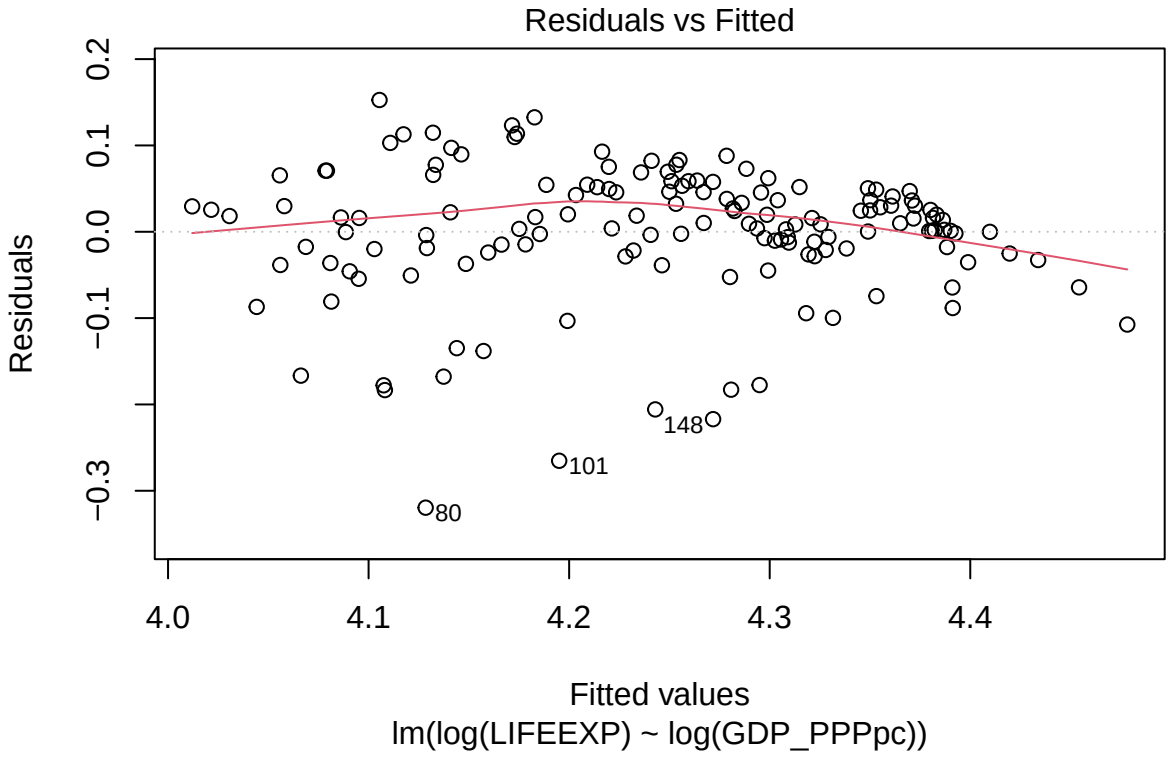
```
plot(model2)
```

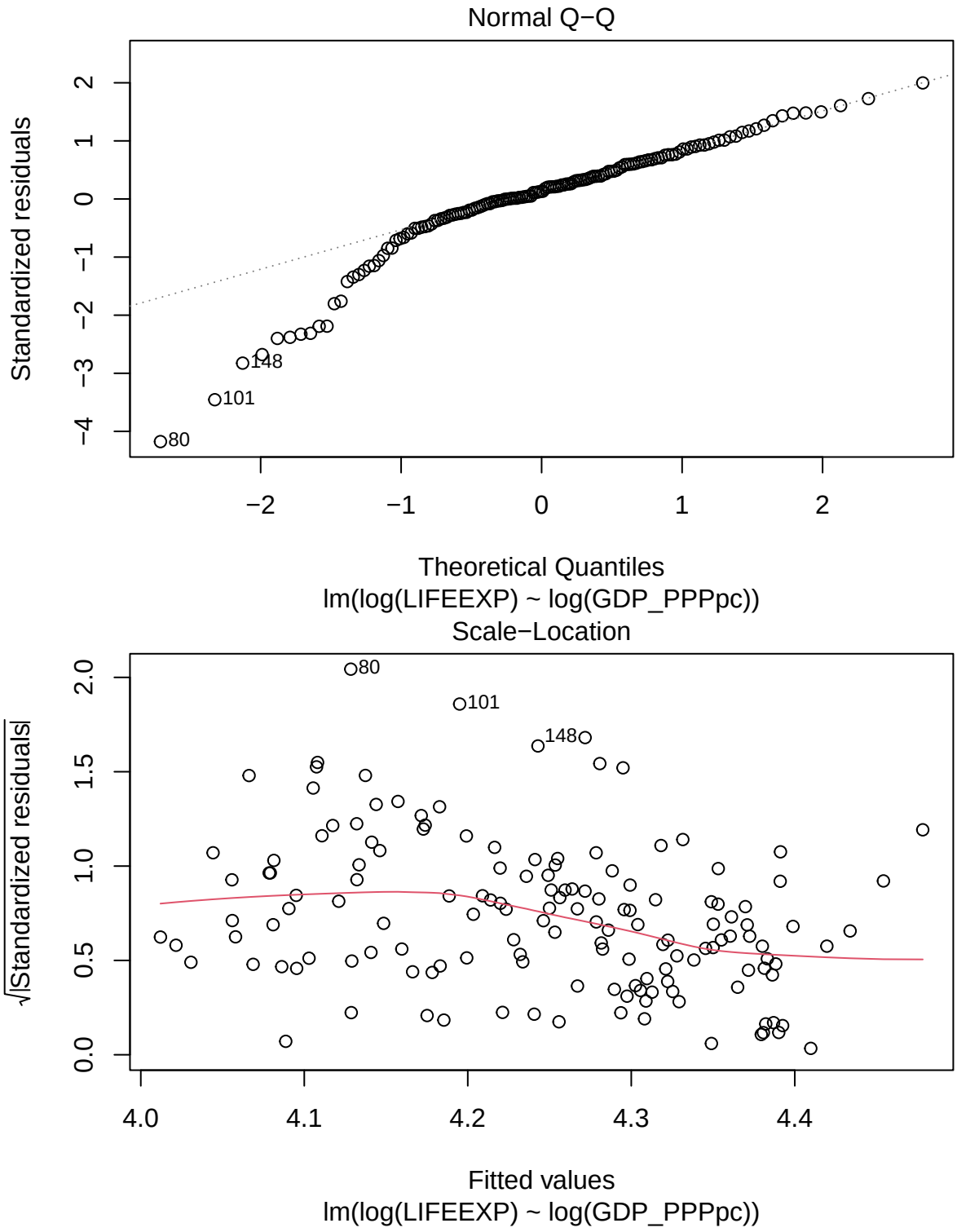

Residuals vs Fitted
Residuals
Fitted values
lm(LIFEEXP ~ log(GDP_PPPpc))
80
101
148
55
60
65
70
75
80
85
10
5
0
−5
−10
−20

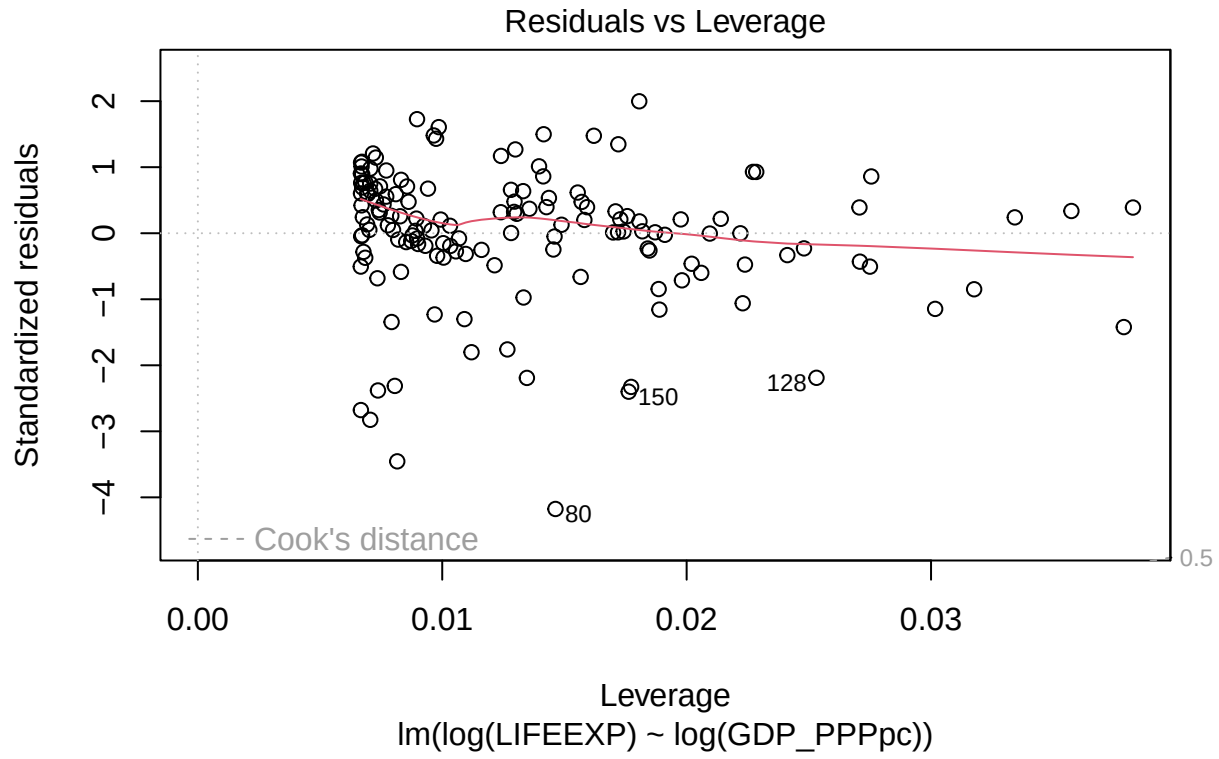

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(LIFEEXP ~ log(GDP_PPPpc))
Scale−Location
√|Standardized residuals|
Fitted values
lm(LIFEEXP ~ log(GDP_PPPpc))
80
101
148


Residuals vs Leverage
Standardized residuals
2
1
0
−1
−2
−3
−4
0.00
0.01
0.02
0.03
Cook's distance
101
80
128
Leverage
lm(LIFEEXP ~ log(GDP_PPPpc))

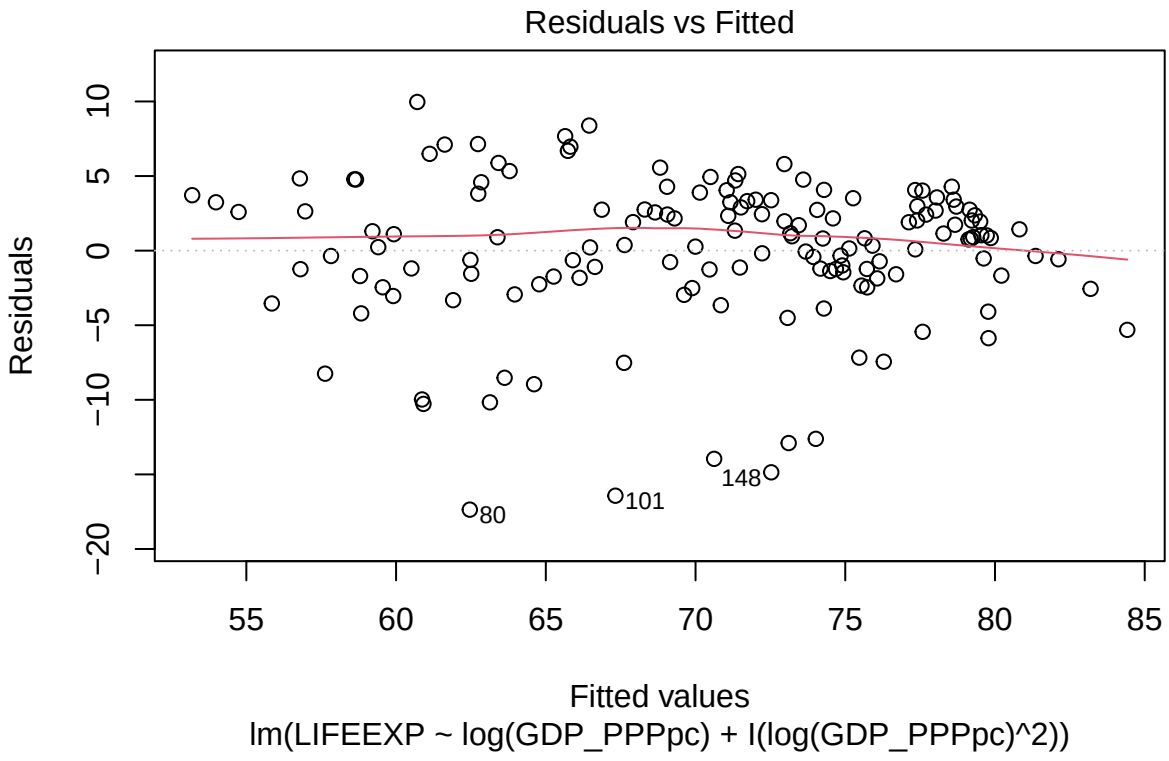
```
plot(model3)
```

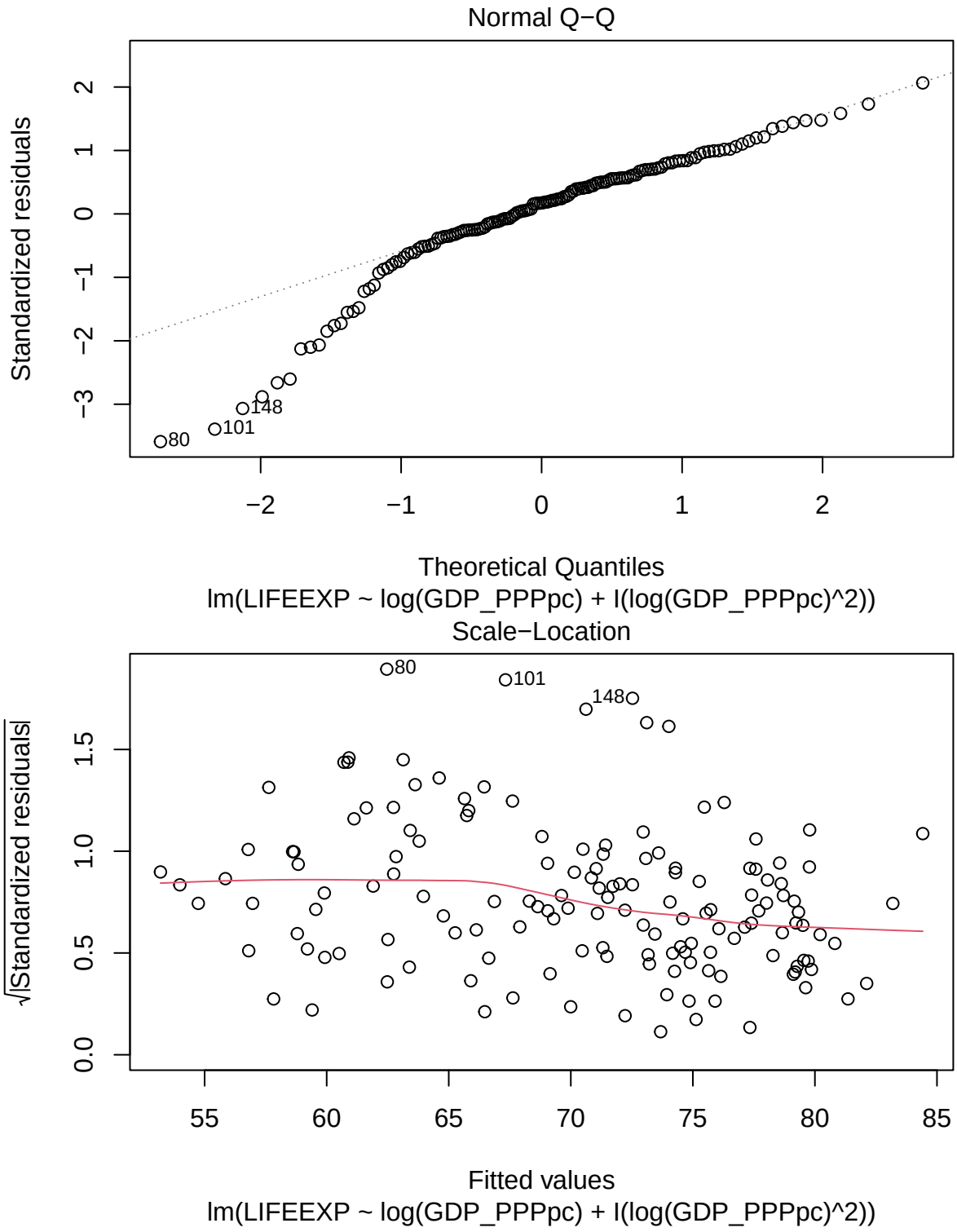

Residuals vs Fitted
Residuals
0.2
0.1
0.0
−0.1
−0.3
4.0
4.1
4.2
4.3
4.4
148
101
80
Fitted values
lm(log(LIFEEXP) ~ log(GDP_PPPpc))

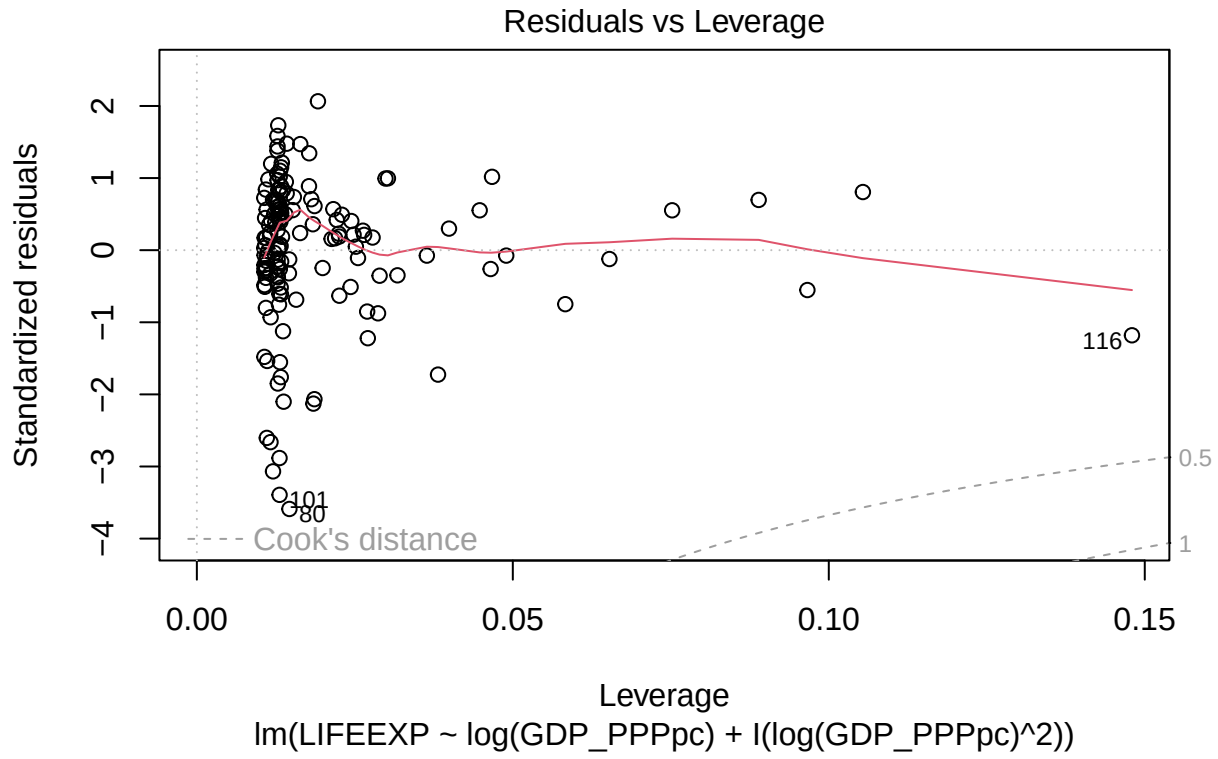

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(log(LIFEEXP) ~ log(GDP_PPPpc))
Scale−Location
√|Standardized residuals|
Fitted values
lm(log(LIFEEXP) ~ log(GDP_PPPpc))
80
101
148


Residuals vs Leverage
Standardized residuals
2
1
0
−1
−2
−3
−4
150
128
80
Cook's distance
0.5
0.00
0.01
0.02
0.03
Leverage
lm(log(LIFEEXP) ~ log(GDP_PPPpc))

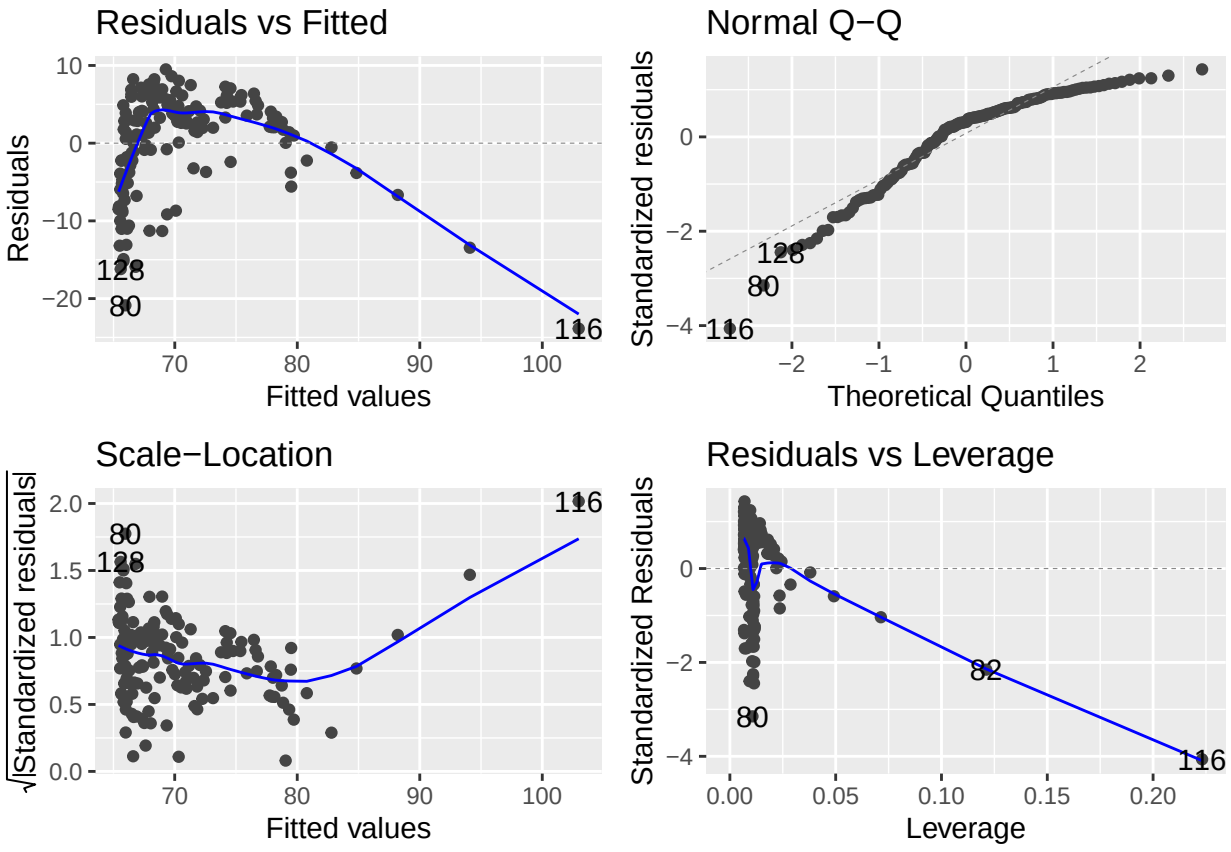
```
plot(model4)
```


Residuals vs Fitted
Residuals
10
5
0
−5
−10
−20
148
101
80
55
60
65
70
75
80
85
Fitted values
lm(LIFEEXP ~ log(GDP_PPPpc) + I(log(GDP_PPPpc)^2))


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(LIFEEXP ~ log(GDP_PPPpc) + I(log(GDP_PPPpc)^2))
Scale−Location
√|Standardized residuals|
Fitted values
lm(LIFEEXP ~ log(GDP_PPPpc) + I(log(GDP_PPPpc)^2))
80
101
148




```
library(ggfortify)
autoplot(model1)
```



Test auf Hetereoskedastie #### Breusch-Pagan Test - H_0 = Homoskedastie

```
library(lmtest)
bptest(model1)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model1
## BP = 3.4002, df = 1, p-value = 0.06519
```

```
bptest(model2)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model2
## BP = 4.7918, df = 1, p-value = 0.0286
```

```
bptest(model3)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model3
## BP = 7.1428, df = 1, p-value = 0.007527
```

```
bptest(model4)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model4
## BP = 7.3991, df = 2, p-value = 0.02474
```

Goldfeld-Quandt Test

- H_0 = Homoskedastie

```
gqtest(model1, alternative = "less")
```

```
##
##  Goldfeld-Quandt test
##
## data:  model1
## GQ = 2.088, df1 = 73, df2 = 73, p-value = 0.999
## alternative hypothesis: variance decreases from segment 1 to 2
```

```r
gqtest(model1, alternative = "greater") # Default
```

```
##
##  Goldfeld-Quandt test
##
## data:  model1
## GQ = 2.088, df1 = 73, df2 = 73, p-value = 0.0009676
## alternative hypothesis: variance increases from segment 1 to 2
```

```r
gqtest(model1, alternative = "two.sided")
```

```
##
##  Goldfeld-Quandt test
##
## data:  model1
## GQ = 2.088, df1 = 73, df2 = 73, p-value = 0.001935
## alternative hypothesis: variance changes from segment 1 to 2
```

```r
gqtest(model2)
```

```
##
##  Goldfeld-Quandt test
##
## data:  model2
## GQ = 1.8364, df1 = 73, df2 = 73, p-value = 0.005111
## alternative hypothesis: variance increases from segment 1 to 2
```

```r
gqtest(model3)
```

```
##
##  Goldfeld-Quandt test
##
## data:  model3
## GQ = 2.1784, df1 = 73, df2 = 73, p-value = 0.0005293
## alternative hypothesis: variance increases from segment 1 to 2
```

```r
gqtest(model4)
```

```
##
##  Goldfeld-Quandt test
##
## data:  model4
## GQ = 1.7924, df1 = 72, df2 = 72, p-value = 0.007141
## alternative hypothesis: variance increases from segment 1 to 2
```

Reaktion auf Heteroskedastie Siehe Skript!

```r
var_covar_matrix <- sandwich::vcovHC(model2, type = "HC1")
lmtest::coeftest(model2, vcov. = var_covar_matrix)
```

```
##
## t test of coefficients:
##
##                Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    15.11241    2.95323  5.1172  9.48e-07 ***
## log(GDP_PPPpc)  6.07485    0.30545 19.8879 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(model2)
```

```
##
## Call:
## lm(formula = LIFEEXP ~ log(GDP_PPPpc), data = lifeexp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.3420  -1.7187   0.6502   3.0746   9.7919
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     15.1124     3.0494   4.956 1.95e-06 ***
## log(GDP_PPPpc)   6.0748     0.3322  18.287  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.876 on 148 degrees of freedom
## Multiple R-squared:  0.6932, Adjusted R-squared:  0.6911
## F-statistic: 334.4 on 1 and 148 DF,  p-value: < 2.2e-16
```

Übungsaufgabe: Regression Lebenserwartung und Gini

Experimentiere mit unterschiedlichen Regressionsmodellen.

Modelle

```r
model5 <- lm(LIFEEXP ~ GINI_post, lifeexp_data)
summary(model5)
```

```
##
## Call:
## lm(formula = LIFEEXP ~ GINI_post, data = lifeexp_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -20.853  -5.839   1.694   5.828  12.863
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  91.5636     3.2164  28.468  < 2e-16 ***
## GINI_post    -0.5569     0.0830  -6.709 3.88e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.709 on 148 degrees of freedom
## Multiple R-squared:  0.2332, Adjusted R-squared:  0.228
## F-statistic: 45.01 on 1 and 148 DF,  p-value: 3.88e-10
```

```
model6 <- lm(LIFEEXP ~ log(GINI_post), lifeexp_data)
summary(model6)
```

```
##
## Call:
## lm(formula = LIFEEXP ~ log(GINI_post), data = lifeexp_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.429  -5.474   1.779   5.902  12.793
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     146.921     11.294  13.009  < 2e-16 ***
## log(GINI_post)  -21.152      3.117  -6.786 2.59e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.689 on 148 degrees of freedom
## Multiple R-squared:  0.2373, Adjusted R-squared:  0.2322
## F-statistic: 46.05 on 1 and 148 DF,  p-value: 2.591e-10
```

```
model7 <- lm(log(LIFEEXP) ~ GINI_post, lifeexp_data)
summary(model7)
```

```
##
## Call:
## lm(formula = log(LIFEEXP) ~ GINI_post, data = lifeexp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.34381 -0.07779  0.03034  0.08779  0.18807
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.559278   0.048782  93.463  < 2e-16 ***
## GINI_post   -0.008248   0.001259  -6.552 8.84e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.1169 on 148 degrees of freedom
## Multiple R-squared:  0.2248, Adjusted R-squared:  0.2196
## F-statistic: 42.92 on 1 and 148 DF,  p-value: 8.842e-10
```

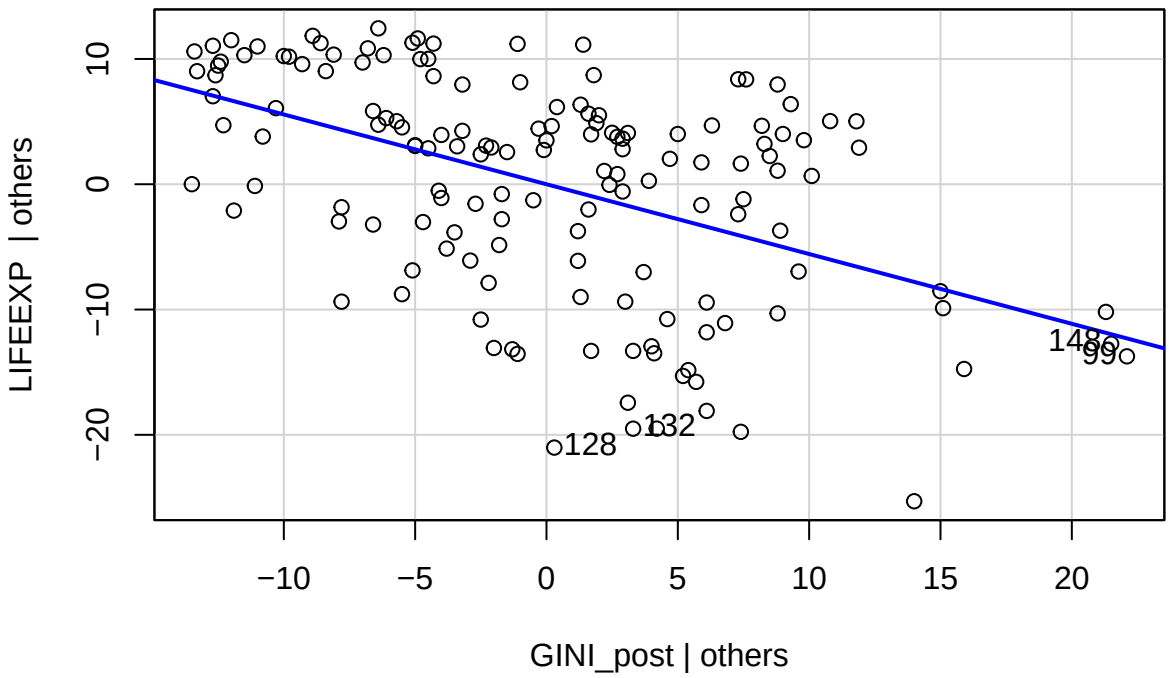
```
model8 <- lm(log(LIFEEXP) ~ log(GINI_post), lifeexp_data)
summary(model8)
```

```
##
## Call:
## lm(formula = log(LIFEEXP) ~ log(GINI_post), data = lifeexp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.33757 -0.07700  0.03081  0.09048  0.18680
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     5.37547    0.17147  31.349  < 2e-16 ***
## log(GINI_post) -0.31226    0.04733  -6.598 6.95e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1167 on 148 degrees of freedom
## Multiple R-squared:  0.2273, Adjusted R-squared:  0.2221
## F-statistic: 43.53 on 1 and 148 DF,  p-value: 6.948e-10
```

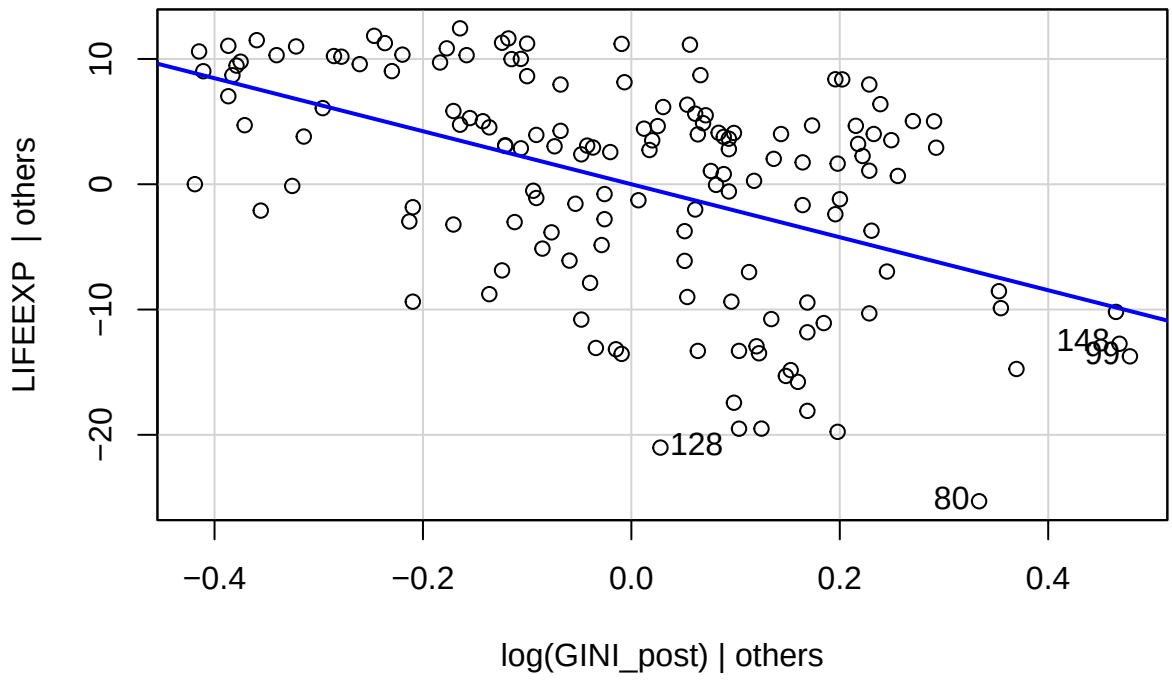
Regressionsdiagnostik

Plots

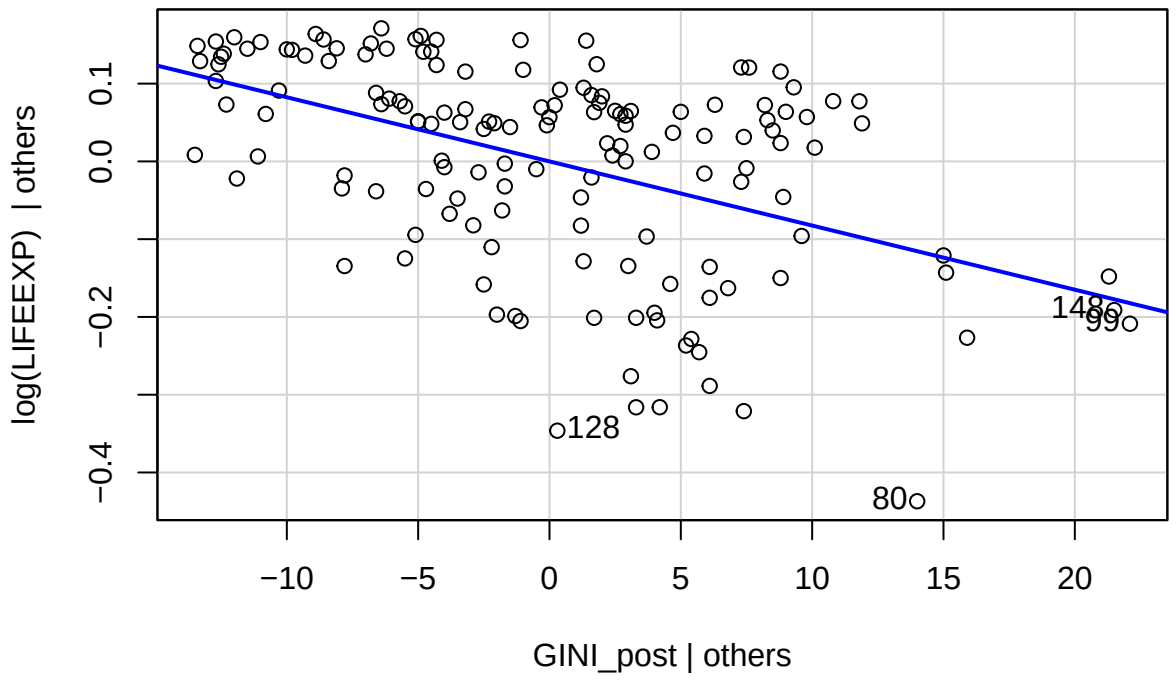
```
avPlots(model5)
```



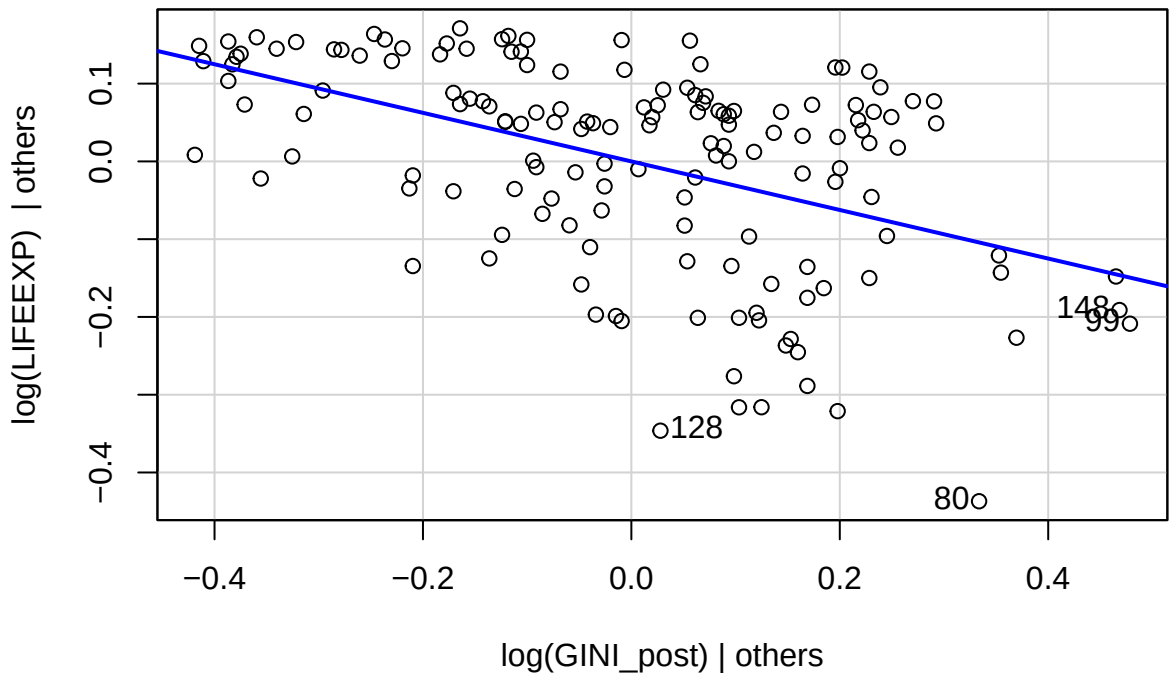

```
avPlots(model6)
```




```
avPlots(model7)
```

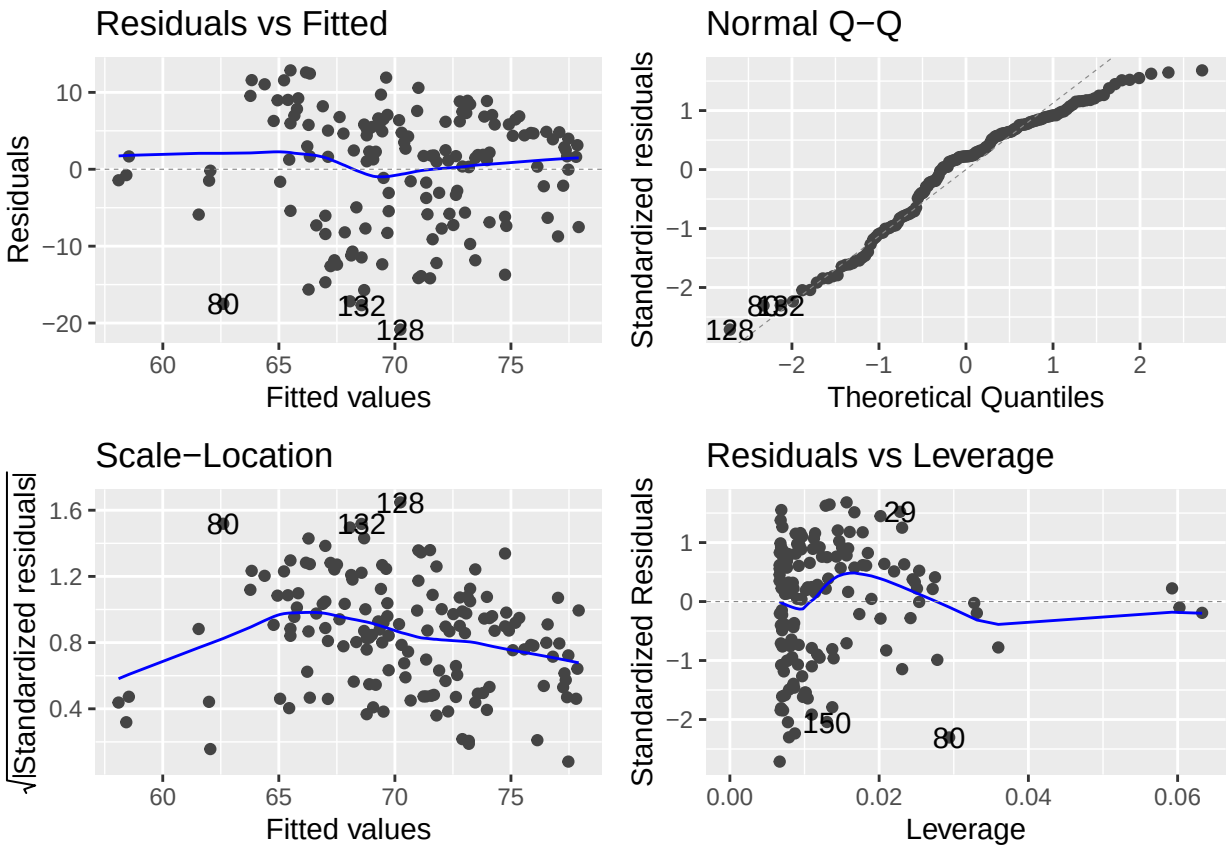



```
avPlots(model8)
```

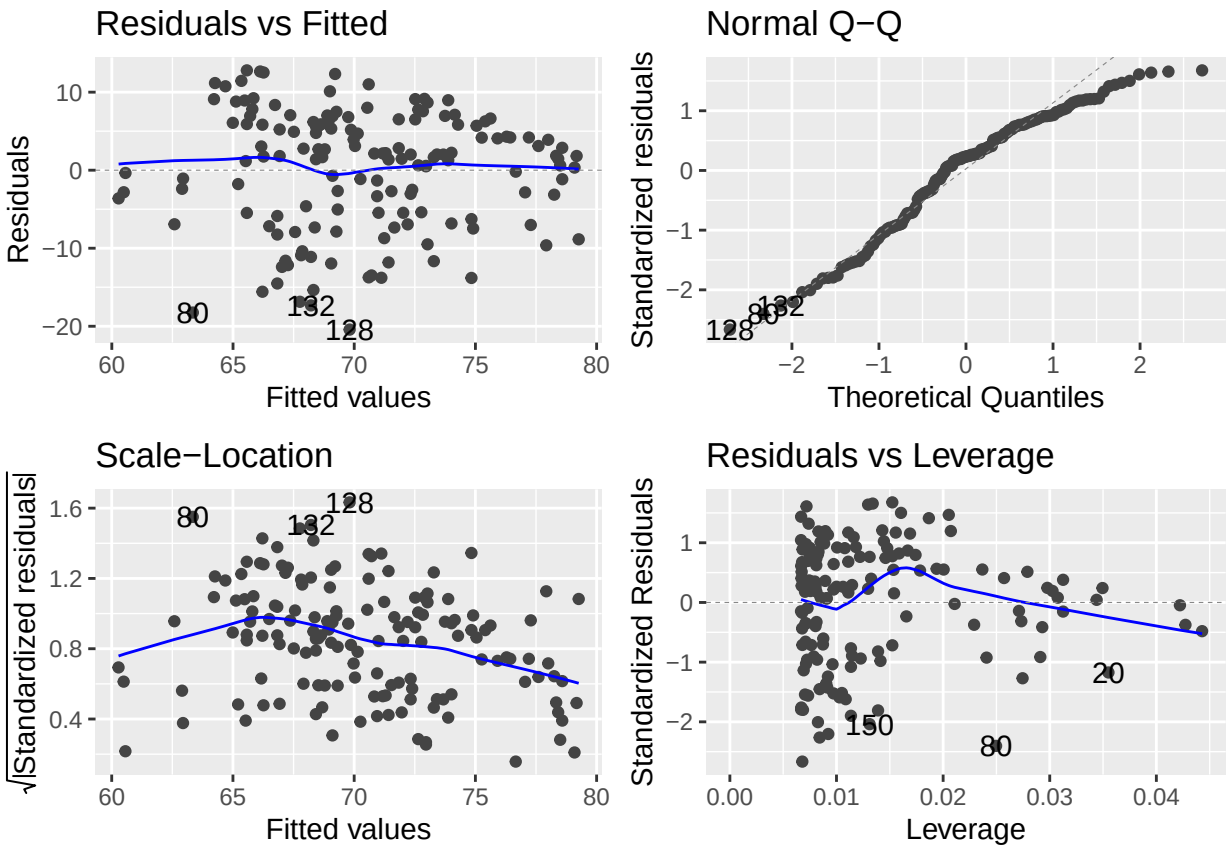


AV-Plot, QQ-Plot, etc.

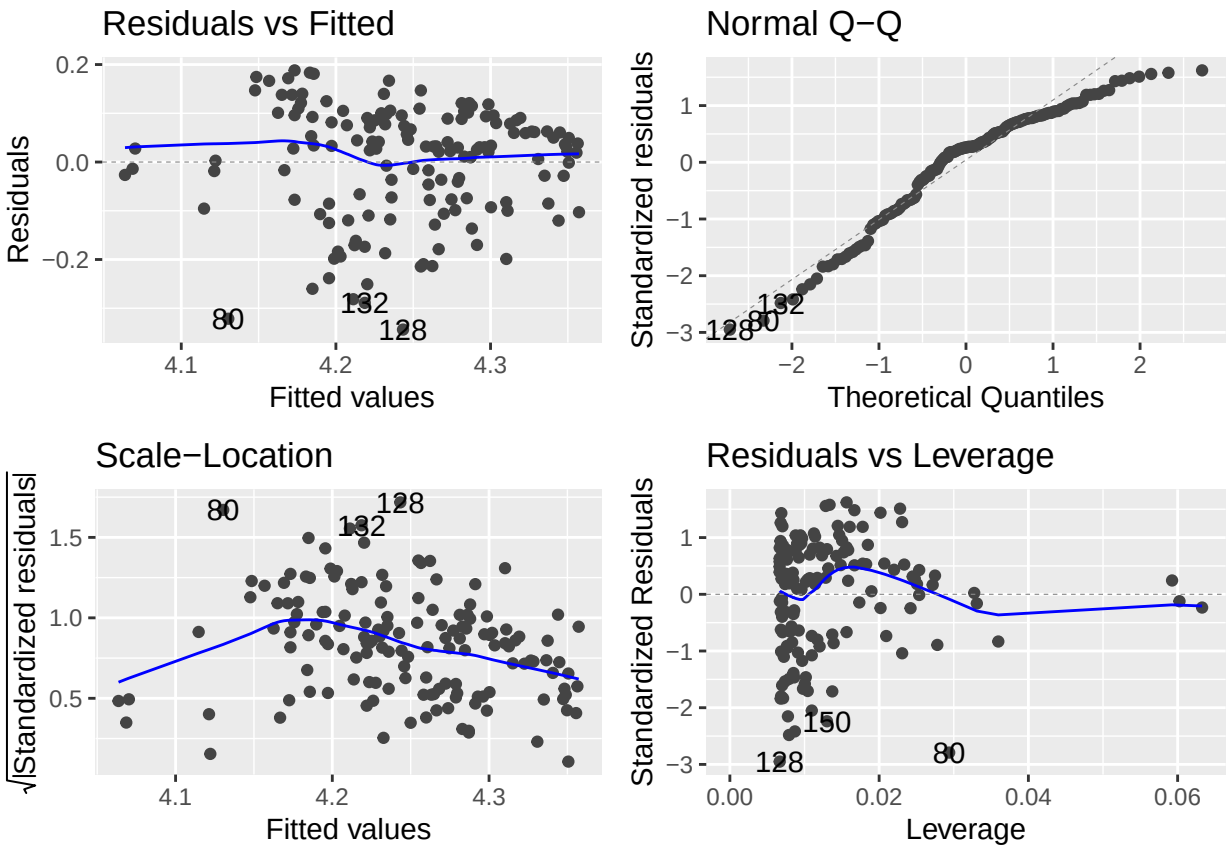
```
autoplot(model5)
```



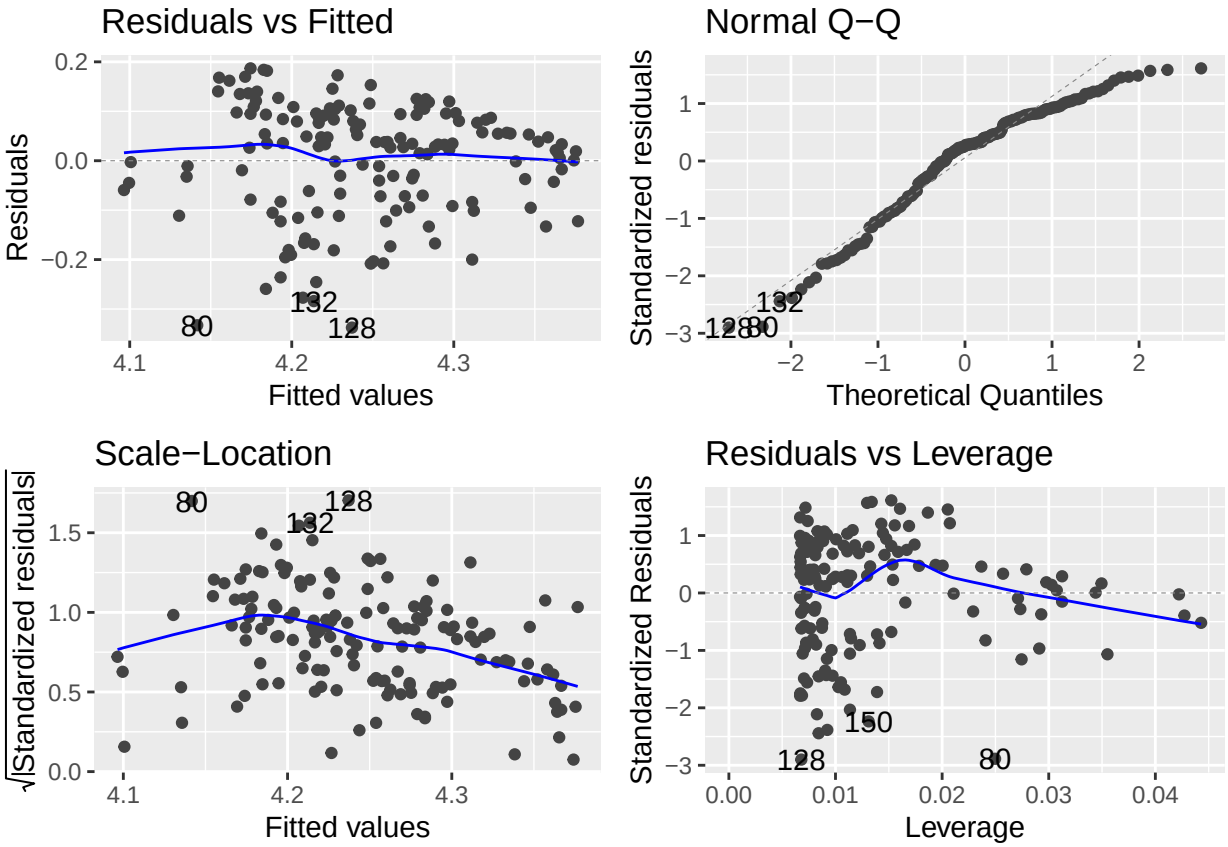

```
autoplot(model6)
```




```
autoplot(model7)
```




```
autoplot(model8)
```



Regressionen: Lebenserwartung und Mordrate

Modelle

```
model9 <- lm(LIFEEXP ~ MORTRATE, lifeexp_data)
summary(model9)
```

```
##
## Call:
## lm(formula = LIFEEXP ~ MORTRATE, data = lifeexp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.7510  -1.8849  -0.0202   2.7103  24.0910
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 77.660903   0.451427  172.03   <2e-16 ***
## MORTRATE    -0.197703   0.008393  -23.56   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.04 on 148 degrees of freedom
```

```
## Multiple R-squared:  0.7894, Adjusted R-squared:  0.788
## F-statistic: 554.9 on 1 and 148 DF,  p-value: < 2.2e-16
```

```
model10 <- lm(LIFEEXP ~ log(MORTRATE), lifeexp_data)
summary(model10)
```

```
##
## Call:
## lm(formula = LIFEEXP ~ log(MORTRATE), data = lifeexp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.2422  -2.4686   0.6188   2.6438   7.2300
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    90.9512     0.8155  111.53   <2e-16 ***
## log(MORTRATE)  -6.8608     0.2538  -27.03   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.613 on 148 degrees of freedom
## Multiple R-squared:  0.8316, Adjusted R-squared:  0.8304
## F-statistic: 730.7 on 1 and 148 DF,  p-value: < 2.2e-16
```

```
model11 <- lm(log(LIFEEXP) ~ MORTRATE, lifeexp_data)
summary(model11)
```

```
##
## Call:
## lm(formula = log(LIFEEXP) ~ MORTRATE, data = lifeexp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.24680 -0.02441  0.00357  0.03596  0.37191
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.3558780  0.0066897  651.13   <2e-16 ***
## MORTRATE    -0.0029960  0.0001244  -24.09   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05987 on 148 degrees of freedom
## Multiple R-squared:  0.7968, Adjusted R-squared:  0.7954
## F-statistic: 580.3 on 1 and 148 DF,  p-value: < 2.2e-16
```

```
model12 <- lm(log(LIFEEXP) ~ log(MORTRATE), lifeexp_data)
summary(model12)
```

```
##
## Call:
```

```
## lm(formula = log(LIFEEXP) ~ log(MORTRATE), data = lifeexp_data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.273439 -0.035280  0.006534  0.045288  0.111400
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.54976    0.01343  338.81   <2e-16 ***
## log(MORTRATE) -0.10146    0.00418  -24.27   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0595 on 148 degrees of freedom
## Multiple R-squared:  0.7993, Adjusted R-squared:  0.7979
## F-statistic: 589.3 on 1 and 148 DF,  p-value: < 2.2e-16
```

```
model13 <- lm(LIFEEXP ~ log(MORTRATE) + I(log(MORTRATE)**2), lifeexp_data)
summary(model13)
```

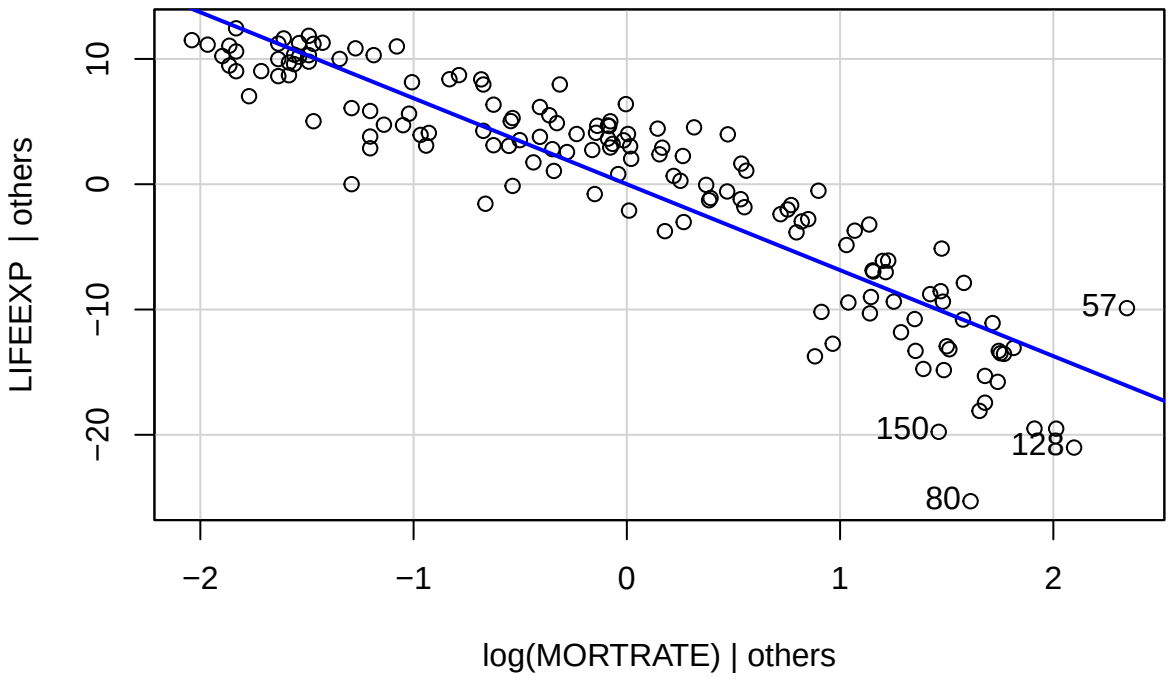
```
##
## Call:
## lm(formula = LIFEEXP ~ log(MORTRATE) + I(log(MORTRATE)^2), data = lifeexp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.5721  -1.9768   0.7967   2.0409  11.8451
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          80.3630     1.7784  45.188  < 2e-16 ***
## log(MORTRATE)         1.4221     1.2914   1.101    0.273
## I(log(MORTRATE)^2)   -1.3777     0.2115  -6.513  1.1e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.194 on 147 degrees of freedom
## Multiple R-squared:  0.8693, Adjusted R-squared:  0.8675
## F-statistic: 488.8 on 2 and 147 DF,  p-value: < 2.2e-16
```

Regressionsdiagnostik

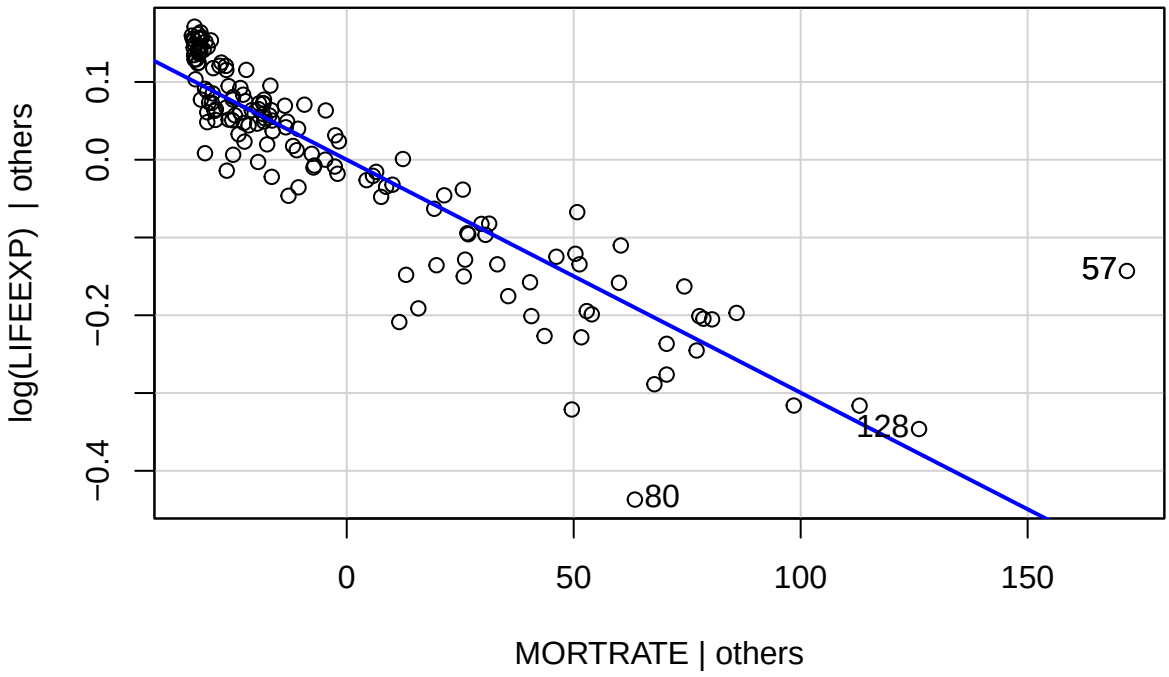
Scatter Plots

```
avPlots(model9)
```

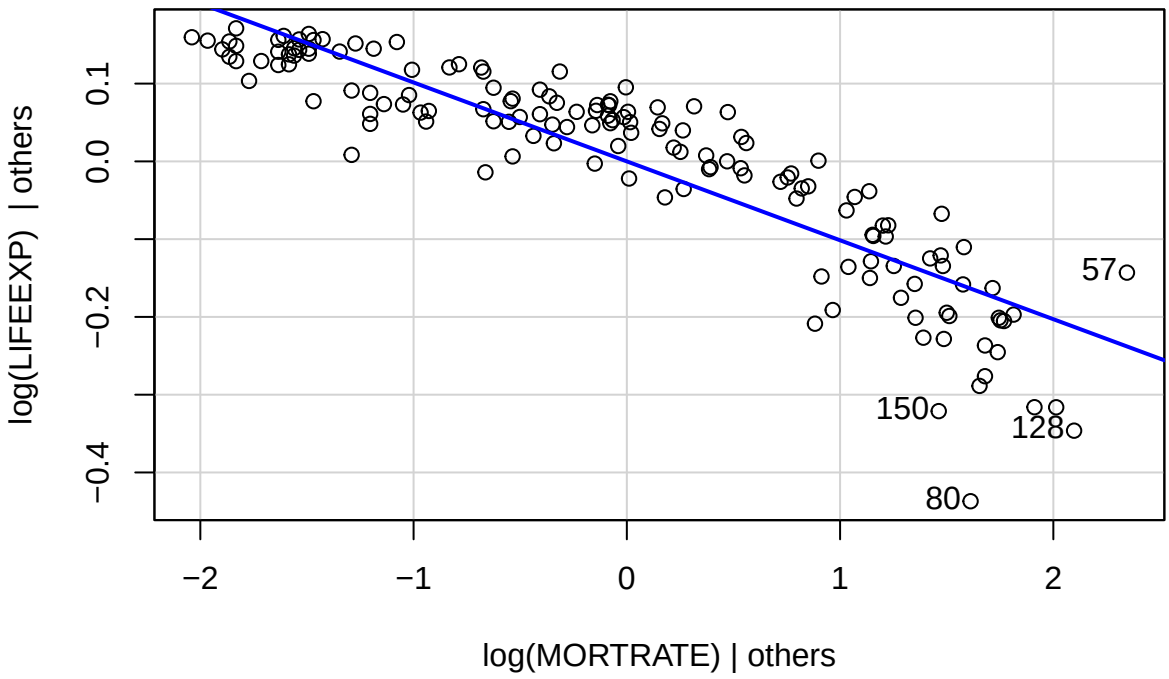

```
avPlots(model10)
```



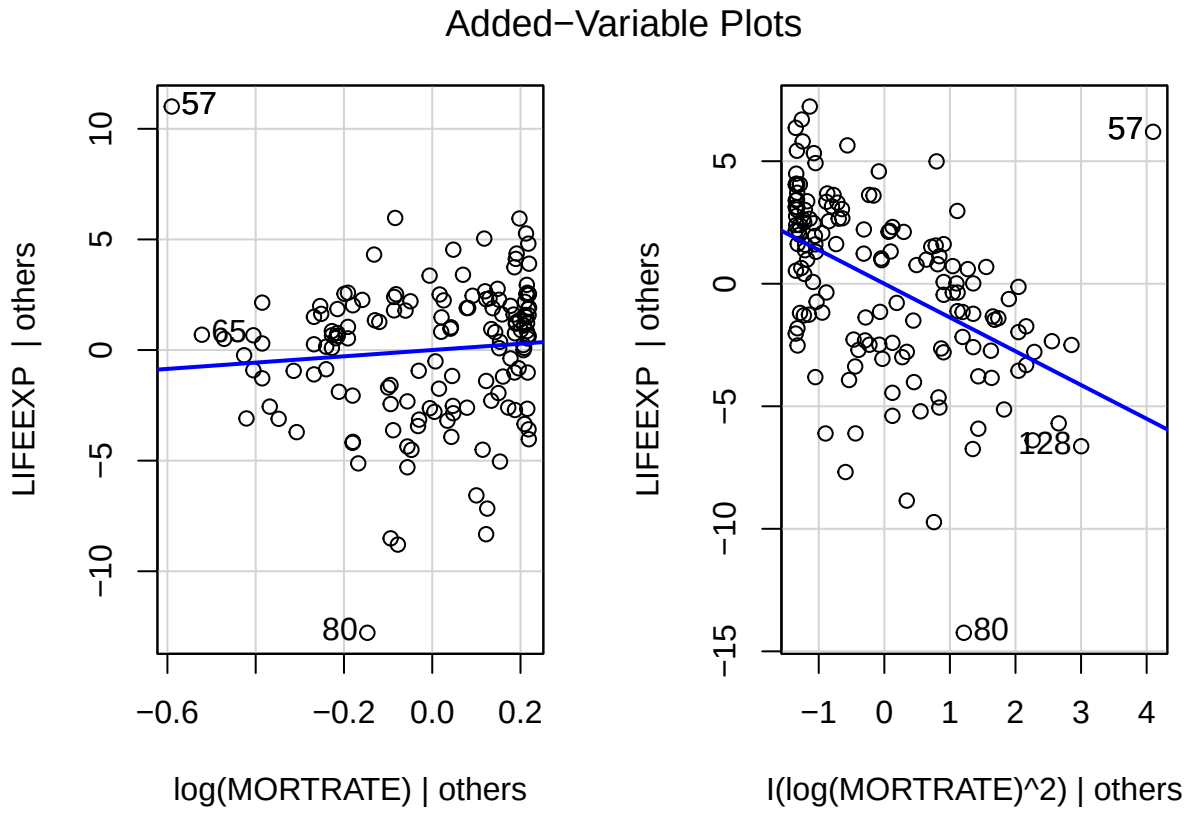

```
avPlots(model11)
```



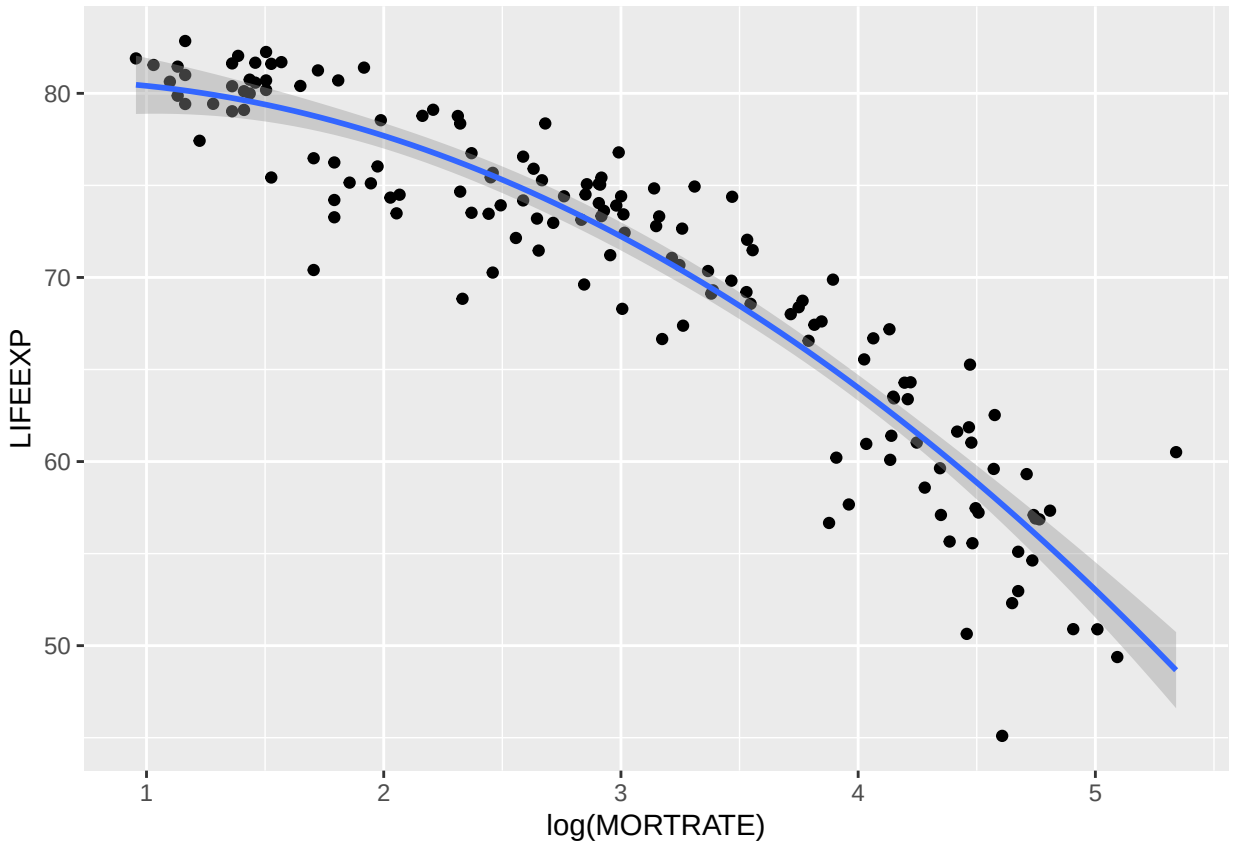

```
avPlots(model12)
```




```
avPlots(model13)
```

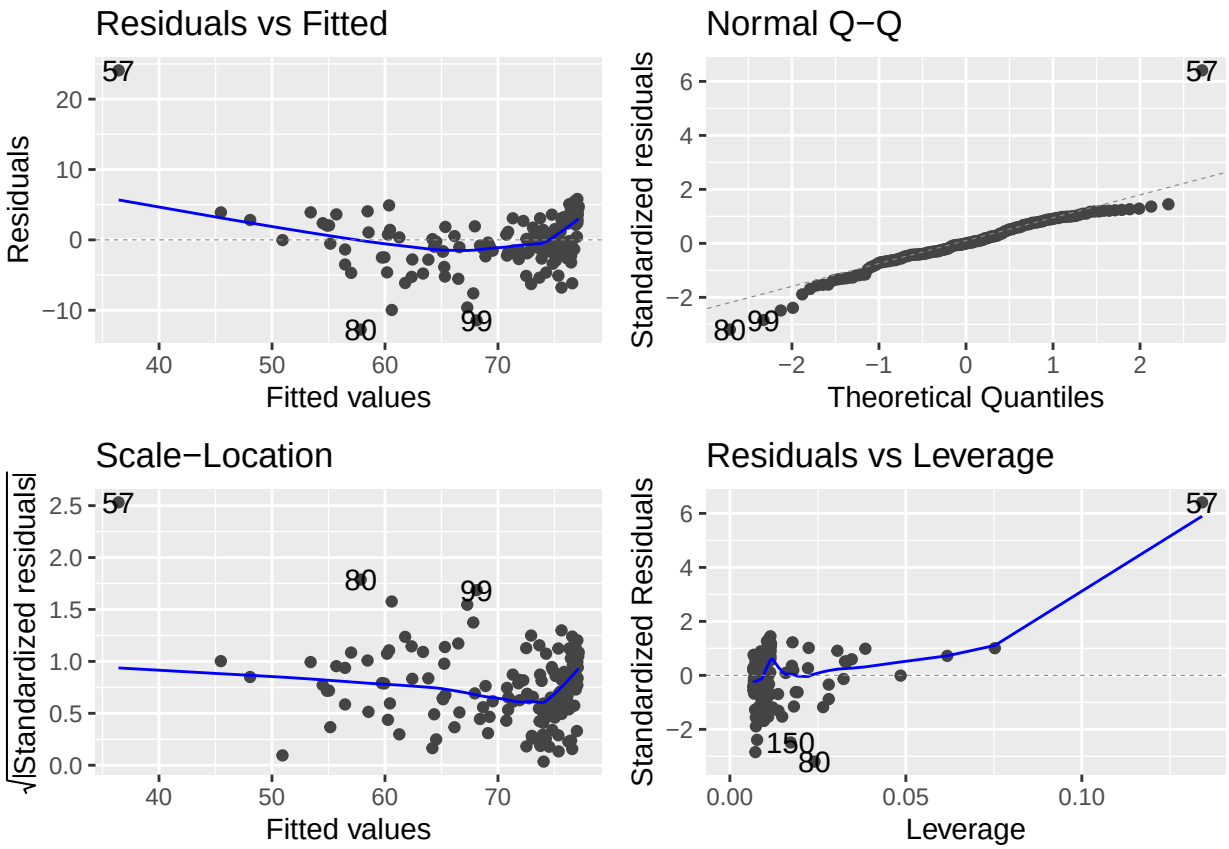



```
ggplot(lifeexp_data, aes(x = log(MORTRATE), y = LIFEEXP))+
  geom_point()+
  geom_smooth(method = "lm", formula = y ~ x + I(x**2))
```

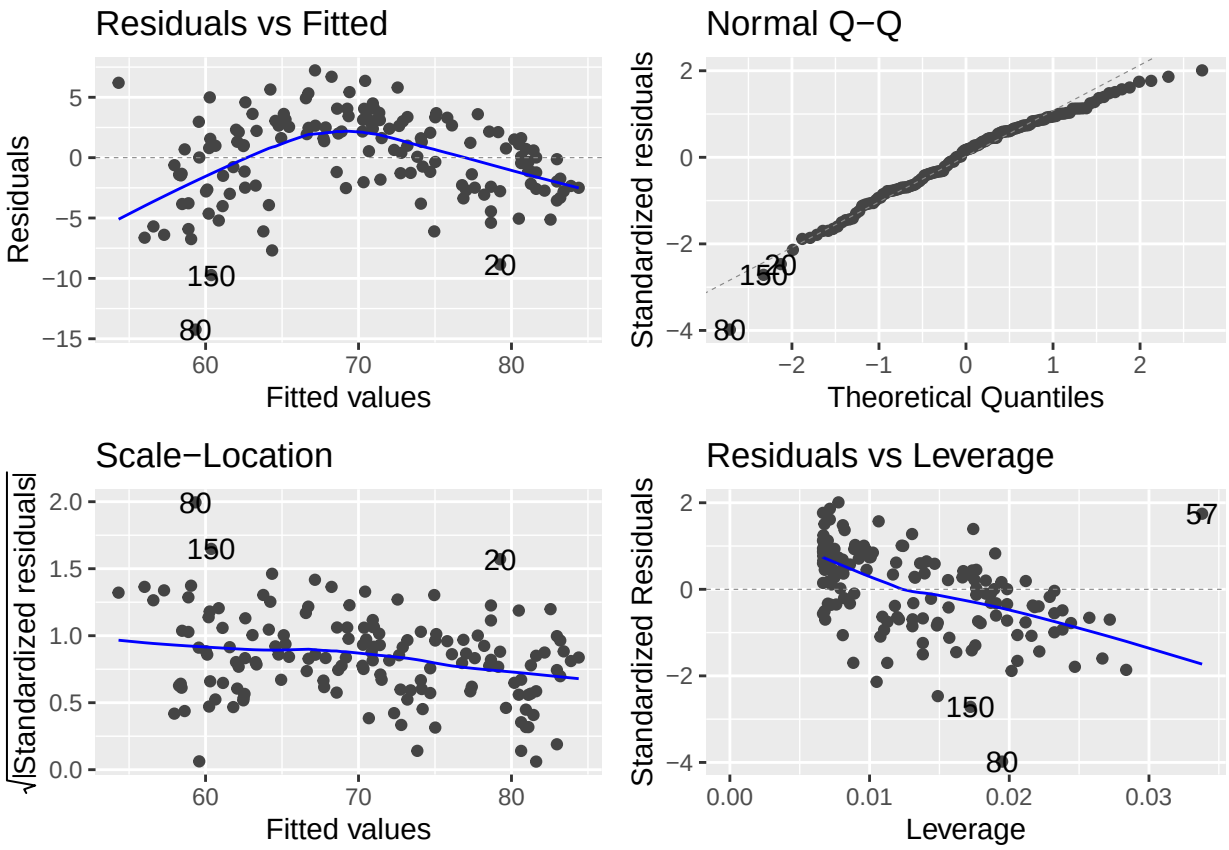


TA-, QQ-Plots, etc.

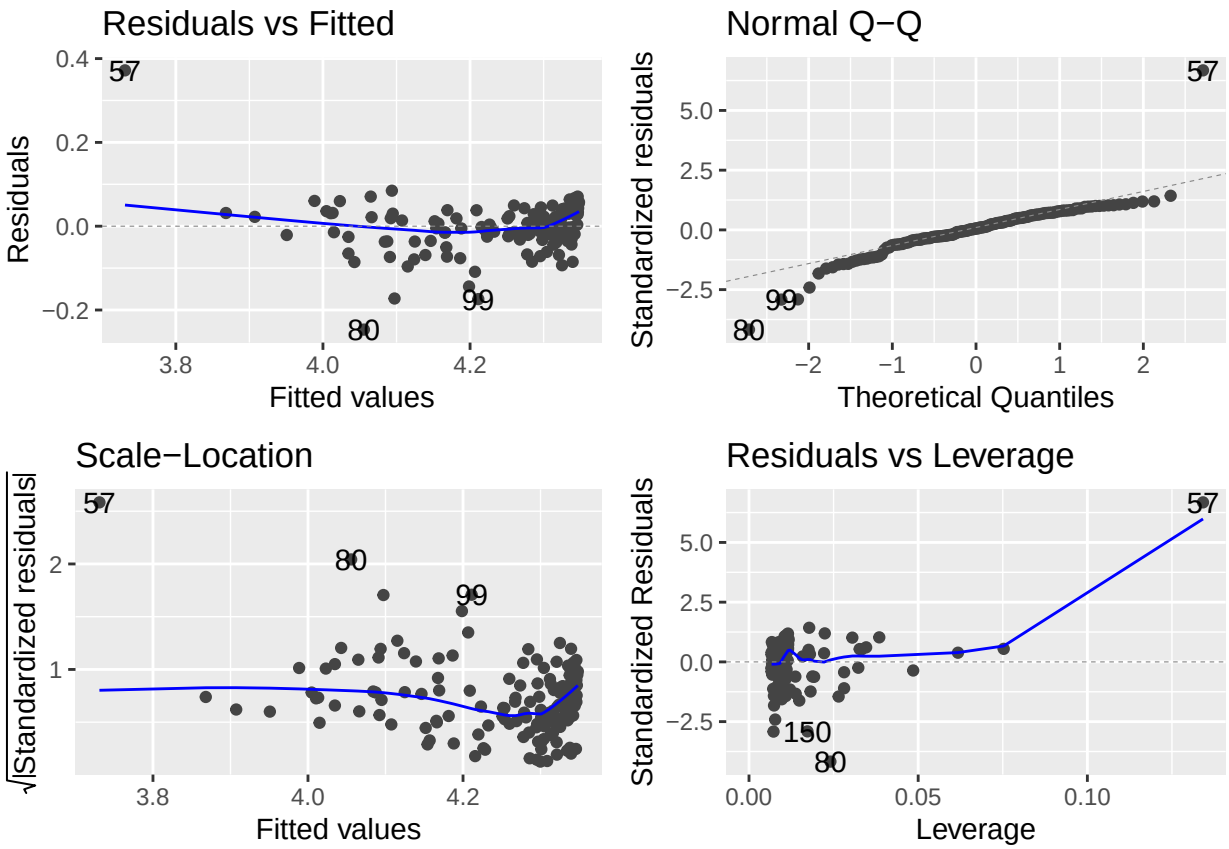
```
autoplot(model9)
```



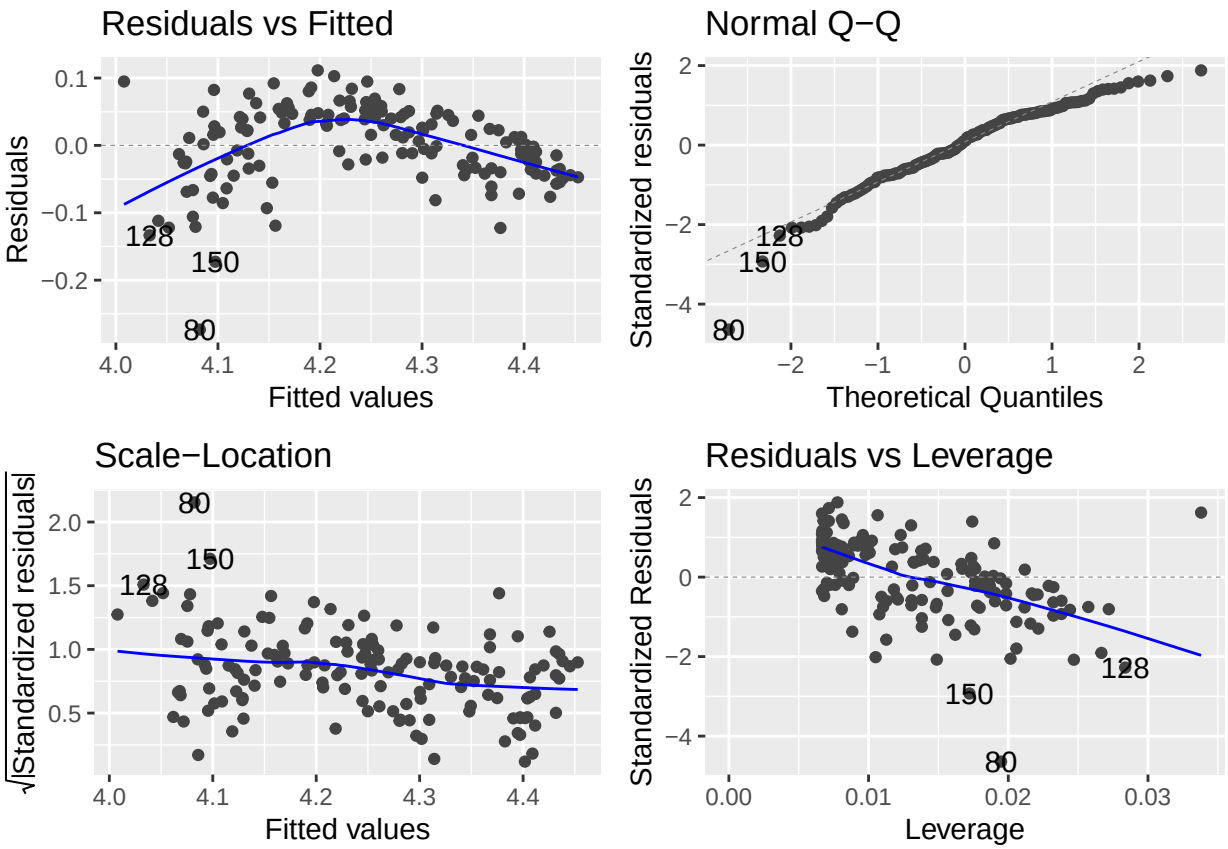

```
autoplot(model10)
```



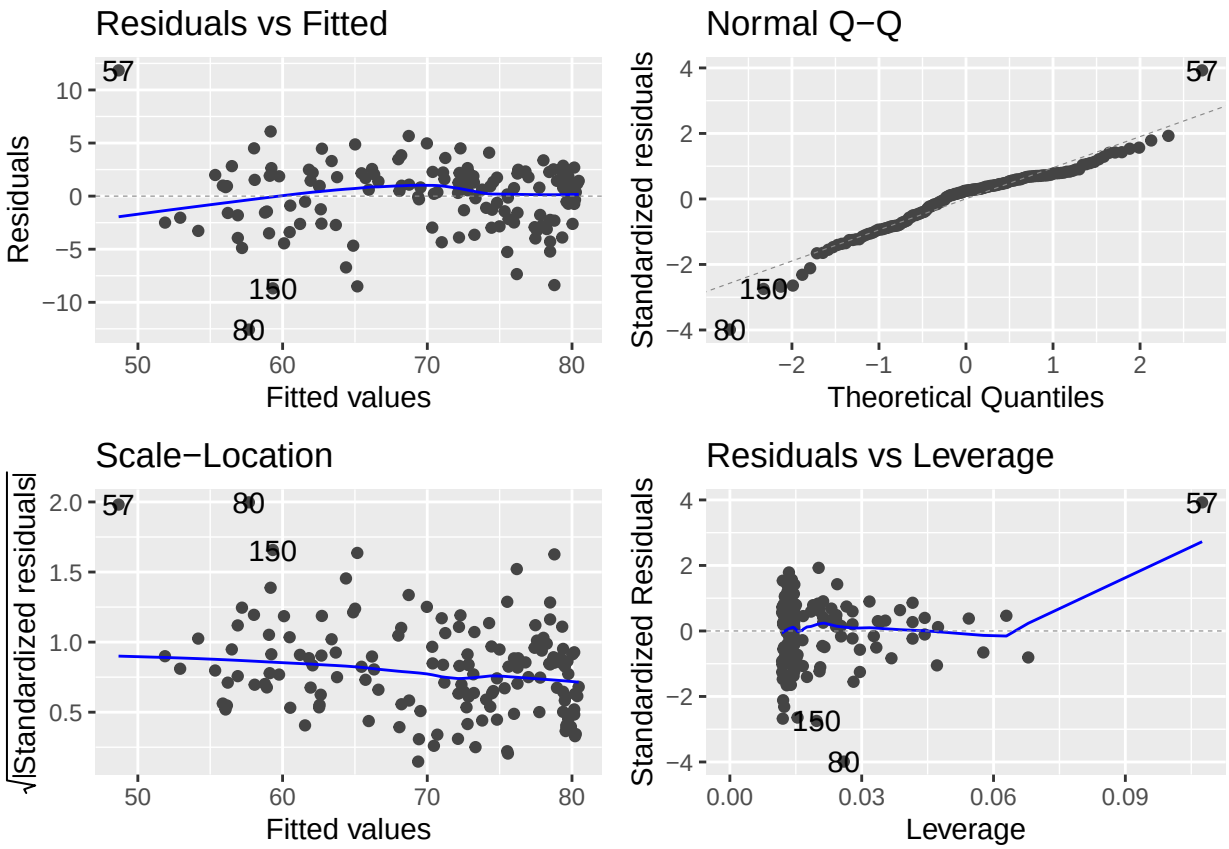

```
autoplot(model11)
```




```
autoplot(model12)
```




```
autoplot(model13)
```



Multivariate Regression

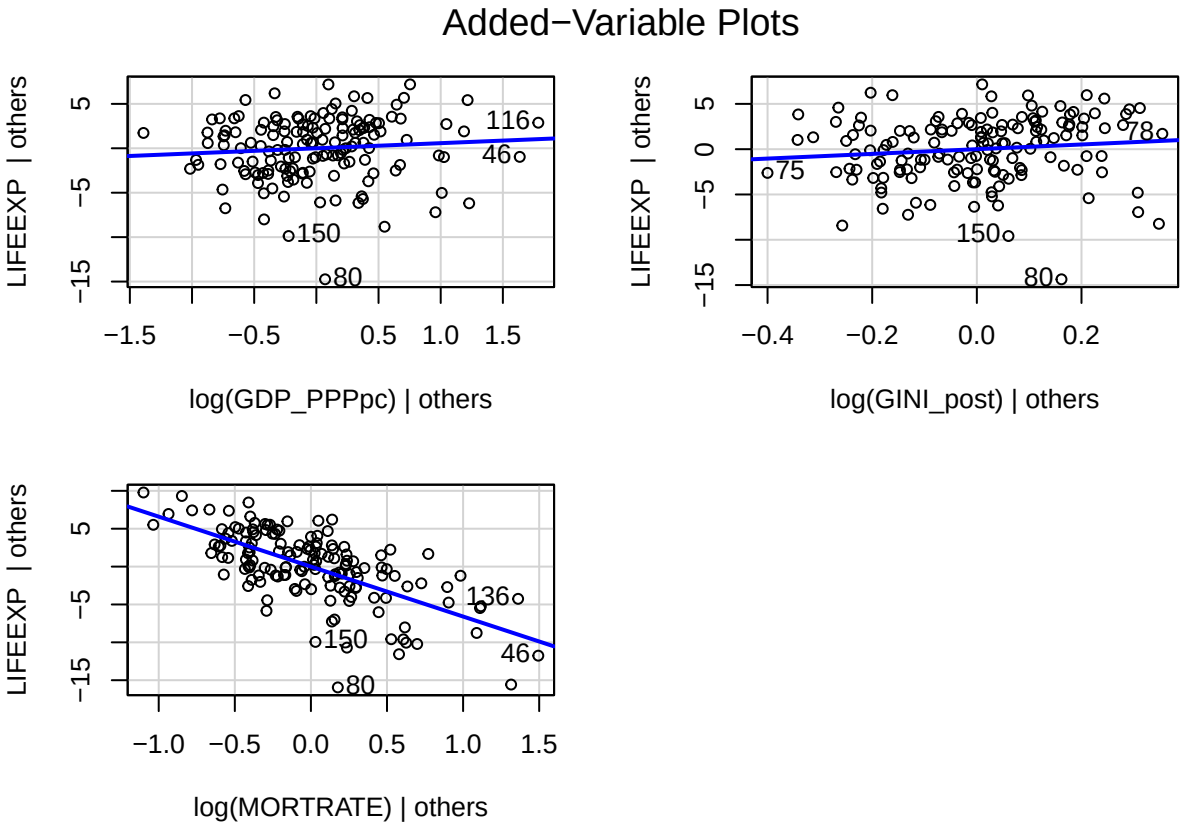
Modelle und Diagnostik

```
model14 <- lm(LIFEEXP ~ log(GDP_PPPpc) + log(GINI_post) + log(MORTRATE), lifeexp_data)
summary(model14)
```

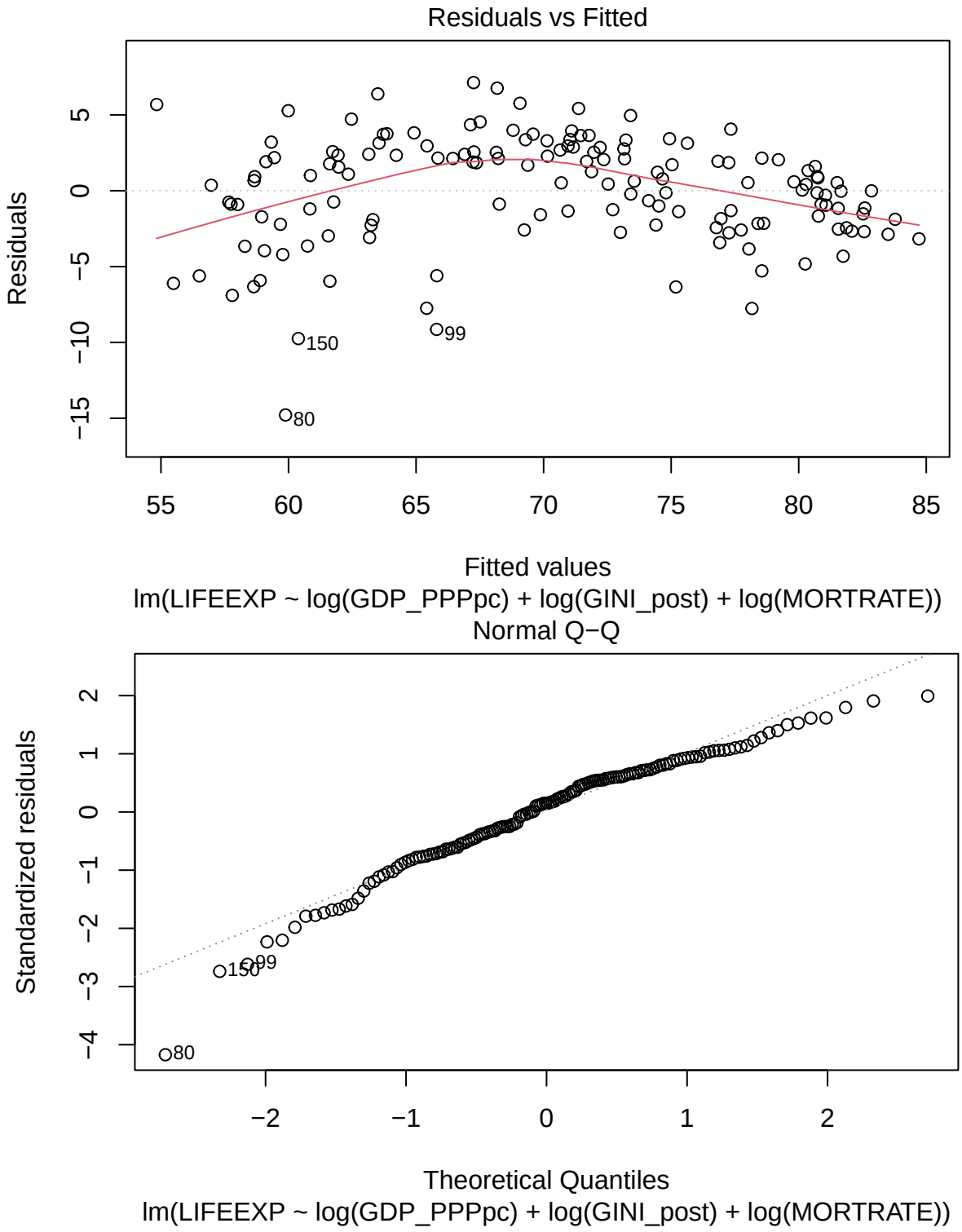
```
##
## Call:
## lm(formula = LIFEEXP ~ log(GDP_PPPpc) + log(GINI_post) + log(MORTRATE),
##     data = lifeexp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.7835  -2.2006   0.5298   2.4959   7.1272
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     75.5242     8.0831   9.343   <2e-16 ***
## log(GDP_PPPpc)   0.5763     0.5473   1.053    0.294
## log(GINI_post)   2.5931     1.8210   1.424    0.157
## log(MORTRATE)   -6.5934     0.6189 -10.654   <2e-16 ***
## ---
```

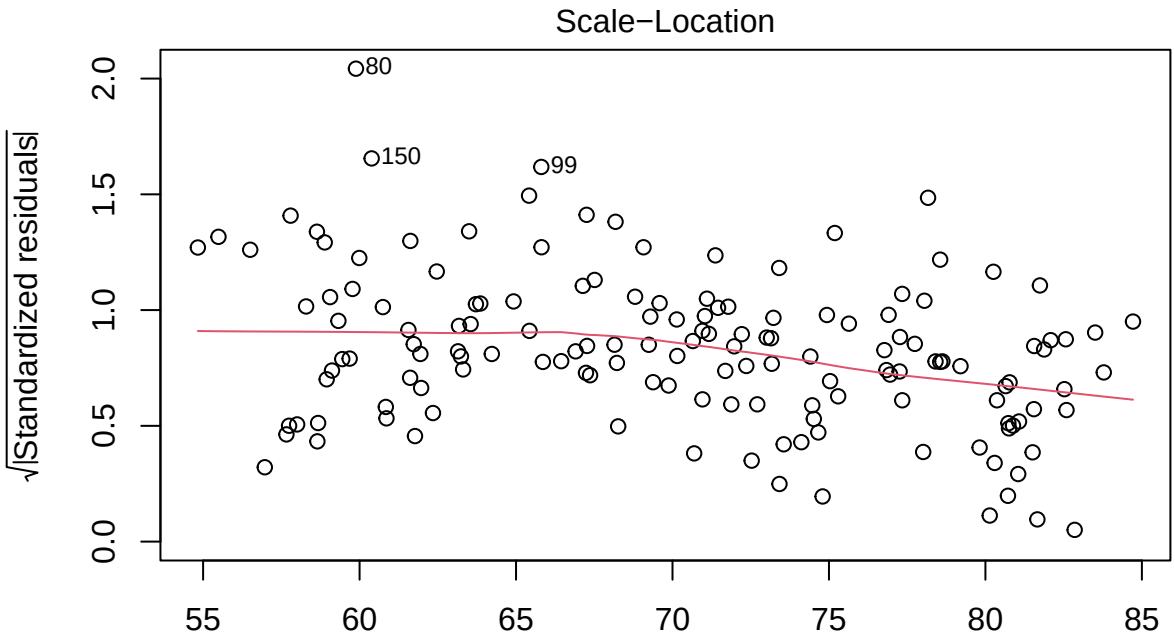
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.592 on 146 degrees of freedom
## Multiple R-squared:  0.8358, Adjusted R-squared:  0.8324
## F-statistic: 247.8 on 3 and 146 DF,  p-value: < 2.2e-16
```

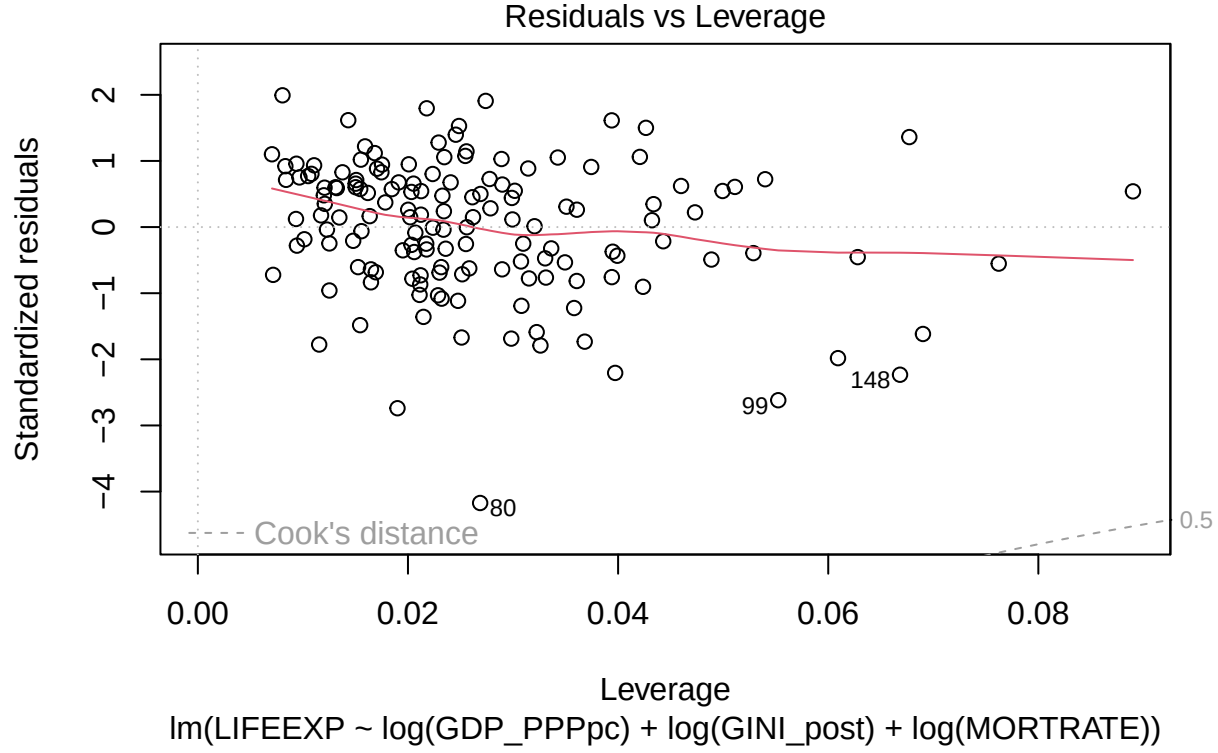
```
avPlots(model14)
```




```
plot(model14)
```


Residuals vs Fitted
Residuals
5
0
−5
−10
−15
150
99
80
55
60
65
70
75
80
85
Fitted values
lm(LIFEEXP ~ log(GDP_PPPpc) + log(GINI_post) + log(MORTRATE))
Normal Q−Q
Standardized residuals
2
1
0
−1
−2
−3
−4
150
99
80
−2
−1
0
1
2
Theoretical Quantiles
lm(LIFEEXP ~ log(GDP_PPPpc) + log(GINI_post) + log(MORTRATE))



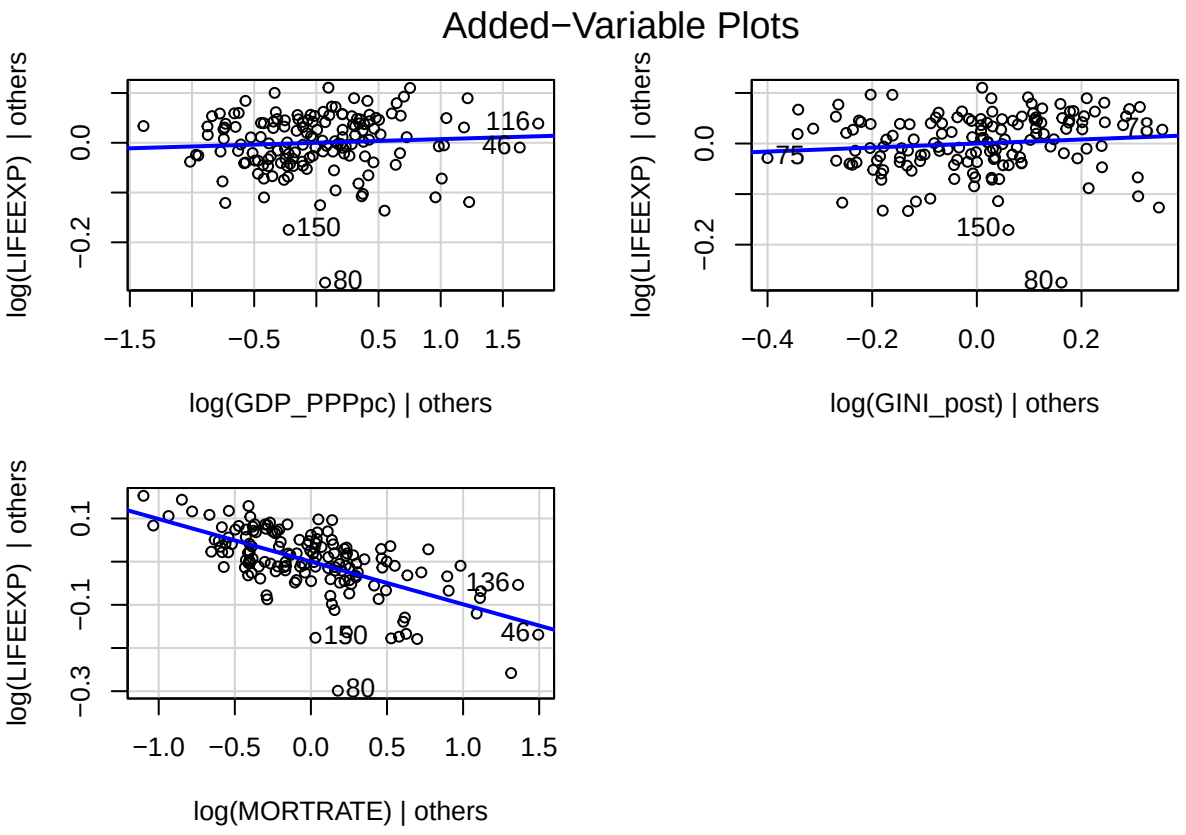



```
model15 <- lm(log(LIFEEXP) ~ log(GDP_PPPpc) + log(GINI_post) + log(MORTRATE), lifeexp_data)
summary(model15)
```

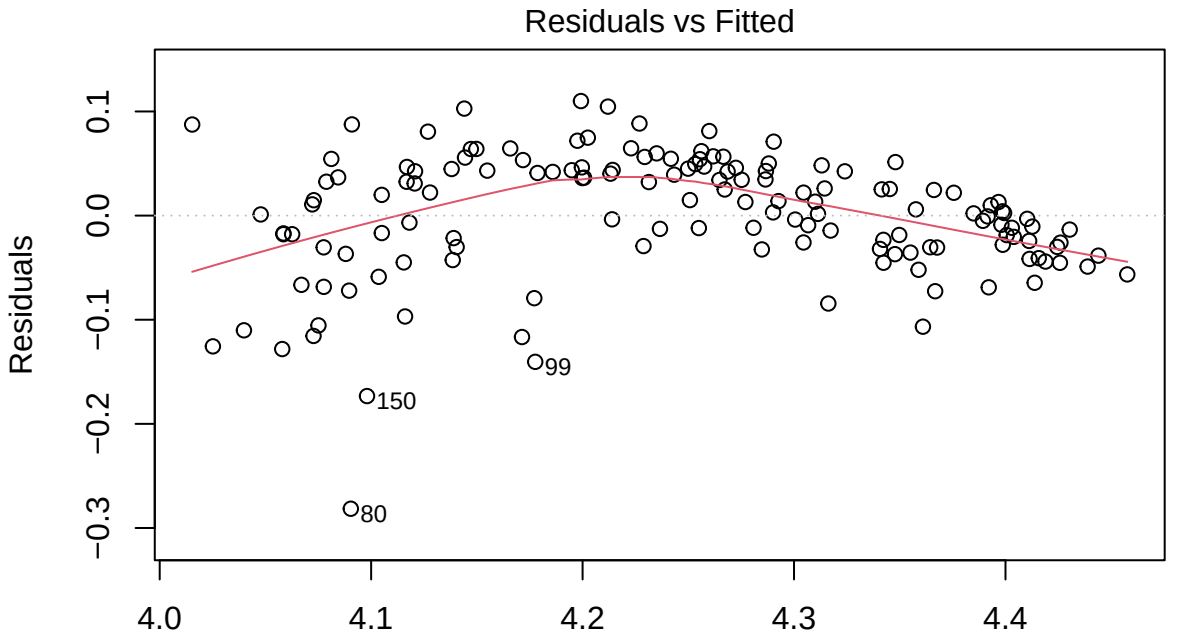
```
##
## Call:
```

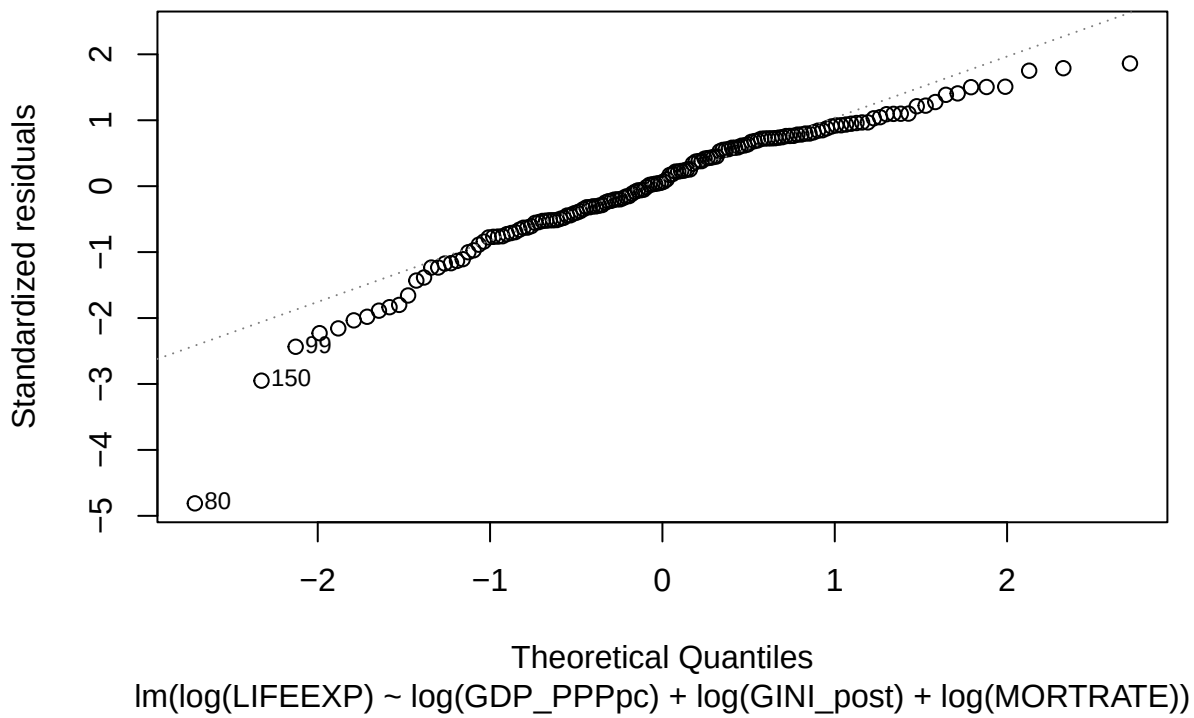
```
## lm(formula = log(LIFEEXP) ~ log(GDP_PPPpc) + log(GINI_post) +
##     log(MORTRATE), data = lifeexp_data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.281505 -0.030490  0.003568  0.043083  0.109980
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     4.330392   0.133503  32.437   <2e-16 ***
## log(GDP_PPPpc)  0.007357   0.009039   0.814    0.417
## log(GINI_post)  0.039869   0.030076   1.326    0.187
## log(MORTRATE)  -0.098729   0.010221  -9.659   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05932 on 146 degrees of freedom
## Multiple R-squared:  0.8032, Adjusted R-squared:  0.7991
## F-statistic: 198.6 on 3 and 146 DF,  p-value: < 2.2e-16
```

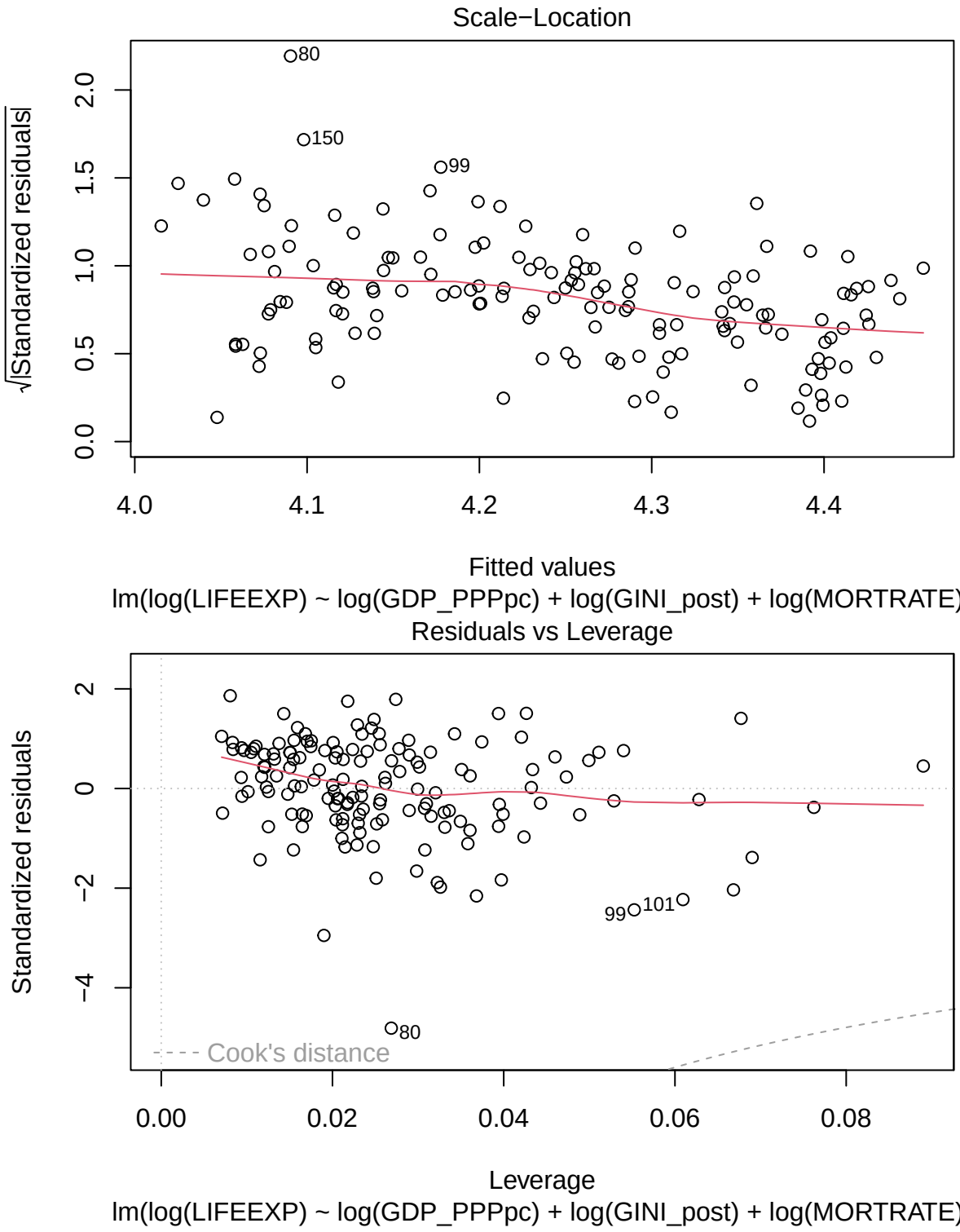
```
avPlots(model15)
```



```
plot(model15)
```





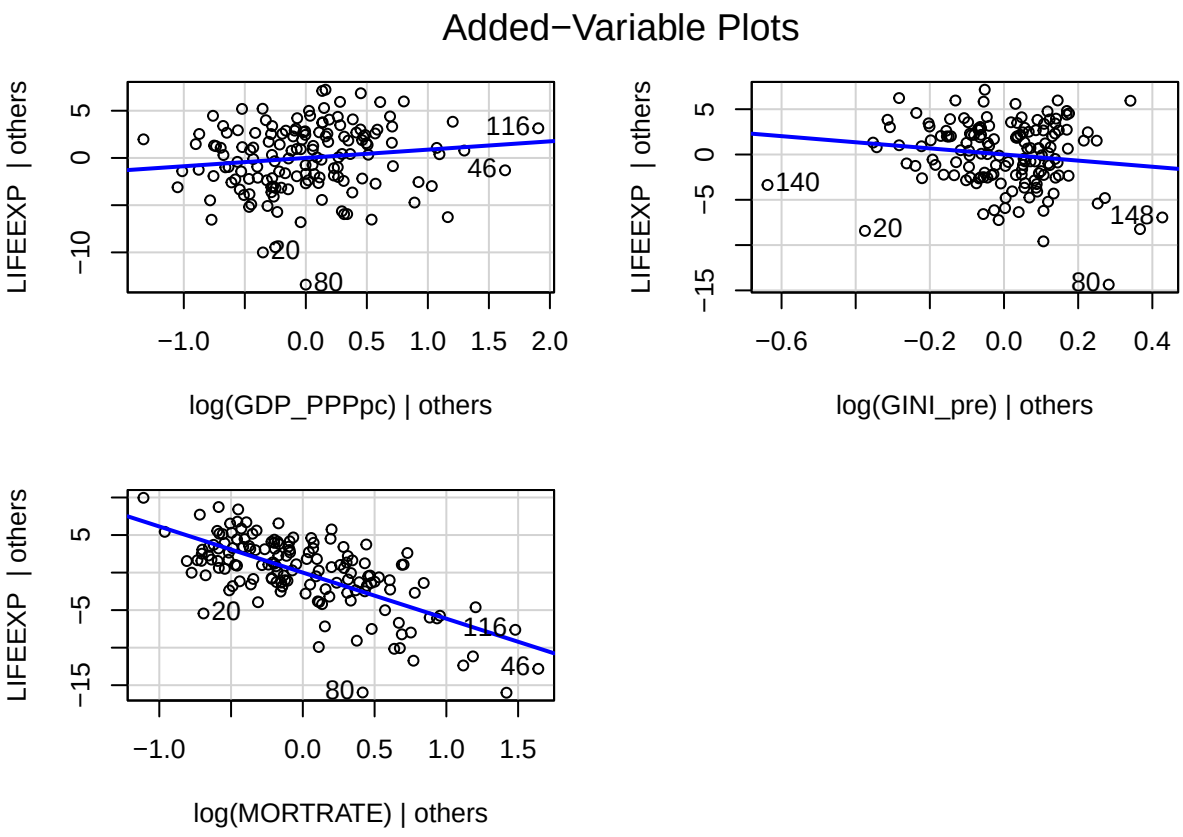



```
model16 <- lm(LIFEEXP ~ log(GDP_PPPpc) + log(GINI_pre) + log(MORTRATE), lifeexp_data)
summary(model16)
```

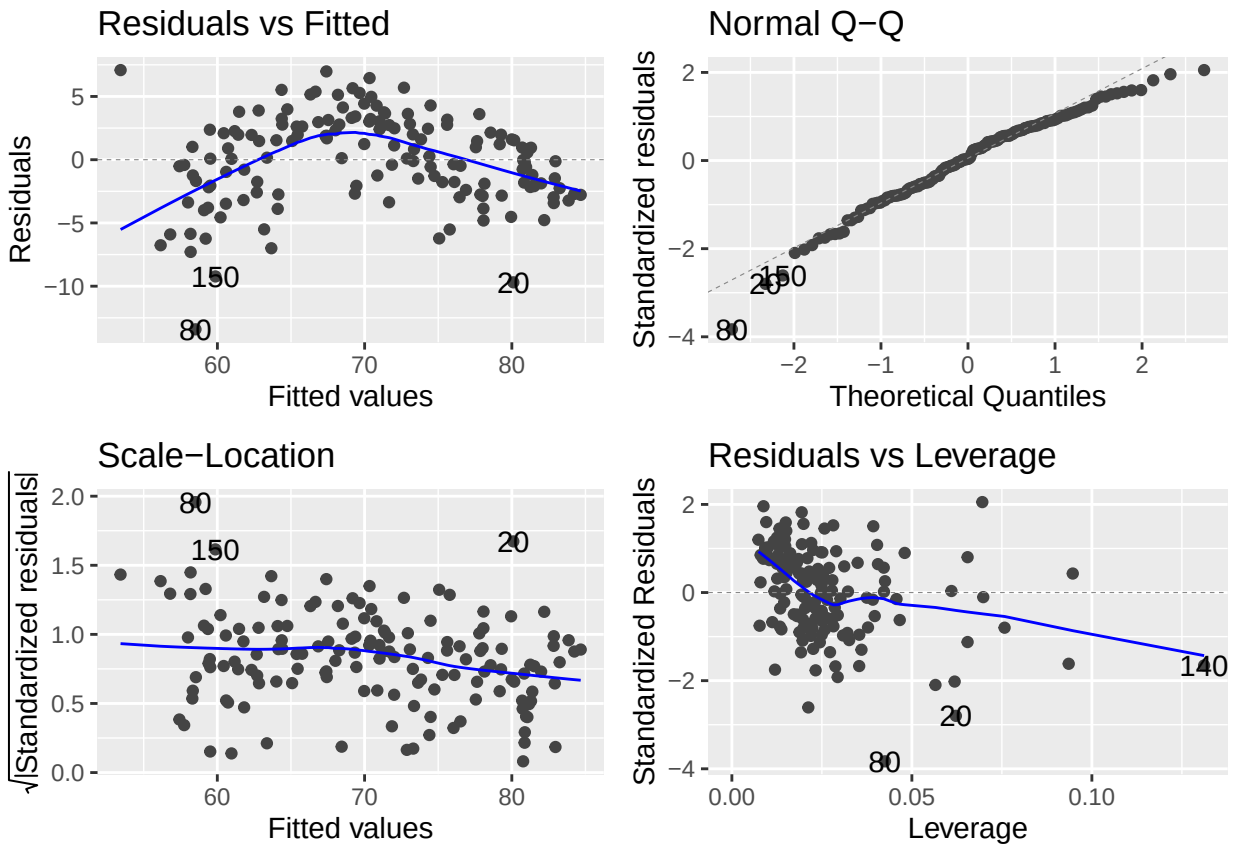
```
##
## Call:
```

```
## lm(formula = LIFEEXP ~ log(GDP_PPPpc) + log(GINI_pre) + log(MORTRATE),
##     data = lifeexp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.4107  -2.2339   0.1399   2.6224   7.0845
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      93.5269     8.9175  10.488   <2e-16 ***
## log(GDP_PPPpc)    0.8783     0.5445   1.613    0.109
## log(GINI_pre)    -3.3764     1.8956  -1.781    0.077 .
## log(MORTRATE)    -6.1423     0.5545 -11.078   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.578 on 146 degrees of freedom
## Multiple R-squared:  0.8371, Adjusted R-squared:  0.8337
## F-statistic:   250 on 3 and 146 DF,  p-value: < 2.2e-16
```

```
avPlots(model16)
```



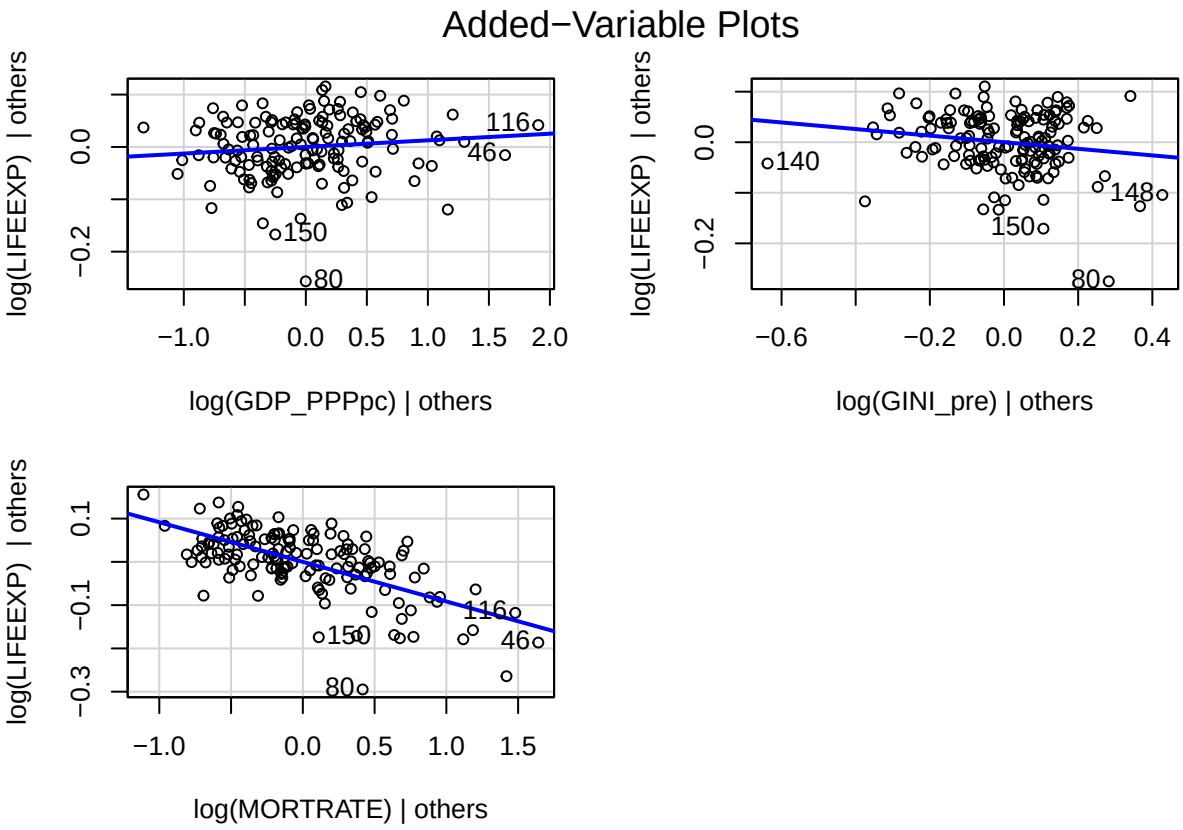
```
autoplot(model16)
```



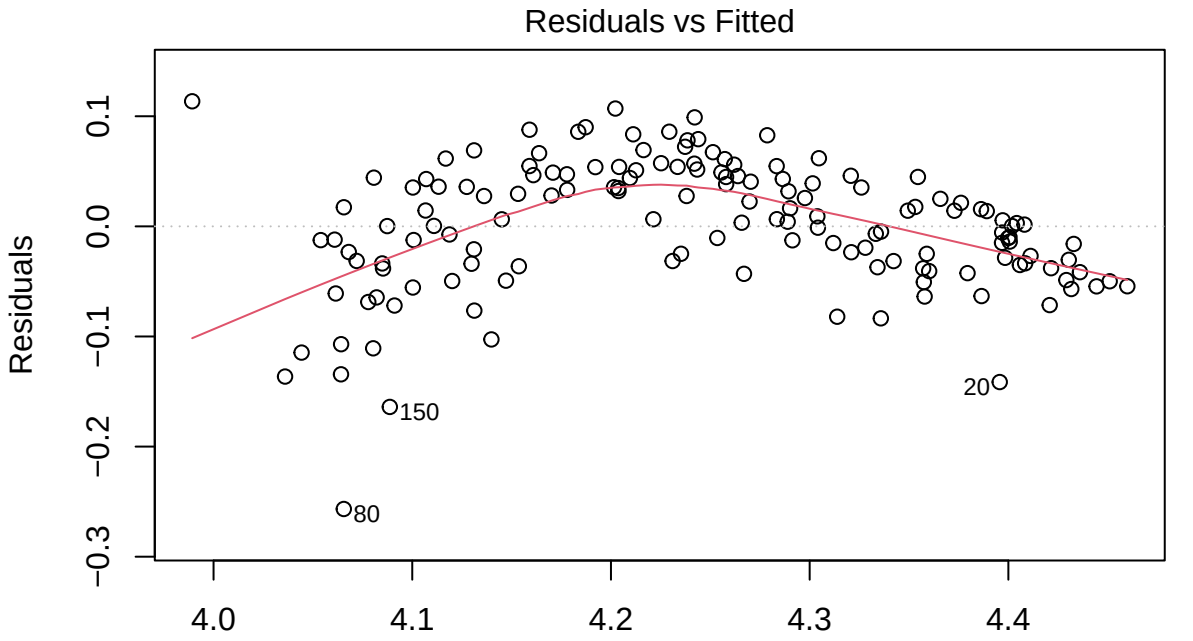

```
model17 <- lm(log(LIFEEXP) ~ log(GDP_PPPpc) + log(GINI_pre) + log(MORTRATE), lifeexp_data)
summary(model17)
```

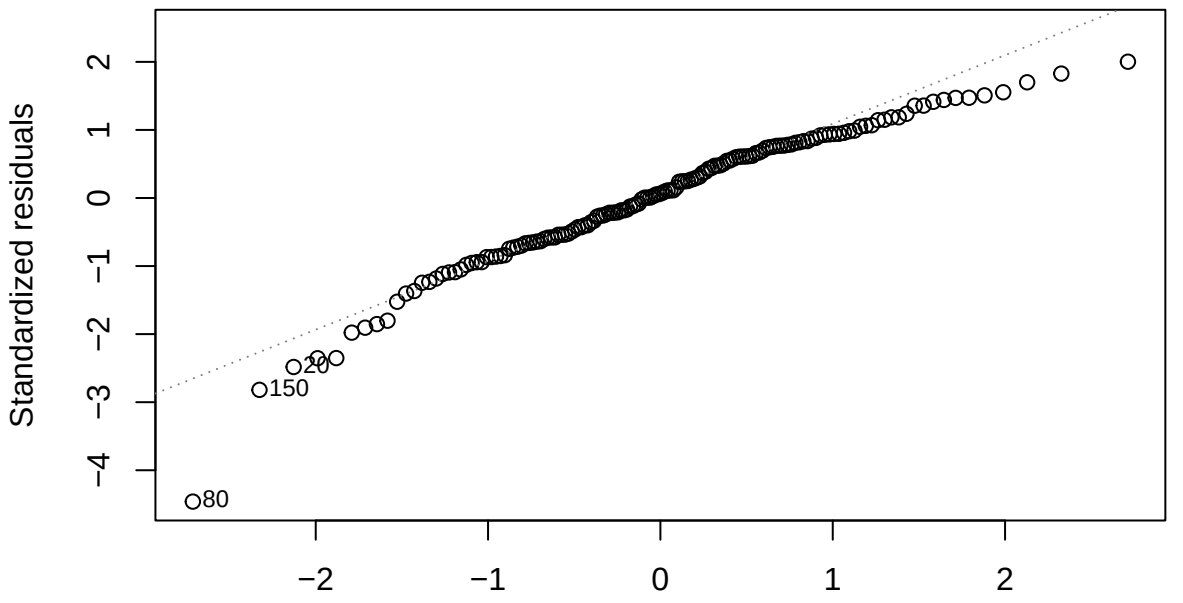
```
##
## Call:
## lm(formula = log(LIFEEXP) ~ log(GDP_PPPpc) + log(GINI_pre) +
##     log(MORTRATE), data = lifeexp_data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.256659 -0.034799  0.003779  0.044156  0.113580
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     4.650472   0.146561  31.731   <2e-16 ***
## log(GDP_PPPpc)  0.012645   0.008949   1.413   0.1598
## log(GINI_pre)  -0.065151   0.031155  -2.091   0.0382 *
## log(MORTRATE)  -0.091548   0.009113 -10.046   <2e-16 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 0.0588 on 146 degrees of freedom
## Multiple R-squared:  0.8066, Adjusted R-squared:  0.8026
## F-statistic: 202.9 on 3 and 146 DF,  p-value: < 2.2e-16
```

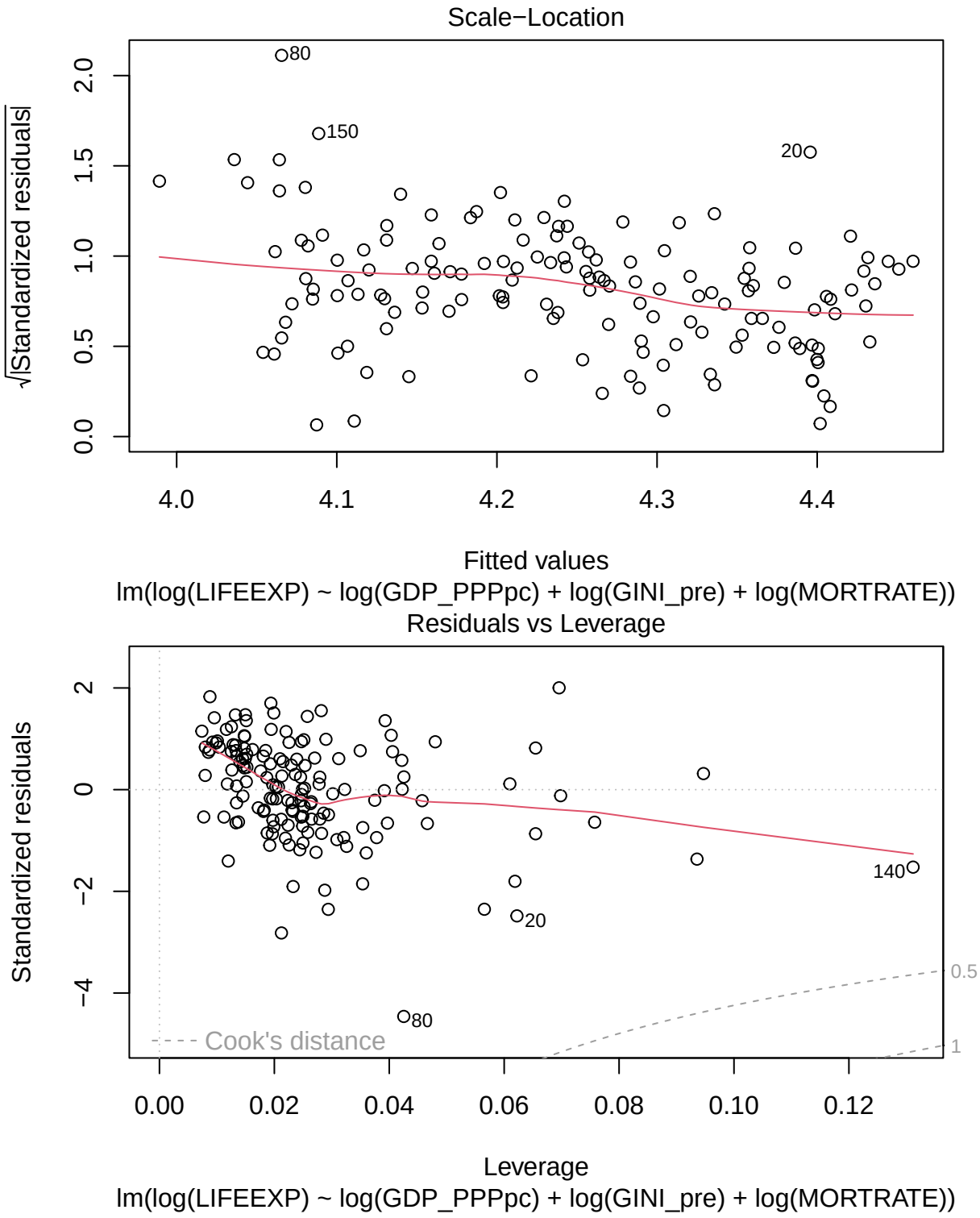
```
avPlots(model17)
```




```
plot(model17)
```


Residuals vs Fitted
Residuals
0.1
0.0
−0.1
−0.2
−0.3
150
80
20
4.0
4.1
4.2
4.3
4.4
Fitted values
lm(log(LIFEEXP) ~ log(GDP_PPPpc) + log(GINI_pre) + log(MORTRATE))


Normal Q−Q
Standardized residuals
2
1
0
−1
−2
−3
−4
150
20
80
−2
−1
0
1
2
Theoretical Quantiles
lm(log(LIFEEXP) ~ log(GDP_PPPpc) + log(GINI_pre) + log(MORTRATE))

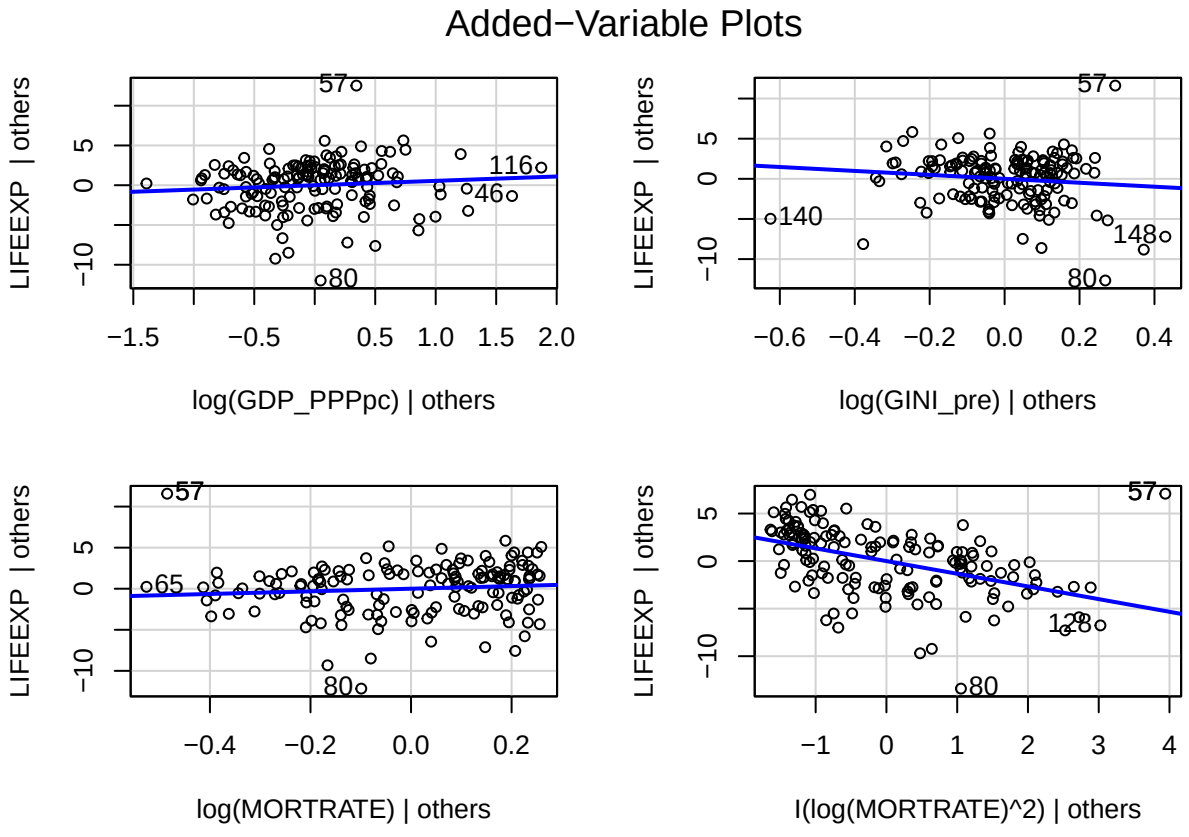



```
model18 <- lm(LIFEEXP ~ log(GDP_PPPpc) + log(GINI_pre) + log(MORTRATE) + I(log(MORTRATE)**2), lifeexp_da
summary(model18)
```

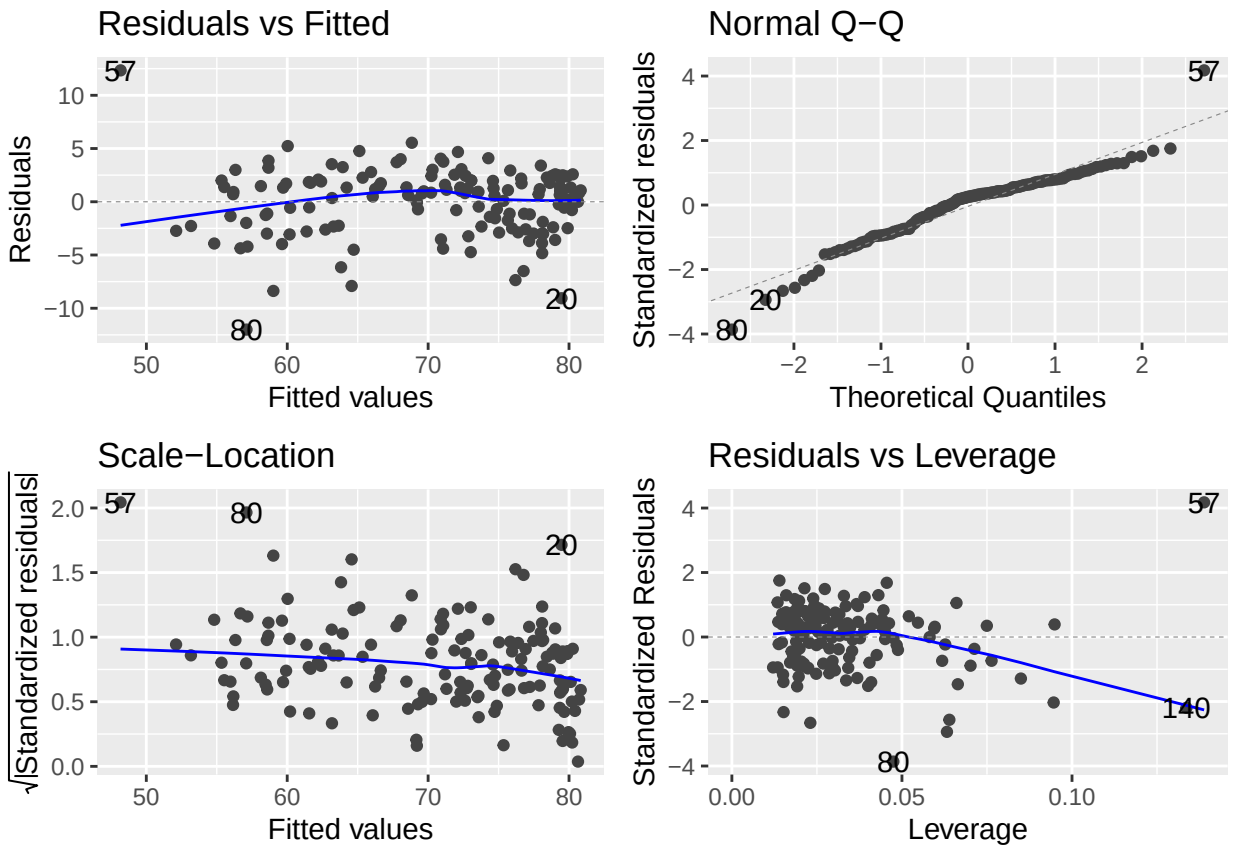
```
##
## Call:
```

```
## lm(formula = LIFEEXP ~ log(GDP_PPPpc) + log(GINI_pre) + log(MORTRATE) +
##     I(log(MORTRATE)^2), data = lifeexp_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.0064  -2.1861   0.7992   1.9784  12.3395
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          83.6320     8.0921  10.335  < 2e-16 ***
## log(GDP_PPPpc)        0.5493     0.4874   1.127    0.262
## log(GINI_pre)        -2.4538     1.6936  -1.449    0.150
## log(MORTRATE)         1.5955     1.3291   1.200    0.232
## I(log(MORTRATE)^2)   -1.3333     0.2127  -6.270 3.91e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.184 on 145 degrees of freedom
## Multiple R-squared:  0.8718, Adjusted R-squared:  0.8683
## F-statistic: 246.6 on 4 and 145 DF,  p-value: < 2.2e-16
```

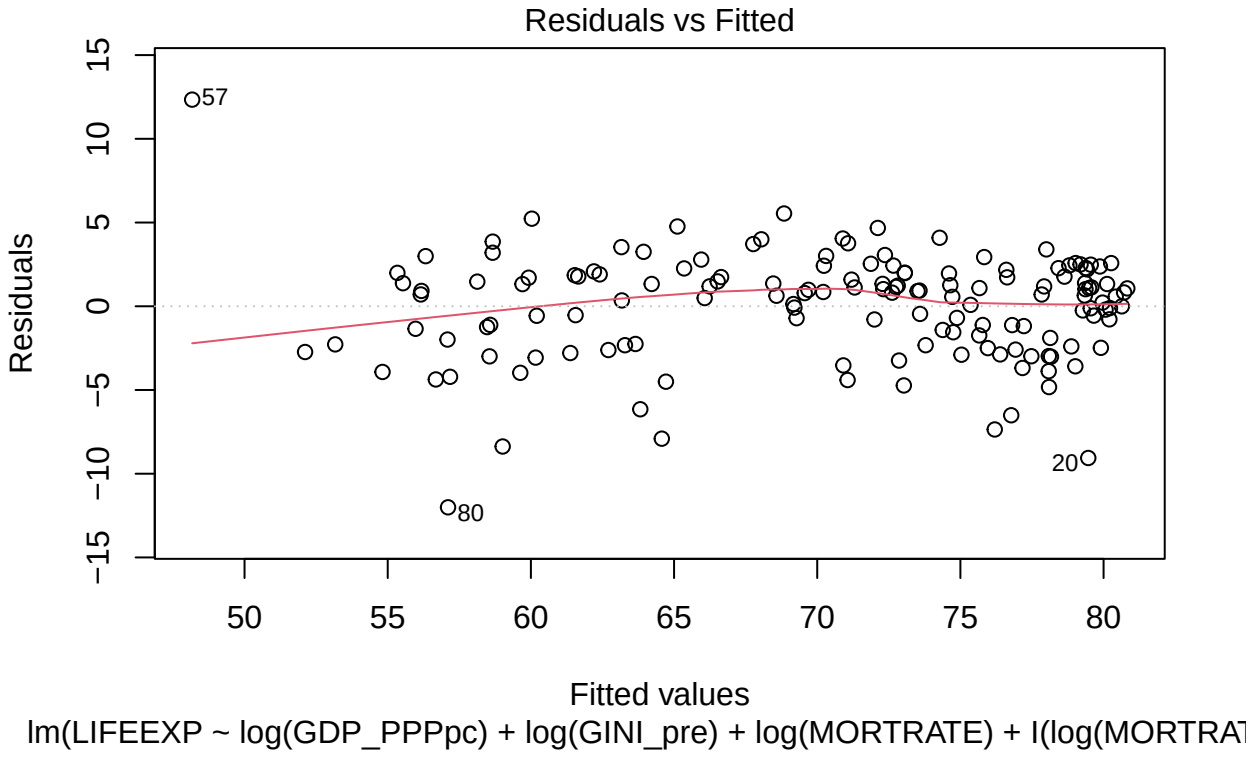
```
avPlots(model18)
```

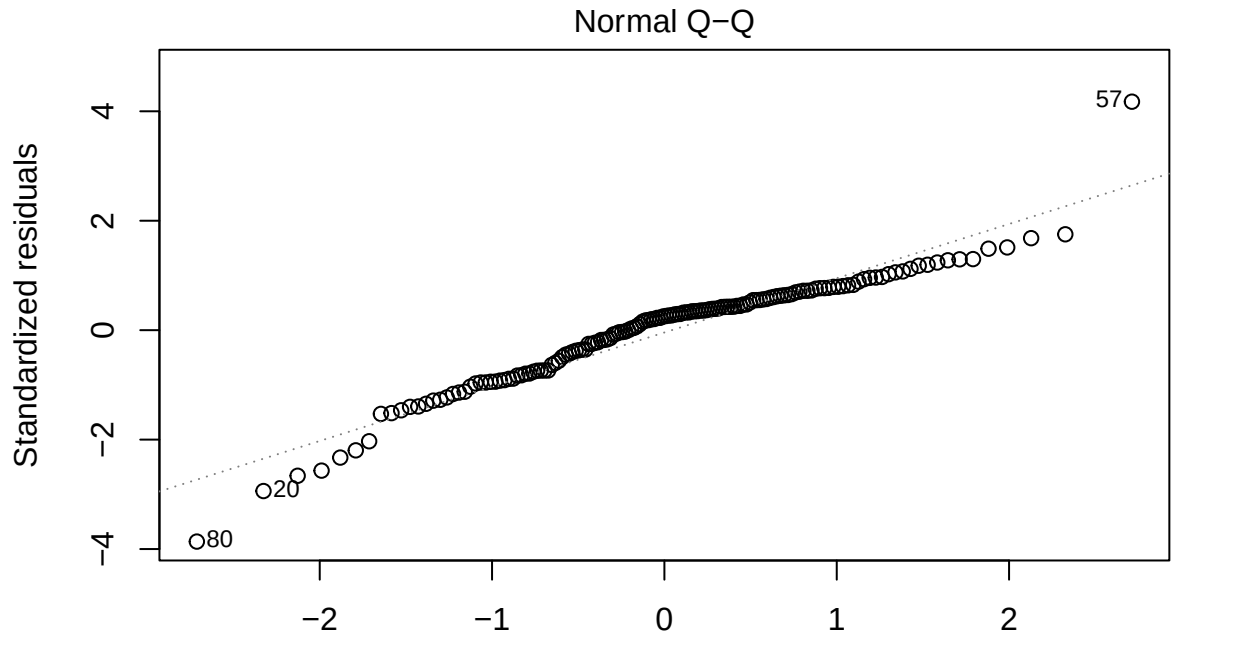


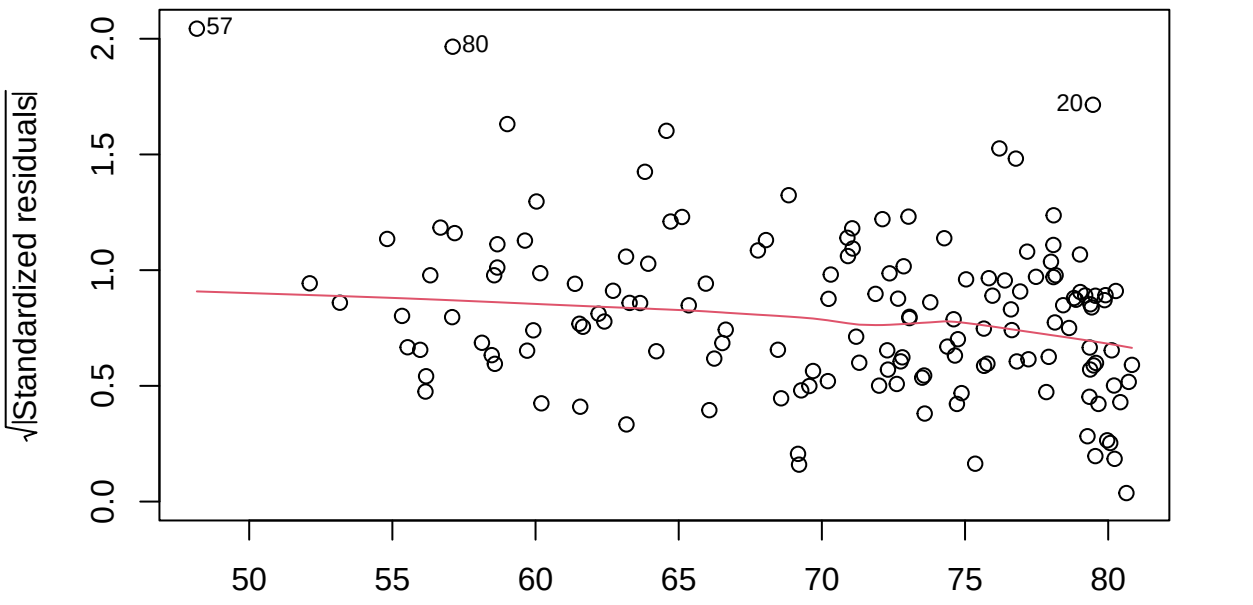

```
autoplot(model18)
```

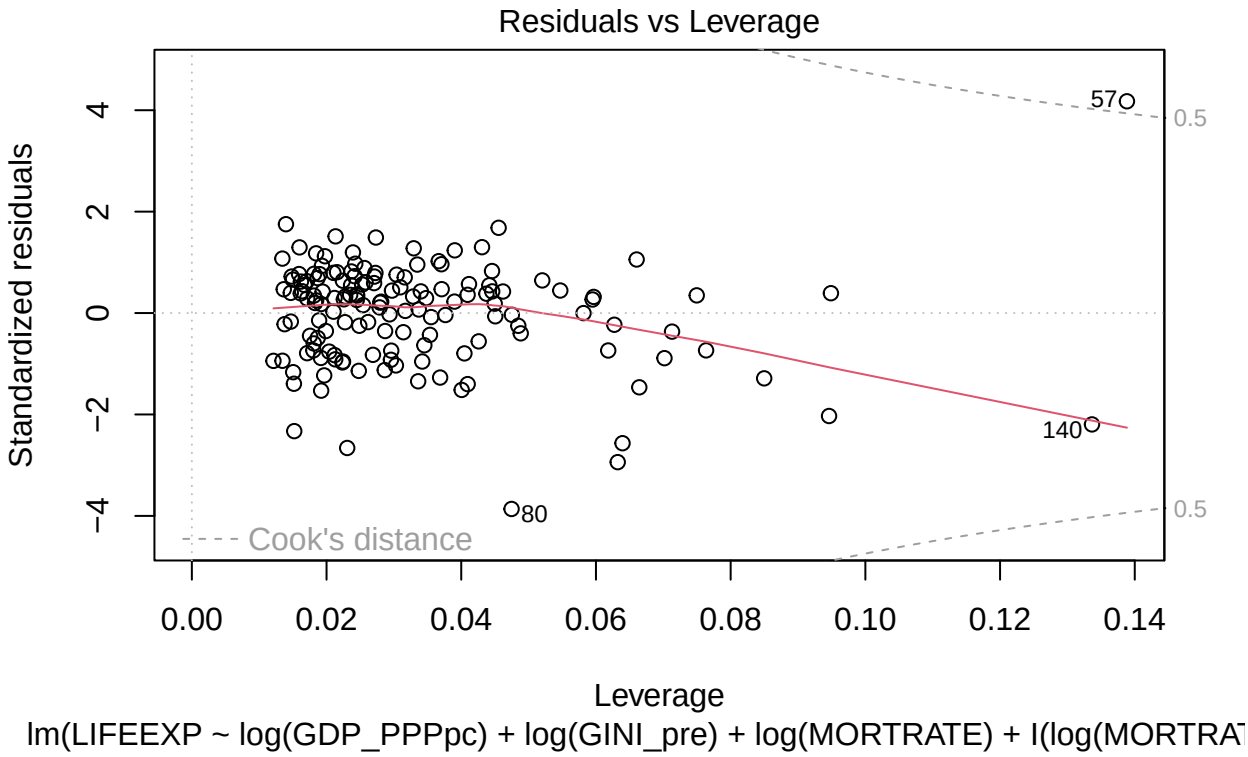



```
plot(model18)
```




Normal Q−Q
Standardized residuals
57
20
80
−4
−2
0
2
4
−2
−1
0
1
2
Theoretical Quantiles
lm(LIFEEXP ~ log(GDP_PPPpc) + log(GINI_pre) + log(MORTRATE) + I(log(MORTRA


Scale−Location
√|Standardized residuals|
57
80
20
0.0
0.5
1.0
1.5
2.0
50
55
60
65
70
75
80
Fitted values
lm(LIFEEXP ~ log(GDP_PPPpc) + log(GINI_pre) + log(MORTRATE) + I(log(MORTRA


Residuals vs Leverage
Standardized residuals
4
2
0
−2
−4
57
140
80
0.5
0.5
Cook's distance
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Leverage
lm(LIFEEXP ~ log(GDP_PPPpc) + log(GINI_pre) + log(MORTRATE) + I(log(MORTRA