第三次作业 R 语言部分参考解答

3.15

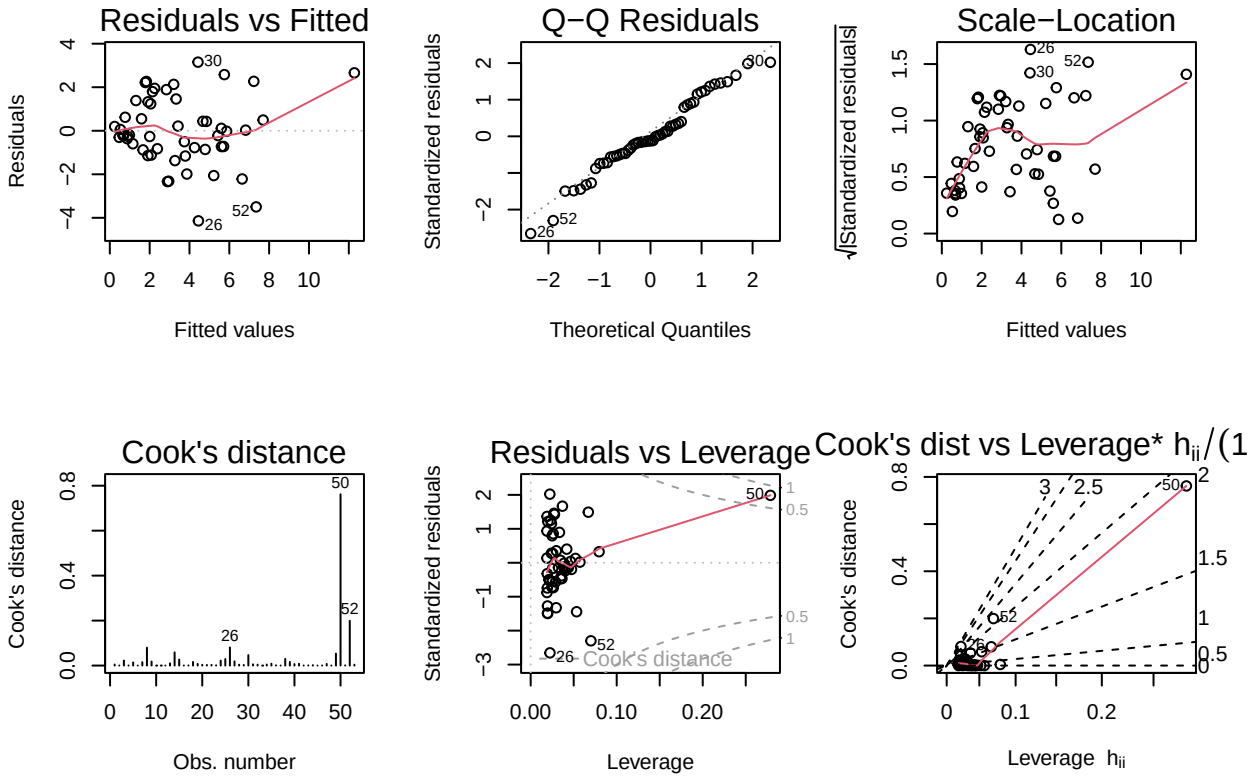
```
X <- c(679,292,1012,493,582,1156,997,2189,1097,2078,1818,1700,747,2030,1643,414,354,1276,745,
       435,540,874,1543,1029,710,1434,837,1748,1381,1428,1255,1777,370,2316,1130,463,770,724,
       808,790,783,406,1242,658,1746,468,1114,413,1787,3560,1495,2221,1526)
Y <- c(0.79,0.44,0.56,0.79,2.70,3.64,4.73,9.50,5.34,6.85,5.84,5.21,3.25,4.43,3.16,0.50,0.17,
       1.88,0.77,1.39,0.56,1.56,5.28,0.64,4.00,0.31,4.20,4.88,3.48,7.58,2.63,4.99,0.59,8.19,4.79,
       0.51,1.74,4.10,3.94,0.96,3.29,0.44,3.24,2.14,5.71,0.64,1.90,0.51,8.33,14.94,5.11,3.85,3.93)
```

```
fit_1 <- lm(Y ~ X)
coefficients(fit_1)
```

```
##  (Intercept)            X
## -0.831303660  0.003682843
```

(a) 拟合回归方程为 $Y = -0.8313 + 0.0037X$

```
par(mfrow=c(2,3))
plot(fit_1,which=1:6)
```



(c) Residuals vs Fitted 图无非线性趋势、误差方差稳定，则线性假设成立；Normal Q-Q 图中的点基本落在斜率为 1 的直线上，则正态假设成立；Scale-Location 图中的点不是一个随机分布，则 G-M 假设中的等方差性不成立。

(d) 由 Residuals vs Leverage 图，50 是高杠杆点；由 Cook's distance 图，26、50、52 是强影响点。

注：根据图像判断各种假设是否成立、进行回归诊断是很重要的，可能会考。

```
library(car)
```

```
## Loading required package: carData
```

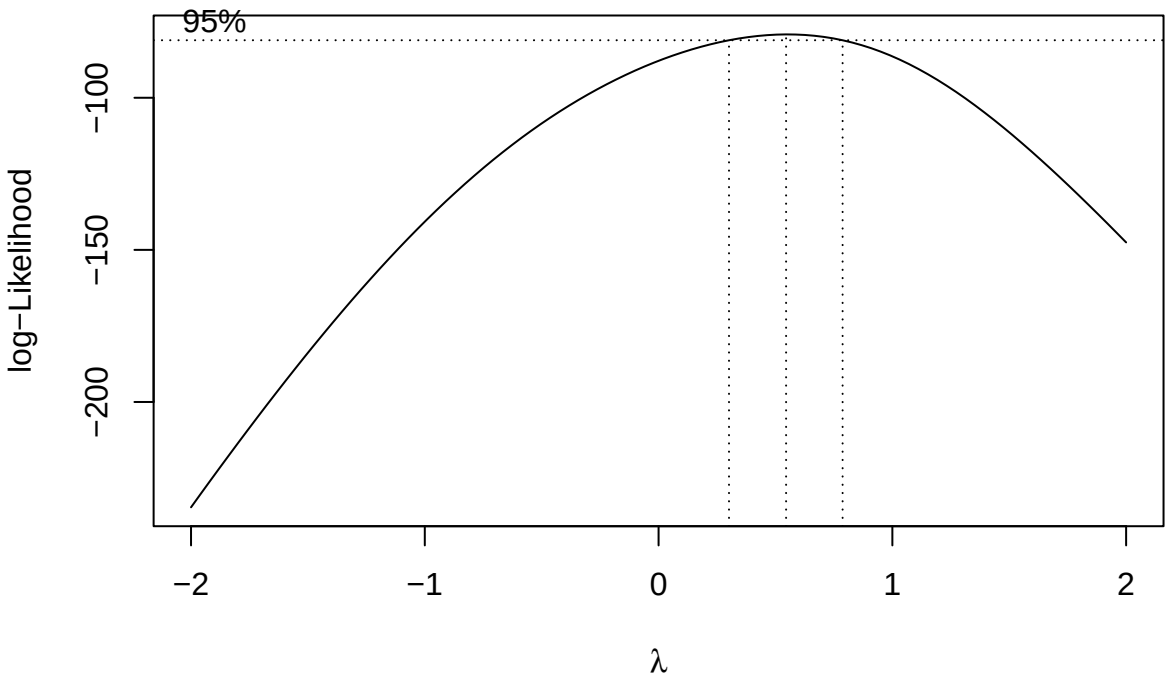
```
outlierTest(fit_1)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##    rstudent unadjusted p-value Bonferroni p
## 26 -2.83152          0.0066579      0.35287
```

(e) 无异常值案例。

(f) Box-Cox 变换

```
library(MASS)
bc1 <- boxcox(fit_1)
```



若想得到 Box-Cox 变换参数的精确取值，可用下面的程序。

```
lambda1 <- bc1$x[which(bc1$y==max(bc1$y))]
lambda1
```

```
## [1] 0.5454545
```

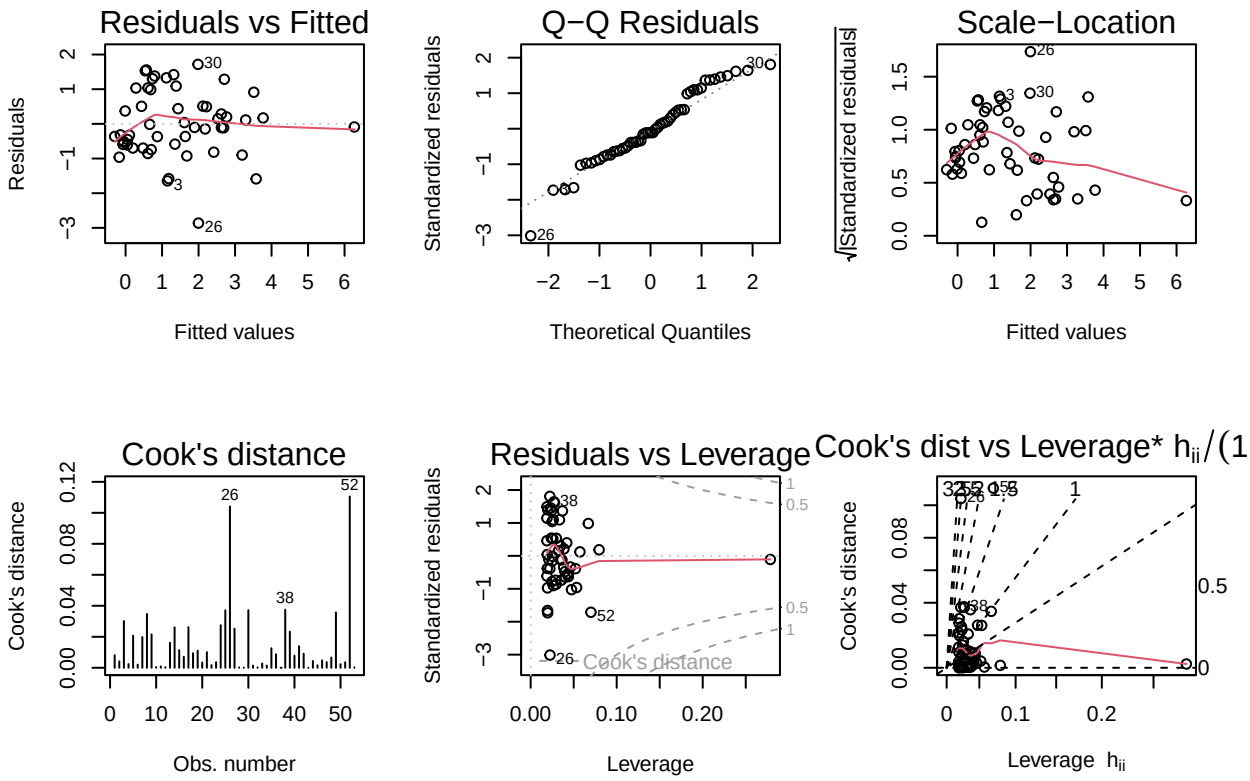
Box-Cox 变换参数 $\lambda = 0.5454545$。

```
fit1_bc <- lm((Y^lambda1-1)/lambda1 ~ X)
coefficients(fit1_bc)
```

```
##  (Intercept)            X
## -0.885541528  0.002009792
```

拟合回归方程为 $Y = -0.8855 + 0.0020X$

```
par(mfrow=c(2,3))
plot(fit1_bc,which=1:6)
```



线性假设成立；正态假设成立；Gauss-Markov 假设的等方差性成立。

3.16

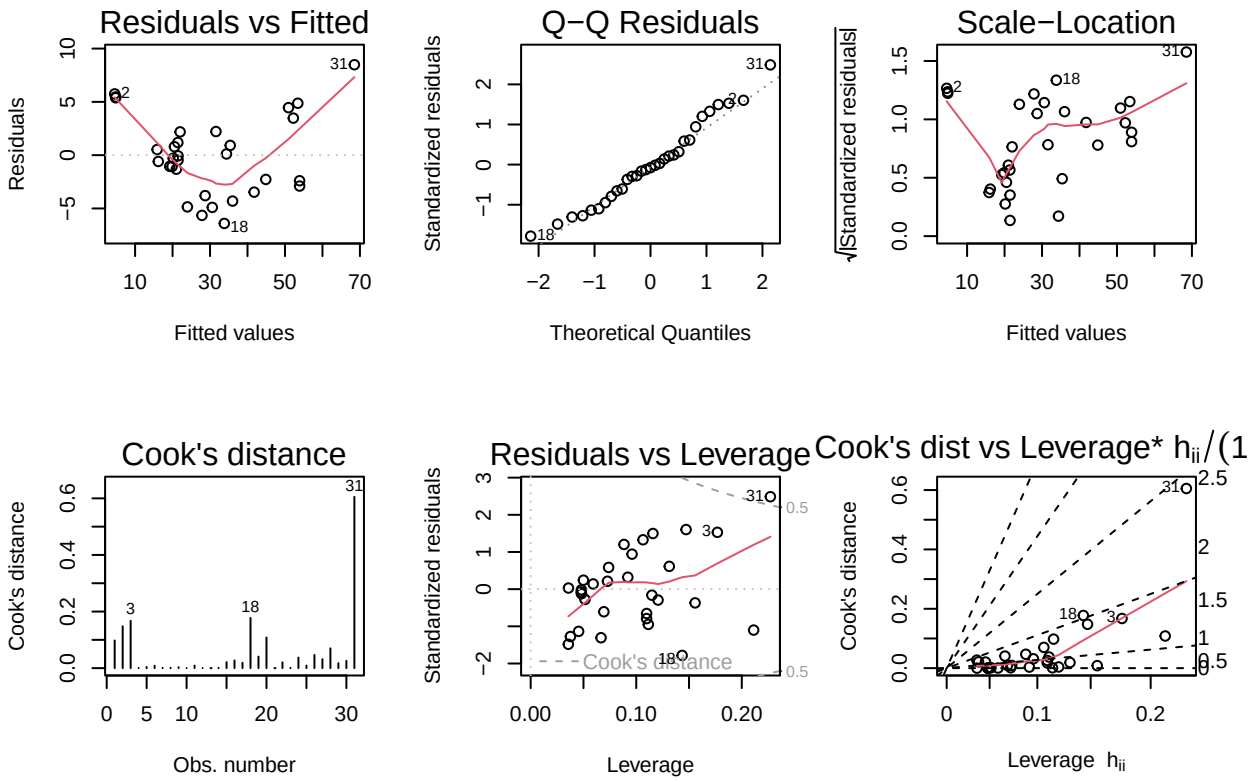
```
x1 <- c(8.3,8.6,8.8,10.5,10.7,10.8,11.0,11.0,11.1,11.2,11.3,11.4,11.4,11.7,12.0,12.9,12.9,
        13.3,13.7,13.8,14.0,14.2,14.5,16.0,16.3,17.3,17.5,17.9,18.0,18.0,20.6)
x2 <- c(70,65,63,72,81,83,66,75,80,75,79,76,76,69,75,74,85,86,71,64,78,80,74,72,77,81,82,
        80,80,80,87)
y <- c(10.3,10.3,10.2,16.4,18.8,19.7,15.6,18.2,22.6,19.9,24.2,21.0,21.4,21.3,19.1,22.2,
       33.8,27.4,25.7,24.9,34.5,31.7,36.3,38.3,42.6,55.4,55.7,58.3,51.5,51.0,77.0)
```

```
fit2 <- lm(y ~ x1 + x2)
coefficients(fit2)
```

```
## (Intercept)          x1          x2
## -57.9876589   4.7081605   0.3392512
```

(a) 拟合回归方程为 $Y = -57.99 + 4.70X_1 + 0.34X_2$

```
par(mfrow=c(2,3))
plot(fit2,which=1:6)
```



(c) Residuals vs Fitted 图中有一个清楚的曲线关系，则线性假设不成立；Normal Q-Q 图中的点基本落在斜率为 1 的直线上，则正态假设成立；Scale-Location 图中的点不是一个随机分布，则 G-M 假设中的等方差性不成立。

(d) 由 Residuals vs Leverage 图，31 是高杠杆点；由 Cook's distance 图，3、18、31 是强影响点。

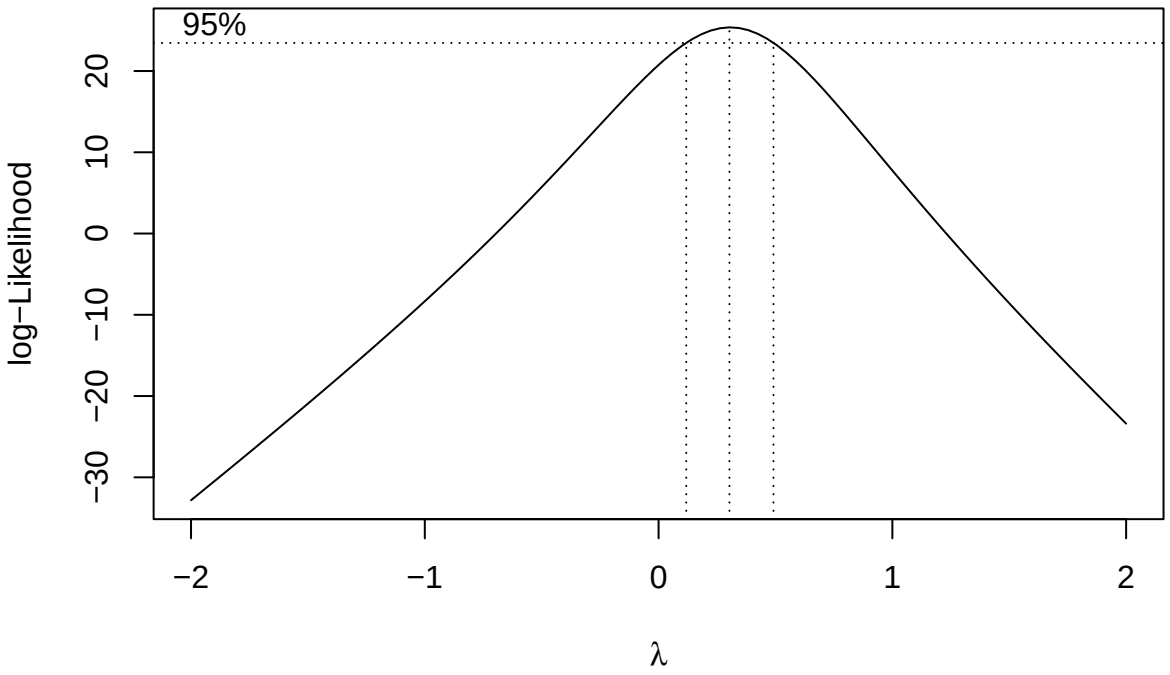
```
library(car)
outlierTest(fit2)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##    rstudent unadjusted p-value Bonferroni p
## 31 2.765603           0.010122      0.31377
```

(e) 无异常值案例。

(f) Box-Cox 变换

```
library(MASS)
bc2 <- boxcox(fit2)
```




```
lambda2 <- bc2$x[which(bc2$y==max(bc2$y))]
lambda2
```

```
## [1] 0.3030303
```

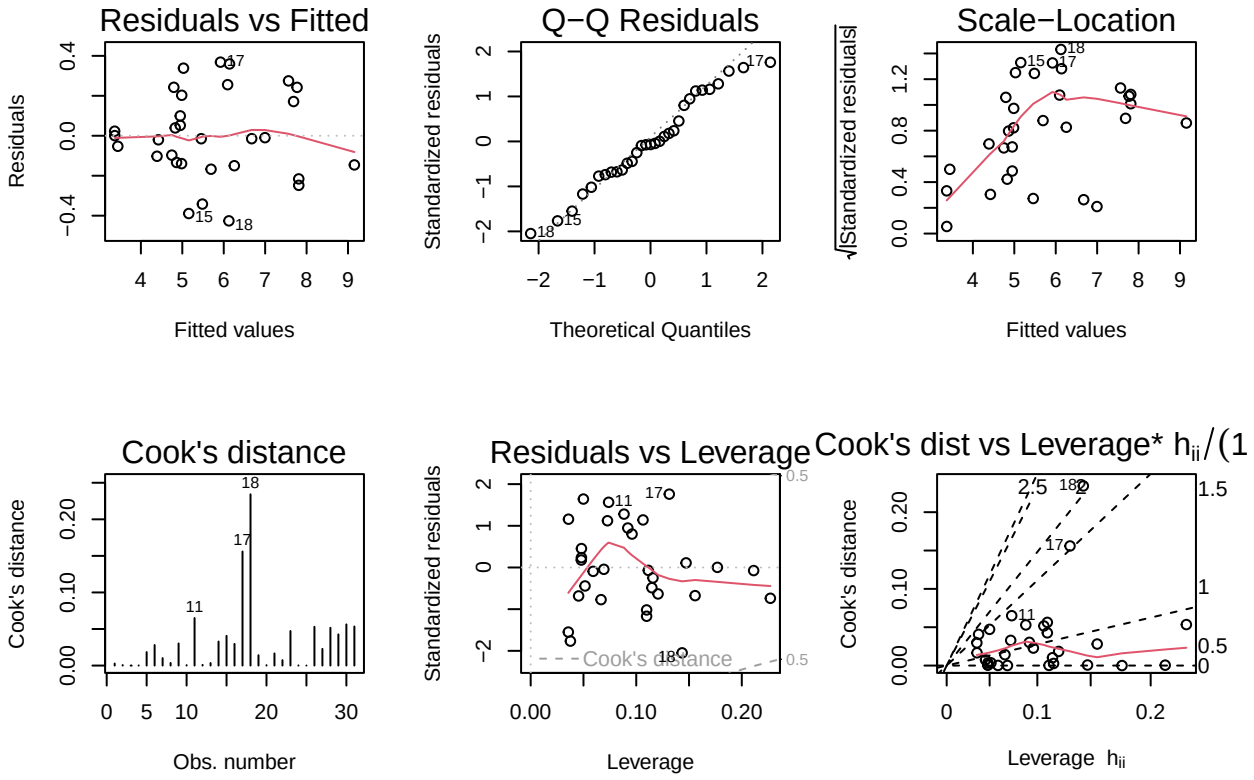
Box-Cox 变换参数 $\lambda = 0.3030303$。

```
fit2_bc <- lm((y^lambda2-1)/lambda2 ~ x1 + x2)
coefficients(fit2_bc)
```

```
## (Intercept)          x1          x2
## -2.73354218  0.40944844  0.03968502
```

拟合回归方程为 $Y = -2.64 + 0.41X_1 + 0.04X_2$

```
par(mfrow=c(2,3))
plot(fit2_bc,which=1:6)
```



线性假设成立；正态假设成立；Gauss-Markov 假设的等方差性不成立。

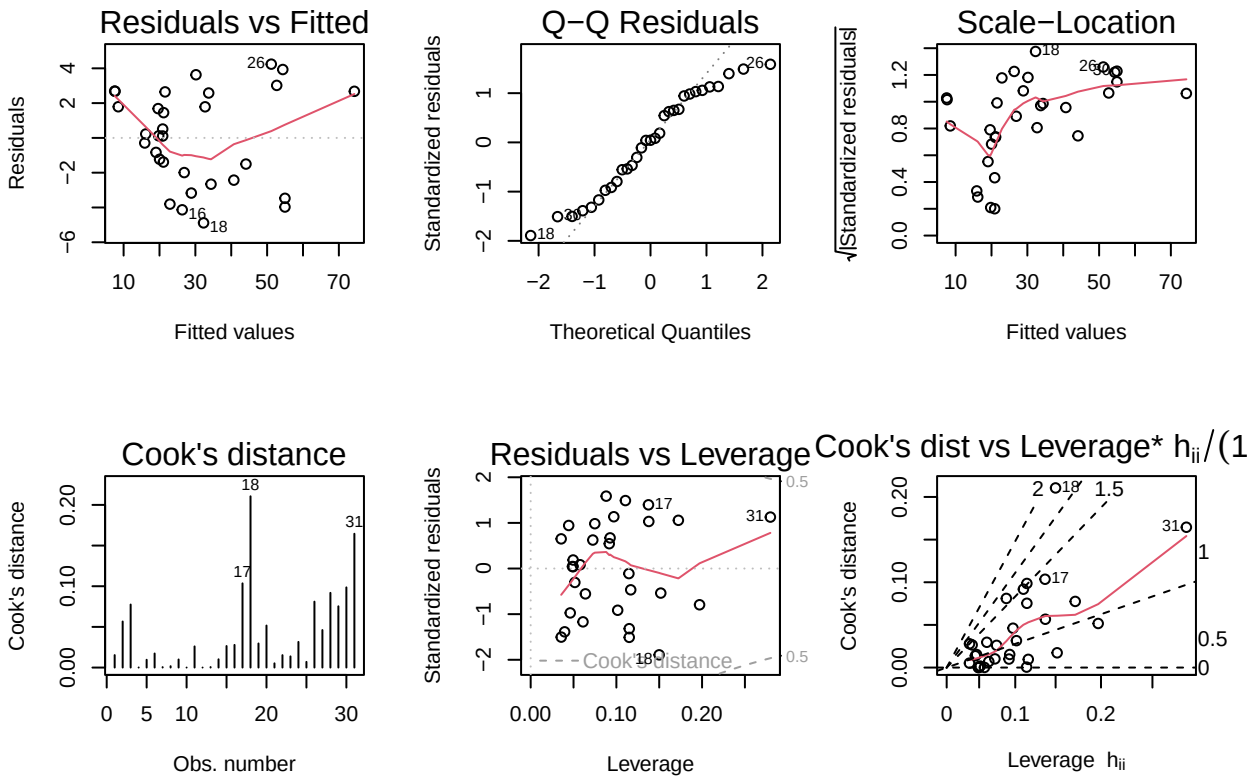
3.17

```
fit3 <- lm(y ~ I(x1^2)+x2)
coefficients(fit3)
```

```
## (Intercept)     I(x1^2)          x2
## -27.5116027   0.1684577   0.3488088
```

(a) 拟合回归方程为 $Y = -27.51 + 0.1684X_1^2 + 0.35X_2$

```
par(mfrow=c(2,3))
plot(fit3,which=1:6)
```



(c) Residuals vs Fitted 图中有一个清楚的曲线关系，则线性假设不成立；Normal Q-Q 图中的点与斜率为 1 的直线偏差较多，则正态假设不成立；Scale-Location 图中的点不是一个随机分布，则 G-M 假设中的等方差性不成立。

(d) 由 Residuals vs Leverage 图，31 是高杠杆点；由 Cook's distance 图，17、18、31 是强影响点。

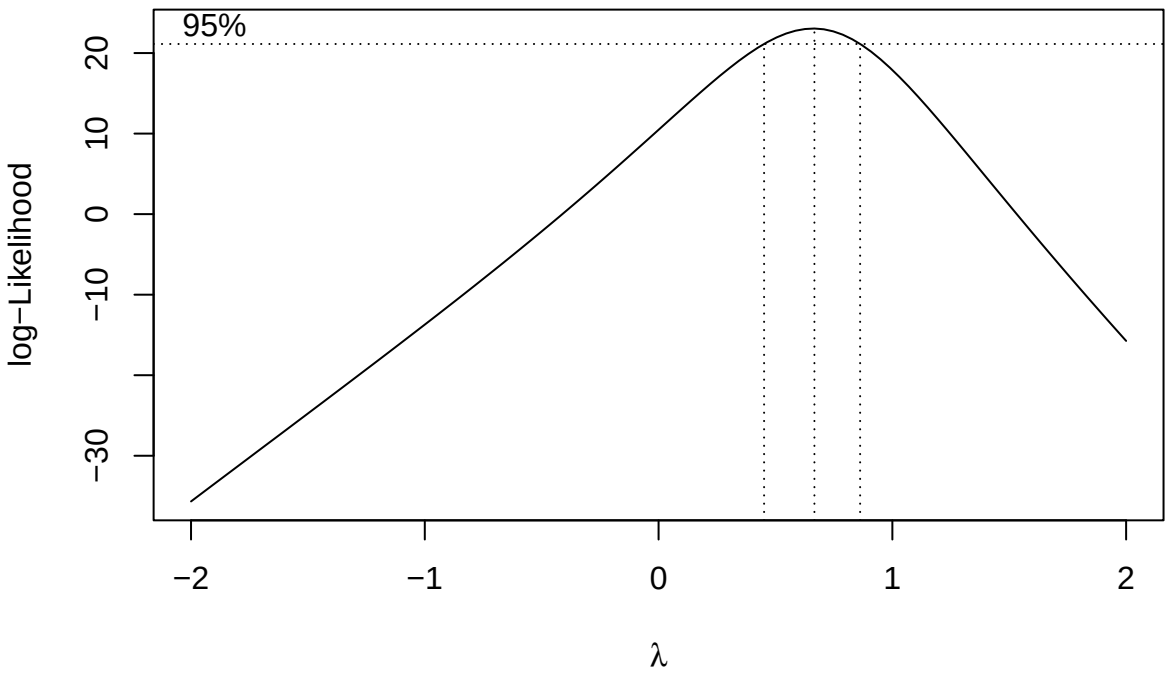
```
library(car)
outlierTest(fit3)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##     rstudent unadjusted p-value Bonferroni p
## 18 -1.989932           0.056813           NA
```

(e) 无异常值案例。

(f) Box-Cox 变换

```
library(MASS)
bc3 <- boxcox(fit3)
```




```
lambda3 <- bc3$x[which(bc3$y==max(bc3$y))]
lambda3
```

```
## [1] 0.6666667
```

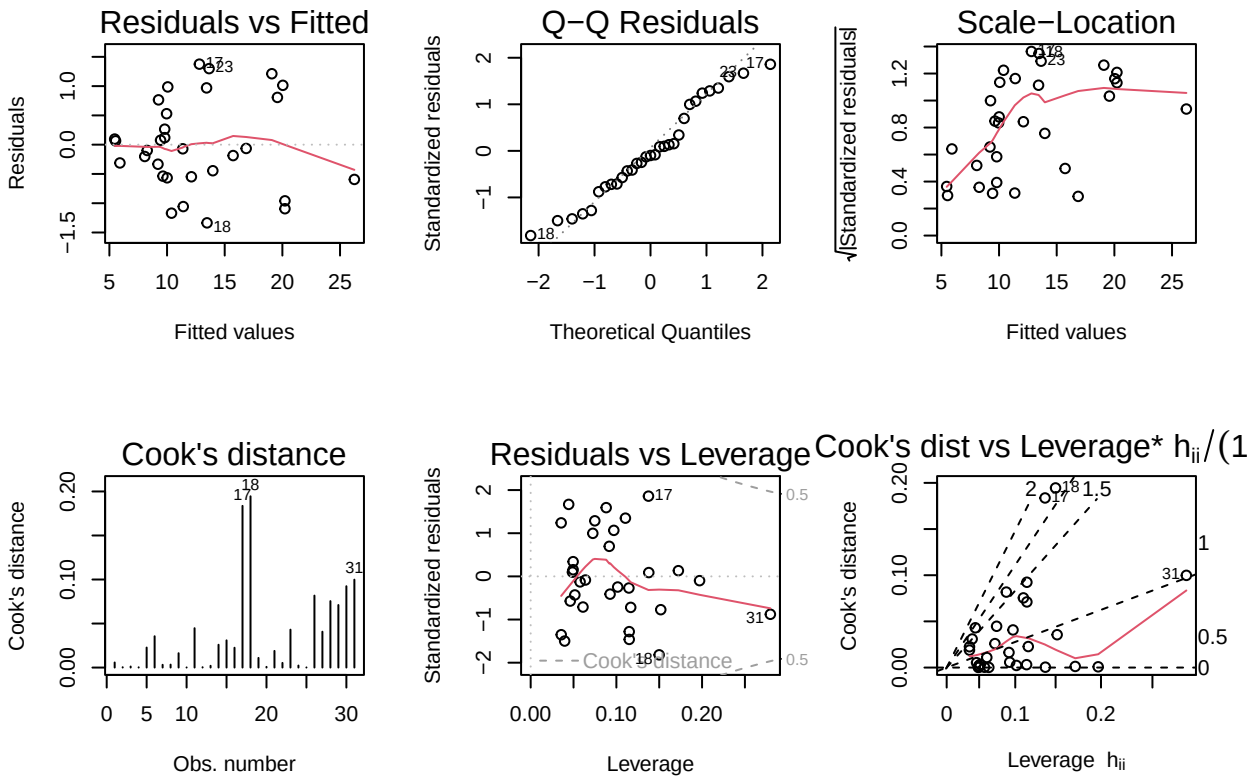
Box-Cox 变换参数 $\lambda = 0.6666667$。

```
fit3_bc <- lm((y^lambda3-1)/lambda3 ~ I(x1^2) + x2)
coefficients(fit3_bc)
```

```
## (Intercept)     I(x1^2)          x2
## -6.51548661  0.05110411  0.12725747
```

拟合回归方程为 $Y = -6.52 + 0.05X_1^2 + 0.13X_2$

```
par(mfrow=c(2,3))
plot(fit3_bc,which=1:6)
```



线性假设成立；正态假设成立；Gauss-Markov 假设的等方差性不成立。