Multivariate Analysis for the Behavioral Sciences, Second Edition (Chapman and Hall/CRC, 2019)

Examples of Chapter 3: Simple Linear and Locally Weighted Regression

Kimmo Vehkalahti and Brian S. Everitt

1 October 2018

Contents

Examples

Table 3.1: Pulse Rates and Heights Data

```
heights <- c(160,167,162,175,185,162,173,167,170,170,163,158,157,
             160,170,177,166,170,148,175,160,153,185,165,165,172,
             185,163,177,165,182,162,172,177,168,178,182,167,170,
             160,182,168,155,175,168,180,175,145,170,175)
#
pulse <- c(68,80,84,80,80,80,92,92,80,80,80,80,80,
           78,90,80,72,80,82,76,84,70,80,82,84,116,
           80,95,80,76,100,88,90,90,90,80,76,80,84,
           80,80,80,80,104,80,68,84,64,84,72)

hp <- cbind(heights, pulse)
head(hp)
```

```
##      heights pulse
## [1,]     160    68
## [2,]     167    80
## [3,]     162    84
## [4,]     175    80
## [5,]     185    80
## [6,]     162    80
```

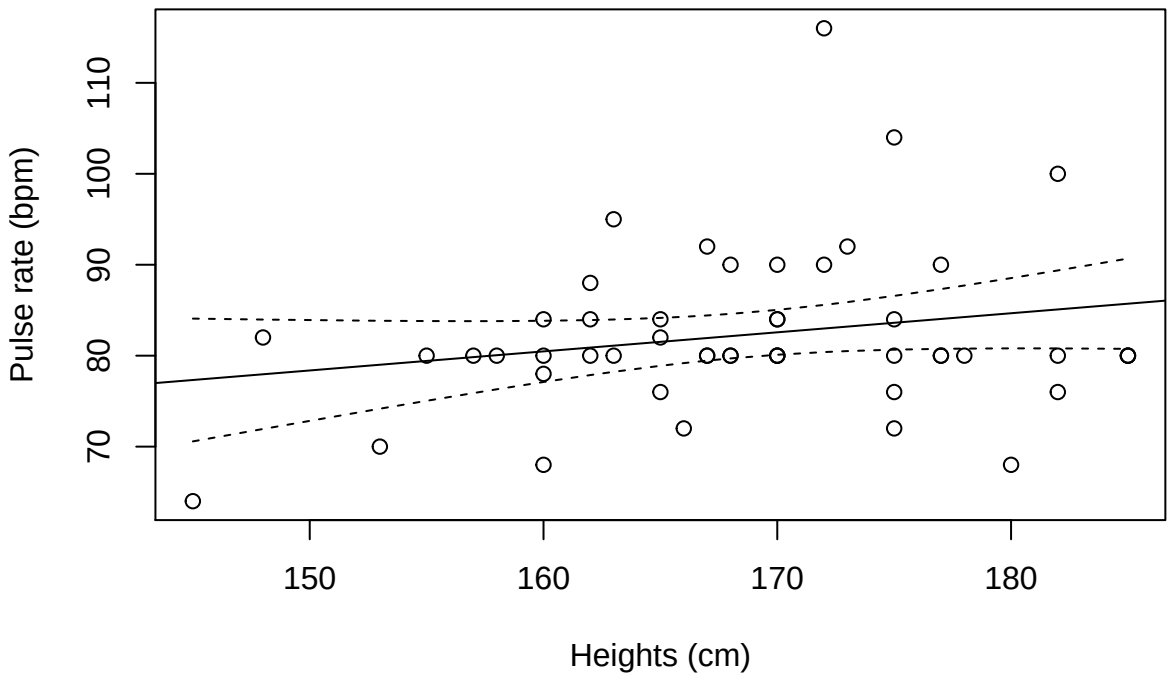
Table 3.2

```
summary(lm(pulse ~ heights))
```

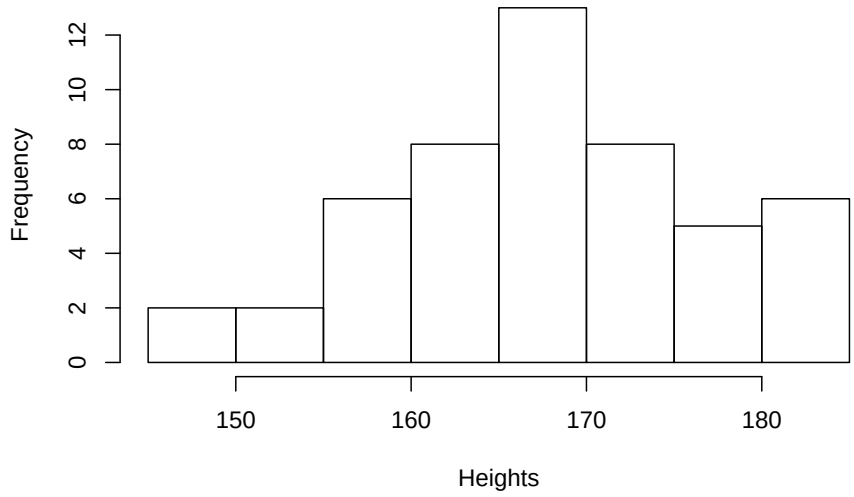
```
##
## Call:
## lm(formula = pulse ~ heights)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.666  -4.876  -1.520   3.424  33.012
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  46.9069    22.8793   2.050   0.0458 *
## heights       0.2098     0.1354   1.549   0.1279
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.811 on 48 degrees of freedom
## Multiple R-squared:  0.04762,    Adjusted R-squared:  0.02778
## F-statistic:   2.4 on 1 and 48 DF,  p-value: 0.1279
```

Figure 3.1

```
# layout(matrix(c(2, 0, 1, 3), 2, 2, byrow=TRUE), c(2,1), c(1,2), TRUE)
plot(pulse ~ heights, xlab = "Heights (cm)", ylab = "Pulse rate (bpm)")
reg <- lm(pulse ~ heights)
abline(reg)
pred <- predict(reg, se.fit = TRUE)
fitval <- pred$fit
se <- pred$se.fit
index <- order(heights)
y <- fitval[index]
se <- se[index]
yu <- y + 1.96*se
yl <- y - 1.96*se
lines(heights[index], yu, lty=2)
lines(heights[index], yl, lty=2)
```



```r
hist(heights, ylab = "Frequency", xlab = "Heights", main = "")
```




```r
boxplot(pulse, ylab = "Heights")
```

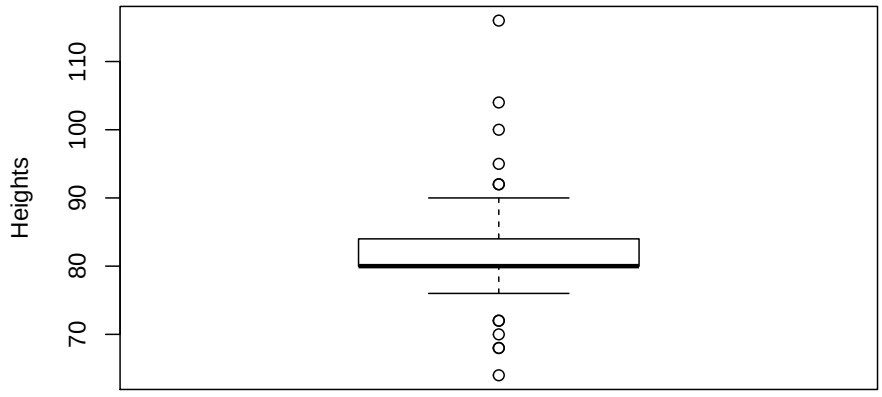


Figure 3.5

```
reg <- lm(pulse ~ heights)
residuals <- reg$residuals
fitval <- predict(reg)
par(mfrow = c(2,2))
boxplot(residuals, ylab = "Residuals")
title(sub="(a)")
qqnorm(residuals, main="")
title(sub="(b)")
plot(residuals ~ heights, xlab = "Heights (cm)", ylab = "Residuals")
abline(h=0)
title(sub="(c)")
plot(residuals ~ fitval, xlab = "Fitted pulse rates", ylab = "Residuals")
abline(h=0)
title(sub="(d)")
```

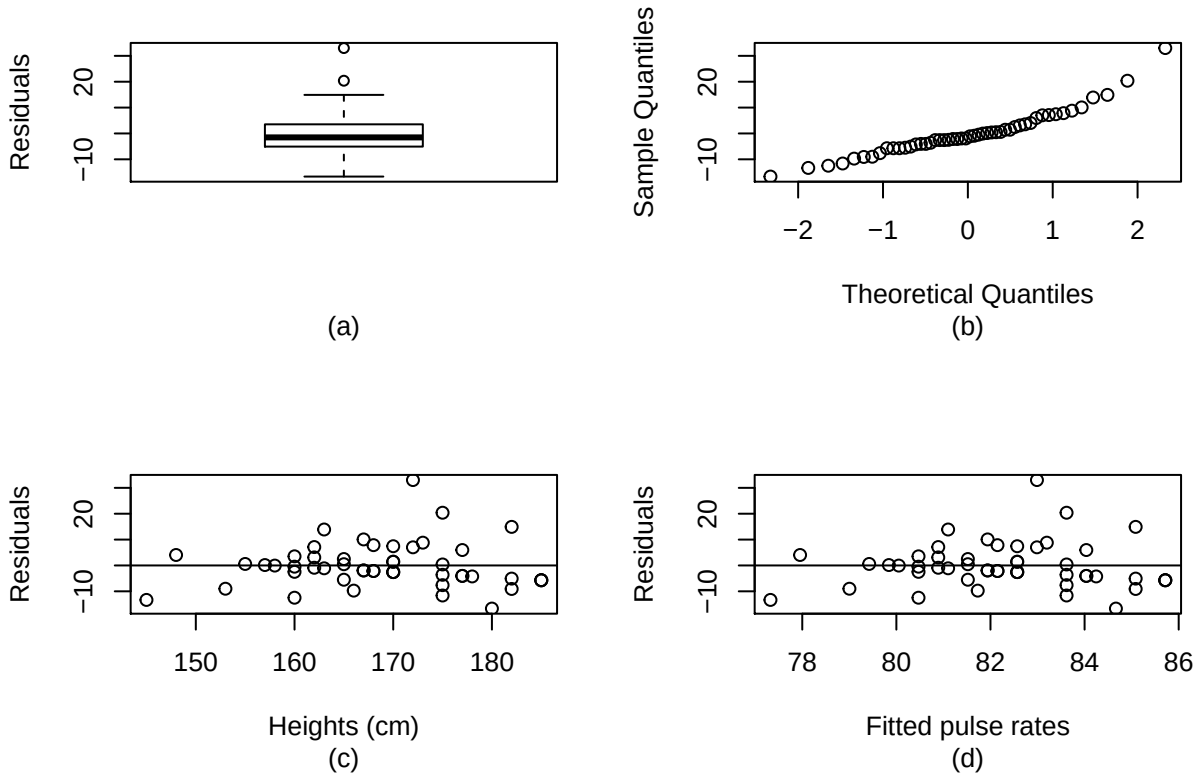


Figure 3.8 and Figure 3.10

```
par(mfrow = c(1,1))
plot(pulse ~ heights, xlab = "Height (cm)", ylab = "Pulse rate (bpm)")
abline(lm(pulse ~ heights))
lines(lowess(pulse ~ heights), lty=2)
lines(smooth.spline(heights, pulse), lty=3)
legend("topleft", c("Linear regression fit", "Lowess fit", "Spline fit"), lty=1:3)
```

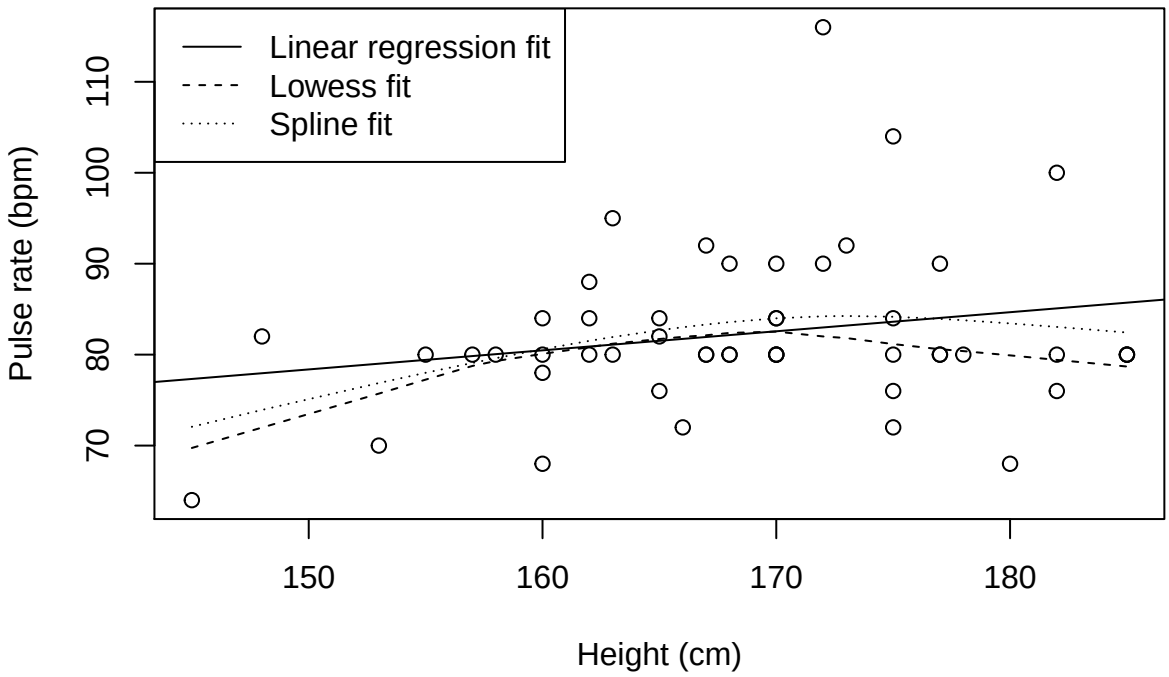


Table 3.3: Data on Oxygen Uptake and Expired Volume

```
# kinesiology data
oxygen <- c(574,592,664,667,718,770,927,947,1020,1096,1277,1323,1330,1599,
            1639,1787,1790,1794,1874,2049,2132,2160,2292,2312,2475,2489,2490,2577,
            2766,2812,2893,2957,3052,3151,3161,3266,3386,3452,3521,3543,3676,3741,
            3844,3878,4002,4114,4152,4252,4290,4331,4332,4390,4393)

ev <- c(21.9,18.6,18.6,19.1,19.2,16.9,18.3,17.2,19.0,19.0,18.6,22.8,24.6,24.9,
        29.2,32.0,27.9,31.0,30.7,35.4,36.1,39.1,42.6,39.9,46.2,50.9,46.5,46.3,
        55.8,54.5,63.5,60.3,64.8,69.2,74.7,72.9,80.4,83.0,86.0,88.9,96.8,89.1,
        100.9,103.0,113.4,111.4,119.9,127.2,126.4,135.5,138.9,143.7,144.8)
oxyev <- cbind(oxygen, ev)
head(oxyev)
```

```
##      oxygen   ev
## [1,]    574 21.9
## [2,]    592 18.6
## [3,]    664 18.6
## [4,]    667 19.1
## [5,]    718 19.2
## [6,]    770 16.9
```

Figure 3.2

```
plot(ev ~ oxygen, xlab = "Oxygen uptake", ylab = "Expired ventilation",
     ylim = c(5,150), xlim = c(500,4500))
abline(lm(ev ~ oxygen))
```

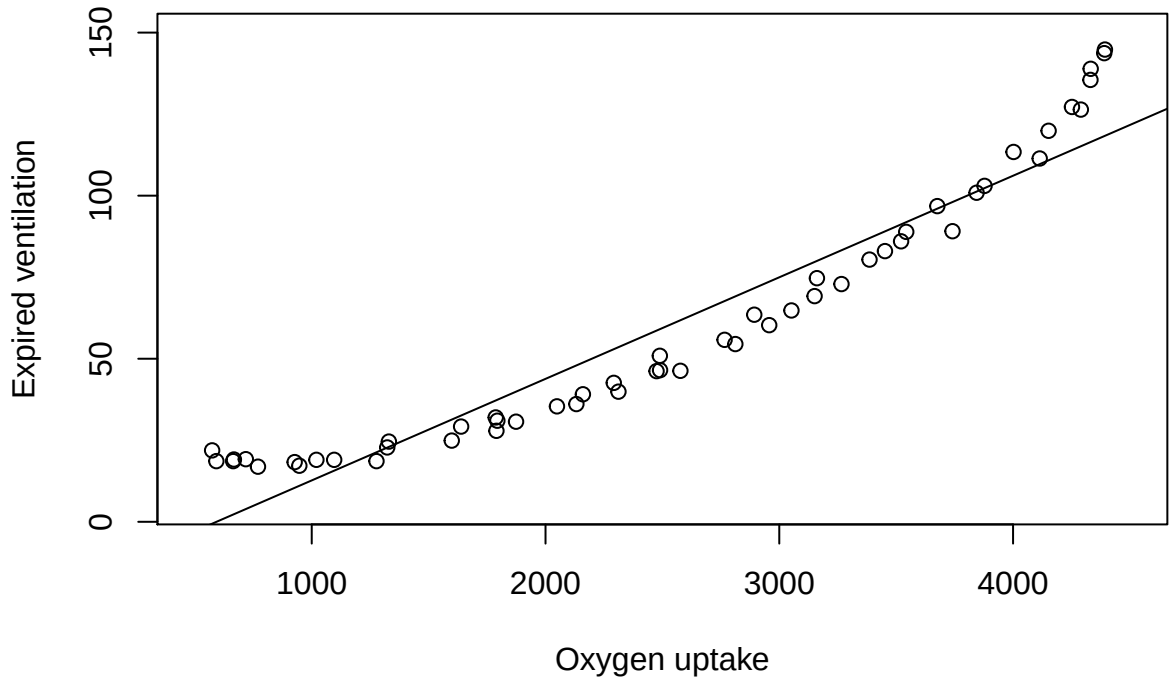


Table 3.4

```
summary(lm(ev ~ oxygen))
```

```
##
## Call:
## lm(formula = ev ~ oxygen)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.502  -9.716  -3.391   7.881  26.446
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -18.448734   3.815196  -4.836 1.26e-05 ***
## oxygen        0.031141   0.001355  22.987  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.96 on 51 degrees of freedom
## Multiple R-squared:  0.912,  Adjusted R-squared:  0.9103
## F-statistic: 528.4 on 1 and 51 DF,  p-value: < 2.2e-16
```

Figure 3.3

```
plot(ev ~ oxygen, xlab = "Oxygen uptake", ylab = "Expired ventilation",
     ylim = c(5,150), xlim = c(500,4500))
o2 <- oxygen*oxygen
reg <- lm(ev ~ oxygen + o2)
lines(oxygen, reg$fit)
```

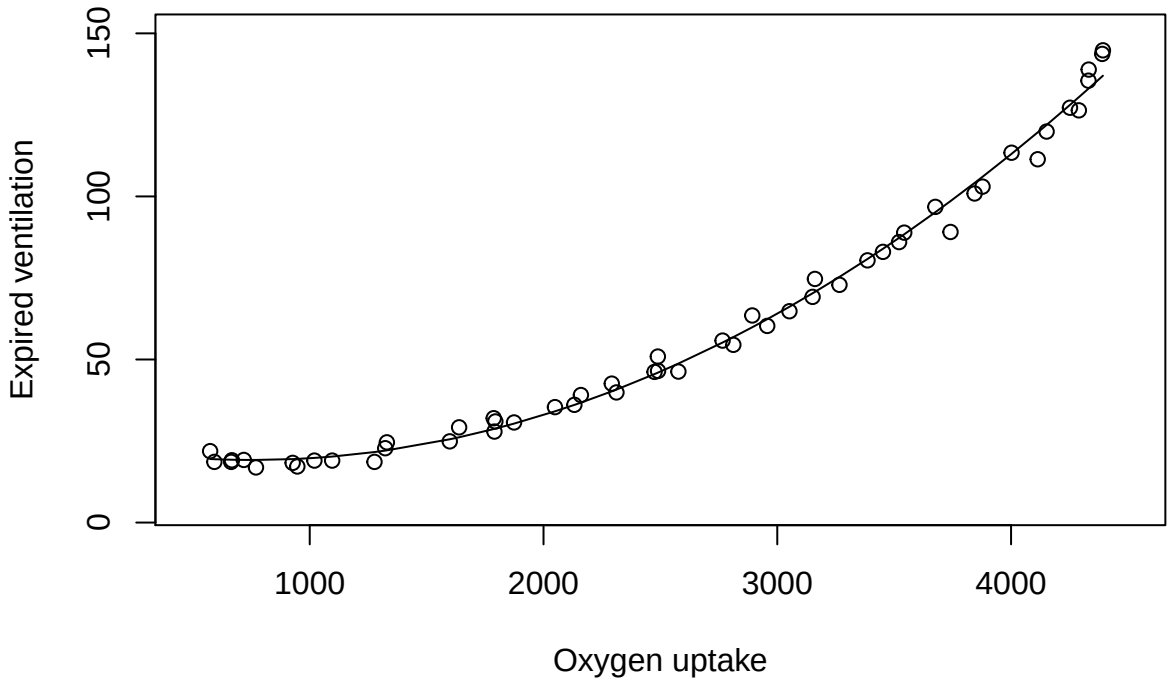


Table 3.5

```
summary(reg)
```

```
##
## Call:
## lm(formula = ev ~ oxygen + o2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.4713 -1.3675 -0.4201  2.1925  7.7817
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.427e+01  1.940e+00  12.509  < 2e-16 ***
## oxygen      -1.344e-02  1.762e-03  -7.628 6.27e-10 ***
## o2           8.902e-06  3.444e-07  25.850  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.186 on 50 degrees of freedom
## Multiple R-squared:  0.9939, Adjusted R-squared:  0.9936
## F-statistic:  4055 on 2 and 50 DF,  p-value: < 2.2e-16
```

Figure 3.6

```
reg <- lm(ev ~ oxygen)
residuals <- reg$residuals
fitval <- predict(reg)
par(mfrow = c(2,2))
boxplot(residuals, ylab = "Residuals")
title(sub="(a)")
qqnorm(residuals, main="")
title(sub="(b)")
plot(residuals ~ oxygen, xlab = "Oxygen uptake", ylab = "Residuals")
abline(h=0)
title(sub="(c)")
plot(residuals ~ fitval, xlab = "Fitted expired ventilation", ylab = "Residuals")
abline(h=0)
title(sub="(d)")
```

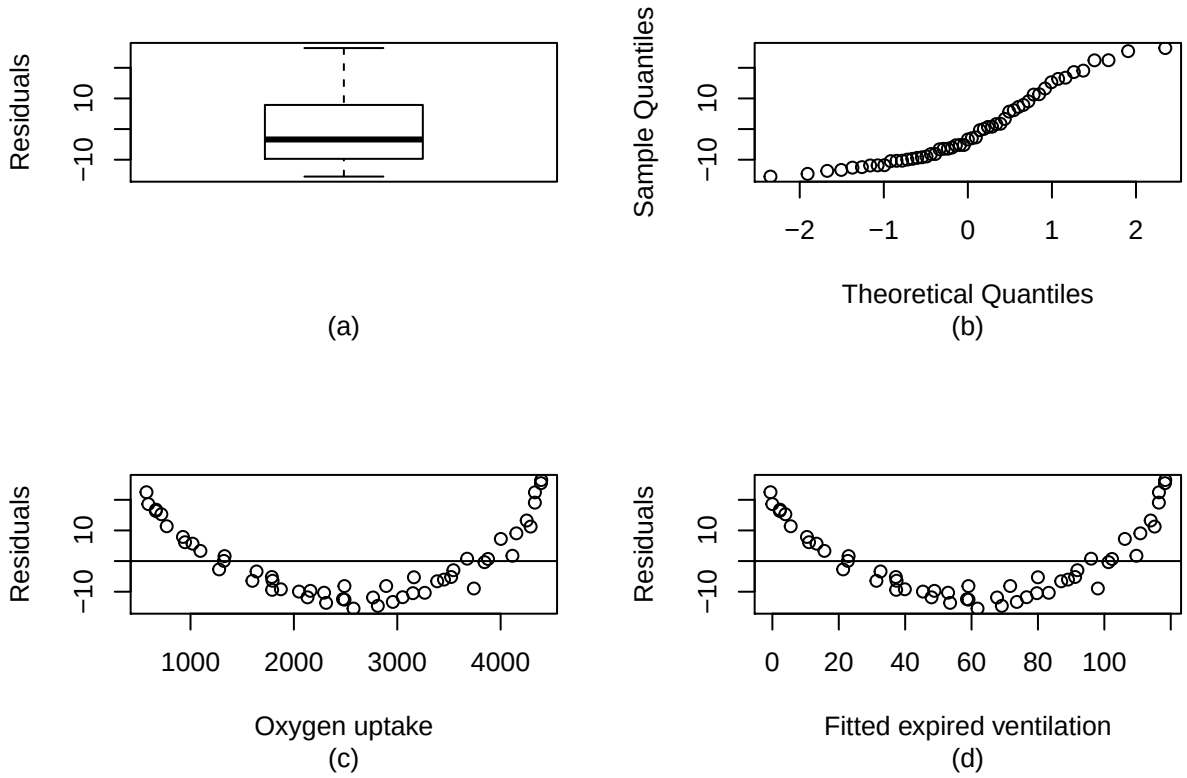


Figure 3.7

```
o2 <- oxygen*oxygen
reg <- lm(ev ~ oxygen + o2)
residuals <- reg$residuals
fitval <- predict(reg)
par(mfrow = c(2,2))
boxplot(residuals, ylab = "Residuals")
title(sub="(a)")
qqnorm(residuals, main="")
title(sub="(b)")
plot(residuals ~ oxygen, xlab = "Oxygen uptake", ylab = "Residuals")
abline(h=0)
title(sub="(c)")
plot(residuals ~ fitval, xlab = "Fitted expired ventilation", ylab = "Residuals")
abline(h=0)
title(sub="(d)")
```

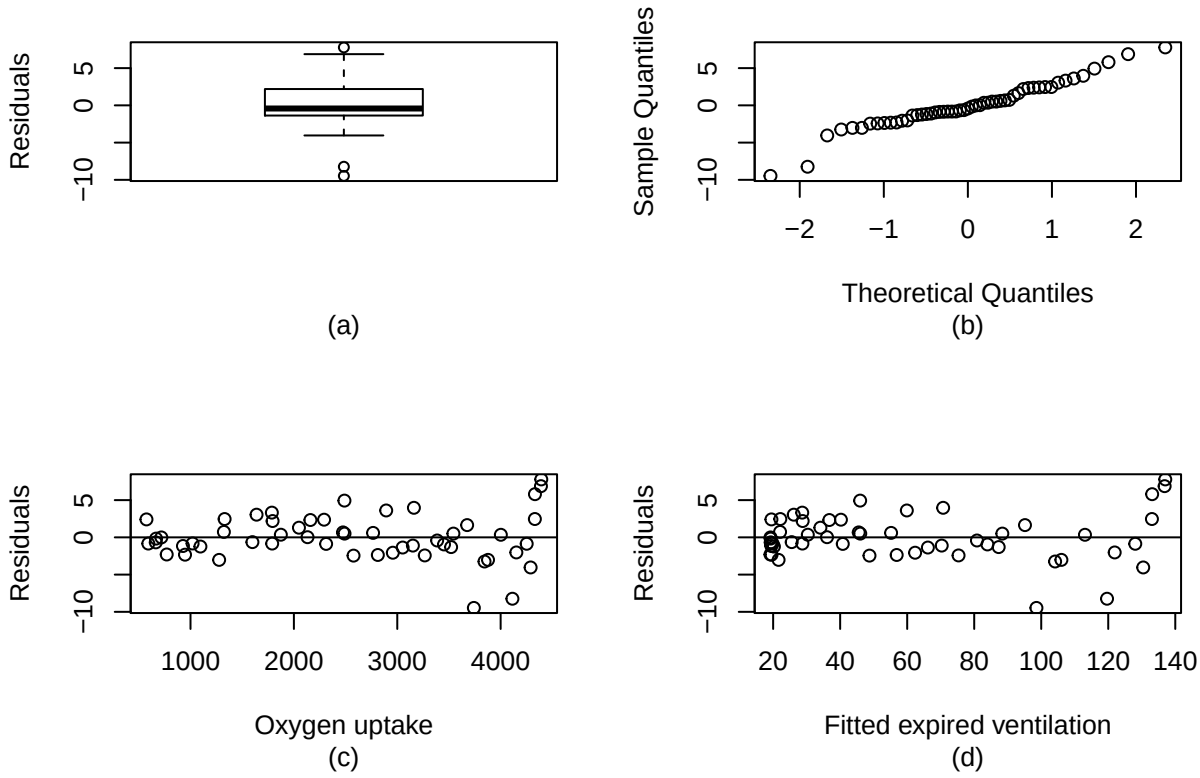


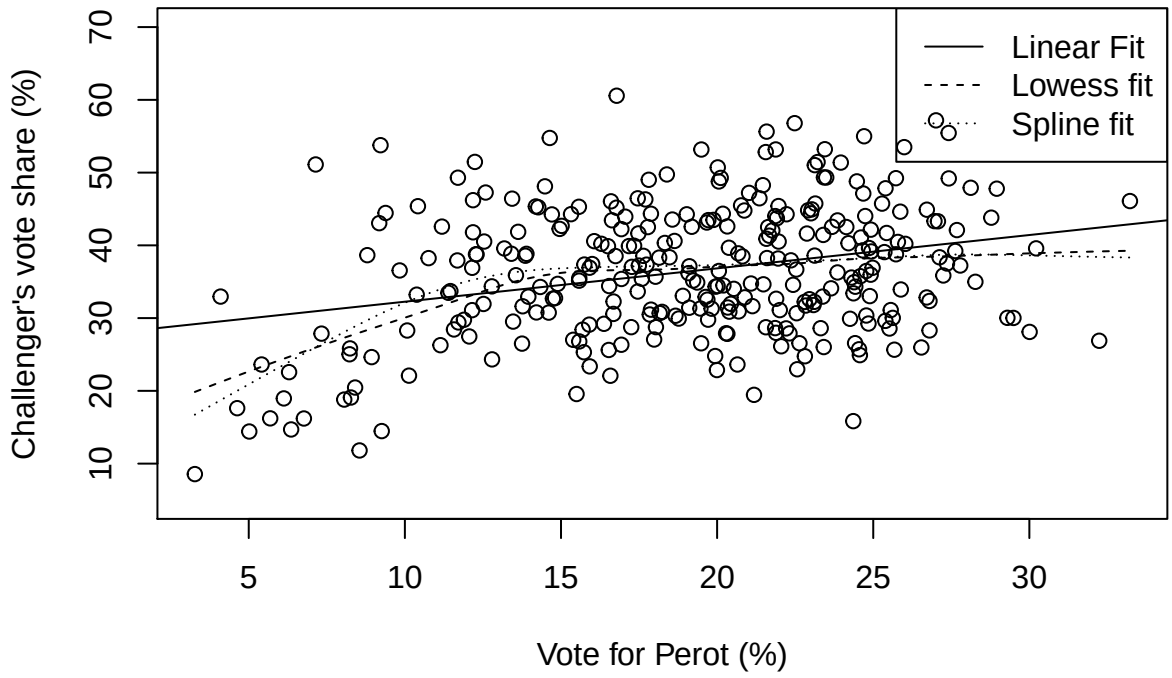
Table 3.6: Data of Challenger Vote and Perot Vote

```
vote <- read.table("data/vote.txt", sep = "\t", header = TRUE)
head(vote)
```

```
##      chall    perot
## 1 37.92675 11.68032
## 2 38.24330 10.75909
## 3 29.76948 11.89173
## 4 32.75101 14.80878
## 5 53.76603  9.22018
## 6 47.78949 28.95518
```

Figure 3.11 and Figure 3.13

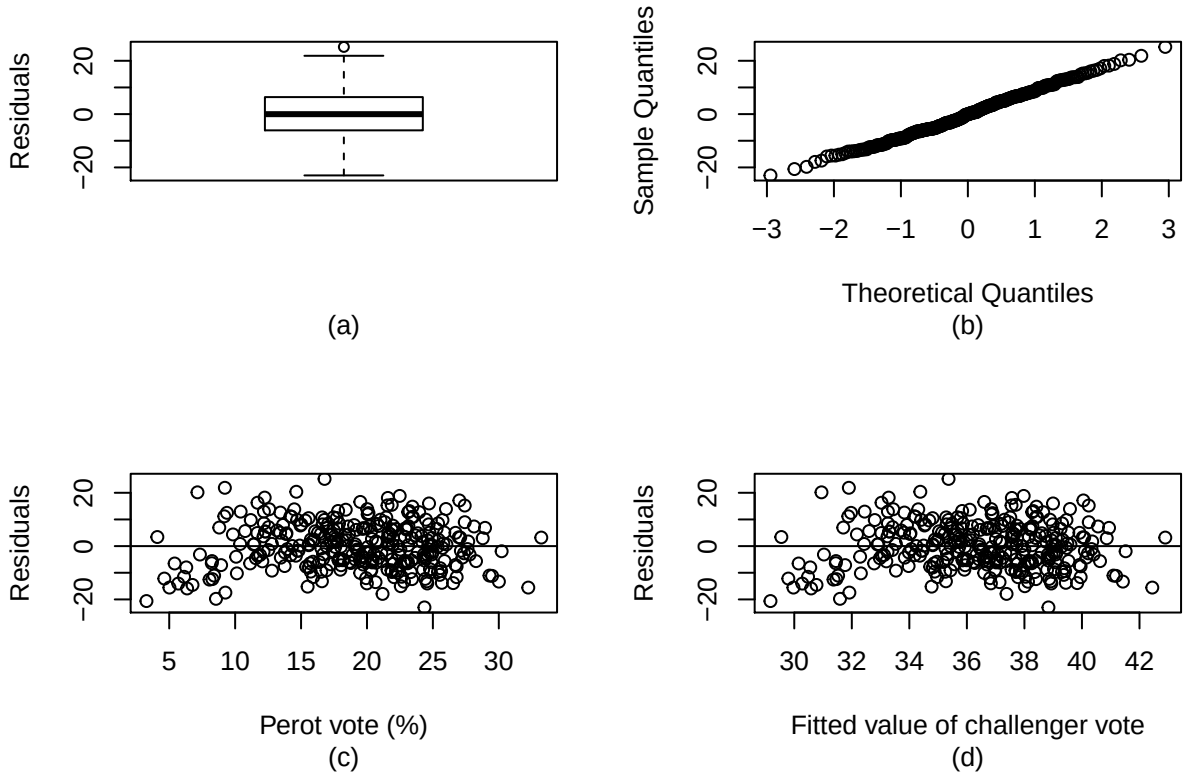
```
attach(vote)
par(mfrow = c(1,1))
plot(chall ~ perot, xlab = "Vote for Perot (%)", ylab = "Challenger's vote share (%)",
     ylim = c(5,70))
abline(lm(chall ~ perot))
lines(lowess(chall ~ perot), lty=2)
lines(smooth.spline(perot, chall), lty=3)
legend("topright", c("Linear Fit", "Lowess fit", "Spline fit"), lty=1:3)
```



```
detach(vote)
```

Figure 3.12

```
attach(vote)
reg <- lm(chall ~ perot)
residuals <- reg$residuals
fitval <- predict(reg)
par(mfrow = c(2,2))
boxplot(residuals, ylab = "Residuals")
title(sub="(a)")
qqnorm(residuals, main="")
title(sub="(b)")
plot(residuals ~ perot, xlab = "Perot vote (%)", ylab = "Residuals")
abline(h=0)
title(sub="(c)")
plot(residuals ~ fitval, xlab = "Fitted value of challenger vote", ylab = "Residuals")
abline(h=0)
title(sub="(d)")
```




```
detach(vote)
```