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

Exercises of Chapter 4: Multiple Linear Regression

Kimmo Vehkalahti and Brian S. Everitt

9 October 2018

Exercises

Exercise 4.1

Use the usapol data (see below) and modify the R code given in the **Examples of Chapter 4** to create suitable graphics and to analyse the data as instructed. Use the procedure involving the AIC described in the text to search for a more parsimonious model for the data. For the final model chosen, use some regression diagnostics to investigate the assumptions made in fitting the model.

```
usapol <- structure(list(

SO2 = c(10, 13, 12, 17, 56, 36, 29, 14, 10, 24, 110, 28, 17, 8, 30, 9, 47,
        35, 29, 14, 56, 14, 11, 46, 11, 23, 65, 26, 69, 61, 94, 10, 18, 9,
        10, 28, 31, 26, 29, 31, 16),

Temp = c(70.3, 61, 56.7, 51.9, 49.1, 54, 57.3, 68.4, 75.5, 61.5, 50.6, 52.3,
         49, 56.6, 55.6, 68.3, 55, 49.9, 43.5, 54.5, 55.9, 51.5, 56.8, 47.6,
         47.1, 54, 49.7, 51.5, 54.6, 50.4, 50, 61.6, 59.4, 66.2, 68.9, 51,
         59.3, 57.8, 51.1, 55.2, 45.7),

Manuf = c(213, 91, 453, 454, 412, 80, 434, 136, 207, 368, 3344, 361, 104, 125,
          291, 204, 625, 1064, 699, 381, 775, 181, 46, 44, 391, 462, 1007, 266,
          1692, 347, 343, 337, 275, 641, 721, 137, 96, 197, 379, 35, 569),

Pop = c(582, 132, 716, 515, 158, 80, 757, 529, 335, 497, 3369, 746, 201, 277,
        593, 361, 905, 1513, 744, 507, 622, 347, 244, 116, 463, 453, 751, 540,
        1950, 520, 179, 624, 448, 844, 1233, 176, 308, 299, 531, 71, 717),

Wind = c(6, 8.2, 8.7, 9, 9, 9, 9.3, 8.8, 9, 9.1, 10.4, 9.7, 11.2, 12.7, 8.3,
         8.4, 9.6, 10.1, 10.6, 10, 9.5, 10.9, 8.9, 8.8, 12.4, 7.1, 10.9, 8.6,
         9.6, 9.4, 10.6, 9.2, 7.9, 10.9, 10.8, 8.7, 10.6, 7.6, 9.4, 6.5, 11.8),

Precip = c(7.05, 48.52, 20.66, 12.95, 43.37, 40.25, 38.89, 54.47, 59.8, 48.34,
           34.44, 38.74, 30.85, 30.58, 43.11, 56.77, 41.31, 30.96, 25.94, 37,
           35.89, 30.18, 7.77, 33.36, 36.11, 39.04, 34.99, 37.01, 39.93, 36.22,
           42.75, 49.1, 46, 35.94, 48.19, 15.17, 44.68, 42.59, 38.79, 40.75, 29.07),

Days = c(36, 100, 67, 86, 127, 114, 111, 116, 128, 115, 122, 121, 103, 82, 123,
         113, 111, 129, 137, 99, 105, 98, 58, 135, 166, 132, 155, 134, 115, 147,
         125, 105, 119, 78, 103, 89, 116, 115, 164, 148, 123)),
```

```
.Names = c("SO2", "Temp", "Manuf", "Pop", "Wind", "Precip", "Days"),

row.names = c("Phoenix", "Little Rock", "San Francisco", "Denver", "Hartford",
              "Wilmington", "Washington", "Jacksonville", "Miami", "Atlanta",
              "Chicago", "Indianapolis", "Des Moines", "Wichita", "Louisville",
              "New Orleans", "Baltimore", "Detroit", "Minneapolis-St. Paul",
              "Kansas City", "St. Louis", "Omaha", "Alburquerque", "Albany",
              "Buffalo", "Cincinnati", "Cleveland", "Columbus", "Philadelphia",
              "Pittsburgh", "Providence", "Memphis", "Nashville", "Dallas",
              "Houston", "Salt Lake City", "Norfolk", "Richmond", "Seattle",
              "Charleston", "Milwaukee"), class = "data.frame")
```

```
head(usapol, n = 10)
```

```
##               SO2 Temp Manuf Pop Wind Precip Days
## Phoenix        10 70.3   213 582  6.0   7.05   36
## Little Rock    13 61.0    91 132  8.2  48.52  100
## San Francisco  12 56.7   453 716  8.7  20.66   67
## Denver         17 51.9   454 515  9.0  12.95   86
## Hartford       56 49.1   412 158  9.0  43.37  127
## Wilmington     36 54.0    80  80  9.0  40.25  114
## Washington     29 57.3   434 757  9.3  38.89  111
## Jacksonville   14 68.4   136 529  8.8  54.47  116
## Miami          10 75.5   207 335  9.0  59.80  128
## Atlanta        24 61.5   368 497  9.1  48.34  115
```

```
tail(usapol, n = 10)
```

```
##               SO2 Temp Manuf  Pop Wind Precip Days
## Memphis        10 61.6   337  624  9.2  49.10  105
## Nashville      18 59.4   275  448  7.9  46.00  119
## Dallas          9 66.2   641  844 10.9  35.94   78
## Houston        10 68.9   721 1233 10.8  48.19  103
## Salt Lake City 28 51.0   137  176  8.7  15.17   89
## Norfolk        31 59.3    96  308 10.6  44.68  116
## Richmond       26 57.8   197  299  7.6  42.59  115
## Seattle        29 51.1   379  531  9.4  38.79  164
## Charleston     31 55.2    35   71  6.5  40.75  148
## Milwaukee      16 45.7   569  717 11.8  29.07  123
```

Exercise 4.2

Use the quality data (see below) and modify the R code given in the **Examples of Chapter 4** to create suitable graphics and to analyse the data. Pay careful attention to how the categorical explanatory variables with more than two categories are coded.

```
quality <- structure(list(

age = structure(c(1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 2L, 2L, 2L, 1L, 1L, 1L, 2L, 2L, 2L, 2L,
                  1L, 2L, 2L, 2L, 1L), .Label = c("A5-6", "A8-9"), class = "factor"),

sex = structure(c(1L,2L, 1L, 2L, 1L, 2L, 2L, 2L, 1L, 2L, 1L, 1L, 2L, 1L, 2L, 1L, 2L, 2L,
                  2L, 1L, 1L, 1L, 1L), .Label = c("Male", "Female"), class = "factor"),

location = structure(c(3L, 2L, 1L, 2L, 3L, 3L, 4L, 2L, 3L, 2L, 3L, 1L, 3L, 2L, 2L, 4L,
                       2L, 3L, 4L, 2L, 4L, 4L, 4L),
                     .Label = c("Home", "School", "Room", "Kroom"), class = "factor"),

coherence = c(3.81, 1.63, 3.54, 4.21, 3.3, 2.32, 4.51, 3.18, 3.02, 2.77, 3.35, 2.66,
              4.7, 4.31, 2.16, 1.89, 1.94, 2.86, 3.11, 2.9, 2.41, 2.32, 2.78),

maturity = c(3.62, 1.61, 3.63, 4.11, 3.12, 2.13, 4.31, 3.08, 3, 2.71, 3.07, 2.72,
             4.98, 4.21, 2.91, 1.87, 1.99, 2.93, 3.01, 2.87, 2.38, 2.33, 2.79),

delay = c(45, 27, 102, 39, 41, 70, 72, 41, 71, 56, 88, 13, 29, 39, 10, 15, 46, 57,
          26, 14, 45, 19, 9),

prosecute = structure(c(1L, 2L, 1L, 1L, 1L, 2L, 1L, 1L, 1L, 2L, 2L, 1L, 1L, 2L, 1L, 2L,
                        2L, 1L, 2L, 1L, 1L, 2L, 2L),
                      .Label = c("No", "Yes"), class = "factor"),

qualityct = c(34.11, 36.59, 37.23, 39.65, 42.07, 44.91, 45.23, 47.53, 54.64, 57.87,
              57.07, 45.81, 49.38, 49.53, 67.08, 83.15, 80.67, 78.47, 77.59, 76.28,
              59.64, 68.44, 65.07)),
              .Names = c("age", "sex", "location", "coherence", "maturity", "delay",
                         "prosecute", "qualityct"),

row.names = c(NA, -23L), class = "data.frame")

str(quality)
```

```
## 'data.frame':    23 obs. of  8 variables:
##  $ age      : Factor w/ 2 levels "A5-6","A8-9": 1 1 1 1 1 1 1 1 2 2 ...
##  $ sex      : Factor w/ 2 levels "Male","Female": 1 2 1 2 1 2 2 2 1 2 ...
##  $ location : Factor w/ 4 levels "Home","School",..: 3 2 1 2 3 3 4 2 3 2 ...
##  $ coherence: num  3.81 1.63 3.54 4.21 3.3 2.32 4.51 3.18 3.02 2.77 ...
##  $ maturity : num  3.62 1.61 3.63 4.11 3.12 2.13 4.31 3.08 3 2.71 ...
##  $ delay    : num  45 27 102 39 41 70 72 41 71 56 ...
##  $ prosecute: Factor w/ 2 levels "No","Yes": 1 2 1 1 1 2 1 1 1 2 ...
##  $ qualityct: num  34.1 36.6 37.2 39.6 42.1 ...
```

```
quality
```

```
##     age    sex location coherence maturity delay prosecute qualityct
## 1  A5-6   Male     Room      3.81     3.62    45         No     34.11
## 2  A5-6 Female   School      1.63     1.61    27        Yes     36.59
## 3  A5-6   Male     Home      3.54     3.63   102         No     37.23
## 4  A5-6 Female   School      4.21     4.11    39         No     39.65
## 5  A5-6   Male     Room      3.30     3.12    41         No     42.07
## 6  A5-6 Female     Room      2.32     2.13    70        Yes     44.91
## 7  A5-6 Female    Kroom      4.51     4.31    72         No     45.23
## 8  A5-6 Female   School      3.18     3.08    41         No     47.53
## 9  A8-9   Male     Room      3.02     3.00    71         No     54.64
## 10 A8-9 Female   School      2.77     2.71    56        Yes     57.87
## 11 A8-9   Male     Room      3.35     3.07    88        Yes     57.07
## 12 A5-6   Male     Home      2.66     2.72    13         No     45.81
## 13 A5-6 Female     Room      4.70     4.98    29         No     49.38
## 14 A5-6   Male   School      4.31     4.21    39        Yes     49.53
## 15 A8-9 Female   School      2.16     2.91    10         No     67.08
## 16 A8-9   Male    Kroom      1.89     1.87    15        Yes     83.15
## 17 A8-9 Female   School      1.94     1.99    46        Yes     80.67
## 18 A8-9 Female     Room      2.86     2.93    57         No     78.47
## 19 A5-6 Female    Kroom      3.11     3.01    26        Yes     77.59
## 20 A8-9   Male   School      2.90     2.87    14         No     76.28
## 21 A8-9   Male    Kroom      2.41     2.38    45         No     59.64
## 22 A8-9   Male    Kroom      2.32     2.33    19        Yes     68.44
## 23 A5-6   Male    Kroom      2.78     2.79     9        Yes     65.07
```

Exercise 4.3

Use the quartet data (see below) and modify the R code given in the **Examples of Chapter 4**.

For more information, see the Wikipedia article:

https://en.wikipedia.org/wiki/Anscombe%27s_quartet

```
quartet <- structure(list(
  x1 = c(10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5),
  y1 = c(8.04, 6.95, 7.58, 8.81, 8.33, 9.96, 7.24, 4.26, 10.84, 4.82, 5.68),

  x2 = c(10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5),
  y2 = c(9.14, 8.14, 8.74, 8.77, 9.26, 8.1, 6.13, 3.1, 9.13, 7.26, 4.74),

  x3 = c(10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5),
  y3 = c(7.46, 6.77, 12.74, 7.11, 7.81, 8.84, 6.08, 5.39, 8.15, 6.42, 5.73),

  x4 = c(8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 19),
  y4 = c(6.58, 5.76, 7.71, 8.84, 8.47, 7.04, 5.25, 5.56, 7.91, 6.89, 12.5)),

.Names = c("x1", "y1", "x2", "y2", "x3", "y3", "x4", "y4"),
row.names = c(NA, -11L), class = "data.frame"
)

quartet
```

```
##    x1    y1 x2   y2 x3    y3 x4    y4
## 1  10  8.04 10 9.14 10  7.46  8  6.58
## 2   8  6.95  8 8.14  8  6.77  8  5.76
## 3  13  7.58 13 8.74 13 12.74  8  7.71
## 4   9  8.81  9 8.77  9  7.11  8  8.84
## 5  11  8.33 11 9.26 11  7.81  8  8.47
## 6  14  9.96 14 8.10 14  8.84  8  7.04
## 7   6  7.24  6 6.13  6  6.08  8  5.25
## 8   4  4.26  4 3.10  4  5.39  8  5.56
## 9  12 10.84 12 9.13 12  8.15  8  7.91
## 10  7  4.82  7 7.26  7  6.42  8  6.89
## 11  5  5.68  5 4.74  5  5.73 19 12.50
```

Exercise 4.4

Use the fat data (see below) and modify the R code given in the **Examples of Chapter 4**.

```
fat <- structure(list(

  Sex = structure(c(1L, 2L, 1L, 1L, 2L, 2L, 1L, 2L, 2L, 2L,
                    2L, 1L, 2L, 2L, 2L, 1L, 2L, 2L, 2L, 2L),
                  .Label = c("F", "M"), class = "factor"),

    Age = c(23L, 23L, 27L, 27L, 39L, 41L, 45L, 49L, 50L, 53L,
            53L, 54L, 54L, 56L, 57L, 57L, 58L, 58L, 60L, 61L),

    Pcfat = c(9.5, 27.9, 7.8, 17.8, 31.4, 25.9, 27.4, 25.2, 31.1,
              34.7, 42, 20, 29.1, 32.5, 30.3, 21, 33, 33.8, 41.1, 34.5)),

.Names = c("Sex", "Age", "Pcfat"), row.names = c(NA, -20L ), class = "data.frame")

fat
```

```
##    Sex Age Pcfat
## 1    F  23   9.5
## 2    M  23  27.9
## 3    F  27   7.8
## 4    F  27  17.8
## 5    M  39  31.4
## 6    M  41  25.9
## 7    F  45  27.4
## 8    M  49  25.2
## 9    M  50  31.1
## 10   M  53  34.7
## 11   M  53  42.0
## 12   F  54  20.0
## 13   M  54  29.1
## 14   M  56  32.5
## 15   M  57  30.3
## 16   F  57  21.0
## 17   M  58  33.0
## 18   M  58  33.8
## 19   M  60  41.1
## 20   M  61  34.5
```

Exercise 4.5

Use the blood data (see below) and modify the R code given in the **Examples of Chapter 4**.

```
blood <- structure(list(
  History = structure(c(1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L,
                        1L, 1L, 1L, 1L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L,
                        2L, 2L, 2L, 2L, 2L, 2L),
                      .Label = c("Present", "Absent"), class = "factor"),

  Smoking = structure(c(1L, 1L, 1L, 1L, 1L, 1L, 1L, 2L, 2L, 2L, 2L, 2L, 2L, 3L, 3L, 3L,
                        3L, 3L, 3L, 3L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 2L, 2L, 2L, 2L,
                        3L, 3L, 3L, 3L, 3L, 3L),
                      .Label = c("NonS", "ExS", "CS"), class = "factor"),

  Bloodp = c(125, 156, 103, 129, 110, 128, 135, 114, 107, 134, 140, 120, 115, 135, 120,
             123, 113, 165, 145, 120, 114, 110, 91, 136, 105, 125, 103, 110, 110, 128,
             105, 90, 140, 125, 123, 108, 113, 160)),

.Names = c("History", "Smoking", "Bloodp"), row.names = c(NA, -38L), class = "data.frame")

str(blood)
```

```
## 'data.frame':    38 obs. of  3 variables:
##  $ History: Factor w/ 2 levels "Present","Absent": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Smoking: Factor w/ 3 levels "NonS","ExS","CS": 1 1 1 1 1 1 1 2 2 2 ...
##  $ Bloodp : num  125 156 103 129 110 128 135 114 107 134 ...
```

```
head(blood)
```

```
##   History Smoking Bloodp
## 1 Present    NonS    125
## 2 Present    NonS    156
## 3 Present    NonS    103
## 4 Present    NonS    129
## 5 Present    NonS    110
## 6 Present    NonS    128
```

```
tail(blood)
```

```
##    History Smoking Bloodp
## 33  Absent      CS    140
## 34  Absent      CS    125
## 35  Absent      CS    123
## 36  Absent      CS    108
## 37  Absent      CS    113
## 38  Absent      CS    160
```

Exercise 4.6

Use the oestrogen data (see below) and modify the R code given in the **Examples of Chapter 4**.

```
oestrogen <- structure(list(
  Treatment = structure(c(1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L,
                          2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L),
            .Label = c("Placebo", "Active"), class = "factor"),

  BL1 = c(18, 25, 24, 19, 22, 27, 21, 26, 20, 24, 24, 27, 19, 25,
          19, 21, 21, 25, 25, 15, 27),

  BL2 = c(18, 27, 17, 15, 20, 28, 16, 26, 19, 20, 22, 27, 15, 28,
          18, 20, 21, 24, 25, 22, 26),

  Depression = c(15, 10, 12, 5, 5, 9, 11, 13, 6, 18, 10, 7, 8, 2,
                 6, 11, 5, 11, 6, 6, 10)),

.Names = c("Treatment", "BL1", "BL2", "Depression"),
row.names = c(NA, -21L), class = "data.frame")

str(oestrogen)
```

```
## 'data.frame':    21 obs. of  4 variables:
##  $ Treatment : Factor w/ 2 levels "Placebo","Active": 1 1 1 1 1 1 1 1 1 1 ...
##  $ BL1       : num  18 25 24 19 22 27 21 26 20 24 ...
##  $ BL2       : num  18 27 17 15 20 28 16 26 19 20 ...
##  $ Depression: num  15 10 12 5 5 9 11 13 6 18 ...
```

```
oestrogen
```

```
##    Treatment BL1 BL2 Depression
## 1    Placebo  18  18         15
## 2    Placebo  25  27         10
## 3    Placebo  24  17         12
## 4    Placebo  19  15          5
## 5    Placebo  22  20          5
## 6    Placebo  27  28          9
## 7    Placebo  21  16         11
## 8    Placebo  26  26         13
## 9    Placebo  20  19          6
## 10   Placebo  24  20         18
## 11   Placebo  24  22         10
## 12    Active  27  27          7
## 13    Active  19  15          8
## 14    Active  25  28          2
## 15    Active  19  18          6
## 16    Active  21  20         11
## 17    Active  21  21          5
## 18    Active  25  24         11
## 19    Active  25  25          6
## 20    Active  15  22          6
## 21    Active  27  26         10
```