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

# Solutions to Exercises of Chapter 4: Multiple Linear Regression

*Kimmo Vehkalahti and Brian S. Everitt*

*9 October 2018*

## Solutions

### Exercise 4.1

```
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",
```

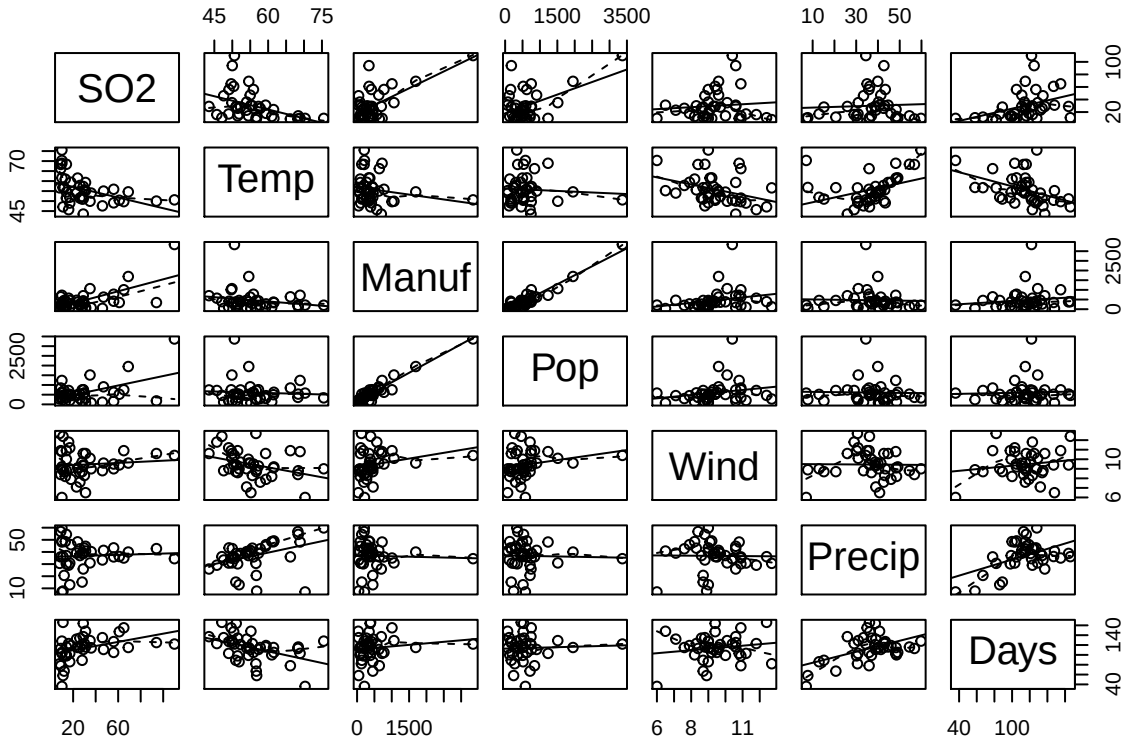
```
              "Wilmington", "Washington", "Jacksonville", "Miami", "Atlanta",
              "Chicago", "Indianapolis", "Des Moines", "Wichita", "Louisville",
              "New Orleans", "Baltimore", "Detroit", "Minneapolis-St. Paul",
              "Kansas City", "St. Louis", "Omaha", "Alburquerque", "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
```

```
attach(usapol)
# Scatterplot matrix with fitted linear and locally weighted regressions
pairs(usapol, panel = function(x, y) {
                        points(x, y)
                        abline(lm(y ~ x))
                        lines(lowess(y ~ x), lty = 2)
                      }
      )
```




```
usapol_reg <- lm(SO2 ~ Temp + Manuf + Pop + Wind + Precip + Days)
summary(usapol_reg)
```

```
##
## Call:
## lm(formula = SO2 ~ Temp + Manuf + Pop + Wind + Precip + Days)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -23.004  -8.542  -0.991   5.758  48.758
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 111.72848   47.31810   2.361 0.024087 *
## Temp         -1.26794    0.62118  -2.041 0.049056 *
## Manuf         0.06492    0.01575   4.122 0.000228 ***
## Pop          -0.03928    0.01513  -2.595 0.013846 *
## Wind         -3.18137    1.81502  -1.753 0.088650 .
## Precip        0.51236    0.36276   1.412 0.166918
## Days         -0.05205    0.16201  -0.321 0.749972
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.64 on 34 degrees of freedom
## Multiple R-squared:  0.6695, Adjusted R-squared:  0.6112
## F-statistic: 11.48 on 6 and 34 DF,  p-value: 5.419e-07
```

```r
# Run regressions to get Rsq values
s1 <- summary(lm(Temp ~ Manuf + Pop + Wind + Precip + Days))
s2 <- summary(lm(Manuf ~ Temp + Pop + Wind + Precip + Days))
s3 <- summary(lm(Pop ~ Manuf + Temp + Wind + Precip + Days))
s4 <- summary(lm(Wind ~ Manuf + Pop + Temp + Precip + Days))
s5 <- summary(lm(Precip ~ Manuf + Pop + Wind + Temp + Days))
s6 <- summary(lm(Days ~ Manuf + Pop + Wind + Precip + Temp))
#
# VIFs
rsq <- c(s1$r.squared, s2$r.squared, s3$r.squared, s4$r.squared, s5$r.squared, s6$r.squared)
rsq
```

```
## [1] 0.7343249 0.9319897 0.9302690 0.2035167 0.7063074 0.7096105
```

```r
1/(1 - rsq)
```

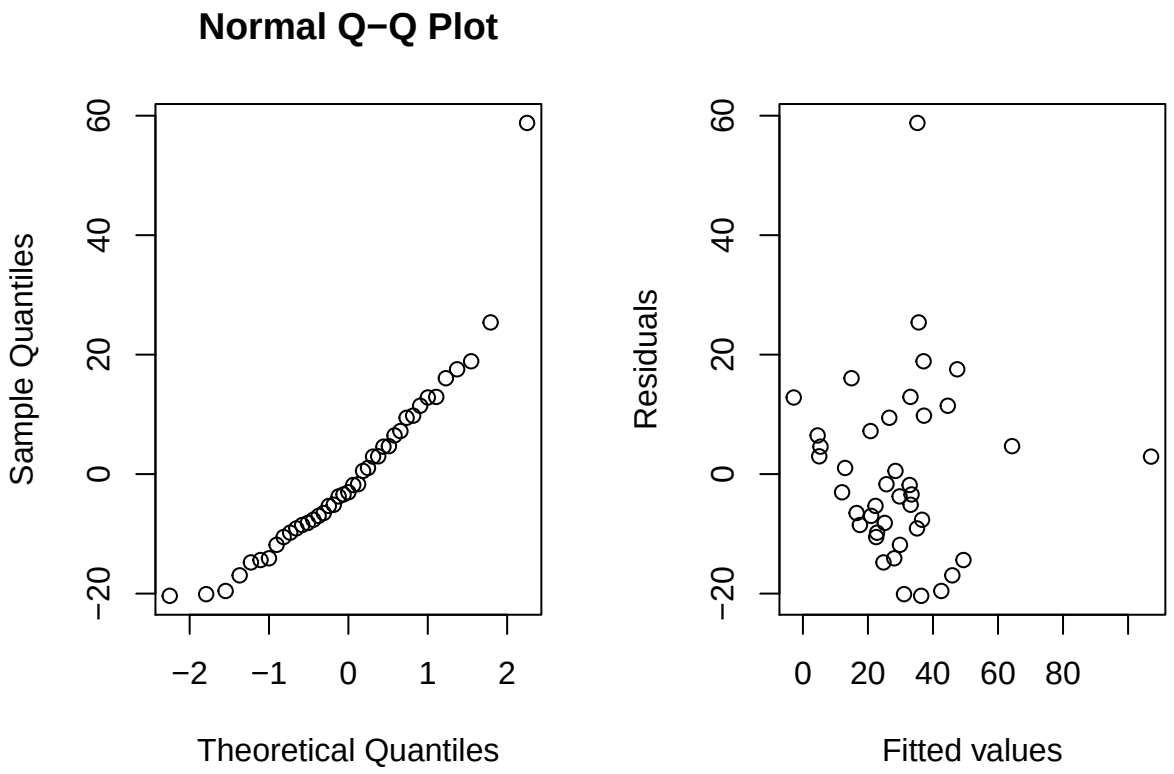
```
## [1]  3.763996 14.703652 14.340833  1.255519  3.404921  3.443651
```

```r
# Drop population size
usapol_reg1 <- lm(SO2 ~ Temp + Manuf + Wind + Precip + Days)
step(usapol_reg1, method = "backwards")
```

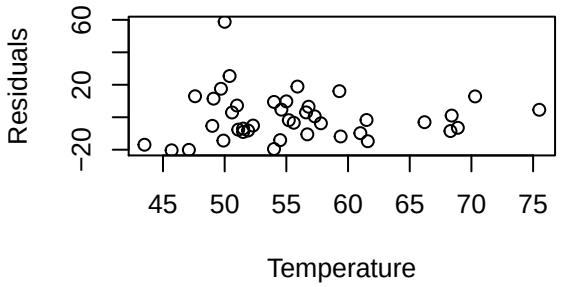
```
## Start:  AIC=231.78
## SO2 ~ Temp + Manuf + Wind + Precip + Days
##
##          Df Sum of Sq     RSS    AIC
## - Days    1      26.6  8752.9 229.91
## <none>                 8726.3 231.78
## - Precip  1     647.1  9373.4 232.72
## - Wind    1     921.4  9647.7 233.90
## - Temp    1    1930.3 10656.6 237.97
## - Manuf   1    7692.0 16418.4 255.70
##
## Step:  AIC=229.91
## SO2 ~ Temp + Manuf + Wind + Precip
##
##          Df Sum of Sq     RSS    AIC
## <none>                 8752.9 229.91
## - Wind    1     894.8  9647.7 231.90
## - Precip  1    1269.7 10022.6 233.46
## - Temp    1    3919.0 12671.9 243.08
## - Manuf   1    7665.8 16418.7 253.70
##
## Call:
## lm(formula = SO2 ~ Temp + Manuf + Wind + Precip)
##
## Coefficients:
## (Intercept)         Temp        Manuf         Wind       Precip
##   123.11833     -1.61144      0.02548     -3.63024      0.52423
```

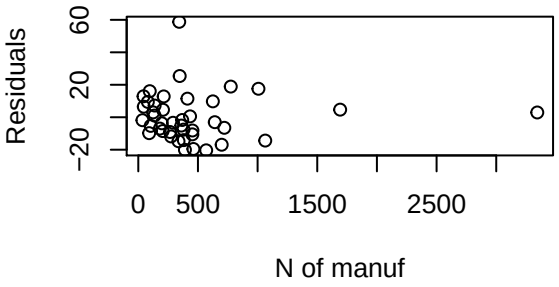
```r
#
# Drop Days
usapol_reg2 <- lm(SO2 ~ Temp + Manuf + Wind + Precip)
pred <- predict(usapol_reg2)
resid <- residuals(usapol_reg2)
```

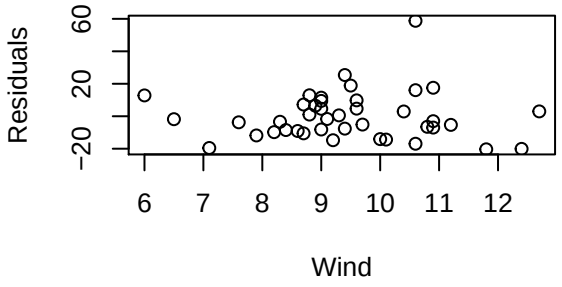
```r
# Normal probability plot and plot of residuals against fitted values
par(mfrow = c(1,2))
qqnorm(resid)
plot(resid ~ pred, xlab = "Fitted values", ylab = "Residuals")
```

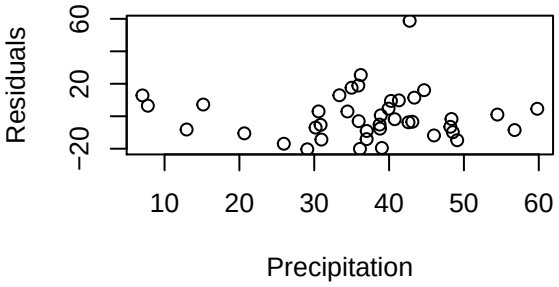



```r
# Plot of residuals against each explanatory variable in the final model
par(mfrow = c(2,2))
plot(resid ~ Temp, xlab = "Temperature", ylab = "Residuals")
plot(resid ~ Manuf, xlab="N of manuf", ylab = "Residuals")
plot(resid ~ Wind, ylab = "Residuals")
plot(resid ~ Precip, xlab = "Precipitation", ylab = "Residuals")
```




```
detach(usapol)
```

### Exercise 4.2

```
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)
```

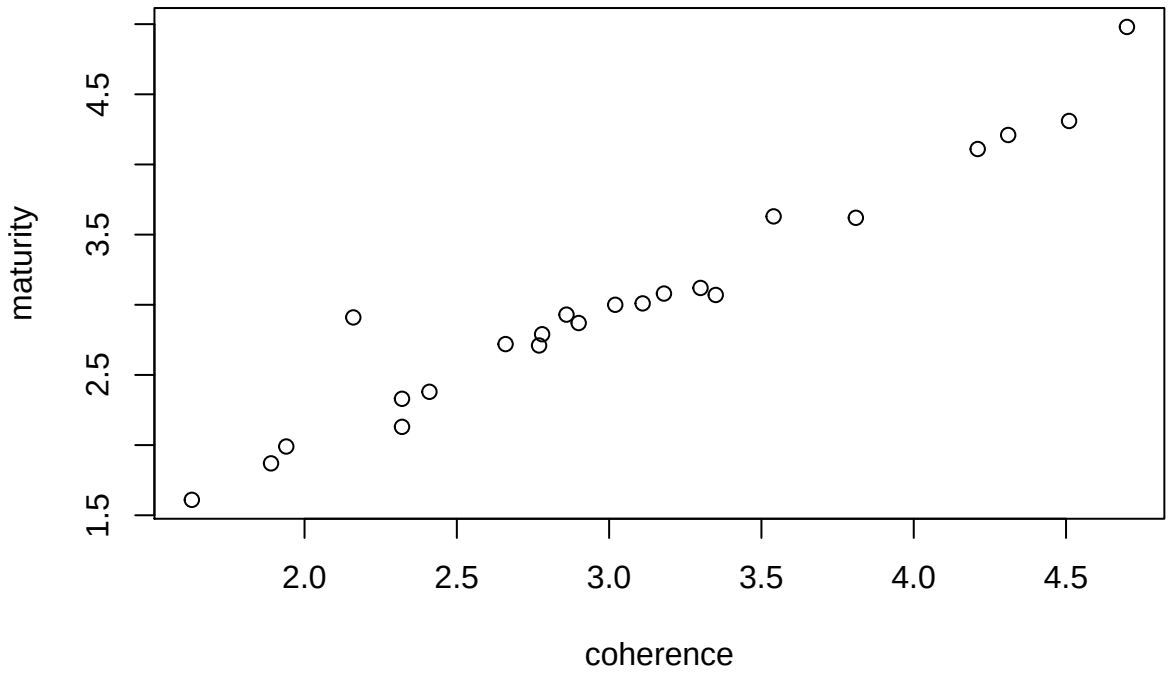
```
## 'data.frame':    23 obs. of  8 variables:
##  $ age      : Factor w/ 2 levels "A5-6","A8-9": 1 1 1 1 1 1 1 1 2 2 ...
##  $ sex      : Factor w/ 2 levels "Male","Female": 1 2 1 2 1 2 2 2 1 2 ...
##  $ location : Factor w/ 4 levels "Home","School",..: 3 2 1 2 3 3 4 2 3 2 ...
##  $ coherence: num  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 ...
```

```
head(quality, n = 10); tail(quality, n = 4)
```

```
##     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
```

```
##     age  sex location coherence maturity delay prosecute qualityct
## 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
```

```
attach(quality)
plot(coherence, maturity)
```




```
cor(coherence, maturity)
```

```
## [1] 0.9706296
```

```r
quality_reg <- lm(qualityct ~ age + location + maturity + delay + prosecute)

# Show structure of dummy variables
contrasts(age)
```

```
##      A8-9
## A5-6    0
## A8-9    1
```

```r
contrasts(sex)
```

```
##        Female
## Male        0
## Female      1
```

```r
contrasts(location)
```

```
##        School Room Kroom
## Home        0    0     0
## School      1    0     0
## Room        0    1     0
## Kroom       0    0     1
```

```r
contrasts(prosecute)
```

```
##     Yes
## No    0
## Yes   1
```

```r
summary(quality_reg)
```

```
##
## Call:
## lm(formula = qualityct ~ age + location + maturity + delay +
##     prosecute)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.595  -4.888  -1.113   3.104  18.101
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     46.0882    13.4626   3.423 0.003772 **
## ageA8-9         21.6144     5.2254   4.136 0.000879 ***
## locationSchool  -3.0069     9.4591  -0.318 0.754956
## locationRoom    -1.4705     8.8403  -0.166 0.870112
## locationKroom    5.1984    10.0058   0.520 0.610969
## maturity         2.1288     3.5699   0.596 0.559852
## delay           -0.1970     0.1019  -1.933 0.072303 .
## prosecuteYes     6.9170     5.4837   1.261 0.226439
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.26 on 15 degrees of freedom
## Multiple R-squared:  0.7008, Adjusted R-squared:  0.5612
## F-statistic: 5.019 on 7 and 15 DF,  p-value: 0.00426
```

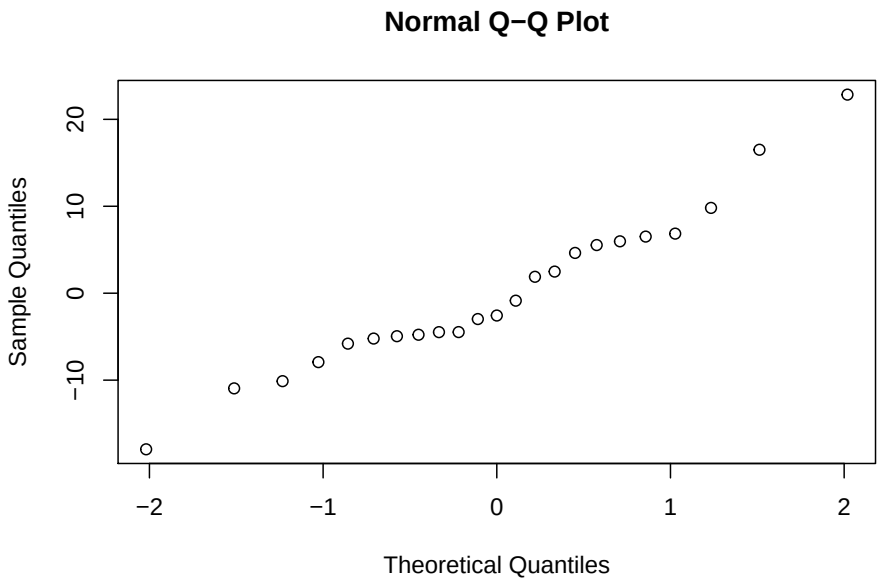
```r
step(quality_reg, method = "backwards")
```

```
## Start:  AIC=113.25
## qualityct ~ age + location + maturity + delay + prosecute
##
##             Df Sum of Sq    RSS    AIC
## - location   3    234.74 1812.4 110.44
## - maturity   1     37.40 1615.1 111.79
## <none>                   1577.7 113.25
## - prosecute  1    167.35 1745.1 113.57
## - delay      1    393.12 1970.8 116.37
## - age        1   1799.65 3377.4 128.75
##
## Step:  AIC=110.44
## qualityct ~ age + maturity + delay + prosecute
##
##             Df Sum of Sq    RSS    AIC
## - maturity   1     36.51 1849.0 108.90
## <none>                   1812.4 110.44
## - prosecute  1    276.03 2088.5 111.70
## - delay      1    628.01 2440.5 115.28
## - age        1   2045.16 3857.6 125.81
##
## Step:  AIC=108.9
## qualityct ~ age + delay + prosecute
##
##             Df Sum of Sq    RSS    AIC
## <none>                   1849.0 108.90
## - prosecute  1    243.82 2092.8 109.75
## - delay      1    591.82 2440.8 113.28
## - age        1   2273.37 4122.3 125.34
```

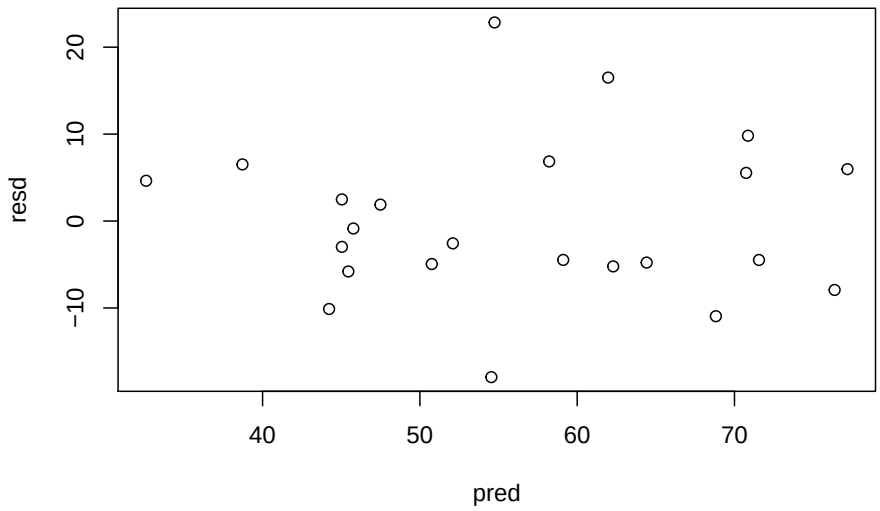
```
##
## Call:
## lm(formula = qualityct ~ age + delay + prosecute)
##
## Coefficients:
##  (Intercept)       ageA8-9         delay  prosecuteYes
##       53.408        20.190        -0.204         6.645
```

```r
quality1_reg <- lm(qualityct ~ age + delay + prosecute)

pred <- predict(quality1_reg)
resd <- residuals(quality1_reg)
```

```
qqnorm(resd)
```




```
plot(pred, resd)
```




```
# etc,etc
detach(quality)
```

Maturity and coherence are highly correlated, so coherence is dropped from regression. Backwards search also drops maturity; residual plots look okay. You need to interpret the estimated regression coefficients.

### Exercise 4.3

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
```

```
attach(quartet)

# (Exceptionally) fit regressions first:

lm1 <- lm(y1 ~ x1); summary(lm1)
```

```
##
## Call:
## lm(formula = y1 ~ x1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.92127 -0.45577 -0.04136  0.70941  1.83882
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.0001     1.1247   2.667  0.02573 *
## x1            0.5001     0.1179   4.241  0.00217 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.237 on 9 degrees of freedom
## Multiple R-squared:  0.6665, Adjusted R-squared:  0.6295
## F-statistic: 17.99 on 1 and 9 DF,  p-value: 0.00217
```

```
lm2 <- lm(y2 ~ x2); summary(lm2)
```

```
##
## Call:
## lm(formula = y2 ~ x2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9009 -0.7609  0.1291  0.9491  1.2691
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.001      1.125   2.667  0.02576 *
## x2             0.500      0.118   4.239  0.00218 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.237 on 9 degrees of freedom
## Multiple R-squared:  0.6662, Adjusted R-squared:  0.6292
## F-statistic: 17.97 on 1 and 9 DF,  p-value: 0.002179
```

```
lm3 <- lm(y3 ~ x3); summary(lm3)
```

```
##
## Call:
## lm(formula = y3 ~ x3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.1586 -0.6146 -0.2303  0.1540  3.2411
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.0025     1.1245   2.670  0.02562 *
## x3            0.4997     0.1179   4.239  0.00218 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.236 on 9 degrees of freedom
## Multiple R-squared:  0.6663, Adjusted R-squared:  0.6292
## F-statistic: 17.97 on 1 and 9 DF,  p-value: 0.002176
```

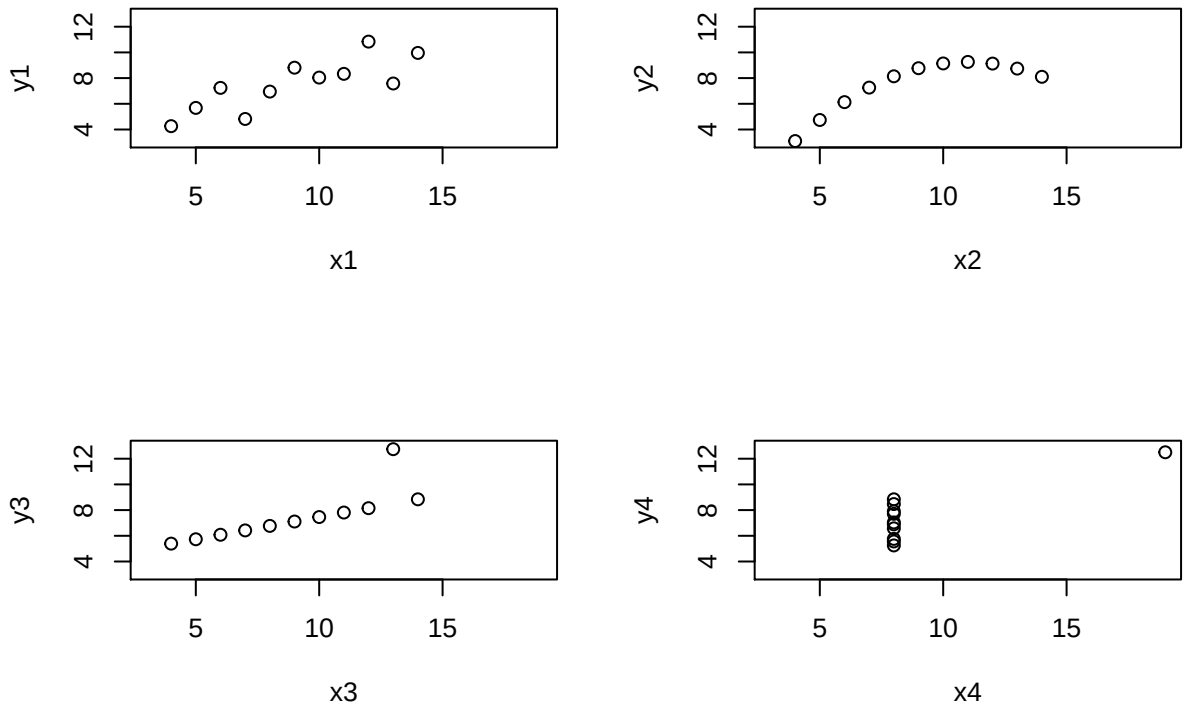
```
lm4 <- lm(y4 ~ x4); summary(lm4)
```

```
##
## Call:
## lm(formula = y4 ~ x4)
##
## Residuals:
```

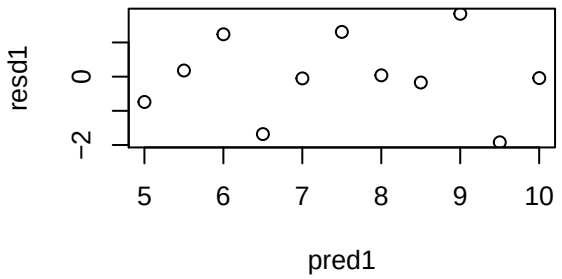
```
##    Min     1Q Median     3Q    Max
## -1.751 -0.831  0.000  0.809  1.839
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.0017     1.1239   2.671  0.02559 *
## x4            0.4999     0.1178   4.243  0.00216 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.236 on 9 degrees of freedom
## Multiple R-squared:  0.6667, Adjusted R-squared:  0.6297
## F-statistic:    18 on 1 and 9 DF,  p-value: 0.002165
```

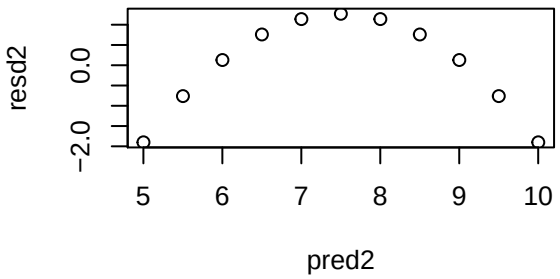
We get the same results on all data sets! How about some graphs?

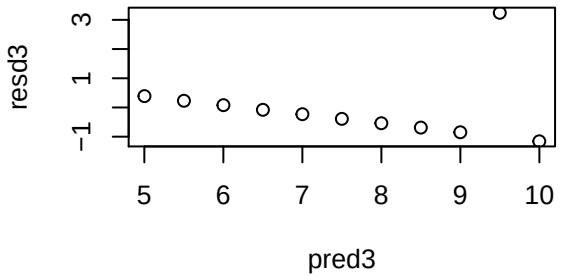
```r
par(mfrow = c(2,2))
plot(y1 ~ x1, xlim = c(3, 19), ylim = c(3, 13))
plot(y2 ~ x2, xlim = c(3, 19), ylim = c(3, 13))
plot(y3 ~ x3, xlim = c(3, 19), ylim = c(3, 13))
plot(y4 ~ x4, xlim = c(3, 19), ylim = c(3, 13))
```

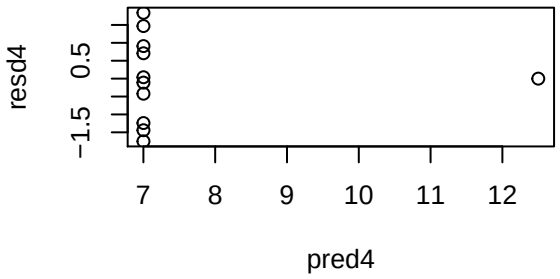



```r
pred1 <- predict(lm1)
pred2 <- predict(lm2)
pred3 <- predict(lm3)
pred4 <- predict(lm4)

resd1 <- residuals(lm1)
resd2 <- residuals(lm2)
resd3 <- residuals(lm3)
resd4 <- residuals(lm4)

plot(pred1, resd1)
plot(pred2, resd2)
plot(pred3, resd3)
plot(pred4, resd4)
```




```
detach(quartet)
```

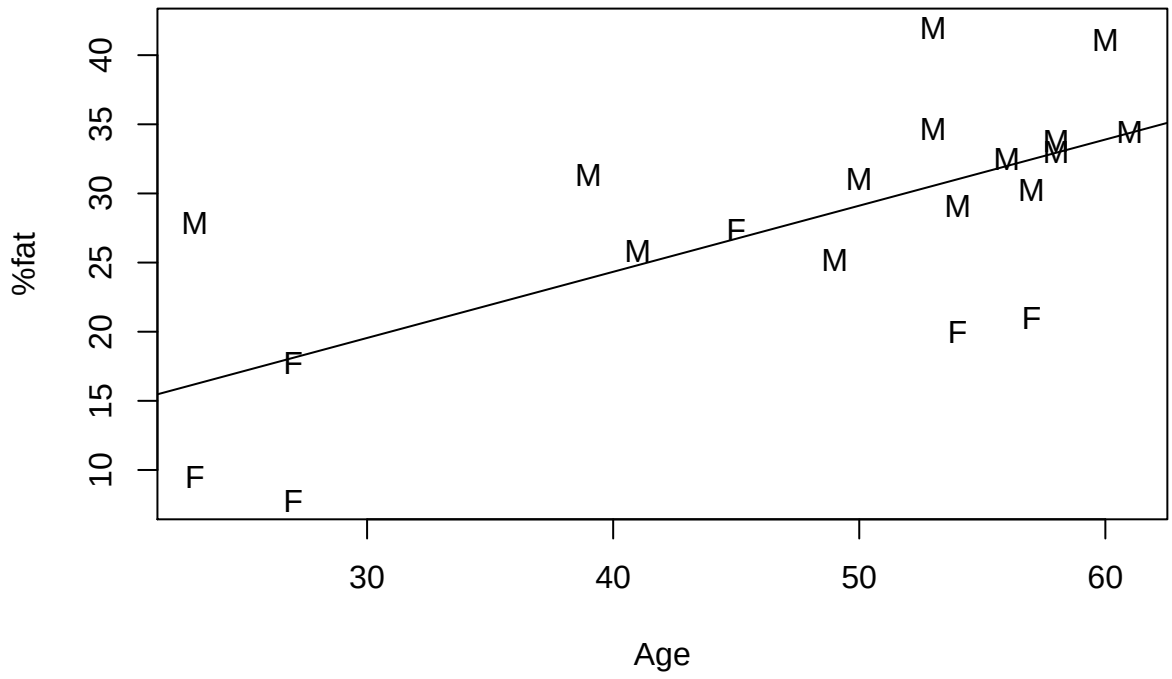
### Exercise 4.4

```
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
```

```
attach(fat)
```

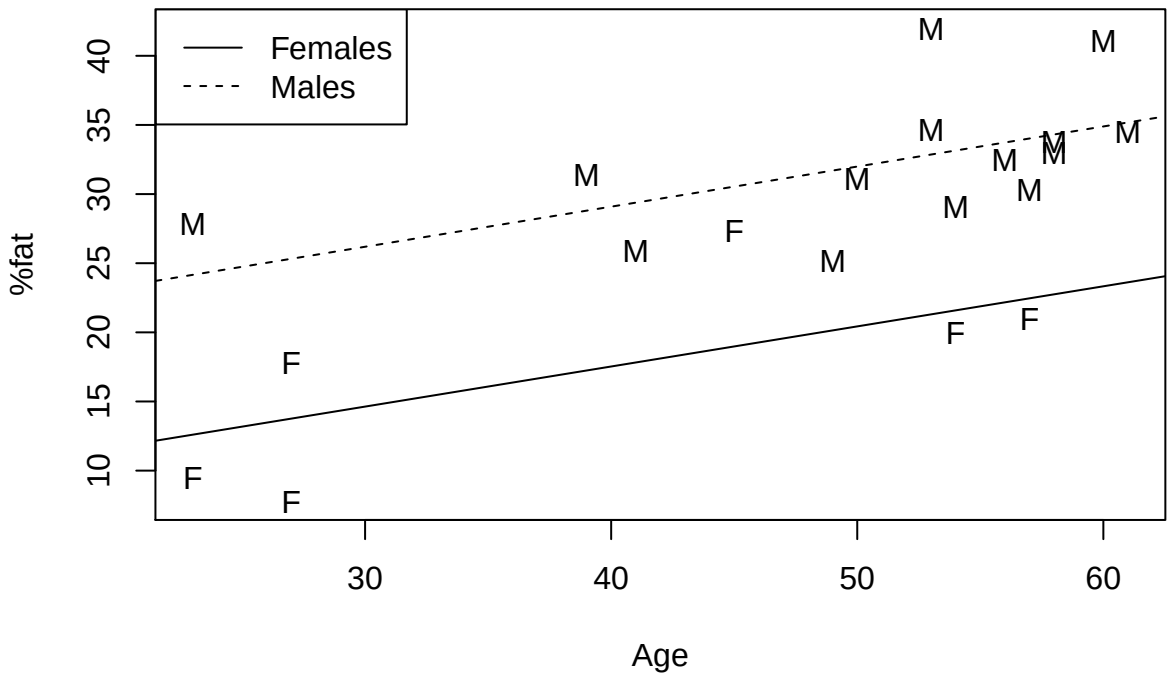
```r
fat_reg <- lm(Pcfat ~ Age)
plot(Age, Pcfat, xlab = "Age", ylab = "%fat", type = "n")
text(Age, Pcfat, labels = Sex)
abline(fat_reg)
```




```r
summary(lm(Pcfat ~ Age + Sex))
```

```
##
## Call:
## lm(formula = Pcfat ~ Age + Sex)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.580 -3.238 -1.146  2.825  9.054
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.93107    4.09520   1.448 0.165727
## Age          0.29147    0.09327   3.125 0.006165 **
## SexM        11.56679    2.54447   4.546 0.000286 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.681 on 17 degrees of freedom
## Multiple R-squared:  0.7582, Adjusted R-squared:  0.7298
## F-statistic: 26.65 on 2 and 17 DF,  p-value: 5.744e-06
```

```r
plot(Age, Pcfat, xlab = "Age", ylab = "%fat", type = "n")
text(Age, Pcfat, labels = Sex)
# Use figures from summary to find slope and intercepts of lines for men and women
# 5.93+11.56=17.49
abline(a = 5.93, b = 0.29)
abline(a = 17.49, b = 0.29, lty = 2)
legend("topleft", c("Females", "Males"), lty = 1:2)
```

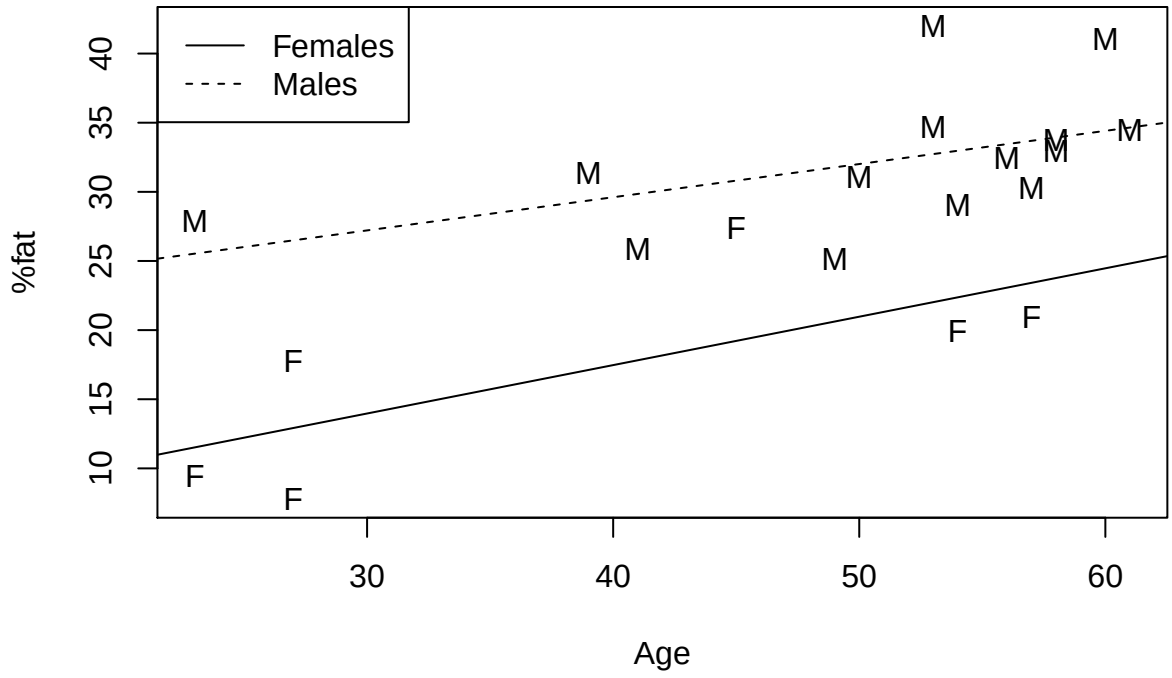



```r
summary(lm(Pcfat ~ Age * Sex))
```

```
##
## Call:
## lm(formula = Pcfat ~ Age * Sex)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.676 -2.895 -1.026  2.011  9.164
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.4740     5.8489   0.594   0.5608
## Age           0.3547     0.1420   2.498   0.0238 *
## SexM         16.6376     8.8436   1.881   0.0783 .
## Age:SexM     -0.1147     0.1912  -0.600   0.5571
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 4.772 on 16 degrees of freedom
## Multiple R-squared:  0.7635, Adjusted R-squared:  0.7192
## F-statistic: 17.22 on 3 and 16 DF,  p-value: 2.901e-05
```

```
plot(Age, Pcfat, xlab = "Age", ylab = "%fat", type = "n")
text(Age, Pcfat, labels = Sex)
abline(a = 3.47, b = 0.35)
abline(a = 20.01, b = 0.24, lty = 2)
legend("topleft", c("Females", "Males"), lty = 1:2)
```




```
detach(fat)
```

### Exercise 4.5

```
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, 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)
```

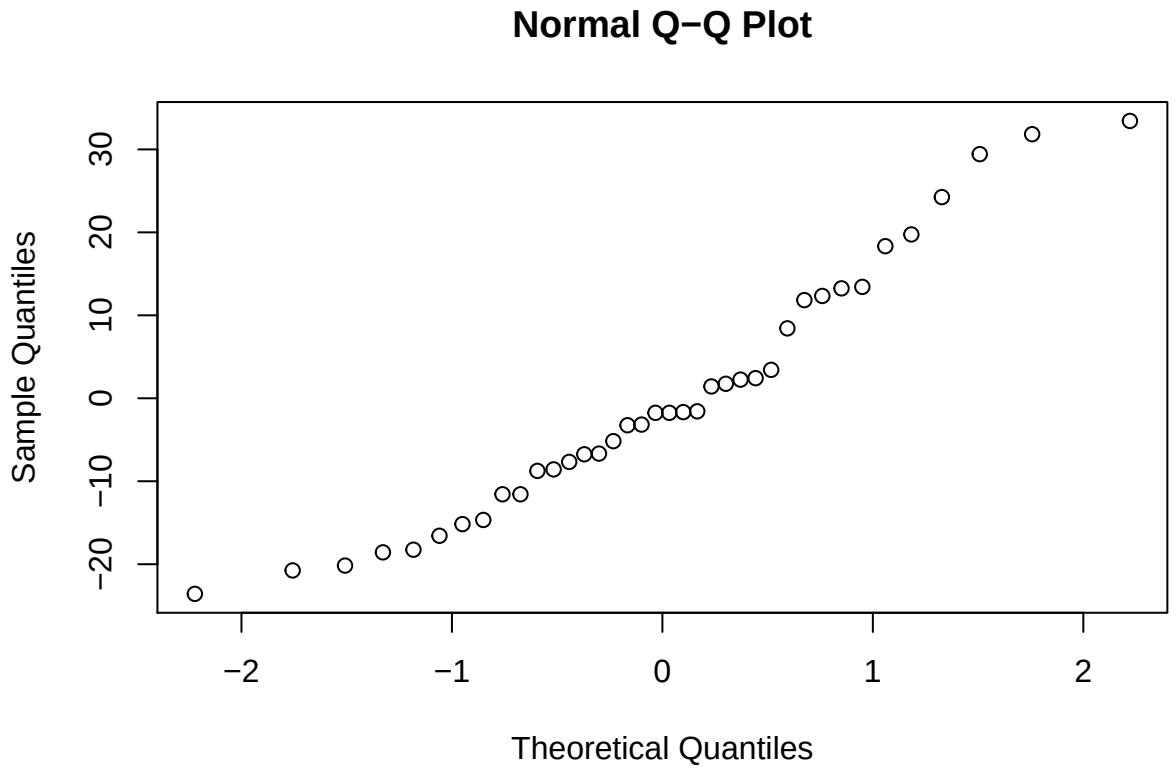
```
##    History Smoking Bloodp
## 33  Absent      CS    140
## 34  Absent      CS    125
## 35  Absent      CS    123
## 36  Absent      CS    108
## 37  Absent      CS    113
## 38  Absent      CS    160
```

```
attach(blood)
blood_reg <- lm(Bloodp ~ History * Smoking)
summary(blood_reg)
```

```
##
## Call:
## lm(formula = Bloodp ~ History * Smoking)
##
## Residuals:
##    Min     1Q Median     3Q    Max
```

```
## -23.57 -10.87  -1.75  10.98  33.43
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                126.571      6.151  20.579   <2e-16 ***
## HistoryAbsent              -14.821      8.422  -1.760    0.088 .
## SmokingExS                  -4.905      9.053  -0.542    0.592
## SmokingCS                    5.000      8.698   0.575    0.569
## HistoryAbsent:SmokingExS     1.405     13.464   0.104    0.918
## HistoryAbsent:SmokingCS     11.417     12.365   0.923    0.363
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.27 on 32 degrees of freedom
## Multiple R-squared:  0.2357, Adjusted R-squared:  0.1163
## F-statistic: 1.973 on 5 and 32 DF,  p-value: 0.1096
```

```r
qqnorm(residuals(blood_reg))
```




```r
detach(blood)
```

### Exercise 4.6

```
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
```

```
attach(oestrogen)
oestrogen_reg <- lm(Depression ~ Treatment + BL1 + BL2)
summary(oestrogen_reg)
```

```
##
## Call:
## lm(formula = Depression ~ Treatment + BL1 + BL2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.3668 -1.6591 -0.0062  1.9088  6.8615
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       6.2774     5.5297   1.135    0.272
## TreatmentActive  -2.4768     1.7072  -1.451    0.165
## BL1               0.4436     0.3650   1.215    0.241
## BL2              -0.2892     0.2927  -0.988    0.337
##
## Residual standard error: 3.619 on 17 degrees of freedom
## Multiple R-squared:  0.2441, Adjusted R-squared:  0.1107
## F-statistic:  1.83 on 3 and 17 DF,  p-value: 0.1799
```

```
detach(oestrogen)
#
treatCI <- c(-2.477 - 2*1.707, -2.477 + 2*1.707)
treatCI
```

```
## [1] -5.891  0.937
```

CI contains the value 0, so there is no evidence of a treatment effect.