Multivariate Analysis for the Behavioral Sciences, Second Edition (Chapman and Hall/CRC, 2019) **Solutions to Exercises of Chapter 6: Applying Logistic Regression**

Kimmo Vehkalahti and Brian S. Everitt

10 November 2018

Solutions

Exercise 6.1

```
womensrole <- structure(list(
    education = c(0L, 1L, 2L, 3L, 4L, 5L, 6L, 7L, 8L, 9L, 10L, 11L, 12L, 13L, 14L, 15L,
                  16L, 17L, 18L, 19L, 20L, 0L, 1L, 2L, 3L, 4L, 5L, 6L, 7L, 8L, 9L, 10L,
                  11L, 12L, 13L, 14L, 15L, 16L, 17L, 18L, 19L, 20L),
  sex = structure(c(1L, 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, 2L, 2L, 2L),
                  .Label = c("Male", "Female"), class = "factor"),
  agree = c(4, 2, 4, 6, 5, 13, 25, 27, 75, 29, 32, 36, 115, 31, 28, 9, 15, 3, 1, 2, 3, 4,
            1, 0, 6, 10, 14, 17, 26, 91, 30, 55, 50, 190, 17, 18, 7, 13, 3, 0, 1, 2),
  disagree = c(2, 0, 0, 3, 5, 7, 9, 15, 49, 29, 45, 59, 245, 70, 79, 23, 110, 29, 28, 13, 20,
               2, 0, 0, 1, 0, 7, 5, 16, 36, 35, 67, 62, 403, 92, 81, 34, 115, 28, 21, 2, 4)),
.Names = c("education", "sex", "agree", "disagree"),

row.names = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10", "11", "12", "13", "14", "15",
"16", "17", "18", "19", "20", "21", "22", "23", "24", "25", "26", "27", "28", "29", "30", "31",
"32", "33", "34", "35", "36", "37", "38", "39", "40", "41", "42"), class = "data.frame")

head(womensrole)
```

```
##   education  sex agree disagree
## 1         0 Male     4        2
## 2         1 Male     2        0
## 3         2 Male     4        0
## 4         3 Male     6        3
## 5         4 Male     5        5
## 6         5 Male    13        7
```

```
attach(womensrole)
#main effects model
womensrole_glm <- glm(cbind(agree,disagree) ~ sex + education, family = "binomial")
summary(womensrole_glm)
```

```
##
## Call:
## glm(formula = cbind(agree, disagree) ~ sex + education, family = "binomial")
```

```
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.72544  -0.86302  -0.06525   0.84340   3.13315
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.50937    0.18389  13.646   <2e-16 ***
## sexFemale   -0.01145    0.08415  -0.136    0.892
## education   -0.27062    0.01541 -17.560   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 451.722  on 40  degrees of freedom
## Residual deviance:  64.007  on 38  degrees of freedom
## AIC: 208.07
##
## Number of Fisher Scoring iterations: 4
```

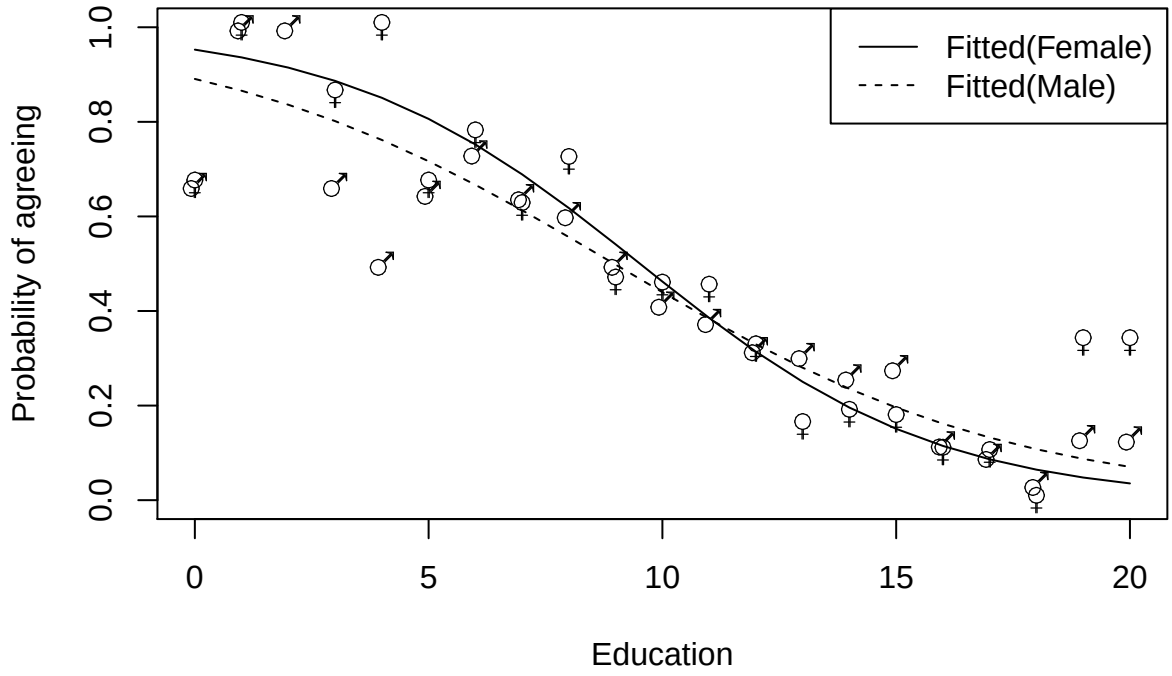
```
#interaction model
womensrole_glm1 <- glm(cbind(agree,disagree) ~ sex * education, family = "binomial")
summary(womensrole_glm1)
```

```
##
## Call:
## glm(formula = cbind(agree, disagree) ~ sex * education, family = "binomial")
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.39097  -0.88062   0.01532   0.72783   2.45262
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)          2.09820    0.23550   8.910  < 2e-16 ***
## sexFemale            0.90474    0.36007   2.513  0.01198 *
## education           -0.23403    0.02019 -11.592  < 2e-16 ***
## sexFemale:education -0.08138    0.03109  -2.617  0.00886 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 451.722  on 40  degrees of freedom
## Residual deviance:  57.103  on 37  degrees of freedom
## AIC: 203.16
##
## Number of Fisher Scoring iterations: 4
```

```
#interaction significant

fitted <- predict(womensrole_glm1, type = "response")
fittedF <- fitted[sex == "Female"]
fittedM <- fitted[sex != "Female"]
```

```
pobsv <- agree / (agree + disagree)
plot(education, pobsv, type = "n", xlab = "Education", ylab = "Probability of agreeing")
text(education, pobsv, ifelse(sex == "Female", "\\VE", "\\MA"), vfont = c("serif", "plain"), cex = 1.25)
lines(education[sex == "Female"], fittedF)
lines(education[sex != "Female"], fittedM, lty = 2)
legend("topright", c("Fitted(Female)", "Fitted(Male)"), lty = 1:2)
```




```
detach(womensrole)
```

The interaction shows that, for fewer years of education, women have a higher probability of agreeing with the statement than men, but when the years of education exceed about 10, then this situation reverses.

Exercise 6.3

```
accidents <- structure(list(

  weight = structure(c(1L, 1L, 1L, 1L, 2L, 2L, 2L, 2L),
                     .Label = c("Small", "Standard"), class = "factor"),

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

  type = structure(c(1L, 2L, 1L, 2L, 1L, 2L, 1L, 2L),
                   .Label = c("Collision", "Rollover"), class = "factor"),

  severely = c(150, 112, 23, 80, 1022, 404, 161, 265),

  nseverely = c(350, 60, 26, 19, 1878, 148, 111, 22)),

.Names = c("weight", "eject", "type", "severely", "nseverely"),
row.names = c(NA, -8L), class = "data.frame")

head(accidents)
```

```
##     weight eject      type severely nseverely
## 1    Small    No Collision      150       350
## 2    Small    No  Rollover      112        60
## 3    Small   Yes Collision       23        26
## 4    Small   Yes  Rollover       80        19
## 5 Standard    No Collision     1022      1878
## 6 Standard    No  Rollover      404       148
```

```
attach(accidents)
accident_glm <- glm(cbind(severely,nseverely) ~ weight + eject + type, family = "binomial")
summary(accident_glm)
```

```
##
## Call:
## glm(formula = cbind(severely, nseverely) ~ weight + eject + type,
##     family = "binomial")
##
## Deviance Residuals:
##       1        2        3        4        5        6        7        8
##  0.9445  -0.4631  -0.7447  -1.1072  -0.1301  -0.3219  -0.4458   2.0255
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.94006    0.08284 -11.348  < 2e-16 ***
## weightStandard  0.33667    0.08612   3.909 9.26e-05 ***
## ejectYes        1.03036    0.09891  10.417  < 2e-16 ***
## typeRollover    1.63859    0.08281  19.787  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 737.894  on 7  degrees of freedom
## Residual deviance:   7.309  on 4  degrees of freedom
## AIC: 61.976
##
## Number of Fisher Scoring iterations: 4
```

```
detach(accidents)
```

Exercise 6.4

```
statistics <- structure(list(

  result = structure(c(1L, 1L, 2L, 1L, 2L, 2L, 2L, 2L, 2L, 2L,
                       2L, 2L, 2L, 2L, 2L, 1L, 2L, 1L, 1L, 2L),
                     .Label = c("Fail", "Pass"), class = "factor"),

  test = c(525, 533, 545, 582, 581, 576, 572, 609, 559, 543,
           576, 525, 574, 582, 574, 471, 595, 557, 557, 584),

  grade = structure(c(2L, 3L, 2L, 1L, 3L, 4L, 2L, 1L, 3L, 4L,
                      2L, 1L, 4L, 3L, 2L, 2L, 3L, 1L, 1L, 1L),
                    .Label = c("A", "B", "C", "D"), class = "factor")),

.Names = c("result", "test", "grade"), row.names = c(NA, -20L), class = "data.frame")

head(statistics)
```

```
##   result test grade
## 1   Fail  525     B
## 2   Fail  533     C
## 3   Pass  545     B
## 4   Fail  582     A
## 5   Pass  581     C
## 6   Pass  576     D
```

```
tail(statistics)
```

```
##    result test grade
## 15   Pass  574     B
## 16   Fail  471     B
## 17   Pass  595     C
## 18   Fail  557     A
## 19   Fail  557     A
## 20   Pass  584     A
```

```
attach(statistics)
statistics_glm <- glm(result ~ test + grade, family = "binomial")
summary(statistics_glm)
```

```
##
## Call:
## glm(formula = result ~ test + grade, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4921  -0.4241   0.3154   0.4158   2.2950
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -32.71260   16.98723  -1.926   0.0541 .
## test           0.05744    0.02974   1.931   0.0535 .
## gradeB         2.35718    1.77681   1.327   0.1846
## gradeC         1.79580    1.67943   1.069   0.2849
```

```
## gradeD        19.11562 3487.42366   0.005   0.9956
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 24.435  on 19  degrees of freedom
## Residual deviance: 14.382  on 15  degrees of freedom
## AIC: 24.382
##
## Number of Fisher Scoring iterations: 17
```

```
#find prediction of pass and fail from model and cross tab with actual result
pred <- predict(statistics_glm, type = "response")
pred[pred > 0.5] <- "P"
pred[pred <= 0.5] <- "F"

table(result, pred)
```

```
##       pred
## result  F  P
##   Fail  5  1
##   Pass  1 13
```

```
detach(statistics)
```

Exercise 6.5

```
menstruation <- structure(list(
  age = c(11.08, 11.33, 11.58, 11.83, 12.08, 12.33, 12.58, 12.83, 13.08,
          13.33, 13.58, 14.08, 14.33, 14.58, 15.08, 15.33, 15.58, 17.58),

  bmens = c(2, 5, 10, 17, 16, 29, 39, 51, 47, 67, 81, 79, 90, 93, 117, 107, 92, 1049),

  n = c(120, 88, 105, 111, 100, 93, 100, 108, 99, 106, 117,
        98, 97, 100, 122, 111, 94, 1049)),

.Names = c("age", "bmens", "n"), row.names = c(NA, -18L), class = "data.frame")

menstruation
```

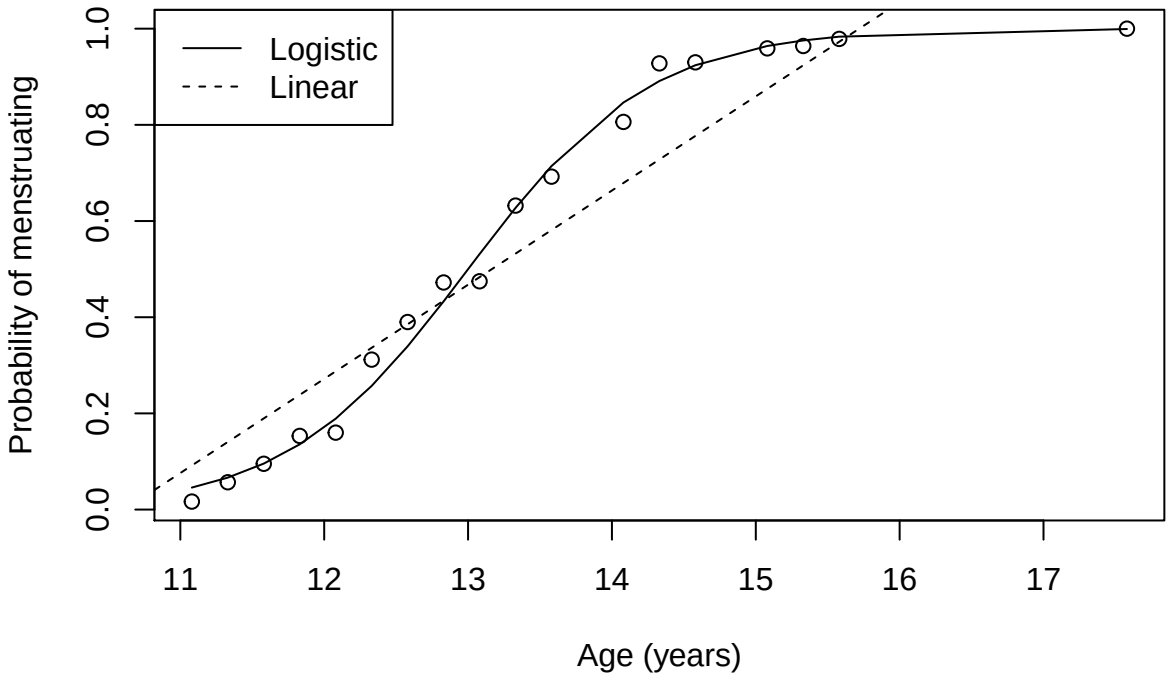
```
##      age bmens    n
## 1  11.08     2  120
## 2  11.33     5   88
## 3  11.58    10  105
## 4  11.83    17  111
## 5  12.08    16  100
## 6  12.33    29   93
## 7  12.58    39  100
## 8  12.83    51  108
## 9  13.08    47   99
## 10 13.33    67  106
## 11 13.58    81  117
## 12 14.08    79   98
## 13 14.33    90   97
## 14 14.58    93  100
## 15 15.08   117  122
## 16 15.33   107  111
## 17 15.58    92   94
## 18 17.58  1049 1049
```

```
attach(menstruation)
menstruation_reg <- glm(cbind(bmens, n-bmens) ~ age, family = "binomial")
summary(menstruation_reg)
```

```
##
## Call:
## glm(formula = cbind(bmens, n - bmens) ~ age, family = "binomial")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7447  -0.6822  -0.1522   0.7488   1.2276
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -20.56351    0.88838  -23.15   <2e-16 ***
## age           1.58172    0.06823   23.18   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1990.93  on 17  degrees of freedom
## Residual deviance:   13.72  on 16  degrees of freedom
## AIC: 88.204
##
## Number of Fisher Scoring iterations: 4
```

```
plot(age, bmens/n, xlab = "Age (years)", ylab = "Probability of menstruating")
abline(lm(bmens/n ~ age), lty = 2)
lines(age, predict(menstruation_reg, type = "response"))
legend("topleft", c("Logistic", "Linear"), lty = 1:2)
```




```
detach(menstruation)
```

This shows why the linear model is useless and the logistic model is not.