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

Examples of Chapter 5: Generalized Linear Models

Kimmo Vehkalahti and Brian S. Everitt

9 November 2018

Contents

Examples

Table 5.1: The Colonic Polyps Data Giving the Number of Polyps for Two Treatments

```
POLYPS <- read.table("data/polyps.txt", header = TRUE, sep = '\t')

# to make sure that the factor is coded as it should be:
POLYPS <- within(POLYPS,
      Treatment <- factor(Treatment, levels = c("placebo", "drug"))
)
str(POLYPS)
```

```
## 'data.frame':    20 obs. of  3 variables:
##  $ Number   : int  63 2 28 17 61 1 7 15 44 25 ...
##  $ Treatment: Factor w/ 2 levels "placebo","drug": 1 2 1 2 1 2 1 1 1 2 ...
##  $ Age      : int  20 16 18 22 13 23 34 50 19 17 ...
```

```
# list the whole (small) data:
POLYPS
```

```
##    Number Treatment Age
## 1      63   placebo  20
## 2       2      drug  16
## 3      28   placebo  18
## 4      17      drug  22
## 5      61   placebo  13
## 6       1      drug  23
## 7       7   placebo  34
## 8      15   placebo  50
## 9      44   placebo  19
## 10     25      drug  17
## 11      3      drug  23
## 12     28   placebo  22
## 13     10   placebo  30
## 14     40   placebo  27
## 15     33      drug  23
## 16     46   placebo  22
## 17     50   placebo  34
## 18      3      drug  23
## 19      1      drug  22
## 20      4      drug  42
```

Figure 5.1

```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
p1 <- ggplot(POLYPS, aes(x = Age, y = Number, shape = Treatment))
p2 <- p1 + geom_point(size = 3) + scale_shape_manual(values = c(2,1))
p3 <- p2 + theme_bw() + theme(panel.grid.major = element_blank(),
                              panel.grid.minor = element_blank())
p4 <- p3 + theme(legend.position = c(0.8, 0.8))
p5 <- p4 + scale_y_continuous(name = "Number of polyps") +
           scale_x_continuous(name = "Age of patient")
p5
```

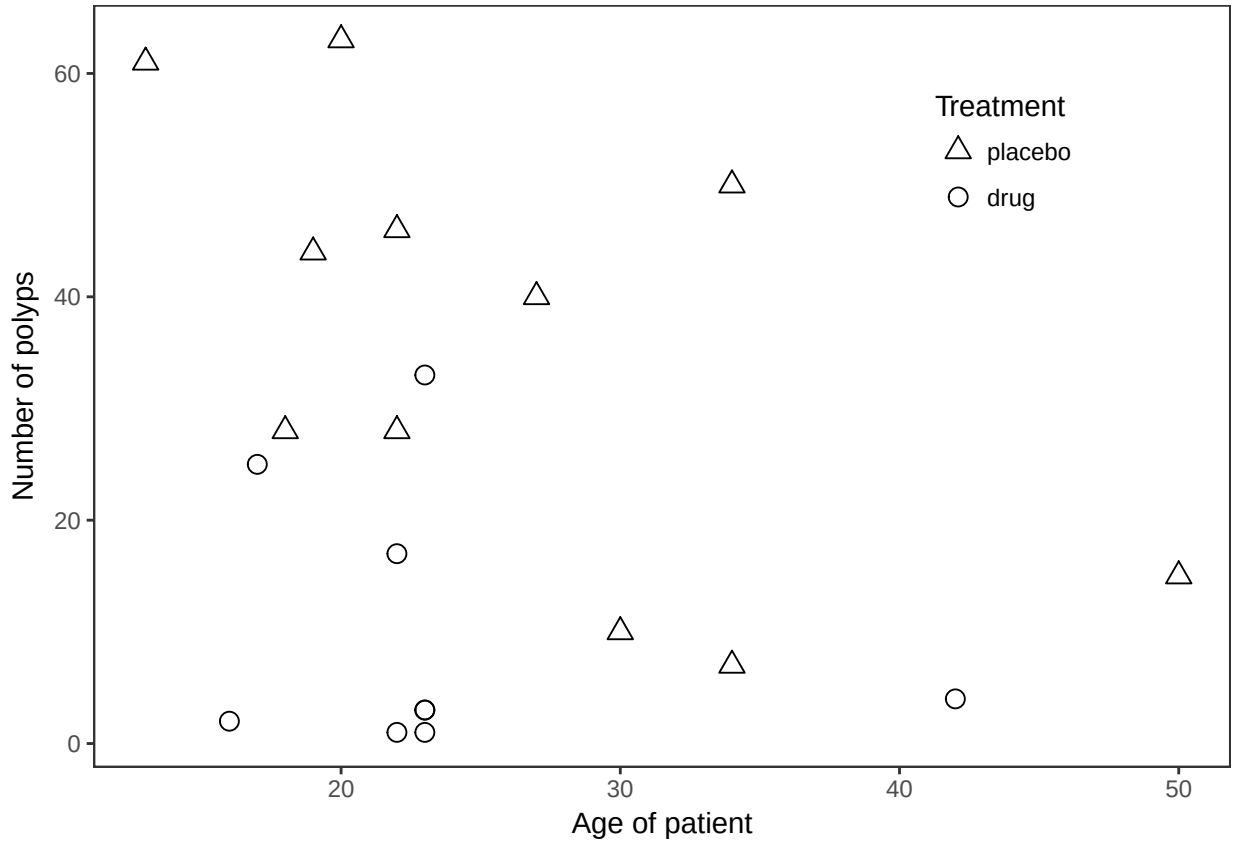

Table 5.2

```
polyps_fit1 <- glm(Number ~ Treatment + Age, data = POLYPS,
                   family = poisson(link = "log"))
summary(polyps_fit1)
```

```
##
## Call:
## glm(formula = Number ~ Treatment + Age, family = poisson(link = "log"),
##     data = POLYPS)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.2212  -3.0536  -0.1802   1.4459   5.8301
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    4.529024   0.146872   30.84  < 2e-16 ***
## Treatmentdrug -1.359083   0.117643  -11.55  < 2e-16 ***
## Age           -0.038830   0.005955   -6.52 7.02e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 378.66  on 19  degrees of freedom
## Residual deviance: 179.54  on 17  degrees of freedom
## AIC: 273.88
##
## Number of Fisher Scoring iterations: 5
```

Table 5.3

```
polyps_fit2 <- glm(Number ~ Treatment + Age, data = POLYPS,
                   family = quasipoisson(link = "log"))
summary(polyps_fit2)
```

```
##
## Call:
## glm(formula = Number ~ Treatment + Age, family = quasipoisson(link = "log"),
##     data = POLYPS)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.2212  -3.0536  -0.1802   1.4459   5.8301
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.52902    0.48106   9.415 3.72e-08 ***
## Treatmentdrug -1.35908    0.38533  -3.527  0.00259 **
## Age           -0.03883    0.01951  -1.991  0.06284 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 10.72805)
##
##     Null deviance: 378.66  on 19  degrees of freedom
## Residual deviance: 179.54  on 17  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```