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

Solutions to Exercises of Chapter 5: Generalized Linear Models

Kimmo Vehkalahti and Brian S. Everitt

10 November 2018

Exercises

Exercise 5.1

```
bladder <- read.table("data/bladder.txt", header = T)
bladder <- within(bladder,
             X <- factor(X, labels = c("< 3cm", "> 3cm"))
)
bladder
```

```
##    Time     X n
## 1    11 > 3cm 1
## 2     2 < 3cm 1
## 3     3 < 3cm 1
## 4     6 < 3cm 1
## 5     8 < 3cm 1
## 6     9 < 3cm 1
## 7    10 < 3cm 1
## 8    11 < 3cm 1
## 9    13 < 3cm 1
## 10   14 < 3cm 1
## 11   16 < 3cm 1
## 12   21 < 3cm 1
## 13   22 < 3cm 1
## 14   24 < 3cm 1
## 15   26 < 3cm 1
## 16   27 < 3cm 1
## 17    7 < 3cm 2
## 18   13 < 3cm 2
## 19   15 < 3cm 2
## 20   18 < 3cm 2
## 21   23 < 3cm 2
## 22   20 < 3cm 3
## 23   24 < 3cm 4
## 24    1 > 3cm 1
## 25    5 > 3cm 1
## 26   17 > 3cm 1
## 27   18 > 3cm 1
## 28   25 > 3cm 1
## 29   18 > 3cm 2
```

```
## 30   25 > 3cm 2
## 31    4 > 3cm 3
## 32   19 > 3cm 4
```

```
# (solution still in process - check that it corresponds to the comments below!)
```

The estimated model is

$$\log \lambda = -2.339 + 0.229x$$

So for smaller tumors ($x = 0$), the estimated (baseline) rate is exp(-2.339) = 0.096 and for larger tumors ($x = 1$), the estimated rate is exp(-2.339 + 0.229) = 0.12. The rate for larger tumors is estimated as 0.12 / 0.096 = 1.25 times the rate for smaller tumors. In terms of waiting times between recurrences the means are 1 / 0.096 = 10.42 months for smaller tumors and 1 / 0.12 = 8.33 months for larger tumors. But the regression coefficient for the dummy variable coding tumor size is seen to be non-significant so the data give no evidence that rates or waiting times for large and small tumors are different. This becomes apparent if we construct a confidence interval for the rate for larger tumors from the confidence limits given in the output as [exp(-2.339 - 0.371), exp(-2.339 + 0.829)], i.e., [0.067, 0.221]. This interval contains the rate for smaller tumors. There is no evidence that size of primary tumor is associated with number of recurrent tumors.

Exercise 5.2

```
CHDrisks <- read.table("data/CHDrisks.txt", header = T)
CHDrisks <- within(CHDrisks,
            {
             Smoking <- factor(Smoking, labels = c("non-smoker", "1-10", "11-20", "20+"))
             Press <- factor(Press, labels = c("< 140", ">= 140"))
             Behavior <- factor(Behavior, labels = c("Type B", "Type A"))
            }
)
str(CHDrisks)
```

```
## 'data.frame':    16 obs. of  5 variables:
##  $ Years   : num  5268 2542 1141 615 4451 ...
##  $ Smoking : Factor w/ 4 levels "non-smoker","1-10",..: 1 2 3 4 1 2 3 4 1 2 ...
##  $ Press   : Factor w/ 2 levels "< 140",">= 140": 1 1 1 1 1 1 1 1 2 2 ...
##  $ Behavior: Factor w/ 2 levels "Type B","Type A": 1 1 1 1 2 2 2 2 1 1 ...
##  $ nCHD    : int  20 16 13 3 41 24 27 17 8 9 ...
```

```
# (solution still in process)
```