

uebung9.r

Janina

Sun Jul 02 13:58:54 2017

```
# Uebungsblatt 9
# Namen: Janina Schoenberger, Benjamin Weigner
# Tutorin: Gergana Stanilova
# Uebung: Mi 12-14 Uhr
```

```
# Aufgabe 19
# Fuer Wahrscheinlichkeiten die zugehoerigen Odds
p_a = 0.3
p_a/(1-p_a)
```

```
## [1] 0.4285714
```

```
p_a = 0.8
p_a/(1-p_a)
```

```
## [1] 4
```

```
p_a = 0.5
p_a/(1-p_a)
```

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

```
# Fuer Odds die zugehoerigen Wahrscheinlichkeiten
r_a = 0.1
r_a/(1+r_a)
```

```
## [1] 0.09090909
```

```
r_a = 100
r_a/(1+r_a)
```

```
## [1] 0.990099
```

```
r_a = 5
r_a/(1+r_a)
```

```
## [1] 0.8333333
```

```
# Aufgabe 20

# A
# setwd
Unfaelle <- read.table("Unfaelle.txt", header=TRUE)
head(Unfaelle)
```

```
##      Lnr Unfall Geschlecht      Beruf Alter Fahrpraxis
## 1      1      0           1 Physiker   31      12
## 2      2      0           2 Physiker   37      18
## 3      3      0           2 Physiker   38      18
## 4      4      0           2 Physiker   44      25
## 5      5      0           2 Physiker   52      33
## 6      6      0           1 Physiker   41      22
```

```
attach(Unfaelle)
# Alter&Fahrpraxis in Jahren, Geschlecht m=1 w=2, Unfall n=0 j=1

# B
# Logistisches Regressionsmodell (Modellgleichung)
# mit abhaengiger Variable Y: Unfall und
# unabhaengigen Variablen x_1: Geschlecht, x_2(1,2): Beruf, x_3: Alter, x_4: Fahrpraxis
# (Beruf hat 2 Auspraegungen -> 2 Dummy Variablen. x21=Physiker, x22=Zahnarzt)
# Logit(Y=1/X=x_i) = b_0 + b_1*x_1 + b_21*x_21 + b_22*x_22 + b_3*x_3 + b_4*x_4

# Geeignete Kodierung fuer Beruf und Geschlecht (as.factor, relevel) -> Referenzkategorie Zahnarzt
Unfaelle$Beruf=as.factor(Unfaelle$Beruf)
Unfaelle=within(Unfaelle,Beruf<-relevel(Beruf, ref="Zahnarzt"))
Unfaelle$Geschlecht = as.factor(Unfaelle$Geschlecht)
supply(Unfaelle, class)
```

```
##          Lnr      Unfall Geschlecht      Beruf      Alter Fahrpraxis
## "integer" "integer"  "factor"    "factor"  "integer" "integer"
```

```
# C
# Model M mit Intercept fuer die Haupteffekte (glm, family=binomial)
modell <- glm(Unfall~Geschlecht+Alter+Beruf+Fahrpraxis, family="binomial")
summary(modell)
```

```
##
## Call:
## glm(formula = Unfall ~ Geschlecht + Alter + Beruf + Fahrpraxis,
##      family = "binomial")
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.87398  -0.13103  -0.03759  -0.00979   2.85497
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -4.1685     3.0317  -1.375   0.1691
## Geschlecht      1.1699     0.2451   4.773 1.81e-06 ***
## Alter          0.2635     0.1546   1.704   0.0884 .
## BerufPhysiker  -0.6434     0.3024  -2.127   0.0334 *
## BerufZahnarzt   1.4881     0.2487   5.984 2.17e-09 ***
## Fahrpraxis     -0.6385     0.1574  -4.056 4.98e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
## Null deviance: 1207.21 on 3499 degrees of freedom
## Residual deviance: 602.17 on 3494 degrees of freedom
## AIC: 614.17
##
## Number of Fisher Scoring iterations: 9
```

```
# D
# Welche Berufsgruppe, welches Geschlecht hat hoeheres Unfallrisiko?
# Odds Ratio = Odds(X+1)/Odds(X)

# Odds Ratio zwischen Physikern und Zahnaerzten
#  $OR(\text{Beruf}=\text{Physiker}/\text{Beruf}=\text{Zahnarzt}) = e^{(b_{21}+b_{22})}/e^{(b_{21})} = e^{(b_{22})}$ 
#  $b_{22}$  ist Koeffizient von Variable BerufZahnarzt
exp(coefficients(modell)[5])
```

```
## BerufZahnarzt
## 4.428746
```

```
# OR > 1 -> Odds der ersten Gruppe ist kleiner
# --> Hoeheres Unfallrisiko (Unfall=1) bei Beruf Zahnarzt
# Odds Ratio zwischen Maennern und Frauen
#  $OR(\text{Geschlecht}=1/\text{Geschlecht}=2) = e^{(b_1)}$ 
#  $b_1$  ist Koeffizient von Geschlecht
exp(coefficients(modell)[2])
```

```
## Geschlecht
## 3.221624
```

```
# OR > 1 -> Odds der ersten Gruppe ist kleiner
# --> Hoeheres Unfallrisiko (Unfall=1) bei Geschlecht Frau
```

```
# E
# Kreuztabelle fuer Unfall und Geschlecht
tab <- table(Unfall, Geschlecht)
tab
```

```
##      Geschlecht
## Unfall    1    2
##      0 1374 1981
##      1   35  110
```

```
# Chi-Quadrat Test
chisq.test(tab, correct=FALSE)
```

```
##
## Pearson's Chi-squared test
##
## data:  tab
## X-squared = 16.342, df = 1, p-value = 5.288e-05
```

```
# Aus Kreuztabelle Odds Ratio zw Maennern und Frauen  
library(vcd)
```

```
## Warning: package 'vcd' was built under R version 3.3.3
```

```
## Loading required package: grid
```

```
exp(oddsratio(tab)$coefficients)
```

```
## 0:1/1:2
```

```
## 2.179851
```

```
# F  
# Vergleich Odds Ratio mit dem aus dem logistischen Regressionsmodell  
# Warum nicht gleich?  
# -> aus E: 2.18, aus D: 3.22  
# -> Bei der Kreuztabelle sind im Modell nur Unfall und Geschlecht enthalten  
# waehrend im logisitischen Regressionsmodell "modell" alle Variablen  
# enthalten sind.  
# -> Daraus ergeben sich andere Koeffizienten, mit denen die OR berechnet wird  
  
# G  
# Erweitere M zu M_int um Interaktion zwischen Geschlecht und Beruf  
modell12 <- glm(Unfall~Geschlecht+Alter+Beruf+Fahrpraxis+Geschlecht*Beruf, family="binomial")  
# Likelihood Test fuer M und M_int (Teststatistik X, p-Wert)  
library(epiDisplay)
```

```
## Warning: package 'epiDisplay' was built under R version 3.3.3
```

```
## Loading required package: foreign
```

```
## Loading required package: survival
```

```
## Warning: package 'survival' was built under R version 3.3.3
```

```
## Loading required package: MASS
```

```
## Warning: package 'MASS' was built under R version 3.3.3
```

```
## Loading required package: nnet
```

```
lrtest (modell, modell12)
```

```
## Likelihood ratio test for MLE method
```

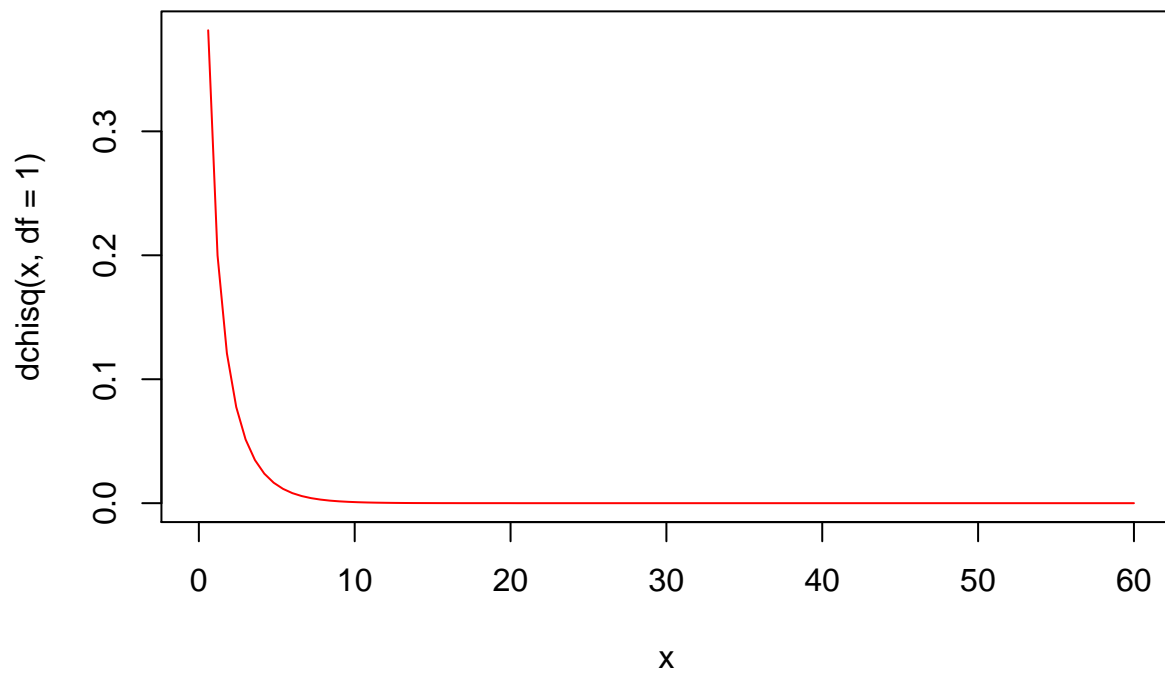
```
## Chi-squared 2 d.f. = 0.5751044 , P value = 0.7500974
```

```
# H_0 = p(modell)=p(modell12)  
# Verteilung: Chi Quadrat Verteilt  
# Plot Dichte der Verteilung. Markiere X und Grenzen des Ablehnungsbereichs fuer alpha=0.05  
X <- anova(modell,modell12)  
summary(X)
```

```
##      Resid. Df      Resid. Dev      Df      Deviance
## Min.   :3492    Min.   :601.6    Min.   :2    Min.   :0.5751
## 1st Qu.:3492    1st Qu.:601.7    1st Qu.:2    1st Qu.:0.5751
## Median :3493    Median :601.9    Median :2    Median :0.5751
## Mean   :3493    Mean   :601.9    Mean   :2    Mean   :0.5751
## 3rd Qu.:3494    3rd Qu.:602.0    3rd Qu.:2    3rd Qu.:0.5751
## Max.   :3494    Max.   :602.2    Max.   :2    Max.   :0.5751
##                                     NA's   :1    NA's   :1
```

```
x <- X$Deviance[2]
curve( dchisq(x, df=1), col='red', main = "Chi-Square Density Graph",
      from=0,to=60)
```

Chi-Square Density Graph



```
# Schwellenwert
qchisq(X$Deviance,df=1)
```

```
## [1]      NA 0.6367349
```

```
# Dichte
dchisq(X$Deviance,df=1)
```

```
## [1]      NA 0.3945976
```

```
# p-Wert
1-pchisq(X$Deviance,df=1)
```

```
## [1] NA 0.4482377
```

```
# Entscheidung fuer ein Modell: Die Interaktion zwischen Geschlecht und Beruf bringt  
# keine neuen Informationen fuer das Modell. Deshalb sollte sich fuer das erste Modell  
# entschieden werden
```

```
# H
# Fuer M Wkt und Odds fuer Unfall einer 25-jaehrigen Physikerin mit 7 Jahren Fahrpraxis
nd <- data.frame("Geschlecht" = 2, "Beruf"=as.factor("Physiker"), "Alter"=25, "Fahrpraxis"=7)
p_a <- predict(modell, nd, type="response")
p_a
```

```
## 1
## 0.4124044
```

```
# -> Wahrscheinlichkeit fuer Unfall fuer die Person = ca. 41%
p_a/(1-p_a)
```

```
## 1
## 0.7018508
```

```
# -> Odds fuer Unfall = 1:0.7
```

```
# I
# Fuer welche Individuen aus dem Datensatz sagt das Modell die hoechste und niedrigste
# unfallwahrscheinlichkeit voraus? Wie hoch sind die Wahrscheinlichkeiten?
```

```
Unfaelle <- read.table("Unfaelle.txt", header=TRUE)
p = c()
for (i in 1:(length(Unfaelle))){
  newdata <- data.frame(Unfaelle[i,c(3,4,5,6)])
  p[i] = predict(modell, newdata, type="response")
}
p
```

```
## [1] 4.167501e-02 1.455403e-02 1.885881e-02 1.068934e-03 5.328551e-05
## [6] 1.021974e-03
```

```
# Hoechste Unfallwahrscheinlichkeit:
max(p)
```

```
## [1] 0.04167501
```

```
# bei Individuum Nr.
which(p==max(p))
```

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

```
# Niedrigste Unfallwahrscheinlichkeit:  
min(p)
```

```
## [1] 5.328551e-05
```

```
# bei Individuum Nr.  
which(p==min(p))
```

```
## [1] 5
```

```
detach(Unfaelle)
```