4050_A4_Q2.R

SOPH

Sun Mar 26 18:30:35 2017

```r
library(faraway)
```

```
## Warning: package 'faraway' was built under R version 3.2.3
```

```r
data(death)
attach(death)

ct<-xtabs(y~penalty+defend, death)
ct.cond<-xtabs(y~penalty+defend+victim, death)

# Combined Data
freqyb = ct[2,1]/(ct[1,1]+ct[2,1])
freqyw = ct[2,2]/(ct[1,2]+ct[2,2])
freqnb = ct[1,1]/(ct[1,1]+ct[2,1])
freqnw = ct[1,2]/(ct[1,2]+ct[2,2])

penalty = c('no','yes')
black.c = c(freqnb,freqyb)
white.c = c(freqnw,freqyw)
combined<-data.frame(penalty,black.c,white.c)

# Conditional on Victim Race
# Victim White
ct.vw<-xtabs(y~penalty+defend, subset=(victim == 'w'), data = death)

freqyb.vw = ct.vw[2,1]/(ct.vw[1,1]+ct.vw[2,1])
freqyw.vw = ct.vw[2,2]/(ct.vw[1,2]+ct.vw[2,2])
freqnb.vw = ct.vw[1,1]/(ct.vw[1,1]+ct.vw[2,1])
freqnw.vw = ct.vw[1,2]/(ct.vw[1,2]+ct.vw[2,2])

black.vw = c(freqnb.vw,freqyb.vw)
white.vw = c(freqnw.vw,freqyw.vw)
victim.white<-data.frame(penalty,black.vw,white.vw)

# Victim Black
ct.vb<-xtabs(y~penalty+defend, subset=(victim == 'b'), data = death)

freqyb.vb = ct.vb[2,1]/(ct.vb[1,1]+ct.vb[2,1])
freqyw.vb = ct.vb[2,2]/(ct.vb[1,2]+ct.vb[2,2])
freqnb.vb = ct.vb[1,1]/(ct.vb[1,1]+ct.vb[2,1])
freqnw.vb = ct.vb[1,2]/(ct.vb[1,2]+ct.vb[2,2])

black.vb = c(freqnb.vb,freqyb.vb)
white.vb = c(freqnw.vb,freqyw.vb)
victim.black<-data.frame(penalty,black.vb,white.vb)

combined
```

```
##   penalty   black.c white.c
## 1      no 0.8975904 0.88125
## 2     yes 0.1024096 0.11875
```

victim.white

```
##   penalty  black.vw  white.vw
## 1      no 0.8253968 0.8741722
## 2     yes 0.1746032 0.1258278
```

victim.black

```
##   penalty   black.vb white.vb
## 1      no 0.94174757        1
## 2     yes 0.05825243        0
```

ct.cond

```
## , , victim = b
##
##        defend
## penalty   b   w
##     no   97   9
##     yes   6   0
##
## , , victim = w
##
##        defend
## penalty   b   w
##     no   52 132
##     yes  11  19
```

summary(ct)

```
## Call: xtabs(formula = y ~ penalty + defend, data = death)
## Number of cases in table: 326
## Number of factors: 2
## Test for independence of all factors:
##  Chisq = 0.22145, df = 1, p-value = 0.6379
```

summary(ct.cond)

```
## Call: xtabs(formula = y ~ penalty + defend + victim, data = death)
## Number of cases in table: 326
## Number of factors: 3
## Test for independence of all factors:
##  Chisq = 122.4, df = 4, p-value = 1.642e-25
```

```r
# This is Simpson's paradox because the observed trend within
# conditional grouped data is opposite than the trend observed
# in marginal data.

# 2b) The most appropriate model would be a conditional independence
# model.

# 2c) Fit a binomial model
ww<-subset(death, defend =='w' & victim =='w')
bb<-subset(death, defend =='b' & victim =='b')
wb<-subset(death, defend =='w' & victim == 'b')
bw<-subset(death, defend =='b' & victim == 'w')
penalty.y<-c(ww$y[1], wb$y[1], bw$y[1], bb$y[1])
penalty.n<-c(ww$y[2], wb$y[2], bw$y[2], bb$y[2])
total<-(penalty.y+penalty.n)

Victim<-c(0,1,0,1)
Defend<-c(0,0,1,1)

DP<-data.frame(penalty.y,total,Victim,Defend)

binmod<-glm(cbind(penalty.y,total-penalty.y)~Defend+Victim+Victim*Defend, family = binomial, data = DP)
summary(binmod)
```

```
##
## Call:
## glm(formula = cbind(penalty.y, total - penalty.y) ~ Defend +
##     Victim + Victim * Defend, family = binomial, data = DP)
##
## Deviance Residuals:
## [1]  0  0  0  0
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.9384     0.2454  -7.900  2.8e-15 ***
## Defend          0.3850     0.4127   0.933    0.351
## Victim        -23.0875 54958.9002   0.000    1.000
## Defend:Victim  21.8579 54958.9002   0.000    1.000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 8.1316e+00  on 3  degrees of freedom
## Residual deviance: 2.4362e-10  on 0  degrees of freedom
## AIC: 20.314
##
## Number of Fisher Scoring iterations: 22
```