STAT537: Statistics for Research I: HW#10

Due on Nov. 10, 2016

Dr. Schmidhammer TR 11:10am – 12:25pm

Wenqiang Feng

Contents

Problem 1

Ott Exercise 12.45abcd.

Solution. (a) **Compute the odds ratio for receiving the death penalty for each of the aggravation levels of the crime.** Since the formula of the odds ratio is as follows:

$$\text{Odds ration} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}. \tag{1}$$

Therefore, the odds ratio for receiving the death penalty for each of the aggravation levels can be calculate as follows:

- **Aggravation level 1:**

$$p_1 = \frac{2}{62}, \quad p_2 = \frac{1}{182}$$

$$\text{Odds ration} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = 6.033333. \tag{2}$$

- **Aggravation level 2:**

$$p_1 = \frac{2}{17}, \quad p_2 = \frac{1}{22}$$

$$\text{Odds ration} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = 2.8. \tag{3}$$

- **Aggravation level 3:**

$$p_1 = \frac{6}{13}, \quad p_2 = \frac{2}{11}$$

$$\text{Odds ration} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = 3.857143. \tag{4}$$

- **Aggravation level 4:**

$$p_1 = \frac{9}{12}, \quad p_2 = \frac{2}{6}$$

$$\text{Odds ration} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = 6. \tag{5}$$

- **Aggravation level 5:**

$$p_1 = \frac{9}{9}, \quad p_2 = \frac{4}{7}$$

$$\text{Odds ration} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = \infty. \tag{6}$$

- **Aggravation level 6:**

$$p_1 = \frac{17}{17}, \quad p_2 = \frac{4}{4}$$

$$\text{Odds ration} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = \infty. \tag{7}$$

(b) **Use a software package to fit the logistic regression model for the variables:** According to the fitted results, the model can be formulated as

$$\log(y) = -4.8653 + 1.5397 \cdot \text{Aggravation} - 1.8106 \cdot \text{Race}.$$

The fitted results are as follows:

```
Call:
glm(formula = cbind(Yes, No) ~ Aggravation + Race, family = binomial(logit),
    data = data)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.93570  -0.22548   0.05142   0.65620   1.01444

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.8653     0.6004  -8.103 5.37e-16 ***
Aggravation   1.5397     0.1867   8.246  < 2e-16 ***
Race1        -1.8106     0.5361  -3.377 0.000732 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 212.2838  on 11  degrees of freedom
Residual deviance:   3.8816  on  9  degrees of freedom
AIC: 31.747

Number of Fisher Scoring iterations: 4
```

(c) **Is there an association between the severity of the crime and the probability of receiving the death penalty?** From the following fitted model and ANOVA test, we get the difference $D = 212.284 - 16.685 = 195.6$ with degree of freedom 1, hence the p-value < 0.05. Hence we reject H_0 which is the coefficient of Aggravation is 0. Therefore, we may conclude that there is an significant association between the severity of the crime and the probability of receiving the death penalty. More specifically

$$\log(y) = -5.7102 + 1.5628 \cdot \text{Aggravation}.$$

- **fitted model with Aggravation as variable:**

```
Call:
glm(formula = cbind(Yes, No) ~ Aggravation, family = binomial(logit),
    data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1621  -0.8024   0.5766   0.9270   1.5198

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.7102     0.5685 -10.044   <2e-16 ***
Aggravation   1.5628     0.1757   8.894   <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: 212.284  on 11  degrees of freedom
Residual deviance:  16.685  on 10  degrees of freedom
AIC: 42.55

Number of Fisher Scoring iterations: 5
```

- **ANOVA test**

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: cbind(Yes, No)

Terms added sequentially (first to last)


            Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                           11    212.284
Aggravation  1    195.6        10     16.685 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(d) **Is the association between the severity of the crime and the probability of receiving the death penalty different for the two races?** From the fitted model for White and Black victims separately, we get

$$\text{For White victim:} \log(y) = -5.2531 + 1.6811 \cdot \text{Aggravation}.$$
$$\text{For Black victim:} \log(y) = -6.2319 + 1.4067 \cdot \text{Aggravation}.$$

Moreover, from the ANOVA test we have

- Since the p-value are less than 0.05, hence both of the races are significant impact on the probability of receiving the death penalty.
- One unit increase in severity of the crime will lead to 1.6811 percentage increases in the probability of receiving the death penalty for the White victims and one unit increase in severity of the crime will lead to 1.4067 percentage increases in the probability of receiving the death penalty for the Black victims.
- The Race Black has 0.2744 percentage lower impact on the probability of receiving the death penalty.

(i). **For White:**

- The fitted model for White victim is as follows:

```
Call:
glm(formula = cbind(Yes, No) ~ Aggravation, family = binomial(logit),
    data = dataWhite)

Deviance Residuals:
      1         3         5         7         9        11
 0.2315  -0.1673   0.1000  -0.5411   0.8681   0.5192

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.2531     0.8480  -6.194 5.85e-10 ***
Aggravation   1.6811     0.2842   5.916 3.30e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 106.2819  on 5  degrees of freedom
Residual deviance:   1.4074  on 4  degrees of freedom
AIC: 16.236

Number of Fisher Scoring iterations: 4
```

- ANOVA test

```
Terms added sequentially (first to last)


            Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                            5    106.282
Aggravation  1   104.87         4      1.407 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(ii). **For Black:**

- The fitted model for Black victim is as follows:

```
Call:
glm(formula = cbind(Yes, No) ~ Aggravation, family = binomial(logit),
```

```
    data = dataBlack)

Deviance Residuals:
      2        4        6        8       10       12
-0.3966   0.3456   0.6146  -0.1022  -0.6614   0.9132

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.2319     0.9015  -6.913 4.75e-12 ***
Aggravation   1.4067     0.2466   5.705 1.17e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 57.5843  on 5  degrees of freedom
Residual deviance:  1.9364  on 4  degrees of freedom
AIC: 16.973

Number of Fisher Scoring iterations: 4
```

- ANOVA test

```
Terms added sequentially (first to last)


            Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                            5     57.584
Aggravation  1   55.648         4      1.936 8.669e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(e) **Compute the probability of receiving the death penalty for an aggravation level of 3 separately for a white and then for a black victim. Place 95% confidence intervals on the two probabilities.**

- **For White:**The estimated probability for Black victim is 0.4615385, and the 95% confidence intervals is [0.2040175, 0.7387967].

```
> prop.test(6,6+7, alternative="two.sided",conf.level = 0.95)

1-sample proportions test with continuity correction

data:  6 out of 6 + 7, null probability 0.5
X-squared = 0, df = 1, p-value = 1
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.2040175 0.7387967
sample estimates:
        p
```

```
0.4615385
```

- **For Black:** The estimated probability for Black victim is 0.1818182, and the 95% confidence intervals is [0.03213862, 0.52245041].

```
> prop.test(2,2+9, alternative="two.sided",conf.level = 0.95)

1-sample proportions test with continuity correction

data:  2 out of 2 + 9, null probability 0.5
X-squared = 3.2727, df = 1, p-value = 0.07044
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.03213862 0.52245041
sample estimates:
        p
0.1818182
```

□

Appendix

R code for HW#10

Listing 1: Source code for problem 1

```
# reference: http://www.stat.columbia.edu/~martin/W2024/R3.pdf
rm(list = ls())
# set the path or envirionment
setwd("/home/feng/Dropbox/UTK_Course/Stat537/Excel/CH12")


oddratio<-function(p1,p2){
  oddr = (p1/(1-p1))/(p2/(1-p2))
  return(oddr)
}

# (a)
# level 1
p1 = 2/62
p2 = 1/182

oddr = oddratio(p1,p2)
oddr

# level 2
p1 = 2/17
p2 = 1/22
```

```
oddr = oddratio(p1,p2)
oddr

# level 3
p1 = 6/13
p2 = 2/11

oddr = oddratio(p1,p2)
oddr

# level 4
p1 = 9/12
p2 = 2/6

oddr = oddratio(p1,p2)
oddr

# level 5
p1 = 9/9
p2 = 4/7

oddr = oddratio(p1,p2)
oddr

# level 6
p1 = 17/17
p2 = 4/4

oddr = oddratio(p1,p2)
oddr

# (b)
#method 1
#install.packages("readxl") # CRAN version
library(readxl)
#install.packages("moments")
library(moments)
rawdata = read_excel("ex12-45.xls",sheet = 1)
attach(rawdata)

rawdata[rawdata=='Yes'] <- 1.0
rawdata[rawdata=='No'] <- 0.0
rawdata[rawdata=='Black'] <- 1
rawdata[rawdata=='White'] <- 0

rawdata$Race =as.factor(rawdata$Race)
rawdata$DeathPenalty =as.factor(rawdata$DeathPenalty)
#rawdata$AggLevel =factor(rawdata$AggLevel)
dvp=glm(DeathPenalty ~ AggLevel+Race, binomial(link = "logit"),data=rawdata);
summary(dvp)
exp(dvp$fitted.values)
predict(dvp, rawdata, type="response")
```

```
## method 2
library(readxl)
data <- read_excel("hw10.xls",sheet = 1)
attach(data)
# (b)
#Race =as.factor(Race)
#Race =relevel(Race,ref="White")
data[data=='Black'] <- 1
data[data=='White'] <- 0
fit1 = glm(cbind(Yes,No) ~ Aggravation+Race, family=binomial(logit), data=data)
summary(fit1)

#(c)
fit2 = glm(cbind(Yes,No) ~ Aggravation, family=binomial(logit), data=data)
summary(fit2)

anova(fit2, test="Chisq")

# (d)
data <- read_excel("hw10.xls",sheet = 1)
attach(data)
dataWhite = data[Race=="White", ]
dataBlack = data[Race=="Black", ]

fit3 = glm(cbind(Yes,No) ~ Aggravation, family=binomial(logit), data=dataWhite)
summary(fit3)
anova(fit3,test="Chisq")

fit4 = glm(cbind(Yes,No) ~ Aggravation, family=binomial(logit), data=dataBlack)
summary(fit4)
anova(fit4,test="Chisq")

# (e)
prop.test(6,6+7, alternative="two.sided",conf.level = 0.95)
prop.test(2,2+9, alternative="two.sided",conf.level = 0.95)
```