STA442 Assignment 10

Ruize Luo, 999307539

Wednesday, November 19, 2014

```
setwd("G:/Dropbox/2014 Fall/STA442/Assignments/Assignment 10")
```

Q1

```
q1data <- read.table("http://www.utstat.toronto.edu/~brunner/appliedf14/code_n_data/hw/awards.data", hea
q1data$prog <- factor(q1data$prog, labels=c("General", "Academic", "Vocational"))
attach(q1data)
```

```
table(awards, prog)
```

```
##       prog
## awards General Academic Vocational
##      0      36       48         40
##      1       9       32          8
##      2       0       11          2
##      3       0        9          0
##      4       0        2          0
##      5       0        2          0
##      6       0        1          0
```

Q1/(a)

It looks like one side of a normal distribution

Q1/(b)

i.

```
contrasts(prog)
```

```
##            Academic Vocational
## General           0          0
## Academic          1          0
## Vocational        0          1
```

Table 1: Dummy Variable Coding Scheme

Program	d_1	d_2
General	0	0
Academic	1	0
Vocational	0	1

ii.

Model: $E(log(Y)|X = x) = \beta_0 + \beta_1 * x + \beta_2 * d_1 + \beta_3 * d_2$

Table 2: Expected Number of Awards Given Score in Each Program

Program	d_1	d_2	Expected number of awards given score
General	0	0	$e^{\beta_0+\beta_1*x}$
Academic	1	0	$e^{\beta_0+\beta_1*x+\beta_2}$
Vocational	0	1	$e^{\beta_0+\beta_1*x+\beta_3}$

iii. e^{β_3}

iv. e^{β_2}

v. $e^{\beta_2-\beta_3}$

vi.

Because the mean is proportional to some other variables

vii.

A. $log(E(Y|X = x)) = \beta_0 + \beta_1 * x + \beta_2 * d_1 + \beta_3 * d_2 + \beta_4 * x * d_1 + \beta_5 * x * d_2$

B. $H_0 : \beta_4 = \beta_5 = 0$

C. $log(E(Y|X = x)) = \beta_0 + \beta_1 * x + \beta_2 * d_1 + \beta_3 * d_2$

D. $df = 2$

Q1/(c)

```
q1cmodel <- glm(awards~score+prog, data=q1data, family=poisson)
summary(q1cmodel)
```

```
##
## Call:
## glm(formula = awards ~ score + prog, family = poisson, data = q1data)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -2.204  -0.844  -0.511   0.256   2.680
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -6.6502     0.8481   -7.84  4.5e-15 ***
## score            0.0702     0.0106    6.62  3.6e-11 ***
## progAcademic     1.0839     0.3583    3.03   0.0025 **
## progVocational   0.3698     0.4411    0.84   0.4018
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 287.67  on 199  degrees of freedom
```

```
## Residual deviance: 189.45  on 196  degrees of freedom
## AIC: 373.5
##
## Number of Fisher Scoring iterations: 6
```

i. Yes. Controlling for program, participants with a **higher** score are expected to get **more** awards.

ii. Yes. Controlling for score on the test, students in the Academic program get **more** awards on average than students in the General program

iii. No. Controlling for score on the test, there is **no enough evidence** to show that students in the Vocational program get more awards on average than students in the General program

iv.

```r
q1cnullmodel <- glm(awards~1, data=q1data, family=poisson)
summary(q1cnullmodel)
```

```
##
## Call:
## glm(formula = awards ~ 1, family = poisson, data = q1data)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.122  -1.122  -1.122   0.429   4.038
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.4620     0.0891   -5.19  2.1e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 287.67  on 199  degrees of freedom
## Residual deviance: 287.67  on 199  degrees of freedom
## AIC: 465.7
##
## Number of Fisher Scoring iterations: 6
```

```r
anova(q1cmodel, q1cnullmodel, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: awards ~ score + prog
## Model 2: awards ~ 1
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       196        189
## 2       199        288 -3    -98.2   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Yes.

$p - value \approx 0$

v.

```
source("http://www.utstat.utoronto.ca/~brunner/Rfunctions/Wtest.txt")
q1cL <- matrix(c(0,0,1,-1), nrow=1)
q1ccoef <- q1cmodel$coefficients
Wtest(L=q1cL, Tn=q1ccoef, Vn=vcov(q1cmodel))
```

```
##       W      df p-value
## 4.97871 1.00000 0.02566
```

Tes. Controlling for score on the test, students in the Vocational program get **less** awards on average than students in the Academic program.

vi.

```
exp(q1ccoef[4])
```

```
## progVocational
##          1.447
```

The expected number of awards for a student in the Vocational program is estimated to be **1.447458** times as great as the expected number of awards for a student in the General program with the same score on the general knowledge test.

vii.

```
exp(q1ccoef[3])
```

```
## progAcademic
##        2.956
```

The expected number of awards for a student in the Academic program is estimated to be **2.956065** times as great as the expected number of awards for a student in the General program with the same score on the general knowledge test.

viii.

```
exp(t(q1ccoef)%*%c(0,0,1,-1))
```

```
##       [,1]
## [1,] 2.042
```

The expected number of awards for a student in the Academic program is estimated to be **2.042245** times as great as the expected number of awards for a student in the Vocational program with the same score on the general knowledge test.

ix.

```
xx <- matrix(c(1, 80, 1, 0), ncol=1)
xpred <- t(xx)%*%q1ccoef
sepred <- sqrt((t(xx)%*%vcov(q1cmodel)%*%xx))
zcrit <- qnorm(0.975)
CI <- exp(c("lower"=xpred-sepred*zcrit, "predicted"=xpred, "upper"=xpred+sepred*zcrit)); CI
```

```
##     lower predicted     upper
##    0.8614    1.0469    1.2725
```

```
expCI <- exp(CI); expCI
```

```
##     lower predicted     upper
##     2.366     2.849     3.570
```