STA303
A3 Part 2

Manomi Koroth
1002298220
March 28, 2019

Solutions

Question 1(a)

For this 2-by-2 contingency table that looks at variables **sex** and **like** and we can test for association between these two variables by applying the tests: *Difference of proportions* and *Pearson's TOI.*

Tab 1: Sex and Like

R Function	P-values
`chisq.test()`	`6.704e-12`
`prop.test()`	`6.704e-12`

This test gives us evidence to answer the following question:

H_0: There is no relationship between **sex** and **like**

H_a: There is a relationship between the two variables

The p-value produced by both function as shown above in Tab 1 is **6.704e-12** which is <0.0001 concluding that there is significant evidence that there is an association between sex and like.

The two proportions studied are:

P_1: Males who like Video Games = **0.8079**

P_2: Females who like Video Games= **0.4597**

Question 1(b)

This table was created to test whether the expected grade of a student had any implication on the results of part (a). This time we are testing the **grade** variable against the **sex** and **like** variable.

Expected Grade: **A+**

Tab 2

R Function	P-values
`chisq.test()`	`0.003861`
`prop.test()`	`0.003861`

This test gives us evidence to answer the following question:

H_0: There is no change between **sex** and **like** with respect to **grade A+**

H_a: There is a change between the two variables with respect to **grade A+**

The p-value produced by both function as shown above in Tab 2 is **0.003861** which is <0.05 concluding that there is significant evidence that there is a change between the association of sex and like when the student's grade is A+

The two proportions studied are:

P_1: Males who like Video Games and get A+ = **0.7442**

P_2: Females who like Video Games and get A+ = **0.4561**

Expected Grade: **Not A+**

Tab 3

R Function	P-values
`chisq.test()`	`2.877e-10`
`prop.test()`	`2.877e-10`

This test gives us evidence to answer the following question:

H_0: There is no change between **sex** and **like** with respect to **grade is not A+**

H_a: There is a change between the two variables with respect to **grade is not A+**

The p-value produced by both function as shown above in Tab 2 is **2.877e-10** which is <0.0001 concluding that there is significant evidence that there is a change between the association of sex and like when the student's grade is not A+

The two proportions studied are:

P_1: Males who like Video Games and don't get A+ = **0.8333**

P_2: Females who like Video Games and don't get A+ = **0.4607**

Question 2 (a)

Model 2.1 With interaction terms

$$\log(Y_i) = -0.1574(\beta_0) + 1.7668(\beta_1)I_{,i\ sex} - 0.0185(\beta_2)I_{,i\ grade} - 0.5231(\beta_3)I_{,i\ sex*grade} + \varepsilon_i$$

Y_i: Student who like playing video games

β_0: Intercept

β_1: Sex is Male if i=0 and Female if i=1

β_2: Grade is A+ if i=0 and not A+ if i=1

β_3: Sex is Male & Grade is A+ if i=0 and Female & not A+ if i=1

ε_i: Uncorrelated error terms

Model 2.2 Without interaction terms

$$\log(Y_i) = -0.1189(\beta_0) + 1.6111(\beta_1)I_{,i\ sex} - 0.1871(\beta_2)I_{,i\ grade} + \varepsilon_i$$

Y_i: Student who like playing video games

β_0: Intercept

β_1: Sex is Male if i=0 and Female if i=1

β_2: Grade is A+ if i=0 and not A+ if i=1

ε_i: Uncorrelated error terms

The two tests conducted to check the accuracy between the two models is: (1) Wald Test, and (2) LRT

i) The Wald Test for model 2.1:

p-value = **8.5e-09:** This p-value indicates that the interaction term sex*grade effect is statisticall y significant because <0.0001

ii) The Wald Test for Model 2.2

p-value = **2.2e-11**

This p value indicates that the **grade** term affects the response variable and is statistically significant because it is <0.0001

2) LRT results

P value= **0.3264 :** The p-value is less 0.05 which mean that the addition of the interaction term between sex and grade does not affect the response

Question 2 (b)

Practical Implications: The LRT test gives us evidence that the interaction between sex and grade does not help change the response of the model –which mean that that students who have A+ will continue to play video games if they like regardless of their sex.

In conclusion: Model 2.2 is better than Model 2.1

Question 3 (a)

<u>Model 3.1</u>

log (μ-hat) = -2.430e+01(β_0) + 4.400e-14(β_1)$I_{,\text{i}\ likes}$ + 4.400e-14 (β_2)$I_{,\text{i}\ sexes}$ + 4.400e-14(β_3)$I_{,\text{i}\ grades}$ - 4.400e-14 (β_4)$I_{,\text{i}\ likes*sexes}$ - 4.400e-14 (β_5)$I_{,\text{i}\ likes*grades}$ - 4.400e-14(β_6)$I_{,\text{i}\ sexes*grades}$+4.400e-14($\beta_7$)$I_{,\text{i}\ likes*grades*sexes}$ + ε_i

β_0: Intercept

β_1: Likes Playing if i=0 and Does not if i=1

β_2: Male if i=0 and Female i=1

β_3: Grade is A+ if i=0 and not A+ if i=1

β_4: Male & Like Games if i=0 and Female & doesn't like Games if i=1

β_5: Likes & has A+ if i=0 and Does not Like & does not have A+ if 1=1

β_6: Male & Grade is A+ if i=0 and Female & not A+ if i=1

β_7: Male & Like Games & Grade is A+ if i=0 and Female & doesn't like Games and not A+ if i=1

ε_i: Uncorrelated error terms

Model 3.2

log (μ-hat) = -2.430e+01(β_0) + 1.440e-14(β_1)$I_{,i\ likes}$ + 1.440e-14 (β_2)$I_{,i\ sexes}$ + 1.440e-14(β_3)$I_{,i\ grades}$ - 1.241e-14 (β_4)$I_{,i\ likes*sexes}$ – 1.241e-14 (β_5)$I_{,i\ likes*grades}$ – 1.241e-14(β_6)$I_{,i\ sexes*grades}$ + ε_i

β_0: Intercept

β_1: Likes Playing if i=0 and Does not if i=1

β_2: Male if i=0 and Female i=1

β_3: Grade is A+ if i=0 and not A+ if i=1

β_4: Male & Like Games if i=0 and Female & doesn't like Games if i=1

β_5: Likes & has A+ if i=0 and Does not Like & does not have A+ if 1=1

β_6: Male & Grade is A+ if i=0 and Female & not A+ if i=1

ε_i: Uncorrelated error terms

Question 3 (b)

i) Deviance

Testing model difference between the two models with LRT

Deviance= **-0.96302**

P value= **0.3264**

ii) Wald Test

Wald test for model 3.1 → p value = **1.1e-15**

Wald test for model 3.2 → p value = **0.0**

Table 4:Comparison between Q2(b) and Q3(b)

iii) Interpretation

The p value from the LRT test is high which gives us evidence that Model 3.2 is better 3.1 which means that the three-way interaction term does not affect the model response counts

The Wald test for Model 3.1 is <0.0001 which gives us evidence that the three-way interaction term between sex, grade and like is statistically significant

The Wald Test for Model 3.2 is also <0.001 which gives us evidence that the three two-way interactions are statistically significant

Below is Table to show the result of logistic regression model and the Poisson regression model side by side. And the results of the LRT are the same and the Wald test have different numerical results but conclude the same evidence.

Model	Wald Test	LRT
2.1	***8.5e-09***	**0.3264**
2.2	**2.2e-11**	
3.1	**1.1e-15**	**0.3264**
3.2	**0.0**	

STA303
A3 Part 2

Manomi Koroth
1002298220
March 28, 2019

Solutions-Appendix

```
data<-read.csv("C:\\Users\\Surface\\Documents\\STA303\\A3\\A3#2\\a3dat
a.csv", header=T)
attach(data)
```

##Question 1

##Q1(a)

```
M <- length(like[sex=="Male"])  #total numbers of males
F <- length(like[sex == "Female"]) #otal number of females

GamerM<-like[like== "1" & sex=="Male" ] #male gamers
NonGamerM <- like[like=="0" & sex=="Male"] #male non gamers
GamerF<- like[like=="1" & sex=="Female"] #female gamers
NonGamerF <- like[like=="1" & sex=="Female"] #female non gamers

GM <-length(like[like== "1" & sex=="Male"])   #no. of gamer males
NGM <-length(like[like== "0" & sex=="Male"])  #no. of non gamer males
GF <-length(like[like== "1" & sex=="Female"]) #no. of female gamer
NGF<-length(like[like== "0" & sex=="Female"]) #no.of female no gamer

GM/M #proporion of gamer males

NGM/M #proporion of non gamer males

GF/F #proporion of gamer females

NGF/F #proporion of non gamer females

table<- matrix(c(GM, NGM, GF, NGF), nrow=2,byrow=TRUE)
dimnames(table)<- list(c("Male", "Female"), c("Likes", "Does not Like"
))
names(dimnames(table))<- c("Sex", "Video Games")
table
```

```
##         Video Games
## Sex      Likes Does not Like
##   Male     122            29
##   Female   114           134
```

```r
chisq.test(table, correct=FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table
## X-squared = 47.112, df = 1, p-value = 6.704e-12
```

```r
prop.test(table, correct=FALSE)
```

```
##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  table
## X-squared = 47.112, df = 1, p-value = 6.704e-12
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.2599816 0.4365576
## sample estimates:
##    prop 1    prop 2
## 0.8079470 0.4596774
```

##Q1(b)

```r
# A+
GamerMAP<- like[like== "1" & sex=="Male" & grade =="1" ] #gamer males
who get A+
NonGamerMAP<- like[like== "0" & sex=="Male" & grade =="1"]  #gamer mal
es who  get A+
GamerFAP<- like[like=="1" & sex=="Female" & grade=="1"]  #gamer female
s who get A+
NonGamerFAP<- like[like=="0" & sex=="Female" & grade=="1"] #nongamer f
emales who get A+

MAP <- length(GamerMAP) + length(NonGamerMAP)
FAP <- length(GamerFAP) + length(NonGamerFAP)

GMAP<- length(like[like== "1" & sex=="Male" & grade =="1" ])
```

```r
NGMAP<-length(like[like== "0" & sex=="Male" & grade =="1" ])
GFAP<-length(like[like=="1" & sex=="Female" & grade=="1"])
NGFAP<-length(like[like=="0" & sex=="Female" & grade=="1"])







table1<-matrix(c(GMAP, NGMAP, GFAP, NGFAP), nrow=2,byrow=TRUE)
dimnames(table1)<-list(c("Male", "Female"), c("Likes", "Does not Like"
))
names(dimnames(table1)) <-c("Sex", "Video Games")
table1

##         Video Games
## Sex      Likes Does not Like
##   Male      32            11
##   Female    26            31

chisq.test(table1, correct=FALSE)

## 
##  Pearson's Chi-squared test
## 
## data:  table1
## X-squared = 8.3481, df = 1, p-value = 0.003861

prop.test(table1, correct=FALSE)

## 
##  2-sample test for equality of proportions without continuity
##  correction
## 
## data:  table1
## X-squared = 8.3481, df = 1, p-value = 0.003861
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.1043991 0.4716923
## sample estimates:
##    prop 1    prop 2 
## 0.7441860 0.4561404
```

```r
#with no A+
GamerMNA <- like[like== "1" & sex=="Male" & grade =="0" ]   #gamer mal
es who dont get A+
NonGamerMNA <- like[like== "0" & sex=="Male" & grade =="0" ] #non game
r males who dont get A+
GamerFNA<-like[like=="1" & sex=="Female" & grade=="0"] #gamer females
who dont get A+
NonGamerFNA<-like[like=="0" & sex=="Female" & grade=="0"] #nongamer fe
males who dont get A+

GMNA<-length(like[like== "1" & sex=="Male" & grade =="0" ])
NGMNA<-length(like[like== "0" & sex=="Male" & grade =="0" ])
GFNA<- length(like[like=="1" & sex=="Female" & grade=="0"])
NGFNA<-length(like[like=="0" & sex=="Female" & grade=="0"])


MNAP <- length(GamerMNA) + length(NonGamerMNA)
FNAP <- length(GamerFNA) + length(NonGamerFNA)

table2<-matrix(c(GMNA, NGMNA, GFNA, NGFNA), nrow=2,byrow=TRUE)
dimnames(table2)<-list(c("Male", "Female"), c("Likes", "Does not Like"
))
names(dimnames(table2))<-c("Sex", "Video Games")
table2

##          Video Games
## Sex       Likes Does not Like
##   Male      90            18
##   Female    88           103

chisq.test(table2, correct=FALSE)

##
##  Pearson's Chi-squared test
##
## data:  table2
## X-squared = 39.757, df = 1, p-value = 2.877e-10

prop.test(table2, correct=FALSE)

##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  table2
## X-squared = 39.757, df = 1, p-value = 2.877e-10
## alternative hypothesis: two.sided
```

```
## 95 percent confidence interval:
##  0.2729147 0.4722860
## sample estimates:
##    prop 1    prop 2
## 0.8333333 0.4607330
```

##Question 2

```
#Model 2.1
mod1 <-glm(like~ sex + grade + sex*grade, data= data, family=binomial)
summary(mod1)
```

```
##
## Call:
## glm(formula = like ~ sex + grade + sex * grade, family = binomial,
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8930  -1.1114   0.6039   1.2449   1.2530
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.1574     0.1452  -1.084    0.278
## sexMale         1.7668     0.2962   5.965 2.45e-09 ***
## grade          -0.0185     0.3030  -0.061    0.951
## sexMale:grade  -0.5231     0.5297  -0.987    0.323
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 539.70  on 398  degrees of freedom
## Residual deviance: 488.41  on 395  degrees of freedom
## AIC: 496.41
##
## Number of Fisher Scoring iterations: 4
```

```
#Model 2.2
mod2 <- glm(like ~ sex + grade, data=data, family= binomial)
summary(mod2)
```

```
##
## Call:
## glm(formula = like ~ sex + grade, family = binomial, data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.8412  -1.1273   0.6369   1.2283   1.3098
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.1189     0.1397  -0.851    0.395
## sexMale       1.6111     0.2438   6.610 3.85e-11 ***
## grade        -0.1871     0.2519  -0.743    0.458
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 539.70  on 398  degrees of freedom
## Residual deviance: 489.37  on 396  degrees of freedom
## AIC: 495.37
##
## Number of Fisher Scoring iterations: 4
```

##Q2(a)

```
##model 1
## Y = 0.1574 -1.7668B1ISex -0.0185B2IGrade -0.5231B3ISexGrade + ei

#tests: Wald tests and LRT

library(aod)

## Warning: package 'aod' was built under R version 3.5.3

wald.test(Sigma=vcov(mod1), b=coef(mod1), Terms=1:3)

## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 40.5, df = 3, P(> X2) = 8.5e-09
```

```r
wald.test(Sigma=vcov(mod2), b=coef(mod2), Terms=1:2)

## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 49.1, df = 2, P(> X2) = 2.2e-11



#anova(mod1, test="Chisq")
#anova(mod2, test="Chisq")   not sure

anova(mod1, mod2,  test="LRT") #lrt pval=0.3264 which means that the i
nteraction doesnt improve the model in addition to sex and grade

## Analysis of Deviance Table
##
## Model 1: like ~ sex + grade + sex * grade
## Model 2: like ~ sex + grade
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       395     488.41
## 2       396     489.37 -1 -0.96302   0.3264
```

##Question 3

```r
###counts

##No like/female/a+
NGFAP<- length(like[like=="0" & sex=="Female" & grade=="1"])
##no like/female/no a+
NGFNA <- length(like[like=="0" & sex=="Female" & grade=="0"])
##no like/male/ a+
NGMAP<-length(like[like== "0" & sex=="Male" & grade =="1" ])
##no like/male/no a+
NGMNA<-length(like[like== "0" & sex=="Male" & grade =="0" ])

#like/female/a+
GFAP<-length(like[like=="1" & sex=="Female" & grade=="1"])
#like/female/no a+
GFNA <- length(like[like=="1" & sex=="Female" & grade=="0"])
#like/male/a+
GMAP<-length(like[like== "1" & sex=="Male" & grade =="1" ])
#like/male/ no a+
```

```
GMNA<-length(like[like== "1" & sex=="Male" & grade =="0" ])

counts <- c(NGFAP, NGFNA, NGMAP, NGMNA, GFAP, GFNA, GMAP, GMNA)
likes <- c("no", "no", "no", "no", "yes", "yes", "yes", "yes")
sexes <- c("female", "female", "male","male", "female", "female", "mal
e", "male")
grades <- c("A+", "not A+", "A+", "not A+", "A+", "not A+", "A+","not
A+")

table4<- data.frame(counts, likes, sexes, grades, stringsAsFactors = F
ALSE) ##table from the A3 sheet
```

##Q3(a)

```
#model 3.1 with three way interaction
mod3 <-glm(counts ~ likes + sexes + grades + likes*sexes + likes*grade
s + sexes*grades + likes*sexes*grades, family= poisson, data=table4)
summary(mod3)

##
## Call:
## glm(formula = counts ~ likes + sexes + grades + likes * sexes +
##     likes * grades + sexes * grades + likes * sexes * grades,
##     family = poisson, data = table4)
##
## Deviance Residuals:
## [1]  0  0  0  0  0  0  0  0
##
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|
)
## (Intercept)                         3.4340     0.1796  19.120  < 2e-1
6 ***
## likesyes                           -0.1759     0.2659  -0.661  0.5083
5
## sexesmale                          -1.0361     0.3509  -2.952  0.0031
5 **
## gradesnot A+                        1.2007     0.2049   5.861 4.59e-0
9 ***
## likesyes:sexesmale                  1.2437     0.4392   2.832  0.0046
3 **
## likesyes:gradesnot A+               0.0185     0.3030   0.061  0.9513
1
## sexesmale:gradesnot A+             -0.7083     0.4341  -1.632  0.1027
6
## likesyes:sexesmale:gradesnot A+     0.5231     0.5297   0.987  0.3234
```

```
1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1.9388e+02  on 7  degrees of freedom
## Residual deviance: 3.9968e-15  on 0  degrees of freedom
## AIC: 59.808
##
## Number of Fisher Scoring iterations: 3
```

```
#model 3.2 without three way interaction
mod4<-glm(counts ~ likes + sexes + grades + likes*sexes + likes*grades
+ sexes*grades, family= poisson, data=table4)
summary(mod4)
```

```
##
## Call:
## glm(formula = counts ~ likes + sexes + grades + likes * sexes +
##     likes * grades + sexes * grades, family = poisson, data = table
4)
##
## Deviance Residuals:
##       1        2        3        4        5        6        7
8
## -0.3220   0.1812   0.5849  -0.4170   0.3672  -0.1935  -0.3171   0.1
940
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)               3.4913     0.1652  21.131  < 2e-16 ***
## likesyes                 -0.3061     0.2329  -1.314    0.189
## sexesmale                -1.2751     0.2704  -4.715 2.42e-06 ***
## gradesnot A+              1.1256     0.1865   6.034 1.60e-09 ***
## likesyes:sexesmale        1.6111     0.2438   6.610 3.85e-11 ***
## likesyes:gradesnot A+     0.1871     0.2519   0.743    0.458
## sexesmale:gradesnot A+   -0.3547     0.2523  -1.406    0.160
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 193.87673  on 7  degrees of freedom
## Residual deviance:   0.96302  on 1  degrees of freedom
## AIC: 58.771
```

```
##
## Number of Fisher Scoring iterations: 4
```

##Q3(b)
##i) Deviance

```
anova(mod3, mod4, test="LRT")

## Analysis of Deviance Table
##
## Model 1: counts ~ likes + sexes + grades + likes * sexes + likes *
grades +
##     sexes * grades + likes * sexes * grades
## Model 2: counts ~ likes + sexes + grades + likes * sexes + likes *
grades +
##     sexes * grades
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         0    0.00000
## 2         1    0.96302 -1 -0.96302   0.3264
```

##ii) Wald test

```
wald.test(Sigma=vcov(mod3), b=coef(mod3), Terms=4:7)

## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 76.2, df = 4, P(> X2) = 1.1e-15

wald.test(Sigma=vcov(mod4), b=coef(mod4), Terms=3:6)

## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 151.2, df = 4, P(> X2) = 0.0
```