STA 303 Assignment 3

Yongwen Tan, 1002158979

Part 2

1.(a).

Like / Sex	Not like	Like
Female	134	114
Male	29	122

46% of female like to play video games and 81% of male like to play video games. The odds of preferring to play video games for females is 0.85 and for males is 4.21. From the fisher's exact test, the p-value is 2.515e-12 which is very significant. From the pearson's chi-sq test of independence, the p-value is 6.704e-12 which is also significant.

Hence, there is strong statistical evidence to show that sex is not independent of a student's preference for playing video games.

(b).Grade A+

Like / Sex	Not like	Like
Female	31	26
Male	11	32

Grade NA+

Like / Sex	Not like	Like
Female	103	88
Male	18	90

In the students who expected grade A+, 45.6% females like playing video games, 74.4% males like playing video games. The odds of preferring to play games for females is 0.84 and 2.91 for males. From the chi-square test, p-value is 0.003861 which is significant. The p-value is 0.004462 is also significant from fisher's exact test. Hence, there is strong statistical evidence to show that preference for playing video games is associated with sex of the student.

46% females like playing video games and 83% males like playing video games for the students who did not expect grade A+. The odds of preferring to play games for females is 0.85and 5 for males The p-value from Pearson's chi-squared test is 2.877e-10. The p-value from the fisher's exact test is 1.048e-10. Both p-value are significant. Hence, there is strong statistical evidence to show that preference for playing video games is associated with sex of the student.

Overall, the preference for playing video games is associated with sex for students with the different expected grades.

2.(a). Model 2.1 $\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 I_{(sex=male)} + \beta_2 I_{(grade=A+)} + \beta_3 I_{(\text{sex=male})} * I_{(grade=A+)}$

Model 2.2 $\log\left(\frac{\pi}{1-\pi}\right) = \alpha_0 + \alpha_1 I_{(sex=male)} + \alpha_2 I_{(grade=A+)}$

π is the probability of having a preference for playing video games.

$I_{(sex=male)}$: 1 is male, otherwise is female.

$I_{(grade=A+)}$: 1 if expected grade was A+, otherwise expected grade was not A+.

$H_0: \beta_3 = 0$

$H_a: \beta_3 \neq 0$

1. Wald test

The test statistic $\frac{\beta_3}{SE(\beta_3)} \sim Z(0,1)$. The p-value of β_3 is 0.323 which is not significant. Hence, there is strong evidence to show the coefficient of the interaction term β_3 equal to 0.

2. Deviance test

Compare to the model 2.1 with model 2.2, the p-value is 0.3264 which is also not significant. We can conclude that there is strong evidence to show the coefficient of the interaction term β_3 equal to 0.

Hence, we conclude that model 2.2 is better than model 2.1. We do not consider the model with the interaction term.

(b). The reduced model (2.2) is better than the saturated model (2.1) because there is no association between the preference of playing video games and the interaction of expected grade and sex. This is consistent with the conclusion in question 1 (b). It can be concluded that the effect of sex on the odds of preference for playing video games does not associated with the expect grade.

3.(a)

Model 3.1 $\log(u_{ijk}) = \beta_0 + \beta_1 I_{(like=yes),ijk} + \beta_2 I_{(sex=male),ijk} + \beta_3 I_{(grade=nA+)ijk} + \beta_4 I_{(like=yes),ijk} * I_{(sex=male),ijk} + \beta_5 I_{(like=yes),ijk} * I_{(grade=nA+),ijk} + \beta_6 I_{(sex=male),ijk} * I_{(grade=nA+)ijk} + \beta_7 I_{(like=yes),ijk} * I_{(sex=male),ijk} * I_{(grade=nA+),ijk}$

Model 3.2 $\log(u_{ijk}) = \alpha_0 + \alpha_1 I_{(like=yes),ijk} + \alpha_2 I_{(sex=male),ijk} + \alpha_3 I_{(grade=nA+)ijk} + \alpha_4 I_{(like=yes),ijk} * I_{(sex=male),ijk} + \alpha_5 I_{(like=yes),ijk} * I_{(grade=nA+),ijk} + \alpha_6 I_{(sex=male),ijk} * I_{(grade=nA+)ijk}$

u_{ijk}: mean of the number of students of the poisson model
$I_{(sex=male)}$: 1 is male, otherwise is female.
$I_{(grade=nA)}$: 1 if expected grade was not A+, otherwise expected grade was A+.
$I_{(like=yes)}$: 1 if the student like to play video games, otherwise does not like to play video games.

(b).

i. The deviance for full model (3.1) is -8.65974e-15 which is very and closed to 0. The deviance for the reduced model (3.2) is 0.963. The according to the null hypothesis, the difference between the deviances should follow X_1^2 distribution. The p-value is 0.6735. We failed to reject the null hypothesis. We conclude that there is strong evidence to show the reduced model is a better fit.
Hence, the result is consistent with the results in the previous parts which the interaction term between expected grade and sex is not necessary.

ii. The p-value of wald test is 0.3234 which is non-significant and fail to reject the null hypothesis ($\beta_7 = 0$) which means there is no three-way interaction effect between the variables. Also this is consistent with the previous results.

iii. By checking the logistics model and poisson model, the results show that the interactions are not needed. We concluded that from the results above, the two-way interaction and the three-way interactions are not necessary which means the reduced model is better.

<u>Appendix</u>

```
> data=read.csv("~/desktop/STA303/A3/a3data.csv")
> attach(data)
The following objects are masked from data (pos = 3):

    grade, Grade, like, Like, sex

The following objects are masked from data (pos = 4):

    grade, Grade, like, Like, sex

The following objects are masked from a3data (pos = 5):

    grade, Grade, like, Like, sex


The following objects are masked from video (pos = 35):

    grade, like, sex

> #1a
> table=xtabs(~sex+like)
> table
        like
sex        0   1
  Female 134 114
  Male    29 122
> chisq.test(table, correct=FALSE)

    Pearson's Chi-squared test

data:  table
X-squared = 47.112, df = 1, p-value = 6.704e-12

> fisher.test(table)
```

```
        Fisher's Exact Test for Count Data

data:  table
p-value = 2.515e-12
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 3.008412 8.248768
sample estimates:
odds ratio
  4.924757

>
>
> #.b.
> gradeA=table(sex[grade=="1"], like[grade=="1"])
> gradenA=table(sex[grade=="0"], like[grade=="0"])
> gradeA

          0  1
  Female 31 26
  Male   11 32
> gradenA

           0   1
  Female 103  88
  Male    18  90
> chisq.test(gradeA, correct=FALSE)

        Pearson's Chi-squared test

data:  gradeA
X-squared = 8.3481, df = 1, p-value = 0.003861

> fisher.test(gradeA)

        Fisher's Exact Test for Count Data
```

```
data:  gradeA
p-value = 0.004462
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.36147 9.09891
sample estimates:
odds ratio
  3.423749

> chisq.test(gradenA, correct=FALSE)

	Pearson's Chi-squared test

data:  gradenA
X-squared = 39.757, df = 1, p-value = 2.877e-10

> fisher.test(gradenA)

	Fisher's Exact Test for Count Data

data:  gradenA
p-value = 1.048e-10
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
  3.185573 11.085730
sample estimates:
odds ratio
  5.817501

>
> #2.
> model2.1=glm(like~sex*grade, family=binomial())
> model2.2=glm(like~sex+grade, family = binomial())
> summary(model2.1)
```

```
Call:
glm(formula = like ~ sex * grade, family = binomial())

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

> summary(model2.2)

Call:
glm(formula = like ~ sex + grade, family = binomial())

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

> anova(model2.1, model2.2, test="Chisq")
Analysis of Deviance Table

Model 1: like ~ 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
>
>
>
> #3.
> count=c(31,103,11,18,26,88,32,90)
> like=as.factor(c("no","no","no","no","yes","yes","yes","yes"))
> sex=as.factor(c("female","female","male","male","female","female","male","male"))
> grade=as.factor(c("A+","not A+","A+","not A+","A+","not A+","A+","not A+"))
> model3.1=glm(count~sex*like*grade,family=poisson())
> summary(model3.1)

Call:
glm(formula = count ~ sex * like * grade, family = poisson())

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-16 ***
sexmale                     -1.0361     0.3509  -2.952  0.00315 **
likeyes                     -0.1759     0.2659  -0.661  0.50835
gradenot A+                  1.2007     0.2049   5.861 4.59e-09 ***
sexmale:likeyes              1.2437     0.4392   2.832  0.00463 **
sexmale:gradenot A+         -0.7083     0.4341  -1.632  0.10276
likeyes:gradenot A+          0.0185     0.3030   0.061  0.95131
sexmale:likeyes:gradenot A+  0.5231     0.5297   0.987  0.32341
---
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: -8.6597e-15  on 0  degrees of freedom
AIC: 59.808

Number of Fisher Scoring iterations: 3

> model3.2=glm(count~sex+like+grade+sex:like+sex:grade+like:grade,family=poisson())
> summary(model3.2)

Call:
glm(formula = count ~ sex + like + grade + sex:like + sex:grade +
    like:grade, family = poisson())

Deviance Residuals:
      1        2        3        4        5        6        7        8
-0.3220   0.1812   0.5849  -0.4170   0.3672  -0.1935  -0.3171   0.1940

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)              3.4913     0.1652   21.131   < 2e-16 ***
sexmale                 -1.2751     0.2704   -4.715 2.42e-06 ***
likeyes                 -0.3061     0.2329   -1.314    0.189
gradenot A+              1.1256     0.1865    6.034 1.60e-09 ***
sexmale:likeyes          1.6111     0.2438    6.610 3.85e-11 ***
sexmale:gradenot A+     -0.3547     0.2523   -1.406    0.160
likeyes:gradenot A+      0.1871     0.2519    0.743    0.458
---
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

>
> #i
> deviance(model3.1)
[1] -8.65974e-15
> deviance(model3.2)
[1] 0.9630241
> deviance(model2.1)
[1] 488.4063
> pchisq(deviance(model3.2),df=1)
[1] 0.6735739
>
```