Analysis on Student Gaming Preferences Data

Introduction

With gaming becoming an increasingly popular past time activity for students around the world, it is no surprise that we would be interested in studying what factors affect a student's gaming preferences. More specifically, in this study, I hope to answer 2 questions: Does a student's sex plays a role in their preference for games? Does the relationship between a student's preference for games and their sex change depending on their expected grades. In other words, is there some sort of interaction effect going on between a student's expected grade and other variables which might affect their preferences for gaming.

Preliminary Analysis

Game Preference	Sex: Male	Sex: Female
Likes Games	122	114
Doesn't Like Games	29	134

From the table above, we can tell that the proportion of students who enjoy playing games that are male is P_{male} = 122/236, and for females that proportion is P_{female} = 114/236. Similarly, the proportion of male and female students who don't enjoy playing games are q_{male} = 29/163 and q_{female}=134/163. Furthermore, using the difference of proportions test, we arrive at a p-value of 6.704×10^{-12}. And using the Fisher test, we arrived at a p-value of 2.515×10^{-12}. It should be noted that the chi-square test's p-value was not used since it is the same as the one in the proportions test. Given these 2 p-values at a 10% significance level, we conclude that there is strong evidence that a student's sexual orientation is dependent on their preference on games.

Below are contingency tables for the two grade types:

Grade

Game Preference	Sex: Male	Sex: Female
Likes Games	32	26
Doesn't Like Games	11	31

A+

Game Preference	Sex: Male	Sex: Female
Likes Games	90	88
Doesn't Like Games	18	103

Grade of NOT A+

Methods

For this particular dataset, it contains 399 different observation. The data was collected through a student survey conducted in the class of STA303 by Professor Shivon Sue Chee. In addition, logistic models and Poisson models were fitted to the data in R. Both these models were used in order to show two different approaches of answering the study question. A logistic model was chosen as the data consisted of a discrete categorical variable. It should also be noted that the Poisson model was fitted to data which consisted of the counts of students who liked and did not like video games (The data used can be found in the *Appendix*).

<u>Logistic Models:</u>

Let X= grade, W = sex. Where X represents the 2 levels of the variable 'grade' (i.e. 1= student's grade is A+, and 0 = student's grade is not A+), and W represents the 2 levels of the variable 'sex' (i.e. 1= male, 0 = Female). In addition, π_i, represents the probability of a student liking to play video games.

Model 2.1: $\text{logit}(\pi_i) = \beta_0 + \beta_1 X + \beta_2 W + \beta_3 XW$ for i = 1,…,399

$$= -0.1574 + 1.7668X - 0.0185W - 0.5231XW$$

Model 2.2: $\text{logit}(\pi_i) = \alpha_0 + \alpha_1 X + \alpha_2 W$ for i = 1,…,399

$$= -0.1189 + 1.6111X - 0.1871W$$

<u>Poisson Model:</u>

Let X represent variable 'like' (X= 1for likes game, X=0 otherwise). Let W represent the variable 'sex' (W =1 for male, W=0 otherwise). Let Z represent the variable 'grade' (Z=1 for grade of A+, Z=0 otherwise). Let

Model 3.1: $\log(\mu_{ijk}) = \beta_0 + \beta_1 X + \beta_2 W + \beta_3 Z + \beta_4 XW + \beta_5 XZ + \beta_6 WZ + \beta_7 XWZ$

$$= 3.4340 - 0.1759X - 1.0361W + 1.2007Z + 1.2437XW + 0.0185XZ - 0.7083WZ + 0.5231XWZ$$

Model 3.2: $\log(\mu_{ijk}) = \beta_0 + \beta_1 X + \beta_2 W + \beta_3 Z + \beta_4 XW + \beta_5 XZ + \beta_6 WZ$

$$= 3.413 - 0.361X - 1.2751W + 1.1256Z + 1.6111XW + 0.1871XZ - 0.3547WZ$$

Model Selection

In order to choose the best logistic model, I performed both a Likelihood Ratio Test and a Wald's Test on models 2.1 and 2.2. The results from the tests can be seen below:

Test 1: The first test is the Likelyhood Ratio Test (LRT).

Null hypothesis (H_o):	The reduced model (Model 2.2) is appropriate
Alternative hypothesis (H_A):	The full model (Model 2.1) is better
Test Statistics:	G^2= 489.37- 488.41 = 0.96 Following a chi-square distribution with 1 degree of freedom
P-Value:	0.3272 (see appendix for code)
Conclusions:	Since we are using a significance level of 0.10, we would fail to reject the null hypothesis since our p-value of 0.3272 is greater than our significance level. In other words, the reduced model (Model 2.2) is an appropriate fit for our data.

Test 2: The second test is a Wald's test

Null hypothesis (H_o):	$\beta_3=0$
Alternative hypothesis (H_A):	$\beta_3\neq0$
Test Statistics:	$Z^2=$ 0.9741 Following a chi-square distribution with 1 degree of freedom
P-Value:	0.3230
Conclusions:	Since we are using a significance level of 0.10, we would still fail to reject the null hypothesis since our p-value of 0.3236 is greater than our significance level. In other words, we have enough evidence saying that grade and sex don't have an effect on the odds of liking games. Thus we should stick to model 2.2.

Regarding the Poisson models, I chose to perform just the Wald's Test on the 2 models. The results indicated that model 3.2 (a reduced model) was the better choice for this study.

Conclusion

In conclusion, from the fitted logistic models above, we see that there is no significant interaction effect between a student's sex and expected grades on their preference for playing video games; this was previously shown by the high p-value of 0.3272 which indicates a lack of an interaction effect. Similarly, the fitted Poisson models also appears to show no signs of an interaction effect between the student's sex and grades on gaming preference. However, when looking at the results from both of the chosen reduced models (i.e. Model 2.2 & Model 3.2), it would appear that a student's sex plays a crucial part in their preference for gaming. That is, if a student is male, there are higher odds of that student preferring to game in contrast to a female student.

Appendix

Pre-code:

```
studData <- read.csv("a3data.csv")
apSubset <- subset(studData, grade == "1")
notapSubset <- subset(studData, Grade == "A " | Grade == "B " | Grade == "C ")
```

Below is code for the 2x2 table:

```
gamerMale <- subset(studData, like == "1" & sex == "Male")
gamerFemale <- subset(studData, like == "1" & sex == "Female")

notGamerMale  <- subset(studData, like == "0" & sex == "Male")
notGamerFemale <- subset(studData, like == "0" & sex == "Female")

contTable <- matrix(c(nrow(gamerMale),nrow(gamerFemale), nrow(notGamerMale),
nrow(notGamerFemale)), nrow= 2, byrow = TRUE)
dimnames(contTable) <- list(c("Likes Games", "Doesn't Like Games"), c("Male", "Female"))
names(dimnames(contTable)) <- c("Game Preference", "Sex")
```

Code below for proportion and fisher tests

```
diffPropTest <- prop.test(contTable, correct = FALSE
```

```
> diffPropTest

	2-sample test for equality of proportions without continuity correction

data:  contTable
X-squared = 47.112, df = 1, p-value = 6.704e-12
alternative hypothesis: two.sided
95 percent confidence interval:
 0.2523654 0.4257047
sample estimates:
   prop 1    prop 2
0.5169492 0.1779141
```

```
fisherTest <- fisher.test(contTable)
```

```
> fisherTest

        Fisher's Exact Test for Count Data

data:  contTable
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
```

Logistic Models

```
like <- studData$like
sex <- studData$sex
grades <- studData$grade

model2.1 <- glm(like ~ sex * grades, family = binomial, data = studData)
model2.2 <- glm(like ~ sex + grades, family = binomial, data = studData)


summary1 <- summary(model2.1)
```

```
> summary1

Call:
glm(formula = like ~ sex * grades, family = binomial, data = studData)

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 ***
grades          -0.0185     0.3030  -0.061    0.951
sexMale:grades  -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
```

```
> pchisq(0.96, 1, lower.tail = FALSE)
[1] 0.3271869
```

summary2 <- summary(model2.2)

```
> summary2

Call:
glm(formula = like ~ sex + grades, family = binomial, data = studData)

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 ***
grades       -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
```

Poisson Models

Generating Count data to feed into poisson model

```
count <- c(31,103,11,18,26,88,32,90)
like1 <- as.factor(c("no", "no", "no","no", "yes","yes","yes","yes"))
sex1 <- as.factor(c("female" , "female", "male", "male", "female", "female", "male", "male"))
grade1 <- as.factor(c("A+", "not A+", "A+", "not A+","A+", "not A+","A+", "not A+"))
```

```
pModel1 <- glm(count ~ like1 * sex1 * grade1, family = poisson)
pModel2 <- glm(count ~ (like1 + sex1 + grade1)^ 2, family = poisson)
```

```
> summary(pModel1)

Call:
glm(formula = count ~ like1 * sex1 * grade1, 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 ***
like1yes                       -0.1759     0.2659  -0.661  0.50835    
sex1male                       -1.0361     0.3509  -2.952  0.00315 ** 
grade1not A+                    1.2007     0.2049   5.861 4.59e-09 ***
like1yes:sex1male               1.2437     0.4392   2.832  0.00463 ** 
like1yes:grade1not A+           0.0185     0.3030   0.061  0.95131    
sex1male:grade1not A+          -0.7083     0.4341  -1.632  0.10276    
like1yes:sex1male:grade1not 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: 4.6629e-15  on 0  degrees of freedom
AIC: 59.808
```

```
> summary(pModel2)

Call:
glm(formula = count ~ (like1 + sex1 + grade1)^2, 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 ***
like1yes               -0.3061     0.2329  -1.314    0.189    
sex1male               -1.2751     0.2704  -4.715 2.42e-06 ***
grade1not A+            1.1256     0.1865   6.034 1.60e-09 ***
like1yes:sex1male       1.6111     0.2438   6.610 3.85e-11 ***
like1yes:grade1not A+   0.1871     0.2519   0.743    0.458    
sex1male:grade1not 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
```