Effects of Video Games and Expectations on Student Grades

Ted Si Yuan Cheng

2018-03-17

Part 1 - Contingency Tables

1. a) Construct a 2 x 2 contingency table of sex by like. Determine if there's evidence taht sex is independet of a student's preference for playing video games. b) Examine the sex and like relationship separately for each grade type expected.

(a)

```
##      sex
## like  female male
##   no      12    8
##   yes     26   44
```

Using the Chi-Square test setting Yate's Continuity Correction to false, we get that the p-value is 0.06797, which is > 0.05, so we can not reject the null hypothesis that there is no relationship between sex and preference for video games.

Using Fisher's Exact test, we get that the p-value is 0.07824, which is still > 0.05, so we can not reject the null hypothesis that there is no relationship between sex and preference for video games.

We see that the 95% CI is [0.82, 8.11] from Fisher's exact test, which contains 1. This is more supporting evidence that the p-value is > 0.05 and that there is evidence of association between the variables.

(b)

From the Chi-Square test for the contingency table where the grades are A, we see that the p-value is $0.001102 < 0.05$, meaning we can reject the null hypothesis that there is no relationship between sex and preference for video games with students who expect to get A's and say that there is an association them.

For grade = NA, p-value is $0.9421 > 0.05$, so we can not reject the null hypothesis that there is no relationship between sex and preference for video games with students who don't expect to get A's.

2. Carry out a logistic regression analysis to determine the effect of sex and grade expected on the odds of liking gaming. Model 2.1 be the model including interaction between sex and grade, model 2.2 be the model without interaction.

(a) Determine which model to use

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

The saturated model 2.1 would be $log(\pi/1-\pi) = \beta_0 + \beta_1 I_{sex=male} + \beta_2 I_{grade=nA} + \beta_3 I_{sex=male} * I_{grade=nA}$, where β_0 is the intercept of the logistic regression line (the expected value of like when all other parameters are 0), $\beta_1 I_{sex}$ is the coefficient if sex is 1 (female = 0, male = 1), $\beta_2 I_{grade}$ is the coefficient if they expect an A is 1 (nA = 0, A = 1), and $\beta_3 Isex * I_{grade}$ is the interaction coefficient of sex and grade if there is interaction.

The independent model 2.2 would be $log(\pi/1-\pi) = \beta_0 + \beta_1 I_{sex=male} + \beta_2 I_{grade=nA}$, where the parameters are the same as the saturated model 2.1 but $\beta_3 I_{sex} * I_{grade} = 0$ (no interaction).

Using the likelihood ratio test, we see that p-value = 0.0087 < 0.05, and the Wald Test produces a p-value = 0.018 < 0.05. From the likelihood ratio test we see that there is evidence to reject that the additive model is better. From the wald test we also see that there is evidence that the interaction between sex and grade expected is significant.

(b) Pracitcal implications of the model in part (a)

A practical way of utilizing the model is to see if males or females who expect an A or not will prefer playing video games or not. We can see if the interaction between being a certain gender and expecting a an A or not has correlation to whether or not they prefer playing video games. We do see that the interaction model is adequate from the likelihood ratio test, and that there is interaction between sex and grade expected for those who expect an A, supporting 1(b) that there is evidence of association between sex and preference for those who expect to get A's

3. Analysis using Poissson Regression

(a) Model counts as Poisson variables and fit two models: Model 3.1 - with explaantory variables sex, grade and like, the three two-way terms and the three-way interaction. Model 3.2 - the three-way interaction term removed

The saturated model 3.1 would be $log(\mu_i) = \beta_0 + \beta_1 I_{sex=male} + \beta_2 I_{grade=nA} + \beta_3 I_{like=yes} + \beta_4 I_{sex} * I_{grade}$ $+ \beta_5 I_{sex} * I_{like} + \beta_6 Igrade * I_{like} + \beta_7 I_{sex} * I_{grade} * I_{like}$, where β_0 is the intercept of the log-linear model line (the expected value of count when all other parameters are 0), $\beta_1 I_{sex}$ is the coefficient if sex is 1 (female = 0, male = 1), $\beta_2 I_{grade}$ is the coefficient if grade is 1 (expect an A = 1, nA = 0), $\beta_3 I_{like}$ is the coefficient if like is 1 (no = 0, yes = 1), and $\beta_4 I_{sex} * I_{grade}$ is the coefficient of sex and grade if there is an interaction. $\beta_5 I_{sex} * I_{like}$ is coefficient of sex and like if there is interaction, $\beta_6 Igrade * I_{like}$ is the coefficient of like and grade if there is an interaction, and $\beta_7 I_{sex} * I_{grade} * I_{like}$ is the coefficient of sex, grade, and like if there is an interaction with all 3.

The reduced model 3.2 has the same parameters as model 3.1 but without $B_7(Sex : Grade : Like)$ (no interaction). It would be $log(\mu_i) = \beta_0 + \beta_1 I_{sex=male} + \beta_2 I_{grade=nA} + \beta_3 I_{like=yes} + \beta_4 I_{sex} * I_{grade} +$ $\beta_5 I_{sex} * I_{like} + \beta_6 Igrade * I_{like}$\$.

(b) Compare the results of the Poisson regression models to the Logistic regression modelling in terms of (i) deviance, (ii) Wald tests, (iii) interpretation.

(i)

Using the likelihood ratio test, the deviance for models 2.1 and 2.2 compared to models 3.1 and 3.2 are the same, as they both have a value of 6.8788

(ii)

The wald tests match for 3-way interaction term of 3.1 and 2-way interaction term of 2.1 (p=0.0185) 2-way interaction between sex and like in 3.2 and the main effect of sex in 2.2 (p=0.0826)

(iii)

The interpretations differ slightly; the Logistic models in part 2 predict the log odds of 'like' while the Poisson (log-linear) models predict the log of the mean counts.

Appendix

```
knitr::opts_chunk$set(echo = TRUE)
#Read video data
video <- read.csv("C:\\Users\\Ted\\Documents\\R Projects\\3\\video.csv")
#2x2 contingency table of sex by likes
sexlike <- matrix(c(12,8,26,44), nrow=2, byrow=TRUE)
dimnames(sexlike) <- list(c("no","yes"), c("female","male"))
names(dimnames(sexlike)) <- c("like", "sex")
sexlike
```

```
##      sex
## like  female male
##   no      12    8
##   yes     26   44
```

```
#Test for independency with Chi-Square and Fisher's test
chisq.test(sexlike, correct=FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  sexlike
## X-squared = 3.3314, df = 1, p-value = 0.06797
```

```
fisher.test(sexlike)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  sexlike
## p-value = 0.07824
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.8195672 8.1070182
## sample estimates:
## odds ratio
##   2.511179
```

```
table(video)
```

```
## , , grade = A
##
##      sex
```

```
## like  female male
##   no       5    1
##   yes      4   21
##
## , , grade = nA
##
##      sex
## like  female male
##   no       7    7
##   yes     22   23
```

```r
#Grade A
gradeA <- matrix(c(5,1,4,21), nrow=2, byrow=TRUE)
dimnames(gradeA) <- list(c("no","yes"), c("female","male"))
names(dimnames(gradeA)) <- c("like", "sex")

#Grade NA
gradeNA <- matrix(c(7,7,22,23), nrow=2, byrow=TRUE)
dimnames(gradeNA) <- list(c("no","yes"), c("female","male"))
names(dimnames(gradeNA)) <- c("like", "sex")

#Chi-Square test for independency
chisq.test(gradeA, correct=FALSE)
```

```
## Warning in chisq.test(gradeA, correct = FALSE): Chi-squared approximation
## may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  gradeA
## X-squared = 10.648, df = 1, p-value = 0.001102
```

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

```
##
##  Pearson's Chi-squared test
##
## data:  gradeNA
## X-squared = 0.0052746, df = 1, p-value = 0.9421
```

```r
#Sex and grade models

attach(video)
```

```
## The following objects are masked _by_ .GlobalEnv:
##
##     grade, like, sex
```

```
## The following objects are masked from videodf:
##
##     grade, like, sex
```

```r
sex = as.factor(sex)
like <- as.factor(like)
grade <- as.factor(grade)
model_2.1 <- glm(like ~ sex*grade, family = binomial, data = video)
model_2.2 <- glm(like ~ sex+grade, family = binomial, data = video)
summary(model_2.1)
```

```
##
## Call:
## glm(formula = like ~ sex * grade, family = binomial, data = video)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4864   0.3050   0.7290   0.7433   1.2735
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -0.2231     0.6708  -0.333  0.73940
## sexmale           3.2677     1.2237   2.670  0.00758 **
## gradenA           1.3683     0.7989   1.713  0.08679 .
## sexmale:gradenA  -3.2232     1.3682  -2.356  0.01848 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 95.347  on 89  degrees of freedom
## Residual deviance: 85.152  on 86  degrees of freedom
## AIC: 93.152
##
## Number of Fisher Scoring iterations: 5
```

```r
summary(model_2.2)
```

```
##
## Call:
## glm(formula = like ~ sex + grade, family = binomial, data = video)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9533   0.5668   0.5861   0.8512   0.8774
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.8288     0.5586   1.484   0.1379
## sexmale       0.9183     0.5291   1.736   0.0826 .
## gradenA      -0.0727     0.5679  -0.128   0.8981
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 95.347  on 89  degrees of freedom
## Residual deviance: 92.031  on 87  degrees of freedom
## AIC: 98.031
##
## Number of Fisher Scoring iterations: 4
```

```
#LRT and Wald test
library(aod)
anova(model_2.1, model_2.2, 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        86     85.152
## 2        87     92.031 -1  -6.8788 0.008723 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
wald.test(Sigma = vcov(model_2.1), b=coef(model_2.1), Terms = 4)
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 5.5, df = 1, P(> X2) = 0.018
```

```
videodf <- data.frame(table(video))
detach(video)
attach(videodf)
```

```
## The following objects are masked _by_ .GlobalEnv:
##
##     grade, like, sex
```

```
## The following objects are masked from videodf (pos = 3):
##
##     Freq, grade, like, sex
```

```
#Saturated and no reduced models
model_3.1 <- glm(Freq~sex*grade*like, family = poisson, data=videodf)
model_3.2 <- glm(Freq~sex+grade+like+sex:grade+sex:like+grade:like, family = poisson, data=videodf)
summary(model_3.1)
```

```
##
## Call:
## glm(formula = Freq ~ sex * grade * like, family = poisson, data = videodf)
##
## Deviance Residuals:
## [1]  0  0  0  0  0  0  0  0
```

```
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)              1.6094     0.4472   3.599  0.00032 ***
## sexmale                 -1.6094     1.0954  -1.469  0.14177
## gradenA                  0.3365     0.5855   0.575  0.56554
## likeyes                 -0.2231     0.6708  -0.333  0.73940
## sexmale:gradenA          1.6094     1.2189   1.320  0.18670
## sexmale:likeyes          3.2677     1.2238   2.670  0.00758 **
## gradenA:likeyes          1.3683     0.7989   1.713  0.08679 .
## sexmale:gradenA:likeyes -3.2232     1.3683  -2.356  0.01849 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance:  5.4112e+01  on 7  degrees of freedom
## Residual deviance: -5.7732e-15  on 0  degrees of freedom
## AIC: 47.169
##
## Number of Fisher Scoring iterations: 3
```

```
summary(model_3.2)
```

```
##
## Call:
## glm(formula = Freq ~ sex + grade + like + sex:grade + sex:like +
##     grade:like, family = poisson, data = videodf)
##
## Deviance Residuals:
##       1        2        3        4        5        6        7        8
##  1.2260  -0.9698  -1.4709   0.5132  -0.7780   0.5005   0.9711  -0.4576
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)       1.0061     0.5122   1.964   0.0495 *
## sexmale           0.1771     0.5684   0.312   0.7553
## gradenA           1.2201     0.5480   2.227   0.0260 *
## likeyes           0.8288     0.5586   1.484   0.1379
## sexmale:gradenA  -0.8484     0.4819  -1.761   0.0783 .
## sexmale:likeyes   0.9183     0.5291   1.736   0.0826 .
## gradenA:likeyes  -0.0727     0.5679  -0.128   0.8981
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 54.1125  on 7  degrees of freedom
## Residual deviance:  6.8788  on 1  degrees of freedom
## AIC: 52.048
##
## Number of Fisher Scoring iterations: 5
```

```r
#LRT and Wald Test for Models 3.1 and 3.2
anova(model_3.1, model_3.2, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: Freq ~ sex * grade * like
## Model 2: Freq ~ sex + grade + like + sex:grade + sex:like + grade:like
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         0     0.0000
## 2         1     6.8788 -1  -6.8788 0.008723 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
wald.test(Sigma = vcov(model_3.1), b=coef(model_3.1), Terms = 8)
```

```
## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 5.5, df = 1, P(> X2) = 0.018
```