ASSIGNEMNT 3 – DAVID VEITCH 1004657917

1. Selected Field of Interest
Education.

2. Reference to the article
Dhuey, E. & Smith, J (2018). How school principals influence student learning. *Empir Econ* (2018) 54: 851. https://doiorg.myaccess.library.utoronto.ca/10.1007/s00181-017-1259-9

3. UofT Department
Elizabeth Dhuey – Department of Management – University of Toronto Scarborough

4. Soft-Copy
https://doiorg.myaccess.library.utoronto.ca/10.1007/s00181-017-1259-9
http://rdcu.be/JymP

5. Statistical Software
The author does not specify which statistical software was used.

6. Data Source
The data comes from an observational study, namely "administrative records of the North Carolina Department of Public Instruction" which includes data on schools, teachers, students, and principals. The authors also used census data in this study. The authors have a cleaned dataset of 5,388,543 observations of student test scores between 1998 and 2010.

7. Summary Statistics/Plots
Yes the article presents summary statistics in two ways. First it presents summary descriptive statistics of the data used (e.g. student test scores) and then it presents tables of its results (i.e. principal effect on student achievement). Two examples below:

How school principals influence student learning 861

Table 1 Descriptive statistics for students in analysis sample

	Mean	SD
Math scores		
Grade 3	0.067	0.960
Grade 4	0.038	0.986
Grade 5	0.041	0.986
Grade 6	0.054	0.983
Grade 7	0.055	0.986
Grade 8	0.052	0.984
Reading scores		
Grade 3	0.063	0.958
Grade 4	0.029	0.987
Grade 5	0.032	0.983
Grade 6	0.047	0.975
Grade 7	0.048	0.975
Grade 8	0.046	0.973
Student demographic characteristics		
Male	0.501	0.500
Black	0.286	0.452
Hispanic	0.054	0.227
White	0.606	0.489
Other race	0.053	0.224
Special education	0.092	0.289
Gifted	0.140	0.347
Learning disabled in math	0.020	0.140
Learning disabled in reading	0.041	0.199
Number of students	1,664,158	
Number of observations	5,388,543	

Test scores are standardized to have a mean of zero and standard deviation of 1 in the population of test takers within a subject, grade, and year, prior to sample exclusions. Other race includes all races except the three listed in the table

Table 7 Relationship between principal quality and education

	Math	Reading
	(1)	(2)
Panel A: Graduate degrees		
Graduate degree (doctorate or advanced certificate)	0.0034	0.0109**
	(0.0059)	(0.0044)
Adjusted *R*-squared	0.0001	0.0014
Panel B: Doctorate or advanced certificate		
Doctorate	0.0094	0.0068
	(0.0091)	(0.0069)
Advanced certificate	−0.0005	0.0120**
	(0.0063)	(0.0047)
Adjusted *R*-squared	0.0002	0.0021
Panel C: Competitive or non-competitive		
Competitive master's school	−0.0061	0.0016
	(0.0083)	(0.0058)
Competitive doctorate school	−0.0169	−0.0089
	(0.0177)	(0.0128)
Non-competitive doctorate school	0.0201**	0.0122
	(0.0102)	(0.0079)
Competitive advanced certificate school	0.0062	0.0220**
	(0.0131)	(0.0107)
Non-competitive advanced certificate school	−0.0024	0.0103**
	(0.0068)	(0.0051)
Adjusted *R*-squared	0.0012	0.0026
Number of observations	4289	4289

Education variables are indicator variables equal to 1 if the principal ever had the particular degree. Standard errors are robust to heteroskedasticity
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

8. Test Statistics/P-Values/Confidence Intervals
The authors conduct a regression analysis of a number of factors principals have versus student achievement. In presenting the results as shown above the authors highlight factors which are statistically significant (having a P-Value less than 0.10, 0.05, and 0.01).

9. How many decimal places were values recorded.
Regression coefficients were reported to four decimal places. Summary statistics (e.g. mean/standard deviation) were reported to three decimal places.

10. Statistical Method to Analyze Data
The authors use ordinary least squares regression to estimate the effect principals have on students. Description of the model below.

4 Empirical specifications

4.1 Estimating principal effects

To estimate the principal effects, we use the following value-added model of students' test scores:

$$y_{it} = \beta_0 + y_{it-1}\beta_1 + x'_{it}\beta_2 + z'_{s(i,t)t}\beta_3 + p'_{p(i,t)t}\beta_4 + \delta_{p(i,t)} + \varphi_{s(i,t)} + \eta_t + \varepsilon_{it}, \tag{1}$$

where y_{it} is the math or reading score for student i at time t; y_{it-1} is the student's 1-year lagged math or reading score; x'_{it} is a vector of student-level demographic

[4] While principals possess the power of evaluating teachers, the superintendent or the board of directors (for regional schools) have the power to employ and contract with teachers along with the duty to maintain personnel files and participate in the firing and demoting.

characteristics; $z'_{s(i,t)t}$ is a vector of school-level demographic characteristics for the school that student i attends at time t; $p'_{p(i,t)t}$ is a vector of principal-level, time-changing characteristics for student i's principal at time t; $\delta_{p(i,t)}$, $\varphi_{s(i,t)}$, and η_t are time-invariant principal, school, and year effects; and ε_{it} is an idiosyncratic error term.[5]

PART 2 CONTINGNEGY TABLES

1a.

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

Contingency table: sex, like

There is evidence that sex is not independent of a student's preference for playing video games. This is evident from the p-values obtained from a difference of proportions test (p-value 0.06797) and the Fisher test (p-value 0.07824). These p-values are below the 10% level of significance suggesting there is evidence sex is not independent of preference for playing video games.

In practical terms this means if a given subject is male they are much more likely to like playing video games than a female. In this case 84% of males liked playing video games compared to only 68% of females).

1b.

It does appear that the sex and like relationship changes with grade expected.

For students expecting an A in a statistics course there appears to be a relationship between sex and preference for playing video games (males much more likely to like playing video games). This can be seen from the difference of proportion test (p-value 0.001102) and Fisher test (p-value 0.003879). This relationship does not seem to exist for students not expecting an A; the difference of proportion tests (p-value 0.9421) and Fisher test (p-value 1) appear to suggest no relationship.

2a.

We fit two models to the data, Model 2_1 with an interaction term, and Model 2_2 without interaction.

Model 2_1 – **logit(like)** = $\beta_{intercept} + I_{male} * \beta_{sexmale} + I_{gradenA} * \beta_{gradenA} + I_{sexmale:gradenA} * \beta_{sexmale:gradenA}$

Model 2_2 – **logit(like)** = $\beta_{intercept} + I_{male} * \beta_{sexmale} + I_{gradenA} * \beta_{gradenA}$

Model Term	Description
logit(like)	predicted log odds that individual likes video games
$\beta_{intercept}$	predicted log odds for sex = female, grade = A
$\beta_{sexmale}$	change in log odds if sex = male
$\beta_{gradenA}$	change in log odds if grade = nA
$\beta_{sexmale:gradenA}$	change in log odds if sex = male and grade = nA
I_{male}	indicator variable, 1 if sex = male, 0 if sex = female, 0 otherwise
$I_{gradenA}$	indicator variable, 1 if grade = nA, 0 if grade = A, 0 otherwise
$I_{sexmale:gradenA}$	indicator variable, 1 if grade = nA and sex = male, 0 otherwise

Model 2_1, the model with interaction, appears to be the better fit the data.

First we conduct a likelihood ratio test between the two models, comparing their residual deviance. From a chi-square distribution of one degree of freedom (the models only differ by one parameter) we get a p-value of 0.0087 suggesting the model with interaction is a better fit.

Second we conduct a Wald chi-square test on the $\beta_{sexmale:gradenA}$ term in Model 2_1. Observing the p-value (0.01848) from the summary output of the logistic regression model we can conclude that the model with $\beta_{sexmale:gradenA}$ better fits the data.

2b.

The model selected in part 2a), the model with interaction, agrees with our analysis in question 1b). The practical implications of this is that when modelling whether a student likes video games one must take into account that males with high grade expectations have a different preference for video games than males with low expectations, and females with high grade expectations have a different preference for video games than females with low expectations.. Therefore to properly model this relationship one should have an interaction term in the model.

3a.

Model 3_1 – **log(Freq)** = $\beta_{intercept} + I_{male} * \beta_{sexmale} + I_{gradenA} * \beta_{gradenA} + I_{likeyes} * \beta_{likeyes} + I_{sexmale:gradenA} * \beta_{sexmale:gradenA} + I_{sexmale:likeyes} * \beta_{sexmale:likeyes} + I_{gradenA:likeyes} * \beta_{gradenA:likeyes} + I_{sexmale:gradenA:likeyes} * \beta_{sexmale:gradenA:likeyes}$

Model 3_2 – **log(Freq)** = $\beta_{intercept} + I_{male} * \beta_{sexmale} + I_{gradenA} * \beta_{gradenA} + I_{likeyes} * \beta_{likeyes} + I_{sexmale:gradenA} * \beta_{sexmale:gradenA} + I_{sexmale:likeyes} * \beta_{sexmale:likeyes} + I_{gradenA:likeyes} * \beta_{gradenA:likeyes}$

Model Term	Description
log(Freq)	predicted log(Freq) where Freq represents the number of individuals in a category
$\beta_{intercept}$	predicted frequency for sex = female, grade = A, like = no
$\beta_{sexmale}$	change in log(Freq) if sex = male
$\beta_{gradenA}$	change in log(Freq) if grade = nA
$\beta_{likeyes}$	change in log(Freq) if like = yes
$\beta_{sexmale:gradenA}$	change in log(Freq) if sex = male and grade = nA
$\beta_{sexmale:likeyes}$	change in log(Freq) if sex = male and like = yes
$\beta_{gradenA:likeyes}$	change in log(Freq) if grade = nA and like = yes
$\beta_{sexmale:gradenA:likeyes}$	change in log(Freq) if sex = male and grade = nA and like = yes
I_{male}	indicator variable, 1 if sex = male, 0 if sex = female
$I_{gradenA}$	indicator variable, 1 if grade = nA, 0 if grade = A
$I_{likeyes}$	indicator variable, 1 if like = yes, 0 if like = no
$I_{sexmale:gradenA}$	indicator variable, 1 if grade = nA and sex = male, 0 otherwise
$I_{sexmale:likeyes}$	indicator variable, 1 if sex = male and like = yes, 0 otherwise
$I_{gradenA:likeyes}$	indicator variable, 1 if grade = nA and like = yes, 0 otherwise
$I_{sexmale:gradenA:likeyes}$	indicator variable, 1 if sex = male and grade = nA and like = yes, 0 otherwise

3b.

Deviance

Model	Residual Deviance
2_1	85.152
2_2	92.031
3_1	0
3_2	6.8788

The deviance of Model 3_1 is 0. This is unsurprising given it is a saturated model and has one predictor variable for every category so it perfectly fits the data. Model 3_2 has a much lower residual deviance than model 2_1 and 2_2 which is unsurprising given Model 3_2 has seven β predictors (including intercept) as opposed to 2_1 which has four β predictors and 2_2 with three β predictors.

Wald Tests

The p-values of all non-intercept predictors in 2_1 are statistically significant (p-value below 0.1), while in 2_2 only one of two predictors is statistically significant ($\beta_{sexmale}$).

This contrasts with 3_1 and 3_2 which have a number of statistically insignificant predictors (according to a Wald test). Only three of seven predictors in 3_1 have p-values below 0.1, and only three of six predictors in 3_2 have p-values below 0.1. Given this, it may be worthwhile to drop some insignificant predictors from 3_1 and 3_2 in order to simplify the model and improve interpretability.

Interpretation

In interpreting model 2_1 and 2_2 we are able to see that there is a definite change in the relationship between sex and liking video games when grade is taken into account.

Comparing model 3_1 to 2_1 (the models with the lowest residuals for each type of regression) in terms of residual deviance we see that model 3_1 is statistically significantly (p-value of 0 on a deviance goodness of fit test) more accurate. Interpreting this would seem to suggest that using a three-way interaction is a superior way to model the relationship between grade, likes video games, and sex. This suggests there are many interactions occurring at once, and it would be too simplistic to use a model with only one interaction (such as model 2_1).

R CODE APPENDIX - Assn3R.R

David

Thu Mar 22 11:10:18 2018

```
# Assignment 3 David Veitch 1004657917

workdir<-("C:/Users/David/Google Drive/Documents/UofT/NonDegree/303/Assignmen
t3")
setwd(workdir)
dat <- read.csv("video.csv")

# 1A

# Create contingency table
cont_tab <- xtabs(~like+sex, data=dat)
cont_tab

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

# Run difference of proportions test and chi-sq test
prop.test(cont_tab, correct=FALSE)

## 
##  2-sample test for equality of proportions without continuity
##  correction
## 
## data:  cont_tab
## X-squared = 3.3314, df = 1, p-value = 0.06797
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.01414205  0.47128491
## sample estimates:
##    prop 1    prop 2 
## 0.6000000 0.3714286

fisher.test(cont_tab)

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

```r
## odds ratio
##   2.511179

# 1B

# Create contingency tables for different grades

# Contingency tables & testsfor grade == A
cont_tab_grade_A <- xtabs(~like+sex, data=subset(dat, grade=="A"))
cont_tab_grade_A

##      sex
## like  female male
##   no       5    1
##   yes      4   21

prop.test(cont_tab_grade_A, correct=FALSE)

## Warning in prop.test(cont_tab_grade_A, correct = FALSE): Chi-squared
## approximation may be incorrect

##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  cont_tab_grade_A
## X-squared = 10.648, df = 1, p-value = 0.001102
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.3423132 1.0000000
## sample estimates:
##    prop 1    prop 2
## 0.8333333 0.1600000

fisher.test(cont_tab_grade_A)

##
##  Fisher's Exact Test for Count Data
##
## data:  cont_tab_grade_A
## p-value = 0.003879
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##     1.877541 1284.874786
## sample estimates:
## odds ratio
##   22.44903

# Contingency tables & tests for grade == nA
cont_tab_grade_nA <- xtabs(~like+sex, data=subset(dat, grade=="nA"))
cont_tab_grade_nA
```

```
##      sex
## like  female male
##   no       7    7
##   yes     22   23
```

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

```
##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  cont_tab_grade_nA
## X-squared = 0.0052746, df = 1, p-value = 0.9421
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.2887695  0.3109917
## sample estimates:
##    prop 1    prop 2
## 0.5000000 0.4888889
```

```r
fisher.test(cont_tab_grade_nA)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  cont_tab_grade_nA
## p-value = 1
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.2637887 4.1391671
## sample estimates:
## odds ratio
##   1.044688
```

```r
# 2 A

# Create logistic regression models

# Model 2.1 to include interaction between sex and grade
model_2_1 <- glm(like~sex+grade+sex*grade,family=binomial, data=dat)
summary(model_2_1)
```

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

```
# Model 2.2 to not include interaction between sex and grade
model_2_2 <- glm(like~sex+grade,family=binomial, data=dat)
summary(model_2_2)
```

```
##
## Call:
## glm(formula = like ~ sex + grade, family = binomial, data = dat)
##
## 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
```

```
# Conduct Likelihood Ratio test using residual deviance
# Test is chi-squre with one degree of freedom
p_val = (1-pchisq(model_2_2$deviance - model_2_1$deviance,1))
p_val
```

```
## [1] 0.008722588
```

```r
# 3 A

# Create a 3-way contingency table and convert to a dataframe
cont_tab_3_way <- xtabs(~like+sex+grade, data=dat)
df_3_way <- as.data.frame(cont_tab_3_way)
df_3_way
```

```
##   like    sex grade Freq
## 1   no female     A    5
## 2  yes female     A    4
## 3   no   male     A    1
## 4  yes   male     A   21
## 5   no female    nA    7
## 6  yes female    nA   22
## 7   no   male    nA    7
## 8  yes   male    nA   23
```

```r
# Create models with counts as Poisson variables

# Model 3.1 three two-way interaction terms and one three-way interaction
model_3_1 <- glm(Freq~sex+grade+like+sex*grade*like,family=poisson, data=df_3
_way)
summary(model_3_1)
```

```
##
## Call:
## glm(formula = Freq ~ sex + grade + like + sex * grade * like,
##     family = poisson, data = df_3_way)
##
## 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
```

```r
# Model 3.2 three two-way interaction terms, no three-way interaction
model_3_2 <- glm(Freq~sex+grade+like+ sex:grade + sex:like + grade:like,famil
y=poisson, data=df_3_way)
summary(model_3_2)
```

```
##
## Call:
## glm(formula = Freq ~ sex + grade + like + sex:grade + sex:like +
##     grade:like, family = poisson, data = df_3_way)
##
## 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
# 3 B

# Compare models 2_2 and 3_1, chisq with 4 df
# Model 3 statistically significantly more accurate
p_val = (1-pchisq(model_2_1$deviance - model_3_1$deviance, 4))
p_val
```

```
## [1] 0
```