STA 138 Project 1
Are Teacher Assistants (TAs) Evaluated Fairly?

Christopher Wong
ID: 999234204

Introduction

Teacher Assistant (TA) evaluations are an important form of anonymous feedback from students. They allow students to freely speak their mind and rate the teaching assistant on a variety of topics. They are also taken into account when the department allocates TAships in the future and since they are paid, can negatively impact a TA's financial sitaution. The goal of this report is to find if TAs are being fairly evaluated or if there are external factors active.

Material and Methods

My data is taken from: https://archive.ics.uci.edu/ml/datasets/Teaching+Assistant+Evaluation

A description from the website follows: The data consist of evaluations of teaching performance over three regular semesters and two summer semesters of 151 teaching assistant (TA) assignments at the Statistics Department of the University of Wisconsin-Madison.

Attribute Information:

1. Whether of not the TA is a native English speaker (binary)
   1=English speaker, 2=non-English speaker
2. Course instructor (categorical, 25 categories)
3. Course (categorical, 26 categories)
4. Summer or regular semester (binary) 1=Summer, 2=Regular
5. Class size (numerical)
6. Class attribute (categorical) 1=Low, 2=Medium, 3=High

Since Course instructor and Course contained 25 and 26 levels respectively, I left them out of my data for ease of analysis and because I was not interested in whether or not certain TAs were evaluated better than others. So, the data I did use consisted of 151 observations with the four attributes remaining: native English speaker, Summer or regular semester, Class size, and Class attribute.

Since Class size was a numerical variable, I categorized it according to the criteria:

- (0, 20] = Small
- [20, 40] = Medium
- [41, 66] = Large

To analyze my data I used loglinear models to fit my data followed by 2-dimensional contingency tables for testing independence/associations between pairs of factors.

Results

To start my analysis, I fit a homogeneous model to my data:

	Estimate	Std. Error	z value	Pr(>|z|)
EnglishYes:TypeSummer	21.0778523	2.956478e+04	7.129379e-04	0.99943116
EnglishYes:SizeMedium	-1.2047651	8.752406e-01	-1.376496e+00	0.16866816
EnglishYes:SizeSmall	-3.1445198	1.393518e+00	-2.256533e+00	**0.02403726**
EnglishYes:AttributeLow	-2.5902672	1.287763e+00	-2.011447e+00	**0.04427824**
EnglishYes:AttributeMedium	-2.7080502	1.282359e+00	-2.111772e+00	**0.03470598**
TypeSummer:SizeMedium	19.0393548	2.956478e+04	6.439877e-04	0.99948617
TypeSummer:SizeSmall	22.6139265	2.956478e+04	7.648942e-04	0.99938970
TypeSummer:AttributeLow	-63.1591736	4.330945e+04	-1.458323e-03	0.99883643
TypeSummer:AttributeMedium	-19.3194407	3.898107e+04	-4.956108e-04	0.99960456
SizeMedium:AttributeLow	-0.3420133	7.605664e-01	-4.496824e-01	0.65293946
SizeSmall:AttributeLow	-0.7291564	8.049782e-01	-9.058089e-01	0.36503703
SizeMedium:AttributeMedium	-0.8448470	7.636365e-01	-1.106347e+00	0.26857631
SizeSmall:AttributeMedium	-0.7166416	7.888914e-01	-9.084161e-01	0.36365845
EnglishYes:TypeSummer:SizeMedium	-19.6756758	2.956478e+04	-6.655107e-04	0.99946900
EnglishYes:TypeSummer:SizeSmall	-19.1384633	2.956478e+04	-6.473400e-04	0.99948350
EnglishYes:TypeSummer:AttributeLow	43.0547735	2.975049e+04	1.447196e-03	0.99884531
EnglishYes:TypeSummer:AttributeMedium	-1.2127840	1.734400e+00	-6.992529e-01	0.48439397
TypeSummer:SizeMedium:AttributeLow	-0.4304490	4.929475e+04	-8.732147e-06	0.99999303
TypeSummer:SizeSmall:AttributeLow	39.8395763	3.785713e+04	1.052366e-03	0.99916033
TypeSummer:SizeMedium:AttributeMedium	20.6872034	3.898107e+04	5.306987e-04	0.99957656
TypeSummer:SizeSmall:AttributeMedium	17.9805688	3.898107e+04	4.612641e-04	0.99963196
EnglishYes:SizeMedium:AttributeLow	0.7993000	1.559716e+00	5.124652e-01	0.60832546
EnglishYes:SizeSmall:AttributeLow	-18.1689996	2.103677e+04	-8.636781e-04	0.99931088
EnglishYes:SizeMedium:AttributeMedium	1.5910693	1.511066e+00	1.052945e+00	0.29236619
EnglishYes:SizeSmall:AttributeMedium	3.0117102	1.905269e+00	1.580727e+00	0.11394038

The bolded p-values are significant at level $\alpha = 0.05$. This suggests that there is dependence between English and Attribute and maybe between English and Size. The p-value for this homogeneous model is 0.7251979 so it is a good model for our data, but its interpretation is meaningless.

I fit a model with the interaction between English and Attribute, and English and Size.

```
##
## Call:
## glm(formula = count ~ English + Type + Size + Attribute + English:Size +
##     English:Attribute, family = poisson, data = tae.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4582  -0.8557  -0.2796   0.2937   3.6204
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 1.5528     0.2691   5.771 7.89e-09 ***
## EnglishYes                 -0.1156     0.4661  -0.248 0.804215
## TypeSummer                 -1.7165     0.2265  -7.579 3.47e-14 ***
## SizeMedium                  1.0116     0.2611   3.874 0.000107 ***
## SizeSmall                   0.8544     0.2670   3.200 0.001373 **
## AttributeLow                0.2578     0.2283   1.129 0.258836
## AttributeMedium             0.2578     0.2283   1.129 0.258836
```

```
## EnglishYes:SizeMedium       -0.6061     0.5258  -1.153 0.249040
## EnglishYes:SizeSmall        -0.7366     0.5544  -1.329 0.183967
## EnglishYes:AttributeLow     -1.5388     0.5547  -2.774 0.005537 **
## EnglishYes:AttributeMedium  -1.3564     0.5238  -2.590 0.009608 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 237.194  on 35  degrees of freedom
## Residual deviance:  64.072  on 25  degrees of freedom
## AIC: 166.28
##
## Number of Fisher Scoring iterations: 6
```

If we look at the p-values of coefficients for this model, we find that the interaction between English and Size is not significant anymore. That is, we can remove this interaction effect and maybe get a better model. Also, the p-value of this model is 2.8078075×10^{-5}.

Now I fit a model with the interaction between English and Attribute, and three-way interactions interpreted as conditional independence of Type and Size given English and Attribute.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	7.911e-01	5.864e-01	1.349	0.1773
EnglishYes	5.952e-01	7.337e-01	0.811	0.4173
TypeSummer	-1.609e+00	1.095e+00	-1.469	0.1418
SizeMedium	1.099e+00	1.155e+00	0.951	0.3414
SizeSmall	6.931e-01	1.225e+00	0.566	0.5714
AttributeLow	1.288e+00	6.848e-01	1.881	0.0599 .
AttributeMedium	1.285e+00	6.767e-01	1.900	0.0575 .
EnglishYes:AttributeLow	-3.185e+00	1.340e+00	-2.376	0.0175 *
EnglishYes:AttributeMedium	-2.854e+00	1.298e+00	-2.198	0.0279 *
EnglishNo:TypeSummer:AttributeHigh	5.878e-01	1.162e+00	0.506	0.6131
EnglishYes:TypeSummer:AttributeHigh	9.163e-01	1.204e+00	0.761	0.4467
EnglishNo:TypeSummer:AttributeLow	-1.928e+01	3.135e+03	-0.006	0.9951
EnglishYes:TypeSummer:AttributeLow	1.204e+00	1.426e+00	0.844	0.3985
EnglishNo:TypeSummer:AttributeMedium	-4.447e-01	1.194e+00	-0.372	0.7096
EnglishYes:TypeSummer:AttributeMedium	NA	NA	NA	NA
EnglishNo:SizeMedium:AttributeHigh	3.677e-01	1.320e+00	0.278	0.7806
EnglishYes:SizeMedium:AttributeHigh	-9.445e-01	1.282e+00	-0.737	0.4612
EnglishNo:SizeSmall:AttributeHigh	1.099e+00	1.374e+00	0.799	0.4241
EnglishYes:SizeSmall:AttributeHigh	-8.755e-01	1.366e+00	-0.641	0.5217
EnglishNo:SizeMedium:AttributeLow	9.015e-16	1.225e+00	0.000	1.0000
EnglishYes:SizeMedium:AttributeLow	-4.055e-01	1.683e+00	-0.241	0.8096
EnglishNo:SizeSmall:AttributeLow	-2.877e-01	1.307e+00	-0.220	0.8258
EnglishYes:SizeSmall:AttributeLow	1.118e-15	1.732e+00	0.000	1.0000
EnglishNo:SizeMedium:AttributeMedium	-4.055e-01	1.225e+00	-0.331	0.7406
EnglishYes:SizeMedium:AttributeMedium	NA	NA	NA	NA
EnglishNo:SizeSmall:AttributeMedium	-5.716e-02	1.292e+00	-0.044	0.9647
EnglishYes:SizeSmall:AttributeMedium	NA	NA	NA	NA

If we look at the p-values of the interaction between English and Attribute, we find that Attribute is dependent on English. That is, the evaulation score of a TA is dependent on whether or not they are a native English speaker or not. Also, the p-value of this model is 8.9232104×10^{-4} which is larger than corresponding p-value of the model above indicating that this is a better fit. Note that while the p-value is less than $\alpha = 0.05$, this is the best model we can fit outside of the homogeneous model.

I use the AIC criterion to select the best model of the three presented. For AIC we have

```
##                    df      AIC
## homogeneous.model 32 146.2617
## tae.model1        11 166.2765
## tae.model4        24 161.4306
```

Our fitted model above with the interaction between English and Attribute, and three-way interactions interpreted as conditional independence of Type and Size given English and Attribute has the smallest AIC of the practical interpretation models (not homogeneous model). Hence, I choose the fitted model above.

We look at the contingency table of English and Attribute.

```
##         Attribute
## English  1  2  3
##       1  5  6 18
##       2 44 44 34
```

I perform a test for independence and find the p-value to be $0.0022546 < \alpha = 0.05$. So, we reject the null hypothesis of independence and conclude that English and Attribute are dependent.

I also examine why the model with both English and Attribute, and English and Size is a worse fit than our selected model. The contingency table of English and Size is given below

```
##         Size
## English Small Medium Large
##       1     9     12     8
##       2    47     55    20
```

and we see that the p-value $= 0.3660786 > \alpha = 0.05$. So, English and Size are independent of one another hence why the interaction between them did not work well with our model.

I also looked at my model's residuals and found that they are in the interval [-2, 2] indicating that the model fits.

```
##             1             2             3             4             5
##  1.500000e+00 -1.000000e+00 -1.000000e+00 -1.332268e-15 -1.000000e+00
##             6             7             8             9            10
## -1.000000e+00  6.666667e-01  2.960595e-16 -1.000000e+00 -1.000000e+00
##            11            12            13            14            15
##  6.666667e-01  2.220446e-16  2.000000e+00  5.529412e-01 -4.000000e-01
##            16            17            18            19            20
## -7.088989e-02 -1.000000e+00 -2.222222e-02  2.000000e-01  2.849003e-03
##            21            22            23            24            25
## -1.000000e+00 -1.000000e+00  2.000000e-01  1.282051e-01  1.400000e+00
##            26            27            28            29            30
##  8.888889e-01 -7.000000e-01 -3.200000e-01 -5.714286e-01 -1.000000e+00
##            31            32            33            34            35
##  2.857143e-01  3.600000e-01 -5.000000e-01 -1.000000e+00  2.500000e-01
##            36
##  3.600000e-01
```

Conclusion and Discussion

In summary, my chosen model for this data contains the main effects, interaction between English and Attribute, and three-way interactions interpreted as conditional independence of Type and Size given English and Attribute.

$$log(\mu_{ijkl}) = \lambda + \lambda_i^{English} + \lambda_j^{Type} + \lambda_k^{Size} + \lambda_l^{Attribute} + \lambda_{il}^{English,Attribute} + \lambda_{ijl}^{English,Type,Attribute} + \lambda_{ikl}^{English,Size,Attribute}$$

That is, the evaulation score of a TA is dependent on whether or not they are a native English speaker or not. In the case of TAs in the Statistics Department of the University of Wisconsin-Madison, we can conclude that native English speaking TAs are going to be rated higher than non-native English speaking TAs.

Of course, this project does not give a definitive answer of the general question of whether ALL TAs, regardless of school, department, etc., are evaluated fairly because of its shortcomings (small sample size, few factors, single sample from one university, small p-value of fitted model, counts of 0, etc.).

Code Appendix

```
setwd("C:/Users/Christopher/Desktop/STA 138/Project 1")

# https://archive.ics.uci.edu/ml/datasets/Teaching+Assistant+Evaluation
tae <- read.csv("C:/Users/Christopher/Desktop/STA 138/Project 1/tae.data", header=FALSE)

colnames(tae) <- c("English", "Instructor", "Course", "Type", "Size", "Attribute")

# We will only deal with English, Type, Size (categorized), and Attribute
tae <- tae[, -c(2, 3)]

# Small = 0-20, Medium = 21-40, Large = 41+
tae$Size <- cut(tae$Size, c(0, 20, 40, max(tae$Size)), labels = c("Small", "Medium", "Large"))

# English: 1 = native English speaker
# Type: 1 = Summer session
# Size: categorized into small, medium, large
# Attribute: 1 = Low, 2 = Medium, 3 = High

tae.data <- data.frame(expand.grid(English = factor(c("Yes", "No")),
                                   Type = factor(c("Summer", "Semester")),
                                   Size = factor(c("Small", "Medium", "Large")),
                                   Attribute = factor(c("Low", "Medium", "High"))),
                       count = c(2, 0, 0, 12,
                                 0, 0, 2, 24,
                                 0, 0, 1, 8,
                                 1, 3, 1, 14,
                                 0, 2, 3, 16,
                                 0, 0, 1, 9,
                                 4, 9, 1, 9,
                                 1, 0, 6, 13,
                                 1, 0, 5, 3))

homogeneous.model <- glm(count~English+Type+Size+Attribute+
                         English:Type+English:Size+English:Attribute+
                           Type:Size+Type:Attribute+Size:Attribute+
                         English:Type:Size+English:Type:Attribute+
                           Type:Size:Attribute+English:Size:Attribute,
                         family = poisson,
                         data = tae.data)

tae.model1 <- glm(count~English+Type+Size+Attribute+
                    English:Size+English:Attribute,
                  family = poisson,
                  data = tae.data)
summary(tae.model1)

tae.model4 <- glm(count~English+Type+Size+Attribute+
                    English:Attribute+English:Type:Attribute+English:Size:Attribute,
                  family = poisson,
                  data = tae.data)
```

```
AIC(homogeneous.model, tae.model1, tae.model4)

fit.table <- xtabs(~English+Attribute, data = tae)
fit.table

fit.table2 <- xtabs(~English+Size, data = tae)
fit.table2

tae.model4$residuals

### Extra code for more models
tae.model2 <- glm(count~English+Type+Size+Attribute+English:Attribute,
                  family = poisson,
                  data = tae.data)

tae.model3 <- glm(count~English+Type+Size+Attribute+
                    English:Size+English:Attribute+Size:Attribute+English:Size:Attribute,
                  family = poisson,
                  data = tae.data)

tae.model5 <- glm(count~English+Type+Size+Attribute+
                    English:Type:Size+Type:Size:Attribute,
                  family = poisson,
                  data = tae.data)
```