**Project 1: Does Being a Boy Scout Prevent Trouble?**

Spencer Phillips, Wade Lewis, Calvin Hsu

***Introduction***

In American society, we often see teenage males sent to juvenile prisons for engaging in criminal activity. As time went on, people started to question what factored into teenage males getting arrested so frequently. Many believe that socioeconomic status (SES) and/or being a Boy Scout factors into whether or not a teenage male turns into a delinquent in the future. The people want to know what truly factors into teenage males for committing crimes in order to raise the future males of America right.

***Materials and Methods***

The data used in this report was collected from a sample of 800 boys classified according to their SES, whether the boy belongs to the Boy Scouts, and the boys’ juvenile delinquency status. We hope to use this data to analyze if being a Boy Scout has any association with being a delinquent, independent of SES.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Socioeconomic Status (SES) | | |
| Delinquent | Scout | Low | Medium | High |
| No | No | 169 | 132 | 59 |
| Yes | 43 | 104 | 196 |
| Yes | No | 42 | 20 | 2 |
| Yes | 11 | 14 | 8 |

The data set we are using is categorical because the data consists of categorical variables which represents types of data that may be divided into groups. In our data set, we have three factors, which include socioeconomic status (an ordinal variable), Boy Scout status, and delinquency (both of which are nominal variables). This is important to analyze because this data could be used as a way to advertise that Boy Scout’s are statistically upstanding citizens.

Before we fit the data to the best log-linear models, we have to test if the data is independent first. If the data is independent, we try to fit the data into 8 different possible models and chose the model that has the lowest AIC and BIC as our best fit log linear model. The final model will be used to calculate the 95% confidence interval of the conditional odds ratio to help explain the meaning of the selected model in the context of the data.

***Results***

Independence:

We tested for mutual independence between all 3 factors using a chi-squared test. The null hypothesis is that all three variables are independent of one another. We found a chi-squared value of 196.36 with a degree of freedom of 6, (p-value < 2.2x10-16). We reject the null hypothesis, because there is strong evidence (based on the p-value) that the three factors are NOT mutually independent.

We separated the data by varying levels of SES creating three 2x2 tables (Low, Medium, High SES) to test for conditional independence. We then performed a chi-squared test, with a null hypothesis that Boy Scout status and delinquency are independent of one another. For each level of SES, we found that the other two factors are independent of each other.

Models

With the lowest AIC of 82.688 and the lowest BIC of 87.05 with a degree of freedom of 9, the best log-linear model includes all the main effects (Boy Scout, delinquency, SES) and the conditional association between delinquency and SES and the conditional association between being a boy scout and SES.

**Ln(μijk) = λ + λiB + λjD + λkS + λikBS + λjkDS**

B = Boy Scout effect

D = Delinquency effect

S = Socioeconomic Status effect

BS = Boy Scout x SES interaction effect

DS = Delinquency x SES interaction effect

The next closest model (AIC=84.67, BIC=89.52) is the above model but it also included the association between being a boy scout and delinquency. This model wasn’t included because the association between boy scouts and delinquency was not statistically significant (ANOVA Test between the two models yields p-value = 0.92). It was sufficient to account for the (SES x Delinquency) association and the (Boy Scout x SES) association.

95% Confidence Interval of the Conditional Odds Ratio

We took our prefered model and created a 95% confidence interval on the odds ratio in the interaction between high SES and being a Boy Scout. When compared to Low SES, the high SES bracket had an odds ratio of 8.640068 to 19.763366. This can be interpreted as those in the high income bracket are 8.6 to 19.7 times more likely to be a Boy Scout than those in the low income bracket: there is a strong positive association between being a Boy Scout and high socioeconomic status. We found this effect again when creating a 95% confidence interval for the odds ratio of the interaction between medium SES and being a Boy Scout. We found a 95% CI of 2 to 4.45. This is interpreted by saying that those in the middle class are two to 4.45 times more likely to be a Boy Scout than those in the lower SES bracket.

***Conclusion and Discussion***

We found our hypothesis that being a Boy Scout would lead to a reduction in delinquency rates is incorrect. As we stated above, the model was better without accounting for the (Boy Scout x Delinquency) association. However, we did find from our 95% Confidence Intervals of the odds ratio that those higher in SES they have higher odds of being a Boy Scout in comparison to other boys who are less well off, or are in the low SES category. This may be due to racial factors not measured in this data.

To better conclude that SES is the biggest contributor to teenage male delinquency, we might need to collect and test more reliable data that includes other factors. In this report, we used data collected from an unknown source from 800 different teenage males. We don’t know if these 800 teenagers are from one specific area of the United States or if these teenagers were from all over the US. There’s also many other factors which can play into teenage males becoming delinquents such as age, ethnicity, and the city the teenage males grew up in. In the data we used, there were only 2 factors that came into play. But if more and better data with many other factors were collected, maybe we can finally find the truth to raising our boys right.

**R Code**

### Input the table

deliquent=c("no","yes")

scout=c("no", "yes")

SES=c("low", "med","high")

table=expand.grid(deliquent=deliquent,scout=scout,SES=SES)

count=c(169,42,43,11,132,20,104,14,59,2,196,8)

table=cbind(table,count=count)

table

temp=xtabs(count~deliquent+scout+SES,table)

temp

table =ftable(temp)

base =glm(count~SES\*scout\*deliquent,family=poisson, data=table)

fit.SD.SB.DB=update(base,.~.-SES:deliquent:scout)

fit.SB.DB=update(fit.SD.SB.DB,.~.-SES:deliquent)

fit.SD.B=update(fit.SD.SB.DB,.~.-SES:scout-deliquent:scout)

fit.S.D.B=update(fit.SD.B,.~.-SES:deliquent)

fit.SB.SD = update(fit.SD.SB.DB,.~.-deliquent:scout)

AIC(fit.S.D.B,fit.SD.B,fit.SB.DB,fit.SD.SB.DB,base,fit.SB.SD)

BIC(fit.S.D.B,fit.SD.B,fit.SB.DB,fit.SD.SB.DB,base,fit.SB.SD)

summary(fit.SB.SD)

fitted(fit.SB.SD)

exp(coef(fit.SB.SD)["SESmed:deliquentyes"])

exp(coef(fit.SB.SD)["SEShigh:deliquentyes"])

est\_SD= coef(summary(fit.SB.SD))["SESmed:scoutyes","Estimate"]

se\_SD= coef(summary(fit.SB.SD))["SESmed:scoutyes","Std. Error"]

CI\_SD=c(exp(est\_SD-(1.96\*se\_SD)),exp(est\_SD+(1.96\*se\_SD)))

#evidence of a slight positive association between scouting and delinquency given SES

anova(fit.SB.SD,fit.SD.SB.DB, test="Chi")

#weak evidence of conditional association between scouting and delinquency given SES

basefit = glm(count~SES+scout+delinquent, family=poisson, data=table)#[1][2][3]  
fit.12 = glm(count~SES+scout+delinquent+delinquent:scout, family = poisson, data=table) #  
fit.23 = glm(count~SES+scout+delinquent+scout:SES, family=poisson, data=table)   
fit.13 = glm(count~SES+scout+delinquent+delinquent:SES, family=poisson, data=table)  
fit.12.23 = glm(count~SES+scout+delinquent+delinquent:scout+scout:SES, family=poisson, data=table)  
fit.12.13 = glm(count~SES+scout+delinquent+delinquent:scout+delinquent:SES, family=poisson, data=table)  
fit.13.23 = glm(count~SES+scout+delinquent+delinquent:SES+scout:SES, family=poisson, data=table)  
fit.12.23.13 = glm(count~SES+scout+delinquent+delinquent:SES+scout:SES+delinquent:scout, family=poisson, data=table)  
AIC(basefit,fit.12,fit.23,fit.13,fit.12.23,fit.12.13,fit.13.23,fit.12.23.13)  
BIC(basefit,fit.12,fit.23,fit.13,fit.12.23,fit.12.13,fit.13.23,fit.12.23.13)  
fit.SD.SB.DB=update(base,.~.-SES:delinquent:scout) #[12][23][13]