Relationships Between Alcohol, Cigarette, and Marijuana Use Among High School Students

Daniel Pinedo

Psych 308d: Assignment 2

Relationships Between Alcohol, Cigarette, and Marijuana Use Among High School Students **Results**

Data analysis is in Appendix A. Observations did not contain any missing parameters in the dataset. Analysis continued with tests of assumptions of adequate expected variable category frequency counts which passed with each category cell having over 5 observations, and independence of observations between variables which also passed.

Hypothesis 1 tested if there was a relationship between alcohol and cigarette use which was significant, χ2 (1) = 451.00 , *p* < .001, *Cramér’s V* = .45, indicating a medium effect size of discrepancy between compared expected and observed frequencies. The largest discrepancy was between the observed amount of students who did smoke cigarettes (*observed* = 46, *expected* = 215) but did not drink alcohol, which indicated that more students smoked cigarettes and drank alcohol than expected values if there was no relationship.

Hypothesis 2 tested if there was a relationship between alcohol and marijuana use which was significant, χ2 (1) = 259.00, *p* < .001, *Cramér’s V* = .34, indicating a medium effect size of discrepancy between compared expected and observed frequencies. The largest discrepancy was between the observed amount of students who did smoke marijuana (*observed* = 5, *expected* = 138) but did not drink alcohol, which indicates that more students smoked marijuana and drank alcohol than expected values if there was no relationship.

Hypothesis 3 tested if there was a relationship between cigarette and marijuana use which was significant, χ2 (1) = 642.00, *p* < .001, *Cramér’s V* = .53, indicating a large effect size of discrepancy between compared expected and observed frequencies. The largest discrepancy was between the observed amount of students who did smoke marijuana (*observed* = 46, *expected* = 329) but did not smoke cigarettes, which indicates that more students smoked marijuana and cigarettes than expected values if there was no relationship.

Hypothesis 4 tested if there was a three-way relationship between alcohol, cigarette, and marijuana use using generalized log-linear models. Since all pairwise comparisons were significant, the two models tested and compared were Model 1.) the combination of pairwise comparisons and Model 2.) three-way relationship between all variables (i.e. saturated model). Model 1 was a fit for observed scores, χ2 (1) = 0.37, *p* = .541, retain null hypothesis that expected frequencies satisfy our model of expected values. Model 2 was a fit for observed scores, χ2 (0) = 0.00, *p* = .999, retain null hypothesis that expected frequencies satisfy our model of expected values. Comparison of both models indicated that the saturated model (Model 2) was not a significantly better fit than Model 1, Δχ2 (1) = 0.37, *p* = .541. These results indicated that, for example, a student who used alcohol had estimated odds of having tried marijuana that are 19.81 times the estimated odds for students who did not use alcohol. The respective odds ratio for cigarettes and marijuana was 17.25, and alcohol and cigarettes was 7.80. Students who used cigarettes and alcohol had a 62% chance of using marijuana. Conversely, students who did not use cigarettes or alcohol had a 99% chance of not using marijuana.

**Discussion**

The results in support of Model 1 support neither the strict nor the linear view of substance use. The strict view was tested using Model 2, that alcohol use would lead to an increase in both cigarette and marijuana use, or a three-way relationship, which was not supported by these results. The lenient view was also not supported, because each of the possible pairwise relationships was significant contrary to lenient view claims. Results notwithstanding, the outcomes and interventions suggested for the strict view would likely be the best candidates as using one substance will likely lead to using another substance, even if the observed outcomes are not as extreme as those holding the strict view.

The major limitations of this study is that it was both cross-sectional and does not establish causality or directionality. We do not know if using one substance leads to using another, and we do not know which substance students tend to try before the others, if any. It would be very difficult to complete an experimental study as it would not be ethical, but substance use behaviors can be tracked over several time points. Future directions related to substance use in student populations should focus not just on use and prevention, but also environmental stressors that may lead to excessive use, and also decriminalizing use.

Appendix A

**Statistical Analysis in R**

Daniel Pinedo

April 9, 2019

At a recent school board meeting, concerns were raised about alcohol, cigarette, and marijuana use among local high school students. After a heated discussion, two extreme views about alcohol and drugs emerged: a strict view and a lenient view. Advocates of the strict view called for a no tolerance policy because they believe that use of one substance will lead to use of other substances. For example, in their view, students who drink are also more likely to smoke cigarettes and use marijuana. Advocates of the lenient view disagreed and don’t believe that the use of alcohol is related to the use of cigarettes or marijuana (or that the use of cigarettes is related to the use of marijuana).

So, the board designed a survey and asked high school seniors whether they had ever used alcohol, cigarettes, or marijuana.

The board has tasked you to examine the research questions below. They are interested in publishing what you find so they ask that you write up the results and discussion section to look at a sample of the report.

**Research Questions to Investigate:** *RQ1:* Investigation of three relationships proposed: (a) Is there a relationship between alcohol use and cigarette use? (b) Is there a relationship between alcohol use and marijuana use? (c) Is there a relationship between cigarette use and marijuana use?

*RQ2:* Is there statistical support for the strict view (students who drink are also more likely to smoke cigarettes and use marijuana) and/or the lenient view (use of alcohol is NOT related to the use of cigarettes or marijuana + use of cigarettes is NOT related to the use of marijuana)?

*RQ3:* Which model best explains the results? (This can be strict view, lenient view, or a view in between the two)?

*Please report all relevant statistics per APA format and write for a professional audience.*

#Load in your data  
AMCtable <- array(data = c(911, 44, 538, 456, 3, 2, 43, 279),   
 dim = c(2,2,2),   
 dimnames = list("cigarette" = c("yes","no"),  
 "marijuana" = c("yes","no"),  
 "alcohol" = c("yes","no")))  
  
dat <- as.data.frame(as.table(AMCtable))  
dat

## cigarette marijuana alcohol Freq  
## 1 yes yes yes 911  
## 2 no yes yes 44  
## 3 yes no yes 538  
## 4 no no yes 456  
## 5 yes yes no 3  
## 6 no yes no 2  
## 7 yes no no 43  
## 8 no no no 279

Load libraries

library(pacman)

## Warning: package 'pacman' was built under R version 3.5.3

p\_load(vcd, vcdExtra, MASS, jmv)

Transform data

# in order to run jmv analyses, must change contingency table to case form  
dat.case <- vcdExtra::expand.dft(dat)  
dim(dat)

## [1] 8 4

class(dat)

## [1] "data.frame"

dim(dat.case)

## [1] 2276 3

class(dat.case)

## [1] "data.frame"

Frequency analysis Assumptions

# 2x2x2 table  
# independence of observations assumption met  
# adequate expected cell counts assumption met  
desc <- jmv::descriptives(data = dat.case,   
 vars = c('cigarette', 'marijuana', 'alcohol'),   
 freq = TRUE)  
desc

##   
## DESCRIPTIVES  
##   
## Descriptives   
## ------------------------------------------------   
## cigarette marijuana alcohol   
## ------------------------------------------------   
## N 2276 2276 2276   
## Missing 0 0 0   
## Mean   
## Median   
## Minimum   
## Maximum   
## ------------------------------------------------   
##   
##   
## FREQUENCIES  
##   
## Frequencies of cigarette   
## --------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## --------------------------------------------------   
## no 781 34.3 34.3   
## yes 1495 65.7 100.0   
## --------------------------------------------------   
##   
##   
## Frequencies of marijuana   
## --------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## --------------------------------------------------   
## no 1316 57.8 57.8   
## yes 960 42.2 100.0   
## --------------------------------------------------   
##   
##   
## Frequencies of alcohol   
## --------------------------------------------------   
## Levels Counts % of Total Cumulative %   
## --------------------------------------------------   
## no 327 14.4 14.4   
## yes 1949 85.6 100.0   
## --------------------------------------------------

*RQ1: Tests of Independence* (a) Is there a relationship between alcohol use and cigarette use? YES

jmv::contTables(dat = dat.case,  
 rows = 'alcohol',  
 cols = 'cigarette',  
 exp = TRUE,  
 phiCra = TRUE)

##   
## CONTINGENCY TABLES  
##   
## Contingency Tables   
## -----------------------------------------------   
## alcohol no yes Total   
## -----------------------------------------------   
## no Observed 281 46 327   
## Expected 112 215   
##   
## yes Observed 500 1449 1949   
## Expected 669 1280   
##   
## Total Observed 781 1495 2276   
## Expected 781 1495   
## -----------------------------------------------   
##   
##   
## <U+03C7>² Tests   
## -------------------------------   
## Value df p   
## -------------------------------   
## <U+03C7>² 451 1 < .001   
## N 2276   
## -------------------------------   
##   
##   
## Nominal   
## ----------------------------   
## Value   
## ----------------------------   
## Phi-coefficient 0.445   
## Cramer's V 0.445   
## ----------------------------

1. Is there a relationship between alcohol use and marijuana use? YES

jmv::contTables(dat = dat.case,  
 rows = 'alcohol',  
 cols = 'marijuana',  
 exp = TRUE,  
 phiCra = TRUE)

##   
## CONTINGENCY TABLES  
##   
## Contingency Tables   
## -----------------------------------------------   
## alcohol no yes Total   
## -----------------------------------------------   
## no Observed 322 5 327   
## Expected 189 138   
##   
## yes Observed 994 955 1949   
## Expected 1127 822   
##   
## Total Observed 1316 960 2276   
## Expected 1316 960   
## -----------------------------------------------   
##   
##   
## <U+03C7>² Tests   
## -------------------------------   
## Value df p   
## -------------------------------   
## <U+03C7>² 259 1 < .001   
## N 2276   
## -------------------------------   
##   
##   
## Nominal   
## ----------------------------   
## Value   
## ----------------------------   
## Phi-coefficient 0.337   
## Cramer's V 0.337   
## ----------------------------

1. Is there a relationship between cigarette use and marijuana use? YES

jmv::contTables(dat = dat.case,  
 rows = 'cigarette',  
 cols = 'marijuana',  
 exp = TRUE,  
 phiCra = TRUE)

##   
## CONTINGENCY TABLES  
##   
## Contingency Tables   
## -------------------------------------------------   
## cigarette no yes Total   
## -------------------------------------------------   
## no Observed 735 46 781   
## Expected 452 329   
##   
## yes Observed 581 914 1495   
## Expected 864 631   
##   
## Total Observed 1316 960 2276   
## Expected 1316 960   
## -------------------------------------------------   
##   
##   
## <U+03C7>² Tests   
## -------------------------------   
## Value df p   
## -------------------------------   
## <U+03C7>² 642 1 < .001   
## N 2276   
## -------------------------------   
##   
##   
## Nominal   
## ----------------------------   
## Value   
## ----------------------------   
## Phi-coefficient 0.531   
## Cramer's V 0.531   
## ----------------------------

*RQ2:* Is there statistical support for the strict view (loglinear model of three-way relationship is a good fit) and/or the lenient view (loglinear model of all two way relationships not a good fit)?

Model1: null model - H0: all variables are orthogonal - NO

# Null hypothesis means that expected frequencies satisfy our model of expected values  
# Alternative Hypothesis means that difference between expected and observed frequencies is significant (indicates our model does not fit)  
  
# Observed = mytable  
mytable<- xtabs(dat$Freq ~ dat$cigarette + dat$marijuana + dat$alcohol) # table of observed values  
mytable

## , , dat$alcohol = yes  
##   
## dat$marijuana  
## dat$cigarette yes no  
## yes 911 538  
## no 44 456  
##   
## , , dat$alcohol = no  
##   
## dat$marijuana  
## dat$cigarette yes no  
## yes 3 43  
## no 2 279

# Expected = loglm  
model1 <- loglm(~dat$cigarette + dat$marijuana + dat$alcohol, mytable)  
summary(model1)

## Formula:  
## ~dat$cigarette + dat$marijuana + dat$alcohol  
## attr(,"variables")  
## list(dat$cigarette, dat$marijuana, dat$alcohol)  
## attr(,"factors")  
## dat$cigarette dat$marijuana dat$alcohol  
## dat$cigarette 1 0 0  
## dat$marijuana 0 1 0  
## dat$alcohol 0 0 1  
## attr(,"term.labels")  
## [1] "dat$cigarette" "dat$marijuana" "dat$alcohol"   
## attr(,"order")  
## [1] 1 1 1  
## attr(,"intercept")  
## [1] 1  
## attr(,"response")  
## [1] 0  
## attr(,".Environment")  
## <environment: R\_GlobalEnv>  
##   
## Statistics:  
## X^2 df P(> X^2)  
## Likelihood Ratio 1286.020 4 0  
## Pearson 1411.386 4 0

Model 2: H0: *Each two-way* relationship in pairs are best model fit - NO

model2a<- loglm(~dat$alcohol\*dat$cigarette + dat$alcohol\*dat$marijuana, mytable)  
summary(model2a)

## Formula:  
## ~dat$alcohol \* dat$cigarette + dat$alcohol \* dat$marijuana  
## attr(,"variables")  
## list(dat$alcohol, dat$cigarette, dat$marijuana)  
## attr(,"factors")  
## dat$alcohol dat$cigarette dat$marijuana  
## dat$alcohol 1 0 0  
## dat$cigarette 0 1 0  
## dat$marijuana 0 0 1  
## dat$alcohol:dat$cigarette dat$alcohol:dat$marijuana  
## dat$alcohol 1 1  
## dat$cigarette 1 0  
## dat$marijuana 0 1  
## attr(,"term.labels")  
## [1] "dat$alcohol" "dat$cigarette"   
## [3] "dat$marijuana" "dat$alcohol:dat$cigarette"  
## [5] "dat$alcohol:dat$marijuana"  
## attr(,"order")  
## [1] 1 1 1 2 2  
## attr(,"intercept")  
## [1] 1  
## attr(,"response")  
## [1] 0  
## attr(,".Environment")  
## <environment: R\_GlobalEnv>  
##   
## Statistics:  
## X^2 df P(> X^2)  
## Likelihood Ratio 497.3693 2 0  
## Pearson 443.7611 2 0

model2b<- loglm(~dat$alcohol\*dat$cigarette + dat$cigarette\*dat$marijuana, mytable) # - lowest chi-square  
summary(model2b)

## Formula:  
## ~dat$alcohol \* dat$cigarette + dat$cigarette \* dat$marijuana  
## attr(,"variables")  
## list(dat$alcohol, dat$cigarette, dat$marijuana)  
## attr(,"factors")  
## dat$alcohol dat$cigarette dat$marijuana  
## dat$alcohol 1 0 0  
## dat$cigarette 0 1 0  
## dat$marijuana 0 0 1  
## dat$alcohol:dat$cigarette dat$cigarette:dat$marijuana  
## dat$alcohol 1 0  
## dat$cigarette 1 1  
## dat$marijuana 0 1  
## attr(,"term.labels")  
## [1] "dat$alcohol" "dat$cigarette"   
## [3] "dat$marijuana" "dat$alcohol:dat$cigarette"   
## [5] "dat$cigarette:dat$marijuana"  
## attr(,"order")  
## [1] 1 1 1 2 2  
## attr(,"intercept")  
## [1] 1  
## attr(,"response")  
## [1] 0  
## attr(,".Environment")  
## <environment: R\_GlobalEnv>  
##   
## Statistics:  
## X^2 df P(> X^2)  
## Likelihood Ratio 92.01836 2 0  
## Pearson 80.81482 2 0

model2c<- loglm(~dat$alcohol\*dat$marijuana + dat$cigarette\*dat$marijuana, mytable)  
summary(model2c)

## Formula:  
## ~dat$alcohol \* dat$marijuana + dat$cigarette \* dat$marijuana  
## attr(,"variables")  
## list(dat$alcohol, dat$marijuana, dat$cigarette)  
## attr(,"factors")  
## dat$alcohol dat$marijuana dat$cigarette  
## dat$alcohol 1 0 0  
## dat$marijuana 0 1 0  
## dat$cigarette 0 0 1  
## dat$alcohol:dat$marijuana dat$marijuana:dat$cigarette  
## dat$alcohol 1 0  
## dat$marijuana 1 1  
## dat$cigarette 0 1  
## attr(,"term.labels")  
## [1] "dat$alcohol" "dat$marijuana"   
## [3] "dat$cigarette" "dat$alcohol:dat$marijuana"   
## [5] "dat$marijuana:dat$cigarette"  
## attr(,"order")  
## [1] 1 1 1 2 2  
## attr(,"intercept")  
## [1] 1  
## attr(,"response")  
## [1] 0  
## attr(,".Environment")  
## <environment: R\_GlobalEnv>  
##   
## Statistics:  
## X^2 df P(> X^2)  
## Likelihood Ratio 187.7543 2 0  
## Pearson 177.6149 2 0

Model 3: H0: *All two-way* relationships are best model fit i.e. alternative hypothesis is lenient model

model3 <- loglm(~dat$alcohol\*dat$cigarette + dat$alcohol\*dat$marijuana + dat$cigarette\*dat$marijuana, mytable)  
summary(model3)

## Formula:  
## ~dat$alcohol \* dat$cigarette + dat$alcohol \* dat$marijuana +   
## dat$cigarette \* dat$marijuana  
## attr(,"variables")  
## list(dat$alcohol, dat$cigarette, dat$marijuana)  
## attr(,"factors")  
## dat$alcohol dat$cigarette dat$marijuana  
## dat$alcohol 1 0 0  
## dat$cigarette 0 1 0  
## dat$marijuana 0 0 1  
## dat$alcohol:dat$cigarette dat$alcohol:dat$marijuana  
## dat$alcohol 1 1  
## dat$cigarette 1 0  
## dat$marijuana 0 1  
## dat$cigarette:dat$marijuana  
## dat$alcohol 0  
## dat$cigarette 1  
## dat$marijuana 1  
## attr(,"term.labels")  
## [1] "dat$alcohol" "dat$cigarette"   
## [3] "dat$marijuana" "dat$alcohol:dat$cigarette"   
## [5] "dat$alcohol:dat$marijuana" "dat$cigarette:dat$marijuana"  
## attr(,"order")  
## [1] 1 1 1 2 2 2  
## attr(,"intercept")  
## [1] 1  
## attr(,"response")  
## [1] 0  
## attr(,".Environment")  
## <environment: R\_GlobalEnv>  
##   
## Statistics:  
## X^2 df P(> X^2)  
## Likelihood Ratio 0.3739859 1 0.5408396  
## Pearson 0.4010998 1 0.5265218

Model 4: All two-way relationships *and the three-way* relationship i.e. strict model

#saturated model or "overfit model  
# this takes us one step past parsimony  
# this means that the three-way relationship does not add to the model  
  
# i.e. Chi-squared is zero  
# e.g., no degrees of freedom  
model4 <- loglm(~dat$cigarette\*dat$marijuana\*dat$alcohol, mytable)  
summary(model4)

## Formula:  
## ~dat$cigarette \* dat$marijuana \* dat$alcohol  
## attr(,"variables")  
## list(dat$cigarette, dat$marijuana, dat$alcohol)  
## attr(,"factors")  
## dat$cigarette dat$marijuana dat$alcohol  
## dat$cigarette 1 0 0  
## dat$marijuana 0 1 0  
## dat$alcohol 0 0 1  
## dat$cigarette:dat$marijuana dat$cigarette:dat$alcohol  
## dat$cigarette 1 1  
## dat$marijuana 1 0  
## dat$alcohol 0 1  
## dat$marijuana:dat$alcohol  
## dat$cigarette 0  
## dat$marijuana 1  
## dat$alcohol 1  
## dat$cigarette:dat$marijuana:dat$alcohol  
## dat$cigarette 1  
## dat$marijuana 1  
## dat$alcohol 1  
## attr(,"term.labels")  
## [1] "dat$cigarette"   
## [2] "dat$marijuana"   
## [3] "dat$alcohol"   
## [4] "dat$cigarette:dat$marijuana"   
## [5] "dat$cigarette:dat$alcohol"   
## [6] "dat$marijuana:dat$alcohol"   
## [7] "dat$cigarette:dat$marijuana:dat$alcohol"  
## attr(,"order")  
## [1] 1 1 1 2 2 2 3  
## attr(,"intercept")  
## [1] 1  
## attr(,"response")  
## [1] 0  
## attr(,".Environment")  
## <environment: R\_GlobalEnv>  
##   
## Statistics:  
## X^2 df P(> X^2)  
## Likelihood Ratio 0 0 1  
## Pearson 0 0 1

## Compare models

stats::anova(model1, model2a, model2b, model2c, model3, model4)

## LR tests for hierarchical log-linear models  
##   
## Model 1:  
## ~dat$cigarette + dat$marijuana + dat$alcohol   
## Model 2:  
## ~dat$alcohol \* dat$cigarette + dat$alcohol \* dat$marijuana   
## Model 3:  
## ~dat$alcohol \* dat$cigarette + dat$cigarette \* dat$marijuana   
## Model 4:  
## ~dat$alcohol \* dat$marijuana + dat$cigarette \* dat$marijuana   
## Model 5:  
## ~dat$alcohol \* dat$cigarette + dat$alcohol \* dat$marijuana + dat$cigarette \* dat$marijuana   
## Model 6:  
## ~dat$cigarette \* dat$marijuana \* dat$alcohol   
##   
## Deviance df Delta(Dev) Delta(df) P(> Delta(Dev)  
## Model 1 1286.0199544 4   
## Model 2 497.3692752 2 788.6506792 2 0.00000  
## Model 3 92.0183606 2 405.3509146 0 0.00000  
## Model 4 187.7543029 2 -95.7359423 0 1.00000  
## Model 5 0.3739859 1 187.3803170 1 0.00000  
## Model 6 0.0000000 0 0.3739859 1 0.54084  
## Saturated 0.0000000 0 0.0000000 0 1.00000

JMV Model comparisons

# note the similarities between 'Deviance' values and the model comparison stats with the loglm output.   
# the top table output is unknown - so look it up   
  
jmv::logLinear(  
 data = dat.case,  
 counts = NULL,  
 factors = c('cigarette', 'marijuana', 'alcohol'),  
 blocks = list(  
 list(  
 'cigarette', 'marijuana', 'alcohol'), # Model 1: null model  
 list(  
 c('alcohol', 'cigarette'), # Model 2b: alcohol and marijuana are independent  
 c('cigarette', 'marijuana')), # but alcohol/cigarette and cigarette/ marijuana are related  
 list(  
 c('alcohol', 'marijuana')), # Model 3: all two-way relationships - best fit  
 list(  
 c('cigarette', 'marijuana', 'alcohol'))), # Model 4: saturated model  
 refLevels = list(  
 list(  
 var = 'cigarette',  
 ref = 'no'),  
 list(  
 var = 'marijuana',  
 ref = 'no'),  
 list(  
 var = 'alcohol',  
 ref = 'no')),  
 modelTest = TRUE)

##   
## LOG-LINEAR REGRESSION  
##   
## Model Fit Measures   
## ------------------------------------------------------------------   
## Model Deviance AIC R²-McF <U+03C7>² df p   
## ------------------------------------------------------------------   
## 1 1286.020 1343.1 0.549 1565 3 < .001   
## 2 92.018 153.1 0.968 2759 5 < .001   
## 3 0.374 63.4 1.000 2851 6 < .001   
## 4 -1.17e-13 65.0 1.000 2851 7 < .001   
## ------------------------------------------------------------------   
##   
##   
## Model Comparisons   
## ---------------------------------------------------   
## Model Model <U+03C7>² df p   
## ---------------------------------------------------   
## 1 - 2 1194.002 2 < .001   
## 2 - 3 91.644 1 < .001   
## 3 - 4 0.374 1 0.541   
## ---------------------------------------------------   
##   
##   
## MODEL SPECIFIC RESULTS  
##   
## MODEL 1  
##   
## Model Coefficients   
## -------------------------------------------------------   
## Predictor Estimate SE Z p   
## -------------------------------------------------------   
## Intercept 4.173 0.0650 64.23 < .001   
## cigarette:   
## yes  no 0.649 0.0442 14.71 < .001   
## marijuana:   
## yes  no -0.315 0.0424 -7.43 < .001   
## alcohol:   
## yes  no 1.785 0.0598 29.87 < .001   
## -------------------------------------------------------   
##   
##   
## MODEL 2  
##   
## Model Coefficients   
## -------------------------------------------------------------------   
## Predictor Estimate SE Z p   
## -------------------------------------------------------------------   
## Intercept 5.578 0.0603 92.46 < .001   
## cigarette:   
## yes  no -2.694 0.1626 -16.57 < .001   
## marijuana:   
## yes  no -2.771 0.1520 -18.23 < .001   
## alcohol:   
## yes  no 0.576 0.0746 7.73 < .001   
## alcohol:cigarette:   
## (yes  no):(yes  no) 2.874 0.1673 17.18 < .001   
## cigarette:marijuana:   
## (yes  no):(yes  no) 3.224 0.1610 20.03 < .001   
## -------------------------------------------------------------------   
##   
##   
## MODEL 3  
##   
## Model Coefficients   
## -------------------------------------------------------------------   
## Predictor Estimate SE Z p   
## -------------------------------------------------------------------   
## Intercept 5.633 0.0597 94.36 < .001   
## cigarette:   
## yes  no -1.887 0.1627 -11.60 < .001   
## marijuana:   
## yes  no -5.309 0.4752 -11.17 < .001   
## alcohol:   
## yes  no 0.488 0.0758 6.44 < .001   
## alcohol:cigarette:   
## (yes  no):(yes  no) 2.055 0.1741 11.80 < .001   
## cigarette:marijuana:   
## (yes  no):(yes  no) 2.848 0.1638 17.38 < .001   
## alcohol:marijuana:   
## (yes  no):(yes  no) 2.986 0.4647 6.43 < .001   
## -------------------------------------------------------------------   
##   
##   
## MODEL 4  
##   
## Model Coefficients   
## -------------------------------------------------------------------------------   
## Predictor Estimate SE Z p   
## -------------------------------------------------------------------------------   
## Intercept 5.631 0.0599 94.060 < .001   
## cigarette:   
## yes  no -1.870 0.1638 -11.414 < .001   
## marijuana:   
## yes  no -4.938 0.7096 -6.959 < .001   
## alcohol:   
## yes  no 0.491 0.0760 6.464 < .001   
## alcohol:cigarette:   
## (yes  no):(yes  no) 2.035 0.1758 11.580 < .001   
## cigarette:marijuana:   
## (yes  no):(yes  no) 2.275 0.9275 2.453 0.014   
## alcohol:marijuana:   
## (yes  no):(yes  no) 2.600 0.7270 3.576 < .001   
## cigarette:marijuana:alcohol:   
## (yes  no):(yes  no):(yes  no) 0.590 0.9424 0.626 0.532   
## -------------------------------------------------------------------------------

Model performance - expected values, deviations, and odds glm model

dat[,-4] <- lapply(dat[,-4], relevel, ref = "no") # relevel reference group (intercept) to "no"  
  
mod1 <- glm(Freq ~ alcohol + marijuana + cigarette, data = dat, family = poisson) # orthogonal model  
mod3 <- glm(Freq ~ alcohol\*cigarette + alcohol\*marijuana + cigarette\*marijuana, data = dat, family = poisson) # best fit model  
summary(mod3)

##   
## Call:  
## glm(formula = Freq ~ alcohol \* cigarette + alcohol \* marijuana +   
## cigarette \* marijuana, family = poisson, data = dat)  
##   
## Deviance Residuals:   
## 1 2 3 4 5 6 7   
## 0.02044 -0.09256 -0.02658 0.02890 -0.33428 0.49134 0.09452   
## 8   
## -0.03690   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 5.63342 0.05970 94.361 < 2e-16 \*\*\*  
## alcoholyes 0.48772 0.07577 6.437 1.22e-10 \*\*\*  
## cigaretteyes -1.88667 0.16270 -11.596 < 2e-16 \*\*\*  
## marijuanayes -5.30904 0.47520 -11.172 < 2e-16 \*\*\*  
## alcoholyes:cigaretteyes 2.05453 0.17406 11.803 < 2e-16 \*\*\*  
## alcoholyes:marijuanayes 2.98601 0.46468 6.426 1.31e-10 \*\*\*  
## cigaretteyes:marijuanayes 2.84789 0.16384 17.382 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 2851.46098 on 7 degrees of freedom  
## Residual deviance: 0.37399 on 1 degrees of freedom  
## AIC: 63.417  
##   
## Number of Fisher Scoring iterations: 4

fittedmod3 <- as.data.frame(fitted(mod3))  
  
tab3 <- cbind(mod3$data, fittedmod3)  
tab3$Dev <- tab3$Freq - tab3$`fitted(mod3)`  
tab3

## cigarette marijuana alcohol Freq fitted(mod3) Dev  
## 1 yes yes yes 911 910.38317 0.6168303  
## 2 no yes yes 44 44.61683 -0.6168303  
## 3 yes no yes 538 538.61683 -0.6168303  
## 4 no no yes 456 455.38317 0.6168303  
## 5 yes yes no 3 3.61683 -0.6168303  
## 6 no yes no 2 1.38317 0.6168303  
## 7 yes no no 43 42.38317 0.6168303  
## 8 no no no 279 279.61683 -0.6168303

# odds someone used [X] using orthogonal model (Z to 1 ratio)  
exp(coef(mod1)[2])

## alcoholyes   
## 5.960245

exp(coef(mod1)[3])

## marijuanayes   
## 0.7294833

exp(coef(mod1)[4])

## cigaretteyes   
## 1.914213

# for X:Y output Z  
# Student who used X have estimated odds of having tried Y that are Z times the estimated odds for students who did not use X.  
exp(coef(mod3)["alcoholyes:cigaretteyes"])

## alcoholyes:cigaretteyes   
## 7.803201

exp(coef(mod3)["alcoholyes:marijuanayes"])

## alcoholyes:marijuanayes   
## 19.80658

exp(coef(mod3)["cigaretteyes:marijuanayes"])

## cigaretteyes:marijuanayes   
## 17.25133

#1/exp(coef(mod3))

## Proportions broken down by alcohol use

prop.table(as.table(AMCtable), margin = c(1,3))

## , , alcohol = yes  
##   
## marijuana  
## cigarette yes no  
## yes 0.628709455 0.371290545  
## no 0.088000000 0.912000000  
##   
## , , alcohol = no  
##   
## marijuana  
## cigarette yes no  
## yes 0.065217391 0.934782609  
## no 0.007117438 0.992882562

from students who used cigarettes and alcohol, 62% used marijuana. Conversely, from students who did not use cigarettes or alcohol, 99% did not use marijuana.