ANLY 545 - Homework #8

Navneet Dhull

August 5, 2018

Exercise 9.2

Consider the data set DaytonSurvey (described in Example 2.6), giving results of a survey of use of alcohol (A), cigarettes (C), and marijuana (M) among high school seniors. For this exercise, ignore the variables sex and race, by working with the marginal table Dayton.ACM, a $2 \times 2 \times 2$ table in frequency data frame form. Find the following models: ??? (a) Joint independence, [AC][M]. ??? (b) Conditional independence [AM][CM] ??? (c) Homogenous model [AC][AM][CM]. ??? (d) Prepare a table giving the goodness-of-fit tests for these models, as well as the model of mutual independence, [A][C][M], and the saturated model, [ACM]. Hint: anova() is useful here. Which model appears to give the most reasonable fit? Explain your interpretation.

```
library(vcd)
```

```
## Loading required package: grid
```

```
library(vcdExtra)
```

```
## Loading required package: gnm
```

```
library(effects)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'carData'
```

```
## The following object is masked from 'package:vcdExtra':
##
##     Burt
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```
library(VGAM)
```

```
## Loading required package: stats4
```

```
## Loading required package: splines
```

```
library(nnet)
library(car)
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:VGAM':
##
##     logit
```

```
library(MASS)


data("DaytonSurvey",package = "vcdExtra")
str(DaytonSurvey)
```

```
## 'data.frame':    32 obs. of  6 variables:
##  $ cigarette: Factor w/ 2 levels "Yes","No": 1 2 1 2 1 2 1 2 1 2 ...
##  $ alcohol  : Factor w/ 2 levels "Yes","No": 1 1 2 2 1 1 2 2 1 1 ...
##  $ marijuana: Factor w/ 2 levels "Yes","No": 1 1 1 1 2 2 2 2 1 1 ...
##  $ sex      : Factor w/ 2 levels "female","male": 1 1 1 1 1 1 1 1 2 2 ...
##  $ race     : Factor w/ 2 levels "white","other": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Freq     : num  405 13 1 1 268 218 17 117 453 28 ...
Dayton.ACM <-aggregate(Freq~cigarette +alcohol +marijuana, data = DaytonSurvey, FUN = sum)

#a
Dayton.loglm <- loglm(Freq ~ (cigarette *alcohol) + marijuana, data= Dayton.ACM, param= TRUE, fitted= T
Dayton.loglm

## Call:
## loglm(formula = Freq ~ (cigarette * alcohol) + marijuana, data = Dayton.ACM,
##     param = TRUE, fitted = TRUE)
##
## Statistics:
##                     X^2 df P(> X^2)
## Likelihood Ratio 843.8266  3        0
## Pearson          704.9071  3        0
mosaic(Dayton.loglm, main="Model [AC] [M]", labeling=labeling_residuals)
```
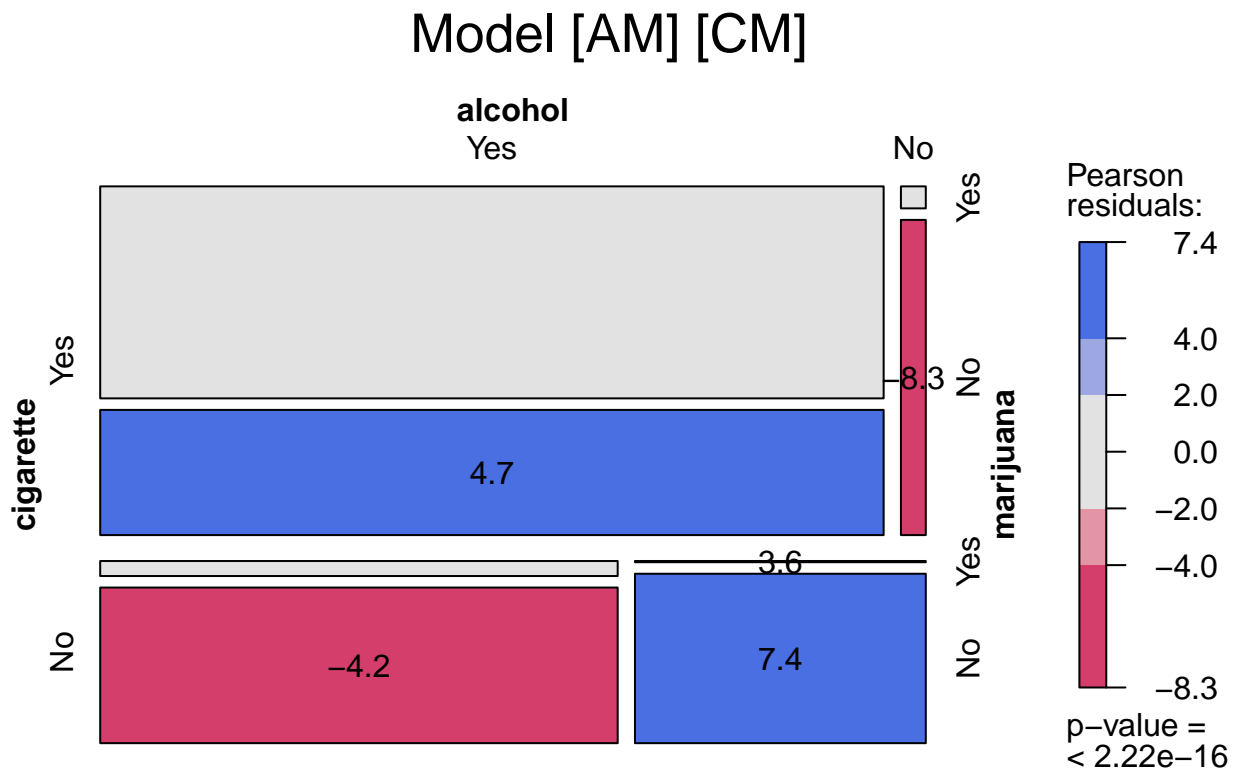

Model [AC] [M]
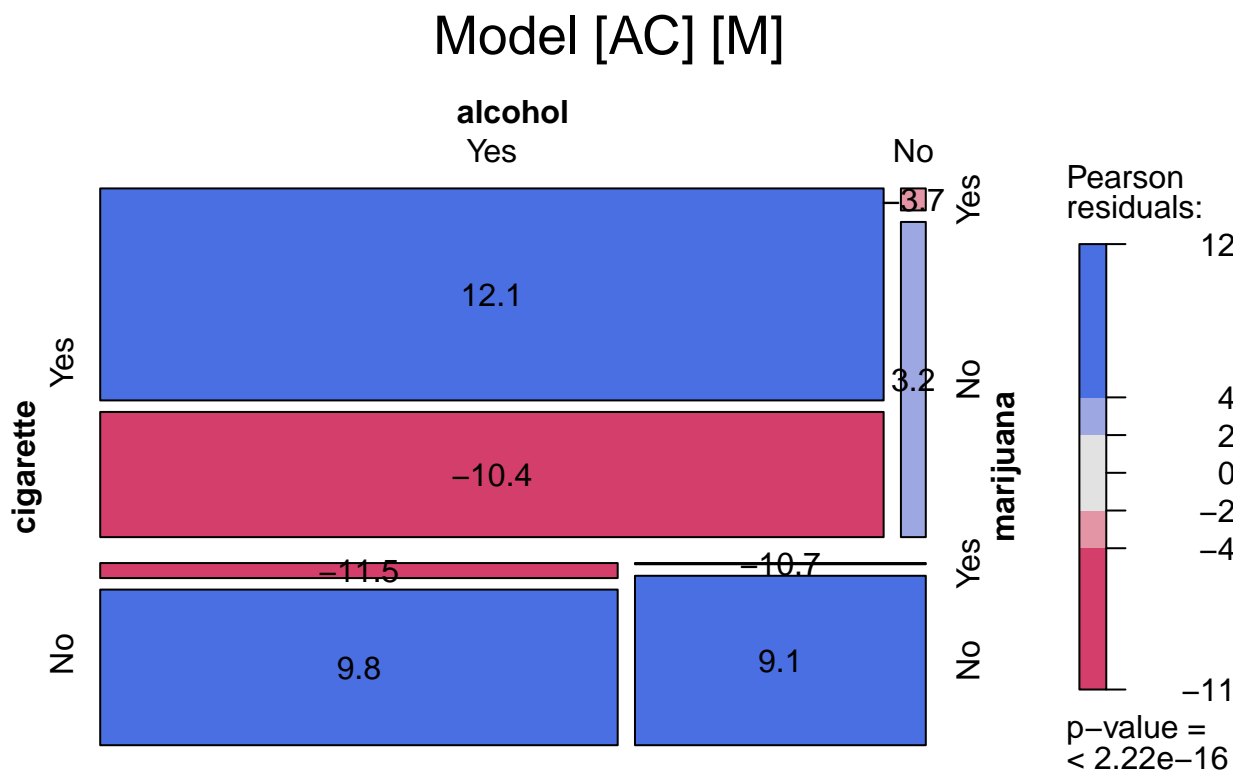
```
#b
Dayton.loglm1 <- loglm(Freq ~ (cigarette + alcohol)*marijuana, data= Dayton.ACM, param= TRUE, fitted= T
Dayton.loglm1
```
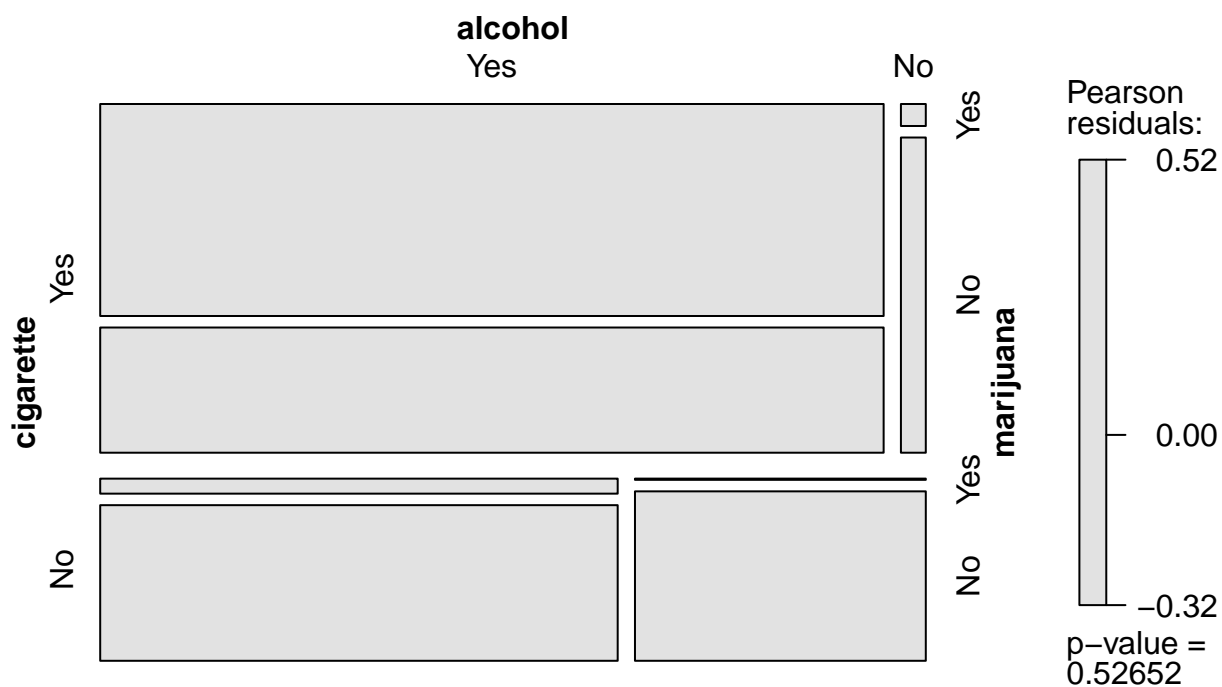
```
## Call:
## loglm(formula = Freq ~ (cigarette + alcohol) * marijuana, data = Dayton.ACM,
##      param = TRUE, fitted = TRUE)
##
## Statistics:
##                        X^2 df P(> X^2)
## Likelihood Ratio 187.7543  2        0
## Pearson          177.6149  2        0
```

```
mosaic(Dayton.loglm1, main="Model [AM] [CM]", labeling=labeling_residuals)
```


Model [AM] [CM]

```
#c
Dayton.loglm2 <- loglm(Freq ~ (cigarette *alcohol) + marijuana*(alcohol +cigarette), data= Dayton.ACM,
Dayton.loglm2
```

```
## Call:
## loglm(formula = Freq ~ (cigarette * alcohol) + marijuana * (alcohol +
##      cigarette), data = Dayton.ACM, param = TRUE, fitted = TRUE)
##
## Statistics:
##                        X^2 df  P(> X^2)
## Likelihood Ratio 0.3739859  1 0.5408396
## Pearson          0.4010998  1 0.5265218
```

```
mosaic(Dayton.loglm2, main="Model AC] [AM] [CM]",labeling=labeling_residuals)
```
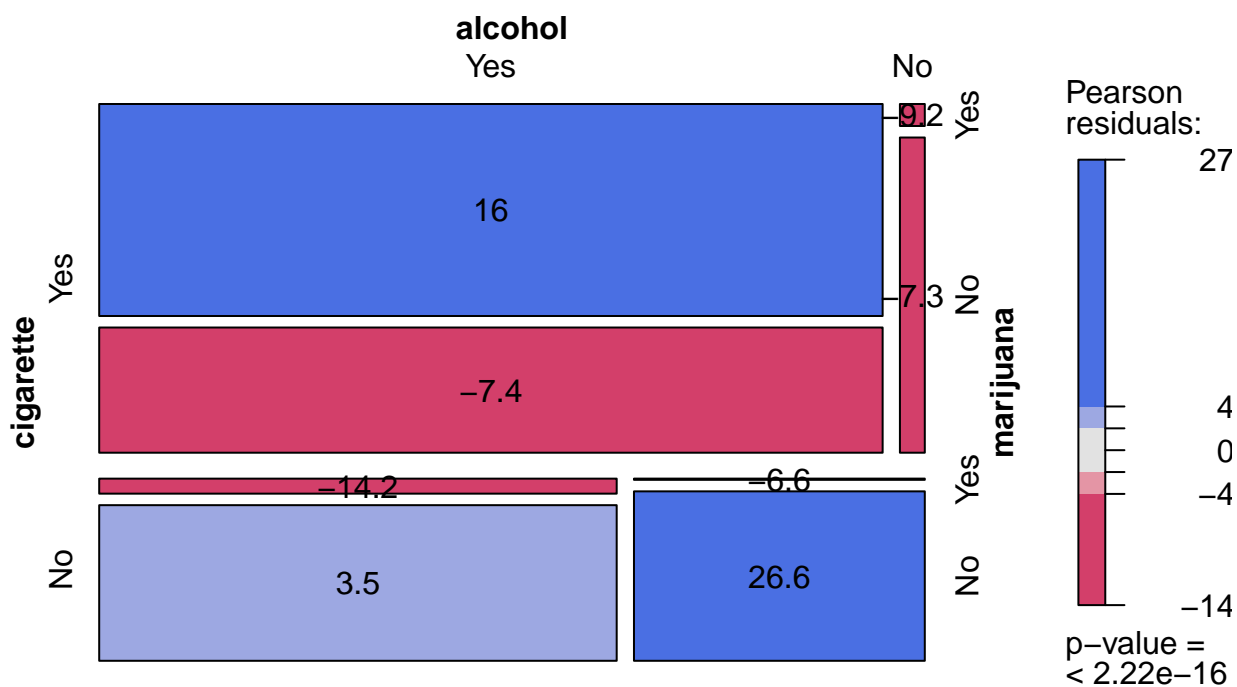
Model AC] [AM] [CM]


```
#d
Dayton.loglm3 <- loglm(Freq ~ cigarette + alcohol + marijuana, data= Dayton.ACM, param= TRUE, fitted= T
Dayton.loglm3

## Call:
## loglm(formula = Freq ~ cigarette + alcohol + marijuana, data = Dayton.ACM,
##     param = TRUE, fitted = TRUE)
##
## Statistics:
##                       X^2 df P(> X^2)
## Likelihood Ratio 1286.020  4        0
## Pearson          1411.386  4        0

mosaic(Dayton.loglm3, main="Model [A] [C] [M]",labeling=labeling_residuals)
```

Model [A] [C] [M]


```r
Dayton.loglm4 <- loglm(Freq ~ (cigarette*alcohol*marijuana), data= Dayton.ACM, param= TRUE, fitted= TRU
Dayton.loglm4
```

```
## Call:
## loglm(formula = Freq ~ (cigarette * alcohol * marijuana), data = Dayton.ACM,
##     param = TRUE, fitted = TRUE)
##
## Statistics:
##                   X^2 df P(> X^2)
## Likelihood Ratio   0   0        1
## Pearson            0   0        1
```

```r
mosaic(Dayton.loglm4, main="Model [ACM]",labeling=labeling_residuals)
```

```
## Warning in legend(residuals, gpfun, residuals_type): All residuals are
## zero.
```

Model [ACM]

alcohol

Yes   No

Pearson
residuals:

— 0

p–value =
1
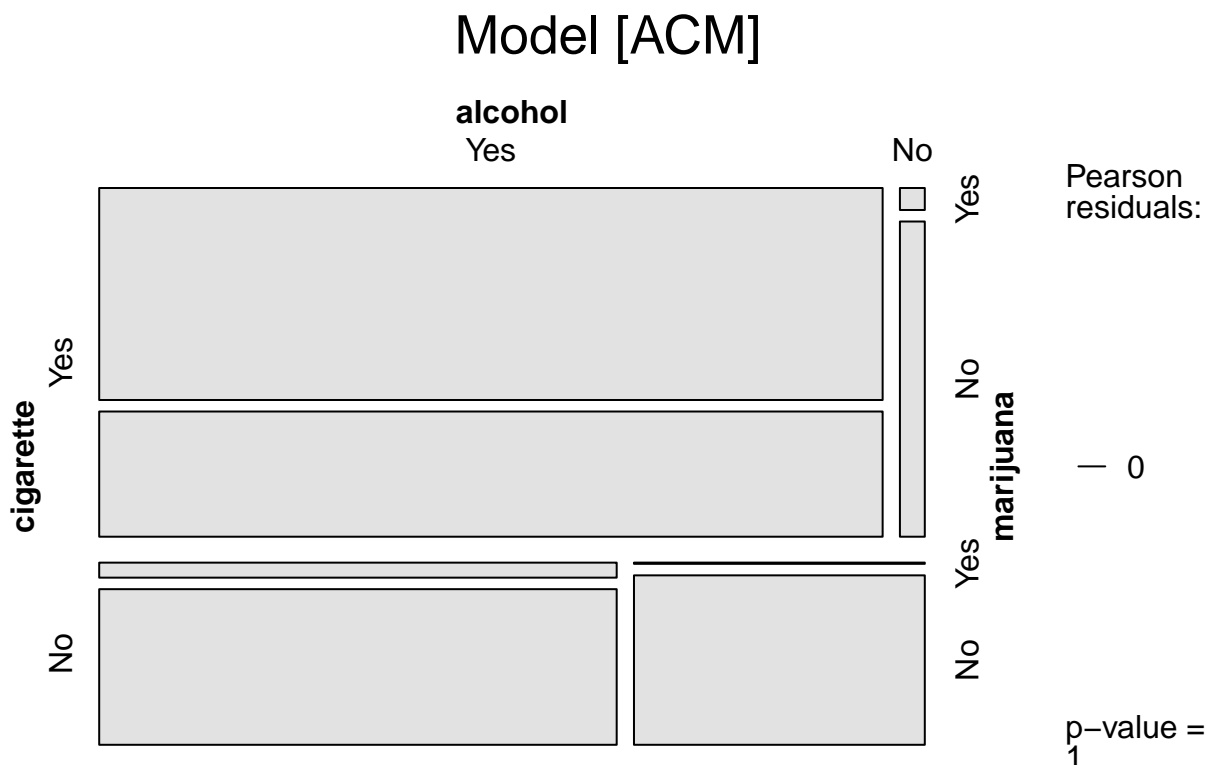
```r
anova(Dayton.loglm, Dayton.loglm1, Dayton.loglm2, Dayton.loglm3, Dayton.loglm4)
```

```
## LR tests for hierarchical log-linear models
##
## Model 1:
##  Freq ~ cigarette + alcohol + marijuana
## Model 2:
##  Freq ~ (cigarette * alcohol) + marijuana
## Model 3:
##  Freq ~ (cigarette + alcohol) * marijuana
## Model 4:
##  Freq ~ (cigarette * alcohol) + marijuana * (alcohol + cigarette)
## Model 5:
##  Freq ~ (cigarette * alcohol * marijuana)
##
##              Deviance df  Delta(Dev) Delta(df) P(> Delta(Dev)
## Model 1    1286.0199544  4
## Model 2     843.8266437  3 442.1933108         1        0.00000
## Model 3     187.7543029  2 656.0723408         1        0.00000
## Model 4       0.3739859  1 187.3803170         1        0.00000
## Model 5       0.0000000  0   0.3739859         1        0.54084
## Saturated     0.0000000  0   0.0000000         0        1.00000
```

```r
#Results indicate that Homogenous model [AC][AM][CM] and  model of mutual independence, [A][C][M] and s
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that
generated the plot.