ANLY545-91-0 HW4

Group 1: Laxman Panthi, Fuyan Li, Akhilesh Mannava, Harsh Patel

9/18/2019

# HOMEWORK 4

## Exercise #1

Bertin (1983, pp. 30-31) used a 4-way table of frequencies of traffic accident victims in France in 1958 to illustrate his scheme for classifying data sets by numerous variables, each of which could have various types and could be assigned to various visual attributes. His data are contained in Accident in vcdExtra, a frequency data frame representing his 5 × 2 × 4 × 2 table of the variables age, result (died or injured), mode of transportation, and gender.

data('Accident', package='vcdExtra')  
str(Accident)

## 'data.frame': 80 obs. of 5 variables:  
## $ age : Ord.factor w/ 5 levels "0-9"<"10-19"<..: 5 5 5 5 5 5 5 5 5 5 ...  
## $ result: Factor w/ 2 levels "Died","Injured": 1 1 1 1 1 1 1 1 2 2 ...  
## $ mode : Factor w/ 4 levels "4-Wheeled","Bicycle",..: 4 4 2 2 3 3 1 1 4 4 ...  
## $ gender: Factor w/ 2 levels "Female","Male": 2 1 2 1 2 1 2 1 2 1 ...  
## $ Freq : int 704 378 396 56 742 78 513 253 5206 5449 ...

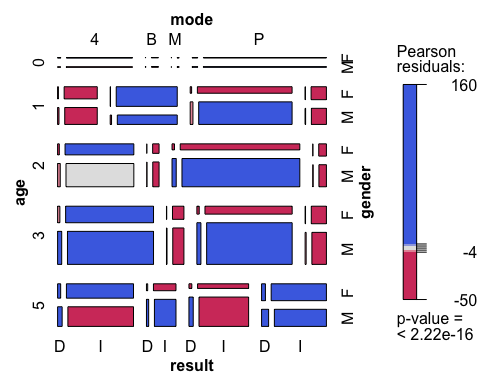
1. Use loglm() to fit the model of mutual independence, Freq ~ age+mode+gender+result to this data set.

f.lm <-loglm(Freq ~ age + mode+ gender+result, data=Accident)  
summary(f.lm)

## Formula:  
## Freq ~ age + mode + gender + result  
## attr(,"variables")  
## list(Freq, age, mode, gender, result)  
## attr(,"factors")  
## age mode gender result  
## Freq 0 0 0 0  
## age 1 0 0 0  
## mode 0 1 0 0  
## gender 0 0 1 0  
## result 0 0 0 1  
## attr(,"term.labels")  
## [1] "age" "mode" "gender" "result"  
## attr(,"order")  
## [1] 1 1 1 1  
## attr(,"intercept")  
## [1] 1  
## attr(,"response")  
## [1] 1  
## attr(,".Environment")  
## <environment: R\_GlobalEnv>  
## attr(,"predvars")  
## list(Freq, age, mode, gender, result)  
## attr(,"dataClasses")  
## Freq age mode gender result   
## "numeric" "ordered" "factor" "factor" "factor"   
##   
## Statistics:  
## X^2 df P(> X^2)  
## Likelihood Ratio 60320.05 70 0  
## Pearson 76865.31 70 0

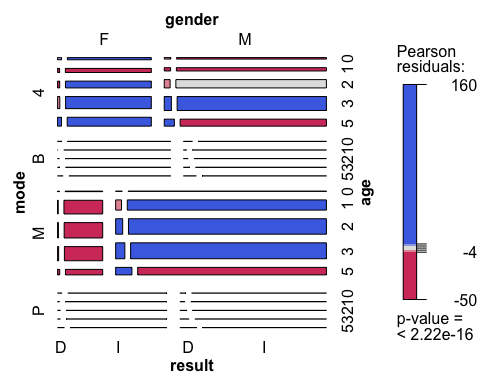
1. Use mosaic() to produce an interpretable mosaic plot of the associations among all variables under the model of mutual independence.

f.m1<-mosaic(f.lm , abbreviate\_labs=TRUE, clip=FALSE)

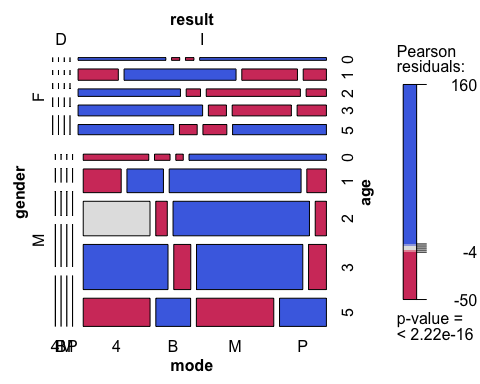


1. Try different orders of the variables in the mosaic. (Hint: the abbreviate component of the labeling\_args argument to mosaic() will be useful to avoid some overlap of the category labels.)

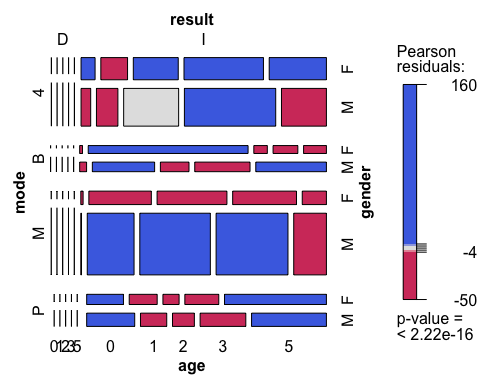
f.lm2 <- loglm(Freq ~ mode+gender+age+result, data=Accident)  
f.m2 <- mosaic(f.lm2, abbreviate\_labs=TRUE)



f.lm3 <- loglm(Freq ~ gender+result+age+mode, data=Accident)  
f.m3 <- mosaic(f.lm3, abbreviate\_labs=TRUE)



f.lm4 <- loglm(Freq ~ mode+result+gender+age, data=Accident)  
f.m4 <- mosaic(f.lm4, abbreviate\_labs=TRUE)



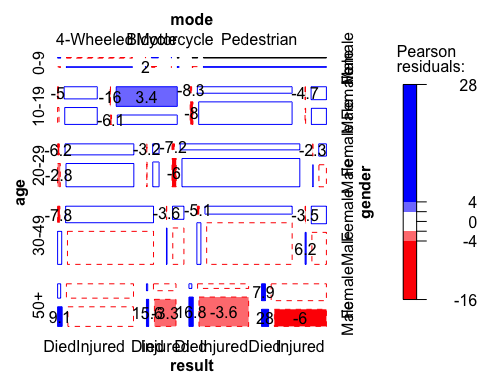
1. Treat result (“Died” vs. “Injured”) as the response variable, and fit the model Freq ~ age mode gender + result that asserts independence of result from all others jointly.

loglm(Freq ~ age \* mode \* gender + result, data = Accident)

## Call:  
## loglm(formula = Freq ~ age \* mode \* gender + result, data = Accident)  
##   
## Statistics:  
## X^2 df P(> X^2)  
## Likelihood Ratio 2217.72 39 0  
## Pearson 2347.60 39 0

1. Construct a mosaic display for the residual associations in this model. Which combinations of the predictor factors are more likely to result in death?

mosaic(loglm(Freq ~ age \* mode \* gender + result, data = Accident), gp = shading\_Friendly, labeling = labeling\_residuals)



# Exercise #2

The data set caith in MASS (Ripley, 2015) gives another classic 4×5 table tabulating hair color and eye color, this for people in Caithness, Scotland, originally from Fisher (1940). The data is stored as a data frame of cell frequencies, whose rows are eye colors and whose columns are hair colors.

data("caith", package="MASS")  
caith

## fair red medium dark black  
## blue 326 38 241 110 3  
## light 688 116 584 188 4  
## medium 343 84 909 412 26  
## dark 98 48 403 681 85

1. The loglm() and mosaic() functions don’t understand data in this format, so use Caith <- as.matrix(caith) to convert to array form. Examine the result, and use names(dimnames(Caith))<-c() to assign appropriate names to the row and column dimensions.

Caith <- as.matrix(caith)  
names(dimnames(Caith))<-c("Eye\_Color","Hair\_Color")   
Caith

## Hair\_Color  
## Eye\_Color fair red medium dark black  
## blue 326 38 241 110 3  
## light 688 116 584 188 4  
## medium 343 84 909 412 26  
## dark 98 48 403 681 85

1. Fit the model of independence to the resulting matrix using loglm().

model=loglm(Freq ~ Eye\_Color + Hair\_Color, data = Caith)  
anova(model)

## Call:  
## loglm(formula = Freq ~ Eye\_Color + Hair\_Color, data = Caith)  
##   
## Statistics:  
## X^2 df P(> X^2)  
## Likelihood Ratio 1218.314 12 0  
## Pearson 1240.039 12 0

1. Calculate and display the residuals for this model.

LRstats(model)

## Likelihood summary table:  
## AIC BIC LR Chisq Df Pr(>Chisq)   
## model 1368 1375.9 1218.3 12 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

1. Create a mosaic display for this data.

mosaic(Caith, shade=TRUE, gp = shading\_Friendly, labeling\_args = list(rot\_labels=c(60, 60, 60, 60)))

