DDAR: Solutions and Hints for Exercises

2015-11-19

## Chapter 1

These exercises are all conceptual. There are no hints or solutions.

## Chapter 2

### Exercise 2.2

### Exercise 2.3

The data set DanishWelfare in vcd gives a 4-way, 3 x 4 x 3 x 5 table as a data frame in frequency form, containing the variable Freq and four factors, Alcohol, Income, Status and Urban. The variable Alcohol can be considered as the response variable, and the others as possible predictors.

1. Find the total number of cases represented in this table.

This is a data set in the form of a frequency data.frame, so sum the Freq variable

data("DanishWelfare", package="vcd")  
sum(DanishWelfare$Freq)

## [1] 5144

1. In this form, the variables Alcohol and Income should arguably be considered ordered factors. Change them to make them ordered.

Use ordered() or as.ordered() on the factor variable. str() will then show them as Ord.factor.

levels(DanishWelfare$Alcohol)

## [1] "<1" "1-2" ">2"

DanishWelfare$Alcohol <- as.ordered(DanishWelfare$Alcohol)  
DanishWelfare$Income <- as.ordered(DanishWelfare$Income)  
str(DanishWelfare)

## 'data.frame': 180 obs. of 5 variables:  
## $ Freq : num 1 4 1 8 6 14 8 41 100 175 ...  
## $ Alcohol: Ord.factor w/ 3 levels "<1"<"1-2"<">2": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Income : Ord.factor w/ 4 levels "0-50"<"50-100"<..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Status : Factor w/ 3 levels "Widow","Married",..: 1 1 1 1 1 2 2 2 2 2 ...  
## $ Urban : Factor w/ 5 levels "Copenhagen","SubCopenhagen",..: 1 2 3 4 5 1 2 3 4 5 ...

1. Convert this data frame to table form, DanishWelfare.tab, a 4-way array containing the frequencies with appropriate variable names and level names.

Use xtabs() with Freq as the response.

DanishWelfare.tab <-xtabs(Freq ~ ., data = DanishWelfare)  
str(DanishWelfare.tab)

## xtabs [1:3, 1:4, 1:3, 1:5] 1 3 2 8 1 3 2 5 2 42 ...  
## - attr(\*, "dimnames")=List of 4  
## ..$ Alcohol: chr [1:3] "<1" "1-2" ">2"  
## ..$ Income : chr [1:4] "0-50" "50-100" "100-150" ">150"  
## ..$ Status : chr [1:3] "Widow" "Married" "Unmarried"  
## ..$ Urban : chr [1:5] "Copenhagen" "SubCopenhagen" "LargeCity" "City" ...  
## - attr(\*, "class")= chr [1:2] "xtabs" "table"  
## - attr(\*, "call")= language xtabs(formula = Freq ~ ., data = DanishWelfare)

1. The variable Urban has 5 categories. Find the total frequencies in each of these. How would you collapse the table to have only two categories, City, Non-city?

margin.table() handles the first part; collapse.table() is designed for the second part. It is arguable whether SubCopenhagen should be considered City or NonCity.

margin.table(DanishWelfare.tab, 4)

## Urban  
## Copenhagen SubCopenhagen LargeCity City Country   
## 552 614 594 1765 1619

DW2 <- vcdExtra::collapse.table(DanishWelfare.tab, Urban=c("City","NonCity","City","City","NonCity"))  
head(ftable(DW2))

##   
## "Urban" "City" "NonCity"  
## "Alcohol" "Income" "Status"   
## "<1" "0-50" "Widow" 10 10  
## "Married" 155 183  
## "Unmarried" 14 10  
## "50-100" "Widow" 29 7  
## "Married" 338 306  
## "Unmarried" 36 32