R_Assignment2

Nilesh Harekal Nayak 21200475

14/11/2021

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com (http://rmarkdown.rstudio.com).

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
#install.packages("tidyverse")
#install.packages("dplyr")
#install.packages("ggplot2")
options(warn=-1)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
options(warn=0)
library(dplyr)
library(ggplot2)
```

Task 1

#Question 1

```
#loading 1995 survey
data_survey_1995 <- read.table("C:/Users/dell/Desktop/Ireland Study/R Prog/week 8/Assignment 2/s50_1995.txt", sep=" ",
header=TRUE)
head(data_survey_1995)
```

```
##   alcohol drugs smoke sport
## 1       3     1     2     2
## 2       2     2     3     1
## 3       2     1     1     1
## 4       2     1     1     2
## 5       3     1     1     2
## 6       4     1     1     2
```

```
class(data_survey_1995)
```

```
## [1] "data.frame"
```

```
#converting each variable to factors
data_survey_1995$alcohol = factor(data_survey_1995$alcohol, levels=c("1","2","3","4","5"), labels = c("not", "once or
 twice a year", "once a month", "once a week", "more than once a week"), ordered=TRUE)
data_survey_1995$drugs = factor(x = data_survey_1995$drugs , levels=c("1","2","3","4"), labels = c("not","tried once",
"occasional", "regular"), ordered=TRUE)
data_survey_1995$smoke = factor(x = data_survey_1995$smoke , levels=c("1","2","3"), labels = c("not","ocassional", "re
gular") , ordered = TRUE)
data_survey_1995$sport = factor(x = data_survey_1995$sport , levels=c("1","2"), labels = c("not regular","regular") ,
 ordered = TRUE)

head(data_survey_1995,3)
```

```
##               alcohol      drugs      smoke       sport
## 1         once a month        not ocassional     regular
## 2 once or twice a year tried once    regular not regular
## 3 once or twice a year        not        not not regular
```

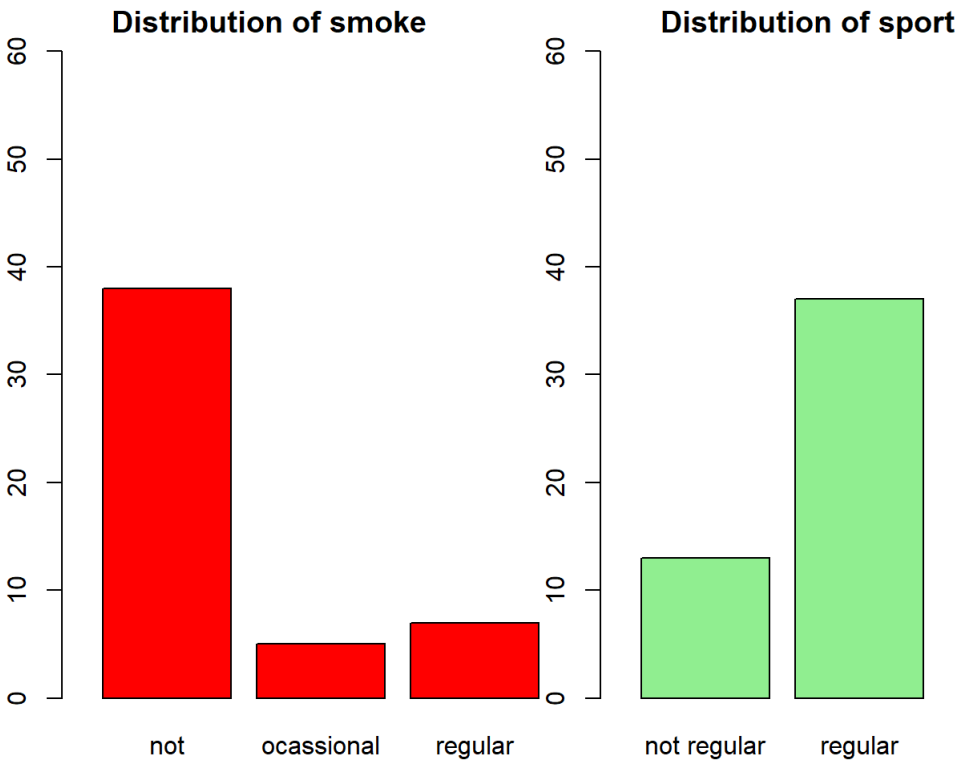
```
str(data_survey_1995)
```

```
## 'data.frame':    50 obs. of  4 variables:
##  $ alcohol: Ord.factor w/ 5 levels "not"<"once or twice a year"<..: 3 2 2 2 3 4 4 4 2 4 ...
##  $ drugs  : Ord.factor w/ 4 levels "not"<"tried once"<..: 1 2 1 1 1 1 3 3 1 1 ...
##  $ smoke  : Ord.factor w/ 3 levels "not"<"ocassional"<..: 2 3 1 1 1 1 1 3 1 1 ...
##  $ sport  : Ord.factor w/ 2 levels "not regular"<..: 2 1 1 2 2 2 1 2 2 2 ...
```

The structure of the dataset is a data frame with 50 observations of 4 variables

#Question 2

```
par(mfrow=c(1, 2), mar = c(2, 2, 2, 2))
barplot(table(data_survey_1995$smoke), xlab = "smoke" , ylab = "count", main = "Distribution of smoke", col = "red", x
lim = c(0, 3),ylim = c(0, 60))
barplot(table(data_survey_1995$sport), xlab = "sport" , ylab = "count", main = "Distribution of sport", col = "lightgr
een", xlim = c(0, 3),ylim = c(0, 60))
```



In a cohort of 50 pupils aged 13 in 1995 in a school present in West of Scotland, more than 35 students do not smoke

Also, more than 35 involve themselves in sport regularly

However we cannot deduce from the above graph that the ones who do not smoke are the same ones who go for sports regularly. For this we need to see that each pupil level

#Question 3

```
#proportion of pupils who smoke at least occasionally?
#this implies who smoke at least ocassionally + regular
newdata_1 <- subset(data_survey_1995, data_survey_1995$smoke == 'ocassional' | data_survey_1995$smoke == 'regular')
length(newdata_1$smoke) / length(data_survey_1995$smoke)
```

```
## [1] 0.24
```

```
#proportion of pupils who regularly practiced sport and smoke at least occasionally
#this implies who go for sport regularly and (smoke at least ocassionally + regular)
newdata_2 <- subset(data_survey_1995, data_survey_1995$sport == 'regular' & (data_survey_1995$smoke == 'ocassional' |
 data_survey_1995$smoke == 'regular'))
length(newdata_2$smoke) / length(data_survey_1995$smoke)
```

```
## [1] 0.18
```

24% of the pupils smoke at least occasionally

18% pupils regularly practice sport and smoke at least occasionally

#Question 4

```
class(data_survey_1995) <- "s50survey"
str(data_survey_1995)
```

```
## List of 4
##  $ alcohol: Ord.factor w/ 5 levels "not"<"once or twice a year"<..: 3 2 2 2 3 4 4 4 2 4 ...
##  $ drugs  : Ord.factor w/ 4 levels "not"<"tried once"<..: 1 2 1 1 1 1 3 3 1 1 ...
##  $ smoke  : Ord.factor w/ 3 levels "not"<"ocassional"<..: 2 3 1 1 1 1 1 3 1 1 ...
##  $ sport  : Ord.factor w/ 2 levels "not regular"<..: 2 1 1 2 2 2 1 2 2 2 ...
##  - attr(*, "row.names")= int [1:50] 1 2 3 4 5 6 7 8 9 10 ...
##  - attr(*, "class")= chr "s50survey"
```

```
summary.s50survey<-function(data)
{
cat ("The proportions for alcohol is : \n" )
print (prop.table(table(data$alcohol)))
cat ("\n" )
cat ("The proportions for drugs is : \n" )
print (prop.table(table(data$drugs)))
cat ("\n" )
cat ("The proportions for smoke is : \n" )
print (prop.table(table(data$smoke)))
cat ("\n" )
cat ("The proportions for sport is : \n" )
print (prop.table(table(data$sport)))
cat ("\n" )
}

summary(data_survey_1995)
```

```
## The proportions for alcohol is :
##
##                   not  once or twice a year          once a month
##                  0.10                  0.32                  0.24
##           once a week more than once a week
##                  0.28                  0.06
##
## The proportions for drugs is :
##
##        not tried once occasional    regular
##       0.72       0.12       0.14       0.02
##
## The proportions for smoke is :
##
##        not ocassional    regular
##       0.76       0.10       0.14
##
## The proportions for sport is :
##
## not regular     regular
##        0.26        0.74
```

#Question 5

```
prop.table(table(data_survey_1995$drugs))[1]
```

```
##  not
## 0.72
```

Based on output of question #4 and #5, 72% of the pupils did not use cannabis

#Question 6

```
#reading 1997 dataset + converting each variable to ordered factors
data_survey_1997 <- read.table("C:/Users/dell/Desktop/Ireland Study/R Prog/week 8/Assignment 2/s50_1997.txt", sep=" ",
header=TRUE)
data_survey_1997$alcohol = factor(data_survey_1997$alcohol, levels=c("1","2","3","4","5"), labels = c("not", "once or
 twice a year", "once a month", "once a week", "more than once a week"), ordered=TRUE)
data_survey_1997$drugs = factor(x = data_survey_1997$drugs , levels=c("1","2","3","4"), labels = c("not","tried once",
"occasional", "regular"), ordered=TRUE)
data_survey_1997$smoke = factor(x = data_survey_1997$smoke , levels=c("1","2","3"), labels = c("not","ocassional", "re
gular") , ordered = TRUE)
data_survey_1997$sport = factor(x = data_survey_1997$sport , levels=c("1","2"), labels = c("not regular","regular") ,
 ordered = TRUE)


#displaying 1997 dataset after ordered factors applied to each variable.
head(data_survey_1997,3)
```

```
##                alcohol      drugs   smoke       sport
## 1         once a month        not     not not regular
## 2 once or twice a year occasional regular not regular
## 3         once a month        not     not not regular
```

```
#assigning class to 1997 dataset
class(data_survey_1997) <- "s50survey"
str(data_survey_1997)
```

```
## List of 4
##  $ alcohol: Ord.factor w/ 5 levels "not"<"once or twice a year"<..: 3 2 3 2 4 4 3 4 2 4 ...
##  $ drugs  : Ord.factor w/ 4 levels "not"<"tried once"<..: 1 3 1 1 3 1 2 3 1 1 ...
##  $ smoke  : Ord.factor w/ 3 levels "not"<"ocassional"<..: 1 3 1 1 1 3 3 3 1 2 ...
##  $ sport  : Ord.factor w/ 2 levels "not regular"<..: 1 1 1 1 2 2 2 2 1 2 ...
##  - attr(*, "row.names")= int [1:50] 1 2 3 4 5 6 7 8 9 10 ...
##  - attr(*, "class")= chr "s50survey"
```

```
#Test the summary S3 method on 1997 dataset
summary(data_survey_1997)
```

```
## The proportions for alcohol is :
##
##                   not  once or twice a year          once a month
##                  0.02                  0.18                  0.34
##           once a week more than once a week
##                  0.34                  0.12
##
## The proportions for drugs is :
##
##        not tried once occasional    regular
##       0.52       0.14       0.34       0.00
##
## The proportions for smoke is :
##
##        not ocassional    regular
##       0.62       0.04       0.34
##
## The proportions for sport is :
##
## not regular     regular
##        0.62        0.38
```

#Question 7

Based on interpretation between Students practising sport regularly decreased from 74% in the year 1995 to 38% in the year 1997