MATH 3070 Lab Project 4

Aini Liang

September 8, 2016

Remember: I expect to see commentary either in the text, in the code with comments created using `#`*, or (preferably) both!* ***Failing to do so may result in lost points!***

Problem 1 (Verzani problem 2.43)

The `time` *variable in the* `nym.2002` *data set (**UsingR**) contains the time to finish the 2002 New York City Marathon for a random sample of the finishers.*

1. *What percent ran the race in under 3 hours?*

```
# Your code here

#install.packages("UsingR")
library(UsingR)
```

```
## Loading required package: MASS
```

```
## Loading required package: HistData
```

```
## Loading required package: Hmisc
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
## Loading required package: ggplot2
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units
```

```
##
## Attaching package: 'UsingR'
```

```
## The following object is masked from 'package:survival':
##
##     cancer
```

```
#View(nym.2002)
attach(nym.2002)
```

```
## The following object is masked from package:UsingR:
##
##     home
```

```
#sum(time <3*60)/length(time)
mean(time <60*3)
```

```
## [1] 0.026
```

```
detach (nym.2002)
```

2. *Given that the slower runners will have higher running times; cut off the highest 10% and the highest 25%.*

```
# Your code here
# use quantile (time,prob)
#datanme#$colum name
attach(nym.2002)
```

```
## The following object is masked from package:UsingR:
##
##     home
```

```
quantile(nym.2002$time,c(1-0.1,1-0.25))
```

```
##      90%      75%
## 331.7500 294.8542
```

3. *Given that the faster runners will have lower running times, cut off the lowest 10%.*

```
# Your code here
# use quantile (time,prob)
attach(nym.2002)
```

```
## The following objects are masked from nym.2002 (pos = 3):
##
##     age, gender, home, place, time
```

```
## The following object is masked from package:UsingR:
##
##     home
```

```
quantile(nym.2002$time,0.1)
```

```
##     10%
## 208.695
```

Problem 2 (Verzani problem 4.1)

*The data set `UScereal` (**MASS**) contains data on cereals sold in the United States in 1993. For this data set, answer the following questions using R (i.e. MUST answer questions using a code. Do not count by hand):*

1. *How many rows does the data frame have? Columns?*

```
# Your code here
#install.packages("MASS")
library(MASS)

#View(UScereal)
length(UScereal)
```

```
## [1] 11
```

```
sum(row(UScereal))
```

```
## [1] 23595
```

```
nrow(UScereal)
```

```
## [1] 65
```

```
#dim(UScereal)
# row:65  colums:11
```

2. *How many different manufacturers are included?*

```
# Your code here
#install.packages("MASS")
#library(MASS)
#View(UScereal)
#length(levels(mfr))
length(unique(UScereal$mfr))
```

```
## [1] 6
```

```
#length(levels())=length(unique())
```

3. *How many vitamin categories are included?*

```
# Your code here
#install.packages("MASS")
library(MASS)
#View(UScereal)
#length(unique(vitamins))
length(unique(UScereal$vitamins))
```

```
## [1] 3
```

4. *How many cereals have a sugar level above 10?*

```
# Your code here
#install.packages("MASS")
#library(MASS)
#attach(UScereal)
#View(UScereal)
sum (UScereal$sugars >10)
```

```
## [1] 39
```

5. *What is the mean calorie value for cereals with more than 5 grams of fat? Less than or equal to 5?*

```
# Your code here
#install.packages("MASS")
library(MASS)
attach(UScereal)
```

```
## The following object is masked from package:UsingR:
##
##     fat
```

```
#View(UScereal)
UScereal$fat>5
```

```
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [12] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE
## [23] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE
## [34] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [45] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [56] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
UScereal$fat<5
```

```
##  [1]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [12]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
## [23]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE
## [34]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [45]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## [56]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
```

```
x <- c (UScereal$fat>5)
mean(calories,na.rm = FALSE)
```

```
## [1] 149.4083
```

6. *What is the mean calorie value for cereals on the middle shelf (2)?*

```
# Your code here

(mean(shelf==2))
```

```
## [1] 0.2769231
```

Problem 3

Create a data frame containing the data in the following table:

First	Last	Age
Marcus	Holstein	23
Samuel	Adams	56
Gus	McPherson	43
Margaret	Olsen	41
Zim	Newbold	95

```
# Your code here

df1 <- data.frame(First = c("Marcus", "Samuel", "Gus", "Margaret", "Zim"), Last = c("Holstein", "Adams", "M
cPherson", "Olsen", "Newbold"), Age=c(23,56,43,41,95))
df1
```

```
##      First      Last Age
## 1   Marcus  Holstein  23
## 2   Samuel     Adams  56
## 3      Gus McPherson  43
## 4 Margaret     Olsen  41
## 5      Zim   Newbold  95
```