MATH 3070 Lab Project 4

Prachi Aswani

September 19, 2024

Contents

4 *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? (note that the data is in minutes)*

```
library(UsingR)
```

```
## Warning: package 'UsingR' was built under R version 4.3.3
```

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

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

```
## Warning: package 'HistData' was built under R version 4.3.3
```

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

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

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

```r
data(nym.2002)

# Extracting time column
time_in_minutes <- nym.2002$time

# Calculate the percentage
under_3_hours <- sum(time_in_minutes < 180)
total_runners <- length(time_in_minutes)
percentage_under_3_hours <- (under_3_hours / total_runners) * 100

percentage_under_3_hours
```

```
## [1] 2.6
```

2. *What is the time cutoff for the top 10%? The top 25%?*

```r
# Notice that these are the 10th and 25th percentiles, respectively
# 10th and 25th percentiles
cutoff_top_10 <- quantile(time_in_minutes, 0.10)
cutoff_top_25 <- quantile(time_in_minutes, 0.25)


cutoff_top_10
```

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

```r
cutoff_top_25
```

```
##     25%
## 233.775
```

3. *What time cuts off the bottom 10%?*

```r
# Bottom 10% percentile
cutoff_bottom_10 <- quantile(time_in_minutes, 0.10)

cutoff_bottom_10
```

```
##     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?*

```r
library(MASS)
data(Uscereal)
```

```
## Warning in data(Uscereal): data set 'Uscereal' not found
```

```r
# Number of rows and columns
num_rows <- nrow(UScereal)
num_columns <- ncol(UScereal)

num_rows
```

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

```r
num_columns
```

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

2. *How many different manufacturers are included?*

```r
# Number of unique manufacturers
num_manufacturers <- length(unique(UScereal$mfr))

num_manufacturers
```

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

3. *How many vitamin categories are included?*

```r
# Number of unique vitamin categories
num_vitamin_categories <- length(unique(UScereal$vitamins))

num_vitamin_categories
```

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

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

```r
# Cereals with sugar level above 10
num_high_sugar <- sum(UScereal$sugars > 10)

num_high_sugar
```

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

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

```r
# Mean calorie values
mean_calories_high_fat <- mean(UScereal$calories[UScereal$fat > 5])
mean_calories_low_fat <- mean(UScereal$calories[UScereal$fat <= 5])

mean_calories_high_fat
```

```
## [1] 291.8182
```

```r
mean_calories_low_fat
```

```
## [1] 144.8873
```

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

```r
# Mean calorie value for cereals on shelf 2
mean_calories_shelf_2 <- mean(UScereal$calories[UScereal$shelf == 2])

mean_calories_shelf_2
```

```
## [1] 129.8162
```

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

```r
# Creating the data frame
data_frame <- data.frame(
  First = c("Marcus", "Samuel", "Gus", "Margaret", "Zim"),
  Last = c("Holstein", "Adams", "McPherson", "Olsen", "Newbold"),
  Age = c(23, 56, 43, 41, 95)
)

data_frame
```

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