S2S Lab 3

Further Exercises Solutions

Exercise 1

The file "*NHSScotland.txt*" contains data on the number of patients attending A&E every month in each of the 14 Scottish NHS boards, from 2007 up to 2023. This data set has the following variables:

- "Date": the end of the month that patient numbers are aggregated over.
- "NHSBoard": the Scottish NHS board the patients are from.
- "TotalAttendances": the total number of patients attending A&E in a given month and NHS board.
- "Within4Hours": the number of patients whose wait time was less than 4 hours.
- "Over4Hours": the number of patients whose wait time was greater than 4 hours.
- "Over8Hours": the number of patients whose wait time was greater than 8 hours.
- "Over12Hours": the number of patients whose wait time was greater than 12 hours.

a. i. Read "*NHSScotland.txt*" into R and save it as a data frame called `nhs`.

ii. Change the column "NHSBoard" to be a factor. (**Hint**: you can see the names of all the Scottish NHS boards using the code `unique(nhs$NHSBoard)`.)

iii. Add an additional column to `nhs` which calculates the percentage of total patients in A&E whose wait time is less than 4 hours. Call this new variable "PercentageWithin4Hours".

iv. What is the average percentage of patients who had to wait less than 4 hours in each of the 14 Scottish NHS boards? (**Hint**: think how you can use the `tapply()` function.)

b. i. Create a new data frame, called `glasgow`, which is a subset of `nhs`. This data set should only show observations from NHS Greater Glasgow & Clyde, as well as only having the variables "Date", "TotalAttendances" and "Over4Hours".

ii. Sort `glasgow` in order of decreasing number of patients who had to wait more than 4 hours in A&E. When did the greatest number of patients have to wait for longer than 4 hours?

c. The file "*HBPopulation.csv*" contains data relating to the population size (in 2021) of each of the 14 Scottish NHS boards. Read this file into R and save it as a data frame called `population`. Merge `nhs` and `population` so that A&E attendance and the health board population size can be seen in the same data frame.

Solution

a.

```
nhs <- read.table(file = "NHSScotland.txt", header = TRUE, na.strings = "*")

nhs$NHSBoard <- factor(x = nhs$NHSBoard, levels = unique(nhs$NHSBoard))

nhs$PercentageWithin4Hours <- (nhs$Within4Hours/nhs$TotalAttendances)*100

tapply(X = nhs$PercentageWithin4Hours, INDEX = list(nhs$NHSBoard), FUN = mean)
```

```
##        NHS Ayrshire & Arran                NHS Borders
```

```
##                     90.46348                     91.55099
##     NHS Dumfries & Galloway                     NHS Fife
##                     92.98376                     92.47211
##            NHS Forth Valley                NHS Grampian
##                     89.60296                     92.68313
## NHS Greater Glasgow & Clyde                NHS Highland
##                     90.20233                     95.81951
##             NHS Lanarkshire                 NHS Lothian
##                     88.26992                     88.97148
##                  NHS Orkney                NHS Shetland
##                     97.36735                     97.36067
##                  NHS Tayside           NHS Western Isles
##                     97.56968                     98.42613
```

b.

```
glasgow <- subset(x = nhs, subset = (NHSBoard == "NHS Greater Glasgow & Clyde"),
                  select = c("Date", "TotalAttendances", "Over4Hours"))

glasgow <- glasgow[order(glasgow$Over4Hours, decreasing = TRUE), ]
glasgow[1, "Date"]
```

```
## [1] "2022-12-31"
```

c.

```
population <- read.csv(file = "HBPopulation.csv", header = TRUE)

nhs_full <- merge(x = nhs, y = population, by = "NHSBoard", all = TRUE)
```

Exercise 2

Draw a random sample of 100 values from the Poisson(3) distribution and save these in a vector `x`.

If any of these random values are less than 3, use a for loop containing an if statement, to change these value to be equal to 3.

Solution

```
x <- rpois(n = 100, lambda = 3)

for(i in 1:100){
  if(x[i] < 3){
    x[i] <- 3
  }
}
```

Exercise 3

The volume of a cylinder is calculated as $V_{\text{cylinder}} = \pi r^2 h$ where r is the radius of the cylinder and h is the height.

a. Write a function called `cyl.vol` which takes the arguments `r` and `h` and returns the volume of cylinder with radius `r` and height `h`.

   Use your function to find the volume of a cylinder with which has radius 2.8cm and height 24cm.

b. Use your function `cyl.vol` to write another function which can be used to calculate the total volume of n cylinders of the same size. This second function should take the arguments `r` for the radius, `h` for the height and `n` for the number of cylinders.

Solution

a.

```
cyl.vol <- function(r, h){
  pi*r^2*h
}

cyl.vol(r = 2.8, h = 24)
```

```
## [1] 591.1221
```

b.

```
n.cyl.vol <- function(r, h, n){
  n*cyl.vol(r, h)
}
```