## Question: Introduction to R

This homework sheet will test your knowledge of basics commands in R. 18

3

a) Use R to determine which of the following equations are correct.

- $2 \cdot 4 \cdot 8 \neq 4^3$
- $52 \cdot 81 \cdot 22 \leq 45^3$
- $5 \cdot \log_{10} \sqrt{57^5} > \log_{10} 1000^3$

Solution:

```
2 * 4 * 8 != 4^3
```

```
## [1] FALSE
```

```
52 * 81 * 22 <= 45^3
```

```
## [1] FALSE
```

```
5 * log10(sqrt(57^5)) > log10(1000^3)
```

```
## [1] TRUE
```

3

b) Use the variables `price`, `net_price` and `vat` to construct a formula to calculate the price given a net price of $57$ and a VAT of $0.19$.

Solution:

```
net_price <- 57
vat <- 1 + 0.19
price <- net_price * vat
```

3

c) Create a vector `odd` that contains all odd numbers between 1 and 100 and a vector `even` that contains all even numbers between 1 and 100.

Which of those two vectors has the higher mean and which one has the higher variance?

Solution:

```
odd <- seq(from=1, to=99, by=2)
even <- seq(from=2, to=100, by=2)

# mean of even is higher
mean(odd)

## [1] 50

mean(even)

## [1] 51

#
var(odd)

## [1] 850

var(even)

## [1] 850
```

1

d) Use your vectors `odd` and `even` from the previous task to generate a matrix `mat` in which the first column consists of the odd numbers and the second column of the even numbers. Which number is at the matrix entry in row 17 of column 1? How can one select the first even number?

Solution:

```
mat <- as.data.frame(cbind(odd, even))
mat[17, 1] #

## [1] 33

mat[1, 2]

## [1] 2
```

3

e) Store the following snippet as a CSV file and then load the file into R. This file is missing its header names. The data is in the following order: name, height, shoe size and age. Calculate the mean for height and shoe size.

Note: to return a column of a matrix as a vector you can use the following notation: `matrix[,col_index]`, e. g. persons[,1] for the first column.

**File snippet** `persons.csv`

```
Mathew,153,44,26
Shaun,186,44,26
Vader,172,45,21
Bruce,184,43,21
Katy,142,41,26
Sandy,186,40,25
Sophie,184,46,23
```

Solution:

```
persons <- read.csv("data/persons.csv", header=FALSE, sep=",")
persons
```

```
##       V1  V2 V3 V4
## 1 Mathew 153 44 26
## 2  Shaun 186 44 26
## 3  Vader 172 45 21
## 4  Bruce 184 43 21
## 5   Katy 142 41 26
## 6  Sandy 186 40 25
## 7 Sophie 184 46 23
```

```
# mean_height
mean(persons[,2])
```

```
## [1] 172.4286
```

```
# mean_shoesize
mean(persons[,3])
```

```
## [1] 43.28571
```

2

f) In order to simplify data operations, one usually assigns column names to data frames. Utilize data from the previous task and then give the matrix the corresponding column names (name, height, shoe size, age). Then repeat the calculation of the mean of height and shoe size with the new columns.

Solution:

```
colnames(persons) <- c("name", "height", "shoesize", "age")

mean(persons$height)

## [1] 172.4286

mean(persons$shoesize)

## [1] 43.28571
```

2

g) Add a column `gender` to the persons matrix containing the correct data. Afterwards, calculate the standard deviation for the height of the women.

Solution:

```
persons[["gender"]] <- c("m", "m", "m", "m", "f", "f", "f")
sd(persons[persons$gender == "f", "height"])

## [1] 24.84619
```

1

h) What are the values of `x`, `y`, and `z` after executing the following code?

```
x <- 3
x <- x * 2
y <- 1
z <- x - y
x <- x * x
z <- z + x
```

Solution:

```
x
```

```
## [1] 36
```

```
y
```

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

```
z
```

```
## [1] 41
```