Homework 2

Stefan Konigorski

October 26, 2022

Download this R Markdown file, save it on your computer, and perform all the below tasks by inserting your answer in text or by inserting R chunks below. After you are done, upload this file with your solutions on Moodle.

Exercise 1

a) Create an R chunk here to insert R code. Add R code in this R chunk to perform a simple calculation (e.g. calculate the sum of 1 and 2).

```
a = 1
b=2
c= a+b
c
```

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

b) Create an R chunk with a basic calculation (e.g. 1+1). Try out the different ways how to include this in the knitted report. Suppose,
x = 9 and
y = 99

1. The result of division of x by y is 11.

2. x + y equals 108.

3. x * y yields 891.

c) Knit this Rmd file to html and to pdf.

The html and pdf is knitted and included with the assignment.

Exercise 2: Manipulating variables and data frames

Load the Pima Indian dataset:

```r
dat_ex2 <- read.csv(file = url("https://www.dropbox.com/s/tqrauwuxyi03kee/Pima_diabetes.csv?dl=1"))
```

and answer the following questions:

```r
# How many women have Glucose levels 0?

sum(dat_ex2$Glucose == 0)
```

```
## [1] 5
```

```r
# How many women have Insulin levels 0?

sum(dat_ex2$Insulin == 0)
```

```
## [1] 374
```

```r
# How many women have both Glucose levels as well as Insulin levels 0?

sum(dat_ex2$Glucose == 0 & dat_ex2$Insulin ==0)
```

```
## [1] 4
```

```r
# How many women have either Glucose levels or Insulin levels 0?

sum(dat_ex2$Glucose == 0 | dat_ex2$Insulin ==0)
```

```
## [1] 375
```

```r
# How many women have missing BMI values?

sum(is.na(dat_ex2$BMI))
```

```
## [1] 0
```

```r
# How many women have BMI larger than 40?

sum(dat_ex2$BMI > 40)
```

```
## [1] 96
```

```r
# Build a dataset that only includes the women with BMI>40

dat_sub = dat_ex2[dat_ex2$BMI>40,]
head(dat_sub)
```

```
##     Pregnancies Glucose BloodPressure SkinThickness Insulin  BMI
## 5             0     137            40            35     168 43.1
## 17            0     118            84            47     230 45.8
## 19            1     103            30            38      83 43.3
```

```
## 42              7      133              84               0        0 40.2
## 44              9      171             110              24      240 45.4
## 46              0      180              66              39        0 42.0
##     DiabetesPedigreeFunction Age Outcome
## 5                      2.288  33       1
## 17                     0.551  31       1
## 19                     0.183  33       0
## 42                     0.696  37       0
## 44                     0.721  54       1
## 46                     1.893  25       1
```

```r
# Create a new variable named BMIOutlier, which has the value 0 if a women has BMI smaller or equal 50,

dat_sub["BMIOutlier"] = ifelse(dat_sub$BMI<=50,0,1)
head(dat_sub)
```

```
##     Pregnancies Glucose BloodPressure SkinThickness Insulin  BMI
## 5             0     137            40            35     168 43.1
## 17            0     118            84            47     230 45.8
## 19            1     103            30            38      83 43.3
## 42            7     133            84             0       0 40.2
## 44            9     171           110            24     240 45.4
## 46            0     180            66            39       0 42.0
##     DiabetesPedigreeFunction Age Outcome BMIOutlier
## 5                      2.288  33       1          0
## 17                     0.551  31       1          0
## 19                     0.183  33       0          0
## 42                     0.696  37       0          0
## 44                     0.721  54       1          0
## 46                     1.893  25       1          0
```

Exercise 3 (optional)

Explore merging two datasets.

As a preparation, execute the following code to create different data frames

```r
# import data
dat_ex3 <- read.csv(file = url("https://www.dropbox.com/s/tqrauwuxyi03kee/Pima_diabetes.csv?dl=1"))

# extract two smaller data sets
dat3_1 <- dat_ex3[1:100, 1:3]
dat3_2 <- dat_ex3[101:300, 1:3]

dat3_3 <- dat_ex3[1:100, 1:3]
dat3_4 <- dat_ex3[1:100, 4:6]
```

Task 3a: Think about how you can use the [.] operator to respectively piece dat3_1 and dat3_2, and dat3_3 and dat3_4 together into one data frame.

Task 3b: Explore the help of the merge() function in R in order to achieve the same goal of combining dat3_3 and dat3_4 together into one data frame. Hint: first create an ID variable in each data frame, then use this in the "by" argument.

Exercise 4 (optional): Times and dates in R

Create an Excel file with 5 observations of 2 variables. Variable 1 is just an ID variable (number 1-5 or character string etc.), and variable 2 is a date/time variable. Use variable 2 to describe the time (and day) you had lunch in the last 5 days. Then try to import the Excel file with both variables into R and/or transform the variables in R to Date or POSIXct variables.