Exercises for Lecture 8

Statistical Computing with R, 2023-24

Exercise 1

In lecture 3 (exercise 2) you created 3 functions that respectively centered, scaled and normalized an input vector v.

1. Retrieve the functions that you created during lecture 3.
2. Modify these functions in such a way that:

- if the input is not a numeric vector, an error message saying "function input should be a numeric vector" is generated;
- if the input is a numeric vector which contains one or more `NA`s, a warning message saying "the supplied vector contains missing values" is generated.

3. Apply your functions to the following objects:

```
obj1 = matrix(1:10, 5, 2)
obj2 = c(5, 7, 10, -25)
obj3 = c(42, NA, 3, 7)
obj4 = c(pi, 42, 'apple', sqrt(3))
```

Do the functions behave as requested in (2)?

Exercise 2

In lecture 3 (exercise 4) you created your own function to compute the skewness of a given input vector x.

1. Retrieve the function that you created during lecture 3.
2. Modify the function in such a way that:

- if the input is a numeric vector containing `NA`s, an information message saying "the supplied vector contains missing values" is generated;
- if the input is a matrix, the function computes the skewness of each column separately, returning a vector of skewness values. Moreover, the function should return a warning that says: "supplied input is a matrix. Skewness by column returned";
- if the input is neither a numeric vector nor a matrix, an error message saying "wrong type of input" is generated.

Exercise 3

Consider again the `heights` and `wages` datasets from the `brolgar` package:

```
library(brolgar)
data(heights)
heights = as.data.frame(heights)
data(wages)
wages = as.data.frame(wages)
```

1. Select the data from 1980 in the `heights` data frame.

- Draw a boxplot that compares the distribution of heights in different continents. Create both a vertical and a horizontal boxplot.
- Create a violin plot that displays the same type of information.

2. Select observations from workers with at most 2 years of work experience in the `wages` data frame. Draw a violin plot that compares the distribution of the log-wage for the different values of `high_grade`.

Exercise 4

The `R` package `mvtnorm`, published on `CRAN`, allows to simulate random numbers from a multivariate normal distribution. Execute the following code:

```
library(mvtnorm)
set.seed(8)
Sigma1 = matrix(c(1, 0.5, -0.3,
                  0.5, 1, -0.6,
                  -0.3, -0.6, 1), 3, 3)
m1 = rmvnorm(500, mean = rep(0, 3), sigma = Sigma1)
Sigma2 = matrix(c(1, 0.4,
                  0.4, 1), 2, 2)
m2 = rmvnorm(500, mean = rep(0, 2), sigma = Sigma2)
all.vars = cbind(m1, m2)
colnames(all.vars) = paste('X', 1:5, sep = '')
```

1. Can you guess what the code above is doing? If so, explain it!
2. Estimate the correlations between X_1, X_2, X_3, X_4 and X_5.
3. Can you relate the values of the correlations computed at point (2) to the code above? Is there a relationship between the two?
4. Draw a correlogram to display the correlations that exist between these 5 variables.
5. Change the colouring scheme used in (4) so that positive correlations are displayed in `darkgreen`, and negative correlations in `red`.