# Assignment11

## FedericaFusi

### 3/29/2021

## ASSIGNMENT 11

### General instructions

This assignment will propose you a set of exercises to get used to write if_else/ifelse statements and use loops and apply functions.

Please submit your output as an **rmarkdown PDF** showing both your code and the results.

Submit both the code and the PDF to Blackboard.

### Exercises

1. Write a loop that iterates over the numbers 1 to 10 and adds 5 to each of them before storing the results in a new vector called output.

2. Create the matrix **mat_x** below and perform the following operations:

a. Make a for-loop that calculates the sum for each column of the matrix
b. Use the apply function to perform the same operation

```
mat_x <- matrix(1:120, nrow = 20, ncol = 6)
```

3. Create a data frame with two numeric and one character vector (if you get stuck on this step, you can use the code below).

Write a loop that loops over the columns and reports the mean of the column if it is numeric and the number of observations if it's a character vector.

```
vector1 <- 1:10
vector2 <- c(letters[1:10])
vector3 <- rnorm(10, sd = 10)

df <- data.frame(vector1, vector2, vector3, stringsAsFactors = FALSE)
```

4. Using a loop, generate 10 random normals from distributions with means of -10, 0, 10, and 100. If you remember from the first assignment, the function is **rnorm(n, mean)** whereas n is the number of observation and mean is the mean. Store them in a new matrix called **mat_distributions**.

5. Use a simple 'ifelse' statement to add a new column 'male.teen' to the data frame. This is a dummy variable, indicating 1 if the observation is a male younger than 20 years and 0 otherwise.

```
respondent.df = data.frame(name = c("Sue", "Eva", "Henry", "Jan", "Mary", "John"),
                           sex = c("f", "f", "m", "m", "f", "m"),
                           years = c(21,31,29,19, 23, 33))
```

6. Using the same **respondent.df** dataset, create a new column where values greater than 30 in the column **year** are coded as NA. Other values should stay the same Use both **ifelse** and **if_else**.

7. Use tapply to calculate the average, minimum, and maximum number of years for female and male respondents.

8. Copy and paste the following code to create a small dataset for this exercise.

Create a loop that will number the publications for each authors. "Tidy" the data so that each row represents one author only.

```
author = c("Author1", "Author1",
           "Author2", "Author3", "Author3",
           "Author3", "Author4", "Author5")

pub = c("Pub1", "Pub2", "Pub3", "Pub4",
        "Pub5", "Pub6", "Pub7", "Pub8")

type = c("preprint", "article", "preprint",
         "article", "article", "preprint", "preprint", "article")

data = as_tibble(cbind(author, pub, type))
```

The output will look like this in the long format:

| author | pub | type | PubNum |
| --- | --- | --- | --- |
| Author1 | Pub1 | preprint | 1 |
| Author1 | Pub2 | article | 2 |
| Author2 | Pub3 | preprint | 1 |
| Author3 | Pub4 | article | 1 |
| Author3 | Pub5 | article | 2 |
| Author3 | Pub6 | preprint | 3 |
| Author4 | Pub7 | preprint | 1 |
| Author5 | Pub8 | article | 1 |

And in the wide format:

| author | pub1 | type1 | pub2 | type2 | pub3 | type3 |
| --- | --- | --- | --- | --- | --- | --- |
| Author1 | Pub1 | preprint | Pub2 | article | | |
| Author2 | Pub3 | preprint | | | | |
| Author3 | Pub4 | article | Pub5 | article | Pub6 | preprint |
| Author4 | Pub7 | preprint | | | | |
| Author5 | Pub8 | article | | | | |