

# Lab 3: Programming Fundamentals

INSERT YOUR NAME HERE (INSERT YOUR UW NETID HERE)

Due by 23:59pm on Jan 29, 2024

**Total Points: 35**

## Part 1. Loops and Iterations (5+5 pts)

1. Generate a vector `norm_vec` with length  $n = 1000$  from the normal distribution with mean 1 and variance 5.
  - Write a `for()` loop with `if()` statement to count the number of entries in `norm_vec` that are strictly larger than 0.
  - Re-do the above task without using any `for()` loop (i.e., apply the build-in function in R with logical vector operation).
  - Is the execution time of using `for()` loop longer than the vectorization? (Return a logical TRUE/FALSE).

```
set.seed(123) ## Don't change this line. It makes the result reproducible.
# Your code here
```

2. Keep generating a pair of integers from the binomial distribution `Bin(30,0.3)`. How many times did you generate the pair of integers until both integers are equal? (Implement it in two methods: `while` and `repeat`.)

```
set.seed(123) ## Don't change this line. It makes the result reproducible.
# Your code here (use `while`)
```

```
set.seed(123) ## Don't change this line. It makes the result reproducible.
# Your code here (use `repeat`)
```

## Part 2. Apply Operations (2+5 pts)

1. Create a vector `a` that contains all the even numbers ranging from -4 to 10. Compute the cosine value of each entry using the `sapply()` function.

```
# Your code here
```

2. Run the code below to obtain a list of results. Here, `lm()` function is used to run a linear regression.

```
linearMod = lm(dist ~ speed, data = cars)
```

- What are the names of the elements in the list `linearMod`?
- Compute the mean and variance of each column of the `model` element in the list `linearMod` using `apply()` function.
- Compute the length of each element in the list `linearMod`. How many of them have length 2?

```
# Your code starts from here
```

### Part 3. More Data Frames and Apply (4+3+2+3+2+4 pts)

We will examine data from the 2016 Summer Olympics in Rio de Janeiro, originally taken from <https://github.com/flother/rio2016>. Below, we read in the data and store it as `rio`. All the following questions will be answered based on `rio`.

```
rio =  
  read.csv(url("https://github.com/zhangyk8/zhangyk8.github.io/raw/master/_teaching/file_stat302/Data/rio2016.csv"),  
           header = TRUE)
```

1. What kind of object is `rio`? What are its dimensions and columns names of `rio`? Is there any missing data (i.e., NA)?

- Answer in words: What does each row represent?

```
# Your code starts from here
```

2. How many athletes competed the 2016 Summer Olympics? How many countries were represented? How many athletes had duplicate name? (Hint: Look at the function `duplicated()`.)

```
# Your code starts from here
```

3. Which country brought the most athletes, and how many was this? (Hint: for a factor variable `f`, you can use `table(f)` to see how many elements in `f` are in each level of the factor.)

```
# Your code starts from here
```

4. How many medals of each type – gold, silver, bronze – were awarded at this Olympics? Are they equal? (Output the logical TRUE/FALSE.)

- Answer in words: Is this result surprising, and can you explain what you are seeing?

```
# Your code starts from here
```

5. Create another column called `total` in the data frame `rio` which adds the number of gold, silver, and bronze medals for each athlete. Which athlete had the most number of medals?

```
# Your code starts from here
```

6. Using `tapply()`, calculate the total medal count for each country. Save the result as `tot_nat`. Which country had the most number of medals, and how many was this? How many countries had zero medals?

```
# Your code starts from here
```