

CIND 123

Data Analytics: Basic Methods

Assignment 1 (10%)

[Insert your full name]

[Insert course section & student number]

Instructions

This assignment can be submitted using either Python or R, whichever you prefer.

- If using R, you must submit an RMD file with its knitted file (PDF or HTML). To learn more about knitting and R markdown, visit [R Markdown](#).
- If using Python, you must submit an IPYNB file and its exported PDF/HTML with clearly printed/shown answers.

Failing to submit both files ((RMD + knitted PDF/HTML) OR (IPYNB + PDF/HTML)) will be subject to a 30% mark deduction.

NOTE: IF YOU USE R STUDIO, YOU SHOULD NEVER HAVE `install.packages` IN YOUR CODE; OTHERWISE, THE `Knit` OPTION WILL RAISE AN ERROR. COMMENT OUT ALL PACKAGE INSTALLATIONS BUT KEEP `library()` CALLS.

NOTE: If you answer the questions in R, all your answers should be in R (ignore Python questions). If you answer the questions in Python, all your answers should be in Python (ignore R questions). You are not allowed to switch languages in this assignment.

Question 1 (50 points)

Q1a (15 points)

Create and print a vector `x` (or list in Python) with all odd numbers from 5 to 120 and a vector `y` containing multiples of 5 in the same range.

HINT: use `seq()` in R or `range()` in Python.

In []:

Calculate the sum of the lengths of the vectors `x` and `y`.

HINT: use `length()` in R or `len()` in Python.

In []:

Q1b (15 points)

Create a new vector (or list in Python), `y_cube`, with the cube of elements at indices 2, 4, 6, 10, 15, 18, 21, and 25 from the variable `y`.

HINT: Use indexing rather than a for loop.

In []:

Calculate the mean and median of the LAST five values from `y_cube`.

In []:

Q1c (10 points)

For a given factor variable

`factorVar <- factor(c("2.5", "7.1", "6.3", "4.8"))` in R or

`factorVar = ["2.5", "7.1", "6.3", "4.8"]` in Python,

would it be correct to use the following commands to convert factor to number?

`as.numeric(factorVar)` in R or `map(float, factorVar)` in Python.

If not, explain your answer and provide the correct one.

```
In [ ]: # Explain it in plain text. If you think it is not correct, what is the correct one?
```

Q1d (10 points)

A comma-separated values file `dataset.csv` consists of missing values represented by Not A Number (null) and question mark (?). Read this type of file in R or Python and preprocess the missing values.

NOTE: Please ensure you have saved the `dataset.csv` file in your current working directory.

```
In [ ]:
```

Print the head of the file to ensure everything is correct. **DO NOT** print the whole file.

```
In [ ]:
```

Question 2 (50 points)

Q2a (15 points)

Compute:

$$\sum_{n=4}^{25} \frac{(-1)^{n+1}}{(2n+1)!}$$

HINT: Use `factorial(n)` in R or `math.factorial(n)` in Python to compute $n!$.

Q2b (15 points)

Compute:

$$\prod_{n=2}^6 \left(3n + \frac{2}{3^n} \right)$$

NOTE: The symbol \prod represents multiplication.

Q2a and Q2b Guidance

For Q2a and Q2b, you'll need to apply basic mathematical concepts in R or Python. If you understand how to code in R or Python and the basic math operations (like factorial, powers, sums, and products), you're all set!

Example for Understanding: Let's start with a simpler example: Compute the sum of the first 5 square numbers. In mathematical terms, this is:

$$\sum_{n=1}^5 (n)^2$$

which translates to $1^2 + 2^2 + 3^2 + 4^2 + 5^2$.

In R, you could solve this by:

```
numbers <- 1:5           # Create a vector of numbers 1 to 5
squared_numbers <- numbers^2 # Square each element of the vector
sum_squares <- sum(squared_numbers) # Sum up the squared numbers
print(sum_squares)
```

In Python, you could solve this by:

```
numbers = range(1, 6) # Create a List of numbers from 1 to 5
squared_numbers = [n**2 for n in numbers] # Square each element of the List
sum_squares = sum(squared_numbers) # Sum up the squared numbers
print(sum_squares)
```

Now, come back to Q2a and Q2b. Your task in those questions is similar. You'll use loops to iterate over a range of numbers, applying the specified mathematical operations at each step. Remember to use built-in functions like `factorial()`, `sum()`, and `prod()` in R, or `math.factorial()`, `sum()`, and `math.prod()` in Python when needed.

```
In [ ]: # Q2a
```

```
In [ ]: # Q2b
```

Q2c (10 points)

Describe what the following R command does: `rep(1:3, each = 2)`

NOTE: If you answer the questions in Python, provide an equivalent command and explain its behavior.

In []:

Q2d (10 points)

Describe the purpose of `is.logical()`, `is.character()`, `is.numeric()`, and `is.na()` functions in R.

In Python, describe the purpose of `isinstance()`, `type()`, and `pd.isna()` functions from the pandas library.

Please use `x <- c("a", FALSE, "b", NA, 2, TRUE)` in R or `x = ["a", False, "b", None, 2, True]` in Python to explain your description.

In []:

In []: