



### Case 3 Writing Functions

#### Part 1. Correlation between Two Vectors

The linear relationship of two size  $n$  vectors is called correlation. Given two vectors  $x$  and  $y$ , to compute for the correlation  $\rho(x, y)$ , we first standardize each vectors:

$$x = \frac{(x_i - \text{mean}(x))}{\text{std}(x)} \quad y = \frac{(y_i - \text{mean}(y))}{\text{std}(y)}$$

Then, we calculate for the dot product of the standardized vectors and divide it by  $n - 1$ .

$$\rho(x, y) = \frac{\sum_{i=1}^n x_i y_i}{n - 1}$$

Create a function that can take in any two sized  $n$  vector and return its correlation value. Using the delivery time dataset (deliverytime.csv). Calculate the correlation between deltime and ncases as well as the correlation between deltime and distance.

#### Part 2. Continuous Functions

Consider the continuous function:

$$f(x) = \begin{cases} x^2 + 2x + 3 & \text{if } x < 0 \\ x + 3 & \text{if } 0 \leq x < 2 \\ x^2 + 4x - 7 & \text{if } 2 \leq x \end{cases}$$

Write a function tmpFn which takes a single argument xVec. The function should return the vector of values of the function  $f(x)$  evaluated at the values in xVec.

What is  $f(x)$  is  $x = -1, 0, 2, 5$ ?

#### Part 3. Data Processing Functions

Daily call center data is summarized in three CSV files, one for 2013, 2014 and 2015. Read the three files into R Studio. Create a single function that answers each of the following business questions:

1. What are the total number of calls per month from 2013 to 2015?
2. What is the average number of calls per Month?
3. For three years, how many invalid calls were made? How many were made on each type?
4. How many times on average did a subscriber call all throughout the period?
5. What were the top transactions availed by the subscribers when calling?

The input of the function would just be all data from 2013 to 2015, and the question number. The output would be a data frame showing the results of the summarization of the selected question number.