QMB 6358: Software Tools for Business Analytics
Executive Development Center
College of Business
University of Central Florida
Fall 2021

Assignment 2

Due Tuesday, September 14, 2021 at 11:59 PM
in your GitHub repo.

Instructions:

Complete this assignment within the space on your GitHub repo in a folder called `assignment_02`. In this folder, save your answer to Question 1 in a file called `Q1_functions.R`. In the same folder, save a copy of the sample file called `Q2_testing.R` that will contain your R code for Question 2.

When you are finished, upload your code to your GitHub repo using the interface in a browser. You are free to discuss your approach to each question with your classmates but you must upload your own work.

Question 1:

Create functions to perform the following calculations. Insert your function definitions in the file `Q1_functions.R` from the `assignment_02` folder in the QMB6358F21 course repository.

For each function, there are at least two solutions. One is to use a `for` loop and the other is to use arithmetic operations on the inputs in vector form. Following the function design recipe, create three examples to test each of your functions. For your examples, a vector can be created with the `c()` function, as in `x <- c(1, 2, 3, 2, 2)`. It is fine to choose simple examples to test your function that you can work out by hand, as long as they test that the function works correctly.

a) Write a function `var_x` that takes in a vector `x`, calculates the average $\bar{x}$, and returns the variance of the numbers in `x`. That is, calculate the following formula:

$$VAR(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

b) Write a function that calculates the covariance between two vectors `y` and `x`, of equal length n, and return this as the output from the function `covar_y_x`. That is, calculate the following formula, in which $\bar{x}$ and $\bar{y}$ are the averages of the values in `x` and `y`:

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

c) Write a function that calculates the Ordinary Least Squares (OLS) slope coefficient between two vectors `y` and `x`, of equal length n, and return this as the output from the function `slope_y_x`. That is, calculate the following formula, in which $\bar{x}$ and $\bar{y}$ are the averages of the values in `x` and `y`:

$$SLOPE(y, x) = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

You may use other functions that you have already created to simplify this function.

d) Write a function `sum_sq_resid` that takes in a vector `x`, a vector `y`, and two numbers `beta_0` and `beta_1`, and returns the sum of the squared residuals from the numbers in `x`, `y` and the number `beta`. That is, calculate the following formula:

$$SSR(x, y, \beta_0, \beta_1) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

Note that β can be any number and is not necessarily the optimal slope coefficient from part (c).

Question 2:

Using the three examples you created in the function design recipes for each of the functions in Question 1, test your library of functions.

1. Enter the examples in the script called `Q2_testing.R` that reads in your library of functions from the script `Q1_functions.R`.

2. Define the functions one-by-one, by running the blocks of code in `Q1_functions.R` that define, for example, the function `sum_sq_dev`.

3. Test the functions one-by-one, by running the block of code in `Q2_testing.R` that each function.

4. Check whether the results are correct. If there are any errors or incorrect calculations, make adjustments to `Q1_functions.R` and run the tests in `Q2_testing.R` again.

5. Once they are correct, push the files to your GitHub repository to submit the corrected version. If you made the changes within your folder in the corresponding folder for your GitHub repository, you can use GitHub Desktop to commit those changes and push the changes to your remote repository. To verify that the changes were made, refresh your browser on the Webpage with the GitHub repository.