ISTA 370: Homework 1

Due by Monday, 11:59 PM, September 14th

September 7, 2015

The goal of this homework is to review some statistical concepts (using R), you will need to know for this class. First, we will go over the concept of functions in R and shall do some exercises and then we will look at the concept of variability. If you are new to R, I recommend you go over the R reference sheet on D2L to look for various R commands and functions.

1 Functions in R

One of the major strengths of R is the user's ability to add functions. A function in R or in any programming language is like a black box. Once a function has been defined, one does not need to worry about it's structure but only the inputs you give to the function and the output it produces. The structure of an R fucntion is given as follows:

```
myfunction <- function(arg1, arg2, ... ){
statements
return(object)
}
```

If you carefully look at the structure, you can break it down into following pieces:

- Name of the function: *myfunction*. This is user defined and you can name it anything based on R guidelines.

- Function inputs (also called function arguments): *arg1, arg2, etc.* You can have any number of arguments to a function. The arguments can be anything. It can be a number, a string, a vector, a matrix, etc.
- Function body (*statements*): This is where you define your function using R statements.
- Function output (*object*): This is the result your function returns.

Now let's consider an example of a function which calculates the area of a circle for any given radius. Read the comments carefully to see what's happening in the function definition.

```
areaofcircle <- function(radius){
aofc = pi*(radius^2) # Here we define the area calculation
return(aofc) # Function returns the calculated area
}
areaofcircle(3); # Invoking the function by passing in the radius.

## [1] 28.27433
```

1. (10 Points) Based on what you learnt about functions from the above example, write a function in R which outputs the slope of a line for any two given points. *Hint*: The arguments you pass to the function are the coordinates of the two points.
2. (2.5 Points) Using the function you just defined, find the slope of a line joining the two points: $(-11, 13)$ and $(8, -17)$.
3. (2.5 Points) Now find the slope of a line joining the two points $(-7, -23)$ and $(19, 31)$, using the function you defined.
4. (10 Points) Now let's consider a scenario where you pass in a vector as an input to a function, which returns a number. Write a function in R, which takes in a vector of numbers as an input and outputs the product of the minimum and maximum numbers in the vector.
5. (4 points) Consider a vector of 10 numbers: $99, 78, 468, 35, 26, 290, 556, 98, 78, 146$. Using the function you defined in the last question, find the product of the minimum and maximum numbers in this vector.

6. (1 Points) Manually check if your function is generating the answer you expect. Answer yes or no. Explain why.

If you were able to finish this exercise without much trouble, you are in a good shape! If you had a tough time doing this exercise, you need more practice with R. I recommend you review R tutorials yourself and ask for help when needed. We will keep building on these concepts over the course of this semester. It's important you understand the basics.

2 Variability

7. Using R, type $data()$ to see which datasets are loaded. Look for $mtcars$. If you find it, type $attach(mtcars)$. Otherwise install the *datasets* package and then find and attach $mtcars$. Type $names(mtcars)$ to get the names of all the variables in the dataset. Type $View(mtcars)$ to see the data table.

   (a) (8 Points) Make a histogram of mpg.

   (b) (12 Points) Calculate the mean and population variance of mpg. Remember to use the *popvar* function discussed in class (Lecture 4, Slide 33).