# Exercise 1 : Beginning R and descriptive statistics

Grading system: Total obtainable points are 5.One point each for Task 3,4,5,6, and 7. One point for bonus task.

Task 0: Introduction:

Motivation for software

Excel vs R vs SPSS vs Matlab

Task 1: Install R and R Studio or use <https://www.tutorialspoint.com/execute_r_online.php>

Task 2: Get familiar with the R environment

Task 3: A hello world program in R .

Task 4: Given a random dataset, plot a graph for a column of the data and save it .

Task 5: Using that same dataset, summarize a random dataset with mean, median, variance, standard deviation.

Task 6: Play with that dataset by changing the values, plotting and summarizing it.

Task 7: Use the builtin iris dataset, plotting and summarizing it.

Task 8: Copy all relevant descriptions, code and results for task 3,4,5, 6, and 7 into a word document and upload to canvas, assignment 1 section.

Bonus task: Find the errors in the code, edit it, and get the output and paste it into your document.

**Task 3:**

print(“hello world”)

**Task 4:**

x <- c( 2,3,4,5,21,6,3,5)

print(x)

plot(x)

hist(x)

**Task 5:**

x <- c( 2,3,4,5,21,6,3,5)

print(x)

mean(x)

median(x)

var(x)

sqrt(var(x))

summary(x)

**Task 6:**

x <- c( 2 ,3 ,4 ,5, 6, 21, 6, 3, 4, 5, 100 , -20 , -20)

print(x)

hist(x)

plot(x)

summary(x)

**Task 7:**

print(iris)

print(iris$Sepal.Length)

plot(iris$Sepal.Length)

hist(iris$Sepal.Length)

summary(iris)

**Bonus Task:**

x <- c( 2,3,4,5,21,6,3,5)

print(x)

mean(x)

median(y)

var(x)

sqrt(var(x))

summary(X)

Statistical analysis in R is performed by using many in-built functions. Most of these functions are part of the R base package. These functions take R vector as an input along with the arguments and give the result.

The functions we are discussing in this exercise are mean, median and variance.

## Mean

It is calculated by taking the sum of the values and dividing with the number of values in a data series.

The function **mean()** is used to calculate this in R.

### Syntax

The basic syntax for calculating mean in R is −

mean(x)

* **x** is the input vector.

Example

# Create a vector.   
x <- c(12,7,3,4.2,18,2,54,-21,8,-5)  
  
# Find Mean.  
result.mean <- mean(x)  
print(result.mean)

When we execute the above code, it produces the following result −

[1] 8.22

## Median

The middle most value in a data series is called the median. The **median()** function is used in R to calculate this value.

### Syntax

The basic syntax for calculating median in R is −

median(x)

Following is the description of the parameters used −

* **x** is the input vector.

### Example

# Create the vector.  
x <- c(12,7,3,4.2,18,2,54,-21,8,-5)  
  
# Find the median.  
median.result <- median(x)  
print(median.result)

## Variance and Standard deviation

The variance**()** function is used in R to calculate this value.

### Syntax

The basic syntax for calculating variance in R is −

var(x)

Following is the description of the parameters used −

* **x** is the input vector.

### Example

# Create the vector.  
x <- c(12,7,3,4.2,18,2,54,-21,8,-5)  
  
# Find the variance.  
variance.result <- var(x)  
print(variance.result)

print(sqrt(var(y)))