R and Python Statistics

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.

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.

```
# Create a vector.
x <- c(12,7,3,4.2,18,2,54,-21,8,-5)

# Find Mean.
result <- mean(x)
print(result)
```

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

you first count the number of data points (n) and arrange the
data points in increasing order.
If the number of data points is uneven, you add 1 to the number
of points and divide the results by 2 to get the rank of the
data point whose value is the median.

```
# Create the vector.
x <- c(5,7,4,8,6)

# Find the median.
median.result <- median(x)
print(median.result)
```

median =6

```
# Create the vector.
x <- c(5,8,4,9)
# Find the median.
median.result <- median(x)
print(median.result)
```

median =6.5

Mode

The mode is the value that has highest number of occurrences
in a set of data.
mode can have both numeric and character data.
R does not have a standard in-built function
to calculate mode.
So we create a user function to calculate mode

of a data set in R.

This function takes the vector as input and

gives the mode value as output.

```
# Create the function.
getmode <- function(v) {
uniqv <- unique(v)
uniqv[which.max(tabulate(match(v, uniqv)))]
}

# Create the vector with numbers.
v <- c(2,1,2,3,1,2,3,4,1,5,5,3,2,3)

# Calculate the mode using the user function.
result <- getmode(v)
print(result)

# Create the vector with characters.
charv <- c("h","hit","they","a","a")

# Calculate the mode using the user function.
result <- getmode(charv)
print(result)
```

Python Statistics

```
import numpy

speed = [12,7,3,4.2,18,2,54,-21,8,-5]
```

```
x = numpy.mean(speed)

print(x)
```

```
import numpy

speed = [5,7,4,8,6]

x = numpy.median(speed)

print(x)
```

Mode

The Mode value is the value that appears the most number of times:

The SciPy module has a method for this.

Use the SciPy mode() method to find the number that appears the most:

```
from scipy import stats

speed = [99,86,87,88,111,86,103,87,94,78,77,85,86]

x = stats.mode(speed)

print(x)
```