

## Overview of Assignment 1

For this assignment, I was tasked with creating a C++ program. This program would read a csv file; store data in numeric vectors; and display various statistics.

Regarding data exploration, C++ and R provided two different experiences. I had to implement different formulas via C++ functions. This required me to iterate through vectors and extract information. In contrast, I only called R's built-in functions to find the same statistics.

The statistical measures—mean, median, and range—provide useful information on datasets. The mean indicates the typical value for all the observations. The median returns the value from the “center” of the dataset. The range measures the spread of the data. Data exploration can leverage these statistics to summarize datasets.

The covariance and correlation statistics measure the relationship between two variables. Covariance measures how changes in variables affect each other. Correlation measures the strength of the relationship between two variables. For machine learning, these statistics can help model different relationships between features.

```
Opening file Boston.csv.  
Reading line 1  
heading: rm,medv  
New length 506  
Closing file Boston.csv.  
Number of records: 506
```

```
Stats for rm  
Sum: 3180.03  
Mean: 6.28463  
Median: 6.2085  
Range: [3.561, 8.78]
```

```
Stats for medv  
Sum: 11401.6  
Mean: 22.5328  
Median: 21.2  
Range: [5, 50]
```

```
Covariance = 4.49345
```

```
Correlation = 0.69536
```

**Figure 1** Program's output