

Alejo Vinluan

Abv210001

CS 4375.004

## Program 1 Overview

Output:

```
Data from rm vector:
    Sum: 3180.03
    Mean: 6.28463
    Median: 6.209
    Range: 5.219

Data from medv vector:
    Sum: 11401.6
    Mean: 22.5328
    Median: 21.2
    Range: 45

Covariance of rm and medv vector: 4.49345
Correlation of rm and medv vector: 0.69536
```

This program simulated utilizing built in functions in R, by creating our own functions in C++. I believe that utilizing the built in functions of R help for code readability and allow for easier to maintain, scalable code. However, by building these functions in C++, I was able to understand how the code works and was effective.

The statistical measures of mean, median, and range are all important in finding context within data and analyzing the given data. Mean is important as it calculates the average of a given data set. The median finds where the middle value of a given data set is. For example, if the median was significantly higher than the mean, this suggests that there may be low-value outliers within the dataset. Finally, range is important as it shows how large the differences of values occur within the dataset.

Covariance is important because it measures how changes in one variable are associated with changes in another variable [1]. This is important for models such as Linear Regression, where we can see how each of the variables are correlated. Correlation is covariance scaled to  $[-1, 1]$  [1]. It is important because to understand correlation because it gives an understanding of how the data relates with each other and the predictions that can be given. A correlation model near 1 shows that if the value in 1 vector increases, the value in the other vector is likely to increase as well [1]. Both of these values are important in Machine Learning since they can both be used for models such as Linear Regression. The model could utilize both covariance and correlation in order to make more accurate predictions.

## References

- [1] Maridi, K. (2020). Linear Regression. In Machine Learning Handbook (2nd ed., pp. 72–77). essay, Creative Commons License CC BY-NC-ND 4.0.