

A assignment of statistics using C++.

The writing here about generic stats math functions in C++ should not be used as factual information, but as an expression of my opinions.

To obtain a pdf of this page, click [here](#). Please note that the pdf is not a 1 to 1 representation of this page and is rendered using `pandoc` and `texlive`.

The code is located [here](#). This requires GNU make and a G++ with C++11 support.

Building functions

Programming the functions to read CSV and do math on the resulting vectors was not complex at all, bar some confusing instructions. C++ doesn't lend well to easy text processing in my experience, but it does have the basics like splitting and iterators. The math functions are not hyper optimized, but they exist and give the right answers. They mostly depend on the compiler to optimize the code.

Statistical description

Mean, median, and range all share the same general purpose: showing variance. Mean and median do this by showing the distance between each other. If the mean is greater than the median then that means the data has more high values than low values, and vice versa. Range shows the distance between the highest and lowest values in a dataset, meaning that a higher range shows more inherent variance in the dataset, or at the least outliers.

Covariance shows the directional relationship between two variables, whether they are inverse or proportional. Proportionality is determined by a positive number, while inverse-ness (is that even a word?) is determined by a negative number. Correlation shows how those variables actually pair up and actually relate to each other, using a range between $[0.0, 1.0]$ where being closer to one means higher correlation.

The output

Rm:

```
Sum:      3180.03
Mean:     6.28463
Median:   6.208
Range:    5.219
```

Medv:

```
Sum:      11401.6
Mean:     22.5328
Median:   21.2
```

Range: 45
Covariance:4.49345
Correlation:0.69536