Project #1 – Statistics

|  |  |
| --- | --- |
| Course | INFO-1156 Object-Oriented Programming in C++ |
| Professor | Garth Santor, Lianne Wong, and Janice Manning |
| Assigned | February 3rd 2020 |
| Due | February 28th 2020 |
| Weight | 9% |
| Student Name |  |

# Project Description

Create a C console application to compile the following statistics on a list of real number pairs:

* minimum value;
* maximum value;
* median value;
* arithmetic mean;
* [mean absolute deviation – (mean, median, mode)](http://en.wikipedia.org/wiki/Absolute_deviation)
* [variance](http://en.wikipedia.org/wiki/Variance#Normal_distribution) (of a discrete random variable);
* [standard deviation](http://en.wikipedia.org/wiki/Standard_deviation) (of a finite population);
* [mode](http://en.wikipedia.org/wiki/Mode_(statistics)) (including multi-modal lists).
* least squares regression line
* outliers

Your program must handle any length of list. The list will be input (or piped) from the console, or read from a file. The list is terminated with *end-of-stream* (^Z) or *non-numeric* input. A sample input list is posted on FOL.

Keep the output clean and minimal. A sample output file is posted on FOL.

You are to implement your own sort (it must sort two arrays in parallel, or an array of structures). Your grade is dependent upon the sorting algorithm that you choose (to get full marks, implement the quick, merge sort or heap sort). To have a reference about the different sorting algorithms, see the [GATS Companion](http://gats.ca/wp-content/uploads/2020/02/GATS_Companion_to_Searching_and_Sorting.pdf).

A sample of the executable, input file format and output file format are posted on FOL. Your output should be formatted as the output file. All statistics numbers calculated should be displayed to 3 decimal places, where applicable. You should create more input files to verify that all your statistics are done properly.

## Least Squares Formulae

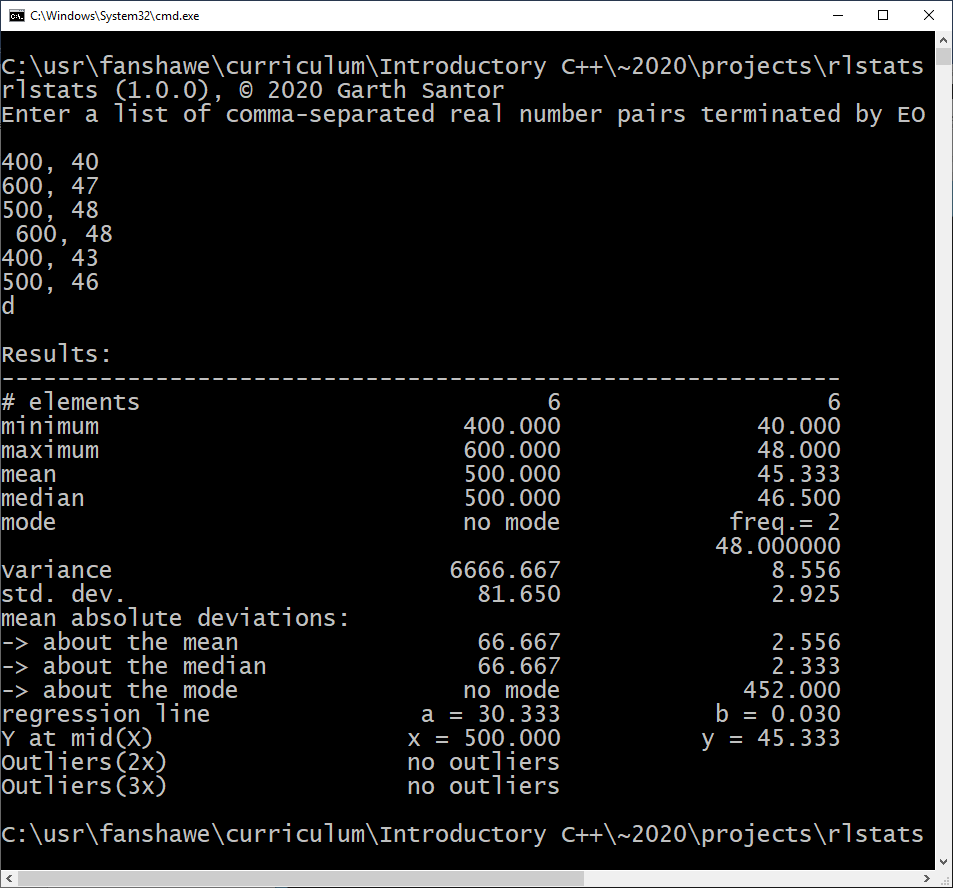
## Outliers

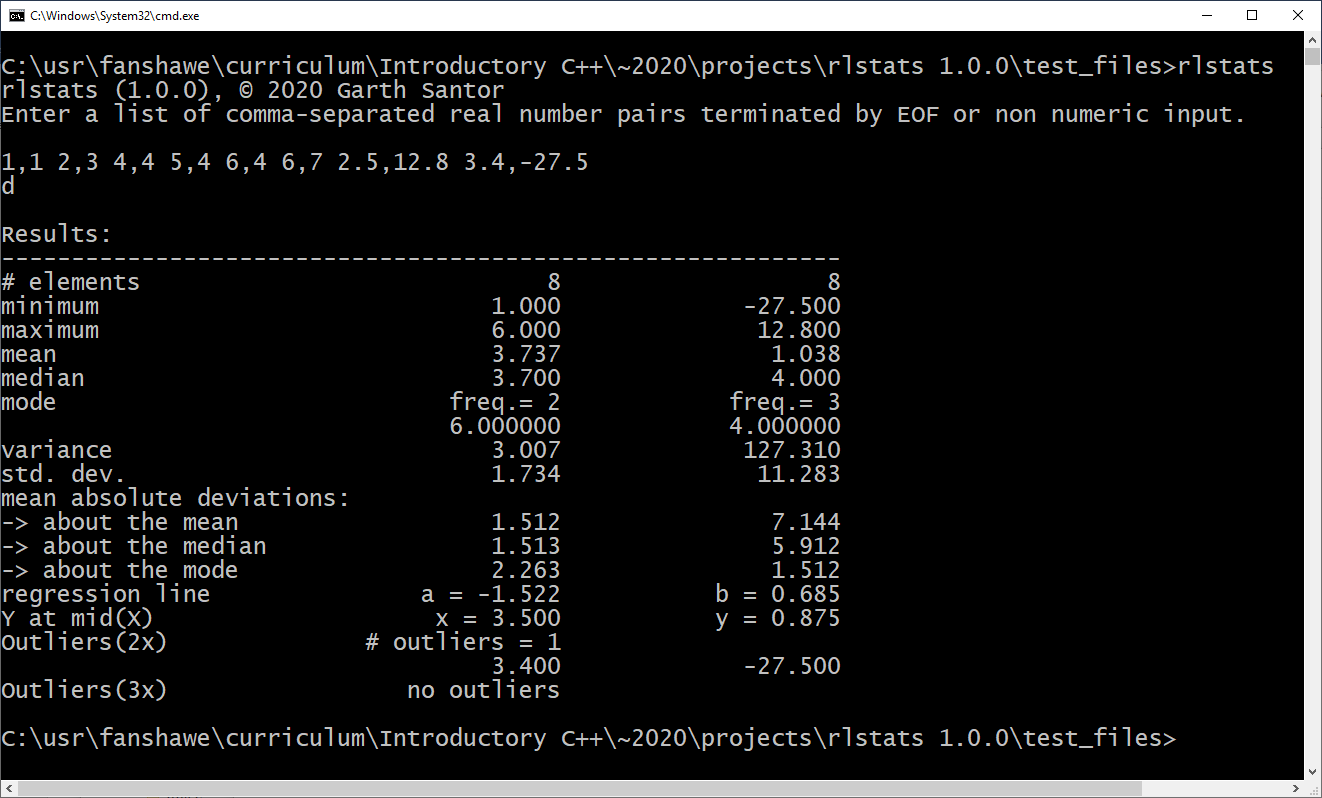
Outliers are values that are a significant distance from the mean. When the absolute difference between a sample and its mean exceeds the standard deviation, it is a 1x outlier. Make a list of 2x and 3x outliers – considering the Y values, the Y mean, and the Y standard deviation.

## Note!

Note that both the modes and the outliers can have multiple results. Additionally, there can be a different number of result for the X data set, and the Y data set, but not so for the outliers as you will be reporting all the outliers as (x, y) pairs (x in the first column, matching y in the second).

## Examples:





1,1 2,3 4,4 5,4 6,4 6,7 2.5,12.8 3.4,-27.5

# Grading Criteria

|  |  |  |
| --- | --- | --- |
| **Functional Requirements** |  |  |
| Number of samples reported (including no samples) | 5% | 5% |
| Implement your own sorting algorithm  Put a comment at the top of your sorting algorithm to specify which sorting algorithm is used (no comment -5)  1 Mark – Bubble Sort  5 Marks – Insertion/Selection Sort  10 Marks – Quick/Merge/Heap Sort  -5000 Marks – (Deducted) if implement BOGO (aka BOZO sort) | 10% | 10% |
| Minimum / maximum value correctly reported in all cases | 10% | 10% |
| Median correctly reported in all cases | 5% | 5% |
| Mean correctly reported in all cases | 5% | 5% |
| Mean absolute distribution about the mean. | 5% | 5% |
| Mean absolute distribution about the median. | 5% | 5% |
| Variance correctly reported in all cases | 5% | 5% |
| Standard deviation correctly reported in all cases | 5% | 5% |
| Mode   * Unique mode correctly reported in all cases * Multi-mode correctly reported in all cases * No-mode correctly reported in all cases | 5%  5%  5% | 5%  5%  5% |
| Mean absolute distribution about the mode (if unimodal). | 5% | 5% |
| Regression Line   * Y-intercept * Slope * Interpolated Y from midpoint of X-range | 5%  5%  5% | 5%  5%  5% |
| Outliers | 10% | 10% |
| **Non-functional requirements** |  |  |
| Executables program are **not** named ‘stats.exe’. | -10% |  |
| Non-trivial calculations are not done in functions | Up to -10% |  |
| Output is not properly formatted   * No fixed point with three decimal places * Dynamic aligned columns | Up to -10%  Up to -10%  **Up to +5%[[1]](#endnote-1)** |  |
| Penalties from *C & C++ Grading Guide* ***v2.2.0*** | various |  |
| Late submission   * One to five days late * More than five days late | -10%/day  -100% |  |
| **Total** | **100%** | **100%** |

# Submission Requirements

1. Submit **entire Visual Studio project directory** to Fanshawe Online
   1. Delete ***all*** debug and release directories.[[2]](#endnote-2)
   2. Submit in a .ZIP, .7z archive file.

1. Total grade cannot exceed 100% [↑](#endnote-ref-1)
2. Alternatively, you can ‘clean’ your project for submission by downloading ‘vsclean’ a Visual Studio Solution Cleaner from [www.gats.ca](http://www.gats.ca) . [↑](#endnote-ref-2)