Project #2 – Lies, Damn Lies, and Statistics

Course	INFO-1156 Object-Oriented Programming in C++
Professor	Garth Santor, Janice Manning, and Lynn Koudsi
Assigned	Tuesday, February 9th, 2021
Due	Monday, March 8th, 2021 by 11:59 pm
Weight	6%

Project Description

Create a C17 console application to compile the following statistics on a list of real numbers:

- Number of values
- minimum value;
- maximum value;
- median value;
- arithmetic mean;
- mean absolute deviation – (mean, median, mode)
- variance (of a discrete random variable);
- standard deviation (of a finite population);
- mode (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 ^D on Linux) 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). Sorting algorithms can be found at GATS Companion to Searching and Sorting.

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

$$ss_{xx} = \left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2$$

$$ss_{xy} = \left(\sum_{i=1}^{n} x_i y_i\right) - n\bar{x}\bar{y}$$

$$b = \frac{ss_{xy}}{ss_{xx}}$$

$$a = \bar{y} - b\bar{x}$$

Where **b** is the slope, and **a** is the y-intercept.

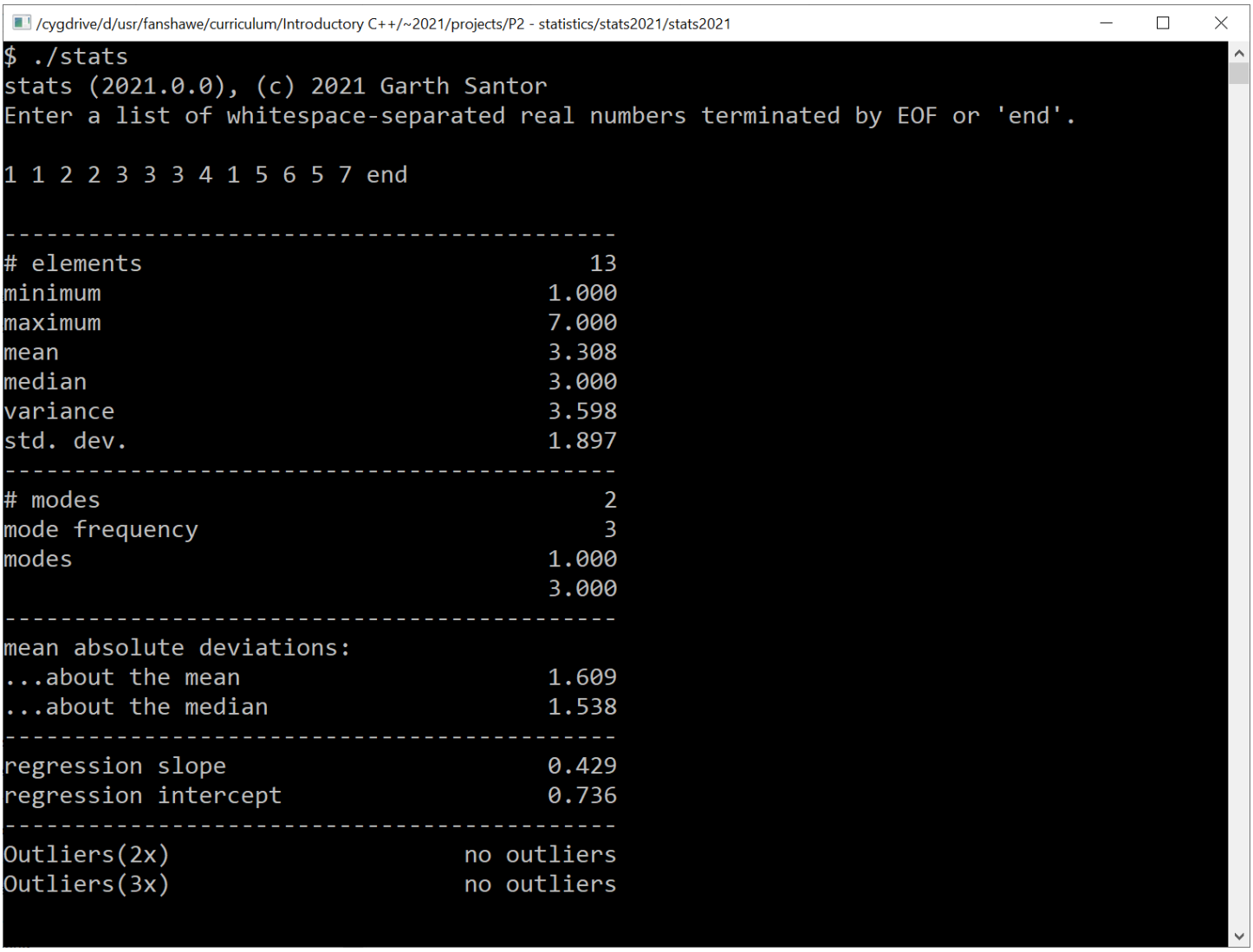
Outliers

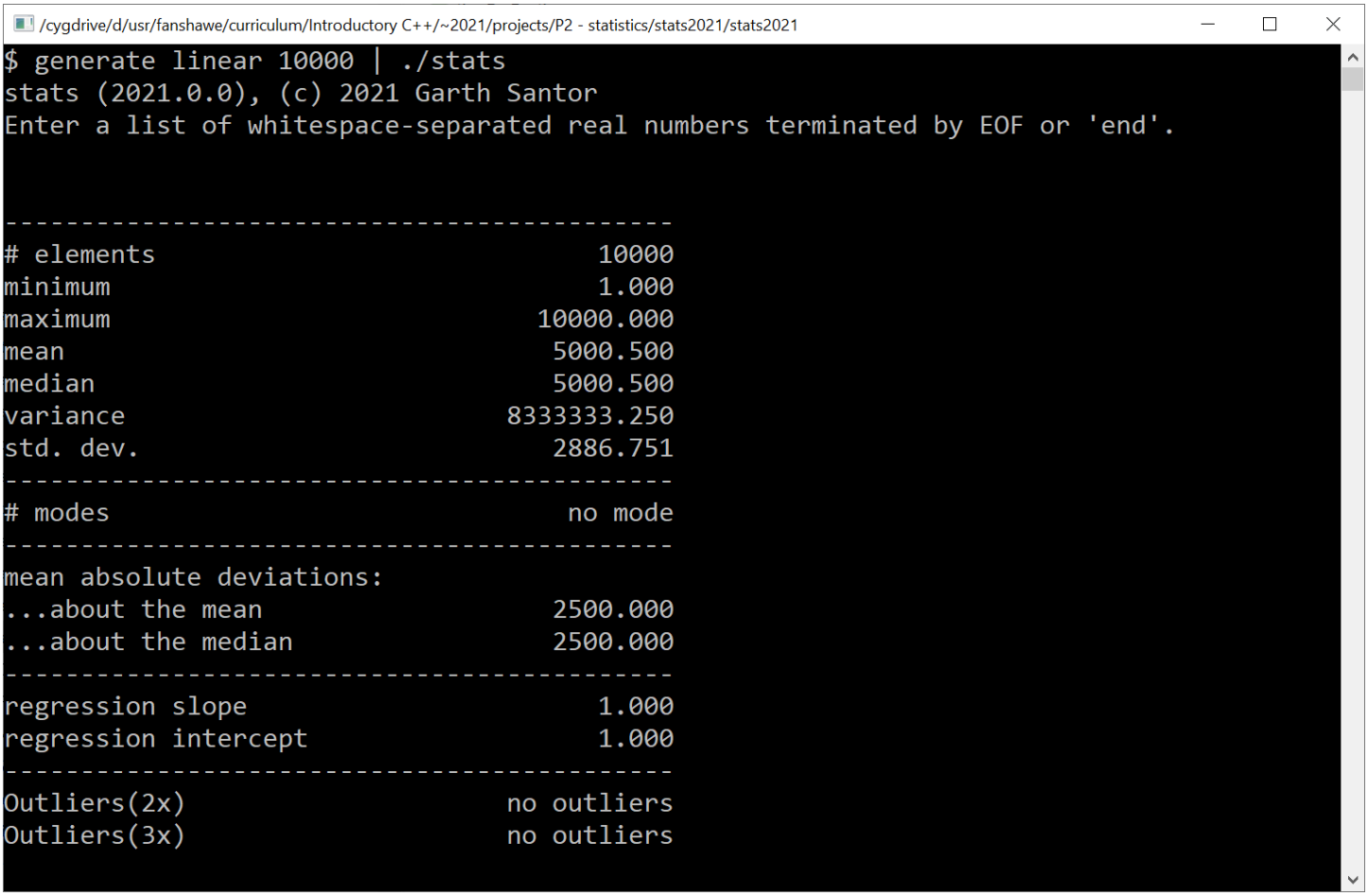
Outliers are values that are a significant distance from the mean. When the absolute difference between a sample and its mean exceeds the standard deviation, it is a 1x outlier. Make a list of 2x and 3x outliers – considering the values, the mean, and the standard deviation.

Note!

1. Note that both the modes and the outliers can have multiple results.
2. A data set with all values occurring the same number of times **has no mode**!

Examples:





Grading Criteria

Input Requirements	Weight	Points	Awarded	Grade
Reads double precision real numbers	1%	1	1	1%
Terminates input on EOF	4%	1	1	4%
Doesn't handle any size data set (array is **NOT** heap allocated)	**-25%**	1		0%
Skips bad inputs (only terminates on 'end' or EOF)	5%	1	1	5%
Input from keyboard	1%	1	1	1%
Input from file named on command line	5%	1	1	5%
Reports bad filename	4%	1	1	4%
Statistics Requirements				
Number of values reported	2%	1	1	2%
Minimum value correctly reported in all cases	2%	1	1	2%
Maximum value correctly reported in all cases	3%	1	1	3%
Mean correctly reported in all cases	5%	1	1	5%
Median correctly reported in all cases	5%	1	1	5%
Variance correctly reported in all cases	5%	1	1	5%
Standard deviation correctly reported in all cases	5%	1	1	5%

Mode: reports number of occurrences	3%	1	1	3%
Mode: Unique mode correctly reported in all cases	4%	1	1	4%
Mode: Multi-mode correctly reported in all cases	4%	1	1	4%
Mode: No-mode correctly reported in all cases	2%	1	1	2%
Mean absolute distribution about **mean**	4%	1	1	4%
Mean absolute distribution about **median**	2%	1	1	2%
Mean absolute distribution about **mode** (if unimodal)	1%	1	1	1%
Regression Line: slope	5%	1	1	5%
Regression Line: intercept	5%	1	1	5%
Outliers: reports no outliers	2%	1	1	2%
Outliers: reports number of outliers	2%	1	1	2%
Outliers: reports 2x outliers	3%	1	1	3%
Outliers: reports 3x outliers	3%	1	1	3%
Non-functional requirements				
Sort is: bubble	1%	1		0%
Sort is: insert or selection	5%	1		0%
Sort is: <stdlib.h> qsort	2%	1		0%
Sort is: quick/merge/heap	10%	1	1	10%
Sort is: Bogo/Bozo	**-5000%**	1		0%
Labels are **not** left justified	**-5%**	1		0%
Values are **not** right justified	**-5%**	1		0%
Values are not rounded to 3 decimal places	**-5%**	1		0%
Multi-file solution	3%	1	1	3%
Penalties				
Penalties from *C & C++ Grading Guide* ***v2.2.0***	**-5%**	1	0	0%
Late submission:	**-10%**	1	0	0%
Total				**100%**

Difficulties
Moderate
Harder
Hardest

Submission Requirements

1. Submit **entire Visual Studio project directory** to Fanshawe Online
    a. Delete ***all*** debug and release directories.[i]
    b. Submit in a .ZIP, .7z archive file.

[i] Alternatively, you can 'clean' your project for submission by downloading 'vsclean' a Visual Studio Solution Cleaner from www.gats.ca .