## Project #2 - Lies, Damn Lies, and Statistics

Course	INFO-1156 Object-Oriented Programming in C++
Professor	Garth Santor, and Lynn Koudsi
_	Tuesday, June 1 <sup>st</sup> , 2021
Due	Monday, Jun 28 <sup>th</sup> , 2021 by 11:59 pm
Weight	9%

## **Project Description**

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

For the following formulae x is the datum, X is the dataset, where  $x \in X$  (x is an element of X).

Metric	Formula
Number of values	The count of the elements in the dataset.
	X
Minimum value	The value that is less than or equal to all values in the set.
	$min \le x, \forall x \in X$
Maximum value	The value that is greater than or equal to all values in the set.
	$max \ge x, \forall x \in X$
Range	$max \ge x, \forall x \in X$ The range of values of the dataset.
	max - min
Statistical median	The middle value of a sorted data set of odd length, or the arithmetic mean of the two closest values to the middle of a sorted data set of even length.
	$Y = sorted\{x_1 \cdots x_n\}$
	$\left( Y_{(N+1)/2}^{(N+1)/2}  if \ N \ is \ odd \right)$
	$Y = sorted\{x_1 \cdots x_n\}$ $\tilde{x} = \begin{cases} Y_{(N+1)/2} & \text{if N is odd} \\ \frac{1}{2} (Y_{N/2} + Y_{1+N/2}) & \text{if N is even} \end{cases}$
Arithmetic mean	The sum of all the values divided by the number of values.
	$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$
Variance	The mean of the squared differences of each sample from the arithmetic mean.
	$variance = s_N^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2$
Standard deviation	The square root of the variance.

	standard deviation = $s_N = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2}$
Frequency table	A table listing the interval, and the count of data set values in that interval, and the percentage of the dataset represented by count of that interval.  Our frequency table will use ten buckets.
	A [ 14.0 40.9) = 17 : 1.70% B [ 40.9 67.8) = 39 : 3.90% C [ 67.8 94.7) = 88 : 8.80% D [ 94.7 121.6) = 144 : 14.40% E [ 121.6 148.5) = 222 : 22.20% F [ 148.5 175.4) = 180 : 18.00% G [ 175.4 202.3) = 159 : 15.90% H [ 202.3 229.2) = 97 : 9.70% I [ 229.2 256.1) = 47 : 4.70% J [ 256.1 283.0] = 7 : 0.70%
Histogram	A graphical depiction of a frequency table – typically a horizontal bar chart.  Labels indicate percentages.
	Grp % 0 2.5 5.0 7.5 10.0 12.5 15.0 17.5 20.0 22.5 25.0  A 1.70% B 3.90% C 8.80% D 14.40% E 22.20% F 18.00%
Logarithmic	G 15.90% H 9.70% I 4.70% J 0.70% A logarithmically transformed dataset is defined as: $x' = \ln x \ \forall x \in X$
transform	This means that the transformed dataset has the natural logarithm of each datum in the original dataset.  Provide both the frequency table and histogram of the log-transformed data.

Your program must handle any length of list (**potentially billions**!) 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 the word *end*. Bad input is to be skipped. Sample input lists are posted on FOL.

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

You will have to sort the data to compute the median and are to implement your own sort. 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.

## **Grading Criteria**

Input Requirements  Reads double precision real numbers Terminates input on EOF Doesn't handle any size data set (array is NOT heap allocated) Skips bad inputs (only terminates on 'end' or EOF) Input from keyboard Input from file named on command line Reports bad filename  Statistics Requirements  Number of values reported Minimum value correctly reported in all cases Maximum value correctly reported in all cases Mean correctly reported in all cases Median correctly reported in all cases Fopulation variance correctly reported in all cases Simple for the first of the firs					
Terminates input on EOF Doesn't handle any size data set (array is NOT heap allocated) Skips bad inputs (only terminates on 'end' or EOF) Input from keyboard Input from file named on command line Reports bad filename Statistics Requirements  Number of values reported Minimum value correctly reported in all cases Maximum value correctly reported in all cases Manage value correctly reported in all cases Mean correctly reported in all cases Median correctly reported in all cases Fopulation variance correctly reported in all cases Median correctly reported in all cases Median correctly reported in all cases Fopulation variance correctly reported in all cases Fopulation variance correctly reported in all cases Fopulation standard deviation correctly reported in all cases Formulation variance variation varian	Input Requirements	Weight	Points	Awarded	Grade
Doesn't handle any size data set (array is NOT heap allocated)  Skips bad inputs (only terminates on 'end' or EOF)  Input from keyboard  Input from keyboard  Input from file named on command line  Reports bad filename  Statistics Requirements  Number of values reported  Minimum value correctly reported in all cases  Range value correctly reported in all cases  Range value correctly reported in all cases  Median correctly reported in all cases  Median correctly reported in all cases  Population variance correctly reported in all cases  Population variance correctly reported in all cases  Frequency Table Requirements (must complete statistics first)  Group labels  Minimum and Maximum ranges  Count is displayed and correct  Frequency Table Requirements (must complete statistics first)  Group labels  Minimum Requirements (must complete Frequency table first)  Group labels  Frequency Table Requirements (must complete Frequency table first)  Group labels  Frequency Table Requirements (must complete Frequency table first)  Group labels  Frequency Table Requirements (must complete Frequency table first)  Bars are correct length  Bars scales are adjusted to next highest multiple of 5% greater  3% 1 1 3%  Bar scales are adjusted to next highest multiple of 5% greater	Reads double precision real numbers	1%	1	1	1%
Skips bad inputs (only terminates on 'end' or EOF) Input from keyboard Input from keyboard Input from keyboard Input from file named on command line Reports bad filename Statistics Requirements  Number of values reported Ninimum value correctly reported in all cases Maximum value correctly reported in all cases Maximum value correctly reported in all cases Median correctly reported in all cases Median correctly reported in all cases Median correctly reported in all cases Population variance correctly reported in all cases Population variance correctly reported in all cases Frequency Table Requirements (must complete statistics first)  Group labels Minimum and Maximum ranges Count is displayed and correct Percentage is displayed and correct Frequency Table Requirements (must complete statistics first)  Group labels Frequency Table Requirements (must complete statistics first)  Group labels Frequency Table Requirements (must complete statistics first)  Group labels Frequency Table Requirements (must complete statistics first)  Group labels Frequency Table Requirements (must complete statistics first)  Group labels Frequency Table Requirements (must complete statistics first)  Group labels Frequency Table Requirements (must complete statistics first)  Bars are correct length Bars are correct length Bars scales are adjusted to next highest multiple of 5% greater	Terminates input on EOF	4%	1	1	4%
Input from keyboard Input from file named on command line Reports bad filename 4% 1 1 4%  Statistics Requirements  Number of values reported Ninimum value correctly reported in all cases Maximum value correctly reported in all cases Range value correctly reported in all cases Median correctly reported in all cases Median correctly reported in all cases Median correctly reported in all cases Fopulation variance correctly reported in all cases Population standard deviation correctly reported in all cases Frequency Table Requirements (must complete statistics first)  Frequency Table Requirements (must complete statistics first)  Group labels Percentage is displayed and correct Frequency Table Requirements (must complete statistics first)  Group labels Frequency Table Requirements (must complete statistics first)  Group labels Frequency Table Requirements (must complete statistics first)  Group labels Frequency Table Requirements (must complete statistics first)  Group labels Frequency Table Requirements (must complete statistics first)  Group labels Frequency Table Requirements (must complete statistics first)  Bars are correct length Fercentages rounded to 2 decimal places Frequency table first)  Bars are correct length Fercentages are adjusted to next highest multiple of 5% greater  3% 1 1 3%	Doesn't handle any size data set (array is <b>NOT</b> heap allocated)	-25%	1		0%
Input from file named on command line Reports bad filename 4% 1 1 4%  Statistics Requirements  Number of values reported Minimum value correctly reported in all cases Maximum value correctly reported in all cases 3% 1 1 3%  Range value correctly reported in all cases Mean correctly reported in all cases 5% 1 1 5%  Median correctly reported in all cases Median correctly reported in all cases Fopulation variance correctly reported in all cases Fopulation variance correctly reported in all cases Fopulation standard deviation correctly reported in all cases Frequency Table Requirements (must complete statistics first)  Frequency Table Requirements (must complete statistics first)  Group labels Count is displayed and correct Frequency Table Requirements (must complete statistics first)  Group labels Count is displayed and correct Frequency Table Requirements (must complete statistics first)  Group labels Thistogram Requirements (must complete Frequency table first)  Group labels Frequency table first  Group labels Thistogram Requirements (must complete Frequency table first)  Bars are correct length Bars are correct length Bars scales are adjusted to next highest multiple of 5% greater  3% 1 1 3%	Skips bad inputs (only terminates on 'end' or EOF)	5%	1	1	5%
Reports bad filename4%114%Statistics RequirementsNumber of values reported2%112%Minimum value correctly reported in all cases2%112%Maximum value correctly reported in all cases3%113%Range value correctly reported in all cases3%113%Mean correctly reported in all cases5%115%Median correctly reported in all cases5%115%Population variance correctly reported in all cases5%115%Population standard deviation correctly reported in all cases5%115%Logarithmic transform5%115%Frequency Table Requirements (must complete statistics first)Group labels2%112%Minimum and Maximum ranges5%115%Count is displayed and correct5%115%Percentage is displayed and correct5%115%Histogram Requirements (must complete Frequency table first)Group labels1%111%Percentages rounded to 2 decimal places1%111%Bars are correct length2%112%Bars scales are adjusted to next highest multiple of 5% greater3%113%	Input from keyboard	1%	1	1	1%
Number of values reported  Number of values reported  Ninimum value correctly reported in all cases  Maximum value correctly reported in all cases  Range value correctly reported in all cases  Range value correctly reported in all cases  Mean correctly reported in all cases  Median correctly reported in all cases  Population variance correctly reported in all cases  Population standard deviation correctly reported in all cases  Population standard deviation correctly reported in all cases  Frequency Table Requirements (must complete statistics first)  Frequency Table Requirements (must complete statistics first)  Group labels  Count is displayed and correct  Percentage is displayed and correct  Frequency Table Requirements (must complete statistics first)  Group labels  Count is displayed and correct  Frequency Table Requirements (must complete statistics first)  Count is displayed and correct  Frequency Table Requirements (must complete statistics first)  Count is displayed and correct  Frequency Table first)  Count is displayed and correct  Frequency Table first)  Count is displayed and correct  Sw 1 1 5%  Percentage is displayed and correct  Frequency Table first)  Group labels  Requirements (must complete Frequency table first)  Bars are correct length  Bars scales are correct  Sw 1 1 2%  Bars scales are adjusted to next highest multiple of 5% greater  3% 1 1 3%	Input from file named on command line	5%	1	1	5%
Number of values reported  Minimum value correctly reported in all cases  Maximum value correctly reported in all cases  Range value correctly reported in all cases  Mean correctly reported in all cases  Median correctly reported in all cases  Median correctly reported in all cases  Median correctly reported in all cases  Population variance correctly reported in all cases  Population standard deviation correctly reported in all cases  Population standard deviation correctly reported in all cases  Frequency Table Requirements (must complete statistics first)  Group labels  Minimum and Maximum ranges  Count is displayed and correct  Percentage is displayed and correct  Mistogram Requirements (must complete Frequency table first)  Frequency Tables  Group labels  Percentages rounded to 2 decimal places  1% 1 1 1%  Percentages rounded to 2 decimal places  1% 1 1 1%  Bars are correct length  Bars scales are adjusted to next highest multiple of 5% greater  3% 1 1 3%	Reports bad filename	4%	1	1	4%
Minimum value correctly reported in all cases 3% 1 1 3% Range 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% Population variance correctly reported in all cases 5% 1 1 5% Population standard deviation correctly reported in all cases 5% 1 1 5% Logarithmic transform 5% 1 1 5% Frequency Table Requirements (must complete statistics first)  Group labels 2% 1 1 2% Minimum and Maximum ranges 5% 1 1 5% Count is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Percentages rounded to 2 decimal places 1% 1 1 1% Percentages rounded to 2 decimal places 1% 1 1 1 1% Percentages rounded to 2 decimal places 1% 1 1 1 1% Percentages are adjusted to next highest multiple of 5% greater 1% 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Statistics Requirements				
Maximum value correctly reported in all cases3%113%Range value correctly reported in all cases3%113%Mean correctly reported in all cases5%115%Median correctly reported in all cases5%115%Population variance correctly reported in all cases5%115%Population standard deviation correctly reported in all cases5%115%Logarithmic transform5%115%Frequency Table Requirements (must complete statistics first)Group labels2%112%Minimum and Maximum ranges5%115%Count is displayed and correct5%115%Percentage is displayed and correct5%115%Histogram Requirements (must complete Frequency table first)Group labels1%111%Percentages rounded to 2 decimal places1%111%Bars are correct length2%112%Bars scales are adjusted to next highest multiple of 5% greater3%113%	Number of values reported	2%	1	1	2%
Range value correctly reported in all cases  Mean correctly reported in all cases  Median correctly reported in all cases  Median correctly reported in all cases  Population variance correctly reported in all cases  Population standard deviation correctly reported in all cases  Population standard deviation correctly reported in all cases  Logarithmic transform  Sw. 1 1 5%  Frequency Table Requirements (must complete statistics first)  Group labels  Count is displayed and correct  Percentage is displayed and correct  Percentage is displayed and correct  Wistogram Requirements (must complete Frequency table first)  Frequency table first  Group labels  Percentages rounded to 2 decimal places  1% 1 1 1%  Percentages rounded to 2 decimal places  1% 1 1 1%  Bars are correct length  Bars scales are correct  3% 1 1 3%  Bar scales are adjusted to next highest multiple of 5% greater  3% 1 1 3%	Minimum value correctly reported in all cases	2%	1	1	2%
Mean correctly reported in all cases5%115%Median correctly reported in all cases5%115%Population variance correctly reported in all cases5%115%Population standard deviation correctly reported in all cases5%115%Logarithmic transform5%115%Frequency Table Requirements (must complete statistics first)Group labels2%112%Minimum and Maximum ranges5%115%Count is displayed and correct5%115%Percentage is displayed and correct5%115%Histogram Requirements (must complete Frequency table first)Group labels1%111%Percentages rounded to 2 decimal places1%111%Bars are correct length2%112%Bars scales are correct3%113%Bar scales are adjusted to next highest multiple of 5% greater3%113%	Maximum value correctly reported in all cases	3%	1	1	3%
Median correctly reported in all cases Population variance correctly reported in all cases Population standard deviation correctly reported in all cases Population standard deviation correctly reported in all cases Frequency Table Requirements (must complete statistics first)  Frequency Table Requirements (must complete statistics first)  Group labels Population standard deviation correctly reported in all cases First Frequency Table Requirements (must complete statistics first)  Group labels First	Range value correctly reported in all cases	3%	1	1	3%
Population variance correctly reported in all cases Population standard deviation correctly reported in all cases Frequency Table Requirements (must complete statistics first)  Frequency Table Requirements (must complete statistics first)  Group labels Minimum and Maximum ranges Count is displayed and correct Percentage is displayed and correct Frequency table first  Histogram Requirements (must complete Frequency table first)  Group labels Percentages rounded to 2 decimal places  Bars are correct length Bars scales are adjusted to next highest multiple of 5% greater  3% 1 1 3%  Bar scales are adjusted to next highest multiple of 5% greater	Mean correctly reported in all cases	5%	1	1	5%
Population standard deviation correctly reported in all cases 5% 1 1 5% Logarithmic transform 5% 1 1 5% Frequency Table Requirements (must complete statistics first)  Group labels 2% 1 1 2% Minimum and Maximum ranges 5% 1 1 5% Count is displayed and correct 5% 1 1 5% Percentage is displayed and correct 5% 1 1 5% Thistogram Requirements (must complete Frequency table first)  Group labels 1% 1 1 1% Percentages rounded to 2 decimal places 1% 1 1 1% 1% 1 1 1% 1% 1 1 1% 1% 1 1 1% 1%	Median correctly reported in all cases	5%	1	1	5%
Logarithmic transform5%115%Frequency Table Requirements (must complete statistics first)Group labels2%112%Minimum and Maximum ranges5%115%Count is displayed and correct5%115%Percentage is displayed and correct5%115%Histogram Requirements (must complete Frequency table first)Group labels1%111%Percentages rounded to 2 decimal places1%111%Bars are correct length2%112%Bars scales are correct3%113%Bar scales are adjusted to next highest multiple of 5% greater3%113%	Population variance correctly reported in all cases	5%	1	1	5%
Frequency Table Requirements (must complete statistics first)  Group labels 2% 1 1 2%  Minimum and Maximum ranges 5% 1 1 5%  Count is displayed and correct 5% 1 1 5%  Percentage is displayed and correct 5% 1 1 5%  Histogram Requirements (must complete Frequency table first)  Group labels 1% 1 1 1%  Percentages rounded to 2 decimal places 1% 1 1 1%  Bars are correct length 2% 1 1 2%  Bars scales are correct 3% 1 1 3%  Bar scales are adjusted to next highest multiple of 5% greater 3% 1 1 3%	Population standard deviation correctly reported in all cases	5%	1	1	5%
Group labels2%112%Minimum and Maximum ranges5%115%Count is displayed and correct5%115%Percentage is displayed and correct5%115%Histogram Requirements (must complete Frequency table first)Group labels1%111%Percentages rounded to 2 decimal places1%111%Bars are correct length2%112%Bars scales are correct3%113%Bar scales are adjusted to next highest multiple of 5% greater3%113%	Logarithmic transform	5%	1	1	5%
Group labels2%112%Minimum and Maximum ranges5%115%Count is displayed and correct5%115%Percentage is displayed and correct5%115%Histogram Requirements (must complete Frequency table first)Group labels1%111%Percentages rounded to 2 decimal places1%111%Bars are correct length2%112%Bars scales are correct3%113%Bar scales are adjusted to next highest multiple of 5% greater3%113%	Frequency Table Requirements (must complet	e statis	tics	first)	
Count is displayed and correct 5% 1 1 5%  Percentage is displayed and correct 5% 1 1 5%  Histogram Requirements (must complete Frequency table first)  Group labels 1% 1 1 1%  Percentages rounded to 2 decimal places 1% 1 1 1%  Bars are correct length 2% 1 1 2%  Bars scales are correct 3% 1 1 3%  Bar scales are adjusted to next highest multiple of 5% greater 3% 1 1 3%	Group labels	2%	1	1	2%
Percentage is displayed and correct5%115%Histogram Requirements (must complete Frequency table first)Group labels1%111%Percentages rounded to 2 decimal places1%111%Bars are correct length2%112%Bars scales are correct3%113%Bar scales are adjusted to next highest multiple of 5% greater3%113%	Minimum and Maximum ranges	5%	1	1	5%
Histogram Requirements (must complete Frequency table first)  Group labels 1% 1 1 1%  Percentages rounded to 2 decimal places 1% 1 1 1%  Bars are correct length 2% 1 1 2%  Bars scales are correct 3% 1 1 3%  Bar scales are adjusted to next highest multiple of 5% greater 3% 1 1 3%	Count is displayed and correct	5%	1	1	5%
Group labels Percentages rounded to 2 decimal places 1% 1 1 1% Bars are correct length Bars scales are correct 3% 1 1 3% Bar scales are adjusted to next highest multiple of 5% greater 3% 1 1 3%	Percentage is displayed and correct	5%	1	1	5%
Percentages rounded to 2 decimal places 1% 1 1 1%  Bars are correct length 2% 1 1 2%  Bars scales are correct 3% 1 1 3%  Bar scales are adjusted to next highest multiple of 5% greater 3% 1 1 3%	Histogram Requirements (must complete Freq	uency t	able	first	)
Bars are correct length  Bars scales are correct  Bars scales are adjusted to next highest multiple of 5% greater  3% 1 1 3%  1 1 3%	Group labels	1%	1	1	1%
Bars scales are correct 3% 1 1 3%  Bar scales are adjusted to next highest multiple of 5% greater 3% 1 1 3%	Percentages rounded to 2 decimal places	1%	1	1	1%
Bar scales are adjusted to next highest multiple of 5% greater  3% 1 1 3%	Bars are correct length	2%	1	1	2%
3% 1 1 3%	Bars scales are correct	3%	1	1	3%
than maximum percentage	Bar scales are adjusted to next highest multiple of 5% greater	3%	1	1	2%
	than maximum percentage	370			370
Non-functional requirements	Non-functional requirements				
Sort is: bubble 0% 1 0%	Sort is: bubble	0%	1		0%
Sort is: insertion sort 5% 1 0%	Sort is: insertion sort	5%	1		0%
Sort is: <stdlib.h> qsort 2% 1 0%</stdlib.h>	Sort is: <stdlib.h> qsort</stdlib.h>	2%	1		0%
Sort is: quick/merge/heap 10% 1 1 10%	Sort is: quick/merge/heap	10%	1	1	10%
Sort is: Bogo/Bozo -5000% 1 0%	Sort is: Bogo/Bozo	-5000%	1		0%
Labels are <b>not</b> left justified -5% 1 0%	Labels are <b>not</b> left justified	-5%	1		0%
Values are <b>not</b> right justified -5% 1 0%	Values are <b>not</b> right justified	-5%	1		0%
Values are not rounded to 3 decimal places -5% 1 0%	Values are not rounded to 3 decimal places	-5%	1		0%
Frequency table code is contained in a function 3% 1 1 3%	Frequency table code is contained in a function	3%	1	1	3%

Histogram code is contained in a function	2%	1	1	2%
Multi-file solution	3%	1	1	3%
Penalties				
Penalties from C & C++ Grading Guide v2.3.0	-5%	1	0	0%
Late submission:	-10%	1	0	0%
Total				100%

## **Submission Requirements**

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

<sup>&</sup>lt;sup>i</sup> Alternatively, you can 'clean' your project for submission by downloading 'vsclean' a Visual Studio Solution Cleaner from <a href="https://www.gats.ca">www.gats.ca</a>.