Part E - Standard Template Library

STL Algorithms

Workshop 8

In this workshop, you use the Standard Template Library's algorithms to evaluate sample data.

Learning Outcomes

Upon successful completion of this workshop, you will have demonstrated the abilities to

- copy data from a file into a sequential container
- accumulate data values using the STL's numeric library
- specify an operation on each value in a data set using a lambda expression
- sort the data values in a data set using the STL's algorithm
- reflect on what you have learned in this workshop

Introduction to Statistical Analysis

Statistical analysis uses standard measures to make predictions based on a small sample of the actual data:

- sample mean - the average of all values in the sample
- sample standard deviation - the spread of the numbers away from their mean
- sample median - the middle number in the sorted set of the values (that is, the value separating the lower and upper halves of the data in a sorted set)

The formula for sample mean is

$$z_{mean} = (\Sigma_i z_i) / n$$

The symbol Σ denotes 'sum of', i refers to an element in the set, and n refers to the number of elements in the set.

The formula for sample standard deviation (ssd) is

$$ssd = \sqrt{ [\Sigma_i (z_i - z_{mean})^2] / (n - 1) }$$

Regression Line

A regression line relates a set of independent (x) values to a corresponding set of dependent (y) values. The number of values in each set is the same in both sets. Each value in the independent set

has one corresponding value in the dependent set.

The regression line best fits the pairs of data values. It is the line that passes through the data points drawn on a two-dimensional (x,y) system of coordinates as close as possible to the data points. The line's coefficients are:

- slope - the slope of the line in the x-y plane
- y_intercept - the y value of the line where it crosses the y-axis

The formulas for these two coefficients are:

$$\text{slope} = [\, n\,(\,\Sigma_i x_i y_i\,) - \Sigma_i x_i\ \Sigma_i y_i\,] \,/\, [\, n\,(\Sigma_i x_i^2) - (\Sigma_i x_i)^2\,]$$

$$\text{y\_intercept} = [\, \Sigma_i y_i - \text{slope} * \Sigma_i x_i\,] \,/\, n$$

You can find a linear regression calculator here.

Specifications

Code a class template named DataTable for performing statistical analysis on data stored in text files.

Main Program

The main() program listed below analyzes data using your DataTable template.

```
// Workshop 8 - STL Algorithms
// w8.cpp

#include <iostream>
#include <iomanip>
#include <fstream>
#include <string>
#include "DataTable.h"

const int FW = 7;
const int ND = 2;

int main(int argc, char** argv) {
    std::cout << "\nCommand Line : ";
    for (int i = 0; i < argc; i++) {
        std::cout << argv[i] << ' ';
    }
    std::cout << std::endl;
    if (argc != 2 && argc != 3) {
        std::cerr << argv[0] << ": incorrect number of arguments\n";
        std::cerr << "Usage: " << argv[0] << " file_name [p]\n";
        return 1;
    }
    std::ifstream dataFile(argv[1]);
    if (!dataFile) {
        std::cerr << "\n***Failed to open file " << argv[1] << "***\n";
        return 2;
    }

    try {
        w8::DataTable<float> data(dataFile, FW, ND);
        if (argc == 3) {
            std::cout << "\nData Values\n==========\n";
            std::cout << data << std::endl;
        }
        float m, c;
        data.regression(m, c);
        std::cout << "\nStatistics\n==========\n";
        std::cout << std::fixed << std::setprecision(ND);
        std::cout << "y mean    = " << std::setw(FW) << data.mean() <<
         "\ny sigma   = " << std::setw(FW) << data.sigma() <<
         "\ny median  = " << std::setw(FW) << data.median() << std::endl;
        std::cout << "slope     = " << std::setw(FW) << m <<
```

```
            "\nintercept = " << std::setw(FW) << c << std::endl;
    }
    catch (std::string& msg) {
        std::cout << msg << std::endl;
        return 3;
    }

    std::cout << "\nPress any key to continue ... ";
    std::cin.get();
}
```

The first command-line argument is the name of the file to analyze. The second argument, if present, requests a listing of all data values retrieved by the object.

Data Files

The data for this workshop is stored in the following three files:

- Simple.dat - test 1
- Flat.dat - test 2
- HS_College_GPA.dat - high school and college GPA comparisons

Each record in each file contains two fields. The first field holds the independent value (x coordinate). The second field holds the dependent value (y coordinate). For example, the file named Simple.dat contains

```
2.1 8
2.5 12
4.0 14
3.6 10
```

DataTable Class Template

Upon instantiation, a DataTable object receives a reference to the file stream that holds the data values, the field width for displaying the data and the number of decimals to display. The object retrieves the data values from the file and stores them in its instance variables.

Your design includes the following queries:

- T mean() const - returns the mean value of the dependent coordinate
- T sigma() const - returns the standard deviation of the dependent coordinates
- T median() const - returns the median value of the dependent coordinate
- void regression(T& slope, T& y_intercept) const - returns the slope and intercept for the data set
- void display(std::ostream&) const - displays the data pairs as shown below

Your design also overloads the insertion operator as a helper for the DataTable class:

- std::ostream& operator<<(std::ostream&, const DataTable&) - inserts the data generated by display() into the output stream

Code your member functions to meet all of the learning outcomes.

Test Runs

The results of your test runs on the data files should be similar to those listed below.

Simple.dat

```
Command Line : w8 Simple.dat p

Data Values
===========
      x      y
   2.10   8.00
   2.50  12.00
   4.00  14.00
```

```
   3.60  10.00


 Statistics
 ==========
 y mean    =   11.00
 y sigma   =    2.58
 y median  =   12.00
 slope     =    1.91
 intercept =    5.18

 Press any key to continue ...
```

Flat.dat

```
 Command Line : w8 Flat.dat p

 Data Values
 ===========
       x      y
    1.00   1.00
    2.00   1.00
    3.00   1.00
    4.00   1.00


 Statistics
 ==========
 y mean    =    1.00
 y sigma   =    0.00
 y median  =    1.00
 slope     =    0.00
 intercept =    1.00

 Press any key to continue ...
```

HS_College_GPA.dat

```
 Command Line : w8 HS_College_GPA.dat

 Statistics
 ==========
 y mean    =    3.12
 y sigma   =    0.51
 y median  =    3.21
 slope     =    0.78
 intercept =    0.73

 Press any key to continue ...
```

Submission

Typescript

On matrix, create a typescript of your complete solution using the following commands:

```
 + At the prompt, type: script w8.txt
 + At the prompt, type: whoami
 + At the prompt, type: cat DataTable.h
 + At the prompt, type: g++ -o w8 w8.cpp
 + At the prompt, type: w8 Simple.dat p
 + At the prompt, type: w8 Flat.dat p
 + At the prompt, type: w8 HS_College_GPA.dat
 + At the prompt, type: exit
```

These commands will produce a file named w8.txt.

Download your typescript file to your local computer. Submit your solution according to your professor's instructions.

Possible Submission Instructions (confirm with your instructor)

MySeneca

- Login to
- Select OOP345 if necessary
- Select Assignments or Workshops
- Select W8
- Press "Browse My Computer" to upload your typescript
- Press "Edit"
- Summarize to your instructor the concepts that you have learned in doing this particular workshop. Add any other comments you wish to make in the comment box provided.
- Press "Submit" IMPORTANT: If you "Save As Draft" your instructor does not receive your submission unitl you press "Submit"

Printer Friendly Version of this Page print this page Top Go Back to the Top of this Page

-
- Home
-
- Timeline

- Notes
-
-
- Workshops
- Assignments
- Instructor