TCP1101- Programming Fundamentals

(Bachelor in Computer Science)

Basic Data Analysis Application

CHENG JIA PAO	**1191101533**
TAN YONG QI	**1191101570**
TANG NING KANG	**1191101645**
NG BEI SHENG	**1191101917**

Lecturer: Bau Yoon Teck

Faculty of Computing & Informatics

Multimedia University

TCP1101 Assignment
Trimester 1, 2020/2021
by <<TAN YONG QI>>
Team Leader: TAN YONG QI, 1191101570@student.mmu.edu.my
Team members:
NG BEI SHENG, 1191101917@student.mmu.edu.my
TANG NING KANG, 1191101645@student.mmu.edu.my
CHENG JIA PAO, 1191101533@student.mmu.edu.my

Table of Contents

TCP1101 Assignment
Trimester 1, 2020/2021
by <<TAN YONG QI>>
Team Leader: TAN YONG QI, 016-7472963, 1191101570@student.mmu.edu.my
Team members:
NG BEI SHENG, 016-5244368, 1191101917@student.mmu.edu.my
TANG NING KANG, 011-31740339, 1191101645@student.mmu.edu.my
CHENG JIA PAO, 011-7524222, 1191101533@student.mmu.edu.my

Acknowledgement

During the development of the project, the cooperation and teamwork amongst the group members who are Cheng Jia Pao, Ng Bei Sheng, Tan Yong Qi and Tang Ning Kang are very much appreciated as these elements are essential to ensure the project runs smoothly.

All group members are thankful to the lecturer, Mr. Bau Yoon Teck for providing guidance and suggestions.

TCP1101 Assignment
Trimester 1, 2020/2021
by <<TAN YONG QI>>
Team Leader: TAN YONG QI, 1191101570@student.mmu.edu.my
Team members:
NG BEI SHENG, 1191101917@student.mmu.edu.my
TANG NING KANG, 1191101645@student.mmu.edu.my
CHENG JIA PAO, 1191101533@student.mmu.edu.my

Part 1: Introduction

1.1 Problem Statement

This project is initiated to solve the problem of computing statistics using given data loaded from a CSV file by creating a simple data analysis program in C++. The program includes functions of computing statistics such as minimum, maximum, mean, median, variance and standard deviation of selected data. It can also compute Pearson's correlation and linear regression line, display histogram and table and finally generate a report in text file and html file.

1.2 Functional Requirement

This section includes the functional requirement of the program. More functions are expected to be created during actual coding to improve the program and this is common in software development.

Table 1.1: Functional Requirement

Function ID	Function Name	Function Explanation
1	showMenu()	Display the main menu.
2	loadCSVFile()	Load the CSV file which contains the marks of three subjects of students.
2.1	findLastComma()	Find the last comma position in a string and return the position as an integer.
2.2	getLastColumn()	Find and return last column before the last comma in a record string.
2.3	constructVectors()	Construct a vector by concatenating data in front of a comma into a counter, push back counter into vector
2.4	checkColumn()	Check the format of column
2.5	checkRow()	Check the format of row
2.6	convertStringToInt()	Convert string data to int data
3	getInput()	Read the user input and process input validation.

4	showSubmenu()	Display the submenu which computes the statistics (min, max, median, mean, variance and standard deviation).
5	storeResult_Txt()	Store all of the six statistics results in text file.
6	getMin()	Compute the minimum of the selected subject.
7	getMax()	Compute the maximum of the selected subject.
8	getMedian()	Compute the median of the selected subject.
8.1	swap()	Swap larger data with smaller data
8.2	sortVectorAscendingly()	Sort the element in the vector ascendingly
9	getMean()	Compute the mean of the selected subject.
10	getVariance()	Compute the variance of the selected subject.
11	getStandardDeviation()	Compute the standard deviation of the selected subject.
12	displayMarkFrequencyDistribution()	Display the distinct numbers with no repetition and the frequency of each number.
13	plotHistogram()	Plot histogram for the data in textual form.
14	displayTable_NumberAboveMean()	Display tables with the data items above the mean and display the number of those data
15	displayTable_NumberBelowMean()	Display tables with the data items below the mean and display the number of those data
16	getPearsonCorrelation()	Compute the Pearson's correlation between any two selected columns and show the table of results.
17	getLinearRegression()	Compute the linear regression line for a selected column or columns of the data.
18	displayStore_Data_Ascendingly()	Display and store data sorted ascendingly using a selected column
19	displayStore_Data_Descendingly()	Display and store data sorted descendingly using a selected column

20	storeReport_Txt()	Store a properly formatted statistical report that includes all the statistical computed values above in a text file
21	storeReport_Html()	Generate a properly formatted statistical report that includes all the statistical computed values above in a html file

1.3 Menu System: Design and Motivation

A user-friendly and easy to navigate menu system is important for good user experience. The menu system in this program is designed to fit the requirements.

Firstly, the user is asked to enter the file name in order to load the desired CSV file. If the file format is correct, then the user can proceed to further option. If the user chooses to compute statistic, he can proceed to view by subject, view all subjects report or compare subjects. The user can choose any three subjects at once to compute the minimum, maximum, mean, median, variance, standard deviation, display marks frequency distribution, plot histogram or display marks above and below mean. If view all subjects is chosen, the minimum, maximum, mean, median, variance, standard deviation of every subjects will be displayed. The user can compute the linear regression or Pearson's Correlation if he chooses compare subjects.

The user can choose to generate HTML report if any statistics is computed in the first place. Exit option is available if the user wishes to leave the program. Please refer to hierarchy chart for visual presentation of menu design.

1.4 Hierarchy Chart

This section includes the hierarchy chart which presents the menu design of the data analysis program.

Team Leader: TAN YONG QI, 1191101570@student.mmu.edu.my
Team members:
NG BEI SHENG, 1191101917@student.mmu.edu.my
TANG NING KANG, 1191101645@student.mmu.edu.my
CHENG JIA PAO, 1191101533@student.mmu.edu.my

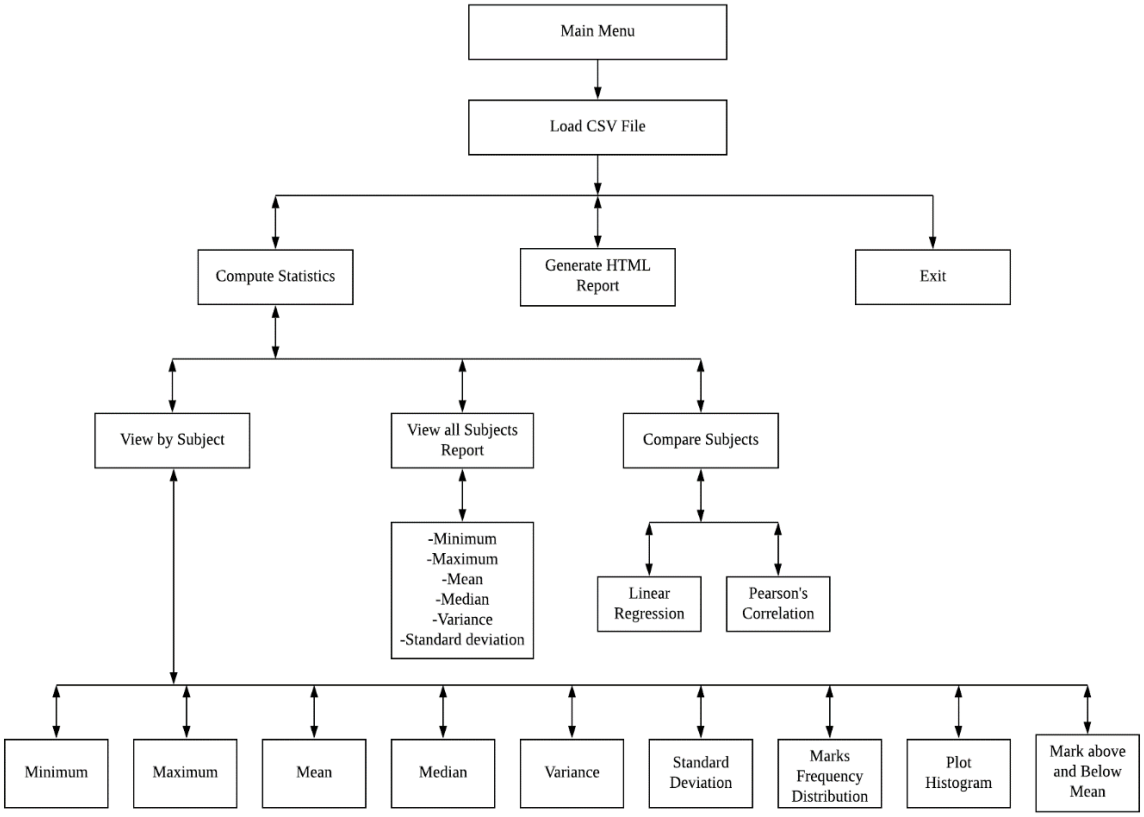

Figure 1.1: Hierarchy Chart

1.5 Flowchart and Pseudocode

This section includes all major flowcharts and its pseudocode used in the program. The actual coding might have difference from the design as it is to improve the program and fix the flaws found in the design.