IEE 380 Team Project **Fall 2017**

Austin Bren

Team # 11

Name Team Member 1: Hunter Hardwick	PIN: 119	Major: Computer Science
Name Team Member 2: Yifan Li	PIN: 158	Major: Computer Science
Name Team Member 3: Raffi Shahbazian	PIN: 217	Major: Computer Science
Name Team Member 4: Nikola Uzelac	PIN: 236	Major: Computer Science

Description

This experiment was conducted to determine if sorting times were affected by data types in the Java programming language. The treatments are the 5 specific data types: Integer, Long, Float, Double, and String. The blocks are the five selected sorting algorithms: bubble sort, insertion sort, merge sort, quicksort, and selection sort. The null hypothesis states that within each block, the treatments, regardless of the data type, will have the same mean sorting time. The alternative hypothesis states that at least one of the treatments had a mean sorting time that differed from the others.

H_o: $\mu_I = \mu_L = \mu_F = \mu_D = \mu_S$

H_1: $\mu_i \neq \mu_j$ for one or more pair i,j

Data and Excel Output

The data table below indicates for each data type, the time in nanoseconds it took to sort an array of size 1000 using a specific sorting algorithm. Ex: Integers take 5,218,000ns to bubble sort, 2,862,000ns to insertion sort, and 780,000ns to merge sort.

Sorting Algo	Integer	Long	Float	Double	String
Bubble (ns)	5,218,000	5,890,000	5,658,000	5,305,000	11,409,000
Insertion (ns)	2,862,000	3,134,000	2,860,000	3,460,000	4,808,000
Merge (ns)	780,000	940,000	1,139,000	877,000	1,056,000
Quick (ns)	1,368,000	999,000	659,000	747,000	440,000
Selection (ns)	3,052,000	3,684,000	4,129,000	3,379,000	7,742,000

The table below is the output from a Two-Factor ANOVA data analysis test conducted on the model using Excel.

Anova: Two-Factor Without Replication						
SUMMARY	*Count*	*Sum*	*Average*	*Variance*		
Bubble	5	33480000	6696000	7.01E+12		
Insertion	5	17124000	3424800	6.59E+11		
Merge	5	4792000	958400	2.02E+10		
Quick	5	4213000	842600	1.26E+11		
Selection	5	21986000	4397200	3.65E+12		
Integer	5	13280000	2656000	2.99E+12		
Long	5	14647000	2929400	4.27E+12		
Float	5	14445000	2889000	4.31E+12		
Double	5	13768000	2753600	3.74E+12		
String	5	25455000	5091000	2.12E+13		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	1.21E+14	4	3.03E+13	19.57764	5.24E-06	3.006917
Columns	2.11E+13	4	5.28E+12	3.404995	0.033915412	3.006917
Error	2.48E+13	16	1.55E+12			
Total	1.67E+14	24				

Hypotheses Tests and P-Values

With a P-Value of .0339, at $\alpha = .05$ we will reject the null hypothesis. This concludes that at least one of the data types mean sorting times differs from any other.

Applicability

The primary audience which this data would appeal to is programmers and mathematicians. Understanding the speed of a specific sorting algorithm depending on the data type of the input is a powerful tool for a programmer. For example, given an an array of strings that could be converted to an integer ex) [“1243”, “3543”, “7234”, “8542”...], it may be advantageous to convert the result to an integer array before sorting.