Project 2

1 Problem Statement

Quickselect Algorithm- Median of Medians Method

2 Theoretical Analysis

It is a method for finding the k-th smallest (or largest) member in an unsorted array. When used with the median-of-medians pivot selection method, it ensures worst-case linear time complexity, which means that the time required to identify the k-th element is proportional to the size of the input array (n). This is a substantial benefit over algorithms like sorting-based approaches, which have a worst-case time complexity of O(n log n).
As a result, Quickselect using the median-of-medians approach is O(n), not O(n log n) in terms of asymptotic analysis. Quickselect's asymptotic time complexity with the median-of-medians pivot selection approach is O(n), not O(n log n).

The given equation **T(n) = cn + T(n/5) + T(7n/10)** describes a recursive algorithm's time complexity, where T(n) represents the time it takes to process a problem of size n.
The substitution method can be used to prove that the equation

T(n/5): This represents the time it takes to process a subproblem of size n/5.
T(7n/10): This represents the time it takes to process another subproblem of size 7n/10.

$T(n/5) \leq 2cn$
T(n/5) is bounded by 2cn. This means that the time to process a subproblem of size n/5 is at most 2 times c times (n/5).

$T(7n/10) \leq 7cn$
T(7n/10) is bounded by 7cn. This means that the time to process a subproblem of size 7n/10 is at most 7 times c times (7n/10).

T(n) = cn + T(n/5) + T(7n/10) has a linear time complexity, particularly O(n). We compute with an inductive hypothesis that $T(m) \leq 10\ cm$ for all values of $m < n$. As a result, we can conclude:
$T(n/5) \leq 2cn$
$T(7n/10) \leq 7cn$
$T(n) \leq cn + 2cn + 7cn = 10\ cn$

T(n) is O(n) because it is bounded above by a linear function of n.

3 Experimental Analysis

3.1 Program Listing

Code: https://github.com/AdoniaSequeira/DAA_EXP2

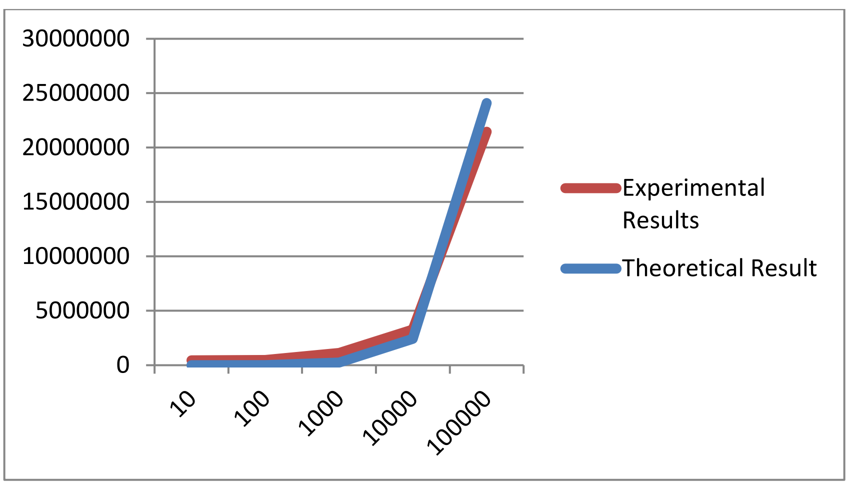
Data Normalization Notes

Do you normalize the values by some constant? How did you derive that constant?
I used a scaling constant in this case. This is calculated by taking the averages of the two columns of the tables.

3.2 Output Numerical Data

n	Experimental Result, in ns	Theoretical Result	Scaling Constant	Adjusted Theoretical Result
10	449680	10		2409.69922
100	481420	100		24096.9922
1000	1115040	1000		240969.922
10000	3268649	10000		2409699.22
100000	21459379	100000		24096992.2
	5354833.6	22222	240.969922	

3.3 Graph



3.4 Graph Observations

Hence a slight difference is observed in the theoretical and experimental values

4 Conclusions

Hence a difference is observed in the theoretical and experimental values