

# Math 343 - Project Proposal

---

Preston Duffield

Western Washington University

May 23, 2023

## Project Description

The experiment we are conducting aims to investigate the influence of various factors on the performance of a C program that reads a text file and performs a word frequency count operation. We are interested in the time efficiency of the operation, which will be measured as the response variable. This variable, the time taken to complete the algorithm in milliseconds, will be recorded for each run of the experiment.

### Factors

1. **Buffer Size** This is the amount of data the program will read from the file at once. The levels for this factor are 512 bytes and 4096 bytes.
2. **Algorithm** This is the specific approach used to perform the word frequency count. We have two levels for this factor, which are the Hash Map approach and the Sorting approach.
3. **CPU Threads** This factor corresponds to the threading strategy used by the program. The two levels for this factor are Single Threaded and Multi-Threaded.

Given that we have 3 factors each with 2 levels, we have a total of  $2^3 = 8$  treatment combinations. We plan to collect a sample size of 2 for every treatment combination. This means that we will run each combination of factor levels twice, resulting in a total of 16 runs for the experiment.

### Influence

We hypothesize that the response variable, the time taken to complete the algorithm, will be influenced by our chosen factors.

The buffer size might influence the time efficiency because larger buffer sizes allow the program to read more data at once, potentially speeding up the process, but they also require more memory, which could slow down the process if memory becomes scarce.

The algorithm type also might affect the response because different algorithms have different computational complexities. The Hash Map approach typically offers quicker insert and lookup operations, but the actual performance may depend on factors such as the implementation of the hash function and the handling of hash collisions.

The CPU threading strategy could influence the time efficiency because multi-threading could potentially allow the program to perform multiple operations simultaneously, speeding up the process. However, this depends on the nature of the operation and how well it can be parallelized.

Given these considerations, we believe that examining the effects of these factors and their interaction will provide valuable insights into the performance characteristics of the C program.