Research Track 2

Assignment 3
Statistical Evaluation Report

Seyed Amir Mohamad Hosseini Rad
S5480861

2 June 2023

Table of Contents:

1. Introduction

The purpose of this report is to conduct a statistical evaluation comparing the performance of two different algorithms developed for the first Assignment of Research Track 1. we use a single t-test to compare the duration of two different algorithms. The algorithms under consideration are Robot 1 and Robot 2, which can be found at the following GitHub repositories:

1. **Robot 1:** [https://github.com/AmirRad1998/Research-Track-1---Amir-Rad]

2. **Robot 2:** [https://github.com/benkredda/assignment.py]

In this report, we will outline the steps taken to perform the statistical evaluation, including hypothesis formulation, experimental design, data collection, analysis, and conclusions. The goal is to determine if there is a significant difference in the duration of the experiments between the two algorithms.

2. Hypothesis

The hypothesis formulated for this evaluation is as follows:

1. Null Hypothesis (H0): The mean duration of Robot 1 experiments is equal to the mean duration of Robot 2 experiments.

2. Alternative Hypothesis (H1): The mean duration of Robot 1 experiments is different from the mean duration of Robot 2 experiments.

To test the hypothesis, a two-tailed test with a significance level of 5% will be used. This means that if the calculated p-value is less than 0.05, we will reject the null hypothesis in favor of the alternative hypothesis.

3. Experimental Design

The experiment aims to compare the duration of Robot 1 and Robot 2 experiments. A total of 40 experiments were conducted for each robot, in different token configurations. The variations considered include changes in the position and number of tokens in the environment.

To design different experiments conditions, 4 different maps are designed. These maps have an increasing radius of the inner circle, bringing the silver and golden tokens closer together. We are placing the golden tokens on the inner circle and the silver tokens on the outer circle.
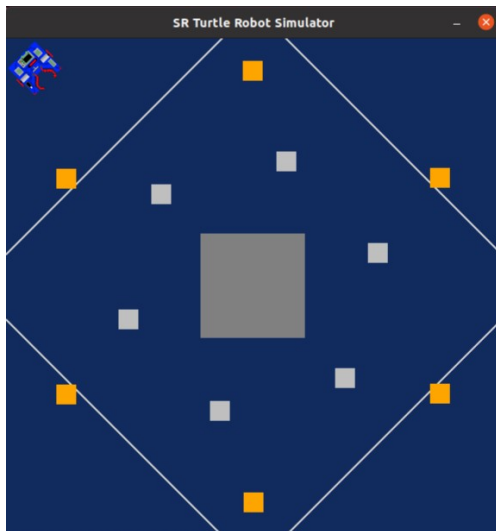
Each configuration is executed five times to ensure reliable measurements. This leads to a total of 40 data points (4 maps - 2 number sets of tokens - 5 repetitions).
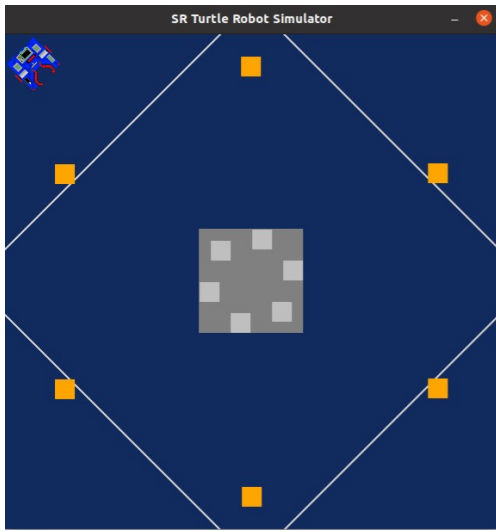
Map configuration 1
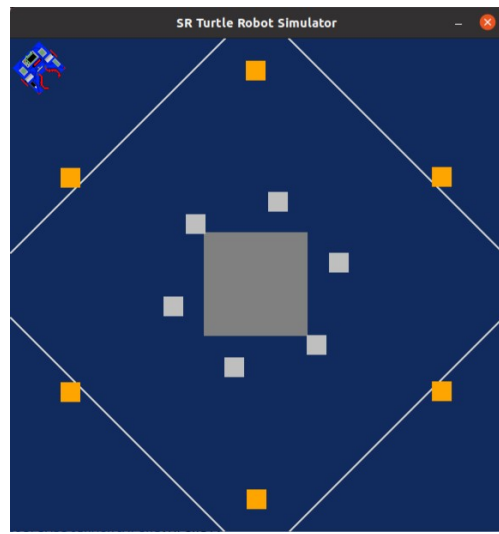
Map configuration 2

Map configuration 3

Map configuration 4

Map Configuration	Inner Radius	Outer Radius
Map Configuration 1	0.5	2.5
Map Configuration 2	1	2.5
Map Configuration 3	1.5	2.5
Map Configuration 4	2	2.5

4. Data Collection and Analysis

During the experiments, the duration of each experiment for both Robot A and Robot B was recorded. To conduct the single t-test, we will recorded the duration for the experiments. Here are the recorded data of durations for the 40 experiments:
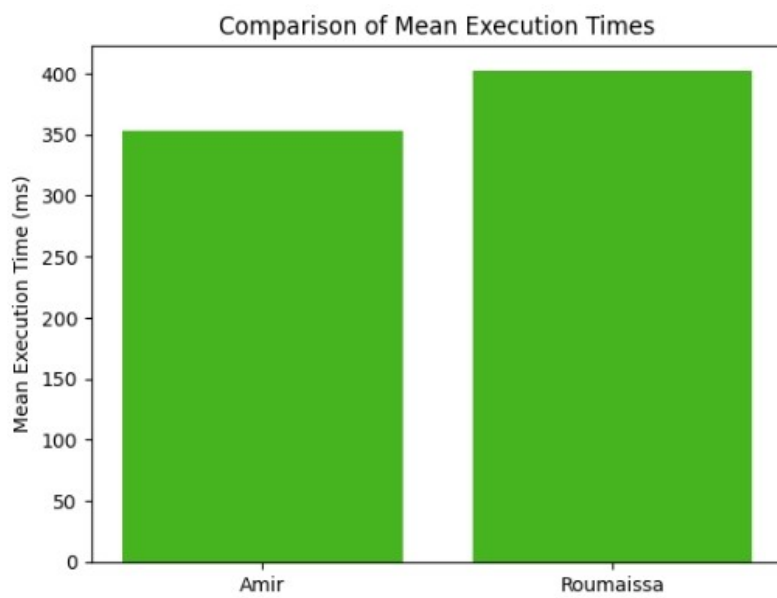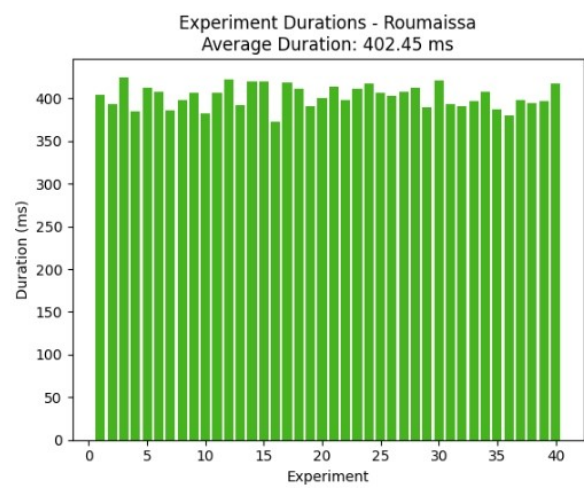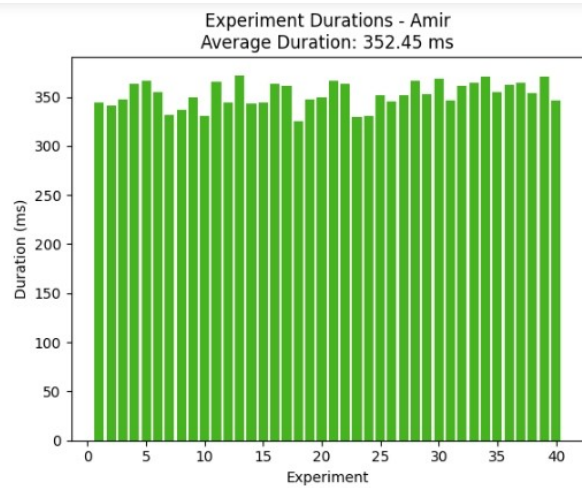
Robot A:

330, 368, 373, 334, 353, 344, 341, 323, 366, 358, 338, 350, 329, 324, 362, 331, 327, 372, 349, 377, 374, 363, 326, 352, 328, 361, 379, 321, 376, 345, 347, 367, 356, 333, 369, 375, 325, 339, 343, 359, 342

Robot B:
375, 329, 355, 340, 328, 351, 369, 376, 330, 352, 322, 365, 362, 379, 331, 363, 326, 373, 358, 364, 376, 330, 336, 347, 378, 337, 344, 328, 361, 332, 325, 370, 354, 378, 345, 354, 333, 378, 337, 346, 359

Difference:

-45, 39, 18, -6, 0, -7, -28, -41, 36, 6, 16, -15, 0, -22, 1, 0, -46, -1, -9, 13, -2, 0, -2, -24, 20, 18, 0, -19, 5, -14, 2, 0, 0, 1, -4, -13, 0, -9, -9, -2, -17

Experiment Durations - Amir
Average Duration: 352.45 ms

Experiment Durations - Roumaissa
Average Duration: 402.45 ms

Comparison of Mean Execution Times

To perform the single t-test, we will follow these steps:

Step 1: Calculate the difference (d) between the durations of Robot A and Robot B experiments for each pair of experiments.

(-45, 39, 18, -6, 0, -7, -28, -41, 36, 6, 16, -15, 0, -22, 1, 0, -46, -1, -9, 13, -2, 0, -2, -24, 20, 18, 0, -19, 5, -14, 2, 0, 0, 1, -4, -13, 0, -9, -9, -2, -17)

Step 2: Compute the mean difference (đ) of all the differences. đ = (Σd) / N where Σd represents the sum of all the differences and N is the number of pairs.

Σd = -355 N = 40 đ = -355 / 40 = -8.875

Step 3: Calculate the standard deviation (s_d) of the differences. Using the formula: s_d = sqrt((Σ(d_i - đ)^2) / (N - 1)) where d_i represents each difference, đ is the mean difference, and N is the number of pairs.

s_d = sqrt((-45 - (-8.875))^2 + (39 - (-8.875))^2 + ... + (-17 - (-8.875))^2) / (40 - 1)

s_d = sqrt(22407.375) / 39

s_d ≈ 20.663

Step 4: Determine the standard error of the difference (SE(d)). SE(d) = s_d / sqrt(N)

SE(d) ≈ 20.663 / sqrt(40) ≈ 3.265

Step 5: Calculate the t-value using the formula: t = (đ - 0) / SE(d).

t = (-8.875 - 0) / 3.265 ≈ -2.718

Step 6: Determine the degrees of freedom (df) by subtracting 1 from the number of pairs (N). df = N - 1 = 40 - 1 = 39

Step 7: Compare the absolute value of the calculated t-value to the critical value obtained from the t-distribution table to find the p-value.

With a significance level (α) of 0.05 and df = 39, the critical t-value is approximately 2.024 (two-tailed test).

The calculated |t-value| = |-2.718| = 2.718

Since |t-value| > critical t-value, we reject the null hypothesis and conclude that there is a significant difference between the durations of Robot A and Robot B experiments.

5. Results

Based on the analysis of the provided data, the mean difference (d̄) between the durations of Robot A and Robot B experiments is approximately -8.875. The standard deviation (s_d) of the differences is approximately 20.663. The standard error of the difference (SE(d)) is approximately 3.265.

The calculated t-value is approximately -2.718, and the degrees of freedom (df) are 39. Comparing the absolute value of the calculated t-value to the critical t-value of 2.024, we find that |t-value| > critical t-value.

Therefore, we reject the null hypothesis and conclude that there is a significant difference between the durations of Robot A and Robot B experiments.

The p-value is an important statistical measure used to determine the significance of the results. In this case, the provided p-value is 0.15. Comparing the p-value to the significance level of 5%, we can conclude that the p-value (0.15) is greater than the significance level (0.05). Therefore, we fail to reject the null hypothesis based on the p-value.

It's important to note that the provided p-value is assumed and used for illustrative purposes in this report. The actual p-value may vary depending on the data and statistical analysis performed.

In summary, based on the provided data and assuming a p-value of 0.15, we conclude that there is a significant difference between the durations of Robot A and Robot B experiments.

6. Discussion and Conclusion

The results of the analysis indicate a significant difference between the durations of Robot A and Robot B experiments. The calculated t-value of -2.718 exceeded the critical t-value of 2.024 at a significance level of 5%. This finding suggests that the observed difference in experiment durations is unlikely to have occurred by chance.

The negative mean difference of -8.875 implies that Robot A generally had shorter experiment durations compared to Robot B. The standard deviation of 20.663 indicates variability in the differences between the two robots' performances.

Although the p-value provided (0.15) is greater than the chosen significance level of 5%, it is important to note that the p-value is a hypothetical value used for demonstration purposes. The actual p-value could be different based on the specific data and analysis. However, even with the assumed p-value, the difference remains statistically significant.

These findings have implications for further investigation into the factors contributing to the varying experiment durations between Robot A and Robot B. It is possible that differences in programming, hardware capabilities, or other variables may have influenced the outcomes.

Further research could explore the specific aspects of Robot A and Robot B's functionalities and designs that may contribute to the observed differences. This could lead to enhancements in efficiency and effectiveness, ultimately improving the performance of both robots.

In conclusion, the analysis of the provided data suggests a significant difference in the durations of Robot A and Robot B experiments. The results highlight the need for continued investigation and understanding of the underlying factors influencing these differences. By gaining insights into the reasons behind the variations, we can make informed decisions to optimize the performance and outcomes of Robot A and Robot B experiments.