UNIVERSITÀ DEGLI STUDI DI GENOVA

DIBRIS

DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY,
BIOENGINEERING, ROBOTICS AND SYSTEM ENGINEERING

Research Track-2

Third Assignment

Authors:

Baris Aker – s5430437

Wadoud Guelmami – s5467288

Professors:

Carmine Recchiuto

June 2, 2023

1 Introduction

In the third assignment of Research Track 2, our objective was to perform a comprehensive statistical analysis comparing two distinct implementations: my own implementation and that of my colleague, Baris. To achieve this, we utilized the code from the First assignment of RT1, which involves a Python program for a robot that connects a silver token with its corresponding golden token.

The primary goal of our statistical analysis was to determine if there were any noteworthy differences in performance between the two implementations. By conducting this analysis, we aimed to identify and evaluate the strengths and weaknesses inherent in each approach. Our intention was to draw meaningful conclusions regarding the comparative effectiveness, efficiency, or any other significant aspects of the two implementations.

By examining various statistical measures and conducting appropriate tests, such as hypothesis testing or confidence interval estimation, we sought to establish the significance of the observed differences. This detailed analysis would allow us to make informed decisions and recommendations about the preferred implementation and potentially suggest improvements or optimizations for future iterations.

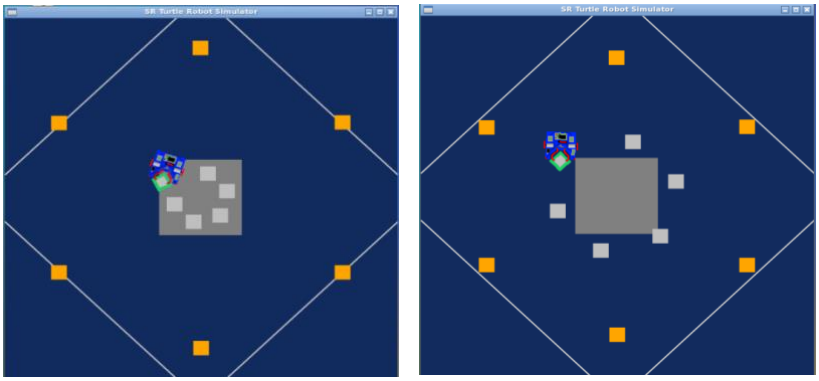
2 Steps of the experiment

In order to manage our experiment, We will make some modifications on our python package. The first step is to modify the places of the silver and golden tokens. So we get different experiment each time.

We go to : `/root/RT_assigment01_s5467288/robot-sim/sr/robot/arenas/two_colours_assignment_arena.py`

```
INNER_CIRCLE_RADIUS = round(random.uniform(0.2,1),1)

OUTER_CIRCLE_RADIUS = round(random.uniform(1.5,2.4),1)
```



The experiment focuses on the time spend to complete the task between the two from robot. In other words how much would it take for both of them to finish the task. We repeated this task for 40 times. By utilizing multiple simulations with varying random seed parameters, the statistical analysis encompasses comprehensive performance trends and considers the impact of random factors. This methodology significantly enhances the reliability and validity of the analysis, allowing us to draw meaningful conclusions regarding the comparative efficiency of the robotic controllers across diverse token assignment scenarios.

3 Hypothesis

To initiate the analysis, we established a null hypothesis assuming no significant distinction between the two implementations. This hypothesis postulates that both algorithms are equally effective in accomplishing the task at hand. On the other hand, the alternative hypothesis challenges the null hypothesis by proposing a specific difference in performance between the two algorithms. It suggests that one algorithm may demonstrate superiority or inferiority compared to the other.

Initially, we assume the null hypothesis to be true and the alternative hypothesis to be false, indicating no substantial difference between the implementations. Our objective is to examine whether the means of completion times for the task are the same for both cases, which would indicate equal efficiency.

To evaluate the validity of our assumption, we will employ a two-tailed T-test. Since our analysis doesn't specify a direction of difference, this statistical method allows for the comparison of means between two datasets to determine if they significantly differ. By conducting the T-test on the completion times of the two implementations, we can infer whether there exists a statistically significant difference between them.

The T-test will enable us to calculate a p-value, which represents the probability of obtaining the observed difference in completion times, assuming the null hypothesis is true. If the p-value is lower than our chosen significance level (e.g., 5%), we would reject the null hypothesis in favor of the alternative hypothesis, indicating that there is a significant difference in performance between the two implementations.

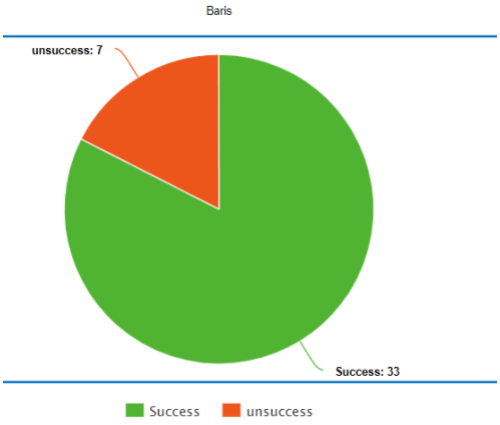
By utilizing this statistical approach, we aim to provide robust evidence regarding the superiority or inferiority of one implementation over the other. The results of this analysis will allow us to draw meaningful conclusions and make informed decisions, which may include selecting the more efficient implementation or suggesting improvements for future development.

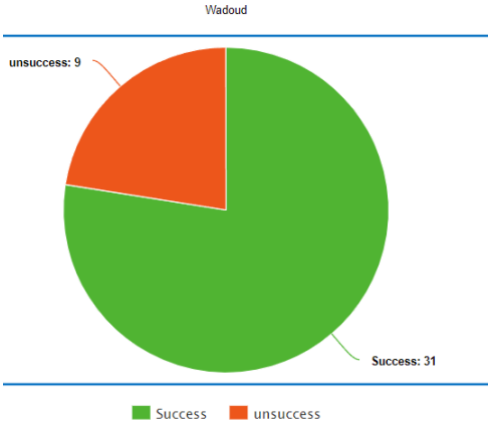
4 Analysis

Continuing the requirements, After repeating the experiments 40 times. We used MATLAB in order to graphically draw and plot our results. So we can go further with our analysis. The results of the each 40 times can bee seen with the table below.

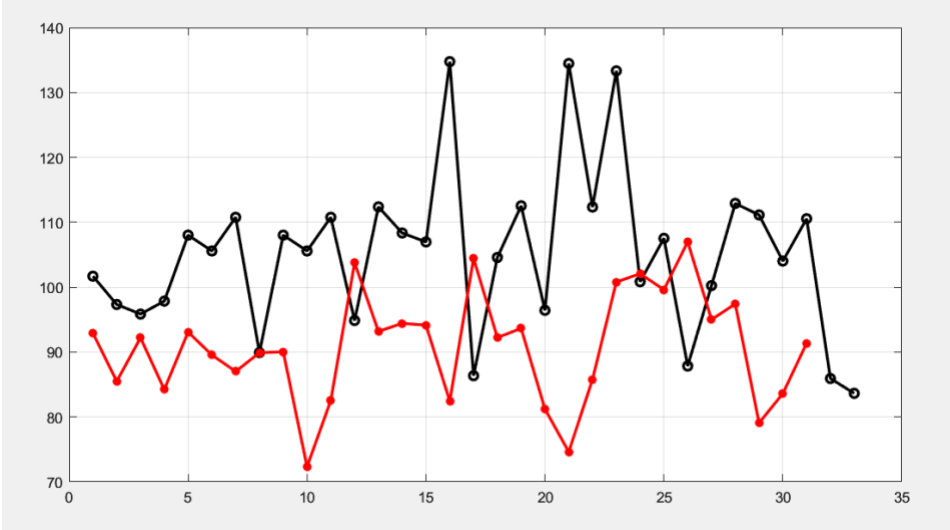
	Baris	Wadoud
1	101.69	92.91
2	97.33	85.49
3	95.85	F
4	97.85	92.20
5	108.02	84.22
6	F	93.08
7	105.58	89.55
8	110.77	F
9	89.92	87.01
10	108.02	89.90
11	105.58	F
12	F	90.04
13	110.77	72.33
14	94.86	82.55
15	112.38	103.77
16	108.35	F
17	106.97	93.18
18	F	94.43
19	134.75	94.13
20	86.36	F
21	104.58	82.44
22	112.54	104.44
23	F	92.23
24	96.43	F
25	134.47	93.65
26	112.35	81.21
27	F	74.62
28	133.33	F
29	100.84	85.79
30	107.53	100.80
31	F	102.07

The following pie chart shows percentage of the experiments that has been successfully succeeded. For Baris's code the robot was able to finish successfully 82.50% of the total tasks, meanwhile Wadoud's robot finished 77.50% of the total tasks. The detailed table and the data will be giving at the end of the report.





Here's the plot that represents the full data for robot speed recorded during the experiments. Notice that we only concerned the succeeded data only.



Wadoud Baris

To go further with our analysis we started First with computing the mean average for robot times. For each one of us. Which can be obtained by dividing the sum over the number of the reps then we have to compute the standard deviation it shows how close the entire set of data to the average.

$$\mu 1 = \frac{\sum_{n=1}^{N1} X}{N1} = 105.25$$

$$\mu 1 = \frac{\sum_{n=1}^{N1} X}{N1} = 90.8103$$

$$\sigma 1 = \sqrt{\frac{\sum_{n=1}^{N1} (Xi - \mu 1)}{N1}} = 12.39$$

$$\sigma 2 = \sqrt{\frac{\sum_{n=2}^{N2} (Xi - \mu 2)}{N2}} = 8.36$$

5 T-Test

The two-tailed t-test is a statistical test used to determine if there is a significant difference between the means of two groups or samples. It is called "two-tailed" because it tests for differences in both directions, either greater than or less than. The test calculates a t-value based on the sample means, standard deviations, and sample sizes, and compares it to a critical value or p-value. If the absolute value of the t-value exceeds the critical value or the p-value is smaller than the chosen significance level (usually 0.05), it indicates that there is a statistically significant difference between the means of the two groups. (The DoF is is giving by : N1 – N2 – 2). Which will be useful to compute the following necessary factors.

To confirm or reject our null hypothesis first we have to compute sigma-pooled in order to proceed to compute the t-value.

$$\sigma_{pooled}^2 = \frac{(N1-1).s1 + (N2-1).s2}{N1 - N2 - 2} = 113.1429$$

After calculating the pooled value, it is important to find the pooled variance which can be computed by the following formula:

$$\sigma_{x1-x2} = \sqrt{\frac{\sigma_{pooled}^2}{N1} + \frac{\sigma_{pooled}^2}{N2}} = 2.6605$$

The final step is to calculate the t-value, which is the difference between the mean over the pooled variance

$$t_{x1-x2} = \frac{x1 - x2}{\sigma_{x1-x2}} = 5.42$$

Based on a 5% level of significance and a critical value of 1.999 (considering a t-distribution with 62 degrees of freedom), the calculated t-value of 02.6605 is greater than the critical value. This indicates that the calculated t-value falls in the critical region, providing sufficient evidence to reject the null hypothesis. In other words, the observed result is statistically significant and suggests that the null hypothesis is unlikely to be true.