
# Research Track 2 - Data Analysis

**Author Name:** Subhransu Sourav Priyadarshan, ROBOTICS ENGG

**Prof.:** Carmine Recchiuto

**Matricola:** 5302091

# Contents

**1 INTRODUCTION** **1**

**2 TOOLS FOR ANALYSIS AND THE ARENA DESCRIPTION** **1**
2.1 Data Collection . . . . . 2

**3 RESULTS** **2**
3.1 Hypothesis . . . . . 2
3.2 Mean Distance Data . . . . . 2
3.3 Average time elapse . . . . . 3
3.4 Comparison of Lap times . . . . . 4
3.5 T-Test . . . . . 5

**4 Conclusion and Comparison** **6**

# 1 INTRODUCTION

The final part of the Research Track 2 assignment deals with **data analysis**. We needed to perform *statistical analysis* on the first assignment of the Research Track 1 course, choosing some parameters in order to compare my implementation with an already available solution. My own implementation was turned in at the end of the project and published on GitHub at the following repository: `https://github.com/subhransu10/RT1_Assignment1`. The Professor's implementation is cloned at the following repository: `https://github.com/CarmineD8/python_simulator/tree/rt2`. The aim of this assignment was to code a python script capable of making a robot sequentially grasp and released some targets that were silver tokens, inside an arena composed of golden tokens which was the wall of the arena (see the figure below).

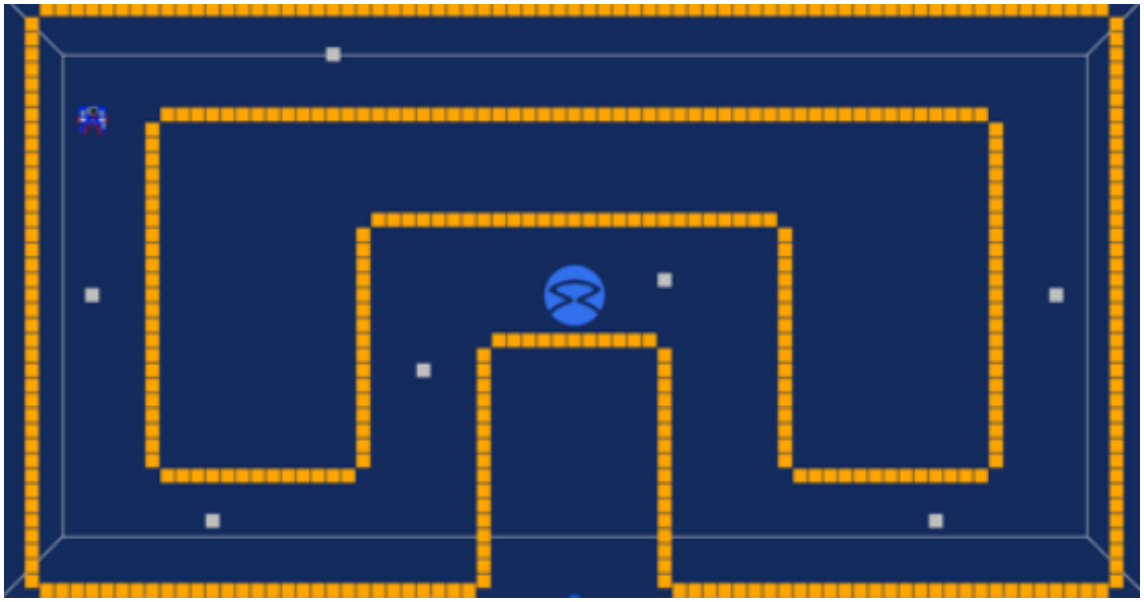


**Figure 1:** Arena

# 2 TOOLS FOR ANALYSIS AND THE ARENA DESCRIPTION

I created the code in MATLAB 2022a to do a good and trustworthy statistical analysis. This code provided all of the necessary data for acceptable statistical testing and the construction of eye-catching visualizations. Instead of having the silver tokens' positions determined by chance, I decided to employ four distinct arena configurations, each with a different number of tokens. I specifically created four arenas:

1. 6 tokens
2. 7 tokens (standard arena)
3. 8 tokens
4. 9 tokens

The tokens' positions are always fixed, so each time I changed the configuration, I simply inserted a new token while maintaining the positions of the others.

## 2.1 Data Collection

As explained in the above section, we compare both projects and the data collected regarding the following:

1. **Elapsed time:** the amount of time it takes the robot to complete a lap of the circuit using a control loop in the assignment code, depending on the environment.

2. **Distances:** The distances between the robot and the walls were represented by golden coins.

The robot's distance from golden tokens was calculated at each iteration of the control cycle while taking into account how many circuit laps it had completed. I also calculated the robot's lap time five times, indicating that the robot completes five laps. To be more precise, the robot's lap time starts when it collects the first silver token and ends when it collects the last one. Distances and times for each arena arrangement were measured.

# 3 RESULTS

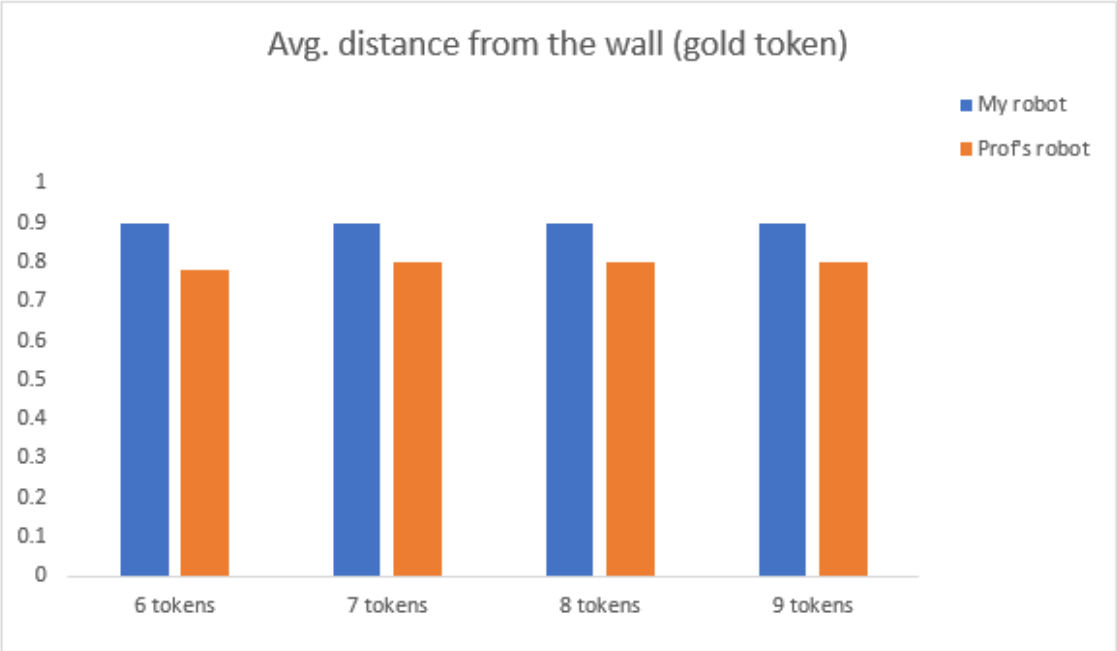
## 3.1 Hypothesis

A hypothesis is one of the strongest tools for statistical analysis. Although it might not be proven absolutely if it has passed critical testing then it can be accepted.

We compare two projects (as mentioned already) in terms of superiority and then proceed assuming that both approaches are equally good. This is known as *null hypothesis*.In contrast, we can think that our approach is better and Professor's approach is not or vice-versa, in which case we are putting out a different theory. If our sample data reject this null hypothesis, we should draw another conclusion. We arrive at the *alternative hypothesis* after rejecting the null hypothesis. The collection of alternatives to the null hypothesis is referred to as the alternative hypothesis in other words. If we accept one approach we reject the other.

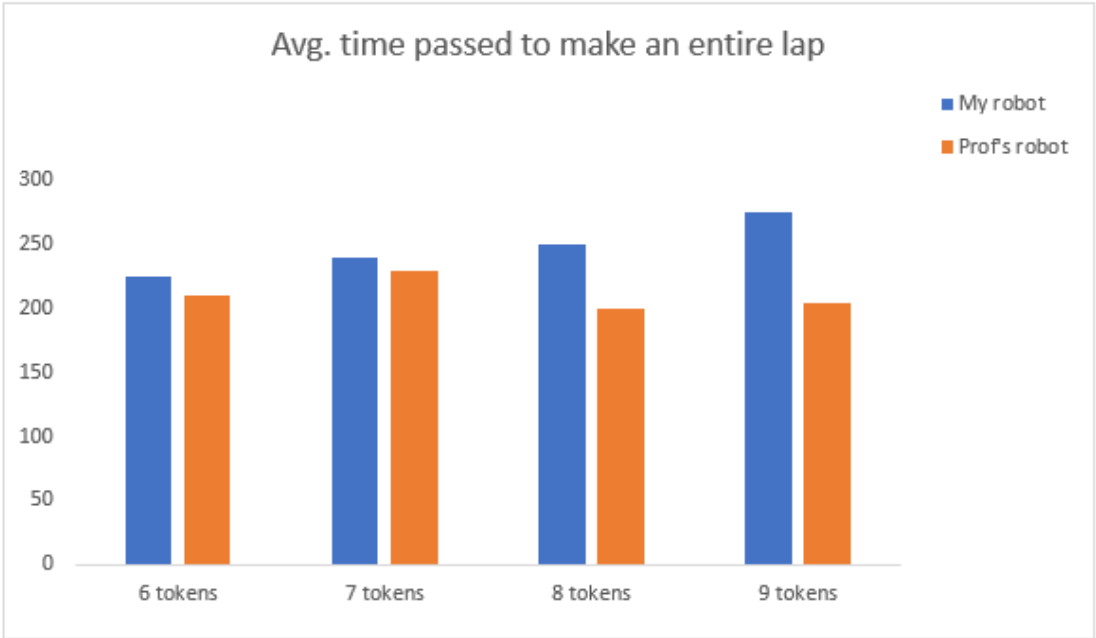
## 3.2 Mean Distance Data

After gathering all of the distance data, all of the averages for each configuration of the arena circuit were calculated. I decided to compare the average distance from the walls for each arrangement and present the results with a graph (see the figure below). It is evident that in each configuration, my robot's distance from the walls is on average greater than the offered answer, implying that my robot stays further away from the walls.




## 3.3 Average time elapse

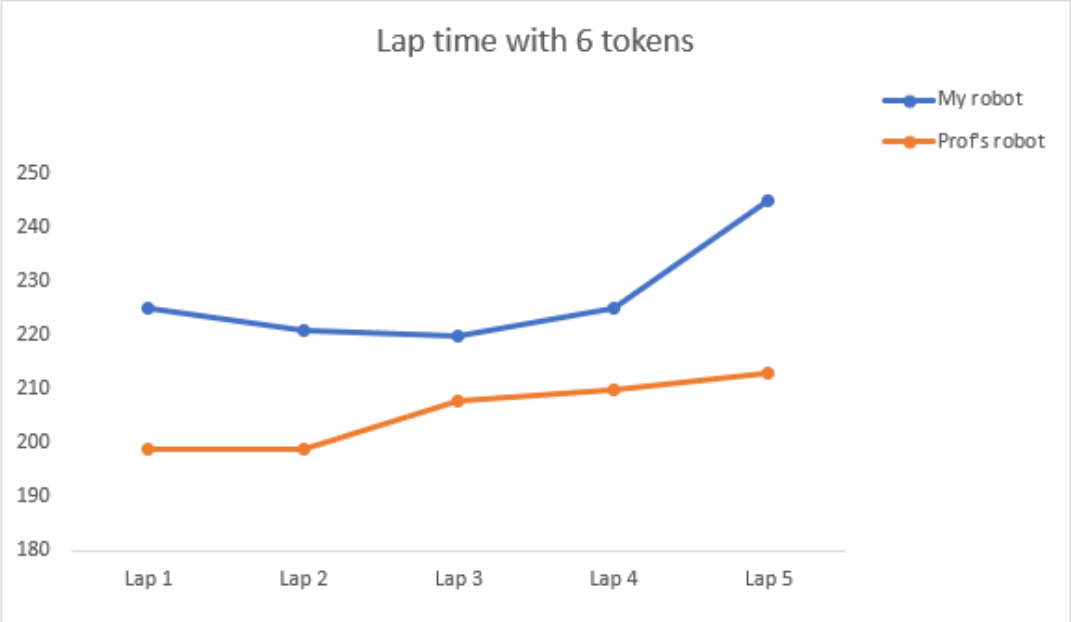
The comparison is done and is plotted in a line graph and the results are analyzed.

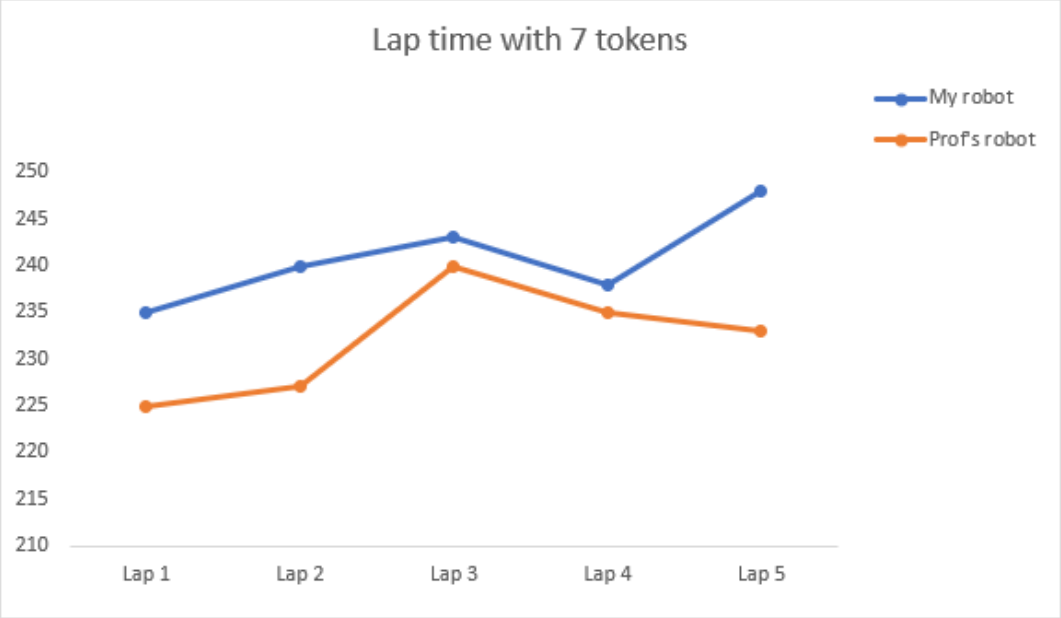



The graph clearly highlights that my robot is always slower than the one provided. In particular, while the time is similar in the 6-token and 7-token configuration, my time gets higher and higher as run with the 9-token and 10-token configuration. So I can strongly conclude that my robot is slower than one of the given solutions by the Professor.

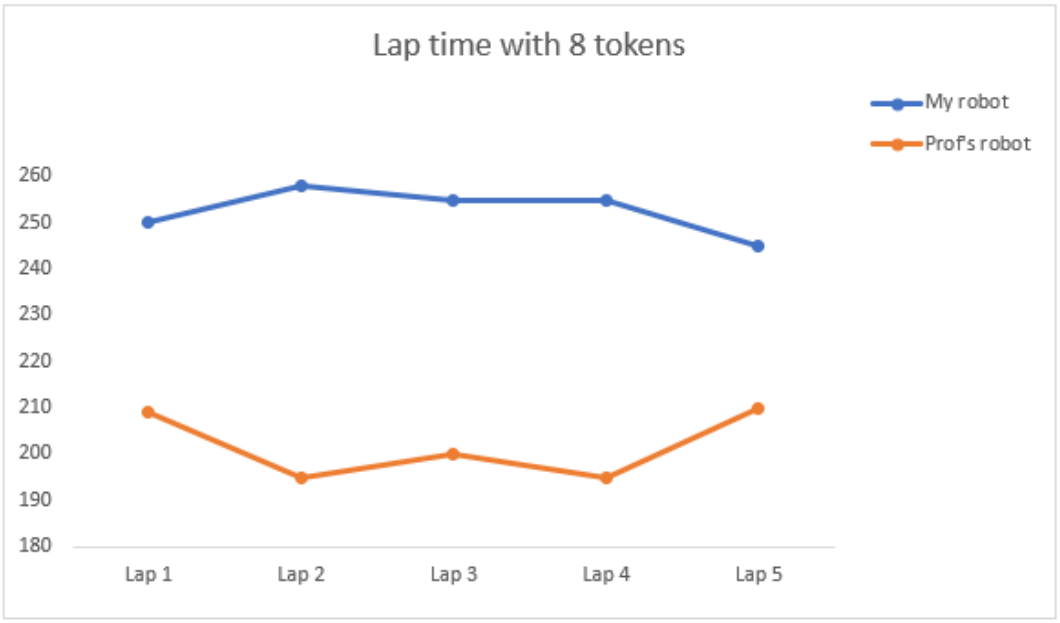
## 3.4 Comparison of Lap times

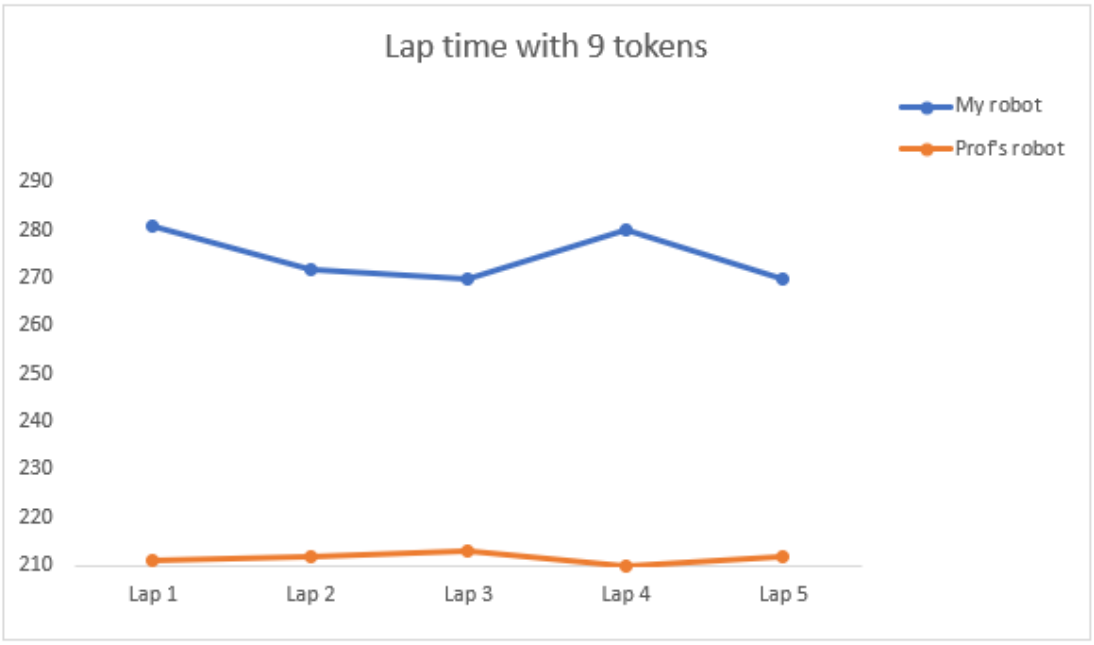
The data is plotted using a *line graph* while comparing with the given solution. We will also briefly discuss the observations below along with with the graphs.




If we see the two graphs above, we can notice that the two lines (in both graphs) are close which means that lap times are almost similar.




In the above two graphs, we notice that the difference is much between the two lines. We can conclude after seeing the results of all four cases that my robot is **comparatively slower** than the given solution.

## 3.5 T-Test

A t-test is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether a process or treatment actually has an effect on the population of interest, or whether two groups are different from one another. If the groups come

from a single population (e.g., measuring before and after an experimental treatment), perform a paired t-test. Hence, we perform the same due to our similar situation.

I did two different **t-tests** based on the total **average distance** from the wall. Average time and distance from the wall for one lap, all 3 per arena configuration. In particular, the first t-test compares the average distance from each of the four configurations (scenarios). For the second t-test, I needed to compare two data sets that contained the **average lap time** for each configuration.
The results can be found below:

| | Value |
|---|---|
| *H* | 1 |
| *p* | 0.00783 |

| | Value |
|---|---|
| *H* | 0 |
| *p* | 1.98e-05 |

Now, coming to the final part of the t-test I chose the **mean distances** for each lap and also each configuration. Hence, in every t-test, the code compared a data set with the dimension of 5(in all 4 t-tests). The results can be found below:

| | 6 tokens | 7 tokens | 8 tokens | 9 tokens |
|---|---|---|---|---|
| *H* | 1 | 1 | 1 | 1 |
| *p* | 0.00050 | 0.01572 | 0.00712 | 0.00479 |

## 4 Conclusion and Comparison

A number of conclusions can be made from the tests and statistical analysis regarding the behavior of my robot and the given solution to the Professor, which can be seen below:

| **Parameters** | **My Robot** | **Given Solution** |
|---|---|---|
| Distance from walls | Further from the walls | closer to the walls |
| Time taken to complete a lap | More | Less than my robot |
| Dependencies on the no. of tokens | more time for more tokens | independent |

*All the conclusions and comparisons made are valid and can be supported with all the tests done and explained in the previous sections. Hence, it can be said from the statistical analysis that all the* ***conclusions are valid****.*