UNIVERSITÀ DEGLI STUDI DI GENOVA

MASTER DEGREE COURSE IN ROBOTICS ENGINEERING

Research Track II

Statistical Analysis

Manuel Delucchi (S4803977)

May, 2023

1 Introduction

The final task of the *Research Track II* assignment is to conduct a statistical analysis of the first assignment from the *Research Track I* course. This analysis focuses on comparing two different implementations: the one I created and another developed by a colleague. The objective is to determine which implementation performs better when different numbers of silver and gold tokens are placed in the environment.

1.1 Tools Used

I used *MATLAB R2022b* as my primary working environment to ensure optimal statistical analysis. In this environment, I developed the necessary code to collect all the relevant data, perform a valid statistical test and generate meaningful graphs.

1.2 Hypotesis Testing

Hypothesis Testing is a statistical tool commonly used in scientific research to make decisions based on data. The basic idea behind hypothesis testing is to use data to test a claim or hypothesis about a *population parameter* (e.g. *mean*, *variance*, etc.). This claim is usually stated in terms of two competing hypotheses: the ***null hypothesis*** and the ***alternative hypothesis***.

The ***null hypothesis*** (H_0) is the claim we want to test and usually represents the default assumption. The ***alternative hypothesis*** (H_a) is the claim that we want to accept if the null hypothesis is rejected. The two hypotheses are ***mutually exclusive***, meaning that if one is true, the other must be false.

In the current analysis the two hypotheses used to assess the performance of the algorithms are:

- ***Null Hypothesis***: The results obtained from both algorithms are highly similar, and it is not feasible to determine which algorithm performs better overall. This hypothesis assumes that there is no significant difference in performance between the two algorithms.
- ***Alternative Hypothesis***: The two algorithms exhibit substantial differences in performance. This hypothesis suggests that there is a statistically significant distinction between the performance of the two algorithms.

To test the null hypothesis, we collect data from a sample and calculate a ***test statistic*** (e.g. a *t-statistic*, *z-statistic*, *chi-square statistic*, etc.) that measures the distance between the sample estimate and the hypothesised population parameter. We then use this test statistic to calculate a ***p-value***, which is the probability of getting a test statistic that is as extreme or more extreme than the one we observed, assuming the null hypothesis is true.

If the p-value is less than a pre-specified ***significance level*** (for example, α = *0.05*), we reject the null hypothesis and accept the alternative hypothesis. This means that we have evidence to support the claim made by the alternative hypothesis. If the p-value is greater than the *significance level*, we do not reject the null hypothesis, which means that we do not have sufficient evidence to support the claim made by the alternative hypothesis.

In summary, hypothesis testing is a powerful statistical tool that allows researchers to make decisions based on data. By specifying the null and alternative hypotheses, calculating a test statistic, and calculating a p-value, we can determine whether or not the evidence supports the claim made by the alternative hypothesis.

2 Experiment

2.1 Collecting Data

I conducted an experiment using different arena configurations, varying the number of silver and gold tokens from 2 to 8. The data collection involved running two projects: ***SimulationA***, which was my own project, and ***SimulationB***, developed by a colleague. Each project was executed five times for each token configuration, resulting in a total of N = *35* trials per project. In the analysis, I measured the *total time taken by the robot to complete its task*. To capture this data, I incorporated simple commands utilizing the *time* library:

```
import time
start = time.time()

### Control Loop ###

end = time.time()

elapsed_time = end - start
print("Time: ", elapsed_time)
```

Note: To modify the number of tokens in the arena, you simply need to adjust the value of the *"TOKENS-PER-CIRCLE"* variable. You can find this variable inside the *"arenas"* folder, specifically in the *"two-colours-assignment-arena.py"* file.

2.2 Comparison

After collecting all the data, the next step involved comparing the two simulations to perform a statistical analysis. Using *MATLAB*, I created a *bar graph* to visually depict the comparison between the average time taken to complete the task for each configuration in the arena. The graph provides a clear representation of the results, as shown in Fig.1.

When comparing the two simulations for each token configuration, we can observe that they exhibit almost similar behavior. It is difficult to say whether one algorithm is more efficient than the other. However, for certain configurations, one algorithm appears to be more efficient than the other, which is probably due to the way in which the algorithms for moving the robot in the environment have been implemented. It's worth noting that a significant difference is observed in the configuration with 5 and 8 tokens, where *SimulationB* has a slower average completion time compared to *SimulationA*.

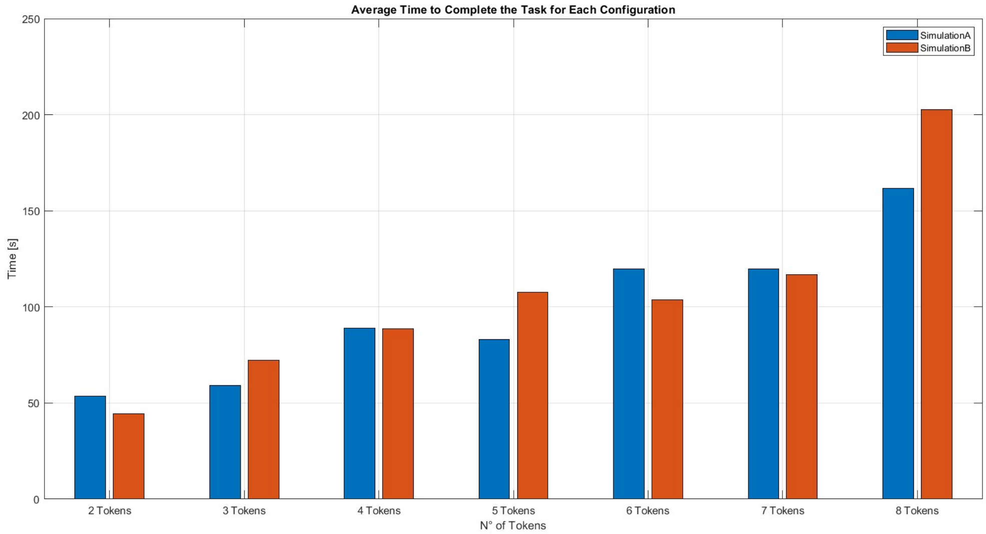

Figura 1: *"The graph displays the average time taken to complete the task for each configuration in the arena. The data is represented in blue for my own project (SimulationA) and in red for my colleague's project (SimultaionB)."*

I also did a comparison between each configuration for both simulations that can be seen in Fig.2

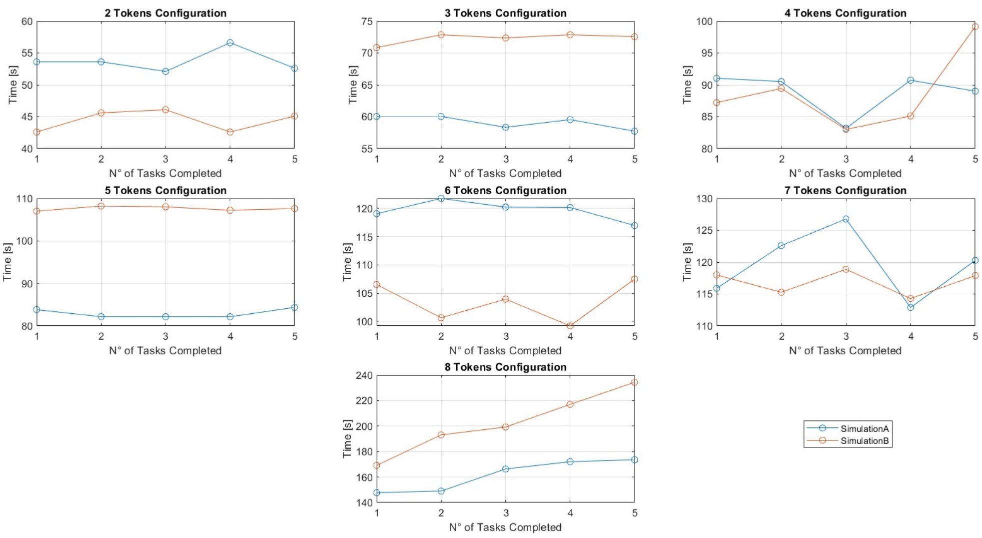

Figura 2: *"The image shows a comparison considering different token configurations. Each project was executed five times for each configuration."*

3 T-Test

3.1 What is it?

The ***t-test*** (also known as ***Student's T-Test***) is a statistical hypothesis test used to determine if there is a significant difference between the means of two groups or populations. It is commonly employed when working with small sample sizes or when the population standard deviation is unknown.

The t-test calculates the *t-statistic*, which measures the difference between the means of the two groups relative to the variation within each group. By comparing the t-statistic to a critical value derived from the t-distribution, we can determine whether the observed difference in means is statistically significant or simply due to random chance.

The t-test assumes that the data follows a (approximately) *normal distribution* and that the *samples* being compared are *independent* and *randomly selected*. It also assumes that the *variance* within each group is *approximately equal*.

When choosing a t-test, you will need to consider two things: whether the groups being compared come from a single population or two different populations, and whether you want to test the difference in a specific direction.

One-sample, two-sample, or paired t-test?

- If the groups come from a single population, perform a ***paired t-test***;
- If the groups come from two different populations, perform a ***two-sample t-test*** (or ***independent t-test***);
- If there is one group being compared against a standard value, perform a ***one-sample t-test***.

3.2 In Practice...

With a total of 35 samples for each of the two populations, I am able to employ a t-test for comparing their means. Specifically, I can conduct a ***two-sample t-test***, as previously mentioned.

Having 35 samples in each population grants a reasonably adequate sample size, which adheres to the ***Central Limit Theorem***. This characteristic allows for more robust statistical analysis. It is worth noting that a larger sample size generally leads to more precise estimates and enhances the ability to identify significant differences between the groups.

In terms of the t-test output, we can evaluate the outcome based on the value of H. When H equals 0, it indicates a failure to reject the null hypothesis with the given data. Conversely, when H equals 1, it signifies a rejection of the null hypothesis based on the available evidence.

Another crucial aspect provided by the t-test is the ***p-value***. This value assists in determining whether the difference between the observed and hypothesized results is merely due to random sampling variability or if it holds statistical significance. A statistically significant difference implies that the observed disparity is unlikely to be explained solely by chance introduced through the sampling process.

After performing a comparison between two datasets in *MATLAB*, I have obtained some valuable insights:

	H	***p-value***
Value	0	0.4780

3.3 Conclusion

Upon examining the results obtained, we observe that the values indicate $H = 0$ and $p = 0.4780 >> 0.05$. This outcome implies that we accept the null hypothesis based on the available evidence.

The result $H = 0$ signifies a failure to reject the null hypothesis. In other words, the data does not provide sufficient evidence to support a significant difference between the compared populations. It suggests that any observed disparities between the means of the populations are likely due to random sampling variability rather than true differences.

Additionally, the p-value, which quantifies the probability of obtaining results as extreme as or more extreme than the observed data, is 0.4780. This value is substantially higher than the commonly used significance level of 0.05. Consequently, we conclude that the observed difference between the populations is not statistically significant and can be attributed to sampling variability.

Accepting the null hypothesis does not imply that the populations are identical or have identical means. It simply suggests that the evidence does not provide substantial support for claiming a significant difference between them.