Università degli Studi di Genova

DIBRIS

Department of Computer Science and Technology,
Bioengineering, Robotics and System Engineering

Research Track II

Third Assignment

Statistical Analysis on Research Track I's First Assignment

Author:
Ouassim Milous

Student ID:
s5938924

Professors:
Carmine Tommaso Recchiuto

May 31, 2024

Contents

1 Assignment Description

This report presents a Statistical Analysis of the Execution time of the first RT1 assignment.

1.1 Goal

The goal of this assignment is to perform a statistical analysis to compare two different implementation algorithms of a robot's movement through an arena to pick up and regroup tokens that are placed randomly. We analyze the performance of my implementation and compare it to that of one of my colleagues.

2 Task Description

2.1 Hypotheses Formulation

- **Null Hypothesis ($\mathbf{H_0}$)**: Represents the default assumption or the statement of no difference. It assumes that there is no significant difference or relationship between the variables under investigation. The null hypothesis is symbolized as H_0.
- **Alternative Hypothesis ($\mathbf{H_a}$)**: It suggests that there is a significant difference or relationship between the variables. It represents the possibility of an effect or difference existing. The alternative hypothesis is symbolized as H_a.

2.2 Experimental Setup

- Experiment Scenario: We conduct the experiment with a robot trying to regroup golden tokens placed randomly in the environment.
- Number of Repetitions: We perform the experiment 50 times to obtain sufficient and accurate data for statistical analysis.
- Number of Tokens: The number of tokens will be set to 6.

2.3 Performance Evaluation Metrics

We evaluate the performance of the algorithms by the Average Execution Time. We measure the average time taken by each implementation to complete the task of retrieving all tokens. This metric will provide insights into the efficiency of the algorithms in terms of time complexity.

2.4 Statistical Analysis

1. Hypotheses Formulation: We clearly state the null and alternative hypotheses.
2. Experimental Setup: We provide a detailed description of the experimental setup and the number of repetitions for the experiment.
3. Results And Statistical Analysis: We present the collected execution time required for each algorithm. We analyze the results obtained from the statistical tests, discussing significant differences observed between the algorithms.
4. Conclusion: Summarize the overall findings and state whether the hypothesis was proven or rejected based on the statistical analysis.

3 Hypotheses Formulation

We formulate the H_0 and the H_a to later conduct statistical analysis to determine whether we accept H_0 or do we reject it in favor of an H_a:

3.1 Null Hypothesis H_0

"There is no significant difference in performance between my regrouping algorithm and my colleague's algorithm."

We propose the null hypothesis to assume both algorithms perform alike, under the same conditions (without the effect of an external condition, for example: randomness in token placement settings), which means our algorithm and my colleague's algorithm perform equally well.

3.2 Alternative Hypothesis H_a

"There is a significant difference in performance between my regrouping algorithm and my colleague's algorithm."

We propose the Alternative hypothesis to assume one algorithm performs better than the other, under the same conditions (which means it cannot be explained with the existence of external condition, for example: randomness in token placement settings).

4 Experimental Setup

The experiment is conducted with 6 tokens placed randomly throughout a square arena. The number 6 was chosen to fit best the size of the arena and to get a number high enough for us to distinguish a good algorithm from a worse one. The randomness was suggested to ensure the algorithm is going to face all possible challenges that might be a factor of impacting the chosen criteria (execution speed) without being biased.

Also, the experiment will be looped 50 times, ensuring going through a sufficient number of tests to come up with conclusions.

5 Results And Statistical Analysis

Here we will do a T-Test analysis on the data that we get after collecting 50 test samples of both algorithms.

5.1 Results

After running the codes, we got the following data:

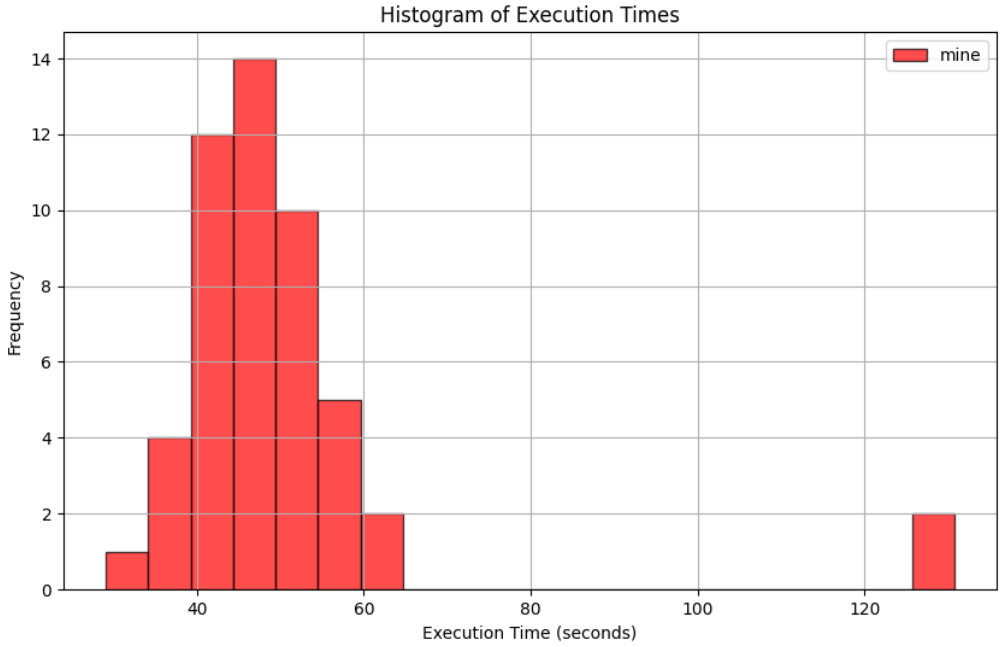
5.1.1 My Results

The 50 measured data in seconds:

48.13187289237976	128.55767393112183	52.73160648345947	49.56117129325867	47.65165567398071
45.95206665992737	49.05328416824341	34.532639026641846	46.23897361755371	39.87034296989441
42.61948275566101	40.004302740097046	34.87387037277222	54.4050714969635	49.826955795288086
52.18607521057129	62.33963871002197	51.85245394706726	45.34819412231445	59.50053095817566
45.60940980911255	37.43087601661682	38.43876266479492	53.158101320266724	43.89826250076294
29.116892099380493	46.10402297973633	130.74855208396912	41.37489700317383	48.44751977920532
42.99095940589905	44.416335582733154	43.84186100959778	42.38340473175049	50.54225206375122
43.96574068069458	55.473950147628784	52.764405488967896	58.51520133018494	50.264798402786255
43.79601788520813	42.11333775520325	43.78447341918945	56.08797025680542	58.40659785270691
49.24154782295227	46.13136649131775	46.54616951942444	63.27558660507202	46.53680920600891

Mean of approximately 50.61 seconds and a Standard deviation of approximately 17.73 seconds.

The data can also be visualized in the graph:



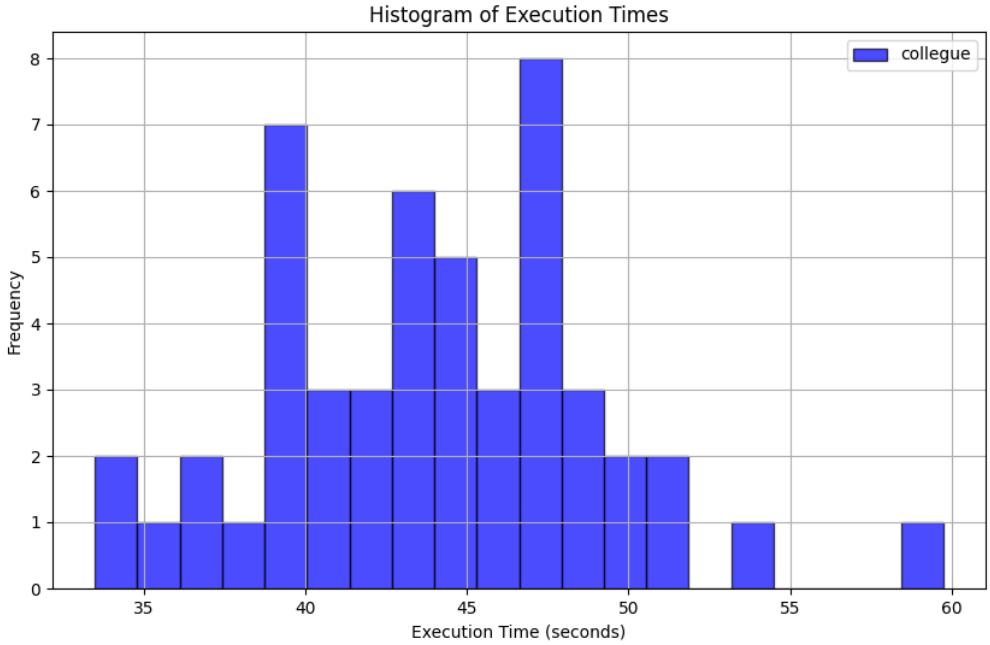
5.1.2 My Colleague's Results

The 50 measured data in seconds:

39.17593717575073,	35.67468976974487,	47.991981983184814,	37.851794481277466,	47.25392508506775,
37.184141635894775,	39.37392497062683,	59.74052977561951,	39.738651275634766,	37.18262791633606,
41.81641340255737,	33.480597734451294,	51.30228519439697,	42.94026231765747,	44.47332048416138,
39.913494348526,	50.38465619087219,	49.49549412727356,	48.684606075286865,	39.47776746749878,
48.667150259017944,	47.27253079414368,	40.27370882034302,	44.91829180717468,	40.93896961212158,
46.674877405166626,	34.41720652580261,	45.45862793922424,	44.71782922744751,	43.4380521774292,
47.4894483089447,	45.998939752578735,	42.18389630317688,	40.76705265045166,	43.237879037857056,
41.98980379104614,	39.369401693344116,	53.61984872817993,	43.21582889556885,	44.08116388320923,
43.10384225845337,	47.55291128158569,	43.69208264350891,	51.42169117927551,	47.52693295478821,
46.18242621421814,	47.771875858306885,	47.632659673690796,	44.91829180717468,	39.47776746749878

This results in a Mean of approximately 44.02 seconds and a Standard deviation of approximately 5.11 seconds.

The data can also be visualized in the graph:



5.2 Statistical Analysis

5.2.1 Choosing a Significance Level

We choose 0.05 %5 to be the significance level, which means we will have a %5 probability of rejecting H_0 when it's true. It allows us to have a threshold to accept or reject H_0.

5.2.2 Calculating the T-Value

the T-Value is calculated using the following formula:

$$t = \frac{mean1 - mean2}{\sqrt{\frac{s1^2}{n1} + \frac{s2^2}{n2}}} \tag{1}$$

We do numerical application:

$$t = \frac{50.61287889480591 - 44.02296180725098}{\sqrt{\frac{17.7337982280017352}{50} + \frac{5.11001036581950352}{50}}} \tag{2}$$

$$= \frac{6.58991708755493}{\sqrt{\frac{314.4676678638767}{50} + \frac{26.11103131831287}{50}}} \tag{3}$$

$$= \frac{6.58991708755493}{\sqrt{6.289353357277534 + 0.5222206263662574}} \tag{4}$$

$$= \frac{6.58991708755493}{\sqrt{6.811573983643791}} \approx \frac{6.58991708755493}{2.610878618488511} \approx 2.522928270048885 \tag{5}$$

After numerical application, we get a T-Value of approximately 2.52.

5.2.3 Calculating Degrees of Freedom

The degrees of freedom (df) for the t-test are found using the formula:

$$df = n_1 + n_2 - 2$$

which gives us $50 + 50 - 2 = 98$.

5.2.4 Obtaining Critical Value

According to the t-distribution table for a two-tailed test with a significance level of 0.05 and using the T-Statistic and the degrees of freedom calculated, the critical value is approximately ± 1.984.

5.2.5 Comparing The T-Value and The Critical Value

The T-value we got (approximately 2.52) exceeds the critical value of approximately 1.984. Therefore, we reject the Null Hypothesis H_0 and conclude that we should accept the Alternative Hypothesis (H_a) instead.

6 Conclusions

The conducted analysis provides us with an evaluation of the performance of our algorithm compared to my colleague's. We accepted the Alternative Hypothesis H_a, indicating a significant difference in performance between my regrouping algorithm and my colleague's algorithm.

Hence, the t-test analysis revealed that my colleague's algorithm outperformed mine in terms of execution speed. Based on the results, it is evident that further improvements should be made to our algorithm to enhance performance. It is suggested to study both algorithms to understand what enhancements can be made.