UNIVERSITÀ DEGLI STUDI DI GENOVA

DIBRIS

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

RESEARCH TRACK 2

Professor:
Carmine Tommaso Recchiuto

Third Assignment

Statistical Analysis

Author:
Saeed Abdollahi Taromsari

Student ID:
5397691

May 2023

Contents

1 Assignment Description

The aim of this task is to conduct a statistical analysis that compares two distinct token placement algorithms. Your primary focus will be to evaluate the performance of your own implementation against a colleague's implementation when silver and golden tokens are randomly placed within the environment. The goal of the analysis is to determine which algorithm delivers better results based on the established performance metrics. Moreover, you will need to carefully plan the experiments, select an appropriate statistical methodology, and prepare a report that summarizes your findings.

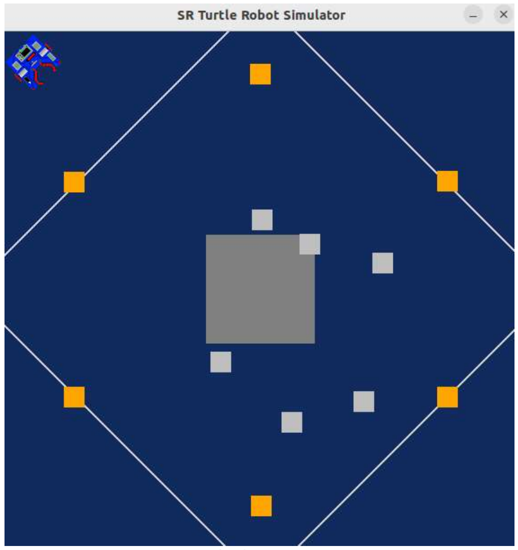


Figure 1: Starting point.

2 Steps of a Statistical Analysis

To evaluate the efficiency of two distinct token placement algorithms, a statistical analysis can be carried out. The process involves designing an experiment and thoroughly examining the outcomes. It is crucial to plan and report each step of the analysis carefully. These are the necessary steps required to successfully conduct the evaluation:

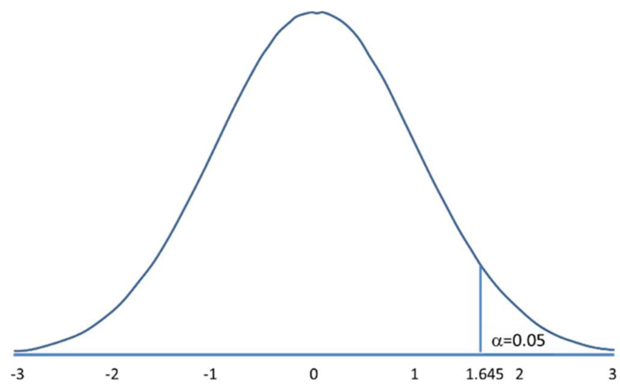
1. **Define the Hypothesis:** The hypothesis we want to test is whether one algorithm performs better than the other when silver and gold tokens are randomly placed in the environment.
2. **Experimental Design:** We will design the experiment to compare the two implementations while varying the number of boxes. This will allow us to observe how the algorithms behave under different conditions.
3. **Variables:**
    - **Independent Variable:** The implementation of the algorithm (yours and your colleague's)
    - **Dependent Variables:** ‘Average time required to finish the task’ and ‘Number of successes/failures’.
4. **Data Collection:** Conduct a series of experiments using both implementations. For each experiment, measure the average time required to finish the task and record whether it was a success or failure.
5. **Determine the Sample Size:** Decide on the number of experiments to perform for each implementation and the number of boxes. A larger sample size will yield more reliable results.
6. **Statistical Analysis:** To compare the two implementations, you can use various statistical techniques, such as:
    - **Average Time Analysis:** Calculate the mean average time for each implementation and compare them using a t-test.
    - **Success/Failure Analysis:** Calculate the proportion of successes and failures for each implementation and perform a chi-square test or Fisher's exact test to determine if there is a significant difference in performance.

To determine when to reject the null hypothesis, a statement called the decision rule is used. The decision rule relies on three factors:

- **Alternative Hypothesis**
- **Test statistics**
- **Level of Significance**

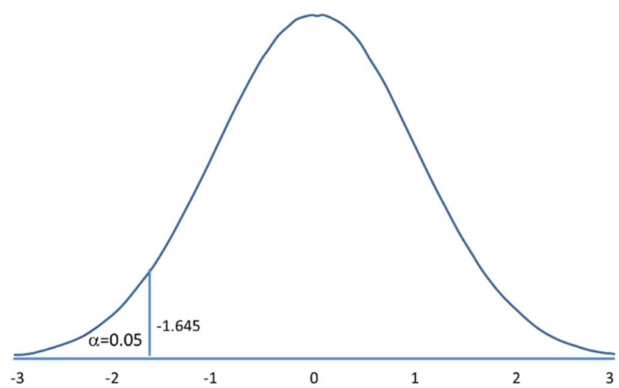
The type of test (upper-tailed, lower-tailed, or two-tailed) determines whether the decision rule requires investigators to reject the null hypothesis if the test statistic is larger than, smaller than, or extreme. The test statistic's exact form also plays a role in determining the decision rule, with the standard normal distribution being used for Z tests and the t distribution used for T-tests. The level of significance chosen in Step 1 determines the critical value to be used.

Below each figure, you can find the decision rules for upper-tailed, lower-tailed, and two-tailed T-tests with an α of 0.05. The rejection regions are shown in the upper, lower, and both tails of the curves, respectively. Take note that the decision rules are clearly stated to avoid confusion:



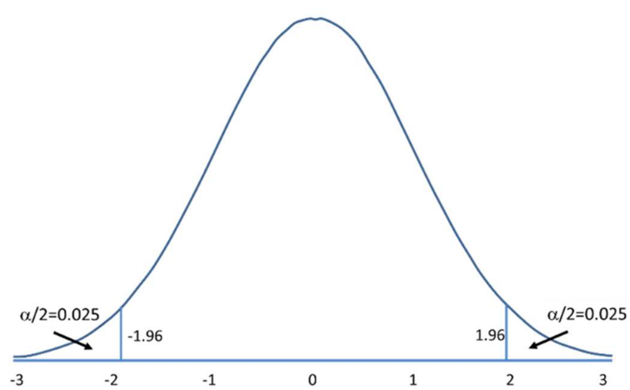
Upper-Tailed Test	
α	Z
0.10	1.282
0.05	1.645
0.025	1.960
0.010	2.326
0.005	2.576
0.001	3.090
0.0001	3.719

Figure 2 Rejection Region for Upper-Tailed Z Test (H1: μ > μ0) with α=0.05
The decision rule is: Reject H0 if Z > 1.645.



Lower-Tailed Test	
a	Z
0.10	-1.282
0.05	-1.645
0.025	-1.960
0.010	-2.326
0.005	-2.576
0.001	-3.090
0.0001	-3.719

Figure 3 Rejection Region for Lower-Tailed Z Test (H1: μ < μ0) with α =0.05
The decision rule is: Reject H0 if Z < 1.645.



Two-Tailed Test	
α	**Z**
0.20	1.282
0.10	1.645
0.05	1.960
0.010	2.576
0.001	3.291
0.0001	3.819

Figure 4 Rejection Region for Two-Tailed Z Test (H1: μ ≠ μ0) with α =0.05
The decision rule is: Reject H0 if Z < -1.960 or if Z > 1.960.

3 Implementation of Statistical Analysis for Evaluation of the Token Placement Algorithms

3.1 Define the Hypothesis

- **Null Hypothesis (H0)** postulates that there is no significant difference between the performance of the two programs.
- **Alternative Hypothesis (Ha)** assumes that the two programs differ significantly, with one algorithm performing better than the other algorithm. So, the alternative hypothesis can take one of three forms:
    1. **H1:** $\mu > \mu_0$, where μ_0 is the comparator or null value and an increase is hypothesized - this type of test is called an upper-tailed test.
    2. **H1:** $\mu < \mu_0$, where a decrease is hypothesized, and this is called a lower-tailed test; or
    3. **H1:** $\mu \neq \mu_0$, where a difference is hypothesized, and this is called a two-tailed test.

The specific shape of the alternative hypothesis is determined by the researcher's understanding of the parameter of interest and whether it may have risen, declined, or deviated from the null value. The alternative hypothesis is established by the researcher prior to the collection of any data. **In the context of the token placement algorithm, the alternative assumption is that the time efficiency of my algorithm surpasses that of my peers ($\mu > \mu_0$).**

3.2 Experimental Design

In order to evaluate the time efficiency of the two algorithms, we employ a method of randomly placing silver tokens on the scene. Figure 2 displays four examples of randomly distributed tokens in a scene.

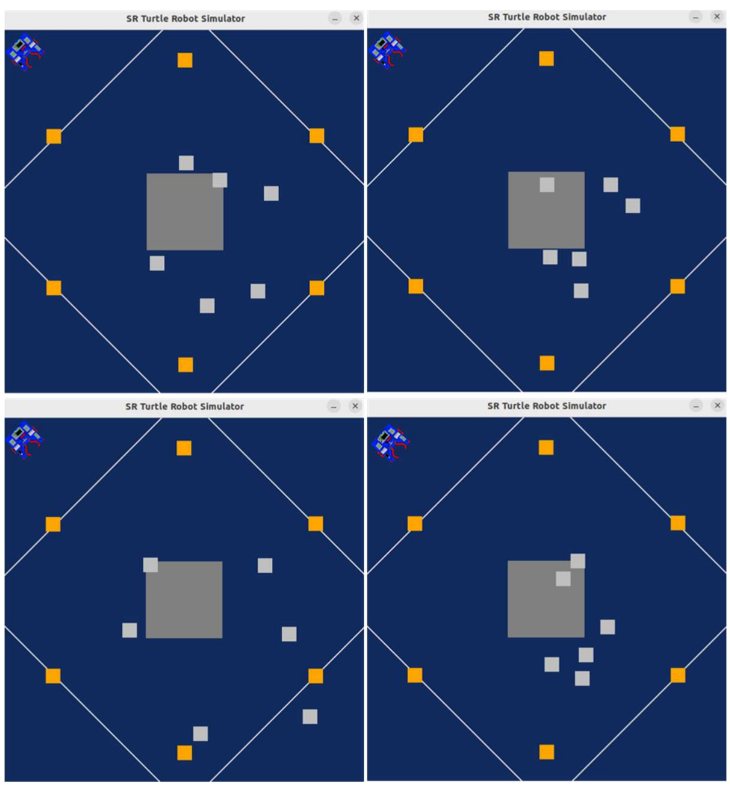


Figure 5 Four examples of the tokens being randomly distributed in a scene.

3.3 Variables, Sample Size and Data Collection

As performance evaluators, we consider the **Average Time Required** that assesses the meantime each implementation takes to retrieve all tokens. It offers valuable insights into the efficiency of the algorithms in relation to time complexity. Also, to ensure that we have enough data for statistical analysis, **we conduct each experiment type 50 times**.

In order to gather information, we initially generated 50 sets of silver tokens arranged in random patterns. Subsequently, we applied the algorithms to these arrangements and noted the time taken for each step. The data collected during this process for the 50 steps are presented in the following chart:

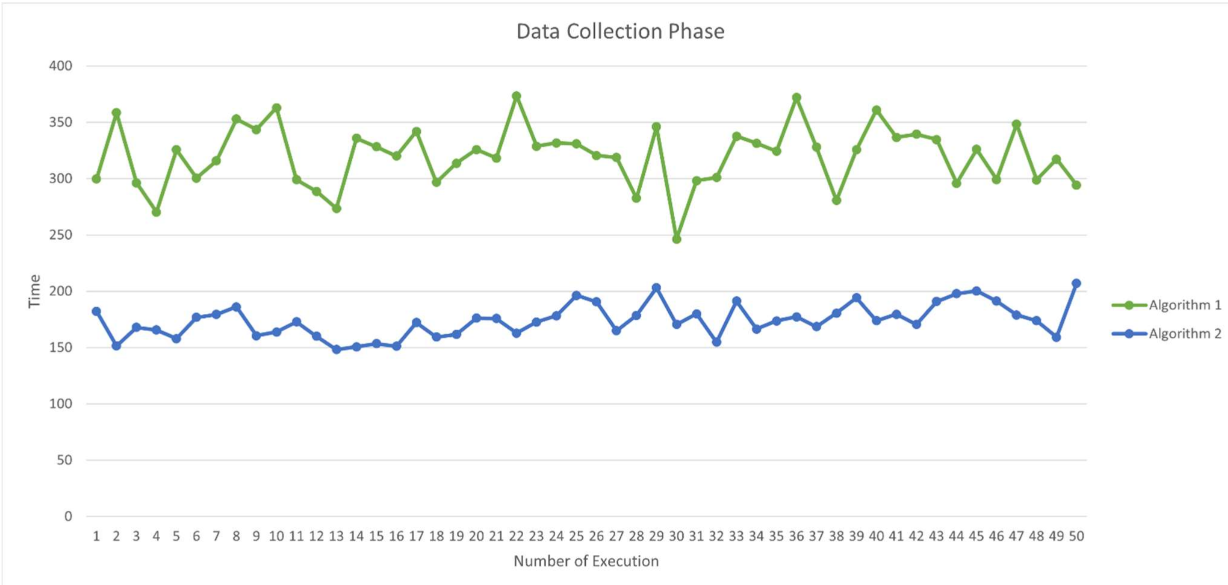


Figure 6 Data Collection Phase.

3.4 Statistical Analysis

When the true variance is not available, we can use the t-test to compare the time average of algorithms. Two types of t-tests are available:

- **A one-sample t-test** is used to compare a single population to a standard value (for example, to determine whether the average lifespan of a specific town is different from the country average).
- **A paired t-test** is used to compare a single population before and after some experimental intervention or at two different points in time (for example, measuring student performance on a test before and after being taught the material).

If the goal is not to check the impact of token arrangement on algorithm execution, the paired t-test is unnecessary. In our experiment, we want to compare the time efficiency of two algorithms under the same conditions**. Although the silver token arrangement is randomly changed in each experiment, their arrangement in each step is the same for both algorithms. Hence, we require a one-sample t-test.**

***t* Table**

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Figure 6 The t-test values for different tail, degree of freedom and confidence level

The type of t-test we should use depends on the alternative hypothesis we have. Because our hypothesis is directional ($\mu > \mu_0$), we should opt for an **upper one-sided t-test**.

The next step is computing the **t-value** and the **critical value** for a **confidence level of 5% (** $\alpha = 0.05$ **):**

- **t-value (Based on Calcualtion in the Exel File):**

$$\mu_1 = 319.9752558$$
$$\mu_2 = 174.070797$$
$$\sigma_1^{\;2} = 728.6906172$$
$$\sigma_2^{\;2} = 214.2275887$$
$$n_1 = 50$$
$$n_2 = 50$$
$$t - value = \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{\sigma_1^{\;2}}{n_1} + \frac{\sigma_2^{\;2}}{n_2}}} = 33.59824211$$

- **critical value** ($\alpha = 0.05$):

$$\alpha = 0.05$$
$$dof = n_1 + n_2 - 2 = 50 + 50 - 2 = 98$$
$$t_{Critical} = 1.6600$$

According to the table below, the critical value for a 5% confidence level and 98 degrees of freedom is 1.6600.

***t* Table**

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

4 Conclusion

We conducted a t-test to compare the average run times of two algorithms and found a t-value of 33.59824211. To determine whether the two algorithms are significantly different, we compared the t-value with the critical value of 1.6600 for our specific case at a 5% confidence level. Since the t-value we obtained is much greater than the critical reference value, we can conclude that the difference in performance between the two algorithms is statistically significant. As a result, we can reject the null hypothesis that there are no significant differences between the two algorithms and support the alternative hypothesis that one algorithm performs better than the other.