



UNIVERSITÀ DEGLI STUDI DI GENOVA

DIBRIS

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

RESEARCH TRACK 2

Statistical Analysis

Professor:

Carmine Recchiuto

Author:

Peyman Peyvandi Pour

Student ID:

s5573284

CONTENT

Abstract	3
Hypotheses.....	3
Method	3
Data Collection.....	4
Analysis Procedure.....	6
Calculation of Paired Differences:	6
Formulation of Hypotheses.....	6
Computation of Test Statistic:	6
Determination of Significance:	7
Results.....	7
Conclusion.....	7

ABSTRACT

This report details the statistical analysis process used to compare the effectiveness of two programs created for the Research Track 1 course assignment. The algorithms under examination were created by Ali Rabani Doost and me. Our algorithms have been specifically created to allow a robot to match the golden tokens with the silver tokens that are present in the arena. This experiment's main goal is to compare the effectiveness of the two scripts while varying the number of tokens and the radius at which they are generated. By doing this, we hope to investigate how the programs adjust to various spatial arrangements and assess their general success in carrying out the desired goal.

HYPOTHESES

The hypotheses formulated for this analysis were as follows:

Null Hypothesis (H_0): There is no significant difference in the performance of the two programs.

Alternative Hypothesis (H_1): There is a significant difference in the performance of the two programs.

METHOD

Testing the hypotheses was done using a paired T-test. In this case, the performances from running both programs on the same token configuration with varying radius and number of tokens were compared using this statistical test for comparing the means of two sets of paired observations. We will use a T-distribution, and in order to apply the central limit theorem to the sampling distribution, we will need a number of samplings greater than 30 in order to get the shape of the distribution more similar to a T-distribution.

DATA COLLECTION

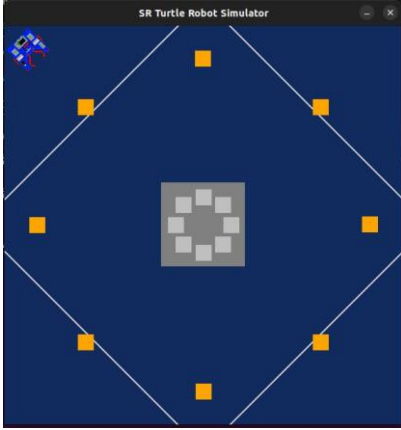


Figure 1: set radius to 0.4

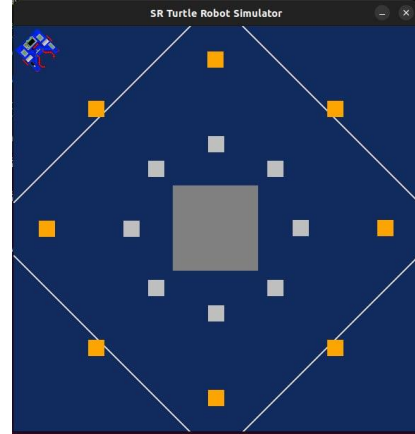


Figure 2: set radius to 1.2

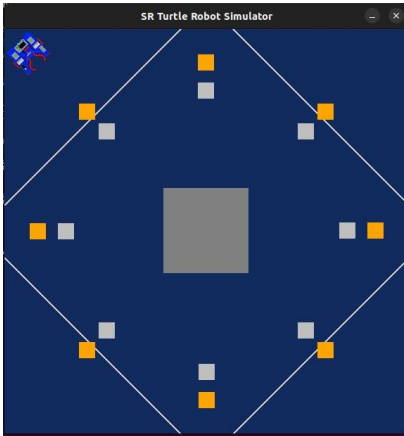


Figure 3: set radius to 2

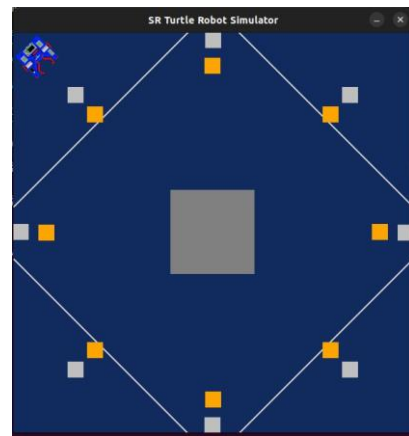


Figure 4: set radius to 2.8

We measured performance for both programs while systematically varying the token radius and number. Measurements from one program were paired with measurements from the other. By comparing these paired observations, we were able to analyze the programs' performance in different experimental conditions. Using 5 runs of the algorithms for each configuration (position-number of tokens) ensures a more accurate measurement of success. Based on the evaluation of a total of 40 overall states, which exceed the minimum threshold of 30, we can apply the central limit theorem. As a result, we can conclude that the sample distribution follows a T-distribution shape. The table below displays the obtained results.

			Success rate		
Sample number	Radius	Number of Tokens	Peyman's algorithm	Rafael's algorithm	Difference
1	0.4	5	1	1	0
2	0.4	5	1	1	0
3	0.4	5	1	1	0
4	0.4	5	1	1	0
5	0.4	5	1	1	0
6	0.4	8	0.75	0.875	0.125
7	0.4	8	0.75	0.75	0
8	0.4	8	0.75	0.875	0.125
9	0.4	8	0.875	0.875	0
10	0.4	8	0.75	0.875	0.125
11	1.2	5	0.8	1	0.2
12	1.2	5	1	1	0
13	1.2	5	1	1	0
14	1.2	5	1	1	0
15	1.2	5	1	1	0
16	1.2	8	0.75	1	0.25
17	1.2	8	0.75	1	0.25
18	1.2	8	0.75	1	0.25
19	1.2	8	0.75	0.875	0.125
20	1.2	8	0.75	0.875	0.125
21	2	5	0.6	1	0.4
22	2	5	0.8	1	0.2
23	2	5	0.6	1	0.4
24	2	5	0.6	0.8	0.2
25	2	5	0.6	1	0.4
26	2	8	0.5	0.875	0.375
27	2	8	0.5	0.75	0.25
28	2	8	0.25	0.875	0.625
29	2	8	0.25	0.625	0.375
30	2	8	0.5	0.625	0.125
31	2.8	5	0.8	0.8	0
32	2.8	5	0.6	1	0.4

33	2.8	5	0.8	0.8	0
34	2.8	5	0.8	1	0.2
35	2.8	5	0.6	1	0.4
36	2.8	8	0.625	0.875	0.25
37	2.8	8	0.75	0.875	0.125
38	2.8	8	0.875	0.75	-0.125
39	2.8	8	0.625	0.625	0
40	2.8	8	0.875	0.875	0

ANALYSIS PROCEDURE

The paired T-test was employed to examine the performance differences between the two programs across different token count and radius configurations. The action steps listed below were carried out:

Calculation of Paired Differences:

The performance differences (d) between the two programs for each token count and radius setting were calculated by subtracting the performance measurement of one program from the corresponding measurement of the other program.

Formulation of Hypotheses:

The null hypothesis (H_0) stated that there is no significant difference in performance between the two programs. The alternative hypothesis (H_1) proposed that there was a significant difference in performance between the programs.

Computation of Test Statistic:

The mean difference (\bar{d}) and standard deviation ($SE(\bar{d})$) of the paired differences were calculated. Using these values, the T-value was computed, measuring the difference between the observed mean difference and the hypothesized mean

difference under the null hypothesis.

Determination of Significance:

The calculated T-value was compared to the critical T-value at a chosen significance level (5%). If the calculated T-value exceeded the critical T-value, the null hypothesis was rejected in favor of the alternative hypothesis, indicating a significant difference in performance.

We have computed the following values as part of our analysis:

1. Calculation of the difference between the two observations on each pair is done in the last column of the table.
2. Calculation of the mean of the difference: $\bar{d} = 0.154375$
3. Calculation of the standard deviation of the difference: $S_d = 0.169169114$
4. Calculation of the standard error of the difference: $SE(\bar{d}) = 0.026747985$
5. Calculation of the T-value: $T = 5.77146268$

Results:

The paired T-test was performed on the collected data, yielding a calculated T-value of 5.77146268. With a significance level of $\alpha = 0.05$ (5%) and degrees of freedom ($df = N - 1 = 39$), where N represents the number of paired observations (40), the critical T-value at a two-sided test was 2.021.

Conclusion

Based on the results, the calculated T-value (5.77146268) exceeded the critical T-value (2.021). Therefore, we reject the null hypothesis and conclude that there is a significant difference in the performance of the two programs with varied token count and radius settings.

The findings suggest that the performance of the programs significantly differs

when the token count and radius are varied. This indicates that the programs' effectiveness in matching silver and golden tokens is influenced by the number of tokens and the spatial distribution within the arena.