UNIVERSITÀ DEGLI STUDI DI GENOVA

DIBRIS

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

RESEARCH TRACK 2

Third Assignment

Author:

Rocca Giovanni

Student ID:

s4802954

Professor:

Carmine Recchiuto

June 1, 2023

Contents

1 Assignment Description

Perform a statistical analysis on the first assignment, considering two different implementations (yours, and a solution of one of your colleagues) and testing which one performs better, when silver and golden tokens are randomly placed in the environment.

As performance evaluators you may possibly consider:

- the average time required to finish the task
- the number of success / failures
- . . .

Possibly, you can also vary the number of boxes to see how the algorithm behaves. You can add obstacles by modifying the file: robot-sim/sr/robot/arenas/two_colours_assignment_arena.py

token.location = (cos(angle) * radius, sin(angle) * radius)

It's very important here to clearly define what is the hypothesis that you want to test. Do you want to test which algorithm does perform better when the token are placed on the two circles? Or when the token are randomly placed in the environment?

- Based on the hypothesis, you need to design the experiments accordingly Carefully plan the number of experiments and choose a suitable statistical approach.

Write a report composed of:

- Hypotheses made (null hypothesis and alternative hypothesis)
- Description and motivation of the experimental setup (types of experiments, number of repetitions)
- Results
- Discussion of the results with statistical analysis
- Conclusion (is the hypothesis proven?)

2 Introduction

In order to work properly on this assignment I have collaborated with Michele Moriconi.

We have focused on observing which of our algorithms works best as the environment changes, to be able to carry out statistical analysis on the speed of completion of the tasks.

3 Hypotheses Made

Null Hypothesis: I know that the two algorithms are running well, noting that both in Research Track 1 assignment 1 quickly completed the task without a hitch (taking more or less the same amount of time) and thus we assume that they complete the tasks at the same speed on average.

Alternative Hypothesis: I take as hypothesis the one that, being the two algorithms different, they behave differently and therefore that one of them is faster than the other, clearly not in all cases, but going to observe the average behavior.

4 Description and Motivation of the Experimental Setup

The main goal of my experiment is observing the behavior of the two algorithms in terms of speed of finishing the tasks and compare these speeds in order to understand which one of the two algorithms performs better on average.

The first thing I have done is to change the environment where the robot works, in particular the position of the golden tokens.

I have been able to change the position of the tokens changing 30 times the angle of disposition along a circle of the golden tokens, modifying a seed and passing it into a function that generates a random number that corresponds to the angle (in the file robot-sim/sr/robot/arenas/two_colours_assignment_arena.py).

As statistical test I use the Paired T-Test, which is used to compare two population means where there are two samples in which observations in one sample can be paired with observations in the other sample. It is very useful because it helps us determine if there is a significant difference between the averages of the two sets of measurements.

5 Results

The table below shows the results I have obtained running the two algorithms 30 times.

Number	Moriconi	Rocca	Seed	Difference
1	113.2568562	99.31476092	10	13.94209528
2	84.685815	92.33062196	11	-7.6448071
3	94.732783	94.38754201	12	0.34524107
4	92.821157	151.74904418	13	-58.9278872
5	84.209522	91.35502696	14	-7.145504952
6	82.673911	98.55895519	15	-15.88504434
7	82.176113	104.0243602	16	-21.84824729
8	85.170658	155.7155468	17	-70.54488873
9	88.190233	85.33834004	18	2.851892948
10	90.208793	83.81672406	19	6.392068863
11	105.762834	137.593051	20	-31.83021712
12	74.680162	90.78722215	21	-16.10706019
13	78.673735	95.52155805	22	-16.8478229
14	103.792728	143.052345	23	-39.25961709
15	89.696134	100.8769982	24	-11.18086433
16	86.680724	105.44752	25	-18.76679587
17	90.673854	92.78944707	26	-2.115592957
18	94.730558	98.7431879	27	-4.012629747
19	82.203786	88.69668794	28	-6.492902041
20	75.666856	83.683676	29	-8.016819954
21	80.158795	80.27108097	30	-0.112285852
22	89.190372	101.2363181	31	-12.04594612
23	91.191099	91.69763303	32	-0.5065341
24	80.172738	84.69903183	33	-4.526293755
25	77.666526	82.18091893	34	-4.514392853
26	76.642797	80.68608212	35	-4.043285131
27	89.407257	91.69872308	36	-2.291466236
28	105.632253	93.2082839	37	12.42396903
29	92.8466649	95.2108779	38	-2.36421299
30	78.3020508	82.1743731	39	-3.87232232

As it's possible to see in the table the first column represents the number of the experiments, the second column and the third represents the time to complete the tasks respectively of my colleague's algorithm and of mine. The fourth column indicates the seed used for that specific experiment and the last column contains all the computed differences between the second and the third columns.

The mean of the differences is: -11.16493913.

I have also calculated the standard deviation: 18.4548198

The standard error of the mean difference is: 3.369373699

The last thing I have computed is the absolute value of the t-statistic: 3.313654148

6 Discussion of the Results with Statistical Analysis

In order to achieve my results I have followed these steps:

1. I calculated the difference between the two observations on each pair, distinguishing between positive and negative differences.
2. I computed the mean difference. ($\overline{d}$)
3. I calculated the standard deviation of the differences using excel. (S_d)
4. I use the standard deviation S_d to calculate the standard error of the mean difference ($SE(\overline{d})$), using the formula: $SE(\overline{d}) = \frac{S_d}{\sqrt{n}}$, where n is the number of experiments.
5. I computed the t-statistic (T), which is given by $T = \frac{\overline{d}}{SE(\overline{d})}$.
6. The last step is using tables of the t-distribution to compare the value for T to the t_{n-1} distribution. This will give the p-value for the paired t-test.

	P						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.12	2.583	2.921	3.686	4.015
17	1.333	1.74	2.11	2.567	2.898	3.646	3.965
18	1.33	1.734	2.101	2.552	2.878	3.61	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.85
21	1.323	1.721	2.08	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.5	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.06	2.485	2.787	3.45	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.689
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.66
30	1.31	1.697	2.042	2.457	2.75	3.385	3.646
60	1.296	1.671	2	2.39	2.66	3.232	3.46
120	1.289	1.658	1.98	2.358	2.617	3.16	3.373
1000	1.282	1.646	1.962	2.33	2.581	3.098	3.3
Inf	1.282	1.645	1.96	2.326	2.576	3.091	3.291

Figure 1: t-table

7 Conclusion

Having chosen as an alternative hypothesis the fact that one algorithm is faster than the other on average, without specifying which of the two is the faster one, I used the two-tailed method to analyze the table.

To analyze it, I looked at the row corresponding to n-1, in my case 29 (30 experiments - 1), and I took as a reference column the one where I found the value that best approximates by default the previously calculated t-statistic (2.756) and finally I looked at what value it matched in the two-tails row.

As can be seen, the corresponding value is: 0.01, which means I can accept the alternative hypothesis with 99% confidence.

In the end I can conclude that the null hypothesis is rejected, while the alternative one is allowed.