UNIVERSITA DEGLI STUDI DI GENOVA

DIBRIS

DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY, BIOENGINEERING AND SYSTEM ENGINEERING

Research Track II

Third Assignment

Statistical Analysis

Author

Lidia Loubar

s5989908

May 2024

1. Introduction

In this report, a statistical analysis of the performance of two different implementations of a task, referred to as "My Code" and "Colleague's Code", is presented. The task involves placing tokens in the environment, and the aim is to determine which implementation performs better under varying conditions. The performance metrics considered include the average time required to finish the task and the success rate.

To conduct the analysis, experiments were designed in which both implementations were tested with varying numbers of tokens in the environment. The elapsed time for each implementation to complete the task was recorded, and the number of successes and failures for each was tracked.

2. Hypotheses and Experimental Setup

a. Hypotheses

The hypotheses tested in this analysis are as follows:

- Null Hypothesis (H_0): There is no significant difference in performance between the "My Code" and "Colleague's Code" implementations.
- Alternative Hypothesis (H_1): One implementation performs better than the other in terms of the average time required to finish the task and/or the success rate.

b. Experimental Setup

To test these hypotheses, experiments were designed with the following characteristics:

- Task Description: The task involves placing tokens randomly in the environment. Both implementations, "My Code" and "Colleague's Code", were evaluated on their ability to complete this task efficiently and accurately.
- Variation of Conditions: The experiments were conducted with different numbers of tokens (3, 4, 5, 6, and 7 golden boxes) in the environment. This variation allowed for the assessment of performance under varying task complexities.
- Environment Creation: The same environments were created for both algorithms to ensure a fair comparison. The positions of the boxes were fixed within each environment configuration, eliminating variability due to random generation.
- Repetitions: Each algorithm was executed 6 times in each environment configuration, resulting in a total of 30 simulations for each algorithm. This repetition helped ensure reliable and consistent results.
- Data Collection: For each experiment, data were collected on the elapsed time for each implementation to complete the task and the corresponding success or failure outcome. This data collection process allowed for a detailed evaluation of performance metrics.
- Statistical Analysis: Statistical methods, including calculation of means, standard deviations, and t-tests, were employed to analyze the data and test the hypotheses. These analyses provide quantitative insights into the performance differences between the two implementations.

By creating identical environments for each algorithm and validating the comparison without using random generation of boxes, the comparability and reliability of the results were ensured. This approach allowed for a rigorous assessment of the performance differences between the two implementations.

3. Performance Analysis

Repetition	Number fo Tokens	My Code's Elapsed Time (s)	Colleauge Code's Elapsed Time (s)	Difference in Time (s)	Success (My Code)	Success (Colleague Code)
1	3	73.18	72.6	0.58	Yes	Yes
2	3	78.73	74.59	4.14	Yes	Yes
3	3	73.2	71.41	1.79	Yes	Yes
4	3	73.68	69.22	4.46	Yes	Yes
5	3	73.71	70.6	3.11	Yes	Yes
6	3	71.69	72.7	-1.01	Yes	Yes
7	4	107.37	131.67	-24.3	Yes	Yes
8	4	106.82	130.18	-23.36	Yes	Yes
9	4	129.94	130.19	-0.25	Yes	Yes
10	4	108.27	130.19	-21.92	Yes	Yes
11	4	108.83	128.13	-19.3	Yes	Yes
12	4	108.87	129.81	-20.94	Yes	Yes
13	5	148.9	146.81	2.09	Yes	Yes
14	5	137.54	142.81	-5.27	Yes	Yes
15	5	137.43	146,91	-9.48	Yes	Yes
16	5	137.74	141.53	-3.79	Yes	Yes
17	5	137.23	141.01	-3.78	Yes	Yes
18	5	138.26	141.53	-3.27	Yes	Yes
19	6	165.26	179.76	-14.5	Yes	Yes
20	6	167.06	168.26	-1.2	Yes	Yes
21	6	173.36	186.29	-12.93	Yes	Yes
22	6	163.81	174.28	-10.39	Yes	Yes
23	6	162.34	167.27	-4.93	Yes	Yes
24	6	163.8	167.77	-3.97	Yes	Yes
25	7	190.84	211.42	-20.58	Yes	Yes
26	7	192.91	211.8	-18.89	Yes	Yes
27	7	178.88	208.66	-29.78	Yes	Yes
28	7	184.01	211.22	-27.21	Yes	Yes
29	7	172.97	211.24	-38.27	Yes	Yes
30	7	203.09	210.77	-7.68	Yes	Yes

Mean		139.5667	145.0947	-7.21		
Standard Deviation		60.0335	102.185	19.477		
Standard Error		10.96057	18.65634	3.556		
t		12.733	7.78	-2.02		

Table 1: Comparative Performance Analysis of My Code and Colleague's Code in Gathering Golden Boxes

- Key Metrics Comparison:

Table 1 provides a comprehensive overview of the performance comparison between "My Code" and "Colleague's Code" implementations. In addition to mean elapsed time and success rate, the table includes the mean difference, standard deviation, standard error, and t-value, offering a detailed analysis of performance metrics.

- Interpretation of Results:

The computed t-value for the paired T-test was -2.02, calculated by dividing the mean difference between paired observations by the standard error of the differences. With a degree of freedom of 29 and a confidence level of 99.5%, the critical value of 1.699 was obtained from a one-tailed t-distribution table (Table 2), specifically for the given degrees of freedom and confidence level.

- Comparison with Critical Value:

The absolute value of -2.02 was compared with the critical value of 1.699. This critical value represents the threshold beyond which the observed difference in mean elapsed times between "My Code" and "Colleague's Code" implementations is considered statistically significant at the 99.5% confidence level.
If the t-value we computed is larger than 1.699, then the H_0 hypothesis is rejected and H_a hypothesis is accepted. Since the absolute value of the computed t-value (-2.02) exceeded the critical value (1.699), we rejected the null hypothesis, indicating a statistically difference in performance between the two implementations.

t-test 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										

Table 2: T-test Table (One-tail & Two-tail)

4. Conclusion

In this study, we conducted a T-test to compare the performance of "My Code" and "Colleague's Code" implementations in gathering golden boxes in various environments. The analysis yielded several key findings :

- **Statistical Significance:** The computed t-value of -2.02, when compared to the critical value of 1.699 at a 99.5% confidence level, revealed a statistically significant difference in mean elapsed times between the two implementations.
- **Performance Discrepancy:** "My Code" demonstrated superior performance compared to "Colleague's Code," as evidenced by its faster mean elapsed times across multiple repetitions and environments.

- **Implications for Algorithm Optimization:** The observed performance discrepancy suggests potential areas for algorithm optimization and improvement. Further investigation into the specific factors contributing to the performance differences could inform future development efforts.

Overall, this study contributes valuable insights into the comparative performance of different algorithm implementations and underscores the importance of rigorous statistical analysis in algorithm evaluation and optimization effort