

UniGe

RESEARCH TRACK 2

Professor:
Carmine Recchiuto

Author:
Ehsan Fatemi
s5546280

June 1st

Contents

Abstract.....	3
Introduction	3
Experimental Design	3
1. Time Efficiency Analysis	4
1.2. Hypothesis Testing - Time Efficiency.....	5
1.2.1 Null Hypothesis (H0)	5
1.2.2 Alternative Hypothesis (H1)	5
1.3. Calculations - Time Efficiency.....	5
1.4. Results - Time Efficiency	6
2. Success/Failure Rates Analysis.....	7
2.1. Hypothesis Testing - Success/Failure Rates	7
2.1.1 Null Hypothesis (H0)	7
2.1.2 Alternative Hypothesis (H1)	7
2.2. Calculations - Success/Failure Rates	8
2.3 Results - Success/Failure Rates	9
Conclusion.....	9
Time Efficiency	9
Success/Failure Rates.....	9

Statistical Analysis

Abstract

This report aims to compare the time efficiency and success/failure rates between two approaches including my colleague's, which is called 'Amir', and my solution, which is known as 'Mine', in a series of experiments. The data for both factors has been collected and analysed to draw meaningful conclusions. The significance level used for hypothesis testing is 5%.

Introduction

The report aims to evaluate hypothesis and compare the performance of the "Amir" and "Mine" approaches in terms of time efficiency and success/failure rates. The experimental design and data collection methodology are explained, followed by the hypothesis testing process.

Experimental Design

The experiment involves four different map configurations (the tokens were placed on the two circles, not randomly) and two different numbers of tokens. Each configuration was tested five times, resulting in a total of 40 experiments. The execution times for each algorithm are recorded for all the experiments.

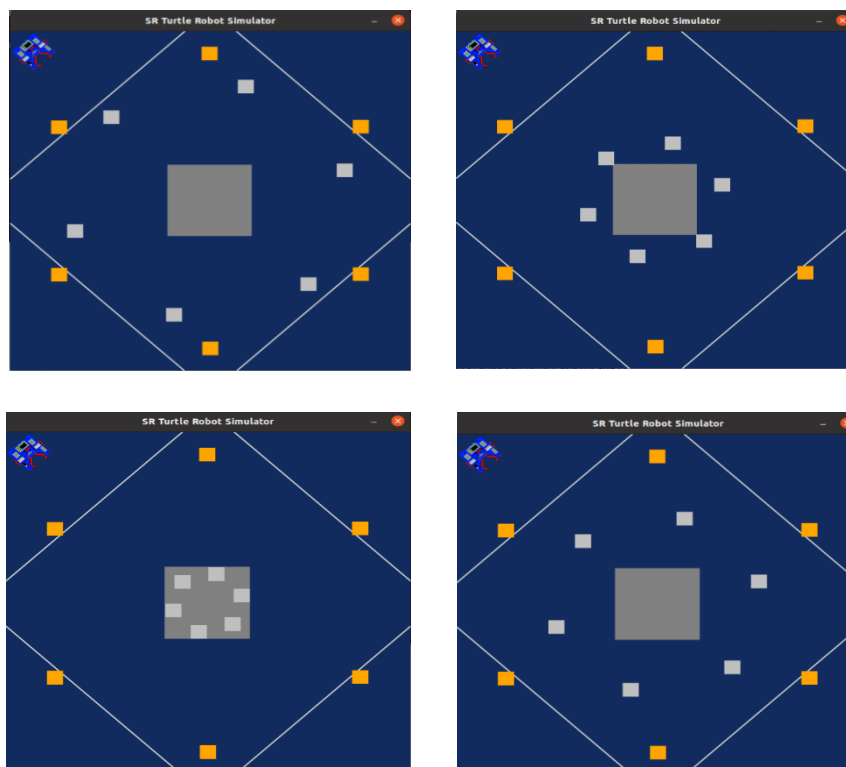


Figure 1 The Map Configuration of Environment

1. Time Efficiency Analysis

For the time efficiency analysis, I have collected data on the time taken by Amir and Mine in multiple experiments. The sample sizes, means, and standard deviations for each performance's time data are as follows:

Statistical parameters:

Amir:

- Sample size (n_1): 40
- Mean time (μ_1): 369.27
- Standard deviation (σ_1): 8.767

Mine:

- Sample size (n_2): 40
- Mean time (μ_2): 294.27
- Standard deviation (σ_2): 8.398

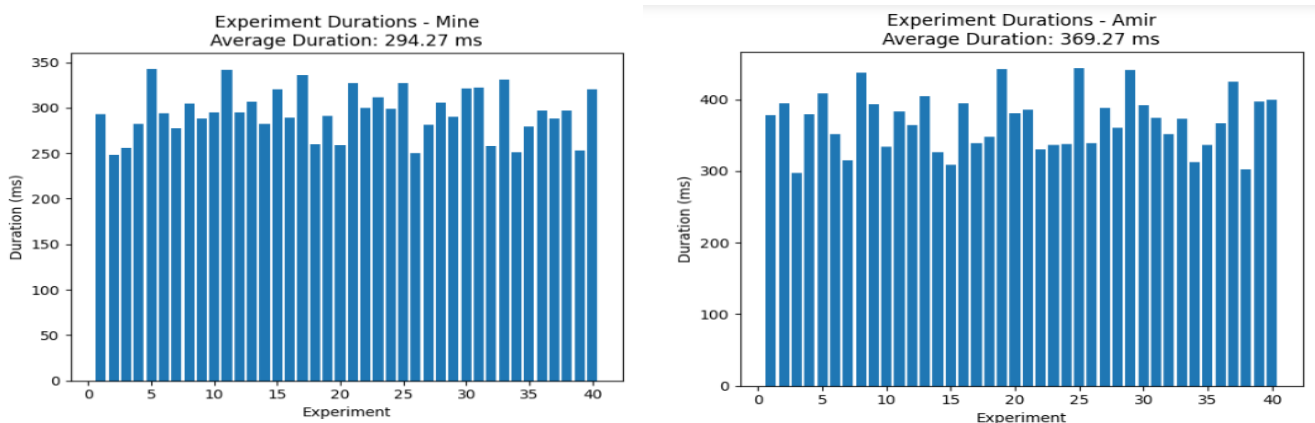


Figure 2 Time Duration of Experiment

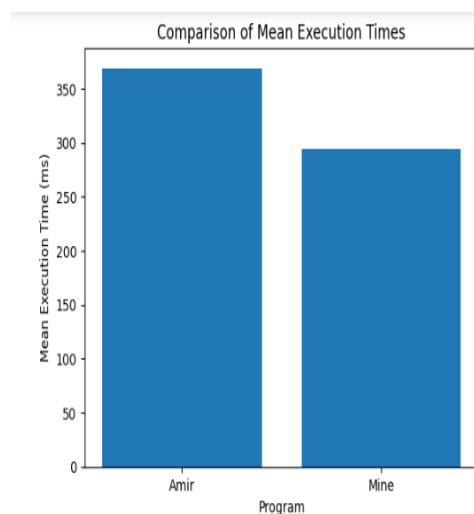


Figure 3 Mean of Time Execution

Using these values, I performed a two-sample t-test to determine if there is a significant difference in the time efficiency between Amir and Mine.

1.2. Hypothesis Testing- Time Efficiency

1.2.1 Null Hypothesis (H0)

There is no significant difference in the time efficiency between ‘Amir’ and ‘Mine’.

1.2.2 Alternative Hypothesis (H1)

There is a significant difference in the time efficiency between ‘Amir’ and ‘Mine’.

To test the hypotheses, I calculated the t-value, degrees of freedom and standard error.

1.3. Calculations- Time Efficiency

1. Calculate the pooled standard deviation:

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\hat{\sigma}_{pooled}^2 \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}$$

- Pooled Standard Deviation (sp) ≈ 8.579

2. Calculate the standard error of the difference:

$$SE(\bar{d}) = \frac{s_d}{\sqrt{n}}$$

- Standard Error ≈ 1.211

3. Calculate the t-value:

- t-value = $(\mu_1 - \mu_2) / \text{Standard Error} = (369.27 - 294.27) / 1.211 \approx 61.004$

4. Determine the degrees of freedom:

- Degrees of Freedom (df) = $n_1 + n_2 - 2 = 40 + 40 - 2 = 78$

5. Compare the calculated t-value with the critical t-value:

- For a significance level of 5% and 78 degrees of freedom, the critical t-value is approximately ± 1.990 (obtained from the t-distribution table)

t Table

cum. prob one-tail two-tails	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
	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 4 t-table

1.4. Results- Time Efficiency

The results of the time efficiency analysis indicate a significant difference between the "Amir" and "Mine" approaches. The calculated t-value (61.004) exceeds the critical t-value (± 1.990), leading to the rejection of the null hypothesis. This finding suggests that the "Mine" approach exhibits superior time efficiency compared to the "Amir" approach.

2. Success/Failure Rates Analysis

In addition to time efficiency, I have also collected data on the success and failure rates for both ‘Amir’ and ‘Mine’ approaches. The counts of successes and failures are as follows:

Statistical parameters:

Amir:

- Successes: 32
- Failures: 8

Mine:

- Successes: 35
- Failures: 5

	Successes	Failures
Amir	32	8
Mine	35	5

Figure 5 Table of Success/Failure rate

Using the gotten results, I performed a hypothesis test to determine if there was a significant difference in success rates between Amir and Mine.

2.1. Hypothesis Testing- Success/Failure Rates

2.1.1 Null Hypothesis (H0)

There is no significant difference in success rates between ‘Amir’ and ‘Mine’.

2.1.2 Alternative Hypothesis (H1)

There is a significant difference in success rates between Amir and Mine.

To test the hypotheses, I calculated the observed success rate difference, standard error, Chi-Square value, and degrees of freedom.

2.2. Calculations- Success/Failure Rates

1. Calculate the success rates:

- Amir's success rate (p1): $32 / (32 + 8) = 0.8$
- Mine's success rate (p2): $35 / (35 + 5) = 0.875$

2. Calculate the observed success rate difference:

- Difference = Mine's success rate - Amir's success rate = $0.875 - 0.8 = 0.075$

3. Calculate the standard error of the difference:

- Standard Error ≈ 0.0828

4. Calculate the Chi-Square value:

- Chi-Square value = $(\text{Difference} / \text{Standard Error})^2 \approx 0.904$

5. Determine the degrees of freedom:

- Degrees of Freedom (df) = $n_1 + n_2 - 2 = 40 + 40 - 2 = 78$

6. Compare the calculated Shi-Square value with the critical Chi-Square value:

- For a significance level of 5% and 78 degree of freedom, the critical chi-square value is approximately 100.93 (obtained from the chi-square distribution table).

Degrees of Freedom	Chi-Square (χ^2) Distribution Area to the Right of Critical Value									
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.173	53.540	57.153	60.201	64.278	96.578	101.879	106.629	112.329	116.321

Figure 6 Chi-Square Value Table

2.3 Results- Success/Failure Rates

Since the calculated chi-square value (0.904) is not greater than the critical chi-square value of 100.93, null hypothesis is rejected. This suggests that there is no significant difference in success rates between 'Amir' and 'Mine' based on the provided data.

Conclusion

Based on the analysis conducted on time efficiency and success/failure rates, the following conclusions can be drawn:

Time Efficiency

There is a significant difference in time efficiency between Amir and Mine. Mine demonstrates significantly faster performance compared to Amir, as indicated by the calculated t-value (61.004).

Success/Failure Rates

There is no significant difference in success rates between Amir and Mine. The calculated chi-square value (0.904) suggests that the success rates of Amir and Mine are not significantly different.

These findings highlight the contrasting performance of Amir and Mine in terms of time efficiency while suggesting no substantial variation in success rates. The results suggest that Mine performs better in terms of time efficiency but does not necessarily lead to higher success rates compared to Amir.