UNIVERSITÀ DEGLI STUDI DI GENOVA

MASTER DEGREE COURSE IN ROBOTICS ENGINEERING

Research Track 2

Statistical Analysis Report

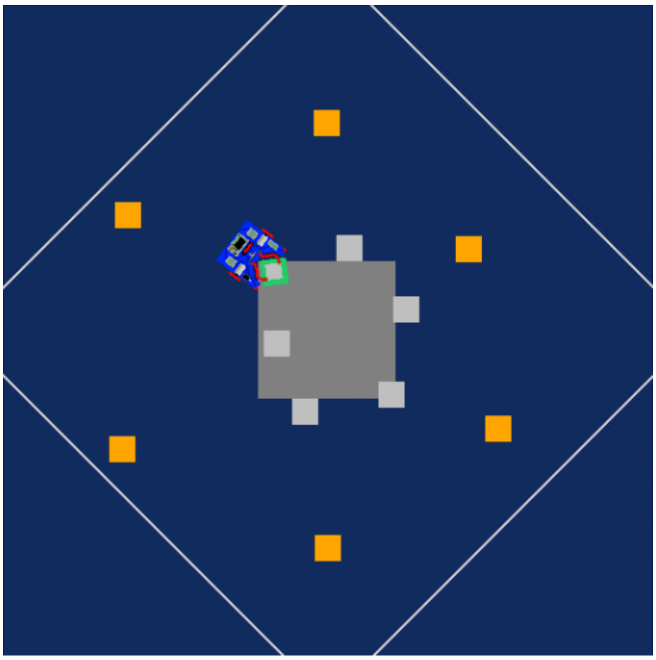
Claudio Demaria S5433737

May, 2023

1 Introduction

This work is the third part of the ***Research Track 2*** course assignment.
In this task we have to perform statistical analysis on the first assignment (related to ***Research Track 1***), considering two different implementations and testing which one performs better, when silver and gold tokens are randomly placed in the environment.
We can choose to consider as performance evaluators what we prefer, even modifying the algorithm's behaves.
The professor stressed the importance of clearly defining the hypothesis we want to test, designing the experiments accordingly, carefully planning the number of experiments, and choosing a suitable statistical approach.
Finally, we have to write a report composed of the following:

- Hypotheses made
- Description and motivation of the experimental setup
- Results
- Discussion of the results with statistical analysis
- Conclusion



2 Description

In this chapter I briefly show the steps I followed to complete the statistical analysis.

- ***Set up the experiment:*** I have to define the specific task or tasks I want to evaluate and choose the metrics to measure. In this case, I want to measure the average time required to finish the tasks, considering a particular situation when the token are placed at a random distance from the center of the environment, but with a fixed angle between each of them: to do that, I modified the *TwoColoursAssignmentArena.py* file, simply putting a random radius for the placement of the tokens in the environment.
- ***Implement the two projects:*** Implement the two different versions of the turtlesim project and instrument these codes to record the time they take to complete each task. To make it easier for the compiler (which runs on docker) to compile, I've added one more small change: I've set a new goal, now the robot has to put only three of the six silver tokens close to the gold tokens, after that the elapsed time is computed and the program ends.
- ***Design the experiment:*** Determine the number of trials or repetitions you want to perform for each implementation (more trials generally lead to more reliable results): I've chosen to run 50 times each project and record the results.
- ***Execute the experiment:*** Run each implementation for the specified number of trials and record the time taken for each trial.
- ***Calculate statistics:*** Compute the mean and the standard deviation of the recorded times for each trial (it is very simple using spreadsheets).
- ***Perform statistical analysis:*** I've chosen to use the t-test, properly described in the next chapter.
- ***Show conclusions:*** Based on the statistical analysis results, I can determine if one implementation performs significantly better or worse than the other.

3 Analysis

3.1 Hypotesis

I started defining my null hypothesis H_0 and alternative hypothesis H_1.
My null hypothesis assumes that there is no significant difference between the two groups of conditions, while the alternative hypothesis suggests that there is a significant difference.

We are in a particular scenario where there are six silver tokens and six gold tokens placed with a constant angular distance from the center of the arena, but with a random distance between the center.
Furthermore, the robot should grab and put only three of the six tokens: the execution time of the task is strictly bound to the placement of the tokens in the environment.
So, finally, I modified the codes to record the execution times.

3.2 Experiments

I've chosen to run 50 times each project and record the results: so, with a little patience, I had the Python script run 50 times, and every time the compiler finished running, the *ElapsedTime* was written directly to a text file contained in the same folder.
Once I had all the execution times, I import the data into a spreadsheet table and I calculated the mean and standard deviation.

I repeated the same operations also for the second version of the turtlesim project.

So, I decided to perform the statistical analysis using a ***T-Test***.

3.3 T-Test

The steps to performing a t-test are simple, and the first things to do involve preparing the hypotheses and data. So initially we need to have well-defined null hypothesesH_0 and alternative hypotheses H_1.
After that, we have to collect the data: calculate the means and standard deviations, and determine the sample size of each group.

The next steps are:

- Compute the ***t-value***. The t-value is calculated using the formula:

$$t = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{1}$$

  where μ_1 and μ_2 are the means of the two groups, sd_1 and sd_2 are the standard deviations of the two groups, and n_1 and n_2 are the sample sizes of the two groups.

- Determine the ***degrees of freedom df***. The degrees of freedom depends on the sample sizes of the two groups and can be calculated as follows:

$$df = n_1 + n_2 - 2 \tag{2}$$

- Determine the ***significance level*** α for your test. This value represents the threshold for determining if the difference between the groups is statistically significant. I choose

$$\alpha = 0.05 \tag{3}$$

- Consult the ***t-distribution table*** to find the ***critical t-value*** corresponding to your desired significance level and degrees of freedom.

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

- ***Make a decision:*** If the calculated t-value is greater than the critical t-value, reject the null hypothesis and conclude that there is a significant difference between the groups. If the calculated t-value is not greater than the critical t-value, fail to reject the null hypothesis.

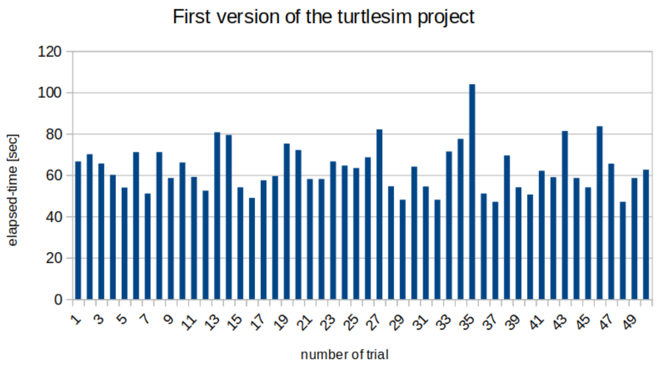
4 Results

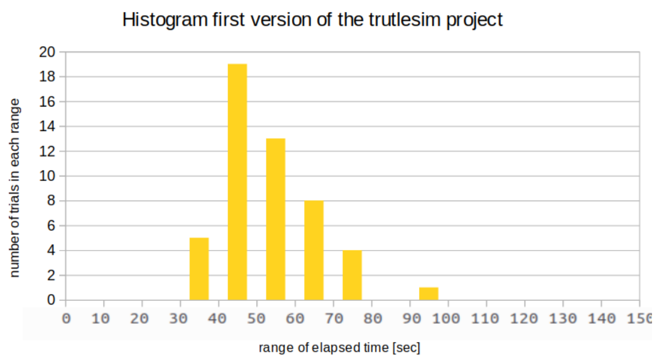
From my experiments, I have imported into my spreadsheet table two datasets of fifty measurements.
I have calculated the mean and standard deviation, and I have printed the data first on a graph, and then on a histogram in order to fully understand the elapsed-time trend.

These are the results:

First version of the turtlesim project:

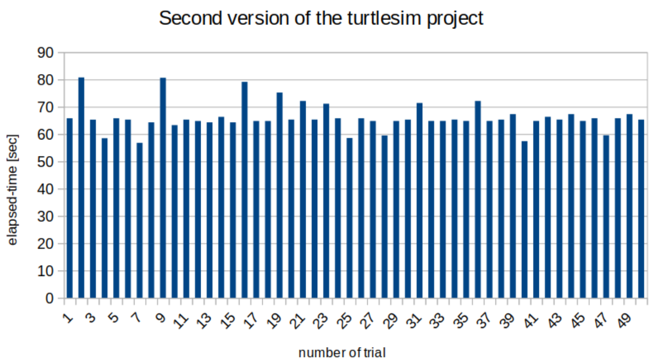
$$\mu_1 = 63,23 \quad and \quad \sigma_1 = 11,57 \tag{4}$$

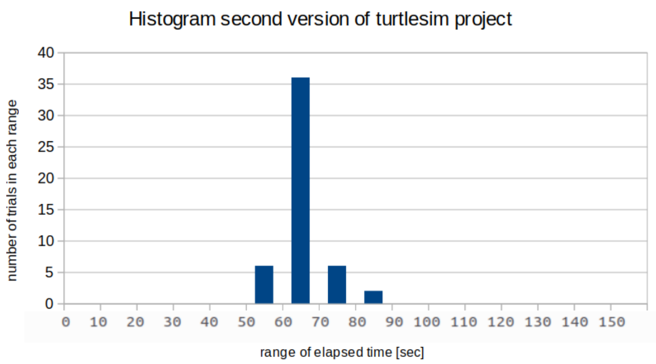




Second version of the turtlesim project:

$$\mu_2 = 66.07 \quad and \quad \sigma_2 = 4.98 \tag{5}$$





Now we can calculate the *t-value* and the *degrees of freedom*:

$$t = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = 1.5933 \tag{6}$$

$$df = n_1 + n_2 - 2 = 98 \tag{7}$$

5 Conclusion

Remembering the ***significance level*** $\alpha = 0.05$ and the ***degrees of freedom*** $df = 98$ we can search the corresponding ***t-value*** in the ***t-table***:

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										

We can see that we obtain a *t-value* $\cong 1.660$, and we can simply notice that:

$$t = 1.5933 \leq 1.660 = t_{value} \tag{8}$$

The result is really interesting.

We have found a t slightly minor than the t_{value}, so this result tells us that we have failed to reject the null hypothesis: we cannot say that there is a significant difference between the groups.

On the other hand, the two values are really similar, and considering that the *degrees of freedom* takes on the *t-table* were slightly less than the real ones (98 instead of 100), and considering the approximation that we must do when we take some measurements, we can conclude that the two values are almost the same.

Another important consideration is the difference between the data of the two versions of the project.
The first set has a mean slightly minor than the second, and in terms of *elapsed-time*, this means that the first version of the project works better than the second on average.
But it's even true that the standard deviation of the first project is very high compared to the second one. This means that the second version is substantially more robust than the first, and we can see these differences even in the graphs on the previous page.