UNIVERSITA DEGLI STUDI DI GENOVA

RESEARCH TRACK II

Third Assignment Report

Fundaments of Statistics

DIBRIS

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

Author

Pisano Davide

May 2023

Contents

1 Introduction

For the third assignment of Research Track 2, our objective was to conduct a statistical analysis on the initial assignment of Research Track 1, focusing on two distinct implementations: my own implementation and that of my colleague, Alessio Mura (s4861320). I will refer to Alessio Mura's implementation as "2" and mine as "1".
The experiments involved modifying the appearance of the arena setup to examine the performance of two different robotic controllers. By comparing the outcomes of my implementation with that of Alessio Mura's, we can obtain valuable insights into the efficiency of various approaches in solving the task. Through statistical analysis, we can identify any significant performance differences between the two implementations.

2 Hypotheses made

The following report aims to examine and evaluate the performance of two distinct programs that solve the same problem. The first step is to compose and formulate the two hypotheses that will be tested and refuted through experiments. The first hypothesis, the null hypothesis, suggests that the two algorithms have similar performances. The second hypothesis, the alternative hypothesis, posits that the first program is faster than the second. This analysis aims to provide an objective and statistical evaluation of the differences between the two algorithms, based on a mathematical approach and data collected during the study.

- **Null hypothesis**:
  The null hypothesis states that there are no significant differences between the two algorithms in terms of performance. This implies that the average execution time is substantially similar for both algorithms. In other words, the null hypothesis assumes that there is no clear preference for either of the two algorithms in terms of speed.

- **Alternative hypothesis**:
  The alternative hypothesis suggests that the first algorithm exhibits greater speed compared to the second algorithm. This implies that the average execution time of the first algorithm is lower than that of the second algorithm. In short, the alternative hypothesis argues that there is a significant speed advantage for the first algorithm over the second algorithm.

3 Description and motivation of the experimental setup

The experiments conducted for the statistical analysis focused on measuring the average time required to complete the silver-gold token pairing task. To ensure a robust analysis, 30 simulations were performed for each token assignment. Each simulation used a different random seed parameter, resulting in unique token placements and environmental setups. In this comparative analysis, a representative sample of data (consisting of 30 numbers) was selected, and both algorithms were iterated multiple times on the same dataset. To introduce variation in the simulations, the angle-offset parameter was modified, impacting the arrangement of the golden tokens on the outer concentric circle.

Here are the execution times of the assignments that I have chosen to examine. The first column concerns the code that I have implemented, the second concerns the one implemented by Alessio Mura.

Angle-Offset	Execution time 1	Execution time 2
value 1	124.152205944	123.747475147
value 2	159.195235014	116.804877996
value 3	164.201638937	123.395332098
value 4	221.377575874	233.639571905
value 5	162.199706078	123.160826921
value 6	149.176033974	120.351647854
value 7	132.163321972	118.255588055
value 8	161.22991395	120.699299097
value 9	121.148519039	122.128663063
value 10	147.165704012	117.783178091
value 11	179.233612061	116.354319096
value 12	169.184362888	125.264173031
value 13	181.251331806	125.269979
value 14	117.128409147	119.554562092
value 15	174.235455036	121.574173927
value 16	169.245977879	122.968338013
value 17	168.253568172	172.331265926
value 18	143.16601181	127.462956905
value 19	199.246370792	117.754606962
value 20	150.157207012	121.04955101
value 21	114.144917011	124.151800156
value 22	179.222509146	122.35408783
value 23	125.145918131	124.468497038
value 24	197.22224021	121.451179028
value 25	147.151827812	125.957015038
value 26	139.158722878	120.75873518
value 27	134.208396196	125.157833099
value 28	162.207538128	124.248339891
value 29	201.222958088	124.968111992
value 30	164.163652897	119.761034012

Table 3.1: Different angle-offset value configuration with correspond execution times.

4 Results and Discussion with statistical analysis

As mentioned earlier, the experiments focused on measuring the average time needed to complete the task. To conduct the analysis, we utilized Microsoft Excel. Once the time data was collected, I calculated the mean value (x) by dividing the sum of the observed values by the number of observations (n). The formula used in both cases (Davide: 1, Alessio: 2) is as follows:

$$x_1 = \frac{\sum_{n=1}^{N_1} X_i}{N_1} = 158{,}5653613965 \qquad x_2 = \frac{\sum_{n=1}^{N_1} X_i}{N_2} = 127{,}4275673151$$

Then I also computed the standard deviation. The standard deviation gives an idea of how close the

entire set of data is to the average value. The formula in the two cases is:

$$\sigma_1 = \sqrt{\frac{\sum_{n=1}^{N_1}(X_i-\mu_1)^2}{N_1}} = 26{,}6915511 \qquad \sigma_2 = \sqrt{\frac{\sum_{n=1}^{N_1}(X_i-\mu_2)^2}{N_2}} = 22{,}24373644$$

where N1, N2 = 30 and Xi are the values of the times.
Now we proceed by conduct a statistical analysis, such as the two-sample T-test, to determine if this observed difference is statistically significant. It is important to perform the statistical analysis to obtain a more accurate conclusion about the equality or difference between the two implementations. I have also calculated the pooled variance, which is:

$$\sigma^2_{pooled} = \frac{(N_1-1)\cdot s_1+(N_2-1)\cdot s_2}{N_1+N_2-2} = 645{,}2397249$$

The pooled variance leads to the pooled, estimated Standard Error of the sampling distribution of the difference of means. The formula is:

$$\sigma_{\bar{x_1}-\bar{x_2}} = \sqrt{\frac{\sigma^2_{pooled}}{N_1} + \frac{\sigma^2_{pooled}}{N_2}} = 6{,}558657001$$

We are interested in the differences, thus the t-statistics turns into:

$$t^2_{\bar{x_1}-\bar{x_2}} = \frac{x_1-x_2}{\sigma_{x_1-x_2}} = 4{,}747586903$$

5 Conclusion

Knowing that there are 30 degrees of freedom, we can calculate the critical value of the t-score and determine if the observed difference is statistically significant.

Since we have a large enough sample (N = 30) to perform the t-test safely and the data appear to approximate a normal distribution, we can use a two-tailed t-test to determine statistical significance. Considering 30 degrees of freedom, the critical value of the t score for a significance level of 5% ($\alpha = 0.05$) is approximately ± 2.042 (these data are taken from the T-test table).

The calculated t-score in your case is 4.747586903. Since the absolute value of the t-score is greater than 2.042, we can conclude that the observed difference between the two implementations is statistically significant. Thus, based on the data provided in the report, we can reject the null hypothesis in favor of the alternative hypothesis, which suggests that the first algorithm is significantly faster than the second algorithm.