UNIVERSITÀ DEGLI STUDI DI GENOVA

DIBRIS

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

RESEARCH TRACK II

ASSIGNMENT

Statistical Analysis

Author:

Pablo Moreno Moreno

Student ID:

s5646698

May 23, 2023

Contents

1 Introduction

This report corresponds to the final part of the assignment of Research Track II. We will perform a statistical analysis on the first assignment of Research Track I, considering two different implementations: mine and the solution of one of my colleagues (Miguel Ángel Sempere Vicente - S5646720). From now on, we will call my solution as Implementation A and the solution of my colleague as Implementation B. The Python scripts of both implementations can be found by clicking on the previous links.

We remember that the algorithm consists on a robot that has to pair silver tokens with golden tokens, which are initially spread over the environment. Once the robot has paired all the tokens, the task is considered as completed.

2 Hypotheses made

First of all, we will clearly define what are the hypotheses we want to test:

Null hypothesis (H_0): *"Implementation A performs the task of pairing 6 pairs of tokens randomly placed in the environment as fast as Implementation B performs the same task".*

$$H_0 : \mu_A = \mu_B$$

Alternative hypothesis (H_a): *"Implementation A does not perform the task of pairing 6 pairs of tokens randomly placed in the environment as fast as Implementation B performs the same task".*

$$H_a : \mu_A \neq \mu_B$$

The conclusion of the statistical analysis will be the acceptance of one of the previous hypothesis and, hence, the rejection of the other.

3 Test selection

We will take 30 measurements of the time that each implementation takes to complete the task. Therefore, according to the Central Limit Theorem, we can consider that the samples are taken from a normal distribution and we will be able to apply a parametric test.

Among the different types of parametric tests, we decided to select the **two sample T-test**. We made this decision because it is one of the most frequently used tests in science to compare how two different independent groups differ from each other, and we will be able to determine all the information that is required to perform it.

We cannot perform a Z-test or a one sample T-test because we are not comparing the statistics of our implementations with any known value (we have no information about the mean or the variance of the population distribution). On the other hand, the reason why we did not choose a paired T-test is because the tokens will be placed randomly in the environment, differently for each implementation, so it will not be possible to pair observations of both samples with each other.

Regarding the significance level, we will stick to the standard value $\alpha = 0.05$. Since the tokens will be placed randomly in the environment, there will be a considerable variability in the measurements within samples, so we decided not to adopt a lower value such as $\alpha = 0.01$. We remember that selecting a significance level of $\alpha = 0.05$ implies a 5% chance of rejecting the null hypothesis when it happens to be true.

4 Experiment

The experiment will consist in measuring the time that each implementation takes to complete the task of pairing the 6 pairs of tokens randomly placed in the environment. In order to initialize the position of the tokens randomly over the environment, we have modified the file:

sim/sr/robot/arenas/two_colours_assignment_arena.py

using the line

token.location = (random.uniform(−2.5, 2.5), random.uniform(−2.5, 2.5))

to set the initial location of the tokens.

On the other hand, to measure the time that it takes for each implementation to complete the task in different trials, we have measured the elapsed time since the program starts until all the tokens are paired.

As mentioned earlier, we decided to take 30 measurements with each implementation, so that we can perform a parametric test. In Figure 1, we can see the time measured in each trial, for both implementations, in seconds.

Implementation A	Implementation B
138.3582582	102.84025
99.28670406	103.2253928
106.2572417	115.414563
103.3192282	106.2078209
107.9485159	97.12596011
103.3635683	111.513813
135.4140859	115.292906
101.2744823	113.6530721
110.2705941	98.90742207
100.2597423	90.30225992
120.292026	129.040606
109.242312	118.453088
100.2198262	101.0282849
164.4395938	107.356695
85.26940584	99.542781
113.2568579	120.0297479
96.24014378	93.60954795
148.0291754	116.9098241
102.3350814	109.2150969
85.23550987	104.5685454
132.1097526	97.3875486
142.0876254	119.3586357
97.93440296	91.2207548
112.0294703	94.3616864
153.0029375	111.2742795
127.2860022	98.4075087
97.43131502	123.3697544
118.1660297	115.864686
142.1397446	93.60392502
106.289782	107.4757946

Figure 1: Results of the experiment: time (in seconds) measured in each trial for both implementations

From the previous data, we have computed the average and the standard deviation of the measurements for each implementation, obtaining the values shown in Figure 2.

	Implementation A	Implementation B
Average	$\bar{x}_A$ = 115.29	σ_A = 106.86
Standard deviation	$\bar{x}_B$ = 20.49	σ_B = 10.37

Figure 2: Average and standard deviation of the measurements for each implementation

5 Two sample T-test

Now, we proceed to perform the selected test: the two sample T-test. First, we compute the pooled, estimated variance of the sampling distribution of the difference of the means, according to the following expression:

$$\hat{\sigma}^2_{pooled} = \frac{(N_A - 1)\sigma_A^2 + (N_B - 1)\sigma_B^2}{N_A + N_B - 2}$$

where N_A and N_B are the number of measurements taken with each implementation, that is, 30 for both implementations. We obtain a value of $\hat{\sigma}^2_{pooled} = 263.83$.

Then, we compute the pooled, estimated standard deviation of the sampling distribution of the difference of means, using the following expression:

$$\hat{\sigma}_{\bar{x}_A - \bar{x}_B} = \sqrt{\hat{\sigma}^2_{pooled}(\frac{1}{N_A} + \frac{1}{N_B})}$$

obtaining a value of $\hat{\sigma}_{\bar{x}_A - \bar{x}_B} = 4.194$.

Finally, we compute the t-statistics using the following expression:

$$t = \frac{\bar{x}_A - \bar{x}_B}{\hat{\sigma}_{\bar{x}_A - \bar{x}_B}}$$

obtaining a value of $t = 2.0047$.

On the other hand, using the table of the T-distribution, we determine the maximum value of the t-statistics that we could have to achieve a significance level of 95%.

We remember that, since our hypothesis is of type $H_0 : \mu_A = \mu_B$, we are using a two-tailed test. So, entering the table with a degree of freedom of $DOF = N_A + N_B - 2 = 58$, and a significance level of 95% for a two-tailed test, we get a value of $t' = 2.0000$.

6 Conclusion

Since the computed t-statistics t is greater than the maximum value t' we could have to achieve the significance level of 95%, we have to reject the null hypothesis H_0 and accept the alternative hypothesis H_a. This means that the difference between both groups is considered to be statistically significant and, hence, we conclude that Implementation A does not perform the specified task as fast as Implementation B does.

Finally, it is worth mentioning that we are committing an error of 5% when we state this conclusion. That is, there is a 5% chance that H_0 is in fact true, even if we are concluding that it is false, due to the fact that we are considering a level of significance of 95%.