Research Track 2
Report assignment 1 Part 3

Gabriele Nicchiarelli

25 May 2023

Contents

1 Introduction

In the third part of the assignment of the Research Track 2 course, we were asked to make a statistical analysis on the first assignment of the past course, Research Track 1.
The research consisted in collecting data from two different implementations of the assignment and making a statistical comparison of the two. For this experiment, my own implementation and my colleague's Pisano Davide implementation were considered. In particular, we want to observe the statistical difference in terms of elapsed time to finish the task.
The task considered for our analysis is similar to the past assignment: collect all the silver tokens in the environment and place each silver token near a gold token, but this time the gold tokens are randomly placed in the environment. In particular, they are placed varying the disposition angle in a random manner.

2 Hypotheses

In statistics, the "null hypothesis", H_0, is the assumption that there is no significant difference between the two algorithms. On the other hand, the "alternative hypothesis", H_a, states that there is a significant difference between the two implementations. Thus H_0 and H_a are mutually exclusive, accepting one of them means rejecting the other hypothesis.
For our analysis, we started considering the null hypothesis, which states that $\mu_{my\_assignment} = \mu_{other\_assignment}$. Our goal is to reject this hypothesis by demonstrating one of the two algorithms is better than the other. Thus we will try to prove H_a by rejecting H_0.
Using the measured time as comparing variable, we want to see if one algorithm performs "better" than the other performing a *T-test*. If the *t-value* that we obtain from calculations is greater than a certain critical t-value, then we can reject H_0 and accept H_a.

3 Methodology

3.1 Data acquisition

First, we retrieved the data executing both algorithms *30* times and registered the elapsed time (in seconds) required to finish the task in the *results.csv* file. It is important to note that both algorithms sometimes fail, thus the number of total executions was actually higher than 30. For simplicity, we already prepared the data in *csv* file in order to store only the valid usable data for the t-test.

3.2 Code implementation

First things first, the file containing the arena (*two_colours_assignment_arena.py*) was modified in order to randomize the angle of the gold tokens. Also, the file containing the implementation of the assignment (*assignment.py*) was slightly modified to allow computing the execution time of the program.
After these preliminary steps, a Jupyter notebook was created (*statistics.ipynb* to plot the Gaussian distribution of the two sets of data and to calculate the t-value necessary for the T-test.
The script first computes the mean ($\bar{x}$) and the standard deviation (σ) for both sets of data. Then, plots the Gaussian distributions shown in Figure 1.

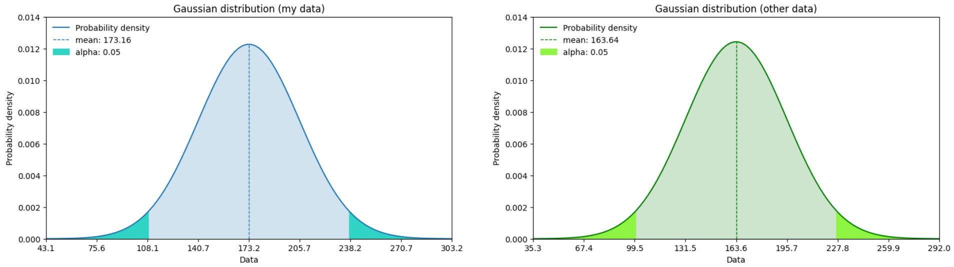


Figure 1: Gaussian distributions

I decided to use a two-tailed test with 95% confidence level, thus choosing a significance level $\alpha = 5\%$. This last value represents the threshold for determining if the difference between the groups is statistically significant.
Then the t-value was computed following these steps:

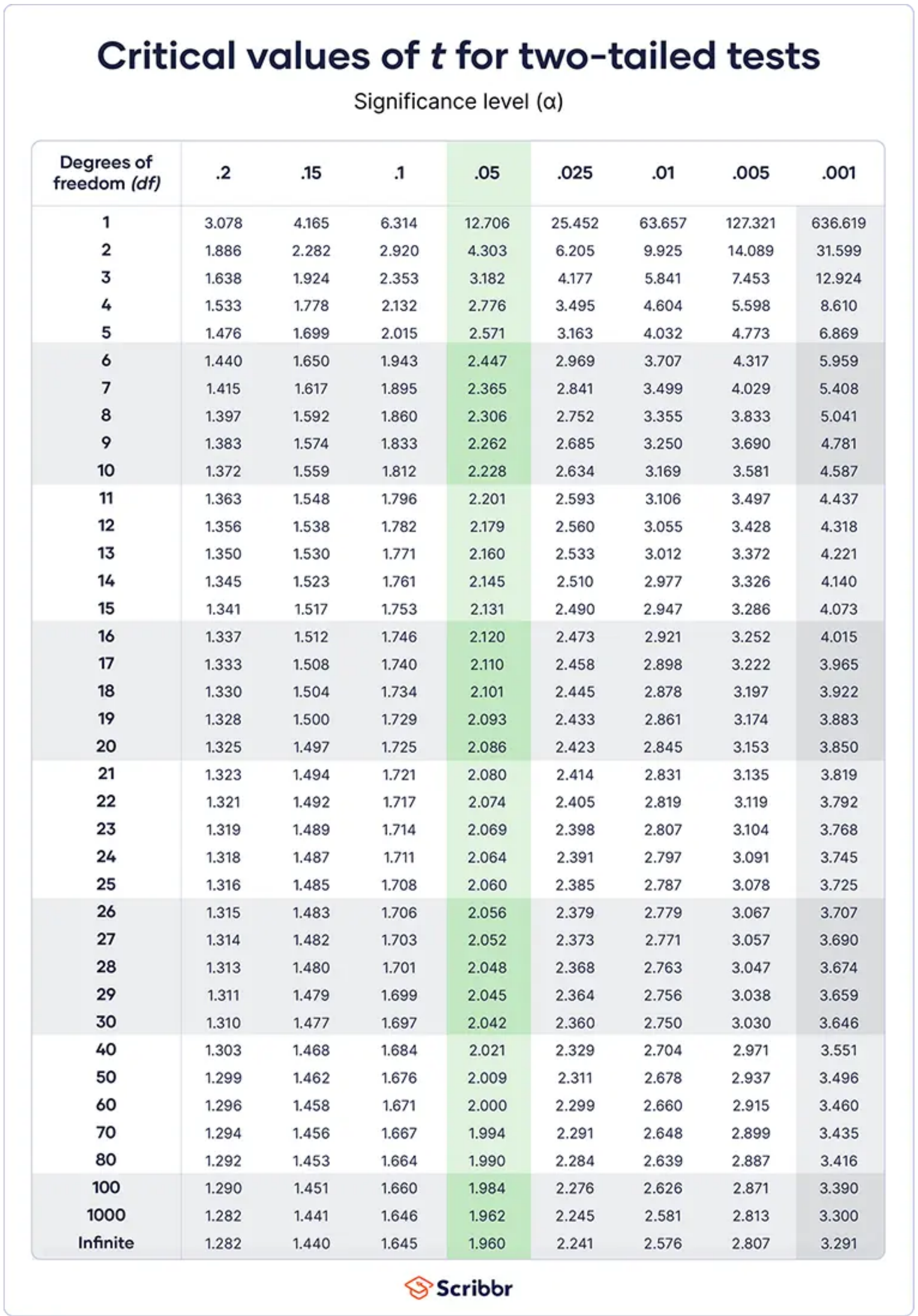
1. Determine the degrees of freedom (dof):
   they depend on the sample size of the two groups and are respectively $N_1 - 1$ and $N_2 - 1$, where N_1 and N_2 are the number of samples in the two groups.
   The total degrees of freedom of the system are:
   $$dof = N_1 + N_2 - 2$$
2. Next step is to compute the pooled variance:
   $$\hat{\sigma}^2_{pooled} = \frac{(N_1 - 1) \cdot \sigma_1^2 + (N_2 - 1) \cdot \sigma_2^2}{N_1 + N_2 - 2}$$
3. Then compute the pooled estimated Standard Error:
   $$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sigma^2_{pooled} \cdot (\frac{1}{N_1} + \frac{1}{N_2})}$$
4. Now we have all the components to compute the t-value:
   $$t_{\bar{x}_1 - \bar{x}_2} = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} = 1.141$$

5. Finally we can compare the value obtained with the critical t-value retrieved from the t-distribution table (Figure 2), based on the degrees of freedom and on the significance level.
   Since we have 58 dof and $\alpha = 0.05$, then our critical t-value is $t_{crit} = 2.001$.
   So we can conclude by stating:

   $$t_{\bar{x}_1 - \bar{x}_2} = 1.141 < t_{crit} = 2.001$$

4 Results

The objective of the t-test is to evaluate if there is a significant difference between the means of the two groups.
Since the computed t-value is less than the critical t-value we can't reject the null hypothesis, thus concluding that the performances of the two algorithms are equivalent. In other words, there isn't a particular difference between the two algorithms, in terms of execution time.
However, it is important to note that even though there aren't significant diversities in the execution time of the two algorithms, it doesn't necessarily imply that they are identical, but for sure we can say that the current data does not provide sufficient evidence to conclude otherwise.



Critical values of *t* for two-tailed tests

Significance level (α)

Degrees of freedom *(df)*	.2	.15	.1	.05	.025	.01	.005	.001
1	3.078	4.165	6.314	12.706	25.452	63.657	127.321	636.619
2	1.886	2.282	2.920	4.303	6.205	9.925	14.089	31.599
3	1.638	1.924	2.353	3.182	4.177	5.841	7.453	12.924
4	1.533	1.778	2.132	2.776	3.495	4.604	5.598	8.610
5	1.476	1.699	2.015	2.571	3.163	4.032	4.773	6.869
6	1.440	1.650	1.943	2.447	2.969	3.707	4.317	5.959
7	1.415	1.617	1.895	2.365	2.841	3.499	4.029	5.408
8	1.397	1.592	1.860	2.306	2.752	3.355	3.833	5.041
9	1.383	1.574	1.833	2.262	2.685	3.250	3.690	4.781
10	1.372	1.559	1.812	2.228	2.634	3.169	3.581	4.587
11	1.363	1.548	1.796	2.201	2.593	3.106	3.497	4.437
12	1.356	1.538	1.782	2.179	2.560	3.055	3.428	4.318
13	1.350	1.530	1.771	2.160	2.533	3.012	3.372	4.221
14	1.345	1.523	1.761	2.145	2.510	2.977	3.326	4.140
15	1.341	1.517	1.753	2.131	2.490	2.947	3.286	4.073
16	1.337	1.512	1.746	2.120	2.473	2.921	3.252	4.015
17	1.333	1.508	1.740	2.110	2.458	2.898	3.222	3.965
18	1.330	1.504	1.734	2.101	2.445	2.878	3.197	3.922
19	1.328	1.500	1.729	2.093	2.433	2.861	3.174	3.883
20	1.325	1.497	1.725	2.086	2.423	2.845	3.153	3.850
21	1.323	1.494	1.721	2.080	2.414	2.831	3.135	3.819
22	1.321	1.492	1.717	2.074	2.405	2.819	3.119	3.792
23	1.319	1.489	1.714	2.069	2.398	2.807	3.104	3.768
24	1.318	1.487	1.711	2.064	2.391	2.797	3.091	3.745
25	1.316	1.485	1.708	2.060	2.385	2.787	3.078	3.725
26	1.315	1.483	1.706	2.056	2.379	2.779	3.067	3.707
27	1.314	1.482	1.703	2.052	2.373	2.771	3.057	3.690
28	1.313	1.480	1.701	2.048	2.368	2.763	3.047	3.674
29	1.311	1.479	1.699	2.045	2.364	2.756	3.038	3.659
30	1.310	1.477	1.697	2.042	2.360	2.750	3.030	3.646
40	1.303	1.468	1.684	2.021	2.329	2.704	2.971	3.551
50	1.299	1.462	1.676	2.009	2.311	2.678	2.937	3.496
60	1.296	1.458	1.671	2.000	2.299	2.660	2.915	3.460
70	1.294	1.456	1.667	1.994	2.291	2.648	2.899	3.435
80	1.292	1.453	1.664	1.990	2.284	2.639	2.887	3.416
100	1.290	1.451	1.660	1.984	2.276	2.626	2.871	3.390
1000	1.282	1.441	1.646	1.962	2.245	2.581	2.813	3.300
Infinite	1.282	1.440	1.645	1.960	2.241	2.576	2.807	3.291

Scribbr

Figure 2: t-distribution table