Statistical Analysis on the 1st Assignment of RT1

In this document I will report a statistical analysis of the 1st assignment of the Research Track 1 course, which basically was about moving a mobile robot around a circuit.
It had to avoid certain objects (golden boxes) and interact with some other (silver boxes).

I will compare a given algorithm **(N. 1)** and the one I developed **(N. 2),** measuring the **time required to complete one lap** in such conditions.

---

1. Samples design and values

To properly compare the two algorithms, I have considered **30 different scenarios** where what changes is the **position and number of obstacles**, and took the duration time of a single lap, in seconds unit.

Sample size for both methods is N = 30; thanks to the central limit theorem, that assures a normal distribution when sample size is large and at least N = 30, we can assume that our sample comes from a normally distributed population, without the need of applying a normality test.

The samples values are shown in a table at the end of the document.

---

2. Paired T-Test

A paired t-test is used to compare two population means where there are two samples, in which observations in one sample can be paired with observations in the other sample, as it is in our case.

We want to **test the null hypothesis**, sustaining that the mean difference of durations between alg.1 and alg.2 is equal or less than 0, with a **confidence level of 99%**: it is a **one-tail** test.

- $H_0 : \mu_1 - \mu_2 \leq 0$
- $H_1 : \mu_1 - \mu_2 > 0$

First step is to calculate the difference between each pair of observations and get the **mean of the differences**. Then we must calculate the **standard deviation of the differences** and, from that, the **standard error of the mean difference.**

$$\bar{d} = 69.53 \qquad s_d = 4.73 \qquad SE(\bar{d}) = \frac{s_d}{\sqrt{n}} = 0.86$$

Finally, we can get the **t-statistic** and compare it to the $\boldsymbol{t_{n-1}}$ **distribution** with **29 DoF** (sample size -1)**,** using the **t-table** and a **confidence level of 99%.** (We get $\boldsymbol{t_{value} = 2.462}$)

$$T = \frac{\bar{d}}{SE(\bar{d})} = 80.43 \gg 2.462$$

This means that we can reject the null hypothesis, with a confidence level of 99%, and accept the alternative hypothesis: we can sustain that the second implementation performs better, as it takes less time to complete a lap.

We can also construct a 99% confidence interval in which the mean difference of any pair of samples will lay, where the margin of error is $\pm\, t_{value} \cdot SE(\bar{d}) = 2{,}18$.

$$\bar{d} \pm \boldsymbol{t_{value}} \cdot SE(\bar{d}) = \bar{d} \pm 2{,}18 = \quad [67{,}35 \;\; ; \;\; 71{,}71]$$

Case NO.	Time required (seconds)		DIFFERENCE
	Algorithm 1	Algorithm 2	
1	140	63	77
2	141	70	71
3	143	65	78
4	152	79	73
5	154	73	81
6	155	92	63
7	159	82	77
8	160	86	74
9	161	91	70
10	162	91	71
11	163	98	65
12	163	100	63
13	165	97	68
14	166	100	66
15	167	103	64
16	168	99	69
17	170	101	69
18	173	106	67
19	179	115	64
20	181	113	68
21	182	110	72
22	186	118	68
23	189	118	71
24	192	124	68
25	203	131	72
26	209	147	62
27	213	144	69
28	219	151	68
29	223	150	73
30	228	154	74
MEAN	178.03	108.5	69.53