

DATA ANALYTIC ASSIGNMENT

- 1.) A football coach is frustrated with his team's lack of speed. He measures each player's 30-yard dash speed and then sends all of them to a speed and agility camp. He then measures their times again after. The data is below. Is there sufficient evidence to say that the camp helped the players speed?

Table :

Before	After
7.95	8.45
6.92	7.66
8.58	8.97
6.44	6.82
3.76	3.88
9.23	9.53
7.09	7.43
6.55	6.76
8.34	8.08
5.53	5.02
7.08	6.46
6.42	6.17
5.02	4.67
7.67	6.98
5.41	5.87

Is there evidence that the team gets significantly better performance after camp? Use alpha 0.05 level of significance.

a. Write an appropriate hypothesis test for this situation and state the appropriate testing procedure.

Answer :

Hypothesis test →

Null hypothesis → There's no significant difference between player's 30-yard dash speed before and after. ($\mu_1 = \mu_2$)

Alternative Hypothesis → There is a significant difference between player's 30-yard dash speed before and after. ($\mu_1 \neq \mu_2$)

Testing procedure →

Paired T-test → Hence, these two columns of player's speed are related to each other and the population standard deviation is unknown for this data sample. Thus, we consider this as a Paired T-test.

b. Compute the necessary summary statistics for the test in part (a).

Answer :

Before	After	Difference (d)
7.95	8.45	-0.50
6.92	7.66	-0.74
8.58	8.97	-0.39
6.44	6.82	-0.38
3.76	3.88	-0.12
9.23	9.53	-0.30
7.09	7.43	-0.34
6.55	6.76	-0.21
8.34	8.08	0.26
5.53	5.02	0.51
7.08	6.46	0.62
6.42	6.17	0.25
5.02	4.67	0.35

7.67	6.98	0.69
5.41	5.87	-0.46

Difference (d) :

$\Sigma d = (-0.50) + (-0.74) + (-0.39) + (-0.38) + (-0.12) + (-0.30) + (-0.34) + (-0.21) + (0.26) + (0.51) + (0.62) + (0.25) + (0.35) + (0.69) + (-0.46)$

$\Sigma d = -0.77$

Standard deviation of the differences (sd):

Before	After	Difference (d)	d ²
7.95	8.45	-0.50	0.25
6.92	7.66	-0.74	0.5476
8.58	8.97	-0.39	0.1521
6.44	6.82	-0.38	0.1444
3.76	3.88	-0.12	0.0144
9.23	9.53	-0.30	0.09
7.09	7.43	-0.34	0.1156
6.55	6.76	-0.21	0.0441
8.34	8.08	0.26	0.0676
5.53	5.02	0.51	0.2601
7.08	6.46	0.62	0.3844
6.42	6.17	0.25	0.0625
5.02	4.67	0.35	0.1225
7.67	6.98	0.69	0.4761
5.41	5.87	-0.46	0.2116

--	--	--	--

$$\Sigma d^2 = (-0.50)^2 + (-0.74)^2 + (-0.39)^2 + (-0.38)^2 + (-0.12)^2 + (-0.30)^2 + (-0.34)^2 + (-0.21)^2 + (0.26)^2 + (0.51)^2 + (0.62)^2 + (0.25)^2 + (0.35)^2 + (0.69)^2 + (-0.46)^2$$

$$\Sigma d^2 = 2.9531$$

T- score :

To calculate , **t-score** using the formula ,

$$t = \frac{\Sigma d}{\sqrt{\frac{n(\Sigma d^2) - (\Sigma d)^2}{n-1}}}$$

We have ,

$$\Sigma d = -0.77$$

$$\Sigma d^2 = 2.9531$$

$$df = n - 1 = 15 - 1 = 14$$

Now,

$$t = -0.77 / \text{SQRT} (15 (2.9531) - (-1.75)^2) / 14$$

$$t = -0.77 / (41.027) / 14$$

$$t = -0.77 / 2.93$$

$$t \approx -0.458$$

According to the t-table we have got ,

$$T \text{ critical} = 1.769$$

Hence ,

the t-statistical is less the t - critical ($-0.458 < 1.769$) .Thus , ***we fail to reject the null hypothesis (h0).***

C. Perform the t-test and report the p-value (For Excel and IBM SPSS)

→ Google sheets :

The screenshot shows a Google Sheets spreadsheet titled "Untitled spreadsheet". The spreadsheet contains two columns, "Before" and "After", with 10 rows of data. A formula in cell E3 calculates the t-test result: $\text{using T.TEST} = 0.3365788491$. To the right of the spreadsheet, a Google account overlay is visible, showing the user's profile picture, name "Hi, Hari!", and options to "Manage your Google Account", "Add account", or "Sign out". The overlay also indicates that 20% of 15 GB is used.

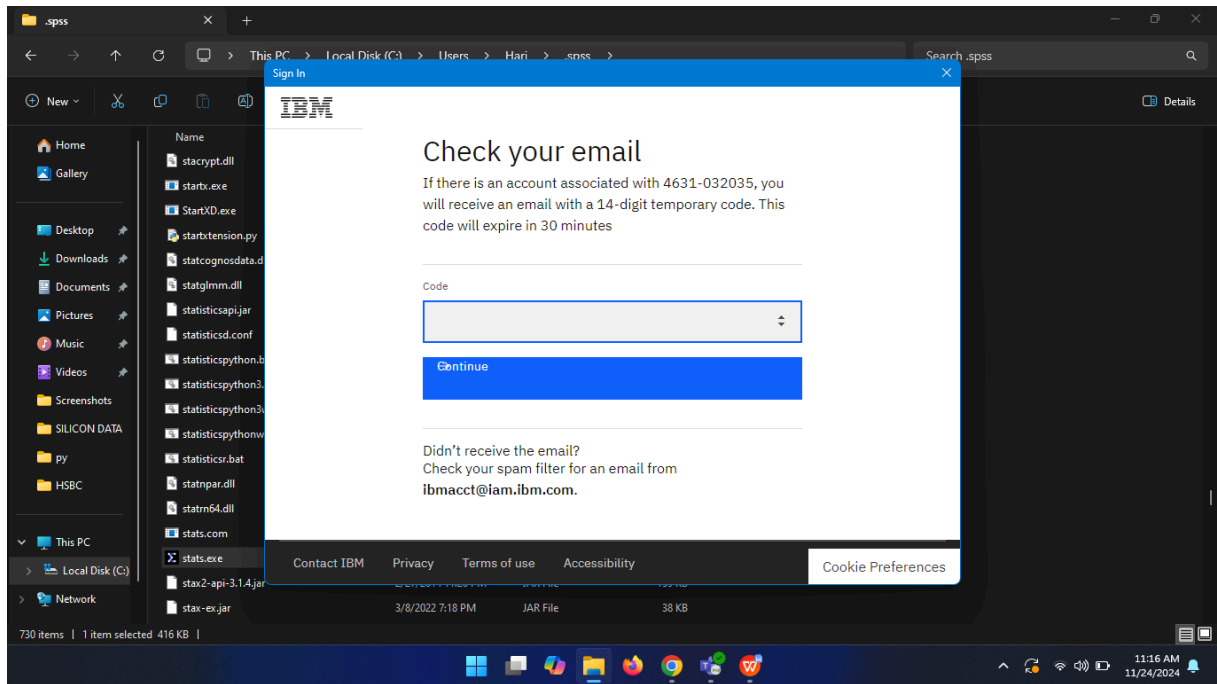
	Before	After
3		
4	7.95	8.45
5	6.92	7.66
6	8.58	8.97
7	6.44	6.82
8	3.76	3.88
9	9.23	9.53
10	7.09	7.43
11	6.55	6.76
12	8.34	8.08
13	5.53	5.02
14	7.08	6.46
15	6.42	6.17
16	5.02	4.67
17	7.67	6.98
18	5.41	5.87

	Variable 1	Variable 2
Mean	6.799333333	6.85
Variance	2.142335238	2.518942857
Observations	15	15
Pearson Correla	0.9586260885	
Hypothesized M	0	
df	14	
t Stat	-0.4308209248	
P(T<=t) one-tail	0.3365788487	
t Critical one-tail	1.761310115	
P(T<=t) two-tail	0.6731576974	
t Critical two-tail	2.144786681	

As I have used **google sheets** instead of **excel** (i don't have a subscription) , as proof of the genuine work i have to pop the account details of mine in the right corner . kindly , consider this .

→ IBM spss :

The screenshot shows the IBM account creation page. The page title is "Welcome to IBM" and the subtitle is "Create an account to access demos and services." Below the title, there is a section "Create an IBMid" with a link "Already have an IBM account? Log in". The "Account information" section is expanded, showing a "Verify email" step. A message states: "We emailed a 7 digit code to hariharan.chita@ue-germany.de. This code will expire in 30 minutes." Below this, there is a "Verification token" field with a red border and a red exclamation mark icon, indicating an error. A small overlay window in the top right corner shows a green checkmark and the text: "Email sent! We sent a verification code to hariharan.chita@ue-germany.de. Timestamp [11:11:19 AM]". The background shows a Windows file explorer window with various files and folders.



After the mail verification , I've got my IDMid . But , since IBM is not sending me any mail for confirmation code to access further . I can't login to IBM spss and work on it (i've tried it more than 10 times) . Kindly, consider this exception .

d. Interpret your results in the conclusion

P-value = 0.3365788491

Significant level = 0.05

Since, the p-value is greater than the significant level ($0.3365788491 > 0.05$) . we **fail to reject the null hypothesis (h_0)** . That means we have **not enough proof to conclude the speed of players from the camp** .

2.) A teacher wants to test the effectiveness of a new textbook. She believes that this new textbook is easier to read, and that her students should have better grades on their tests this year than they have in the past. She took a random sample of test scores from last year's classes, and then a random sample of test scores from this year's classes. Assume normal populations for both years. Test her theory at $\alpha = 0.05$.

Old book					New book				
7 9	7 8	9 2	8 9	8 2	7 8	8 7	8 6	9 0	8 5
8 8	8 5	8 3	8 8	7 0	8 4	8 8	8 8	7 9	7 5
6 5	7 6	7 2	9 0	8 0	6 8	7 1	6 6	9 4	7 3

Answer :

The screenshot shows a Google Sheets spreadsheet titled 'Untitled spreadsheet'. It contains two columns of data: 'Old book' and 'New book'. The 'Old book' column has values: 79, 88, 65, 78, 85, 76, 92, 83, 88, 70, 80, 82. The 'New book' column has values: 86, 88, 66, 87, 88, 71, 86, 90, 85, 75, 73, 73. Below the data, there is a section for 'Using levene's test' with the formula `F.TEST = 0.7611863221`. A note states: 'Thus, the p-value is greater than the significant level we confirm this is equal variance.' To the right, there is a section for 't-Test: Two-Sample Assuming Equal Variances' with the following results:

	Variable 1	Variable 2
Mean	81.13333333	80.8
Variance	62.98095238	74.31428571
Observations	15	15
Pooled Variance	68.64761905	
Hypothesized Mean Difference	0	
df	28	
t Stat	0.1101784447	
P(T<=t) one-tail	0.456526884	
t Critical one-tail	1.701130908	
P(T<=t) two-tail	0.913053768	
t Critical two-tail	2.048407115	

From the given table , I understand that this is a **one- tailed test** . To progress further more ,

Hypothesis test :

Null hypothesis (H0) : There is no difference in the test score of student from last year

Alternative hypothesis (H1) : The test score of student's is higher than last year.

1. I have done a '**=F.TEST**' method to perform a **levене's test** to understand the data variance .
2. Since , from the test my p-value is **greater than** my significant level . I confirm that this is an **equal variance** .
3. Henceforth , I performed a **t-test : two-sample assuming equal variance t-test** using google sheets to obtain the final result.

P-value = 0.456526884

Since the **p-value is greater than the significant level (0.456526884 > 0.05) , we fail to reject the null hypothesis.** There is not enough proof to prove that the new textbook is effective .

3.) The Chapin Social Insight Test is a psychological test designed to measure how accurately a person appraises other people. The possible scores on the test range from 0 to 41. During the development of the test, it was given to several groups of people. Here are the results for male and female college students at a liberal arts college: Does the data support the contention that female and male students differ in average social insight? Use 98% confidence to make your conclusion.

n		avg.	std.dev
Male	15	23.78	5.05
Female	15	26.83	5.10

Answer :

Hypothesis test :

Null - Hypothesis (h0) → there is no difference in average total insight score of male and female students .

Alternative - Hypothesis (h1) → there is significant difference in average total insight of male and female students.

Significant level → 0.02

T-statical :

$$t = (x1bar - x2bar) / [(s1^2/n1) + (s2^2/n2)]$$

We have ,

$$X1bar = 23.73$$

$$X2bar = 26.83$$

$$S1 = 5.05$$

$$S2 = 5.10$$

N1 & N2 = 15

Therefore,

$$T = (23.73 - 26.83) / \{\text{SQRT}((5.05^2/15) + (5.10^2/15))\}$$

$$T = -1.6458$$

$$\text{ABS}(t) = 1.6458$$

To calculate degree of freedom (df) we use ,

$$Df = \{[(s1^2/n1) + (s2^2/n2)]^2 / \{[(s1^2/n1)^2/(n1-1) + (s2^2/n2)^2/(n2-1)]\}$$

$$df = \{[1.7005 + 1.744]^2 / [0.203 + 0.211]\}$$

$Df = 28.0792$

To calculate t-critical we use “=T>INV.2T” in sheets ,

We get,

$t = 2.4671432$

To calculate p-value , we use “T>DIST.2T” in sheets,

We get,

$P\text{-value} = 0.11129$

Thus,

T-statical < t-critical (-1.6458 < 2.4671432) and ,

P-value is greater than significant level ($0.11129 > 0.02$)

We conclude that , **we fail to reject the null hypothesis** , there is no difference between average total insight of male and female students .