

Assignment # 1 (Data Analytics)

Name: Turuvekere Satish Sagar

Matriculation No. 26400540

Q1:

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 to improve their speed and run faster? Run a test. (12 marks)

NOTE: Perform this test by all approaches Excel, SPSS, and manually

Before	After
5.08	5.45
3.92	4.08
7.09	7.97
8.77	6.96
5.69	5.88
9.23	8.53
7.44	7.86
8.22	8.76
4.34	4.08
8.53	9.02
6.50	6.99
8.42	8.87
8.33	8.67
2.67	2.98
7.83	7.87

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.

Ans. The appropriate testing procedure for this would be a paired T-test as we are comparing same value after a certain period and after trying to improve performance.

It is a one tailed test as we are checking if the performance has improved.

Null Hypothesis $\rightarrow H_0 - \text{Mean}_{\text{before}} = \text{Mean}_{\text{after}}$

Alternative Hypothesis $\rightarrow H_1 \text{ Mean}_{\text{before}} < \text{Mean}_{\text{after}}$

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

```

1 #Turuvekere Satish Sagar
2 import math
3 before = [5.08,3.92,7.09,8.77,5.69,9.23,7.44,8.22,4.34,8.53,6.5,8.42,8.33,2.67,7.83]
4 after = [5.45,4.08,7.97,6.96,5.88,8.53,7.86,8.76,4.08,9.02,6.99,8.87,8.67,2.98,7.87]
5 diff = []
6 diffsquare = []
7 n = len(before)
8
9 for i in range(n):
10     diffsquare.append((before[i] - after[i])**2)
11     diff.append(before[i] - after[i])
12
13 sum1 = sum(diff)
14 print("The sum of differences",sum1)
15 sum2 = sum(diffsquare)
16 print("The sum of squares of differences",sum2)
17
18 t = sum1/(math.sqrt(((n*sum2) - (sum1**2))/(n-1)))
19
20 print("Test statistic of paired T-test",t)
21
22
23 |

```

PROBLEMS ① OUTPUT DEBUG CONSOLE TERMINAL PORTS Filter Code ▾ ≡ 🔍 ⌂ ⌃ ⌄

```
[Running] python -u "e:\sagar\Study\MTech\DSS\x.py"
The sum of differences -1.90999999999998
The sum of squares of differences 6.17069999999998
Test statistic of paired T-test -0.757907523977007
```

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

SPSS

VAR00 001	VAR00 002	VAR00 003	VAR00004
5.08	5.45	.	Turuvekere Satish Sagar
3.92	4.08	.	
7.09	7.97	.	
8.77	6.96	.	
5.69	5.88	.	
9.23	8.53	.	
7.44	7.86	.	
8.22	8.76	.	
4.34	4.08	.	
8.53	9.02	.	
6.50	6.99	.	
8.42	8.87	.	
8.33	8.67	.	
2.67	2.98	.	
7.83	7.87	.	

➔ Turuvekere Satish Sagar

T-Test

[DataSet1]

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1	VAR00001	6.8040	15	2.01775
	VAR00002	6.9313	15	1.98284

Paired Samples Correlations

N	Correlation	Significance	
		One-Sided p	Two-Sided p
Pair 1	VAR00001 & VAR00002	.947	<.001

Paired Samples Test

	Paired Differences			95% Confidence Interval of the Difference			Significance			
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	One-Sided p	Two-Sided p	
Pair 1	VAR00001 - VAR00002	-.12733	.65069	.16801	-.48767	.23300	-.758	.14	.231	.461

Paired Samples Test: One-Sided p

- The average difference is not significant for the paired variable(s): (VAR00001 - VAR00002)
- The average difference is significant for the paired variable(s): (None)

Paired Samples Test: Two-Sided p

- The average difference is not significant for the paired variable(s): (VAR00001 - VAR00002)
- The average difference is significant for the paired variable(s): (None)

Note: Curated Help is calculated based on actual cell values, not the formatted values.

Turuvekere Satish Sagar

T-Test

[DataSet1]

Pivot Table Paired Samples Test

File Edit View Insert Pivot Format Help

Pair 1

	df	One-Sided p	Two-Sided p
Pair 1	VAR00001 - VAR00002	0.230537222320995	.461

General Area Format Borders Cell Format Notes

General

Hide empty rows and columns

Row Dimension Labels

In Corner

Nested

Column Widths

Minimum width for column labels: 0.94

Maximum width for column labels: 1.36

Minimum width for row labels: 0.68

Maximum width for row labels: 2.28

Paired

Paired

The average difference is not significant for the paired variable(s): (None)

Paired Samples Test: Two-Sided p

- The average difference is not significant for the paired variable(s): (VAR00001 - VAR00002)
- The average difference is significant for the paired variable(s): (None)

Note: Curated Help is calculated based on actual cell values, not the formatted values.

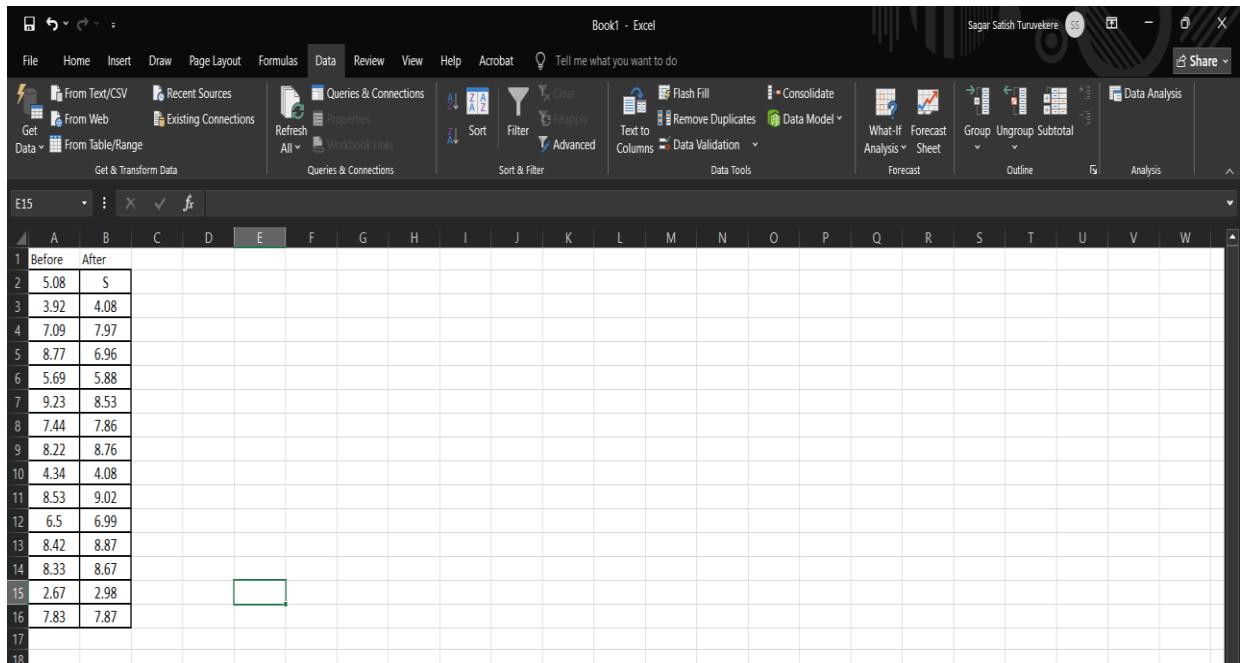
IBM SPSS Statistics Processor is ready

Unicode:ON

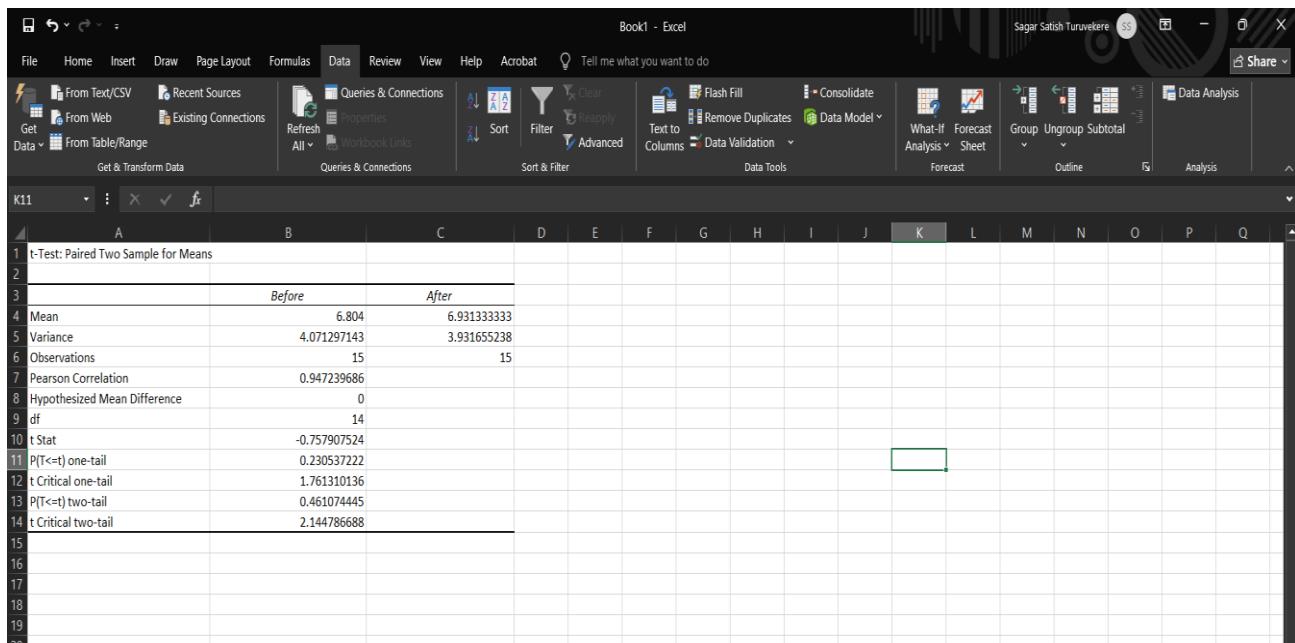
Classic

H: 126, W: 956 pt.

Excel



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Before	After																					
2	5.08	5																					
3	3.92	4.08																					
4	7.09	7.97																					
5	8.77	6.96																					
6	5.69	5.88																					
7	9.23	8.53																					
8	7.44	7.86																					
9	8.22	8.76																					
10	4.34	4.08																					
11	8.53	9.02																					
12	6.5	6.99																					
13	8.42	8.87																					
14	8.33	8.67																					
15	2.67	2.98																					
16	7.83	7.87																					
17																							
18																							



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	t-Test: Paired Two Sample for Means																
2																	
3		Before	After														
4	Mean		6.804		6.931333333												
5	Variance		4.071297143		3.931655238												
6	Observations		15		15												
7	Pearson Correlation		0.947239686														
8	Hypothesized Mean Difference		0														
9	df		14														
10	t Stat		-0.757907524														
11	P(T<=t) one-tail		0.230537222														
12	t Critical one-tail		1.761310136														
13	P(T >t) two-tail		0.461074445														
14	t Critical two-tail		2.144786688														
15																	
16																	
17																	
18																	
19																	

d) Interpret your results in the conclusion

By performing the tests, we can see that the p value for one tailed is greater than the alpha value (confidence value 95%) 0.05 hence we reject the null hypothesis.

Hence, the speed and agility camp did make a difference and their speeds improved

Q2:

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 $\alpha= 0.05$.

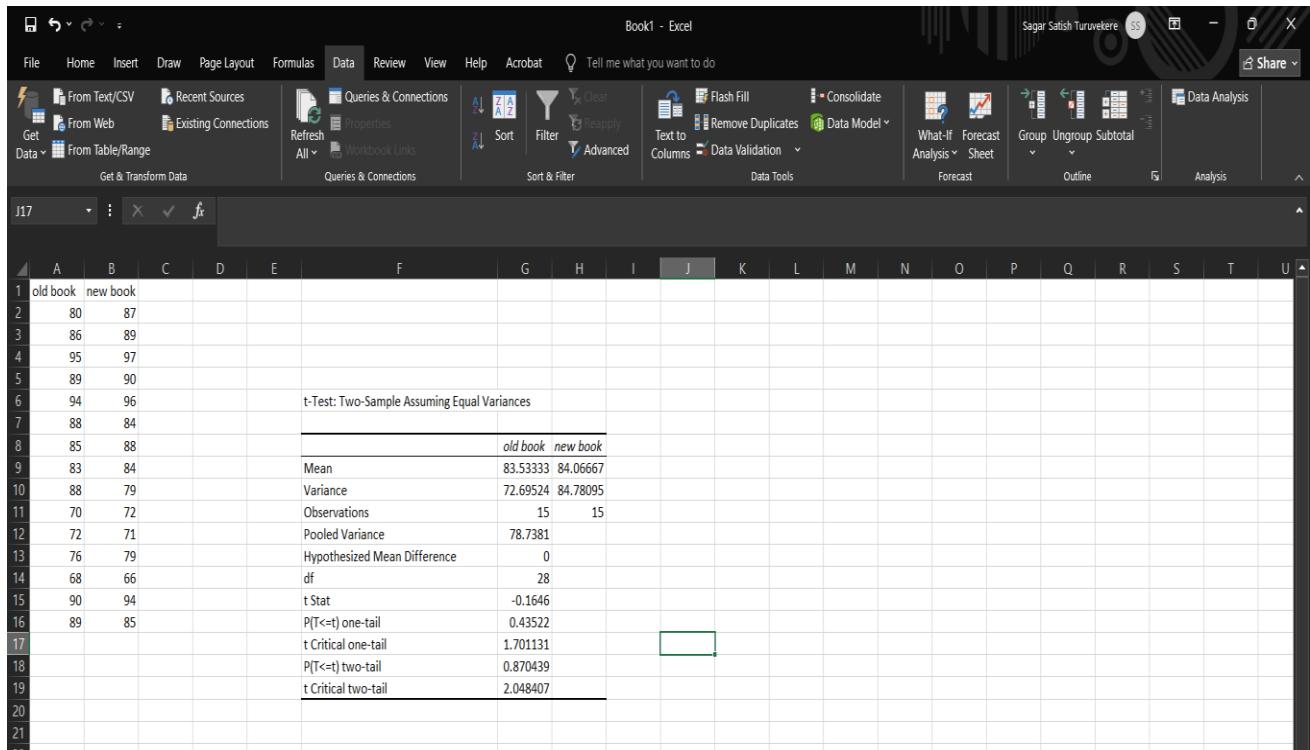
Note: You can solve this question by any (SPSS, EXCEL or Manually) (3 Marks)

Old book					New book				
80	86	95	89	94	87	89	97	90	96
88	85	83	88	70	84	88	84	79	72
72	76	68	90	89	71	79	66	94	85

Null Hypothesis: The mean test scores with old book = The mean test scores with new book

Alternative Hypothesis: The mean test scores with old book < The mean test scores with new book

Using two sample independent t – test (one tail)



The above has a Less than symbol in the Alternative Hypothesis hence it's a one tail test.

As the p value is greater than alpha which is 0.05 so we fail to reject the null hypothesis

Q3:

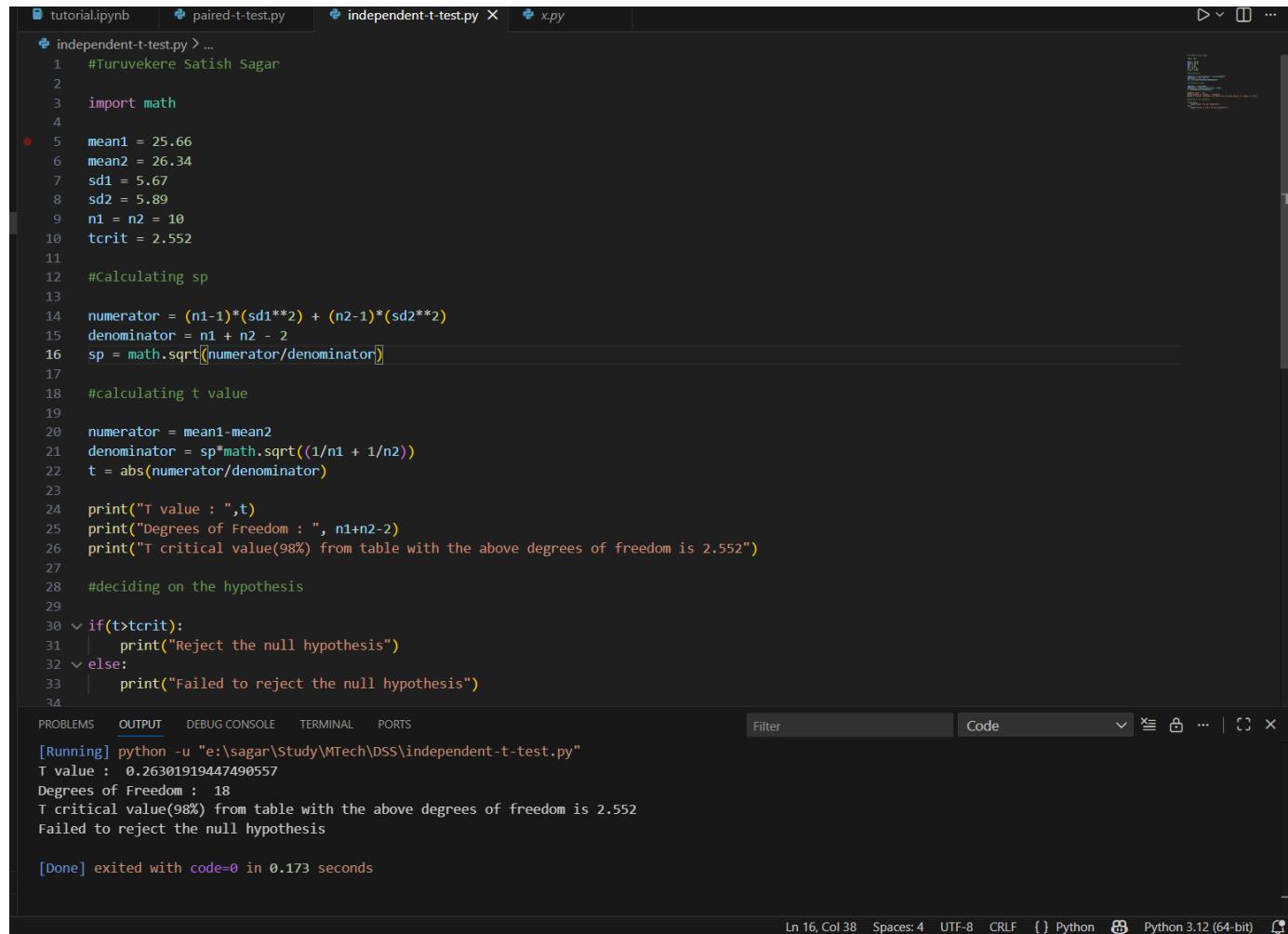
The Chapin Social Insight Test is a psychological test designed to measure how accurately a person appraises other people. The possible scores on the test range from 0 to 41. During the development of the test, it was given to several groups of people. Here are the results for male and female college 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. **(3 Marks)**

	n	avg.	std.dev
Male	10	25.66	5.67
Female	10	26.34	5.89

Null Hypothesis: Average value for male = Average value for female

Alternative Hypothesis: Average value for male \neq Average value for female

Using two sample independent t – test (two tail)



The screenshot shows a Jupyter Notebook interface with the following details:

- File Tabs:** tutorial.ipynb, paired-t-test.py, independent-t-test.py (active), x.py
- Code Cell Content:**

```
#Turuvekere Satish Sagar
import math
mean1 = 25.66
mean2 = 26.34
sd1 = 5.67
sd2 = 5.89
n1 = n2 = 10
tcrit = 2.552

#Calculating sp
numerator = (n1-1)*(sd1**2) + (n2-1)*(sd2**2)
denominator = n1 + n2 - 2
sp = math.sqrt(numerator/denominator)

#calculating t value
numerator = mean1-mean2
denominator = sp*math.sqrt(1/n1 + 1/n2)
t = abs(numerator/denominator)

print("T value : ",t)
print("Degrees of Freedom : ", n1+n2-2)
print("T critical value(98%) from table with the above degrees of freedom is 2.552")

#deciding on the hypothesis
if(t>tcrit):
    print("Reject the null hypothesis")
else:
    print("Failed to reject the null hypothesis")
```
- Output Cell Content:**

```
[Running] python -u "e:\sagar\Study\MTech\PSS\independent-t-test.py"
T value : 0.26301919447490557
Degrees of Freedom : 18
T critical value(98%) from table with the above degrees of freedom is 2.552
Failed to reject the null hypothesis

[Done] exited with code=0 in 0.173 seconds
```
- Bottom Status Bar:** Ln 16, Col 38, Spaces: 4, UTF-8, CRLF, Python 3.12 (64-bit)

As we use unequal symbol in the Alternative Hypothesis we perform a two tail test

Q4:

The distribution of scores of students taking the LSATs is claimed to have a mean of 500. We take a sample of 20 incoming Harvard Law School freshman LSAT scores and find a mean of 550 and a standard deviation of 37. Since Harvard is an Ivy League school, they think their freshmen are smarter than average law students. Test this theory (that Harvard students score higher than average on the LSATs) at the 0.05 significance level. **(2 Marks)**

The screenshot shows a Jupyter Notebook interface with a code cell containing Python code for a hypothesis test. The code imports math, defines population and sample means, calculates the test statistic, and prints the results. It also includes an if-else statement to determine whether to reject the null hypothesis based on the test statistic and critical value.

```
x.py > ...
1 #Turuvekere Satish Sagar
2 import math
3
4 print("Null Hypothesis H0 Mean score of population = Mean score of Harvard")
5 print("Alternative Hypothesis H1 Mean score of population < Mean score of Harvard")
6
7 popmean = 500
8 n = 20
9 sampmean = 550
10 sd = 37
11
12 teststat = (sampmean - popmean)/(sd/math.sqrt(n))
13
14 print("Test Statistic ",teststat)
15
16 print("\n Confidence Interval 98%")
17
18 print("\n Degrees of freedom is n - 1 here it is 19 as n = 20")
19 print("The critical test value from T-test table for df = 19 is 2.539")
20
21 if(teststat > 2.539):
22     print("Test statistic is greater than critical value hence we can reject the Null Hypothesis")
23 else:
24     print("Test statistic is less than the critical value hence we fail to reject the null Hypothesis")
25
26
```

PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL PORTS Filter Code

```
[Running] python -u "e:\sagar\Study\MTech\DS\Python\Hypothesis Testing\x.py"
Null Hypothesis H0 Mean score of population = Mean score of Harvard
Alternative Hypothesis H1 Mean score of population < Mean score of Harvard
Test Statistic 6.043426966215649

Confidence Interval 98%

Degrees of freedom is n - 1 here it is 19 as n = 20
The critical test value from T-test table for df = 19 is 2.539
Test statistic is greater than critical value hence we can reject the Null Hypothesis
```