# 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 procdeure 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 imporved.

Null Hypothesis –> $H_0$ – $Mean_{before} = Mean_{after}$

Alternative Hypothesis -> $H_1$ $Mean_{before} < Mean_{after}$

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

```python
#Turuvekere Satish Sagar
import math
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]
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]
diff = []
diffsquare = []
n = len(before)

for i in range(n):
    diffsquare.append((before[i] - after[i])**2)
    diff.append(before[i] - after[i])

sum1 = sum(diff)
print("The sum of differences",sum1)
sum2 = sum(diffsquare)
print("The sum of squares of differences",sum2)


t = sum1/(math.sqrt(((n*sum2) - (sum1**2))/(n-1)))

print("Test statistic of paired T-test",t)

```

```
PROBLEMS 1  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS
[Running] python -u "e:\sagar\Study\MTech\DSS\x.py"
The sum of differences -1.909999999999998
The sum of squares of differences 6.170699999999998
Test statistic of paired T-test -0.757907523977007
```

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

**SPSS**

| VAR00001 | VAR00002 | VAR00003 | 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 | .52098 |
| | VAR00002 | 6.9313 | 15 | 1.98284 | .51197 |

**Paired Samples Correlations**

| | | N | Correlation | Significance One-Sided p | Significance Two-Sided p |
|---|---|---|---|---|---|
| Pair 1 | VAR00001 & VAR00002 | 15 | .947 | <.001 | <.001 |

**Paired Samples Test**

| | | Paired Differences | | | | | t | df | Significance One-Sided p | Significance Two-Sided p |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference Lower | Upper | | | | |
| 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)*

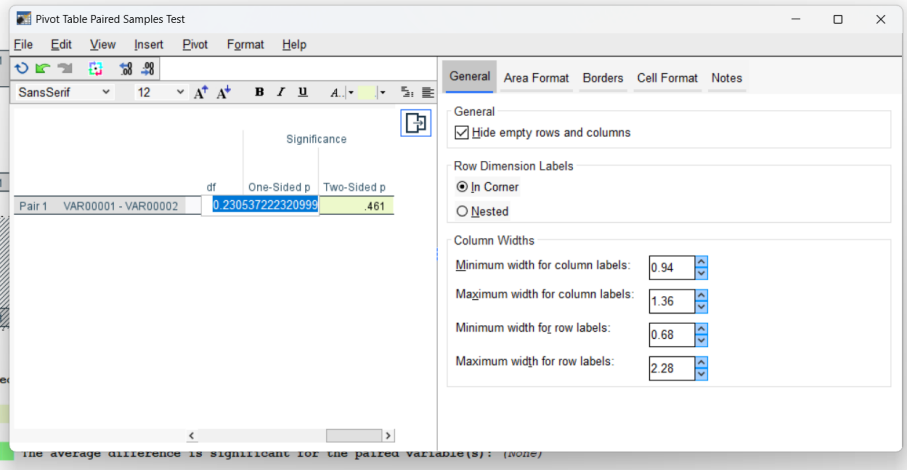
**The average difference is significant for the paired variable(s):** *(None)*

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

**Turuvekere Satish Sagar**

**T-Test**

[DataSet1]



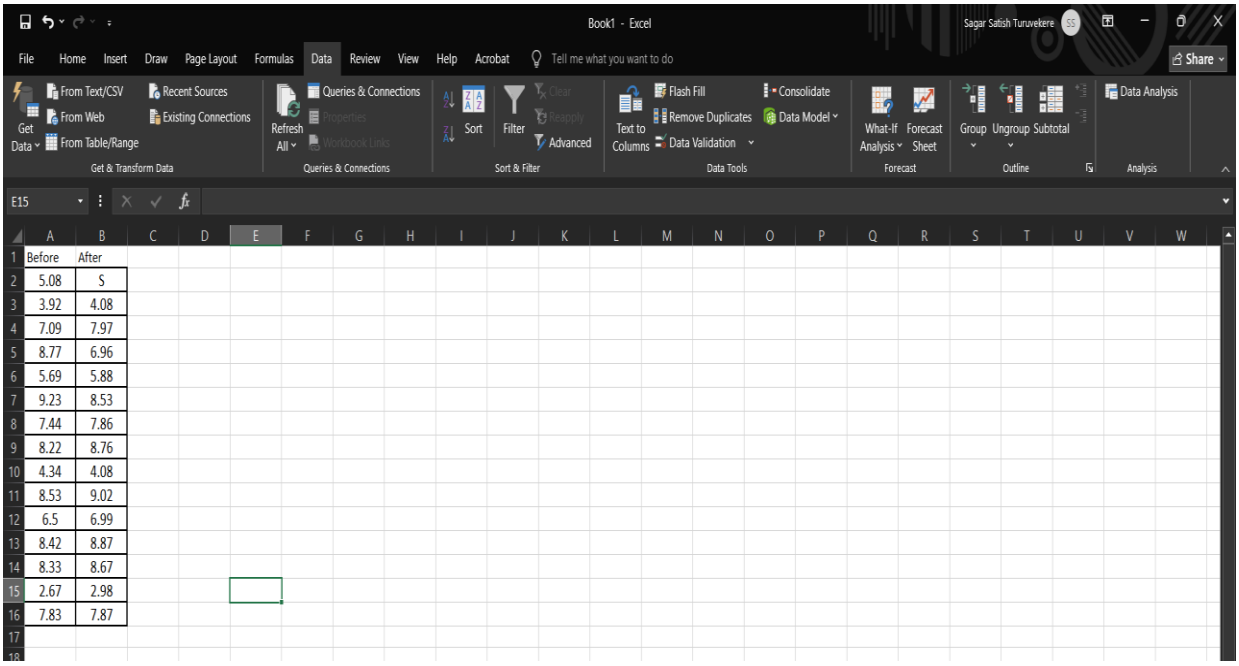

**Paired Samples Test: Two-Sided p**

**The average difference is not significant for the paired variable(s):** *(VAR00001 - VAR00002)*

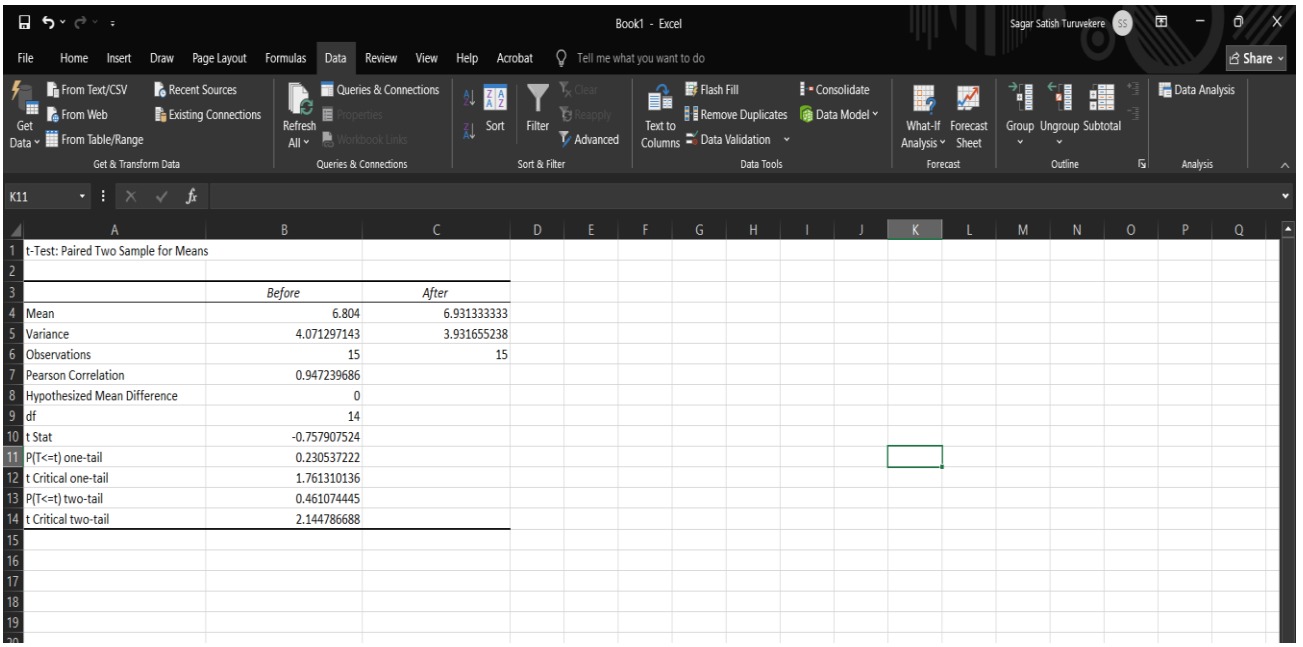
**The average difference is significant for the paired variable(s):** *(None)*

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

**Excel**



| | A | B |
|---|---|---|
| 1 | Before | After |
| 2 | 5.08 | S |
| 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 |



| | A | B | C |
|---|---|---|---|
| 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 | |

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 α= 0.05.

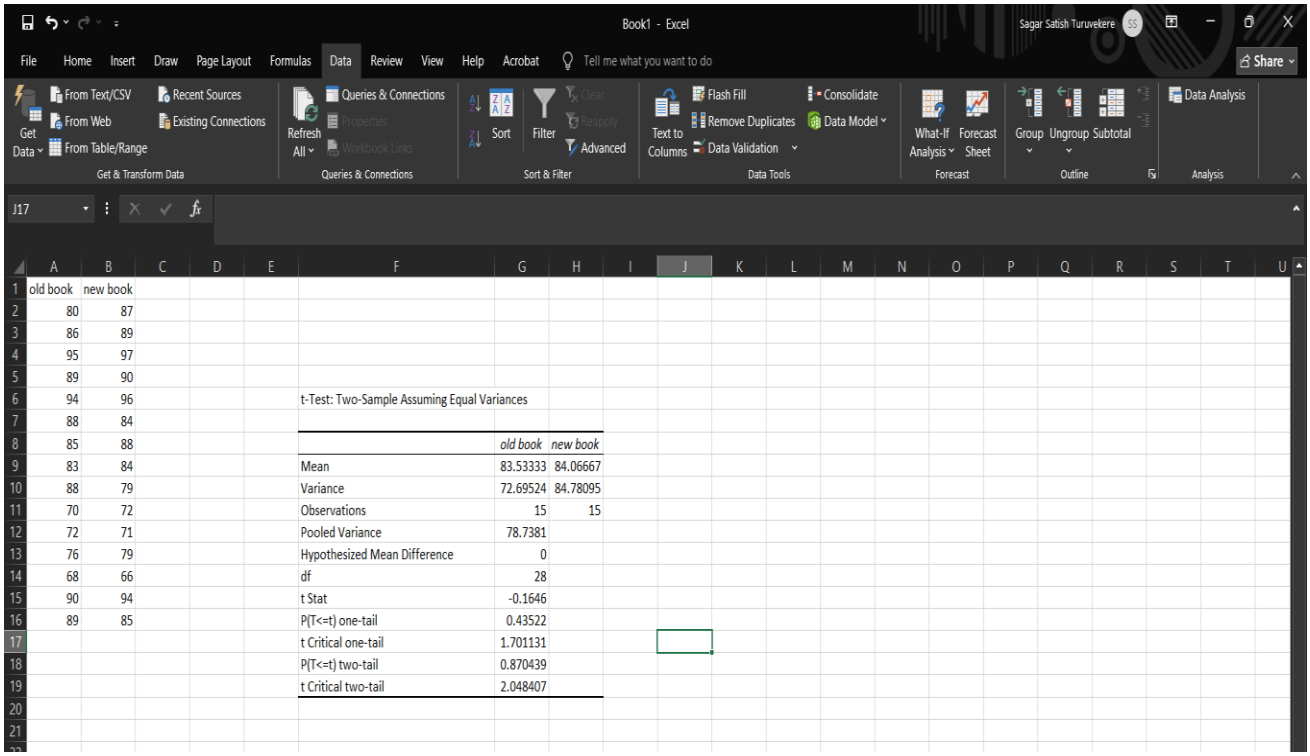
**Note:** You can solve this question by any (SPSS, EXCEL or Manually) **(3 Marks)**

| 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:

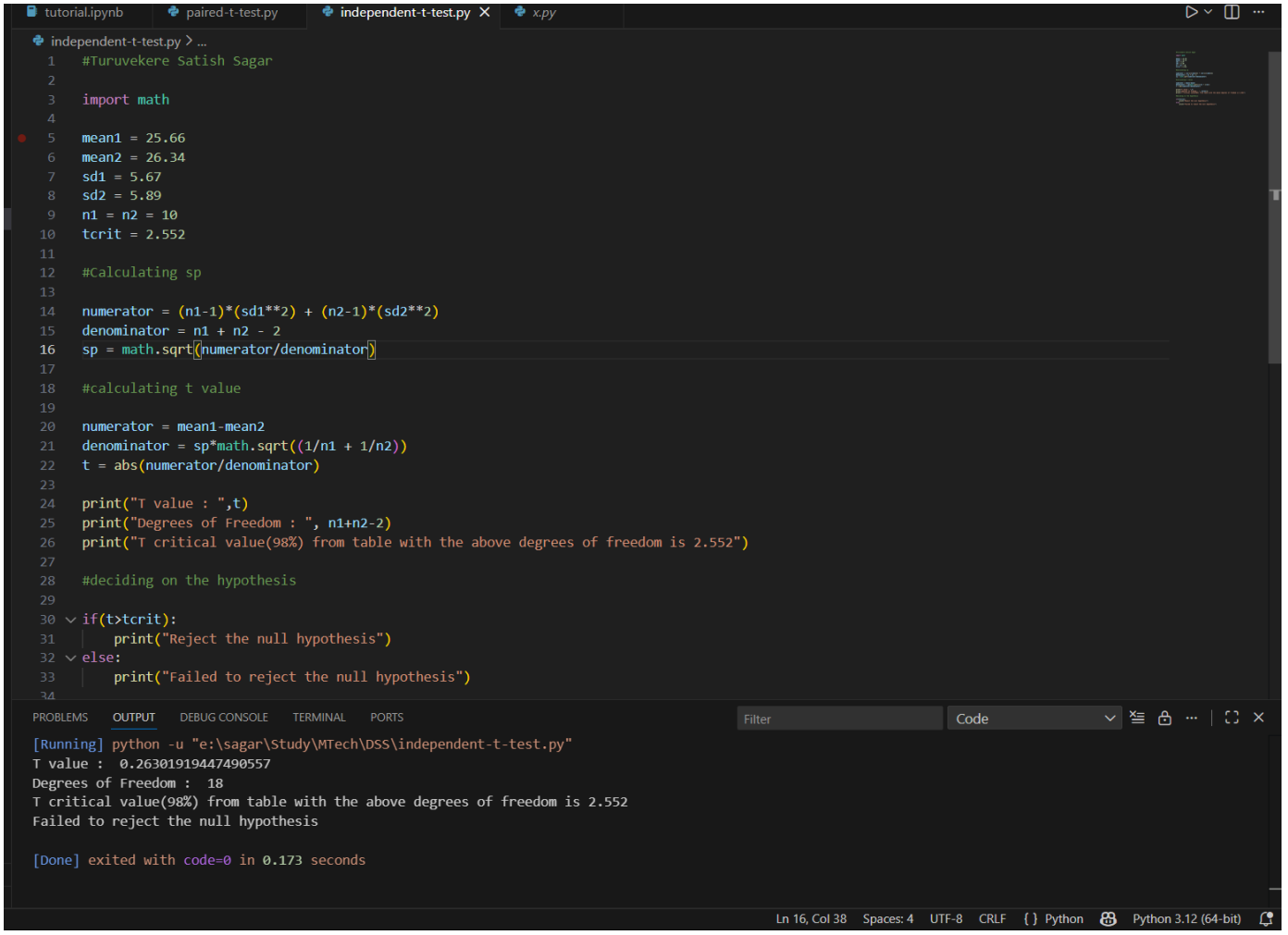
The Chapin Social Insight Test is a psychological test designed to measure how accurately a person appraises other people. The possible scores on the test range from 0 to 41. During the development of the test, it was given to several groups of people. Here are the results for male and female college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

**Using two sample independent t – test (two tail)**




**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)**