Data Analysis Y3- Week 10 Lab

- In this lab, you will parctice how you get calculate the correaltion, and answer hypothsese asumptions using Hypothsis testing
- Use the the Exercise and Fitness Metrics Dataset "exercise_dataset" in Moodel.

1. Import te imortant packages (pandas, scipy.stats, ttest, ttest_1samp, ttest_ind, ttest_rel, etc.,)

```
#Import packages
import scipy.stats as stats
from scipy.stats import ttest_1samp
from scipy.stats import ttest_ind
from scipy.stats import ttest_rel
import pandas as pd
```

2. Load and read the Exercise and Fitness Metrics Dataset "exercise_dataset"

```
#load and read the dataset
df =pd.read_csv('exercise_dataset.csv')
df.head()
```

	ID	Exercise	Calories Burn	Dream Weight	Actual Weight	Age	Gender	Duration	Heart Rate	BMI	Weather Conditions	Exercise Intensity
0	1	Exercise 2	286.959851	91.892531	96.301115	45	Male	37	170	29.426275	Rainy	5
1	2	Exercise 7	343.453036	64.165097	61.104668	25	Male	43	142	21.286346	Rainy	5
2	3	Exercise 4	261.223465	70.846224	71.766724	20	Male	20	148	27.899592	Cloudy	4
3	4	Exercise 5	127.183858	79.477008	82.984456	33	Male	39	170	33.729552	Sunny	10
4	5	Exercise	416.318374	89.960226	85.643174	29	Female	34	118	23.286113	Cloudy	3

Task 1: Using Correlation Test (Pearson's Correlation Coefficient):

- Calculate Pearson correlation coefficient between 'duration' and 'Calories Burn' and confrim the bewlow hypothses:

Hypotheses

- H0: There is no significant correlation between exercise time and calories burned.
- H1: There is a significant correlation between exercise time and calories burned.

```
stats.pearsonr(df['Duration'], df['Calories Burn'])
```

```
PearsonRResult(statistic=0.02179310988586642, pvalue=0.17560694580651773)
```

There is no Correlation between the two as the result was 0, so I have confidence there is none.

Task 2: Use One-Sample T-test to answer the wuestion below

Assuming you want to test if the mean weight of participants is significantly different from the population mean weight (70 kg).

Hypothisis

- H0: The mean weight of participants is equal to the population mean weight (70 kg).
- H1: The mean weight of participants is significantly different from the population mean weight (70 kg).

```
stats.ttest_1samp(df['Actual Weight'], 70)
```

```
TtestResult(statistic=21.725021808925085, pvalue=7.838353179052936e-99, df=3863)
```

My Hypothsis Answer: H1 - The P Value is Less Than 0.05 due to the e-99 in the result

Task 3: Use two-Sample T-test to answer the question below

Assuming you want to test if there is a significant difference in exercise time between male and female participants.

Hypotheses

- H0: There is no significant difference in exercise time between male and female participants.
- H1: There is a significant difference in exercise time between male and female participants.

```
group1 = df[df['Gender']=='Male']
group2 = df[df['Gender']=='Female']
```

```
ttest_ind(group1['Duration'], group2['Duration'])
```

```
TtestResult(statistic=-0.1391572031254142, pvalue=0.8893331821165864, df=3862.0)
```

Your Hypothsis Answer: H0 as there is no significant difference in exercise time between Men and Women, P Value is greater than 0.05

Task 4: Use Paired T-test to answer the below question

Assuming you want to test if there is a significant difference in weight before and after exercise.

- Hint (look at the dataset for the varaible that represnt the "after exercise")

Hypotheses

- H0: There is no significant difference in weight before and after exercise.
- H1: There is a significant difference in weight before and after exercise.

```
ttest_rel(df['Actual Weight'], df['Dream Weight'])
```

```
TtestResult(statistic=1.0719225710999716, pvalue=0.2838217725306073, df=3863)
```

Your Hypothsis Answer: H0 - P Value is Greater than 0.05

Note: Upload your lab to Moodel