Paired t-Test

This lesson will focus on how to perform the paired t-test in Python.

WE'LL COVER THE FOLLOWING

- Paired t-test

Paired t-test

A **paired t-test** checks whether the means of the same sample differ at two different times. We can use the function `ttest_rel` from the `scipy.stats` module to perform the paired t-test.

We will check the mean of grades for two exams, `G1` and `G3`. `G1` is the result of the exam after the first quarter and `G3` is the result of exam at the end of the year. We will check if the mean grades differ for the same sample at different times.

Null hypothesis H_0: $\bar{x_1} = \bar{x_2}$

Alternate hypothesis H_a: $\bar{x_1} \neq \bar{x_2}$

We choose α to be at 95% confidence level which means $\alpha = 1 - (conf.level) = 1 - 0.95 = 0.05$

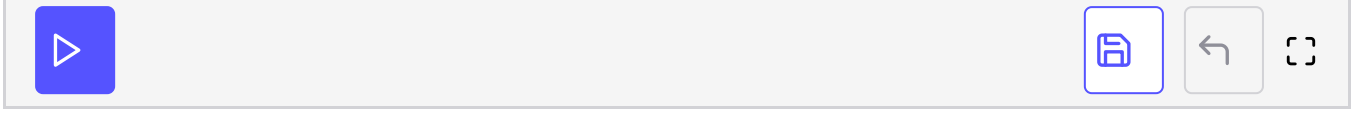
```
import pandas as pd
import scipy.stats as st

df = pd.read_csv('student-mat.csv')
# Samples
sample_1 = df['G1']
sample_2 = df['G3']

sample_1_grade_mean = sample_1.mean()
sample_2_grade_mean = sample_2.mean()

print('Sample 1 Grade Mean',sample_1_grade_mean)
print('Sample 2 Grade Mean',sample_2_grade_mean)
# Test
```

```
result = st.ttest_rel(sample_1,sample_2)
print(result)
```



We take the grades for `G1` and `G3` in **lines 6-7** in `sample_1` and `sample_2` respectively. We calculate their means in **lines 9-10**.

We use the `ttest_rel` function in **line 15** that expects both samples, to test the hypothesis. From the results, we can see that the p-value is less than the threshold of 0.05 which means that we reject the null hypothesis and the drop in performance in `G3` is statistically significant.

This brings the end of this chapter. We have learned how to construct confidence intervals and test our hypotheses. In the next chapter, we will move towards the last stage in the data science lifecycle, i.e., *prediction*.