# Two Sample t-Test

This lesson will focus on how to perform a two-sample t-test in Python.

WE'LL COVER THE FOLLOWING

- Two-sample t-test

## Two-sample t-test

A **two-sample t-test** checks whether means of two independent samples differ from each other. We can use the function `ttest_ind` from the `scipy.stats` module to perform the two-sample t-test.

We will be using the Student Alcohol Consumption Dataset. We will divide the data into two groups based on alcohol consumption. Then we will find the *mean* grade for both groups. We will check if the mean grades for both groups differ from each other or not.

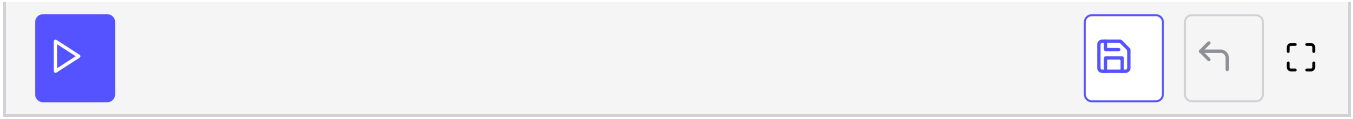
**Null hypothesis** $H_0$: $\bar{x_1} = \bar{x_2}$

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

We choose $\alpha$ 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')
df['grade'] = df['G1'] + df['G2'] + df['G3']
df['alc'] = df['Walc'] + df['Dalc']
# Samples
sample_1 = df[df['alc'] >5 ]['grade']
sample_2 = df[df['alc'] <=5  ]['grade']

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_ind(sample_1,sample_2)

print(result)
```



We combine the grades in **line 5** as we did in the previous example. We also combine the weekend (`Walc`) and weekday (`Dalc`) consumption to form a new column, `alc`. We will divide the dataset into two samples. One group with `alc` values greater than 5 and the other with `alc` values less than or equal to 5. We filter in **lines 8-9**. Since we are interested in the `grade`, we separate out the `grade` columns. We take the mean grades of both samples in **lines 11 and 12** and print these.

We use the function `ttest_ind` which expects both samples as arguments. From the result, we can see that the p-value is less than the significance value of 0.05, which means that we can safely reject the hypothesis and the difference between the mean grades of both samples is statistically significant. Therefore, we can conclude that the mean grade of students who consume less alcohol is greater than those who consume more alcohol.

In the next lesson, we will look at the *paired t-test*.