One Sample t-Test

This lesson will focus on how to perform one sample t-tests in Python.

WE'LL COVER THE FOLLOWING

One-sample t-test

A **One sample t-test** checks whether a population mean differs from the sample mean or not. We can use the function `ttest_1samp` from the `scipy.stats` module.

We will be using the Student Alcohol Consumption Dataset. We will take a sample from the data and find the *mean* grade. We will check if the population mean grade (μ) differs from the sample mean grade ($\bar{x}$) or not.

Null hypothesis H_0: $\bar{x} = \mu$

Alternate hypothesis H_a: $\bar{x} \neq \mu$

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')
df['grade'] = df['G1'] + df['G2'] + df['G3']

# population
pop_grades = df['grade']
pop_grade_mean = pop_grades.mean()

# sample
sample_grades = pop_grades.sample(n=100,random_state = 6)
sample_grade_mean = sample_grades.mean()

print('Population Grade Mean',pop_grade_mean)
print('Sample Grade Mean',sample_grade_mean)
```

```
# Test
result = st.ttest_1samp(sample_grades,pop_grade_mean)
print(result)
```

Since there were three separate grades, we combined all of the grades in one column and named it `grade` in **line 5**. In **line 8**, we filter for the grade column since we are only concerned with the grades at this time. We consider this as our population. We find the mean of the population in the next line. In **line 12**, we sample from the population and find the sample mean in the next line. We print both means in **lines 15-16** to see the difference between the two means. Now our test will tell us whether this difference was by chance or if it is statistically significant.

In **line 19**, we use the `ttest_1samp` function. It expects two arguments; the sample and the population mean. We provide both the arguments and store the result in `result`. Then we print the results in the next line.

`ttest_1samp` returns the test statistic and the p-value. The test statistic tells us how much the sample mean deviates from the population mean. `ttest_1samp` conducts a **two-tailed test** meaning that the p-value returned to us caters for both positive and negative differences in both the means. From the output, we can see that the p-value = 0.40, which is **greater** than our significance level α. Therefore, we cannot reject the null hypothesis and the difference in both the means is by chance and not statistically significant. This implies that if we were to construct a 95% confidence interval with the sample mean, the population mean would be captured in it.

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