

CT4031 Maths for Data Science

Week 9 - Practical



Statistical test

```
from scipy.stats import ttest_ind
import pandas
data = pandas.read_csv('data.csv')
data1 = data.value[data.names == 'sample1']
data2 = data.value[data.names == 'sample2']
stat, p = ttest_ind(data1, data2)
print('P-value = ', p)
confidence = round((1-p)*100)
if p>0.05:
    print('Similar, confidence of ', confidence, '%')
else:
    print('Different, confidence of ', confidence, '%')
```





Statistical test

```
from statsmodels.stats import weightstats as stests
import numpy as np
income = np.genfromtxt("income.csv", delimiter=',')
#ztest, income mean is equals to 188888
ztest,pval = stests.ztest(income[:,1], value=188888)
print("p-values",pval)
if pval<0.05:
    print("Z-test => we are rejecting hypothesis")
else:
    print("Z-test => we are accepting hypothesis")
```



