T- TEST

A T-test is a statistical method of comparing the means or proportions of two samples gathered from either the same group or different categories.

One-sample, two-sample, paired, equal, and unequal variance are the types of T-tests users can use for mean comparisons.

A T-test studies a set of data gathered from **two similar or different groups** to determine the probability of the difference in the result than what is usually obtained. The accuracy of the test depends on various factors, including the distribution patterns used and the variants influencing the collected samples. Depending on the parameters, the test is conducted, and a **T-value** is obtained as the statistical inference of the probability of the usual resultant being driven by chance.

Two of such statistical hypothesis testing techniques are **paired t-test and unpaired t-test**. The key difference between both of them is that in

- **paired t-test** you compare the paired measures that match deliberately.
- Whereas in **unpaired t-test** you compare the means of two samples that have no natural pairing.

Working of paired and unpaired t-test with an example

For this comparison, we will take an example: a doctor wants to see how a group of people respond to the same drug. Let's take a look at the data:

1) To perform a paired t-test: the doctor will let the group of patients take drug #1 for one month and then measure the recovery. Then have them use drug #2 and then again measure the recovery equally.

Drug #1

Patient	Recovery in %
Patient 1	77
Patient 2	79
Patient 3	83
Patient 4	84
Patient 5	84
Patient 6	87
Patient 7	89
Patient 8	90
Patient 9	94
Patient 10	95

Drug #2

Patient	Recovery in %
Patient 1	79
Patient 2	84
Patient 3	80
Patient 4	83
Patient 5	83
Patient 6	82
Patient 7	80
Patient 8	91
Patient 9	92
Patient 10	87

Since each patient takes both the drugs, the doctor will take paired t-test and compare the mean of two tables to find out which drug is more effective.

2) To perform an unpaired t-test: the doctor will take 20 patients and randomly split them into two groups and make them take drug #1 and drug #2 separately.

Drug #1

Patient	Recovery in %
Patient 1	77
Patient 14	79
Patient 5	83
Patient 8	84
Patient 20	84
Patient 17	87
Patient 13	89
Patient 19	90
Patient 2	94
Patient 4	95

Drug #2

Patient	Recovery in %
Patient 3	84
Patient 15	78
Patient 6	80
Patient 9	76
Patient 10	88
Patient 18	89
Patient 12	92
Patient 11	93
Patient 16	90
Patient 18	86

Since the patient in **the two groups is totally different and independent**, the doctor will use an unpaired t-test to see which of the group has the higher mean and determine which drug will be more effective.

For example,

➤ p value of 5% is 0.05.

➤ **Low p-values are good**; They indicate your data did not occur by chance.

➤ For example, a p-value of .01 means there is only a 1% probability that the results from an experiment happened by chance.

➤ In most cases, a p-value of 0.05 (5%) is accepted to mean the data is valid.

UnPaired T-Test

```python
from scipy.stats import ttest_ind
dataset=dataset.dropna()
male = dataset[dataset['gender']=='M']['salary']
female = dataset[dataset['gender']=='F']['salary']
ttest_ind(male, female)
```

TtestResult(statistic=2.0937842507001165, pvalue=0.03746230397067219, df=213.0)

The t-statistic of 2.0937842507001165 suggests that there is a difference between the two groups, and the p-value of 0.03746230397067219 suggests that there is a relatively low probability of observing a t-statistic as extreme as the one calculated if the null hypothesis is true. If the significance level is set to 0.05, then the null hypothesis would be rejected, and it would be concluded that there is a statistically significant difference between the two groups

Paired T test

Dependant Sample-Paired T_Test

Same Group(Male) but Different Condition(ssc_p,hsc_p)

```
from scipy.stats import ttest_rel
#dataset=dataset.dropna()
male = dataset[dataset['gender']=='M']['ssc_p']
male1 = dataset[dataset['gender']=='M']['hsc_p']
ttest_rel(male, male1)
```

```
TtestResult(statistic=0.6842940392688153, pvalue=0.4949370716000775, df=138)
```

In the given search results, the T-test result is TtestResult(statistic=0.6842940392688153, pvalue=0.4949370716000775, df=138). This means that the statistic value is 0.684, the p-value is 0.4949, and the degrees of freedom is 138. Based on this result, the null hypothesis of no significant difference between the means of the two groups cannot be rejected, as the p-value is greater than the significance level of 0.05.