

How can a two sample t-test be conducted in Python?

Authored by
stats writer

April 17, 2024

RECOMMENDED CITATION

stats writer (2024). *How can a two sample t-test be conducted in Python?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=136173>

A two sample t-test in Python is a statistical method used to compare the means of two independent groups. It is conducted by first importing the appropriate packages, then organizing the data into two separate arrays or data frames. Next, the t-test function is applied to the two groups, which calculates the t-statistic and p-value. The results can then be interpreted to determine if there is a significant difference between the means of the two groups. This test is useful for analyzing the effectiveness of a certain treatment or intervention, or for comparing the performance of two different groups in a study.

Conduct a Two Sample T-Test in Python

A two sample t-test is used to test whether or not the means of two populations are equal.

This tutorial explains how to conduct a two sample t-test in Python.

Example: Two Sample t-Test in Python

Researchers want to know whether or not two different species of plants have the same mean height. To test this, they collect a simple random sample of 20 plants from each species.

Use the following steps to conduct a two sample t-test to determine if the two species of plants have the same height.

Step 1: Create the data.

First, we'll create two arrays to hold the measurements of each group of 20 plants:

```
import numpy as np
```

```
group1 = np.array()
```

```
group2 = np.array()
```

Step 2: Conduct a two sample t-test.

Next, we'll use the ttest_ind() function from the `scipy.stats` library to conduct a two sample t-test, which uses the following syntax:

```
ttest_ind(a, b, equal_var=True)
```

where:

a: an array of sample observations for group 1

b: an array of sample observations for group 2

equal_var: if True, perform a standard independent 2 sample t-test that assumes equal population variances. If False, perform Welch's t-test, which does not assume equal population variances. This is True by default.

Before we perform the test, we need to decide if we'll

assume the two populations have equal variances or not. As a rule of thumb, we can assume the populations have equal variances if the ratio of the larger sample variance to the smaller sample variance is less than 4:1.

```
#find variance for each group  
print(np.var(group1), np.var(group2))
```

7.73 12.26

The ratio of the larger sample variance to the smaller sample variance is $12.26 / 7.73 = 1.586$, which is less than 4. This means we can assume that the population variances are equal.

Thus, we can proceed to perform the two sample t-test with equal variances:

```
import scipy.stats as stats  
  
#perform two sample t-test with equal variances  
stats.ttest_ind(a=group1, b=group2, equal_var=True)  
  
(statistic=-0.6337, pvalue=0.53005)
```

The t test statistic is -0.6337 and the corresponding two-sided p-value is 0.53005.

Step 3: Interpret the results.

The two hypotheses for this particular two sample t-test are as follows:

H0: $\mu_1 = \mu_2$ (the two population means are equal)

HA: $\mu_1 \neq \mu_2$ (the two population means are *not* equal)

Because the p-value of our test (0.53005) is greater than alpha = 0.05, we fail to reject the null hypothesis of the test. We do not have sufficient evidence to say that the mean height of plants between the two populations is different.

How to Conduct a One Sample T-Test in Python

How to Conduct a Paired Samples T-Test in Python