

Can you demonstrate how to perform t-tests in Pandas using three different examples?

Authored by
stats writer

June 28, 2024

RECOMMENDED CITATION

stats writer (2024). *Can you demonstrate how to perform t-tests in Pandas using three different examples?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=156394>

This formal description introduces the concept of performing t-tests in Pandas and explains the purpose of demonstrating three different examples. It highlights the main objective of the task and provides a brief overview of what the audience can expect to learn.

"The demonstration of performing t-tests in Pandas using three different examples aims to showcase the functionality and versatility of this statistical analysis tool. Through comprehensive examples, the audience will learn how to perform t-tests in Pandas, a popular data analysis and manipulation library in Python. These examples will cover various scenarios and use cases, providing a comprehensive understanding of the t-test function in Pandas. By the end of the demonstration, the audience will be equipped with the knowledge and skills to confidently conduct t-tests using Pandas in various data analysis projects."

Perform t-Tests in Pandas (3 Examples)

The following examples show how to perform three different t-tests using a pandas DataFrame:

**Independent Two Sample t-Test
Welch's Two Sample t-Test
Paired Samples t-Test**

Example 1: Independent Two Sample t-Test in Pandas

An is used to determine if two population means are equal.

For example, suppose a professor wants to know if two different studying methods lead to different mean exam scores.

To test this, he recruits 10 students to use method A and 10 students to use method B.

The following code shows how to enter the scores of each student in a pandas DataFrame and then use the function from the SciPy library to perform an independent two sample t-test:

```
import pandas as pd
from scipy.stats import ttest_ind

#create pandas DataFrame
df = pd.DataFrame({'method': ,
'score': })

#view first five rows of DataFrame
df.head()

method score
0 A 71
1 A 72
2 A 72
3 A 75
4 A 78

#define samples
group1 = df=='A']
group2 = df=='B']
```

```
#perform independent two sample t-test  
ttest_ind(group1, group2)
```

```
Ttest_indResult(statistic=-2.6034304605397938,  
pvalue=0.017969284594810425)
```

From the output we can see:

t test statistic: -2.6034p-value: 0.0179

Since the p-value is less than .05, we reject the null hypothesis of the t-test and conclude that there is sufficient evidence to say that the two methods lead to different mean exam scores.

Example 2: Welch's t-Test in Pandas

is similar to the independent two sample t-test, except it does not assume that the two populations that the samples came from have .

To perform Welch's t-test on the exact same dataset as the previous example, we simply need to specify `equal_var=False` within the `ttest_ind()` function as follows:

```
import pandas as pd
from scipy.stats import ttest_ind

#create pandas DataFrame
df = pd.DataFrame({'method': ,
'score': })

#define samples
group1 = df=='A']
group2 = df=='B']

#perform Welch's t-test
ttest_ind(group1, group2, equal_var=False)

Ttest_indResult(statistic=-2.603430460539794,
pvalue=0.02014688617423973)
```

From the output we can see:

t test statistic: -2.6034p-value: 0.0201

Example 3: Paired Samples t-Test in Pandas

A is used to determine if two population means are equal in which each observation in one sample can be paired with an observation in the other sample.

For example, suppose a professor wants to know if two different studying methods lead to different mean exam scores.

To test this, he recruits 10 students to use method A and then take a test. Then, he lets the same 10 students used method B to prepare for and take another test of similar difficulty.

Since all of the students appear in both samples, we can perform a paired samples t-test in this scenario.

The following code shows how to enter the scores of each student in a pandas DataFrame and then use the function from the SciPy library to perform a paired samples t-test:

```
import pandas as pd  
from scipy.stats import ttest_rel  
  
#create pandas DataFrame  
df = pd.DataFrame({'method': ,  
'score': })  
  
#view first five rows of DataFrame
```

```
df.head()
```

```
method score
```

```
0 A 71
```

```
1 A 72
```

```
2 A 72
```

```
3 A 75
```

```
4 A 78
```

```
#define samples
```

```
group1 = df=='A']
```

```
group2 = df=='B']
```

```
#perform independent two sample t-test
```

```
ttest_rel(group1, group2)
```

```
Ttest_relResult(statistic=-6.162045351967805,  
pvalue=0.0001662872100210469)
```

From the output we can see:

t test statistic: -6.1620p-value: 0.0001

Since the p-value is less than .05, we reject the null hypothesis of the paired samples t-test and conclude that there is sufficient evidence to say that the two

methods lead to different mean exam scores.

Additional Resources

The following tutorials explain how to perform other common tasks in Python:

ARABPSYCHOLOGY.COM