What is the process for calculating standard deviation in Pandas and can you provide some examples?

Authored by

stats writer

May 12, 2024

The process for calculating standard deviation in Pandas involves using the "std()" function, which is a built-in method that can be applied to a Pandas dataframe or series. This function takes in the data as an input and returns the standard deviation as an output. The formula used to calculate standard deviation in Pandas is the same as the one used in statistics, which involves finding the square root of the variance. Some examples of calculating standard deviation in Pandas include finding the standard deviation for a specific column in a dataframe, or for a subset of data using conditional statements. This can be useful for analyzing the variability or spread of a dataset.

Calculate Standard Deviation in Pandas (With Examples)

You can use the function to calculate the standard deviation of values in a pandas DataFrame.

You can use the following methods to calculate the standard deviation in practice:

Method 1: Calculate Standard Deviation of One Column

df.std()

Method 2: Calculate Standard Deviation of Multiple Columns

df].std()

Method 3: Calculate Standard Deviation of All Numeric Columns

df.std()

Note that the std() function will automatically ignore any NaN values in the DataFrame when calculating the standard deviation.

The following examples shows how to use each method with the following pandas DataFrame:

```
import pandas as pd

#create DataFrame
df = pd.DataFrame({'team': ,
'points': ,
'assists': ,
'rebounds': })

#view DataFrame
print(df)
```

team points assists rebounds
0 A 25 5 11
1 A 12 7 8
2 B 15 7 10
3 B 14 9 6

4 B 19 12 6

5 B 23 9 5

6 C 25 9 9

7 C 29 4 12

Method 1: Calculate Standard Deviation of One Column

The following code shows how to calculate the standard deviation of one column in the DataFrame:

#calculate standard deviation of 'points' column
df.std()

6.158617655657106

The standard deviation turns out to be 6.1586.

Method 2: Calculate Standard Deviation of Multiple Columns

The following code shows how to calculate the standard deviation of multiple columns in the DataFrame:

#calculate standard deviation of 'points' and 'rebounds' columns
df].std()

points 6.158618
rebounds 2.559994
dtype: float64

The standard deviation of the 'points' column is 6.1586 and the standard deviation of the 'rebounds' column is 2.5599.

Method 3: Calculate Standard Deviation of All Numeric Columns

The following code shows how to calculate the standard deviation of every numeric column in the DataFrame:

```
#calculate standard deviation of all numeric columns
df.std()
points 6.158618
assists 2.549510
rebounds 2.559994
dtype: float64
```

Notice that pandas did not calculate the standard deviation of the 'team' column since it was not a numeric column.

The following tutorials explain how to perform other common operations in pandas: