How can I calculate the standard deviation by group in Pandas?

Authored by

stats writer

June 26, 2024

The process of calculating the standard deviation by group in Pandas involves using the groupby function to group the data by a specific variable, and then using the std() method to calculate the standard deviation for each group. This allows for the analysis of how data varies within different groups, providing valuable insights into the overall distribution of the data. By utilizing this technique, users can efficiently calculate and compare standard deviations for multiple groups within a dataset in a single step.

Calculate Standard Deviation by Group in Pandas

You can use the following methods to calculate the standard deviation by group in pandas:

Method 1: Calculate Standard Deviation of One Column Grouped by One Column

df.groupby().std()

Method 2: Calculate Standard Deviation of Multiple Columns Grouped by One Column

df.groupby().std()

Method 3: Calculate Standard Deviation of One Column Grouped by Multiple Columns

df.groupby().std()

The following examples show how to use each method in practice with the following pandas DataFrame:

```
import pandas as pd

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

#view DataFrame
print(df)

team position points assists
0 A G 30 4
1 A F 22 3
2 A F 19 7
3 A G 14 7
4 B F 14 12
5 B F 11 15
6 B G 20 8
7 B G 28 4
```

Example 1: Calculate Standard Deviation of One Column Grouped by One Column

The following code shows how to calculate the standard deviation of the points column, grouped by the team column:

```
#calculate standard deviation of points grouped by team
df.groupby('team').std()
```

```
team
A 6.70199
B 7.50000
Name: points, dtype: float64
```

From the output we can see:

The standard deviation of points for team A is 6.70199.The standard deviation of points for team B is 7.5.

Example 2: Calculate Standard Deviation of Multiple Columns Grouped by One Column

The following code shows how to calculate the

standard deviation of the points column and the standard deviation of the assists column, grouped by the team column:

#calculate standard deviation of points and assists grouped by team
df.groupby('team')].std()

points assists
team
A 6.70199 2.061553
B 7.50000 4.787136

Example 3: Calculate Standard Deviation of One Column Grouped by Multiple Columns

The following code shows how to calculate the standard deviation of the points column, grouped by the team and position columns:

#calculate standard deviation of points, grouped by team and position
df.groupby().std()

team position
A F 2.121320

G 11.313708

B F 2.121320

G 5.656854

Name: points, dtype: float64

From the output we can see:

The standard deviation of points for players on team A and position F is 2.12.The standard deviation of points for players on team A and position G is 11.31.The standard deviation of points for players on team B and position F is 2.12.The standard deviation of points for players on team B and position G is 5.65.

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