

How do I calculate the mean, median, and mode in Pandas?

Authored by
stats writer

June 25, 2024

RECOMMENDED CITATION

stats writer (2024). *How do I calculate the mean, median, and mode in Pandas?*.
PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=152329>

Pandas is a popular Python library used for data analysis. It offers powerful tools for performing various statistical calculations, including mean, median, and mode. To calculate the mean, median, and mode in Pandas, you can use the built-in functions `mean()`, `median()`, and `mode()`, respectively. These functions can be applied to a Pandas series or dataframe, which contain the data that you want to analyze. The mean is the average of all the values in the data, the median is the middle value when the data is arranged in ascending order, and the mode is the most frequently occurring value. By using Pandas, you can easily and efficiently calculate these measures of central tendency for your data.

Calculate Mean, Median and Mode in Pandas

You can use the following functions to calculate the mean, median, and mode of each numeric column in a pandas DataFrame:

```
print(df.mean(numeric_only=True))
print(df.median(numeric_only=True))
print(df.mode(numeric_only=True))
```

The following example shows how to use these functions in practice.

Example: Calculate Mean, Median and Mode in Pandas

Suppose we have the following pandas DataFrame that contains information about points scored by various basketball players in four different games:

```
import pandas as pd
```

```
#create DataFrame
```

```
df = pd.DataFrame({'player': ,  
'game1': ,  
'game2': ,  
'game3': ,  
'game4': })
```

```
#view DataFrame
```

```
print(df)
```

```
player game1 game2 game3 game4  
0 A 18 5 11 9  
1 B 22 7 8 8  
2 C 19 7 10 10  
3 D 14 9 6 9  
4 E 14 12 6 14  
5 F 11 9 5 15  
6 G 20 9 9 10  
7 H 28 4 12 11
```

We can use the following syntax to calculate the mean value of each numeric column:

```
#calculate mean of each numeric column
```

```
print(df.mean(numeric_only=True))
```

```
game1 18.250
```

```
game2 7.750
```

```
game3 8.375
```

```
game4 10.750
```

```
dtype: float64
```

From the output we can see:

The mean value in the game1 column is 18.25. The mean value in the game2 column is 7.75. The mean value in the game3 column is 8.375. The mean value in the game4 column is 10.75.

We can then use the following syntax to calculate the median value of each numeric column:

```
#calculate median of each numeric column
```

```
print(df.median(numeric_only=True))
```

```
game1 18.5
```

```
game2 8.0
```

```
game3 8.5
```

```
game4 10.0
```

dtype: float64

From the output we can see:

The median value in the game1 column is 18.5. The median value in the game2 column is 8. The median value in the game3 column is 8.5. The median value in the game4 column is 10.

We can then use the following syntax to calculate the mode of each numeric column:

```
#calculate mode of each numeric column  
print(df.mode(numeric_only=True))
```

```
game1 game2 game3 game4  
0 14.0 9.0 6.0 9  
1 NaN NaN NaN 10
```

From the output we can see:

The mode in the game1 column is 14. The mode in the game2 column is 9. The mode in the game3 column is 6. The mode in the game4 column is 9 and 10

Note that the game4 column had two modes since there were two values that occurred most frequently in that column.

Note: You can also use the function in pandas to generate more descriptive statistics for each column.

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

ARABPSYCHOLOGY.COM