

How can I calculate percentiles in Python, and what are some examples?

Authored by
stats writer

April 21, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate percentiles in Python, and what are some examples?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=137833>

Calculating percentiles in Python involves using a statistical function or method to determine the value that divides a dataset into two parts: the specified percentage of the data falls below the percentile, and the remaining percentage falls above it. This is a useful tool for understanding the distribution and spread of a dataset.

To calculate percentiles in Python, one can use the built-in statistics module or the numpy library. The function or method used will depend on the type of data and the specific percentile desired. For example, the `percentile()` function in the statistics module can be used to calculate any percentile from a list of values, while the `percentile()` method in numpy can be used to calculate multiple percentiles at once from a numpy array.

An example of calculating percentiles in Python could be determining the 75th percentile of a dataset representing the heights of a group of people. Using the statistics module, the `percentile()` function can be applied to the list of heights to find the value where 75% of the heights fall below and 25% fall above. Similarly, using the `percentile()` method in numpy, multiple percentiles (e.g. 25th, 50th, and 75th) can be calculated from a numpy array of heights.

In conclusion, calculating percentiles in Python is a simple and efficient way to analyze and understand data. By using the appropriate function or method, percentiles can be easily calculated for various types of data, providing valuable insights into its distribution and variability.

Calculate Percentiles in Python (With Examples)

The n th percentile of a dataset is the value that cuts off the first n percent of the data values when all of the values are sorted from least to greatest.

For example, the 90th percentile of a dataset is the value that cuts off the bottom 90% of the data values from the top 10% of data values.

We can quickly calculate percentiles in Python by using the `numpy.percentile()` function, which uses the

following syntax:

numpy.percentile(a, q)

where:

a: Array of values
q: Percentile or sequence of percentiles to compute, which must be between 0 and 100 inclusive.

This tutorial explains how to use this function to calculate percentiles in Python.

How to Find Percentiles of an Array

The following code illustrates how to find various percentiles for a given array in Python:

```
import numpy as np
```

```
#make this example reproducible
```

```
np.random.seed(0)
```

```
#create array of 100 random integers distributed  
between 0 and 500
```

```
data = np.random.randint(0, 500, 100)
```

#find the 37th percentile of the array

```
np.percentile(data, 37)
```

173.26

#Find the quartiles (25th, 50th, and 75th percentiles) of the array

```
np.percentile(data, )
```

```
array()
```

How to Find Percentiles of a DataFrame Column

The following code shows how to find the 95th percentile value for a single pandas DataFrame column:

```
import numpy as np  
import pandas as pd#create DataFrame  
df = pd.DataFrame({'var1': ,  
'var2': ,  
'var3': })#find 90th percentile of var1 column  
np.percentile(df.var1, 95)
```

34.1

How to Find Percentiles of Several DataFrame Columns

The following code shows how to find the 95th percentile value for a several columns in a pandas DataFrame:

```
import numpy as np
import pandas as pd#create DataFrame
df = pd.DataFrame({'var1': ,
'var2': ,
'var3': })#find 95th percentile of each column
df.quantile(.95)
```

```
var1 34.10
```

```
var2 14.55
```

```
var3 14.65
```

```
#find 95th percentile of just columns var1 and var2
df].quantile(.95)
```

```
var1 34.10
```

```
var2 14.55
```

Note that we were able to use the pandas quantile() function in the examples above to calculate percentiles.