

Variance and Standard Deviation

Standard Deviation

The standard deviation is a measure of a dataset's spread. It is calculated by taking the square root of the variance of a data set. The resulting value has the same units as the original data.

Calculating Variance in Python

In Python, we can calculate the variance of an array using the NumPy `var()` function.

```
import numpy as np
values = np.array([1, 3, 4, 2, 6, 3, 4, 5])

# calculate variance of values
variance = np.var(values)
```

Standard Deviation Units

Because standard deviation is in the same units as the original data set, it is often used to provide context for the mean of the dataset. For example, if the data set is `[3, 5, 10, 14]`, the standard deviation is 4.301 units, and the mean is 8.0 units. By using the standard deviation, we can fairly easily see that the data point 14 is more than one standard deviation away from the mean.

Calculating Standard Deviation in Python

We can calculate standard deviation in Python using the NumPy `std()` function.

```
import numpy as np
values = np.array([1, 3, 4, 2, 6, 3, 4, 5])

# calculate standard deviation of values
variance = np.std(values)
```

Interpretation of Variance

A larger variance means the data is more spread out and values tend to be far away from the mean. A variance of 0 means all values in the dataset are the same.

Variance

Variance is a measure of spread. It is calculated by finding the average of the squared differences between every observation and the mean. The resulting value is in units squared.