

How to Calculate Cohen's Kappa in Python

Authored by
stats writer

November 29, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Calculate Cohen's Kappa in Python*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=101561>

Cohen's Kappa is a measure of inter-rater agreement between two raters, usually used in fields such as psychology or medicine. In Python, one can calculate Cohen's Kappa using the `cohen_kappa_score` function from the Scikit-learn library. This function requires two lists of ratings, one from each rater, as its parameters and returns a score between 0 and 1 that represents the level of agreement between the two raters.

In statistics, is used to measure the level of agreement between two raters or judges who each classify items into mutually exclusive categories.

The formula for Cohen's kappa is calculated as:

$$k = (po - pe) / (1 - pe)$$

where:

po: Relative observed agreement among raters

pe: Hypothetical probability of chance agreement

Rather than just calculating the percentage of items that the raters agree on, Cohen's Kappa attempts to account for the fact that the raters may happen to agree on some items purely by chance.

The value for Cohen's Kappa always ranges between 0 and 1 where:

0 indicates no agreement between the two raters

1 indicates perfect agreement between the two raters

The following table summarizes how to interpret different values for Cohen's Kappa:

Cohen's Kappa	Interpretation
0	No agreement
0.10 - 0.20	Slight agreement
0.21 - 0.40	Fair agreement
0.41 - 0.60	Moderate agreement
0.61 - 0.80	Substantial agreement
0.81 - 0.99	Near perfect agreement
1	Perfect agreement

The following example shows how to calculate Cohen's Kappa in Python.

Example: Calculating Cohen's Kappa in Python

Suppose two art museum curators are asked to rate 15 paintings on whether they're good enough to be shown in a new exhibit.

The following code shows how to use the `cohen_kappa_score()` function from the **sklearn** library to calculate Cohen's Kappa for the two raters:

```
from sklearn.metrics import cohen_kappa_score
```

```
#define array of ratings for both raters
```

```
rater1 =
```

```
rater2 =
```

```
#calculate Cohen's Kappa
```

```
cohen_kappa_score(rater1, rater2)
```

```
0.33628318584070793
```

Cohen's Kappa turns out to be **0.33628**.

Based on the table from earlier, we would say that the two raters only had a "fair" level of agreement.

You can use the function from the **statsmodels** library to calculate this metric.

Note: You can find the complete documentation for the `cohen_kappa_score()` function .

The following tutorials offer additional resources on Cohen's Kappa: