

# How can Cohen's Kappa be calculated in Python?

Authored by  
**stats writer**

June 28, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can Cohen's Kappa be calculated in Python?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=156388>

Cohen's Kappa is a statistical measure used to evaluate the level of agreement between two raters or evaluators. In order to calculate Cohen's Kappa in Python, the following steps can be followed:

1. Import the necessary libraries: The first step is to import the necessary libraries such as scikit-learn, numpy, and pandas.
2. Prepare the data: The data must be prepared in a format that can be used by the scikit-learn library. This includes converting the data into a numpy array or a pandas dataframe.
3. Calculate the observed agreement: Using the scikit-learn library, the observed agreement between the two raters can be calculated by passing the actual values and the predicted values.
4. Calculate the expected agreement: The expected agreement is calculated by creating a contingency table using the actual and predicted values and then computing the expected agreement using a formula.
5. Calculate Cohen's Kappa: Finally, Cohen's Kappa can be calculated by using the formula, which takes into account the observed and expected agreements.

By following these steps, one can easily calculate Cohen's Kappa in Python, providing a reliable measure of agreement between two raters or evaluators.

## Calculate Cohen's Kappa in Python

**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:**

$$k = (p_o - p_e) / (1 - p_e)$$

**where:**

**$p_o$ : Relative observed agreement among raters**  
 **$p_e$ :**

## 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 .**

#### **Additional Resources**

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

ARABPSYCHOLOGY.COM