

🏠 > API-интерфейс SciPy > Статистические функции (`scipy.stats`) > `f_oneway`

scipy.статистика.

f_oneway

`f_oneway` (* *выборки* , *ось* = 0 , *nan_policy* = 'распространять' , *keepdims* = Ложь)

[\[источник\]](#)

Проведите однофакторный дисперсионный анализ.

Однофакторный дисперсионный анализ проверяет нулевую гипотезу о том, что две или более групп имеют одинаковое среднее значение совокупности. Тест применяется к выборкам из двух или более групп, возможно, с разными размерами.

Параметры:

образец1, образец2, ... : *массив_подобный*

Измерения выборки для каждой группы. Должно быть не менее двух аргументов. Если массивы многомерные, то все измерения массива должны быть одинаковыми, за исключением *оси* .

ось : *int* или *None*, по умолчанию: 0

Если *int*, ось ввода, вдоль которой вычисляется статистика. Статистика каждого среза оси (например, строки) ввода появится в соответствующем элементе вывода. Если *None* , ввод будет разложен перед вычислением статистики.

nan_policy : {'распространять', 'опускать', 'повышать'}

Определяет, как обрабатывать входные NaN.

- `propagate` : если в срезе оси (например, строке), по которой вычисляется статистика, присутствует NaN, соответствующая запись выходных данных будет NaN.
- `omit` : NaN будут опущены при выполнении расчета. Если в срезе оси, по которому вычисляется статистика, остается недостаточно данных, соответствующая запись вывода будет NaN.
- `raise` : если присутствует NaN, `ValueError` будет выведено исключение.

keepdims : *bool*, по умолчанию: Ложь

Если установлено значение True, оси, которые уменьшаются, остаются в результате как измерения с размером 1. С этой опцией результат будет транслироваться правильно по отношению к входному массиву.

Возвраты:

статистика : *плавать*

Вычисленная F-статистика теста.

р-значение : *плавать*

Соответствующее р-значение из F-распределения.

Выдается, если все значения в каждом из входных массивов идентичны. В этом случае статистика F либо бесконечна, либо не определена, поэтому возвращается `np.inf` или `np.nan`

Предупреждение о выполнении

Выдается, если длина любого входного массива равна 0 или если все входные массивы имеют длину 1. `np.nan` В этих случаях возвращается для F-статистики и р-значения.

Примечания

Тест ANOVA имеет важные предположения, которые должны быть выполнены для того, чтобы соответствующее р-значение было действительным.

1. Выборки независимы.
2. Каждый образец взят из нормально распределенной популяции.
3. Стандартные отклонения популяций всех групп равны. Это свойство известно как гомоскедастичность.

Если эти предположения не верны для данного набора данных, все равно можно использовать H-тест Краскела-Уоллиса (`scipy.stats.kruskal`) или тест Александра-Говерна (`scipy.stats.alexandergovern`), хотя и с некоторой потерей мощности.

Длина каждой группы должна быть не менее одного, и должна быть хотя бы одна группа с длиной больше одного. Если эти условия не выполняются, генерируется предупреждение и возвращается (`np.nan` ,). `np.nan`).

Если все значения в каждой группе идентичны и существует хотя бы две группы с разными значениями, функция генерирует предупреждение и возвращает (`np.inf` , 0).

Если все значения во всех группах одинаковы, функция генерирует предупреждение и возвращает (`np.nan` , `np.nan`).

Алгоритм взят из Heiman [2] , стр.394-7.

Начиная с SciPy 1.9, `np.matrix` входные данные (не рекомендуется для нового кода) преобразуются в `np.ndarray` до выполнения вычисления. В этом случае вывод будет скаляром или `np.ndarray` соответствующей формы, а не 2D `np.matrix` . Аналогично, хотя маскированные элементы маскированных массивов игнорируются, вывод будет скаляром или, `np.ndarray` а не маскированным массивом с `mask=False` .

Ссылки

- [1] Р. Лоури, «Концепции и применение инференциальной статистики», Глава 14, 2014, <http://vassarstats.net/textbook/>
- [2] GW Heiman, «Понимание методов исследования и статистики: комплексное введение в психологию», Houghton, Mifflin and Company, 2001.
- [3] GH McDonald, «Справочник по биологической статистике», однофакторный дисперсионный анализ. <http://www.biostathandbook.com/onewayanova.html>

Примеры

```
>>> import numpy as np
>>> from scipy.stats import f_oneway
```

Ниже приведены некоторые данные [3] по измерению раковины (длина отпечатка передней приводящей мышцы, стандартизированная путем деления на длину) мидии *Mytilus trossulus* из пяти мест: Тилламук, штат Орегон; Ньюпорт, штат Орегон; Петербург, Аляска; Магадан, Россия; и Тварминне, Финляндия, взятые из гораздо большего набора данных, использованного в работе Макдональда и др. (1991).

```
>>> tillamook = [0.0571, 0.0813, 0.0831, 0.0976, 0.0817, 0.0859, 0.0735,
...             0.0659, 0.0923, 0.0836]
>>> newport = [0.0873, 0.0662, 0.0672, 0.0819, 0.0749, 0.0649, 0.0835,
...            0.0725]
>>> petersburg = [0.0974, 0.1352, 0.0817, 0.1016, 0.0968, 0.1064, 0.105]
>>> magadan = [0.1033, 0.0915, 0.0781, 0.0685, 0.0677, 0.0697, 0.0764,
...            0.0689]
>>> tvarminne = [0.0703, 0.1026, 0.0956, 0.0973, 0.1039, 0.1045]
>>> f_oneway(tillamook, newport, petersburg, magadan, tvarminne)
F_onewayResult(statistic=7.121019471642447, pvalue=0.0002812242314534544)
```

`f_oneway` принимает многомерные входные массивы. Когда входные данные многомерные и ось не указана, тест выполняется по первой оси входных массивов. Для следующих данных тест выполняется три раза, по одному разу для каждого столбца.

```
>>> a = np.array([[9.87, 9.03, 6.81],
...               [7.18, 8.35, 7.00],
...               [8.39, 7.58, 7.68],
...               [7.45, 6.33, 9.35],
...               [6.41, 7.10, 9.33],
...               [8.00, 8.24, 8.44]])
>>> b = np.array([[6.35, 7.30, 7.16],
...               [6.65, 6.68, 7.63],
...               [5.72, 7.73, 6.72],
...               [7.01, 9.19, 7.41],
...               [7.75, 7.87, 8.30],
...               [6.90, 7.97, 6.97]])
>>> c = np.array([[3.31, 8.77, 1.01],
...               [8.25, 3.24, 3.62],
...               [6.32, 8.81, 5.19],
...               [7.48, 8.83, 8.91],
...               [8.59, 6.01, 6.07],
...               [3.07, 9.72, 7.48]])
>>> F = f_oneway(a, b, c)
>>> F.statistic
array([1.75676344, 0.03701228, 3.76439349])
>>> F.pvalue
array([0.20630784, 0.96375203, 0.04733157])
```

[< Предыдущий](#)
[кстест](#)

[Следующий >](#)
[tukey_hsd](#)