

# Ранговая корреляция (Rank correlation)

Синонимы: Rank-order correlation

Разделы: [Метрики](#)

Ранговая корреляция представляет собой непараметрический (т.е. независящий от предположений о типе распределения исходных данных и его параметров) метод корреляционного анализа, предназначенный для описания зависимости между признаками, представленными в порядковой (ранговой) шкале.

Ранговая корреляция широко применяется при решении задач в области экономики, бизнеса, социальной сферы и других областях, где распространены порядковые признаки, т.е. те, в которых объекты описываются не только значением, но и положением относительно значений признака (порядка) других объектов. Например, клиенты могут упорядочиваться по уровню дохода, трудовому стажу и т.д. Предприятия — по обороту, численности персонала и др. Товары — по цене, объему продаж, запасу на складе.

Ранги — это порядковые номера единиц наблюдения, упорядоченных по значениям исследуемого признака. Если ранжировать набор данных по двум признакам, зависимость между которыми требуется определить, то полное совпадение рангов означает максимально высокую прямую зависимость, а полная противоположность рангов — максимально высокую обратную зависимость. Ранжировать оба признака необходимо в одном и том же порядке: либо от меньших значений признака к большим, либо наоборот.

Рассмотрим пример присвоения рангов признакам. Пусть имеется набор данных, который содержит два признака — возраст клиента сотовой компании, и число звонков, совершенное им в неделю. Задача заключается в том, чтобы определить, связаны ли эти признаки.

<b>Возраст</b>	<b>Число звонков</b>
24	12
51	111
41	41
35	47

<b>Возраст</b>	<b>Число звонков</b>
26	58
26	45
50	61
66	10
36	34
40	79

Произведем их упорядочивание и присвоение рангов.

<b>Ранг</b>	<b>Возраст</b>	<b>Число звонков</b>
1	24	10
2	26	12
3	26	34
4	35	41
5	36	45
6	40	47
7	41	58
8	50	61
9	51	79
10	66	111

Возвращаемся к исходному порядку следования значений с указанием рангов:

<b>Возраст</b>	<b>Ранг</b>	<b>Число звонков</b>	<b>Ранг</b>
24	1	12	2
51	9	111	10
41	7	41	4
35	4	47	6

Возраст	Ранг	Число звонков	Ранг
26	2.5	58	7
26	2.5	45	5
50	8	61	8
66	10	10	1
36	5	34	3
40	6	79	9

Если среди ранжируемых значений попадаются одинаковые, то им присваивается ранг, равный среднему значению рангов, которые были бы им присвоены, если бы значения различались. В нашем примере у признака «Возраст» встречаются два одинаковых значения 26. При упорядочивании им должны быть присвоены ранги 2 и 3, но в соответствии с упомянутым правилом им будут присвоены одинаковые ранги 2.5.

Общая сумма рангов должна совпадать с расчетной, которая определяется по формуле:

$$\sum r = \frac{N \cdot (N+1)}{2},$$

где  $N$  — количество ранжируемых значений.

Основной задачей статистического исследования, проводимого с помощью ранговой корреляции, является вычисление коэффициентов ранговой корреляции — статистических характеристик, являющихся мерой зависимости двух порядковых признаков. Существуют несколько видов коэффициентов ранговой корреляции:

- Спирмена (Spearman),  $\rho$ , ро-коэффициент;
- Кендалла (Kendall),  $\tau$ , тау-коэффициент;
- Гудмана и Краскала (Goodman and Kruskal's),  $\gamma$ , гамма-коэффициент;
- Сомерса (Somers),  $D$ ,  $D$ -коэффициент.

Как и обычный коэффициент корреляции Пирсона, все упомянутые выше коэффициенты корреляции изменяются в диапазоне от -1 до 1, и интерпретируются схожим образом:

- значения, близкие к 1 указывают на высокую положительную связь между ранговыми признаками — если значения одного признака с увеличением ранга растут, то и другого тоже;
- значения, близкие к 0 указывают на отсутствие зависимости — если значения одного признака с увеличением ранга растут, то и другого могут как расти, так и уменьшаться;
- значения, близкие к -1 указывают на высокую отрицательную зависимость — когда значения одного признака с увеличением ранга растут, другого всегда уменьшаются.

В Logipom существует специализированный обработчик Корреляционный анализ с использованием которого можно рассчитывать коэффициенты ранговой корреляции Спирмена и Кендалла.