

Коэффициент ранговой корреляции Спирмена (Spearman's rank correlation coefficient)

Синонимы: Spearman's rank-order correlation coefficient, rho-коэффициент

Разделы: [Метрики](#)

Как и любой коэффициент [ранговой корреляции](#), коэффициент Спирмена используется для обнаружения и описания статистической зависимости между признаками, а также проверки гипотез о наличии этой зависимости. Он был предложен английским статистиком и психологом Чарльзом Спирманом в 1904 году и обычно обозначается греческой буквой ρ или r_s . Получил популярность при решении задач в [бизнес-аналитике](#).

Расчет коэффициента ранговой корреляции Спирмена включает следующие этапы:

1. Присвоить по каждому из признаков, зависимость между которыми требуется определить, порядковый номер (ранг) для каждого наблюдения по возрастанию или убыванию.
2. Определить разности рангов каждой пары сопоставляемых значений (d).
3. Возвести в квадрат каждую разность и суммировать полученные результаты.
4. Вычислить коэффициент ранговой корреляции по формуле:

$$\rho = 1 - \frac{6 \cdot \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \quad (1)$$

где d_i — разность между рангами $d_i = R(X_i) - R(Y_i)$, n — число наблюдений, по которым вычисляется коэффициент.

Следует отметить, что данная формула может использоваться только если все n рангов являются различными целыми числами. В противном случае следует использовать другое выражение:

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}},$$

где \bar{x} и \bar{y} средние значения признаков.

Определить статистическую значимость коэффициента можно с использованием t -критерия [Стьюдента](#), рассчитанного по следующей формуле:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Если вычисленное значение t -критерия меньше табличного при заданном числе степеней свободы, статистическая значимость искомой зависимости низкая. Если больше, то корреляционная связь считается статистически значимой.

При использовании коэффициента ранговой корреляции условно оценивают силу зависимости между признаками следующим образом:

- $\rho < 0.3$ — слабая зависимость;
- $0.3 \leq \rho < 0.7$ — умеренная зависимость;
- $\rho \geq 0.7$ — сильная зависимость.

Также для оценки тесноты связи может использоваться шкала Чеддока:

Абсолютное значение коэффициента Спирмена	Сила корреляционной зависимости
менее 0.3	слабая
от 0.3 до 0.5	умеренная
от 0.5 до 0.7	заметная
от 0.7 до 0.9	высокая
более 0.9	очень высокая

Рассмотрим пример. Пусть задан исходный набор данных:

Признак 1	Признак 2
56	66
75	70
45	40
71	60
62	65
64	56
58	59
80	77
76	67

Признак 1	Признак 2
61	63

Упорядочим значения признаков по убыванию и каждому присвоим ранг

Ранг	Признак 1	Признак 2
1	80	77
2	76	70
3	75	67
4	71	66
5	64	65
6	62	63
7	61	60
8	58	59
9	56	56
10	45	40

Вычислим d и d^2 :

Признак 1	Признак 2	R1	R2	$d=R1-R2$	d^2
56	66	9	4	5	25
75	70	3	2	1	1
45	40	10	10	0	0
71	60	4	7	3	9
62	65	6	5	1	1
64	56	5	9	4	16
58	59	8	8	0	0
80	77	1	1	0	0
76	67	2	3	1	1

Признак 1	Признак 2	R1	R2	d=R1-R2	d^2
61	63	7	6	1	1

Где R_1 и R_2 ранги Признака 1 и Признака 2 соответственно.

Вычисляем сумму квадратов разностей из формулы 1:

$$\sum d^2 = 25 + 1 + 0 + 9 + 1 + 16 + 0 + 0 + 1 + 1 = 54$$

Теперь вычислим коэффициент ранговой корреляции по формуле 1:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2-1)} = \frac{6 \cdot 54}{10(10^2-1)} = 0.67$$

Таким образом в исходном наборе данных имеет место заметная положительная зависимость между признаками.

Преимущество коэффициента ранговой корреляции Спирмена заключается в возможности ранжирования по качественным признакам, которые нельзя выразить численно: можно ранжировать субъективные экспертные оценки, например, популярности товара или услуги, степени лояльности клиента или надежности заемщика. Т.е. он подходит как для непрерывных, так и для дискретных порядковых переменных. Можно ранжировать оценки разных экспертов и найти их корреляции, чтобы затем исключить из рассмотрения оценки эксперта, слабо коррелированные с оценками других экспертов.

Не рекомендуется использовать коэффициент Спирмена, если имеет место неравномерное распределение значений измеряемой величины, т.е. в данных присутствуют аномальные значения и выбросы.

В Loginot существует специализированный обработчик Корреляционный анализ с использованием которого можно рассчитывать коэффициент ранговой корреляции Спирмена.