

Коэффициент корреляции рангов Спирмена

Выполнила Букина Людмила

Студентка группы С-1841

Корреляция

Корреляционный анализ – статистический метод, предназначенный для выявления существования зависимости между 2 и более случайными величинами (переменными), а также ее силы.

Варианты развития

событий

```
graph TD; A[событий] --> B[Наличие корреляции]; A --> C[Отсутствие корреляции];
```

Наличие корреляции

(положительная или отрицательная)

Отсутствие корреляции

(отрицательная)

Существование корреляционной связи **может отобразиться** графически(графики) и с помощью коэффициента (числовое отображение).

Корреляционная связь **характеризуется** силой связи (сильная, средняя, слабая) и направлением (прямая или обратная).

Методы корреляционного анализа

Метод квадратов или Пирсона

- Применим для расчетов, требующих точного определения силы, существующей между переменными
- Изучаемые с его помощью признаки должны выражаться только количественно.

Ранговый метод или Спирмена

- Нет требований в выражении признаков - могут выражаться количественно и качественно (атрибутивно).
- Информация имеет ориентировочный характер.
- В рядах могут быть открытые варианты (н-р, стаж работы более лет).

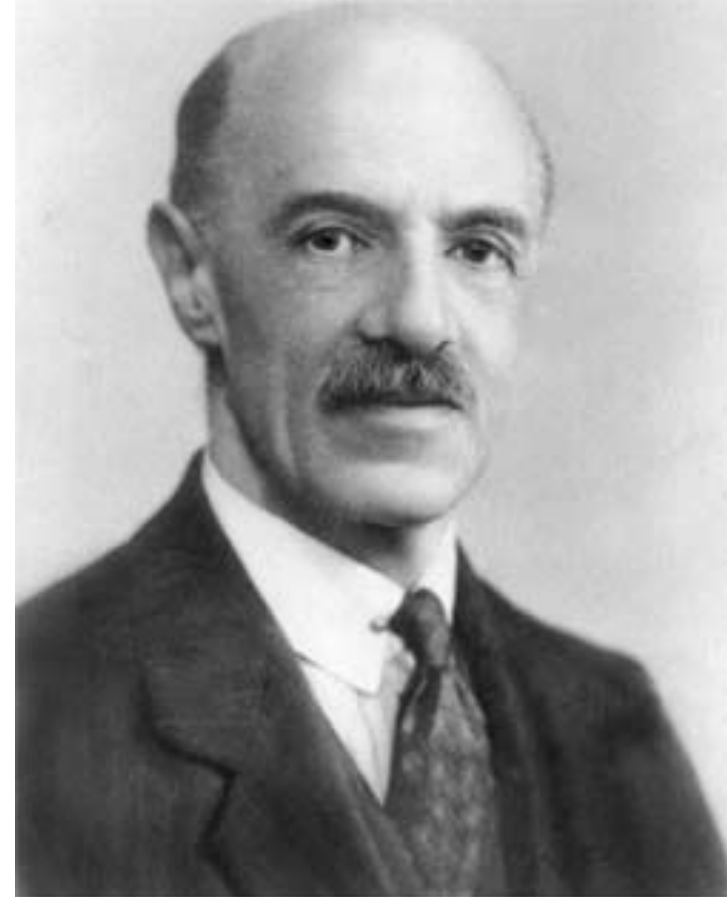
Описание метода ранговой корреляции Спирмена

- Используется при наличии 2 рядов значений, подвергающихся ранжированию;
 - Предполагает проведение ранжирования показателей в отдельности для каждого признака;
 - Наименьшее значение имеет наименьший ранг;
 - Относится к непараметрическому статистическому методу, предназначенному для установления существования связи изучаемых явлений:
1. Определение связи между 2 рядами количественных данных.
 2. Оценка тесноты выявленной связи, выражаемой количественно.

Описание метода ранговой корреляции Спирмена

Ряды могут выражаться:

- Парой признаков, определяемых в одной и той же группе исследуемых объектов
- Парой индивидуальных соподчинённых признаков
- Парой групповых соподчинённых признаков
- Индивидуальной или групповой соподчинённостью признаков



Коэффициент корреляции

Статистическая величина, определяющая характер изменения двух переменных.

В количественном выражении колеблется в пределах от -1 до $+1$.

Наиболее распространены коэффициенты:

- Пирсона – применим для переменных, принадлежащих к интервальной шкале;
- Спирмена – для переменных порядковой шкалы

Ранги Спирмена

- Коэффициент ранговой корреляции Спирмена позволяет статистически установить наличие связи между явлениями. Его расчет предполагает установление для каждого признака порядкового номера – ранга. Ранг может быть возрастающим или убывающим.
- Количество признаков, подвергаемых ранжированию, может быть любым. Затруднения начинаются при достижении 20 признаков.

Формула

$$\rho_{xy} = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

n – отображает количество ранжируемых признаков;
 d - разность между рангами по двум переменным;
 $6\sum(d^2)$ – сумма квадратов разностей рангов.

Алгоритм корреляционного анализа по критерию Спирмена

1. Парные сопоставимые признаки располагаются в 2 ряда, один из которых обозначается с помощью X, а другой Y;
2. Значения ряда X располагаются в порядке возрастания либо убывания;
3. Последовательность расположения значений ряда Y определяется их соответствием значений ряда X;
4. Для каждого значения в ряду X определить ранг – присвоить порядковый номер от минимального значения к максимальному;
5. Для каждого значения в ряду Y также определить ранг (от минимального к максимальному);
6. Вычислить разницу (D) между рангами X и Y, прибегнув к формуле $D=X-Y$;
7. Полученные значения разницы возвести в квадрат;
8. Выполнить суммирование квадратов разниц рангов;
9. Выполнить расчеты по формуле.

Пример корреляции Спирмена

Необходимо установить наличие корреляционной связи между рабочем стажем и показателем травматизма при наличии следующих данных:

Рабочий стаж в годах	Травматизм на 100 работающих
до 1 года	24
1-2	16
3-4	12
5-6	12
7 и более	6

Наиболее подходящим методом анализа является ранговый метод, так как один из признаков представлен в виде открытых вариантов: рабочий стаж до 1 года и рабочий стаж 7 и более лет.

Решение задачи начинается с ранжирования данных, которые сводятся в рабочую таблицу.

В случае появления одинаковых по величине показателей, находится среднее арифметическое значение ранга ($2+3/2=2,5$)

Рабочий стаж	Число травм	Ранги (X)	Ранги (Y)	Разность рангов (D)	Квадрат разности рангов
до 1 года	24	1	5	-4	16
1-2	16	2	4	-2	4
3-4	12	3	2,5	0,5	0,25
5-6	12	4	2,5	1,5	2,5
7 и более	6	5	1	4	16

$$\dot{r}_{xy} = 1 - \frac{6 \times 38,5}{5 - (5^2 - 1)} = 1 - \frac{325}{5 - (25 - 1)} = 1 - \frac{325}{120} = 1 - 1,95 = -0,92$$

Отрицательное значение коэффициента свидетельствует о наличии **обратной связи** между признаками и позволяет утверждать, что небольшой стаж работы сопровождается большим числом травм. Причем, **сила** этих показателей достаточно **большая**.

Задача для самостоятельного решения

Два преподавателя оценили знания 12 учащихся по стобалльной системе и выставили им следующие оценки (в первой строке указано количество баллов, выставленных первым преподавателем, а во второй – вторым):

98, 94, 88, 80, 76, 70, 63, 61, 60, 58, 56, 51.

99, 91, 93, 74, 78, 65, 64, 66, 52, 53, 48, 62.

Найти выборочный коэффициент ранговой корреляции Спирмена между оценками двух преподавателей.

Решение

Оценка 1 преподавателя	Оценка 2 преподавателя	Ранги (X)	Ранги (Y)	Разность рангов (D)	Квадрат разности рангов
98	99	1	1	0	0
94	91	2	3	-1	1
88	93	3	2	1	1
80	74	4	5	-1	1
76	78	5	4	1	1
70	65	6	7	-1	1
63	64	7	8	-1	1
61	66	8	6	2	4
60	52	9	11	-2	4
58	53	10	10	0	0
56	48	11	12	-1	1
51	62	12	9	3	9

- $$r_{xy} = 1 - \frac{6 \times 24}{12^3 - 12} \approx 0,916$$

Положительное значение коэффициента свидетельствует о наличии прямой связи между признаками и позволяет утверждать, что оценки преподавателей имеют высокую связь между собой.