

МЕТОД РАНГОВОЙ КОРРЕЛЯЦИИ И ЕГО ПРИМЕНЕНИЕ

Корнилова А.С.¹, Никонова Р.А.², Дрягина Д.Р.³

¹Корнилова Анастасия Сергеевна - магистрант;

²Никонова Рада Андреевна – магистрант;

³Дрягина Дарья Романовна – магистрант,

кафедра экологии промышленных зон и акваторий,

Федеральное государственное бюджетное образовательное учреждение высшего образования

Санкт-Петербургский государственный морской технический университет,

г. Санкт-Петербург

Аннотация: в данной статье описаны математико-статистические методы исследования и их значение для изучения различных параметров и явлений. Рассматривается корреляционный анализ и его значение, коэффициент корреляции и критерии для его определения. Дано определение ранговой корреляции и ее коэффициента, более подробно изучается метод ранговой корреляции и его применение, а также коэффициент ранговой корреляции Спирмена и его алгоритм вычисления. Приведен пример расчета ранговой корреляции Спирмена.

Ключевые слова: корреляционный анализ, коэффициент корреляции, ранговая корреляция, коэффициент ранговой корреляции Спирмена.

В наше время математико-статистические исследования являются необходимым инструментом для получения более полных знаний о механизме изучаемых явлений. Предметом статистики являются такие явления, которые имеют количественный характер и изучаются с помощью количественных методов. Одним из основных является метод корреляционно-регрессивного анализа. Статистический метод, предназначенный для выявления существования зависимости между двумя и более случайными величинами (переменными), а также ее силы, получил название корреляционного анализа.

При изучении корреляций стараются установить, существует ли какая-то связь между двумя показателями в одной выборке либо между двумя различными выборками, и если эта связь существует, то сопровождается ли увеличение одного показателя возрастанием (положительная корреляция) или уменьшением (отрицательная корреляция) другого.

Корреляционный анализ помогает установить, можно ли предсказывать возможные значения одного показателя, зная величину другого.

Коэффициент корреляции — это величина, которая может изменяться в пределах от +1 до -1. В случае полной положительной корреляции этот коэффициент равен +1, при полной отрицательной — -1.

В случае если эти точки не выстраиваются по прямой линии, а образуют «облако», коэффициент корреляции по абсолютной величине становится меньше единицы и по мере округления этого облака приближается к нулю.

Если же коэффициент корреляции равен 0, обе переменные полностью независимы друг от друга.

Корреляция считается сильной, если коэффициент находится в пределах от ±0,7 до ±1; от ±0,3 до ±0,699 – средней; от 0 до ±0,299 – слабой [1].

Ранговая корреляция – это мера зависимости между случайными величинами (наблюдаемыми признаками, переменными), когда эту зависимость невозможно определить количественно с помощью обычного коэффициента корреляции. Процедура установления ранговой корреляции заключается в упорядочении изучаемых объектов в отношении некоторого признака, т. е. им приписываются порядковые номера — ранги (по два номера в соответствии с двумя наблюдаемыми признаками, между которыми исследуется корреляция). Например, наибольшее значение для переменной обозначается номером 1, второе по величине — номером 2 и т. д. Наиболее распространен коэффициент ранговой корреляции (коэффициент Спирмена).

Коэффициент ранговой корреляции Спирмена позволяет статистически установить наличие связи между событиями. Его расчет предполагает установление для каждого признака порядкового номера – ранга. Ранг может быть убывающим или возрастающим. Количество ранжируемых признаков может быть любым. [2]

Для расчета коэффициента Спирмена пользуются формулой:

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

Проведение корреляционного анализа с использованием метода Спирмена выполняется по следующему алгоритму: парные сопоставимые признаки располагаются в 2 ряда, один из которых обозначается с помощью X, другой Y; значения ряда X располагаются в порядке возрастания либо убывания; последовательность расположения значений ряда Y определяется их соответствием значений ряда X; для каждого значения в ряду X определить ранг — присвоить порядковый номер от минимального значения к максимальному; для каждого из значений в ряду Y также определить ранг (от

минимального к максимальному); вычислить разницу (D) между рангами X и Y, прибегнув к формуле $D=X-Y$; полученные значения разницы возводятся в квадрат; выполнить суммирование квадратов разниц рангов; выполнить расчеты по формуле 1.

Пример расчета корреляции Спирмена. Необходимо установить наличие корреляционной связи между рабочим стажем и показателем травматизма при наличии следующих данных:

Таблица 1. Исходные данные

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

Ранжированные данные сводятся в рабочую таблицу.

Таблица 2. Ранжированные данные

Рабочий стаж, лет	Число травм	Порядковые номера	Ранги	Разность рангов	Квадрат разности рангов
до 1	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
$\Sigma d^2=38,5$					

Появление дробных рангов в колонке связано с тем, что в случае появления вариант одинаковых по величине находится среднее арифметическое значение ранга. В данной задаче показатель травматизма 12 встречается дважды, ему присваиваются ранги 2 и 3, среднее арифметическое этих рангов $(2+3)/2=2,5$ помещается в рабочую таблицу для 2 показателей. После подстановки полученных значений в рабочую формулу и расчетов получаем коэффициент Спирмена равный -0,92.

$$r_{xy} = 1 - \frac{6 \cdot 38,5}{5(5^2 - 1)} = 1 - \frac{325}{5(25 - 1)} = 1 - \frac{325}{120} = 1 - 1,92 = -0,92$$

Отрицательное значение коэффициента указывает на наличие обратной связи между признаками и свидетельствует о том, что небольшой стаж работы сопровождается большим количеством травм. При этом, сила связи данных показателей достаточно большая. [3]

Список литературы

1. Гмурман В.Е. Теория вероятностей и математическая статистика: Учебное пособие для вузов. 10-е издание, стереотипное. Москва: Высшая школа, 2004. 479 с.
2. Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник / Под ред. И. И. Елисеевой. 4-е издание, переработанное и дополненное. Москва: Финансы и Статистика, 2002. 480 с.
3. Общая теория статистики: Учебник / Под ред. Р.А. Шмойловой. 3-е издание, переработанное. Москва: Финансы и Статистика, 2002. 560 с.