



Корреляционный анализ

Описание

Компонент на основании вычисляемых коэффициентов корреляции выявляет взаимосвязь между рядами данных входного набора. Применяется для оценки предполагаемой зависимости факторов.

▼ Пример:

Проведем корреляционный анализ на основе небольшой таблицы, содержащей данные о количестве продаж четырех видов товаров по датам.

Исходная таблица:

Дата	Спагетти	Томатная паста	Макароны	Кофе
02.09.17	10	20	15	25
03.09.17	12	22	12	26
04.09.17	14	25	9	26
05.09.17	13	24	10	25
06.09.17	14	25	9	24
07.09.17	14	25	9	23
08.09.17	12	21	12	24
09.09.17	10	18	14	23
10.09.17	16	24	9	22
11.09.17	13	21	9	23
12.09.17	17	25	7	25

Определим корреляцию товара "Спагетти" с остальными товарами на основе коэффициентов Пирсона (в мастере настройки узла поле "Спагетти" отметим в "Наборе 1", а остальные товары - в "Наборе 2").

Выходная таблица:

Метка	Метка	Коэфф. Пирсона
Спагетти	Томатная паста	0,83

Метка	Метка	Коэфф. Пирсона
Спагетти	Макароны	-0,93
Спагетти	Кофе	-0,12

Как видно из таблицы, ряд продаж для товара "Томатная паста" имеет очень большую положительную корреляцию, а товара "Макароны" — отрицательную. Из этого можно сделать вывод, что "Томатная паста" является сопутствующим товаром, а "Макароны" — заместителем товара "Спагетти". Корреляция продаж товара "Кофе" с товаром "Спагетти" является отрицательной, но при этом абсолютное значение корреляции невелико, и, следовательно, можно говорить об отсутствии взаимосвязи между продажами этих товаров.

Порты

Вход

- Входной источник данных (таблица данных).

Выход

- Выходной набор данных** — таблица, содержащая данные о корреляции между полями. Имеет следующую структуру:
 - Обязательные поля:
 - Поля|Имя** — имя первого поля в корреляционной паре;
 - Поля|Метка** — метка первого поля в корреляционной паре;
 - Поля|Имя** — имя второго поля в корреляционной паре;
 - Поля|Метка** — метка второго поля в корреляционной паре.
 - Поля, наличие которых задается пользователем:
 - Пирсона** — коэффициенты корреляции Пирсона.
 - Экстремум К. Ф.** — экстремумы взаимнокорреляционной функции.
 - Лаг К.Ф.** — величина смещения, при которой экстремум ВКФ был получен. Значение лага выводится только для взаимнокорреляционной функции.
 - Тау-в Кендалла** — коэффициенты ранговой корреляции Кендалла.
 - Спирмена** — коэффициенты корреляции Спирмена.

Мастер настройки

Включает список флагов, позволяющих выбрать коэффициенты для оценки корреляции:

- Коэффициент корреляции Пирсона** — с его помощью можно определить силу и направление линейной зависимости между двумя процессами, происходящими одновременно.

- **Коэффициент Тау-в Кендалла** — коэффициент ранговой корреляции, применяется для выявления количественной взаимосвязи между переменными, если их можно ранжировать. Рекомендуется использовать для категориальных данных.
- **Экстремум взаимнокорреляционной функции** — вычисляет максимальный по модулю из коэффициентов корреляции двух процессов, рассчитанных при всевозможных временных сдвигах. Следует применять, если необходимо узнать линейную зависимость между двумя процессами или частями процессов, происходящими с некоторым временным лагом.
- **Коэффициент корреляции Спирмена** — еще один вариант ранговой корреляции. Для числовых полей для оценки силы связи используются не численные значения, а соответствующие им ранги. Поэтому для любых монотонных последовательностей коэффициент Спирмена будет равен -1 или 1.

В таблице необходимо выбрать ряды для анализа взаимосвязи. Для каждого поля из "Набора 1", отмеченного флагом, будут вычислены коэффициенты корреляции с полями, отмеченными флагами в "Наборе 2".

Примечание: При расчете Коэффициента корреляции Пирсона цифровые данные считаются "непрерывными" вне зависимости от типа входных данных (непрерывные или дискретные). Если на вход подаются текстовые данные, являющиеся "дискретными", то расчет Коэффициента корреляции Пирсона происходит следующим образом:

1. данные сортируются по алфавиту (по возрастанию),
2. каждой уникальной записи присваивается порядковый номер,
3. рассчитывается Коэффициент корреляции Пирсона по полученным уникальным номерам.

Смотри также:

- [Мультиколлинеарность](#)