

Коэффициенты ранговой корреляции

Поиск взаимосвязей в данных

Когда коэффициент Пирсона не применим?

1. Распределение признаков отличается от нормального.
2. Признаки измерены неметрической (порядковой) шкалой.
3. В выборке есть смесь.

Коэффициенты ранговой корреляции

- Коэффициент Спирмена
- Коэффициент тау Кендалла

Так же, как и коэффициент корреляции Пирсона:

- показывают наличие линейной взаимосвязи между признаками, её силу и направленность;
- изменяются в диапазоне $[-1; 1]$.

В отличие от коэффициента корреляции Пирсона **основаны не на абсолютных значениях, а на рангах.**

Коэффициент Пирсона

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

В основе — отклонения каждого значения от среднего по X и по Y.

Коэффициент Спирмена

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

В основе — разница рангов.

d^2 — квадрат разницы рангов x_i и y_i

n — объём выборки

Коэффициент Пирсона: для интервальных нормально распределённых данных.

Коэффициент Спирмена: для неметрических данных и данных с аномалиями.