

Линейные взаимосвязи между признаками. Коэффициенты корреляции

Поиск взаимосвязей в данных

Коэффициент корреляции Пирсона

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

Диапазон значений r : от -1 до 1.

$r = -1$ означает, что все значения лежат вдоль линии **обратной** корреляции.

$r = +1$ означает, что все значения лежат вдоль линии **прямой** корреляции.

Величина r показывает **силу связи**.

Недостатки коэффициента корреляции Пирсона

1. Чувствителен к выбросам.
2. Не показывает нелинейные формы связи.