

Коэффициент корреляции (Correlation coefficient)

Разделы: [Метрики](#)

Loginom: [Корреляционный анализ \(обработчик\)](#), [Статистика \(визуализатор\)](#).

В [математической статистике](#) — показатель, характеризующий силу статистической связи между двумя или несколькими случайными величинами.

Если коэффициент корреляции описывает связь между двумя случайными величинами, то он называется простым, если между одной случайной величиной и их группой, то множественным.

Простой коэффициент корреляции (Пирсона) вычисляется по формуле:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x\sigma_y},$$

где n — число статистических наблюдений, x и y — случайные переменные. Значения коэффициента корреляции всегда расположены в диапазоне от -1 до 1 и интерпретируются следующим образом:

- если коэффициент корреляции близок к 1, то между переменными наблюдается положительная [корреляция](#). Иными словами, отмечается высокая степень связи между переменными. В данном случае, если значения переменной x будут возрастать, то и выходная переменная также будет увеличиваться;
- если коэффициент корреляции близок к -1, это означает, что между переменными имеет место сильная отрицательная корреляция. Иными словами, поведение выходной переменной будет противоположным поведению входной. Если значение x будет возрастать, то y будет уменьшаться, и наоборот;
- промежуточные значения, близкие к 0, будут указывать на слабую корреляцию между переменными и, соответственно, низкую зависимость. Иными словами, поведение переменной x не будет совсем (или почти совсем) влиять на поведение y (и наоборот).

Коэффициент корреляции равен квадратному корню [коэффициента детерминации](#), поэтому может применяться для оценки значимости [регрессионных моделей](#).

Очевидно, что если корреляция между переменными высокая, то, зная поведение входной переменной, проще предсказать поведение выходной, и полученное предсказание будет точнее (говорят, что входная переменная хорошо «объясняет»

выходную). Однако чем выше корреляция наблюдается между переменными, тем очевиднее связь между ними, например, взаимозависимость между ростом и весом людей.

Коэффициент корреляции Пирсона описывает только степень линейной связи и применим к непрерывным величинам. Для дискретных (качественных) данных используются коэффициенты ранговой корреляции Кендалла или Спирмена.

Коэффициент корреляции широко применяется в анализе данных для отбора переменных в аналитические модели и выявления наиболее значимых признаков с точки зрения решаемой задачи.