

**Как делать статстически значимые выводы о зависимости?**

# Статистические тесты

- В общем случае, проверить на зависимость выборки не представляется возможным, так как корреляция  $\neq$  зависимость. Поэтому мы можем проверять только выборки на скоррелированность, и из смысла задачи понимая под этим наличие зависимости или независимость.
- Гипотеза выглядит следующим образом:

$$H_0: \rho(X, Y) = 0$$

$$H_a: \rho(X, Y) \neq 0$$

# Статистические тесты

У коэффициентов корреляции Пирсона и Спирмена имеется известное распределение при верности нулевой гипотезы.

На основании них строятся тесты, реализованные, например, в пакете `scipy.stats` (функции `pearsonr` и `spearmanr`), которые помимо самого значения коэффициента корреляции возвращают `p-value` для статистического теста.

Тесты проверяют только наличие или отсутствие **КОРРЕЛЯЦИИ** в данных!

# Статистические тесты

Но есть одно важное исключение из правила:

Если выборки  $X$  и  $Y$  из нормального распределения, то равенство коэффициента корреляции 0 эквивалентно независимости.

Для нормальных выборок нужно применять тест на основе коэффициента корреляции **Пирсона**, и нулевая гипотеза в нём будет о независимости выборок.

Для остальных распределений для обнаружения монотонной зависимости лучше пользоваться критерием корреляции Спирмена.