



Выборочные коэффициенты корреляции, значимость, интерпретация.

Выборочный коэффициент корреляции является одним из основных показателей тесноты связи между двумя переменными. При изучении зависимости переменной Y от переменной X выборочный коэффициент корреляции обозначается как r_{xy} . При изучении зависимости переменной X от переменной Y выборочный коэффициент корреляции обозначается как r_{yx} .

Пусть при проведении некоторого опыта наблюдаются две случайные величины X и Y , причем одно и то же значение x встречается n_{xy} раз, y - n_{yx} раз, одна и та же пара чисел (x, y) наблюдается n_{xy} раз. Все данные записываются в виде таблицы, которую называют корреляционной.

Выборочная ковариация $k(X, Y)$ величин X и Y определяется формулой

$$k(X, Y) = \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})n_{xy},$$

где $n = \sum n_{xy}$, а \bar{x} , \bar{y} - выборочные средние величин X и Y .

Выборочный коэффициент корреляции находится по формуле

$$r(X, Y) = \frac{k(X, Y)}{\sigma_x^* \cdot \sigma_y^*} = \frac{\sum n_{xy}xy - \bar{x}\bar{y}}{n\sigma_x^* \cdot \sigma_y^*},$$

где σ_x^* , σ_y^* - выборочные средние квадратические отклонения величин X и Y .

Чем ближе $|r(X, Y)|$ к единице, тем сильнее линейная связь между X и Y .

Коэффициент корреляции не рассчитывается:

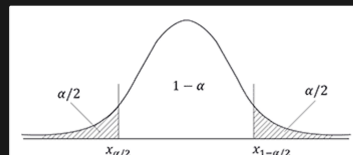
- когда соотношение между двумя переменными не линейное, например, квадратичное;
- в данных имеется больше 1-го наблюдения по каждому случаю;
- имеются аномальные наблюдения (выбросы, «отщепенцы»);
- данные содержат ярко выраженные подгруппы наблюдений.

Значимость коэффициента корреляции

Значимость коэффициента корреляции определяется по формуле:

$$T_{набл} = r_{xy} \frac{\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$$

Для оценки значимости коэффициента корреляции используют критерий Стьюдента. По таблице Стьюдента находим $T_{табл}(n-m-1; \alpha/2)$



Если $T_{набл} > T_{табл}$ (наблюдаемое значение критерия $T_{набл}$ принадлежит критической области), то отклоняем гипотезу о равенстве 0 коэффициента корреляции. Другими словами, коэффициента корреляции статистически - значим.

Интерпретация

- При значении КК равным 1, следует понимать, что при каждом изменении 1-й переменной происходит эквивалентное изменение 2-й переменной в том же направлении.
- Если значение КК равно -1, то при каждом изменении происходит эквивалентное изменение второй переменной в противоположном направлении.

Интерпретация значений коэффициента корреляции	
Значение	Интерпретация
до 0,2	Очень слабая
до 0,5	Слабая
до 0,7	Средняя
до 0,9	Высокая
свыше 0,9	Очень высокая корреляция

<http://cito-web.yspu.org/link1/metod/theory/node41.html>
<http://statistica.ru/theory/znachimost-koeffitsienta-korreljatsii-doveritelnyy-interval/?ysclid=l4snohgz7a664344809>
[https://portal.tpu.ru/SHARED/i/IOM/Tek_kursi/Tab_LK/2 Коппел-пергесс анализ.pdf](https://portal.tpu.ru/SHARED/i/IOM/Tek_kursi/Tab_LK/2%20Коппел-пергесс%20анализ.pdf)
<https://math.semestr.ru/corel/prim5.php?ysclid=l4sody65g740279406>
<https://tocodata.com/marketing/korreljaciya.html?ysclid=l4sojk6ulr759109123>