

## билет 31

**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  $p_x$  раз,  $y$  –  $p_y$  раз, одна и та же пара чисел  $(x, y)$  наблюдается  $p_{xy}$  раз. Все данные записываются в виде таблицы, которую называют корреляционной.

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

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

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

**Выборочный коэффициент корреляции** находится по формуле

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

где  $\sigma_x^*, \sigma_y^*$  – 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.