28. КОРРЕЛЯЦИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

Ранее было введено понятие корреляционного момента двух случайных величин:

$$K_{XY} = M\left[\left(X - m_x\right)\left(Y - m_y\right)\right]$$

Для дискретных СВ он выражается формулой:

$$K_{XY} = \sum_i \sum_j (x_i - m_x)(y_j - m_y) p_{ij}$$

А для непрерывных СВ:

$$K_{XY} = \iint\limits_{-\infty}^{+\infty} (x - m_x)(y - m_y) f(x, y)\, dx\, dy$$

Выясним смысл этой характеристики. Для этого вычислим корреляционный момент для двух независимых величин X и У:

$$K_{XY} = M\left[\left(X - m_x\right)\left(Y - m_y\right)\right] =$$

$$= M\left[XY - m_x Y - m_y X + m_x m_y\right] =$$

По свойству математического ожидания:

$$= M[XY] - m_x M[Y] - m_y M[X] + m_x m_y =$$

$$= M[XY] - m_x m_y =$$

Математическое ожидание произведения независимых случайных величин Х и У равно произведению мат. ожиданий этих величин:

$$M[XY]=M[X]M[Y]$$

Следовательно,

$$= M[X]M[Y] - m_x m_y = 0$$

Корреляционный момент двух независимых величин равен нулю.

Следовательно, если корреляционный момент двух случайных величин отличен от нуля, то это есть признак наличия между ними зависимости.

Из определения корреляционного момента следует, что если одна из величин мало отклоняется от своего мат. ожидания (почти не случайна), то момент будет небольшим, какой бы тесной не была зависимость.

Поэтому для характеристики связи между величинами Х и У переходят к безразмерной величине:

$$k_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y}$$

коэффициент корре

Для независимых СВ он также равен нулю. Такие СВ называются *некоррелированными*.

Некорреляция СВ слабее независимости, т.е. если СВ некоррелированны, то они не обязательно будут независимыми.

Если $K_{xy}>0$, то СВ называются положительно коррелированными.

Если $K_{xy}<0$, то СВ называются отрицательно коррелированными.

Вычислим коэффициент корреляции для СВ X и У из предыдущего примера.

$$m_x = 0.4 \quad D_x = 0.64$$

$$m_y = 0.4 \quad D_y = 0.24$$

$$M[XY] = -1 \cdot 0.2 + 1 \cdot 0.1 = -0.1$$

$$K_{XY} = M[XY] - m_x m_y = -0.1 - 0.4 \cdot 0.4 = -0.26$$

$$k_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y} = \frac{-0.26}{\sqrt{0.6 \cdot 0.24}} = -0.66$$

Коэффициент корреляции характеризует не всякую, а только *линейную зависимость*, при которой возрастание (убывание) одной СВ приводит к возрастанию (убыванию) другой по линейному закону.

Коэффициент корреляции характеризует степень тесноты линейной зависимости между СВ.

Пусть

$$Y = AX + B$$

где A и B – постоянные.

Вычислим корреляционный момент случайных величин *X* и *У*:

$$K_{XY} = M[XY] - m_x m_y = M[X(AX+B)] - m_x(Am_x + B) =$$

$$= M[A\cdot X^2 + B\cdot X] - A\cdot m_x^{\,2} - B\cdot m_x =$$

По свойству математического ожидания:

$$= A\cdot M[X^2] + B\cdot m_x - A\cdot m_x^{\,2} - B\cdot m_x =$$

$$= A\cdot(M[X^2] - m_x^{\,2}) =$$

Выражение, стоящее в скобках, по определению является дисперсией *X*:

$$= A \cdot D_x = A \cdot \sigma_x \cdot \sigma_x$$

С другой стороны, по свойству дисперсии:

$$D_y = D[A \cdot X + B] = A^2 D_x$$

Тогда

$$\sigma_y = |A| \sigma_x$$

Следовательно

$$k_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y} = \frac{A\sigma_x \sigma_x}{|A|\sigma_x \sigma_x} = \pm 1$$

Таким образом, знак коэффициента корреляции определяется знаком постоянной *A*.

Далее, чтобы показать, что абсолютное значение коэффициента корреляции не превосходит единицы, рассмотрим СВ

$$Z = \sigma_y X + \sigma_x Y$$

Найдем дисперсию *Z*:

$$D[Z] = D[\sigma_y X + \sigma_x Y] =$$

$$= \sigma_y^2 \cdot D[X] + \sigma_x^2 \cdot D[Y] \pm 2\sigma_x \cdot \sigma_y \cdot K_{XY} =$$

$$= \sigma_y^2 \sigma_x^2 + \sigma_x^2 \sigma_y^2 \pm 2\sigma_x \sigma_y K_{XY} =$$

$$= 2\sigma_x^2 \sigma_y^2 \pm 2\sigma_x \sigma_y K_{XY} =$$

$$= 2\sigma_x \sigma_y (\sigma_x \sigma_y \pm K_{XY}) \geq 0$$

(т.к. дисперсия всегда неотрицательна).

Тогда

$$|K_{XY}| \leq \sigma_x \sigma_y$$

Следовательно,

$$\boxed{\left|k_{xy}\right| \leq 1}$$

Если случайные величины положительно коррелированы, то возрастанию одной из них соответствует возрастание другой (например, рост и вес человека).

Если корреляция отрицательная, то возрастанию одной СВ соответствует убывание другой (например, время, потраченное студентом на подготовку к контрольной и количество сделанных им в работе ошибок).