

Раздел: Теория вероятностей и математическая
статистика

Тема: *Элементы теории корреляции*

Лектор Рожкова С.В.

2021 г.

§21. Корреляционная зависимость. Условные средние

Зависимость СВ X и Y может быть:

- а) функциональной, т.е $Y = f(X)$;
- б) статистической (изменение одной СВ влечет изменение распределения другой).

Частный случай статистической зависимости – корреляционная зависимость.

Пусть каждому значению x СВ X соответствует несколько значений СВ Y : y_1, y_2, \dots, y_n .

Среднее арифметическое чисел y_1, y_2, \dots, y_n назовем **условным средним** (соответствующим данному значению x):

$$\bar{y}_x = \frac{y_1 + y_2 + \dots + y_n}{n}$$

ОПРЕДЕЛЕНИЕ: *Корреляционной зависимостью* Y от X называют функциональную зависимость условной средней \bar{y}_x от x :

$$\bar{y}_x = f(x) \quad (1)$$

Уравнение (1) называют *уравнением регрессии* Y на X .

Функцию $f(x)$ называют *регрессией* Y на X .

График $f(x)$ называют *линией регрессии* Y на X .

Аналогичным образом вводится условная средняя \bar{x}_y и корреляционная зависимость X от Y .

1-я задача теории корреляции – установить форму корреляционной связи, т.е вид функции регрессии (линейная, квадратичная, показательная).

2-я задача теории корреляции – оценить тесноту корреляционной связи.

§22. Отыскание выборочного уравнения прямой линии регрессии

Уравнение прямой линии регрессии Y на X будем искать в виде

$$\bar{y}_x = \rho_{yx} \cdot x + b$$

Угловым коэффициентом ρ_{yx} прямой линии регрессии Y на X называют **выборочным коэффициентом регрессии Y на X** .

Для параметров уравнения прямой линии регрессии Y на X справедливы условия :

$$\begin{cases} \overline{x^2} \cdot \rho_{yx} + \bar{x} \cdot b = \overline{xy}, \\ \bar{x} \cdot \rho_{yx} + b = \bar{y}. \end{cases} \quad (2)$$

Аналогично находится уравнение прямой линии регрессии X на Y :

$$\bar{x}_y = \rho_{xy} \cdot y + b$$

Параметры уравнения прямой линии регрессии X на Y находятся из системы условий

$$\begin{cases} \overline{y^2} \cdot \rho_{xy} + \bar{y} \cdot b = \overline{xy}, \\ \bar{y} \cdot \rho_{xy} + b = \bar{x}. \end{cases} \quad (3)$$

Найдем решение системы (2). Получим:

$$\rho_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2}, \quad b = \frac{\overline{x^2} \bar{y} - \bar{x} \cdot \overline{xy}}{\overline{x^2} - (\bar{x})^2}.$$

Но: $D_X = \overline{x^2} - (\bar{x})^2$

Следовательно:

$$\Rightarrow \rho_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{D_X} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{(\sigma_X)^2}$$

$$\Rightarrow \rho_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{(\sigma_X)^2} \cdot \frac{\sigma_Y}{\sigma_Y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_X \cdot \sigma_Y} \cdot \frac{\sigma_Y}{\sigma_X}$$

$$\Rightarrow \rho_{yx} = r_e \cdot \frac{\sigma_Y}{\sigma_X}$$

где r_e – выборочный коэффициент корреляции.

Из второго уравнения системы (2) находим:

$$b = \bar{y} - \rho_{yx} \bar{x}$$

⇒ Уравнение прямой линии регрессии Y на X примет вид:

$$\bar{y}_x = r_e \cdot \frac{\sigma_Y}{\sigma_X} \cdot x + \bar{y} - \rho_{yx} \bar{x}$$

$$\Rightarrow \bar{y}_x - \bar{y} = r_e \cdot \frac{\sigma_Y}{\sigma_X} \cdot (x - \bar{x})$$

Аналогично, выборочное уравнение прямой линии регрессии X на Y можно записать в виде:

$$\bar{x}_y - \bar{x} = r_e \cdot \frac{\sigma_X}{\sigma_Y} \cdot (y - \bar{y})$$