Раздел: Теория вероятностей и математическая статистика

Тема: ***Элементы теории корреляции***

Лектор Пахомова Е.Г.

2015 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**: $y_1, y_2, \ldots, y_n$.

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

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

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

$$\overline{y}_x = f(x) \qquad (1)$$

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 $\overline{x}_y$ 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

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

Угловой коэффициент 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 :

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

Аналогично находится уравнение прямой линии регрессии **X** на **Y:**

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

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

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

Найдем решение системы (2). Получим:

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

Но:

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

Следовательно:

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

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

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

где $r_в$ – выборочный коэффициент корреляции.

Из второго уравнения системы (2) находим:

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

$\Rightarrow$ Уравнение прямой линии регрессии **Y** на **X** примет вид:

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

$$\Rightarrow \quad \overline{y}_x - \overline{y} = r_в \cdot \frac{\sigma_Y}{\sigma_X} \cdot (x - \overline{x})$$

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

$$\overline{x}_y - \overline{x} = r_в \cdot \frac{\sigma_X}{\sigma_Y} \cdot (y - \overline{y})$$