

§ 2. Криволинейная корреляция

Если график регрессии — кривая линия, то корреляцию называют *криволинейной*. В частности, в случае параболической корреляции второго порядка выборочное уравнение регрессии Y на X имеет вид

$$\bar{y}_x = Ax^2 + Bx + C.$$

Неизвестные параметры A , B и C находят (например, методом Гаусса) из системы уравнений:

$$\begin{aligned} (\sum n_x x^4) A + (\sum n_x x^3) B + (\sum n_x x^2) C &= \sum n_x \bar{y}_x x^2, \\ (\sum n_x x^3) A + (\sum n_x x^2) B + (\sum n_x x) C &= \sum n_x \bar{y}_x x, \\ (\sum n_x x^2) A + (\sum n_x x) B + nC &= \sum n_x \bar{y}_x. \end{aligned} \quad (*)$$

Аналогично находится выборочное уравнение регрессии X на Y :

$$\bar{x}_y = A_1 y^2 + B_1 y + C_1.$$

Для оценки силы корреляции Y на X служит *выборочное корреляционное отношение* (отношение межгруппового среднего квадратического отклонения к общему среднему квадратическому отклонению признака Y)

$$\eta_{yx} = \sigma_{\text{межгр}} / \sigma_{\text{общ}},$$

или в других обозначениях

$$\eta_{yx} = \sigma_{\bar{y}_x} / \sigma_y.$$

Здесь
$$\sigma_{\bar{y}_x} = \sqrt{D_{\text{межгр}}} = \sqrt{\frac{\sum n_x (\bar{y}_x - \bar{y})^2}{n}}, \quad \sigma_y = \sqrt{D_{\text{общ}}} = \sqrt{\frac{\sum n_y (y - \bar{y})^2}{n}},$$

где n — объем выборки (сумма всех частот); n_x — частота значения x признака X ; n_y — частота значения y признака Y ; \bar{y}_x — условная средняя признака Y ; \bar{y} — общая средняя признака Y .

Аналогично определяется выборочное корреляционное отношение X к Y :

$$\eta_{xy} = \sigma_{\bar{x}_y} / \sigma_x.$$

537. Найти выборочное уравнение регрессии $\bar{y}_x = Ax^2 + Bx + C$ по данным, приведенным в корреляционной табл. 8.

Оценить силу корреляционной связи по выборочному корреляционному отношению.

Таблица 8

| y | x | | | n _y |
|----------------|----|----|----|----------------|
| | 2 | 3 | 5 | |
| 25 | 20 | — | — | 20 |
| 45 | — | 30 | 1 | 31 |
| 110 | — | 1 | 48 | 49 |
| n _x | 20 | 31 | 49 | n = 100 |

Решение. Составим расчетную табл. 9.

Таблица 9

| x | n _x | \bar{y}_x | n _x x | n _x x ² | n _x x ³ | n _x x ⁴ | n _x \bar{y}_x | n _x $\bar{y}_x x$ | n _x $\bar{y}_x x^2$ |
|---|----------------|-------------|------------------|-------------------------------|-------------------------------|-------------------------------|----------------------------|------------------------------|--------------------------------|
| 2 | 20 | 25 | 40 | 80 | 160 | 320 | 500 | 1 000 | 2 000 |
| 3 | 31 | 47,1 | 93 | 279 | 837 | 2 511 | 1 460 | 4 380 | 13 141 |
| 5 | 49 | 108,67 | 245 | 1 225 | 6 125 | 30 625 | 5 325 | 26 624 | 133 121 |
| Σ | 100 | | 378 | 1 584 | 7 122 | 33 456 | 7 285 | 32 004 | 148 262 |

Подставив числа, содержащиеся в последней строке табл. 9, в (*), получим систему уравнений относительно неизвестных коэффициентов A , B , C :

$$\begin{aligned} 33456 A + 7122 B + 1584 C &= 148262, \\ 7122 A + 1584 B + 378 C &= 32004, \\ 1584 A + 378 B + 100 C &= 7285. \end{aligned}$$