ГЛАВА 5
ОСНОВИ ТЕОРІЇ КОРЕЛЯЦІЇ

5.1. Кореляційний момент

При обробці результатів за методом найменших квадратів передбачалось, що величини x та y пов’язані жорсткою функціональною залежністю. Інший полюс – це повна незалежність двох величин. Для цього випадку існує критерій – відповідно до властивостей дисперсії (глава 2) для незалежних величин y та x:

$$D[x+y]=D[x]+D[y] \tag{5.1}$$

Між жорсткою функціональною залежністю та повною незалежністю існує ціла гама залежностей, які називаються *кореляційними*. Наприклад, температура повітря у Києві не визначає температуру повітря в Ташкенті, однак, ці величини не незалежні, влітку і там, і там тепліше, взимку – холодніше, так що ці температури – *скорельовані*. Інший приклад, оцінки студентів (групи, курсу) за різними предметами не обумовлюють одна одну, але вони скорельовані: сильні студенти отримують більш високі оцінки, а слабкіші – низькі.

Мета теорії кореляції полягає у з’ясуванні, який характер зв’язку між вибраними величинами: сильний, слабкий, відсутній.

Для пошуку кількісного критерію взаємодії можна відштовхуватися від дисперсії суми двох величин $D[x+y]$, не вимагаючи від величин x та y незалежності. Тоді отримуємо, використовуючи властивості дисперсії та математичного сподівання, а також формальну трактову моментів як відповідних математичних сподівань:

$$
\begin{aligned}
&D[x+y] = M\{[x+y] - M[x+y]\}^2 = \\
&= M\{(x+y)^2 - 2(x+y)(m_x+m_y) + (m_x+m_y)^2\} = \\
&= M\{x^2 + 2xy + y^2 - 2xm_x - 2xm_y - 2ym_x - 2ym_y + m_x^2 + \\
&+ 2m_x m_y + m_y^2\} = M\{(x^2 - 2m_x x + m_x^2) + (y^2 - 2m_y y + m_y^2) + \\
&+ 2(xy - m_x y - m_y x + m_x m_y)\} = \qquad (5.2) \\
&= M\{(x-m_x)^2 + (y-m_y)^2 + 2(x-m_x)(y-m_y)\} = \\
&= M[(x-m_x)^2] + M[(y-m_y)^2] + \\
&+ 2M[(x-m_x)(y-m_y)] = D[x] + D[y] + 2K_{xy}
\end{aligned}
$$

Величина $K_{xy} = M[(x-m_x)(y-m_y)]$ - центральний змішаний момент, має назву *кореляційний момент*.

Співставляючи (5.1) та (5.2) бачимо, що для незалежних величин x та y кореляційний момент повинен дорівнювати нулю. Дійсно, для незалежних величин $M[xy] = M[x]M[y]$, тоді:

$$
\begin{aligned}
&K_{xy} = M[(x-m_x)(y-m_y)] = M[x-m_x]\cdot M[y-m_y] = \\
&= \{\underbrace{(M[x]-m_x)}_{0}\cdot\underbrace{(M[y]-m_y)}_{0}\} = 0
\end{aligned}
$$

Якщо $K_{xy} \neq 0$, то це є свідченням наявності зв'язку між x та y.

Аналогічно тому, як паралельно дисперсії (центральний другий момент) існує другий початковий момент, так і паралельно кореляційному (центральний змішаний момент) існує *початковий змішаний момент*:

$$\beta_{11}[xy] = M[xy] \qquad (5.3)$$

Знайдемо зв'язок K_{xy} та $\beta_{11}[xy]$:

$$K_{xy} = M\left[(x-m_x)(y-m_y)\right] = M\left[xy - m_x y - x m_y + m_x m_y\right] =$$
$$= M[xy] - m_x M[y] - m_y M[x] + m_x m_y = \beta_{11}[xy] - m_x m_y$$

отже,

$$\beta_{11}[xy] = K_{xy} + m_x m_y \tag{5.4}$$

На практиці замість K_{xy} та $\beta_{11}[xy]$ використовують їх статистичні аналоги:

$$K_{xy}^* = \frac{\sum_{i=1}^{n}(x_i - m_x^*)(y_i - m_y^*)}{n} \tag{5.5}$$

$$\beta_{11}^*[xy] = \frac{\sum_{i=1}^{n} x_i y_i}{n}. \tag{5.6}$$

5.2. Рівняння регресії

Використовуючи величини (5.5) та (5.6), перетворимо рівняння, що були отримані за методом найменших квадратів для лінійної залежності.

За формулою (4.17) маємо:

$$\begin{cases} \sum_{i=1}^{n} x_i y_i = b_1 \sum_{i=1}^{n} x_i^2 + b_0 \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i = b_1 \sum_{i=1}^{n} x_i + n b_0 \end{cases}$$

Розділимо обидві частини на n :

$$\begin{cases} \dfrac{\sum\limits_{i=1}^{n} x_i y_i}{n} = b_1 \dfrac{\sum\limits_{i=1}^{n} x_i^2}{n} + b_0 \dfrac{\sum\limits_{i=1}^{n} x_i}{n} \\ \dfrac{\sum\limits_{i=1}^{n} y_i}{n} = b_1 \dfrac{\sum\limits_{i=1}^{n} x_i}{n} + b_0 \end{cases}$$

Зробимо відповідні заміни:

$$m_x^* = \frac{\sum\limits_{i=1}^{n} x_i}{n}; \quad m_y^* = \frac{\sum\limits_{i=1}^{n} y_i}{n};.$$

$$\alpha_2^*[x] = \frac{\sum\limits_{i=1}^{n} x_i^2}{n}; \quad \beta_{11}^*[xy] = \frac{\sum\limits_{i=1}^{n} x_i y_i}{n}.$$

та отримаємо:

$$\begin{cases} \beta_{11}^*[xy] = a\alpha_2^*[x] + bm_x^* \\ m_y^* = b_1 m_x^* + b_0 \end{cases}$$

Замість початкових моментів $\alpha_2^*[x]$ та $\beta_{11}^*[xy]$ вводимо центральні $D^*[x]$ та K_{xy}^* :

Розв’яжемо цю систему:

$$K_{xy}^* + b_1\left(m_x^*\right)^2 + b_0 m_x^* = b_1 D_x^* + b_1\left(m_x^*\right)^2 + b_0 m_x^*$$

$$K_{xy}^* = b_1 D_x^*$$

$$b_1 = \frac{K_{xy}^*}{D_x^*}$$

та

$$b_0 = m_y^* - b_1 m_x^* = m_y^* - \frac{K_{xy}^*}{D_x^*} m_x^*$$

Розглянемо рівняння $y = b_1 x + b_0$:

$$y_x = \frac{K_{xy}^*}{D_x^*} \cdot x + m_y^* - \frac{K_{xy}^*}{D_x^*} m_x^*$$

$$y_x - m_y^* = \frac{K_{xy}^*}{D_x^*}\left(x - m_x^*\right) \qquad (5.7)$$

Рівняння (5.7) – це пряма, яка проходить через точку (*М*) з координатами $(m_x^*; m_y^*)$ та має тангенс кута нахилу $tg\alpha = \frac{K_{xy}^*}{D_x^*}$ (рис. 5.1, пряма 1). Рівняння (5.7) називається *рівнянням регресії* y за x. Відповідно, метод найменших квадратів часто називають *регресивним аналізом*.

Рівняння (5.7) отримано, виходячи із умови, що всі помилки включені у величину y, а величина x - визначена точно. Назвемо рівняння (5.7) рівнянням регресії y за x, та додамо до змінної, яка включає помилки (y), підстрочником точного аргументу (x).

Якщо y та x пов’язані функціональною залежністю, то функцію та аргумент можна змінювати місцями, тобто можна записати:

$$x = cy + d$$

Коефіцієнти c та d однозначно пов'язані з коефіцієнтами b_1 та b_0 залежності $y = b_1 x + b_0$ при відсутності помилки експерименту. При наявності помилок параметри c та d знаходяться із умови:

$$S = \sum_{i=1}^{n} \left[x_i - c y_i - d \right]^2 = \min \qquad (5.8)$$

тобто зараз припускаємо, що всі помилки зосереджені в x. Мінімізуючи S за c та d і розв'язуючи відповідну систему лінійних рівнянь, отримуємо рівняння регресії x за y:

$$x_y - m_x^* = \frac{K_{xy}^*}{D_y^*}\left(y - m_y^*\right) \qquad (5.9)$$

Рівняння (5.9) відповідає прямій, що проходить також, як рівняння регресії y за x, через точку (M) з координатами $(m_x^*; m_y^*)$ та має котангенс кута нахилу $ctg\beta = \frac{K_{xy}^*}{D_y^*}$, відповідно, $tg\beta = \frac{D_y^*}{K_{xy}^*}$ (рис. 5.1 , пряма 2).

Рівняння регресії приймає більш простий вигляд у координатах:

$$s = y - m_y^*$$

та

$$t = x - m_x^*,$$

тобто в системі координат, де початку координат відповідає точка M .

На величину центральних моментів D_x^*, D_y^*, K_{xy}^* паралельний перенос координат не впливає:

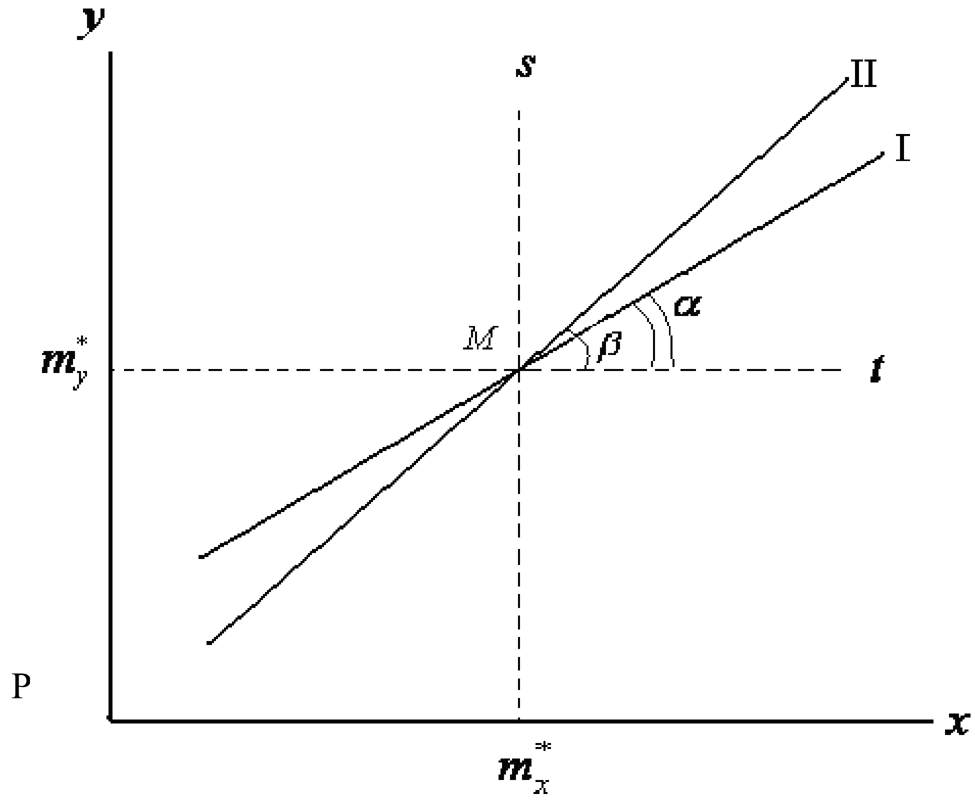

Рис. 5.1. Рівняння регресії y за x (I) та x за y (II).

$$D_x^* = D_t^*;$$

$$D_y^* = D_s^*;$$

$$K_{xy}^* = K_{st}^*;$$

$$\begin{cases} s_t = \dfrac{K_{st}^*}{D_t^*} t = \dfrac{K_{xy}^*}{D_x^*} t \\ t_s = \dfrac{K_{st}^*}{D_s^*} s = \dfrac{K_{xy}^*}{D_y^*} s \end{cases} \qquad (5.10)$$

Ці рівняння використовуються для знаходження меж коефіцієнта кореляції.

5.3. Коефіцієнт кореляції

Кореляційний момент K_{xy} має розмірність, яка дорівнює розмірності добутку випадкових величин x та y. Зручніше користуватися безрозмірною величиною, *коефіцієнтом кореляції* (r):

$$r = \frac{K_{xy}^*}{\sigma_x^* \sigma_y^*}; \tag{5.11}$$

Для незалежних величин x та y $K_{xy} = 0$ та, відповідно, $r = 0$.

Знайдемо, чому дорівнює величина r для жорсткої функціональної залежності за відсутності помилок експерименту. В цьому випадку всі точки ідеально лягають на пряму лінію і прямі регресії I та II (рис. 5.1) повинні тотожно співпасти. Опускаючи підстрочні індекси в рівняннях (5.10), отримуємо:

$$s = \frac{K_{xy}^*}{D_x^*} t = \frac{K_{xy}^*}{D_x^*} \cdot \frac{K_{xy}^*}{D_y^*} s \rightarrow$$

$$\rightarrow \frac{(K_{xy}^*)^2}{D_x^* D_y^*} = 1$$

Добудемо квадратний корінь з обох частин рівняння та отримуємо:

$$\rightarrow \frac{(K_{xy}^*)^2}{D_x^* D_y^*} = 1 \rightarrow$$

$$\to \frac{\sqrt{\left(K_{xy}^*\right)^2}}{\sqrt{D_x^* D_y^*}} = 1 \to$$

$$\to \frac{K_{xy}^*}{\sigma_x^* \sigma_y^*} = r = |1| = \pm 1$$

Тобто точній функціональній залежності відповідає значення $r = \pm 1$.

У загальному вигляді:

$$0 \le |r| \le 1 \tag{5.12}$$

Наближення r до одиниці відповідає функціональній залежності; нулю – відсутності лінійної залежності; проміжне значення r вказує на наявність кореляційної залежності. «Міцність» зв'язку визначається величиною r: чим менше r, тим слабкіший зв'язок між величинами x та y.

Коефіцієнту кореляції r можна дати графічну інтерпретацію (рис. 5 .1):

$$tg\alpha = \frac{K_{xy}^*}{D_x^*} \text{ (пряма I);}$$

$$tg\beta = \frac{D_y^*}{K_{xy}^*} \text{ (пряма II)}$$

Відповідно:

$$r = \frac{K_{xy}^*}{\sqrt{D_x^*}\sqrt{D_y^*}} = \sqrt{\frac{K_{xy}^* K_{xy}^*}{D_x^* D_y^*}} = \sqrt{\frac{tg\alpha}{tg\beta}}. \tag{5.13}$$

тобто, чим більше розходяться між собою рівняння регресії (рис. 5 .1), тим менший коефіцієнт кореляції.

Визначимо коефіцієнт кореляції аналітичним методом:

$$r = \frac{K^*_{xy}}{\sigma^*_x \sigma^*_y} = \frac{\beta^*_{11}[xy] - m^*_x m^*_y}{\sqrt{\left\{\alpha^*_2[x] - \left(m^*_x\right)^2\right\}\left\{\alpha^*_2[y] - \left(m^*_y\right)^2\right\}}};$$

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2} \cdot \sqrt{\frac{\sum_{i=1}^{n} y_i^2}{n} - \left(\frac{\sum_{i=1}^{n} y_i}{n}\right)^2}};$$

Помножимо чисельник та знаменник на n^2 та введемо позначки Гаусса:

$$r = \frac{n[xy] - [x][y]}{\sqrt{n[x^2] - [x]^2} \cdot \sqrt{n[y^2] - [y]^2}}. \qquad (5.14)$$

Приклад 5.1. Знайти коефіцієнт кореляції для даних, наведених в таблиці 5 .1

Таблиця 5.1.

Вихідні дані для розрахунку коефіцієнта кореляції.

x_i	$x_i y_i$	y_i
1	1	1

Продовження таблиці 5.1.

	x_i	$x_i y_i$	y_i
	1	3	3
	5	5	1
	5	15	3
Σ	12	24	8

Розв'язок:

$$m_x^* = \frac{[x]}{n} = \frac{12}{4} = 3;$$

$$\beta_{11}^*[xy] = \frac{[xy]}{n} = \frac{24}{4} = 6;$$

$$m_y^* = \frac{[y]}{n} = \frac{8}{4} = 2;$$

$$K_{xy}^* = \beta_{11}^*[xy] - m_x^* m_y^* = 6 - 6 = 0 \to\to r = 0,$$

тобто кореляція відсутня.

Рівняння регресії в цьому випадку має вигляд:

$$\begin{cases} y_x - 2 = 0(I) \\ x_y - 3 = 0(II) \end{cases}$$

З даних, наведених на рис. 5.2, видно, що в цьому випадку прямі регресії перпендикулярні одна одній, тобто використання регресивного аналізу за відсутністю кореляції не має сенсу.

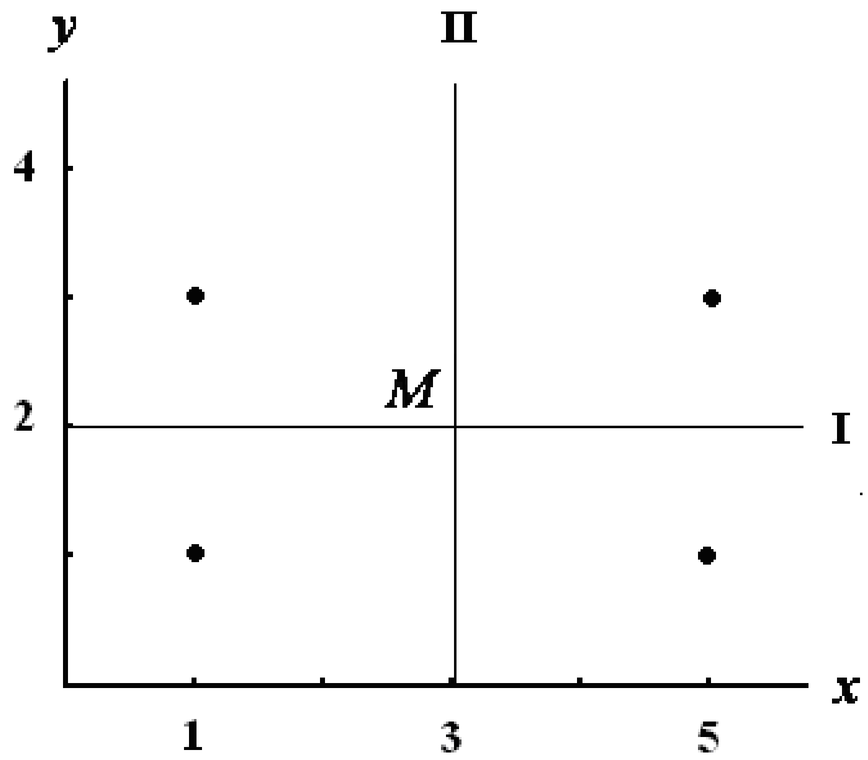

Рис. 5.2. Прямі регресії $y_x(I)$ та $x_y(II)$ за відсутності кореляції між x та y.

Приклад 5.2. Знайти кореляційний коефіцієнт для масиву даних, наведених в таблиці 5.2.

Таблиця 5.2.

Кореляційна залежність

N п/п	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	1	2	2	1	4
2	2	1	2	4	1
3	2	3	6	4	9
4	2	4	8	4	16
5	3	3	9	9	9
6	3	6	18	9	36

Продовження таблиці 5.2

N п/п	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
7	4	2	8	16	4
8	4	4	16	16	16
9	4	6	24	16	36
10	4	8	32	16	64
11	6	3	18	36	9
12	6	4	24	36	16
13	6	6	36	36	36
14	6	8	48	36	64
15	6	9	54	36	81
Σ	59	69	305	275	401

Розв'язок:

$$\alpha_2^*[x] = \frac{275}{15} = 18{,}3\,;\; m_x^* = \frac{59}{15} = 3{,}9\,;\; \beta_{11}^*[xy] = \frac{305}{15} = 20{,}3$$

$$m_y^* = \frac{69}{15} = 4{,}6\,;\;\; \alpha_2^*[y] = \frac{401}{15} = 26{,}7\,;$$

$$D^*[x] = 18{,}3 - 3{,}9^2 = 3{,}09\,;\; K_{xy}^* = 20{,}3 - 3{,}9 \cdot 4{,}6 = 2{,}36$$

$$D^*[y] = 26{,}7 - 4{,}6^2 = 5{,}54\,;$$

$$\sigma_x^* = \sqrt{3{,}09} = 1{,}76\,;\;\; \sigma_y^* = \sqrt{5{,}54} = 2{,}35\,;\;\; r = \frac{2{,}36}{1{,}76 \cdot 2{,}35} = 0{,}57$$

В даному випадку між величинами x та y існує не функціональна, а кореляційна залежність середньої сили. Це наглядно видно з даних, наведених на рис.5.3. Спостерігається тенденція до зростання y із збільшенням x ($r = 0{,}57$). Як видно з рис.5.3 характерна для функціональної залежності лінія перетворюється на широку розмиту смугу.

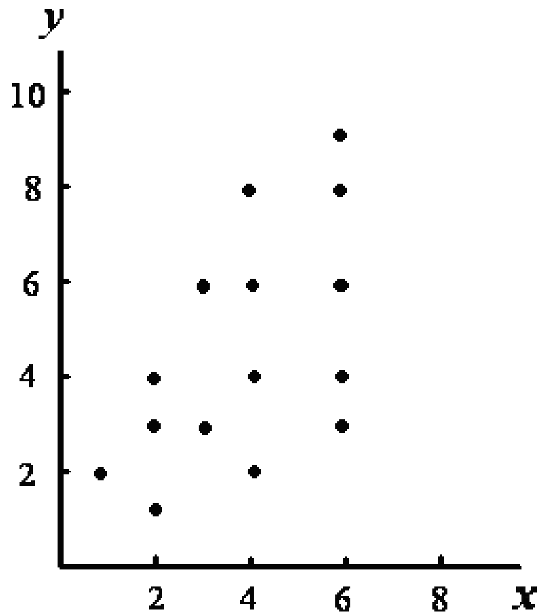

Рис.5 .3. Масив даних для приклада 5.2\.

5.4. Нелінійна кореляція

Все, що було сказано вище, відноситься до так званої *лінійної кореляції*. Значення $r \to |1|$ вказує на лінійну функціональну залежність, але відсутність кореляції ($r \to 0$) не дає гарантії повної незалежності, виключається лише лінійний зв'язок між x та y. При цьому може існувати нелінійний зв’язок між x та y.

У випадку, якщо між x та y передбачається нелінійна залежність, для її характеристики («сили» зв’язку) використовується *кореляційне відношення (θ)*:

$$\theta = \sqrt{1 - \frac{S}{S_y}}; \tag{5.15}$$

де $S=\sum_{i=1}^{n}\left[y_i-\varphi(b_0,b_1,b_2,...x_i)\right]^2$ - відома сума квадратів нев'язок для відповідного рівняння регресії:

$$S_y=D_y^*\cdot n=\sum_{i=1}^{n}\left[y_i-m_y^*\right]^2$$

За ***S=0*** усі точки ідеально лягають на відповідну криву, $\theta=1$ - це випадок жорсткої функціональної залежності.

Величина S_y характеризує розкид ординат навколо m_y^*. У випадку лінійної залежності раніше було показано, що за відсутності кореляції ($K_{xy}^*=0$) рівняння регресії y_x – це пряма, що паралельна осі абсцис на відстані $y=m_y^*=b_0$. В цьому випадку для суми квадратів нев'язок отримуємо:

$$S=\sum_{i=1}^{n}\left[y_i-b_1x_i-b_0\right]^2=\sum_{i=1}^{n}\left[y_i-b_0\right]^2=\sum_{i=1}^{n}\left[y_i-m_y^*\right]^2=S_y\,.$$

При $S=S_y$, $\theta=0$. Це положення зберігається і для нелінійної залежності.

Таким чином, межі для кореляційного відношення θ ($0\le|\theta|\le1$) такі ж самі, що і для лінійного коефіцієнту кореляції.

У випадку лінійної залежності θ та r тотожно співпадають. Покажемо це.

Відповідно до формули (4.61) і (4.62) для лінійної залежності:

$$S = [y^2] - b_1[xy] - b_0[y] =$$
$$= n\left\{\frac{[y^2]}{n} - b_1\frac{[xy]}{n} - b_0\frac{[y]}{n}\right\} =$$
$$= n\left\{\alpha_2^*[y] - b_1\beta_{11}^*[xy] - b_0 m_y^*\right\} =$$
$$= n\left\{D_y^* + \left(m_y^*\right)^2 - b_1 K_{xy}^* - b_1 m_x^* m_y^* - b_0 m_y^*\right\};$$

Враховуючи, що

$$b_1 = \frac{K_{xy}^*}{D_x^*}$$

та

$$m_y^* = b_1 m_x^* + b_0,$$

отримуємо:

$$S = n\left\{D_y^* - \frac{\left(K_{xy}^*\right)^2}{D_x^*} + m_y^*\left[m_y^* - b_1 m_x^* - b_0\right]\right\} =$$
$$= nD_y^*\left\{1 - \frac{\left(K_{xy}^*\right)^2}{D_x^* D_y^*}\right\} = S_y\left(1 - r^2\right);$$

$$\theta = \sqrt{1 - \frac{S}{S_y}} = \sqrt{1 - \left(1 - r^2\right)} = r.$$

Розглянемо приклад розрахунку θ для нелінійної залежності – параболи.

Приклад 5.3. Перевіримо, чи являється параболічна залежність функціональною за такими значеннями:

$$b_2 = 0{,}0233809;$$

$$b_1 = -2{,}606619;$$

$$b_0 = 100{,}79114 .$$

$$\left[x^2 y\right] = 186054{,}3 ;$$

$$\left[xy\right] = 7688{,}9;$$

$$\left[y\right] = 399{,}7;$$

$$\left[y^2\right] = 24594{,}33; \; n = 7 .$$

Розв'язок.

Для розв'язку цього завдання необхідно обрахувати кореляційне відношення θ та співставити його з одиницею.

Для параболи відповідно до формули (4.20) отримуємо:

$$S = \left[y^2\right] - b_2\left[x^2 y\right] - b_1\left[xy\right] - b_0\left[y\right];$$

$$S = 24594{,}33 - 0{,}0233809 \cdot 186054{,}3 + 2{,}606619 \cdot 7688{,}9 -$$
$$-100{,}79114 \cdot 399{,}70 = 0{,}0272;$$

Додатково обраховуємо:

$$S_y = D_y^* \cdot n = \left\{\alpha_2^*\left[y\right] - \left(m_y^*\right)^2\right\} \cdot n =$$
$$= \left\{\frac{\left[y^2\right]}{n} - \left(\frac{\left[y\right]}{n}\right)^2\right\} \cdot n = \left[y^2\right] - \frac{\left[y\right]^2}{n};$$

$$S_y = 24594{,}33 - \frac{399{,}7^2}{7} = 1771{,}5;$$

$$\theta = \sqrt{1 - \frac{S}{S_y}} = \sqrt{1 - \frac{0{,}0272}{1771{,}5}} = 0{,}99999.$$

Таким чином, ця параболічна залежність є жорстко функціональною.

Підсумовуючи, відмітимо, що відносно невисокі значення θ можуть бути і в тому випадку, коли функціональна залежність у принципі існує, але експериментальні дані за рахунок впливу випадкових факторів дуже розкидані, неякісні.

5.5 Аналіз множинної кореляції

Розглянуте вище поняття лінійної кореляції між двома величинами може бути перенесено на випадок більшого числа величин.

Розрахунки зазвичай починають з обчислення з *парних коефіцієнтів кореляції*, що характеризують тісноту взаємозв'язку між двома величинами. При цьому обчислюють два типи парних коефіцієнтів кореляції:

1. r_{yx_j} – Коефіцієнти кореляції, що визначають силу зв'язку між залежною змінною і однієї з незалежних змінних;
2. $r_{x_j x_m}$ – Коефіцієнти кореляції, що визначають силу зв'язку між однією з незалежних змінних x_j та іншими незалежними змінними x_m $(j, m = 1, p)$.

Як приклад розглянемо формули розрахунку для трьох величин *x, y* і *z*, причому для скорочення записів наведемо лише формули, які відносяться до емпіричної регресії *z* на *x* та *y*. У цьому випадку ми маємо справу не з лінією, а з *площиною регресії z* на *x* і *y*, яка має рівняння

$$z - \bar{z} = b_{z|x}(x - \bar{x}) + b_{z|y}(y - \bar{y}) \tag{5.16}$$

де коефіцієнти регресії $b_{z|x}$ та $b_{z|y}$ визначаються через коефіцієнти кореляції між парами величин z та x, z та y, x та y наступним чином:

$$b_{z|x} = \frac{r_{zx} - r_{yx} r_{zy}}{1 - r_{yx}} \cdot \frac{\tilde{\sigma}_z}{\tilde{\sigma}_x},$$

(5.17)

$$b_{z|x} = \frac{r_{zy} - r_{yx} r_{zx}}{1 - r_{yx}} \cdot \frac{\tilde{\sigma}_z}{\tilde{\sigma}_y}.$$

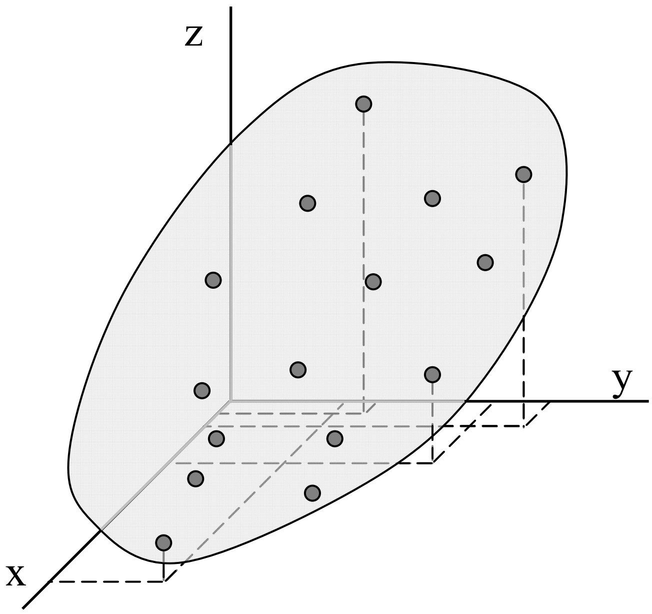

Рис. 5.4. Площина множинної регресії

Тут коефіцієнти кореляції мають таке ж значення, що і для парної кореляції і визначаються з рівнянь:

$$r_{yx} = \frac{1}{\tilde{\sigma}_y \tilde{\sigma}_x} \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}),$$

(5.18)

$$r_{zx} = \frac{1}{\tilde{\sigma}_z \tilde{\sigma}_x} \frac{1}{n-1} \sum_{i=1}^{n} (z_i - \bar{z})(x_i - \bar{x}),$$

$$r_{zy} = \frac{1}{\widetilde{\sigma}_y \widetilde{\sigma}_z} \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \overline{y})(z_i - \overline{z}) ,$$

де n – загальне число точок, а середньоквадратичне відхилення визначаються з виразів:

$$\widetilde{D}_x = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 ,$$

$$\widetilde{D}_y = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \overline{y})^2 , \tag{5.19}$$

$$\widetilde{D}_z = \frac{1}{n-1} \sum_{i=1}^{n} (z_i - \overline{z})^2 .$$

Зведені і частинні коефіцієнти кореляції

Мірою залежності між величиною z і величинами x, y служить зведений коефіцієнт кореляції

$$r = \sqrt{\frac{r_{zx}^2 + r_{zy}^2 - r_{zx} r_{zy} r_{yx}}{1 - r_{yx}^2}} \tag{5.20}$$

Значення зведеного коефіцієнта кореляції r завжди знаходиться між 0 і 1. Якщо величина z не залежить від x і y, то теоретичне значення зведеного коефіцієнта кореляції дорівнює нулю (емпіричне значення R досить мало), отже, між досліджуваними величинами немає лінійної кореляційної залежності (але може бути нелінійна). Зведений коефіцієнт кореляції дорівнює 1 тоді і тільки тоді, коли всі експериментальні точки лежать в площині регресії. Теоретичне значення зведеного коефіцієнта регресії дорівнює одиниці тільки в разі лінійної функціональної залежності між величиною z і величинами x, y.

Для вивчення впливу тільки одного з факторів, наприклад x на величину z, тобто для вивчення кореляції між x і z після усунення змін, викликаних зміною величини y, вводять парціальний (частинний) коефіцієнт кореляції величин x і z по відношенню до величини y:

$$r_{xz|y} = \frac{r_{zx} - r_{yx} r_{zy}}{\sqrt{1 - r_{yx}^2} \cdot \sqrt{1 - r_{yz}^2}} \quad (5.21)$$

Аналогічно визначаються парціальні коефіцієнти кореляції $R_{zy|x}$ і $r_{xy|z}$.

Питання для самостійного повторення

Замість крапок запишіть таке продовження тексту, щоб отримати правильне означення або твердження. Де можливо запишіть відповідну формулу.

1. Кореляційними називаються функціональні залежності …
2. Кореляційний момент – це…
3. Початковий змішаний момент – це…
4. Рівнянням регресії y за x можна записати як…
5. Коефіцієнтом кореляції можна записати як…
6. Коефіцієнту кореляції r можна дати графічну інтерпретацію…
7. Коефіцієнт кореляції можна визначити аналітичним методом…
8. Кореляційне відношення можна визначити як …

Задачі для самостійного розв'язку

1.Знайти коефіцієнт кореляції для залежності $y=b_1x+b_0$, якщо значення x та y наведені в таблиці:

x	1,000	2,500	3,000	4,000	7,000
y	-0,505	3,060	4,170	7,380	15,050

Відповідь: r = 0,999

2. Знайти коефіцієнт кореляції для залежності $y=b_1x+b_0$, якщо значення x та y наведені в таблиці:

x	1,000	1,200	2,000	3,000	3,500
y	3,050	2,800	1,100	0,300	-0,650

Відповідь: r = – 0,988

3. Кількість речовини (y) %, яка залишилася в системі через x хвилин після початку реакції задається рівнянням $y = b_2x^2 + b_1x + b_0$. Знайти кореляційне відношення θ для цієї залежності. .

x	7,000	12,000	17,000	22,000	27,000	32,000	57,000
y	83,700	72,900	63,200	54,700	47,500	41,400	36,300

Відповідь:

θ = 0,9999.