Розділ 12

ЕЛЕМЕНТИ ТЕОРІЇ РЕГРЕСІЇ ТА КОРЕЛЯЦІЇ

Учення про кореляцію (від латинської correlatio – співвідношення, взаємозв'язок) і регресію (від латинської regressio – рух назад) широко використовується при аналізі зв'язків різних явищ. Так, в економіці проводяться дослідження: залежності обсягів виробництва від цілого ряду факторів (розміру основних фондів, їхнього віку і т. д.); залежності продуктивності праці на підприємствах від рівня механізації та електрифікації виробництва, стажу і кваліфікації робітників; залежності попиту і споживання населення від рівня доходу, цін на товари і т. д. Інтенсивно застосовується регресійний аналіз при вивченні залежності врожайності певної сільськогосподарської культури від впливу на неї природних і економічних факторів. Широко застосовуються методи кореляційного і регресійного аналізу в психології, соціології, педагогіці й у ряді інших галузей науки і практики.

Поняття кореляції та регресії з'явилося в середині XIX ст. завдяки роботам Ф. Гальтона[1] і К. Пірсона.

12.1 Функціональна, статистична та кореляційна залежності

У природничих науках здебільшого стикаються з *функціональними залежностями*, у яких кожному можливому значенню аргументу X (незалежної змінної) поставлено у відповідність одне строго визначене значення функції Y (залежної змінної) (наприклад, у математиці при вивченні фізичних законів). Однак набагато частіше в навколишньому світі має місце залежність, коли кожному фіксованому значенню однієї змінної відповідає не одне, а множина значень другої змінної. Це пояснюється тим, що залежна змінна схильна до впливу ряду неконтрольованих або неврахованих факторів, а також численних неконтрольованих випадкових факторів. У цій ситуації залежна

[1]Гальтон Френсіс (1822–1911) – англійський статистик, психолог і астролог.

змінна Y є випадковою величиною. Змінна ж X може бути як детермінованою (тобто такою, що набуває певних значень), так і випадковою величиною. Така залежність отримала назву *статистичної* (або *стохастичної*, *імовірнісної*).

Припустимо, що існує стохастична залежність випадкової змінної Y від X. При $X = x$ змінна Y через її стохастичну залежність від X може набути будь-якого значення з деякої множини. Середнє цієї множини називають *груповим генеральним середнім* змінної Y при $X = x$ або умовним математичним сподіванням ВВ Y, яка обчислена за умови, що $X = x$. Ця величина позначається $M(Y \mid X = x)$.

Зокрема, якщо стохастична залежність Y від X проявляється в тому, що при зміні x змінюється умовне математичне сподівання $M(Y \mid X = x)$, то говорять, що має місце *кореляційна залежність* величини Y від X.

Якщо ж $M(Y \mid X = x)$ залишаються незмінними, то говорять, що кореляційна залежність величини Y від X відсутня.

Якщо існує стохастична залежність випадкової змінної X від Y, то, аналогічно до попереднього, можна ввести поняття умовного математичного сподівання $M(X \mid Y = y)$ і кореляційної залежності величини X від Y.

Таким чином, кореляційна залежність може бути представлена в наступних виглядах:

$$M(Y \mid X = x) = \varphi(x); \tag{12.1}$$

$$M(X \mid Y = y) = \psi(y), \tag{12.2}$$

де $\varphi(x) \neq$ const, $\psi(y) \neq$ const.

Рівняння (12.1) і (12.2) називаються *модельними рівняннями регресії* відповідно Y по X (Y на X) і X по Y (X на Y), функції $\varphi(x)$ і $\psi(y)$ – *модельними функціями регресії*, а їхні графіки – *модельними лініями регресії*.

Слід зазначити, що уведені поняття стохастичної і кореляційної залежностей схильні до генеральної сукупності.

Як оцінки умовних математичних сподівань набувають умовних середніх, які знаходяться за даними спостережень (за вибіркою). На практиці дослідник, як правило, має лише вибірки пар значень $(x_i; y_i)$ обмеженого об'єму.

Умовним середнім $\overline{y}_x(\overline{x}_y)$ називають середнє арифметичне спостережуваних значень $Y(X)$, що відповідають $X = x\ (Y = y)$. Оскільки умовні математичні сподівання $M(Y \mid X = x)$ і $M(X \mid Y = y)$ є відповідно функціями від x і y, то їх оцінки, тобто умовні середні $\overline{y}_x$ і $\overline{x}_y$ також є функціями відповідно від x і y, тобто

$$\overline{y}_x = \tilde{\varphi}(x, b_0, b_1, ..., b_k), \tag{12.3}$$

$$\overline{x}_y = \tilde{\psi}(y, c_0, c_1, ..., c_k), \tag{12.4}$$

де $b_0,\ b_1, ..., b_k$ і $c_0,\ c_1, ..., c_k$ – параметри.

Рівняння (12.3) і (12.4) називаються *вибірковими рівняннями регресії* відповідно Y по X і X по Y, функції $\tilde{\varphi}(x, b_0, b_1, ..., b_k)$ і $\tilde{\psi}(y, c_0, c_1, ..., c_k)$ – *вибірковими функціями регресії*, а їхні графіки – *вибірковими лініями регресії*.

Зв'язок або кореляція двох змінних називається *парною*. Якщо в рівняннях (12.3) і (12.4) зі збільшенням x і y змінні $\overline{y}_x$ і $\overline{x}_y$ у середньому зростають (мають тенденцію до зменшення), то така парна кореляція буде *додатною* (*від'ємною*). Нульова кореляція спостерігається за відсутності зв'язку між X і Y.

Діаграма, на якій зображується сукупність значень двох ознак, називається *кореляційним полем*. Кожна точка цієї діаграми має координати x_i, y_i, що відповідають розмірам ознак в i-му спостереженні. Три варіанти розподілу точок на кореляційному полі наведено на рис. 12.1.

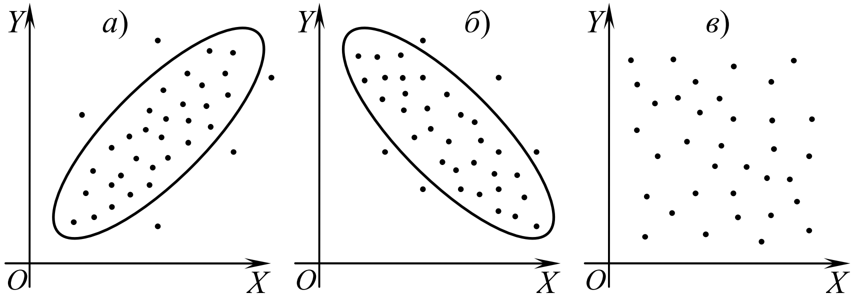

Рис. 12.1

На першому з них основна маса точок укладається в еліпс, більша вісь якого утворює додатний кут з віссю *OX* (додатна кореляція). Другий варіант відповідає від’ємній кореляції. Рівномірний розподіл точок у просторі *OXY* свідчить про відсутність кореляційної залежності (рис. 12.1, в).

Статистичні зв'язки між змінними можна вивчати методами кореляційного і регресійного аналізу. Основною задачею *регресійного аналізу* є встановлення форми залежності за експериментальними даними, визначення функції регресії (процес *вирівнювання*) і вивчення залежності між змінними. Основною задачею *кореляційного аналізу* є виявлення зв'язку між випадковими змінними та оцінювання її тісноти.

12.2 Лінійна парна регресія

Нехай у результаті незалежних спостережень над досліджуваною системою кількісних ознак $(X;Y)$, отримано n пар чисел $(x_1;y_1)$, $(x_2;y_2),...,(x_n;y_n)$. За даними спостережень знайдемо вибіркове рівняння прямої лінії регресії, для визначеності Y по X

$$y=\alpha+\beta x\,. \tag{12.5}$$

У (12.5) $\overline{y}_x$ замінене на y, тому що різні значення x ознаки X і відповідні їм значення y ознаки Y спостерігалися по одному разу і тому групувати дані немає необхідності. Якщо позначити через $\tilde{y}_i=\alpha+\beta x_i$ наближене значення y_i, обчислене з рівняння регресії (12.5), то величина $y_i-\tilde{y}_i$ є відхиленням наближеного значення $\tilde{y}_i$ від точного y_i (рис. 12.2).

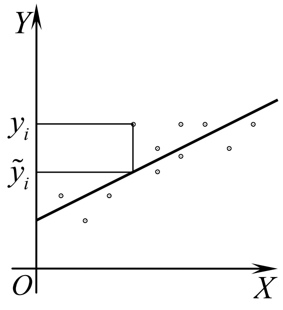

Рис. 12.2

Найбільш часте оцінювання параметрів α і β прямої регресії здійснюють на основі *методу найменших квадратів* (МНК), розробка якого належить К. Гауссу і П. Лапласу. Цей метод набув широкого застосування в економіко-статистичних розрахунках після створення теорії регресії. Згідно з МНК параметри α і β прямої регресії вибирають так, щоб сума квадратів відхилень $y_i - \tilde{y}_i$ була мінімальною, тобто з умови мінімізації функції:

$$S(\alpha;\beta) = \sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 = \sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2.$$

Необхідною умовою існування мінімуму функції $S(\alpha;\beta)$ є рівність нулю частинних похідних за невідомими параметрами α і β. Прирівнюючи частинні похідні S'_α і S'_β до нуля, одержимо систему рівнянь для визначення α і β:

$$\begin{cases} S'_\alpha = \sum_{i=1}^{n} 2(y_i - \alpha - \beta x_i)(-1) = 0; \\ S'_\beta = \sum_{i=1}^{n} 2(y_i - \alpha - \beta x_i)(-x_i) = 0; \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} \beta\sum_{i=1}^{n} x_i^2 + \alpha\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i; \\ \beta\sum_{i=1}^{n} x_i + \alpha n = \sum_{i=1}^{n} y_i. \end{cases} \quad (12.6)$$

Розв’язавши цю систему, знайдемо шукані параметри:

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

$$\beta = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}. \quad (12.7)$$

Якщо потрібно за результатами спостережень одержати лінійне рівняння регресії X по Y, то в рівнянні регресії $y = \alpha + \beta x$ треба поміняти місцями змінні x і y. При цьому одержимо рівняння $x = \alpha' + \beta' y$, де α' і β' обчислюються за формулами:

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

$$\beta' = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}. \qquad (12.8)$$

Зазначимо, що регресійні прямі $y = \alpha + \beta x$ і $x = \alpha' + \beta' y$ різні. Перша пряма виходить у результаті розв'язку задачі про мінімізацію суми квадратів відхилень по вертикалі, а друга – при розв'язку задачі про мінімізацію суми квадратів відхилень по горизонталі.

На практиці для знаходження рівнянь регресії складається наступна таблиця 12.1:

Таблиця 12.1

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	x_1	y_1	x_1^2	y_1^2	$x_1 y_1$
2	x_2	y_2	x_2^2	y_2^2	$x_2 y_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n	x_n^2	y_n^2	$x_n y_n$
$\sum$	$\sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n} y_i$	$\sum_{i=1}^{n} x_i^2$	$\sum_{i=1}^{n} y_i^2$	$\sum_{i=1}^{n} x_i y_i$

В останньому рядку цієї таблиці суми й визначають коефіцієнти α й β або α' й β' у формулах (12.7) або (12.8) відповідно.

Приклад 1. За даними таблиці спостережень

x_i	1	1,5	2	2,5	3
y_i	2,1	2,2	2,7	2,8	2,85

скласти рівняння регресії *Y* по *X* і *X* по *Y*.

○ Складемо таблицю 12.2:

Таблиця 12.2

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	1	2,1	1	4,41	2,1
2	1,5	2,2	2,25	4,84	3,3
3	2	2,7	4	7,29	5,4
4	2,5	2,8	6,25	7,84	7
5	3	2,85	9	8,1225	8,55
$\sum$	10	12,65	22,5	32,5025	26,35

За формулами (12.7) при $n = 5$ одержуємо:

$$\alpha = \frac{22,5 \cdot 12,65 - 10 \cdot 26,35}{5 \cdot 22,5 - 100} \approx 1,69;$$

$$\beta = \frac{5 \cdot 26,35 - 10 \cdot 12,65}{5 \cdot 22,5 - 100} \approx 0,42.$$

Отже, рівняння регресії Y по X є

$$y = \beta x + \alpha = 0,42x + 1,69.$$

Аналогічно за формулами (12.8) знаходимо:

$$\alpha' = \frac{32,5025 \cdot 10 - 12,65 \cdot 26,35}{5 \cdot 32,5025 - (12,65)^2} \approx -3,33;$$

$$\beta' = \frac{5 \cdot 26,35 - 12,65 \cdot 10}{5 \cdot 32,5025 - (12,65)^2} \approx 2,11.$$

Звідси рівняння X по Y є

$$x = \beta' y + \alpha' = 2,11y - 3,33. \bullet$$

Якщо число вимірів велике, то з метою спрощення розрахунків експериментальні дані потрібно групувати, тобто поєднувати в таблицю 12.3, яка називається *кореляційною*.

Пояснимо, як заповнюється кореляційна таблиця.

У першому стовпці (першому рядку) перераховуються у вибірці значення величини $X: x_i,\ i=\overline{1,s}\quad (Y: y_j,\ j=\overline{1,k})$.

Таблиця 12.3

$X \backslash Y$	y_1	y_2	$\cdots$	y_k	n_{x_i}
x_1	n_{11}	n_{12}	$\cdots$	n_{1k}	$n_{x_1}=\sum_{j=1}^{k} n_{1j}$
x_2	n_{21}	n_{22}	$\cdots$	n_{2k}	$n_{x_2}=\sum_{j=1}^{k} n_{2j}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_s	n_{s1}	n_{s2}	$\cdots$	n_{sk}	$n_{x_S}=\sum_{j=1}^{k} n_{sj}$
n_{y_j}	$n_{y_1}=\sum_{i=1}^{s} n_{i1}$	$n_{y_2}=\sum_{i=1}^{s} n_{i2}$	$\cdots$	$n_{y_k}=\sum_{j=1}^{s} n_{ik}$	$n=\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}$

Якщо кількість різних значень величин X і Y велике або ці величини розподілені неперервно, то здійснюється угруповання їхніх значень по інтервалах. У цьому випадку x_i і y_j являють собою середини відповідних інтервалів.

На перетинанні рядків і стовпців указуються частоти n_{ij}, які дорівнюють числу появ у вибірці пари $(x_i; y_j)$. Якщо пари значень ознак $(x_i; y_j)$ не спостерігалися, то у відповідних клітинках таблиці ставиться риска.

В останньому рядку (останньому стовпці) указуються числа $n_{y_j},\ j=\overline{1,k}\ (n_{x_i},\ i=\overline{1,s})$, рівні кількості появ у вибірці значень $y_j\ (x_i)$ незалежно від того, у парі з яким зі значень величин $X(Y)$ воно з'явилося.

Кореляційна таблиця містить всю інформацію, отриману в результаті вибіркових спостережень величин X і Y. Звідси з урахуванням частот появ змінних x_i і y_j, одержимо:

$$\sum_{i=1}^{n} x_i=\sum_{i=1}^{s} n_{x_i} x_i;\qquad \sum_{i=1}^{n} y_i=\sum_{j=1}^{k} n_{y_j} y_j;\qquad \sum_{i=1}^{n} x_i^2=\sum_{i=1}^{s} n_{x_i} x_i^2;$$

$$\sum_{i=1}^{n} y_i^2 = \sum_{j=1}^{k} n_{y_j} y_j^2; \qquad \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} x_i y_j.$$

Підставивши ці суми до формули (12.7) і (12.8), одержимо:

$$\alpha = \frac{\left(\sum_{i=1}^{s} n_{x_i} x_i^2\right)\left(\sum_{j=1}^{k} n_{y_j} y_j\right) - \left(\sum_{i=1}^{s} n_{x_i} x_i\right)\left(\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} x_i y_j\right)}{n\sum_{i=1}^{s} n_{x_i} x_i^2 - \left(\sum_{i=1}^{s} n_{x_i} x_i\right)^2};$$

$$\beta = \frac{n\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} x_i y_j - \left(\sum_{i=1}^{s} n_{x_i} x_i\right)\left(\sum_{j=1}^{k} n_{y_j} y_j\right)}{n\sum_{i=1}^{s} n_{x_i} x_i^2 - \left(\sum_{i=1}^{s} n_{x_i} x_i\right)^2}; \qquad (12.9)$$

$$\alpha' = \frac{\left(\sum_{j=1}^{k} n_{y_j} y_j^2\right)\left(\sum_{i=1}^{s} n_{x_i} x_i\right) - \left(\sum_{j=1}^{k} n_{y_j} y_j\right)\left(\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} x_i y_j\right)}{n\sum_{j=1}^{k} n_{y_j} y_j^2 - \left(\sum_{j=1}^{k} n_{y_j} y_j\right)^2};$$

$$\beta' = \frac{n\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} x_i y_j - \left(\sum_{j=1}^{k} n_{y_j} y_j\right)\left(\sum_{i=1}^{s} n_{x_i} x_i\right)}{n\sum_{j=1}^{k} n_{y_j} y_j^2 - \left(\sum_{j=1}^{k} n_{y_j} y_j\right)^2}. \qquad (12.10)$$

Як відомо, система рівнянь для визначення параметрів α і β рівняння прямої лінії регресії Y на X має вигляд (12.6). Передбачалося, що значення X і відповідні їм значення Y спостерігалися один раз. Тепер же допустимо, що отримано велику кількість даних, серед яких є повторювані, і вони згруповані у вигляді кореляційної таблиці. Запишемо систему (12.6) так, щоб вона відображала дані кореляційної таблиці. Скористаємося тотожностями:

$$\sum_{i=1}^{n} x_i = n\overline{x}; \ \sum_{i=1}^{n} y_i = n\overline{y}; \ \sum_{i=1}^{n} x_i^2 = n\overline{x^2}; \ \sum_{i=1}^{n} x_i y_i = n\overline{xy}, \qquad (12.11)$$

де $\overline{x}$, $\overline{y}$ – вибіркові середні; $\overline{x^2}$ – вибіркове середнє квадрата; $\overline{xy}$ – вибіркове середнє добутку. Нагадаємо, що *вибіркова коваріація* s_{xy} визначається рівністю

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x}\cdot\overline{y} = \overline{xy} - \overline{x}\cdot\overline{y}. \qquad (12.12)$$

Вибіркові дисперсії визначаються співвідношеннями:

$$\begin{aligned} s_x^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 = \overline{x^2} - \overline{x}^2, \\ s_y^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \overline{y}^2 = \overline{y^2} - \overline{y}^2. \end{aligned} \qquad (12.13)$$

За визначенням *вибірковий коефіцієнт кореляції*

$$\rho_{xy} = \frac{s_{xy}}{s_x s_y}. \qquad (12.14)$$

Підставивши праві частини (12.11) до системи (12.6), одержимо

$$\begin{cases} \beta n\overline{x^2} + \alpha n\overline{x} = n\overline{xy}, \\ \beta n\overline{x} + \alpha n = n\overline{y}, \end{cases} \Leftrightarrow \begin{cases} \overline{x^2}\beta + \overline{x}\,\alpha = \overline{xy}, \\ \overline{x}\beta + \alpha = \overline{y}. \end{cases} \qquad (12.15)$$

Розв'язуючи систему (12.15) за формулами Крамера, одержимо:

$$\Delta = \begin{vmatrix} \overline{x^2} & \overline{x} \\ \overline{x} & 1 \end{vmatrix} = \overline{x^2} - \overline{x}^2 = s_x^2; \qquad \Delta_\beta = \begin{vmatrix} \overline{xy} & \overline{x} \\ \overline{y} & 1 \end{vmatrix} = \overline{xy} - \overline{x}\cdot\overline{y} = s_{xy};$$

$$\Delta_\alpha = \begin{vmatrix} \overline{x^2} & \overline{xy} \\ \overline{x} & \overline{y} \end{vmatrix} = \overline{x^2}\,\overline{y} - \overline{x}\;\overline{xy} = \overline{x^2}\,\overline{y} - \overline{x}^2\overline{y} + \overline{x}^2\overline{y} - \overline{x}\;\overline{xy} =$$

$$= \overline{y}(\overline{x^2} - \overline{x}^2) - \overline{x}(\overline{xy} - \overline{x}\cdot\overline{y}) = \overline{y}\cdot s_x^2 - \overline{x}\cdot s_{xy};$$

$$\alpha = \frac{\Delta_\alpha}{\Delta} = \frac{\overline{y}\cdot s_x^2 - \overline{x}\cdot s_{xy}}{s_x^2} = \overline{y} - \frac{s_{xy}}{s_x^2}\overline{x}; \qquad \beta = \frac{\Delta_\beta}{\Delta} = \frac{s_{xy}}{s_x^2}.$$

Підставивши коефіцієнти α і β до рівняння регресії $\overline{y} = \alpha + \beta x$,

одержимо $\overline{y}_x = \overline{y} - \frac{s_{xy}}{s_x^2}\overline{x} + \frac{s_{xy}}{s_x^2}x \;\Leftrightarrow\; \overline{y}_x - \overline{y} = \frac{s_{xy}}{s_x^2}(x - \overline{x}),$

$$\frac{s_{xy}}{s_x^2} = \frac{s_{xy}}{s_x s_y}\cdot\frac{s_y}{s_x} = |12.14| = \rho_{xy}\frac{s_y}{s_x}.$$

Таким чином, рівняння прямої регресії Y по X має вигляд

$$\overline{y}_x - \overline{y} = \rho_{xy} \frac{s_y}{s_x}(x - \overline{x}). \tag{12.16}$$

Аналогічно знайдемо, що рівняння прямої регресії X по Y запишеться у вигляді

$$\overline{x}_y - \overline{x} = \rho_{xy} \frac{s_x}{s_y}(y - \overline{y}). \tag{12.17}$$

З рівнянь (12.16) і (12.17) випливає, що прямі регресії Y по X і X по Y проходять через точку ($\overline{x}$; $\overline{y}$). Ці прямі збігаються, коли $|\,\rho_{xy}\,|^2 = 1$.

Величини $\rho_{xy} \dfrac{s_y}{s_x}$ і $\rho_{xy} \dfrac{s_x}{s_y}$ називаються *вибірковими коефіцієнтами лінійної регресії* та позначаються

$$\rho_{y|x} = \rho_{xy} \frac{s_y}{s_x}; \; \rho_{x|y} = \rho_{xy} \frac{s_x}{s_y}. \tag{12.18}$$

Перемножуючи праві й ліві частини рівностей (12.18), після добування кореня одержимо $\rho_{xy} = \pm\sqrt{\rho_{y|x}\rho_{x|y}}$, тобто коефіцієнт кореляції є середнє геометричне коефіцієнтів лінійної регресії. Знак у правій частині ρ_{xy} збігається зі знаками $\rho_{y|x}$ і $\rho_{x|y}$.

Коефіцієнт регресії Y по X (X по Y) показує, на скільки одиниць у середньому змінюється змінна Y (X) при збільшенні змінної X (Y) на одну одиницю.

Якщо дані спостережень над ознаками X і Y задані у вигляді кореляційної таблиці з рівновіддаленими варіантами, то доцільно перейти до умовних варіант:

$$u_i = (x_i - C_1)/h_1; \quad v_j = (y_j - C_2)/h_2,$$

де C_1 і C_2 – помилкові нулі варіант X і Y відповідно, h_1 і h_2 – кроки, тобто різниці між двома сусідніми варіантами X і Y.

Тоді відповідно до (8.9) і (8.18)

$$\overline{x} = \frac{h_1}{n}\sum_{i=1}^{s} u_i n_{x_i} + C_1 = h_1\overline{u} + C_1, \qquad (12.19)$$

$$\overline{y} = \frac{h_2}{n}\sum_{j=1}^{k} v_j n_{y_j} + C_2 = h_2\overline{v} + C_2, \qquad (12.20)$$

$$s_x^2 = \frac{h_1^2}{n}\sum_{i=1}^{s} u_i^2 n_{x_i} - (\overline{x} - C_1)^2 = h_1^2\overline{u^2} - (\overline{x} - C_1)^2, \qquad (12.21)$$

$$s_y^2 = \frac{h_2^2}{n}\sum_{j=1}^{k} v_j^2 n_{y_j} - (\overline{y} - C_2)^2 = h_2^2\overline{v^2} - (\overline{y} - C_2)^2. \qquad (12.22)$$

Виразимо коваріацію через умовні варіанти. Маємо

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{s}\sum_{j=1}^{k} x_i y_j n_{ij} - \frac{1}{n}\sum_{i=1}^{s} x_i n_{x_i}\frac{1}{n} = \frac{h_1h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{k}\frac{(x_i - C_1 + C_1)}{h_1}\frac{y_j - C_2 + C_2}{h_2}n_{ij} -$$

$$-\frac{h_1}{n}\sum_{i=1}^{s}\frac{(x_i - C_1 + C_{1)}}{h_1}n_{x_i}\cdot\frac{h_2}{n}\sum_{j=1}^{k}\frac{y_j - C_2 + C_2}{h_2}n_{y_j} = \frac{h_1h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{k}\left(u_i + \frac{C_1}{h_1}\right)\times$$

$$\times\left(v_j + \frac{C_2}{h_2}\right)n_{ij} - \frac{h_1h_2}{n^2}\sum_{i=1}^{s}\left(u_i + \frac{C_1}{h_1}\right)n_{x_i}\times\sum_{j=1}^{k}\left(v_j + \frac{C_2}{h_2}\right)n_{y_j} = \frac{h_1h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{k}\Big(u_i v_j +$$

$$+\frac{C_1}{h_1}v_j + \frac{C_2}{h_2}u_i + \frac{C_1C_2}{h_1h_2}\Big)n_{ij} - \frac{h_1h_2}{n^2}\left(\sum_{i=1}^{s}u_i n_{x_i} + \frac{C_1}{h_1}n\right)\left(\sum_{j=1}^{k}v_j n_{y_j} + \frac{C_2}{h_2}n\right) =$$

$$= \frac{h_1h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{k}u_i v_j n_{ij} + \frac{C_1h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{k}v_j n_{ij} + \frac{C_2h_1}{n}\sum_{i=1}^{s}\sum_{j=1}^{k}u_i n_{ij} + C_1C_2 - \frac{h_1h_2}{n^2}\sum_{i=1}^{s}u_i n_{x_i}\times$$

$$\times\sum_{j=1}^{k}v_j n_{y_j} - \frac{C_1h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{k}v_j n_{ij} - \frac{C_2h_1}{n}\sum_{i=1}^{s}\sum_{j=1}^{k}u_i n_{ij} - C_1C_2 = \frac{h_1h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{k}u_i v_j n_{ij} -$$

$$-\frac{h_1h_2}{n^2}\sum_{i=1}^{s}u_i n_{x_i}\sum_{j=1}^{k}v_j n_{y_j} = h_1h_2\,(\overline{uv} - \overline{u}\cdot\overline{v}),$$

де $n = \sum_{i=1}^{s} n_{x_i} = \sum_{j=1}^{k} n_{y_j} = \sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}$,

$$\overline{u} = \frac{1}{n}\sum_{i=1}^{s}u_i n_{x_i}; \qquad \overline{v} = \frac{1}{n}\sum_{j=1}^{k}v_j n_{y_j}; \qquad \overline{uv} = \frac{1}{n}\sum_{i=1}^{s}\sum_{j=1}^{k}u_i v_j n_{ij}.$$

Таким чином,

$$s_{xy} = \frac{h_1h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{k}u_i v_j n_{ij} - \frac{h_1h_2}{n^2}\sum_{i=1}^{s}u_i n_{x_i}\sum_{j=1}^{k}v_j n_{y_j}. \qquad (12.23)$$

Приклад 2. Знайти рівняння прямих регресій Y по X і X по Y за даними кореляційної таблиці 12.4 і побудувати їхні графіки.

○ Перейдемо до умовних варіантів $u_i = (x_i - C_1)/h_1 = (x_i - 20)/5$, $v_j = (y_j - C_2)/h_2 = (y_j - 50)/10$, де як помилкові нулі C_1 і C_2 узяті варіанти відповідно $x = 20$ і $y = 50$, розташовані наближено в середині варіаційних рядів, тобто $C_1 = 20$ і $C_2 = 50$, а $h_1 = 5$ і $h_2 = 10$.

Таблиця 12.4

X \ Y	30	40	50	60	70	n_{x_i}
5	2	–	–	–	–	2
10	6	5	–	–	–	11
15	–	3	7	4	–	14
20	–	–	40	9	4	53
25	–	–	2	6	7	15
30	–	–	–	–	5	5
n_{y_j}	8	8	49	19	16	$n = 100$

Зобразимо кореляційну табл. 12.4 у вигляді табл. 12.5, де два передостанні стовпці та два передостанні рядки містять наступні суми:

$$\sum_{i=1}^{6} u_i n_{x_i} = -17; \quad \sum_{i=1}^{6} u_i^2 n_{x_i} = 111; \quad \sum_{j=1}^{5} v_j n_{y_j} = 27; \quad \sum_{j=1}^{5} v_j^2 n_{y_j} = 123 .$$

Для зручності обчислення суми $\sum_{i=1}^{6}\sum_{j=1}^{5} u_i v_j n_{ij}$, спочатку розраховуємо $u_i v_j$ і проставляємо ці значення у верхньому правому кутку тих клітин, у яких $n_{ij} \neq 0$, а потім знаходимо добутки $u_i v_j n_{ij}$. Підсумовуючи їх за рядками і стовпцями, записуємо отримані результати відповідно в останньому стовпці та останньому рядку табл. 12.5. Підсумовуючи значення останніх стовпця і рядка, одержимо в правій нижній клітинці табл. 12.5 $\sum_{i=1}^{6}\sum_{j=1}^{5} u_i v_j n_{ij} = 85$; їхній збіг свідчить про правильність обчислень.

Використовуючи формули (12.19) − (12.23), одержимо:

$$\overline{x} = 5\cdot(-17)/100 + 20 = 19{,}15; \qquad \overline{y} = 10\cdot 27/100 + 50 = 52{,}7;$$

$$s_x^2 = 25\cdot 111/100 - (19{,}15-20)^2 = 27{,}0275; \quad s_x \approx 5{,}1988;$$

$$s_y^2 = 100\cdot 123/100 - (52{,}7-50)^2 = 115{,}71; \quad s_y \approx 10{,}7569;$$

$$s_{xy} = 5\cdot 10\cdot 85/100 - 5\cdot 10\cdot(-17)27/10000 = 44{,}795.$$

Таблиця 12.5

	Y	30	40	50	60	70	n_{x_i}	$u_i n_{x_i}$	$u_i^2 n_{x_i}$	$\sum_{j=1}^{5} u_i v_j n_{ij}$
X	u_i \ v_j	–2	–1	0	1	2				
5	–3	$2\ \boxed{6}$	–	–	–	–	2	–6	18	12
10	–2	$6\ \boxed{4}$	$5\ \boxed{2}$	–	–	–	11	–22	44	34
15	–1	–	$3\ \boxed{1}$	$7\ \boxed{0}$	$4\ \boxed{-1}$	–	14	–14	14	–1
20	0	–	–	$40\ \boxed{0}$	$9\ \boxed{0}$	$4\ \boxed{0}$	53	0	0	0
25	1	–	–	$2\ \boxed{0}$	$6\ \boxed{1}$	$7\ \boxed{2}$	15	15	15	20
30	2	–	–	–	–	$5\ \boxed{4}$	5	10	20	20
n_{y_j}		8	8	49	19	16	100	–17	111	–
$v_j n_{y_j}$		–16	–8	0	19	32	27	–	–	–
$v_j^2 n_{y_j}$		32	8	0	19	64	123	–	–	–
$\sum_{i=1}^{6} u_i v_j n_{ij}$		36	13	0	2	34	–	–	–	85

За формулою (12.14) знайдемо вибірковий коефіцієнт кореляції

$$\rho_{xy} = 44{,}795/(5{,}1988\cdot 10{,}7569) \approx 0{,}801,$$

а потім, використовуючи (12.16) і (12.17), одержимо шукані рівняння регресії

$$\overline{y}_x - 52,7 = 0,801 \cdot \frac{10,7569}{5,1988}(x - 19,15) \Rightarrow \overline{y}_x = 1,6574x + 20,9616;$$

$$\overline{x}_y - 19,15 = 0,801 \cdot \frac{5,1988}{10,7569}(y - 52,7) \Rightarrow \overline{x}_y = 0,3871y - 1,2502.$$

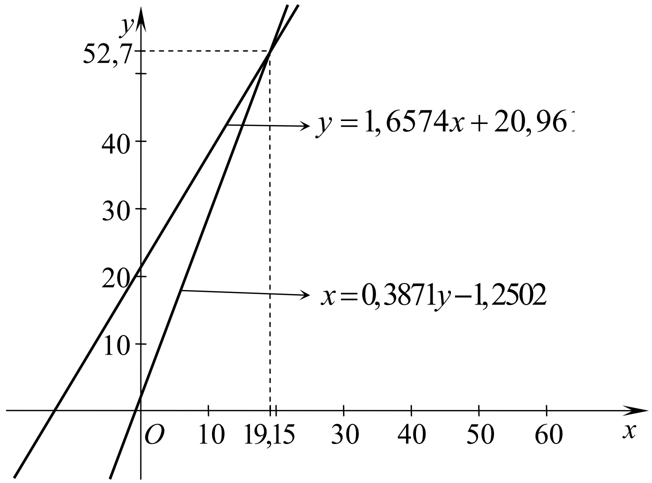

Рис. 12.3

На рис. 12.3 надано графіки отриманих прямих регресії.

12.3 Коефіцієнт кореляції

Для кількісної характеристики тісноти лінійного кореляційного зв'язку між двома величинами X і Y генеральної сукупності, у розділі 6 було введене поняття *коефіцієнта лінійної кореляції*, який визначався зі співвідношення

$$r_{xy} = \frac{M(XY) - m_x m_y}{\sigma_x \sigma_y}, \tag{12.24}$$

де $m_x = M(X)$, $m_y = M(Y)$, $\sigma_x = \sqrt{D(X)}$, $\sigma_y = \sqrt{D(Y)}$.

Відомо, що якщо величини X і Y незалежні, то $r_{xy} = 0$; якщо $r_{xy} = \pm 1$, то X і Y зв'язані лінійною функціональною залежністю, причому, дорівнює $r_{xy}=1$ у випадку зростаючої залежності та $r_{xy}=-1$ у випадку спадної; $|r_{xy}|\le 1$. При цьому $r_{xy}>0$, якщо при зростанні однієї величини (наприклад, X) $M(Y)$ збільшується, і $r_{xy}<0$ у протилежному випадку.

На практиці для оцінювання тісноти лінійного кореляційного зв'язку між величинами X і Y за результатами вибіркових спостережень використовується вибірковий коефіцієнт лінійної кореляції, який визначається за формулою (12.14), тобто

$$\rho_{xy}=\frac{s_{xy}}{s_x s_y}=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{s_x s_y}. \qquad (12.25)$$

Формулу (12.25) зазвичай застосовують у дещо перетвореному вигляді. Так, підставивши до неї розгорнуті вирази для s_x, s_y і s_{xy}, легко одержати наступну формулу:

$$\rho_{xy}=\frac{n\sum_{i=1}^{s}\sum_{j=1}^{k}x_i y_j n_{ij}-\left(\sum_{i=1}^{s}x_i n_{x_i}\right)\left(\sum_{j=1}^{k}y_j n_{y_j}\right)}{\sqrt{n\sum_{i=1}^{s}x_i^2 n_{x_i}-\left(\sum_{i=1}^{s}x_i n_{x_i}\right)^2}\cdot\sqrt{n\sum_{j=1}^{k}y_j^2 n_{y_j}-\left(\sum_{j=1}^{k}y_j n_{y_j}\right)^2}}. \qquad (12.26)$$

Якщо дані не згруповані у вигляді кореляційної таблиці та представляють n пар чисел (x_i ; y_j), то для обчислення коефіцієнта кореляції у формулі (12.26) варто взяти $n_{ij}=n_{x_i}=n_{y_j}=1$, $i=j$, а $\sum_{i=1}^{s}\sum_{j=1}^{k}$ замінити на $\sum_{i=1}^{n}$.

Коефіцієнт кореляції ρ_{xy} є безрозмірною величиною (оскільки розмірності чисельника і знаменника є розмірностю добутку XY); його величина не залежить від вибору одиниць виміру обох змінних; величина ρ_{xy} набуває значення на відрізку [–1; 1]. Наближена до нуля величина коефіцієнта кореляції говорить про відсутність лінійного зв'язку змінних, але не заперечує можливість існування іншої форми залежності між ними.

Хоча вибірковий коефіцієнт кореляції ρ_{xy} являє собою обґрунтовану оцінку для r_{xy}, однак більш надійну оцінку наближеності ρ_{xy} до r_{xy} за даними вибірки можна дати лише в тому випадку, коли розподіл величин X і Y достатньо наближений до нормальної форми. Наприклад, для оцінки r_{xy} нормально розподіленої генеральної сукупності, у випадку великих вибірок $(n \geq 50)$, можна скористатися формулою

$$\rho_{xy} - u_{\gamma}\frac{1-\rho_{xy}^2}{\sqrt{n}} \leq r_{xy} \leq \rho_{xy} + u_{\gamma}\frac{1-\rho_{xy}^2}{\sqrt{n}}, \qquad (12.27)$$

де u_{γ} корінь рівняння $2\Phi(\varepsilon) = \gamma$, що визначається з таблиці (Д 4) за заданою довірчою ймовірністю γ, а величина

$$\delta = u_{\gamma}\frac{1-\rho_{xy}^2}{\sqrt{n}} \qquad (12.28)$$

називається точністю оцінки.

Наведемо схему находження точності δ і довірчих границь, які відповідають надійності γ:

$$\left.\begin{matrix} n \\ \rho_{xy} \\ \gamma \end{matrix}\right\} \to 2\Phi(\varepsilon) = \gamma \xrightarrow{\text{Д 4}} \varepsilon = u_{\gamma} \to \delta = u_{\gamma}\frac{1-\rho_{xy}^2}{\sqrt{n}} \to \begin{cases} \rho_{xy} - \delta, \\ \rho_{xy} + \delta. \end{cases} \qquad (12.29)$$

Приклад 3. Знайти довірчий інтервал для оцінки коефіцієнта кореляції r_{xy} з надійністю $\gamma = 0,95$, якщо $\rho_{xy} = 0,23$, $n = 320$.

○ За схемою (12.29), знаходимо

$$\left.\begin{matrix} n = 320 \\ \rho_{xy} = 0,23 \\ \gamma = 0,95 \end{matrix}\right\} \to 2\Phi(\varepsilon) = 0,95 \to \varepsilon = 1,96 \to$$

$$\to \delta = 1,96\frac{1-(0,23)^2}{\sqrt{230}} \approx 0,104 \to \begin{cases} 0,23 - 0,104, \\ 0,23 + 0,104. \end{cases}$$

Звідси, 0,126 < r_{xy} < 0,334. ●

Застосовуючи коефіцієнт кореляції як міру зв'язку, потрібно мати на увазі, що він отриманий на основі даних вибірки і, отже, схильний до впливу випадковості. Якщо об'єм вибірки невеликий, то знайти вибіркову помилку цієї величини досить складно, тому на практиці зазвичай замість визначення помилки коефіцієнта кореляції перевіряють гіпотезу про його значущість, тобто, чи істотно ρ_{xy} відрізняється від нуля або цю відмінність можна приписати впливу випадковості, пов'язаної з вибіркою. Інакше кажучи, виникає необхідність при заданому рівні значущості α перевірити нульову гіпотезу $H_0 : r_{xy} = 0$ при конкуруючій гіпотезі $H_1 : r_{xy} \neq 0$.

Якщо нульова гіпотеза відкидається, то це означає, що ρ_{xy} істотно відрізняється від нуля, а *X* і *Y* корельовані, тобто зв'язані лінійною залежністю. Якщо нульова гіпотеза буде прийнята, то ρ_{xy} неістотно відрізняється від нуля, а *X* і *Y* некорельовані, тобто не зв'язані лінійною залежністю.

Як критерій перевірки нульової гіпотези візьмемо випадкову величину

$$T = \frac{\rho_{xy}\sqrt{n-2}}{\sqrt{1-\rho_{xy}^2}}, \tag{12.30}$$

яка має розподіл Стьюдента із $k = n-2$ ступенями вільності. Число ступенів вільності менше за число спостережень на 2, оскільки до формули вибіркового коефіцієнта кореляції входять середні вибіркові значення *X* і *Y*, для розрахунку яких використовуються дві лінійні формули їхньої залежності від спостережень випадкових величин. Якщо альтернативна гіпотеза має вигляд $H_1 : r_{xy} \neq 0$, то критична область є двобічною. Перевірку нульової гіпотези будемо здійснювати за наступною схемою:

$$\left.\begin{array}{c} n; k = n-2 \\ \rho_{xy} \\ \alpha \end{array}\right\} \to H_0 : r_{xy} = 0 \to T_{\text{спос}} = \frac{\rho_{xy}\sqrt{n-2}}{\sqrt{1-\rho_{xy}^2}} \to$$

$$\to H_1 : r_{xy} \neq 0 \to t_{\text{кр}}(\alpha; k) \xrightarrow{Д7} t_{\text{кр}}^{\text{дв}} \to \begin{cases} |T_{\text{спос}}| < t_{\text{кр}}^{\text{дв}} \to \mathbf{П}, \\ |T_{\text{спос}}| > t_{\text{кр}}^{\text{дв}} \to \mathbf{В}. \end{cases} \tag{12.31}$$

Приклад 4. За вибіркою об'єму *n* = 62, витягнутої з нормальної двовимірної генеральної сукупності (*X*; *Y*), знайдений вибірковий коефіцієнт кореляції $\rho_{xy} = 0,3$. Потрібно при рівні значущості

$\alpha = 0,01$ перевірити нульову гіпотезу $H_0 : r_{xy} = 0$ при альтернативній гіпотезі $H_1 : r_{xy} \neq 0$.

○ Використовуючи схему (12.31), одержимо

$$\left.\begin{array}{l} n = 62; k = 60 \\ \rho_{xy} = 0,3 \\ \alpha = 0,01 \end{array}\right\} \to H_0 : r_{xy} = 0 \to T_{\text{спос}} = \frac{0,3 \cdot \sqrt{60}}{\sqrt{1 - 0,09}} \approx 2,436 \to$$

$$\to H_1 : r_{xy} \neq 0 \to t_{\text{кр}}(0,01;60) \xrightarrow{Д7} t_{\text{кр}}^{\text{дв}} = 2,66 \to |T_{\text{спос}}| < t_{\text{кр}}^{\text{дв}} \to \mathbf{П}.$$

Таким чином, немає підстав відкидати нульову гіпотезу. Інакше кажучи, ρ_{xy} незначно відрізняється від нуля, тобто *X* і *Y* некорельовані, тобто не зв'язані лінійною залежністю. ●

12.4 Поняття про множинну кореляцію

У тих випадках, коли досліджується кореляційний зв'язок між величинами, число яких більше за два, уводять поняття *множинної кореляції*. Так, при дослідженні кореляційного зв'язку між трьома кількісними ознаками *X*, *Y* і *Z* можна ввести рівняння регресії

$$\overline{z}_{xy} = f(x; y),$$

де $\overline{z}_{xy}$ – середнє значення величини *Z*, що відповідає відповідним значенням *x* і *y*. Геометричною інтерпретацією цього рівняння є деяка поверхня в прямокутній системі координат тривимірного простору. У найбільш простому випадку лінійної кореляційної залежності ознак *X*, *Y* і *Z* вибіркове рівняння регресії має вигляд

$$\overline{z}_{xy} = ax + by + c.$$

Нехай у результаті незалежних спостережень над досліджуваною системою кількісних ознак (*X*; *Y*; *Z*), отримано *n* сукупностей чисел $(x_1; y_1; z_1)$, $(x_2; y_2; z_2)$,…, $(x_n; y_n; z_n)$. Припустимо, що функція регресії лінійна, тобто

$$z = ax + by + c. \tag{12.32}$$

У (12.32) $\overline{z}_{xy}$ замінене на z, тому що різні значення *x* і *y* ознак *X* і *Y*, і відповідні їм значення *z* ознаки *Z* спостерігалися один раз.

Якщо позначити $\tilde{z}_i = ax_i + by_i + c$ наближене значення z_i, обчислене з рівняння регресії (12.32), то величина $z_i - \tilde{z}_i$ є відхиленням наближеного значення $\tilde{z}_i$ від точного z_i. Коефіцієнти a, b і c рівняння регресії (12.32) знайдемо з вимоги МНК:

$$S(a;b;c) = \sum_{i=1}^{n}(z_i - \tilde{z}_i)^2 = \sum_{i=1}^{n}(z_i - ax_i - by_i - c)^2 \to \min.$$

Необхідні умови мінімуму функції S утворять систему

$$\begin{cases} \dfrac{\partial S}{\partial a} = 2\sum\limits_{i=1}^{n}(z_i - ax_i - by_i - c)(-x_i) = 0, \\ \dfrac{\partial S}{\partial b} = 2\sum\limits_{i=1}^{n}(z_i - ax_i - by_i - c)(-y_i) = 0, \\ \dfrac{\partial S}{\partial c} = 2\sum\limits_{i=1}^{n}(z_i - ax_i - by_i - c)(-1) = 0, \end{cases}$$

яка в результаті тотожних перетворень набуде вигляду

$$\begin{cases} a\sum\limits_{i=1}^{n}x_i^2 + b\sum\limits_{i=1}^{n}x_iy_i + c\sum\limits_{i=1}^{n}x_i = \sum\limits_{i=1}^{n}x_iz_i, \\ a\sum\limits_{i=1}^{n}x_iy_i + b\sum\limits_{i=1}^{n}y_i^2 + c\sum\limits_{i=1}^{n}y_i = \sum\limits_{i=1}^{n}y_iz_i, \\ a\sum\limits_{i=1}^{n}x_i + b\sum\limits_{i=1}^{n}y_i + cn = \sum\limits_{i=1}^{n}z_i. \end{cases} \tag{12.33}$$

Розв'язуючи систему (12.33) відносно невідомих параметрів a, b і c, а потім підставляючи їхні значення до (12.32), одержимо шукане рівняння регресії.

Зручніше, однак, знайти готові формули для розрахунку параметрів, а не розв'язувати систему щоразу. Коефіцієнти регресії a, b і параметр c у цьому випадку обчислюються за наступними формулами:

$$a = \frac{\rho_{xz} - \rho_{yz}\rho_{xy}}{1-\rho_{xy}^2}\frac{s_z}{s_x}; \quad b = \frac{\rho_{yz} - \rho_{xz}\rho_{xy}}{1-\rho_{xy}^2}\frac{s_z}{s_y}; \quad c = \overline{z} - a\overline{x} - b\overline{y}. \tag{12.34}$$

Тут ρ_{xy}, ρ_{xz}, ρ_{yz} – вибіркові коефіцієнти лінійної кореляції між відповідними ознаками; s_x, s_y, s_z – середні квадратичні відхилення; $\overline{x}$, $\overline{y}$, $\overline{z}$ – середні значення.

Тіснота зв'язку ознаки Z з ознаками X і Y, оцінюється *вибірковим сукупним коефіцієнтом кореляції*

$$R=\sqrt{(\rho_{xz}^2-2\rho_{xy}\rho_{xz}\rho_{yz}+\rho_{yz}^2)\Big/(1-\rho_{xy}^2)},\text{ причому } 0\le R\le 1. \quad (12.35)$$

Тіснота зв'язку між Z і X (при постійному Y), між Z і Y (при постійному X) оцінюється відповідно *частковими вибірковими коефіцієнтами кореляції*:

$$\rho_{xz(y)}=\frac{\rho_{xz}-\rho_{xy}\rho_{yz}}{\sqrt{(1-\rho_{xy}^2)(1-\rho_{yz}^2)}};\qquad \rho_{yz(x)}=\frac{\rho_{yz}-\rho_{xy}\rho_{xz}}{(1-\rho_{xy}^2)(1-\rho_{xz}^2)}. \qquad (12.36)$$

Ці коефіцієнти показують ступінь лінійної залежності між спостережуваними значеннями Z і X, а також Z і Y, якщо вплив третьої ознаки усунуто.

Приклад 5. Нехай змінна z знаходиться в лінійній залежності від змінних x і y. Для оцінювання параметрів рівняння регресії зібрано наступні дані:

Таблиця 12.6

i	1	2	3	4	5	6	7	8	9	10
z_i	10	12	17	13	15	10	14	12	16	18
x_i	2	2	8	2	6	3	5	3	9	10
y_i	1	2	10	4	8	4	7	3	10	11

Знайти рівняння лінійної регресії, виходячи із системи (12.33) і формул (12.34), а також сукупний і частинні вибіркові коефіцієнти кореляції.

○ Складемо розрахункову табл. 12.7

Таблиця 12.7

i	x_i	x_i^2	y_i	y_i^2	z_i	z_i^2	x_iy_i	x_iz_i	y_iz_i
1	2	4	1	1	10	100	2	20	10
2	2	4	2	4	12	144	4	24	24
3	8	64	10	100	17	289	80	136	170
4	2	4	4	16	13	169	8	26	52
5	6	36	8	64	15	225	48	90	120
6	3	9	4	16	10	100	12	30	40
7	5	25	7	49	14	196	35	70	98
8	3	9	3	9	12	144	9	36	36
9	9	81	10	100	16	256	90	144	160
10	10	100	11	121	18	324	110	180	198
$\sum$	50	336	60	480	137	1947	398	756	908

Побудуємо систему рівнянь (12.33), використовуючи розрахунки табл. 12.7

$$\begin{cases} 336a + 398b + 50c = 756, \\ 398a + 480b + 60c = 908, \\ 50a + 60b + 10c = 137. \end{cases}$$

Розв'язок системи відносно невідомих параметрів дає шукані оцінки: $a = 0,1285$; $b = 0,6117$; $c = 9,3872$. Рівняння регресії має вигляд $z = 0,1285a + 0,6117b + 9,3872$.

Для розв'язку цього ж прикладу за допомогою виразу (12.34) знайдемо: $\overline{x} = 5$; $\overline{y} = 6$; $\overline{z} = 13,7$; $\overline{x^2} = 33,6$; $\overline{y^2} = 48$; $\overline{z^2} = 194,7$; $\overline{xy} = 39,8$; $\overline{xz} = 75,6$; $\overline{yz} = 90,8$; $s_x = \sqrt{\overline{x^2} - \overline{x}^2} = \sqrt{33,6 - 25} \approx 2,9325757$;

$$s_y = \sqrt{\overline{y^2} - \overline{y}^2} = \sqrt{48 - 36} \approx 3,4641016;$$

$$s_z = \sqrt{\overline{z^2} - \overline{z}^2} = \sqrt{194,7 - 187,69} \approx 2,6476405;$$

$$s_{xy} = \overline{xy} - \overline{x} \cdot \overline{y} = 39,8 - 30 = 9,8; \quad s_{xz} = \overline{xz} - \overline{x} \cdot \overline{z} = 75,6 - 68,5 = 7,1;$$

$$s_{yz} = \overline{yz} - \overline{y} \cdot \overline{z} = 90,8 - 82,2 = 8,6;$$

$$\rho_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{9,8}{2,9325757 \cdot 3,4641016} \approx 0,9646865;$$

$$\rho_{xz} = \frac{s_{xz}}{s_x s_z} = \frac{7,1}{2,9325757 \cdot 2,6476405} \approx 0,9144292;$$

$$\rho_{yz} = \frac{s_{yz}}{s_y s_z} = \frac{8,6}{3,4641016 \cdot 2,6476405} \approx 0,9376674;$$

$$a = \frac{0,9144292 - 0,9376674 \cdot 0,9646865}{1 - (0,9646865)^2} \cdot \frac{2,6476405}{2,9325757} \approx 0,1285;$$

$$b = \frac{0,9376674 - 0,9144292 \cdot 0,9646865}{1 - (0,9646865)^2} \cdot \frac{2,6476405}{3,4641016} \approx 0,6117;$$

$$c = 13,7 - 0,1285 \cdot 5 - 0,6117 \cdot 6 \approx 9,3872.$$

Тіснота зв'язку ознаки Z з ознаками X і Y, X (при постійному Y), Y (при постійному X) визначається за формулами (12.35) і (12.36):

$$R = \sqrt{(0,836181 - 1,654303 + 0,879220)/0,069380} \approx 0,9384;$$

$$\rho_{xz(y)} = \frac{0,9144292 - 0,904555}{\sqrt{0,069380 \cdot 0,120780}} \approx 0,1079;$$

$$\rho_{yz(x)} \frac{0,9376674 - 0,8821375}{\sqrt{0,069380 \cdot 0,163819}} \approx 0,5209. \bullet$$

Питання для самоперевірки

1. Дайте визначення функціональної, статистичної та кореляційної залежностей між ознаками *X* і *Y*.

2. Що називається умовним середнім $\overline{y}_x (\overline{x}_y)$ спостережуваних значень *Y*(*X*)?

3. Що називається кореляційним полем?

4. Сформулюйте основні задачі регресійного і кореляційного аналізу.

5. Запишіть модель лінійної парної регресії *Y* по *X* і *X* по *Y*.

6. Запишіть формули для знаходження параметрів α і β (α' і β') прямої регресії.

7. Чим вирізняються регресійні прямі $y = \alpha + \beta x$ і $x = \alpha' + \beta' y$?

8. Поясніть, як заповнюється кореляційна таблиця.

9. Чому дорівнює вибірковий коефіцієнт кореляції?

10.Чому дорівнюють вибіркові коефіцієнти лінійної регресії?

11. Установіть зв'язок r_{xy} з $r_{y|x}$ і $r_{x|y}$.

12. Для чого використовується вибірковий коефіцієнт лінійної кореляції? Наведіть формулу.

13. Запишіть модель лінійної регресійної залежності ознак *X*, *Y* і *Z*.

14. Запишіть формули для знаходження параметрів a, b і c лінійного рівняння регресії.

15. Для чого використовуються вибіркові сукупний і частинний коефіцієнти кореляції? Наведіть формули.

Індивідуальні завдання 12. Скласти кореляційну таблицю для двовимірної вибірки *XY*, згідно з відповідним варіантом табл. 12.8, знайти рівняння прямих регресій *Y* по *X* і *X* по *Y*, побудувати кореляційне поле й графіки прямих регресій.

Таблиця 12.8

№	***Варіанти***											
	1				***2***				***3***			
	X	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*
1	67	–201	88	–259	48	99	47	97	46	50	42	46
2	68	–199	86	–251	40	83	50	102	55	57	48	50
3	70	–206	71	–207	52	106	46	101	57	61	60	63
4	76	–221	87	–257	50	107	56	112	55	58	62	64
5	80	–238	77	–227	39	79	42	93	51	51	46	47
6	87	–256	73	–214	47	100	41	84	62	70	51	59
7	75	–222	82	–243	38	80	55	112	43	43	69	76
8	79	–230	74	–214	46	96	40	80	64	71	60	60
9	79	–234	67	–196	47	98	50	105	56	64	57	61
10	73	–217	82	–239	44	97	49	99	65	67	62	65
11	86	–253	72	–210	45	92	54	113	56	63	58	66
12	78	–228	74	–221	44	90	45	99	51	58	54	60
13	79	–230	72	–212	53	108	36	75	58	60	57	63
14	67	–201	68	–203	52	107	41	91	42	47	44	45
15	79	–237	86	–252	45	96	46	101	46	54	55	62
16	82	–237	85	–251	42	86	38	83	54	60	50	54
17	70	–209	71	–206	45	98	44	89	62	67	63	67
18	83	–243	72	–214	45	97	35	79	57	58	44	48
19	80	–239	76	–227	61	128	47	99	68	68	51	60
20	76	–221	90	–269	42	88	44	93	47	56	47	55
21	81	–238	95	–279	55	117	57	120	69	74	58	62
22	80	–238			52	110	53	107	65	72	57	57
23	76	–223			44	94	51	108	59	59	46	49
24	70	–207			41	82	48	100	57	60	54	63
25	79	–237			47	97	46	92	48	52	67	72
26	74	–216			43	91	43	89	41	47	60	64
27	77	–228			55	118	32	67	57	57	53	53
28	65	–193			43	87	48	97	50	57	57	65
29	80	–234			49	104	40	82	64	67	41	45
30	79	–229			42	89	40	81	43	45	58	61
31	78	–226			31	71	46	92	57	62	43	45
32	75	–218			40	86	41	90	54	56	55	57
33	82	–245			47	97	54	110	50	51	40	43
34	85	–247			43	89	48	96	59	66	67	74
35	68	–201			48	101	53	110	48	51	57	57
36	72	–215			44	93			45	54	48	54
37	71	–209			52	112			51	53	56	57

№	*Варіанти*											
	4				***5***				***6***			
	X	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***
1	67	207	71	214	31	318	30	309	89	–262	87	–261
2	57	171	59	185	28	280	25	257	90	–261	92	–273
3	82	246	64	196	30	305	31	316	91	–268	90	–270
4	62	191	68	208	23	234	25	251	87	–255	89	–266
5	57	172	71	220	25	255	27	278	89	–261	90	–264
6	83	255	76	231	25	258	26	268	93	–275	91	–270
7	69	213	77	234	25	258	27	273	90	–263	95	–285
8	86	261	73	225	27	272	28	280	94	–279	92	–268
9	66	203	65	198	31	317	24	244	92	–269	92	–268
10	79	242	70	217	25	252	24	243	93	–276	91	–272
11	82	248	74	225	28	280	26	269	94	–274	91	–269
12	73	221	73	224	25	258	25	255	92	–272	92	–273
13	65	195	77	237	30	304	22	229	97	–284	97	–283
14	55	172	77	236	28	289	27	272	93	–276	92	–269
15	61	186	49	154	28	288	26	263	94	–275	96	–284
16	64	197	63	189	31	317	24	249	92	–276	92	–270
17	59	182	71	222	23	231	25	252	85	–250	93	–271
18	75	225	69	214	28	288	22	229	95	–278	92	–271
19	68	207	52	160	26	269	27	272	92	–267	97	–289
20	65	201	67	201	26	267	29	292	99	–288	99	–294
21	80	241	70	215	28	284	24	245	89	–266	92	–271
22	71	219	64	197	27	277	27	274	94	–275	91	–271
23	61	183	66	206	25	253	25	256	98	–285	93	–279
24	77	239	78	241	26	268	29	291	89	–259	93	–277
25	69	209	65	197	27	270	29	290	88	–256	93	–270
26	64	201	63	196	28	269	25	257	96	–288	91	–264
27	68	208	69	216	27	278	25	258	92	–269	92	–269
28	52	159	59	180	29	295			92	–276	90	–268
29	59	183			24	246			91	–266	90	–262
30	63	197			23	238			85	–248	94	–280
31	56	176			27	279			94	–281	90	–263
32	62	192			27	273			89	–266	97	–289
33	67	204			28	284			94	–282	95	–278
34	62	195			26	265			94	–281	94	–279
35	71	220			29	298			90	–265	91	–265
36	67	205			26	264			96	–288	93	–273
37	66	207			23	232			97	–290	95	–283

Продовження таблиці 12.8

№	*Варіанти*											
	7				***8***				***9***			
	X	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***
1	45	–177	43	–168	46	279	39	234	77	619	77	631
2	38	–146	45	–174	40	245	49	298	76	626	67	542
3	58	–226	43	–172	58	354	44	264	62	505	73	587
4	46	–184	53	–204	35	212	49	302	83	680	67	542
5	40	–152	29	–110	53	323	53	323	83	673	83	681
6	41	–155	40	–152	47	285	44	272	79	637	67	544
7	58	–230	41	–161	40	240	47	284	77	629	76	616
8	41	–155	44	–167	60	361	51	315	81	649	71	579
9	51	–199	28	–109	39	235	56	340	71	568	74	604
10	59	–227	48	–186	41	246	39	236	83	669	73	594
11	62	–244	53	–203	58	357	46	285	73	591	75	601
12	44	–172	26	–99	59	361	68	415	81	648	80	659
13	47	–184	43	–168	57	343	47	282	82	663	72	583
14	43	–170	42	–159	65	390	40	242	77	616	66	529
15	47	–183	68	–266	34	208	35	210	82	672	80	647
16	55	–218	42	–159	50	301	60	360	83	677	72	585
17	61	–240	60	–237	45	277	37	227	80	655	68	561
18	47	–180	56	–222	50	306	41	248	77	619	76	619
19	58	–232	39	–147	55	334	46	284	69	568	81	667
20	40	–158	47	–188	57	342	46	278	78	624	69	567
21	47	–183	55	–211	45	274	53	320	74	605	83	667
22	46	–184	35	–138	55	336	43	260	77	632	77	634
23	42	–162	41	–157	46	277	53	321	69	554	69	565
24	50	–198	47	–183	53	320	54	328	72	594	81	664
25	45	–171	56	–217	45	278	50	304	76	612	65	528
26	38	–145	42	–167	43	266	45	278	69	556	71	579
27	34	–135	53	–208	48	294	50	303	84	679	69	564
28	44	–172	41	–162	50	308	25	159	77	629	85	688
29	47	–181	40	–151	44	271	44	266	72	582	81	656
30	65	–259	30	–111	38	228	54	331	76	611	72	578
31	34	–133	38	–143	68	414	49	301	78	626	84	675
32	42	–161	51	–201	54	325	56	338	87	703	69	557
33	48	–189	45	–180	57	348	48	297	79	650	77	619
34	54	–208	43	–167	34	211	44	268	80	655	67	549
35	55	–220	44	–176	59	356	49	294	70	577	69	568
36	50	–191	46	–176	57	350	50	309	80	657	66	530
37	37	–141	59	–234	57	350	46	284	75	613	71	568

№	Варіанти											
	10				11				12			
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
1	72	152	69	139	7	23	10	37	40	209	41	218
2	72	152	73	147	11	38	14	47	40	201	39	210
3	73	155	68	139	10	34	12	42	42	214	40	219
4	68	144	67	135	4	16	6	25	41	222	40	204
5	71	145	69	139	11	42	6	22	42	210	41	205
6	77	155	80	168	12	44	8	28	41	219	40	217
7	74	154	77	157	9	33	6	26	40	208	41	217
8	67	136	74	149	5	18	6	25	41	208	40	202
9	68	142	78	161	8	29	13	46	41	206	41	219
10	71	147	68	137	13	42	11	42	39	200	39	213
11	74	153	73	153	12	39	9	32	38	209	41	219
12	69	145	64	131	10	36	10	36	39	202	41	208
13	69	143	67	139	9	27	8	31	42	212	40	201
14	81	168	67	143	9	30	12	43	40	206	41	216
15	68	143	70	144	11	36	14	44	40	216	39	212
16	67	134	67	142	9	31	14	49	41	220	41	217
17	76	152	67	137	10	38	7	28	40	212	40	216
18	67	139	74	153	10	35	16	50	39	204	42	225
19	68	141	66	139	13	45	11	34	41	224	42	219
20	69	146	65	136	15	48	11	40	42	223	40	215
21	70	146	75	159	4	18	12	42	41	215	40	214
22	69	142	73	152	9	30	11	42	42	214	41	223
23	69	140	75	152	14	50	9	35	40	210	40	209
24	70	145	80	164	6	19	12	45	41	207	41	220
25	73	155	71	151	5	18	10	32	42	215	41	205
26	68	140	75	158	10	31	7	30	40	203	40	212
27	70	144	79	164	10	31	8	32	41	207		
28	69	145	76	153	13	43	9	30	40	212		
29	71	150	73	152	15	49	9	34	42	226		
30	71	142	65	137	6	26	11	36	40	204		
31	71	142	70	141	11	42	13	42	42	221		
32	65	139	66	141	10	31	9	27	41	220		
33	67	137	68	143	14	45	9	29	41	224		
34	68	144	74	156	8	28	10	38	40	213		
35	66	139	74	151	14	49	13	41	41	216		
36	71	144	71	149	6	23	7	21	40	213		
37	72	148	70	140	16	57	9	32	41	209		

Продовження таблиці 12.8

№	*Варіанти*											
	13				***14***				***15***			
	X	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***
1	48	200	66	265	59	420	62	436	76	304	88	361
2	46	188	58	236	71	509	65	457	85	340	79	317
3	62	249	63	261	61	435	66	475	77	315	81	325
4	63	261	58	234	67	469	66	474	74	303	81	332
5	51	213	70	282	62	449	62	452	83	341	83	339
6	61	253	59	239	62	450	60	435	73	297	90	368
7	62	252	62	255	61	437	64	465	77	316	73	300
8	56	230	51	204	59	422	60	431	78	318	80	323
9	50	204	61	245	65	463	60	432	85	348	75	303
10	54	225	71	286	63	455	63	446	82	329	77	313
11	53	220	59	242	65	472	62	444	81	325	80	320
12	47	192	65	262	62	448	62	437	85	349	85	341
13	66	273	65	267	62	443	64	458	79	320	83	338
14	58	236	61	244	65	462	64	453	83	337	90	362
15	52	216	64	260	67	484	62	443	81	330	84	343
16	65	261	60	246	63	442	56	410	87	357	78	315
17	65	265	54	224	58	419	65	473	84	337	75	303
18	49	196	67	277	64	456	66	464	80	321	80	325
19	69	276	57	237	64	451	64	458	81	328	85	344
20	55	229			61	435	61	431	81	324	77	312
21	47	192			63	457	62	434	74	303	85	340
22	61	246			59	422	69	486	89	364	77	313
23	53	212			64	454	65	465	78	318	93	373
24	58	237			63	458	62	439	87	353	78	313
25	68	273			62	449	66	466	83	332	74	298
26	48	192			68	486	63	453	82	331	78	312
27	52	214			65	468	57	408	83	332	86	351
28	52	212			68	478	55	396	84	341	75	307
29	65	269			65	463	67	487	81	324	79	317
30	70	281			62	441	63	460	91	364	93	372
31	58	232			68	491	65	462	85	343	84	339
32	70	287			64	450	64	456	80	320	82	337
33	63	258			60	432	68	485	86	345	85	343
34	60	249			64	453	70	490	84	345	80	328
35	77	311			67	478	57	409	88	355	75	305
36	63	261			68	481	65	472	87	356	83	341
37	49	199			62	438	69	502	82	333	75	309

Продовження таблиці 12.8

№	Варіанти											
	16				17				18			
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
1	25	233	43	392	13	71	11	49	47	482	51	521
2	47	432	27	245	11	62	16	76	49	504	51	527
3	42	385	49	445	18	91	17	70	50	513	53	540
4	38	347	35	323	13	60	15	77	50	500	49	506
5	30	274	36	331	15	79	14	72	52	526	50	515
6	30	276	15	140	21	85	13	57	49	493	51	511
7	32	294	33	303	17	79	17	70	49	505	59	598
8	38	348	27	247	13	54	15	64	54	552	52	522
9	34	315	43	387	16	71	14	61	50	503	51	515
10	34	307	47	426	13	67	12	67	52	529	52	521
11	41	377	33	304	19	90	12	51	54	546	51	527
12	37	338	28	257	17	81	20	96	58	590	52	524
13	39	353	41	377	15	61	15	64	55	550	51	526
14	36	332	59	535	12	54	16	74	49	491	52	531
15	34	307	23	208	17	77	16	69	55	567	47	486
16	29	264	39	357	20	82	10	40	50	512	54	558
17	31	285	37	337	17	72	15	61	51	512	50	510
18	41	377	49	447	18	91	23	102	56	564	55	556
19	15	137	26	238	17	81	15	63	50	515	50	506
20	27	245	20	182	17	87	17	70	53	538	49	494
21	46	421	48	437	16	70	17	83	49	492	53	541
22	34	313	32	292	17	80	17	76	52	529	52	521
23	39	359	52	473	16	67	19	86	47	479	55	561
24	49	443	43	392	17	68	16	73	53	532	51	519
25	22	201	49	442	13	67	11	52	57	580	54	556
26	48	440	39	356	18	86	13	55	48	491	50	503
27	30	274	42	386	14	57	17	87	51	516	56	563
28	29	261	37	338	13	67	20	88	46	465	54	547
29	29	270	30	277	11	48	19	85	57	578	50	516
30	38	345	27	247	15	60	15	71	50	507	48	496
31	37	334	18	170	10	51	13	60	53	539	54	548
32	55	498	42	380	15	63	16	78	51	510	59	606
33	38	342	24	221	14	67	15	65	55	566	52	534
34	27	250	31	287	9	54	22	95	53	533	57	586
35	44	403	27	244	19	85	14	69	51	528	54	554
36	48	434	34	313	18	87	15	77	54	549	52	537
37	35	320	24	219	15	61	12	49	53	537	48	492

№	*Варіанти*											
	19				***20***				***21***			
	X	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***
1	42	299	48	337	28	298	31	327	17	117	15	90
2	52	373	48	340	29	291	33	343	16	103	16	110
3	46	325	34	247	30	301	27	270	17	108	17	121
4	39	277	53	374	28	283	30	304	16	102	16	96
5	54	383	41	296	32	337	29	295	15	91	18	112
6	41	291	50	354	31	311	28	293	17	110	15	98
7	47	331	21	148	29	306	30	307	17	111	16	111
8	34	240	51	359	30	307	32	332	16	111	17	120
9	62	435	40	284	28	293	30	302	17	109	19	128
10	46	327	46	323	29	291	30	309	16	96	15	99
11	35	251	50	355	29	307	29	303	16	96	16	98
12	43	310	48	337	30	306	30	303	18	108	15	95
13	21	152	28	198	29	298	27	283	15	97	18	127
14	44	310	55	386	30	310	32	324	16	108	15	96
15	28	198	60	420	29	292	30	303	16	102	16	106
16	52	365	48	342	30	311	29	297	17	109	15	108
17	59	416	45	320	28	298	31	310	15	97	16	100
18	59	413	47	329	30	315	29	309	18	112	17	110
19	50	353	40	283	29	303	30	303	16	98	14	86
20	57	404	58	414	31	329	29	299	18	114	17	110
21	58	409	57	400	29	309	30	303	18	113	15	101
22	41	288	35	252	30	306	29	293	15	99	18	110
23	41	288	34	238	29	292	31	328	18	121	17	112
24	43	310	43	301	29	297	31	310	17	103	16	99
25	41	295	33	239	29	308	30	315	18	120	18	108
26	41	289	33	233	30	310	30	306	15	95		
27	36	254	45	316	30	308	30	301	17	116		
28	39	273	41	295	32	326	32	338	17	116		
29	32	224	30	216	30	305	31	318	16	112		
30	45	320	43	302	30	316	30	216	18	117		
31	46	331	45	222	29	308	30	306	16	104		
32	36	260	57	399	31	328	30	318	18	121		
33	40	281	35	254	30	300	30	304	16	107		
34	52	373	48	345	29	299	31	320	16	113		
35	36	252	58	408	29	290	30	306	16	111		
36	56	401	34	239	29	293	29	294	17	120		
37	41	288	39	279	30	303	30	305	18	127		

Продовження таблиці 12.8

№	Варіанти											
	22				23				24			
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
1	37	159	36	154	58	122	65	147	70	702	72	434
2	39	174	31	133	62	138	61	128	72	739	72	725
3	30	133	29	125	65	137	62	138	75	760	70	717
4	32	137	41	182	61	133	62	137	68	699	76	779
5	40	173	27	109	66	132	66	141	68	688	80	800
6	34	145	27	124	67	143	73	160	71	713	65	662
7	33	150	37	154	67	137	70	151	69	700	70	710
8	40	162	38	157	65	131	64	131	71	719	70	718
9	38	157	47	198	62	142	69	141	69	706	70	702
10	35	140	40	176	68	138	70	157	68	696	80	804
11	37	157	33	149	64	143	59	121	68	696	65	655
12	35	149	36	154	55	114	65	145	69	703	72	739
13	26	114	36	150	68	136	65	149	75	755	70	707
14	29	118	41	179	64	130	64	137	83	838	75	752
15	37	166	41	172	63	143	64	130	73	748	69	702
16	31	134	35	150	64	137	63	144	71	715	73	747
17	38	154	39	169	67	138	68	148	82	831	73	744
18	38	159	33	133	69	140	67	145	69	709	72	729
19	33	151	40	169	70	153	63	131	73	734	78	785
20	36	144	35	147	68	152	63	143	73	736	76	777
21	37	158	24	115	69	153	68	145	72	736	71	711
22	48	202	37	160	68	149	77	158	70	715	78	780
23	41	182	42	168	70	158	62	138	73	730	77	784
24	31	127	38	153	63	126	59	123	71	712	74	755
25	44	191	35	159	63	133	67	152	71	716	67	684
26	38	159	38	162	68	148	63	139	69	700	70	706
27	28	126	32	146	64	144	72	151	71	722	75	756
28	42	181	38	154	70	154	61	137	66	662	67	678
29	31	132	37	167	63	140	58	131	74	752	79	808
30	26	116	41	182	67	153	63	128	76	765	75	756
31	32	140	33	139	65	148	65	148	78	798	67	683
32	31	143	32	128	71	150	70	151	74	746	74	741
33	40	163	42	178	64	145	58	130	79	797	69	696
34	31	126	28	129	62	124	62	138	78	787	70	705
35	34	139	44	178	64	135	66	140	72	737	71	719
36	45	185	42	180	72	160	64	137	70	708	74	751
37	36	161	35	142	66	135	64	137	75	753	68	682

№	***Варіанти***											
	25				***26***				***27***			
	X	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***	***X***	***Y***
1	32	177	42	218	89	542	90	550	93	298	83	268
2	40	201	27	145	89	549	89	550	88	281	88	266
3	36	189	41	206	89	550	89	537	91	283	88	282
4	40	213	40	204	90	556	88	528	83	254	89	273
5	36	180	32	172	89	550	90	553	89	271	92	284
6	34	185	40	219	90	540	90	551	86	260	88	267
7	33	177	36	197	91	553	90	553	85	262	90	274
8	35	175	32	173	88	544	88	535	88	275	91	288
9	36	189	40	211	89	541	89	545	85	265	84	264
10	43	229	33	177	90	556	88	530	87	277	82	261
11	29	153	36	187	89	541	90	542	82	256	83	263
12	41	208	32	179	90	552	89	547	91	280	86	276
13	44	221	43	224	89	548	90	546	86	263	84	262
14	36	186	38	205	89	534	92	568	86	264	86	258
15	37	185	40	205	89	544	90	555	87	279	92	295
16	40	212	33	181	91	564	90	544	83	251	92	280
17	39	208	29	147	88	535	86	519	87	273	80	243
18	28	142	38	202	90	559			91	284	87	261
19	39	196	39	210	91	564			87	266	88	271
20	28	143	35	181	90	555			87	265	85	270
21	32	162	34	187	88	545			91	280		
22	40	200			89	535			86	274		
23	37	191			89	543			93	286		
24	30	163			90	554			89	278		
25	32	177			89	548			85	269		
26	34	187			91	553			87	268		
27	40	208			89	536			87	269		
28	40	213			90	553			88	264		
29	36	196			88	536			84	263		
30	36	199			90	549			90	279		
31	34	180			90	542			89	270		
32	34	180			91	546			81	261		
33	33	177			88	545			87	276		
34	31	159			90	556			88	266		
35	33	182			88	546			89	278		
36	42	223			87	527			89	268		
37	26	130			92	559			82	265		

№	Варіанти											
	28				29				30			
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
1	39	242	32	206	16	96	12	66	90	192	79	161
2	36	235	31	205	16	86	15	84	83	171	91	193
3	35	228	27	174	16	96	14	71	81	162	85	186
4	20	136	41	252	13	68	16	86	83	175	90	187
5	31	204	44	274	16	92	14	85	86	178	89	185
6	41	249	22	142	10	54	13	78	87	185	91	193
7	24	160	33	200	16	81	13	75	90	185	80	169
8	22	140	38	236	15	89	17	97	85	189	86	190
9	38	229	30	184	14	78	12	67	99	214	90	199
10	35	213	30	186	13	71	15	83	91	192	91	192
11	26	172	31	205	12	77	13	74	92	189	90	193
12	34	207	28	178	13	79	15	93	98	201	92	191
13	16	98	27	166	15	81	19	109	87	178	84	174
14	35	225	22	143	14	81	17	99	89	179	77	156
15	22	141	38	242	13	66	15	76	108	225	88	190
16	34	215	28	181	14	85	15	77	82	180	89	183
17	31	198	30	199	17	92	14	70	79	159	92	189
18	34	209	36	225	13	68	16	92	93	200	81	169
19	36	225	32	201	18	104	15	79	91	195	92	201
20	19	132	26	159	15	83	17	101	98	203	93	196
21	34	218	26	163	13	81	15	93	90	182	108	229
22	28	174	29	181	17	102	13	74	92	186	78	169
23	30	197			15	92	14	86	79	158	98	201
24	24	162			13	75	16	86	82	176	99	210
25	17	104			16	80	16	80	103	211	90	191
26	27	175			15	92	15	80	93	188	93	193
27	29	176			17	86	15	88	89	194	92	195
28	33	203			15	77	13	74	93	188	76	169
29	32	210			16	99	17	86	91	186	90	193
30	30	188			17	94	14	80	83	174	90	190
31	27	165			17	104	13	79	104	211	87	177
32	31	196			16	86	15	92	95	199	85	186
33	28	183			12	74	14	77	97	204	88	185
34	30	196			13	75	14	81	100	204	103	211
35	38	229			13	84	16	87	92	187	91	190
36	34	205			14	88	11	71	92	184	94	200
37	35	221			12	67	13	81	84	175	93	205

Розв'язок варіанта ІЗ

Скласти кореляційну таблицю для двовимірної вибірки XY об'єму n = 78, наведеної в табл. 12.9, знайти рівняння прямих регресії Y по X і X по Y, побудувати кореляційне поле і графіки прямих регресії.

Таблиця 12.9

X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
73	–291	63	–243	74	–290	70	–277	79	–309	63	–248
69	–270	80	–313	68	–266	61	–241	77	–300	63	–243
72	–279	71	–278	69	–270	62	–243	78	–310	67	–264
72	–282	74	–292	71	–283	63	–245	66	–255	68	–267
65	–254	68	–271	60	–237	71	–282	63	–252	55	–213
67	–264	65	–256	56	–222	65	–252	69	–274	56	–218
56	–216	73	–291	71	–281	70	–276	74	–291	58	–223
70	–276	57	–219	68	–269	70	–276	68	–264	70	–278
63	–248	71	–281	66	–257	63	–246	62	–240	59	–236
64	–253	66	–262	60	–235	73	–284	70	–277	68	–263
70	–276	76	–302	70	–275	68	–271	70	–279	69	–268
67	–262	70	–275	69	–276	59	–227	65	–253	63	–243
60	–234	68	–267	72	–282	64	–256	70	–275	70	–271

○ Знайдемо мінімальні та максимальні значення випадкових величин X і Y:

$$x_{\min} = 55; \qquad y_{\min} = -313;$$
$$x_{\max} = 80; \qquad y_{\max} = -213.$$

Використовуючи формулу Стерджеса (8.5), знайдемо величини часткових інтервалів h_1 і h_2, за допомогою яких розбиваються розмахи варіювання спостережуваних значень випадкових величин відповідно X і Y.

Маємо

$$h_1 = \frac{25}{1 + 3{,}322 \lg 78} = \frac{25}{7{,}2856} \approx 3{,}43,$$

$$h_2 = \frac{100}{7{,}2856} \approx 13{,}73.$$

Для зручності подальших обчислень візьмемо $h_1 = 4$, $h_2 = 14$. Відкладаємо інтервали зміни x_i по вертикалі, а y_i – по горизонталі.

Складемо кореляційну таблицю 12.10, уважаючи, що початок інтервалу входить до інтервалу, а кінець – не входить.

Таблиця 12.10

X \ Y	-320...-306	-306...-292	-392...-278	-278...-264	-264...-250	-250...-236	-236...-222	-222...-208	n_{x_i}
53–57								4	4
57–61						1	5	1	7
61–65					3	10			13
65–69				6	12				18
69–73			8	17					25
73–77		1	6						7
77–81	3	1							4
n_{y_j}	3	2	14	23	15	11	5	5	78

Подамо кореляційну табл. 12.10 у вигляді табл. 12.11, у якій, замість інтервалів значень, запишемо середини відповідних інтервалів, а також уведемо умовні варіанти

$$u_i = (x_i - C_1)/h_1 = (x_i - 67)/4,$$

$$v_j = (y_j - C_2)/h_2 = (y_j + 271)/14,$$

де $C_1 = 67$ і $C_2 = -271$ – помилкові нулі, $h_1 = 4$, $h_2 = 14$.

Використовуючи формули (12.19) – (12.23), одержимо:

$$\overline{x} = 4 \cdot 12/78 + 67 \approx 67,6154;$$

$$\overline{y} = 14 \cdot 45/78 - 271 \approx 262,9231;$$

$$s_x^2 = 16 \cdot 166/78 - (67,6154 - 67)^2 \approx 33,6726; \qquad s_x \approx 5,8028;$$

$$s_y^2 = 196 \cdot 233/78 - (-262,9231 + 271)^2 \approx 520,2509; \qquad s_y \approx 22,809;$$

$$s_{xy} = 4 \cdot 14 \cdot (-170)/78 - 4 \cdot 14 \cdot 12 \cdot 45/6084 \approx -127,0217;$$

Таблиця 12.11

X \ Y		–313	–299	–285	–271	–257	–243	–229	–215	n_{x_i}	$u_i n_{x_i}$	$u_i^2 n_{x_i}$	$\sum_{j=1}^{8} u_i v_j n_{ij}$
X	u_i \ v_j	–3	–2	–1	0	1	2	3	4				
55	–3	–	–	–	–	–	–	–	4 $\boxed{-12}$	4	–12	36	–48
59	–2	–	–	–	–	–	1 $\boxed{-4}$	5 $\boxed{-6}$	1 $\boxed{-8}$	7	–14	28	–42
63	–1	–	–	–	–	3 $\boxed{-1}$	10 $\boxed{-2}$	–	–	13	–13	13	–23
67	0	–	–	–	6 $\boxed{0}$	12 $\boxed{0}$	–	–	–	18	0	0	0
71	1	–	–	8 $\boxed{-1}$	17 $\boxed{0}$	–	–	–	–	25	25	25	–8
75	2	–	1 $\boxed{-4}$	6 $\boxed{-2}$	–	–	–	–	–	7	14	28	–16
79	3	3 $\boxed{-9}$	1 $\boxed{-6}$	–	–	–	–	–	–	4	12	36	–33
n_{y_j}		3	2	14	23	15	11	5	5	78	12	166	–
$v_j n_{y_j}$		–9	–4	–14	0	15	22	15	20	45	–	–	–
$v_j^2 n_{y_j}$		27	8	14	0	15	44	45	80	233	–	–	–
$\sum_{i=1}^{7} u_i v_j n_{ij}$		–27	–10	–20	0	–3	–24	–30	–56	–	–	–	–170

За формулою (12.14) знайдемо вибірковий коефіцієнт кореляції

$$\rho_{xy} = -127,0217/(5,8028 \cdot 22,809) \approx -0,9597,$$

а потім, використовуючи (12.16) і (12.17), одержимо шукані рівняння регресій:

$$\overline{y}_x + 262,9231 = -0,9597\frac{22,809}{5,8028}(x - 67,6154);$$

$$\overline{y}_x = -3,7723x - 7,8588;$$

$$\overline{x}_y - 67,6154 = -0,9597\frac{5,8028}{22,809}(y + 262,9231);$$

$$\overline{x}_y = -0,2442y + 3,4212.$$

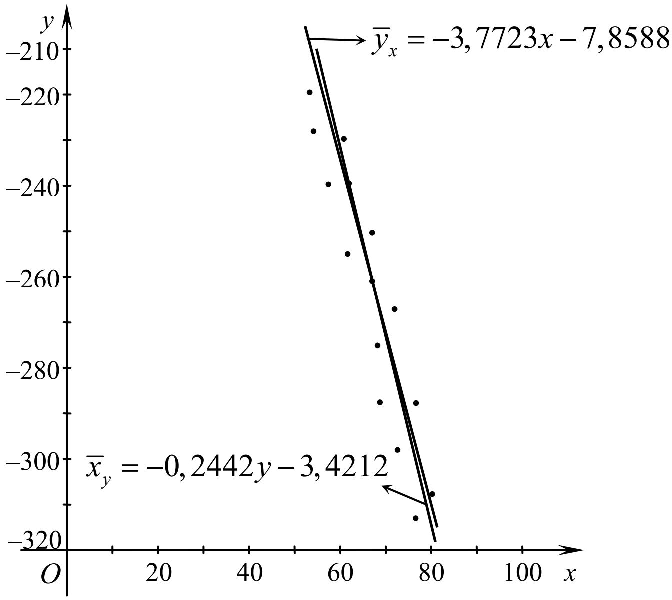

Рис. 12.4

Кореляційне поле і графіки прямих регресій зображено на рис. 12.4. ●