

Державний комітет зв'язку та інформатизації України  
Одеська Національна Академія зв'язку ім. О. С. Попова  
Кафедра вищої математики

## **МАТЕМАТИЧНА СТАТИСТИКА**

Методичні вказівки та контрольні завдання для студентів, що  
навчаються за прискореною формою навчання  
(спеціальності економіка і менеджмент)

Затверджено  
радою факультету  
заочного навчання  
Протокол №7  
від 10 квітня 2001

Одеса-2001

УДК – 519.21

План НМВ 2001 рік

Укладач **І.Г. Кальмушевский**Рецензент **В.М. Орлов**Комп'ютерне верстання  
та макетування **Є.С. Корнійчук**

Приведені основні положення математичної статистики і методи аналізу статистичних даних: вибірковий метод, метод довірчих інтервалів, кореляційні методи. По кожній темі надано завдання, що залежить від номеру варіанта студента.

Схвалено  
на засіданні кафедри  
вищої математики  
й рекомендовано до друку.  
Протокол №6  
від 17 березня 2001

Математична статистика – розділ математики, присвячений математичним методам систематизації, опрацювання й використання статистичних даних для наукових та практичних висновків.

Перша ціль математичної статистики – вказати способи збору і групування статистичних відомостей, одержаних в результаті спостережень чи в результаті спеціально поставлених експериментів.

Друге завдання математичної статистики – розробити методи аналізу статистичних даних в залежності від мети дослідження.

## 1 Вибірковий метод

Вибірковою сукупністю, або просто вибіркою називають сукупність випадково відібраних об'єктів. Генеральною сукупністю називають сукупність об'єктів, з яких проводиться вибірка. Об'ємом сукупності (вибіркової або генеральної) називають кількість об'єктів цієї сукупності.

Нехай із генеральної сукупності добута вибірка, причому  $x_1$  спостерігалось  $n_1$  разів,  $x_2$  –  $n_2$  разів, ...,  $x_k$  –  $n_k$  разів і  $\sum n_i = n$  – об'єм вибірки. Спостережувані значення  $x_i$  називають варіантами, а послідовність варіант, записаних в зростаючому порядку, – варіаційним рядом.

Кількості спостережень називають частотами, а їх відношення до об'єму вибірки  $n_i/n = w_i$  – відносними частотами.

Статистичним розподілом вибірки називають перелік варіант і відповідних їм частот або відносних частот. Статистичний розподіл можна задавати також у вигляді послідовності інтервалів і відповідних їм частот.

Емпіричною функцією розподілу називають функцію  $F^*(x)$ , яка визначає для кожного значення  $x$  відносну частоту події  $X < x$ , де  $X$  – деяка кількісна ознака досліджуваного явища. Таким чином  $F^*(x) = n_x/n$ , де  $n_x$  – число варіант, менших  $x$ ;  $n$  – об'єм вибірки. На відміну від емпіричної функції розподілу вибірки, функцію розподілу  $F(x)$  генеральної сукупності називають теоретичною функцією розподілу.

Полігоном частот називають ламану, відрізки якої з'єднують точки  $(x_1; n_1)$ ,  $(x_2; n_2)$ , ...,  $(x_k; n_k)$ . Полігоном відносних частот називають ламану, відрізки якої з'єднують точки  $(x_1, w_1)$ ,  $(x_2, w_2)$ , ...,  $(x_k; w_k)$ .

У випадку неперервної ознаки доцільно будувати гістограму, для чого інтервал, в якому містяться всі спостережені значення ознаки, розбивають на декілька часткових інтервалів довжиною  $h$  і знаходять для кожного часткового інтервалу  $n_i$  – суму частот варіант, що попали в  $i$ -ий інтервал.

Гістограмою частот називають східчасту фігуру, що складається із прямокутників, основами яких слугують часткові інтервали довжиною  $h$ , а висоти дорівнюють відношенню  $n_i/h$  (густина частот). Площа гістограми частот дорівнює об'єму вибірки.

Гістограмою відносних частот називають східчасту фігуру, що складається із прямокутників, основами яких слугують часткові

інтервали довжиною  $h$ , а висоти дорівнюють відношенню  $w_i / h$  (густина відносної частоти).

### Завдання 1

Задано розподіл частот вибірки об'ємом  $n = 100$

|       |          |          |    |          |          |
|-------|----------|----------|----|----------|----------|
| $x_i$ | 1        | 3        | 5  | 7        | 9        |
| $n_i$ | $10 + N$ | $15 - N$ | 30 | $33 - N$ | $12 + N$ |

Тут, як і в подальшому,  $N$  – номер варіанту, який визначається останньою цифрою номеру залікової книжки студента (у разі нуля взяти  $N = 10$ ).

Слід:

- записати розподіл відносних частот;
- записати емпіричну функцію розподілу і побудувати її графік;
- побудувати полігони частот і відносних частот розподілу.

### Завдання 2

Побудувати гістограми частот і відносних частот розподілу (в верхньому рядку вказані часткові інтервали, в нижньому – сума частот варіант часткового інтервалу).

|         |         |          |          |          |         |         |
|---------|---------|----------|----------|----------|---------|---------|
| 5 – 10  | 10 – 15 | 15 – 20  | 20 – 25  | 25 – 30  | 30 – 35 | 35 – 40 |
| $4 + N$ | $6 + N$ | $16 - N$ | $36 - N$ | $24 - N$ | 10      | $4 + N$ |

## 2 Статистичні оцінки параметрів розподілу

Нехай потрібно вивчити кількісну ознаку генеральної сукупності. Припустимо, що з теоретичних міркувань мали можливість встановити, який розподіл має ознака. Тоді виникає питання оцінювання параметрів, якими визначається цей розподіл. Наприклад, якщо наперед відомо, що ознака, яка вивчається, розподілена в генеральній сукупності нормально, то необхідно оцінити (приблизно знайти) математичне сподівання і середнє квадратичне відхилення, так як саме ці два параметри повністю визначають нормальний розподіл.

При дослідженні, як правило, відомі лише дані вибірки, наприклад, значення кількісної ознаки  $x_1, x_2, \dots, x_n$ , одержані в результаті  $n$  спостережень. Через ці дані і потрібно виразити оцінювані параметри. Для того, щоб статистичні оцінки давали “хороші” наближення таких параметрів, вони повинні задовольнювати певним вимогам.

Нехай  $\theta^*$  – статистична оцінка невідомого параметра  $\theta$  теоретичного розподілу. Незсуненою називають статистичну оцінку  $\theta^*$ , математичне сподівання якої дорівнює параметру  $\theta$  за будь-якого об'єму вибірки, тобто  $M(\theta^*) = \theta$ . Ефективною називають статистичну оцінку, яка (при заданому об'єму вибірки  $n$ ) має найменшу можливу дисперсію. Слушною називають

статистичну оцінку, яка при  $n \rightarrow \infty$  прямує за ймовірністю до оцінюваного параметра.

## 2.1 Оцінка генеральної середньої та генеральної дисперсії

Генеральною середньою  $\bar{x}_r$  називають середнє арифметичне значення ознаки генеральної сукупності. Якщо всі значення  $x_1, x_2, \dots, x_N$  ознаки генеральної сукупності різні, то

$$\bar{x}_r = (x_1 + x_2 + \dots + x_N)/N.$$

Якщо ж значення ознаки  $x_1, x_2, \dots, x_k$  мають відповідно частоти  $N_1, N_2, \dots, N_k$ , причому  $N_1 + N_2 + \dots + N_k = N$ , то

$$\bar{x}_r = (x_1 N_1 + x_2 N_2 + \dots + x_k N_k)/N.$$

Вибірковою середньою  $\bar{x}_b$  називають середнє арифметичне значення ознаки вибіркової сукупності. Якщо всі значення ознаки вибірки об'єму  $n$  різні, то

$$\bar{x}_b = (x_1 + x_2 + \dots + x_n)/n.$$

Якщо ж всі значення  $x_1, x_2, \dots, x_k$  мають відповідно частоти  $n_1, n_2, \dots, n_k$ , причому  $n_1 + n_2 + \dots + n_k = n$ , то

$$\bar{x}_b = (x_1 n_1 + x_2 n_2 + \dots + x_k n_k)/n = \left( \sum_{i=1}^k x_i n_i \right) / n.$$

За оцінку генеральної середньої обирають вибіркoву середню. При цьому  $\bar{x}_b$  – незсунена оцінка.

Для того, щоб охарактеризувати розсіяння спостережуваних значень кількісної ознаки вибірки біля свого середнього значення  $\bar{x}_b$ , вводять зведену характеристику – вибіркoву дисперсію. Вибірковою дисперсією  $D_b$  називають середнє арифметичне квадратів відхилення спостережуваних значень ознаки від їх середнього значення  $\bar{x}_b$ .

Якщо всі значення  $x_1, x_2, \dots, x_n$ , ознаки вибірки об'єму  $n$  різні, то

$$D_b = \left( \sum_{i=1}^n (x_i - \bar{x}_b)^2 \right) / n.$$

Якщо ж значення ознаки  $x_1, x_2, \dots, x_k$  мають відповідні частоти  $n_1, n_2, \dots, n_k$ , причому  $n_1 + n_2 + \dots + n_k = n$ , то

$$D_b = \left( \sum_{i=1}^k n_i (x_i - \bar{x}_b)^2 \right) / n.$$

Вибірковим середнім квадратичним відхиленням називають квадратний корінь із вибіркової дисперсії  $\sigma_b = \sqrt{D_b}$ .

Подібним чином вводиться поняття генеральної дисперсії та генерального середнього квадратичного відхилення.

Обчислення дисперсії, байдуже – вибіркової чи генеральної, можна спростити, використовуючи співвідношення:

$$D = \overline{x^2} - [\bar{x}]^2,$$

тобто дисперсія дорівнює середньому квадратів значень ознаки мінус квадрат загальної середньої, де

$$\overline{x^2} = (\sum n_i x_i^2) / n, \quad \bar{x} = (\sum n_i x_i) / n.$$

Якщо за оцінку генеральної дисперсії взяти вибіркиму дисперсію, то ця оцінка спричинятиме до систематичних помилок. Пояснюється це тим, що вибіркима дисперсія є зсуненою оцінкою  $D_B$ , при цьому

$$M[D_B] = \frac{n-1}{n} D_G.$$

Тому для оцінювання генеральної дисперсії використовують виправлену дисперсію, яку позначають через  $S^2$  і визначають рівністю

$$S^2 = \frac{n}{n-1} D_B = \frac{\sum_{i=1}^k n_i (x_i - \bar{x}_B)^2}{n-1}.$$

Виправлена дисперсія є, звичайно, незсуненою оцінкою генеральної дисперсії.

Виправлене середнє квадратичне відхилення дорівнює кореневі квадратному із виправленої дисперсії й не є незсуненою оцінкою.

Відмітимо, що за великих значень  $n$  об'єму вибірки вибіркима і виправлена дисперсії відрізняються мало. Практично виправленою дисперсією користуються при  $n < 30$ .

### Завдання 3

*Нижче наведено результати вимірювання зросту (у сантиметрах) випадково відібраних 100 студентів.*

| Зріст           | 154 -158 | 158-162 | 162-166 | 166-170 | 170-174 | 174-178 | 178-182 |
|-----------------|----------|---------|---------|---------|---------|---------|---------|
| Число студентів | 10 + N   | 14      | 26 - N  | 28 - N  | 12      | 8       | 2 + N   |

*Знайти вибіркиму середню, вибіркиму дисперсію, вибіркове середнє квадратичне відхилення, виправлену дисперсію.*

Вказівка. *Знайти середини інтервалів і прийняти їх за значення варіант.*

### 2.2 Довірча ймовірність (надійність). Довірчий інтервал

Точковою називають оцінку, яка визначається одним числом. Всі оцінки, розглянуті раніше, – точкові. Інтервальною називають оцінку, яка визначається двома числами – границями інтервалу.

Нехай знайдена за даними вибірки статистична характеристика  $\theta^*$  є оцінка невідомого параметра  $\theta$ . Додатне число  $\delta$  характеризує точність оцінки, якщо  $|\theta - \theta^*| < \delta$ . Статистичні методи не дозволяють категорично стверджувати, що оцінка  $\theta^*$  задовольняє цій нерівності; можна лише говорити про ймовірність  $\gamma$ , з якою ця нерівність здійснюється.

Надійністю (довірчою ймовірністю) оцінки  $\theta$  по  $\theta^*$  називають ймовірність  $\gamma$ , з якою виконується нерівність  $|\theta - \theta^*| < \delta$ .

Звичайно надійність оцінки задається наперед, причому вона близька до одиниці. Найчастіше задають надійність, що дорівнює 0,95; 0,99; 0,999. Рівність  $P(|\theta - \theta^*| < \delta) = \gamma$  слід розуміти так: ймовірність того, що інтервал  $(\theta^* - \delta, \theta^* + \delta)$  включає невідомий параметр  $\theta$  (покриває параметр  $\theta$ ), дорівнює  $\gamma$ . Інтервал  $(\theta^* - \delta, \theta^* + \delta)$  називається довірчим інтервалом.

### 2.3 Довірчі інтервали для оцінювання математичного сподівання нормального розподілу

Нехай кількісна ознака  $X$  генеральної сукупності розподілена нормально, причому середнє квадратичне відхилення  $\sigma$  цього розподілу відоме. Потрібно оцінити невідоме математичне сподівання  $a$  за вибірковою середньою  $\bar{x}$ . Знайдемо довірчі інтервали, що покривають параметр  $a$  з надійністю  $\gamma$ .

Розв'язок задачі задається формулою

$$P(\bar{x} - t\sigma / \sqrt{n} < a < \bar{x} + t\sigma / \sqrt{n}) = 2\Phi(t) = \gamma.$$

Число  $t$  визначають з рівності

$$2\Phi(t) = \gamma, \text{ або } \Phi(t) = \gamma / 2;$$

у таблиці значень функції Лапласа знаходять аргумент  $t$ , якому відповідає значення функції Лапласа, рівне  $\gamma / 2$ .

**Приклад 1** Випадкова величина  $X$  має нормальний розподіл з відомим середнім квадратичним відхиленням  $\sigma = 3$ . Знайти довірчий інтервал для оцінювання невідомого математичного сподівання  $a$  за вибірковою середньою  $\bar{x} = 4,1$ , якщо об'єм вибірки  $n = 36$  і задана надійність оцінки  $\gamma = 0,95$ .

Розв'язок. Знайдемо  $t$ . Із співвідношення  $2\Phi(t) = 0,95$  одержимо  $\Phi(t) = 0,475$ . У таблиці знаходимо  $t = 1,96$ . Визначаємо точність оцінки:

$$\delta = t\sigma / \sqrt{n} = (1,96 \cdot 3) / \sqrt{36} = 0,98.$$

Одже, довірчий інтервал такий:  $(\bar{x} - 0,98; \bar{x} + 0,98)$  або  $(3,12; 5,08)$ .

#### Завдання 4

Випадкова величина  $X$  має нормальний розподіл з відомим середнім квадратичним відхиленням  $\sigma = 3 + \frac{N+1}{50}$ .

Знайти довірчий інтервал для оцінювання невідомого математичного сподівання  $a$  за вибірковою середньою  $\bar{x} = 5,1 + N$ , якщо об'єм вибірки  $n = 36 \cdot (N + 1)^2$  і задана надійність оцінки  $\gamma = 0,95$ .

### 3 Елементи теорії кореляції

Дуже часто виникає потреба встановити і оцінити залежність визначуваної випадкової величини  $Y$  від іншої випадкової величини  $X$

(або декількох випадкових величин). Дві випадкові величини можуть бути пов'язані функціональною чи статистичною залежністю або бути незалежними. Жорстка функціональна залежність реалізується рідко. Частіше виникає статистична залежність.

Статистичною називають залежність, за якої зміна однієї з величин тягне за собою зміну розподілу іншої. Зокрема, статистична залежність проявляється в тому, що зі зміною однієї з величин змінюється середнє значення іншої. В цьому разі статистичну залежність називають кореляційною.

### 3.1 Вибірковий коефіцієнт кореляції

Вибірковий коефіцієнт кореляції визначається рівністю

$$r_B = \frac{\sum n_{xy}xy - n\bar{x} \cdot \bar{y}}{n\bar{\sigma}_x \cdot \bar{\sigma}_y},$$

де  $x, y$  – варіанти (спостережувані значення) ознак  $X$  і  $Y$ ,  $n_{xy}$  – частоти пари варіант  $(x, y)$ ,  $n$  – об'єм вибірки (сума всіх частот);  $\bar{\sigma}_x, \bar{\sigma}_y$  – вибіркові середні квадратичні відхилення;  $\bar{x}, \bar{y}$  – вибіркові середні.

Якщо величини  $X$  і  $Y$  незалежні, то коефіцієнт кореляції  $r = 0$ ; якщо  $r = \pm 1$ , то  $X$  і  $Y$  зв'язані лінійною функціональною залежністю. Таким чином, коефіцієнт кореляції  $r$  вимірює ступінь лінійного зв'язку між  $X$  і  $Y$ .

### 3.2 Визначення параметрів вибіркового рівняння прямої лінії регресії за згрупованими даними

Якщо вибірковий коефіцієнт кореляції відмінний від нуля, то кореляційна залежність між  $X$  і  $Y$  визначається рівнянням

$$\bar{y}_x - \bar{y} = r_B \frac{\bar{\sigma}_y}{\bar{\sigma}_x} (x - \bar{x}),$$

яке називається вибірковим рівнянням прямої лінії регресії  $Y$  на  $X$ , а її графік – вибірковою лінією регресії  $Y$  на  $X$ . Вибіркове рівняння прямої лінії регресії  $X$  на  $Y$  має вигляд

$$\bar{x}_y - \bar{x} = r_B \frac{\bar{\sigma}_x}{\bar{\sigma}_y} (y - \bar{y}).$$

Графік її є вибірковою лінією регресії  $X$  на  $Y$ .

**Приклад 2** Знайти вибіркове рівняння прямої лінії регресії  $Y$  на  $X$  за даними кореляційної таблиці 3.1.

Таблиця 3.1

| Y              | X  |    |    |    |    |    | n <sub>y</sub> |
|----------------|----|----|----|----|----|----|----------------|
|                | 10 | 20 | 30 | 40 | 50 | 60 |                |
| 15             | 5  | 7  | –  | –  | –  | –  | 12             |
| 25             | –  | 20 | 23 | –  | –  | –  | 43             |
| 35             | –  | –  | 30 | 47 | 2  | –  | 79             |
| 45             | –  | –  | 10 | 11 | 20 | 6  | 47             |
| 55             | –  | –  | –  | 9  | 7  | 3  | 19             |
| n <sub>x</sub> | 5  | 27 | 63 | 67 | 29 | 9  | n=200          |

Розв'язок. Перейдемо до умовних варіант  $u_i = (x_i - c_1)/h_1 = (x_i - 40)/10$  (за хибний нуль  $c_1$  взято варіанту  $x=40$ , що міститься приблизно в середині варіаційного ряду; крок  $h_1$  дорівнює різниці між двома сусідніми варіантами  $20-10=10$ ) і  $v_j = (y_j - c_2)/h_2 = (y_j - 35)/10$ .

Складемо кореляційну таблицю в умовних варіантах. Практично це можна зробити так: в першій колонці замість хибного нуля  $c_2$  (варіанти 35) пишуть 0; над нулем послідовно записують -1, -2; під нулем пишуть 1, 2. В першій лінії замість хибного нуля  $c_1$  (варіанти 40) пишуть 0; зліва від нуля пишуть -1, -2, -3; справа пишуть 1, 2. Всі інші дані переписують із початкової кореляційної таблиці. В результаті одержимо кореляційну таблицю 2 в умовних варіантах.

Таблиця 3.2

| V              | u  |    |    |    |    |   | n <sub>v</sub> |
|----------------|----|----|----|----|----|---|----------------|
|                | -3 | -2 | -1 | 0  | 1  | 2 |                |
| -2             | 5  | 7  | –  | –  | –  | – | 12             |
| -1             | –  | 20 | 23 | –  | –  | – | 43             |
| 0              | –  | –  | 30 | 47 | 2  | – | 79             |
| 1              | –  | –  | 10 | 11 | 20 | 6 | 47             |
| 2              | –  | –  | –  | 9  | 7  | 3 | 19             |
| n <sub>u</sub> | 5  | 27 | 63 | 67 | 29 | 9 | n=200          |

В умовних варіантах вибіркового коефіцієнт кореляції обчислюють за формулою  $r_B = \frac{\sum n_{uv}uv - n\bar{u} \cdot \bar{v}}{n\bar{\sigma}_u \cdot \bar{\sigma}_v}$ .

Для обчислення шуканої суми  $\sum n_{uv}uv$  складемо розрахункову таблицю 3.3. Складають її слідуючим чином:

Таблиця 3.3

| <b>v</b>            | <b>u</b>                |                          |                          |                       |                        |                      | $U = \sum n_{uv} u$ | $vU$                                             |
|---------------------|-------------------------|--------------------------|--------------------------|-----------------------|------------------------|----------------------|---------------------|--------------------------------------------------|
|                     | -3                      | -2                       | -1                       | 0                     | 1                      | 2                    |                     |                                                  |
| -2                  | 5 <sup>-15</sup><br>-10 | 7 <sup>-14</sup><br>14   | —                        | —                     | —                      | —                    | -29                 | 58                                               |
| -1                  | —                       | 20 <sup>-40</sup><br>-20 | 23 <sup>-23</sup><br>-23 | —                     | —                      | —                    | -63                 | 63                                               |
| 0                   | —                       | —                        | 30 <sup>-30</sup><br>0   | 47 <sup>0</sup><br>0  | 2 <sup>2</sup><br>0    | —                    | -28                 | 0                                                |
| 1                   | —                       | —                        | 20 <sup>-10</sup><br>10  | 20 <sup>0</sup><br>11 | 20 <sup>20</sup><br>20 | 6 <sup>12</sup><br>6 | 22                  | 22                                               |
| 2                   | —                       | —                        | —                        | 9 <sup>0</sup><br>18  | 7 <sup>7</sup><br>14   | 3 <sup>6</sup><br>6  | 13                  | 26                                               |
| $V = \sum n_{uv} v$ | -10                     | -34                      | -13                      | 29                    | 34                     | 12                   |                     | $\sum vU = 169$                                  |
| $uV$                | 30                      | 68                       | 13                       | 0                     | 34                     | 24                   | $\sum uV = 169$     | $\overset{\uparrow}{\text{контроль}} \leftarrow$ |

1. В кожній клітині, в якій частота  $n_{uv} \neq 0$ , записують в правому верхньому куті добуток частоти  $n_{uv}$  на варіанту  $u$ .

Наприклад, в першій лінії записано добутки:  $5(-3)=-15$ ;  $7(-2)=-14$ .

2. Додають всі числа, розміщені в правих кутах клітин однієї лінії і їх суму записують в клітину цієї ж лінії колонки  $U$ . Наприклад, для першої лінії  $U=-15+(-14)=-29$ .

3. Множать варіанту  $v$  на  $U$  і одержаний добуток записують в останній клітині розгляненої лінії, тобто в клітину колонки  $vU$ . Наприклад, в першій лінії таблиці  $v=-2$ ,  $U=-29$ ; значить,

$$vU=(-2) \times (-29)=58.$$

4. Нарешті, додавши всі числа колонки  $vU$ , одержимо суму  $\sum_v vU$ , що дорівнює шуканій сумі  $\sum n_{uv} uv$ . В нашому випадку одержуємо  $\sum_v vU = 169$ .

Для контролю аналогічні обчислення проводять по колонкам. Одержують суму  $\sum_u uV$ , яка також дорівнює  $\sum n_{uv} uv$ .

Знаходимо  $\bar{u}$  і  $\bar{v}$ :

$$\bar{u} = (\sum n_u u) / n = [5 \cdot (-3) + 27 \cdot (-2) + 63 \cdot (-1) + 29 \cdot 1 + 9 \cdot 2] / 200 = -0,425$$

$$\bar{v} = (\sum n_v v) / n = [12 \cdot (-2) + 43 \cdot (-1) + 47 \cdot 1 + 19 \cdot 2] / 200 = 0,19$$

Знаходимо допоміжну величину  $\bar{u}^2$ , а потім  $\sigma_u$ :

$$\overline{u^2} = (\sum n_u \cdot u^2) / n = (5 \cdot 9 + 27 \cdot 4 + 63 \cdot 1 + 29 + 1 + 9 \cdot 4) / 200 = 1,405$$

$$\overline{\sigma_u} = \sqrt{\overline{u^2} - (\overline{u})^2} = \sqrt{1,405 - (0,425)^2} = 1,106.$$

Аналогічно одержимо  $\overline{\sigma_v} = 1,209$ .

Знаходимо вибіркового коефіцієнт кореляції

$$r_B = \frac{\sum n_{uv} u \cdot v - n \cdot \overline{u} \cdot \overline{v}}{n \cdot \overline{\sigma_u} \cdot \overline{\sigma_v}} = \frac{169 - 200 \cdot (-0,425) \cdot 0,09}{200 \cdot 1,106 \cdot 1,209} = 0,603.$$

Для складення рівняння прямої лінії регресії знаходимо

$$\overline{x} = \overline{u} h_1 + c_1 = -0,425 \cdot 10 + 40 = 35,75$$

$$\overline{y} = \overline{v} h_2 + c_2 = 0,09 \cdot 10 + 35 = 35,9$$

$$\overline{\sigma_x} = \overline{\sigma_u} h_1 = 1,106 \cdot 10 = 11,06$$

$$\overline{\sigma_y} = \overline{\sigma_v} h_2 = 1,209 \cdot 10 = 12,09$$

Підставивши знайдені величини в рівняння лінії регресії, одержимо

$$\overline{y_x} - 35,9 = 0,603 \cdot \frac{12,09}{11,06} (x - 35,75)$$

або

$$\overline{y_x} = 0,659x + 12,34.$$

Порівняємо умовні середні, обчислені:

- а) за цим рівнянням;
- б) за даними кореляційної таблиці 1.

Наприклад, при  $x = 30$ , маємо:

$$\text{а) } \overline{y_{30}} = 0,659 \cdot 30 + 12,34 = 32,11$$

$$\text{б) } \overline{y_{30}} = \frac{(23 \cdot 25 + 30 \cdot 35 + 10 \cdot 45)}{63} = 32,94.$$

Як видно, узгодженість розрахункової і спостереженої середніх задовільна.

### Завдання 5

Задана кореляційна таблиця системи кількісних ознак (X, Y).

Знайти:

- 1) вибіркове рівняння прямої лінії регресії Y на X;
- 2) зрівняти для  $x = 30$  умовні середні, обчислені:
  - а) по цьому рівнянню;
  - б) за даними кореляційної таблиці;
- 3) вибіркове рівняння прямої лінії регресії X на Y;
- 4) зрівняти для  $y = 35 + N$  умовні середні, обчислені:
  - а) по цьому рівнянню;
  - б) за даними кореляційної таблиці;
- 5) побудувати графіки одержаних ліній регресії.

***Література***

1. Гмурман В.Е. Теория вероятностей и математическая статистика. Учебное пособие для вузов. М., Высшая школа, 1977.
2. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. М., Высшая школа, 1975.
3. Колемаев В.А. и др. Теория вероятностей и математическая статистика: Учеб. пособие для экон. спец. вузов. – М.: Высшая школа, 1991.
4. Чистяков В.П. Курс теории вероятностей. – М.: Наука, 1982.