МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ

В. І. Жлуктенко, С. І. Наконечний, С. С. Савіна

Теорія ймовірностей і математична статистика

Частина II

Математична статистика

Навчально-методичний посібник

У двох частинах

Рекомендовано Міністерством освіти і науки України

Київ 2001

ББК 22.17
Ж 76

Рецензенти:

О. П. Суслов, д-р екон. наук, проф.
(НДЕІ М-ва економіки України)

О. А. Корольов, д-р екон. наук, проф.
(Київ. нац. торг.-екон. ун-т)

Гриф надано Міністерством освіти і науки України
Лист № 14/18.2-183 від 27.02.01

Жлуктенко В. І., Наконечний С. І., Савіна С. С.

Ж 76 Теорія ймовірностей і математична статистика: Навч.-метод. посібник: У 2-х ч. — Ч. ІІ. Математична статистика. — К.: КНЕУ, 2001. — 336 с.

ISBN 966–574–265–5

У другій частині пропонованого навчального посібника розглянуто основи математичної статистики як науки, що вивчає ймовірнісну природу статистичних оцінок параметрів генеральної сукупності, та закони їх розподілу. Ці закони застосовуються з метою побудови довірчих інтервалів параметрів генеральних сукупностей, а також для перевірки правильності параметричних і непараметричних статистичних гіпотез обробкою результатів вибірки. Докладно висвітлюються теоретичні основи дисперсійного та регресійного аналізу.

До кожної теми наведено розв'язування типових задач із поясненнями, а наприкінці кожної теми — перелік теоретичних питань та блок прикладів для тестування в аудиторних і домашніх умовах.

Посібник розрахований на самостійне вивчення курсу «Математична статистика» студентами економічних вузів усіх форм навчання.

ББК 22.17

ISBN 966–574–265–5

Навчальне видання

ЖЛУКТЕНКО Володимир Іванович
НАКОНЕЧНИЙ Степан Ількович
САВІНА Світлана Станіславівна

ТЕОРІЯ ЙМОВІРНОСТЕЙ І МАТЕМАТИЧНА СТАТИСТИКА

Навчальний посібник

У двох частинах

Частина II

МАТЕМАТИЧНА СТАТИСТИКА

Редактор *Н. Царик*
Художник обкладинки *Т. Зябліцева*
Технічний редактор *Т. Піхота*
Коректор *А. Бородавко*
Верстка *Т. Мальчевської*

Підписано до друку 2.08.2001. Формат 60×84/16. Папір офсетний №1.
Гарнітура Тип Таймс. Друк офсетний. Умов. друк. арк. 23,25.
Умов. фарбовідб. 23,69. Обл.-вид. арк. 26,46. Наклад 8500 прим. Зам. № 20-2077

Видавництво КНЕУ
03680, м. Київ, проспект Перемоги, 54/1

Свідоцтво про реєстрацію №235 від 07.11.2000

Тел./факс: (044) 458-00-66; (044) 446-64-58
E-mail: publish@kneu.kiev.ua

ВСТУП

Основним змістом математичної статистики є систематизація, обробка і використання статистичної інформації для виявлення статистичних закономірностей ознаки або ознак певної сукупності елементів.

Оскільки суцільна обробка всіх елементів сукупності практично неможлива, то, як правило, застосовується вибірковий метод. Отже, розрізняють генеральну і вибіркову сукупності.

Множина Ω однотипних елементів, яким притаманні певні кількісні ознаки (розміри, вага, маса тощо), утворює генеральну сукупність. Кількість усіх елементів генеральної сукупності називають її обсягом і позначають символом N, значення якого здебільшого невідоме.

Кожна непорожня підмножина A множини Ω ($A \subset \Omega$) випадково вибраних елементів із генеральної сукупності називається вибіркою. Кількість усіх елементів вибірки називають її обсягом і позначають символом n. Його значення відоме, причому воно набагато менше за обсяг генеральної сукупності ($n << N$).

Математична статистика розв’язує дві категорії задач:

1) статистичне оцінювання (точкове, інтервальне) параметрів генеральної сукупності;

2) перевірка правдивості статистичних гіпотез про значення параметрів генеральної сукупності або про закон розподілу ознаки генеральної сукупності на підставі обробки результатів вибірки.

РОЗДІЛ V

СТАТИСТИЧНІ РОЗПОДІЛИ ВИБІРКИ. СТАТИСТИЧНІ ОЦІНКИ ПАРАМЕТРІВ ГЕНЕРАЛЬНОЇ СУКУПНОСТІ

ТЕМА 12. СТАТИСТИЧНІ РОЗПОДІЛИ ВИБІРОК ТА ЇХ ЧИСЛОВІ ХАРАКТЕРИСТИКИ

1. Загальна інформація

Кількісні ознаки елементів генеральної сукупності можуть бути одновимірними і багатовимірними, дискретними і неперервними.

Коли реалізується вибірка, кількісна ознака, наприклад X, набуває конкретних числових значень ($X = x_i$), які називають *варіантою*.

Зростаючий числовий ряд варіант називають *варіаційним*.

Кожна варіанта вибірки може бути спостереженою n_i раз ($n_i \geq 1$), число n_i називають *частотою варіанти* x_i.

При цьому

$$n = \sum_{i=1}^{k} n_i \ , \tag{350}$$

де k — кількість варіант, що різняться числовим значенням;

n — обсяг вибірки.

Відношення частоти n_i варіанти x_i до обсягу вибірки n називають її *відносною частотою* і позначають через W_i , тобто

$$W_i = \frac{n_i}{n} . \tag{351}$$

Для кожної вибірки виконується рівність

$$\sum_{i=1}^{k} W_i = 1 . \tag{352}$$

Якщо досліджується ознака генеральної сукупності X, яка є неперервною, то варіант буде багато. У цьому разі варіаційний ряд — це певна кількість рівних або нерівних частинних інтервалів чи груп варіант зі своїми частотами.

Такі частинні інтервали варіант, які розміщені у зростаючій послідовності, утворюють *інтервальний варіаційний ряд*.

На практиці для зручності, як правило, розглядають інтервальні варіаційні ряди, у котрих інтервали є рівними між собою.

2. Дискретний статистичний розподіл вибірки та її числові характеристики

Перелік варіант варіаційного ряду і відповідних їм частот, або відносних частот, називають *дискретним статистичним розподілом вибірки*.

У табличній формі він має такий вигляд:

$X = x_i$	x_1	x_2	x_3	…	x_k
n_i	n_1	n_2	n_3	…	n_k
W_i	W_1	W_2	W_3	…	W_k

Дискретний статистичний розподіл вибірки можна подати емпіричною функцією $F^*(x)$.

Емпірична функція $F^*(x)$ та її властивості. Функція аргументу x, що визначає відносну частоту події $X < x$, тобто

$$F^*(x) = W(X < x) = \frac{n_x}{n}, \qquad (353)$$

називається *емпіричною*, або *комулятою*.

Тут n — обсяг вибірки;

n_x — кількість варіант статистичного розподілу вибірки, значення яких менше за фіксовану варіанту x;

$F^*(x)$ — називають ще *функцією нагромадження відносних частот*.

Властивості $F^*(x)$:

1) $0 \le F^*(x) \le 1$;

2) $F(x_{\min}) = 0$, де $x_{\min}$ є найменшою варіантою варіаційного ряду;

3) $F(x)\big|_{x>x_{\max}} = 1$, де $x_{\max}$ є найбільшою варіантою варіаційного ряду;

4) $F(x)$ є неспадною функцією аргументу x, а саме: $F(x_2) \ge F(x_1)$ при $x_2 \ge x_1$.

Полігон частот і відносних частот. Дискретний статистичний розподіл вибірки можна зобразити графічно у вигляді ламаної лінії, відрізки якої сполучають координати точок (x_i; n_i), або (x_i; W_i).

У першому випадку ламану лінію називають *полігоном частот*, у другому — *полігоном відносних частот*.

Приклад. За заданим дискретним статистичним розподілом вибірки

$X = x_i$	–6	–4	–2	2	4	6
n_i	5	10	15	20	40	10
W_i	0,05	0,1	0,15	0,2	0,4	0,1

потрібно:
1. Побудувати $F^*(x)$ і зобразити її графічно;
2. Накреслити полігони частот і відносних частот.

Розв'язання. Згідно з означенням та властивостями $F^*(x)$ має такий вигляд:

$$F^*(x) = W(X < x) = \frac{n_x}{n} = \begin{cases} 0 & x \le -6, \\ 0{,}05 & -6 < x \le -4, \\ 0{,}15 & -4 < x \le -2, \\ 0{,}3 & -2 < x \le 2, \\ 0{,}5 & 2 < x \le 4, \\ 0{,}9 & 4 < x \le 6, \\ 1, & x > 6. \end{cases}$$

Графічне зображення $F^*(x)$ подано на рис. 106.

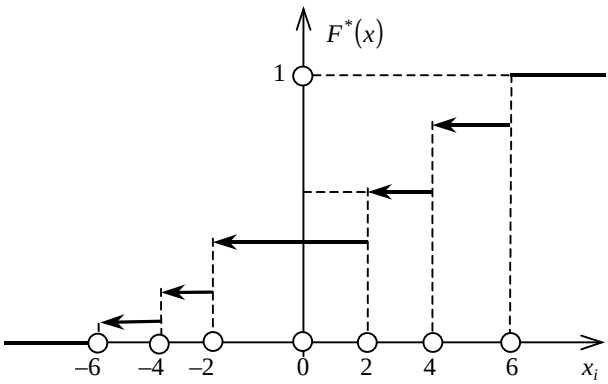


Рис. 106

Полігони частот та відносних частот зображено на рис.107, 108.

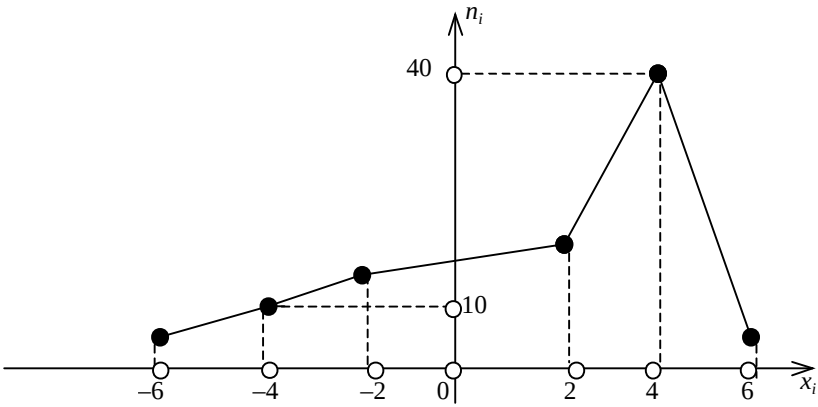


Рис. 107

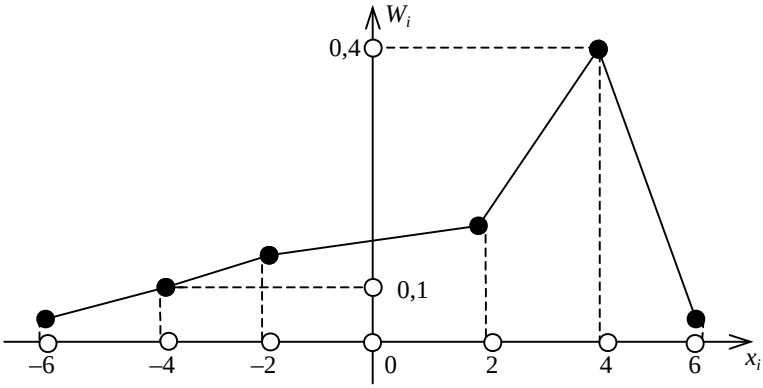


Рис. 108

***Числові характеристики*:**

1) *вибіркова середня величина* $\bar{x}_{\text{В}}$. Величину, яка визначається формулою

$$\bar{x}_{\text{В}} = \frac{\sum x_i n_i}{n}, \qquad (354)$$

називають *вибірковою середньою величиною дискретного статистичного розподілу вибірки*.

Тут x_i — варіанта варіаційного ряду вибірки;

n_i — частота цієї варіанти;

n — обсяг вибірки ($n = \sum n_i$).

Якщо всі варіанти з'являються у вибірці лише по одному разу, тобто $n_i = 1$, то

$$\bar{x}_{\text{В}} = \frac{\sum x_i}{n}; \qquad (355)$$

2) *відхилення варіант*. Різницю $(x_i - \bar{x}_{\text{В}})n_i$ називають відхиленням варіант.

При цьому

$$\sum (x_i - \bar{x}_{\text{В}})n_i = \sum x_i n_i - \sum \bar{x}_{\text{В}} n_i = n \cdot \bar{x}_{\text{В}} - n \cdot \bar{x}_{\text{В}} = 0 .$$

Отже, сума відхилень усіх варіант варіаційного ряду вибірки завжди дорівнює нулеві;

3) *мода* (Мо*). *Модою дискретного статистичного розподілу вибірки* називають варіанту, що має найбільшу частоту появи.

Мод може бути кілька. Коли дискретний статистичний розподіл має одну моду, то він називається *одномодальним*, коли має дві моди — *двомодальним* і т. д.;

4) *медіана* (Ме*). *Медіаною дискретного статистичного розподілу вибірки* називають варіанту, яка поділяє варіаційний ряд на дві частини, рівні за кількістю варіант;

5) *дисперсія*. Для вимірювання розсіювання варіант вибірки відносно $\bar{x}_{\text{В}}$ вибирається дисперсія.

Дисперсія вибірки — це середнє арифметичне квадратів відхилень варіант відносно $\bar{x}_{\text{В}}$, яке обчислюється за формулою

$$D_{\text{В}} = \frac{\sum (x_i - \bar{x}_{\text{В}})^2 n_i}{n} \qquad (356)$$

або

$$D_{\text{В}} = \frac{\sum x_i^2 n_i}{n} - (\bar{x}_{\text{В}})^2; \qquad (357)$$

6) *середнє квадратичне відхилення вибірки* $\sigma_{\text{В}}$. При обчисленні $D_{\text{В}}$ відхилення підноситься до квадрата, а отже, змінюється одиниця виміру ознаки X, тому на основі дисперсії вводиться середнє квадратичне відхилення

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}}, \qquad (358)$$

яке вимірює розсіювання варіант вибірки відносно $\bar{x}_{\text{В}}$, але в тих самих одиницях, в яких вимірюється ознака X;

7) *розмах* (R). Для грубого оцінювання розсіювання варіант відносно $\bar{x}_{\text{В}}$ застосовується величина, яка дорівнює різниці між

найбільшою $x_{\max}$ і найменшою $x_{\min}$ варіантами варіаційного ряду. Ця величина називається *розмахом*

$$R = x_{\max} - x_{\min}; \qquad (359)$$

8) *коефіцієнт варіації V.* Для порівняння оцінок варіацій статистичних рядів із різними значеннями $\bar{x}_{\text{В}}$, які не дорівнюють нулеві, вводиться коефіцієнт варіації, який обчислюється за формулою

$$V = \frac{\sigma_{\text{В}}}{\bar{x}_{\text{В}}} 100\%. \qquad (360)$$

Приклад. За заданим статистичним розподілом вибірки

$X = x_i$	2,5	4,5	6,5	8,5	10,5
n_i	10	20	30	30	10

потрібно:
1) обчислити $\bar{x}_{\text{В}}$, $D_{\text{В}}$, $\sigma_{\text{В}}$;
2) знайти Мо^*, Ме^*;
3) обчислити *R*, *V*.

Розв'язання. Оскільки $n = \sum n_i = 100$, то згідно з формулами (354), (357), (358) дістанемо:

$$\bar{x}_{\text{В}} = \frac{\sum x_i n_i}{n} = \frac{2{,}5 \cdot 10 + 4{,}5 \cdot 20 + 6{,}5 \cdot 30 + 8{,}5 \cdot 30 + 10{,}5 \cdot 10}{100} = 6{,}7;$$

$$\bar{x}_{\text{В}} = 6{,}7.$$

Для обчислення $D_{\text{В}}$ визначається

$$\frac{\sum x_i^2 n_i}{n} = \frac{(2{,}5)^2 \cdot 10 + (4{,}5)^2 \cdot 20 + (6{,}5)^2 \cdot 30 + (8{,}5)^2 \cdot 30 + (10{,}5)^2 \cdot 10}{100} = 50{,}05.$$

Тоді

$$D_{\text{В}} = \frac{\sum x_i^2 n_i}{n} - (\bar{x}_{\text{В}})^2 = 50{,}05 - (6{,}7)^2 = 50{,}05 - 44{,}89 = 5{,}16.$$

$$D_{\text{В}} = 5{,}16.$$

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}} = \sqrt{5{,}16} \approx 2{,}27.$$

$$\sigma_{\text{В}} = 2{,}27.$$

$$\text{Мо}^* = 6{,}5;\ 8{,}5.$$

Отже, наведений статистичний розподіл вибірки буде двомодальним. $\text{Ме}^* = 6{,}5$, оскільки варіанта $x = 6{,}5$ поділяє варіаційний ряд

2,5; 4,5; **6,5**; 8,5; 10,5 на дві частини: 2,5; 4,5 і 8,5; 10,5, які мають однакову кількість варіант.

$$R = x_{\max} - x_{\min} = 10{,}5 - 2{,}5 = 8.$$

$$V = \frac{\sigma_{\text{В}}}{\bar{x}_{\text{В}}} 100\% = \frac{2{,}27}{6{,}7} 100\% = 33{,}88\%.$$

3. Інтервальний статистичний розподіл вибірки та його числові характеристики

Перелік часткових інтервалів і відповідних їм частот, або відносних частот, називають *інтервальним статистичним розподілом вибірки*.

У табличній формі цей розподіл має такий вигляд:

h	$x_1 - x_2$	$x_2 - x_3$	$x_3 - x_4$	…	$x_{k-1} - x_k$
n_i	n_1	n_2	n_3	…	N_k
W_i	W_1	W_2	W_3	…	W_k

Тут $h = x_i - x_{i-1}$ є довжиною часткового i-го інтервалу. Як правило, цей інтервал береться однаковим.

Інтервальний статистичний розподіл вибірки можна подати графічно у вигляді гістограми частот або відносних частот, а також, як і для дискретного статистичного розподілу, емпіричною функцією $F^*(x)$ (комулятою).

Гістограма частот та відносних частот. Гістограма частот являє собою фігуру, яка складається з прямокутників, кожний з яких має основу h і висоту $n_i \dfrac{1}{h}$.

Гістограма відносних частот є фігурою, що складається з прямокутників, кожний з яких має основу завдовжки h і висоту, що дорівнює $W_i \dfrac{1}{h}$.

Приклад. За заданим інтервальним статистичним розподілом вибірки

$h = 8$	0—8	8—16	16—24	24—32	32—40	40—48
n_i	10	15	20	25	20	10
W_i	0,1	0,15	0,2	0,25	0,2	0,1

потрібно побудувати гістограму частот і відносних частот.

Розв'язання. Гістограми частот і відносних частот наведені на рис. 109, 110.

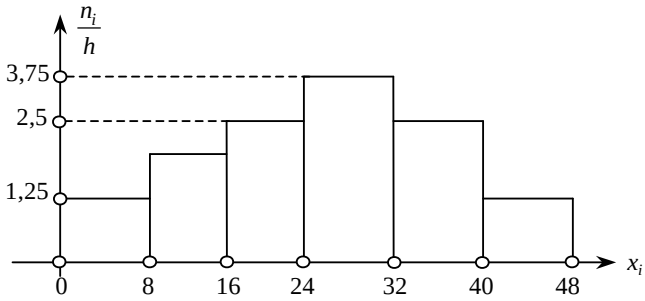


Рис. 109

Площа гістограми частот $S = \sum h \frac{n_i}{h} = \sum n_i = n = 100.$

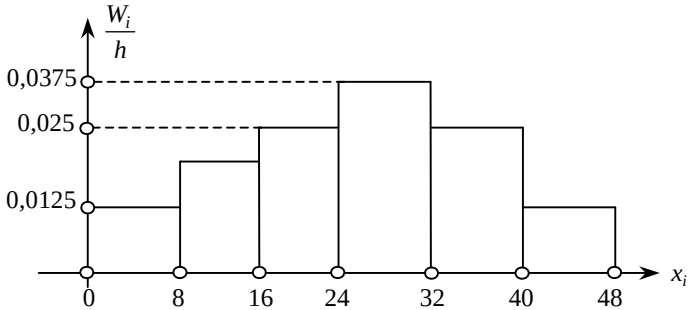


Рис. 110

Площа гістограми відносних частот

$$S = \sum h \frac{W_i}{h} = \sum W_i = 1.$$

Емпірична функція $F^*(x)$ (комулята). При побудові комуляти $F^*(x)$ для інтервального статистичного розподілу вибірки за основу береться припущення, що ознака на кожному частинному інтервалі має рівномірну щільність імовірностей. Тому комулята матиме вигляд ламаної лінії, яка зростає на кожному частковому інтервалі і наближається до одиниці.

Приклад. Для заданого інтервального статистичного розподілу вибірки

$h = 10$	0—10	10—20	20—30	30—40	40—50	50—60
n_i	5	15	20	25	30	5

побудувати $F^*(x)$ і подати її графічно.

Розв'язання.

$$F^*(x) = W(X < x) = \frac{n_x}{n} = \begin{cases} 0, & x \le 0, \\ 0{,}05 & 0 < x \le 10, \\ 0{,}2 & 10 < x \le 20, \\ 0{,}4 & 20 < x \le 30, \\ 0{,}65 & 30 < x \le 40, \\ 0{,}95 & 40 < x \le 50, \\ 1 & 50 < x \le 60. \end{cases}$$

Графік $F^*(x)$ зображено на рис. 111.

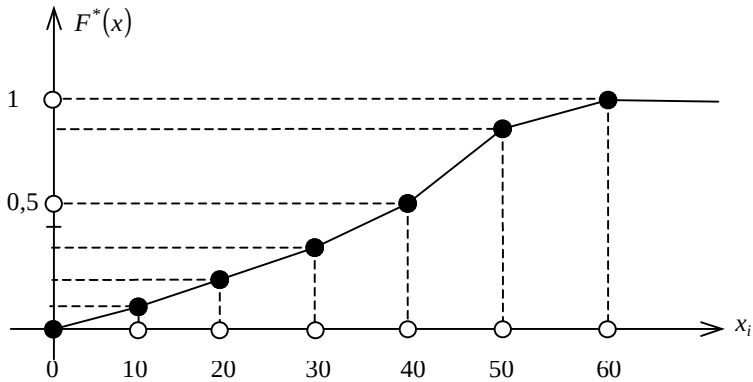


Рис. 111

Аналогом емпіричної функції $F^*(x)$ у теорії ймовірностей є інтегральна функція $F(x) = P(X < x)$.

Медіана. Для визначення медіани інтервального статистичного розподілу вибірки необхідно визначити медіанний частковий інтервал. Якщо, наприклад, на i-му інтервалі $[x_{i-1} - x_i]$ $F^*(x_{i-1}) < 0{,}5$ і $F^*(x_i) > 0{,}5$, то, беручи до уваги, що досліджувана ознака X є неперервною і при цьому $F^*(x)$ є неспадною функцією, всередині інтервалу $[x_{i-1} - x_i]$ неодмінно існує таке значення $X = \text{Me}$, де $F^*(\text{Me}) = 0{,}5$.

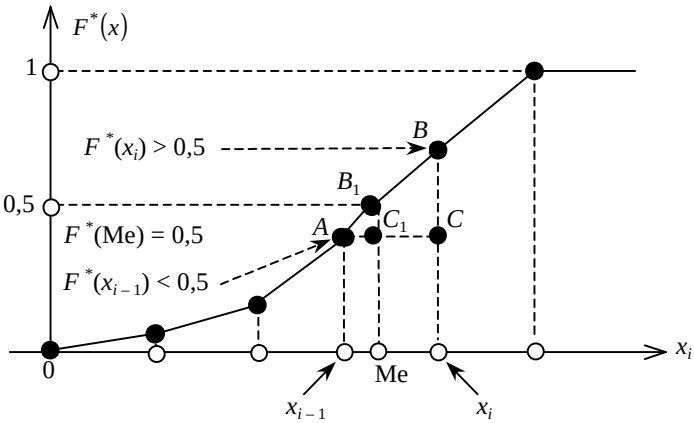


Рис. 112

З подібності трикутників ΔABC і ΔAB_1C_1, зображених на рис. 112, маємо:

$$\frac{x_i - x_{i-1}}{\text{Me}^* - x_{i-1}} = \frac{F^*(x_i) - F^*(x_{i-1})}{0,5 - F^*(x_{i-1})} \to \text{Me}^* = x_{i-1} + \frac{0,5 - F^*(x_{i-1})}{F^*(x_i) - F^*(x_{i-1})} h, \quad (361)$$

де $h = x_i - x_{i-1}$ називають *кроком*.

Мода. Для визначення моди інтервального статистичного розподілу необхідно знайти модальний інтервал, тобто такий частинний інтервал, що має найбільшу частоту появи.

Використовуючи лінійну інтерполяцію, моду обчислимо за формулою

$$\text{Mo}^* = x_{i-1} + \frac{n_{\text{Mo}} - n_{\text{Mo}-1}}{2n_{\text{Mo}} - n_{\text{Mo}-1} - n_{\text{Mo}+1}} h, \quad (362)$$

де x_{i-1} — початок модального інтервалу;
h — довжина, або крок, часткового інтервалу;
n_{Mo} — частота модального інтервалу;
$n_{\text{Mo}-1}$ — частота домодального інтервалу;
$n_{\text{Mo}+1}$ — частота післямодального інтервалу.

Приклад. За заданим інтервальним статистичним розподілом вибірки

$h = 4$	0—4	4—8	8—12	12—16	16—20	20—24
n_i	6	14	20	25	30	5

побудувати гістограму частот і $F^*(x)$.
Визначити Mo^*, Me^*.

Розв'язання. Гістограма частот зображена на рис. 113.

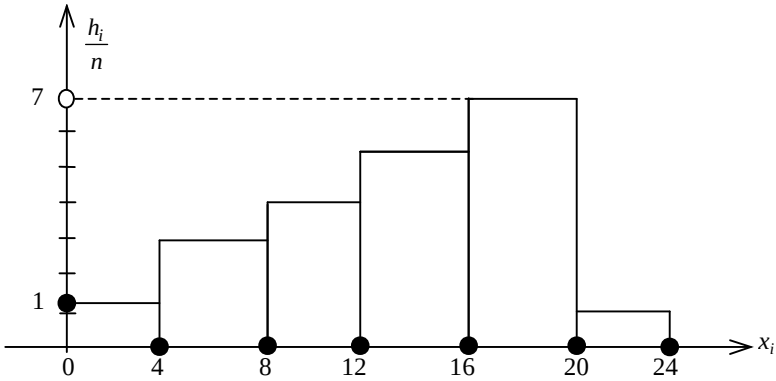


Рис. 113

Графік $F^*(x)$ зображено на рис. 114.

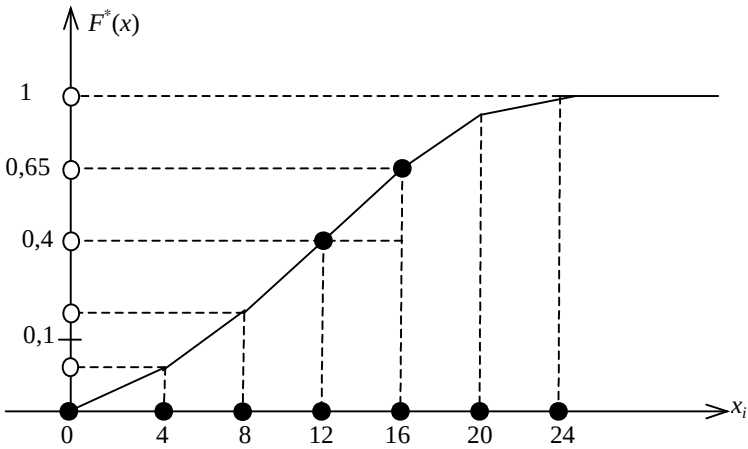


Рис. 114

З рис. 113 визначається модальний інтервал, який дорівнює 16—20.

Застосовуючи (362) і беручи до уваги, що $n_{\text{Mo}} = 30$, $n_{\text{Mo}-1} = 25$, $n_{\text{Mo}+1} = 5$, $h = 4$, $x_{i-1} = 16$, дістанемо

$$\text{Mo}^* = x_{i-1} + \frac{n_{\text{Mo}} - n_{\text{Mo}-1}}{2n_{\text{Mo}} - n_{\text{Mo}-1} - n_{\text{Mo}+1}} h;$$

$$\text{Mo}^* = 16 + \frac{30-25}{60-25-5} 4 = 16 + \frac{5}{30} = 16{,}17.$$

Отже, $\text{Mo}^* = 16{,}17$.

З графіка $F^*(x)$ визначається медіанний інтервал, який дорівнює 12—16.

Беручи до уваги, що $F(12) = 0{,}4$, $F(16) = 0{,}65$, $h = 4$ і застосовуючи (361), дістанемо:

$$\text{Me}^* = x_{i-1} + \frac{0{,}5 - F^*(x_{i-1})}{F^*(x_i) - F^*(x_{i-1})} h = 12 + \frac{0{,}5-0{,}4}{0{,}65-0{,}4} 4 = 12 + \frac{0{,}1}{0{,}25} 4 = 13{,}6.$$

Отже, $\text{Me}^* = 13{,}6$.

$\bar{x}_{\text{В}}, D_{\text{В}}, \sigma_{\text{В}}$ для інтервального статистичного розподілу вибірки. Для визначення $\bar{x}_{\text{В}}, D_{\text{В}}, \sigma_{\text{В}}$ перейдемо від інтервального розподілу до дискретного, варіантами якого є середина часткових інтервалів $x_i^* = x_{i-1} + \frac{h}{2} = x_i - \frac{h}{2}$ і який має такий вигляд:

$x_i^* = x_i - \frac{h}{2} = x_{i-1} + \frac{h}{2}$	x_1^*	x_2^*	x_3^*	…	x_k^*
h_i	h_1	h_2	h_3	…	h_k

Тоді $\bar{x}_{\text{В}}, D_{\text{В}}, \sigma_{\text{В}}$ обчислюються за формулами:

$$\bar{x}_{\text{В}} = \frac{\sum x_1^* n_i}{h}; \tag{363}$$

$$D_{\text{В}} = \frac{\sum (x_1^*)^2 n_i}{h} - (\bar{x}_{\text{В}})^2; \tag{364}$$

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}}. \tag{365}$$

Приклад. За заданим інтервальним статистичним розподілом вибірки, в якому наведено розподіл маси новонароджених x_i,

$X = x_i$, кг	1—1,2	1,2—1,4	1,4—1,6	1,6—1,8	1,8—2	1,8—2	2—2,2	2,4—2,6	2,6—2,8	2,8—3	3—3,2
n_i	5	12	18	22	36	24	19	15	11	9	2

обчислити $\bar{x}_{\text{В}}, D_{\text{В}}, \sigma_{\text{В}}$.

Розв'язання. Побудуємо дискретний статистичний розподіл за заданим інтервальним. Оскільки $h = 0{,}2$, то дістанемо:

$x_i^* = x_i - \frac{h}{2} = x_{i-1} + \frac{h}{2}$	1,1	1,3	1,5	1,7	1,9	2,1	2,3	2,5	2,7	2,9	3,1
h_i	5	12	18	22	36	24	19	15	11	9	2

Беручи до уваги (363), (364), (365) і те, що $n = 173$, дістанемо:

$$\bar{x}_{\text{В}} = \frac{\sum x_1^* n_i}{n} = \frac{5{,}5 + 15{,}6 + 27 + 37{,}4 + 68{,}4 + 50{,}4 + 43{,}7}{173} +$$

$$+ \frac{37{,}5 + 29{,}7 + 26{,}1 + 6{,}2}{173} = \frac{347{,}5}{173} \approx 2{,}008671 \text{ кг}.$$

Отже, $\bar{x}_{\text{В}} = 2{,}008671$ кг.

$$\frac{\sum (x_1^*)^2 n_i}{n} = \frac{6{,}05 + 20{,}29 + 40{,}5 + 63{,}58 + 129{,}96 + 105{,}84 + 100{,}51}{173} +$$

$$+ \frac{93{,}75 + 80{,}19 + 75{,}69 + 19{,}22}{173} = \frac{735{,}58}{173} = 4{,}251908.$$

$$D_{\text{В}} = \frac{\sum \left(x_1^*\right)^2 n_i}{n} - (\bar{x}_B)^2 = 4{,}251908 - (2{,}008671)^2 =$$

$$= 4{,}251908 - 4{,}034759 = 0{,}217149.$$

$$D_{\text{В}} = 0{,}217149.$$

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}} = \sqrt{0{,}217149} \approx 0{,}466.$$

Отже, $\sigma_{\text{В}} = 0{,}466$ кг.

4. Двовимірний статистичний розподіл вибірки та його числові характеристики

Перелік варіант $Y = y_i$, $X = x_j$ та відповідних їм частот n_{ij} спільної їх появи утворюють *двовимірний статистичний розподіл вибірки*, що реалізована з генеральної сукупності, елементам цієї вибірки притаманні кількісні ознаки X і Y.

У табличній формі цей розподіл має такий вигляд:

$Y = y_i$	$X = x_j$					
	x_1	x_2	x_3	…	x_m	n_{y_i}
y_1	n_{11}	n_{12}	n_{13}	…	n_{1m}	n_{y_1}
y_2	n_{21}	n_{22}	n_{23}	…	n_{2m}	n_{y_2}
y_3	n_{31}	n_{32}	n_{33}	…	n_{3m}	n_{y_3}
…	…	…	…	…	…	…
y_k	n_{k1}	n_{k2}	n_{k3}	…	n_{km}	n_{y_k}
n_{x_j}	n_{x_1}	n_{x_2}	n_{x_3}	…	n_{x_m}	

Тут n_{ij} — частота спільної появи варіант

$$Y = y_i, \quad X = x_j;$$

$$n_{y_i} = \sum_{j=1}^{m} n_{ij}, \qquad n_{x_j} = \sum_{i=1}^{k} n_{ij};$$

$$n = \sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij} = \sum_{i=1}^{k} n_{y_i} = \sum_{j=1}^{m} n_{x_j} .$$

Загальні числові характеристики ознаки X:
загальна середня величина ознаки X

$$\bar{x} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} x_j\, n_{ij}}{n} = \frac{\sum_{j=1}^{m} x_j\, n_{x_j}}{n}; \qquad (366)$$

загальна дисперсія ознаки X

$$D_x = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} x_j^2\, n_{ij}}{n} - (\bar{x})^2 = \frac{\sum_{j=1}^{m} x_j^2\, n_{x_j}}{n} - (\bar{x})^2; \qquad (367)$$

загальне середнє квадратичне відхилення ознаки X

$$\sigma_x = \sqrt{D_x}. \qquad (368)$$

Загальні числові характеристики ознаки Y:
загальна середня величина ознаки Y

$$\bar{y} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} y_i\, n_{ij}}{n} = \frac{\sum_{i=1}^{k} y_i\, n_{y_i}}{n}; \qquad (369)$$

загальна дисперсія ознаки Y

$$D_y = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} y_i^2\, n_{ij}}{n} - (\bar{y})^2 = \frac{\sum_{i=1}^{k} y_i^2\, n_{y_i}}{n} - (\bar{y})^2; \qquad (370)$$

загальне середнє квадратичне відхилення ознаки Y

$$\sigma_y = \sqrt{D_y}. \qquad (371)$$

Умовні статистичні розподіли та їх числові характеристики

Умовним статистичним розподілом ознаки Y при фіксованому значенні $X = x_i$ називають перелік варіант ознаки Y та відповідних їм частот, узятих при фіксованому значенні X.

$$Y / X = x_j.$$

$Y = y_i$	y_1	y_2	y_3	…	y_k
n_{ij}	n_{1j}	n_{2j}	n_{3j}	…	n_{kj}

Тут $\sum_{i=1}^{k} n_{ij} = n_{x_j}$.

Числові характеристики для такого статистичного розподілу називають *умовними*. До них належать:
умовна середня ознаки Y

$$\bar{y}_{X=x_j} = \frac{\sum_{i=1}^{k} y_i\, n_{ij}}{\sum_{i=1}^{k} n_{ij}} = \frac{\sum_{i=1}^{k} y_i\, n_{ij}}{n_{x_j}}; \qquad (372)$$

умовна дисперсія ознаки Y

$$D(Y / X = x_j) = \frac{\sum_{i=1}^{k} y_i^2\, n_{ij}}{n_{x_j}} - (\bar{y}_{x=x_j})^2; \qquad (373)$$

умовне середнє квадратичне відхилення ознаки Y

$$\sigma(Y / X = x_j) = \sqrt{D(Y / X = x_j)}. \qquad (374)$$

$D(Y / X = x_j)$, $\sigma(Y / X = x_j)$ вимірюють розсіювання варіант ознаки Y щодо умовної середньої величини $\bar{y}_{x=x_j}$.

Умовним статистичним розподілом ознаки X при $Y = y_i$ називають перелік варіант $X = x_j$ та відповідних їм частот, узятих при фіксованому значенні ознаки $Y = y_i$.

$$X / Y = y_i .$$

$X = x_j$	x_1	x_2	x_3	…	x_m
n_{ij}	n_{i1}	n_{i2}	n_{i3}	…	n_{im}

Тут $\sum_{j=1}^{m} n_{ij} = n_{y_i}$.

Умовні числові характеристики для цього розподілу:
умовна середня величина ознаки X

$$\bar{x}_{Y=y_j} = \frac{\sum_{j=1}^{m} x_i \; n_{ij}}{\sum_{j=1}^{m} n_{ij}} = \frac{\sum_{j=1}^{m} x_i \; n_{ij}}{n_{y_i}} ; \qquad (375)$$

умовна дисперсія ознаки X

$$D(X / Y = y_i) = \frac{\sum_{j=1}^{m} x_i^2 \; n_{ij}}{n_{y_i}} - (\bar{x}_{y=y_j})^2 ; \qquad (376)$$

умовне середнє квадратичне відхилення ознаки X

$$\sigma((X / Y = y_i)) = \sqrt{D((X / Y = y_i))} . \qquad (377)$$

При відомих значеннях умовних середніх $\bar{y}_{x_j}$, $\bar{x}_{y_i}$ загальні середні ознаки X та Y можна обчислити за формулами:

$$\bar{y} = \frac{\sum_{j=1}^{n} y_{x_j} \; n_{x_j}}{n} ; \qquad (378)$$

$$\bar{x} = \frac{\sum_{i=1}^{m} x_{y_i} \; n_{y_i}}{n} . \qquad (379)$$

Кореляційний момент, вибірковий коефіцієнт кореляції

Під час дослідження двовимірного статистичного розподілу вибірки постає потреба з’ясувати наявність зв’язку між ознаками X

і Y, який у статистиці називають кореляційним. Для цього обчислюється емпіричний кореляційний момент K_{xy}^{*} за формулою

$$K_{xy}^{*} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{m} y_i \, x_i \, n_{ij}}{n} - \bar{x} \cdot \bar{y} \, . \qquad (380)$$

Якщо $K_{xy}^{*} = 0$, то кореляційного зв'язку між ознаками X і Y немає. Якщо ж $K_{xy}^{*} \neq 0,$ то цей зв'язок існує.

Отже, кореляційний момент дає лише відповідь на запитання: є зв'язок між ознаками X і Y, чи його немає.

Для вимірювання тісноти кореляційного зв'язку обчислюється вибірковий коефіцієнт кореляції $r_{\text{В}}$ за формулою

$$r_{\text{В}} = \frac{K_{xy}^{*}}{\sigma_x \sigma_y} \, . \qquad (381)$$

Як і в теорії ймовірностей, $|r_{\text{В}}| \leq 1, \quad -1 \leq r_{\text{В}} \leq 1.$

Приклад. За заданим двовимірним статистичним розподілом вибірки ознак X і Y

$Y = y_i$	$X = x_j$				
	10	20	30	40	n_{y_i}
2	—	2	4	4	10
4	10	8	6	6	30
6	5	10	5	—	20
8	15	—	15	10	40
n_{x_j}	30	20	30	20	

потрібно:

1) обчислити K_{xy}^{*}, $r_{\text{В}}$;

2) побудувати умовні статистичні розподіли $Y / X = 30$, $X / Y = 4$ й обчислити умовні числові характеристики.

Розв'язання. 1) Щоб обчислити K_{xy}^{*}, $r_{\text{В}}$ визначимо $\bar{x}, \sigma_x, \bar{y}, \sigma_y$. Оскільки $n = \sum\sum n_{ij} = 100,$ то

$$\bar{x} = \frac{\sum x_j \; n_{x_j}}{n} = \frac{10 \cdot 30 + 20 \cdot 20 + 30 \cdot 30 + 40 \cdot 20}{100} =$$

$$= \frac{300 + 400 + 900 + 800}{100} = \frac{2400}{100} = 24.$$

$$\bar{x} = 24.$$

$$\frac{\sum x_j^2 \; n_{x_j}}{n} = \frac{(10)^2 \cdot 30 + (20)^2 \cdot 20 + (30)^2 \cdot 30 + (40)^2 \cdot 20}{100} =$$

$$= \frac{3000 + 8000 + 27000 + 32000}{100} = \frac{70000}{100} = 700.$$

$$D_x = \frac{\sum x_j^2 \; n_{x_j}}{n} - (\bar{x})^2 = 700 - (24)^2 = 700 - 576 = 124.$$

$$\sigma_x = \sqrt{D_x} = \sqrt{124} \approx 11{,}14.$$

Отже, $\sigma_x = 11{,}14$.

$$\bar{y} = \frac{\sum y_i \; n_{y_i}}{n} = \frac{2 \cdot 10 + 4 \cdot 30 + 6 \cdot 20 + 8 \cdot 40}{100} =$$

$$= \frac{20 + 120 + 120 + 320}{100} = 5{,}8.$$

Отже, $\bar{y} = 5{,}8$.

$$\frac{\sum y_i^2 \; n_{y_i}}{n} = \frac{(2)^2 \cdot 10 + (4)^2 \cdot 30 + (6)^2 \cdot 20 + (8)^2 \cdot 40}{100} =$$

$$= \frac{40 + 480 + 720 + 2560}{100} = \frac{3800}{100} = 38.$$

$$D_y = \frac{\sum y_i^2 n_{y_i}}{n} - (\bar{y})^2 = 38 - (5{,}8)^2 = 38 - 33{,}64 = 4{,}36,$$

$$\sigma_y = \sqrt{D_y} = \sqrt{4{,}36} \approx 2{,}1.$$

Для визначення K_{xy}^* обчислюють

$$\sum\sum y_i x_j n_{ij} = 2 \cdot 10 \cdot 0 + 2 \cdot 20 \cdot 2 + 2 \cdot 30 \cdot 4 + 2 \cdot 40 \cdot 4 + 4 \cdot 10 \cdot 10 + 4 \cdot 20 \cdot 8 +$$

$$+ 4 \cdot 30 \cdot 6 + 4 \cdot 40 \cdot 6 + 6 \cdot 10 \cdot 5 + 6 \cdot 20 \cdot 10 + 6 \cdot 30 \cdot 5 + 6 \cdot 40 \cdot 0 + 8 \cdot 10 \cdot 15 +$$

$$+ 8 \cdot 20 \cdot 0 + 8 \cdot 30 \cdot 15 + 8 \cdot 40 \cdot 10 = 0 + 80 + 240 + 320 + 400 + 640 + 720 +$$

$$+ 960 + 300 + 1200 + 900 + 0 + 1200 + 0 + 3600 + 3200 = 13760.$$

Тоді

$$K_{xy}^{*}=\frac{\sum\sum y_i x_j n_{ij}}{n}-\bar{x}\cdot\bar{y}=\frac{13760}{100}-24\cdot 5{,}8=137{,}6-139{,}2=-1{,}6.$$

Отже, $K_{xy}^{*}=-1{,}6,$ а це свідчить про те, що між ознаками X і Y існуватиме від'ємний кореляційний зв'язок.

Для вимірювання тісноти цього зв'язку обчислимо вибірковий коефіцієнт кореляції.

$$r_{\text{В}}=\frac{K_{xy}^{*}}{\sigma_x\sigma_y}=\frac{-1{,}6}{11{,}14\cdot 2{,}1}=\frac{-1{,}6}{23{,}394}\approx -0{,}068.$$

Отже, $r_{\text{В}}=-0{,}068,$ тобто тіснота кореляційного зв'язку між ознаками X та Y є слабкою.

Умовний статистичний розподіл $X/Y=30$ матиме такий вигляд:

$Y=y_i$	2	4	6	8
n_{i3}	4	6	5	15

Обчислюються умовні числові характеристики для цього розподілу:

Умовна середня величина

$$\bar{y}_{X=30}=\frac{\sum\limits_{j=1}^{n} y_i\ n_{i3}}{\sum\limits_{j=1}^{n} n_{i3}}=\frac{2\cdot 4+4\cdot 6+6\cdot 5+8\cdot 15}{30}=\frac{8+24+30+120}{30}=\frac{182}{30}=6{,}07.$$

Умовна дисперсія та середнє квадратичне відхилення

$$\frac{\sum\limits_{j=1}^{n} y_i^2\ n_{i3}}{\sum n_{i3}}=\frac{(2)^2\cdot 4+(4)^2\cdot 6+(6)^2\cdot 5+(8)^2\cdot 15}{30}=\frac{16+96+180+960}{30}=$$

$$=\frac{1252}{30}=41{,}73;$$

$$D\big(X/Y=30\big)=\frac{\sum y_i^2\ n_{i3}}{\sum n_{i3}}-\big(\bar{y}_{X=30}\big)^2=41{,}73-36{,}8449\approx 4{,}89\ ;$$

$$\sigma\big(Y/X=30\big)=\sqrt{D_{(Y/X=30)}}=\sqrt{4{,}89}\approx 2{,}21\,.$$

Отже, $\sigma\big(Y/X=30\big)\approx 2{,}21\,.$

Умовний статистичний розподіл $X/Y=4$ матиме такий вигляд:

$X = x_j$	10	20	30	40
n_{2j}	10	8	6	6

Обчислюються умовні числові характеристики.

Умовна середня величина

$$\bar{x}_{Y=4} = \frac{\sum_{j=1}^{m} x_i\, n_{2j}}{\sum_{j=1}^{m} n_{2j}} = \frac{10 \cdot 10 + 20 \cdot 8 + 30 \cdot 6 + 40 \cdot 6}{30} =$$

$$= \frac{100 + 160 + 180 + 240}{30} = \frac{680}{30} \approx 22{,}7.$$

Отже, $\bar{x}_{y=4} \approx 22{,}7$.

Умовна дисперсія та середнє квадратичне відхилення

$$\frac{\sum_{j=1}^{m} x_i^2\, n_{2j}}{\sum_{j=1}^{m} n_{2j}} = \frac{(10)^2 \cdot 10 + (20)^2 \cdot 8 + (30)^2 \cdot 6 + (40)^2 \cdot 6}{30} =$$

$$= \frac{1000 + 3200 + 5400 + 9600}{30} = \frac{19200}{30} = 640.$$

$$D(X / y = 4) = \frac{\sum_{j=1}^{m} x_i\, n_{2j}}{\sum_{j=1}^{m} n_{2j}} - (\bar{x}_{y=4})^2 = 640 - (22{,}7)^2 = 640 - 515{,}29 = 124{,}71.$$

$$\sigma(X / y = 4) = \sqrt{D_{(X / y=4)}} = \sqrt{124{,}71} \approx 11{,}17\,.$$

Отже, $\sigma(X / y = 4) \approx 11{,}17$.

5. Парний статистичний розподіл вибірки та його числові характеристики

Якщо частота спільної появи ознак X і Y $n_{ij} = 1$ для всіх варіант, то в цьому разі двовимірний статистичний розподіл набуває такого вигляду:

$Y = y_i$	y_1	y_2	y_3	y_4	…	y_n
$X = x_j$	x_1	x_2	x_3	x_4	…	x_k

Його називають *парним статистичним розподілом вибірки*. Тут кожна пара значень ознак X і Y з'являється лише один раз.

Обсяг вибірки в цьому разі дорівнює кількості пар, тобто n.

Числові характеристики ознаки X:
середня величина

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}; \qquad (382)$$

дисперсія

$$D_x = \frac{\sum_{i=1}^{n} x_i^2}{n} - (x)^2; \qquad (383)$$

середнє квадратичне відхилення

$$\sigma_x = \sqrt{D_x}. \qquad (384)$$

Числові характеристики ознаки Y:
середня величина

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}; \qquad (385)$$

дисперсія

$$D_y = \frac{\sum_{i=1}^{n} y_i^2}{n} - (y)^2; \qquad (386)$$

середнє квадратичне відхилення

$$\sigma_y = \sqrt{D_y}; \qquad (387)$$

емпіричний кореляційний момент

$$K_{xy}^* = \frac{\sum y_i \ x_i}{n} - \bar{x} \cdot \bar{y}; \qquad (388)$$

вибірковий коефіцієнт кореляції

$$r_{\text{В}} = \frac{K_{xy}^*}{\sigma_x \sigma_y}. \qquad (389)$$

Приклад. Залежність кількості масла y_i, що його споживає певна особа за місяць, від її прибутку в гривнях x_i наведена в таблиці:

y_i, грн.	10,5	15,8	17,8	19,5	20,4	21,5	22,2	24,3	25,3	26,5	28,1	30,1	35,2	36,4	37	38,5	39,5	40,5	41	42,5
x_i, грн.	70	75	82	89	95	100	105	110	115	120	125	130	135	140	145	150	155	160	165	170

Потрібно обчислити K_{xy}^{*}, r_B.

Розв'язання. Оскільки обсяг вибірки n = 20, то маємо:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{70+75+82+89+95+100+105+110+115+120+125+}{20}$$

$$\frac{+130+135+140+145+150+155+160+165+170}{20} = \frac{2436}{20} = 121{,}8.$$

Отже, $\bar{x} = 121{,}8$.

$$\frac{\sum x_i^2}{n} = \frac{4900+5625+6724+7921+9025+10000+11025+}{20}$$

$$\frac{+12100+13225+14400+15625+16900+18225+19600+21025+}{20}$$

$$\frac{+22500+24025+25600+27225+28900}{20} = \frac{314570}{20} = 15728{,}5.$$

$$D_x = \frac{\sum x_i^2}{n} - (\bar{x})^2 = 15728{,}5 - (121{,}8)^2 = 15728{,}5 - 14835{,}24 = 893{,}26.$$

$$D_x = 893{,}26.$$

$$\sigma_x = \sqrt{D_x} = \sqrt{893{,}26} = 29{,}89.$$

$$\sigma_x = 29{,}89.$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{10{,}5+15{,}8+17{,}8+19{,}5+20{,}4+21{,}5+22{,}2+24{,}3+25{,}3+}{20}$$

$$\frac{+26{,}5+28{,}1+30{,}1+35{,}2+36{,}4+37+38{,}5+39{,}5+40{,}5+41+42{,}5}{20} =$$

$$= \frac{572{,}6}{20} = 28{,}63.$$

$$\bar{y} = 28{,}63.$$

$$\frac{\sum y_i^2}{n} = \frac{110{,}25+249{,}64+316{,}84+380{,}25+416{,}16+462{,}25+}{20}$$

$$\frac{+492{,}84+590{,}49+640{,}09+702{,}25+789{,}61+906{,}01+1239{,}04+}{20}$$

$$\frac{+1324{,}96+1369+1482{,}25+1560{,}25+1640{,}25+1681+1806{,}25}{20}=$$

$$=\frac{18159{,}68}{20}=907{,}98.$$

$$D_y=\frac{\sum y_i^2}{n}-(\overline{y})^2=907{,}98-819{,}68=88{,}3.$$

$$D_y=88{,}3\,.$$

$$\sigma_y=\sqrt{D_y}=\sqrt{88{,}3}\approx 9{,}4\,.$$

$$\sigma_y=9{,}4.$$

$$\sum\sum y_i x_i=10{,}5\cdot 70+15{,}8\cdot 75+17{,}8\cdot 82+19{,}5\cdot 89+20{,}4\cdot 95+$$
$$+21{,}5\cdot 100+22{,}2\cdot 105+24{,}3\cdot 110+25{,}3\cdot 115+26{,}5\cdot 120+28{,}1\cdot 125+$$
$$+30{,}1\cdot 130+35{,}2\cdot 135+36{,}4\cdot 140+37\cdot 145+38{,}5\cdot 150+39{,}5\cdot 155+$$
$$+40{,}5\cdot 160+41\cdot 165+42{,}5\cdot 170=735+1185+1459{,}6+1735{,}5+1938+$$
$$+2150+2331+2673+2909{,}5+3180+3512{,}5+3913+4752+5096+$$
$$+5365+5775+6122{,}5+6480+6765+7225=75302{,}6.$$

$$K_{xy}^*=\frac{\sum y_i x_i}{n}-\overline{x}\cdot\overline{y}=\frac{75302{,}6}{20}-121{,}8\cdot 28{,}63=$$

$$=3765{,}13-3487{,}13=278.$$

$$K_{xy}^*=278\,.$$

$$r_{\text{В}}=\frac{K_{xy}^*}{\sigma_x\sigma_y}=\frac{278}{29{,}89\cdot 9{,}4}=\frac{278}{280{,}966}=0{,}989\,.$$

$$r_{\text{В}}\approx 0{,}989\,.$$

Оскільки значення $r_{\text{В}}$ близьке до одиниці, то звідси випливає, що залежність між кількістю масла, споживаного певною особою, та її місячним прибутком майже функціональна.

6. Емпіричні моменти

Початкові емпіричні моменти. Середнє зважене значення варіант у степені k (k = 1, 2, 3,…) називають *початковим емпіричним моментом k-го порядку* ν_k^*, який обчислюється за формулою

$$\nu_k^*=\frac{\sum x_k\, n_i}{n}\,. \qquad (390)$$

При k = 1 дістанемо початковий момент першого порядку:

$$\nu_1^* = \frac{\sum x_i \; n_i}{n} = \bar{x}_{\text{В}} \,. \tag{391}$$

При k = 2 обчислимо початковий момент другого порядку:

$$\nu_2^* = \frac{\sum x_i^2 n_i}{n} \,. \tag{392}$$

Отже, дисперсію вибірки можна подати через початкові моменти першого та другого порядків, а саме:

$$D_{\text{В}} = \nu_2^* - (\nu_1^*)^2 \,. \tag{393}$$

Центральний емпіричний момент *k*-го порядку. Середнє зважене відхилення варіант у степені k (k = 1, 2, 3,…) називають центральним емпіричним моментом k-го порядку

$$\mu_k^* = \frac{\sum (x_i - \bar{x}_{\text{В}})^k n_i}{n} \,. \tag{394}$$

При k = 1 дістанемо:

$$\mu_1^* = \frac{\sum (x_i - \bar{x}_{\text{В}}) n_i}{n} = \frac{\sum x_i \, n_i}{n} - \bar{x}_{\text{В}} \cdot \frac{\sum n_i}{n} = \bar{x}_{\text{В}} - \bar{x}_{\text{В}} = 0 \,.$$

При k = 2 маємо:

$$\mu_2^* = \frac{\sum (x_i - \bar{x}_{\text{В}})^2 n_i}{n} = D_{\text{В}}.$$

На практиці найчастіше застосовуються центральні емпіричні моменти третього та четвертого порядків, що обчислюються за формулами:

$$\mu_3^* = \frac{\sum (x_i - \bar{x}_{\text{В}})^3 n_i}{n}, \tag{395}$$

$$\mu_4^* = \frac{\sum (x_i - \bar{x}_{\text{В}})^4 n_i}{n} \,. \tag{396}$$

Підносячи до третього та четвертого степеня відхилення варіант, подаємо μ_3^* та μ_4^* через відповідні початкові моменти:

$$\mu_3^* = \nu_3^* - 3\nu_2^* \cdot \nu_1^* + 2(\nu_1^*)^2, \tag{397}$$

$$\mu_4^* = \nu_4^* - 4\nu_3^* \cdot \nu_1^* + 6\nu_2^*(\nu_1^*)^2 - 3(\nu_1^*)^4. \tag{398}$$

Коефіцієнт асиметрії A_s^*. Центральний емпіричний момент третього порядку застосовується для обчислення коефіцієнта асиметрії:

$$A_s^* = \frac{\mu_3^*}{\sigma_{\text{В}}^3}. \qquad (399)$$

Якщо варіанти статистичного розподілу вибірки симетрично розміщені відносно $\bar{x}_{\text{В}}$, то в цьому разі $A_s = 0$, оскільки $\mu_3^* = 0$.

При $A_s < 0$ варіанти статистичного розподілу $x_i < \bar{x}_{\text{В}}$ переважають варіанти $x_i > \bar{x}_{\text{В}}$. Таку асиметрію називають *від'ємною*. При $A_s > 0$ варіанти $x_j > \bar{x}_{\text{В}}$ переважають варіанти $x_i < \bar{x}_{\text{В}}$, і таку асиметрію називають *додатною*.

Ексцес. Центральний емпіричний момент четвертого порядку застосовується для обчислення ексцесу:

$$E_s^* = \frac{\mu_4^*}{\sigma_{\text{В}}^4} - 3. \qquad (400)$$

E_s^*, як правило, використовується при дослідженні неперервних ознак генеральних сукупностей, оскільки він оцінює крутизну закону розподілу неперервної випадкової величини порівняно з нормальним. Для нормального закону розподілу, як відомо, $E_s^* = 0$.

Приклад. Оцінки в балах x_i, одержані абітурієнтами на вступних іспитах з математики, наведені у вигляді дискретного розподілу:

x_i	15	25	35	45	55	65	75	85
n_i	5	10	15	20	25	15	8	2

Обчислити A_s^*.

Розв'язання. Використовуючи наведені вище формули і враховуючи, що $n = \sum n_i = 100$, обчислимо

$$\bar{x}_{\text{В}} = \frac{\sum x_i n_i}{n} = \frac{15\cdot 5 + 25\cdot 10 + 35\cdot 15 + 45\cdot 20 + 55\cdot 25 + 65\cdot 15 + 75\cdot 8 +}{100}$$

$$\frac{+85\cdot 2}{100} = \frac{75 + 250 + 525 + 900 + 1375 + 975 + 600 + 170}{100} = \frac{4870}{100} = 48{,}7.$$

Звідси $\bar{x}_{\text{В}} = 48{,}7$.

$$\mu_3^* = \frac{\sum (x_i - \bar{x}_{\text{В}})^3 n_i}{n} =$$

$$= \frac{(15-48{,}7)^3 \cdot 5 + (25-48{,}7)^3 \cdot 10 + (35-48{,}7)^3 \cdot 15 + (45-48{,}7)^3 \cdot 20 +}{100}$$

$$\frac{+(55-48{,}7)^3 \cdot 25 + (65-48{,}7)^3 \cdot 15 + (75-48{,}7)^3 \cdot 8 + (85-48{,}7)^3 \cdot 2}{100} =$$

$$= \frac{-191363{,}765 - 133120{,}53 - 38570{,}295 - 1013{,}06 + 6251{,}175 +}{100}$$

$$\frac{+64961{,}205 + 145531{,}576 + 95664{,}294}{100} = -516{,}594.$$

$$\frac{\sum x_i^2 n_i}{n} = \frac{(15)^2 \cdot 5 + (25)^2 \cdot 10 + (35)^2 \cdot 15 + (45)^2 \cdot 20 + (55)^2 \cdot 25 +}{100}$$

$$\frac{+(65)^2 \cdot 15 + (75)^2 \cdot 8 + (85)^2 \cdot 2}{100} = \frac{1125 + 6250 + 18375 + 40500 +}{100}$$

$$\frac{+75625 + 63375 + 45000 + 14450}{100} = \frac{264700}{100} = 2647.$$

$$D_{\text{В}} = \frac{\sum x_i^2 n_i}{n} - (\bar{x}_{\text{В}})^2 = 2647 - (48{,}7)^2 = 2647 - 2371{,}69 = 275{,}31.$$

Остаточно маємо: $D_{\text{В}} = 275{,}31$.

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}} = \sqrt{275{,}31} = 16{,}59.$$

$$A_s = \frac{\mu_3}{\sigma_{\text{В}}^3} = -\frac{516{,}594}{(16{,}59)^3} = -\frac{516{,}594}{4566{,}034} = -0{,}11.$$

Отже, дістанемо: $A_s = -0{,}11$.

Оскільки A_s порівняно малий, то статистичний розподіл близький до симетричного.

Приклад. Довжина заготівок x_i, виготовлених робітником за зміну, та частоти цих довжин n_i наведені у вигляді статистичного розподілу:

x_i, мм	6,5	8,5	10,5	12,5	14,5	16,5
n_i	4	16	20	30	24	6

Визначити E_s^*.

Розв'язання. Обчислюється значення $\bar{x}_{\text{В}}, \sigma_{\text{В}}$, $\bar{x}_{\text{В}} = \dfrac{\sum x_i\ n_i}{n}$. Оскільки $n = \sum n_i = 100$, то дістанемо:

$$\bar{x}_{\text{В}} = \frac{6{,}5 \cdot 4 + 8{,}5 \cdot 16 + 10{,}5 \cdot 20 + 12{,}5 \cdot 30 + 14{,}5 \cdot 24 + 16{,}5 \cdot 6}{100} =$$

$$= \frac{26 + 136 + 210 + 375 + 348 + 99}{100} = \frac{1194}{100} = 11{,}94.$$

Отже, $\bar{x}_{\text{В}} = 11{,}94$.

$$\frac{\sum x_i^2\, n_i}{n} = \frac{(6{,}5)^2 \cdot 4 + (8{,}5)^2 \cdot 16 + (10{,}5)^2 \cdot 20 + (12{,}5)^2 \cdot 30 + (14{,}5)^2 \cdot 24 +}{100}$$

$$\frac{+(16{,}5)^2 \cdot 6}{100} = \frac{169 + 1156 + 2205 + 4687{,}5 + 5046 + 1633{,}5}{100} = \frac{14897}{100} = 148{,}97.$$

$$D_{\text{В}} = \frac{\sum x_i^2\, n_i}{n} - (\bar{x}_{\text{В}})^2 = 148{,}97 - (11{,}94)^2 = 148{,}97 - 142{,}564 = 6{,}406.$$

$$D_{\text{В}} = 6{,}406.$$

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}} = \sqrt{6{,}406} \approx 2{,}53.$$

$$D_{\text{В}} = 2{,}53.$$

Обчислимо центральний емпіричний момент четвертого порядку.

$$\mu_4^* = \frac{\sum (x_i - \bar{x}_{\text{В}})^4 n_i}{n} = \frac{(6{,}5 - 11{,}94)^4 \cdot 4 + (8{,}5 - 11{,}94)^4 \cdot 16 +}{100}$$

$$\frac{+(10{,}5 - 11{,}94)^4 \cdot 20 + (12{,}5 - 11{,}94)^4 \cdot 30 + (14{,}5 - 11{,}94)^4 \cdot 24 +}{100}$$

$$\frac{+(16{,}5 - 11{,}94)^4 \cdot 6}{100} = \frac{3503{,}125 + 2240{,}55 + 85{,}996 + 2{,}95 + 1030{,}79 +}{100}$$

$$\frac{+2594{,}24}{100} = \frac{9457{,}651}{100} = 94{,}58.$$

$$E_S^* = \frac{\mu_4^*}{\sigma_{\text{В}}^4} - 3 = \frac{94{,}58}{(2{,}53)^4} - 3 = \frac{94{,}58}{40{,}9715} - 3 = 2{,}308 - 3 = -0{,}692.$$

$$E_S^* = -0{,}692.$$

Оскільки $E_S^* < 0$, то вершина закону розподілу випадкової величини, заданого щільністю ймовірностей, буде плоскою, тобто це так званий *туповершинний розподіл*.

Теоретичні запитання до теми ?

1. Дати визначення генеральної та вибіркової сукупності.
2. Що називається варіантою, варіаційним рядом?
3. Що таке частота, відносна частота варіант?
4. Дати визначення дискретного статистичного розподілу вибірки.
5. $\bar{x}_{\text{В}}$, $D_{\text{В}}$, $\sigma_{\text{В}}$ для дискретного статистичного розподілу вибірки.
6. Що таке медіана, мода дискретного статистичного розподілу?
7. Що називається емпіричною функцією (комулятою)?
8. Властивість $F^*(x)$.
9. Що називається інтервальним статистичним розподілом вибірки?
10. $\bar{x}_{\text{В}}$, $D_{\text{В}}$, $\sigma_{\text{В}}$ для інтервального статистичного розподілу.
11. Як визначається Me^* для інтервального статистичного розподілу?
12. Як визначається Mo^* для інтервального статистичного розподілу?
13. Що являє собою полігон частот і відносних частот?
14. Що називається гістограмою частот і відносних частот?
15. Що таке початковий момент *k*-го порядку?
16. Що таке центральний момент *k*-го порядку?
17. Асиметрія і ексцес статистичного розподілу вибірки.
18. Що називається розмахом, коефіцієнтом варіації?
19. $F^*(x)$ для інтервального статистичного розподілу вибірки.
20. Що називається двовимірним статистичним розподілом вибірки?
21. Формули для обчислення основних числових характеристик ознак *X* і *Y* для двовимірного статистичного розподілу вибірки.
22. Емпіричний кореляційний момент K^*_{xy} та його властивості.
23. Вибірковий коефіцієнт кореляції $r_{\text{В}}$ та його властивості.
24. Що називається умовним статистичним розподілом $Y/X = y_i$?
25. Умовні числові характеристики для умовного статистичного розподілу $X/Y = y_i$, $Y/X = x_i$.
26. Що називається умовним статистичним розподілом $X/Y = y_i$?
27. Умовні числові характеристики для умовного статистичного розподілу $X/Y = y_i$.

Задачі до теми

1. При вивченні випадкової величини X у результаті 40 незалежних спостережень дістали вибірку:

10, 13, 10, 9, 9, 12, 12, 6, 7, 9, 8, 9, 11, 9, 14, 13, 9, 8, 8, 7, 10, 10, 11, 11, 11, 12, 8, 7, 9, 10, 14, 13, 8, 8, 9, 10, 11, 11, 12, 12.

Потрібно:

1. Побудувати дискретний статистичний розподіл для цієї вибірки, а також полігон частот і $F^*(x)$.

2. Обчислити $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$, R, V.

3. Знайти Mo^*, Me^*.

Відповідь. $\bar{x}_{\text{В}} = 10$, $\sigma_{\text{В}} = 2$.

2. П'ятдесят абітурієнтів на вступних іспитах з інформатики дістали таку кількість балів:

12, 14, 19, 15, 14, 18, 13, 16, 17, 12,
20, 17, 15, 13, 17, 16, 20, 14, 14, 13,
17, 16, 15, 19, 16, 15, 18, 17, 15, 14,
16, 15, 15, 18, 15, 15, 19, 14, 16, 18,
18, 15, 15, 17, 15, 16, 16, 14, 14, 17.

Потрібно:

1. Побудувати дискретний статистичний розподіл. Полігон частот, $F^*(x)$.

2. Обчислити $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$, R, V.

3. Знайти Mo^*, Me^*.

Відповідь. $\bar{x}_{\text{В}} = 15{,}78$, $\sigma_{\text{В}} = 1{,}93$.

3. Через кожну годину вимірювалась напруга в електромережі. Результати вимірювання напруги у вольтах наведені у вигляді статистичного ряду:

222, 219, 224, 220, 218, 217, 221, 220, 215, 218, 223, 225,
220, 226, 221, 216, 211, 219, 220, 221, 222, 218, 221, 219.

Потрібно:

1. Побудувати дискретний статистичний розподіл, полігон частот і $F^*(x)$.

2. Обчислити $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$, R, V.

3. Знайти Mo^*, Me^*.

Відповідь. $\bar{x}_{\text{В}} = 220{,}25$, $\sigma_{\text{В}} = 2{,}66$.

4. З допомогою радіодальноміра було здійснено 16 вимірювань однієї і тієї самої відстані. Результати вимірювання в метрах наведені у вигляді сатистичного ряду:

201, 195, 207, 203, 191, 208, 198, 210, 204, 192, 195, 211, 206, 196, 208, 197.

Потрібно:

1. Побудувати дискретний статистичний розподіл, полігон відносних частот і $F^*(x)$.

2. Обчислити $\bar{x}_В$, $\sigma_В$, R, V.

3. Знайти Mo^*, Me^*.

Відповідь. $\bar{x}_В = 201$ м, $\sigma_В = 13{,}85$ м.

5. Для обчислення середньої врожайності озимої пшениці x_i кооперативне поле площею 2000 га було поділено на 20 рівних ділянок n_i. Фактичний урожай на кожній ділянці наведено в таблиці:

x_i, ц/га	25	30	35	40	45
n_i	2	3	8	4	3

Потрібно:

1. Побудувати полігон відносних частот і комуляту.

2. Обчислити $\bar{x}_В$, $\sigma_В$, R, V.

3. Знайти Mo^*, Me^*.

Відповідь. $\bar{x}_В = 35{,}75$ ц/га, $\sigma_В = 5{,}76$ ц/га.

6. На кожну сотню деталей, що їх виготовляє цех, у середньому припадає дві браковані. Було перевірено 10 партій по 100 деталей у кожній. Відхилення кількості виявлених бракованих деталей від середнього x_i наведено в таблиці:

Номер партії	1	2	3	4	5	6	7	8	9	10
x_i	–1	0	1	1	–1	1	0	–2	2	1

Потрібно:

1. Побудувати дискретний статистичний розподіл, полігон частот і $F^*(x)$.

2. Обчислити $\bar{x}_В$, $\sigma_В$, R, V.

3. Знайти Mo^*, Me^*.

Відповідь. $\bar{x}_В = 0{,}2$, $\sigma_В = 1{,}233$.

7. Результати вимірювання максимальної місткості x_i 20-ти конденсаторів у пікофарадах наведено в таблиці:

Номер конденсатора	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_i, пФ	4,4	4,31	4,4	4,4	4,65	4,56	4,71	4,54	4,36	4,56	4,31	4,42	4,6	4,35	4,5	4,4	4,43	4,48	4,42	4,45

Потрібно:

1. Побудувати дискретний статистичний розподіл, полігон відносних частот і $F^*(x)$.

2. Обчислити $\bar{x}_В$, $\sigma_В$, R, V.

3. Знайти Мо*, Ме*.

Відповідь. $\bar{x}_В = 4{,}47$ пФ, $\sigma_В = 1{,}108$ пФ.

8. Із партії однотипних амперметрів N для контролю відібрано 10 штук. Вимірювання показали такі відхилення x_i від номіналу в міліамперах:

Номер партії	1	2	3	4	5	6	7	8	9	10
x_i, мА	1	3	–2	2	4	2	5	3	–2	4

Потрібно:

1. Побудувати дискретний статистичний розподіл, полігон частот і $F^*(x)$.

2. Обчислити $\bar{x}_В$, $\sigma_В$, R, V.

3. Знайти Мо*, Ме*.

Відповідь. $\bar{x}_В = 2$ мА, $\sigma_В = 2{,}23$ мА.

9. 200 однотипних деталей n_i після шліфування були піддані контрольним вимірюванням, результати яких наведено в таблиці x_i:

x_i, мм	3,7	3,8	3,9	4	4,1	4,2	4,3	4,4
n_i	1	22	40	79	27	26	4	1

Потрібно:

1. Побудувати полігон відносних частот, комуляту, $F^*(x)$.

2. Обчислити $\bar{x}_В$, $\sigma_В$, R, V;

3. Знайти Мо*, Ме*.

Відповідь. $\bar{x}_В = 14{,}34$, $\sigma_В = 0{,}039$.

10. Із партії сталевих кілець, виготовлених заводом, була здійснена вибірка обсягом 200, і результати вимірювання товщини цих кілець x_i наведено у вигляді дискретного статистичного розподілу:

x_i, мм	14,41	14,43	14,45	14,47	14,49	14,51	14,53	14,55	14,57	14,59	14,61	14,63
n_i	2	2	8	9	9	14	41	76	21	11	4	3

Потрібно:

1. Побудувати полігон відносних частот і комуляту.

2. Обчислити $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$, R, V.

3. Знайти Мо^*, Ме^*.

Відповідь. $\bar{x}_{\text{В}} = 14{,}34$; $\sigma_{\text{В}} = 0{,}039$.

11. У відділі технічного контролю було виміряно діаметри 200 валиків із партії, виготовленої одним верстатом-автоматом. Відхилення виміряних діаметрів від номіналу наведено як інтервальний статистичний розподіл, де $X = x_i$ вимірюється в мікронах:

h = 5, мк	–20…–15	–15…–10	–10…–5	–5…0	0…5	5…10	10…15	15…20	20…25	25…30
n_i	7	11	15	24	49	41	26	17	7	3

Потрібно:

1. Побудувати гістограму частот і $F^*(x)$.

2. Обчислити $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$, A_s^*, E_s^*, Мо^*, Ме^*.

Відповідь. $\bar{x}_{\text{В}} = 4{,}3$ мк; $\sigma_{\text{В}} = 9{,}79$ мк; $A_s^* = -0{,}128$ мк; $E_s^* = -0{,}16$; $\text{Мо}^* = 3{,}79$ мк; $\text{Ме}^* = -1{,}46$ мк.

12. Із партії однотипних сталевих болтів, виготовлених заводом, була здійснена вибірка обсягом 200 шт. і результати вимірювання їх діаметрів x_i наведено у вигляді інтервального статистичного розподілу:

x, мм h = 2 мм	14,40—14,42	14,42—14,44	14,44—14,46	14,46—14,48	14,48—14,50	14,50—14,52	14,52—14,54	14,54—14,56	14,56—14,58	14,58—14,60	14,60—14,62	14,62—14,64
n_i	2	2	8	9	9	14	41	76	21	11	4	3

Потрібно:

1. Побудувати гістограму частот і $F^*(x)$.

2. Обчислити $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$, E_S^*, A_s^*, Мо^*, Ме^*.

Відповідь. $\bar{x}_{\text{В}} = 14{,}34$ мм; $\sigma_{\text{В}} = 0{,}039$ мм; $A_s^* = 0{,}311$ мм; $E_s^* = 1{,}549$ мм, $\text{Мо}^* = 15{,}34$ мм, $\text{Ме}^* = 15{,}39$ мм..

13. Після шліфування 200 однотипних деталей були піддані контрольним вимірюванням, результати яких наведено у формі інтервального статистичного розподілу:

$h = 1$	3,65—3,75	3,75—3,85	3,85—3,95	3,95—4,05	4,05—4,15	4,15—4,25	4,25—4,35	4,35—4,45
n_i	1	22	40	79	27	26	4	1

Потрібно:

1. Побудувати гістограму відносних частот і $F^*(x)$.

2. Обчислити $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$, E_s^*, A_s^*, Мо^*, Ме^*.

Відповідь. $\bar{x}_{\text{В}} = 4{,}004$ мм; $\sigma_{\text{В}} = 0{,}126$ мм; $A_s^* = 0{,}311$ мм; $E_s^* = -0{,}117$ мм, $\text{Мо}^* = 4{,}38$ мм, $\text{Ме}^* = 4{,}875$ мм.

14. У 100 осіб було виміряно зріст x_i. Результати вимірювання наведено як інтервальний статистичний розподіл:

x_i, см $h = 4$ см	168—172	172—176	176—180	180—184	184—188	192—196	196—166
n_i	10	20	30	25	10	3	2

Потрібно:

1. Побудувати гістограму частот і $F^*(x)$.

2. Обчислити $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$, E_s^*, A_s^*, Мо^*, Ме^*.

Відповідь. $\bar{x}_{\text{В}} = 178{,}88$ см, $\sigma_{\text{В}} = 98{,}87$ см, $A_s^* = 0{,}0063$ см, $E_s^* = -2{,}9999$ см, $\text{Мо}^* = 178{,}7$ см, $\text{Ме}^* = 178{,}6$ см.

15. Для дослідження розподілу маси новонароджених x_i була зібрана інформація про 100 дітей. Ця інформація подана як інтервальний статистичний розподіл, що має такий вигляд у табличній формі:

x_i, кг $h = 0{,}5$	1,0—1,5	1,5—2,0	2,0—2,5	2,5—3,0	3,0—3,5	3,5—4,0	4,0—4,5	4,5—5,0
n_i	2	8	10	30	40	6	3	1

Потрібно:

1. Побудувати гістограму відносних частот і $F^*(x)$.

2. Обчислити $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$, E_s^*, A_s^*, Мо^*, Ме^*.

Відповідь. $\bar{x}_{\text{В}} = 2{,}915$ кг, $\sigma_{\text{В}} = 0{,}625$ кг, $A_s^* = -0{,}26$, $E_s^* = 0{,}73$, $\text{Мо}^* = 3{,}11$ кг, $\text{Ме}^* = 3$ кг.

16. Число розладнань у роботі верстатів-автоматів заводу протягом року наведено у вигляді інтервального статистичного розподілу:

x_i $h = 2$	0—2	2—4	4—6	6—8	8—10	10—12	12—14	14—16	16—18	18—20	20—22	22—24
n_i	2	5	7	11	15	18	26	20	14	10	6	3

Потрібно:

1. Побудувати гістограму частот і комуляту $F^*(x)$.

2. Обчислити $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$, E_s^*, A_s^*, Mo^*, Me^*.

Відповідь. $\bar{x}_{\text{В}} = 12{,}58$, $\sigma_{\text{В}} = 4{,}88$, $A_s^* = -0{,}13$, $E_s^* = -0{,}39$, $\text{Mo}^* = 13$, $\text{Me}^* = 11$.

17. Відсоток виконання плану підприємства за рік та кількість підприємств, що виконують цей план, наведено у вигляді інтервального статистичного розподілу:

x_i, % $h = 10$ %	10—20	20—30	30—40	40—50	50—60	60—70	70—80	80—90	90—100	100—110	110—120
n_i	2	6	13	16	25	12	10	8	5	3	1

Потрібно:

1. Побудувати гістограму відносних частот і комуляту $F^*(x)$.

2. Обчислити $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$, E_s^*, A_s^*, Mo^*, Me^*.

Відповідь. $\bar{x}_{\text{В}} = 58{,}2$ %, $\sigma_{\text{В}} = 21{,}21$ %, $A_s^* = 0{,}0043$ %, $E_s^* = -0{,}22$ %, $\text{Mo}^* = 54{,}1$ %, $\text{Me}^* = 57{,}2$ %.

18. Кількість виготовленого полотна за зміну x_i ткачами наведено у формі інтервального статистичного розподілу:

x_i, м $h = 10$	52,8—62,8	62,8—72,8	72,8—82,8	82,8—92,8	92,8—102,8	102,8—112,8	112,8—122,8	122,8—132,8	132,8—142,8
n_i	8	12	25	39	26	18	12	9	4

Потрібно:

1. Побудувати гістограму відносних частот і комуляту $F^*(x)$.

2. Обчислити $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$, E_s^*, A_s^*, Mo^*, Me^*.

Відповідь. $\bar{x}_{\text{В}} = 93{,}13$ м; $\sigma_{\text{В}} = 19{,}1$ м; $E_s^* = 0{,}38$; $A_s^* = 0{,}31$; $\text{Mo}^* = 88$ м; $\text{Me}^* = 84{,}12$ м.

19. Рівень води x_i в річці відносно номіналу вимірювався протягом 50-ти років навесні. Результати вимірювання наведено у формі інтервального статистичного розподілу.

x_i, см $h = 24$	0—24	24—48	48—72	72—96	96—120	120—144	144—168	168—192	192—216
n_i	0	2	4	6	12	16	6	3	1

Потрібно:

1. Побудувати гістограму частот і комуляту $F^*(x)$.

2. Обчислити $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$, E_s^*, Δ_s^*, Mo^*.

Відповідь. $\bar{x}_{\text{В}} = 118{,}1$ см, $\sigma_{\text{В}} = 36{,}2$ см, $\Delta_s^* = 0{,}38$, $E_s^* = -0{,}12$, $\text{Mo}^* = 126{,}86$.

20. Залежність урожайності x_i ярової пшениці від кількості опадів протягом весни зображено у формі інтервального статистичного розподілу:

x_i, ц/га $h = 2$	4,2—6,2	6,2—8,2	8,2—10,2	10,2—12,2	12,2—14,2	14,2—16,2	16,2—18,2	18,2—20,2	20,2—22,2
n_i	5	15	20	25	30	18	8	2	1

Потрібно:

1. Побудувати гістограму частот і комуляту $F^*(x)$.

2. Обчислити $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$, E_s^*, A_s^*, Mo^*, Me^*.

Відповідь. $\bar{x}_{\text{В}} = 11{,}9$ ц/га, $\sigma_{\text{В}} = 40{,}8$ ц/га, $A_s^* = 0{,}0002$, $E_s^* = -2{,}9999$, $\text{Mo}^* = 12{,}79$ ц/га, $\text{Me}^* = 11$ ц/га.

21. Залежність річної заробітної плати Y від загального виробітку X показано у вигляді двовимірного статистичного розподілу:

Y	X					
	1,5	2,5	3,5	4,5	5,5	n_{yi}
0,82	1	3	—		—	
0,86	—	3	2	1	—	
0,9	—	2	5	9	3	
0,94	—	—	—	6	4	
0,98	—	—	—	—	2	
n_{xj}						

Обчислити $r_{\text{В}}$, $\bar{y}_x = 4{,}5$, $\bar{x}_y = 0{,}80$.

Відповідь. $r_{\text{В}} = 0{,}783$; $\bar{y}_{x=4{,}5} = 0{,}913$; $\bar{x}_{y=0{,}86} = 3{,}17$.

22. Зі старших класів ліцею було відібрано групу учнів. Дані про їх середньорічні оцінки з математики x_i та решти дисциплін n_i (за стобальною системою) наведено в таблиці:

n_i	45	30	48	50	52	54	51	60	62	63	65	70	71	74	76	68	79	85
x_i	30	35	40	44	48	55	52	65	69	72	78	82	84	86	90	91	92	95

Обчислити K_{xy}, $r_{\text{В}}$.

Відповідь. $K_{xy} = 252{,}62$; $r_{\text{В}} = 0{,}903$.

23. Виготовлені в цеху втулки сортувалися за відхиленням внутрішнього діаметра X і зовнішнього Y. Спільний статистичний розподіл ознак X і Y наведено в таблиці:

$X = x_j$, мм	$Y = y_i$, мм				
	0,002	0,004	0,006	0,008	n_{yi}
0,01	1	3	4	2	
0,02	2	2	24	10	
0,03	4	15	8	3	
0,04	4	6	8	2	
n_{xj}					

Обчислити $r_{\text{В}}$, $\bar{y}_{x=0{,}03}$, $\bar{x}_{y=0{,}04}$.

Відповідь. $r_{\text{В}} = 0{,}141$; $\bar{y}_{x=0{,}03} = 0{,}0047$ мм; $\bar{x}_{y=0{,}04} = 0{,}029$ мм.

24. Залежність річної продуктивності праці в розрахунку на одного робітника Y від енергомісткості праці на підприємстві певної галузі показано в таблиці:

y_i, тис. грн/робітн.	11,0	11,6	12,1	12,7	13,2	13,9	14,1	14,6	14,9	15,4
x_i, кВт/робітн.	5,2	5,8	5,9	6,2	6,9	7,2	7,5	8,5	8,8	9,4

Продовження табл.

y_i, тис. грн/робітн.	11,0	11,6	12,1	12,7	13,2	13,9	14,1	14,6	14,9	15,4
x_i, кВт/робітн.	5,2	5,8	5,9	6,2	6,9	7,2	7,5	8,5	8,8	9,4

Обчислити K_{xy}, $r_{\text{В}}$.

Відповідь. $K_{xy} = 6{,}945$; $r_{\text{В}} = 0{,}681$.

25. При аналізі руди дістали такі дані про відсотковий вміст у ній свинцю та срібла. Результати аналізу наведено в таблиці:

$Y = y_j$	$X = x_j$									
	2,5	7,5	12,5	17,5	22,5	27,5	32,5	37,5	42,5	n_{yi}
2	119	9	—	—	—	—	—	—	—	
6	9	59	7	—	—	—	—	—	—	
10	1	4	28	3	—	—	—	—	—	
14	—	—	8	12	4	—	—	—	—	
18	—	—	1	6	7	1	1	—	—	
22	—	—	—	1	1	8	3	—	—	
26	—	—	—	—	—	2	1	—	—	
30	—	—	—	—	—	—	3	2	1	
34	—	—	—	—	—	—	—	—	—	
38	—	—	—	—	—	—	—	—	1	
n_{xj}										

Обчислити $r_{\text{В}}$, $\bar{y}_{x=12,5}$; $\bar{x}_{y=14}$.

Відповідь. $r_{\text{В}} = 0{,}865$; $\bar{y}_{x=12,5} = 3{,}32\,\%$; $\bar{x}_{y=14} = 50\,\%$.

26. Залежність урожайності озимої пшениці y_i від кількості внесених добрив x_i показано в таблиці:

y_i, ц/га	10	12	14	16	18	20	22	24	26	28	30	32	34
x_i, кг/га	10	30	40	50	60	70	80	90	100	110	120	130	140

Обчислити K_{xy}, $r_{\text{В}}$.

Відповідь. $K_{xy} = 289{,}23$, $r_{\text{В}} = 0{,}998$.

27. Залежність міцності бавовняного волокна від його довжини відображено у вигляді двовимірного статистичного розподілу вибірки (в умовних одиницях):

$Y = y_j$	$X = x_j$								
	4100	4300	4500	4700	4900	5100	5300	5500	n_{yi}
6,75	—	2	—	—	—	—	—	—	2
6,25	1	4	4	2	—	—	—	—	11
5,75	—	2	5	6	8	2	3	—	26
5,25	—	3	8	10	2	1	—	—	24
4,75	—	—	4	5	5	3	2	1	20
4,25	—	—	—	—	—	—	1	1	3
3,75	—	—	—	—	—	—	1	1	3
n_{xj}	1	11	21	23	16	7	7	3	89

Обчислити $r_{\text{В}}$, $\bar{y}_{x=4300}$, $\bar{x}_{y=6,25}$.

Відповідь. $r_{\text{В}} = -0,62$, $\bar{y}_{x=4300} = 5,98$; $\bar{x}_{y=6,25} = 4427,3$.

28. Результати проведеного аналізу залежності кількості проданих пар чоловічого взуття y_i від його розміру x_i наведено в таблиці:

y_i	25	38	65	95	120	140	152	160	165	175	180	185	190	200
x_i	45	43	42	41	40	39	38,5	39	37,5	37	36,5	36	35,5	35

Обчислити K_{xy}, $r_{\text{В}}$.

Відповідь. $K_{xy} = -157,43$, $r_{\text{В}} = -0,98$.

29. Залежність урожайності ячменю y_i від кількості внесених добрив на 1 га x_i наведено у вигляді двовимірного статистичного розподілу вибірки:

$Y = y_i$, ц/га	$X = x_j$, га/т					
	0,5	1	1,5	2	2,5	n_{yi}
15,5	1	2	—	—	—	
16,5	2	4	1	—	—	
17,5	—	3	6	1	—	
18,5	—	—	4	1	1	
19,5	—	—	1	2	1	
n_{xj}						

Обчислити $r_{\text{В}}$, $\bar{y}_{x=1,5}$, $\bar{x}_{y=16,5}$.

Відповідь. $\bar{y}_{x=1,5} = 17{,}83$; $\bar{x}_{y=16,5} = 2{,}29$.

30. Результати вимірювання чутливості відео y_i та звукового каналів x_i наведено в таблиці:

y_i	250	200	180	160	140	110	100	95	90
x_i	180	230	240	250	300	320	330	340	350

Продовження табл.

y_i	85	80	75	80	70	65	60	55
x_i	360	370	380	390	400	410	420	430

Обчислити K_{xy}, $r_{\text{В}}$.

Відповідь. $K_{xy} = -3456{,}9$, $r_{\text{В}} = -0{,}97$.

РОЗДІЛ VI

СТАТИСТИЧНІ ОЦІНКИ. СТАТИСТИЧНІ ГІПОТЕЗИ

ТЕМА 13. СТАТИСТИЧНІ ОЦІНКИ ПАРАМЕТРІВ ГЕНЕРАЛЬНОЇ СУКУПНОСТІ

1. Загальна інформація

Інформація, яку дістали на основі обробки вибірки про ознаку генеральної сукупності, завжди міститиме певні похибки, оскільки вибірка становить лише незначну частину від неї ($n < N$), тобто обсяг вибірки значно менший від обсягу генеральної сукупності.

Тому слід організувати вибірку так, щоб ця інформація була найбільш повною (вибірка має бути репрезентативною) і забезпечувала з найбільшим ступенем довіри про параметри генеральної сукупності або закон розподілу її ознаки.

Параметри генеральної сукупності $M(x) = \overline{X}_{\text{Г}}, D_{\text{Г}}, \sigma_{\text{Г}}, \text{Mo}, \text{Me}, r_{xy}$ є величинами сталими, але їх числове значення невідоме. Ці параметри оцінюються параметрами вибірки: $\overline{x}_{\text{В}}, D_{\text{В}}, \sigma_{\text{В}}, \text{Mo}^*, \text{Me}^*, r_{\text{В}}$, які дістають при обробці вибірки. Вони є величинами непередбачуваними, тобто випадковими. Схематично це можна показати так (рис. 115).

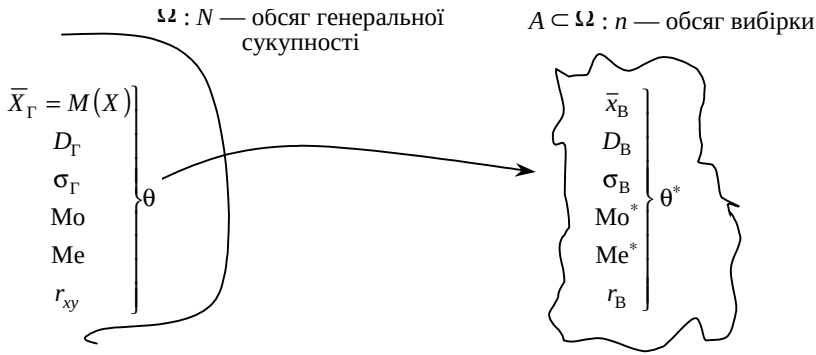


Рис. 115

Тут через θ позначено оцінювальний параметр генеральної сукупності, а через θ^* — його статистичну оцінку, яку називають ще *статистикою*. При цьому θ = const, а θ^* — випадкова величина, що має певний закон розподілу ймовірностей. Зауважимо, що до реалізації вибірки кожну її варіанту розглядають як випадкову величину, що має закон розподілу ймовірностей ознаки генеральної сукупності з відповідними числовими характеристиками:

$$M(x_i) = \overline{X}_\Gamma = M(x), \quad D(x_i) = D_\Gamma, \quad \sigma(x_i) = \sigma_\Gamma.$$

2. Точкові статистичні оцінки параметрів генеральної сукупності

Статистична оцінка θ^*, яка визначається одним числом, точкою, називається *точковою*. Беручи до уваги, що θ^* є випадковою величиною, точкова статистична оцінка може бути зміщеною і незміщеною: коли математичне сподівання цієї оцінки точно дорівнює оцінювальному параметру θ, а саме:

$$M\left(\theta^*\right) = \theta, \tag{401}$$

то θ^* називається *незміщеною*; в противному разі, тобто коли

$$M\left(\theta^*\right) \neq \theta, \tag{402}$$

точкова статистична оцінка θ^* називається *зміщеною відносно параметра генеральної сукупності* θ.

Різниця

$$\theta^* - \theta = \delta \tag{403}$$

називається *зміщенням статистичної оцінки* θ^*.

Оцінювальний параметр може мати кілька точкових незміщених статистичних оцінок, що можна зобразити так (рис. 116):

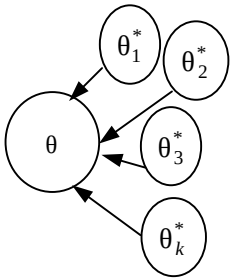


Рис. 116

Наприклад, нехай $\theta = M(X)$, яка має дві незміщені точкові статистичні оцінки — θ_1^* і θ_2^*. Тоді щільності ймовірностей для θ_1^*, θ_2^* матимуть такий вигляд (рис. 117):

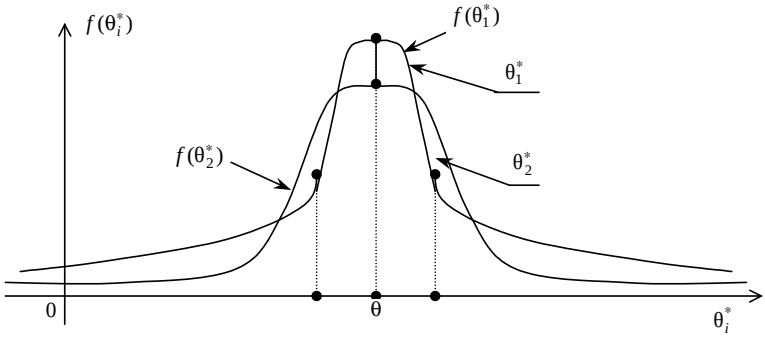


Рис. 117

Із графіків щільностей бачимо, що оцінка θ_1^* порівняно з оцінкою θ_2^* має ту перевагу, що в малому околі параметра θ, $f(\theta_1^*) > f(\theta_2^*)$. Звідси випливає, що оцінка θ_1^* частіше набуватиме значення в цьому околі, ніж оцінка θ_2^*.

Але на «хвостах» розподілів маємо іншу картину: більші відхилення від θ будуть спостерігатися для статистичної оцінки θ_1^* частіше, ніж для θ_2^*. Тому, порівнюючи дисперсії статистичних оцінок θ_1^*, θ_2^* як міру розсіювання, бачимо, що θ_2^* має меншу дисперсію, ніж оцінка θ_1^*.

Точкова статистична оцінка називається *ефективною*, коли при заданому обсязі вибірки вона має мінімальну дисперсію. Отже, оцінка θ_2^* буде незміщеною й ефективною.

Точкова статистична оцінка називається *ґрунтовною*, якщо у разі необмеженого збільшення обсягу вибірки θ^* наближається до оцінювального параметра θ, а саме:

$$\lim_{n\to\infty} P\left(\left|\theta^* - \theta\right| < \delta\right) = 1. \qquad (404)$$

3. Методи визначення точкових статистичних оцінок

Існують три методи визначення точкових статистичних оцінок для параметрів генеральної сукупності.

Метод аналогій. Цей метод базується на тому, що для параметрів генеральної сукупності вибирають такі самі параметри

вибірки, тобто для оцінки $\overline{X}_{\Gamma} = M(X)$, D_{Γ} вибирають аналогічні статистики — $\overline{x}_{\text{В}}$, $D_{\text{В}}$.

Метод найменших квадратів. Згідно з цим методом статистичні оцінки визначаються з умови мінімізації суми квадратів відхилень варіант вибірки від статистичної оцінки θ^*.

Отож, використовуючи метод найменших квадратів, можна, наприклад, визначити статистичну оцінку для $\overline{X}_{\Gamma} = M(X)$. Для цього скористаємося функцією $u = \sum_{i=0}^{n} \left(x_i - \theta^*\right)^2 n_i$. Використовуючи умову екстремуму, дістанемо:

$$\frac{\partial u}{\partial \theta^*} = -2\sum_{i=0}^{n} \left(x_i - \theta^*\right) n_i = 0 \rightarrow$$

$$\rightarrow \sum_{i=1}^{n} x_i n_i - \sum_{i=1}^{n} n_i \theta^* = 0 \rightarrow \theta^* = \frac{\sum_{i=1}^{n} x_i n_i}{n} = \overline{x}_{\text{В}}.$$

Звідси для $\theta = \overline{X}_{\Gamma}$ точковою статистичною оцінкою буде $\theta^* = \overline{x}_{\text{В}}$ — вибіркова середня.

Метод максимальної правдоподібності. Цей метод посідає центральне місце в теорії статистичного оцінювання параметрів θ. На нього свого часу звертав увагу К. Гаусс, а розробив його Р. Фішер. Цей метод розглянемо докладніше.

Нехай ознака генеральної сукупності X визначається лише одним параметром θ і має щільність імовірностей $f(x; \theta)$. У разі реалізації вибірки з варіантами $x_1, x_2, ..., x_n$ щільність імовірностей вибірки буде такою:

$$f\left(x_1, x_2, ..., x_n, \theta^*\right) = f\left(x_1, \theta^*\right) \cdot f\left(x_2, \theta^*\right) \cdot ... \cdot f\left(x_n, \theta^*\right). \qquad (405)$$

При цьому варіанти розглядаються як незалежні випадкові величини, котрі мають один і той самий закон розподілу, що й ознака генеральної сукупності X.

Суть цього методу полягає в тому, що, фіксуючи значення варіант $x_1, x_2, ..., x_n$, визначають таке значення параметра θ^*, при якому функція (405) максимізується. Вона називається *функцією максимальної правдоподібності* і позначається так: $L = L\left(\theta^*\right)$.

Наприклад, коли ознака генеральної сукупності X має нормальний закон розподілу, то функція максимальної правдоподібності набере такого вигляду:

$$f\left(x_1, x_2, ..., x_n, \theta_1^*, \theta_2^*\right) = \frac{1}{(2\pi\theta_2)^{\frac{n}{2}}} \cdot e^{-\frac{\sum\limits_{i=1}^{n}\left(x_i-\theta_1^*\right)^2}{2\theta_2^*}}. \qquad (406)$$

При цьому за статистичні оцінки θ_1^*, θ_2^* вибирають ті їх значення, за яких задана вибірка буде найімовірнішою, тобто функція (406) досягає максимуму.

На практиці зручно від функції (406) перейти до її логарифма, а саме:

$$\ln f(x_1, x_2, ..., x_n, \theta_1^*, \theta_2^*) =$$

$$= L\left(x_1, x_2, ..., x_n, \theta_1^*, \theta_2^*\right) = -\frac{n}{2}\left(\ln\pi + \ln\theta_2^*\right) - \frac{\sum\left(x_i - \theta_1^*\right)^2}{2\theta_2^*}.$$

Згідно з необхідною умовою екстремуму для цієї функції дістанемо:

$$\begin{cases} \dfrac{\partial L}{\partial \theta_1^*} = -\dfrac{1}{\theta_2^*}\sum\limits_{i=1}^{n}\left(x_i - \theta_1^*\right) = 0, \\ \dfrac{\partial L}{\partial \theta_2^*} = -\dfrac{n}{2\theta_2^*} + \dfrac{1}{2\left(\theta_2^*\right)^2} \cdot \sum\limits_{i=1}^{n}\left(x_i - \theta_1^*\right)^2 = 0. \end{cases} \qquad (407)$$

З першого рівняння системи (407) дістанемо:

$$\theta_1^* = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \bar{x}_{\text{В}}; \qquad (408)$$

з другого рівняння системи (407) маємо:

$$\theta_2^* = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x}_{\text{В}})^2 = D_{\text{В}}. \qquad (409)$$

Отже, для $\overline{X}_{\Gamma} = M(X)$ точковою статистичною оцінкою є $\bar{x}_{\text{В}}$ для $D_{\Gamma} - D_{\text{В}}$.

Властивості $\bar{x}_{\text{В}}$, $D_{\text{В}}$. Виправлена дисперсія, виправлене середнє квадратичне відхилення. Точковою незміщеною статистичною оцінкою для $\overline{X}_{\Gamma} = M(X)$ є $\bar{x}_{\text{В}}$.

І справді,

$$M(\bar{x}_{\text{В}}) = M\left(\frac{\sum\limits_{i=1}^{n} x_i}{n}\right) = \frac{\sum M(x_i)}{n} = \left|\begin{matrix}\text{Ураховуючите, що} \\ M(x_i) = \overline{X}_{\Gamma} = a\end{matrix}\right| = \frac{\sum\limits_{i=1}^{n} a}{n} = \frac{na}{n} = a.$$

Отже, $M\left(\bar{x}_{\text{В}}\right)=\bar{X}_{\Gamma}$.

Перевіримо на незміщеність статистичну оцінку $D_{\text{В}}$.

$$M\left(D_{\text{В}}\right)=M\left(\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}_{\text{В}}\right)^2}{n}\right)=M\left(\frac{\sum_{i=1}^{n}\left(\left(x_i-a\right)-\left(\bar{x}_{\text{В}}-a\right)\right)^2}{n}\right)=$$

$$=M\frac{\sum_{i=1}^{n}\left(\left(x_i-a\right)^2-2\left(x_i-a\right)\left(\bar{x}_{\text{В}}-a\right)+\left(\bar{x}_{\text{В}}-a\right)^2\right)}{n}=$$

$$=M\frac{\sum_{i=1}^{n}\left(x_i-a\right)^2-2\sum_{i=1}^{n}\left(x_i-a\right)\left(\bar{x}_{\text{В}}-a\right)+\sum_{i=1}^{n}\left(\bar{x}_{\text{В}}-a\right)^2}{n}=$$

$$=M\frac{\sum_{i=1}^{n}\left(x_i-a\right)^2-2\left(\bar{x}_{\text{В}}-a\right)\sum_{i=1}^{n}\left(x_i-a\right)+\left(\bar{x}_{\text{В}}-a\right)^2 n}{n}=$$

$$=M\frac{\sum_{i=1}^{n}\left(x_i-a\right)^2-2\left(\bar{x}_{\text{В}}-a\right)\left(\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}a\right)+n\left(\bar{x}_{\text{В}}-a\right)^2}{n}=$$

$$=M\frac{\sum_{i=1}^{n}\left(x_i-a\right)^2-2\left(\bar{x}_{\text{В}}-a\right)\left(n\bar{x}_{\text{В}}-na\right)+n\left(\bar{x}_{\text{В}}-a\right)^2}{n}=$$

$$=M\frac{\sum_{i=1}^{n}\left(x_i-a\right)^2-2n\left(\bar{x}_{\text{В}}-a\right)^2+n\left(\bar{x}_{\text{В}}-a\right)^2}{n}=$$

$$=M\frac{\sum_{i=1}^{n}\left(x_i-a\right)^2}{n}-M\left(\bar{x}_{\text{В}}-a\right)^2=$$

$$=\left|\begin{array}{c}\text{Оскільки } M\left(x_i-a\right)^2=D_{\Gamma}\text{, і}\\ M\left(\bar{x}_{\text{В}}-a\right)^2=D\left(\bar{x}_{\text{В}}\right)=D\left(\frac{\sum_{i=1}^{n}x_i}{n}\right)=\frac{\sum_{i=1}^{n}D\left(x_i\right)}{n^2}=\frac{D_{\Gamma}}{n},\\ \text{дістанемо:}\end{array}\right|=$$

$$= \frac{\sum_{i=1}^{n} M\left(x_i - a\right)^2}{n} - M\left(\bar{x}_{\text{В}} - a\right)^2 = \frac{\sum_{i=1}^{n} D_{\text{Г}}}{n} - \frac{D_{\text{Г}}}{n} =$$

$$= \frac{nD_{\text{Г}}}{n} - \frac{D_{\text{Г}}}{n} = D_{\text{Г}} - \frac{1}{n} D_{\text{Г}} = \left(1 - \frac{1}{n}\right) D_{\text{Г}} = \frac{n-1}{n} D_{\text{Г}}.$$

Таким чином, маємо:

$$M\left(D_{\text{В}}\right) = \frac{n-1}{n} D_{\text{Г}}.$$

Отже, $D_{\text{В}}$ є точковою зміщеною статистичною оцінкою для $D_{\text{Г}}$, де $\frac{n-1}{n}$ — коефіцієнт зміщення, який зменшується зі збільшенням обсягу вибірки *n*.

Коли $D_{\text{В}}$ помножити на $\frac{n}{n-1}$, то дістанемо $\frac{n}{n-1} D_{\text{В}}$.

Тоді

$$M\left(\frac{n}{n-1} D_{\text{В}}\right) = \frac{n}{n-1} M\left(D_{\text{В}}\right) = \frac{n}{n-1} \cdot \frac{n-1}{n} D_{\text{Г}} = D_{\text{Г}}.$$

Отже, $\frac{n}{n-1} D_{\text{В}}$ буде точковою незміщеною статистичною оцінкою для $D_{\text{Г}}$. Її назвали *виправленою дисперсією* і позначили через S^2.

Звідси точковою незміщеною статистичною оцінкою для $D_{\text{Г}}$ є виправлена дисперсія $S^2 = \frac{n}{n-1} D_{\text{В}}$ або

$$S^2 = \frac{n}{n-1} \cdot \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}_{\text{В}}\right)^2}{n} = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}_{\text{В}}\right)^2}{n-1}. \qquad (410)$$

Величину

$$S = \sqrt{\frac{n}{n-1} D_{\text{В}}} \qquad (411)$$

називають *виправленим середнім квадратичним відхиленням*.

Виправлене середнє квадратичне відхилення, слід наголосити, буде зміщеною точковою статистичною оцінкою для $\sigma_{\text{Г}}$, оскільки

$$M(S)=\sqrt{\frac{2}{k}}\cdot\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}\sigma_{\Gamma}, \qquad (412)$$

де $k = n - 1$ є кількістю ступенів свободи;

$\sqrt{\frac{2}{k}}\cdot\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}$ — коефіцієнт зміщення.

Приклад. 200 однотипних деталей були піддані шліфуванню. Результати вимірювання наведені як дискретний статистичний розподіл, поданий у табличній формі:

x_i, мм	3,7	3,8	3,9	4,0	4,1	4,2	4,3	4,4
n_i	1	22	40	79	27	26	4	1

Знайти точкові незміщені статистичні оцінки для $\overline{X}_{\Gamma}=M(x)$, D_{Γ}.

Розв'язання. Оскільки точковою незміщеною оцінкою для $\overline{X}_{\Gamma}$ є $\overline{x}_{\text{В}}$, то обчислимо

$$\overline{x}_{\text{В}}=\frac{\sum x_i n_i}{n}=$$

$$=\frac{3{,}7\cdot1+3{,}8\cdot22+3{,}9\cdot40+4{,}0\cdot79+4{,}1\cdot27+4{,}2\cdot26+4{,}3\cdot4+4{,}4\cdot1}{200}=$$

$$=\frac{3{,}7+83{,}6+156+316+110{,}7+109{,}2+17{,}2+4{,}4}{200}=\frac{808{,}8}{200}=4{,}004\text{ мм.}$$

Для визначення точкової незміщеної статистичної оцінки для D_{Γ} обчислимо $D_{\text{В}}$:

$$\frac{\sum x_i^2 n_i}{n}=\frac{(3{,}7)^2\cdot1+(3{,}8)^2\cdot22+(3{,}9)^2\cdot40+(4{,}0)^2\cdot79+}{200}=$$

$$\frac{+(4{,}1)^2\cdot27+(4{,}2)^2\cdot26+(4{,}3)^2\cdot4+(4{,}4)^2\cdot1}{200}=$$

$$=\frac{13{,}69+317{,}68+608{,}4+1264+453{,}87+458{,}64+73{,}96+19{,}36}{200}=$$

$$=\frac{3209{,}6}{200}=16{,}048.$$

$$D_{\text{В}}=\frac{\sum x_i^2 n_i}{n}-(\bar{x}_{\text{В}})^2=16{,}048-(4{,}004)^2=$$

$$=16{,}048-16{,}032016=0{,}015984.$$

Тоді точкова незміщена статистична оцінка для $D_{\text{Г}}$ дорівнюватиме:

$$S^2=\frac{n}{n-1}D_{\text{В}}=\frac{200}{200-1}\cdot 0{,}015984=\frac{200}{199}\cdot 0{,}015984=0{,}01606\ \text{мм}^2.$$

Приклад. Граничне навантаження на сталевий болт x_i, що вимірювалось в лабораторних умовах, задано як інтервальний статистичний розподіл:

x_i, км/мм^2	4,5—5,5	5,5—6,5	6,5—7,5	7,5—8,5	8,5—9,5	9,5—10,5	10,5—11,5	11,5—12,5	12,5—13,5	13,5—14,5
n_i	40	32	28	24	20	18	16	12	8	4

Визначити точкові незміщені статистичні оцінки для $\bar{X}_{\text{Г}}=M(x),\ D_{\text{Г}}$.

Розв'язання. Для визначення точкових незміщених статистичних оцінок $\bar{x}_{\text{В}}$, S^2 перейдемо від інтервального статистичного розподілу до дискретного, який набирає такого вигляду:

$x_i^*=x_{i-1}+\frac{h}{2}$	5	6	7	8	9	10	11	12	13	14
n_i	40	32	28	24	20	18	16	12	8	4

Обчислимо $\bar{x}_{\text{В}}$: $n=\sum n_i=202$,

$$\bar{x}_{\text{В}}=\frac{\sum x_i^* n_i}{n}=\frac{5\cdot 40+6\cdot 32+7\cdot 28+8\cdot 24+9\cdot 20+10\cdot 18+}{202}$$

$$\frac{+11\cdot 16+12\cdot 12+13\cdot 8+14\cdot 4}{202}=\frac{1620}{202}=8{,}02\ \text{кг/мм}^2.$$

Отже, точкова незміщена статистична оцінка для $\overline{X}_{\text{Г}} = M(x)$, $\overline{x}_{\text{В}} = 8{,}02$ кг/мм^2.

Для визначення S^2 обчислимо $D_{\text{В}}$:

$$\frac{\sum (x_i^*)^2 n_i}{n} = \frac{(5)^2 \cdot 40 + (6)^2 \cdot 32 + (7)^2 \cdot 28 + (8)^2 \cdot 24 + (9)^2 \cdot 20 +}{202}$$

$$\frac{+(10)^2 \cdot 18 + (11)^2 \cdot 16 + (12)^2 \cdot 12 + (13)^2 \cdot 8 + (14)^2 \cdot 4}{202} = \frac{14280}{202} \approx 70{,}69.$$

$$D_{\text{В}} = \frac{\sum (x_i^*)^2 n_i}{n} - (\overline{x}_{\text{В}})^2 = 70{,}69 - (8{,}02)^2 = 70{,}69 - 64{,}32 \approx 6{,}37 \text{ кг/мм}^2.$$

$$S^2 = \frac{n}{n-1} D_{\text{В}} = \frac{202}{202-1} \cdot 6{,}37 = \frac{202}{201} \cdot 6{,}37 \approx 6{,}4\,.$$

Звідси точкова незміщена статистична оцінка для $D_{\text{Г}}$ є $S^2 = 6{,}4$ кг/мм^2.

4. Закони розподілу ймовірностей для $\overline{x}_{\text{В}}$, S^2, S

Як уже зазначалося, числові характеристики вибірки є випадковими величинами, що мають певні закони розподілу ймовірностей. Так, $\overline{x}_{\text{В}}$ (вибіркова середня) на підставі центральної граничної теореми теорії ймовірностей (теореми Ляпунова) матиме нормальний закон розподілу з числовими характеристиками

$$M(\overline{x}_{\text{В}}) = M\left(\frac{\sum x_i n_i}{n}\right) = \frac{1}{n} \sum M(x_i) \cdot n_i = \frac{1}{n} \sum a \cdot n_i = a \frac{\sum n_i}{n} = a\,;$$

$$\left(a = M(x) = \overline{X}_{\text{Г}}\right);$$

$$D(\overline{x}_{\text{В}}) = D\left(\frac{\sum x_i n_i}{n}\right) = \frac{D_{\text{Г}}}{n};$$

$$\sigma(\overline{x}_{\text{В}}) = \frac{\sigma_{\text{Г}}}{\sqrt{n}}.$$

Отже, випадкова величина $\overline{x}_{\text{В}}$ має закон розподілу $N\left(a; \frac{\sigma_{\text{Г}}}{\sqrt{n}}\right)$.

Щоб визначити закон розподілу для S^2, необхідно виявити зв'язок між S^2 і розподілом χ^2.

Нехай ознака генеральної сукупності X має нормальний закон розподілу $N(a;\sigma)$. При реалізації вибірки кожну з варіант $X = x_i$ розглядають як випадкову величину, що також має закон розподілу $N(a;\sigma)$. При цьому варіанти вибірки є незалежними, тобто $K_{ij} = 0$, а випадкова величина $z = \dfrac{x_i - a}{\sigma}$ відповідно матиме закон розподілу $N(0;1)$.

Розглянемо випадок, коли варіанти вибірки мають частоти $n_i = 1$, тоді

$$S^2 = \frac{1}{n-1} \cdot \sum (x_i - \bar{x})^2, \text{ де } \bar{x} = \bar{x}_{\text{В}} = \frac{\sum x_i}{n_i}.$$

Перейдемо від випадкових величин $x_1, x_2, ..., x_n$ до випадкових величин $y_1, y_2, ..., y_n$, які лінійно виражаються через x_i, а саме:

$$y_1 = \frac{1}{\sqrt{1 \cdot 2}}(x_1 - x_2);$$

$$y_2 = \frac{1}{\sqrt{2 \cdot 3}}(x_1 + x_2 - 2x_3);$$

$$y_3 = \frac{1}{\sqrt{3 \cdot 4}}(x_1 + x_2 + x_3 - 3x_4);$$

$$y_4 = \frac{1}{\sqrt{4 \cdot 5}}(x_1 + x_2 + x_3 + x_4 - 4x_5);$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$y_{n-1} = \frac{1}{\sqrt{(n-1)n}}(x_1 + x_2 + x_3 + ... + x_{n-1} - (n-1)x_n),$$

$$y_n = \frac{1}{\sqrt{n}}(x_1 + x_2 + x_3 + ... + x_{n-1} + x_n) = \frac{\sum x_i}{n} = \frac{n\bar{x}}{\sqrt{n}} = \sqrt{n} \cdot \bar{x}.$$

Оскільки випадкові величини y_i є лінійними комбінаціями випадкових величин x_i, то $y_1, y_2, ..., y_n$ теж матимуть нормальний закон розподілу з числовими характеристиками:

$$M(y_1) = M\left(\frac{1}{\sqrt{1 \cdot 2}}(x_1 - x_2)\right) = \frac{1}{\sqrt{1 \cdot 2}}(M(x_1) - M(x_2)) = \frac{1}{\sqrt{1 \cdot 2}}(a - a) = 0,$$

$$D(y_1)=D\left(\frac{1}{\sqrt{1\cdot 2}}(x_1-x_2)\right)=\frac{1}{2}(D(x_1)+D(x_2))=\frac{1}{2}\left(\sigma^2+\sigma^2\right)=\sigma^2=D_\Gamma;$$

$$M(x_2)=M\left(\frac{1}{\sqrt{2\cdot 3}}(x_1+x_2-2x_3)\right)=\frac{1}{\sqrt{2\cdot 3}}(M(x_1)+M(x_2)-2M(x_3))=$$

$$=\frac{1}{\sqrt{2\cdot 3}}(a+a-2a)=0,$$

$$D(y_2)=D\left(\frac{1}{\sqrt{2\cdot 3}}(x_1+x_2-2x_3)\right)=\frac{1}{6}(D(x_1)+D(x_2)+4D(x_3))=$$

$$=\frac{1}{6}\left(\sigma^2+\sigma^2+4\sigma^2\right)=\sigma^2=D_\Gamma.$$

.....

.....

$$M(y_{n-1})=M\left(\frac{1}{\sqrt{(n-1)\cdot n}}(x_1+x_2+x_3+....+x_{n-1}-(n-1)x_n)\right)=$$

$$=\frac{1}{\sqrt{(n-1)n}}(M(x_1)+M(x_2)+M(x_3)+...+M(x_{n-1})-(n-1)M(x_n))=$$

$$=\frac{1}{\sqrt{(n-1)n}}(a+a+a+...+a-(n-1)\cdot a)=$$

$$=\frac{1}{\sqrt{(n-1)\cdot n}}((n-1)\cdot a-(n-1)a)=0.$$

$$D(y_{n-1})=D\left(\frac{1}{\sqrt{(n-1)\cdot n}}(x_1+x_2+x_3+....+x_{n-1}-(n-1)x_n)\right)=$$

$$=\frac{1}{(n-1)n}\left(D(x_1)+D(x_2)+D(x_3)+...+D(x_{n-1})+(n-1)^2D(x_n)\right)=$$

$$=\frac{1}{(n-1)\cdot n}\left(\sigma^2+\sigma^2+\sigma^2+...+\sigma^2+(n-1)^2\cdot\sigma^2\right)=$$

$$=\frac{1}{(n-1)\cdot n}\left((n-1)\cdot\sigma^2+(n-1)^2\sigma^2\right)=$$

$$=\frac{n-1}{(n-1\cdot n)}\left(\sigma^2+(n-1)\sigma^2\right)=\frac{(n-1)\cdot n}{(n-1)\cdot n}\sigma^2=\sigma^2=D_\Gamma.$$

Отже, випадкові величини y_i $\left(i=1,\overline{n-1}\right)$ мають закон розподілу $N\left(0;\sigma^2\right)$.

Побудуємо матрицю *A*, елементами якої є коефіцієнти при x_i у лінійних залежностях для y_i:

$$A=\begin{pmatrix} \frac{1}{\sqrt{1\cdot 2}} & -\frac{1}{\sqrt{1\cdot 2}} & 0 & 0 & ... & 0 & 0 \\ \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{23}} & -\frac{2}{\sqrt{2\cdot 3}} & 0 & ... & 0 & 0 \\ \frac{1}{\sqrt{3\cdot 4}} & \frac{1}{\sqrt{3\cdot 4}} & \frac{1}{\sqrt{3\cdot 4}} & -\frac{3}{\sqrt{3\cdot 4}} & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ \frac{1}{\sqrt{(n-1)\cdot n}} & \frac{1}{\sqrt{(n-1)\cdot n}} & \frac{1}{\sqrt{(n-1)\cdot n}} & \frac{1}{\sqrt{(n-1)\cdot n}} & ... & \frac{1}{\sqrt{(n-1)\cdot n}} & -\frac{n-1}{\sqrt{(n-1)\cdot n}} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & ... & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} \end{pmatrix}.$$

Транспонувавши матрицю *A*, дістанемо:

$$A'=A^{\mathrm{T}}=\begin{pmatrix} \frac{1}{\sqrt{1\cdot 2}} & \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{3\cdot 4}} & \frac{1}{\sqrt{4\cdot 5}} & ... & \frac{1}{\sqrt{(n-1)\cdot n}} & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{1\cdot 2}} & \frac{1}{\sqrt{2\cdot 3}} & \frac{1}{\sqrt{3\cdot 4}} & ... & ... & ... & \frac{1}{\sqrt{n}} \\ 0 & 0 & \frac{1}{\sqrt{3\cdot 4}} & ... & ... & ... & \frac{1}{\sqrt{n}} \\ 0 & 0 & .... & ... & ... & ... & \frac{1}{\sqrt{n}} \\ ... & ... & ... & ... & ... & ... & .... \\ 0 & 0 & 0 & 0 & .... & \frac{-(n-1)}{\sqrt{(n-1)\cdot n}} & \frac{1}{\sqrt{n}} \end{pmatrix}.$$

Якщо перемножити матриці *A* і A^{T}, то матимемо:

$$A\cdot A^{\mathrm{T}}=I,$$

де *I* є одинична матриця.

Отже, випадкові величини $y_1, y_2, ..., y_n$ визначені ортогональним перетворенням випадкових величин $x_1, x_2, ..., x_n$. У векторно-матричній формі це можна записати так:

$$\vec{Y}=A\cdot\vec{X}\ ,\quad \vec{Y}=\begin{pmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{pmatrix},\quad \vec{X}=\begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{pmatrix}.$$

З курсу алгебри відомо, що під час ортогональних перетворень вектора зберігається його довжина, тобто

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i^2.$$

Тоді з формули для S^2 дістанемо:

$$(n-1)S^2 = \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 = \sum_{i=1}^{n} x_i^2 - n \cdot \left(\overline{x}\right)^2.$$

Оскільки $y_n = \sqrt{n} \cdot \overline{x}$,
далі обчислимо:

$$(n-1)S^2 = \sum_{i=1}^{n} x_i^2 - n\left(\overline{x}\right)^2 = \sum_{i=1}^{n} y_i^2 - y_n^2 = \sum_{i=1}^{n-1} y_i^2 + y_n^2 - y_n^2 = \sum_{i=1}^{n-1} y_i^2.$$

Отже, маємо

$$(n-1)S^2 = \sum_{i=1}^{n-1} \left(y_i^2\right). \qquad (413)$$

Коли поділимо ліву і праву частини (413) на σ^2, то дістанемо

$$\frac{n-1}{\sigma^2} \cdot S^2 = \sum_{i=1}^{n-1} \left(\frac{y_i}{\sigma}\right)^2.$$

Оскільки y_i має закон розподілу $N(0; \sigma)$, то $\frac{y_i}{\sigma}$ матиме закон розподілу $N(0; 1)$, тобто нормований нормальний закон.

Тоді випадкова величина

$$\frac{n-1}{\sigma^2} \cdot S^2 = \sum_{i=1}^{n-1} \left(\frac{y_i}{\sigma}\right)^2$$

матиме розподіл χ^2 із $k = n - 1$ ступенями свободи.

Звідси випливає, що випадкова величина $\frac{\sqrt{n-1}}{\sigma} S$ матиме розподіл χ із $k = n - 1$ ступенями свободи.

Таким чином, доведено:

випадкова величина $\overline{x}_B \sim N(a; b)$, тут символ ~ потрібно читати «розподілена як»;

випадкова величина $S^2 \sim \frac{\chi^2(n-1)}{n-1} \sigma^2$;

випадкова величина $S \sim \frac{\chi(n-1)}{\sqrt{n-1}} \sigma$.

5. Інтервальні статистичні оцінки для параметрів генеральної сукупності

Точкові статистичні оцінки θ^* є випадковими величинами, а тому наближена заміна θ на θ^* часто призводить до істотних похибок, особливо коли обсяг вибірки малий. У цьому разі застосовують інтервальні статистичні оцінки.

Статистична оцінка, що визначається двома числами, кінцями інтервалів, називається *інтервальною*.

Різниця між статистичною оцінкою θ^* та її оцінювальним параметром θ, взята за абсолютним значенням, називається *точністю оцінки*, а саме:

$$\left|\theta^* - \theta\right| < \delta, \tag{414}$$

де δ є точністю оцінки.

Оскільки θ^* є випадковою величиною, то і δ буде випадковою, тому нерівність (414) справджуватиметься з певною ймовірністю.

Імовірність, з якою береться нерівність (414), тобто

$$P\left(\left|\theta^* - \theta\right| < \delta\right) = \gamma, \tag{415}$$

називають *надійністю*.

Рівність (415) можна записати так:

$$P\left(\theta^* - \delta < \theta < \theta^2 + \delta\right) = \gamma. \tag{416}$$

Інтервал $\left[\theta^* - \delta;\ \theta^* + \delta\right]$, що покриває оцінюваний параметр θ генеральної сукупності з заданою надійністю γ, називають *довірчим*.

6. Побудова довірчого інтервалу для $\overline{X}_Г$ при відомому значенні $\sigma_Г$ із заданою надійністю γ

Нехай ознака X генеральної сукупності має нормальний закон розподілу. Побудуємо довірчий інтервал для $\overline{X}_Г$, знаючи числове значення середнього квадратичного відхилення генеральної сукупності $\sigma_Г$, із заданою надійністю γ. Оскільки $\overline{x}_В$ як точкова незміщена статистична оцінка для $\overline{X}_Г = M(x)$ має нормальний закон розподілу з числовими характеристиками $M(\overline{x}_В) = \overline{X}_Г = a$, $\sigma(\overline{x}_В) = \dfrac{\sigma_Г}{\sqrt{n}}$, то, скориставшись (416), дістанемо

$$P\left(\left|\bar{x}_{\text{В}} - a\right| < \delta\right) = \gamma\,. \qquad (417)$$

Випадкова величина $\bar{x}_{\text{В}} - a$ має нормальний закон розподілу з числовими характеристиками

$$M\left(\bar{x}_{\text{В}} - a\right) = M\left(\bar{x}_{\text{В}}\right) - a = a - a = 0;$$

$$D\left(\bar{x}_{\text{В}} - a\right) = D\left(\bar{x}_{\text{В}}\right) = \frac{D_{\text{Г}}}{n};$$

$$\sigma\left(\bar{x}_{\text{В}}\right) = \frac{\sigma_{\text{Г}}}{\sqrt{n}}.$$

Тому $\dfrac{\bar{x}_{\text{В}} - a}{\dfrac{\sigma_{\text{Г}}}{\sqrt{n}}}$ матиме нормований нормальний закон розподілу $N(0;\ 1)$.

Звідси рівність (417) можна записати, назначивши $\dfrac{\delta}{\dfrac{\sigma_{\text{Г}}}{\sqrt{n}}} = x,$ так:

$$P\left(\left|\frac{\bar{x}_{\text{В}} - a}{\dfrac{\sigma_{\text{Г}}}{\sqrt{n}}}\right| < x\right) = \gamma \qquad (418)$$

або

$$P\left(\bar{x}_{\text{В}} - \frac{x \cdot \sigma_{\text{Г}}}{\sqrt{n}} < a < \bar{x}_{\text{В}} + \frac{x \cdot \sigma_{\text{Г}}}{\sqrt{n}}\right) = \gamma.$$

Згідно з формулою нормованого нормального закону

$$P\left(\left|X - a\right| < \delta\right) = 2\Phi\left(\delta\right)$$

для (418) вона набирає такого вигляду:

$$P\left(\left|\frac{\bar{x}_{\text{В}} - a}{\dfrac{\sigma_{\text{Г}}}{\sqrt{n}}}\right| < x\right) = 2\Phi(x) = \gamma. \qquad (419)$$

З рівності (419) знаходимо аргументи x, а саме:

$$2\Phi\left(x\right) = \gamma \rightarrow \Phi\left(x\right) = 0{,}5\gamma.$$

Аргумент x знаходимо за значенням функції Лапласа, яка дорівнює 0,5 γ за таблицею (додаток 2).

Отже, довірчий інтервал дорівнюватиме:

$$\bar{x}_{\text{В}} - \frac{x \cdot \sigma_{\text{Г}}}{\sqrt{n}} < a < \bar{x}_{\text{В}} + \frac{x \cdot \sigma_{\text{Г}}}{\sqrt{n}}, \qquad (420)$$

що можна зобразити умовно на рис. 118.

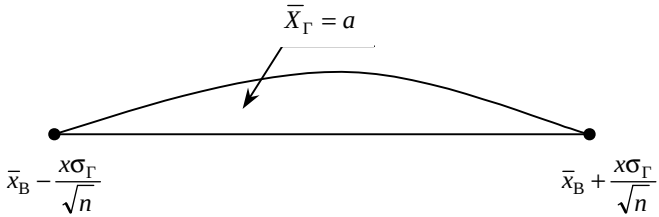


Рис. 118

Величина $\frac{x \cdot \sigma_{\text{Г}}}{\sqrt{n}}$ називається *точністю оцінки*, або *похибкою вибірки*.

Приклад. Вимірявши 40 випадково відібраних після виготовлення деталей, знайшли вибіркову середню, що дорівнює 15 см. Із надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для середньої величини всієї партії деталей, якщо генеральна дисперсія дорівнює $0{,}09\ \text{см}^2$.

Розв'язання. Для побудови довірчого інтервалу необхідно знати: $\bar{x}_{\text{В}}$, $\sigma_{\text{Г}}$, n, x.

З умови задачі маємо: $\bar{x}_{\text{В}} = 15$ см, $\sigma_{\text{Г}} = \sqrt{D_{\text{Г}}} = \sqrt{0{,}09\ \text{см}^2} = 0{,}3$ см, $n = 40 \quad \rightarrow \quad \sqrt{n} = \sqrt{40} = 6{,}32$. Величина x обчислюється з рівняння

$$\Phi(x) = 0{,}5\gamma = 0{,}5 \cdot 0{,}99 = 0{,}495.$$

$$\Phi(x) = 0{,}495 \rightarrow x = 2{,}58 \ [\text{за таблицею значень функції Лапласа}].$$

Знайдемо числові значення кінців довірчого інтервалу:

$$\bar{x}_{\text{В}} - \frac{\sigma_{\text{Г}} \cdot x}{\sqrt{n}} = 15 - \frac{0{,}3 \cdot 2{,}58}{6{,}32} = 15 - 0{,}12 = 14{,}88 \text{ см}.$$

$$\bar{x}_{\text{В}} + \frac{\sigma_{\text{Г}} \cdot x}{\sqrt{n}} = 15 + \frac{0{,}3 \cdot 2{,}58}{6{,}32} = 15 + 0{,}12 = 15{,}12 \text{ см}.$$

Таким чином, маємо:

$$14{,}88 < \bar{X}_{\text{Г}} < 15{,}12.$$

Отже, з надійністю 0,99 (99% гарантії) оцінюваний параметр $\overline{X}_{\text{Г}}$ перебуває усередині інтервалу [14,87; 15,13].

Приклад. Маємо такі дані про розміри основних фондів (у млн грн.) на 30-ти випадково вибраних підприємствах:
4,2; 2,4; 4,9; 6,7; 4,5; 2,7; 3,9; 2,1; 5,8; 4,0;
2,8; 7,8; 4,4; 6,6; 2,0; 6,2; 7,0; 8,1; 0,7,; 6,8;
9,4; 7,6; 6,3; 8,8; 6,5; 1,4; 4,6; 2,0; 7,2; 9,1.

Побудувати інтервальний статистичний розподіл із довжиною кроку h = 2 млн грн.

З надійністю $\gamma = 0{,}999$ знайти довірчий інтервал для $\overline{X}_{\text{Г}}$, якщо $\sigma_{\text{Г}}$ = 5 млн грн.

Розв'язання. Інтервальний статистичний розподіл буде таким:

h = 2 млн грн.	2—4	4—6	6—8	8—10
n_i	9	7	10	4

Для визначення $\overline{x}_{\text{В}}$ необхідно побудувати дискретний статистичний розподіл, що має такий вигляд:

x_i^*	3	5	7	9
n_i	9	7	10	4

$n = \sum n_i = 30$.

Тоді

$$\overline{x}_{\text{В}} = \frac{\sum x_i^* n_i}{n} = \frac{3\cdot 9 + 5\cdot 7 + 7\cdot 10 + 9\cdot 4}{30} = \frac{27+35+70+36}{30} =$$

$$= \frac{168}{30} = 5{,}6 \text{ млн грн.}$$

Для побудови довірчого інтервалу із заданою надійністю $\gamma = 0{,}999$ необхідно знайти x:

$$\Phi(x) = 0{,}5\gamma = 0{,}5\cdot 0{,}999 = 0{,}4995 \to x \approx 3{,}4.$$

Обчислюємо кінці інтервалу:

$$\overline{x}_{\text{В}} - \frac{x\sigma_{\text{Г}}}{\sqrt{n}} = 5{,}6 - \frac{3{,}4\cdot 5}{\sqrt{30}} = 5{,}6 - \frac{3{,}4\cdot 5}{5{,}5} = 5{,}6 - 3{,}1 = 2{,}5 \text{ млн грн.}$$

$$\overline{x}_{\text{В}} + \frac{x\sigma_{\text{Г}}}{\sqrt{n}} = 5{,}6 + \frac{3{,}4\cdot 5}{\sqrt{30}} = 5{,}6 + \frac{3{,}4\cdot 5}{5{,}5} = 5{,}6 + 3{,}1 = 8{,}7 \text{ млн грн.}$$

Отже, довірчий інтервал для $\overline{X}_{\text{Г}}$ буде $2{,}5 < \overline{X}_{\text{Г}} < 8{,}7$.

Приклад. Якого значення має набувати надійність оцінки γ, щоб за обсягу вибірки n = 100 похибка її не перевищувала 0,01 при $\sigma_{\text{Г}} = 5$.

Розв'язання. Позначимо похибку вибірки

$$\frac{x \cdot \sigma_{\text{Г}}}{\sqrt{n}} = \varepsilon \to x = \frac{\varepsilon\sqrt{n}}{\sigma_{\text{Г}}} = \frac{0{,}01\sqrt{100}}{5} = \frac{0{,}01 \cdot 10}{5} = 0{,}02.$$

Далі маємо:

$$P\left(\left|\frac{\overline{x}_{\text{В}} - a}{\frac{\sigma_{\text{Г}}}{\sqrt{n}}}\right| < x\right) = 2\Phi(x) = 2\Phi(0{,}02) = 2 \cdot 0{,}008 = 0{,}016.$$

Як бачимо, надійність мала.

Приклад. Визначити обсяг вибірки n, за якого похибка $\varepsilon = 0{,}01$ гарантується з імовірністю 0,999, якщо $\sigma_{\text{Г}} = 5$.

Розв'язання. За умовою задачі $P\left(\left|\frac{\overline{x}_{\text{В}} - a}{\frac{\sigma_{\text{Г}}}{\sqrt{n}}}\right| < x\right) = \gamma = 0{,}999.$ Оскільки $\frac{x\sigma_{\text{Г}}}{\sqrt{n}} = \varepsilon,$ то дістанемо: $n = \frac{x^2\sigma_{\text{Г}}^2}{\varepsilon^2}$. Величину x знаходимо з рівності $\Phi(x) = 0{,}5\gamma = 0{,}5 \cdot 0{,}999 = 0{,}4995 \to x \approx 3{,}4.$ Тоді $n = \frac{(3{,}4)^2 \cdot 5^2}{(0{,}01)^2} = 2\,890\,000.$

7. Побудова довірчого інтервалу для $\overline{X}_{\text{Г}}$ при невідомому значенні $\sigma_{\text{Г}}$ із заданою надійністю γ

Для малих вибірок, з якими стикаємося, досліджуючи різні ознаки в техніці чи сільському господарстві, для оцінювання $\overline{X}_{\text{Г}} = a$ при невідомому значенні $\sigma_{\text{Г}}$ неможливо скористатися нормальним законом розподілу. Тому для побудови довірчого інтервалу застосовується випадкова величина

$$t = \frac{\bar{x}_{\text{В}} - a}{\frac{S}{\sqrt{n}}}, \qquad (421)$$

що має розподіл Стьюдента з $k = n - 1$ ступенями свободи.

Тоді (421) набирає такого вигляду:

$$P\left(\left|\frac{\bar{x}_{\text{В}} - a}{\frac{S}{\sqrt{n}}}\right| < t_\gamma\right) = P\left(\bar{x}_{\text{В}} - \frac{t_\gamma \cdot S}{\sqrt{n}} < a < \bar{x}_{\text{В}} + \frac{t_\gamma S}{\sqrt{n}}\right) = 2\int_0^{t_\gamma} f(t) = \gamma,$$

оскільки $f(t)$ для розподілу Стьюдента є функцією парною.

Обчисливши за даним статистичним розподілом $\bar{x}_{\text{В}}$, S і визначивши за таблицею розподілу Стьюдента значення t_γ, будуємо довірчий інтервал

$$\bar{x}_{\text{В}} - \frac{t_\gamma \cdot S}{\sqrt{n}} < a < \bar{x}_{\text{В}} + \frac{t_\gamma S}{\sqrt{n}}. \qquad (422)$$

Тут $t_\gamma(\gamma, k = n - 1)$ обчислюємо за заданою надійністю γ і числом ступенів свободи $k = n - 1$ за таблицею (додаток 3).

Приклад. Випадково вибрана партія з двадцяти приладів була випробувана щодо терміну безвідказної роботи кожного з них t_i. Результати випробувань наведено у вигляді дискретного статистичного розподілу:

t_i	100	170	240	310	380
n_i	2	5	10	2	1

З надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для «*a*» (середнього часу безвідказної роботи приладу).

Розв'язання. Для побудови довірчого інтервалу необхідно знайти середнє вибіркове і виправлене середнє квадратичне відхилення.

Обчислимо $\bar{x}_{\text{В}}$:

$$\bar{x}_{\text{В}} = \frac{\sum t_i n_i}{n} = \frac{100 \cdot 2 + 170 \cdot 5 + 240 \cdot 10 + 310 \cdot 2 + 380 \cdot 1}{20} = \frac{4450}{20} = 222{,}5.$$

Отже, дістали $\bar{x}_{\text{В}} = 222{,}5$ год.

Визначимо $D_{\text{В}}$:

$$\frac{\sum t_i^2 n_i}{n} = \frac{100^2 \cdot 2 + 170^2 \cdot 5 + 240^2 \cdot 10 + 310^2 \cdot 2 + 380^2 \cdot 1}{20} = \frac{1\,077100}{20} = 53855$$

$$D_{\text{В}} = \frac{\sum t_i^2 n_i}{n} - \left(\bar{x}_{\text{В}}\right)^2 = 53\,855 - \left(222{,}5\right)^2 = 53\,855 - 49\,506{,}25 = 4348{,}75.$$

Отже, $D_{\text{В}} = 4348{,}75$.

Виправлене середнє квадратичне відхилення дорівнюватиме:

$$S = \sqrt{\frac{n}{n-1} D_{\text{В}}} = \sqrt{\frac{20}{20-1} \cdot 4348{,}75} \approx 67{,}66 \text{ год.}$$

За таблицею значень $\int_0^t f(x)\,dt = \gamma = 0{,}99$ (додаток 3) розподілу Стьюдента за заданою надійністю $\gamma = 0{,}99$ і числом ступенів свободи $k = n - 1 = 20 - 1 = 19$ знаходимо значення $t(\gamma = 0{,}99,\ k = 19) = 2{,}861$.

Обчислимо кінці довірчого інтервалу:

$$\bar{x}_{\text{В}} - \frac{t_\gamma S}{\sqrt{n}} = 222{,}5 - \frac{2{,}861 \cdot 67{,}66}{\sqrt{20}} = 222{,}5 - \frac{2{,}861 \cdot 67{,}66}{4{,}472} = 179{,}2 \text{ год.}$$

$$\bar{x}_{\text{В}} + \frac{t_\gamma S}{\sqrt{n}} = 222{,}5 + \frac{2{,}861 \cdot 67{,}89}{\sqrt{20}} = 222{,}5 + \frac{2{,}861 \cdot 67{,}66}{4{,}472} = 265{,}8 \text{ год.}$$

Отже, з надійністю $\gamma = 0{,}99$ можна стверджувати, що $\overline{X}_{\text{Г}} = a$ буде міститися в інтервалі

$$179{,}2 < a < 265{,}8\,.$$

При великих обсягах вибірки, а саме: $n > 30$, на підставі центральної граничної теореми теорії ймовірностей (теореми Ляпунова) розподіл Стьюдента наближається до нормального закону. У цьому разі t_γ знаходиться за таблицею значень функції Лапласа.

Приклад. У таблиці наведено відхилення діаметрів валиків, оброблених на верстаті, від номінального розміру:

h = 5 мк	0 — 5	5 — 10	10 — 15	15 — 20	20 — 25
n_i	15	75	100	50	10

Із надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для $\overline{X}_{\text{Г}} = a$.

Розв'язання. Для побудови довірчого інтервалу необхідно знайти $\bar{x}_{\text{В}}$, S.

Для цього від інтервального статистичного розподілу, наведеного в умові задачі, необхідно перейти до дискретного, а саме:

x_i^*	2,5	7,5	12,5	17,5	22,5
n_i	15	75	100	50	10

Обчислимо $\bar{x}_{\text{В}}$:

$$\bar{x}_{\text{В}} = \frac{\sum x_i^* n_i}{n} = \left| \text{Оскільки } n = \sum n_i = 250 \right| =$$

$$= \frac{2{,}5 \cdot 15 + 7{,}5 \cdot 75 + 12{,}5 \cdot 100 + 17{,}5 \cdot 50 + 22{,}5 \cdot 10}{250} =$$

$$= \frac{37{,}5 + 562{,}5 + 1250 + 875 + 225}{250} = \frac{2950}{250} = 11{,}8.$$

Отже, $\bar{x}_{\text{В}} = 11{,}8$ мк.

Визначимо $D_{\text{В}}$:

$$\frac{\sum \left(x_i^*\right)^2 n_i}{n} = \frac{(2{,}5)^2 \cdot 15 + (7{,}5)^2 \cdot 75 + (12{,}5)^2 \cdot 100 + (17{,}5)^2 \cdot 50 + (22{,}5)^2 \cdot 10}{250} =$$

$$= \frac{93{,}75 + 4218{,}75 + 15625 + 15312{,}5 + 5062{,}5}{250} = \frac{40312{,}5}{250} = 161{,}25.$$

$$D_{\text{В}} = \frac{\sum \left(x_i^*\right)^2 n_i}{n} - \left(\bar{x}_{\text{В}}\right)^2 = 161{,}25 - (11{,}8)^2 = 161{,}25 - 139{,}24 = 22{,}01.$$

Обчислимо виправлене середнє квадратичне відхилення S:

$$S = \sqrt{\frac{n}{n-1} D_{\text{В}}} = \sqrt{\frac{250}{250-1} \cdot 22{,}01} \approx 4{,}7 \text{ мк.}$$

З огляду на великий ($n = 250$) обсяг вибірки можна вважати, що розподіл Стьюдента близький до нормального закону. Тоді за таблицею значення функції Лапласа

$$\Phi(t_\gamma) = 0{,}495 \to t_\gamma = 2{,}58.$$

Обчислимо кінці інтервалів:

$$\bar{x}_{\text{В}} - \frac{t_\gamma S}{\sqrt{n}} = 11{,}8 - \frac{2{,}58 \cdot 4{,}7}{\sqrt{250}} = 11{,}8 - \frac{2{,}58 \cdot 4{,}7}{15{,}8} = 11{,}8 - 0{,}77 = 11{,}03 \text{ мк.}$$

$$\bar{x}_{\text{В}} + \frac{t_\gamma S}{\sqrt{n}} = 11{,}8 + \frac{2{,}58 \cdot 4{,}7}{\sqrt{250}} = 11{,}8 + \frac{2{,}58 \cdot 4{,}7}{15{,}8} = 11{,}8 + 0{,}77 = 12{,}57 \text{ мк.}$$

Отож, довірчий інтервал для середнього значення відхилень буде таким:

$$11{,}03 < a < 12{,}57 .$$

Звідси з надійністю $\gamma = 0{,}99$ (99%) можна стверджувати, що $a \in$ [11,03 мк; 12,57 мк].

8. Побудова довірчих інтервалів із заданою надійністю γ для $D_{\text{Г}}$, $\sigma_{\text{Г}}$

У разі, коли ознака X має нормальний закон розподілу, для побудови довірчого інтервалу із заданою надійністю γ для $D_{\text{г}}$, $\sigma_{\text{г}}$ застосовуємо випадкову величину

$$\chi^2 = \frac{n-1}{\sigma_{\text{Г}}^2} S^2, \tag{423}$$

що має розподіл χ^2 із $k = n-1$ ступенями свободи.

Оскільки випадкові події

$$A\left(\chi_1^2 < \chi^2 < \chi_2^2\right) \text{ і } B\left(\frac{1}{\chi_2^2} < \frac{1}{\chi^2} < \frac{1}{\chi_1^2}\right)$$

є рівноймовірними, тобто їх імовірності рівні $\left(P(A) = P(B)\right)$, маємо:

$$P\left(\chi_1^2 < \chi^2 < \chi_2^2\right) = P\left(\frac{1}{\chi_2^2} < \frac{1}{\chi^2} < \frac{1}{\chi_1^2}\right). \tag{424}$$

Підставляючи в (424) $\chi^2 = \frac{n-1}{\sigma^2} S^2$, дістанемо

$$P\left(\frac{1}{\chi_2^2} < \frac{1}{\chi^2} < \frac{1}{\chi_1^2}\right) = P\left(\frac{1}{\chi_2^2} < \frac{1}{\frac{n-1}{\sigma_{\text{Г}}^2} S^2} < \frac{1}{\chi_1^2}\right) =$$

$$= P\left(\frac{1}{\chi_2^2} < \frac{\sigma_{\Gamma}^2}{(n-1)S^2} < \frac{1}{\chi_1^2}\right) = P\left(\frac{(n-1)S^2}{\chi_2^2} < \sigma_{\text{Г}}^2 < \frac{(n-1)S^2}{\chi_1^2}\right) = \gamma.$$

Отже, довірчий інтервал для $\sigma_{\Gamma}^2 = D_{\Gamma}$ матиме вигляд:

$$\frac{(n-1)S^2}{\chi_2^2} < D_{\Gamma} < \frac{(n-1)S^2}{\chi_1^2}. \qquad (425)$$

Тоді довірчий інтервал для σ_{Γ} випливає із (425) і буде таким:

$$\frac{S\sqrt{n-1}}{\chi_2} < \sigma_{\Gamma} < \frac{S\sqrt{n-1}}{\chi_1}. \qquad (426)$$

Значення χ_1^2, χ_2^2 знаходимо за таблицею (додаток 4) згідно з рівностями:

$$P\left(\chi^2 > \chi_1^2\right) = 1 - \frac{\alpha}{2}; \qquad (427)$$

$$P\left(\chi^2 > \chi_2^2\right) = \frac{\alpha}{2}, \qquad (428)$$

де $\alpha = 1 - \gamma$.

Приклад. Перевірена партія однотипних телевізорів x_i на чутливість до відеопрограм n_i, дані перевірки наведено як дискретний статистичний розподіл:

n_i, мкв	200	250	300	350	400	450	500	550
x_i	2	5	6	7	5	2	2	1

З надійністю $\gamma = 0{,}99$ побудувати довірчі інтервали для $D_{\Gamma}, \sigma_{\Gamma}$.

Розв'язання. Для побудови довірчих інтервалів необхідно знайти значення S^2, S.

Обчислимо значення $\bar{x}_{\text{В}}$:

$$\bar{x}_{\text{В}} = \frac{\sum x_i n_i}{n} = \left|\text{Так як } n = \sum n_i = 30\right| =$$

$$= \frac{200\cdot 2 + 250\cdot 5 + 300\cdot 6 + 350\cdot 7 + 400\cdot 5 + 450\cdot 2 + 500\cdot 2 + 550\cdot 1}{30} =$$

$$= \frac{400 + 1250 + 1800 + 2450 + 2000 + 900 + 1000 + 550}{30} = \frac{10350}{30} = 345 \text{ мкв.}$$

Обчислимо $D_{\text{В}}$:

$$\frac{\sum x_i^2 n_i}{n} = \frac{(200)^2 \cdot 2 + (250)^2 \cdot 5 + (300)^2 \cdot 6 + (350)^2 \cdot 7 +}{30}$$

$$\frac{+(400)^2 \cdot 5 + (450)^2 \cdot 2 + (500)^2 \cdot 2 + (550)^2 \cdot 1}{30} =$$

$$= \frac{80\ 000 + 312\ 500 + 540\ 000 + 857\ 500 + 800\ 000 + 405\ 000 +}{30}$$

$$\frac{+500\ 000 + 302\ 500}{30} = \frac{3797500}{30} = 126583{,}3\,.$$

$$D_{\text{В}} = \frac{\sum x_i^2 n_i}{n} - \left(\bar{x}_{\text{В}}\right)^2 = 126583{,}3 - (345)^2 = 126583{,}3 - 119025 = 7558{,}3.$$

Отже, $D_{\text{В}} = 7558{,}3\,[\text{мкв}]^2$.

Виправлена дисперсія і виправлене середнє квадратичне відхилення дорівнюватимуть:

$$S^2 = \frac{n}{n-1} D_{\text{В}} = \frac{30}{30-1} \cdot 7558{,}3 = \frac{30}{29} \cdot 7558{,}3 = 7818{,}9\,[\text{мкв}]^2;$$

$$S = \sqrt{\frac{n}{n-1} D_{\text{В}}} = \sqrt{7818{,}9} \approx 88{,}42 \text{ мкв.}$$

Оскільки $\alpha = 1 - \gamma = 1 - 0{,}99 = 0{,}01$, то згідно з (427), (428) знаходимо значення χ_1^2, χ_2^2, а саме:

$$P\left(\chi^2 > \chi_1^2\right) = 1 - \frac{\alpha}{2} = 1 - \frac{0{,}01}{2} = 1 - 0{,}005 = 0{,}995.$$

$$P\left(\chi^2 > \chi_2^2\right) = \frac{\alpha}{2} = \frac{0{,}01}{2} = 0{,}005.$$

За таблицею (додаток 4) знаходимо:

$$\chi_1^2\left(0{,}995;\ k = m - 1\right) = \chi_1^2\left(0{,}995;\ k = 29\right) = 14{,}3.$$

$$\chi_2^2\left(0{,}005;\ k = 29\right) = 52{,}5.$$

Обчислимо кінці довірчого інтервалу для D_{Γ}:

$$\frac{n-1}{\chi_2^2} S^2 = \frac{29}{52{,}5} \cdot 7818{,}9 = 4319{,}01;$$

$$\frac{n-1}{\chi_1^2}S^2 = \frac{29}{14{,}3}\cdot 7818{,}9 = 15856{,}5\,.$$

Отже, довірчий інтервал для D_Γ буде таким:

$$4319{,}0 < D_\Gamma < 15856{,}5.$$

Довірчий інтервал для σ_Γ становить

$$68{,}3 < \sigma_\Gamma < 130{,}83.$$

Довірчий інтервал для σ_Γ можна побудувати із заданою надійністю γ, узявши розподіл χ.

Оскільки

$$P\left(\left|\sigma_\Gamma - S\right| < \delta\right) = \gamma, \qquad (429)$$

то рівність (429) можна записати так:

$$P(S - \delta < \sigma_\Gamma < S + \delta) = \gamma$$

або

$$P\left(S\left(1-\frac{\delta}{S}\right) < \sigma_\Gamma < S\left(1+\frac{\delta}{S}\right)\right) = \gamma.$$

Позначивши $\frac{\delta}{S} = q,$ дістанемо

$$P\left(S(1-q) < \sigma_\Gamma < S(1+q)\right) = \gamma,$$

щоб знайти *q*, візьмемо випадкову величину

$$\chi = \frac{S}{\sigma_\Gamma}\sqrt{n-1}, \qquad (430)$$

що має розподіл χ (хі-розподіл).

Ураховуючи те, що події

$$A\left(S(1-q) < \sigma_\Gamma < S(1+q)\right) \text{ і } B\left(\frac{1}{S(1+q)} < \frac{1}{\sigma_\Gamma} < \frac{1}{S(1-q)}\right)$$

при $q < 1$ є рівноймовірними, маємо:

$$P\left(S(1-q) < \sigma_\Gamma < S(1+q)\right) = P\left(\frac{1}{S(1+q)} < \frac{1}{\sigma_\Gamma} < \frac{1}{S(1-q)}\right).$$

Якщо помножити всі члени подвійної нерівності $\frac{1}{S(1+q)} < \frac{1}{\sigma_\Gamma} < \frac{1}{S(1-q)}$ на $S\sqrt{n-1}$, то дістанемо:

$$P\left(S(1-q) < \sigma_\Gamma < S(1+q)\right) = P\left(\frac{\sqrt{n-1}}{S(1+q)} < \frac{S}{\sigma_\Gamma}\sqrt{n-1} < \frac{\sqrt{n-1}}{S(1-q)}\right) =$$

$$= P\left(\frac{\sqrt{n-1}}{1+q} < \chi < \frac{\sqrt{n-1}}{1-q}\right) = \int_{\frac{\sqrt{n-1}}{1+q}}^{\frac{\sqrt{n-1}}{1-q}} f(t)\,dt = \gamma.$$

Звідси маємо:

$$P(S(1-q) < \sigma_\Gamma < S(1+q)) =$$

$$= P\left(\frac{\sqrt{n-1}}{1+q} < \chi < \frac{\sqrt{n-1}}{1-q}\right) = \int_{\frac{\sqrt{n-1}}{1+q}}^{\frac{\sqrt{n-1}}{1-q}} f(t)\,dt = \gamma. \qquad (431)$$

З рівняння (431) за заданою надійністю γ і обсягом вибірки n знаходимо за таблицею (додаток 5) значення величини $q(\gamma;\, n)$.

Довірчий інтервал буде таким:

$$S\left(1 - q(\gamma; n)\right) < \sigma_\Gamma < S\left(1 + q(\gamma; n)\right). \qquad (432)$$

Приклад. З надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для σ_Γ. Якщо S = 45, n = 30.

Розв'язання. Для побудови довірчого інтервалу обчислимо значення $q(\gamma;\ n)$ за таблицею (додаток 5). ($\gamma = 0{,}99$; n = 30) = 0,43.

Визначимо кінці інтервалу:

$$S\left(1 - q(\gamma; n)\right) = 4{,}5(1 - 0{,}43) = 4{,}5 \cdot 0{,}57 = 2{,}565;$$

$$S\left(1 + q(\gamma; n)\right) = 4{,}5(1 + 0{,}43) = 4{,}5 \cdot 1{,}43 = 6{,}435.$$

Отже, довірчий інтервал для σ_Γ з надійністю $\gamma = 0{,}99$ буде таким:

$$2{,}565 < \sigma_\Gamma < 6{,}435.$$

9. Побудова довірчого інтервалу для r_{xy} генеральної сукупності із заданою надійністю γ

Як величина, одержана за результатами вибірки, $r_{\text{В}}$ є випадковою і являє собою точкову незміщену статистичну оцінку для r_{xy} $\left(M\left(r_{\text{В}}\right)=r_{xy}\right)$.

Виправлене середнє квадратичне відхилення для $r_{\text{В}}$

$$S=\frac{1-r_{\text{В}}^2}{\sqrt{n}}. \qquad (433)$$

Для побудови довірчого інтервалу для r_{xy} використовується випадкова величина

$$x_\gamma=\frac{r_B-r_{xy}}{\sigma\left(r_{\text{В}}\right)}=\frac{r_{\text{В}}-r_{xy}}{\dfrac{1-r_{\text{В}}^2}{\sqrt{n}}}, \qquad (434)$$

що має нормований нормальний закон розподілу $N(0;\,1)$.

Скориставшись (434), дістанемо

$$P\left(\left|\frac{r_{\text{В}}-r_{xy}}{\dfrac{1-r_{\text{В}}^2}{\sqrt{n}}}\right|<x_\gamma\right)=P\left(r_{\text{В}}-t_\gamma\frac{1-r_{\text{В}}^2}{\sqrt{n}}<r_{xy}<r_{\text{В}}-t_\gamma\frac{1-r_{\text{В}}^2}{\sqrt{n}}\right)=\gamma=2\Phi\left(x_\gamma\right).$$

Отже, довірчий інтервал для r_{xy} буде таким:

$$r_{\text{В}}-t_\gamma\frac{1-r_{\text{В}}^2}{\sqrt{n}}<r_{xy}<r_{\text{В}}+t_\gamma\frac{1-r_{\text{В}}^2}{\sqrt{n}}, \qquad (435)$$

де t_γ знаходимо з рівності

$$\Phi\left(x_\gamma\right)=0{,}5\gamma$$

за таблицею значень функції Лапласа.

Приклад. Випадково вибраних студентів із потоку університету було піддано тестуванню з математики і хімії. Результати цих тестувань подано двовимірним статистичним розподілом, де $X=x_i$ — оцінки з математики, $Y=y_i$ — із хімії. Відповіді оцінювалися за десятибальною системою:

$Y = y_i$	$X = x_i$					
	1	3	5	7	9	n_{y_i}
1	2	2	1	—	—	5
3	1	1	1	1	—	4
5	—	—	1	2	3	6
7	—	—	1	1	4	6
9	—	—	2	3	4	9
n_{x_j}	3	3	6	7	11	

Необхідно:

1) з надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для $\overline{X}_\Gamma$, якщо $\sigma_\Gamma = 5$;

2) з надійністю $\gamma = 0{,}999$ побудувати довірчі інтервали для σ_Γ, $\overline{Y}_\Gamma$, r_{xy}.

Розв'язання. Обчислимо основні числові характеристики ознак X і Y, а також K^*_{xy}, $r_\text{В}$. Оскільки $n = \sum\sum n_{ij} = 30$, дістанемо:

$$\overline{x} = \frac{\sum x_i n_{x_j}}{n} = \frac{1\cdot 3 + 3\cdot 3 + 5\cdot 6 + 7\cdot 7 + 9\cdot 11}{30} = 6{,}33;$$

$$\frac{\sum x_i^2 n_{x_j}}{n} = \frac{1^2\cdot 3 + 3^2\cdot 3 + 5^2\cdot 6 + 7^2\cdot 7 + 9^2\cdot 11}{30} = 47{,}13.$$

$$D_x = \frac{\sum x_i n_{x_j}}{n} - (\overline{x})^2 = 47{,}13 - (6{,}33)^2 = 47{,}13 - 40{,}07 = 7{,}06;$$

$$\sigma_x = \sqrt{D_x} = \sqrt{7{,}06} \approx 2{,}66.$$

$$S_x = \sqrt{\frac{n}{n-1} D_x} = \sqrt{\frac{30}{29} 7{,}06} \approx 2{,}7.$$

$$\overline{y} = \frac{\sum y_i n_{y_i}}{n} = \frac{1\cdot 5 + 3\cdot 4 + 5\cdot 6 + 7\cdot 6 + 9\cdot 9}{30} = 5{,}67.$$

$$\frac{\sum y_i^2 n_{y_i}}{n} = \frac{1^2\cdot 5 + 3^2\cdot 4 + 5^2\cdot 6 + 7^2\cdot 6 + 9^2\cdot 9}{30} = 40{,}47.$$

$$D_y = \frac{\sum y_i^2 n_{y_i}}{n} - (\overline{y})^2 = 40{,}47 - (5{,}67)^2 = 40{,}47 - 32{,}15 = 8{,}32.$$

$$\sigma_y = \sqrt{D_y} = \sqrt{8{,}32} \approx 2{,}88.$$

$$S_y = \sqrt{\frac{n}{n-1} D_y} = \sqrt{\frac{30}{29} 8{,}32} \approx 2{,}93.$$

$$\frac{\sum\sum y_i\, x_i\, n_{y_i}}{n} = \frac{1\cdot 2\cdot 1 + 1\cdot 2\cdot 3 + 1\cdot 1\cdot 5 + 3\cdot 1\cdot 1 + 3\cdot 1\cdot 3 + 3\cdot 1\cdot 5 + 3\cdot 1\cdot 7 +}{30}$$

$$\frac{+5\cdot 1\cdot 5 + 5\cdot 2\cdot 7 + 5\cdot 3\cdot 9 + 7\cdot 1\cdot 5 + 7\cdot 1\cdot 7 + 7\cdot 4\cdot 9 +}{30}$$

$$\frac{+9\cdot 2\cdot 5 + 9\cdot 3\cdot 7 + 9\cdot 4\cdot 9}{30} = 41.$$

$$K^*_{xy} = \frac{\sum\sum y_i\, x_i\, n_{iy}}{n} - \bar{x}\cdot\bar{y} = 41 - 6{,}33\cdot 5{,}67 = 41 - 35{,}89 = 5{,}11.$$

$$r_{\text{В}} = \frac{K^*_{xy}}{\sigma_x \sigma_y} = \frac{5{,}11}{2{,}66\cdot 2{,}88} = \frac{5{,}11}{7{,}661} \approx 0{,}667.$$

1. Побудуємо довірчий інтервал з надійністю $\gamma = 0{,}99$ для $\overline{X}_{\text{Г}}$, якщо $\sigma_{\text{Г}} = 5$.

$$\bar{x}_{\text{В}} - \frac{x\sigma_{\text{Г}}}{\sqrt{n}} < \overline{X}_{\text{Г}} < \bar{x}_{\text{В}} + \frac{x\sigma_{\text{Г}}}{\sqrt{n}}.$$

Нам відомі значення $\bar{x}_{\text{В}} = \bar{x} = 6{,}33$, $\sigma_{\text{Г}} = 5$, $\sqrt{n} = \sqrt{30} = 5{,}48$. Значення x обчислюємо з рівняння

$$\Phi(x) = 0{,}5\gamma = 0{,}5\cdot 0{,}99 = 0{,}495,$$

де $x = 2{,}58$ знаходимо за таблицею значень функції Лапласа.

Визначимо кінці інтервалу:

$$\bar{x}_{\text{В}} - \frac{x\sigma_{\text{Г}}}{\sqrt{n}} = \bar{x}_{\text{В}} - \frac{x\sigma_{\text{Г}}}{\sqrt{n}} = 6{,}33 - \frac{2{,}58\cdot 5}{5{,}48} = 6{,}33 - 2{,}35 = 3{,}98;$$

$$\bar{x}_{\text{В}} + \frac{x\sigma_{\text{Г}}}{\sqrt{n}} = \bar{x}_{\text{В}} + \frac{x\sigma_{\text{Г}}}{\sqrt{n}} = 6{,}33 + \frac{2{,}58\cdot 5}{5{,}48} = 6{,}33 + 2{,}35 = 8{,}68.$$

Отже, довірчий інтервал для $\overline{X}_{\text{Г}}$ буде таким:

$$3{,}98 < \overline{X}_{\text{Г}} < 8{,}68.$$

2. Побудуємо довірчий інтервал з надійністю $\gamma = 0{,}999$ для $\overline{Y}_{\text{Г}}$.

Оскільки σ_{Γ} нам не відоме, то довірчий інтервал у цьому разі визначається так:

$$\bar{y}_{\text{В}} - \frac{t_{\gamma} S_y}{\sqrt{n}} < \bar{Y}_{\Gamma} < \bar{y}_{\text{В}} + \frac{t_{\gamma} S_y}{\sqrt{n}}.$$

Нам відоме значення $\bar{y}_{\text{В}} = \bar{y} = 5{,}67$, $S_y = 2{,}93$, t_{γ} знаходимо за таблицею розподілу Стьюдента (додаток 3).

$$t(\gamma = 0{,}999,\ k = 29) = 3{,}659.$$

Обчислимо кінці довірчого інтервалу:

$$\bar{y} - \frac{t_{\gamma} S_y}{\sqrt{n}} = 5{,}67 - \frac{3{,}659 \cdot 2{,}93}{5{,}5} = 5{,}67 - 1{,}95 = 3{,}72;$$

$$\bar{y} + \frac{t_{\gamma} S_y}{\sqrt{n}} = 5{,}67 + \frac{3{,}659 \cdot 2{,}93}{5{,}5} = 5{,}67 + 1{,}95 = 7{,}62.$$

Таким чином, довірчий інтервал для $\bar{Y}_{\Gamma}$ буде в таких межах:

$$3{,}72 < \bar{Y}_{\Gamma} < 7{,}62.$$

Довірчий інтервал з надійністю γ = 0,999 для σ_{Γ} буде таким:

$$S_y(1 - q(\gamma; n)) < \sigma_{\Gamma} < S_y(1 - q(\gamma; n)).$$

Нам відоме значення $S_y = 2{,}93$. Враховуючи, що γ = 0,999, n = 30, знайдемо за таблицею (додаток 5) значення $q(\gamma = 0{,}999,\ n = 30) = 0{,}63$.

Визначимо кінці довірчого інтервалу:

$$S_y(1 - q(\gamma; n)) = 2{,}93(1 - 0{,}63) = 2{,}93 \cdot 0{,}37 = 1{,}084;$$

$$S_y(1 + q(\gamma; n)) = 2{,}93(1 + 0{,}63) = 2{,}93 \cdot 1{,}63 = 4{,}776.$$

Отже, довірчий інтервал для σ_{Γ} подається такою нерівністю:

$$1{,}084 < \sigma_{\Gamma} < 4{,}776.$$

Довірчий інтервал для r_{xy} із заданою надійністю γ = 0,999 буде таким:

$$r_{\text{В}} - t_{\gamma} \frac{1 - r_{\text{В}}^2}{\sqrt{n}} < r_{xy} < r_{\text{В}} + t_{\gamma} \frac{1 - r_{\text{В}}^2}{\sqrt{n}}.$$

Нам відомі значення $r_{\text{В}} = 0{,}67$, $\sqrt{n} = \sqrt{30} \approx 5{,}48$, t_γ визначаємо за таблицею значень функції Лапласа $\Phi(x_\gamma) = 0{,}5\gamma = 0{,}5 \cdot 0{,}999 = 0{,}4995$, де $x_\gamma = 3{,}2$.

Визначимо кінці довірчого інтервалу:

$$r_{\text{В}} - x_\gamma \frac{1 - r_{\text{В}}^2}{\sqrt{n}} = 0{,}67 - 3{,}2 \frac{1 - (0{,}67)^2}{5{,}48} = 0{,}67 - \frac{3{,}2 \cdot 0{,}5511}{5{,}48} =$$
$$= 0{,}67 - 0{,}322 = 0{,}348;$$

$$r_{\text{В}} + x_\gamma \frac{1 - r_{\text{В}}^2}{\sqrt{n}} = 0{,}67 + 3{,}2 \frac{1 - (0{,}67)^2}{5{,}48} = 0{,}67 + \frac{3{,}2 \cdot 0{,}5511}{5{,}48} =$$
$$= 0{,}67 + 0{,}322 = 0{,}992.$$

Таким чином, довірчий інтервал для r_{xy} буде в таких межах:

$$0{,}348 < r_{xy} < 0{,}992.$$

10. Побудова довірчого інтервалу для $\overline{X}_{\text{Г}}$ за допомогою нерівності Чебишова із заданою надійністю γ

У разі, коли відсутня інформація про закон розподілу ознаки генеральної сукупності X, оцінювання ймовірності події $|\overline{x}_{\text{В}} - a| < \delta$, де $a = \overline{X}_{\text{Г}}$, та побудова довірчого інтервалу для $\overline{X}_{\text{Г}}$ із заданою надійністю γ виконуються з використанням нерівності Чебишова за умови, що відоме значення $\sigma_{\text{Г}}$, а саме:

$$P\left(|\overline{x}_{\text{В}} - a| < \delta\right) \geq 1 - \frac{\sigma_{\text{Г}}^2}{n\delta^2} = \gamma. \qquad (436)$$

Із (436) визначаємо величину δ:

$$1 - \frac{\sigma_{\text{Г}}^2}{n\delta^2} = \gamma \;\rightarrow \delta = \frac{\sigma_{\text{Г}}^2}{\sqrt{(1-\gamma)\,n}}. \qquad (437)$$

Довірчий інтервал подається такою нерівністю:

$$\overline{x}_{\text{В}} - \frac{\sigma_{\text{Г}}}{\sqrt{(1-\gamma)\,n}} < a < \overline{x}_{\text{В}} + \frac{\sigma_{\text{Г}}}{\sqrt{(1-\gamma)\,n}}. \qquad (438)$$

Коли $\sigma_Г$ невідоме, застосовуємо виправлену дисперсію S^2, і довірчий інтервал набирає такого вигляду:

$$\bar{x}_В - \frac{S}{\sqrt{(1-\gamma)\,n}} < a < \bar{x}_В + \frac{S}{\sqrt{(1-\gamma)\,n}}. \qquad (439)$$

Приклад. Одержано дані зі 100 навмання вибраних підприємств щодо зростання виробітку на одного робітника x_i (у % відносно попереднього року), які мають такий інтервальний статистичний розподіл:

x_i, %; h = 10	80—90	90—100	100—110	110—120	120—130
n_i	3	14	60	20	4

Використовуючи нерівність Чебишова, побудувати довірчий інтервал для $\bar{X}_Г$, якщо відоме значення $\sigma_Г = 5$ % з надійністю γ = 0,99.

Розв'язання. Для побудови довірчого інтервалу з допомогою нерівності Чебишова необхідно обчислити $\bar{x}_В$, δ. Щоб визначити $\bar{x}_В$, перейдемо від інтервального до дискретного статистичного розподілу, а саме:

x_i	85	95	105	115	125
n_i	3	14	60	20	4

Тоді маємо:

$$\bar{x}_В = \frac{\sum x_i n_i}{n} = \left| n = \sum n_i = 101 \right| = \frac{85\cdot 3 + 95\cdot 14 + 105\cdot 60 + 115\cdot 20 + 125\cdot 4}{101} =$$

$$= \frac{255 + 1330 + 6300 + 2300 + 500}{101} = \frac{10685}{101} = 105{,}8\%.$$

Використовуючи (437), обчислимо δ:

$$\delta = \frac{\sigma_Г}{\sqrt{(1-\gamma)\,n}} = \frac{5}{\sqrt{(1-0{,}99)101}} = \frac{5}{\sqrt{0{,}01\cdot 101}} = 4{,}98\%.$$

Таким чином, довірчий інтервал для $\bar{X}_Г$ подається такими нерівностями:

$$\bar{x}_В - \delta < \bar{X}_Г < \bar{x}_В + \delta,$$

або

$$100{,}8 < \bar{X}_Г < 110{,}8.$$

Приклад. Задані розміри основних фондів x_i на 30-ти підприємствах дискретним статистичним розподілом:

x_i, млн грн.	3	5	7	9
n_i	9	7	10	4

Використовуючи нерівність Чебишова з надійністю γ = 0,99, побудувати довірчий інтервал для $X_{\text{Г}}$.

Розв'язання. Для побудови довірчого інтервалу для $X_{\text{Г}}$ за допомогою нерівності Чебишова необхідно обчислити $\bar{x}_{\text{В}}$, S:

$$\bar{x}_{\text{В}} = \frac{\sum x_i\, n_i}{n} = \frac{3\cdot 9 + 5\cdot 7 + 7\cdot 10 + 9\cdot 4}{30} =$$
$$= \frac{27+35+70+36}{30} = \frac{168}{30} = 5{,}6 \text{ млн грн.}$$

Отже, $\bar{x}_{\text{В}} = 5{,}6$ млн грн.

$$\frac{\sum x_i^2 n_i}{n} = \frac{9\cdot 9 + 25\cdot 7 + 49\cdot 10 + 81\cdot 4}{30} = \frac{81+175+490+324}{30} = \frac{1070}{30} = 35{,}7.$$

$$D_{\text{В}} = \frac{\sum x_i^2\, n_i}{n} - \left(\bar{x}_{\text{В}}\right)^2 = 35{,}7 - \left(5{,}6\right)^2 = 35{,}7 - 31{,}36 = 4{,}34.$$

$$S = \sqrt{\frac{n}{n-1} D_{\text{В}}} = \sqrt{\frac{30}{29} 4{,}34} \approx 2{,}12 \text{ млн грн}.$$

Визначимо кінці довірчого інтервалу:

$$\bar{x}_{\text{В}} - \frac{S}{\sqrt{(1-\gamma)30}} = 5{,}6 - \frac{2{,}12}{\sqrt{(1-0{,}99)30}} = 5{,}6 - 3{,}87 = 1{,}73 \text{ млн грн.};$$

$$\bar{x}_{\text{В}} + \frac{S}{\sqrt{(1-\gamma)30}} = 5{,}6 + \frac{2{,}12}{\sqrt{(1-0{,}99)30}} = 5{,}6 + 3{,}87 = 9{,}47 \text{ млн грн.}$$

Отож, довірчий інтервал для $\bar{X}_{\text{Г}}$ подається нерівностями

$$1{,}73 < \bar{X}_{\text{Г}} < 9{,}47.$$

Теоретичні запитання до теми ?

1. Що називається точковою статистичною оцінкою?
2. Що таке незміщена точкова статистична оцінка?
3. Що таке зміщена точкова статистична оцінка?
4. Що називають ефективною точковою статистичною оцінкою?
5. Що називають ґрунтовною точковою статистичною оцінкою?
6. У чому сутність методу аналогій?
7. У чому сутність методу найменших квадратів?
8. У чому сутність методу максимальної правдоподібності?
9. Що є точковою незміщеною статистичною оцінкою для $\overline{X}_{\Gamma}$?
10. Що означає точкова незміщена статистична оцінка для D_{Γ}?
11. Що називається виправленою дисперсією, виправленим середнім квадратичним відхиленням?
12. Довести, що $M\left(\overline{x}_{\text{В}}\right)=...$
13. Довести, що $M\left(D_{\text{В}}\right)=...$
14. Довести, що $M\left(S_{\text{В}}^{2}\right)=...$
15. Який закон розподілу ймовірностей має випадкова величина $\overline{x}_{\text{В}}$?
16. Який закон розподілу має випадкова величина $\frac{n-1}{\sigma_{\Gamma}^{2}}S^{2}$?
17. Який закон розподілу ймовірностей має випадкова величина $\frac{\sqrt{n-1}}{\sigma_{\Gamma}}S$?
18. Визначення інтервальної статистичної оцінки для параметрів генеральної сукупності.
19. Що називають точністю і надійністю оцінки?
20. Що називають довірчим інтервалом?
21. Який закон розподілу ймовірностей має випадкова величина $\frac{\overline{x}_{B}-a}{\frac{\sigma_{\Gamma}}{\sqrt{n}}}$?
22. Як побудувати довірчий інтервал із заданою надійністю γ при відомому значенні σ_{Γ}?

23. Який закон розподілу ймовірностей має випадкова величина $\dfrac{\overline{x}_{\text{В}} - a}{\dfrac{S}{\sqrt{n}}}$?

24. Як побудувати довірчий інтервал для $\overline{X}_{\Gamma}$ із заданою надійністю γ при невідомому значенні σ_{Γ}?

25. Як побудувати довірчий інтервал для $\overline{X}_{\Gamma}$ із заданою надійністю γ при невідомому значенні σ_{Γ} і при обсягах вибірки $n > 30$?

26. Як побудувати довірчий інтервал із заданою надійністю γ для $D_{\Gamma}, \sigma_{\Gamma}$ при обсягах вибірки $n < 30$?

27. Як побудувати довірчий інтервал із надійністю γ для σ_{Γ}, використовуючи розподіл χ?

28. Як побудувати довірчий інтервал із заданою надійністю γ для r_{xy}?

29. Як побудувати довірчий інтервал для X_{Γ}, використовуючи нерівність Чебишова?

Задачі до теми

1. У будинку відпочинку випадковим способом було відібрано 20 осіб і виміряно їх зріст x_i. Здобуті результати наведено у вигляді інтервального статистичного розподілу:

x_i, см	165,5—170,5	170,5—175,5	175,5—180,5	180,5—185,5
n_i	4	6	8	2

Із надійністю γ = 0,99 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 2$.

Відповідь. $173{,}81 < \overline{X}_{\Gamma} < 176{,}19$.

2. У 25 осіб було виміряно кров'яний тиск x_i (в умовних одиницях). Результати вимірювання наведено у вигляді дискретного статистичного розподілу:

x_i	1,5	1,8	2,3	2,5	2,9	3,3
n_i	2	3	5	8	4	3

Із надійністю γ = 0,999 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 1$.

Відповідь. $1{,}78 < \overline{X}_{\Gamma} < 3{,}14$.

3. Виміряна максимальна місткість конденсаторів x_i (у пікофарадах). Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i пФ	4,0—4,2	4,2—4,4	4,4—4,6	4,6—4,8	4,8—5,0
n_i	2	4	5	9	5

З надійністю $\gamma = 0{,}999$ побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 0{,}5$.

Відповідь. $4{,}24 < \overline{X}_{\Gamma} < 4{,}92$.

4. У 30 телевізорів була виміряна чутливість x_i. Результати вимірювання подано як дискретний статистичний розподіл:

x_i, мкВ	200	250	300	350	400	450	500
n_i	2	7	6	8	4	2	1

Із надійністю γ = 0,99 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 4$.

Відповідь. $376 < \overline{X}_{\Gamma} < 380{,}27$.

5. У 25 випадково відібраних деталей була виміряна відстань у мікронах від центру маси, що міститься на її осі, до зовнішньої поверхні. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

Межі інтервалів x_i, мк	80—96	96—112	112—128	128—144	144—160
n_i	2	5	8	6	4

З надійністю γ = 0,999 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 3$.

Відповідь. $130{,}58 < \overline{X}_{\Gamma} < 134{,}42$.

6. З партії однотипних запобіжників відібрано 24 шт. Вимірювання відхилення від номіналу в кілоомах x_i наведено як дискретний статистичний розподіл:

x_i, кОм	– 1	– 2	1	2	3	4	5
n_i	3	4	4	5	4	3	1

Із надійністю γ = 0,99 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 3$.

Відповідь. $-0{,}3 < \overline{X}_{\Gamma} < 2{,}96$.

7. З партії однотипних плашок навмання було вибрано 28 шт., і в кожній із них була виміряна глибина пазу (канавки) x_i. Результати вимірювання наведено як інтервальний статистичний розподіл:

x_i, мм	2,4—2,6	2,6—2,8	2,8—3,0	3,0—3,2	3,2—3,4
n_i	5	8	9	5	1

З надійністю γ = 0,999 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 0{,}8$.

Відповідь. $2{,}31 < \overline{X}_{\Gamma} < 3{,}33$.

8. 28 однотипних приладів були випробувані щодо їх безвідмовної роботи x_i. Результати вимірювання наведено як дискретний статистичний розподіл:

x_i, год	100	110	120	130	140	150
n_i	10	6	5	4	2	1

З надійністю γ = 0,99 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 4$.

Відповідь. $112{,}63 < \overline{X}_{\Gamma} < 116{,}65$.

9. У 30 випадково вибраних валиках виміряні відхилення їх діаметрів від номіналу x_i. Результати вимірювань наведено як інтервальний статистичний розподіл:

x_i, мм	5—10	10—15	15—20	20—25	25—30
n_i	2	6	10	8	4

З надійністю γ = 0,999 побудувати довірчий інтервал для $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 0{,}8$ мм.

Відповідь. $18{,}004 < \overline{X}_{\Gamma} < 18{,}996$.

10. Навмання вибрано 29 різців, які випробувані на знос. Результати експерименту наведено у вигляді дискретного статистичного розподілу:

x_i, год	2	3	4	5	6	7	8
n_i	10	8	6	2	1	1	1

З надійністю γ = 0,99 побудувати довірчий інтервал $\overline{X}_{\Gamma}$, якщо $\sigma_{\Gamma} = 2$.

Відповідь. $2{,}41 < \overline{X}_{\Gamma} < 4{,}39$.

11. Залежність собівартості *Y* одного примірника книги від тиражу *X* досліджувалась видавництвом. Результати дослідження наведено у вигляді двовимірного статистичного розподілу:

$X = x_i$, тис. прим.	$Y = y_i$, грн.				
	10,15	5,52	4,08	2,85	n_{yi}
1	10	5	5	—	
2	—	15	10	5	
3	—	—	20	10	
4	—	—	5	15	
n_{xj}					

Потрібно:

1. З надійністю γ = 0,99 побудувати інтервали для $\overline{Y}_{\Gamma}$, σ_{Γ}, r_{xy}.

2. Використовуючи нерівність Чебишова побудувати довірчий інтервал для $\overline{X}_{\Gamma}$.

Відповідь. $4{,}0742 < \overline{y}_{\Gamma} < 5{,}1388$; $1{,}681 < \sigma_{\Gamma} < 2{,}417$;

$$-0{,}8036 < r_{xy} < -0{,}4966;\ 1{,}47 < \overline{X}_{\Gamma} < 3{,}53.$$

12. Залежність річної продуктивності праці в розрахунку на одного робітника y_i від енергомісткості праці x_i на підприємствах однієї галузі наведено в таблиці:

y_i, грн.	5,4	5,6	6,2	6,8	7,1	7,8	8,5	9,1	10,5	10,9
x_i, Вт/робітн.	1,8	2,1	2,8	3,0	3,2	3,8	3,9	4,2	4,5	4,8

Продовження табл.

y_i, грн.	11,0	11,6	12,1	12,7	13,2	13,9	14,1	14,6	14,9	15,4
x_i, Вт/робітн.	5,2	5,8	5,9	6,2	6,9	7,2	7,5	8,5	8,8	9,4

Потрібно:

1. Із надійністю γ = 0,999 побудувати довірчі інтервали для $\overline{Y}_{\Gamma}$, σ_{Γ}, r_{xy}. x_i y_i.

2. Використовуючи нерівність Чебишова, побудувати довірчий інтервал для $\overline{X}_{\Gamma}$.

Відповідь. $7{,}69 < \bar{y}_\Gamma < 13{,}45$; $0{,}39 < \sigma_\Gamma < 5{,}99$;

$0{,}946 < r_{xy} < 1$; $-10{,}82 < \bar{X}_\Gamma < 21{,}38$.

13. Середня температура у квітні у Києві *X* і Донецьку *Y* вимірювалась протягом 40 років. Результати вимірювання наведено у вигляді двовимірного статистичного розподілу:

$Y = y_i$	$X = x_i$					
	10	14	18	22	26	n_{yi}
12	2	3	5	—	—	
16	—	8	2	3	2	
18	1	5	2	2	1	
20	—	2	1	1	1	
n_{xj}						

Потрібно:

1. Із надійністю $\gamma = 0{,}99$ побудувати довірчі інтервали для $\bar{y}_\Gamma$, σ_Γ, r_{xy}.

2. Використовуючи нерівність Чебишова побудувати довірчий інтервал для $\bar{X}_\Gamma$.

Відповідь. $14{,}87 < \bar{y}_\Gamma < 17{,}13$; $1{,}742 < \sigma_\Gamma < 3{,}618$;

$-0{,}129 < r_{xy} < 0{,}655$; $10{,}36 < \bar{X}_\Gamma < 24{,}04$.

14. На підприємствах однієї галузі промисловості була досліджена залежність річної продуктивності праці одного робітника *Y* від енергомісткості виробництва *X*. Результати дослідження наведено у вигляді парного статистичного розподілу:

y_i, тис. грн/робітн.	2,88	2,91	2,92	2,96	3,01	3,11	3,21	3,25
x_i, кВт/робітн.	2,07	2,12	2,41	2,59	2,89	2,92	3,01	3,12

Продовження табл.

y_i, тис. грн/робітн.	3,32	3,36	3,42	3,46	3,58	3,88	4,12
x_i, кВт/робітн.	3,21	3,29	3,31	3,35	3,41	3,48	3,81

Потрібно:

1. З надійністю $\gamma = 0{,}999$ побудувати довірчий інтервал для $\bar{y}_\Gamma$, σ_Γ, r_{xy}.

2. Використовуючи нерівність Чебишова, побудувати довірчий інтервал для $\overline{X}_{\Gamma}$.

Відповідь. $2{,}88 < \bar{y}_{\Gamma} < 3{,}7$; $0{,}19 < \sigma_{\Gamma} < 0{,}57$;

$0{,}53 < r_{xy} < 1$; $0 < \overline{X}_{\Gamma} < 9{,}17$.

15. Проводяться випробування міцності 100 волокон залежно від їх товщини. Результати експериментів задано двовимірним статистичним розподілом:

$Y = y_i$	$X = x_i$, мк					
	4100	4300	4500	4700	4900	n_{yi}
6,75	5	5	10	—	—	
6,25	—	5	10	5	—	
5,75	—	—	5	15	10	
5,25	—	—	5	5	10	
4,75	—	—	—	—	10	
n_{xi}						

Потрібно:

1. З надійністю $\gamma = 0{,}99$ побудувати довірчі інтервали для $\bar{y}_{\Gamma}$, σ_{Γ}, r_{xy}.

2. Використовуючи нерівність Чебишова, побудувати довірчий інтервал для $\overline{X}_{\Gamma}$.

Відповідь. $5{,}68 < \bar{y}_{\Gamma} < 6{,}02$; $0{,}505 < \sigma_{\Gamma} < 6{,}02$;

$0{,}17 < r_{xy} < 0{,}63$; $4308{,}17 < \overline{X}_{\Gamma} < 4941{,}83$.

16. Досліджувалась залежність кількості гризунів y_i, що загинули, від концентрації спожитого яду (в умовних одиницях). Результати досліджень наведено у вигляді парного статистичного розподілу:

y_i	32	36	36	42	46	47	49	55	59
x_i	3	3,5	4	4,5	5	5,5	6	6,5	7

Продовження табл.

y_i	62	68	70	73	75	88	92	94	98
x_i	7,5	8	8,5	9	9,5	10,5	11	11,5	12

Потрібно:

1. З надійністю $\gamma = 0{,}999$ побудувати довірчий інтервал для $\bar{y}_{\Gamma}$, σ_{Γ}, r_{xy}.

2. Використовуючи нерівність Чебишова, побудувати довірчий інтервал для $\overline{X}_{\Gamma}$.

Відповідь. $41{,}48 < \overline{y}_{\Gamma} < 82{,}96$; $0{,}844 < \sigma_{\Gamma} < 41{,}356$;

$0{,}9892 < r_{xy} < 1$; $0 < \overline{X}_{\Gamma} < 29{,}12$.

17. На заводі «Азовсталь» вимірювався вміст кремнію (у %) у чавуні за різних температур шлаку:

$Y = y_i$ ($t°$ C)	S_i (%) $X = x_i$					
	0,27	0,32	0,42	0,51	0,65	n_{yi}
1330	2	1	1	1	—	
1340	—	4	2	3	1	
1345	—	—	3	4	3	
1365	—	—	—	1	4	
n_{xj}						

Потрібно:

1. З надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для $\overline{y}_{\Gamma}$, σ_{Γ}, r_{xy}.

2. Застосовуючи нерівність Чебишова, побудувати довірчий інтервал для $\overline{X}_{\Gamma}$.

Відповідь. $1341{,}75 < \overline{y}_{\Gamma} < 1346{,}65$; $2{,}73 < \sigma_{\Gamma} < 6{,}85$; $0{,}3 < r_{xy} < 0{,}92$; $0 < \overline{X}_{\Gamma} < 1$.

18. У навмання вибраних однотипних телевізорів вимірювалась чутливість відео y_i та звукового x_i каналів. Результати перевірки наведено як парний статистичний розподіл:

y_i	250	200	180	160	140	120	110	100	95
x_i	180	230	240	250	300	310	320	330	340

Продовження табл.

y_i	90	85	80	75	80	70	65	60	55
x_i	350	360	370	380	390	400	410	420	430

Потрібно:

1. З надійністю $\gamma = 0{,}999$ побудувати довірчий інтервал для $\overline{y}_{\Gamma}$, σ_{Γ}, r_{xy}.

2. Застосовуючи нерівність Чебишова, побудувати довірчий інтервал для $\overline{X}_{\Gamma}$.

Відповідь. $59{,}26 < \overline{y}_{\Gamma} < 164{,}6$; $0 < \sigma_{\Gamma} < 108{,}2$; $0{,}915 < r_{xy} < 1$;

$0 < \overline{X}_{\Gamma} < 648{,}94$.

19. У лабораторних умовах здійснювався експеримент з метою визначення залежності кількості речовин, що розчинюється у воді y_i від температури останньої x_i. Результати експерименту наведено у вигляді двовимірного статистичного розподілу:

$Y = y_i$	$X = x_i$, °C					
	10	20	30	40	50	n_{yi}
48	—	2	3	5	—	
60	2	1	1	1	5	
63	1	2	1	1	—	
71	—	—	2	2	1	
n_{xi}						

Потрібно:

1. З надійністю γ = 0,99 побудувати довірчий інтервал для $\overline{y}_{\Gamma}$, σ_{Γ}, r_{xy}.

2. Згідно з нерівністю Чебишова побудувати довірчий інтервал для $\overline{X}_{\Gamma}$.

Відповідь. $53{,}59 < \overline{y}_{\Gamma} < 62{,}67$; $4{,}45 < \sigma_{\Gamma} < 12{,}67$;

$-0{,}432 < r_{xy} < 0{,}588$; $10{,}06 < \overline{X}_{\Gamma} < 56{,}54$.

20. Результати проведеного аналізу залежності числа проданих пар чоловічого взуття y_i від його розміру x_i подано у вигляді парного статистичного розподілу:

y_i	25	38	65	95	120	140	152	160	165	175	180	185	190	200
x_i	45	43	42	41	40	39	38,5	39	37,5	37	36,5	36	35,5	35

Потрібно:

1. З надійністю γ = 0,99 побудувати довірчий інтервал для $\overline{y}_{\Gamma}$, σ_{Γ}, r_{xy}.

2. За нерівністю Чебишова, побудувати довірчий інтервал для $\overline{X}_{\Gamma}$.

Відповідь. $87{,}16 < \overline{y}_{\Gamma} < 182{,}87$; $9{,}85 < \sigma_{\Gamma} < 106{,}03$;

$-1 < r_{xy} < -0{,}95$; $30{,}85 < \overline{X}_{\Gamma} < 47{,}01$.

ТЕМА 14. **СТАТИСТИЧНІ ГІПОТЕЗИ**

1. Загальна інформація

Інформація, яку дістають на підставі вибірки, реалізованої із генеральної сукупності, може бути використана для формулювання певних суджень про всю генеральну сукупність. Наприклад, розпочавши виготовляти покришки нового типу для автомобілів, відбирають певну кількість цих покришок і піддають їх певним тестам.

За результатами тестів можна зробити висновок про те, чи кращі нові покришки від покришок старого типу, чи ні. А це, у свою чергу, дає підставу для прийняття рішення: виготовляти їх чи ні.

Такі рішення називають *статистичними*.

Статистичні рішення мають імовірнісний характер, тобто завжди існує ймовірність того, що прийняті рішення будуть помилковими.

Головна цінність прийняття статистичних рішень полягає в тому, що в межах імовірнісних категорій можна об'єктивно виміряти ступінь ризику, що відповідає тому чи іншому рішенню.

Будь-які статистичні висновки, здобуті на підставі обробки вибірки, називають *статистичними гіпотезами*.

2. Параметричні і непараметричні статистичні гіпотези

Статистичні гіпотези про значення параметрів ознак генеральної сукупності називають *параметричними*.

Наприклад, висувається статистична гіпотеза про числові значення генеральної середньої $\overline{X}_{\text{Г}}$, генеральної дисперсії $D_{\text{Г}}$, генерального середнього квадратичного відхилення $\sigma_{\text{Г}}$ та ін.

Статистичні гіпотези, що висуваються на підставі обробки вибірки про закон розподілу ознаки генеральної сукупності, називаються *непараметричними*. Так, наприклад, на підставі обробки вибірки може бути висунута гіпотеза, що ознака генеральної сукупності має нормальний закон розподілу, експоненціальний закон та ін.

3. Нульова й альтернативна гіпотези

Гіпотезу, що підлягає перевірці, називають *основною*. Оскільки ця гіпотеза припускає відсутність систематичних розбіжнос-

тей (нульові розбіжності) між невідомим параметром генеральної сукупності і величиною, що одержана внаслідок обробки вибірки, то її називають *нульовою гіпотезою* і позначають H_0.

Зміст нульової гіпотези записується так:

$$H_0 : \bar{x}_Г = a ;$$

$$H_0 : \sigma_Г = 2 ;$$

$$H_0 : \ r_{xy} = 0{,}95 .$$

Кожній нульовій гіпотезі можна протиставити кілька альтернативних (конкуруючих) гіпотез, які позначають символом H_α, що заперечують твердження нульової. Так, наприклад, нульова гіпотеза стверджує: $H_0 : \bar{x}_Г = a$, а альтернативна гіпотеза — $H_\alpha : \bar{x}_Г > a$, тобто заперечує твердження нульової.

4. Прості і складні статистичні гіпотези

Проста гіпотеза, як правило, належить до параметра ознак генеральної сукупності і є однозначною.

Наприклад, згідно з простою гіпотезою параметр генеральної сукупності дорівнює конкретному числу, а саме:

$$H_0 : \bar{x}_Г = 4 ;$$

$$H_0 : \sigma_Г = 4 .$$

Складна статистична гіпотеза є неоднозначною. Вона може стверджувати, що значення параметра генеральної сукупності належить певній області ймовірних значень, яка може бути дискретною і неперервною.

Наприклад:

$$H_0 : \bar{x}_Г \in [2; 2{,}1; 2{,}2] \quad \text{або} \quad H_0 : \bar{x}_Г \in [5{,}2 \div 6{,}5] .$$

Нульова гіпотеза може стверджувати як про значення одного параметра генеральної сукупності, так і про значення кількох параметрів, а також про закон розподілу ознаки генеральної сукупності.

5. Статистичний критерій. Емпіричне значення критерію

Для перевірки правильності висунутої статистичної гіпотези вибирають так званий статистичний критерій, керуючись яким відхиляють або не відхиляють нульову гіпотезу. Статистичний

критерій, котрий умовно позначають через K, є випадковою величиною, закон розподілу ймовірностей якої нам заздалегідь відомий. Так, наприклад, для перевірки правильності $H_0 : \overline{X}_{\text{Г}} = a$ як статистичний критерій K можна взяти випадкову величину, яку позначають через $K = Z$, що дорівнює

$$Z = \frac{\overline{x}_{\text{В}} - a}{\sigma(\overline{x}_{\text{В}})}, \qquad (440)$$

і яка має нормований нормальний закон розподілу ймовірностей. При великих обсягах вибірки ($n > 30$) закони розподілу статистичних критеріїв наближатимуться до нормального.

Спостережуване значення критерію, який позначають через K^*, обчислюють за результатом вибірки.

6. Область прийняття гіпотези. Критична область. Критична точка

Множину Ω всіх можливих значень статистичного критерію K можна поділити на дві підмножини A і $\overline{A}$, які не перетинаються.

$$(A \bigcup \overline{A} = \Omega, \quad A \bigcap \overline{A} = \varnothing).$$

Сукупність значень статистичного критерію $K \in A$, за яких нульова гіпотеза не відхиляється, називають *областю прийняття нульової гіпотези*.

Сукупність значень статистичного критерію $K \in \overline{A}$, за яких нульова гіпотеза не приймається, називають *критичною областю*.

Отже, A — область прийняття H_0,

$\overline{A}$ — критична область, де H_0 відхиляється.

Точку або кілька точок, що поділяють множину Ω на підмножини A і $\overline{A}$, називають *критичними* і позначають через $K_{\text{кр}}$.

Існують три види критичних областей:

Якщо при $K < K_{\text{кр}}$ нульова гіпотеза відхиляється, то в цьому разі ми маємо лівобічну критичну область, яку умовно можна зобразити (рис. 119).

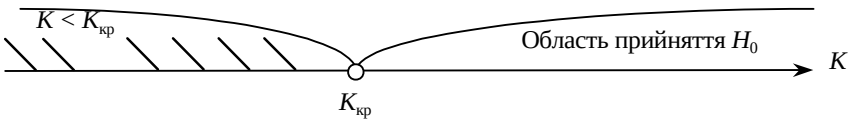


Рис. 119

Якщо при $K > K_{кр}$ нульова гіпотеза відхиляється, то в цьому разі маємо правобічну критичну область (рис. 120).

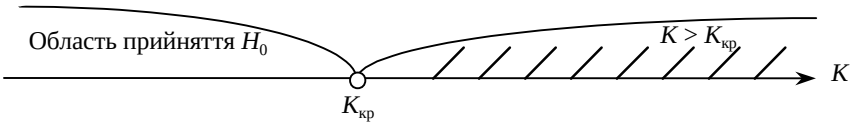


Рис. 120

Якщо ж при $K < K'_{кр}$ і при $K > K''_{кр}$ нульова гіпотеза відхиляється, то маємо двобічну критичну область (рис. 121).

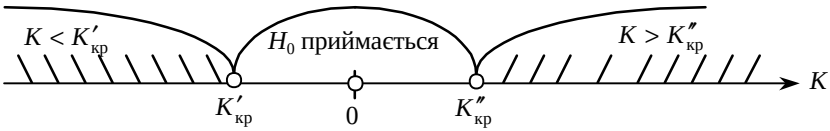


Рис. 121

Лівобічна і правобічна області визначаються однією критичною точкою, двобічна критична область — двома критичними точками, симетричними відносно нуля.

7. Загальний алгоритм перевірки правильності нульової гіпотези

Для перевірки правильності H_0 задається так званий *рівень значущості* α.

α — це мала ймовірність, якою наперед задаються. Вона може набувати значення $\alpha = 0{,}005$; 0,01; 0,001.

В основу перевірки H_0 покладено принцип $P(K \in \overline{A}) = \alpha$, тобто ймовірність того, що статистичний критерій потрапляє в критичну область $\overline{A}$, дорівнює малій імовірності α. Якщо ж виявиться, що $K \in \overline{A}$, а ця подія малоймовірна і все ж відбулася, то немає підстав приймати нульову гіпотезу.

Пропонується такий алгоритм перевірки правильності H_0:

1. Сформулювати H_0 й одночасно альтернативну гіпотезу H_α.
2. Вибрати статистичний критерій, який відповідав би сформульованій нульовій гіпотезі.

3. Залежно від змісту нульової та альтернативної гіпотез будується правобічна, лівобічна або двобічна критична область, а саме:

нехай $H_0 : \bar{x}_r = a$, тоді, якщо

$H_\alpha : \bar{x}_r > a$, то вибирається правобічна критична область, якщо

$H_\alpha : \bar{x}_r < a$, то вибирається лівобічна критична область і коли

$H_\alpha : \bar{x}_r \neq a$, то вибирається двобічна критична область.

4. Для побудови критичної області (лівобічної, правобічної чи двобічної) необхідно знайти критичні точки. За вибраним статистичним критерієм та рівнем значущості α знаходяться критичні точки.

5. За результатами вибірки обчислюється спостережуване значення критерію $K^*_{\text{сп}}$.

6. Відхиляють чи приймають нульову гіпотезу на підставі таких міркувань:

у разі, коли $K^* \in \overline{A}$, а це є малоймовірною випадковою подією, $P(K^* \in \overline{A}) = \alpha$ і, незважаючи на це, вона відбулася, то в цьому разі H_0 відхиляється:

для лівобічної критичної області

$$P(K^*_{\text{сп}} < K_{\text{к р}}) = \alpha\,; \tag{441}$$

для правобічної критичної області

$$P(K^*_{\text{сп}} > K_{\text{кр}}) = \alpha\,; \tag{442}$$

для двобічної критичної області

$$P(K^*_{\text{сп}} < K'_{\text{кр}}) + P(K^*_{\text{сп}} > K''_{\text{кр}}) = \alpha \tag{443}$$

або

$$P(K^*_{\text{сп}} < K'_{\text{кр}}) = P(K^*_{\text{сп}} > K''_{\text{кр}}) = \frac{\alpha}{2}\,, \tag{444}$$

ураховуючи ту обставину, що критичні точки $K'_{\text{кр}}$ і $K''_{\text{кр}}$ симетрично розташовані відносно нуля.

8. Помилки першого та другого роду. Потужність критерію

Якою б не була малою величина α, потрапляння спостережуваного значення $K^*_{\text{сп}}$ у критичну область $\left(K^*_{\text{сп}} \in \overline{A}\right)$ ніколи не буде подією абсолютно неможливою. Тому не виключається той ви-

падок, коли H_0 буде правильною, а $K^*_{сп} \in \overline{A}$, а тому нульову гіпотезу буде відхилено.

Отже, при перевірці правильності H_0 можуть бути допущені помилки. Розрізняють при цьому помилки першого і другого роду.

Якщо H_0 є правильною, але її відхиляють на основі її перевірки, то буде допущена помилка першого роду.

Якщо H_0 є неправильною, але її приймають, то в цьому разі буде допущена помилка другого роду.

Між помилками першого і другого роду існує тісний зв'язок.

Нехай, для прикладу, перевіряється $H_0 : \overline{X}_Г = a$. При великих обсягах вибірки n $\overline{x}_В$, як випадкова величина, закон розподілу ймовірностей якої асимптотично наближатиметься до нормального з числовими характеристиками:

$$M(\overline{x}_В) = a = \overline{X}_Г, \ \sigma(\overline{x}_B) = \frac{\sigma_Г}{\sqrt{n}}.$$

Тому, коли гіпотеза H_0 є правдивою, $M(\overline{x}_В) = a$. Цей розподіл має такий вигляд (рис. 122, крива $f(x; a)$).

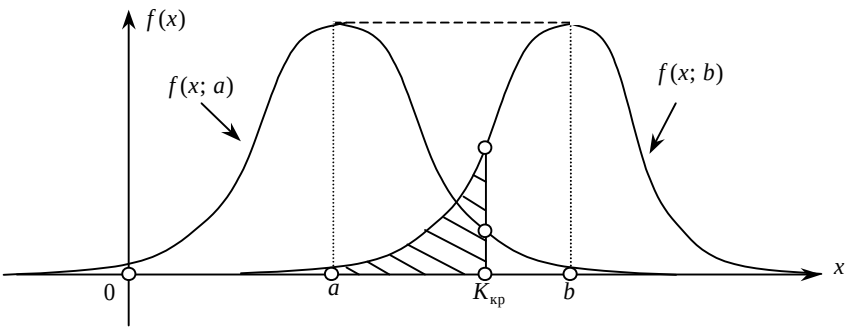


Рис. 122

Коли альтернативна гіпотеза заперечує H_0 і стверджує $H_\alpha : \overline{X}_Г = b > a$, то в цьому разі нормальна крива буде зміщена праворуч (на рис. 122, крива $f(x; b)$).

За вибраним рівнем значущості α визначається критична область (рис. 122).

Коли $\overline{x}_В > K_{кр}$, то H_0 відхиляється з імовірністю помилки першого роду:

$$P(\overline{x}_В > K_{кр}) = \int_{K_{кр}}^{\infty} f(x; a)dx = \alpha. \qquad (445)$$

Коли $\bar{x}_{\text{В}} < K_{\text{кр}}$, то H_0 не відхиляється, хоча може бути правильною альтернативна гіпотеза H_α.

Отже, в цьому разі припускаються помилки другого роду.

Імовірність цієї помилки, яку позначають символом β, може бути визначена на кривій $f(x;\ b)$, а саме:

$$\beta = \int_{-\infty}^{K_{\text{кр}}} f(x;\ b)dx\,. \qquad (446)$$

Ця ймовірність на рис. 122 показана штрихуванням площі під кривою $f(x;\ b)$, що міститься ліворуч $K_{\text{кр}}$.

Якщо з метою зменшення ризику відхилити правильну гіпотезу H_0 зменшуватимемо значення α, то в цьому разі критична точка $K_{\text{кр}}$ зміщуватиметься праворуч, що, у свою чергу, спричинює збільшення ймовірності помилки другого роду, тобто величини β.

Різницю $\pi = 1 - \beta$ називають *імовірністю* обґрунтованого відхилення H_0, або *потужністю* критерію.

Під час розв'язування практичних завдань може виникнути потреба вибору статистичного критерію з їх певної множини. У цьому разі вибирають той критерій, якому притаманна найбільша потужність.

9. Параметричні статистичні гіпотези

9.1. Перевірка правильності нульової гіпотези про значення генеральної середньої

Для перевірки правильності $H_0 : \bar{X}_{\text{Г}} = a,\ (M(x) = a)$, де «*a*» є певним числом, при заданому рівні значущості α насамперед необхідно вибрати статистичний критерій *K*.

Найзручнішим критерієм для цього типу задач є випадкова величина $K = Z$, що має нормований нормальний закон розподілу ймовірностей $N(0;\ 1)$, а саме:

$$Z = \frac{\bar{x}_{\text{В}} - a}{\sigma(\bar{x}_{\text{В}})} = \frac{\bar{x}_{\text{В}} - a}{\dfrac{\sigma_{\text{г}}}{\sqrt{n}}} = \frac{\sqrt{n}\left(\bar{x}_{\text{В}} - a\right)}{\sigma_{\text{г}}}\,. \qquad (447)$$

При розв'язуванні такого класу задач можливий один із трьох випадків:

1) при $H_\alpha : \bar{x}_r > a$ — будується правобічна критична область;

2) при $H_\alpha : \bar{x}_r < a$ — будується лівобічна критична область;

3) при $H_\alpha : \bar{x}_r \neq a$ (тобто може бути $\bar{x}_r < a$, або $\bar{x}_r > a$) — будується двобічна критична область.

Лівобічна і правобічна критичні області визначаються однією критичною точкою, двобічна — двома критичними точками, розташованими симетрично щодо нуля (у цьому разі потужність критерію буде максимальною), будуть рівними між собою за модулем і матимуть протилежні знаки.

Для побудови правобічної критичної області необхідно знайти критичну точку $Z_{\text{кр}}$ за умови $P(Z > z_{\text{кр}}) = \alpha$. Значення $Z_{\text{кр}}$ обчислюємо з рівняння

$$P(0 < Z < z_{\text{кр}}) + P(Z > z_{\text{кр}}) = \frac{1}{2}. \qquad (448)$$

І далі:

$$P(0 < Z < z_{\text{кр}}) + \alpha = \frac{1}{2} \to P(0 < Z < z_{\text{кр}}) = \frac{1-2\alpha}{2} \to$$

$$\to \Phi(z_{\text{кр}}) - \Phi(0) = \frac{1-2\alpha}{2} \to \Phi(z_{\text{кр}}) = \frac{1-2\alpha}{2}, \text{ оскільки } \Phi(0) = 0.$$

За таблицею значень функції Лапласа, скориставшись значенням $\frac{1-2\alpha}{2}$, знаходимо аргумент $x = z_{\text{кр}}$.

Правобічна критична область зображена на рис. 123.

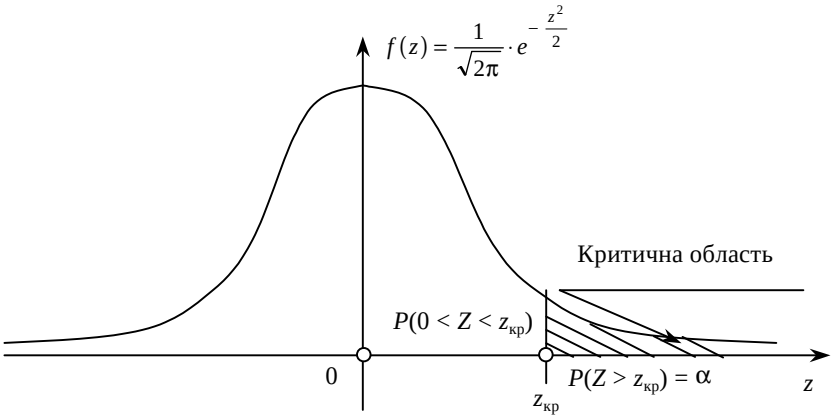


Рис. 123

Для побудови лівобічної критичної області необхідно знайти критичну точку $Z_{кр}$, дотримуючись умови $P(Z < z_{кр}) = \alpha$.

$z_{кр}$ у цьому випадку обчислюється з допомогою рівняння

$$P(Z < z_{кр}) + P(z_{кр} < Z < 0) = \frac{1}{2}. \qquad (449)$$

І далі:

$$\alpha + P(z_{кр} < Z < 0) = \frac{1}{2} \to P(z_{кр} < Z < 0) = \frac{1-2\alpha}{2} \to$$

$$\to \Phi(0) - \Phi(z_{кр}) = \frac{1-2\alpha}{2} \to -\Phi(z_{кр}) = \frac{1-2\alpha}{2} \to \Phi(z_{кр}) = -\frac{1-2\alpha}{2}.$$

Враховуючи ту обставину, що функція Лапласа $\Phi(z)$ є непарною, за таблицею значень $\Phi(z)$ знаходимо аргумент $x = z_{кр}$ і беремо його із знаком «мінус» $(-z_{кр})$. Лівобічна критична область зображена на рис. 124.

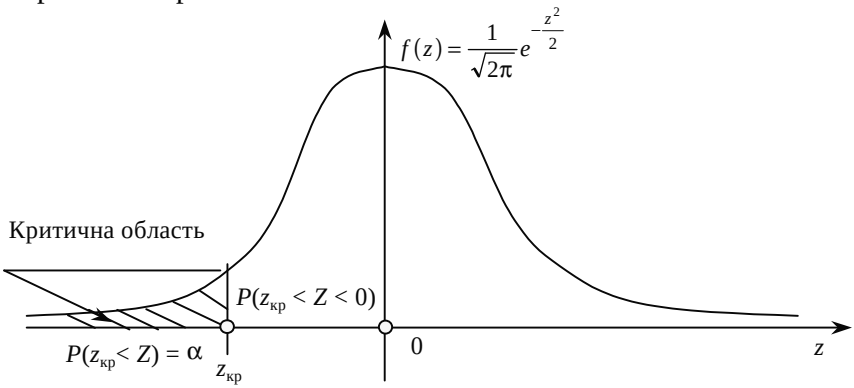


Рис. 124

Для двобічної критичної області необхідно знайти дві критичні точки $z'_{кр}$, $z''_{кр}$ за умови

$$P(Z < z'_{кр}) = \frac{\alpha}{2}, \quad P(Z > z''_{кр}) = \frac{\alpha}{2},$$

де $z'_{кр} = -z''_{кр}$.

Отож, нам необхідно обчислити лише $z''_{кр}$, скориставшись рівнянням

$$P(0 < Z < z''_{кр}) + P(Z > z''_{кр}) = \frac{1}{2}. \qquad (450)$$

І далі:

$$P(0 < Z < z''_{кр}) + \frac{\alpha}{2} = \frac{1}{2} \to P(0 < Z < z''_{кр}) = \frac{1-\alpha}{2} \to$$

$$\to \Phi(z''_{кр}) - \Phi(0) = \frac{1-\alpha}{2} \to \Phi(z''_{кр}) = \frac{1-\alpha}{2},$$

де $z''_{кр}$ знаходимо за таблицею значень функції Лапласа.

Двобічна критична область зображена на рис. 125.

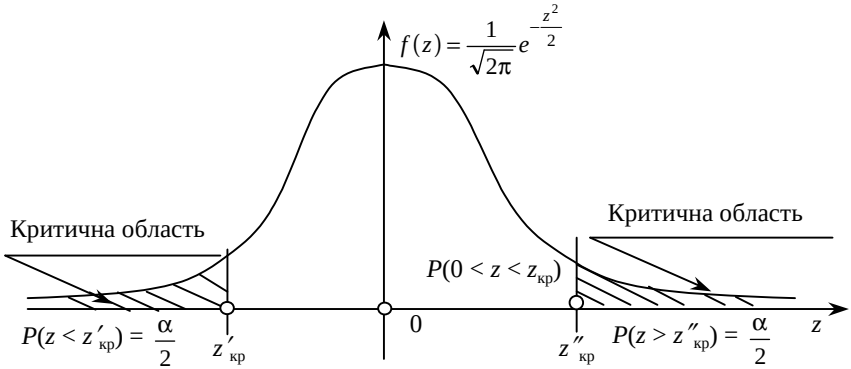


Рис. 125

Розглянутий метод побудови критичних областей придатний лише за умови, коли відоме значення середнього квадратичного відхилення $\sigma_Г$ ознаки генеральної сукупності. При цьому спостережуване значення критерію обчислюється так:

$$z^* = \frac{\bar{x}_В - a}{\frac{\sigma_Г}{\sqrt{n}}}. \qquad (451)$$

У випадку, коли значення $\sigma_Г$ є невідомим, його замінюють статистичною оцінкою

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2 n_i}{n-1}}.$$

Тоді за статистичний критерій вибирається випадкова величина $K = t$, що має розподіл Стьюдента з $k = n - 1$ ступенями свободи, а саме:

$$t = \frac{\bar{x}_В - a}{\frac{S}{\sqrt{n}}}. \qquad (452)$$

Критичні точки у цьому разі визначаються за таблицею (додаток 6) заданим рівнем значущості α та числом ступенів свободи $k = n - 1$. Спостережуване значення критерію обчислюється за формулою

$$t^* = \frac{\bar{x}_{\text{В}} - a}{\frac{S}{\sqrt{n}}}.$$

Приклад. Розбіжність вимірів діаметрів кульок $X = x_i$ є випадковою величиною, що має закон розподілу $N(a;\ 4)$. При рівні значущості $\alpha = 0{,}01$ перевірити правильність $H_0 : a = 240$ мм, якщо альтернативна гіпотеза $H_\alpha : a > 240$ мм, коли відомо, що $\sigma_{\text{Г}} = 4$ мм і вибіркове середнє значення виміряних у 100 однотипних кульок $\bar{x}_{\text{В}} = 225$ мм.

Розв'язання. Оскільки $H_\alpha : a > 240$ мм, будується правобічна критична область. Для цього необхідно знайти критичну точку і побудувати правобічну критичну область. Для знаходження критичної точки застосовуємо відомий вираз:

$$\Phi(z_{\text{кр}}) = \frac{1 - 2\alpha}{2} = \frac{1 - 2 \cdot 0{,}01}{2} = \frac{1 - 0{,}02}{2} = \frac{0{,}98}{2} = 0{,}49.$$

За значенням $\Phi(z_{\text{кр}}) = 0{,}49$ і скориставшись таблицею (додаток 2) знаходимо $z_{\text{кр}} \approx 2{,}34$. Отже, правобічна критична область матиме вигляд, зображений на рис. 126.

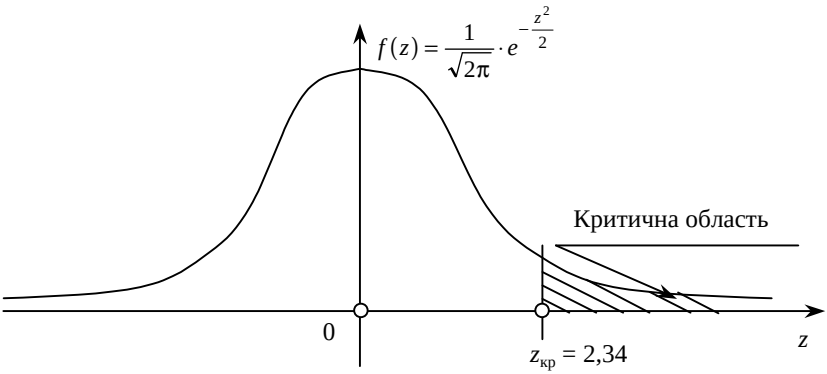


Рис. 126

Обчислимо спостережуване значення критерію за формулою (451) $z^* = \dfrac{\bar{x}_{\text{В}} - a}{\dfrac{\sigma_{\text{Г}}}{\sqrt{n}}}$. Оскільки $\bar{x}_{\text{В}}$ = 225 мм, a = 240 мм, $\sigma_r = 4$ мм, n = 100, маємо

$$z^* = \frac{225 - 240}{\dfrac{4}{\sqrt{100}}} = \frac{-15}{\dfrac{4}{10}} = -\frac{15}{0{,}4} = -\frac{150}{4} = -37{,}5.$$

Висновок. Оскільки $z^* \in]-\infty; 2{,}34]$, то немає підстав для відхилення нульової гіпотези $H_0 : a = 240$ мм. Отже, нульова гіпотеза приймається.

Приклад. Проведено 10 незалежних експериментів над випадковою величиною X, що має нормальний закон розподілу з невідомими значеннями a, σ. Наслідки експериментів подано у вигляді статистичного ряду:

x_i	2,5	2	–2,3	1,9	–2,1	2,4	2,3	–2,5	1,5	–1,7
n_i	1	1	1	1	1	1	1	1	1	1

При рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези
$H_0 : a = 0{,}9$, при альтернативній гіпотезі
$H_\alpha : a < 0{,}9$.

Розв'язання. Запишемо статистичний ряд у вигляді статистичного розподілу й обчислимо $\bar{x}_{\text{В}}$, S:

x_i	–2,5	–2,3	–2,1	–1,7	1,5	1,9	2	2,3	2,4	2,5
n_i	1	1	1	1	1	1	1	1	1	1

$$\bar{x}_{\text{В}} = \frac{\sum x_i}{n} = \frac{-2{,}5 - 2{,}3 - 2{,}1 - 1{,}7 + 1{,}5 + 1{,}9 + 2 + 2{,}3 + 2{,}4 + 2{,}5}{10} = 0{,}4.$$

$$D_{\text{В}} = \frac{\sum x_i^2}{n} - (\bar{x}_{\text{В}})^2 =$$

$$= \frac{6{,}25 + 5{,}29 + 4{,}41 + 2{,}89 + 2{,}25 + 3{,}61 + 4 + 5{,}29 + 5{,}76 + 6{,}25}{10} - (0{,}4)^2 =$$

$$= 4{,}6 - 0{,}16 = 4{,}44\,.$$

$$S^2 = \frac{n}{n-1} D_{\text{В}} = \frac{10}{9} \cdot 4{,}44 = 4{,}933 .$$

$$S = \sqrt{4{,}933} \approx 2{,}22 .$$

При альтернативній гіпотезі $H_\alpha : a < 0{,}9$ будується лівобічна критична область. Для цього необхідно знайти критичну точку, застосовуючи статистичний критерій (451). За таблицею (додаток 6) знаходимо значення

$$t_{\text{кр}}(\alpha = 0{,}001,\ k = n-1 = 10-1 = 9) = t(\alpha = 0{,}001,\ k = 9) = 4{,}78 .$$

Оскільки щільність ймовірностей для розподілу Стьюдента є парною, то $t_{\text{кр}} = -4{,}78$.

Критична область показана на рис. 127.

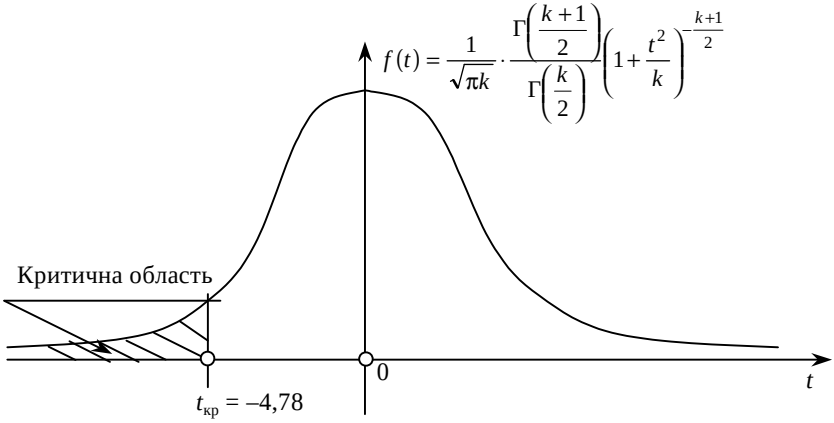


Рис. 127

Обчислимо спостережуване значення критерію:

$$t^* = \frac{\overline{x}_{\text{В}} - a}{\frac{S}{\sqrt{n}}} = \frac{0{,}4 - 0{,}9}{\frac{2{,}22}{\sqrt{10}}} = \frac{0{,}4 - 0{,}9}{\frac{2{,}22}{3{,}16}} = \frac{0{,}4 - 0{,}9}{0{,}702} = -\frac{0{,}5}{0{,}702} = -0{,}712 .$$

Висновок. Оскільки $t^* \in [-4{,}78;\ \infty[$, то немає підстав відхилити $H_0 : a = 0{,}9$.

Приклад. Реалізувавши вибірку з генеральної сукупності, ознака якої X має нормальний закон розподілу, дістали статистичний розподіл:

x_i	6	7	8	9	10	11	12	13	14
n_i	1	3	6	8	6	6	5	3	2

При рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези
$H_0 : a = 8$, якщо альтернативна гіпотеза $H_\alpha : a \neq 8$.

Розв'язання. Обчислимо значення $\bar{x}_В$, S:

$$\bar{x}_В = \frac{\sum x_i\, n_i}{n} = \left| \text{Так як } n = \sum n_i = 40 \right| =$$

$$= \frac{6 + 7\cdot 3 + 8\cdot 6 + 9\cdot 8 + 10\cdot 6 + 11\cdot 6 + 12\cdot 5 + 13\cdot 3 + 14\cdot 2}{40} =$$

$$= \frac{6 + 21 + 48 + 72 + 60 + 66 + 60 + 39 + 28}{40} = 10\,.$$

$$D_В = \frac{\sum x_i^2\, n_i}{n} - (\bar{x}_В)^2 =$$

$$= \frac{36 + 49\cdot 3 + 64\cdot 6 + 81\cdot 8 + 100\cdot 6 + 121\cdot 6 + 144\cdot 5 + 169\cdot 3 + 196\cdot 2}{40} - (10)^2 =$$

$$= \frac{36 + 147 + 384 + 648 + 600 + 726 + 720 + 507 + 392}{40} - 100 =$$

$$= \frac{4160}{40} - 100 = 104 - 100 = 4\,.$$

$$S^2 = \frac{n}{n-1} D_В = \frac{40}{39}\cdot 4 = 4{,}103\,.$$

$$S = \sqrt{4{,}103} \approx 2{,}03\,.$$

При альтернативній гіпотезі $H_\alpha : a \neq 8$ будуємо двобічну критичну область. Враховуючи, що $\sigma_Г$ є невідомою величиною, для побудови цієї області беремо статистичний критерій (452).

Оскільки критичні точки $t'_{кр}$ і $t''_{кр}$ симетричні відносно нуля і при цьому $t'_{кр} = -\,t''_{кр}$, знаходимо за таблицею (додаток 6) $t''_{кр}$:

$$t''_{кр}(\alpha = 0{,}01,\ k = n - 1 = 40 - 1 = 39) = t''_{кр}(\alpha = 0{,}01;\ k = 39) = 2{,}7\,.$$

Тоді $t'_{кр} = -\,2{,}7$.

Двобічна критична область зображена на рис. 128.

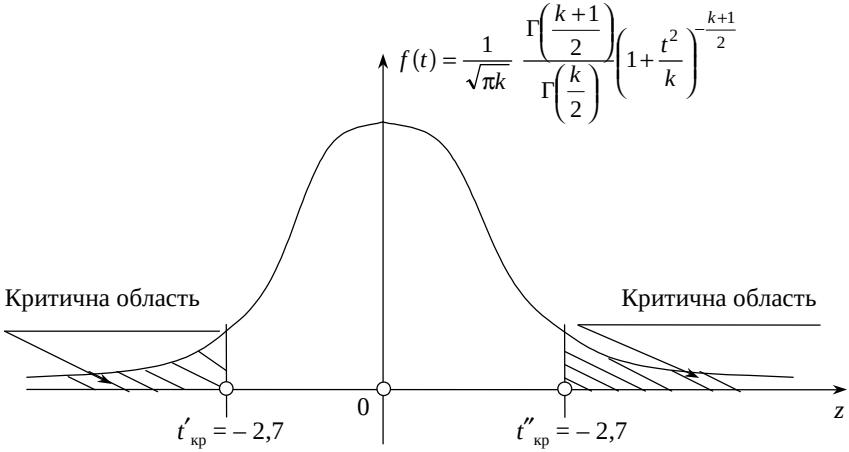


Рис. 128

Обчислимо спостережуване значення критерію:

$$t^*=\frac{\overline{x}_{\text{В}}-a}{\frac{S}{\sqrt{n}}}=\frac{10-8}{\frac{2{,}03}{\sqrt{40}}}=\frac{2}{\frac{2{,}03}{6{,}325}}=\frac{2}{0{,}321}=6{,}23.$$

Висновок. Оскільки $t^*\in[-2{,}708;\ 2{,}708]$, то немає підстав приймати $H_0: a=8$.

Приклад. З генеральної сукупності, ознака якої *X* має закон розподілу *N*(*a*; 5), реалізована вибірка і побудовано статистичний розподіл:

x_i	10,9	11	11,2	11,3	11,5	11,6	11,8	11,9
n_i	2	4	1	3	4	1	2	3

При рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези

$H_0: a=11{,}44$ при альтернативній гіпотезі

$H_\alpha: a\neq 11{,}44$.

Розв'язання. Обчислимо значення $\overline{x}_{\text{В}}$. Оскільки $n=\sum n_i=20$, то дістанемо

$$\bar{x}_{\text{В}} = \frac{\sum x_i n_i}{n} =$$

$$= \frac{10{,}9 \cdot 2 + 11 \cdot 4 + 11{,}2 \cdot 1 + 11{,}3 \cdot 3 + 11{,}5 \cdot 4 + 11{,}6 \cdot 1 + 11{,}8 \cdot 2 + 11{,}9 \cdot 3}{20} =$$

$$= \frac{21{,}8 + 44 + 11{,}2 + 33{,}9 + 46 + 11{,}6 + 23{,}6 + 35{,}7}{20} = \frac{227{,}8}{20} = 11{,}39 .$$

При альтернативній гіпотезі $H_\alpha : a \neq 11{,}44$ будується двобічна критична область. Враховуючи те, що відоме значення $\sigma_{\text{Г}} = 5$, для знаходження критичних точок скористаємося статистичним критерієм $z = \dfrac{\bar{x}_{\text{В}} - a}{\dfrac{\sigma_{\text{Г}}}{\sqrt{n}}}$, що має закон розподілу $N(0; 1)$.

Критична точка $z''_{\text{кр}}$ визначається з рівності

$$\Phi(z''_{\text{кр}}) = \frac{1-\alpha}{2} = \frac{1-0{,}01}{2} = \frac{0{,}99}{2} = 0{,}495 .$$

За значенням функції Лапласа $\Phi(z''_{\text{кр}}) = 0{,}495$ знаходимо $z''_{\text{кр}} = 2{,}58$.

Оскільки $z'_{\text{кр}} = -\ z''_{\text{кр}}$, то маємо $z'_{\text{кр}} = -\ 2{,}58$.

Двобічна критична область зображена на рис. 129.

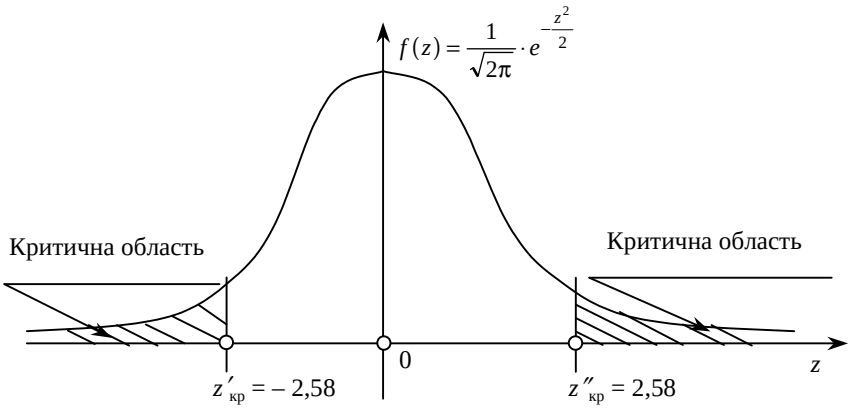


Рис. 129

Обчислимо спостережуване значення критерію

$$z^* = \frac{\bar{x}_{\text{В}} - a}{\dfrac{\sigma_{\text{Г}}}{\sqrt{n}}} = \frac{11{,}39 - 11{,}44}{\dfrac{5}{\sqrt{20}}} = -\frac{0{,}05}{1{,}119} \approx -0{,}045 .$$

Висновок. Оскільки $z^* \in [-2{,}58;\ 2{,}58]$, немає підстав відхиляти $H_0 : a = 11{,}44$.

При великих обсягах вибірки ($n > 40$) статистичний критерій $z = \dfrac{\bar{x}_{\text{В}} - a}{\dfrac{S}{\sqrt{n}}}$, що має закон розподілу Стьюдента з $k = n - 1$ ступенями свободи, наближається асимптотично до нормованого нормального закону $N(0;\ 1)$. У цьому разі критичні точки визначаються з рівностей (448) — (450).

Приклад. Реалізувавши вибірку з генеральної сукупності, елементами якої є однотипні заготівки, довжина яких X є випадковою величиною з нормальним законом розподілу, дістали статистичний розподіл:

x_i	6,5	8,5	10,5	12,5	14,5	16,5
n_i	10	20	30	20	10	10

Якщо рівень значущості $\alpha = 0{,}001$, перевірити правильність $H_0 : a = 15{,}5$ при альтернативній гіпотезі $H_\alpha : a > 15{,}5$.

Розв'язання. Обчислимо значення $\bar{x}_{\text{В}}$, S. Оскільки $n = \sum n_i = 100$, то маємо

$$\bar{x}_{\text{В}} = \frac{\sum x_i n_i}{n} = \frac{6{,}5 \cdot 10 + 8{,}5 \cdot 20 + 10{,}5 \cdot 30 + 12{,}5 \cdot 20 + 14{,}5 \cdot 10 + 16{,}5 \cdot 10}{100} =$$

$$= \frac{65 + 170 + 315 + 250 + 145 + 165}{100} = \frac{1110}{100} = 11{,}1;$$

$$\frac{\sum x_i^2 n_i}{n} =$$

$$= \frac{42{,}25 \cdot 10 + 72{,}25 \cdot 20 + 110{,}25 \cdot 30 + 156{,}25 \cdot 20 + 210{,}25 \cdot 10 + 272{,}25 \cdot 10}{100} =$$

$$= \frac{422{,}5 + 144{,}5 + 3307{,}5 + 3125 + 2102{,}5 + 2722{,}5}{100} = \frac{13125}{100} = 131{,}25.$$

$$D_{\text{В}} = \frac{\sum x_i^2 n_i}{n} - (\bar{x}_{\text{В}})^2 = 131{,}25 - (11{,}1)^2 = 131{,}25 - 123{,}21 = 8{,}04.$$

$$S^2 = \frac{n}{n-1} D_{\text{В}} = \frac{100}{99} \cdot 8{,}04 = 8{,}12 .$$

$$S = \sqrt{8{,}12} \approx 2{,}85 .$$

Оскільки обсяг вибірки великий (n = 100 > 40), статистичний критерій $t = \dfrac{\overline{x}_{\text{В}} - a}{\dfrac{S}{\sqrt{n}}}$ наближатиметься до закону розподілу $N(0; 1)$.

Тому для визначення критичної точки $t_{\text{кр}} = z_{\text{кр}}$ застосовуємо рівність

$$\Phi(z_{\text{кр}}) = \frac{1 - 2\alpha}{2} = \frac{1 - 2 \cdot 0{,}001}{2} = \frac{0{,}998}{2} = 0{,}499 \to z_{\text{кр}} = 3{,}2 .$$

Правобічна критична область матиме такий вигляд (рис. 130):

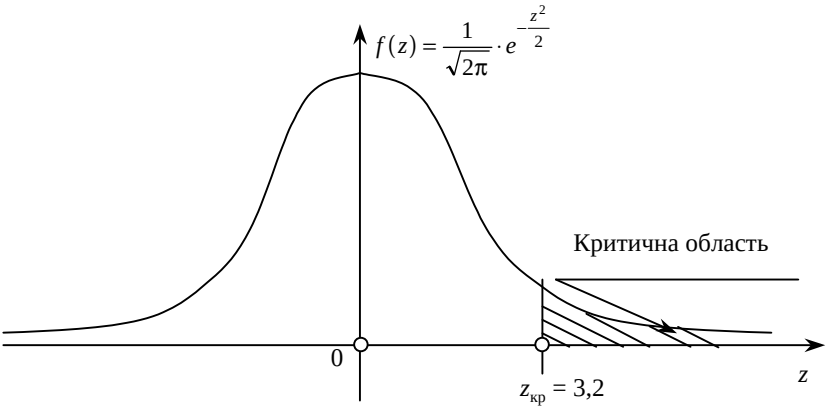


Рис. 130

Обчислимо спостережуване значення критерію

$$z^* = \frac{\overline{x}_{\text{В}} - a}{\dfrac{S}{\sqrt{n}}} = \frac{11{,}1 - 15{,}5}{\dfrac{2{,}85}{\sqrt{100}}} = -\frac{4{,}4}{0{,}285} \approx -15{,}44 .$$

Висновок. Оскільки $z^* \in]-\infty; 5{,}44]$, то $H_0 : a = 15{,}5$ приймається.

9.2. Перевірка правильності нульової гіпотези про рівність двох генеральних середніх (M(X) = M(Y))

Нехай задано дві генеральні сукупності, ознаки яких X і Y мають нормальний закон розподілу і при цьому незалежні одна від одної. Необхідно перевірити правдивість $H_0 : M(X) = M(Y)$ $(\overline{X}_\Gamma = \overline{Y}_\Gamma)$.

Тут можуть спостерігатися два випадки:

Випадок 1. Обсяг вибірки великий ($n > 40$) і відомі значення D_x, D_Γ ознак генеральних сукупностей.

З кожної генеральної сукупності здійснюють вибірку відповідно з обсягами n' і n'' і будують статистичні розподіли:

x_i	x_1	x_2	x_3	.....	x_k
n'_i	n'_1	n'_2	n'_3	.....	n'_k

y_j	y_1	y_2	y_3	.....	y_m
n''_j	n''_1	n''_2	n''_3	.....	n''_m

Тут $n' = \sum n'_i, \quad n'' = \sum n''_j$.

Обчислюються значення

$$\overline{x}_\text{В} = \frac{\sum x_i\, n'_i}{n'}, \qquad \overline{y}_\text{В} = \frac{\sum y_j\, n''_j}{n''}.$$

За статистичний критерій береться випадкова величина

$$Z = \frac{\overline{x}_\text{В} - \overline{y}_\text{В}}{\sigma(\overline{x}_\text{В} - \overline{y}_\text{В})}, \tag{453}$$

що має закон розподілу $N(0; 1)$.

Оскільки $D(\overline{x}_\text{В} - \overline{y}_\text{В}) = \dfrac{D_x}{n'} + \dfrac{D_y}{n''}$, то статистичний критерій (453) набере такого вигляду:

$$Z = \frac{\overline{x}_\text{В} - \overline{y}_\text{В}}{\sqrt{\dfrac{D_x}{n'} + \dfrac{D_y}{n''}}}. \tag{454}$$

Коли $D_x = D_y = D$, дістанемо:

$$Z = \frac{\overline{x}_\text{В} - \overline{y}_\text{В}}{\sigma_\Gamma \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}. \tag{455}$$

Залежно від формулювання альтернативної гіпотези H_α будуються відповідно правобічна, лівобічна та двобічна критичні області.

Спостережуване значення критерію відповідно обчислюється:

$$Z^* = \frac{\bar{x}_В - \bar{y}_В}{\sqrt{\frac{D_x}{n'} + \frac{D_y}{n''}}} \qquad (456)$$

або

$$Z^* = \frac{\bar{x}_В - \bar{y}_В}{\sigma_Г \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}. \qquad (457)$$

Приклад. За заданими статистичними розподілами двох вибірок, реалізованих із двох генеральних сукупностей, ознаки яких мають нормальний закон розподілу зі значенням дисперсій генеральних сукупностей $D_x = 10$; $D_y = 15$,

x_i	12,2	13,2	14,2	15,2	16,2
n'_i	5	15	40	30	10

y_j	8,4	12,4	16,4	20,4	24,4
n''_j	10	15	35	20	20

при рівні значущості $\alpha = 0{,}01$ перевірити правдивість нульової гіпотези

$H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза

$H_\alpha : M(X) > M(Y)$.

Розв'язання. Оскільки $n' = \sum n'_i = 100$; $n'' = \sum n''_j = 100$, обчислимо $\bar{x}_В$, $\bar{y}_В$:

$$\bar{x}_В = \frac{\sum x_i n'_i}{n'} = \frac{12{,}5 \cdot 5 + 13{,}2 \cdot 15 + 14{,}2 \cdot 40 + 15{,}2 \cdot 30 + 16{,}2 \cdot 10}{100} =$$

$$= \frac{62{,}5 + 198 + 568 + 456 + 162}{100} = \frac{1446{,}5}{100} = 14{,}465.$$

$$\bar{y}_В = \frac{\sum y_j n''_j}{n''} = \frac{8{,}4 \cdot 10 + 12{,}4 \cdot 15 + 16{,}4 \cdot 35 + 20{,}4 \cdot 20 + 24{,}4 \cdot 20}{100} =$$

$$= \frac{84 + 186 + 574 + 408 + 488}{100} = \frac{1740}{100} = 17{,}4.$$

Для альтернативної гіпотези $H_\alpha : M(X) > M(Y)$ будується правобічна критична область. Критичну точку $z_{кр}$ знаходимо з рівності

$$\Phi(z_{кр}) = \frac{1-2\alpha}{2} = \frac{1-2\cdot 0{,}01}{2} = \frac{0{,}98}{2} = 0{,}49 \to z_{кр} = 2{,}34 .$$

Правобічна критична область зображена на рис. 131.

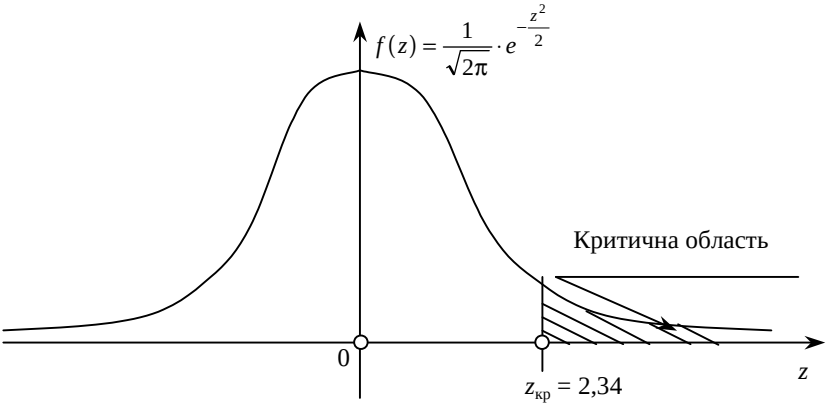


Рис. 131

Обчислимо спостережуване значення критерію

$$Z^* = \frac{\bar{x}_В - \bar{y}_В}{\sqrt{\frac{D_x}{n'} + \frac{D_y}{n''}}} = \frac{14{,}465 - 17{,}4}{\sqrt{\frac{10}{100} + \frac{15}{100}}} = -\frac{2{,}935}{\sqrt{0{,}1 + 0{,}15}} = -\frac{2{,}935}{\sqrt{0{,}25}} =$$

$$= -\frac{2{,}935}{0{,}5} = -5{,}87.$$

Висновок. Оскільки $Z^* \in]-\infty; 2{,}34]$, то $H_0 : M(X) = M(Y)$ не відхиляється.

Приклад. Ознаки X і Y двох генеральних сукупностей, елементами яких є однотипні заклепки, мають нормальний закон розподілу зі значеннями дисперсій $D_x = 2{,}2$ мм^2, $D_y = 2{,}8$ мм^2.

При реалізації двох вибірок із генеральних сукупностей дістали статистичні розподіли:

y_i	9,7	9,8	9,9	10	10,1	10,2
n'_i	2	3	5	4	1	1

x_j	8,9	9,2	9,5	9,8	10,1
n''_j	1	4	5	6	4

При рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези
$H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза
$H_\alpha : M(X) < M(Y)$.

Розв'язання. Ураховуючи, що

$$n' = \sum n'_i = 15; \quad n'' = \sum n''_j = 20 \text{, обчислимо}$$

$$\bar{x}_{\text{В}} = \frac{\sum x_j\, n''_j}{n''} = \frac{8{,}9 \cdot 1 + 9{,}2 \cdot 4 + 9{,}5 \cdot 5 + 9{,}8 \cdot 6 + 10{,}1 \cdot 4}{20} =$$

$$= \frac{8{,}9 + 36{,}8 + 47{,}5 + 58{,}8 + 40{,}4}{20} = \frac{192{,}4}{20} = 9{,}62 \text{ мм.}$$

$$\bar{y}_{\text{В}} = \frac{\sum y_i\, n'_i}{n'} = \frac{9{,}7 \cdot 2 + 9{,}8 \cdot 3 + 9{,}9 \cdot 5 + 10 \cdot 4 + 10{,}1 \cdot 1 + 10{,}2 \cdot 1}{15} =$$

$$= \frac{19{,}4 + 29{,}4 + 49{,}5 + 40 + 10{,}1 + 10{,}2}{15} = \frac{158{,}6}{15} \approx 10{,}57 \text{ мм.}$$

При альтернативній гіпотезі $H_\alpha : M(X) < M(Y)$ будуємо лівобічну критичну область, критичну точку для якої знаходимо з рівності

$$\Phi(z_{\text{кр}}) = -\frac{1 - 2\alpha}{2} = -\frac{1 - 2 \cdot 0{,}001}{2} = -\frac{0{,}998}{2} = -0{,}499 \to z_{\text{кр}} = -3{,}2 .$$

Лівобічна критична область зображена на рис. 132.

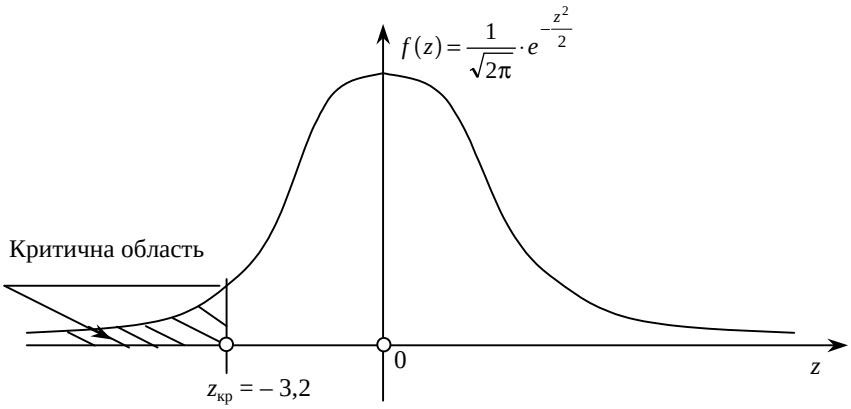


Рис. 132

Обчислюємо спостережуване значення критерію

$$Z^* = \frac{\bar{x}_{\text{В}} - \bar{y}_{\text{В}}}{\sqrt{\frac{D_x}{n''} + \frac{D_y}{n'}}} = \frac{9{,}62 - 10{,}57}{\sqrt{\frac{2{,}2}{20} + \frac{2{,}8}{15}}} = -\frac{0{,}95}{\sqrt{0{,}11 + 0{,}19}} =$$

$$= -\frac{0{,}95}{\sqrt{0{,}3}} = -\frac{0{,}95}{0{,}55} = -1{,}73.$$

Висновок. Оскільки $Z^* \in [-3{,}2; \infty[$, то відсутні підстави для відхилення $H_0 : M(X) = M(Y)$.

Приклад. Для дослідження розтягування певного типу гуми після хімічного оброблення було відібрано шість її мотків, кожний з яких було розділено навпіл і одна його половина була піддана хімічній обробці, а друга — ні.
Потім за допомогою приладу, що вимірює розтягування матеріалу, мотки гуми були виміряні і результати вимірювання наведені у вигляді двох статистичних розподілів ознак X і Y, які мають нормальний закон розподілу з відомими значеннями генеральних дисперсій $D_x = 10$; $D_y = 16$.

y_i	16,7	17,2	17,3	18,1	18,4	19,1
n'_i	1	1	1	1	1	1

x_j	16,2	16,3	17	17,6	18,4
n''_j	1	1	2	1	1

При рівні значущості $\alpha = 0{,}001$ перевірити правдивість нульової гіпотези
$H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза
$H_0 : M(X) \neq M(Y)$.

Розв'язання. Обчислимо значення $\bar{x}_{\text{В}}$, $\bar{y}_{\text{В}}$.
Оскільки $n' = n'' = 6$, то маємо:

$$\bar{y}_{\text{В}} = \frac{\sum y_i}{n'} = \frac{16{,}7 + 17{,}2 + 17{,}3 + 18{,}1 + 18{,}4 + 19{,}1}{6} = \frac{106{,}8}{6} = 17{,}8.$$

$$\bar{x}_{\text{В}} = \frac{\sum x_j n''_j}{n''} = \frac{16{,}2 \cdot 1 + 16{,}3 \cdot 1 + 17 \cdot 2 + 17{,}6 \cdot 1 + 18{,}4 \cdot 1}{6} =$$

$$= \frac{16{,}2 + 16{,}3 + 34 + 17{,}6 + 18{,}4}{6} = \frac{102{,}5}{6} = 17{,}08.$$

При альтернативній гіпотезі $H_\alpha : M(X) \neq M(Y)$ будується двобічна критична область.

Оскільки $z'_{\text{кр}} = -z''_{\text{кр}}$, то $z''_{\text{кр}}$ обчислюємо, використовуючи рівність

$$\Phi(z''_{\text{кр}}) = \frac{1-\alpha}{2} = \frac{1-0{,}001}{2} = \frac{0{,}999}{2} = 0{,}4995 \to z''_{\text{кр}} = 3{,}4 \to z'_{\text{кр}} = -3{,}4\,.$$

Критична область зображена на рис. 133.

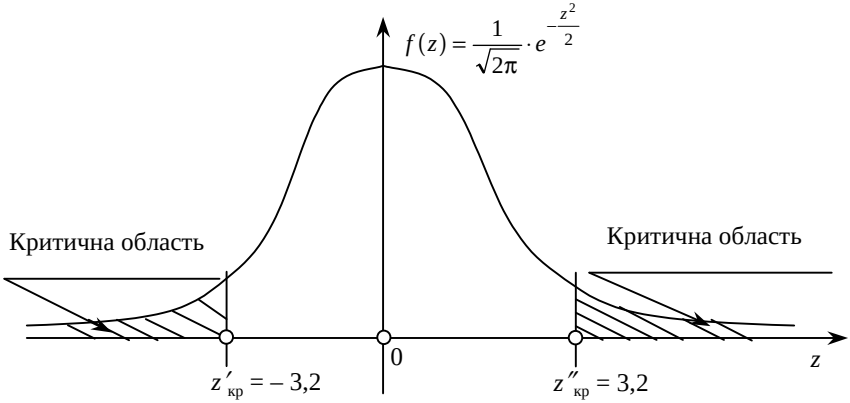


Рис. 133

Обчислимо спостережуване значення критерію

$$Z^* = \frac{\bar{x}_{\text{В}} - \bar{y}_{\text{В}}}{\sqrt{\frac{D_x}{n''} + \frac{D_y}{n'}}} = \frac{17{,}08 - 17{,}8}{\sqrt{\frac{10}{6} + \frac{16}{6}}} = -\frac{0{,}72}{\sqrt{1{,}67 + 2{,}67}} =$$

$$= -\frac{0{,}72}{\sqrt{4{,}34}} = -\frac{0{,}72}{2{,}08} = -0{,}346\,.$$

Висновок. Оскільки $Z^* \in [-3{,}2;\ 3{,}2]$, то немає підстав відхиляти $H_0 : M(X) = M(Y)$.

Випадок 2. Якщо обсяг вибірки великий ($n > 40$), але невідомі значення генеральних дисперсій D_x, D_y, то у цьому випадку застосовують їх точкові незміщені статистичні оцінки, а саме:

$$D(\bar{x}_{\text{В}} - \bar{y}_{\text{В}}) \to S^2 = \frac{\sum (x_j - \bar{x}_{\text{В}}) \cdot n''_j + \sum (y_i - \bar{y}_{\text{В}}) \cdot n'_i}{n' + n'' - 2} =$$

$$= \frac{(n'-1)S_x^2 + (n''-1)S_y^2}{n' + n'' - 2}\,. \qquad (458)$$

При великих обсягах вибірок n', n'' статистичний критерій

$$Z = \frac{\bar{x}_{\text{В}} - \bar{y}_{\text{В}}}{\sqrt{\frac{(n'-1)S_x^2 + (n''-1)S_y^2}{n'+n''-2}} \cdot \sqrt{\frac{1}{n'} + \frac{1}{n''}}} \qquad (459)$$

асимптотично наближається до закону розподілу $N(0; 1)$. Тому для визначення критичних точок застосовується функція Лапласа.

Приклад. З допомогою двох радіовимірних приладів вимірювалась відстань до певного об'єкта. Результати вимірювання наведені у вигляді двох статистичних розподілів ознак: Y — відстань, виміряна першим радіоприладом, та X — другим. При цьому Y і X є незалежними між собою і підпорядковані нормальному закону розподілу. Статистичні розподіли мають такий вигляд:

y_i, км	195	198	201	204	207	210
n'_i	10	20	30	20	15	5

x_j, км	184	188	192	196	200	204
n''_j	5	15	30	40	6	4

При рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези

$H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза

$H_\alpha : M(Y) > M(X)$.

Розв'язання. Значення дисперсій генеральних сукупностей невідомі. Необхідно обчислити $\bar{x}_{\text{В}}$, $\bar{y}_{\text{В}}$, S_x^2, S_y^2.

Оскільки $n' = \sum n'_i = n'' = \sum n''_j = 100$, то

$$\bar{x}_{\text{В}} = \frac{\sum x_j\, n''_j}{n''} = \frac{184 \cdot 5 + 188 \cdot 15 + 192 \cdot 30 + 196 \cdot 40 + 200 \cdot 6 + 204 \cdot 4}{100} =$$

$$= \frac{920 + 2820 + 5760 + 7840 + 1200 + 816}{100} = \frac{19356}{100} = 193{,}56 \text{ км.}$$

$$\frac{\sum x_j^2\, n_j''}{n''} = \frac{184^2 \cdot 5 + 188^2 \cdot 15 + 192^2 \cdot 30 + 196^2 \cdot 40 + 200^2 \cdot 6 + 204^2 \cdot 4}{100} =$$

$$= \frac{3748464}{100} = 37484{,}64.$$

$$D_{\text{В}} = \frac{\sum x_j^2\, n_j'}{n'} - (\bar{x}_{\text{В}})^2 = 37484{,}64 - (193{,}56)^2 =$$

$$= 37484{,}64 - 37465{,}47 = 19{,}17;$$

$$S_x^2 = \frac{n''}{n''-1} D_{\text{В}} = \frac{100}{100-1} \cdot 19{,}17 = 19{,}36\,;$$

$$S_x = \sqrt{19{,}36} \approx 4{,}4\,.$$

$$\bar{y}_{\text{В}} = \frac{\sum y_i\, n_i'}{n'} = \frac{195 \cdot 10 + 198 \cdot 20 + 201 \cdot 30 + 204 \cdot 20 + 207 \cdot 15 + 210 \cdot 5}{100} =$$

$$= \frac{1950 + 3960 + 6030 + 4080 + 3105 + 1050}{100} = \frac{20175}{100} = 201{,}75 \text{ км.}$$

$$\frac{\sum y_i^2\, n_i'}{n'} = \frac{195^2 \cdot 10 + 198^2 \cdot 20 + 201^2 \cdot 30 + 204^2 \cdot 20 + 207^2 \cdot 15 + 210^2 \cdot 5}{100} =$$

$$= \frac{4071915}{100} = 40719{,}15\,;$$

$$D_{\text{В}} = \frac{\sum y_i^2\, n_i'}{n'} - (\bar{y}_{\text{В}})^2 = 40719{,}15 - (201{,}75)^2 =$$

$$= 40719{,}15 - 40703{,}0625 = 16{,}0875\,;$$

$$S_y^2 = \frac{n'}{n'-1} D_{\text{В}} = \frac{100}{100-1} 16{,}0875 = 16{,}25\,;$$

$$S_y = \sqrt{16{,}25} \approx 4{,}03\,.$$

При альтернативній гіпотезі $H_\alpha : M(X) > M(Y)$ будуємо правобічну критичну область, критична точка якої, ураховуючи те, що обсяг вибірки великий, знаходиться з рівності

$$\Phi(z_{\text{кр}}) = \frac{1-2\alpha}{2} = \frac{1-2 \cdot 0{,}01}{2} = \frac{0{,}98}{2} = 0{,}49 \to z_{\text{кр}} = 2{,}34\,.$$

Критична область зображена на рис. 134.

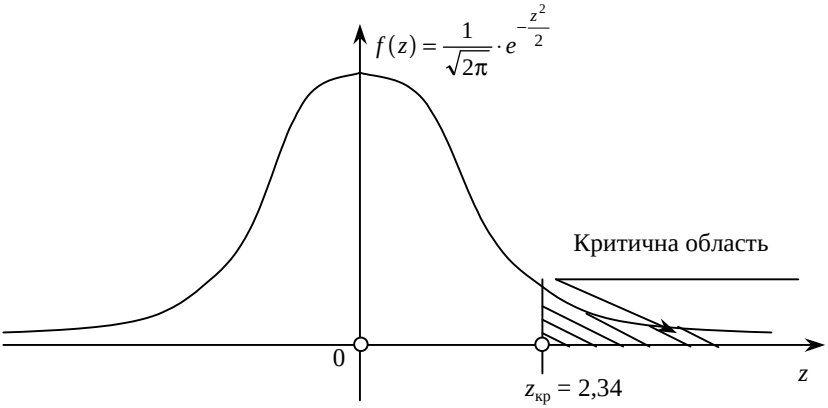


Рис. 134

Спостережуване значення критерію обчислюється так:

$$Z^* = \frac{\bar{x}_В - \bar{y}_В}{\sqrt{\frac{(n'-1)S_x^2 + (n''-1)S_y^2}{n'+n''-2}} \cdot \sqrt{\frac{1}{n'} + \frac{1}{n''}}} =$$

$$= \frac{193{,}56 - 201{,}75}{\sqrt{\frac{99\cdot 19{,}36 + 99\cdot 16{,}25}{100+100-2}} \cdot \sqrt{\frac{1}{100} + \frac{1}{100}}} = -\frac{8{,}19}{\sqrt{4{,}215\cdot 0{,}02}} = -\frac{8{,}19}{0{,}29} = -28{,}24.$$

Висновок. Оскільки $Z^* \in \left]-\infty;\, 2{,}34\right]$, то відсутні підстави для відхилення $H_0 : M(X) = M(Y)$.

9.3. Малий обсяг вибірки ($n' < 40$, $n'' < 40$) і невідомі значення дисперсій генеральної сукупності

При малих обсягах вибірок статистичний критерій

$$z = \frac{\bar{x}_В - \bar{y}_В}{\sqrt{\frac{(n'-1)S_x^2 + (n''-1)S_y^2}{n'+n''-2}} \cdot \sqrt{\frac{1}{n'} + \frac{1}{n''}}}$$

матиме розподіл Стьюдента з $k = n' + n'' - 2$ ступенями свободи. У цьому разі для побудови критичних областей критичні точки знаходять за таблицею (додаток 6).

Приклад. Протягом доби двома приладами вимірювали напругу в електромережі. Результати вимірювання наведено у вигляді статистичних розподілів

y_i	223	227	229	230	235
n_i'	1	2	6	2	1

x_j	216	217	219	228	236
n_j''	2	3	5	1	1

Припускаючи, що випадкові величини X і Y (напруга у вольтах) є незалежними і мають нормальний закон розподілу ймовірностей, за рівня значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези

$H_0 : M(X) = M(Y)$ при альтернативних гіпотезах:

1) $H_\alpha : M(X) > M(Y)$;

2) $H_\alpha : M(X) \neq M(Y)$.

Розв'язання. Обсяги вибірок відповідно дорівнюють $n' = \sum n_i' = 12$, $n'' = \sum n_j'' = 12$.

Обчислимо значення $\bar{x}_{\text{В}}$, $\bar{y}_{\text{В}}$, S_x^2, S_y^2:

$$\bar{y}_{\text{В}} = \frac{\sum y_i\, n_i'}{n'} = \frac{223\cdot 1 + 227\cdot 2 + 229\cdot 6 + 230\cdot 2 + 235\cdot 1}{12} =$$

$$= \frac{223 + 454 + 1374 + 460 + 235}{12} = \frac{2746}{12} = 228{,}83;$$

$$\frac{\sum y_i^2\, n_i'}{n'} = \frac{223^2\cdot 1 + 227^2\cdot 2 + 229^2\cdot 6 + 230^2\cdot 2 + 235^2\cdot 1}{12} =$$

$$= \frac{628458}{12} \approx 52371{,}5;$$

$$D_{\text{В}} = \frac{\sum y_i^2\, n_i'}{n'} - (\bar{y}_B)^2 = 52371{,}5 - (228{,}8)^2 =$$

$$= 52371{,}5 - 52349{,}44 = 22{,}06;$$

$$S_y^2 = \frac{n'}{n'-1} D_{\text{В}} = \frac{12}{12-1}\cdot 22{,}06 \approx 24{,}1;$$

$$\bar{x}_{\text{В}} = \frac{\sum x_j\, n''_j}{n''} = \frac{216\cdot 2 + 217\cdot 3 + 219\cdot 5 + 228\cdot 1 + 236\cdot 1}{12} =$$

$$= \frac{432 + 651 + 1095 + 228 + 236}{12} = \frac{2642}{12} \approx 220{,}17\,;$$

$$\frac{\sum x_j^2\, n''_j}{n''} = \frac{216^2\cdot 2 + 217^2\cdot 3 + 219^2\cdot 5 + 228^2\cdot 1 + 236^2\cdot 1}{12} =$$

$$= \frac{582064}{12} \approx 48505{,}3\,;$$

$$D_{\text{В}} = \frac{\sum x_j^2\, n''_j}{n''} - (\bar{x}_{\text{В}})^2 = 48505{,}3 - (220{,}17)^2 =$$

$$= 48505{,}3 - 48474{,}83 \approx 30{,}47\,;$$

$$S_x^2 = \frac{n''}{n''-1} D_{\text{В}} = \frac{12}{12-1}\cdot 30{,}47 \approx 33{,}24\,.$$

1) Для перевірки правильності нульової гіпотези $H_0 : M(X) = M(Y)$ при альтернативній гіпотезі $H_\alpha : M(X) > M(Y)$ будуємо правобічну критичну область. Ураховуючи, що статистичний критерій має розподіл Стьюдента з $k = n' + n'' - 2 = 12 + 12 - 2 = 22$ та рівнем значущості $\alpha = 0{,}001$, за таблицею (додаток 6) знаходимо критичну точку $z_{\text{кр}}(\alpha = 0{,}001;\ k = 22) = 3{,}79$.

Правобічна критична область зображена на рис. 135.

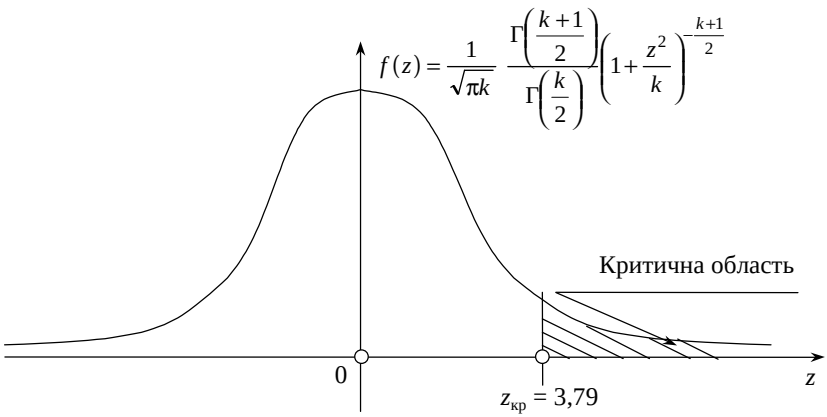


Рис. 135

За формулою (459) обчислюємо спостережуване значення критерію

$$z^* = \frac{\overline{x}_{\text{В}} - \overline{y}_{\text{В}}}{\sqrt{\dfrac{(n'-1)S_x^2 + (n''-1)S_y^2}{n'+n''-2}} \cdot \sqrt{\dfrac{1}{n'} + \dfrac{1}{n''}}} =$$

$$= \frac{220{,}17 - 228{,}8}{\sqrt{\dfrac{11 \cdot 33{,}24 + 11 \cdot 24{,}1}{12+12-2}} \cdot \sqrt{\dfrac{1}{12} + \dfrac{1}{12}}} = \frac{8{,}63}{\sqrt{\dfrac{365{,}64 + 265{,}1}{22}}\sqrt{0{,}17}} =$$

$$= -\frac{8{,}63}{\sqrt{28{,}67 \cdot 0{,}17}} = -\frac{8{,}63}{\sqrt{4{,}8739}} = -\frac{8{,}63}{2{,}21} \approx -3{,}91 .$$

Висновок. Оскільки $z^* \in [-\infty; 3{,}79]$, то $H_0 : M(X) = M(Y)$ приймається.

2) Для альтернативної гіпотези $H_\alpha : M(X) \neq M(Y)$ будується двобічна критична область. Беручи до уваги, що $z'_{\text{кр}} = -z''_{\text{кр}}$, а $z''_{\text{кр}} = 3{,}79$, тоді $z'_{\text{кр}} = -3{,}79$. Двобічна критична область зображена на рис. 136.

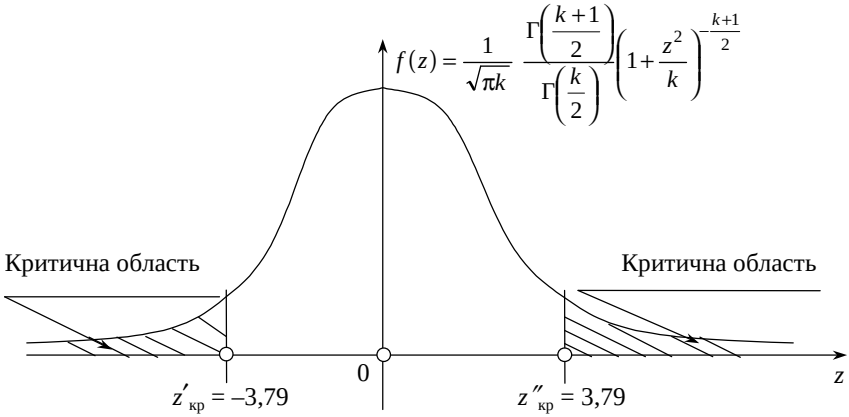


Рис. 136

З попередніх обчислень маємо $z^* = -3{,}91$.

Висновок. Оскільки $z^* \overline{\in}]-3{,}79; 3{,}79]$, то в цьому разі немає підстав для прийняття $H_0 : M(X) = M(Y)$.

Приклад. З двох вибірок обсягом $n' = 16$, $n'' = 14$, реалізованих із двох генеральних сукупностей, ознаки яких X і Y є незалежними і мають нормальний закон розподілу, обчислені значення $\bar{x}_{\text{В}} = 6{,}2$, $\bar{y}_{\text{В}} = 8{,}5$, $S_x^2 = S_y^2 = 4{,}2$.

При рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези

$H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза

$H_\alpha : M(X) > M(Y)$.

Розв'язання. Статистичний критерій у цьому разі є випадковою величиною

$$z = \frac{\bar{x}_{\text{В}} - \bar{y}_{\text{В}}}{\sqrt{\dfrac{(n'-1)S_x^2 + (n''-1)S_y^2}{n'+n''-2}} \cdot \sqrt{\dfrac{1}{n'} + \dfrac{1}{n''}}} =$$

$$= \frac{\bar{x}_{\text{В}} - \bar{y}_{\text{В}}}{\sqrt{\dfrac{S^2(n'-1+n''-1)}{n'+n''-2}} \cdot \sqrt{\dfrac{1}{n'} + \dfrac{1}{n''}}} = \frac{\bar{x}_{\text{В}} - \bar{y}_{\text{В}}}{S \cdot \sqrt{\dfrac{1}{n'} + \dfrac{1}{n''}}},$$

тобто

$$z = \frac{\bar{x}_{\text{В}} - \bar{y}_{\text{В}}}{S \cdot \sqrt{\dfrac{1}{n'} + \dfrac{1}{n''}}},$$

що має розподіл Стьюдента з $k = n' + n'' - 2$ ступенями свободи.

Для альтернативної гіпотези $H_\alpha : M(X) < M(Y)$ будуємо лівобічну критичну точку. $z_{\text{кр}}$ знаходимо за таблицею (додаток 6). Звідси маємо, що $z_{\text{кр}}(\alpha = 0{,}001,\ k = 28) = -3{,}67$.

Критична область зображена на рис. 137.

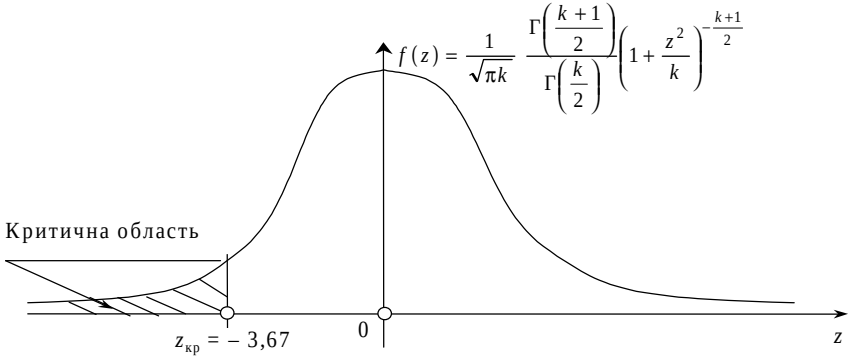


Рис. 137

Обчислимо спостережуване значення критерію

$$z^* = \frac{\bar{x}_{\text{В}} - \bar{y}_{\text{В}}}{S \cdot \sqrt{\frac{1}{n'} + \frac{1}{n''}}} = \frac{6{,}2 - 8{,}5}{4{,}2 \cdot \sqrt{\frac{1}{16} + \frac{1}{14}}} = -\frac{2{,}3}{4{,}2 \cdot \sqrt{0{,}134}} =$$

$$= \frac{2{,}3}{4{,}2 \cdot 0{,}37} = -\frac{2{,}3}{1{,}554} = -1{,}48 .$$

Висновок. Оскільки $z^* \in [-3{,}67; \infty[$, то $H_0 : M(X) = M(Y)$ приймається.

9.4. Перевірка правильності нульової гіпотези про рівність двох дисперсій

Одним із важливих завдань математичної статистики є порівняння двох або кількох вибіркових дисперсій. Таке порівняння дає можливість визначити, чи можна вважати вибіркові дисперсії статистичними оцінками однієї і тієї самої дисперсії генеральної сукупності. Воно застосовується передусім при обчисленні дисперсій за результатами технологічних вимірювань.

Порівняння дисперсій D_x, D_y здійснюється зіставленням виправлених дисперсій S_x^2, S_y^2, які відповідно мають закон розподілу χ^2 із $k_1 = n' - 1$, $k_2 = n'' - 1$ ступенями свободи, де n' і n'' є обсяги першої і другої вибірок.

Нехай перша вибірка здійснена з генеральної сукупності з ознакою *Y*, дисперсія якої дорівнює D_y, друга — з генеральної сукупності з ознакою *X*, дисперсія якої дорівнює D_x. Необхідно перевірити правильність нульової гіпотези

$$H_0 : D_x = D_y .$$

За статистичний критерій береться випадкова величина $F = \frac{S_\delta^2}{S_m^2}$, яка має розподіл Фішера—Снедекора із k_1 і k_2 ступенями свободи, де S_δ^2 є більшою з виправлених дисперсій, одержаною внаслідок обробки результатів вибірок, S_m^2 є меншою з виправлених дисперсій.

Щільність імовірностей розподілу Фішера—Снедекора

$$f(F)=\frac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)}\cdot\left(\frac{k_2}{k_1}\right)^{\frac{k_2}{2}}(F)^{\frac{k_2}{2}-1}\left(1+\frac{k_2}{k_1}F\right)^{-\frac{k_1+k_2}{2}},\ F\geq 0$$

визначена лише на додатній півосі, тобто $0\leq F<\infty$.

Приклад. Під час дослідження стабільності температури в термостаті дістали такі результати: 21,2; 21,8; 21,3; 21,0; 21,4; 21,3.
З метою стабілізації температури було використано удосконалений пристрій, після цього заміри температури показали такі результати: 37,7; 37,6; 37,6; 37,4. Чи можна за рівня значущості $\alpha=0{,}01$ вважати використання удосконаленого пристрою до стабілізатора температури ефективним?

Розв'язання. Очевидно, що ефективність стабілізаторів без удосконаленого пристрою і з ним залежить від дисперсій вимірюваних ними температур. Отже, задача звелась до порівняння двох дисперсій.

Обчислимо виправлені вибіркові дисперсії

$$\bar{y}_{\text{В}}=\frac{\sum y_i\,n_i'}{n'}=\frac{21{,}2+21{,}4+21{,}0+21{,}3\cdot 2+21{,}8}{6}=21{,}333\,;$$

$$\frac{\sum y_i^2\,n_i'}{n'}=\frac{21{,}2^2\cdot 1+21{,}4^2\cdot 1+21{,}0^2\cdot 1+21{,}3^2\cdot 2+21{,}8^2\cdot 1}{6}=$$

$$=\frac{2731{,}02}{6}=455{,}17\,;$$

$$D_{\text{В}}=\frac{\sum y_i^2\,n_i'}{n'}-(\bar{y}_{\text{В}})^2=455{,}17-(21{,}333)^2=455{,}17-455{,}097=0{,}073\,;$$

$$S_y^2=\frac{n'}{n'-1}D_{\text{В}}=\frac{6}{6-1}\cdot 0{,}073=0{,}0876\,;$$

$$\bar{x}=\frac{\sum x_j\,n_j''}{n''}=\frac{37{,}7+37{,}6\cdot 2+37{,}4}{4}=\frac{37{,}7+75{,}2+37{,}4}{4}=$$

$$=\frac{150{,}3}{4}=37{,}575\,;$$

$$\frac{\sum x_j^2\,n_j''}{n''}=\frac{37{,}7^2\cdot 1+37{,}6^2\cdot 2+37{,}4^2\cdot 1}{4}=\frac{5647{,}57}{4}=1411{,}8925\,;$$

$$D_{\text{В}} = \frac{\sum x_j^2 n_j''}{n''} - (\bar{x}_{\text{В}})^2 = 1411{,}8925 - (37{,}575)^2 =$$

$$= 1411{,}8925 - 1411{,}880625 = 0{,}011875\,;$$

$$S_x^2 = \frac{n''}{n''-1} D_{\text{В}} = \frac{4}{4-1} \cdot 0{,}011875 = 0{,}01583\,.$$

Обчислимо спостережуване значення критерію

$$F^* = \frac{S_\delta^2}{S_m^2} = \frac{0{,}0876}{0{,}01583} = 5{,}534\,.$$

Число ступенів свободи для більшої виправленої дисперсії $S_\delta^2 = S_y^2$, $k_1 = n' - 1 = 5$, для меншої $S_m^2 = S_x^2$, $k_2 = n'' - 1 = 3$.

Оскільки удосконалення стабілізатора температур може тільки зменшити дисперсію, то будуємо правобічну критичну область. Отже,

$$H_\alpha : S_y^2 > S_x^2\,.$$

Критичну точку знаходимо за таблицею (додаток 7) відповідно до заданого рівня значущості $\alpha = 0{,}01$ і числа ступенів свободи $k_1 = 5$, $k_2 = 3$, $F_{\text{кр}}(\alpha = 0{,}01;\ k_1 = 5;\ k_2 = 3) = 28{,}2$.

Схематично правобічна критична область зображена на рис. 138.

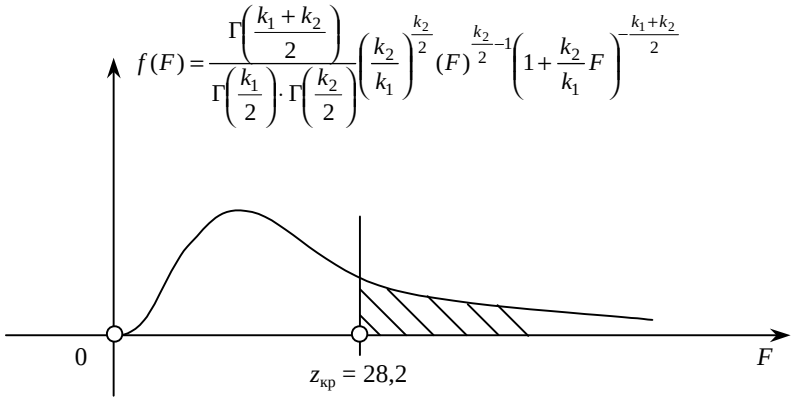


Рис. 138

Висновок. Оскільки $F^* \in]0;\,28{,}5]$, дані спостережень не дають підстав відхилити нульову гіпотезу, тобто вдосконалення термостабілізатора є ефективним.

Приклад. За заданими статистичними розподілами вибірок, які реалізовано з генеральних сукупностей, ознаки яких X і Y є незалежними і мають нормальний закон розподілу,

y_i	1,2	2,2	3,2	4,2	5,2
n_i'	1	2	4	2	3

x_j	0,8	1,6	2,4	3,2	4
n_j''	2	6	1	1	2

при рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези

$H_0 : D_x = D_y$, якщо альтернативна гіпотеза

$H_\alpha : D_x > D_y$.

Розв'язання. Обчислимо значення S_x^2, S_y^2:

$$\bar{y} = \frac{\sum y_i\, n_i'}{n'} = \frac{1{,}2 \cdot 1 + 2{,}2 \cdot 2 + 3{,}2 \cdot 4 + 4{,}2 \cdot 2 + 5{,}2 \cdot 3}{12} =$$

$$= \frac{1{,}2 + 4{,}4 + 12{,}8 + 8{,}4 + 15{,}6}{12} = \frac{42{,}4}{12} \approx 3{,}53;$$

$$\frac{\sum y_i^2\, n_i'}{n'} = \frac{1{,}2^2 \cdot 1 + 2{,}2^2 \cdot 2 + 3{,}2^2 \cdot 4 + 4{,}2^2 \cdot 2 + 5{,}2^2 \cdot 3}{12} = \frac{168{,}48}{12} = 14{,}04;$$

$$D_{\text{В}} = \frac{\sum y_i^2\, n_i'}{n'} - (\bar{y})^2 = 14{,}04 - (3{,}53)^2 = 14{,}04 - 12{,}4609 = 1{,}5791;$$

$$S_y^2 = \frac{n'}{n'-1} D_{\text{В}} = \frac{12}{12-1} \cdot 1{,}5191 = 1{,}723;$$

$$\bar{x} = \frac{\sum x_j\, n_j''}{n''} = \frac{0{,}8 \cdot 2 + 1{,}6 \cdot 6 + 2{,}4 \cdot 1 + 3{,}2 \cdot 1 + 4 \cdot 2}{12} =$$

$$= \frac{1{,}6 + 9{,}6 + 2{,}4 + 3{,}2 + 8}{12} = \frac{24{,}8}{12} = 2{,}067;$$

$$\frac{\sum x_j^2\, n_j''}{n''} = \frac{0{,}8^2 \cdot 2 + 1{,}6^2 \cdot 6 + 2{,}4^2 \cdot 1 + 3{,}2^2 \cdot 1 + 4^2 \cdot 2}{12} = \frac{64{,}64}{12} = 5{,}39.$$

$$D_{\text{В}} = \frac{\sum x_j^2\, n_j''}{n''} - (\bar{x})^2 = 5{,}39 - (2{,}067)^2 = 5{,}39 - 4{,}272489 = 1{,}1175;$$

$$S_x^2 = \frac{n''}{n''-1} D_{\text{В}} = \frac{12}{12-1} \cdot 1{,}1175 \approx 1{,}22 .$$

Обчислимо спостережуване значення критерію

$$F^* = \frac{S_\delta^2}{S_m^2} = \frac{1{,}723}{1{,}22} = 1{,}41 .$$

Для альтернативної гіпотези $H_\alpha : D_x > D_y$ будуємо правобічну критичну область. Знайдемо за таблицею (додаток 7) критичну точку

$$F_{\text{кр}}(\alpha = 0{,}01,\ k_1 = 12-1 = 11,\ k_2 = 12-1 = 11) =$$
$$= F_{\text{кр}}(0{,}01;\ k_1 = 11;\ k_2 = 11) = 4{,}4.$$

Критична область зображена на рис. 139.

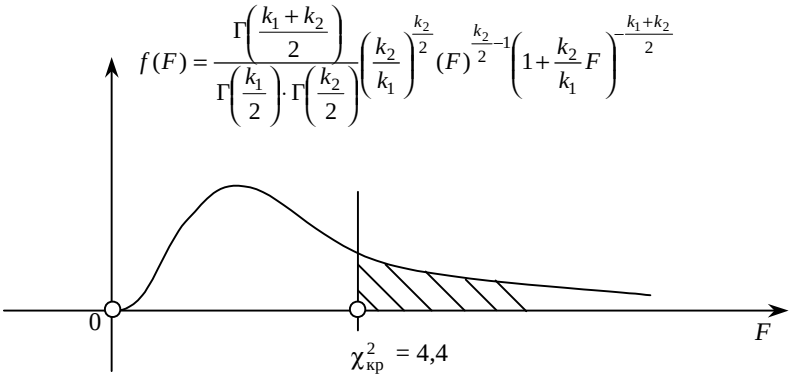


Рис. 139

Висновок. Оскільки $F^* \in [0;\ 4{,}4]$, нульова гіпотеза $H_0 : D_x = D_y$ є правильною.

10. Перевірка правильності непараметричних статистичних гіпотез

Усі перевірки параметричних статистичних гіпотез ґрунтувалися на припущенні, що ознака генеральної сукупності має нормальний закон розподілу ймовірностей і що за іншого розподілу висновки щодо статистичних гіпотез можуть бути хибними.

Тому використання в наведених методах перевірки гіпотез можливе у разі достатньої упевненості, що спостережувана озна-

ка генеральної сукупності має нормальний закон розподілу або близький до нормального.

Основою для висунення гіпотези про закон розподілу ознаки генеральної сукупності може бути наявність теоретичних передумов про характер зміни ознаки. До них, зокрема, відносять виконання умов, що є підґрунтям теореми Ляпунова. У деяких випадках підставою для висунення гіпотези про закон розподілу ознаки генеральної сукупності можуть бути певні формальні властивості здобутого статистичного розподілу, а саме: рівність нулю A_s і E_s для нормального розподілу, рівність вибіркової середньої і вибіркового середнього квадратичного відхилення для експоненціального розподілу.

Інколи підґрунтям для висновків про характер гіпотетичного розподілу можуть бути форми полігону, гістограми.

Приклад. За заданим статистичним розподілом вибірки ознаки X:

$h = 6$	0—6	6—12	12—18	18—24	24—30	30—36
n_i	8	12	30	36	10	4

гіпотетично визначити закон розподілу ознаки генеральної сукупності X.

Розв'язання. Побудуємо гістограму частот для заданого статистичного розподілу вибірки, яка має такий вигляд (рис. 140).

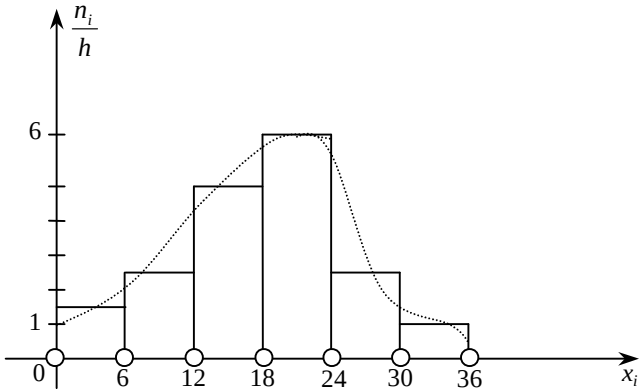


Рис. 140

Якщо з'єднати пунктирною лінією середини кожного прямокутника гістограми, то дістанемо криву лінію, яка певною мірою подібна до

графіка щільності для нормального закону з ненульовим математичним сподіванням. Це може бути підставою для висунення гіпотези про нормальний закон розподілу ознаки X генеральної сукупності. Але цю гіпотезу необхідно перевірити на її правильність.

Приклад. За заданим статистичним розподілом вибірки ознаки X:

$h = 4$	0—4	4—8	8—12	12—16	16—20	20—24
n_i	40	24	16	12	8	4

гіпотетично визначити закон розподілу ознаки генеральної сукупності.

Розв'язання. Побудуємо гістограму частот, записавши статистичний розподіл у такому вигляді:

$h = 4$	0—4	4—8	8—12	12—16	16—20	20—24
$\frac{n_i}{h}$	10	6	4	3	2	1

Гістограма частот має такий вигляд (рис. 141).

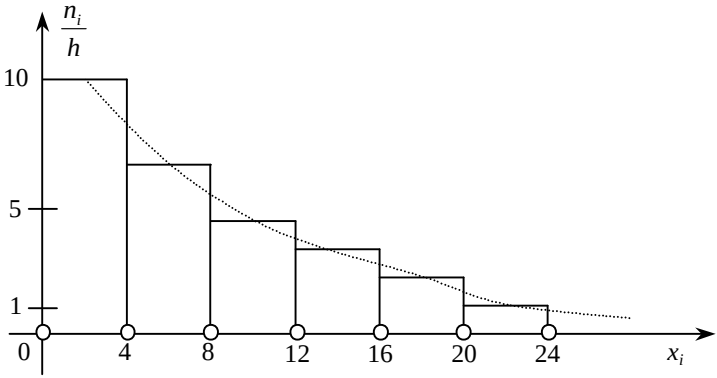


Рис. 141

Коли з'єднаємо послідовно середини кожного прямокутника пунктирною лінією, то дістанемо криву, що в деякому наближенні подібна до графіка щільності ймовірностей для експоненціального закону розподілу. Це дає нам підстави для висунення нульової гіпотези про експоненціальний закон розподілу ознаки генеральної сукупності X, котру, звичайно, необхідно перевірити на правильність. А для цього необхідно мати значення емпіричних і теоретичних частот.

Емпіричними називаються частоти, які спостерігаються при реалізації вибірки, а *теоретичними* — які обчислюються за формулами.

Дискретний закон розподілу. Теоретичні частоти для дискретної випадкової величини обчислюємо за формулою

$$n_i' = nP_i\,, \tag{460}$$

де n — обсяг вибірки;

P_i — імовірність спостережуваного значення $X = x_i$, яка обчислюється за умови, що ознака X має взятий за припущенням закон розподілу ймовірностей.

Приклад. За результатами вибірки, реалізованої з генеральної сукупності, ознака якої X за припущенням має пуассонівський закон розподілу ймовірностей, дістали такий статистичний розподіл:

x_j	0	2	4	6	8
n_i	45	20	15	12	8

Необхідно знайти теоретичні частоти n_i'.

Розв'язання. Для обчислення теоретичних частот застосовуємо формулу Пуассона

$$P_n(k) = \frac{\lambda^k}{k!}e^{-\lambda} \;\to P_n(k) = \frac{a^k}{k!}e^{-a}\,, \tag{461}$$

де $\lambda = a$.

Оскільки для математичного сподівання, тобто для параметра $\lambda = a$, точковою незміщеною статистичною оцінкою є вибіркова середня величина $\overline{x}_{\text{В}}$, обчислимо її значення

$$\overline{x}_{\text{В}} = \frac{\sum x_i\, n_i}{n} = \frac{0\cdot 45 + 2\cdot 20 + 4\cdot 15 + 6\cdot 12 + 8\cdot 8}{100} =$$

$$= \frac{40 + 60 + 72 + 64}{100} = \frac{236}{100} = 2{,}36\,.$$

Отже, $\lambda = 2{,}36 = a$.

Обчислимо ймовірності $P_{100}(k)$, де $k = 0, 2, 4, 6, 8$.

$$P_{100}(0) = \frac{\lambda^0}{0!}e^{-\lambda} = e^{-2{,}36} = 0{,}094\,;$$

$$P_{100}(2)=\frac{\lambda^2}{2!}e^{-\lambda}=\frac{(2{,}36)^2}{2}e^{-2{,}36}=2{,}7848\cdot 0{,}094=0{,}262\,;$$

$$P_{100}(4)=\frac{\lambda^4}{4!}e^{-\lambda}=\frac{(2{,}36)^4}{24}e^{-2{,}36}=1{,}2925\cdot 0{,}094=0{,}121\,;$$

$$P_{100}(6)=\frac{\lambda^6}{6!}e^{-\lambda}=\frac{(2{,}36)^6}{720}e^{-2{,}36}=0{,}240\cdot 0{,}094=0{,}022\,;$$

$$P_{100}(8)=\frac{\lambda^8}{8!}e^{-\lambda}=\frac{(2{,}36)^8}{40320}e^{-2{,}36}=0{,}02386\cdot 0{,}094=0{,}0022\,.$$

Тоді теоретичні частоти будуть такі:

$n_1'=n\cdot P_{100}(0)=100\cdot 0{,}094=9$;

$n_2'=n\cdot P_{100}(2)=100\cdot 0{,}262=26$;

$n_3'=n\cdot P_{100}(4)=100\cdot 0{,}121=12$;

$n_4'=n\cdot P_{100}(6)=100\cdot 0{,}022=2$;

$n_5'=n\cdot P_{100}(8)=100\cdot 0{,}0022=0{,}22\approx 0$.

У підсумку маємо:

Емпіричні частоти n_i	45	20	15	12	8
Теоретичні частоти $n_i'=nP_{100}(k)$	9	26	12	2	0

Як бачимо, велика розбіжність між емпіричними та теоретичними частотами ставить під сумнів припущення про пуассонівський закон розподілу ознаки X генеральної сукупності.

Неперервний закон розподілу. Якщо ознака X генеральної сукупності має неперервний розподіл імовірностей, то теоретичні частоти обчислюються за формулою

$$n_i'=n\ P_i\,,$$

де n — обсяг вибірки,

а P_i — імовірність того, що випадкова величина X потрапить в i-й частковий інтервал. Вона обчислюється за формулами того закону розподілу, який припускаємо на основі обробки статистичного розподілу вибірки.

Так, наприклад, якщо є підстави для припущення, що ознака генеральної сукупності X має нормальний закон розподілу, то теоретичні частоти в цьому разі можна обчислювати за формулами:

$$n_i' = \frac{nh}{\sigma_В} \cdot \varphi(u_i) = \frac{nh}{\sigma_В} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - \bar{x}_B)^2}{2\sigma_B^2}}, \qquad (462)$$

де n — обсяг вибірки;

h — довжина *часткового* інтервалу;

$\bar{x}_В$ — вибіркова середня величина;

$\sigma_В$ — вибіркове середнє квадратичне відхилення;

$\varphi(u_i)$ — щільність імовірностей для загального нормального закону розподілу

або

$$n_i' = n \cdot \left(\Phi\left(\frac{x_{i+1} - \bar{x}_В}{\sigma_В} \right) - \Phi\left(\frac{x_i - \bar{x}_В}{\sigma_В} \right) \right), \qquad (463)$$

де $\Phi\left(\frac{x_{i+1} - \bar{x}_В}{\sigma_В} \right)$, $\Phi\left(\frac{x_i - \bar{x}_В}{\sigma_В} \right)$ — функції Лапласа.

Приклад. У ВТК (відділ технічного контролю) були виміряні 400 валиків із партії, які виготовляє завод. Результати вимірів наведено в таблиці:

x_j, мм	10,4—10,6	10,6—10,8	10,8—11,0	11,0—11,2	11,2—11,4
n_i	40	100	200	40	20

Припускаючи, що ознака X має нормальний закон розподілу ймовірностей, обчислити теоретичні частоти за формулою (461).

Розв'язання. Для обчислення n' необхідно знайти $\bar{x}_В$, $\sigma_В$.

За заданим інтервальним статистичним розподілом будуємо дискретний статистичний розподіл, варіантами якого є середини частинних інтервалів, а саме:

x_j	10,5	10,7	10,9	11,1	11,3
n_i	40	100	200	40	20

Тепер обчислюємо:

$$\bar{x}_В = \frac{\sum x_i n_i}{n} = \frac{10{,}5 \cdot 40 + 10{,}7 \cdot 100 + 10{,}9 \cdot 200 + 11{,}1 \cdot 40 + 11{,}3 \cdot 20}{400} =$$

$$= \frac{420 + 1070 + 2180 + 444 + 226}{400} = \frac{4340}{400} = 10{,}85 \text{ мм};$$

$$\frac{\sum x_i^2 n_i}{n} = \frac{(10{,}5)^2 \cdot 40 + (10{,}7)^2 \cdot 100 + (10{,}9)^2 \cdot 200 + (11{,}1)^2 \cdot 40 + (11{,}3)^2 \cdot 20}{400} =$$

$$= \frac{4410 + 11449 + 23762 + 4928{,}4 + 2553{,}8}{400} = \frac{47103{,}2}{400} = 117{,}758\,;$$

$$D_{\text{В}} = \frac{\sum x_i^2 n_i}{n} - (\bar{x}_{\text{В}})^2 = 117{,}758 - (10{,}85)^2 = 117{,}758 - 117{,}7225 = 0{,}0355;$$

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}} = \sqrt{0{,}0355} \approx 0{,}1884\,.$$

Обчислення теоретичних частот за формулою (463), показаною в таблиці:

x_i	n_i	$u_i = \frac{x_i - 10{,}85}{0{,}1884}$	$\varphi(u_i)$	$n_i' = \frac{nh}{\sigma_{\text{В}}}\varphi(u_i) = 10{,}55 \cdot \varphi(u_i)$
10,5	40	– 1,858	0,0707	30
10,7	100	– 0,796	0,2897	123
10,9	200	0,265	0,3847	163
11,1	40	1,327	0,1647	70
11,3	20	2,388	0,0258	11

Великі розбіжності між емпіричними і теоретичними частотами дають підстави зробити висновок, що припущення про нормальний закон розподілу ознаки *X* генеральної сукупності не має підстав.

Приклад. За поданим інтервальним статистичним розподілом вибірки

$h = 10$	80—90	90—100	100—110	110—120	120—130
n_i	2	14	60	20	4

скориставшись формулою (462), обчислити теоретичні частоти на підставі припущення, що ознака *X* генеральної сукупності *X* має нормальний закон розподілу.

Розв'язання. Значення $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$ обчислюємо за дискретним статистичним розподілом

x_j	95	95	105	115	125
n_i	2	14	60	20	4

$$\bar{x}_{\text{В}} = \frac{\sum x_i\, n_i}{n} = \frac{95\cdot 2 + 95\cdot 14 + 105\cdot 60 + 115\cdot 20 + 125\cdot 4}{100} = \frac{10620}{100} = 106{,}2\,;$$

$$\frac{\sum x_i^2\, n_i}{n} = \frac{18050 - + 126350 + 661500 + 264500 + 62500}{100} = \frac{1132900}{100} = 11329;$$

$$D_{\text{В}} = \frac{\sum x_i^2\, n_i}{n} - (\bar{x}_B)^2 = 11329 - (106{,}2)^2 = 11329 - 11278{,}44 = 50{,}56\,;$$

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}} = \sqrt{50{,}56} \approx 7{,}11\,.$$

Обчислення теоретичних частот за формулою (463) показано в таблиці:

x_i	x_{i+1}	n_i	$z_i = \frac{x_i - \bar{x}_{\text{В}}}{\sigma_{\text{В}}}$	$z_{i+1} = \frac{x_{i+1} - \bar{x}_{\text{В}}}{\sigma_{\text{В}}}$	$\Phi(z_i)$	$\Phi(z_{i+1})$	$n_i' = n(\Phi(z_{i+1}) - \Phi(z_i))$
80	90	2	– 3,68	– 2,28	– 0,499968	– 0,4837	2
90	100	14	– 2,28	– 0,87	– 0,4887	– 0,3078	20
100	110	60	– 0,87	0,53	– 0,3078	0,2019	49
110	120	20	0,53	1,94	0,2019	0,4732	27
120	130	4	1,94	3,35	0,4738	0,49966	3

Результати обчислень дають можливість зробити висновок, що ознака генеральної сукупності гіпотетично має нормальний закон розподілу, оскільки розбіжності між емпіричними та теоретичними частотами є, але вони порівняно незначні. Однак це твердження необхідно ще перевірити, скориставшись відповідними методами математичної статистики.

Приклад. За поданим інтервальним статистичним розподілом вибірки

$h = 8$	0—8	8—16	16—24	24—32	32—40
n_i	40	30	20	8	2

знайти теоретичні частоти, виходячи з припущення, що ознака X генеральної сукупності має експоненціальний закон розподілу.

Розв'язання. Теоретичні частоти обчислюються так:

$$n_i' = nP_i\,,$$

де $P_i = F(x_{i+1}) - F(x_i) = \left(1 - e^{-\lambda x_{i+1}}\right) - \left(1 - e^{-\lambda x_i}\right) = e^{-\lambda x_i} - e^{-\lambda x_{i+1}}$.

Далі маємо

$$n_i' = n\left(e^{-\lambda x_i} - e^{-\lambda x_{i+1}}\right). \qquad (464)$$

Таким чином, для обчислення n_i' необхідно знайти числове значення параметра λ. Ураховуючи, що $M(X) = \frac{1}{\lambda}$ для експоненціального закону, маємо

$$\lambda = \frac{1}{M(X)}. \qquad (465)$$

Отже, для визначення λ необхідно обчислити його точкову незміщену статистичну оцінку $\bar{x}_{\text{В}}$. Тоді $\lambda = \frac{1}{\bar{x}_{\text{В}}}$.

За дискретним статистичним розподілом

x_j	4	12	18	22	34
n_i	40	30	20	8	2

$$\bar{x}_{\text{В}} = \frac{\sum x_i n_i}{n} = \frac{4 \cdot 40 + 12 \cdot 30 + 18 \cdot 20 + 22 \cdot 8 + 34 \cdot 2}{100} =$$

$$= \frac{160 + 360 + 360 + 176 + 68}{100} = \frac{1124}{100} = 11{,}24.$$

Звідси маємо

$$\lambda = \frac{1}{\bar{x}_{\text{В}}} = \frac{1}{11{,}24} = 0{,}089.$$

Обчислення теоретичних частот подано в таблиці:

x_i	x_{i+1}	n_i	$e^{-\lambda x_i}$	$e^{-\lambda x_{i+1}}$	$e^{-\lambda x_i} - e^{-\lambda x_{i+1}}$	$n_i' = n\left(e^{-\lambda x_i} - e^{-\lambda x_{i+1}}\right)$
0	8	40	1	0,491	0,509	51
8	16	30	0,491	0,241	0,25	25
16	24	20	0,241	0,118	0,123	12
24	32	8	0,118	0,058	0,060	6
32	40	2	0,058	0,0028	0,0552	6
40	∞	—	0,0028	0	0,0028	0

Беручи до уваги порівняно невеликі розбіжності між емпіричними і теоретичними частотами, гіпотетично можна стверджувати про експоненціальний закон розподілу ознаки генеральної сукупності X.

Оскільки всі припущення про закон розподілу ознаки генеральної сукупності, розглянуті в прикладах, наведених вище, мають риси гіпотез, а не категоричних тверджень, то вони мають бути перевірені з допомогою критерію узгодженості.

Критерій узгодженості Пірсона. Критерій узгодженості Пірсона є випадковою величиною, що має розподіл χ^2, який визначається за формулою

$$\chi^2 = \sum_{i=1}^{q} \frac{(n_i - np_i)^2}{np_i}, \qquad (466)$$

і має $k = q - m - 1$ ступенів свободи,

де q — число часткових інтервалів інтервального статистичного розподілу вибірки;

m — число параметрів, якими визначається закон розподілу ймовірностей генеральної сукупності згідно з нульовою гіпотезою. Так, наприклад, для закону Пуассона, який характеризується одним параметром λ, $m = 1$, для нормального закону $m = 2$, оскільки цей закон визначається двома параметрами $a = M(X)$ і σ.

Якщо $n_i = np_i$ (усі емпіричні частоти збігаються з теоретичними), то $\chi^2 = 0$, у противному разі $\chi^2 > 0$. Визначивши при заданому рівні значущості α і числу ступенів свободи критичну точку $\chi^2_{\text{кр}}(\alpha; k = q - m - 1)$, за таблицею (додаток 8) будується правобічна критична область. Якщо виявиться, що спостережуване значення критерію $\chi^2_{\text{сп}} > \chi^2_{\text{кр}}$, то H_0 про закон розподілу ознаки генеральної сукупності відхиляється. У противному разі $\left(\chi^2_{\text{сп}} < \chi^2_{\text{кр}}\right)$ H_0 приймається.

Приклад. За заданим інтервальним статистичним розподілом випадкової величини X — маса новонароджених дітей —

$h = 0{,}5$	1—1,5	1,5—2	2—2,5	2,5—3	3—3,5	3,5—4	4—4,5
n_i	10	20	50	35	28	15	12

при рівні значущості $\alpha = 0{,}01$ перевірити правильність H_0 про нормальний закон розподілу ознаки X — маси новонароджених дітей.

Розв'язання. Для визначення теоретичних частот $n_i' = np_i$ необхідно обчислити $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$.

Дискретний статистичний розподіл буде таким:

x_i	1,25	1,75	2,25	2,75	3,25	3,75	4,25
n_i	10	20	50	35	28	15	12

$n = \sum n_i = 170$.

$$\bar{x}_{\text{В}} = \frac{\sum x_i n_i}{n} =$$

$$= \frac{1{,}25 \cdot 10 + 1{,}75 \cdot 20 + 2{,}25 \cdot 50 + 2{,}75 \cdot 35 + 3{,}25 \cdot 28 + 3{,}75 \cdot 15 + 4{,}25 \cdot 12}{170} =$$

$$= \frac{12{,}5 + 35 + 112{,}5 + 96{,}25 + 91 + 56{,}25 + 51{,}0}{170} = \frac{454{,}5}{170} = 2{,}67;$$

$$\frac{\sum x_i^2 n_i}{n} = \frac{1{,}25^2 \cdot 10 + 1{,}75^2 \cdot 20 + 2{,}25^2 \cdot 50 + 2{,}75^2 \cdot 35 + 3{,}25^2 \cdot 28 +}{170}$$

$$\frac{+ 3{,}75^2 \cdot 15 + 4{,}25^2 \cdot 12}{170} = \frac{15{,}625 + 61{,}25 + 253{,}125 + 264{,}6875 + 295{,}75 +}{170}$$

$$\frac{+ 210{,}9375 + 216{,}75}{170} = \frac{1318{,}125}{170} = 7{,}75;$$

$$D_{\text{В}} = \frac{\sum x_i^2 n_i}{n} - (\bar{x}_{\text{В}})^2 = 7{,}75 - (2{,}67)^2 = 7{,}75 - 7{,}1289 = 0{,}6211;$$

$\sigma_{\text{В}} = \sqrt{D_{\text{В}}} = \sqrt{0{,}6211} \approx 0{,}79$.

Обчислення теоретичних частот подано в таблиці:

x_i	x_{i+1}	n_i	$z_i = \frac{x_i - \bar{x}_{\text{В}}}{\sigma_{\text{В}}}$	$z_{i+1} = \frac{x_{i+1} - \bar{x}_{\text{В}}}{\sigma_{\text{В}}}$	$\Phi(z_i)$	$\Phi(z_{i+1})$	$n_i' = n(\Phi(z_{i+1}) - \Phi(z_i))$
1	1,5	10	– 2,11	– 1,48	– 0,4821	– 0,4306	9
1,5	2	20	– 1,48	– 0,85	– 0,4306	– 0,3023	22
2	2,5	50	– 0,85	– 0,22	– 0,3023	0,0871	37
2,5	3	35	– 0,22	0,42	– 0,0871	0,1628	43
3	3,5	28	0,42	1,05	0,1628	0,3531	32
3,5	4	15	1,05	1,68	0,3531	0,4535	17
4	4,5	12	1,68	2,32	0,4535	0,4898	6

Обчислення спостережуваного значення статистичного критерію $\chi^2_{\text{сп}}$ дається нижче в таблиці:

n_i	np_i	$n_i - np_i$	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
10	9	– 1	1	0,11
20	22	– 2	4	0,18
50	37	13	169	4,57
35	43	– 12	144	3,35
28	32	– 4	16	0,5
15	17	– 2	4	0,24
12	6	6	36	6

Отже, маємо

$$\chi^2_{\text{сп}} = \sum \frac{(n_i - np_i)^2}{np_i} = 14{,}95 .$$

За таблицею (додаток 8) знаходимо значення

$$\chi^2_{\text{кр}}(\alpha = 0{,}01;\ k = 7 - 2 - 1 = 4) = \chi^2_{\text{кр}}(0{,}01;\ 4) = 13{,}3 .$$

Правобічна критична область показана на рис. 142.

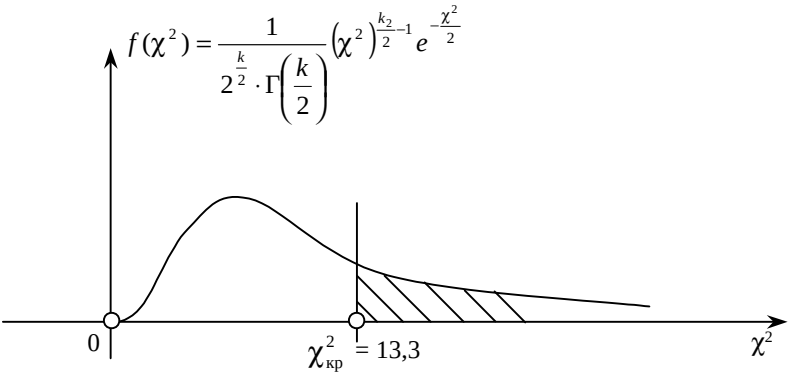


Рис. 142

Висновок. Оскільки $\chi^2_{\text{сп}} \overline{\in} [0; 13{,}3]$, то не маємо підстав для прийняття H_0 про нормальний закон розподілу ознаки генеральної сукупності X.

Приклад. Вимірювання зросту юнаків віком 17 років дав такі результати:

$h = 4$ см	154—158	158—162	162—166	166—170	170—174	174—178	178—182	182—186
n_i	8	14	20	32	12	8	4	2

Визначити гіпотетично, який закон розподілу має ознака X — зріст юнака. При рівні значущості $\alpha = 0{,}01$ перевірити правильність висунутої нульової гіпотези.

Розв'язання. Для заданого статистичного розподілу побудуємо гістограму частот (рис. 143).

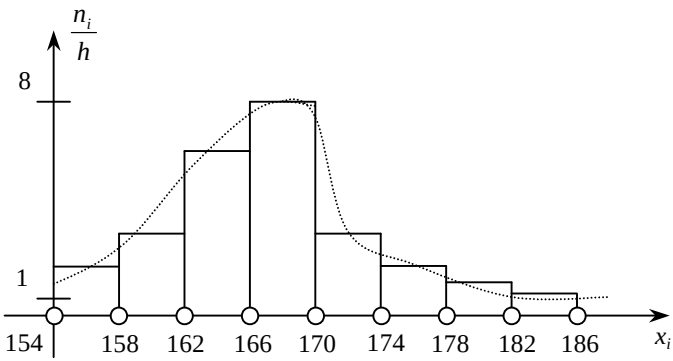


Рис. 143

За формою гістограми частот можемо припустити, що ознака X має нормальний закон розподілу. Отже, висуваємо нульову гіпотезу H_0: ознака X має нормальний закон розподілу ймовірностей.

Для перевірки правильності H_0 використаємо критерій узгодженості Пірсона.

Отже, необхідно обчислити теоретичні частоти, а для цього знайдемо значення $\bar{x}_{\text{В}}$, $\sigma_{\text{В}}$, побудувавши дискретний розподіл за заданим інтервальним, а саме:

x_i	156	160	164	168	172	176	180	184
n_i	8	14	20	32	12	8	4	2

$$\bar{x}_{\text{В}} = \frac{\sum x_i\, n_i}{n} = \frac{156 \cdot 8 + 160 \cdot 14 + 164 \cdot 20 + 168 \cdot 32 +}{100}$$

$$\frac{+172 \cdot 12 + 176 \cdot 8 + 180 \cdot 4 + 184 \cdot 2}{100} = \frac{16704}{100} = 167{,}04 \text{ см};$$

$$\frac{\sum x_i^2 n_i}{n} = \frac{156^2 \cdot 8 + 160^2 \cdot 14 + 164^2 \cdot 20 + 168^2 \cdot 32 + 172^2 \cdot 12 + 176^2 \cdot 8 +}{100}$$

$$\frac{+180^2 \cdot 4 + 184^2 \cdot 2}{100} = \frac{2794304}{100} = 2794304;$$

$$D_{\text{В}} = \frac{\sum x_i^2 n_i}{n} - (\bar{x}_{\text{В}})^2 = 27943{,}04 - (167{,}04)^2 =$$

$$= 27943{,}04 - 27902{,}3616 = 40{,}68;$$

$\sigma_{\text{В}} = \sqrt{D_{\text{В}}} = \sqrt{40{,}68} \approx 6{,}38$ см.

Обчислення теоретичних частот наведено в таблиці:

x_i	x_{i+1}	n_i	$z_i = \frac{x_i - \bar{x}_{\text{В}}}{\sigma_{\text{В}}}$	$z_{i+1} = \frac{x_{i+1} - \bar{x}_{\text{В}}}{\sigma_{\text{В}}}$	$\Phi(z_i)$	$\Phi(z_{i+1})$	$n_i' = n(\Phi(z_{i+1}) - \Phi(z_i))$
154	158	8	– 2,04	– 1,42	– 0,4793	– 0,4222	6
158	162	14	– 1,42	– 0,79	– 0,4222	– 0,2852	14
162	166	20	– 0,79	– 0,16	– 0,2852	– 0,0636	22
166	170	32	– 0,16	0,464	– 0,0636	0,1772	24
170	174	12	0,464	1,09	0,1772	0,3621	19
174	178	8	1,09	1,72	0,3621	0,4573	10
178	182	4	1,72	2,34	0,4573	0,4904	3
182	186	2	2,34	2,97	0,4904	0,4986	1

Обчислення спостережуваного значення $\chi^2_{\text{сп}}$ наведено в таблиці:

n_i	np_i	$n_i - np_i$	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
8	6	2	4	0,667
14	14	0	0	0
20	22	– 2	4	0,182
32	24	8	64	2,667
12	19	– 7	49	2,579
8	10	– 2	4	0,4
4	3	1	1	0,333
2	1	1	1	1

$$\chi^2_{\text{сп}} = \sum_{i=1}^{8} \frac{(n_i - np_i)^2}{np_i} = 7{,}828\,.$$

За таблицею (додаток 8) знаходимо значення

$$\chi^2_{\text{сп}}(\alpha = 0{,}01;\ k = 8 - 2 - 1) = \chi^2_{\text{сп}}(0{,}01; 5) = 15{,}1.$$

Критична область зображена на рис. 144.

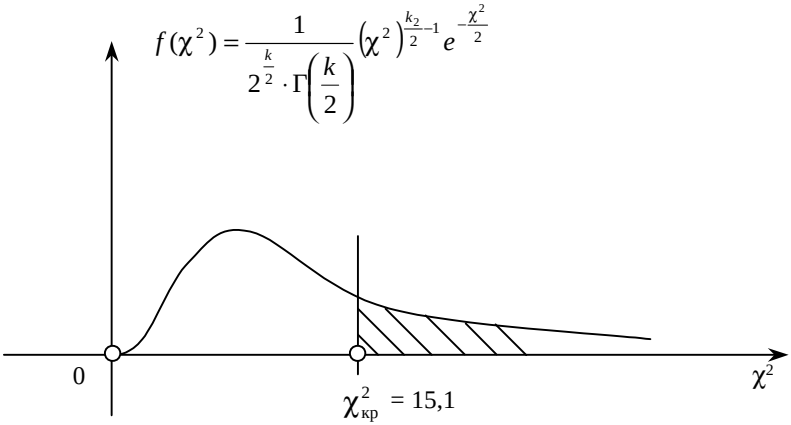


Рис. 144

Висновок. Оскільки $\chi^2_{\text{сп}} \in [0; 15{,}1]$, немає підстав для відхилення нульової гіпотези H_0 про нормальний закон розподілу ймовірностей ознаки X.

Приклад. За заданим статистичним розподілом вибірки:

h = 4 см	0—10	10—20	20—30	30—40	40—50
n_i	40	30	20	6	4

з'ясувати гіпотетично закон розподілу ймовірностей випадкової величини X. При рівні значущості $\alpha = 0{,}01$ перевірити правильність цього припущення.

Розв'язання. Для визначення закону розподілу ознаки X побудуємо гістограму частот (рис. 145).

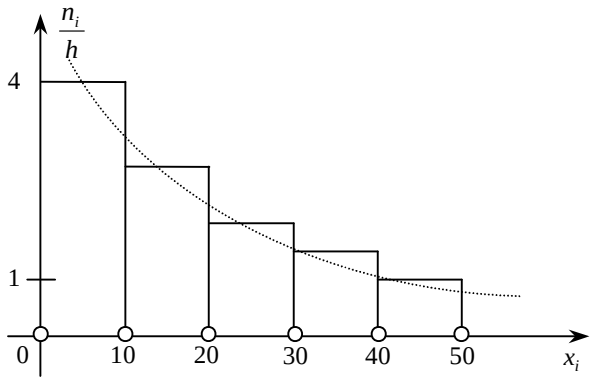


Рис. 145

За формою гістограми частот можна гіпотетично стверджувати, що ознака *X* має експоненціальний закон розподілу ймовірностей.

Для перевірки правильності цього твердження використаємо критерій узгодженості Пірсона. Теоретичні частоти в цьому разі обчислюються за формулою

$$n_i' = n\left(e^{-\lambda x_i} - e^{-\lambda x_{i+1}}\right),$$

де $\lambda = \dfrac{1}{\overline{x}_{\text{В}}}$.

Отже, необхідно обчислити $\overline{x}_{\text{В}}$, побудувавши дискретний статистичний розподіл за наведеним інтервальним, а саме:

x_i	5	15	25	35	45
n_i	40	30	20	6	4

Оскільки $n = \sum n_i = 100$, то

$$\overline{x}_{\text{В}} = \frac{\sum x_i\, n_i}{n} = \frac{5\cdot 40 + 15\cdot 30 + 25\cdot 20 + 35\cdot 6 + 45\cdot 4}{100} =$$

$$= \frac{200 + 450 + 500 + 910 + 180}{100} = 15{,}4\,.$$

Тоді $\lambda = \dfrac{1}{\overline{x}_{\text{В}}} = \dfrac{1}{15{,}4} = 0{,}065$.

Обчислення теоретичних частот наведено в таблиці:

x_i	x_{i+1}	n_i	$e^{-\lambda x_i}$	$e^{-\lambda x_{i+1}}$	$n_i' = n\left(e^{-\lambda x_i} - e^{-\lambda x_{i+1}}\right)$
0	10	40	1	0,522	48
10	20	30	0,522	0,273	25
20	30	20	0,273	0,142	13
30	40	6	0,142	0,074	7
40	50	4	0,074	0,0039	7

Обчислення спостережуваного значення критерію $\chi^2_{\text{сп}}$ наведено в таблиці:

n_i	np_i	$n_i - np_i$	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
40	48	–8	64	1,33
30	25	5	25	1
20	13	7	49	3,77
6	7	– 1	1	0,14
4	7	– 3	9	1,29

$$\chi^2_{\text{сп}} = \sum \frac{(n_i - np_i)^2}{np_i} = 7{,}53 .$$

За таблицею (додаток 8) знаходимо значення критичної точки

$$\chi^2_{\text{кр}}(\alpha = 0{,}01;\ k = 5 - 1 - 1 = 3) = \chi^2_{\text{кр}}(0{,}01;\ 3) = 11{,}3 .$$

Критичну область зображено на рис. 146.

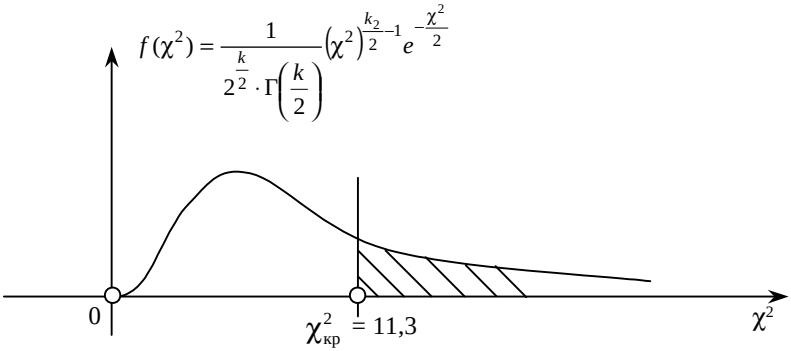


Рис. 146

Висновок. Оскільки $\chi^2_{\text{сп}} \in [0; 11{,}3]$, нульова гіпотеза про експоненціальний закон розподілу ознаки X приймається.

Теоретичні запитання до теми ?

1. Дати визначення нульової та альтернативної гіпотез.
2. Які гіпотези називають параметричними?
3. Які гіпотези називають непараметричними?
4. Що називають простою та складною статистичними гіпотезами?
5. Що називається статистичним критерієм?
6. Що називається емпіричним значенням критерію?
7. Область прийняття нульової гіпотези, критична область, критична точка.
8. Які Ви знаєте критичні області?
9. Загальна методика перевірки правильності нульової гіпотези.
10. Що таке рівень значущості α?
11. Помилки першого та другого роду.
12. Що таке потужність критерію?
13. Перевірка правильності $H_0 : \bar{x}_{\text{Г}} = a$, при альтернативних гіпотезах $H_\alpha : \bar{x}_{\text{Г}} < a;\ \bar{x}_{\text{Г}} > a;\ \bar{x}_{\text{Г}} \neq a$.
14. Який закон розподілу має випадкова величина $z = \dfrac{\bar{x}_{\text{В}} - a}{\sigma(\bar{x}_{\text{В}})}$?
15. Знаходження критичних точок для статистичного критерію $z = \dfrac{\bar{x}_{\text{В}} - a}{\sigma(\bar{x}_{\text{В}})}$.
16. Коли застосовується статистичний критерій $z = \dfrac{\bar{x}_{\text{В}} - a}{\dfrac{S}{\sqrt{n}}}$?
17. Який закон розподілу ймовірностей має статистичний критерій $z = \dfrac{\bar{x}_{\text{В}} - a}{\dfrac{S}{\sqrt{n}}}$?
18. Перевірка правильності $H_0 : M(X) = M(Y)$ при $n > 40$.

19. Який закон розподілу ймовірностей має статистичний критерій $z = \dfrac{\bar{x}_{\text{В}} - \bar{y}_{\text{В}}}{\sigma(\bar{x}_{\text{В}} - \bar{y}_{\text{В}})}$?

20. Який статистичний критерій застосовується для перевірки правдивості $H_0 : M(X) = M(Y)$, якщо $n < 40$?

21. Який закон розподілу ймовірностей має статистичний критерій $z = \dfrac{\bar{x}_{\text{В}} - \bar{y}_{\text{В}}}{\sqrt{\dfrac{(n'-1)S_x^2 + (n''-1)S_y^2}{n' + n'' - 2}}\sqrt{\dfrac{1}{n'} + \dfrac{1}{n''}}}$?

22. Перевірка правильності $H_0 : D_x = D_y$.

23. Який статистичний критерій вибирається для перевірки правильності $H_0 : D_x = D_y$?

24. Який закон розподілу ймовірностей має статистичний критерій $F = \dfrac{S_{\delta}^2}{S_{\text{М}}^2}$?

25. Які існують підстави для висунення гіпотези про закон розподілу ознаки генеральної сукупності?

26. Що називають емпіричними частотами?

27. Що називають теоретичними частотами?

28. Записати формулу для теоретичної частоти, якщо припускається, що ознака *X* має пуассонівський закон розподілу ймовірностей.

29. Записати формулу для обчислення теоретичних частот, якщо припускається, що ознака *X* генеральної сукупності має експоненціальний закон розподілу.

30. Записати формули для обчислення теоретичних частот, якщо припускається, що ознака *X* генеральної сукупності має нормальний закон розподілу.

31. Критерій узгодженості Пірсона.

32. Загальна методика перевірки правильності H_0 про закон розподілу ознаки генеральної сукупності.

Задачі до теми

1. За заданим статистичним розподілом вибірки, реалізованим із генеральної сукупності, ознака *X* якої має нормальний закон розподілу

x_i	4,2	6,2	8,2	10,2	12,2
n_i	6	8	12	8	2

при рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези

$H_0 : M(X) = 10$, якщо альтернативна гіпотеза

$H_\alpha : M(X) > 10$, якщо $\sigma_r = 4$.

Відповідь. $\bar{x}_В = 7{,}78$; $z^* = \dfrac{\bar{x}_В - a}{\dfrac{\sigma_r}{\sqrt{n}}} = \dfrac{7{,}78 - 10}{\dfrac{4}{\sqrt{36}}} = -3{,}33$; $z_{кр} = 2{,}32$.

$z^* \in \left]-\infty; 2{,}32\right]$; $H_0 : M(X) = 10$ приймається.

2. Проведено 25 незалежних вимірювань випадкової величини *X*, що має нормальний закон розподілу зі значенням $\sigma_Г = 2$:

x_i	2,4	5,4	8,4	11,4	14,4	17,4
n_i	2	3	10	6	3	1

При рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези

$H_0 : M(X) = 10{,}5$, якщо альтернативна гіпотеза

$H_0 : M(X) < 10{,}5$.

Відповідь. $z^* = \dfrac{\bar{x}_В - a}{\dfrac{\sigma_r}{\sqrt{n}}} = \dfrac{8{,}92 - 10{,}5}{\dfrac{2}{5}} = -3{,}95$; $z_{кр} = -3$.

$z^* \bar{\in} \left]-\infty; -3\right]$, $z^* \bar{\in} \left[-3; \infty\right[$; $H_0 : M(X) = 10{,}5$ приймається.

3. Маємо дані про розподіл підприємств певної області за зростанням виробітку на одного працівника у відсотках до наступного року:

x_i, %	75	85	95	105	115	125
N_i	5	8	10	5	2	1

Ураховуючи, що ознака має нормальний закон розподілу зі значенням $\sigma_Г = 6$, перевірити правильність нульової гіпотези при $\alpha = 0{,}01$.

$H_0 : M(X) = 90$, якщо альтернативна гіпотеза

$H_\alpha : M(X) \neq 90$.

Відповідь. $z^* = \dfrac{\bar{x}_В - a}{\dfrac{\sigma_Г}{\sqrt{n}}} = \dfrac{96 - 90}{\dfrac{6}{\sqrt{30}}} = \dfrac{6}{1{,}095} = 5{,}48$; $z'_{кр} = -2{,}32$;

$z''_{кр} = 2{,}32$; $z^* \in \left[-2{,}32; 2{,}32\right]$ гіпотеза $H_0 : M(X) = 90$ приймається.

4. У результаті двадцяти незалежних вимірювань певної величини X дістали статистичний розподіл:

x_i	3,4	6,4	9,4	12,4	15,4	18,4
n_i	2	4	8	3	2	1

Припускаючи, що випадкова величина X має нормальний закон розподілу, при рівні значущості $\alpha = 0{,}01$ перевірити правильність

$H_0 : M(X) = 10$, якщо альтернативна гіпотеза

$H_\alpha : M(X) > 10$.

Відповідь. $t^* = \dfrac{\overline{x}_{\text{В}} - a}{\dfrac{S}{\sqrt{n}}} = \dfrac{9{,}7 - 10}{\dfrac{3{,}88}{\sqrt{20}}} = -\dfrac{0{,}3}{0{,}868} = -0{,}346$; $t_{\text{кр}} = 2{,}09$.

$t^* \in]-\infty; 2{,}09]$; $H_0 : M(X) = 10$ приймається.

5. Результати вимірювання зросту дівчаток віком 16 років дали такі показники:

h = 4, см	160—164	164—168	168—172	172—176	176—180
n_i	4	6	20	4	2

Вважаючи, що випадкова величина X — зріст дівчаток — має нормальний закон розподілу, при рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези

$H_0 : M(X) = 180$, якщо альтернативна гіпотеза

$H_\alpha : M(X) \neq 180$.

Відповідь. $t^* = \dfrac{\overline{x}_{\text{В}} - a}{\dfrac{S}{\sqrt{n}}} = \dfrac{169{,}3 - 180}{\dfrac{5{,}17}{6}} = -\dfrac{10{,}7}{0{,}86} = -12{,}42$; $t'_{\text{кр}} = -3{,}65$;

$t''_{\text{кр}} = 3{,}65$ $t^* \overline{\in} [-3{,}65; 3{,}65]$; $H_0 : M(X) = 180$ відхиляється.

6. Рівноточні вимірювання довжини двадцяти однотипних деталей дали такі результати:

x_i, мм	122,8	128,8	134,8	140,8	146,8
n_i	2	6	8	3	1

Вважаючи, що випадкова величина X — довжина деталі — має нормальний закон розподілу, при рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези

$H_0 : M(X) = 144$, якщо альтернативна гіпотеза

$H_\alpha : M(X) < 144$.

Відповідь. $t^* = \dfrac{\bar{x}_В - a}{\dfrac{S}{\sqrt{n}}} = \dfrac{133{,}3 - 144}{\dfrac{6{,}12}{4{,}47}} = -\dfrac{10{,}7}{1{,}37} = -7{,}82$; $t_{кр} = -3{,}88$;

$t^* \overline{\in} [-3{,}88; \infty[$; $H_0 : M(X) = 144$ відхиляється.

7. Вимірювалась швидкість руху автомобілів x_i на певній ділянці шляху. Результати вимірів наведено в таблиці:

x_i, км/год	56	60	64	68	72	70	80
n_i	2	4	6	8	3	1	1

Вважаючи, що X — швидкість автомобіля — є випадковою величиною, яка має нормальний закон розподілу, при рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези

$H_0 : M(X) = 70$, якщо альтернативна гіпотеза

$H_\alpha : M(X) \neq 70$.

Відповідь. $t^* = \dfrac{\bar{x}_В - a}{\dfrac{S}{\sqrt{n}}} = \dfrac{66{,}08 - 70}{\dfrac{5{,}78}{5}} = -\dfrac{3{,}92}{1{,}156} = -3{,}39$; $t'_{кр} = -2{,}8$;

$t''_{кр} = 2{,}8$; $t^* \overline{\in} [-2{,}8; 2{,}8]$; $H_0 : M(X) = 70$ відхиляється.

8. Маса 100 шарикопідшипників наведена у вигляді статистичного розподілу:

x_i, мг	148	150	152	154	156	158	160
n_i	2	4	14	30	40	8	2

Беручи до уваги, що випадкова величина X — маса шарикопідшипників — має нормальний закон розподілу, при рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези

$H_0 : M(X) = 159$, якщо альтернативна гіпотеза

$H_\alpha : M(X) \neq 159$.

Відповідь. $t^* = z^* = \dfrac{\bar{x}_В - a}{\dfrac{S}{\sqrt{n}}} = \dfrac{154{,}68 - 159}{\dfrac{2{,}26}{10}} = -\dfrac{4{,}32}{0{,}226} = -19{,}12$;

$z'_{кр} = -3{,}4$; $z''_{кр} = 3{,}4$; $t^* \overline{\in} [-3{,}4; 3{,}4]$; $H_0 : M(X) = 159$ відхиляється.

9. Вимірювання барометром атмосферного тиску протягом 100 діб дали такі результати:

x_i, мм рт. ст.	744,4	746,4	748,4	750,4	752,4	754,4
n_i	10	20	30	20	15	5

Вважаючи, що X — атмосферний тиск — є випадковою величиною, яка має нормальний закон розподілу, при рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези

$H_0 : M(X) = 749{,}2$, якщо альтернативна гіпотеза

$H_\alpha : M(X) > 749{,}2$.

Відповідь. $t^* = z^* = \dfrac{\bar{x}_В - a}{\dfrac{S}{\sqrt{n}}} = \dfrac{748{,}89 - 749{,}2}{\dfrac{3{,}84}{10}} = -\dfrac{0{,}31}{0{,}384} = -0{,}807$;

$z_{кр} = 2{,}58$; $z^* \in [2{,}58; \infty[$; $H_0 : M(X) = 749{,}2$ мм рт. ст. стверджується.

10. Вимірювання електроопору елементів, виготовлених із молібдену, наведено у вигляді статистичного розподілу:

x_i, мкОм	28,94	32,09	37,72	47,92	52,7	57,32
n_i	8	12	20	50	6	4

Беручи до уваги, що X — електроопір елементів — є випадковою величиною, яка має нормальний закон розподілу, при рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези

$H_0 : M(X) = 50{,}6$, якщо альтернативна гіпотеза

$H_\alpha : M(X) < 50{,}6$.

Відповідь. $t^* = z^* = \dfrac{43{,}1248 - 50{,}6}{0{,}789} = -9{,}47$; $z_{кр} = -3{,}4$;

$z^* \in]-\infty; -3{,}4]$; $H_0 : M(X) = 50{,}6$ приймається.

11. Електролампочки на 220 В виготовлялися двома електроламповими заводами. З першої партії, виготовленої заводом № 1, здійснили вибірку обсягом $n' = 25$, а з другої партії — обсягом $n'' = 36$. Першу і другу партії електролампочок перевірили на тривалість роботи. Результати перевірки наведено у вигляді статистичних розподілів такого вигляду:

y_i	48	50	52	54	56
n'_i	2	3	14	5	1

x_j	53	56	59	62	65
n''_j	4	6	10	12	4

Відомо, що ознаки Y — тривалість роботи електролампочки першого заводу і Y — тривалість роботи електролампочки другого заводу є випадковими величинами, які незалежні між собою і мають нормальний закон розподілу зі значеннями $\sigma_y = 50$, $\sigma_x = 72$. При рівні значущості $\alpha = 0{,}01$ перевірити правильність

$H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза

$H_\alpha : M(X) > M(Y)$.

Відповідь. $z^* = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{D_x}{n''} + \dfrac{D_y}{n'}}} = \dfrac{59{,}5 - 52}{\sqrt{\dfrac{2500}{25} + \dfrac{5184}{30}}} = \dfrac{7{,}5}{16{,}52} = 0{,}45$;

$z_{\text{кр}} = 2{,}58$; $z^* \in]-\infty; 2{,}58]$; $H_0 : M(X) = M(Y)$ приймається.

12. У двох партіях містяться однотипні шарикопідшипники, виготовлені двома заводами. Вимірювання їх діаметрів дали результати, які наведено у вигляді двох статистичних розподілів:

y_i, мм	6,64	6,7	6,74	6,78	6,82
n'_i	2	4	8	6	4

x_j, мм	6,58	6,6	6,8	7	7,2
n''_j	6	8	10	4	2

При рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези

$H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза

$H_\alpha : M(X) \neq M(Y)$,

коли відомі значення $D_x = 50$; $D_y = 60$.

Відповідь. $z^* = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{D_x}{n''} + \dfrac{D_y}{n'}}} = -0{,}05$; $z'_{\text{кр}} = -2{,}58$; $z''_{\text{кр}} = 2{,}58$;

$z^* \in [-2{,}58; 2{,}58]$; $H_0 : M(X) = M(Y)$ приймається.

13. З двох партій монет вартістю 5 коп. було вибрано 50 і 60 штук, які зважували на терезах. Результати цих зважувань подано у вигляді двох статистичних розподілів:

y_i, мг	9,4	9,6	9,8	10	10,2
n'_i	5	15	20	8	2

x_j, мг	9,33	9,63	9,63	10,23	10,53
n''_j	8	12	26	10	4

Припускаючи, що X і Y мають нормальний закон розподілу і незалежні між собою, при рівні значущості $\alpha = 0{,}01$ перевірити

$H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза

$H_\alpha : M(X) < M(Y)$,

коли відомі значення $D_x = 10$; $D_y = 14$.

Відповідь. $z^* = \dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{D_x}{n''} + \dfrac{D_y}{n'}}} = \dfrac{9{,}88 - 9{,}748}{\sqrt{\dfrac{10}{50} + \dfrac{14}{60}}} = 0{,}2$; $z_{\text{кр}} = -3{,}2$;

$z^* \in [-3{,}2; \infty[$; $H_0 : M(X) = M(Y)$ не відхиляється.

14. Вимірювання зросту дітей віком шість років, випадково вибраних із двох дитячих садків, дало такі результати:

y_i, м	0,52	0,58	0,64	0,72	0,8
n'_i	2	5	10	3	1

x_j, м	0,48	0,56	0,64	0,72	0,8
n''_j	1	4	12	6	2

Беручи до уваги, що випадкові величини X і Y є незалежними і мають нормальний закон розподілу, при рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези

$H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза

$H_\alpha : M(X) > M(Y)$.

Відповідь. $t^* = \dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{(n'-1)S_x^2 + (n''-1)S_y^2}{n' + n'' - 2}}} = \dfrac{0{,}6528 - 0{,}633}{\sqrt{\dfrac{21 \cdot 0{,}0057 + 24 \cdot 0{,}00494}{44}}} =$

$= \dfrac{0{,}0198}{0{,}074} = 0{,}269$; $t_{\text{кр}} = 2{,}7$; $t^* \in]-\infty; 2{,}7]$; $H_0 : M(X) = M(Y)$ не відхиляється.

15. Кров'яний тиск було виміряно (в умовних одиницях) у 20 осіб віком 40 років із одного району міста і в 18 осіб того самого віку з іншого району міста. Результати вимірювання подано двома статистичними розподілами:

y_i	114	116	118	120	122	124
n'_i	2	4	6	5	2	1

x_j	115	118	121	124	127	130
n''_j	1	3	6	4	3	1

Припускаючи, що випадкові величини X і Y є незалежними і мають нормальний закон розподілу, при рівні значущості $\alpha = 0{,}001$ перевірити правильність

$H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза

$H_\alpha : M(X) \neq M(Y)$.

Відповідь. $t^* = \dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{(n'-1)S_x^2 + (n''-1)S_y^2}{n' + n'' - 2}}} = \dfrac{122{,}33 - 117{,}4}{\sqrt{\dfrac{19 \cdot 37{,}1 + 17 \cdot 126{,}78}{36}}} =$

$= \dfrac{4{,}93}{8{,}91} \approx 0{,}55$; $t'_{\text{кр}} = -3{,}55$; $t''_{\text{кр}} = 3{,}55$; $t^* \in [-3{,}55; 3{,}55]$; $H_0 : M(X) = M(Y)$ не відхиляється.

16. Пружність вимірювалась на зразках, виготовлених з однієї і тієї самої марки сталі і вибраних із двох партій. Результати вимірювання подано двома статистичними розподілами:

y_i	36,8	38,8	40,8	42,8	44,8
n'_i	2	4	6	5	3

x_j	34,2	38,2	42,2	46,2	50,2
n''_j	2	5	10	4	4

Зважаючи, що ознаки X і Y є незалежними і мають нормальний закон розподілу, при рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези

$H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза

$H_\alpha : M(X) < M(Y)$.

Відповідь. $t^* = \dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{(n'-1)S_x^2 + (n''-1)S_y^2}{n' + n'' - 2}}} = \dfrac{42{,}68 - 41{,}1}{\sqrt{\dfrac{19 \cdot 21{,}76 + 24 \cdot 6{,}01}{43}}} =$

$= \dfrac{1{,}58}{3{,}6} = 0{,}439$; $t_{\text{кр}} = -2{,}7$; $t^* \in [-2{,}7; \infty[$; $H_0 : M(X) = M(Y)$ не відхиляється.

17. Протягом року вимірювалась продуктивність праці (в тис. грн/працівн.) у двох будівельних фірмах. Результати вимірювання подано статистичними розподілами:

y_i	120	150	180	210	240	270
n'_i	10	20	30	20	15	5

x_j	90	130	170	210	250	290
n''_j	10	20	40	20	5	5

Вважаючи, що ознаки X і Y є незалежними і мають нормальний закон розподілу, при рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези

$H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза

$H_\alpha : M(X) > M(Y)$.

Відповідь.

$$t^* = z^* = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{(n'-1)S_x^2 + (n''-1)S_y^2}{n' + n'' - 2}}} = \frac{172 - 187}{\sqrt{\dfrac{99 \cdot 2339{,}7 + 99 \cdot 170{,}46}{198}}} = -\frac{15}{35{,}42} =$$

$= -0{,}42$; $z_{\text{кр}} = 3{,}4$; $z^* \in]-\infty; 3{,}4]$; $H_0 : M(X) = M(Y)$ приймається.

18. Визначався обсяг валової продукції на підприємствах однієї і тієї самої галузі у двох районах України. Результати розрахунків подано двома статистичними розподілами:

y_i, млн грн.	380	400	420	440	460
n'_i	5	15	30	40	10

x_j, млн грн.	360	400	440	480	500	540
n''_j	10	20	30	20	15	5

Ураховуючи, що ознаки X і Y є незалежними і мають нормальний закон розподілу, при рівні значущості $\alpha = 0{,}01$ перевірити правильність

$H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза

$H_\alpha : M(X) \neq M(Y)$.

Відповідь.

$$t^* = z^* = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{(n'-1)S_x^2 + (n''-1)S_y^2}{n' + n'' - 2}}} = \frac{446 - 427}{\sqrt{\dfrac{99 \cdot 2307{,}1 + 99 \cdot 415{,}15}{198}}} = \frac{19}{36{,}89} =$$

$= 0{,}51$; $z'_{\text{кр}} = -2{,}58$; $z''_{\text{кр}} = 2{,}58$; $z^* \in [-2{,}58; 2{,}58]$; $H_0 : M(X) = M(Y)$ не відхиляється.

19. Досліджувався місячний прибуток робітників у гривнях двох заводів однієї і тієї самої галузі виробництва. Результати досліджень подано двома статистичними розподілами:

y_i	150,6	160,6	170,6	180,6	190,6
n'_i	12	28	40	18	2

x_j	140,8	160,8	180,8	200,8	220,8
n''_j	2	6	32	8	2

Ознаки X і Y є незалежними і мають нормальний закон розподілу. При рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези

$H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза

$H_\alpha : M(X) < M(Y)$.

Відповідь.

$$t^* = z^* = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{(n'-1)S_x^2 + (n''-1)S_y^2}{n' + n'' - 2}}} = \frac{181 - 167}{\sqrt{\dfrac{99 \cdot 244{,}25 + 49 \cdot 93{,}94}{148}}} = \frac{14}{13{,}95} \approx$$

$\approx 1{,}004$; $z_{\text{кр}} = -2{,}58$; $z^* \in [-2{,}58; \infty[$; $H_0 : M(X) = M(Y)$ не відхиляється.

20. Вимірювався вміст золи в умовних одиницях в цукрових буряках, що вирощувалися на двох ділянках господарства з однаковим складом добрив у ґрунті. Результати вимірювання подано двома статистичними розподілами:

y_i	0,652	0,692	0,732	0,772	0,812
n'_i	10	20	50	8	2

x_j	0,664	0,684	0,704	0,724	0,744	0,764
n''_j	8	12	50	20	5	5

Ознаки X і Y є незалежними між собою і мають нормальний закон розподілу. При рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези

$H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза

$H_\alpha : M(X) \neq M(Y)$.

Відповідь.

$$t^* = z^* = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{(n'-1)S_x^2 + (n''-1)S_y^2}{n' + n'' - 2}}} = \frac{0{,}7074 - 0{,}72}{\sqrt{\dfrac{89 \cdot 0{,}00052 + 99 \cdot 0{,}00057}{188}}} =$$

$$= \frac{-0{,}0126}{0{,}02337} \approx -0{,}539 \text{ ; } z'_{\text{кр}} = -3{,}4 \text{ ; } z''_{\text{кр}} = 3{,}4 \text{ ; } z^* \in [-3{,}4; 3{,}4];$$

$H_0 : M(X) = M(Y)$ не відхиляється.

21. Визначалась урожайність зеленої маси вівса, зібраного у двох районах області. Результати розрахунків наведено у вигляді статистичних розподілів:

y_i, ц/га	88	92	96	100	104
n'_i	2	4	8	6	4

x_j, ц/га	82	88	94	100	100
n''_j	4	8	6	2	2

Ураховуючи, що ознаки X і Y (урожайність в ц/га) є незалежними і мають нормальний закон розподілу, при рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези

$H_0 : D_x = D_y$, якщо альтернативна гіпотеза

$H_\alpha : D_y > D_x$.

Відповідь. $F^* = \dfrac{S_\delta^2}{S_М^2} = 2{,}44$; $F_{кр}(\alpha = 0{,}01;\ k_1 = 21;\ \ k_2 = 24) = 2{,}9$; $F^* \in [0;\ 2{,}9]$; $H_0 : D_x = D_y$ не відхиляється.

22. Норма витрат на технічне обслуговування і ремонт нових марок тракторів вимірювалась у двох сільських господарствах району. Результати вимірювань показано двома статистичними розподілами:

y_i, грн/га	0,58	0,6	0,62	0,64	0,66
n_i'	2	3	10	4	1

x_j, грн/га	0,56	0,6	0,64	0,7	0,74
n_j''	4	6	3	2	1

Ознаки X і Y (норми витрат) є незалежними випадковими величинами, що мають нормальний закон розподілу. При рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези

$H_0 : D_x = D_y$, якщо альтернативна гіпотеза

$H_\alpha : D_x > D_y$.

Відповідь. $F^* = \dfrac{S_\delta^2}{S_М^2} = 7{,}547$; $F_{кр}(\alpha = 0{,}001;\ k_1 = 15;\ k_2 = 19) = 5$; $F^* \overline{\in} [0;\ 5]$; $H_0 : D_x = D_y$ відхиляється.

23. Визначалися річні середні витрати електроенергії на комунально-побутові вимоги для одного мешканця у двох містах. Результати розрахунків подано двома статистичними розподілами для першого і другого міст:

y_i, Вт/м.	700	708	716	724	732	740
n_i'	5	6	9	6	3	1

x_j, Вт/м.	706	710	714	718	722	726	730
n_j''	8	10	12	5	2	2	1

Ознаки X і Y (річні витрати в кВт/особу) є незалежними між собою і мають нормальний закон розподілу. При рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези.

$H_0 : D_x = D_y$, якщо альтернативна гіпотеза

$H_\alpha : D_x < D_y$.

Відповідь. $F^* = \dfrac{S_\delta^2}{S_М^2} = 1{,}511$; $F_{кр}(\alpha = 0{,}001;\ k_1 = 39;\ k_2 = 39) = 2{,}2$; $F^* \in [0;\ 2{,}2]$; $H_0 : D_x = D_y$ не відхиляється.

24. Вимірювання вмісту азоту в цукрових буряках, які вирощувалися на двох ділянках, розміщених у різних місцях колективного госпо-

дарства, з однаковим складом ґрунту показав результати, що наведені у двох статистичних розподілах:

y_i, умов. од.	1,24	1,28	1,32	1,36	1,4	1,44
n'_i	5	6	8	13	2	1

x_j, умов. од.	714	718	722	726	730
n''_j	4	10	16	10	6

Ознаки X і Y (вміст азоту) є незалежними випадковими величинами, що мають нормальний закон розподілу. При рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези при альтернативній гіпотезі

$H_0 : D_x = D_y$, якщо альтернативна гіпотеза

$H_\alpha : D_x < D_y$.

Відповідь. $F^* = \dfrac{S_\delta^2}{S_\text{М}^2} = 3{,}36$; $F_\text{кр}(\alpha = 0{,}01; k_1 = 45; k_2 = 34) = 2$;

$F^* \overline{\in} [0; 2]$; $H_0 : D_x = D_y$ не приймається.

25. Вимірювання значень наробки на мотор автомобіля, що здійснювався у двох автопарках міста, наведено у вигляді статистичних розподілів:

y_i, тис. км	1,9	2,15	2,4	2,65	2,9	3,15
n'_i	2	4	6	10	5	1

x_j, тис. км	1,8	2	2,2	2,4	2,6	2,8	3
n''_j	4	6	12	16	8	2	1

Ознаки X і Y (наробки в тис. км) є випадковими величинами, що мають нормальний закон розподілу. При рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези

$H_0 : D_x = D_y$, якщо альтернативна гіпотеза

$H_\alpha : D_x < D_y$.

Відповідь. $F^* = \dfrac{S_\delta^2}{S_\text{М}^2} = 4{,}475$; $F_\text{кр}(\alpha = 0{,}001; k_1 = 28; k_2 = 48) = 2{,}2$;

$F^* \overline{\in} [0; 2{,}2]$; $H_0 : D_x = D_y$ не приймається.

26. Заміри довжини волокон вовни, одержаної від овець, що утримувалися на двох фермах, подано двома статистичними розподілами:

y_i, мм	64	66	68	70	72	74
n'_i	2	4	6	8	4	2

x_j, мм	66	68	72	76	80	84
n''_j	4	6	10	12	4	2

Ознаки X і Y (довжини волокон) є незалежними випадковими величинами, що мають нормальний закон розподілу. При рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези

$H_0 : D_x = D_y$, якщо альтернативна гіпотеза

$H_\alpha : D_x < D_y$.

Відповідь. $F^* = \dfrac{S_\delta^2}{S_{\text{М}}^2} = 5{,}727$; $F_{\text{кр}}(\alpha = 0{,}001; k_1 = 37; k_2 = 27) = 2{,}7$;

$F^* \overline{\in} [0; 2{,}7]$; $H_0 : D_x = D_y$ відхиляється.

27. У двох автопарках виміряли витрати палива за годину автомобілем. Результати вимірювання подано двома статистичними розподілами:

y_i, кг/год	35	35,2	35,4	35,6	35,8	36
n'_i	2	8	10	6	4	3

x_j, кг/год	35,4	35,8	36,2	36,6	37
n''_j	4	5	6	15	6

Ознаки X і Y (витрати палива за год) є незалежними випадковими величинами, які мають нормальний закон розподілу ймовірностей. При рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези

$H_0 : D_x = D_y$, якщо альтернативна гіпотеза

$H_\alpha : D_x > D_y$.

Відповідь. $F^* = \dfrac{S_\delta^2}{S_{\text{М}}^2} = 56{,}3$; $F_{\text{кр}}(\alpha = 0{,}01; k_1 = 35; k_2 = 32) = 2$;

$F^* \overline{\in} [0; 2]$; $H_0 : D_x = D_y$ відхиляється.

28. Вимірювалось споживання масла за одну добу одним мешканцем у двох регіонах країни. Результати вимірювання подано двома статистичними розподілами:

y_i, мг	15,99	18,99	21,99	21,99	24,99
n'_i	4	6	20	10	5

x_j, мг	14,55	20,55	26,55	30,55	36,55
n''_j	6	14	16	6	4

Ознаки X і Y (добове споживання масла в мг) є незалежними випадковими величинами, які мають нормальний закон розподілу ймовірностей. При рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези

$H_0 : D_x = D_y$, якщо альтернативна гіпотеза

$H_\alpha : D_x > D_y$.

Відповідь. $F^* = \dfrac{S_\delta^2}{S_М^2} = 7{,}24$; $F_{кр}(\alpha = 0{,}001;\ k_1 = 45;\ \ k_2 = 44) = 2{,}2$;
$F^* \overline{\in} [0;\ 2{,}2]$; $H_0 : D_x = D_y$ відхиляється.

29. Вимірювання маси в грамах пухових волокон від овець подано двома статистичними розподілами:

y_i, г	4,44	4,84	5,24	6,64	6,04
n_i'	2	4	5	8	1

x_j, г	4,36	4,96	5,46	5,96	6,46
n_j''	3	5	8	6	4

Ознаки X і Y (маса волокон в грамах) є незалежними випадковими величинами, які мають нормальний закон розподілу. При рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези
$H_0 : D_x = D_y$, якщо альтернативна гіпотеза
$H_\alpha : D_x > D_y$.

Відповідь. $F^* = \dfrac{S_\delta^2}{S_М^2} = 2{,}07$; $F_{кр}(\alpha = 0{,}01;\ k_1 = 25;\ k_2 = 19) = 2{,}9$;
$F^* \in [0;\ 2{,}9]$; $H_0 : D_x = D_y$ не відхиляється.

30. Вимірювалась жива маса курчат, які відгодовувалися у двох птахофермах. Результати вимірювання подано двома статистичними розподілами:

y_i, г	96,5	99,5	102,5	108,5	111,5
n_i'	5	10	6	4	4

x_j, г	85,5	105,5	125,5	145,5	165,5
n_j''	6	8	12	4	2

Ознаки X і Y (жива маса курчат) є незалежними випадковими величинами, що мають нормальний закон розподілу ймовірностей. Якщо рівень значущості $\alpha = 0{,}01$, перевірте правильність нульової гіпотези
$H_0 : D_x = D_y$, за альтернативної гіпотези
$H_\alpha : D_y < D_x$.

Відповідь. $F^* = \dfrac{S_\delta^2}{S_М^2} = 18{,}87$; $F_{кр}(\alpha = 0{,}01;\ k_1 = 33;\ \ k_2 = 28) = 2{,}1$;
$F^* \overline{\in} [0;\ 2{,}1]$; $H_0 : D_x = D_y$ відхиляється.

Лабораторна робота № 1
до теми «Статистичні гіпотези»

За заданими статистичними розподілами вибірки висунути H_0 про закон розподілу ознаки генеральної сукупності і при рівні значущості $\alpha = 0{,}01$ перевірити її правильність:

1. Вимірювалась маса новонароджених дітей у пологовому будинку. Результати вимірювань подано інтервальним статистичним розподілом:

x_i, кг, $h = 0{,}2$	1—1,2	1,2—1,4	1,4—1,6	1,6—1,8	1,8—2	2—2,2
n_i	5	12	18	22	36	24

Продовження табл.

x_i, кг, $h = 0{,}2$	2,2—2,4	2,4—2,6	2,6—2,8	2,8—3	3—3,2
n_i	19	15	11	9	2

2. Протягом 50 років вимірювався рівень води навесні під час повені відносно умовного нуля. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i, см, $h = 24$	0—24	24—48	48—72	72—96	96—120
n_i	1	2	4	6	12

Продовження табл.

x_i, см, $h = 24$	120—146	146—170	170—196	196—220
n_i	16	6	3	1

3. Результати вимірювання граничного навантаження на сталевий болт наведено інтервальним статистичним розподілом:

x_i, кг/мм^2, $h = 1$	4,5—5,5	5,5—6,5	6,5—7,5	7,5—8,5	8,5—9,5
n_i	40	32	28	24	20

Продовження табл.

x_i, кг/мм^2, $h = 1$	9,5—10,5	10,5—11,5	11,5—12,5	12,5—13,5
n_i	18	16	12	4

4. Точність вимірювання діаметру валиків, виготовлених верстатом-автоматом, показано у вигляді інтервального статистичного розподілу:

x_i, мк, h = 0,01	0,025—0,035	0,035—0,045	0,045—0,055	0,055—0,065
n_i	47	40	36	25

Продовження табл.

x_i, мк, h = 0,01	0,065—0,075	0,075—0,085	0,085—0,095
n_i	18	12	8

5. Результати вимірювання зносу стального різця за певний відтинок часу наведено у вигляді інтервального статистичного розподілу:

x_i, мк, h = 0,001	0,0212—0,0222	0,0222—0,0232	0,0232—0,0242
n_i	30	25	15

Продовження табл.

x_i, мк, h = 0,001	0,0242—0,0252	0,0252—0,0262
n_i	12	10

6. Фіксувалось число розлагоджень верстатів-автоматів протягом року. Результати спостережень наведено у вигляді інтервального статистичного розподілу:

x_i, h = 2	0—2	2—4	4—6	6—8	8—10	10—12	12—14	14—16
n_i	16	12	10	9	7	6	5	1

7. Вимірювання показників виконання річного плану підприємствами певної галузі наведено у вигляді інтервального статистичного розподілу:

x_i, %, h = 10	10—20	20—30	30—40	40—50	50—60
n_i	2	5	13	16	25

Продовження табл.

x_i, %, h = 10	60—70	70—80	80—90	90—100	100—120
n_i	12	10	5	3	1

8. Вимірювався місячний дохід робітника певної галузі. Результати вимірювання подано інтервальним статистичним розподілом:

x_i, грн., $h = 20$	100—120	120—140	140—160	160—180	180—200
n_i	10	15	20	25	30

Продовження табл.

x_i, грн., $h = 20$	200—220	220—240	240—260	260—280
n_i	40	10	4	2

9. Досліджувалась залежність кількості проданих пар чоловічого взуття від його розміру. Результати досліджень наведено у вигляді інтервального статистичного розподілу:

x_i, грн., $h = 20$	75—125	125—175	175—225	225—275	275—325
n_i	2	12	18	24	38

Продовження табл.

x_i, грн., $h = 20$	325—375	375—425	425—475	475—525	525—575	575—625
n_i	21	19	12	10	5	3

10. Вимірювалось відхилення діаметра валика від його номінального розміру. Результати вимірювання наведено як інтервальний статистичний розподіл:

x_i, мк, $h = 0{,}02$	0,228—0,248	0,248—0,268	0,268—0,288	0,288—0,308	0,308—0,328
n_i	6	16	21	36	42

Продовження табл.

x_i, мк, $h = 0{,}02$	0,328—0,348	0,348—0,368	0,368—0,388	0,388—0,408
n_i	32	22	12	8

11. Вимірювання зросту юнаків віком 18—20 років наведено у вигляді інтервального статистичного розподілу:

x_i, см, $h = 6$	154—160	160—166	166—172	172—178
n_i	8	20	30	42

Продовження табл.

x_i, см, h = 6	178—184	184—190	190—196	196—202
n_i	34	21	9	2

12. Випробувалась чутливість каналу x_i 100 телевізорів. Результати випробувань наведено у вигляді інтервального статистичного розподілу:

x_i, умовн. один.; h = 50	75–125	125–175	175–225	225–275	275–325	325—375
n_i	1	3	6	22	36	30

Продовження табл.

x_i, умовн. один.; h = 50	325—375	375—425	425—475	475—525	525—575	575—625	625—675
n_i	30	24	18	12	10	6	2

13. Вимірювалась жирність молока корів x_i із навмання вибраної ферми. Результати вимірювання наведено як інтервальний статистичний розподіл:

x_i, %, h = 0,1	3,45—3,55	3,55—3,65	3,65—3,75	3,75—3,85	3,85—3,95
n_i	10	16	22	30	34

Продовження табл.

x_i, %, h = 0,1	3,95—4,05	4,05—4,15	4,15—4,25	4,25—4,35	4,35—4,45
n_i	20	14	10	6	4

14. Вимірювався вміст фосфору в чавуні. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i, %, h = 0,02	0,36—0,38	0,38—0,4	0,4—0,42	0,42—0,44
n_i	10	16	24	40

Продовження табл.

x_i, %, h = 0,02	0,44—0,46	0,46—0,48	0,48—0,5	0,5—0,52
n_i	32	20	16	5

15. Вимірювалась стабільність температури у виготовлених фірмою термостатах *t*. Результати вимірювання наведено як статистичний розподіл:

t,°C, h = 0,04	20,24—20,28	20,28—20,32	20,32—20,36	20,36—20,4
n_i	40	38	26	18

Продовження табл.

t,°C, h = 0,04	20,4—20,44	20,44—20,48	20,48—20,52
n_i	12	6	2

16. Вимірювався вміст різних речовин у зразках, виготовлених із чавуну. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i, %, h = 0,06	2,22—2,28	2,28—2,34	2,34—2,4	2,4—2,46
n_i	52	44	36	20

Продовження табл.

x_i, %, h = 0,06	2,46—2,52	2,52—2,56	2,56—2,62	2,62—2,68
n_i	18	12	6	2

17. Проводився експеримент з вимірювання потужності відбиття від поверхні моря сигналу на виході радіолокаційного приймача. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i, вт, h = 0,05	0—0,05	0,05—0,1	0,1—0,15	0,15—0,2	0,2—0,25
n_i	88	64	58	42	30

Продовження табл.

x_i, вт, h = 0,05	0,25—0,3	0,3—0,35	0,35—0,4	0,4—0,45
n_i	22	18	6	4

18. За допомогою радіодальноміра вимірювалась відстань до одного і того самого об’єкта. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i, м, h = 0,25	2000—2000,25	2000,25—2000,5	2000,5—2000,75
n_i	12	24	32

Продовження табл.

x_i, м, h = 0,25	2000,75—2001,25	2001,25—2001,75	2001,75—2002,25
n_i	44	38	26

Продовження табл.

x_i, м, h = 0,25	2002,25—2002,75	2002,75—2003,25	2003,25—2003,75
n_i	18	12	6

19. Урожайність рису, який зібрано з поливних ділянок фермерських господарств, наведена у вигляді інтервального статистичного розподілу:

x_i, ц/га, h = 0,5	9,5—10	10—10,5	10,5—11	11—11,5	11,5—12
n_i	5	16	24	32	40

20. Вимірювання діаметрів кульок, виготовлених верстатом-автоматом, наведено як інтервальний статистичний розподіл:

x_i, мм, h = 0,08	6—6,08	6,08—6,16	6,16—6,24	6,24—6,32
n_i	8	18	24	32

Продовження табл.

x_i, мм, h = 0,08	6,32—6,4	6,4—6,48	6,48—6,56	6,56—6,64
n_i	28	21	15	6

21. Вимірювався знос автошин за місяць. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i, мм, h = 0,02	0—0,02	0,02—0,04	0,04—0,06	0,06—0,08
n_i	48	42	34	26

Продовження табл.

x_i, мм, h = 0,02	0,08—0,1	0,1—0,12	0,12—0,14	0,14—0,16
n_i	18	10	6	4

22. Вимірювалась в’язкість нафти, видобутої із свердловин. Результати вимірювання наведено як інтервальний статистичний розподіл:

x_i, ум. од., h = 0,08	4,2—4,28	4,28—4,36	4,36—4,44	4,44—4,52	4,52—4,6
n_i	2	6	10	14	16

Продовження табл.

x_i, ум. од., h = 0,08	4,6—4,68	4,68—4,76	4,76—4,84	4,84—4,92	4,92—5
n_i	8	6	4	2	1

23. Вимірювання опору елементів, виготовлених шляхом використання нової технології, дали результати, які наведено у вигляді інтервального статистичного розподілу:

x_i, Ом, h = 0,04	0,32—0,36	0,36—0,4	0,4—0,44	0,44—0,48	0,48—0,52
n_i	40	36	30	24	20

Продовження табл.

x_i, Ом, h = 0,04	0,52—0,56	0,56—0,6	0,6—0,64	0,64—0,68	0,68—0,72
n_i	18	16	12	8	2

24. Вимірювався час неперервного горіння електролампочок, виготовлених фірмою, до виходу їх з ладу. Результати вимірювання наведено як інтервальний статистичний розподіл:

x_i, год, h = 0,5	22—22,5	22,5—23	23—23,5	23,5—24	24—24,5	24,5—25
n_i	4	12	16	24	36	28

Продовження табл.

x_i, год, h = 0,5	25—25,5	25,5—26	26—26,5	26,5—27	27—27,5
n_i	22	18	16	8	4

25. Вимірювалась продуктивність праці за зміну робітників однакового профілю спеціальності й однакового віку певної галузі. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i, грн., h = 0,8	28,4—29,2	29,2—30	30—30,8	30,8—31,6	31,6—32,4
n_i	60	48	36	24	18

Продовження табл.

x_i, грн., h = 0,8	32,4—33,2	33,2—34	34—34,8	34,8—35,6	35,6—36,4
n_i	14	12	10	4	2

26. Урожайність цукрових буряків вимірювалась у певному районі південного регіону України. Результати вимірювання подано як інтервальний статистичний розподіл:

x_i, ц/га, h = 2,6	340—342,6	342,6—345,2	345,2—347,8	347,8—350,4
n_i	12	18	26	38

Продовження табл.

x_i, ц/га, h = 2,6	350,4—353	353—355,6	355,6—358,2	358,2—360,8
n_i	40	26	16	6

27. Вимірювалась плинність кадрів у відсотках на підприємствах певної галузі виробництва протягом року. Результати вимірювання показано як інтервальний статистичний розподіл:

x_i, %, h = 0,06	0,12—0,18	0,18—0,24	0,24—0,3	0,3—0,36	0,36—0,42
n_i	46	38	32	28	24

Продовження табл.

x_i, %, h = 0,06	0,42—0,48	0,48—0,54	0,54—0,6	0,6—0,66
n_i	18	16	8	6

28. Вимірювались коливання місячної заробітної плати робітника певного профілю спеціальності певної галузі виробництва. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i, грн., h = 5,8	165—170,8	170,8—176,6	176,6—182,4	182,4—188,2
n_i	125	115	104	86

Продовження табл.

x_i, грн., h = 5,8	188,2—194	194—199,8	199,8—205,6	205,6—211,4
n_i	64	36	20	12

29. Вимірювалась кількість опадів, які випали протягом веснянопольових робіт у північних регіонах України. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i, мм, h = 12	440—452	452—464	464—476	476—488	488—500
n_i	24	18	16	14	12

Продовження табл.

x_i, мм, h = 12	500—512	512—524	524—536	536—548	548—560
n_i	10	8	6	4	2

РОЗДІЛ VII

ЕЛЕМЕНТИ ДИСПЕРСІЙНОГО, КОРЕЛЯЦІЙНОГО ТА РЕГРЕСІЙНОГО АНАЛІЗУ

ТЕМА 15. ЕЛЕМЕНТИ ДИСПЕРСІЙНОГО АНАЛІЗУ

1. Загальна інформація

Дисперсійний аналіз був створений спочатку для статистичної обробки агрономічних дослідів. В наш час його також використовують як в економічних експериментах, так і технічних, соціальних.

Сутність цього аналізу полягає в тому, що загальну дисперсію досліджуваної ознаки розділяють на окремі компоненти, які обумовлені впливом певних конкретних чинників. Істотність їх впливу на цю ознаку здійснюється методом дисперсійного аналізу.

Відповідно до дисперсійного аналізу будь-який його результат можна подати у вигляді суми певної кількості компонент. Так, наприклад, якщо досліджується вплив певного чинника на результат експерименту, то модель, що описує структуру останнього, можна подати так:

$$x_{ij} = \bar{x} + \alpha_j + \varepsilon_{ij}, \qquad (467)$$

де x_{ij} — значення ознаки *X*, одержане при *i*-му експерименті на *j*-му рівні фактора. Під рівнем фактора розуміють певну його міру. Наприклад, якщо фактором є добрива, які вносяться в ґрунт з метою збільшення врожайності сільськогосподарської культури, то рівнем фактора в цьому разі є кількість добрива, що вноситься в ґрунт; $\bar{x}$ — загальна середня величина ознаки *X*; α_j — ефект впливу фактора на значення ознаки *X* на *j*-му рівні; ε_{ij} — випадкова компонента, що впливає на значення ознаки *X* в *i*-му експерименті на *j*-му рівні.

При цьому $M(\varepsilon_{ij}) = 0$ і ε_{ij} як випадкові величини мають закон розподілу ймовірностей $N(0;\sigma^2)$ і між собою незалежні ($K_{ij} = 0$).

Складнішою моделлю аналізу є вивчення впливу на результати експерименту кількох факторів. Зокрема при аналізі впливу двох факторів структура моделі набуває такого вигляду:

$$x_{ijk} = \bar{x} + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}\,, \qquad (468)$$

де x_{ijk} — значення ознаки X в i-му експерименті на j-му рівні впливу фактора A і на k-му рівні впливу фактора B; $\bar{x}$ — загальна середня величина ознаки X; α_i — ефект впливу фактора A на i-му рівні; β_j — ефект впливу фактора B на j-му рівні; γ_{ij} — ефект одночасного впливу факторів A і B; ε_{ijk} — випадкова компонента.

У разі проведення дисперсійного аналізу досліджуваний масив даних, одержаних під час експерименту, поділяють на певні групи, які різняться дією на результати експерименту певних рівнів факторів.

Вважається, що досліджувана ознака має нормальний закон розподілу, а дисперсії в кожній окремій групі здобутих значень ознаки однакові. Ці припущення необхідно перевірити.

2. Однофакторний дисперсійний аналіз

Нехай потрібно дослідити вплив на ознаку X певного одного фактора. Результати експерименту ділять на певне число груп, які відрізняються між собою ступенем дії фактора.

Для зручності в проведенні необхідних обчислень результати експерименту зводять в спеціальну таблицю:

Таблиця 1

Ступінь впливу фактора (групи)	Спостережуване значення ознаки X	Групові середні	Загальна середня
1	$x_{11}, x_{21}, x_{31}, ..., x_{n,1}$	$\bar{x}_1 = \dfrac{\sum_{i=1}^{m} x_{i1}}{n_1}$	$\bar{x} = \dfrac{\sum_{i=1}^{n_j}\sum_{j=1}^{p} x_{ij}}{N}$, $N = \sum_{j=1}^{p} n_j$
2	$x_{12}, x_{22}, x_{32}, ..., x_{n,2}$	$\bar{x}_2 = \dfrac{\sum_{i=1}^{m} x_{i2}}{n_2}$	
3	$x_{13}, x_{23}, x_{33}, ..., x_{n,31}$	$\bar{x}_3 = \dfrac{\sum_{i=1}^{m} x_{i3}}{n_3}$	
⋮	.....	.....	
p	$x_{1p}, x_{2p}, x_{3p}, ..., x_{n,p}$	$\bar{x}_p = \dfrac{\sum_{i=1}^{m} x_{ip}}{n_p}$	

Відповідно до моделі однофакторного дисперсійного аналізу необхідно визначити дві дисперсії, а саме: міжгрупову (дисперсію групових середніх), зумовлену впливом досліджуваного фактора на ознаку X, і внутрішньогрупову, зумовлену впливом інших випадкових факторів.

Загальна дисперсія розглядається як сума квадратів відхилень:

$$\sum_{i=1}^{n_j}\sum_{j=1}^{p}(x_{ij}-\bar{x})^2 .$$

Тоді поділ загальної дисперсії на компоненти здійснюється так:

$$\sum_{i=1}^{n_j}\sum_{j=1}^{p}(x_{ij}-\bar{x})^2=\sum_{i=1}^{n_j}\sum_{j=1}^{p}((x_{ij}-\bar{x}_j)+(\bar{x}_j-\bar{x}))^2=$$

$$=\sum_{i=1}^{n_j}\sum_{j=1}^{p}((x_{ij}-\bar{x}_j)^2+2(x_{ij}-\bar{x}_j)(\bar{x}_j-\bar{x})+(\bar{x}_j-\bar{x})^2)=$$

$$=\sum_{i=1}^{n_j}\sum_{j=1}^{p}(x_{ij}-\bar{x}_j)^2+2\sum_{i=1}^{n_j}\sum_{j=1}^{p}(x_{ij}-\bar{x}_j)(\bar{x}_j-\bar{x})+\sum_{i=1}^{n_j}\sum_{j=1}^{p}(\bar{x}_j-\bar{x})^2=$$

$$=\sum_{i=1}^{n_j}\sum_{j=1}^{p}(x_{ij}-\bar{x}_j)^2+2\sum_{j=1}^{p}n_j(\bar{x}_j-\bar{x})^2+\sum_{i=1}^{n_j}\sum_{j=1}^{p}(\bar{x}_j-\bar{x})^2=$$

$$=\left|\begin{array}{l}\text{оскільки}\\ \sum_{i=1}^{n_j}\sum_{j=1}^{p}(x_{ij}-\bar{x}_j)(\bar{x}_j-\bar{x})=\sum_{i=1}^{n_1}(x_{i1}-\bar{x}_1)(\bar{x}_1-\bar{x})+\sum_{i=1}^{n_2}(x_{i2}-\bar{x}_2)(\bar{x}_2-\bar{x})+\\ +....+\sum_{i=1}^{n_p}(x_{ip}-\bar{x}_p)(\bar{x}_p-\bar{x})=(\bar{x}_1-\bar{x})\sum_{i=1}^{n_1}(x_{i1}-\bar{x}_1)+\\ +(\bar{x}_2-\bar{x})\sum_{i=1}^{n_2}(x_{i2}-\bar{x}_2)+.....+(\bar{x}_p-\bar{x})\sum_{i=1}^{n_p}(x_{ip}-\bar{x}_p)=0,\\ \text{враховуючи те, що}\\ \sum_{i=1}^{n_p}(x_{ij}-\bar{x}_j)=0,\ j=\overline{1,p},\ \text{третій член суми дорівнюватиме:}\end{array}\right|$$

$$\sum_{i=1}^{n_j}\sum_{j=1}^{p}(\bar{x}_j-\bar{x})^2=\sum_{j=1}^{p}\sum_{i=1}^{n_i}(\bar{x}_j-\bar{x})^2=\sum_{j=1}^{p}n_j(\bar{x}_j-\bar{x})^2.$$

Таким чином, дістаємо:

$$\sum_{i=1}^{n_j}\sum_{j=1}^{p}(x_{ij}-\bar{x})^2=\sum_{i=1}^{n_p}\sum_{j=1}^{p}(x_{ij}-\bar{x})^2+\sum_{j=1}^{p}n_j(\bar{x}_j-\bar{x})^2 . \qquad (469)$$

Для того щоб мати виправлені дисперсії, необхідно кожну зі здобутих сум поділити на число ступенів свободи.

Так, для загальної дисперсії $\frac{\sum_{i=1}^{n_j}\sum_{j=1}^{p}(x_{ij}-\bar{x})^2}{N}$ виправлена дисперсія дорівнюватиме

$$S^2=\frac{\sum_{i=1}^{n_j}\sum_{j=1}^{p}(x_{ij}-\bar{x})^2}{N-1}. \qquad (470)$$

Виправлена дисперсія S_1^2, що характеризує розсіювання всередині групи, зумовлене впливом випадкових факторів, обчислюється за формулою:

$$S_1^2=\frac{\sum_{i=1}^{n_j}\sum_{j=1}^{p}(x_{ij}-\bar{x}_j)^2}{N-p}, \qquad (471)$$

де $N-p=k_1$ є числом ступенів свободи для S_1^2, оскільки при цьому використовується p співвідношень при обчисленні групових середніх $\bar{x}_j, j=\overline{1,p}$.

Виправлена дисперсія S_2^2, що характеризує розсіювання групових середніх $\bar{x}_j$ відносно загальної середньої $\bar{x}$, яке викликане впливом фактора на результат експерименту ознаки *X*, обчислюється за формулою:

$$S_2^2=\frac{\sum_{j=1}^{p}n_j(\bar{x}_j-\bar{x})^2}{p-1}, \qquad (472)$$

де $p-1=k_2$ — це число ступенів свободи для S_2^2, оскільки групові середні варіюють відносно однієї загальної середньої $\bar{x}$.

Завдання виявлення впливу фактора на наслідки експерименту полягає в порівнянні виправлених дисперсій S_1^2, S_2^2. І справді, якщо досліджуваний фактор не впливає на значення ознаки *X*, то в цьому разі S_1^2 і S_2^2 можна розглядати як незалежні оцінки загальної дисперсії *D*. І навпаки, якщо відношення S_1^2 і S_2^2 істотне, то в цьому разі вибірки слід вважати здійсненими з різних сукупностей, тобто з сукупностей з різним рівнем впливу фактора.

Порівняння двох дисперсій ґрунтується на перевірці правильності нульової гіпотези: $H_0: D_1=D_2$ — про рівність дисперсій двох вибірок.

За статистичний критерій вибирається випадкова величина

$$F=\frac{S_2^2}{S_1^2}\cdot\frac{p-1}{N-p}, \qquad (473)$$

що має розподіл Фішера—Снедекора з $k_1 = N - p$, $k_2 = p - 1$ ступенями свободи.

За значеннями α, $k_1 = N - p$, $k_2 = p - 1$, знаходимо критичну точку (додаток 7).

Спостережуване значення критерію обчислюється за формулою (473).

Якщо $F^* \le F_{\text{кр}}$, то нульова гіпотеза про вплив фактора на результати досліджень відхиляється, а коли $F^* > F_{\text{кр}}$, то цим самим підтверджується вплив фактора на ознаку *X*.

Результати спостережень та обчислення статистичних оцінок зручно подати в упорядкованому вигляді за допомогою табл. 2.

Таблиця 2

Рівень фактора (групи)	Спостережуване значення ознаки *X*	Групові середні	Загальна середня
1	$x_{11}, x_{21}, x_{31}, \ldots, x_{n,1}$	$\bar{x}_1 = \dfrac{\sum_{i=1}^{n_1} x_{i1}}{n_1}$	$\bar{x} = \dfrac{\sum_{i=1}^{n_j}\sum_{j=1}^{p} x_{ij}}{N}$, $N = \sum_{j=1}^{p} n_j$
2	$x_{12}, x_{22}, x_{32}, \ldots, x_{n,2}$	$\bar{x}_2 = \dfrac{\sum_{i=1}^{n_2} x_{i2}}{n_2}$	
⋮	.....	.....	
p	$x_{1p}, x_{2p}, x_{3p}, \ldots, x_{n,p}$	$\bar{x}_p = \dfrac{\sum_{i=1}^{n_p} x_{ip}}{n_p}$	

Вид варіацій ознаки	Сума квадратів відхилень	Число ступенів свободи	Статистичні оцінки дисперсії
внутрішньогрупова	$\sum_{i=1}^{n_j}\sum_{j=1}^{p}(x_{ij} - \bar{x}_j)^2$	$N - p$	$S_1^2 = \dfrac{\sum_{i=1}^{n_j}\sum_{j=1}^{p}(x_{ij} - \bar{x}_j)^2}{N - p}$
міжгрупова	$\sum_{j=1}^{p} n_j(\bar{x}_j - \bar{x})^2$	$p - 1$	$S_2^2 = \dfrac{\sum_{j=1}^{p} n_j(\bar{x}_j - \bar{x})^2}{p - 1}$
загальна	$\sum_{i=1}^{n_j}\sum_{j=1}^{p}(x_{ij} - \bar{x})^2$	$N - 1$	$S^2 = \dfrac{\sum_{i=1}^{n_j}\sum_{j=1}^{p}(x_{ij} - \bar{x})^2}{N - 1}$

Приклад 1. Ступінь впливу каталізатора на кінцевий продукт заданої хімічної реакції, наведеної в таблиці.

Ступінь впливу каталізатора	Кінцевий продукт хімічної реакції
1	3,2; 3,1; 3,1; 2,8; 3,3; 3,0
2	2,6; 3,1; 2,7; 2,9; 2,7; 2,8
3	2,9; 2,6; 3,0; 3,1; 3,0; 2,8
4	3,7; 3,4; 3,2; 3,3; 3,5; 3,3
5	3,0; 3,4; 3,2; 3,5; 2,9; 3,1

З'ясувати, чи істотно впливає каталізатор на кінцевий продукт хімічної реакції при $\alpha = 0{,}001$.

Розв'язання. Використовуючи табл. 2 і виконавши відповідні обчислення, дістанемо

Ступінь впливу каталізатора	Спостережуване значення (кінцевий продукт)	Групові середні	Загальна середня
1	3,2; 3,1; 3,1; 2,8; 3,3; 3,0	$\bar{x}_1 = 3{,}083$	
2	2,6; 3,1; 2,7; 2,9; 2,7; 2,8	$\bar{x}_2 = 2{,}8$	
3	2,9; 2,6; 3,0; 3,1; 3,0; 2,8	$\bar{x}_3 = 2{,}9$	$\bar{x} = 3{,}073$
4	3,7; 3,4; 3,2; 3,3; 3,5; 3,3	$\bar{x}_4 = 3{,}4$	
5	3,0; 3,4; 3,2; 3,5; 2,9; 3,1	$\bar{x}_5 = 3{,}18$	

Вид варіації ознаки	Сума квадратів відхилень	Число ступенів свободи	Статистичні оцінки дисперсій
внутрішньо-групова	$\sum_{i=1}^{n_j}\sum_{j=1}^{5}(x_{ij}-\bar{x}_j)^2 = 0{,}926734$	$k_1 = N - p = 30 - 5 = 25$	$S_1^2 = 0{,}03707$
міжгрупова	$\sum_{j=1}^{5} n_j (x_{ij}-\bar{x}_j)^2 = 1{,}3377$	$k_2 = p - 1 = 5 - 1 = 4$	$S_2^2 = 0{,}3344$

$$F_{\text{кр}}(\alpha = 0{,}001;\ k_1 = 4;\ k_2 = 25) = 6{,}6\,;$$

$$F^* = \frac{S_2^2}{S_1^2} = \frac{0{,}3344}{0{,}03707} = 9{,}0208\,.$$

Висновок. Оскільки $F^* > F_{\text{кр}}$, то вплив каталізатора на кінцевий продукт хімічної реакції є істотним.

3. Двофакторний дисперсійний аналіз

Нехай необхідно визначити вплив двох факторів *A* і *B* на певну ознаку *X*. Для цього необхідно, щоб дослід здійснювався при фіксованих рівнях факторів *A* і *B*, а також їх одночасній дії на ознаку. При цьому дослід здійснюватимемо *n* раз для кожного з рівнів факторів *A* і *B*.

Позначимо через x_{ijk} конкретне значення ознаки *X*, якого вона набуває при *i*-му експерименті, *j*-му рівні фактора *A* і *k*-му рівні фактора *B*.

Результат експерименту зручно подати у вигляді таблиці, яка поділена на блоки, в кожному з яких ураховується на певних рівнях факторів *A* і *B* їх вплив на конкретні значення ознаки $X = x_{ijk}$ (табл. 3).

Виходячи з даних табл., $$\bar{x}_{ij} = \frac{\sum_{i=1}^{n} x_{ijk}}{n} \qquad (474)$$

є середнім значенням ознаки *X* для кожного блока;

$$\bar{z}_j = \frac{\sum_{i=1}^{n}\sum_{j=1}^{q} x_{ijk}}{nq}, \qquad j = \overline{1, p} \qquad (475)$$

є середнім значенням ознаки *X* за стовпцями;

$$\bar{y}_i = \frac{\sum_{i=1}^{n}\sum_{j=1}^{q} x_{ijk}}{np}, \qquad i = \overline{1, q} \qquad (476)$$

є середнім значенням ознаки *X* за рядками;

$$\bar{x} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{q}\sum_{k=1}^{p} x_{ijk}}{npq} \qquad (477)$$

є загальною середньою ознакою *X*;

$$S_1^2 = \frac{np\sum(\bar{y}_i - \bar{x})^2}{p-1} = \frac{Q_1}{p-1} \qquad (478)$$

є виправленою дисперсією, яка зумовлена впливом фактора *A* на ознаку *X*;

$$S_2^2 = \frac{nq\sum(\bar{z}_j - \bar{x})^2}{q-1} = \frac{Q_2}{q-1} \qquad (479)$$

є виправленою дисперсією, яка зумовлена впливом фактора *B* на ознаку *X*;

Таблиця 3

Рівень фактора B	Рівень фактора A							Середня величина за рядками	Загальна середня величина
	A_1		A_2		.....	A_p			
		блочна середня		блочна середня			блочна середня		
B_1	$x_{111}, x_{211}, x_{311}, \ldots, x_{n11}$	$\bar{x}_{11} = \frac{\sum_{i=1}^{n} x_{i11}}{n}$	$x_{112}, x_{212}, x_{312}, \ldots, x_{n12}$	$\bar{x}_{12} = \frac{\sum_{i=1}^{n} x_{i12}}{n}$	.....	$x_{11p}, x_{21p}, x_{31p}, \ldots, x_{n1p}$	$\bar{x}_{1p} = \frac{\sum_{i=1}^{n} x_{i1p}}{n}$	$\bar{y}_1 = \frac{\sum_{i=1}^{n}\sum_{k=1}^{p} x_{i1k}}{np}$	$\bar{x} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{g}\sum_{k=1}^{p} x_{ijk}}{npg}$
B_2	$x_{121}, x_{221}, x_{321}, \ldots, x_{n21}$	$\bar{x}_{21} = \frac{\sum_{i=1}^{n} x_{i21}}{n}$	$x_{122}, x_{222}, x_{322}, \ldots, x_{n22}$	$\bar{x}_{22} = \frac{\sum_{i=1}^{n} x_{i22}}{n}$	.....	$x_{12p}, x_{22p}, x_{32p}, \ldots, x_{n2p}$	$\bar{x}_{2p} = \frac{\sum_{i=1}^{n} x_{i2p}}{n}$	$\bar{y}_2 = \frac{\sum_{i=1}^{n}\sum_{k=1}^{p} x_{i2k}}{np}$	
.....	.....	.....	.....	.....	.....	.....	.....	.....	
B_g	$x_{1g1}, x_{2g1}, x_{3g1}, \ldots, x_{ng1}$	$\bar{x}_{g1} = \frac{\sum_{i=1}^{n} x_{ig1}}{n}$	$x_{1g2}, x_{2g2}, x_{3g2}, \ldots, x_{ng2}$	$\bar{x}_{g2} = \frac{\sum_{i=1}^{n} x_{ig2}}{n}$	.....	$x_{1gp}, x_{2gp}, x_{3gp}, \ldots, x_{ngp}$	$\bar{x}_{gp} = \frac{\sum_{i=1}^{n} x_{igp}}{n}$	$\bar{y}_g = \frac{\sum_{i=1}^{n}\sum_{k=1}^{p} x_{igk}}{np}$	
Середня величина за стовпцями	$\bar{z}_1 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{g} x_{ij1}}{ng}$		$\bar{z}_2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{g} x_{ij2}}{ng}$		.....	$\bar{z}_{gp} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{g} x_{ijp}}{ng}$			

Джерело, що спонукає до розсіювання	Сума квадратів відхилень	Число ступенів свободи	Статистичні оцінки дисперсій (виправлені дисперсії)
Фактор A	$Q_1 = np\sum_{j=1}^{q}(\bar{y}_j - \bar{x})^2$	$p-1$	$S_1^2 = \frac{Q_1}{p-1}$
Фактор B	$Q_2 = np\sum_{j=1}^{q}(\bar{y}_j - \bar{x})^2$	$q-1$	$S_2^2 = \frac{Q_2}{q-1}$
Одночасна дія факторів A і B	$Q_3 = \sum_{j=1}^{q}\sum_{k=1}^{p}(\bar{x}_{jk} - \bar{z}_j - \bar{y}_i + \bar{x})^2$	$(p-1)(q-1)$	$S_3^2 = \frac{Q_3}{(p-1)(q-1)}$
Дія випадкової компоненти ε_{ijk}	$Q_4 = \sum_{i=1}^{n}\sum_{j=1}^{q}\sum_{k=1}^{p}(x_{ijk} - \bar{x}_{jk})^2$	$N-pq$	$S_4^2 = \frac{Q_4}{N-pq}$
Загальне відхилення	$Q = \sum_{i=1}^{n}\sum_{j=1}^{q}\sum_{k=1}^{p}(x_{ijk} - \bar{x})^2$	$N-1$	$S^2 = \frac{Q}{N-1}$

$$S_3^2 = \frac{\sum\sum(\bar{x}_{ij} - \bar{z}_j - \bar{y}_i + \bar{x})^2}{(p-1)(q-1)} = \frac{Q_3}{(p-1)(q-1)} \qquad (480)$$

є виправленою дисперсією, яка зумовлена одночасним впливом на ознаку X факторів A і B;

$$S_4^2 = \frac{\sum\sum\sum(x_{ijk} - \bar{x}_{jk})^2}{N - pq} = \frac{Q_4}{N - pq} \qquad (481)$$

є виправленою дисперсією, яка зумовлена впливом на ознаку X інших, не головних факторів.

Обчислюються спостережувані значення критерію

$$F_A^* = \frac{S_\sigma^2}{S_m^2}; \quad F_B^* = \frac{S_\sigma^2}{S_m^2}; \quad F_{AB}^* = \frac{S_\sigma^2}{S_m^2}.$$

При рівні значущості α визначають критичні точки:

$F_{\text{кр}}(\alpha; k_4, k_1)$, $F_{\text{кр}}(\alpha; k_3, k_1)$, $F_{\text{кр}}(\alpha; k_2, k_1)$.

Якщо:

1) $F_A^* > F_{\text{кр}}(\alpha; k_4, k_1)$, то нульова гіпотеза про відсутність впливу фактора A відхиляється;

2) $F_B^* > F_{\text{кр}}(\alpha; k_3, k_1)$, то нульова гіпотеза про відсутність впливу фактора B відхиляється;

3) $F_{AB}^* > F_{\text{кр}}(\alpha; k_2, k_1)$, то нульова гіпотеза про відсутність спільного впливу факторів A і B відхиляється.

Приклад. У чотирьох різних лабораторіях здійснювався експеримент з опріснення морської води за допомогою трьох опріснювачів. Кожний експеримент для кожного опріснювача в кожній лабораторії проводився тричі. Наслідки опріснювання, виражені у відсотках, наведені в таблиці:

Ступінь впливу B	Ступінь впливу A		
	A_1	A_2	A_3
B_1	3,6; 3,9; 4,1	2,9; 3,1; 3,0	2,7; 2,5; 2,9
B_2	4,2; 4,0; 4,1	3,3; 2,9; 3,2	3,7; 3,5; 3,6
B_3	3,8; 3,5; 3,6	3,6; 3,7; 3,5	3,2; 3,0; 3,4
B_4	3,4; 3,2; 3,2	3,4; 3,6; 3,5	3,6; 3,8; 3,7

При рівні значущості α = 0,05 перевірити, чи існує вплив факторів A і B, а також спільної дії A і B на кінцевий результат експерименту.

Розв'язання. Використовуючи табл. 3, дістанемо табл. 4:

Таблиця 4

Ступінь впливу фактора B		Ступінь впливу фактора A						Середня величина за рядками	Загальна середня
		A_1		A_2		A_3			
		перший рівень	блочна середня	другий рівень	блочна середня	третій рівень	блочна середня		
B_1	Перший рівень	3,6; 3,8; 4,1	$\bar{x}_{11} = 3{,}83$	2,9; 3,1; 3,0	$\bar{x}_{12} = 3$	2,7; 2,5; 2,9	$\bar{x}_{13} = 2{,}7$	$\bar{y}_1 = 3{,}18$	$\bar{x} = 3{,}44$
B_2	Другий рівень	4,2; 4,0; 4,1	$\bar{x}_{21} = 4{,}1$	3,3; 2,9; 3,2	$\bar{x}_{22} = 3{,}13$	3,7; 3,5; 3,6	$\bar{x}_{23} = 3{,}6$	$\bar{y}_2 = 3{,}61$	
B_3	Третій рівень	3,8; 3,5; 3,6	$\bar{x}_{31} = 3{,}63$	3,6; 3,7; 3,5	$\bar{x}_{32} = 3{,}6$	3,2; 3,0; 3,4	$\bar{x}_{33} = 3{,}2$	$\bar{y}_3 = 3{,}48$	
B_4	Четвертий рівень	3,4; 3,2; 3,2	$\bar{x}_{41} = 3{,}27$	3,4; 3,6; 3,5	$\bar{x}_{42} = 3{,}5$	3,6; 3,8; 3,7	$\bar{x}_{43} = 3{,}7$	$\bar{y}_4 = 3{,}49$	
Середня величина за стовпцями		$\bar{z}_1 = 3{,}71$		$\bar{z}_2 = 3{,}31$		$\bar{z}_3 = 3{,}3$			

Джерело, що спонукає до розсіювання	Сума квадратів відхилень	Число ступенів свободи	Статистичні оцінки дисперсії (виправлені дисперсії)
Фактор A	$Q_1 = ng\sum_{k=1}^{p}(\bar{z}_k - \bar{x})^2 = 12\sum_{k=1}^{3}(\bar{z}_k - 3{,}44)^2 = 1{,}3128$	$p - 1 = 2$	$S_1^2 = \frac{Q_1}{p-1} = 0{,}6564$
Фактор B	$Q_2 = np\sum_{j=1}^{g}(\bar{y}_j - \bar{x})^2 = 9\sum_{j=1}^{3}(\bar{y}_j - 3{,}44)^2 = 0{,}9054$	$q - 1 = 3$	$S_2^2 = \frac{Q_2}{g-1} = 0{,}3018$
Одночасний вплив факторів A і B	$Q_3 = n\sum_{j=1}^{g}\sum_{k=1}^{p}(\bar{x}_{jk} - \bar{z}_k - \bar{y}_j + \bar{x})^2 = 3\sum_{j=1}^{4}\sum_{k=1}^{3}(\bar{x}_{jk} - \bar{z}_k - \bar{y}_j + \bar{x})^2 = 2{,}7873$	$(p-1)\times(q-1) = 6$	$S_3^2 = \frac{Q_3}{(p-1)(q-1)} = 0{,}4646$
Вплив випадкових факторів	$Q_4 = \sum_{i=1}^{3}\sum_{j=1}^{4}\sum_{k=1}^{3}(x_{ijk} - \bar{x}_{jk})^2 = 0{,}5668$	$N - pq = 24$	$S_4^2 = \frac{Q_4}{N - pq} = 0{,}02362$
Загальна дисперсія	$Q = \sum_{i=1}^{3}\sum_{j=1}^{4}\sum_{k=1}^{3}(x_{ijk} - \bar{x})^2 = \sum_{i=1}^{3}\sum_{j=1}^{4}\sum_{k=1}^{3}(x_{ijk} - 3{,}44)^2 = 5{,}5723$	$N - 1 = 35$	$S^2 = \frac{Q}{N-1} = 0{,}1675$

Визначимо спостережувані значення статистичних критеріїв за кожним фактором *A*, *B* та їх сумісної дії:

$$F_A^* = \frac{S_1^2}{S_4^2} = \frac{0{,}6564}{0{,}02362} = 27{,}79\,;$$

$$F_B^* = \frac{S_2^2}{S_4^2} = \frac{0{,}3018}{0{,}02362} = 12{,}78\,;$$

$$F_{AB}^* = \frac{S_3^2}{S_4^2} = \frac{0{,}4646}{0{,}02362} = 19{,}67\,;$$

$$\left.\begin{array}{l} F_{\text{кр}}(\alpha = 0{,}05;\ k_1 = 1;\ k_2 = 23) = 4{,}3; \\ F_{\text{кр}}(\alpha = 0{,}05;\ k_1 = 2;\ k_2 = 23) = 3{,}4; \\ F_{\text{кр}}(\alpha = 0{,}05;\ k_1 = 5;\ k_2 = 23) = 2{,}7; \end{array}\right\} \text{ за таблицею (додаток 7).}$$

Висновок. Оскільки $F_A^* > F_{\text{кр}}$, то нульова гіпотеза про відсутність впливу фактора *A* відхиляється, аналогічно $F_B^* > F_{\text{кр}}$, $F_{AB}^* > F_{\text{кр}}$, а це інформує про те, що нульова гіпотеза про відсутність впливу фактора *B*, а також сумісної дії факторів *A* і *B* також відхиляється.

Теоретичні запитання до теми ?

1. У чому сутність дисперсійного аналізу?
2. Записати математичну модель для однофакторного дисперсійного аналізу.
3. Записати математичну модель для двофакторного дисперсійного аналізу.
4. Що таке рівень впливу певного фактора на досліджувану ознаку *X* ?
5. Розкрити сутність x_{ijk} .
6. Властивості випадкових компонент ε_{ij} , ε_{ijk} .
7. Що називають внутрішньогруповою дисперсією?
8. Що називають міжгруповою дисперсією для однофакторного дисперсійного аналізу?
9. Що називають груповою середньою?
10. Групові та загальні середні і формули для їх обчислення.
11. Виправлена дисперсія, що характеризує розсіювання в середні групи, та формула для її обчислення.

12. Виправлена дисперсія, що характеризує вплив фактора, та формула для її обчислення.

13. Число ступенів свободи для S_1^2.

14. Число ступенів свободи для S_2^2.

15. Статистичний критерій для перевірки істотності впливу фактора на досліджувану ознаку *X*.

16. Виправлена дисперсія, що вимірює розсіювання ознаки під впливом фактора *A*, та формула для її обчислення.

17. Число ступенів свободи для S_1^2.

18. Виправлена дисперсія, що вимірює розсіювання ознаки під впливом фактора *B*, та формула для її обчислення.

19. Число ступенів свободи для S_2^2.

20. Виправлена дисперсія, що вимірює розсіювання ознаки під впливом факторів *A* і *B* сумісно, та формула для її обчислення.

21. Виправлена дисперсія, що вимірює розсіювання під впливом інших випадкових факторів, та формула для її обчислення.

22. Формула для обчислення блочних середніх.

23. Формула для обчислення середнього значення ознаки за стовпцями.

24. Формула для обчислення середнього значення ознаки за рядками.

25. Формула для обчислення загального середнього.

26. Обчислення спостережуваних значень критеріїв F_A^*, F_B^*, F_{AI}^*.

Задачі до теми

1. У результаті проведення досліду з метою з'ясування впливу чорного пару на врожайність пшениці з ділянки в 9 га (3 га були під чорним паром; 3 га — під картоплею; 3 га — під кормовими травами) дістали такі результати:

Фактор	Врожайність, ц/га
Чорний пар	26,6; 26,6; 30,6
Площа під картоплею	24,3; 25,2; 25,2
Площа під кормовими травами	26,6; 28,0; 31,0

За рівень значущості береться $\alpha = 0{,}01$.

Відповідь. $F^* = \dfrac{S_\delta^2}{S_m^2} = 11{,}36$; $F_{\text{кр}}(\alpha = 0{,}01,\ k_1 = 2, k_2 = 6) = 10{,}9$; $F^* > F_{\text{кр}}$, вплив чорного пару на врожайність пшениці є істотним.

2. Експериментально досліджувався вплив на зносостійкість колінчатих валів технології їх виготовлення — вплив фактора *A*, який має чотири рівні, тобто застосовувалися чотири технології виготовлення валів.

Одержані результати наведено в таблиці:

Ступінь впливу фактора *A*	Кількість відпрацьованих місяців
A_1	9; 8; 10; 12
A_2	10; 12; 11; 8
A_3	8; 16; 10; 18
A_4	9; 18; 10; 8

При рівні значущості α = 0,01 перевірити вплив технологій на зносостійкість валів.

Відповідь. $F^* = \frac{S_\delta^2}{S_m^2} = 1{,}906$; $F_{кр}(\alpha = 0{,}01,\ k_1 = 3,\ k_2 = 12) = 6{,}0$; $F^* < F_{кр}$, і вплив технологій на зносостійкість є неістотним.

3. Для перевірки впливу методики навчання виробничим навикам на якість підготовки із випускників виробничо-технічного училища навмання вибирають чотири групи учнів, які після закінчення навчання за різними методиками тестувалися на кількість виготовлених однотипних деталей протягом робочої зміни.

Результати тестування наведено в таблиці:

Ступінь впливу фактора *A* (методики)	Кількість виготовлених деталей за робочу зміну
A_1	60, 80, 75, 80, 85, 70
A_2	75, 66, 85, 80, 70, 80, 90
A_3	60, 80, 65, 60, 86, 75
A_4	95, 85, 100, 80

При рівні значущості α = 0,05 з'ясувати вплив методики навчання на якість підготовки учнів.

Відповідь. $F^* = \frac{S_\delta^2}{S_m^2} = 3{,}88$; $F_{кр}(\alpha = 0{,}05,\ k_1 = 3,\ k_2 = 19) = 3{,}1$; оскільки $F^* > F_{кр}$, то вибір методики впливає на виробничі навики підготовлених фахівців.

4. Досліджується залежність урожайності пшениці від сорту пшениці, яких чотири.

Результати досліджень наведені в таблиці:

Ступінь впливу фактора A (сорт пшениці)	Урожайність, ц/га
A_1	28,7; 26,7; 21,6; 25,0; 28,2
A_2	24,5; 28,5; 27,7; 28,7; 32,5
A_3	23,2; 24,7; 20,0; 24,0; 24,0
A_4	29,0; 28,7; 20,5; 28,0; 27,0

При рівні значущості α = 0,01 з'ясувати вплив сортності пшениці на її врожайність.

Відповідь. $F^* = \frac{S_2^2}{S_1^2} = 4{,}11$; $F_{\text{кр}}(\alpha = 0{,}01,\ k_1 = 3,\ k_2 = 15) = 5{,}4$; $F^* < F_{\text{кр}}$, вплив сортності пшениці на її врожайність несуттєвий.

5. Стальні болти з різною добавкою компоненти A в сталі, з якої вони виготовлялися, були піддані випробуванням на міцність.

Результати цих випробувань наведено в таблиці:

Ступінь впливу фактора A (відсоткова добавка)	Міцність, кг/мм^2
A_1	25; 28; 20; 22
A_2	29; 22; 21; 18
A_3	19; 25; 30; 22
A_4	18; 30; 24; 20

При рівні значущості α = 0,01 з'ясувати вплив добавки компоненти на міцність болта.

Відповідь. $F^* = \frac{S_\delta^2}{S_m^2} = 11{,}02$; $F_{\text{кр}}(\alpha = 0{,}01,\ k_1 = 3,\ k_2 = 12) = 6{,}0$; $F^* > F_{\text{кр}}$, вплив добавки компоненти A в сталі є суттєвим.

6. Електролампочки напругою 220 В виготовлялися на трьох заводах із використанням різних технологій. З кожної партії, що надходили в науково-дослідний інститут від кожного заводу, навмання брали по чотири електролампочки і піддавали їх випробуванням на тривалість горіння.

Результати цього експерименту наведено в таблиці:

Ступінь впливу фактора A (технології виготовлення)	Тривалість горіння, год
A_1	90; 85; 105; 110; 95
A_2	80; 110; 115; 90; 105
A_3	75; 120; 110; 90; 85

При рівні значущості α = 0,01 з'ясувати вплив технологій виготовлення на тривалість горіння лампочок.

Відповідь. $F^* = \dfrac{S_\delta^2}{S_m^2} = 5{,}096$; $F_{кр}(\alpha = 0{,}01,\ k_1 = 12,\ k_2 = 2) = 99{,}4$; $F^* < F_{кр}$, не впливає.

7. Рейтинг лівих партій, що вимірювався у навмання вибраних шести районах на Заході України, у центральній її частині і на Сході, дав такі результати:

Ступінь впливу фактора *A*	Рейтинг, %
A_1 (західні райони)	14,5; 5,6; 23,8; 6,4; 26,2; 14,5
A_2 (центральні райони)	22,5; 12,2; 24,8; 16,8; 11,9; 26,6
A_3 (східні райони)	13,4; 20,8; 30,8; 20,8; 6,4; 12,3

При рівні значущості α = 0,001 з'ясувати, чи впливає істотно регіон України на рейтинг лівих партій.

Відповідь. $F^* = \dfrac{S_\delta^2}{S_m^2} = 1{,}82$; $F_{кр}(\alpha = 0{,}01,\ k_1 = 2,\ k_2 = 15) = 6{,}4$; $F^* < F_{кр}$, регіони не істотно впливають на рейтинг.

8. На дослідних ділянках, кожна з яких має площу 6 га, досліджувалась залежність урожайності пшениці від внесення в ґрунт добрив A_1, A_2, A_3, A_4.

Результати експерименту наведено в таблиці:

Ступінь впливу фактора *A* (тип добрива)	Урожайність, ц/га
A_1	25,6; 36,2; 22,8; 30,2; 32,5; 28,4
A_2	28,5; 40,6; 42,8; 36,4; 22,4; 29,6
A_3	24,4; 38,6; 48,4; 50,2; 28,4; 22,8
A_4	29,5; 52,8; 24,2; 22,8; 56,2; 48,4

При рівні значущості α = 0,01 з'ясувати вплив типу добрива, що вноситься в ґрунт, на урожайність пшениці.

Відповідь. $F^* = \dfrac{S_\delta^2}{S_m^2} = 5{,}47$; $F_{кр}(\alpha = 0{,}01,\ k_1 = 3,\ k_2 = 23) = 4{,}7$; $F^* > F_{кр}$, вплив є істотним.

9. З кожної із 8 партій однотипних заготовок навмання бралися заготовки, які оброблялися на трьох верстатах-автоматах різної модифікації. Кількість деталей, виготовлена верстатами, досліджувалася на стандартність.

Результати досліджень подано в таблиці:

Фактор A (тип верстатів-автоматів)	Кількість деталей, виготовлених верстатами-автоматами, що відповідають стандарту
A_1	100; 86; 90; 89; 95; 22; 80; 79
A_2	99; 82; 98; 88; 100; 96; 98; 100
A_3	100; 88; 86; 98; 98; 100; 99; 99

При рівні значущості $\alpha = 0{,}01$ перевірити вплив модифікації верстатів-автоматів на якість виготовлених деталей.

Відповідь. $F^* = \dfrac{S_\delta^2}{S_m^2} = 4{,}12$; $F_{кр}(\alpha = 0{,}01,\ k_1 = 2,\ k_2 = 21) = 5{,}9$; $F^* < F_{кр}$, вплив неістотний.

10. Проводилось дослідження розподілу числа кров'яних тілець у певній одиниці об'єму крові в людей, що перебували певний час у трьох зонах на різній відстані від Чорнобильської АЕС та в зоні, вільній від радіації.

Результати досліджень наведено в таблиці:

Фактор A (зони)	Кількість кров'яних тілець
A_1 (в зоні АЕС)	6; 8; 3; 2; 6; 9
A_2 (на відстані 50 км)	5; 4; 10; 11; 6; 8
A_3 (на відстані 100 км)	5; 4; 13; 12; 10; 15
A_4 (вільна від радіації зона)	18; 16; 21; 20; 22; 21

При рівні значущості $\alpha = 0{,}01$ з'ясувати вплив перебування людини в певній зоні на кількість кров'яних тілець.

Відповідь. $F^* = \dfrac{S_\delta^2}{S_m^2} = 23{,}2$; $F_{кр}(\alpha = 0{,}01,\ k_1 = 3,\ k_2 = 20) = 4{,}9$; $F^* > F_{кр}$, існує вплив.

Лабораторна робота № 2
до теми «Елементи дисперсійного аналізу»

При рівні значущості $\alpha = 0{,}05$ визначити вплив факторів A і B та їх сумісний вплив на досліджувану ознаку:

1. Досліджується вплив на зносостійкісіть деталей таких факторів: A — матеріал для виготовлення деталей (застосовували три види сталі) і B — технологія виготовлення деталей (дві технології).

Результати експерименту наведено в таблиці:

Фактор B	Фактор A		
	A_1	A_2	A_3
B_1	10; 7; 8; 6; 12; 8; 11; 10; 14; 13	8; 14; 6; 10; 16; 14; 13; 12; 11; 15	15; 12; 11; 9; 8; 13; 11; 12; 16; 14
B_2	12; 13; 6; 9; 8; 11; 10; 10; 13; 17	11; 12; 12; 16; 13; 8; 10; 9; 8; 15	13; 12; 14; 8; 6; 8; 16; 12; 14; 16

2. Досліджується вплив на врожайність ячменю таких факторів: *A:* посів здійснюється після чорного пару — A_1; після коренеплодів — A_2; після колосових культур — A_3; B — сортність ячменю (три сорти).

Результати досліджень наведено в таблиці:

Фактор B	Фактор A		
	A_1	A_2	A_3
B_1	34,2; 30,6; 36,8; 35; 32,5; 34,2; 33,4; 36	42,5; 40,4; 44,6; 46,8; 39,4; 38,6; 45,8; 49,3	44,2; 46; 45,6; 48; 49,3; 45,8; 42,3; 40,8; 41,4; 40
B_2	32,5; 30,4; 39,4; 40,3; 36,4; 38,9; 39,8; 42	30,3; 35,3; 36,8; 40,5; 28,4; 33,2; 39,1; 26,9	40,3; 45; 46,8; 30,2; 48,8; 50,2; 39; 38,5
B_3	33,3; 34,8; 39,2; 35; 32,4; 34; 39,8; 40,8	30,4; 36; 40,5; 44,4; 30,8; 42,5; 46; 33,5	32,3; 29,8; 34,3; 42; 34,8; 31,6; 40; 29,6

3. Досліджується вплив на міцність чавуну двох факторів: A — вміст кремнію в чавуні, а саме A_1 — 0,24%; A_2 — 0,42%; A_3 — 0,52%; B — температурний режим плавлення (два режими).

Результати досліджень наведено в таблиці:

Фактор B	Фактор A		
	A_1	A_2	A_3
B_1	40,2; 40,8; 38,2; 39,6; 42,4; 44,5; 40,1; 38,8	42,5; 43,4; 44,5; 46,4; 40,1; 36,5; 40,3; 41,8; 38; 43,5	49,2; 50,2; 48,4; 50; 52,5; 38,4; 49,8; 50,4; 51,8; 49
B_2	33,4; 36,5; 34,4; 40,2; 42; 30,2; 31,8; 35,5; 34; 41,8	31,6; 33,4; 38,4; 35; 38,9; 29,5; 28,4; 30,6; 32,9; 43	29,3; 35,6; 36; 26,8; 38; 28,5; 30,6; 40,2; 33,3

4. Досліджується вплив на врожайність кукурудзи двох факторів: *A* — внесення добрив у ґрунт (три добрива); *B* — глибина поливу землі (три глибини поливу).

Результати досліджень наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	30,2; 30,8; 31,6; 32; 32,6; 28,9; 30,5; 32,6; 33	28,4; 29,9; 30,6; 44,3; 36,2; 42,3; 28,2; 26,5; 34,3; 26,5	40,2; 42,3; 42,7; 43,5; 44; 36,8; 38,9; 45,3; 46,2; 45,4
B_2	44,2; 42,8; 43,7; 46,5; 46,9; 40,5; 45,6; 38,4; 32,5; 44,6	42,4; 43,5; 40,6; 36,8; 40; 36,4; 38,5; 43,2; 34,6; 39,8	42,3; 43,4; 45,2; 44; 36,5; 29,8; 25,4; 43,2; 45; 46,8
B_3	40,2; 36,4; 36,9; 41,8; 40,4; 34,8; 38,5; 35; 38,6; 42,4	38,5; 33,4; 30,2; 29,4; 40,1; 26,2; 25,4; 44,1; 30,6; 34,5	43,2; 44,5; 39,5; 32,5; 45; 40,8; 36,3; 43,5; 47,8; 49

5. Досліджується вплив факторів *A* і *B* на число виготовлених втулок зі ста взятих, які відповідають нормам стандарту: *A* — використали дві технології виготовлення; *B* — заготовки надходили із трьох заводів.

Результати досліджень наведено в таблиці:

Фактор *B*	Фактор *A*	
	A_1	A_2
B_1	90; 88; 90; 96; 98; 76; 80; 95; 85; 80	100; 99; 82; 98; 95; 80; 96; 95; 99; 91; 89; 90
B_2	79; 88; 92; 76; 80; 83; 85; 90; 96; 75	81; 82; 100; 98; 89; 85; 96; 98; 75; 97
B_3	82; 78; 75; 79; 80; 81; 86; 89; 75; 90	80; 86; 90; 91; 78; 76; 75; 82; 73; 82

6. Досліджується вплив факторів *A* і *B* на продуктивність праці підприємства певної галузі промисловості: *A* — фондозабезпеченість (три рівні); *B* — рівень оплати праці робітникам (два рівні).

Результати дослідження наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	14,85; 11,94; 10,5; 12,35; 15,62; 13,2; 10,62; 12,82; 11,48; 13,5	6,42; 5,23; 4,96; 5,6; 9,82; 10,23; 12,44; 16,5; 5,41; 6,32	7,82; 9,63; 12,92; 10,82; 9,36; 5,11; 13,52; 14,2; 8,96; 9,92
B_2	12,5; 13,8; 14,9; 12,6; 10,85; 11,96; 12,6; 13,42; 16; 17,2	10,2; 10,85; 12,34; 11,95; 12,4; 14,92; 9,86; 9,62; 8,36; 13,62	13,62; 12,55; 14,7; 13,25; 14,66; 8,35; 10,96; 11,62; 6,12; 15,66

7. Експериментально досліджувався вплив на зносостійкість деталей факторів *A* і *B*: фактор *A* — тип сталі (три типи); фактор *B* — технологія виготовлення (дві технології).

Результати експерименту наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	10; 8; 6; 9; 5; 12; 5; 8; 10; 11	8; 12; 12; 10; 11; 6; 10; 10; 9; 5	15; 14; 14; 8; 8; 13; 10; 11; 9; 6
B_2	12; 9; 9; 6; 6; 5; 10; 8; 8; 9	12; 13; 13; 14; 15; 8; 9; 10; 11; 11	13; 13; 10; 5; 5; 10; 15; 14; 14; 10

8. Експериментально досліджувався вплив факторів *A* і *B* на опріснення морської води: фактор *A* — тип опріснювача (три типи); фактор *B* — три різні лабораторії.

Результати експерименту, що інформує про відсоток солі в опрісненій воді, наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	3,6; 3,2; 3,4; 4,1; 3,5; 4,2; 3,5; 3,8; 4,2; 3	2,92; 2,84; 2,88; 3,2; 3,45; 3,02; 2,12; 2,26; 2,43; 3,5	2,7; 2,75; 2,97; 3,2; 4,15; 2,63; 2,49; 3,25; 3,4; 4,2
B_2	4,2; 4; 4,25; 4,35; 4,5; 3,6; 3,2; 3,2; 3,6; 3,8	3,33; 3,35; 4,2; 2,93; 2,65; 2,96; 2,25; 3,8; 3,96; 4,2	3,75; 3,87; 3,64; 2,95; 2,25; 3,85; 3,99; 4,2; 4,15; 3,14
B_3	3,46; 3,45; 4,25; 4,8; 4,1; 4,05; 3,81; 3,62; 3,4; 3,02	3,42; 3,49; 2,99; 4,1; 2,65; 3,11; 3,12; 4,41; 4,0; 3,8	3,63; 3,56; 2,99; 3,79; 4,12; 3,12; 2,05; 3,81; 3,79; 4,0

9. Експериментально досліджувався вплив факторів *A* і *B* на врожайність цукрових буряків в ц/га: фактор *A* — добрива (три види); фактор *B* — сума температур за період вегетації (три рівні).

Результати досліджень наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	365; 36; 370; 385; 350; 340; 342; 340; 365; 380	403; 410; 412; 416; 345; 374; 450; 430; 402; 412	452; 440; 403; 395; 382; 444; 410; 420; 433; 390
B_2	379; 381; 390; 420; 400; 402; 380; 340; 410; 390	445; 436; 470; 412; 390; 396; 380; 445; 444; 389	433; 391; 340; 455; 460; 405; 399; 413; 449; 401
B_3	332; 450; 420; 445; 390; 420; 422; 444; 380; 395	330; 413; 425; 449; 385; 399; 440; 412; 405; 382	325; 34; 412; 402; 390; 399; 375; 399; 401; 455

10. Досліджувався вплив факторів *A* і *B* на пружність сталі в умовних одиницях: фактор *A* — % нікелю (три рівні); фактор *B* — % марганцю (два рівні).

Результати досліджень наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	36,4; 38,7; 36,5; 37,5; 39,6; 40,2; 38,5; 42,6; 35,6; 38,4	41,2; 42; 42,8; 44,3; 45,2; 44,3; 42,4; 39,5; 38,4; 39,6	36,5; 39,8; 42,4; 45,8; 48,4; 49,5; 37,2; 38,4; 40,2; 40,5
B_2	39,2; 42,3; 44,5; 40,5; 38,1; 40,8; 45,3; 41,8; 38,7; 42	39,7; 38,4; 42,5; 44,3; 47,2; 48,4; 45,2; 46,4; 49,2; 49,8	40,8; 43,2; 41,8; 44,7; 50; 42,8; 35,6; 38,9; 47,2; 48,2

11. Досліджувався вплив факторів *A* і *B* на твердість сплаву: фактор *A* — % нікелю (три рівні); фактор *B* — % хрому (три рівні).

Результати досліджень наведено в таблиці, де твердість подається в умовних одиницях:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	53,2; 53,9; 54,8; 55,9; 62,2; 66,8; 70; 58,2; 54,4; 52,3	55,4; 66,7; 77,2; 53,2; 65,4; 66,2; 53,2; 58,1; 73,2; 75,4	68,3; 69,8; 74,7; 79,2; 53,4; 61,5; 58,4; 59,8; 76,2; 78,3
B_2	67,2; 66,2; 55,3; 53; 72,3; 52,4; 74,2; 52; 63,2; 53,2	77,2; 65,4; 53,9; 65,1; 63,4; 61,2; 71,4; 74,2; 54,2; 53,8	77,9; 62,3; 68,9; 64,5; 73,2; 53,1; 55,2; 54,4; 76,8; 78,9
B_3	70,2; 72,1; 54,4; 53,1; 73,4; 74,8; 75,2; 53; 54,2; 67,2	69,2; 65,4; 70,4; 55,4; 62,3; 72,5; 74,4; 70,5; 53,1; 54,2	75,5; 76,4; 54,2; 56,1; 62,3; 64,8; 73,4; 75,6; 79,2; 53,5

12. Досліджувався вплив на продуктивність праці факторів *A* і *B*: фактор *A* — плинність кадрів, % (три рівні); фактор *B* — рівень оплати праці, тис. грн. на рік (два рівні).

Результати досліджень наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	15,62; 14,3; 14,25; 15,81; 16,35; 15,61; 14,3; 12,5; 11,2; 6,5	10,83; 10,2; 13,4; 16,25; 12,2; 5,45; 6,41; 8,93; 13,44; 15,66	12,44; 14,5; 7,6; 6,75; 8,96; 16,37; 9,82; 7,83; 10,53; 8,96
B_2	16,52; 14,21; 6,85; 8,7; 10,43; 13,5; 12,8; 11,6; 6,72; 8,9	13,24; 8,16; 9,44; 10,8; 14,56; 12,46; 11,83; 10,99; 16,42; 15,34	6,81; 5,74; 10,36; 14,57; 12,44; 13,47; 15,25; 13,4; 5,07; 6,8

13. Досліджувався вплив факторів *A* і *B* на приріст урожаю картоплі з гектара: фактор *A* — внесення в ґрунт добрива (три рівні за кількістю у відсотках); фактор *B* — поливання ґрунту (два рівні проникнення вологи на глибину ґрунту).

Результати досліджень наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	4,25; 4,5; 5,6; 6,8; 4,05; 4,8; 7,6; 8,2; 4,02; 6,06	6,25; 4,95; 4,26; 8,29; 8,8; 9,25; 7,44; 7,8; 4,82; 5,61	5,44; 5,23; 9,82; 8,9; 4,35; 6,81; 7,84; 6,51; 4,08; 6,52
B_2	7,45; 4,05; 8,25; 9,6; 4,06; 5,25; 6,73; 5,76; 9,21; 4,01	8,28; 6,44; 7,35; 4,9; 4,22; 7,42; 8,82; 9,5; 4,08; 5,8	6,32; 7,81; 8,92; 8,6; 4,02; 5,21; 4,21; 9,47; 9,81; 10,22

14. Досліджувався вплив факторів *A* і *B* на масу корів у кг: фактор *A* — три види соковитих кормів; фактор *B* — два види грубих кормів із добавкою певного відсотка вітамінів.

Результати досліджень наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	455,6; 460,2; 350,2; 500; 521,6; 534,2; 605; 340; 390; 395,5	435,6; 489,6; 572,5; 399; 480; 550,6; 580; 341,5; 382,6; 599,5	331,4; 340,5; 390,6; 405,6; 545,7; 596,2; 320,2; 305,8; 421,6; 399,5
B_2	446,2; 480,5; 620,8; 700; 721,6; 750,2; 440,2; 600; 430,8; 444,6	600; 595,6; 401,8; 321,8; 340,4; 600; 431,8; 549,6; 590; 300,6	443,8; 389,5; 541,3; 590,6; 555,4; 481,6; 405,6; 311,8; 300,6; 375,8

15. Досліджувався вплив факторів *A* і *B* на рейтинг лівих партій у відсотках: фактор *A* — регіони (A_1 — західний, A_2 — південний, A_3 — східний); фактор *B* — віковий рівень опитаних (B_1 — до 30 років, B_2 — 30—55 років і B_3 — 55—70 років).

Результати досліджень наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	9,5; 5,5; 4,2; 6,7; 12,4; 16,8; 2,5; 10,2; 5,8; 6,4	4,2; 10,5; 8,9; 9,6; 12,4; 5,7; 7,3; 8,4; 13,4; 15,5	8,6; 7,5; 4,3; 19,8; 26,4; 3,2; 32,4; 3,8; 4,5; 3,6
B_2	2,5; 3,4; 7,8; 12,4; 2,8; 4,5; 3,9; 6,7; 2,3; 4,9	6,5; 7,2; 13,6; 22,4; 30,5; 4,2; 7,8; 4,8; 7,9; 12,4	12,5; 10,6; 22,4; 8,5; 4,3; 3,3; 7,8; 4,4; 5,6; 9,7
B_3	2,1; 3,3; 7,8; 2,2; 3,2; 4,6; 12,1; 13,1; 6,7; 8,5	4,5; 12,6; 22,5; 40,1; 3,6; 8,5; 31,6; 6,2; 3,2; 5,6	15,8; 35,6; 21,4; 3,2; 4,5; 3,6; 8,4; 9,1; 7,3; 4,2

16. Досліджувався вплив факторів *A* і *B* на вміст протеїну в курячих яйцях, %: фактор *A* — три типи зернових; фактор *B* — два типи комбікормів.

Результати досліджень наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	13,2; 15,6; 18,2; 13,4; 13,5; 16,4; 17,5; 14,9; 19,2; 13,1	13,9; 16,5; 14,4; 18,2; 13,1; 13,9; 19,1; 17,1; 13,2; 14,5	13,4; 18,9; 14,2; 13,5; 16,2; 13,1; 14,1; 19,1; 13,8; 14,2
B_2	14,9; 15,8; 19,2; 19,4; 18,5; 13,2; 16,4; 13,1; 17,6; 16,5	15,4; 13,2; 16,2; 13,1; 19,1; 16,2; 13,5; 14,5; 16,2; 14,1	15,4; 17,2; 18,4; 13,1; 19,3; 14,2; 15,2; 16,4; 13,1; 13,9

17. Досліджувався вплив факторів *A* і *B* на приріст урожаю соняшнику, ц/га: фактор *A* — три види добрив; фактор *B* — глибина зрошення (дві глибини).

Результати досліджень наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	6,2; 6,4; 6,3; 7,2; 8; 6,1; 7,2;7,4; 8,2; 6,3; 6,5	6,8; 7,2; 8,3; 9,2; 6,2; 7,1; 7,5; 6,2; 6,8; 9,4	9,4; 6,2; 6,8; 8,8; 8,5; 6,1; 9,2; 8,2; 7,6; 8,1; 6,3; 6,9
B_2	8,3; 9,1; 6,2; 6,8; 7,4; 8,2; 6,5; 8,3; 9,2; 6,5	7,4; 9,4; 6,5; 6,1; 7,2; 8,3; 9,2; 6,4; 9,4; 8,1	9,2; 7,5; 6,3; 8,9; 7,9; 7,2; 6,3; 9,3; 9,4; 8,2

18. Досліджувався вплив факторів *A* і *B* на рейтинг правих партій (у відсотках): фактор *A* — регіони (A_1 — західний, A_2 — центральний, A_3 — східний); фактор *B* — вік опитаних (B_1 — 20—35 років, B_2 — 35—50 років і B_3 — 50—70 років).

Результати досліджень наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	25,2; 10,2; 5,4; 13,2; 18,2; 5,2; 13,4; 15,2; 4,5; 19,2	10,6; 8,4; 11,2; 4,6; 5,8; 18,2; 16,4; 13,2; 4,8; 8,9	2,5; 6,4; 12,5; 14,8; 12,3; 8,5; 5,9; 8,9; 15,4; 12,8; 4,2; 3,9
B_2	4,3; 10,5; 20,3; 32,4; 5,6; 12,4; 6,2; 9,8; 16,8; 18,4	12,4; 4,3; 13,2; 5,6; 8,9; 14,8; 22,3; 6,8; 7,2; 11,4; 4,2	4,5; 4,9; 12,3; 15,6; 7,9; 8,9; 9,8; 13,9; 4,2; 6,9
B_3	14,3; 10,6; 28,4; 10,8; 7,4; 6,5; 4,5; 26,3; 30,2; 11,8	6,2; 7,5; 3,5; 12,4; 13,5; 16,4; 7,9; 8,9; 15,4; 10,8	14,8; 2,9; 5,9; 10,6; 8,5; 13,4; 2,2; 19,5; 7,9; 9,9

19. Досліджувався вплив факторів *A* і *B* на жирність молока, %: фактор *A* — три види соковитих кормів; фактор *B* — два види біодобавок.

Результати досліджень наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	3,25; 3,45; 3,55; 4,04; 4,08; 4,2; 3,3; 3,8; 3,45; 3,25	4,2; 3,95; 3,33; 4,1; 3,5; 3,42; 3,49; 3,59; 3,68; 3,79	3,99; 3,89; 4,32; 4,23; 4,4; 3,29; 3,25; 3,11; 4,45; 4,05
B_2	3,41; 3,45; 3,5; 4,45; 4,25; 4,33; 4,5; 4,29; 3,42; 3,41	3,3; 3; 3,42; 4,2; 4,29; 4,39; 3,8; 3,92; 3,99; 4,05; 4,11	4,2; 3,21; 3,2; 3,11; 4,29; 4,41; 4,5; 4,48; 3,81; 4,29

20. Досліджувався вплив факторів *A* і *B* на річні надої молока в літрах: фактор *A* — три види соковитих кормів; фактор *B* — два види грубих кормів із різним рівнем частки, %, спеціальних вітамінів.

Результати досліджень наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	5050; 4090; 6000; 6500; 8900; 2900; 2500; 6000; 10000; 9500	9500; 12000; 6300; 4500; 3900; 8500; 8600; 5900; 12400; 6900	12500; 8900; 6500; 7900; 8700; 9200; 10500; 14000; 7200; 5300
B_2	3800; 1050; 12900; 6900; 3950; 8000; 11200; 12400; 4900; 8900	12000; 11500; 8900; 4400; 9800; 6900; 7200; 6200; 10500; 9200	14500; 4300; 6700; 12400; 13200; 8400; 7900; 15200; 3200; 5500

21. Досліджувався вплив факторів *A* і *B* на приріст урожаю з гектара кукурудзи, ц/га: фактор *A* — три рівні внесення азотних добрив (частка), %; фактор *B* — два види глибини зрошення шляхом штучного поливу.

Результати досліджень наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	6,2; 6,53; 6,82; 7,42; 6,55; 8,56; 9,49; 10,25; 9,64; 6,89	7,63; 8,53; 6,92; 9,73; 11,25; 7,33; 6,25; 10,11; 12,55; 8,93	8,35; 9,44; 8,44; 9,89; 10,99; 11,35; 15,21; 14,25; 11,6; 6,2; 6,01
B_2	6,99; 8,49; 12,45; 13,4; 12,45; 6,85; 10,23; 9,51; 7,21; 11,92	9,47; 6,83; 6,74; 13,53; 15,41; 12,36; 8,79; 6,44; 12,35; 10,42	12,53; 14,5; 10,26; 8,96; 7,44; 6,72; 6,34; 14,39; 13,29; 11,95

22. Досліджувався вплив факторів A і B на норми витрат пального трактором на 100 га в умовних одиницях: фактор A — три види обробітку землі (A_1 — під злакові культури; A_2 — картопляні культури; A_3 — кормові трави); фактор B — пори року (B_1 — весна; B_2 — осінь).

Результати досліджень наведено в таблиці:

Фактор B	Фактор A		
	A_1	A_2	A_3
B_1	9,72; 10,55; 9,85; 12,44; 10,24; 11,24; 9,99; 11,95; 9,85; 10,95	9,95; 9,71; 9,99; 16,53; 14,91; 13,86; 12,44; 11,66; 10,31; 11,63	9,81; 10,64; 11,85; 12,44; 10,95; 13,44; 14,53; 9,84; 10,25; 10,96
B_2	11,44; 12,53; 15,64; 9,51; 9,73; 9,56; 10,41; 12,22; 9,34; 16,5; 9,2	12,44; 10,64; 11,53; 9,83; 11,62; 9,71; 13,44; 16,51; 15,32; 14,95	9,96; 10,24; 16,44; 13,53; 14,83; 9,71; 9,51; 10,43; 11,22; 9,7

23. Досліджувався вплив факторів A і B на яйценосність курей у штуках за 5 місяців: фактор A — добавка до корму з моркви A_1; силосу A_2; біокорму A_3; фактор B — два режими температури в птахофермі.

Результати досліджень наведено в таблиці:

Фактор B	Фактор A		
	A_1	A_2	A_3
B_1	180; 195; 205; 184; 196; 220; 245; 192; 210; 230; 260; 184	184; 180; 256; 196; 260; 195; 210; 216; 233; 205	196; 210; 236; 186; 192; 210; 181; 244; 212; 212; 182
B_2	189; 210; 219; 250; 180; 196; 244; 249; 260; 202	196; 199; 188; 199; 244; 236; 259; 212; 189; 204	213; 243; 189; 199; 244; 249; 254; 260; 195; 182

24. Досліджувався вплив факторів A і B на час, який витрачається на виготовлення однієї деталі в хвилинах: фактор A — розряд робітника (A_1 — четвертий, A_2 — п’ятий, A_3 — шостий); фактор B — два сорти сталі (B_1 і B_2).

Результати досліджень наведено в таблиці:

Фактор B	Фактор A		
	A_1	A_2	A_3
B_1	2,2; 3,4; 2,9; 3,5; 4,2; 2,8; 3,5; 3,9; 3,6; 4,6; 5,6; 5,8	2,9; 2,5; 4,4; 4,2; 3,8; 3,9; 5,4; 6,4; 5,2; 6,8	2,5; 6,8; 4,6; 5,9; 6,2; 4,2; 2,4; 2,6; 3,4; 3,8
B_2	5,4; 4,4; 3,8; 4,6; 2,4; 3,5; 6,8; 4,2; 3,6; 5,8	2,9; 5,4; 6,3; 5,4; 3,8; 3,9; 2,5; 4,6; 3,7; 4,8	3,6; 2,2; 2,5; 2,9; 6,2; 4,7; 5,2; 3,2; 3,9; 4,2

25. Досліджувався вплив факторів A і B на відсоток виготовлених бракованих деталей: фактор A — три технології виготовлення; фактор B — два сорти сталі.

Результати досліджень наведено в таблиці:

Фактор B	Фактор A		
	A_1	A_2	A_3
B_1	2,25; 3,45; 4,52; 4,2; 2,42; 2,04; 3,5; 3,9; 4,35; 4,42	2,24; 2,15; 5,12; 4,32; 3,25; 3,06; 3,11; 4,11; 2,99; 3,16	4,15; 3,91; 2,16; 2,99; 3,65; 2,09; 3,12; 4,8; 5,02; 3,09
B_2	2,95; 5,42; 2,6; 4,35; 2,26; 4,72; 5,62; 3,66; 3,66; 3,95	3,26; 3,33; 2,95; 3,96; 4,12; 4,05; 3,85; 3,96; 2,96; 2,06	3,15; 2,98; 2,15; 5,12; 5,56; 2,88; 3,91; 3,16; 4,15; 3,21

26. Досліджувався вплив факторів A і B на тривалість горіння електролампочок на вуличних ліхтарях, год: фактор A — три технології виготовлення електролампочок (A_1, A_2, A_3); фактор B — пори року (B_1 — літо, B_2 — зима).

Результати досліджень наведено в таблиці:

Фактор B	Фактор A		
	A_1	A_2	A_3
B_1	2020; 2010; 1900; 1950; 1990; 2050; 1860; 1800; 2005; 2002	2005; 2010; 1990; 1860; 2010; 2050; 1890; 1810; 2100; 1860	2100; 2090; 1990; 1660; 1960; 1760; 2150; 1960; 1965; 2010
B_2	1400; 1590; 1900; 1850; 1690; 1850; 1790; 1790; 2100; 2095	1850; 1790; 1650; 2005; 2100; 1770; 1890; 1860; 1620; 1610	2150; 2120; 1550; 1560; 1950; 1990; 2000; 1895; 1670; 1790

27. Досліджувався вплив в одиниці об’єму (см^3) факторів A і B на число кров’яних тілець у людей: фактор A — досліджувались люди, які мешкають: A_1 — в сільській місцевості, A_2 — в міській; фактор B — вік людей (B_1 — від 20 до 30 років, B_2 — від 30 до 45 років і B_3 — від 45 до 55 років).

Результати досліджень наведено в таблиці:

Фактор B	Фактор A	
	A_1	A_2
B_1	18; 16; 22; 21; 20; 19; 22; 24; 20; 19	16; 19; 15; 22; 15; 17; 18; 20; 17; 14
B_2	16; 18; 14; 20; 22; 16; 22; 17; 16; 15	19; 20; 18; 14; 22; 20; 21; 14; 15; 16
B_3	14; 22; 16; 19; 15; 16; 21; 16; 20; 15	14; 16; 22; 15; 17; 21; 19; 16; 14; 18

28. Досліджувався вплив факторів *A* і *B* на міцність волокна бавовни, г: фактор *A* — район вирощування бавовни (три райони); фактор *B* — агрономія вирощування бавовни (два види).

Результати досліджень наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	3,75; 4,25; 3,5; 3,95; 4,75; 5,25; 3,65; 6,05; 3,15; 3,95	4,75; 4,25; 5,85; 3,25; 3,65; 5,25; 4,45; 4,05; 5,05; 4,05	5,05; 3,25; 4,05; 3,75; 4,15; 5,25; 4,25; 3,05; 5,65; 4,05
B_2	3,95; 5,55; 4,75; 3,65; 4,25; 4,85; 5,25; 4,05; 3,85; 5,45	4,15; 3,25; 3,75; 4,25; 5,15; 5,65; 5,05; 4,05; 5,75; 4,05	4,95; 4,15; 3,95; 3,15; 5,55; 5,15; 4,85; 4,1; 3,05; 3,65

29. Досліджувався вплив факторів *A* і *B* на відхилення діаметрів валиків від номіналу (за модулем), мкм: фактор *A* — технологія виготовлення (три види технологій); фактор *B* — сталь із різним відсотком вмісту хрому (два сорти сталі).

Результати досліджень наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	15; 18; 20; 22; 5; 8; 10; 8; 12; 6	12; 6; 25; 22; 18; 24; 8; 10; 5; 12	9; 26; 20; 6; 4; 26; 14; 18; 5; 10
B_2	20; 5; 9; 8; 25; 6; 4; 5; 10; 12	12; 4; 6; 4; 25; 8; 5; 12; 16; 9	10; 22; 16; 4; 5; 8; 6; 8; 24; 10

30. Досліджувався вплив факторів *A* і *B* на врожайність пшениці з одного гектара, ц/га: фактор *A* — три сорти добрив, що вносились у ґрунт; фактор *B* — дві технології обробітку ґрунту під посів.

Результати спостережень наведено в таблиці:

Фактор *B*	Фактор *A*		
	A_1	A_2	A_3
B_1	15,5; 20,2; 18,4; 22,3; 16,4; 21,5; 19,8; 21,5; 25,2; 15,4	16,4; 14,2; 22,8; 19,8; 17,3; 18,5; 25,2; 20,4; 26,1; 14,3	19,8; 16,5; 14,9; 22,8; 24,9; 15,3; 18,9; 23,4; 26,8; 17,2
B_2	21,3; 20,4; 16,9; 15,4; 24,8; 23,2; 18,4; 19,9; 17,4; 23,8	18,4; 23,8; 26,2; 14,8; 18,9; 25,2; 20,8; 15,9; 19,9; 16,3	20,2; 21,3; 15,9; 16,4; 18,5; 24,9; 21,4; 19,5; 25,8; 14,8

ТЕМА 16. ЕЛЕМЕНТИ КОРЕЛЯЦІЙНОГО ТА РЕГРЕСІЙНОГО АНАЛІЗУ

1. Загальна інформація

Кожній величині, яку дістають у результаті проведення експерименту, притаманний елемент випадковості, що виявляється більшою чи меншою мірою залежно від її природи.

При сумісній появі двох і більше величин у результаті проведення експерименту дослідник має підстави для встановлення певної залежності між ними, зв'язку.

Ідея зв'язку між змінними величинами має особливе, принципове значення в економетричних дослідженнях, де здійснюється перевірка на адекватність створених математичних моделей реальним економічним процесам, в яких співвідношення між змінними пов'язані функціональною залежністю.

Строгої функціональної залежності між змінними, у буквальному розумінні цього слова, у реальному світі не існує, бо вони перебувають під впливом випадкових факторів, наслідки якого передбачити практично неможливо. Тому між змінними існує особлива форма зв'язку, яку називають стохастичною (про що йшлося в попередніх темах) і яка в математичній статистиці трансформується, не змінюючи своєї сутності, у статистичну залежність.

Наприклад, при дослідженні двох змінних X та Y зміна значень $X = x_i$ призводить до такої зміни значень Y, яку можна розбити на два компоненти: систематичну, що пов'язана із залежністю, котра існує між X та Y, і випадкову, яка зазнає впливу випадкових факторів.

Показником, що вимірює стохастичний зв'язок між змінними, є *коефіцієнт кореляції*, який свідчить з певною мірою ймовірності, наскільки зв'язок між змінними близький до строгої лінійної залежності.

Значно збільшується цінність коефіцієнта кореляції для випадкових змінних, що мають закон розподілу ймовірностей, близький до нормального. Для таких величин відсутність кореляції одночасно означає і відсутність будь-якої залежності між ними.

Крім цього, як і в дисперсійному аналізі, кореляційний аналіз оцінює, наскільки значні невипадкові змінні у випадкових величинах у процесі проведення експерименту.

За наявності кореляційного зв'язку між змінними необхідно виявити його форму функціональної залежності (лінійна чи нелінійна), а саме:

$$y = \beta_0 + \beta_i x\,; \tag{482}$$

$$y = \beta_0 + \beta_1 x + \beta_2 x^2\,; \tag{483}$$

$$y = \beta_0 + \frac{\beta_1}{x}\,. \tag{484}$$

Наведені можливі залежності між змінними X і Y (482), (483), (484) називають *функціями регресії*. Форму зв'язку між змінними X і Y можна встановити, застосовуючи кореляційні поля, які зображені на рисунках 147—149.

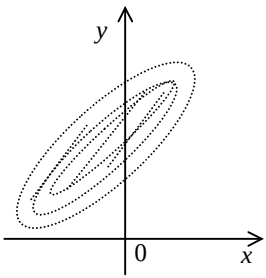


Рис. 147

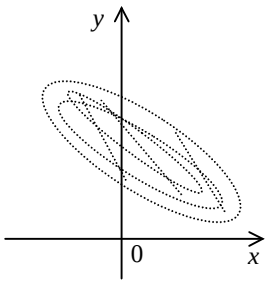


Рис. 148

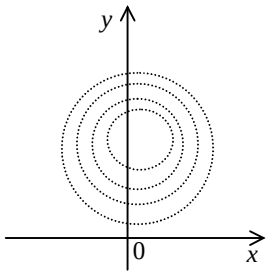


Рис. 149

Тут кожній точці з координатами x_i, y_i відповідає певне числове значення ознак X та Y.

На рис. 147 більшість точок утворюють множину, що має тенденцію при збільшенні значень X зумовлювати збільшення значень ознаки Y.

На рис. 148 множина точок має тенденцію при збільшенні значень X зумовлювати зменшення Y.

На рис. 149 точки рівномірно розміщені на координатній площині $x0y$, що свідчить про відсутність кореляційної залежності між ознаками X і Y.

Отже, на основі розміщення точок кореляційного поля дослідник має підстави для гіпотетичного припущення про лінійні чи нелінійні залежності між ознаками X і Y.

Для двовимірного статистичного розподілу вибірки ознак (X, Y) поняття статистичної залежності між ознаками X та Y має таке визначення:

статистичною залежністю X від Y називають таку, за якої при зміні значень ознаки $Y = y_i$ змінюється умовний статистичний розподіл ознаки X, статистичною залежністю ознаки Y від X називають таку, за якої зі зміною значень ознаки $X = x_i$ змінюється умовний статистичний розподіл ознаки Y.

У разі зміни умовних статистичних розподілів змінюватимуться і умовні числові характеристики.

Звідси випливає визначення кореляційної залежності між ознаками *X* і *Y*. *Кореляційною залежністю* ознаки *X* від *Y* називається функціональна залежність умовного середнього $\bar{y}_{x_j}$ від аргументу *x*, що можна записати так:

$$\bar{y}_x = \alpha(x).$$

Аналогічно кореляційною залежністю ознаки *X* від *Y* називається функціональна залежність умовного середнього $\bar{x}_{y_i}$ від аргументу *y*, що можна записати, так:

$$\bar{x}_y = \beta(y).$$

Між ознаками *X* та *Y* може існувати статистична залежність і за відсутності кореляційної. Але коли існує кореляційна залежність між ознаками *X* та *Y*, то обов'язково між ними існуватиме і статистична залежність.

2. Рівняння лінійної парної регресії

Нехай між змінними *X* та *Y* теоретично існує певна лінійна залежність. Це твердження може ґрунтуватися на тій підставі, наприклад, що кореляційне поле для пар $(x_i; y_i)$ має такий вигляд (рис. 150).

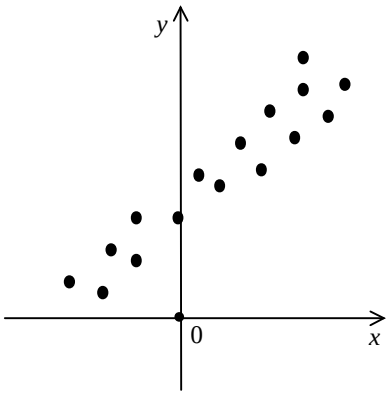


Рис. 150

Як бачимо, насправді між ознаками *X* і *Y* спостерігається не такий тісний зв'язок, як це передбачає функціональна залежність.

Окремі спостережувані значення *y*, як правило, відхилятимуться від передбаченої лінійної залежності під впливом випадкових збудників, які здебільшого є невідомими. Відхилення від передбаченої лінійної форми зв'язку можуть статися внаслідок неправильної специфікації рівняння, тобто ще з самого початку неправильно вибране рівняння, що описує залежність між *X* і *Y*.

Будемо вважати, що специфікація рівняння вибрана правильно. Ураховуючи вплив на значен-

ня Y збурювальних випадкових факторів, лінійне рівняння зв'язку X і Y можна подати в такому вигляді:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad (485)$$

де β_0, β_1 є невідомі параметри регресії, ε_i є випадковою змінною, що характеризує відхилення y від гіпотетичної теоретичної регресії.

Отже, в рівнянні (485) значення «y» подається у вигляді суми двох частин: систематичної $\beta_0 + \beta_1 x_i$ і випадкової ε_i. Параметри β_0, β_1 є невідомими величинами, а ε_i є випадковою величиною, що має нормальний закон розподілу з числовими характеристиками: $M(\varepsilon_i) = 0$, $D(\varepsilon_i) = \sigma^2_{\varepsilon_i} = \text{const}$. При цьому елементи послідовності $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_i$ є некорельованими $(K_{ij} = 0)$.

У результаті статистичних спостережень дослідник дістає характеристики для незалежної змінної x і відповідні значення залежної змінної y.

Отже, необхідно визначити параметри β_0, β_1. Але істинні значення цих параметрів дістати неможливо, оскільки ми користуємося інформацією, здобутою від вибірки обмеженого обсягу. Тому знайдені значення параметрів будуть лише статистичними оцінками істинних (невідомих нам) параметрів β_0, β_1. Якщо позначити параметри β_0^*, β_1^*, які дістали способом обробки вибірки, моделі

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad (486)$$

відповідатиме статистична оцінка

$$y_i = \beta_0^* + \beta_1^* x_i + \varepsilon_i. \qquad (487)$$

2.1. Визначення параметрів β_0^*, β_1^*. Якщо ми прийняли гіпотезу про лінійну форму зв'язку між ознаками X і Y, то однозначно вибрати параметри β_0^*, β_1^*, які є точковими статистичними оцінками відповідно для параметрів β_0, β_1, практично неможливо. І справді, через кореляційне поле (рис. 150) можна провести безліч прямих. Тому необхідно вибрати такий критерій, за яким можна здійснити вибір параметрів β_0^*, β_1^*.

На практиці найчастіше параметри β_0^*, β_1^* визначаються за методом найменших квадратів, розробка якого належить К. Гауссу і П. Лапласу. Цей метод почали широко застосовувати в економіко-статистичних обчисленнях, відколи була створена теорія регресії.

Відповідно до цього методу рівняння лінійної парної регресії $y_i = \beta_0^* + \beta_1^* x_i$ необхідно вибрати так, щоб сума квадратів відхилень спостережуваних значень від лінії регресії була б мінімальною.

Для цього розглянемо графік (рис. 151):

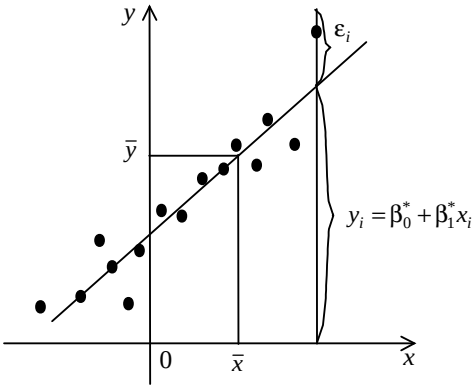


Рис. 151

Через кореляційне поле проведена лінія регресії $y_i = \beta_0^* + \beta_1^* x_i$. Відхилення будь-якої точки із координатами x_i, y_i становить величину ε_i:

$$\varepsilon_i = y_i - \left(\beta_0^* + \beta_1^* x_i\right). \quad (488)$$

Тут: y_i — спостережуване значення ознаки Y, яке дістали внаслідок реалізації вибірки; $\beta_0^* + \beta_1^* x_i$ — значення ознаки Y, обчислене за умови, що $X = x_i$.

Як бачимо, величина ε_i є функцією від параметрів β_0^*, β_1^*. Функція від цих параметрів і буде узагальнюючим показником розсіювання точок навколо прямої, а саме:

$$\sum (\varepsilon_i)^2 . \quad (489)$$

Звідси є сенс узяти критерій, згідно з яким параметри β_0^*, β_1^* необхідно добирати так, щоб сума квадратів відхилень ε_i^2 була мінімальною:

$$\sum (\varepsilon_i)^2 = \min. \quad (490)$$

Позначивши $\sum (\varepsilon_i)^2 = \sum \left(y_i - \left(\beta_0^* + \beta_1^* x_i\right)\right)^2 = \theta\left(\beta_0^*; \beta_1^*\right)$, розглянемо необхідну умову існування мінімуму функції $\theta\left(\beta_0^*; \beta_1^*\right)$:

$$\begin{cases} \dfrac{\partial \theta\left(\beta_0^*; \beta_1^*\right)}{\partial \beta_0^*} = 0 \\ \dfrac{\partial \theta\left(\beta_0^*; \beta_1^*\right)}{\partial \beta_1^*} = 0. \end{cases} \quad (491)$$

Дістанемо лінійне рівняння відносно параметрів β_0^*, β_1^*:

$$\begin{cases} \dfrac{\partial\theta(\beta_0^*;\ \beta_1^*)}{\partial\beta_0^*} = -2\sum\left(y_i - \beta_0^* - \beta_1^* x_i\right) = 0 \\ \dfrac{\partial\theta(\beta_0^*;\ \beta_1^*)}{\partial\beta_1^*} = -2\sum\left(y_i - \beta_0^* - \beta_1^* x_i\right)\ x_i = 0 \end{cases} \rightarrow$$

$$\rightarrow \begin{cases} n\beta_0^* + (x_i)\beta_1^* = \sum y_i \\ (\sum x_i)\beta_0^* + \left(\sum x_i^2\right)\beta_1^* = \sum x_i y_i \end{cases} \rightarrow$$

$$\rightarrow \begin{matrix} \beta_0^* + \dfrac{\sum x_i}{n}\beta_1^* = \dfrac{\sum y_i}{n} \\ \dfrac{\sum x_i}{n}\beta_0^* + \dfrac{\sum x_i^2}{n}\beta_1^* = \dfrac{\sum x_i y_i}{n} \end{matrix} \rightarrow$$

$$\rightarrow \left| \begin{matrix} \text{оскільки } \ \bar{x} = \dfrac{\sum x_i}{n}, \quad \bar{y} = \dfrac{\sum y_i}{n}, \\ \dfrac{\sum x_1^2}{n} - (\bar{x})^2 = \sigma_x^2, \quad K_{xy}^* = \dfrac{\sum x_i y_i}{n} - \overline{xy} \end{matrix} \right| \rightarrow$$

$$\rightarrow \begin{cases} \beta_0^* + \bar{x}\cdot\beta_1^* = \bar{y}, \\ \bar{x}\cdot\beta_0^* + \dfrac{\sum x_i^2}{n}\beta_1^* = \dfrac{\sum x_i y_i}{n}. \end{cases} \tag{492}$$

Розв’язавши систему (492) відносно параметрів β_0^*, β_1^*, знайдемо:

$$\beta_0^* = \bar{y} - \beta_1^*\bar{x}\,; \tag{493}$$

$$\beta_1^* = \frac{\dfrac{\sum x_i y_i}{n} - \overline{xy}}{\dfrac{\sum x_i^2}{n} - (\bar{x})^2} = \frac{K_{xy}}{\sigma_x^2}\,. \tag{494}$$

Помноживши ліву і праву частини (494) на $\dfrac{\sigma_x}{\sigma_y}$, дістанемо:

$$\frac{\sigma_x}{\sigma_y}\beta_1^* = \frac{K_{xy}}{\sigma_x^2}\,\frac{\sigma_x}{\sigma_y} = \frac{K_{xy}}{\sigma_x\ \sigma_y} = r_{xy} \rightarrow \beta_1^* = r_{xy}\,\frac{\sigma_x}{\sigma_y}\,, \tag{495}$$

де r_{xy} —парний коефіцієнт кореляції між ознаками X і Y. Тоді

$$\beta_0^* = \bar{y} - \beta_1^*\bar{x} = \bar{y} - r_{xy}\,\frac{\sigma_x}{\sigma_y}\,\bar{x}\,. \tag{496}$$

З урахуванням (495), (496) рівняння лінійної парної регресії набере такого вигляду:

$$y_i = r_{xy}\frac{\sigma_y}{\sigma_x}\left(x-\bar{x}\right)+\bar{y} \qquad (497)$$

або

$$y_i = \rho_{yx}\left(x-\bar{x}\right)+\bar{y}\,, \qquad (498)$$

де $\rho_{yx} = r_{xy}\dfrac{\sigma_y}{\sigma_x}$ і називають *коефіцієнтом регресії*.

Приклад. Залежність розчинності y_i тіосульфату від температури x_i наведено парним статистичним розподілом вибірки:

$Y = y_i$	33,5	37,0	41,2	46,1	50,0	52,9	56,8	64,3	69,9
$X = x_i$	0	10	20	30	40	50	60	70	80

Потрібно:
1) побудувати кореляційне поле залежності ознаки Y від X;
2) визначити точкові незміщені статистичні оцінки β_0^*, β_1^*. Обчислити r_{xy} ;
3) побудувати графік лінії регресії.

Розв'язання. 1) кореляційне поле залежності ознаки Y від X має такий вигляд (рис. 152).

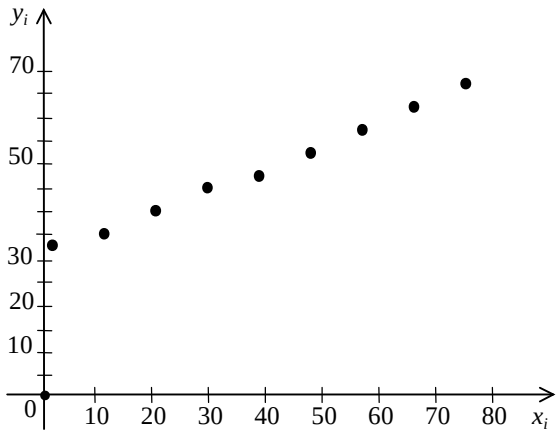


Рис. 152

З рис. 152 бачимо, що зі збільшенням значень ознаки $X = x_i$ залежна зміна $Y = y_i$ має тенденцію до збільшення.

Тому припускаємо, що між ознаками X та Y існує лінійна функціональна залежність

$$y_i = \beta_0^* + \beta_1^* x_i,$$

2) для визначення параметрів β_0^*, β_1^* скористаємося таблицею, що має такий вигляд:

№ з/п	x_i	y_i	x_i^2	$x_i y_i$	y_i^2
1	0	33,5	0	0	1122,25
2	10	37,0	100	307	1369,00
3	20	41,2	400	824	1697,44
4	30	46,1	900	1383	2125,21
5	40	50,0	1000	2000	2500,00
6	50	52,9	2500	2645	2798,41
7	60	56,8	3600	3408	3226,24
8	70	64,3	4900	4501	4134,49
9	80	69,9	6400	5592	4886,01
Σ	360	451,7	20400	20723	23859,05

Скориставшись формулами (494), (496), дістанемо

$$\beta_1^* = \frac{\dfrac{\sum x_i y_i}{n} - \bar{x}\ \bar{y}}{\dfrac{\sum x_i^2}{n} - (\bar{x})^2}, \qquad \beta_0^* = \bar{y} - \beta_1^* \bar{x}.$$

Оскільки $n = 9$, $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{360}{9} = 40$; $\quad \bar{y} = \dfrac{\sum y_i}{n} = \dfrac{451{,}7}{9} = 50{,}19$;

$$\frac{\sum y_i^2}{n} = \frac{23859{,}05}{9} = 2651; \qquad \frac{\sum x_i^2}{n} = \frac{20400}{9} = 2266{,}7;$$

$$\frac{\sum x_i y_i}{n} = \frac{20723}{9} = 2302{,}6;$$

$\bar{x}\,\bar{y} = 40 \cdot 50{,}19 = 2007{,}6$; $(\bar{x})^2 = 1600$, одержимо:

$$\beta_1^* = \frac{2302{,}6 - 2007{,}6}{2266{,}7 - 1600} = \frac{295}{666{,}7} = 0{,}44;$$

$$\beta_1^* = 0{,}44.$$

$$\beta_0^* = 50{,}19 - 0{,}44 \cdot 40 = 50{,}19 - 17{,}6 = 32{,}59.$$

$$\beta_0^* = 32{,}59.$$

Отже, рівняння регресії буде таким:

$$y_i = 32{,}59 + 0{,}44 \cdot x_i.$$

Для обчислення r_{xy} необхідно знайти K_{xy}, σ_x, σ_y.

$$K_{xy}^* = \frac{\sum x_i y_i}{n} - \overline{xy} = 2302{,}6 - 2007{,}6 = 295;$$

$$\sigma_x = \sqrt{\frac{\sum x_1^2}{n} - (\overline{x})^2} = \sqrt{2266{,}7 - 40^2} = \sqrt{666{,}7} = 25{,}8\,;$$

$$\sigma_y = \sqrt{\frac{\sum y_i^2}{n} - (\overline{y})^2} = \sqrt{2651 - (50{,}19)^2} = \sqrt{131{,}96} = 11{,}49\,;$$

$$r_{xy} = \frac{K_{xy}^*}{\sigma_x \sigma_y} = \frac{295}{25{,}8 \cdot 11{,}49} = \frac{295}{296{,}44} = 0{,}995.$$

Як бачимо, коефіцієнт кореляції близький за своїм значенням до одиниці, що свідчить про те, що залежність між X та Y є практично лінійною.

Графік парної лінійної функції регресії подано на рис. 153.

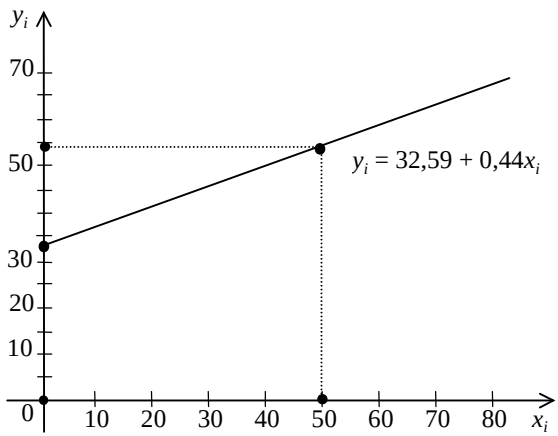


Рис. 153

Якщо параметри β_0, β_1 рівняння (486) — сталі величини, то β_0^*, β_1^*, які обчислені шляхом обробки реалізованої вибірки, є випадковими величинами і виконують функцію точкових статистичних оцінок для них.

2.2. Властивості β_0^*, β_1^*. Точкові статистичні оцінки β_0^*, β_1^* можна подати в такому вигляді:

$$\beta_1^* = \frac{\sum x_i y_i - n\overline{x}\overline{y}}{\sum (x_i - \overline{x})^2} = \frac{\sum x_i y_i - \sum y_i \overline{x}}{\sum (x_i - \overline{x})^2} = \frac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2} = \frac{\sum (x_i - \overline{x})(\beta_0 + \beta_1 x_i + \varepsilon_i)}{\sum (x_i - \overline{x})^2} =$$

$$= \beta_0 \frac{\sum (x_i - \overline{x})}{\sum (x_i - \overline{x})^2} + \beta_1 \frac{\sum (x_i - \overline{x}) x_i}{\sum (x_i - \overline{x})^2} + \frac{\sum (x_i - \overline{x}) \varepsilon_i}{\sum (x_i - \overline{x})^2} =$$

$$= \beta_0 \frac{\sum (x_i - \overline{x})}{\sum (x_i - \overline{x})^2} + \beta_1 \frac{\sum \left(x_i^2 - \overline{x} x_i\right)}{\sum (x_i - \overline{x})^2} + \frac{\sum (x_i - \overline{x}) \varepsilon_i}{\sum (x_i - \overline{x})^2} =$$

$$= \beta_1 \frac{\sum (x_i - \overline{x})^2}{\sum (x_i - \overline{x})^2} + \frac{\sum (x_i - \overline{x}) \varepsilon_i}{\sum (x_i - \overline{x})^2} = \beta_1 + \frac{\sum (x_i - \overline{x}) \varepsilon_i}{\sum (x_i - \overline{x})^2}.$$

Оскільки $\sum \left(x_i - \overline{x}\right) = 0$,

$$\sum \left(x_i - \overline{x}\right) x_i = \sum \left(x_i^2 - \overline{x} x_i\right) = \sum x_i^2 - \overline{x} \sum x_i = \sum x_i^2 - n\left(\overline{x}\right)^2 = \sum \left(x_i - \overline{x}\right)^2.$$

Отже, дістали:

$$\beta_1^* = \beta_1 + \frac{\sum (x_i - \overline{x})\ \varepsilon_i}{\sum (x_i - \overline{x})^2}. \qquad (499)$$

$$\beta_0^* = \overline{y} - \overline{x}\beta_1^* = \frac{\sum y_i}{n} - \overline{x}\left(\beta_1 + \frac{\sum (x_i - \overline{x})\ \varepsilon_i}{\sum (x_i - \overline{x})^2}\right) =$$

$$= \frac{\sum (\beta_0 + \beta_1 x_i + \varepsilon_i)}{n} - \overline{x}\beta_1 - \frac{\sum (x_i - \overline{x})\ \varepsilon_i}{\sum (x_i - \overline{x})^2}\overline{x} =$$

$$= \beta_0 + \frac{\sum x_i}{n}\beta_1 + \frac{\sum \varepsilon_i}{n} - \overline{x}\beta_1 - \frac{\sum (x_i - \overline{x})\ \varepsilon_i}{\sum (x_i - \overline{x})^2}\overline{x} =$$

$$= \beta_0 + \overline{x}\beta_1 + \frac{\sum \varepsilon_i}{n} - \overline{x}\beta_1 - \frac{\sum (x_i - \overline{x}) \varepsilon_i}{\sum (x_i - \overline{x})^2}\overline{x} = \beta_0 + \frac{\sum \varepsilon_i}{n} - \frac{\sum (x_i - \overline{x})\ \varepsilon_i}{\sum (x_i - \overline{x})^2}\overline{x}.$$

Остаточно маємо:

$$\beta_0^* = \beta_0 + \frac{\sum \varepsilon_i}{n} - \frac{\sum (x_i - \overline{x})\ \varepsilon_i}{\sum (x_i - \overline{x})^2}\overline{x}. \qquad (500)$$

Знаходимо числові характеристики для випадкових величин β_0^*, β_1^*:

а) основні числові характеристики для β_0^*

$$M\left(\beta_0^*\right)=M\left(\beta_0+\frac{\sum\varepsilon_i}{n}-\frac{\sum(x_i-\bar{x})\,\varepsilon_i}{\sum(x_i-\bar{x})^2}\bar{x}\right)=$$

$$=M\left(\beta_0\right)+M\left(\frac{\sum\varepsilon_i}{n}\right)-M\left(\frac{\sum(x_i-\bar{x})\,\varepsilon_i}{\sum(x_i-\bar{x})^2}\bar{x}\right)=$$

$$=\beta_0+\frac{\sum M(\varepsilon_i)}{n}-\frac{\sum(x_i-\bar{x})M(\varepsilon_i)}{\sum(x_i-\bar{x})^2}=\beta_0.\quad (M(\varepsilon_i)=0).$$

Отже, ми довели, що β_0^* є точковою незміщеною статистичною оцінкою для параметра β_0,

$$D\left(\beta_0^*\right)=D\left(\beta_0+\frac{\sum\varepsilon_i}{n}-\frac{\sum(x_i-\bar{x})\,\varepsilon_i}{\sum(x_i-\bar{x})^2}\bar{x}\right)=$$

$$=D\left(\beta_0\right)+D\left(\frac{\sum\varepsilon_i}{n}\right)-D\left(-\frac{\sum(x_i-\bar{x})\,\varepsilon_i}{\sum(x_i-\bar{x})^2}\bar{x}\right)=$$

$$=\frac{\sum D(\varepsilon_i)}{n^2}+(\bar{x})^2\,\frac{\sum(x_i-\bar{x})^2D(\varepsilon_i)}{\left[\sum(x_i-\bar{x})^2\right]^2}=$$

$$=\frac{\sum D(\varepsilon_i)}{n^2}+(\bar{x})^2\,\frac{\sum(x_i-\bar{x})^2D(\varepsilon_i)}{\left[\sum(x_i-\bar{x})^2\right]^2}=$$

$$=\frac{\sigma_\varepsilon^2}{n}+(\bar{x})^2\,\frac{\sum(x_i-\bar{x})^2}{\left(\sum(x_i-\bar{x})^2\right)^2}\sigma_\varepsilon^2=$$

$$=\frac{\sigma_\varepsilon^2}{n}+\frac{(\bar{x})^2\sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2}=\frac{\sigma_\varepsilon^2}{n}\left(1+\frac{(\bar{x})^2n}{\sum(x_i-\bar{x})^2}\right)=$$

$$=\frac{\sigma_\varepsilon^2}{n}\cdot\frac{\sum(x_i-\bar{x})^2+(\bar{x})^2n}{\sum(x_i-\bar{x})^2}=\frac{\sigma_\varepsilon^2}{n}\,\frac{\sum x_i^2-n(\bar{x})^2+n(\bar{x})^2}{\sum(x_i-\bar{x})^2}=\frac{\sigma_\varepsilon^2}{n}\,\frac{\sum x_i^2}{\sum(x_i-\bar{x})^2}.$$

Далі маємо:

$$D\left(\beta_0^*\right)=\frac{\sum x_i^2}{n\sum\left(x_i-\bar{x}\right)^2}\sigma_\varepsilon^2,\qquad (501)$$

$$\sigma(\beta_0^*) = \sqrt{\frac{\sum x_i^2}{n\sum (x_i - \bar{x})^2}}\ \sigma_\varepsilon . \qquad (502)$$

б) Основні числові характеристики для параметрів β_i^*

$$M\left(\beta_1 + \frac{\sum (x_i - \bar{x})\varepsilon_i}{\sum (x_i - \bar{x})^2}\right) = \beta_1 + \frac{\sum (x_i - \bar{x}) M(\varepsilon_i)}{\sum (x_i - \bar{x})^2} = \beta_1. \ (M(\varepsilon_i) = 0).$$

Отже, визначили, що β_1^* є точковою незміщеною систематичною оцінкою для параметра β_1

$$M(\beta_1^*) = \beta_1 . \qquad (503)$$

$$D(\beta_1^*) = D\left(\beta_1 + \frac{\sum (x_i - \bar{x})\ \varepsilon_i}{\sum (x_i - \bar{x})^2}\right) = D(\beta_1) + D\left(\frac{\sum (x_i - \bar{x})\ \varepsilon_i}{\sum (x_i - \bar{x})^2}\right) =$$

$$= D(\varepsilon_i)\frac{\sum (x_i - \bar{x})^2}{\left[\sum (x_i - \bar{x})^2\right]^2} = \frac{D(\varepsilon_i)}{\sum (x_i - \bar{x})^2} = \frac{\sigma_\varepsilon^2}{\sum (x_i - \bar{x})^2}.$$

Звідси маємо:

$$D(\beta_1^*) = \frac{\sigma_\varepsilon^2}{\sum (x_i - \bar{x})^2}, \qquad (504)$$

$$\sigma(\beta_1^*) = \frac{\sigma_\varepsilon}{\sqrt{\sum (x_i - \bar{x})^2}}. \qquad (505)$$

Статистичні оцінки β_0^*, β_1^* як випадкові величини впливають на зміщення лінії регресії; так, β_0^* викликає вертикальне зміщення лінії регресії, а β_1^* — зміну кута нахилу її.

З'ясуємо, чи існує кореляційний зв'язок між випадковими величинами β_0^*, β_1^*.

$$K_{\beta_0^*\ \beta_1^*} = M(\beta_0^* - \beta_0)(\beta_1^* - \beta_1) =$$

$$= M\left(\left(\frac{\sum \varepsilon_i}{n} - \frac{\sum (x_i - \bar{x})\ \varepsilon_i}{\sum (x_i - \bar{x})^2}\bar{x}\right) \cdot \frac{\sum (x_i - \bar{x})\ \varepsilon_i}{\sum (x_i - \bar{x})^2}\right) =$$

$$= M\left(\frac{\sum (x_i - \bar{x})\ \varepsilon_i \cdot \sum \varepsilon_i}{n\sum (x_i - \bar{x})^2} - \frac{\sum (x_i - \bar{x})\ \varepsilon_i \cdot \sum (x_i - \bar{x})\ \varepsilon_i}{\left(\sum (x_i - \bar{x})^2\right)^2} \cdot \bar{x}\right) =$$

$$= M\left(\frac{[(x_1-\bar{x})\ \varepsilon_1+(x_2-\bar{x})\ \varepsilon_2+\ldots+(x_n-\bar{x})\ \varepsilon_n]\ [\varepsilon_1+\varepsilon_2+\ldots+\varepsilon_n]}{n\sum(x_i-\bar{x})^2}\right)-$$

$$-M\frac{[(x_1-\bar{x})\varepsilon_1+(x_2-\bar{x})\varepsilon_2+\ldots+(x_n-\bar{x})\varepsilon_n][(x_1-\bar{x})\varepsilon_1+(x_2-\bar{x})\varepsilon_2+\ldots+(x_n-\bar{x})\varepsilon_n]}{\left(\sum(x_i-\bar{x})^2\right)^2}\bar{x}=$$

$$=\left|\begin{array}{c}\text{оскільки } M(\varepsilon_i\varepsilon_j)=0,\ M(\varepsilon_i)=0,\text{ бо}\\ \text{між } \varepsilon_i \text{ і } \varepsilon_j \text{ відсутня кореляція і математичні}\\ \text{сподівання дорівнюють нулеві, тоді при } i=j\\ \text{маємо } M(\varepsilon_i^2)=\sigma_\varepsilon^2=\text{const}\end{array}\right|=$$

$$=\frac{\sum(x_i-\bar{x})}{n\sum(x_i-\bar{x})^2}\sigma_\varepsilon^2-\frac{\sum(x_i-\bar{x})\sigma_\varepsilon^2}{\left(\sum(x_i-\bar{x})^2\right)^2}\bar{x}=-\frac{\bar{x}\cdot\sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2};$$

$$K_{\beta_0^*\beta_1^*}=-\frac{\bar{x}\cdot\sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2}. \qquad (506)$$

Відповідно до (499), (500) β_0^*, β_1^* містять як складову випадковий компонент ε_i з нормальним законом розподілу ймовірностей. Звідси випливає:

β_0^* матиме нормальний закон розподілу із числовими характеристиками:

$$a=\beta_0,\ \ \sigma=\sqrt{\frac{\sum_1^n x_1^2}{n\sum_{i=1}^n(x_i-\bar{x})^2}}\sigma_\varepsilon,\ \text{тобто}\ N\left(\beta_0;\sqrt{\frac{\sum_1^n x_1^2}{n\sum_{i=1}^n(x_i-\bar{x})^2}}\sigma_\varepsilon\right);$$

β_1^* також буде мати нормальний закон розподілу з числовими характеристиками

$$a=\beta_1,\ \ \sigma=\frac{\sigma_2}{\sqrt{\sum_{i=1}^n(x_i-\bar{x})^2}},\ \text{тобто}\ N\left(\beta_1;\frac{\sigma_2}{\sqrt{\sum_{i=1}^n(x_i-\bar{x})^2}}\right).$$

Тоді

$$\frac{\beta_0^*-\beta_0}{\sigma_\varepsilon\sqrt{\frac{\sum x_1^2}{n\sum(x_i-\bar{x})^2}}}\ \text{ і }\ \frac{\beta_1^*-\beta_1}{\frac{\sigma_\varepsilon}{\sqrt{\sum(x_i-\bar{x})^2}}}\ \text{ матимуть закон розподілу } N(0;\,1).$$

Рівняння регресії можна подати в такому вигляді:

$$y_i - \bar{y} = \beta_1^*\left(x_i - \bar{x}\right) = \varepsilon_i^*. \qquad (507)$$

Розглянемо модель (486)

$$y_i = \beta_0 + \beta_i x_i + \varepsilon_i, \qquad (508)$$

яку подамо через середні величини:

$$\sum y_i = \sum(\beta_0 + \beta_i x_i + \varepsilon_i) = n\beta_0 + (\sum x_i)\beta_i + \sum \varepsilon_i \to$$

$$\to \frac{\sum y_i}{n} = \beta_0 + \frac{\sum x_i}{n}\beta_1 + \frac{\sum \varepsilon_i}{n} \to$$

$$\to \bar{y} = \beta_0 + \bar{x}\beta_1 + \bar{\varepsilon} \ \ (\text{тут } \bar{\varepsilon} = \frac{\sum \varepsilon_i}{n}).$$

Отже, дістали

$$\begin{cases} y_i = \beta_0 + \beta_i x_i + \varepsilon_i, \\ \bar{y} = \beta_0 + \beta_i \bar{x} + \bar{\varepsilon}, \end{cases} \to \qquad (509)$$

$$y_i - \bar{y} = \beta_1\left(x_i - \bar{x}\right) + \left(\varepsilon_i - \bar{\varepsilon}\right) = \varepsilon_i^*. \qquad (510)$$

Ураховуючи (507) і (510), ε_i^* можна подати так:

$$\varepsilon_i^* = \beta_1\left(x_i - \bar{x}\right) + \left(\varepsilon_i - \bar{\varepsilon}\right) - \beta_1^*\left(x_i - \bar{x}\right) \to$$

$$\to \varepsilon_i^* = -\left(\beta_1^* - \beta_1\right)(x_i - \bar{x}) + (\varepsilon_i - \bar{\varepsilon}). \qquad (511)$$

А тому $\sum\left(\varepsilon_i^*\right)^2 = \sum\left[(\varepsilon_i - \bar{\varepsilon}) - \left(\beta_1^* - \beta_1\right)(x_i - \bar{x})\right]^2.$

Тоді $M\left(\sum\left(\varepsilon_i^*\right)^2\right) =$

$$= M\left(\sum(\varepsilon_i - \varepsilon)^2 + \left(\beta_1^* - \beta_1\right)^2 \sum(x_i - \bar{x})^2 - 2\left(\beta_1^* - \beta_1\right)\sum(x_i - \bar{x})(\varepsilon_i - \bar{\varepsilon})\right) =$$

$$= M\left(\sum(\varepsilon_i - \bar{\varepsilon})^2\right) + \sum(x_i - \bar{x})^2 M\left(\beta_1^* - \beta_1\right)^2 - M\left[\left(\beta_1^* - \beta_1\right)\sum(x_i - \bar{x})(\varepsilon_i - \bar{\varepsilon})\right] =$$

$$= \left| \begin{array}{c} M\sum(\varepsilon_i - \bar{\varepsilon})^2 = M\sum\left(\varepsilon_i^2 - \left(\frac{\sum \varepsilon_i}{n}\right)^2\right) = \\ = \sum\left(M\left(\varepsilon_i^2\right)\right) - \frac{1}{n^2} M(\sum \varepsilon_i)^2 = n\sigma_\varepsilon^2 - \frac{n^2\sigma_\varepsilon^2}{n^2} = (n-1)\sigma_\varepsilon^2, \\ \text{оскільки } M\left(\varepsilon_i^2\right) = \sigma_\varepsilon^2, M\left(\varepsilon_i \varepsilon_j\right) = 0; \\ M\left(\beta_1^* - \beta_1\right)^2 = \frac{\sigma_\varepsilon^2}{\sum(x_i - \bar{x})^2} \to \sum(x_i - \bar{x})^2 \frac{\sigma_\varepsilon^2}{\sum(x_i - \bar{x})^2} = \sigma_\varepsilon^2; \\ \text{враховуючи, що } \beta_1^* - \beta_1 = \frac{\sum(x_i - \bar{x})\varepsilon_i}{\sum(x_i - \bar{x})^2}, \text{ одержимо} \end{array} \right| =$$

$$= \left| \begin{array}{c} M\left(\left(\beta_1^* - \beta_1\right)\sum(x_i - \bar{x})(\varepsilon_i - \bar{\varepsilon})\right) = \\ M\left(\dfrac{\sum(x_i - \bar{x})\varepsilon_i}{\sum(x_i - \bar{x})^2} \cdot [\sum(x_i - \bar{x})\varepsilon_i - \sum(x_i - \bar{x})\bar{\varepsilon}]\right) = \\ = M\,\dfrac{[\sum(x_i - \bar{x})\varepsilon_i]^2}{\sum(x_i - \bar{x})^2} - \sum(x_i - \bar{x})M(\bar{\varepsilon}) = \\ = \dfrac{\sum(x_i - \bar{x})^2 M\left(\varepsilon_i^2\right)}{\sum(x_i - \bar{x})^2} = \sigma_\varepsilon^2. \quad \text{Тут } \sum(x_i - \bar{x}) = 0 \end{array} \right| .$$

Отже, дістали

$$M\left(\sum\left(\varepsilon_i^*\right)^2\right) = (n-1)\sigma_\varepsilon^2 + \sigma_\varepsilon^2 - 2\sigma_\varepsilon^2 = (n-2)\sigma_\varepsilon^2.$$

Звідси маємо

$$\sigma_\varepsilon^2 = \frac{M\left(\sum\left(\varepsilon_i^*\right)^2\right)}{n-2}. \tag{512}$$

$$\frac{\sum\left(\varepsilon_i^*\right)^2}{n-2} = S_\varepsilon^2 \tag{513}$$

є точковою незміщеною статистичною для σ_ε^2.

Далі, враховуючи (513), дістанемо:

$$D\left(\beta_0^*\right) = \frac{\sum_i^n x_i^2}{n\sum_i^n \left(x_i - \bar{x}\right)^2} S_\varepsilon^2, \tag{514}$$

$$\sigma\left(\beta_0^*\right) = \sqrt{\frac{\sum_i^n x_i^2}{n\sum_i^n \left(x_i - \bar{x}\right)^2}} S_\varepsilon, \tag{515}$$

$$D\left(\beta_1^*\right) = \frac{S_\varepsilon^2}{\sum_i^n \left(x_i - \bar{x}\right)^2}, \tag{516}$$

$$\sigma\left(\beta_1^*\right) = \frac{S_\varepsilon^2}{\sqrt{\sum_i^n \left(x_i - \bar{x}\right)^2}}, \tag{517}$$

$$K_{\beta_0^*\beta_1^*} = -\frac{\bar{x}S_\varepsilon^2}{\sum_i^n \left(x_i - \bar{x}\right)^2}. \qquad (518)$$

З наведених вище перетворень можна зробити висновок, що: випадкова величина

$$\frac{\left(n-2\right) S_\varepsilon^2}{\sigma_\varepsilon^2} = \chi^2 \qquad (519)$$

матиме розподіл χ^2 із $k = n - 2$ ступенями свободи;

випадкові величини:

$$t = \frac{\beta_0^* - \beta_0}{\sqrt{\frac{\sum_i^n x_i^2}{n\sum_i^n \left(x_i - \bar{x}\right)^2}} S_\varepsilon}; \qquad (520)$$

$$t = \frac{\beta_1^* - \beta_0}{\frac{S_\varepsilon}{\sqrt{\sum_i^n \left(x_i - \bar{x}\right)^2}}}$$

матимуть розподіл Стьюдента (*t*-розподіл) із $k = n - 2$ ступенями свободи.

Ураховуючи (519), (520) ми дійшли висновку, що $\beta_0^* + \beta_1^* x_2^*$ буде мати двовимірний нормальний закон на площині, а саме:

$$N\left(\beta_0; \sqrt{\frac{\sum x_1^2}{n\sum \left(x_i - \bar{x}\right)^2}}\sigma_\varepsilon; \beta_1; \frac{\sigma_\varepsilon}{\sqrt{\sum \left(x_i - \bar{x}\right)^2}}; r_{\beta_0^*\beta_1^*} = \frac{K_{\beta_0^*\beta_1^*}}{\sigma_{\beta_0^*}\sigma_{\beta_1^*}}\right).$$

Скориставшись (519), (520) для параметрів β_0^*, β_1^* парної лінійної функції регресії, ми зможемо побудувати довірчі інтервали із заданою надійністю γ, а також перевірити значущість коефіцієнтів лінійної регресії.

2.3. Перевірка значущості коефіцієнтів лінійної регресії. Ґрунтуючись на значеннях дисперсій β_0^*, β_1^*, можна перевірити значущість цих коефіцієнтів на заданому рівні $\alpha = 0{,}05$, викорис-

товуючи при цьому відому схему перевірки правильності нульової гіпотези H_0. У парній лінійній регресії, як правило, на значущість перевіряється коефіцієнт β_1.

Нульова гіпотеза має вигляд H_0: $\beta_1 = 0$. За статистичний критерій беремо випадкову величину

$$t = \frac{\beta_1^* - \beta_0}{\dfrac{S_\varepsilon}{\sqrt{\sum\limits_i (x_i - \bar{x})^2}}} = \frac{\beta_1^*}{\dfrac{S_\varepsilon}{\sqrt{\sum (x_i - \bar{x})^2}}},$$

що має t-розподіл із $k = n - 2$ ступенями свободи (розподіл Стьюдента).

При альтернативній гіпотезі $H\alpha$: $\beta_1 < 0$ — лівобічна критична область і при $H\alpha$: $\beta_1 \neq 0$ — двобічна критична область.

Спостережуване значення критерію обчислюється як

$$t^* = \frac{\beta_1^*}{\sqrt{\sum (x_i - \bar{x})^2}}.$$

2.4. Довірчі інтервали для β_0^*, β_1^*. Побудова довірчого інтервалу для параметра β_1^* із заданою надійністю γ здійснюється використовуючи (520). Отже, маємо:

$$P\left(\left|\frac{\dfrac{\beta_1^* - \beta_1}{S_\varepsilon}}{\sqrt{\sum (x_i - \bar{x})^2}}\right| < t_\gamma(\gamma, k)\right) = \gamma,$$

$$P\left(t_\gamma(\gamma, k) < \left|\frac{\dfrac{\beta_1^* - \beta_1}{S_\varepsilon}}{\sqrt{\sum (x_i - \bar{x})^2}}\right| < t_\gamma(\gamma, k)\right) = \gamma \to$$

$$\to P\left(\beta_1 - \frac{t(\gamma, k) S_\varepsilon}{\sqrt{\sum (x_i - \bar{x})^2}} < \beta_1 < \beta_1^* + \frac{t(\gamma, k) S_\varepsilon}{\sqrt{\sum (x_i - \bar{x})^2}}\right) = \gamma.$$

Отже, довірчий інтервал для параметра β_1 буде

$$\beta_1^* - \frac{t(\gamma,k)\,S_\varepsilon}{\sqrt{\sum\left(x_i-\bar{x}\right)^2}} < \beta_1 < \beta_1^* + \frac{t(\gamma,k)\,S_\varepsilon}{\sqrt{\sum\left(x_i-\bar{x}\right)^2}}, \qquad (521)$$

де $t(\gamma,k)$ знаходимо за таблицею (додаток 3) за заданою надійністю γ і числом ступенів свободи $k = n-2$;

Побудова довірчого інтервалу для параметра β_0^* із заданою надійністю *γ*.

Аналогічно скориставшись (520), маємо

$$P\left(\left|\frac{\beta_0^*-\beta_0}{\sqrt{\frac{\sum x_i^2}{n\sum(x_i-\bar{x})^2}}\,S_\varepsilon}\right| < t(\gamma,k)\right) = \gamma \to$$

$$\to P\left(\beta_0^* - t(\gamma,k)\sqrt{\frac{\sum x_i^2}{n\sum\left(x_i-\bar{x}\right)^2}}S_\varepsilon < \beta_0 < \beta_0^* + t(\gamma,k)\sqrt{\frac{\sum x_i^2}{n\sum\left(x_i-\bar{x}\right)^2}}S_\varepsilon\right) = \gamma.$$

Отже, довірчий інтервал для параметра β_0^* буде таким:

$$\beta_0^* - t(\gamma,k)\sqrt{\frac{\sum x_i^2}{n\sum(x_i-\bar{x})^2}}S_\varepsilon < \beta_0 < \beta_0^* + t(\gamma,k)\sqrt{\frac{\sum x_i^2}{n\sum(x_i-\bar{x})^2}}S_\varepsilon. \qquad (522)$$

2.5. Довірчий інтервал для парної лінійної функції регресії із заданою надійністю γ . Ураховуючи те, що β_0^* і β_1^* є випадковими величинами, то і лінійна функція регресії $\bar{y} + \beta_1^*\left(x_i - \bar{x}\right)$ буде випадковою. Позначимо через y_i^* значення ознаки *Y*, обчислимо за формулою

$$y_i^* = \bar{y} + \beta_1^*\left(x_i - \bar{x}\right). \qquad (523)$$

Тоді

$$D\left(y_i^*\right) = D\left(\bar{y} + \beta_1^*\left(x_i-\bar{x}\right)\right) = D\left(\frac{\sum y_i}{n} + \beta_1^*\left(x_i-\bar{x}\right)\right) =$$

$$= D\left(\frac{\sum y_i}{n}\right) + \left(x_i-\bar{x}\right)^2 D\left(\beta_1^*\right) =$$

$$= D\left(\frac{\beta_0 + \beta_1 x_i + \varepsilon_i}{n}\right) + \left(x_i - \bar{x}\right)^2 D\frac{\sum\left(x_i - \bar{x}\right)\varepsilon_i}{\sum\left(x_i - \bar{x}\right)^2} =$$

$$= \frac{\sum D(\varepsilon_i)}{n^2} + \left(x_i - \bar{x}\right)^2 \frac{\sum\left(x_i - \bar{x}\right)^2 D\left(\varepsilon_i\right)}{\left(\sum\left(x_i - \bar{x}\right)^2\right)^2} = \frac{\sum D\left(\varepsilon_i\right)}{n^2} + \frac{\left(x_i - \bar{x}\right)^2 \sigma_\varepsilon^2}{\sum\left(x_i - \bar{x}\right)^2} =$$

$$= \frac{\sigma_\varepsilon^2}{n} + \frac{\left(x_i - \bar{x}\right)^2 \sigma_\varepsilon^2}{\sum\left(x_i - \bar{x}\right)^2} = \sigma_\varepsilon^2\left(\frac{1}{n} + \frac{\left(x_i - \bar{x}\right)^2}{\sum\left(x_i - \bar{x}\right)^2}\right).$$

Звідси дістали:

$$D\left(y_i^*\right) = \sigma_\varepsilon^2\left(\frac{1}{n} + \frac{\left(x_i - \bar{x}\right)^2}{\sum\left(x_i - \bar{x}\right)^2}\right) \tag{524}$$

або

$$D\left(y_i^*\right) = S_\varepsilon^2\left(\frac{1}{n} + \frac{\left(x_i - \bar{x}\right)^2}{\sum\left(x_i - \bar{x}\right)^2}\right). \tag{525}$$

Випадкова величина

$$t = \frac{y_i^* - y_i}{S_\varepsilon\sqrt{\frac{1}{n} + \frac{\left(x_i - \bar{x}\right)^2}{\sum\left(x_i - \bar{x}\right)^2}}} \tag{526}$$

має *t*-розподіл із $k = n - 2$ ступенями свободи. Ураховуючи (526), можна побудувати довірчий інтервал для лінійної парної функції регресії із заданою надійністю γ, а саме:

$$P\left(\left|\frac{y_i^* - y_i}{S_\varepsilon\sqrt{\frac{1}{n} + \frac{\left(x_i - \bar{x}\right)^2}{\sum\left(x_i - \bar{x}\right)^2}}}\right| < t(\gamma, k)\right) = \gamma. \tag{527}$$

З (527) випливає

$$\beta_0^* + \beta_1^* x_i - t(\gamma, k) S_\varepsilon\sqrt{\frac{1}{n} + \frac{\left(x_i - \bar{x}\right)^2}{\sum\left(x_i - \bar{x}\right)^2}} < \beta_0 + \beta_1 x_i <$$

$$< \beta_0^* + \beta_1^* x_i - t(\gamma, k) S_\varepsilon\sqrt{\frac{1}{n} + \frac{\left(x_i - \bar{x}\right)^2}{\sum\left(x_i - \bar{x}\right)^2}}. \tag{528}$$

2.6. Довірчий інтервал для індивідуальних значень Y = y_i із заданою надійністю γ. Отже, маємо довірчу зону (528), яка визначає розміщення лінії регресії із заданою надійністю γ, але не окремих можливих значень змінної *Y*, які відхиляються від неї.

Тому, щоб визначити довірчі інтервали для окремих прогнозуючих значень ознаки *Y*, необхідно знайти дисперсію для них, а саме:

$$D\left(y_i^* - y_i\right) = D\left(y_i^*\right) + D\left(y_i\right) = D\left(y_i^*\right) + D\left(\beta_0 + \beta_1 x_i + \varepsilon_i\right) =$$

$$= \sigma_\varepsilon^2\left(\frac{1}{n} + \frac{\left(x_i - \bar{x}\right)^2}{\sum\left(x_i - \bar{x}\right)^2}\right) + \sigma_\varepsilon^2 = \sigma_\varepsilon^2\left(1 + \frac{1}{n} + \frac{\left(x_i - \bar{x}\right)^2}{\sum\left(x_i - \bar{x}\right)^2}\right). \qquad (529)$$

Отже, дістали

$$D\left(y_i^* - y_i\right) = \sigma_\varepsilon^2\left(1 + \frac{1}{n} + \frac{\left(x_i - \bar{x}\right)^2}{\sum\left(x_i - \bar{x}\right)^2}\right). \qquad (530)$$

Випадкова величина

$$\frac{y_i^* - y_i}{S_\varepsilon^2\sqrt{1 + \frac{1}{n} + \frac{\left(x_i - \bar{x}\right)^2}{\sum\left(x_i - \bar{x}\right)^2}}} = t \qquad (531)$$

має *t*-розподіл із $k - n - 2$ ступенями свободи. Довірчий інтервал в цьому разі будується згідно з рівністю

$$P\left(\left|\frac{y_i^* - y_i}{S_\varepsilon\sqrt{1 + \frac{1}{n} + \frac{\left(x_i - \bar{x}\right)^2}{\sum\left(x_i - \bar{x}\right)^2}}}\right| < t(\gamma, k)\right) = \gamma. \qquad (532)$$

З (532) дістанемо

$$\beta_0 + \beta_1 x_i + t(\gamma, k)S_p < y_p < \beta_0 + \beta_1 x_i + t(\gamma, k)S_p, \qquad (533)$$

де y_p — прогнозне значення ознаки *Y;*

$S_p = S_\varepsilon\sqrt{1 + \frac{1}{n} + \frac{\left(x_i - \bar{x}\right)^2}{\sum\left(x_i - \bar{x}\right)^2}}$ — прогнозне виправлене середньоквадратичне відхилення.

Приклад. Середня температура взимку вимірювалась протягом 13 років у двох містах *A* і *B*. Результати вимірювання наведено в таблиці:

$A(X = x_i)$	–19,2	–14,8	–19,6	–11,1	–9,4	–16,9	–13,7
$B(Y = y_i)$	–21,8	–15,4	–20,8	–11,3	–11,6	–19,2	–13,0

Продовження табл.

$A(X = x_i)$	–4,9	–13,9	–9,4	–8,3	–7,9	–5,3
$B(Y = y_i)$	–7,4	–15,1	–14,4	–11,1	–10,5	–7,2

Потрібно:

1) ґрунтуючись на гіпотетичному припущенні, що між ознаками X і Y існує лінійна форма зв'язку, визначити β_0^*, β_1^*, r_{xy};

2) визначити $D(\beta_0^*)$, $D(\beta_1^*)$, $K_{\beta_0^*\beta_1^*}$, $K_{\beta_0^*\beta_1^*}$, $r_{\beta_0^*\beta_1^*}$;

3) побудувати довірчі інтервали для β_0, β_1 із заданою надійністю $\gamma = 0{,}95$;

4) побудувати довірчі інтервали для функції регресії $y_i = \beta_0 + \beta_1 x_i$ із заданою надійністю $\gamma = 0{,}95$;

5) побудувати довірчі інтервали для прогнозів індивідуальних значень ознаки $Y = y_i$ із заданою надійністю $\gamma = 0{,}95$;

6) при рівні значущості $\alpha = 0{,}05$ перевірити правильність гіпотези H_0: $\beta_1 = 0$, при альтернативній гіпотезі $H\alpha$: $\beta_0 > 0$.

Розв'язання. 1. Для обчислення значень β_0^*, β_1^* використаємо таблицю:

№ з/п	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	–19,2	–21,8	418,56	368,64	475,24
2	–14,8	–15,4	227,92	219,04	237,16
3	–19,6	–20,8	407,68	384,16	432,64
4	–11,1	–11,3	125,43	123,21	127,69
5	–9,4	–11,6	109,04	88,36	134,56
6	–16,9	–19,2	324,48	285,61	368,64
7	–13,7	–13,0	178,1	187,69	169,0
8	–4,9	–7,4	36,26	24,01	54,76
9	–13,9	–15,1	209,89	193,21	228,01
10	–9,4	–14,4	135,36	88,36	207,36
11	–8,3	–11,1	92,13	68,89	123,21
12	–7,9	–10,5	82,95	62,41	110,25
13	–5,3	–7,2	38,16	28,09	51,84
$\sum$	–154,4	–178,8	2385,96	2121,68	2720,36

Ураховуючи те, що

$$\bar{x} = \frac{\sum x_i}{n} = -\frac{154{,}4}{13} = -11{,}88; \quad \bar{y} = \frac{\sum y_i}{n} = -\frac{178{,}8}{13} = -13{,}75.$$

$$\frac{\sum x_i^2}{n} = \frac{2121{,}68}{13} = 163{,}21; \quad \frac{\sum y_i^2}{n} = \frac{2720{,}36}{13} = 209{,}26.$$

$\frac{\sum x_i y_i}{n} = \frac{2385{,}96}{13} = 183{,}54,$ знаходимо:

$$\beta_1^* = \frac{\frac{\sum x_i y_i}{n} - \overline{xy}}{\frac{\sum x_i^2}{n} - (\bar{x})^2} = \frac{183{,}54 - (-11{,}88)(-13{,}75)}{163{,}21 - (-11{,}88)^2} = \frac{20{,}19}{22{,}08} = 0{,}92;$$

$$\beta_0^* = \bar{y} - \beta_1^* \bar{x} = -13{,}75 - 0{,}92(-11{,}88) = -13{,}75 + 10{,}93 = -2{,}82.$$

Таким чином, дістанемо:

$$\beta_0^* = -2{,}82; \quad \beta_1^* = 0{,}92.$$

Отже, рівняння регресії має вигляд

$$y_i = -2{,}82 + 0{,}92 x_i .$$

Оскільки $r_{xy} = \frac{K_{xy}^*}{\sigma_x \sigma_y}$, то

$$\sigma_x = \sqrt{\frac{\sum x_i^2}{n} - (x)^2} = \sqrt{163{,}21 - (-11{,}88)^2} = \sqrt{22{,}08} = 4{,}7;$$

$$\sigma_y = \sqrt{\frac{\sum y_i^2}{n} - (y)^2} = \sqrt{209{,}26 - \left(-13{,}75^2\right)} = \sqrt{20{,}20} = 4{,}5;$$

$$K_{xy}^* = \frac{\sum x_i y_i}{n} - \overline{xy} = 183{,}54 - (-11{,}88)(-13{,}75) = 20{,}19.$$

Далі обчислимо:

$$r_{xy} = \frac{K_{xy}^*}{\sigma_x \sigma_y} = \frac{20{,}19}{4{,}7 \cdot 4{,}5} = \frac{20{,}19}{21{,}15} = 0{,}96.$$

Оскільки значення коефіцієнта кореляції близьке до одиниці, то звідси випливає, що функція регресії ознак *Y* від *X* близька до лінійної.

2. Для визначення $D(\beta_0^*)$, $D(\beta_1^*)$, $K_{\beta_0^*\beta_1^*}$, $r_{\beta_0^*\beta_1^*}$ необхідно обчислити значення $S_\varepsilon = \dfrac{\sum(\varepsilon_i^*)^2}{n-2}$, що є точковою незміщеною статистичною оцінкою для σ_ε — середньоквадратичного відхилення випадкового фактора ε_i.

Обчислюємо

$$S_\varepsilon = -2{,}82 + 0{,}92x_i = y_i - (-2{,}82 + 0{,}92x_i).$$

x_i	y_i	$\beta_0^* + \beta_1^* x_i$	$y_i - (\beta_0^* + \beta_1^* x_i)$	$\varepsilon_i^2 = [y_i - (\beta_0^* + \beta_1^* x_i)]^2$
–19,2	–21,8	–20,484	–1,316	1,732
–14,8	–15,4	–16,436	1,036	1,073
–19,6	–20,8	–20,852	0,052	0,003
–11,1	–11,3	–13,032	1,732	2,999
–9,4	–11,6	–11,462	–1,138	0,003
–16,9	–19,2	–18,368	–0,832	0,692
–13,7	–13,0	–15,424	2,424	5,876
–4,9	–7,4	–7,328	–0,072	0,005
–13,9	–15,1	–15,608	0,508	0,258
–9,4	–14,4	–11,468	–2,932	8,597
–8,3	–11,1	–10,456	–0,644	0,415
–7,9	–10,5	–10,088	–0,412	0,169
–5,3	–7,2	–7,696	0,496	0,246
–154,4	–178,8			22,068

Отже, дістанемо: $S_\varepsilon^2 = \dfrac{\sum(\varepsilon_i^*)^2}{n-2} = \dfrac{22{,}068}{13-2} = \dfrac{22{,}068}{11} = 2{,}006;$

$$S_\varepsilon = \sqrt{2{,}006} = 1{,}416.$$

Згідно з (514)—(518) маємо:

$$D(\beta_0^*) = \frac{\sum x_i}{\sum(x_i - \bar{x})^2} S_\varepsilon^2 = \frac{\sum x_i^2}{\sum x_i^2 - n(\bar{x})^2} S_\varepsilon^2 = \frac{2121{,}68 \cdot 2{,}006}{2121{,}68 - 13 \cdot (-11{,}88)^2} =$$

$$= \frac{4256{,}1}{2121{,}68 - 1834{,}75} = \frac{4256{,}1}{286{,}93} = 14{,}83.$$

$$D\left(\beta_0^*\right)=14{,}83.$$

$$\sigma_{\left(\beta_0^*\right)}=\sqrt{14{,}83}=3{,}85.$$

$$D\left(\beta_1^*\right)=\frac{S_\varepsilon^2}{\sum\left(x_i-\bar{x}\right)^2}=\frac{2{,}006}{286{,}93}=0{,}007;$$

$$\sigma_{\beta_1^*}=\sqrt{0{,}007}=0{,}084,$$

$$K_{\beta_0^*\beta_1^*}=-\frac{\bar{x}\cdot S_\varepsilon^2}{\sum\left(x_i-\bar{x}\right)^2}=-\frac{\bar{x}\cdot S_\varepsilon^2}{\sum x_i^2-n(\bar{x})^2}=-\frac{11{,}88\cdot 2{,}006}{2121{,}68-13\cdot(-11{,}88)^2}=$$

$$=\frac{11{,}88\cdot 2{,}006}{2121{,}68-1834{,}75}=\frac{23{,}83}{286{,}93}=0{,}083.$$

$$K_{\beta_0^*\beta_1^*}=0{,}083.\quad r_{\beta_0^*\beta_1^*}=\frac{K_{\beta_0^*\beta_1^*}}{\sigma_{\beta_0^*}\sigma_{\beta_1^*}}=\frac{0{,}083}{3{,}85\cdot 0{,}084}=0{,}26.$$

3. З надійністю $\gamma=0{,}95$ побудуємо довірчий інтервал для параметрів β_0, β_1 лінійної регресії ознак *X* і *Y* генеральної сукупності.

Оскільки довірчий інтервал для параметра β_0 дорівнюватиме:

$$\beta_0^*-t(\gamma,k)\,S_\varepsilon\sqrt{\frac{\sum x_i^2}{n\sum\left(x_i-\bar{x}\right)^2}}<\beta_0<\beta_0^*+t(\gamma,k)\,S_\varepsilon\sqrt{\frac{\sum x_i^2}{n\sum\left(x_i-\bar{x}\right)^2}},$$

то нам необхідно знайти $t(\gamma,\,k)$ за таблицею (додаток 3) за заданим значенням $\gamma=0{,}95$ і $k=n-2=13-2=11$. Отже, маємо $t(\gamma=0{,}95;\,k=11)=2{,}201.$ Значення лівого кінця інтервалу дорівнюватиме:

$$\beta_0^*-t(\gamma,k)\,S_\varepsilon\sqrt{\frac{\sum x_i^2}{n\sum\left(x_i-\bar{x}\right)^2}}=\beta_0^*-t(\gamma,k)\,S_\varepsilon\sqrt{\frac{\sum x_i^2}{n\left(\sum x_i^2-(\bar{x})^2\right)}}=$$

$$=-2{,}82-2{,}201\cdot 1{,}416\sqrt{\frac{2121{,}68}{13\left(2121{,}68-(-11{,}88)^2\right)}}=$$

$$=-2{,}82-2{,}201\cdot 1{,}416\sqrt{\frac{2121{,}68}{13\left(2121{,}68-141{,}1344\right)}}=$$

$$=-2{,}82-2{,}201\cdot 1{,}416\sqrt{0{,}0824}=-2{,}82-2{,}201\cdot 1{,}416\cdot 0{,}2871=$$

$$=-2{,}82-0{,}89=-3{,}71.$$

Обчисливши значення правого кінця інтервалу, дістанемо:

$$\beta_0^* + t(\gamma,k)S_\varepsilon\sqrt{\frac{\sum x_i^2}{n\sum(x_i-\bar{x})^2}} = -2{,}82 + 0{,}89 = -1{,}93.$$

Таким чином, маємо

$$-3{,}71 < \beta_0 < -1{,}93\,.$$

Отже, з надійністю $\gamma = 0{,}95$ можна стверджувати, що параметр β_0 перебуватиме в інтервалі $[-3{,}71; -1{,}93]$. Це показано на рис. 154.

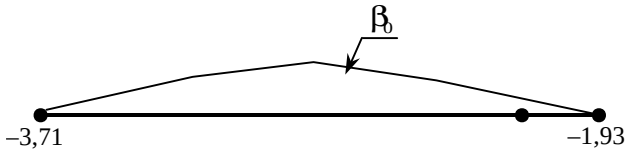


Рис. 154

Довірчий інтервал для параметра β_1 має вигляд:

$$\beta_1^* - \frac{t(\gamma,k)S_\varepsilon}{\sqrt{\sum(x_i-\bar{x})^2}} < \beta_1 < \beta_1^* + \frac{t(\gamma,k)S_\varepsilon}{\sqrt{\sum(x_i-\bar{x})^2}}.$$

Обчислюємо значення кінців довірчого інтервалу:

$$\beta_1^* + \frac{t(\gamma,k)S_\varepsilon}{\sqrt{\sum(x_i-\bar{x})^2}} = 0{,}92 - \frac{2{,}201\cdot 1{,}416}{\sqrt{2121{,}68 - 1834{,}7472}} =$$

$$= 0{,}92 - \frac{3{,}12542}{\sqrt{286{,}9328}} = 0{,}92 - \frac{3{,}12542}{16{,}94} = 0{,}92 - 0{,}25 = 0{,}67.$$

$$\beta_1^* + t(\gamma,k)\frac{S_\varepsilon}{\sqrt{\sum(x_i-\bar{x})^2}} = 0{,}92 + \frac{2{,}201\cdot 1{,}42}{\sqrt{286{,}9328}} = 0{,}92 + 0{,}25 = 1{,}17.$$

Отже, маємо:

$$0{,}67 < \beta_1 < 1{,}17,$$

що можна тлумачити так: із надійністю $\gamma = 0{,}95$ параметр β_1 міститься в інтервалі $[0{,}67; 1{,}17]$, що ілюструє рис. 155.

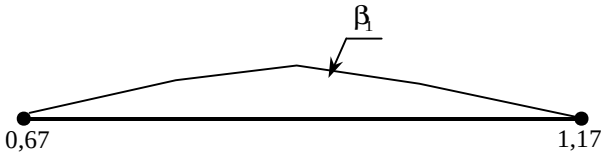


Рис. 155

Побудова довірчого інтервалу для функції регресії $\beta_0 + \beta_1 x_i$. Для побудови довірчого інтервалу функції регресії $y_i = \beta_0 + \beta_1 x_i$ необхідно обчислити дисперсію для $\beta_0^* + \beta_1^* x_i$, а саме:

$$D(y_i^*) = S_\varepsilon^2 \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right),$$

для значень x_i:

$$x_1 = 19{,}2 \to D(y_1^*) = 2{,}006 \left(\frac{1}{13} + \frac{(-19{,}2 + 11{,}88)^2}{\sum x_i^2 - n(\bar{x})^2} \right) =$$

$$= 2{,}006 \left[\frac{1}{13} + \frac{(-19{,}2 + 11{,}88)^2}{2121{,}68 - 13(-11{,}88)^2} \right] = 2{,}006 \left[\frac{1}{13} + \frac{53{,}5824}{286{,}9328} \right] =$$

$$= 2{,}006[0{,}077 + 0{,}187] = 2{,}006 \cdot 0{,}264 = 0{,}53.$$

$$\sigma(y_1^*) = \sqrt{0{,}53} = 0{,}73;$$

$$x_2 = 14{,}8 \to D(y_2^*) = 2{,}006 \left(\frac{1}{13} + \frac{(-14{,}8 + 11{,}88)^2}{286{,}9328} \right) =$$

$$= 2{,}006 \left[0{,}077 + \frac{8{,}5264}{286{,}9328} \right] = 2{,}006[0{,}077 + 0{,}030] = 2{,}006 \cdot 0{,}107 = 0{,}215.$$

$$\sigma(y_2^*) = \sqrt{0{,}215} = 0{,}46;$$

$$x_3 = 19{,}6 \to D(y_3^*) = 2{,}006 \left(0{,}077 + \frac{(-19{,}6 + 11{,}88)^2}{286{,}9328} \right) =$$

$$= 2{,}006[0{,}077 + 0{,}208] = 2{,}006 \cdot 0{,}285 = 0{,}572.$$

$$\sigma(y_3^*) = \sqrt{0{,}572} = 0{,}76;$$

$$x_4 = 11{,}1 \to D(y_4^*) = 2{,}006 \left(0{,}077 + \frac{(-11{,}1 + 11{,}88)^2}{286{,}9328} \right) =$$

$$= 2{,}006[0{,}077 + 0{,}002] = 2{,}006 \cdot 0{,}079 = 0{,}158.$$

$$\sigma(y_4^*) = \sqrt{0{,}158} = 0{,}40;$$

$$x_5 = 9{,}4 \to D(y_5^*) = 2{,}006 \left(0{,}077 + \frac{(-9{,}4 + 11{,}88)^2}{286{,}9328} \right) =$$

$$= 2{,}006[0{,}077 + 0{,}021] = 2{,}006 \cdot 0{,}098 = 0{,}198.$$

$$\sigma(y_5^*) = \sqrt{0{,}198} = 0{,}44;$$

$$x_6 = 16{,}9 \to D\left(y_6^*\right) = 2{,}006[0{,}077 + 0{,}88] = 2{,}006 \cdot 0{,}165 = 0{,}33.$$

$$\sigma\left(y_6^*\right) = 0{,}57;$$

$$x_7 = 13{,}7 \to D\left(y_7^*\right) = 2{,}006[0{,}077 + 0{,}0115] = 2{,}006 \cdot 0{,}0885 = 0{,}178.$$

$$\sigma\left(y_7^*\right) = 0{,}42;$$

$$x_8 = 4{,}9 \to D\left(y_8^*\right) = 2{,}006[0{,}077 + 0{,}170] = 2{,}006 \cdot 0{,}247 = 0{,}495.$$

$$\sigma\left(y_8^*\right) = 0{,}70;$$

$$x_9 = 13{,}9 \to D\left(y_9^*\right) = 2{,}006[0{,}077 + 0{,}014] = 2{,}006 \cdot 0{,}091 = 0{,}183.$$

$$\sigma\left(y_9^*\right) = 0{,}43;$$

$$x_{10} = 9{,}4 \to D\left(y_{10}^*\right) = 2{,}006[0{,}077 + 0{,}021] = 2{,}006 \cdot 0{,}098 = 0{,}198.$$

$$\sigma\left(y_{10}^*\right) = 0{,}44;$$

$$x_{11} = 8{,}3 \to D\left(y_{11}^*\right) = 2{,}006[0{,}077 + 0{,}04] = 2{,}006 \cdot 0{,}117 = 0{,}235.$$

$$\sigma\left(y_{11}^*\right) = 0{,}48;$$

$$x_{12} = 7{,}9 \to D\left(y_{12}^*\right) = 2{,}006[0{,}077 + 0{,}05] = 2{,}006 \cdot 0{,}127 = 0{,}255.$$

$$\sigma\left(y_{12}^*\right) = 0{,}5;$$

$$x_{13} = 5{,}3 \to D\left(y_{13}^*\right) = 2{,}006[0{,}077 + 0{,}151] = 2{,}006 \cdot 0{,}228 = 0{,}457.$$

$$\sigma\left(y_{13}^*\right) = 0{,}68.$$

Довірчий інтервал для парної лінійної функції регресії має вигляд

$$\beta_0^* + \beta_1^* x_i - t(\gamma, k) S_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{n\sum (x_i - \bar{x})^2}} < \beta_0 + \beta_1 x_i <$$

$$< \beta_0^* + \beta_1^* x_i + t(\gamma, k) S_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{n\sum (x_i - \bar{x})^2}}.$$

Обчислюємо довірчі інтервали для послідовності значень аргументу x_i, використовуючи при цьому результати обчислень, наведених у табл. 2.

$x_1 = 19{,}2.$

$$-20{,}484 - 2{,}201 \cdot 0{,}73 < \beta_0 + \beta_1 x_i < -20{,}484 + 2{,}201 \cdot 0{,}73 \to$$

$$\to -20{,}484 - 1{,}629 < \beta_0 + \beta_1(-19{,}2) < -20{,}484 + 1{,}629 \to$$

$$\to 22{,}113 < \beta_0 + \beta_1(-19{,}2) < -18{,}856.$$

$x_2 = -14{,}8.$

$-16{,}436 - 2{,}201 \cdot 0{,}46 < \beta_0 + \beta_1(-14{,}8) < -16{,}436 + 2{,}201 \cdot 0{,}46 \rightarrow$

$\rightarrow -16{,}436 - 1{,}03 < \beta_0 + \beta_1(-14{,}8) < -16{,}436 + 1{,}03.$

$-17{,}466 < \beta_0 + \beta_1(-14{,}8) < -15{,}406.$

$x_3 = -19{,}6.$

$-20{,}852 - 2{,}201 \cdot 0{,}76 < \beta_0 + \beta_1(-19{,}6) < -20{,}852 + 2{,}201 \cdot 0{,}76 \rightarrow$

$\rightarrow -20{,}852 - 1{,}69 < \beta_0 + \beta_1(-19{,}6) < -20{,}852 + 1{,}69 \rightarrow$

$\rightarrow -22{,}542 < \beta_0 + \beta_1(-19{,}6) < -19{,}162.$

$x_4 = -11{,}1.$

$-13{,}032 - 2{,}201 \cdot 0{,}4 < \beta_0 + \beta_1(-11{,}1) < -13{,}032 + 2{,}201 \cdot 0{,}4 \rightarrow$

$\rightarrow -13{,}032 - 0{,}88 < \beta_0 + \beta_1(-11{,}1) < -13{,}032 + 0{,}88 \rightarrow$

$\rightarrow -13{,}912 < \beta_0 + \beta_1(-11{,}1) < -12{,}152.$

$x_5 = -9{,}4.$

$-11{,}462 - 2{,}201 \cdot 0{,}44 < \beta_0 + \beta_1(-9{,}4) < -11{,}462 + 2{,}201 \cdot 0{,}44 \rightarrow$

$\rightarrow -11{,}462 - 0{,}968 < \beta_0 + \beta_1(-9{,}4) < -11{,}462 + 0{,}968 \rightarrow$

$\rightarrow -12{,}43 < \beta_0 + \beta_1(-9{,}4) < -10{,}494.$

$x_6 = -16{,}9.$

$-18{,}368 - 2{,}201 \cdot 0{,}57 < \beta_0 + \beta_1(-19{,}9) < -18{,}568 + 2{,}201 \cdot 0{,}57 \rightarrow$

$\rightarrow -18{,}368 - 1{,}299 < \beta_0 + \beta_1(-19{,}9) < -18{,}368 + 1{,}299 \rightarrow$

$\rightarrow -19{,}667 < \beta_0 - \beta_1(-19{,}9) < -17{,}069.$

$x_7 = -13{,}7.$

$-15{,}424 - 2{,}201 \cdot 0{,}42 < \beta_0 + \beta_1 13{,}7 < -15{,}424 + 2{,}201 \cdot 0{,}42 \rightarrow$

$\rightarrow -15{,}424 - 0{,}946 < \beta_0 - \beta_1 13{,}7 < -15{,}424 + 0{,}946 \rightarrow$

$\rightarrow -16{,}37 < \beta_0 - \beta_1 13{,}7 < -14{,}478.$

$x_8 = -4{,}9.$

$-7{,}328 - 2{,}201 \cdot 0{,}70 < \beta_0 - \beta_1 4{,}9 < -7{,}328 + 2{,}201 \cdot 0{,}70 \rightarrow$

$\rightarrow -7{,}328 - 1{,}519 < \beta_0 - \beta_1 4{,}9 < -7{,}328 + 1{,}519 \rightarrow$

$\rightarrow -8{,}847 < \beta_0 - \beta_1 4{,}9 < -5{,}809.$

$x_9 = -13{,}9.$

$-15{,}608 - 2{,}201 \cdot 0{,}43 < \beta_0 - \beta_1 13{,}9 < -15{,}608 + 2{,}201 \cdot 0{,}43 \rightarrow$

$\rightarrow -15{,}608 - 0{,}946 < \beta_0 - \beta_1 13{,}9 < -15{,}608 + 0{,}946 \rightarrow$

$\rightarrow -16{,}554 < \beta_0 - \beta_1 13{,}9 < -14{,}662.$

$x_{10} = -9{,}4.$

$-11{,}468 - 2{,}201 \cdot 0{,}44 < \beta_0 - \beta_1 9{,}4 < -11{,}468 + 2{,}201 \cdot 0{,}44 \rightarrow$

$\rightarrow -11{,}468 - 0{,}968 < \beta_0 - \beta_1 9{,}4 < -11{,}468 + 0{,}968 \rightarrow$

$\rightarrow -11{,}468 < \beta_0 - \beta_1 9{,}4 < -10{,}5.$

$x_{11} = -8{,}3.$

$-10{,}456 - 2{,}201 \cdot 0{,}48 < \beta_0 - \beta_1 8{,}3 < -10{,}456 + 2{,}201 \cdot 0{,}48 \rightarrow$

$\rightarrow -10{,}456 - 1{,}057 < \beta_0 - \beta_1 8{,}3 < -10{,}456 + 1{,}057 \rightarrow$

$\rightarrow -11{,}513 < \beta_0 - \beta_1 8{,}3 < -9{,}399.$

$x_{12} = -7{,}9.$

$-10{,}088 - 2{,}201 \cdot 0{,}5 < \beta_0 - \beta_1 7{,}9 < -10{,}088 + 2{,}201 \cdot 0{,}5 \rightarrow$

$\rightarrow -10{,}088 - 1{,}1005 < \beta_0 - \beta_1 7{,}9 < -10{,}088 + 1{,}1005 \rightarrow$

$\rightarrow -11{,}1885 < \beta_0 + \beta_1 7{,}9 < -8{,}9875.$

$x_{13} = -5{,}3.$

$-7{,}696 - 2{,}201 \cdot 0{,}68 < \beta_0 - \beta_1 5{,}3 < -7{,}696 + 2{,}201 \cdot 0{,}68 \rightarrow$

$\rightarrow -7{,}696 - 1{,}452 < \beta_0 - \beta_1 5{,}3 < -7{,}696 + 1{,}452 \rightarrow$

$\rightarrow -9{,}148 < \beta_0 - \beta_1 5{,}3 < -6{,}244.$

Графічно довірчий інтервал зображено на рис. 156.

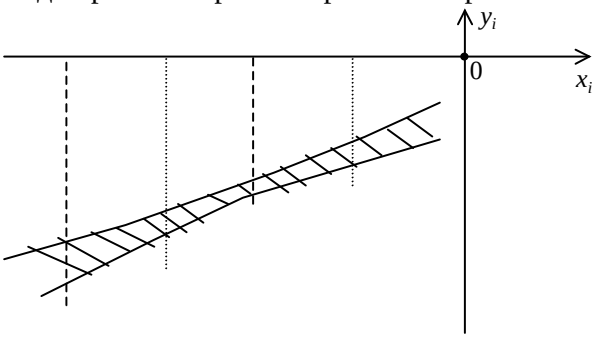


Рис. 156

Побудуємо довірчий інтервал для прогнозів індивідуальних значень $Y = y_i$ із заданою надійністю $\gamma = 0{,}95$. Для побудови цього інтервалу необхідно обчислити прогнозне середнє квадратичне відхилення S_p за формулою

$$S_p = S_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = S_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum x_i^2 - n(\bar{x})^2}}$$

для послідовності значень x_i, а саме:

$x_1 = -19{,}2.$

$S_{p_1}^2 = 2{,}006(1 + 0{,}077 + 0{,}187) = 2{,}006 \cdot 1{,}264 = 2{,}536.$

$S_{p_1} = 1{,}592.$

$x_2 = -14{,}8.$

$S_p^2 = 2{,}006(1 + 0{,}077 + 0{,}030) = 2{,}006 \cdot 1{,}107 = 2{,}221.$

$S_{p_2} = 1{,}490.$

$x_3 = -19{,}6.$

$S_{p_3}^2 = 2{,}006(1 + 0{,}077 + 0{,}208) = 2{,}006 \cdot 1{,}285 = 2{,}578.$

$S_{p_3} = 1{,}606.$

$x_4 = -11{,}1.$

$S_{p_4}^2 = 2{,}006(1 + 0{,}077 + 0{,}002) = 2{,}006 \cdot 1{,}079 = 2{,}164.$

$S_{p_4} = 1{,}471.$

$x_5 = -9{,}4.$

$S_{p_5}^2 = 2{,}006(1 + 0{,}077 + 0{,}021) = 2{,}006 \cdot 1{,}098 = 2{,}203.$

$S_{p_5} = 1{,}484.$

$x_6 = -16{,}9.$

$S_{p_6}^2 = 2{,}006(1 + 0{,}077 + 0{,}088) = 2{,}006 \cdot 1{,}165 = 2{,}337.$

$S_{p_6} = 1{,}529.$

$x_7 = -13{,}7.$

$S_{p_7}^2 = 2{,}006(1 + 0{,}077 + 0{,}0115) = 2{,}006 \cdot 1{,}0885 = 2{,}184.$

$S_{p_7} = 1{,}478.$

$x_8 = -4{,}9.$

$S_{p_8}^2 = 2{,}006(1 + 0{,}077 + 0{,}170) = 2{,}006 \cdot 1{,}247 = 2{,}501.$

$S_{p_8} = 1{,}581.$

$x_9 = -13{,}9.$

$S_{p_9}^2 = 2{,}006(1 + 0{,}077 + 0{,}014) = 2{,}006 \cdot 1{,}091 = 2{,}189.$

$S_{p_9} = 1{,}480.$

$x_{10} = -9{,}4.$

$S_{p_{10}}^2 = 2{,}006(1 + 0{,}077 + 0{,}021) = 2{,}006 \cdot 1{,}098 = 2{,}203.$

$S_{p_{10}} = 1{,}484.$

$x_{11} = -8{,}3.$

$S_{p_{11}}^2 = 2{,}006(1 + 0{,}077 + 0{,}04) = 2{,}006 \cdot 1{,}117 = 2{,}24.$

$S_{p_{11}} = 1{,}497.$

$x_{12} = -7{,}9.$

$S_{p_{12}}^2 = 2{,}006(1 + 0{,}077 + 0{,}05) = 2{,}006 \cdot 1{,}127 = 2{,}26.$

$S_{p_{12}} = 1{,}5.$

$x_{13} = -5{,}3.$

$S_{p_{13}}^2 = 2{,}006(1 + 0{,}077 + 0{,}151) = 2{,}006 \cdot 1{,}228 = 2{,}463.$

$S_{p_{13}} = 1{,}569.$

Оскільки довірчий інтервал для прогнозних значень y_i дорівнює

$$\beta_0^* + \beta_1^* x_i - t(\gamma; k) S_p < \beta_0 + \beta_1 x_i < \beta_0^* + \beta_1^* x_i + t(\gamma; k) S_p,$$

то для послідовних значень x_i дістанемо:

$x_1 = -19{,}2.$

$-20{,}484 - 2{,}201 \cdot 1{,}592 < \beta_0 - \beta_1 19{,}2 < -20{,}484 - 2{,}201 \cdot 1{,}592 \rightarrow$

$\rightarrow -20{,}484 - 3{,}504 < \beta_0 - \beta_1 19{,}2 < -20{,}484 + 3{,}495 \rightarrow$

$\rightarrow -23{,}979 < \beta_0 - \beta_1 19{,}2 < -1{,}6989.$

$x_2 = -14{,}8.$

$-16{,}436 - 2{,}201 \cdot 1{,}490 < \beta_0 - \beta_1 14{,}8 < -16{,}436 - 2{,}201 \cdot 1{,}490 \rightarrow$

$\rightarrow -16{,}436 - 3{,}279 < \beta_0 - \beta_1 14{,}8 < -16{,}436 + 3{,}279 \rightarrow$

$\rightarrow -19{,}722 < \beta_0 - \beta_1 14{,}8 < -13{,}15.$

$x_3 = -19{,}6.$

$-20{,}852 - 2{,}201 \cdot 1{,}606 < \beta_0 - \beta_1 19{,}6 < -20{,}852 - 2{,}201 \cdot 1{,}606 \rightarrow$

$\rightarrow -20{,}852 - 3{,}548 < \beta_0 - \beta_1 19{,}6 < -20{,}852 - 3{,}548 \rightarrow$

$\rightarrow -24{,}4 < \beta_0 - \beta_1 19{,}6 < -17{,}304.$

$x_4 = -11{,}1.$

$-13{,}032 - 2{,}201 \cdot 1{,}471 < \beta_0 - \beta_1 11{,}1 < -13{,}032 - 2{,}201 \cdot 1{,}471 \rightarrow$

$\rightarrow -13{,}032 - 3{,}236 < \beta_0 - \beta_1 11{,}1 < -13{,}032 + 3{,}236 \rightarrow$

$\rightarrow -16{,}268 < \beta_0 - \beta_1 11{,}1 < -9{,}796.$

$x_5 = -9{,}4.$

$-11{,}462 - 2{,}201 \cdot 1{,}484 < \beta_0 - \beta_1 9{,}4 < -11{,}462 + 2{,}201 \cdot 1{,}484 \rightarrow$

$\rightarrow -11{,}462 - 3{,}258 < \beta_0 - \beta_1 9{,}4 < -11{,}462 + 3{,}258 \rightarrow$

$\rightarrow -14{,}72 < \beta_0 - \beta_1 9{,}4 < -8{,}204.$

$x_6 = -16{,}9.$

$-18{,}368 - 2{,}201 \cdot 1{,}529 < \beta_0 - \beta_1 16{,}9 < -18{,}368 - 2{,}201 \cdot 1{,}529 \rightarrow$

$\rightarrow -18{,}368 - 3{,}37 < \beta_0 - \beta_1 16{,}9 < -18{,}368 + 3{,}37 \rightarrow$

$\rightarrow -21{,}738 < \beta_0 - \beta_1 16{,}9 < -14{,}998.$

$x_7 = -13{,}7.$

$-15{,}424 - 2{,}201 \cdot 1{,}478 < \beta_0 - \beta_1 13{,}7 < -15{,}424 - 2{,}201 \cdot 1{,}478 \rightarrow$

$\rightarrow -15{,}424 - 3{,}258 < \beta_0 - \beta_1 13{,}7 < -15{,}424 + 3{,}258 \rightarrow$

$\rightarrow -18{,}682 < \beta_0 - \beta_1 13{,}7 < -12{,}160.$

$x_8 = -4{,}9.$

$-7{,}328 - 2{,}201 \cdot 1{,}581 < \beta_0 - \beta_1 4{,}9 < -7{,}328 - 2{,}201 \cdot 1{,}581 \rightarrow$

$\rightarrow -7{,}328 - 3{,}478 < \beta_0 - \beta_1 4{,}9 < -7{,}328 + 3{,}478 \rightarrow$

$\rightarrow -10{,}806 < \beta_0 - \beta_1 4{,}9 < -3{,}85.$

$x_9 = -13{,}9.$

$-15{,}608 - 2{,}201 \cdot 1{,}480 < \beta_0 - \beta_1 13{,}9 < -15{,}608 - 2{,}201 \cdot 1{,}480 \rightarrow$

$\rightarrow -15{,}608 - 3{,}258 < \beta_0 - \beta_1 13{,}9 < -15{,}608 + 3{,}258 \rightarrow$

$\rightarrow -18{,}866 < \beta_0 - \beta_1 13{,}9 < -12{,}358.$

$x_{10} = -9{,}4.$

$-11{,}468 - 2{,}201 \cdot 1{,}484 < \beta_0 - \beta_1 9{,}4 < -11{,}468 - 2{,}201 \cdot 1{,}484 \rightarrow$

$\rightarrow -11{,}468 - 3{,}258 < \beta_0 - \beta_1 9{,}4 < -11{,}468 + 3{,}258 \rightarrow$

$\rightarrow -14{,}726 < \beta_0 - \beta_1 9{,}4 < -8{,}21.$

$x_{11} = -8{,}3.$

$-10{,}456 - 2{,}201 \cdot 1{,}497 < \beta_0 - \beta_1 8{,}3 < -10{,}456 - 2{,}201 \cdot 1{,}497 \rightarrow$

$\rightarrow -10{,}456 - 3{,}259 < \beta_0 - \beta_1 8{,}3 < -10{,}456 + 3{,}259 \rightarrow$

$\rightarrow -13{,}751 < \beta_0 - \beta_1 8{,}3 < -7{,}161.$

$x_{12} = -7{,}9.$

$-10{,}088 - 2{,}201 \cdot 1{,}503 < \beta_0 - \beta_1 7{,}9 < -10{,}088 - 2{,}201 \cdot 1{,}503 \rightarrow$

$\rightarrow -10{,}088 - 3{,}302 < \beta_0 - \beta_1 7{,}9 < -10{,}088 + 3{,}302 \rightarrow$

$\rightarrow -13{,}39 < \beta_0 - \beta_1 7{,}9 < -6{,}780.$

$x_{13} = -5{,}3.$

$-7{,}696 - 2{,}201 \cdot 1{,}569 < \beta_0 - \beta_1 5{,}3 < -7{,}696 - 2{,}201 \cdot 1{,}569 \rightarrow$

$\rightarrow -7{,}696 - 3{,}442 < \beta_0 - \beta_1 5{,}3 < -7{,}696 + 3{,}442 \rightarrow$

$\rightarrow -11{,}138 < \beta_0 - \beta_1 5{,}3 < -4{,}254.$

Графічно довірчий інтервал для прогнозних значень ознаки Y зображено на рис. 157.

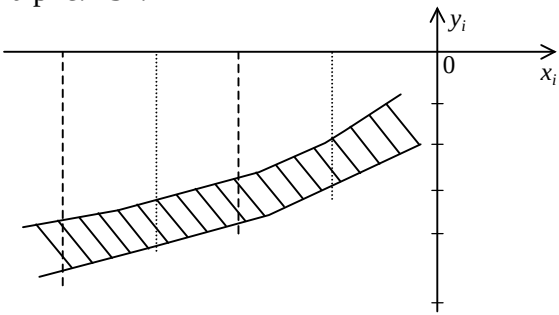


Рис. 157

Для перевірки правдивості статистичної гіпотези $H_0: \beta_1 = 0$ при альтернативній гіпотезі $H_\alpha: \ \beta_1 > 0$ при рівні значущості $\alpha = 0{,}05$. Побудуємо правобічну критичну область.

Оскільки за статистичний критерій береться випадкова величина

$$t = \frac{\beta_1^* - \beta_1}{\sigma(\beta^*)} = \frac{\beta_1^*}{\dfrac{S_\varepsilon}{\sqrt{\sum (x_i - \bar{x})^2}}},$$

яка має t-розподіл (Стьюдента) з $k = n - 2$ ступенями свободи, за таблицею (додаток 6) знаходимо критичну точку $t(\alpha = 0{,}05;\ k = 13 - 2) = t(\alpha = 0{,}05;\ k = 11) = 2{,}2$. Правобічна критична область зображена на рис. 158.

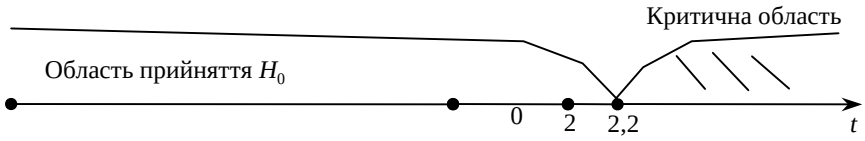


Рис. 158

Обчислюємо спостережуване значення критерію

$$t^* = \frac{\beta_1^*}{\dfrac{S_\varepsilon}{\sqrt{\sum (x_i - \bar{x})^2}}} = \frac{0{,}92}{0{,}024} = 38{,}3.$$

Оскільки $t^* > t_{кр}$, то статистична гіпотеза $\beta_1 = 0$ відхиляється.

Отже, форма зв'язку між ознаками X і Y є лінійною.

3. Множинна лінійна регресія

Загальна інформація. На практиці здебільшого залежна змінна y_i пов'язана з впливом не одного, а кількох аргументів. У цьому разі регресію називають множинною. При цьому якщо аргументи в функції регресії в першій степені, то множинна регресія називається *лінійною*, у противному разі — *множинною нелінійною регресією*.

Деякі елементи матричної алгебри:

а) *норма вектора. Ортогональні вектори і матриці*

Якщо $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_4 \end{pmatrix}$, тоді $\vec{x}' = (x_1, x_2, \dots x_n)$, тоді норма вектора $\vec{x}$ буде число, яке дістанемо за формулою

$$|\vec{x}| = \sqrt{\vec{x}'\,\vec{x}} = \sqrt{\sum_{i=1}^{n} x_i^2}\,. \qquad (534)$$

У разі, коли $\sqrt{\sum x_i^2} = 1$, то вектор $\vec{x}$ називають нормованим. Якщо для квадратної матриці A виконується рівність $A \cdot A' = E = I$ (E — одинична матриця $A' = A^{\mathrm{T}}$), то вона називається ортогональною;

б) *диференціювання векторів*

Нехай задано два вектори $\vec{x}' = (x_1, x_2, \dots x_n)$, $\vec{a}' = (a_1, a_2, \dots a_n)$, тоді

$$\vec{x} \cdot \vec{a} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{pmatrix} = \sum_{i=1}^{n} a_i x_i\,.$$

Частинні похідні за $\vec{x}$ $\left(i = \overline{1,n}\right)$ від добутку $\vec{x}' \cdot \vec{a}$ можна записати так:

$$\frac{\partial}{\partial \vec{x}}\left(\vec{x}'\,\vec{a}\right) = \frac{\partial}{\partial \vec{x}}\begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \vec{a}'.$$

Тоді

$$\frac{\partial}{\partial \vec{x}}(\vec{a}'\,\vec{x}) = \begin{pmatrix} \dfrac{\partial}{\partial \vec{x}_1} \\ \dfrac{\partial}{\partial \vec{x}_2} \\ \dots \\ \dfrac{\partial}{\partial \vec{x}_n} \end{pmatrix} (a_1\ a_2\ \dots a_n) \cdot \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} =$$

$$=\begin{pmatrix} \dfrac{\partial}{\partial \vec{x}_1} \\ \dfrac{\partial}{\partial \vec{x}_2} \\ \dots \\ \dfrac{\partial}{\partial \vec{x}_n} \end{pmatrix} (a_1 x_1 + a_2 x_2 + \dots + a_n x_n) = \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{pmatrix} = \vec{a}.$$

Отже, маємо:

$$\frac{\partial}{\partial \vec{x}} (\vec{x}' \vec{a}) = \vec{a}, \tag{535}$$

$$\frac{\partial}{\partial \vec{x}} (\vec{a}' \vec{x}) = \vec{a}'. \tag{536}$$

Диференціювання добутку $A\vec{x}$ і $\vec{x}'A$.

Нехай задано

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix},$$

тоді

$$\frac{\partial}{\partial \vec{x}} (A\vec{x}) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = \frac{\partial}{\partial \vec{x}} \begin{bmatrix} \sum_{j=1}^{n} a_{1j} x_j \\ \sum_{j=1}^{n} a_{2j} x_j \\ \dots \\ \sum_{j=1}^{n} a_{nj} x_j \end{bmatrix} =$$

$$= \begin{pmatrix} \dfrac{\partial}{\partial x_1} \sum_{j=1}^{n} a_{1j} x_j & \dfrac{\partial}{\partial x_1} \sum_{j=1}^{n} a_{2j} x_j & \dots & \dfrac{\partial}{\partial x_n} \sum_{j=1}^{n} a_{nj} x_j \\ \dfrac{\partial}{\partial x_2} \sum_{j=1}^{n} a_{1j} x_j & \dfrac{\partial}{\partial x_2} \sum_{j=1}^{n} a_{2j} x_j & \dots & \dfrac{\partial}{\partial x_n} \sum_{j=1}^{n} a_{nj} x_j \\ \dots & \dots & \dots & \dots \\ \dfrac{\partial}{\partial x_n} \sum_{j=1}^{n} a_{1j} x_j & \dfrac{\partial}{\partial x_n} \sum_{j=1}^{n} a_{2j} x_j & \dots & \dfrac{\partial}{\partial x_n} \sum_{j=1}^{n} a_{nj} x_j \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{pmatrix} = A'.$$

Отже,

$$\frac{\partial}{\partial \vec{x}}(A\vec{x}) = A'. \qquad (537)$$

Частинні похідні від добутку $\vec{x}A$ знаходимо так:

$$\frac{\partial}{\partial \vec{x}}(A\vec{x}) = \frac{\partial}{\partial \vec{x}}\left(\begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \cdot \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}\right) = \frac{\partial}{\partial \vec{x}}\begin{bmatrix} \sum_{i=1}^{n} a_{i1} x_i \\ \sum_{i=1}^{n} a_{i2} x_i \\ \dots \\ \sum_{i=1}^{n} a_{in} x_i \end{bmatrix} =$$

$$= \begin{pmatrix} \frac{\partial}{\partial x_1}\sum_{i=1}^{n} a_{i1} x_i & \frac{\partial}{\partial x_1}\sum_{i=1}^{n} a_{i2} x_i & \dots & \frac{\partial}{\partial x_n}\sum_{i=1}^{n} a_{in} x_i \\ \frac{\partial}{\partial x_2}\sum_{i=1}^{n} a_{i1} x_i & \frac{\partial}{\partial x_2}\sum_{i=1}^{n} a_{i2} x_i & \dots & \frac{\partial}{\partial x_n}\sum_{i=1}^{n} a_{in} x_i \\ \dots & \dots & \dots & \dots \\ \frac{\partial}{\partial x_n}\sum_{i=1}^{n} a_{i1} x_i & \frac{\partial}{\partial x_n}\sum_{i=1}^{n} a_{i2} x_i & \dots & \frac{\partial}{\partial x_n}\sum_{i=1}^{n} a_{in} x_i \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} = A.$$

$$\frac{\partial}{\partial \vec{x}}(\vec{x}'A) = A. \qquad (538)$$

Диференціювання добутку $\vec{x}'A\vec{x}$.

$$\frac{\partial}{\partial \vec{x}}(\vec{x}'A\vec{x}) = \frac{\partial}{\partial \vec{x}}(\vec{x}'(A\vec{x})) + \frac{\partial}{\partial \vec{x}}((\vec{x}'A)\vec{x}) =$$

$$= \left|\text{Використовуючи (537), (538), дістанемо}\right| = A\vec{x} + (\vec{x}'A)' = A\vec{x} + A'\vec{x}.$$

Отже, маємо:

$$\frac{\partial}{\partial \vec{x}}(\vec{x}'A\vec{x}) = A\vec{x} + A'\vec{x}. \qquad (539)$$

Тут використана властивість транспонування добутку матриць, а саме: якщо A і B є матрицями одного й того самого розміру, то $(A \cdot B)' = B' \cdot A'$.

Для матриць A, B, C маємо $(ABC)' = C'B'A'$ і т. д.

Якщо матриця A є симетричною, то

$$A' = A.$$

Тоді

$$\frac{\partial}{\partial \vec{x}}(\vec{x}'A\vec{x}) = 2A\vec{x}. \qquad (540)$$

Лінійна множинна регресія.

Визначення статистичних точкових оцінок

Розглянемо лінійну залежність y_i від m аргументів $(x_1, x_2, \ldots x_m)$. Лінійна модель у цьому разі набирає такого вигляду

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi} . \qquad (541)$$

Для вибірки обсягу n матимемо систему лінійних рівнянь

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \ldots + \beta_m x_{m1} + \varepsilon_1, \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_m x_{m2} + \varepsilon_2, \\ y_3 &= \beta_0 + \beta_1 x_{31} + \beta_2 x_{32} + \ldots + \beta_m x_{m3} + \varepsilon_3, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n &= \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \ldots + \beta_m x_{mn} + \varepsilon_n , \end{aligned} \qquad (542)$$

де ε_i — випадкова величина, що має нормальний закон розподілу з числовими характеристиками $M(\varepsilon_i) = 0, \quad D(\varepsilon_i) = M(\varepsilon_i^2) = \sigma_\varepsilon^2$ і при цьому $K_{ij} = 0$. У векторно-матричній формі система (542) набирає такого вигляду:

$$\vec{Y} = X\vec{\beta} + \vec{\varepsilon}, \qquad (543)$$

де

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix}; \quad \vec{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \ldots \\ \beta_n \end{pmatrix}; \quad \vec{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix}; \quad X = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1m} \\ 1 & x_{21} & \ldots & x_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & \ldots & x_{nm} \end{pmatrix}.$$

Матрицю X розміром $(m+1) \times n$ називають регресійною, а елементи x_{ij} цієї матриці — регресорами. Параметри рівняння (541) є величинами сталими, але невідомими. Ці параметри оцінювання статистичними точковими оцінками $\beta_0^*, \beta_1^*, \beta_2^*, \ldots \beta_m^*$, які дістають шляхом обробки результатів вибірки, і є величинами випадковими. Таким чином, рівнянню (541) відповідає статистична оцінка

$$y_i = \beta_0^* + \beta_1^* x_{i1} + \beta_2^* x_{i2} + \ldots + \beta_m^* x_{im} + \varepsilon_i . \qquad (544)$$

Статистична оцінка для вектора $\vec{y}$ буде визначатись вектором

$$\vec{y} = X\vec{\beta}^* + \vec{\varepsilon}, \qquad (545)$$

де

$$\vec{y}=\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_m \end{pmatrix}; \quad X=\begin{pmatrix} 1 & x_{11} & \dots & x_{1m} \\ 1 & x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nm} \end{pmatrix}; \quad \vec{\beta}^*=\begin{pmatrix} \beta_0^* \\ \beta_1^* \\ \beta_2^* \\ \dots \\ \beta_m^* \end{pmatrix}; \quad \vec{\varepsilon}=\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix}.$$

Вектор похибок дорівнюватиме

$$\vec{\varepsilon}=\vec{y}-X\cdot\vec{\beta}^*. \tag{546}$$

Для визначення компонентів вектора $\vec{\beta}^*$ (статистичних точкових оцінок компонентів вектора $\vec{\beta}$) застосовується метод найменших квадратів.

Знайдемо суму квадратів усіх похибок:

$$(\vec{\varepsilon})'\vec{\varepsilon}=\left(\vec{y}-X\vec{\beta}^*\right)'\left(\vec{y}-X\vec{\beta}\right)=\left((\vec{y})'-\left(\vec{\beta}^*\right)'X'\right)\left(\vec{y}-X\vec{\beta}\right)=$$

$$=(\vec{y})'\vec{y}-(\vec{y})'X\vec{\beta}^*-\left(\vec{\beta}^*\right)'X'\vec{y}^*+\left(\vec{\beta}^*\right)'X'X\vec{\beta}^*=$$

$$=(\vec{y})'\vec{y}-2\left(\vec{\beta}^*\right)'X'\vec{y}+\left(\vec{\beta}^*\right)'X'X\vec{\beta}^*.$$

Тут застосовано такі рівності: $(\vec{y})'X\vec{\beta}^*=\left(\vec{\beta}^*\right)'X'\vec{y}$; $\left(X\vec{\beta}^*\right)'=\left(\vec{\beta}^*\right)'X'$.

Мінімізуючий добуток $(\vec{\varepsilon})'\vec{\varepsilon}$, а саме:

прирівнюючи частинні похідні за елементами вектора $\vec{\beta}^*$ до нуля, дістаємо:

$$\frac{\partial\vec{\varepsilon}'\cdot\vec{\varepsilon}}{\partial\vec{\beta}^*}=\frac{\partial}{\partial\vec{\beta}^*}\left((\vec{y})'\vec{y}-2\left(\vec{\beta}^*\right)'X'\vec{y}+\left(\vec{\beta}^*\right)'X'X\vec{\beta}^*\right)=$$

$$=-2\frac{\partial}{\partial\vec{\beta}^*}\left(\left(\vec{\beta}^*\right)'X'\vec{y}\right)+\frac{\partial}{\partial\vec{\beta}^*}\left(\left(\vec{\beta}^*\right)'X'X\vec{\beta}^*\right)=$$

$$=-2X'\vec{y}+X'X\vec{\beta}^*+\left(\left(\vec{\beta}^*\right)'X'X\right)'=$$

$$=-2X'\vec{y}+X'X\vec{\beta}^*+X'X\vec{\beta}^*=0\rightarrow X'X\vec{\beta}^*=X'\vec{y}\rightarrow$$

$$\rightarrow\vec{\beta}^*=(X'X)^{-1}X\vec{y}. \tag{547}$$

Довірчий інтервал для множинної лінійної регресії

Матриця X містить m лінійно незалежних векторів-стовпців, а це означає, що ранг її дорівнюватиме m і визначник $|X'X| \neq 0$. Отже, матриця $X'X$ має обернену.

Дисперсії статистичних оцінок $\beta_0^*, \beta_1^*, \beta_2^*, \ldots \beta_m^*$ визначають з допомогою кореляційної матриці для вектора $\vec{\beta}^*$.

$$K(\vec{\beta}^*) = M(\vec{\beta}^* - \vec{\beta})(\vec{\beta}^* - \vec{\beta})'.$$

Оскільки $(X'X)^{-1}(X'X) = E$, то, скориставшись (545), (547), дістанемо

$$\vec{\beta}^* = \vec{\beta} + (X'X)^{-1} X \vec{\varepsilon}. \qquad (548)$$

Тоді маємо

$$\vec{\beta}^* - \vec{\beta} = (X'X)^{-1} X'\vec{\varepsilon}, \qquad (549)$$

$$(\vec{\beta}^* - \vec{\beta})' = \left((X'X)^{-1} X'\vec{\varepsilon}\right)' = (\vec{\varepsilon})' X (X'X)^{-1}. \qquad (550)$$

Скориставшись (549), (550), дістанемо

$$K(\vec{\beta}^*) = M(\vec{\beta}^* - \vec{\beta})(\vec{\beta}^* - \vec{\beta})' = M\left((X'X)^{-1} X'\vec{\varepsilon}\vec{\varepsilon} X X (X'X)^{-1}\right) =$$

$$= M(\vec{\varepsilon}\vec{\varepsilon}')(X'X)^{-1} X'X (X'X)^{-1} = M(\vec{\varepsilon}\vec{\varepsilon}')(X'X)^{-1} =$$

$$= M\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_m \end{pmatrix} (\varepsilon_1 \quad \varepsilon_2 \quad \ldots \quad \varepsilon_m)(X'X)^{-1} =$$

$$= M\begin{pmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \varepsilon_1\varepsilon_3 & \ldots & \varepsilon_1\varepsilon_m \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \varepsilon_2\varepsilon_3 & \ldots & \varepsilon_2\varepsilon_m \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \varepsilon_m\varepsilon_1 & \varepsilon_m\varepsilon_2 & \varepsilon_m\varepsilon_3 & \ldots & \varepsilon_m^2 \end{pmatrix} (X'X)^{-1} =$$

$$= \begin{pmatrix} M(\varepsilon_1^2) & M(\varepsilon_1\varepsilon_2) & M(\varepsilon_1\varepsilon_3) & \ldots & M(\varepsilon_1\varepsilon_m) \\ M(\varepsilon_2\varepsilon_1) & M(\varepsilon_2^2) & M(\varepsilon_2\varepsilon_3) & \ldots & M(\varepsilon_2\varepsilon_m) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ M(\varepsilon_m\varepsilon_1) & M(\varepsilon_m\varepsilon_2) & M(\varepsilon_m\varepsilon_3) & \ldots & M(\varepsilon_m)^2 \end{pmatrix} (X'X)^{-1} =$$

$$=\begin{pmatrix} \sigma_\varepsilon^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_\varepsilon^2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \sigma_\varepsilon^2 \end{pmatrix}(X'X)^{-1}=\left|\begin{array}{c} \text{оскільки } M(\varepsilon_i\varepsilon_j)=K_{ij}=0, \\ M(\varepsilon_i^2)=D(\varepsilon_i^2)=\sigma_\varepsilon^2 \end{array}\right|=$$

$$=\sigma_\varepsilon^2\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}(X'X)^{-1}=\sigma_\varepsilon^2\cdot I\cdot(X'\cdot X)^{-1}=\sigma_\varepsilon^2(X'X)^{-1}.$$

Таким чином, маємо

$$K\left(\vec{\beta}^*\right)=\sigma_\varepsilon^2(X'X)^{-1}. \qquad (551)$$

Оскільки σ_ε^2 є невідомою величиною, то в (551) замість σ_ε^2 підставляють його точкову незміщену статистичну оцінку за аналогією з (513).

$$S_\varepsilon^2=\frac{\sum\left(\varepsilon_i^*\right)^2}{n-m-1}, \qquad (552)$$

де n є кількістю спостережень, а m — кількістю оцінюваних параметрів множинної лінійної регресії.

Значення дисперсії для β_i^* $(i=0,1,2,3,\dots m)$ обчислюють за формулою

$$S_{\beta_i^*}^2=S_\varepsilon^2\,C_{ii}, \qquad (553)$$

де C_{ii} — діагональний елемент матриці $(X'X)^{-1}$.

Розглянемо рівняння лінійної множинної регресії з параметрами β_i^* $(i=0,1,2,3,...,m)$, знайденими за результатами вибірки

$$y_i^*=\beta_0^*+\beta_1^*x_1+\beta_2^*x_2+...+\beta_m^*x_m,$$

де y_i^* — лише одне з можливих значень прогнозної величини для заданих значень x_i.

Ураховуючи те, що β_i^* $(i=0,1,2,3,\dots m)$ є випадковими величинами, то y_i^* буде також випадковою, а тому матиме дисперсію.

Отже,

$$D\left(y_i^*\right)=D\left(\beta_0^*+\beta_1^*x_1+\beta_2^*x_2+...+\beta_m^*x_m\right).$$

Використовуючи властивості дисперсії від суми залежних випадкових величин (випадкові величини β_i є залежними), дістанемо:

$$D\left(y_i^*\right) = D\left(\beta_0^* + \beta_1^* x_1 + \beta_2^* x_2 + \ldots + \beta_m^* x_m\right) =$$
$$D\left(\beta_0^*\right) + x_1^2 D\left(\beta_1^*\right) + x_2^2 D\left(\beta_2^*\right) + \ldots + x_m^2 D\left(\beta_m^*\right) + 2x_1 K\left(\beta_0^* \beta_2^*\right) + \ldots$$
$$\ldots + 2x_m K\left(\beta_0^* \beta_m^*\right) + 2x_1 x_2 K\left(\beta_1^* \beta_2^*\right) + \ldots + 2x_1 x_m K\left(\beta_1^* \beta_m^*\right) + \ldots$$
$$\ldots + 2x_{m-1} x_m K\left(\beta_{m-1}^* \beta_m^*\right) = \vec{x}' K\left(\vec{\beta}^*\right)\vec{x},$$

оскільки це є квадратична форма, яку можна записати у векторно-матричній формі.

Отже, маємо

$$D\left(y_i^*\right) = \vec{x}' K\left(\vec{\beta}^*\right)\vec{x}. \qquad (554)$$

Тоді, використовуючи (551), дістанемо

$$D\left(y_i^*\right) = \sigma_\varepsilon^2 \ \vec{x}'(X'X)^{-1}\vec{x}. \qquad (555)$$

Оскільки σ_ε^2 — невідома величина, то в (555) використовуємо її точкову незміщену статистичну оцінку S_ε^2.

Таким чином, маємо:

$$D\left(y_i^*\right) = S_\varepsilon^2 \cdot \vec{x}'(X'X)^{-1}\vec{x}. \qquad (556)$$

Отже, істинне значення *Y* перебуватиме в інтервалі:

$$y^* - t(\gamma,k)S_\varepsilon\sqrt{\vec{x}'(X'X)^{-1}\vec{x}} < y < y^* + t(\gamma,k)S_\varepsilon\sqrt{\vec{x}'(X'X)^{-1}\vec{x}}, \qquad (557)$$

який називають довірчим.

$t(\gamma, k)$ є випадковою величиною, що має розподіл Стьюдента з $k = n - m - 1$ ступенями свободи і обчислюється за таблицею (додаток 7) за заданою надійністю γ та числом ступенів свободи *k*.

Якщо до значень y_i — прогнозне значення — додати можливі відхилення ознаки *Y* від функції регресії, то до дисперсії $D\left(y^*\right)$ необхідно додати дисперсію випадкової величини ε_i — σ_ε^2, тобто його точкову незміщену статистичну оцінку S_ε^2.

У цьому разі

$$S_y^2 = S_\varepsilon^2(1 + \vec{x}'(X'X))^{-1}\vec{x}. \qquad (558)$$

І довірчий інтервал тепер дорівнюватиме:

$$y^* - t(\gamma, k)S_y < y < y^* + t(\gamma, k)S_y. \qquad (559)$$

Коефіцієнт множинної регресії

Тісноту між ознаками *Y* та *X*, де $X = (x_1, x_2, ... x_m)$, вимірюють з допомогою коефіцієнта множинної кореляції *R*, що є узагальненням парного коефіцієнта кореляції r_{ij} і обчислюється за формулою

$$R = \sqrt{1 - \frac{\sum \varepsilon_i^2}{\sum (y_i - \bar{y})^2}}\ . \qquad (560)$$

Чим ближче значення *R* до ±1, тим краще вибрано функцію регресії

$$y = \alpha(x_1,\ x_2, ..., x_m).$$

Оскільки $\sum \varepsilon_i^2 = \vec{\varepsilon}'\,\vec{\varepsilon}$, то

$$\sum \varepsilon_i^2 = \vec{\varepsilon}'\,\vec{\varepsilon} = {}'\left(\vec{y} - X\ \vec{\beta}^*\right)\left(\vec{y} - X\ \vec{\beta}^*\right) =$$

$$= (\vec{y})'\ \vec{y} - 2\left(\vec{\beta}^*\right)'\ X'\ \vec{y} + \left(\vec{\beta}^*\right)'\ X'\,X\,\vec{\beta}^* =$$

$$= (\vec{y})'\ \vec{y} - 2\left(\vec{\beta}^*\right)'\ X'\ \vec{y} + \left(\vec{\beta}^*\right)'\ X'\,\vec{y} =$$

$$= (\vec{y})'\ \vec{y} - \left(\vec{\beta}^*\right)'\ X'\ \vec{y},$$

оскільки $\left(\vec{\beta}^*\right)'\ X'\,X\,\vec{\beta}^* = \left(\vec{\beta}^*\right)'\ X'\,\vec{y}^*$.

При цьому $\sum (y_i - \bar{y})^2 = \sum y_i^2 - n(\bar{y})^2$, а оскільки $\sum (y_i)^2 = (\vec{y})'\ \vec{y}$, то остаточно маємо

$$R = \sqrt{1 - \frac{\left(\vec{y}^*\right)'\ \vec{y} - \left(\vec{\beta}^*\right)'\ X'\,\vec{y}}{(\vec{y})'\ \vec{y} - n(\bar{y})^2}}\ . \qquad (561)$$

Нормування коефіцієнтів регресії

Множинна лінійна регресія дає змогу порівняти вплив на досліджуваний процес різних чинників. У загальному випадку змінні x_i репрезентують чинники, що мають різні одиниці виміру (кілограми, гривні, метри тощо). Отже, для того щоб порівняти і з'ясувати відносну вагомість кожного з чинників, використовують так звані нормовані коефіцієнти регресії, які визначають за формулою

$$a_j^* = \beta_j^* \frac{S_{x_j}}{S_y}\ \left(j = \overline{1, m}\right), \qquad (562)$$

де a_j — коефіцієнт регресії після нормування; S_{x_j} — виправлене середнє квадратичне відхилення змінної x_j; S_y — виправлене середнє квадратичне відхилення ознаки *Y*.

Приклад 1. Ознака *Y* — лінійно залежна від x_{i1}, x_{i2}, x_{i3}. Результати спостережень наведено в таблиці:

i	y_i	x_{i1}	x_{i2}	x_{i3}
1	6	1	1	2
2	8	2	2	1
3	14	1	0	0
4	20	3	2	1
5	26	5	2	2

Необхідно:

1) знайти компоненти вектора і побудувати лінійну

$$\beta^* = \begin{pmatrix} \beta_0^* \\ \beta_1^* \\ \beta_2^* \\ \beta_3^* \end{pmatrix}$$

функцію регресії $y_i = \beta_0^* + \beta_1^* x_{i1} + \beta_2^* x_{i2} + ... + \beta_3^* x_{i3}$;

2) обчислити *R*;

3) побудувати довірчий інтервал із надійністю $\gamma = 0{,}95$ для множинної лінійної функції регресії та визначити дисперсії для $\beta_0^*, \beta_1^*, \beta_2^*, \beta_3^*$ і оцінити ефективність впливу на ознаку *Y* незалежних змінних x_{i1}, x_{i2}, x_{i3}.

Розв'язання. 1. З умови задачі маємо:

$$X = \begin{pmatrix} 1 & 1 & 1 & 2 \\ 1 & 2 & 2 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 3 & 2 & 1 \\ 1 & 5 & 2 & 2 \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} 6 \\ 8 \\ 14 \\ 20 \\ 26 \end{pmatrix}.$$

Оскільки

$$\beta^* = \begin{pmatrix} \beta_0^* \\ \beta_1^* \\ \beta_2^* \\ \beta_3^* \end{pmatrix} = (X'X)^{-1} X'\vec{y} =$$

$$=\frac{1}{178}\begin{pmatrix}173 & -14 & -39 & -41\\ -14 & 32 & -38 & -8\\ -39 & -38 & 123 & -35\\ -41 & -8 & -35 & 91\end{pmatrix}\begin{pmatrix}1 & 1 & 1 & 1 & 1\\ 1 & 2 & 1 & 3 & 5\\ 1 & 2 & 0 & 2 & 2\\ 2 & 1 & 0 & 1 & 2\end{pmatrix}\begin{pmatrix}6\\ 8\\ 14\\ 20\\ 26\end{pmatrix}=\begin{pmatrix}7{,}98\\ 6{,}34\\ -3{,}78\\ -2{,}58\end{pmatrix}.$$

Отже, дістали: $\beta_0^*=7{,}98$; $\beta_1^*=6{,}34$; $\beta_2^*=-3{,}78$; $\beta_3^*=-2{,}58$.
Рівнянням регресії буде

$$y_i^*=7{,}98+6{,}34x_{i1}-3{,}78x_{i2}-2{,}58x_{i3}.$$

2. Знайдемо R. Для цього необхідно визначити

$$\left(\vec{\beta}^*\right)' X'\vec{y}=(7{,}98\quad 6{,}34\quad -3{,}78\quad -2{,}58)\begin{pmatrix}1 & 1 & 1 & 1 & 1\\ 1 & 2 & 1 & 3 & 5\\ 1 & 2 & 0 & 2 & 2\\ 2 & 1 & 0 & 1 & 2\end{pmatrix}\begin{pmatrix}6\\ 8\\ 14\\ 20\\ 26\end{pmatrix}=1354{,}38;$$

$$\vec{y}=\frac{\sum y_i}{n}=\frac{6+8+14+20+26}{5}=14{,}8;\quad n\left(\vec{y}\right)^2=5\left(14{,}8\right)^2=1095{,}2;$$

$$(\vec{y}')\,\vec{y}-n(\vec{y})^2=1372-1095{,}2=276{,}8;$$

$$\left(\vec{\beta}^*\right)' X'\,\vec{y}-n(\vec{y})^2=1354{,}38-1095{,}2=259{,}18.$$

Тоді

$$R=\sqrt{1-\frac{(\vec{y})'\,\vec{y}-\left(\vec{\beta}^*\right)' X'\vec{y}}{(\vec{y})'\,\vec{y}-n(\vec{y})^2}}=\sqrt{\frac{\left(\vec{\beta}^*\right)' X'\vec{y}-n(\vec{y})^2}{(\vec{y})'\,\vec{y}-n(\vec{y})^2}}=$$

$$=\sqrt{\frac{259{,}18}{276{,}8}}=0{,}968.$$

Для побудови довірчого інтервалу для множинної лінійної функції регресії необхідно обчислити S_ε. Оскільки $S_\varepsilon=\sqrt{\frac{\sum\varepsilon_i^2}{n-m-1}}$, то в цьому разі результати обчислень зручно подати у вигляді таблиці:

i	y_i	x_{i1}	x_{i2}	x_{i3}	$y_i^* = 7{,}98 + 6{,}34x_{i1} - 3{,}78x_{i2} - 2{,}58x_{i3}$	$y_i - y_i^*$	$(\varepsilon_i^*)^2$
1	6	1	1	2	5,38	0,62	0,3844
2	8	2	2	1	10,52	–2,52	6,3504
3	14	1	0	0	14,32	–0,32	0,1024
4	20	3	2	1	16,86	3,14	9,8596
5	26	5	2	2	26,96	–0,96	0,9216
						$\sum \varepsilon_i^2 = 17{,}618$	

Таким чином, дістанемо:

$$S_\varepsilon^2 = \frac{\sum (\varepsilon_i^*)^2}{n-m-1} = \frac{17{,}618}{5-3-1} = 17{,}618.$$

Візьмемо $x_1 = 2;\ x_2 = 6;\ x_3 = 10$ і обчислимо

$$y_i = 7{,}98 + 6{,}34 \cdot 2 - 3{,}78 \cdot 6 - 2{,}58 \cdot 10 = -27{,}82.$$

Знайдемо

$$D(y_i^*) = S_\varepsilon^2\, \vec{x}(X'X)^{-1}\vec{x} =$$

$$= 17{,}618 \cdot \begin{pmatrix} 1 & 2 & 6 & 10 \end{pmatrix} \frac{1}{178} \begin{pmatrix} 173 & -14 & -39 & -41 \\ -14 & 32 & -38 & -8 \\ -39 & -38 & 123 & -35 \\ -43 & -8 & -35 & 91 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 6 \\ 10 \end{pmatrix} = 683{,}992.$$

$$S_{y^*} = \sqrt{683{,}992} \approx 26{,}15.$$

Для побудови довірчого інтервалу знаходимо

$$t(\gamma = 0{,}95,\ k = n-m-1) = t(\gamma = 0{,}95,\ k = 5-3-1) =$$

$$= t(\gamma = 0{,}95,\ k = 1) = 12{,}706.$$

Тоді $t(\gamma, k) S_\varepsilon \sqrt{\vec{x}'(X' \cdot X)^{-1}\vec{x}} = 12{,}706 \cdot 26{,}15 = 332{,}262.$

І довірчий інтервал дорівнюватиме

$$y_i = y_i^* \pm t(\gamma, k) \cdot S_\varepsilon \sqrt{\vec{x}'(X' \cdot X)^{-1}\vec{x}} \rightarrow$$

$$\rightarrow -360{,}28 < y_i < 304{,}242.$$

Оскільки діагональні елементи матриці $(X'X)^{-1}$ відповідно дорівнюють

$$b_{11}=\frac{173}{178};\quad b_{22}=\frac{32}{178};\quad b_{33}=\frac{123}{178};\quad b_{44}=\frac{91}{178},$$

то відповідно дістанемо

$$S^2_{\beta_0^*}=S^2_\varepsilon b_{11}=17{,}618\cdot\frac{173}{178}=17{,}123,\ S_{\beta_0^*}=4{,}138;$$

$$S^2_{\beta_1^*}=S^2_\varepsilon b_{22}=17{,}618\cdot\frac{32}{178}=3{,}167,\ S_{\beta_1^*}=1{,}78;$$

$$S^2_{\beta_2^*}=S^2_\varepsilon b_{33}=17{,}618\cdot\frac{123}{178}=12{,}17,\ S_{\beta_2^*}=3{,}489;$$

$$S^2_{\beta_3^*}=S^2_\varepsilon b_{44}=17{,}618\cdot\frac{91}{178}=9{,}007,\ S_{\beta_3^*}=3{,}001.$$

Обчислимо

$$S_y=\sqrt{\frac{\vec{y}'\vec{y}}{n}-(\bar{y})^2}=\sqrt{53{,}36}\approx 7{,}44.$$

Визначимо нормовані коефіцієнти регресії:

$$a_1=\beta_1^*\frac{S_{\beta_1^*}}{S_y}=6{,}34\cdot\frac{1{,}78}{7{,}44}=1{,}52,$$

$$a_2=\beta_2^*\frac{S_{\beta_2^*}}{S_y}=-3{,}78\cdot\frac{3{,}489}{7{,}44}=-1{,}77,$$

$$a_3=\beta_3^*\frac{S_{\beta_3^*}}{S_y}=-2{,}58\cdot\frac{3{,}001}{7{,}44}=-1{,}04.$$

Отже, для змінної x_{12} вплив на ознаку Y є найефективнішим порівняно з дією змінних x_{i1}, x_{i3}.

4. Нелінійна регресія

Якщо в рівняння множинної регресії змінні x_{ij} входять як x_{ij}^n, то регресія називається *нелінійною*.

У загальному випадку нелінійна регресія записується в такому вигляді:

$$y_i=\beta_0+\beta_1x_{1i}+\beta_2x_2^2+\beta_3x_3^3+\ldots+\beta_mx_m^m+\varepsilon_i, \qquad (569)$$

де параметри β_0, β_1, β_2, β_3, ... β_m є сталими невідомими величинами, які підлягають статистичним оцінкам, а ε_i — випадкова величина, яка має нормальний закон розподілу з числовими характеристиками $M(\varepsilon_i)=0$, $D(\varepsilon_i)=M\left(\varepsilon_i^2\right)=\sigma_\varepsilon^2$, і при цьому випадкові величини ε_1, ε_2, ... ε_m між собою не корельовані. Реалізуючи вибірку обсягом *n*, згідно з (563), дістанемо систему нелінійних рівнянь виду:

$$\begin{aligned}
y_1 &= \beta_0^* + \beta_1^* x_{11} + \beta_2^* x_{12}^2 + \beta_3^* x_{13}^2 + \ldots + \beta_m^* x_{1m}^m + \varepsilon_1^*;\\
y_2 &= \beta_0^* + \beta_1^* x_{21} + \beta_2^* x_{22}^2 + \beta_3^* x_{23}^2 + \ldots + \beta_m^* x_{2m}^m + \varepsilon_2^*;\\
y_3 &= \beta_0^* + \beta_1^* x_{31} + \beta_2^* x_{32}^2 + \beta_3^* x_{33}^2 + \ldots + \beta_m^* x_{3m}^m + \varepsilon_3^*;\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
y_n &= \beta_0^* + \beta_1^* x_{n1} + \beta_2^* x_{n2}^2 + \beta_3^* x_{n3}^2 + \ldots + \beta_m^* x_{nm}^m + \varepsilon_n^*.
\end{aligned} \tag{564}$$

Систему (564) можна подати у векторно-матричній формі так:

$$\vec{y}^* = X\,\vec{\beta}^* + \vec{\varepsilon}, \tag{565}$$

де

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \ldots \\ y_n \end{pmatrix},\quad X = \begin{pmatrix} 1 & x_{11} & x_{12}^2 & x_{22}^2 & \ldots & x_{1m}^m \\ 1 & x_{21} & x_{22}^2 & x_{22}^2 & \ldots & x_{2m}^m \\ 1 & x_{31} & x_{32}^2 & x_{33}^2 & \ldots & x_{2m}^m \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1}^1 & x_{n2}^2 & x_{n3}^3 & \ldots & x_{nm}^m \end{pmatrix},\quad \vec{\beta}^* = \begin{pmatrix} \beta_0^* \\ \beta_1^* \\ \beta_2^* \\ \beta_3^* \\ \ldots \\ \beta_m^* \end{pmatrix},\quad \vec{\varepsilon}^* = \begin{pmatrix} \varepsilon_1^* \\ \varepsilon_2^* \\ \varepsilon_3^* \\ \ldots \\ \varepsilon_n^* \end{pmatrix}.$$

Тут $\beta_0, \beta_1, \beta_2, \beta_3, \ldots \beta_m$ є випадковими величинами, які визначаються шляхом обробки результатів вибірки і є точковими незміщеними статистичними оцінками відповідних параметрів рівняння (563), а саме: $\beta_0, \beta_1, \beta_2, \beta_3, \ldots \beta_m$.

Здійснивши аналогічні перетворення, які були зроблені для лінійної множинної регресії, дістанемо:

$$\vec{\beta}^* = (X'\,X)^{-1}\,X'\,\vec{y}\,. \tag{566}$$

Тіснота зв’язку вимірюється з допомогою кореляційного відношення

$$\eta = \sqrt{1 - \frac{\sum \varepsilon_i^*}{\sum (y_i - \vec{y})^2}}\,, \tag{567}$$

$$0 \le \eta \le 1\,.$$

Приклад 2. Результати спостережень над ознаками X і Y наведено у таблиці:

i	x_i	y_i
1	1	8
2	2	4
3	4	2
4	6	1
5	8	0
6	10	6
7	12	8
8	14	10

Потрібно:

1) визначити точкові незміщені статистичні оцінки для параметрів нелінійної регресії

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x^2;$$

2) обчислити η.

Розв'язання. З результатів вибірки маємо:

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \\ 1 & 6 & 36 \\ 1 & 8 & 64 \\ 1 & 10 & 100 \\ 1 & 12 & 144 \\ 1 & 14 & 196 \end{pmatrix}; \quad \vec{y} = \begin{pmatrix} 8 \\ 4 \\ 2 \\ 1 \\ 0 \\ 6 \\ 8 \\ 10 \end{pmatrix}.$$

Використовуючи (566), дістанемо:

$$\vec{\beta}^* = \begin{pmatrix} \beta_0^* \\ \beta_1^* \\ \beta_2^* \end{pmatrix} = \begin{pmatrix} \frac{291499}{280301} & \frac{-82843}{280301} & \frac{680}{40043} \\ \frac{-82843}{280301} & \frac{94613}{840903} & \frac{-289}{40043} \\ \frac{680}{40043} & \frac{-289}{40043} & \frac{59}{120129} \end{pmatrix} \times$$

$$\times\begin{pmatrix}1 & 1 & 1 & 1 & 1 & 1 & 1 & 1\\ 1 & 2 & 4 & 6 & 6 & 10 & 12 & 14\\ 1 & 4 & 16 & 36 & 64 & 100 & 144 & 196\end{pmatrix}\begin{pmatrix}8\\4\\2\\1\\0\\6\\8\\10\end{pmatrix}=\begin{pmatrix}8{,}807\\-2{,}301\\0{,}178\end{pmatrix}.$$

Таким чином, маємо:

$$\beta_0^* = 8{,}807;\ \beta_1^* = -2{,}301;\ \beta_2^* = 0{,}178.$$

Для визначення η застосовуємо табличний запис:

i	x_i	y_i	$y_i^* = 8{,}807 - 2{,}301x_i + 0{,}178x_i^2$	$(\varepsilon_i^*)^2 = (y_i - y_i^*)^2$
1	1	8	6,684	1,732
2	2	4	4,917	0,841
3	4	2	2,451	0,203
4	6	1	1,409	0,167
5	8	0	1,791	3,208
6	10	6	3,597	5,774
7	12	8	6,827	1,376
8	14	10	11,481	2,193
		39		15,494

Отже, дістали $\sum(\varepsilon_i^*)^2 = \sum(y_i - y_i^*)^2 = 15{,}494.$

Оскільки

$$\bar{y} = \frac{\sum y_i}{n} = \frac{39}{8} = 4{,}875,$$

то

$$\sum(y_i - \bar{y})^2 = \sum(y_i - 4{,}875)^2 = 94{,}875,$$

$$\eta = \sqrt{1 - \frac{\sum(\varepsilon_i^*)^2}{\sum(y_i - \bar{y})^2}} = \sqrt{1 - \frac{15{,}494}{94{,}875}} = \sqrt{1 - 0{,}426} = \sqrt{0{,}574} \approx 0{,}76.$$

Приклад 3. За результатами спостережень ознак генеральної сукупності X і Y:

i	x_i	y_i
1	1	30
2	2	20
3	4	10
4	5	8
5	8	6
6	10	1

Знайти точкові незміщені статистичні оцінки для параметрів β_0, β_1 рівняння нелінійної регресії

$$y_i = \beta_0 + \frac{\beta_1}{x_i}.$$

Обчислити η.

Розв'язання. За результатами вибірки маємо:

$$X = \begin{pmatrix} 1 & 1 \\ 1 & \frac{1}{x_1} \\ 1 & \frac{1}{x_2} \\ 1 & \frac{1}{x_3} \\ 1 & \frac{1}{x_4} \\ 1 & \frac{1}{x_5} \\ 1 & \frac{1}{x_6} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0{,}5 \\ 1 & 0{,}25 \\ 1 & 0{,}2 \\ 1 & 0{,}125 \\ 1 & 0{,}1 \end{pmatrix}; \ \vec{y} = \begin{pmatrix} 30 \\ 20 \\ 10 \\ 8 \\ 6 \\ 1 \end{pmatrix}.$$

З рівняння (566) знаходимо:

$$\vec{\beta}^* = \begin{pmatrix} \beta_0^* \\ \beta_1^* \end{pmatrix} = (X'X)^{-1} X' \vec{y}^* =$$

$$=\begin{pmatrix} 0{,}38955440121559065691 & -0{,}61486271599703169723 \\ -0{,}61486271599703169723 & 1{,}6961730096469839924 \end{pmatrix}=$$

$$=\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0{,}5 & 0{,}25 & 0{,}2 & 0{,}125 & 0{,}1 \end{pmatrix}\begin{pmatrix} 30 \\ 20 \\ 10 \\ 8 \\ 6 \\ 1 \end{pmatrix}=\begin{pmatrix} 1{,}579 \\ 30{,}128 \end{pmatrix}.$$

Отже, маємо:

$$y_i = 1{,}579 + \frac{30{,}128}{x_i}.$$

Для обчислення η застосовуємо таблицю:

i	x_i	y_i	$y_i^* = 1{,}579 + \frac{30{,}128}{x_i}$	$(\varepsilon_i^*)^2 = (y_i - y_i^*)^2$
1	1	30	31,707	2,914
2	2	20	16,643	11,269
3	4	10	9,111	0,790
4	5	8	7,6046	0,156
5	8	6	5,345	0,429
6	10	1	4,5918	12,901
Σ		75		28,459

Таким чином, дістали:

$$\sum(\varepsilon_i^*)^2 = \sum(y_i - y_i^*)^2 = 28{,}459.$$

Оскільки $\bar{y} = \frac{\sum y_i}{n} = \frac{75}{6} = 12{,}5$, $\sum(y_i - \bar{y})^2 = \sum(y_i - 12{,}5)^2 = 562{,}5$,

то

$$\eta = \sqrt{1 - \frac{\sum(\varepsilon_i^*)^2}{\sum(y_i - \bar{y})^2}} = \sqrt{1 - \frac{28{,}459}{562{,}5}} = \sqrt{1 - 0{,}0506} = \sqrt{0{,}9494} \approx 0{,}974.$$

Отже, $\eta \approx 0{,}976$.

5. Нелінійна модель за параметрами

В економічному аналізі розглядають нелінійну регресію за параметрами, що подається в такому найпростішому вигляді:

$$y_i = \beta_0 \, x_{i1}^{\beta_1} \, x_{i2}^{\beta_2} \, . \qquad (568)$$

Такі функції регресії використовують для вимірювання впливу на обсяг виробництва таких чинників, як кількість зайнятих у виробництві робітників, обсяг основних фондів тощо.

У рівнянні β_0, β_1 є невідомими величинами, але сталими, які оцінюються точковими незміщеними статистичними оцінками β_0^*, β_1^*, котрі визначаються обробкою результатів вибірки.

Для врахування впливу випадкових збудників, які відхиляють теоретично прогнозовану регресію, вводиться випадкова величина e^{ε_i}. Тоді нелінійна модель відносно параметрів β_0, β_1 набуває такого вигляду:

$$y_i = \beta_0 \, x_{i1}^{\beta_1} \, x_{i2}^{\beta_2} \, e^{\varepsilon_i} \, . \qquad (569)$$

Статистичною оцінкою рівняння (568) буде

$$y_i^* = \beta_0^* \, x_{i1}^{\beta_1^{-*}} \, x_{i2}^{\beta_2^*} \, e^{\varepsilon_i} \, . \qquad (570)$$

Для визначення точкових незміщених статистичних оцінок β_0^*, β_1^*, β_2^* використовуємо, як і в попередніх моделях, метод найменших квадратів, а для цього рівняння (570) подано в такому вигляді:

$$\ln y_i = \beta_0^* + \beta_1^* \ln x_{i1} + \beta_2^* \ln x_{i2} + \varepsilon_i \, . \qquad (571)$$

Здійснивши вибірку обсягу *n*, дістанемо систему рівнянь, яку у векторно-матричній формі можна записати так:

$$\vec{y}^* = X \, \vec{\beta}^* + \vec{\varepsilon}, \qquad (572)$$

де

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & \ln x_{11} & \ln x_{12} \\ 1 & \ln x_{21} & \ln x_{22} \\ 1 & \ln x_{31} & \ln x_{32} \\ \dots & \dots & \dots \\ 1 & \ln x_{n1} & \ln x_{n2} \end{pmatrix}, \quad \vec{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \dots \\ \varepsilon_n \end{pmatrix}.$$

Тоді, за аналогією з попередніми випадками, компоненти вектора $\vec{\beta}^*$ визначають із рівняння

Комп'ютерна реалізація прикладу № 1.

$$\vec{\beta}^* = (X' X)^{-1} X' \vec{y}^*. \qquad (573)$$

Теоретичні запитання до теми ?

1. Дати визначення статистичної залежності між ознаками *X* та *Y*.
2. Що означає кореляційна залежність між ознаками *X* та *Y*?
3. Записати модель парної лінійної регресії?
4. Чому дорівнює β_0^*?
5. Чому дорівнює β_1^*?
6. Які числові характеристики для β_0^*?
7. Які числові характеристики для β_1^*?
8. Який закон розподілу ймовірностей мають випадкові величини β_0^*, β_1^* для парної лінійної регресії?
9. Чому дорівнює $K_{\beta_0^*, \beta_1^*}$?
10. Який закон розподілу ймовірності мають випадкові величини $\beta_0^* + \beta_1^* x_i$?
11. Який закон розподілу має випадкова величина $\dfrac{\beta_0^* - \beta_0}{\dfrac{S_\varepsilon}{\sqrt{\sum (x_i - \bar{x})^2}}}$?
12. Чому дорівнює S_ε?
13. Чому дорівнює довірчий інтервал для β_1^*?
14. Чому дорівнює довірчий інтервал для β_0^*?
15. Чому дорівнює довірчий інтервал для $y_i = \beta_0^* + \beta_1^* x_i$?
16. Чому дорівнює $\eta_{\beta_0^*, \beta_1^*}$?
17. Записати у векторно-матричній формі модель лінійної множинної регресії.
18. Чому дорівнює $\vec{\beta}^*$ інтервал для множинної регресії?
19. Чому дорівнює довірчий інтервал для множинної регресії?
20. Чому дорівнює довірчий інтервал S_ε для множинної регресії?
21. Чому дорівнює $D(y_i^*)$ для множинної регресії?
22. Чому дорівнює коефіцієнт множинної регресії?

23. Що розуміють під нормуванням коефіцієнта регресії?

24. Чому дорівнює вектор $\vec{\beta}^*$ для нелінійної множинної регресії?

25. Чому дорівнює матриця X для нелінійної множинної регресії?

26. Чому дорівнює нелінійна модель за параметрами β_0, β_1, β_2?

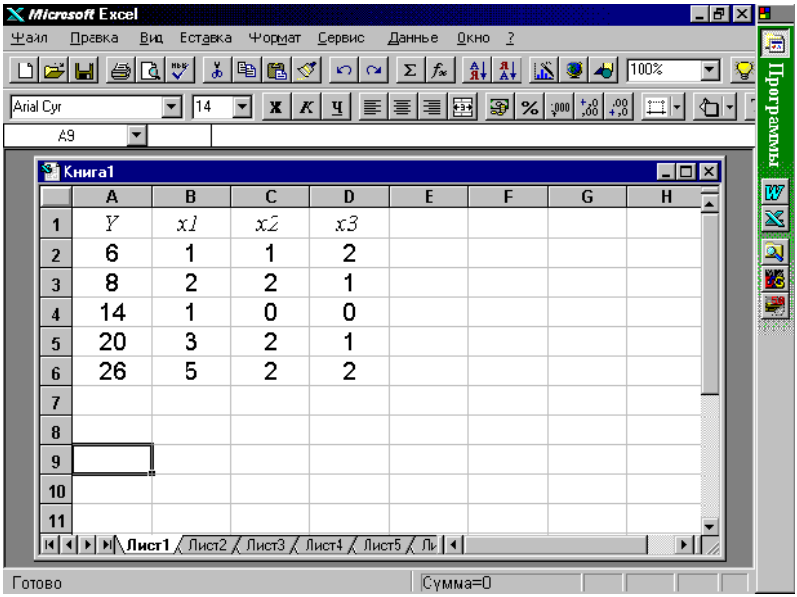
27. Чому дорівнює матриця X для нелінійної за параметрами β_0, β_1, β_2 регресії?

28. Чим вимірюється кореляційний зв'язок для нелінійних моделей регресії?

Додаток до теми 16

Усі розрахунки в прикладах 1—3 можна здійснити, застосовуючи табличний процесор Excel. Розглянемо порядок виконання обчислень для прикладу 1. Табличний процесор Excel пропонує функцію, яка знаходить значення оцінок параметрів лінійної залежності за методом найменших квадратів.

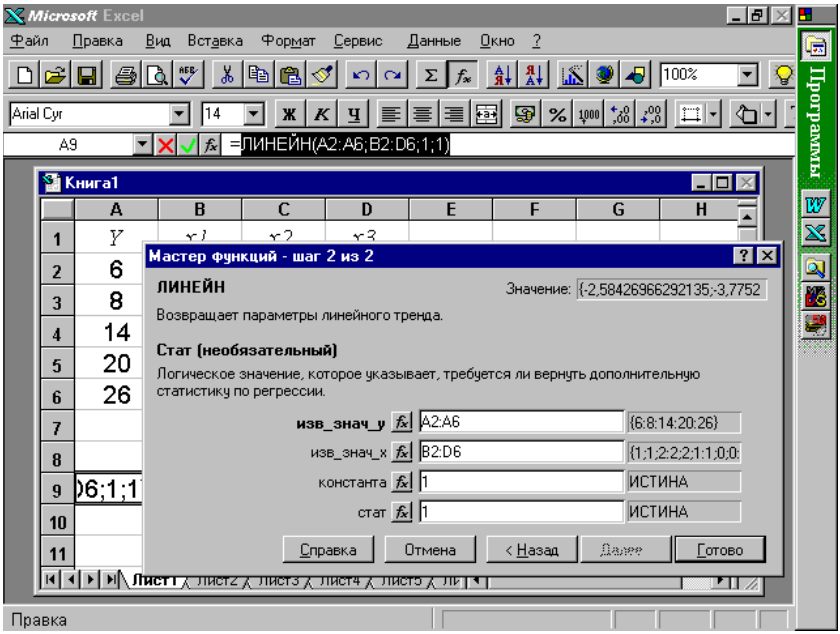
1. Нехай вихідні дані містяться в блоках: матриця X — (B2:D6) та вектор $\vec{Y}$ — (A2:A6). Зауважимо, що в даному разі не потрібно вводити вектор-стовпчик x_0, елементами якого є одиниці.



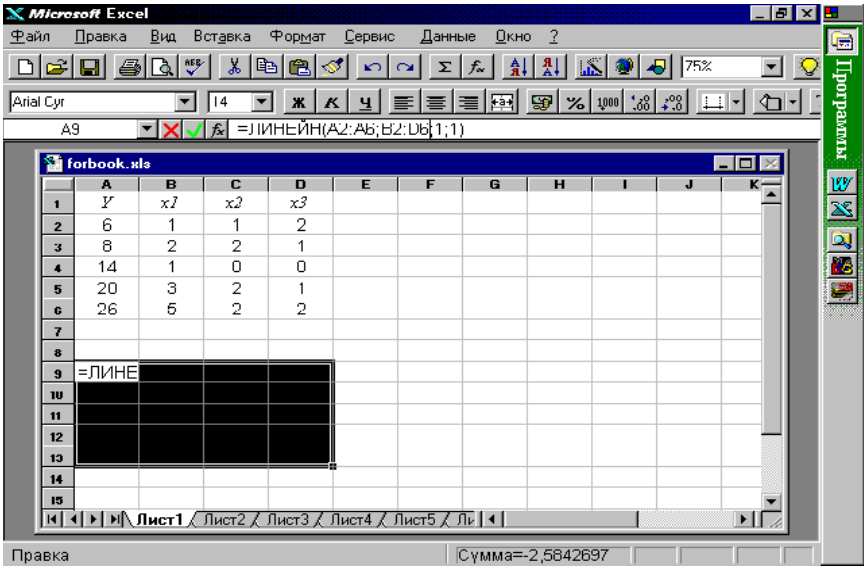
	A	B	C	D
1	Y	x1	x2	x3
2	6	1	1	2
3	8	2	2	1
4	14	1	0	0
5	20	3	2	1
6	26	5	2	2

Результат розрахунку, тобто оцінки параметрів β_0^*, β_1^*, β_2^*, β_3^* будемо знаходити в блоці (A9:D9). Для цього необхідно, установивши курсор у клітині **A9**, викликати «**Вставить функцию»**, і в категорії «**Статистические»** обрати функцію «**ЛИНЕЙН»**.

У вікні запиту необхідно вказати: у першому рядку — «відомі значення *y*», в нашому прикладі вони розташовані в блоці (A2:A6), у другому — «відомі значення *x*», це вся матриця *X*, що в нашому прикладі міститься в блоці (B2:D6), у третьому рядку «константа» вводиться логічне значення «істина» (відповідає числу 1), що вказує на необхідність розрахунку оцінки параметра β_0^*. В останньому рядку «статистика» також має бути логічне значення «істина» (число 1) у тому випадку, коли необхідна додаткова статистична інформація (стандартні похибки оцінок параметрів, коефіцієнт детермінації, залишкова сума квадратів відхилень тощо).

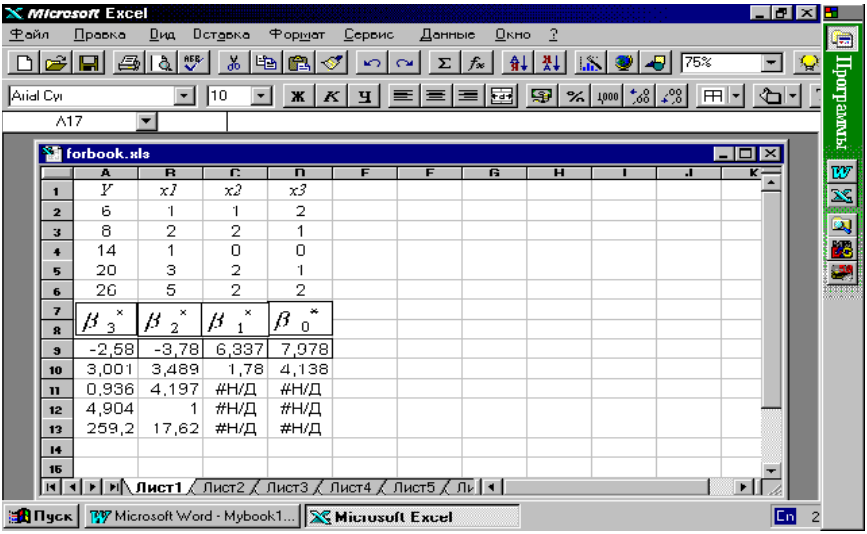


Функція **ЛИНЕЙН** повертає оцінки параметрів, починаючи з останнього, тобто в клітинці **A9** міститься значення оцінки параметра β_3^*. Для того щоб знайти значення всіх параметрів, необхідно, починаючи з клітинки **A9**, виділити блок розмірності $(5 \times (m + 1))$, де m — кількість змінних x_i. У нашому випадку $m = 3$, тому необхідно виділити блок розмірності (5×4) — (A9:D13).



Після цього натиснути клавіш **F2**, а далі комбінацію клавішів **Ctrl + Shift + Enter**.

У результаті у першому рядку блока (A9:D13) отримаємо значення всіх параметрів у зворотному порядку: у клітинці з адресою **A9** — значення оцінки параметра β_3^*, в **B9** — оцінка параметра β_2^*, в **C9** — β_1^* і в **D9** — оцінка параметра β_0^*.



За обчисленим значенням оцінок параметрів запишемо рівняння лінійної множинної регресії:

$$y_i^* = 7{,}978 + 6{,}337x_{1i} - 3{,}78x_{2i} - 2{,}58x_{3i}\,.$$

Додаткова регресійна статистика в масиві (A9:D13) подана в такому порядку:

β_3^*	β_2^*	β_1^*	β_0^*
$S_{\beta_3^*}$	$S_{\beta_2^*}$	$S_{\beta_1^*}$	$S_{\beta_0^*}$
R^2	Стандартна похибка		
F-критерій	Ступені свободи ($n - m - 1$)		
Сума квадратів відхилень, що пояснюється регресією	Сума квадратів відхилень, що пояснюється похибкою ε		

2. Знайдемо *R*.

Оскільки $R = \sqrt{R^2}$, а значення R^2 вказане в наведеній таблиці додаткової статистики за регресією в клітинці **A11**, то для розрахунку необхідно обчислити $\sqrt{0{,}936}$. Отримаємо $R = 0{,}968$.

3. Для побудови довірчого інтервалу для множинної лінійної функції регресії необхідно обчислити $S_\varepsilon = \sqrt{\dfrac{\sum \varepsilon_i^2}{n-m-1}}$, де $\sum \varepsilon_i^2$ — сума квадратів відхилення, що пояснюється похибкою *ε*. Дане значення вказане в клітинці **B13**. Тому $S_\varepsilon^2 = \dfrac{17{,}618}{n-m-1} = \dfrac{17{,}618}{5-3-1} = 17{,}618$.

Подальші обчислення за формулою $D(y_i^*) = S_\varepsilon^2\, \vec{x}(X'X)^{-1}\vec{x}$ здійснюються з допомогою таких функцій:

категорія — **Ссылки и массивы**:

ТРАНСП — повертає транспонований масив;

категорія — **Математические**:

МУМНОЖ — повертає добуток двох матриць;

МОБР — повертає обернену матрицю.

Дістанемо $S_{y^*} = \sqrt{683{,}992} \approx 26{,}15$. Довірчий інтервал перебуває в межах:

$$-360{,}28 < y_i < 304{,}242\,.$$

Значення $S_{\beta_0^*}$, $S_{\beta_1^*}$, $S_{\beta_2^*}$, $S_{\beta_3^*}$ містяться в таблиці з додатковою статистикою в другому рядку, тобто в масиві (A10:D10). Дістаємо:

$$S_{\beta_0^*} = 4{,}138; \quad S_{\beta_1^*} = 1{,}78; \quad S_{\beta_2^*} = 3{,}489; \quad S_{\beta_3^*} = 3{,}001.$$

Підставивши відомі значення у формулу $a_i = \beta_i^* \dfrac{S_{\beta_i^*}}{S_y}$, $i = \overline{1{,}3}$ дістанемо:

$$a_1 = \beta_1^* \frac{S_{\beta_1^*}}{S_y} = 6{,}34 \frac{1{,}78}{7{,}44} = 1{,}52\ ;$$

$$a_2 = \beta_2^* \frac{S_{\beta_2^*}}{S_y} = -3{,}78\ \frac{3{,}489}{7{,}44} = -1{,}77\ ;$$

$$a_3 = \beta_3^* \frac{S_{\beta_3^*}}{S_y} = -2{,}58\ \frac{3{,}001}{7{,}44} = -1{,}04\ .$$

ЛІТЕРАТУРА

1. *Brownlee K. A.* Statistical theory and methodology in science and engineering. — New York; London; Sydney, 1977.
2. *Бондарчук П. І., Скоробагатько В. Я.* Гіллясті ланцюгові дроби та їх застосування. — К.: Наук. думка, 1974.
3. *Ивченко Г. И., Медведев Ю. И.* Статистика. — М.: Высш. школа, 1984.
4. *Searl S. R., Hausman W. H.* Matrix algebra for business and economics. — New York; London; Sydney; Toronto, 1970.

Лабораторна робота № 3
до теми «Елементи кореляційного та регресійного аналізу»

1. Парна лінійна регресія

1. Залежність між зростом Y та масою дітей X наведена в таблиці:

$Y = y_i$, м	0,620	0,580	0,640	0,650	0,670	0,680	0,695	0,699	0,710
Маса $X = x_i$, кг	0,531	0,524	0,541	0,550	0,559	0,620	0,632	0,672	0,682

Продовження табл.

$Y = y_i$, м	0,715	0,725	0,781	0,790	0,795	0,800	0,810	0,850	0,860
Маса $X = x_i$, кг	0,689	0,692	0,694	0,698	0,690	0,710	0,720	0,725	0,730

2. Залежність кількості проданих пар чоловічого взуття Y від його розміру X наведена в таблиці:

$Y = y_i$, шт.	10	25	68	136	152	162	170	180
$X = x_i$	44	43	42	41	40	39	38	37

3. Вимірювання температури в грудні, здійснені у двох містах, що умовно позначені А і В, наведено в таблиці:

Місто А $Y = y_i$,°С	– 10,2	– 11,5	– 12,4	– 12,8	– 13,0	– 13,5	– 14,2	– 14,6
Місто В $X = x_i$,°С	– 20,2	– 20,5	– 21,4	– 21,8	– 22,0	– 22,5	– 22,8	– 22,8

Продовження табл.

Місто А $Y = y_i$,°С	– 14,6	– 15,7	– 16,4	– 17,2	– 17,5	– 18,2	– 18,6	– 18,9
Місто В $X = x_i$,°С	– 23,2,	– 24,1	– 24,5	– 25,1	– 25,8	– 26,0	– 26,5	– 27,0

4. Зі старшого класу навмання вибраної середньої школи було відібрано групу учнів. Дані про їх середньорічні оцінки з математики та решти дисциплін в балах наведено в таблиці:

$Y = y_i$	45	25	48	52	54	51	59	60	62	69
$X = x_i$	30	35	31	38	41	48	50	55	51	58

Продовження табл.

$Y = y_i$	72	78	76	80	82	85	81	90	93
$X = x_i$	60	59	65	73	78	71	79	80	81

5. Конденсатор було заряджено до повної напруги в певний момент часу t, після цього він починає розряджатися. Залежність напруги Y від часу розряджання X наведено в таблиці:

$Y = y_i$	100	85	70	65	60	55	50
$X = x_i$	0	1	2	3	4	5	6

Продовження табл.

$Y = y_i$	45	40	35	30	25	22	20
$X = x_i$	7	8	9	10	11	12	13

6. Залежність урожайності пшениці Y від глибини зволоження X наведено в таблиці:

$Y = y_i$, ц/га	10	12	14	16	18	20	22	24	26	28
$X = x_i$, см	0	5	8	10	12	14	16	18	20	22

Продовження табл.

$Y = y_i$, ц/га	30	32	34	36	38	40	42	44	46	48
$X = x_i$, см	24	26	28	30	32	34	36	38	40	42

7. Показники товарообігу Y та суми витрат X, які досліджувалися в 20-ти магазинах, наведено в таблиці:

$Y = y_i$, грн.	480	510	530	540	555	564	570	575	580	585
$X = x_i$, грн.	30	25	31	32	38	41	40	46	49	54

Продовження табл.

$Y = y_i$, грн.	590	596	605	618	625	635	640	650	660
$X = x_i$, грн.	58	60	64	75	78	82	83	85	90

8. Результати вимірювання чутливості Y відеоканалу та звукового каналу X наведено в таблиці:

$Y = y_i$	240	200	190	180	170	160	150	140	130	120
$X = x_i$	170	180	200	230	240	250	280	300	310	320

Продовження табл.

$Y = y_i$	110	100	90	80	70	65	60	55	50	45
$X = x_i$	330	350	380	400	410	420	430	440	450	460

9. Залежність величини зносу різця Y від тривалості роботи X показано в таблиці:

$Y = y_i$, мм	30,0	29,1	28,4	28,1	28,0	27,7	27,5	27,2	27,0
$X = x_i$, год	6	7	8	9	10	11	12	13	14

Продовження табл.

$Y = y_i$, мм	26,8	26,5	26,3	26,1	25,7	25,3	24,3	24,1	24,0
$X = x_i$, год	15	16	17	18	19	20	21	22	23

10. Залежність кров'яного тиску Y людини (в умовних одиницях) від довжини руки X наведена в таблиці:

$Y = y_i$, умов. од.	115	116	117	118	119	120	121	122	123
$X = x_i$, см	62,1	61,1	61,0	60,5	60,0	59,0	58,5	58,0	57,5

Продовження табл.

$Y = y_i$, умов. од.	124	125	126	127	128	129	130	135	150
$X = x_i$, см	56,5	56,0	55,5	55,0	54,5	54,0	53,5	53,0	52,5

11. Залежність пружності Y сталевих болтів від вмісту в них нікелю X наведена в таблиці:

$Y = y_i$, %	35,4	35,0	35,8	36,2	36,7	36,9	37,3	37,8	38,2
$X = x$, %	2,20	2,35	2,42	2,58	2,65	2,69	2,74	2,88	2,91

Продовження табл.

$Y = y_i$, %	39,1	40,5	42,4	43,8	45,6	46,9	48,5	49,4	50,0
$X = x$, %	2,95	2,99	3,00	3,11	3,21	3,29	3,34	3,44	3,50

12. Результати порівняння нового методу газового аналізу зі старим X наведено в таблиці:

$Y = y_i$, умов. од.	2,88	2,91	2,92	2,96	3,01	3,11	3,21	3,25
$X = x_i$, умов. од.	2,07	2,12	2,11	2,58	2,89	2,92	3,01	3,12

Продовження табл.

$Y = y_i$, умов. од.	3,32	3,36	3,42	3,46	3,58	3,88	4,12
$X = x_i$, умов. од.	3,21	3,29	3,31	3,35	3,41	3,48	3,81

13. Показники річної продуктивності праці в розрахунку на одного робітника Y і енергомісткості праці X на підприємствах однієї галузі наведено в таблиці:

$Y = y_i$, тис. грн.	5,4	5,6	6,2	6,8	7,1	7,8	8,5	9,1	10,5	10,9
$X = x_i$, кВт/робітн.	1,8	2,1	2,8	3,0	3,2	3,8	3,9	4,2	4,5	4,8

Продовження табл.

$Y = y_i$, тис. грн.	11,0	11,6	12,1	12,7	13,2	13,9	14,1	14,6	14,9	15,4
$X = x_i$, кВт/робітн.	5,2	5,8	5,9	6,2	6,9	7,2	7,5	8,5	8,8	9,4

14. Залежність денного споживання масла Y певної особи від розміру її заробітної плати за місяць X наведено в таблиці:

$Y = y_i$, г	10,5	15,8	17,8	19,5	20,4	21,5	22,2	24,3	25,8	26,5
$X = x_i$, грн.	70	75	82	89	95	100	105	110	115	120

Продовження табл.

$Y = y_i$, г	28,1	30,1	35,2	36,4	37,0	38,5	39,5	40,5	41,0	42,5
$X = x_i$, грн.	125	130	135	140	145	150	155	160	165	170

15. Залежність маси монети Y від часу її обігу в роках X наведено в таблиці:

$Y = y_i$, мг	9,35	9,21	9,18	9,50	9,10	9,08	9,05	9,01	9,00
$X = x_i$, років	4,0	5,0	5,5	6,0	6,8	7,5	8,5	10,8	12,0

Продовження табл.

$Y = y_i$, мг	8,98	8,94	8,90	8,88	8,82	8,78	8,75	8,70	8,65
$X = x_i$, років	14,5	15,9	25,0	28,5	30,5	36,8	40,0	45,8	50,0

16. Залежність вмісту кремнію *Y* у чавуні від температури шлаку *X* наведено в таблиці:

$Y = y_i$, %	0,27	0,40	0,36	0,42	0,45	0,51	0,55	0,58	0,61
$X = x_i$, °С	1330	1340	1350	1360	1370	1380	1390	1400	1410

Продовження табл.

$Y = y_i$, %	0,64	0,68	0,72	0,76	0,78	0,82	0,88	0,95	1,20
$X = x_i$, °С	1420	1430	1440	1450	1460	1470	1480	1490	1500

17. Залежність урожайності *Y* пшениці від кількості внесених добрив *X* наведено в таблиці:

$Y = y_i$, ц/га	10	12	14	16	18	20	22	24	26	28	30	32	34
$X = x_i$, кг/га	10	30	40	50	60	70	80	90	100	110	120	130	140

18. Залежність міцності волокна бавовни в умовних одиницях *Y* від його товщини *X* наведено в таблиці:

$Y = y_i$, умов. од.	6,02	6,12	6,22	6,28	6,30	6,35	6,39
$X = x_i$, мкм	0,41	0,48	0,56	0,66	0,72	0,79	0,85

Продовження табл.

$Y = y_i$, умов. од.	6,44	6,48	6,52	6,54	6,56	6,60	6,69
$X = x_i$, мкм	0,86	0,88	0,92	0,94	0,96	0,98	0,99

19. Залежність граничного навантаження на болт *Y* від його твердості *X* наведено в таблиці:

$Y = y_i$, умов. од.	10,10	10,30	10,45	10,90	11,20	11,35	11,90	12,45	12,58
$X = x_i$, умов. од.	50,0	50,2	52,8	53,5	54,0	56,8	58,8	59,5	60,5

Продовження табл.

$Y = y_i$, умов. од.	12,96	13,44	13,60	13,95	14,50	14,98	15,48	15,96	16,50
$X = x_i$, умов. од.	64,8	65,4	68,4	69,2	70,5	74,5	76,8	78,5	80,0

20. Вплив температури Y в °С середовища на добовий хід хронометра X наведено в таблиці:

$Y = y_i$, °С	2,60	2,30	2,11	2,01	1,92	1,82	1,55	1,34	1,30	1,28	1,22
$X = x_i$	5,0	5,5	6,0	6,5	7,0	7,5	8,0	8,5	9,0	9,5	10,0

Продовження табл.

$Y = y_i$, °С	1,18	1,12	1,10	0,98	0,92	0,90	0,89	0,88	0,80	0,79
$X = x_i$	10,5	11,0	11,5	12,0	12,5	13,0	14,0	18,0	24,0	30,0

21. Залежність вмісту свинцю Y в руді від вмісту X срібла наведено в таблиці:

$Y = y_i$, %	2,0	2,5	3,0	3,5	4,0	4,5	5,0	5,5	6,0	6,5
$X = x_i$, %	2,0	7,5	12,5	14,5	16,0	18,5	20,0	20,5	22,0	24,5

Продовження табл.

$Y = y_i$, %	7,0	7,5	8,0	8,5	9,0	10,5	12,5	14,5	15,0	16,5
$X = x_i$, %	26,0	28,5	30,0	32,5	34,0	36,5	38,0	40,5	42,0	45,0

22. Залежність числа гризунів Y, які загинули від наявності отрути в їжі при концентрації X, наведено в таблиці:

$Y = y_i$	32	36	38	42	46	49	55	59	62
$X = x_i$, %	3,0	3,5	4,0	4,5	5,0	6,0	6,5	7,0	7,5

Продовження табл.

$Y = y_i$	68	70	73	75	81	88	92	94	98
$X = x_i$, %	8,0	8,5	9,0	9,5	10,0	10,5	11,0	11,5	12,0

23. Залежність між собівартістю X та кількістю виготовлених виробів Y наведено в таблиці:

$Y = y_i$, тис. грн.	2,2	3,5	3,7	3,8	4,5	5,7
$X = x_i$, тис. шт.	1,5	1,4	1,2	1,1	0,9	0,8

24. Залежність урожайності пшениці Y від глибини оранки X наведено в таблиці:

$Y = y_i$, ц/га	7	8	9	10	11	12
$X = x_i$, см	8,1	8,3	8,2	9,1	10,3	10,8

25. Залежність кількості споживання масла на добу певною категорією пенсіонерів X від розміру Y отриманої місячної пенсії наведено в таблиці:

$Y = y_i$, грн.	29	38	49	54	62	70	79	98
$X = x_i$, г	15,99	19,75	23,10	26,44	29,79	33,13	36,89	44,54

26. Залежність вмісту срібла в руді Y від вмісту свинцю наведено в таблиці:

$Y = y_i$, %	2	6	10	14	18	22	26	30
$X = x_i$, %	2,5	7,5	12,5	17,5	22,5	27,5	32,5	37,5

27. Залежність вмісту кремнію Y від температури X наведено в таблиці:

$Y = y_i$, %	0,27	0,26	0,27	0,28	0,29	0,3	0,31	0,32	0,33
$X = x_i$, °C	1330	1340	1350	1360	1370	1380	1390	1400	1410

28. Залежність міцності волокна бавовни X від граничного навантаження Y наведено в таблиці:

$X = x_i$, умов. од.	4100	4300	4500	4700	4900	5100	5200	5300	5500
$Y = y_i$, г	3,75	4,25	4,75	5,25	5,75	6,25	6,75	7,00	7,25

29. Залежність собівартості від вирощеного врожаю соняшнику наведено в таблиці:

$Y = y_i$, грн/ц	10,36	11,56	13,29	14,51	15,6	14,25	17,36	16,23
$X = x_i$, ц/га	1,23	1,33	1,43	1,53	1,63	1,73	1,83	1,93

30. Залежність урожайності цукрових буряків Y від кількості внесених у ґрунт поживних речовин X наведено в таблиці:

$Y = y_i$, ц/га	369	380	370	395	420	412	436	420
$X = x_i$, кг/га	83	92	112	132	144	154	162	189

Потрібно:

1. Знайти точкові незміщені статистичні оцінки β_0^*, β_1^* для параметрів β_0, β_1 парної лінійної функції регресії $y_i = \beta_0 + \beta_1 x_i$.

2. З надійністю $\gamma = 0{,}99$ побудувати довірчі інтервали для параметрів β_0, β_1.

3. При рівні значущості $\alpha = 0{,}01$ перевірити значущість параметра β_1.

4. З надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для функції регресії $y_i = \beta_0 + \beta_1 x_i$.

5. Обчислити вибірковий коефіцієнт кореляції.

6. З надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для прогнозованих індивідуальних значень.

2. Множинна лінійна регресія

а) Залежність між продуктивністю праці Y та фондозабезпеченістю, тис. грн., X_1, стажем роботи в роках X_2, X_3 — плинністю кадрів (у частках), рівнем оплати праці, тис. грн/рік — X_4 наведено в таблицях:

1.

№ з/п	Y	X_1	X_2	X_3	X_4
1	14,85	60	30	0,15	5,0
2	11,94	48	19	0,02	3,1
3	8,03	39	8	0,14	4,7
4	7,11	28	18	0,11	2,5
5	9,50	45	9	0,12	4,9
6	9,40	37	23	0,10	2,6
7	11,60	58	15	0,13	4,6
8	8,14	27	17	0,09	3,4
9	15,62	58	28	0,07	4,8
10	11,12	47	16	0,12	4,9
11	7,34	38	7	0,08	3,2
12	10,58	44	15	0,11	4,7
13	7,37	23	25	0,15	2,7
14	10,63	57	8	0,13	5,0
15	10,63	38	24	0,07	2,9

2.

№ з/п	Y	X_1	X_2	X_3	X_4
1	11,12	47	16	0,12	4,9
2	7,34	38	7	0,08	3,2
3	10,58	44	15	0,11	4,7
4	7,37	23	25	0,15	2,7
5	10,63	57	8	0,13	5,0
6	10,63	38	24	0,07	2,9
7	7,85	22	15	0,12	4,6
8	5,73	29	7	0,09	2,8
9	14,84	56	27	0,02	3,5
10	10,30	45	15	0,14	4,9
11	7,85	34	9	0,10	4,1
12	9,68	51	14	0,11	3,3
13	9,49	55	5	0,13	4,8
14	12,53	43	26	0,08	4,0
15	10,29	44	27	0,15	2,9

3.

№ з/п	Y	X_1	X_2	X_3	X_4
1	7,85	22	15	0,12	4,6
2	5,73	29	7	0,09	2,8
3	14,84	56	27	0,02	3,5
4	10,30	45	15	0,14	4,9
5	7,85	34	9	0,10	4,1
6	9,68	51	14	0,11	3,3
7	9,49	55	5	0,13	4,8
8	12,53	43	26	0,08	4,0
9	10,29	44	27	0,15	2,9
10	8,99	37	8	0,06	4,3
11	12,28	33	24	0,12	5,0
12	8,00	25	18	0,02	2,9
13	7,27	29	4	0,07	3,5
14	7,47	53	13	0,14	2,7
15	10,86	41	9	0,08	4,9
16	5,23	26	12	0,13	3,4

4.

№ з/п	Y	X_1	X_2	X_3	X_4
1	8,00	25	18	0,02	2,9
2	7,27	29	4	0,07	3,5
3	7,47	53	13	0,14	2,7
4	10,86	41	9	0,08	4,9
5	5,23	26	12	0,13	3,4
6	12,16	32	23	0,10	4,8
7	9,19	59	11	0,13	3,9
8	10,12	48	3	0,09	4,8
9	6,86	51	8	0,12	2,9
10	11,02	43	22	0,15	3,7
11	7,77	29	9	0,02	3,5
12	10,62	37	12	0,08	5,0
13	7,40	49	5	0,14	4,1
14	10,55	57	11	0,11	3,6
15	12,30	46	15	0,06	4,7
16	7,83	29	21	0,15	2,8

5.

№ з/п	Y	X_1	X_2	X_3	X_4
1	10,55	57	11	0,11	3,6
2	12,30	46	15	0,06	4,7
3	7,83	29	21	0,15	2,8
4	11,10	35	18	0,05	4,9
5	7,66	38	10	0,14	3,6
6	9,26	30	22	0,06	3,1
7	11,50	45	6	0,02	5,0
8	14,51	60	20	0,05	4,2
9	6,33	39	7	0,09	2,8
10	12,94	50	21	0,06	4,7
11	13,13	49	15	0,04	4,8

6.

№ з/п	Y	X_1	X_2	X_3	X_4
1	14,85	60	30	0,15	5,0
2	8,03	39	8	0,14	4,7
3	9,50	45	9	0,12	4,9
4	11,61	58	15	0,13	4,6
5	15,62	58	28	0,07	4,8
6	7,34	38	7	0,08	3,2
7	7,37	23	25	0,15	2,7
8	10,63	38	24	0,07	2,9
9	5,73	29	7	0,09	2,8
10	10,30	45	15	0,14	4,9
11	9,68	51	14	0,11	3,3
12	12,53	43	26	0,08	4,0
13	8,99	37	8	0,06	4,3
14	8,00	25	18	0,02	2,9
15	7,47	53	13	0,14	2,7

7.

№ з/п	Y	X_1	X_2	X_3	X_4
1	5,73	29	7	0,09	2,8
2	7,85	34	9	0,10	4,1
3	12,53	43	26	0,08	4,0
4	12,28	33	24	0,12	5,0
5	7,47	53	13	0,14	2,7
6	5,23	26	12	0,13	3,4
7	12,16	32	23	0,10	4,8
8	6,86	51	8	0,12	2,9
9	11,02	43	22	0,15	3,7
10	7,77	29	9	0,02	3,5
11	10,62	37	12	0,08	5,0
12	7,40	49	5	0,14	4,1
13	10,55	57	11	0,11	3,6
14	12,30	46	15	0,06	4,7
15	7,83	29	21	0,15	2,8

8.

№ з/п	Y	X_1	X_2	X_3	X_4
1	8,99	37	8	0,06	4,3
2	12,28	33	24	0,12	5,0
3	8,00	25	18	0,02	2,9
4	7,27	29	4	0,07	3,5
5	7,47	53	13	0,14	2,7
6	10,86	41	9	0,08	4,9
7	5,23	26	12	0,13	3,4
8	12,16	32	23	0,10	4,8
9	9,19	59	11	0,13	3,9
10	10,12	48	3	0,09	4,8
11	6,86	51	8	0,12	2,9
12	11,02	43	22	0,15	3,7
13	7,77	29	9	0,02	3,5
14	10,62	37	12	0,08	5,0
15	7,40	49	5	0,14	4,1

9.

№ з/п	Y	X_1	X_2	X_3	X_4
1	10,58	44	15	0,11	4,7
2	7,37	23	25	0,15	2,7
3	10,63	38	24	0,07	2,9
4	7,85	22	15	0,12	4,6
5	5,73	29	7	0,09	2,8
6	14,84	56	27	0,02	3,5
7	10,30	45	15	0,14	4,9
8	9,68	51	14	0,11	3,3
9	9,49	55	5	0,13	4,8
10	12,53	43	26	0,08	4,0
11	10,29	44	27	0,15	2,9
12	12,28	33	24	0,12	5,0
13	8,00	25	18	0,02	2,9
14	7,27	29	4	0,07	3,5
15	7,47	53	13	0,14	2,7

10.

№ з/п	Y	X_1	X_2	X_3	X_4
1	5,23	26	12	0,13	3,4
2	12,16	32	23	0,10	4,8
3	9,19	59	11	0,13	3,9
4	10,12	48	3	0,09	4,8
5	6,86	51	8	0,12	2,9
6	10,62	37	12	0,08	5,0
7	10,55	57	11	0,11	3,6
8	7,83	29	21	0,15	2,8
9	11,10	35	18	0,05	4,9
10	7,66	38	10	0,14	3,6
11	9,26	30	22	0,06	3,1
12	11,50	45	6	0,02	5,0
13	6,33	39	7	0,09	2,8
14	12,94	50	21	0,06	4,7
15	13,13	49	15	0,04	4,8

11.

№ з/п	Y	X_1	X_2	X_3	X_4
1	9,50	45	9	0,12	4,9
2	8,14	27	17	0,09	3,4
3	7,34	38	7	0,08	3,2
4	7,37	23	25	0,15	2,7
5	10,63	38	24	0,07	2,9
6	5,73	29	7	0,09	2,8
7	10,30	45	15	0,14	4,9
8	9,68	51	14	0,11	3,3
9	12,53	43	26	0,08	4,0
10	8,99	37	8	0,06	4,3
11	7,27	29	4	0,07	3,5
12	11,10	35	18	0,05	4,9
13	7,47	53	13	0,14	2,7
14	9,26	30	22	0,06	3,1
15	12,16	32	23	0,01	4,8
16	9,19	59	11	0,13	3,9

12.

№ з/п	Y	X_1	X_2	X_3	X_4
1	13,13	49	15	0,04	4,8
2	6,33	39	7	0,09	2,8
3	11,50	45	6	0,02	5,0
4	7,66	38	10	0,14	3,6
5	7,83	29	21	0,15	3,8
6	10,55	57	11	0,11	3,6
7	7,40	49	5	0,14	4,1
8	10,62	37	12	0,08	5,0
9	7,77	29	9	0,02	3,5
10	6,86	51	8	0,12	2,9
11	10,12	48	3	0,09	4,8
12	9,19	59	11	0,13	3,9
13	14,85	60	30	0,15	5,0
14	8,03	39	19	0,02	3,1
15	7,11	28	18	0,11	2,5

13.

№ з/п	Y	X_1	X_2	X_3	X_4
1	10,29	44	27	0,15	2,9
2	12,53	43	26	0,08	4,0
3	9,49	55	5	0,13	4,8
4	9,68	51	14	0,11	3,3
5	7,85	34	9	0,10	4,1
6	10,30	45	15	0,14	4,9
7	14,84	56	27	0,02	3,5
8	5,73	29	7	0,09	2,8
9	7,85	22	15	0,12	4,6
10	10,63	57	8	0,13	5,0
11	7,37	23	25	0,15	2,7
12	10,58	44	15	0,11	4,7
13	7,34	38	7	0,08	3,2
14	11,12	47	16	0,12	4,9
15	15,62	58	28	0,07	4,8

14.

№ з/п	Y	X_1	X_2	X_3	X_4
1	7,83	29	21	0,15	2,8
2	12,30	46	15	0,06	4,7
3	10,55	57	11	0,11	3,6
4	7,40	49	5	0,14	4,1
5	10,62	37	12	0,08	5,0
6	7,77	29	9	0,02	3,5
7	11,02	43	22	0,15	3,7
8	5,86	51	8	0,12	2,9
9	10,12	48	3	0,09	4,8
10	9,19	59	11	0,13	3,9
11	10,30	45	15	0,14	4,9
12	7,85	34	9	0,10	4,1
13	9,68	51	14	0,11	3,3
14	9,49	55	5	0,13	4,8
15	12,53	43	26	0,08	4,0
16	10,29	44	27	0,15	2,9

б) За заданим статистичним даним залежності врожайності цукрових буряків Y, ц/га, від: 1) живої праці X_1, людино-днів/га, 2) кількості внесених поживних речовин X_2, кг/га, 3) опадів на рік X_3, мм, 4) суми температур за період активної вегетації X_4, °С:

15.

№ з/п	Y	X_1	X_2	X_3	X_4
1	369	16	83	460	2500
2	457	18	240	503	2621
3	379	13	125	496	2564
4	403	21	86	548	2792
5	439	17	221	472	2672
6	421	12	201	484	2840
7	448	23	217	537	2711
8	407	24	97	461	2638
9	419	18	144	493	2578
10	441	19	205	539	2617
11	418	20	156	526	2835
12	401	15	175	467	2693
13	451	17	189	542	2691
14	381	21	86	472	2532
15	432	18	204	483	2783

16.

№ з/п	Y	X_1	X_2	X_3	X_4
1	439	17	221	472	2672
2	448	23	217	537	2711
3	419	18	144	493	2578
4	418	20	156	526	2835
5	451	17	189	542	2693
6	381	21	86	472	2532
7	439	15	110	538	2627
8	423	17	210	523	2593
9	396	21	125	539	2543
10	412	20	93	471	2682
11	402	15	125	539	2543
12	413	22	87	501	2736
13	389	17	216	463	2639
14	418	18	173	542	2817
15	405	15	214	498	2572
16	399	21	92	498	2735

17.

№ з/п	Y	X_1	X_2	X_3	X_4
1	439	19	217	463	2702
2	423	17	210	523	2593
3	396	21	125	492	2828
4	412	20	93	471	2682
5	402	15	125	539	2543
6	413	22	87	501	2736
7	389	17	216	463	2639
8	418	18	173	542	2817
9	405	15	214	492	2572
10	399	21	92	498	2735
11	403	23	89	483	2720
12	396	17	140	523	2527
13	377	15	96	499	2793
14	427	20	180	471	2815
15	412	17	200	483	2584
16	453	19	171	511	2801

18.

№ з/п	Y	X_1	X_2	X_3	X_4
1	393	15	110	538	2627
2	396	21	125	492	2828
3	402	15	125	539	2543
4	413	22	87	501	2736
5	389	17	216	463	2639
6	389	18	173	542	2817
7	399	21	92	498	2735
8	403	23	89	483	2720
9	396	17	140	523	2527
10	377	15	96	499	2793
11	427	20	180	471	2815
12	412	17	200	483	2584
13	453	19	171	511	2801
14	404	22	163	476	2612
15	397	24	103	516	2643

19.

№ з/п	Y	X_1	X_2	X_3	X_4
1	371	15	170	493	2648
2	478	18	217	510	2573
3	377	17	154	475	2543
4	452	22	180	518	2801
5	439	21	143	478	2562
6	401	17	130	523	2517
7	429	19	160	468	2650
8	366	15	126	474	2628
9	424	26	90	493	2529
10	371	20	115	521	2823
11	429	21	220	464	2730
12	391	18	97	547	2555
13	407	15	225	472	2711
14	449	24	239	517	2784
15	408	25	184	492	2548

20.

№ з/п	Y	X_1	X_2	X_3	X_4
1	413	22	87	501	2736
2	418	18	173	542	2817
3	399	21	92	498	2735
4	396	17	140	523	2527
5	427	20	180	471	2815
6	453	19	171	511	2801
7	397	24	103	516	2643
8	478	18	217	510	2573
9	452	22	180	518	2801
10	401	17	130	523	2517
11	366	15	126	474	2628
12	371	20	115	521	2823
13	391	18	97	547	2555
14	449	24	239	517	2784
15	393	21	85	547	2837
16	407	24	97	461	2638

21.

№ з/п	Y	X_1	X_2	X_3	X_4
1	408	25	184	492	2548
2	407	15	225	472	2711
3	429	21	220	464	2730
4	424	26	90	493	2529
5	429	19	160	468	2650
6	439	21	143	478	2562
7	377	17	154	475	2543
8	371	15	170	493	2648
9	404	22	163	476	2612
10	412	17	200	483	2584
11	377	15	96	499	2793
12	403	23	89	483	2720
13	405	15	214	498	2572
14	389	17	216	463	2639
15	402	15	125	539	2543

22.

№ з/п	Y	X_1	X_2	X_3	X_4
1	452	22	180	518	2801
2	377	17	154	475	2543
3	478	18	217	510	2573
4	371	15	170	493	2648
5	397	24	103	516	2643
6	404	22	163	476	2612
7	427	20	180	471	2815
8	396	17	140	523	2527
9	399	21	92	483	2720
10	418	18	173	542	2817
11	413	22	87	501	2736
12	412	20	93	471	2682
13	423	17	210	523	2593
14	393	15	110	538	2627
15	381	21	86	472	2532
16	401	15	175	467	2693

23.

№ з/п	Y	X_1	X_2	X_3	X_4
1	401	17	130	523	2517
2	452	22	180	518	2801
3	478	18	217	510	2573
4	397	24	103	516	2643
5	453	19	171	511	2801
6	427	20	180	471	2815
7	396	17	140	523	2527
8	399	21	92	498	2735
9	418	18	173	542	2817
10	413	22	87	501	2736
11	412	20	93	471	2682
12	423	17	210	523	2593
13	393	15	110	538	2627
14	381	21	86	472	2532
15	401	15	175	467	2693

24.

№ з/п	Y	X_1	X_2	X_3	X_4
1	405	15	214	498	2572
2	418	18	173	542	2817
3	389	17	216	463	2639
4	413	22	87	501	2736
5	402	15	125	539	2543
6	412	20	93	471	2682
7	396	21	125	492	2828
8	423	17	210	523	2593
9	439	19	217	463	2702
10	393	15	110	538	2627
11	432	18	204	483	2783
12	381	21	86	472	2532
13	451	17	189	542	2691
14	401	15	175	467	2693
15	418	20	156	526	2835

25.

№ з/п	Y	X_1	X_2	X_3	X_4
1	453	19	171	511	2801
2	427	20	180	471	2715
3	377	15	96	499	2793
4	396	17	140	523	2527
5	403	23	89	483	2720
6	399	21	92	498	2735
7	405	15	214	498	2572
8	418	18	173	542	2817
9	389	17	216	463	2639
10	413	22	87	501	2736
11	402	15	125	539	2543
12	412	20	93	471	2682
13	396	21	125	492	2828
14	423	17	210	523	2593
15	439	19	217	463	2702

26.

№ з/п	Y	X_1	X_2	X_3	X_4
1	451	17	189	542	2691
2	381	21	86	472	2532
3	432	18	204	483	2783
4	393	15	110	538	2627
5	439	19	217	463	2702
6	423	17	210	523	2593
7	396	21	125	539	2543
8	412	20	93	471	2682
9	402	15	125	539	2543
10	413	22	87	501	2736
11	389	17	216	463	2639
12	418	18	173	542	2817
13	405	15	214	498	2735
14	399	21	92	498	2735
15	403	23	89	483	2720

27.

№ з/п	Y	X_1	X_2	X_3	X_4
1	439	19	217	463	2702
2	393	15	110	538	2627
3	432	18	204	483	2783
4	381	21	86	472	2532
5	451	17	189	542	2691
6	401	15	175	467	2693
7	418	20	156	526	2835
8	441	19	205	539	2617
9	419	18	144	493	2578
10	407	24	97	461	2638
11	448	23	217	537	2711
12	421	12	201	484	2840
13	439	17	221	472	2672
14	403	21	86	548	2792
15	379	13	125	496	2564

28.

№ з/п	Y	X_1	X_2	X_3	X_4
1	457	18	240	503	2621
2	403	21	86	548	2792
3	421	12	201	484	2840
4	407	24	97	461	2638
5	441	19	205	539	2617
6	401	15	175	467	2693
7	381	21	86	472	2532
8	393	19	217	463	2702
9	423	17	210	523	2593
10	412	20	93	471	2682
11	413	22	87	501	2736
12	418	18	173	542	2817
13	399	21	92	498	2572
14	396	17	140	523	2572

29.

№ з/п	Y	X_1	X_2	X_3	X_4
1	439	19	217	463	2702
2	412	20	93	471	2682
3	389	17	216	463	2639
4	399	21	92	498	2735
5	377	15	96	499	2793
6	453	19	171	511	2801
7	371	15	170	493	2648
8	452	22	180	518	2801
9	429	19	160	468	2650
10	424	26	90	493	2529
11	371	20	115	521	2823
12	391	18	97	547	2555
13	449	24	239	517	2784
14	408	25	184	492	2548
15	393	21	85	547	2837

30.

№ з/п	Y	X_1	X_2	X_3	X_4
1	424	26	90	493	2526
2	429	19	160	468	2650
3	439	21	143	478	2562
4	377	17	154	475	2543
5	378	18	217	510	273
6	371	15	170	493	2648
7	397	24	103	516	2643
8	404	22	163	476	2612
9	453	19	171	511	2801
10	412	17	200	483	2584
11	427	20	180	471	2815
12	377	15	96	499	2793
13	396	17	190	523	2527
14	403	23	89	483	2720
15	399	21	92	498	2735

***Потрібно*:**

1. Визначити статистичні оцінки $\beta_0^*, \beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*$ для параметрів лінійної множинної регресії:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}\,.$$

2. З надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для функції регресії.
3. Обчислити коефіцієнт множинної кореляції R.

3. Множинна нелінійна регресія

1) Підприємство виробляє речі побуту (двері, вікна) і веде торгівлю супутними товарами (замки, дверні ручки тощо). Залежність прибутку Y, тис. грн., від: обсягу виробництва X, тис. грн., обсягу торгівлі Z, тис. грн., наведена статистичними розподілами:

1.

	Y	*X*	*Z*
1	468	1200	600
2	496	1300	650
3	484	1400	630
4	528	1450	620
5	495	1500	610
6	543	1550	590
7	509	1600	580
8	565	1650	560
9	502	1630	570
10	568	1680	540
11	511	1710	520
12	575	1780	510
13	536	1810	500
14	557	1830	490
15	534	1850	430

2.

	Y	*X*	*Z*
1	502	1630	570
2	511	1710	520
3	536	1810	500
4	534	1850	430
5	548	1740	420
6	532	1860	410
7	550	1910	390
8	508	2050	300
9	534	2060	320
10	519	2070	340
11	542	2100	350
12	524	2150	370
13	549	2210	410

3.

	Y	*X*	*Z*
1	536	1810	500
2	557	1830	490
3	534	1850	430
4	548	1740	420
5	532	1860	410
6	550	1910	390
7	508	2050	300
8	534	2060	320
9	519	2070	340
10	542	2100	350
11	524	2150	370
12	549	2210	410
13	534	2300	550
14	542	2350	530
15	531	2340	550

4.

	Y	*X*	*Z*
1	508	2050	300
2	534	2060	320
3	519	2070	340
4	542	2100	350
5	524	2150	370
6	549	2210	410
7	534	2300	550
8	542	2350	530
9	531	2340	550
10	535	2450	490
11	507	2500	350
12	496	2600	330
13	485	2650	350
14	500	2700	410
15	486	2750	440

5.

	Y	*X*	*Z*
1	486	2750	440
2	481	2850	460
3	464	2900	480
4	450	3000	510
5	467	2900	550
6	475	2850	560
7	484	2800	550
8	492	2750	540
9	500	2700	530
10	06	2650	550
11	514	2600	510
12	519	2550	530
13	521	2500	570
14	529	2450	520
15	534	2400	510

6.

	Y	*X*	*Z*
1	496	1300	650
2	528	1450	620
3	543	1550	590
4	565	1650	560
5	568	1680	540
6	575	1780	510
7	557	1830	490
8	548	1740	420
9	550	1910	390
10	534	2060	320
11	542	2100	350
12	549	2210	410
13	542	2350	530
14	535	2450	490
15	496	2600	330
16	500	2700	410

7.

	Y	*X*	*Z*
1	524	2150	370
2	549	2210	410
3	542	2350	530
4	535	2450	490
5	496	2600	330
6	500	2700	410
7	481	2850	460
8	450	3000	510
9	475	2850	560
10	492	2750	540
11	506	2650	550
12	519	2550	530
13	529	2450	520
14	537	2350	530
15	544	2250	500

8.

	Y	*X*	*Z*
1	468	1200	600
2	528	1450	620
3	509	1600	580
4	511	1710	520
5	534	1850	430
6	550	1910	390
7	519	2070	340
8	549	2210	410
9	531	2340	550
10	496	2600	330
11	500	2700	410
12	464	2900	480
13	475	2850	560
14	500	2700	530
15	521	2500	570

9.

	Y	*X*	*Z*
1	511	1710	520
2	536	1810	500
3	534	1850	430
4	532	1860	410
5	508	2050	300
6	519	2070	340
7	524	2150	370
8	534	2300	550
9	531	2340	550
10	507	2500	350
11	485	2650	350
12	486	2750	440
13	464	2900	480
14	467	2900	550
15	484	2800	550

10.

	Y	*X*	*Z*
5	557	1830	490
2	548	1740	420
3	550	1910	390
4	534	2060	320
5	542	2100	350
6	549	2210	410
7	542	2350	530
8	535	2450	490
9	496	2600	330
10	500	2700	410
11	481	2850	460
12	450	3000	510
13	475	2850	550
14	492	2750	540
15	506	2650	550
16	519	2550	530

2) Залежність собівартості *Y* (тис. грн.) від: 1) обсягу виробництва *X* (тис. грн.), 2) обсягу торгівлі *Z* (тис. грн.) подано статистичними розподілами:

11.

	Y	*X*	*Z*
1	1332	1200	600
2	1453	1300	650
3	1546	1400	630
4	1542	1450	620
5	1615	1500	610
6	1597	1550	590
7	1671	1600	580
8	1645	1650	560
9	1698	1630	570
10	1652	1680	540
11	1719	1710	520
12	1715	1780	510
13	1774	1810	500
14	1763	1830	490
15	1746	1850	430

12.

	Y	*X*	*Z*
1	1546	1400	630
2	1615	1500	610
3	1671	1600	580
4	1698	1630	570
5	1719	1710	520
6	1774	1810	500
7	1746	1850	430
8	1612	1740	420
9	1738	1860	410
10	1750	1910	390
11	1842	2050	300
12	1846	2060	320
13	1891	2070	340
14	1908	2100	350
15	1996	2150	370

13.

	Y	*X*	*Z*
1	1746	1850	430
2	1338	1860	410
3	1842	2050	300
4	1891	2070	340
5	1996	2150	370
6	2316	2300	550
7	2359	2340	550
8	2343	2500	350
9	2515	2650	350
10	2704	2750	440
11	2829	2850	360
12	2916	2900	480
13	3060	3000	510
14	2983	2900	550
15	2935	2850	560

14.

	Y	*X*	*Z*
1	1891	2070	340
2	1996	2150	270
3	2316	2300	550
4	2359	2340	530
5	2405	2450	490
6	2434	2600	330
7	2610	2700	410
8	2704	2750	440
9	2916	2900	480
10	2983	2900	550
11	2866	2800	550
12	2730	2700	530
13	2596	2600	510
14	2549	2500	570
15	2376	2400	510
16	2300	2300	540

15.

	Y	*X*	*Z*
1	1746	1850	430
2	1612	1740	420
3	1750	1910	390
4	1842	2050	300
5	1891	2070	340
6	1908	2100	350
7	2316	2300	550
8	2338	2350	530
9	2405	2450	490
10	2343	2500	350
11	2515	2650	350
12	2610	2700	410
13	2829	2850	460
14	2916	2900	480
15	2983	2900	550

16.

	Y	*X*	*Z*
1	1615	1500	610
2	1517	1600	580
3	1698	1630	570
4	1719	1710	520
5	1774	1810	500
6	1746	1850	430
7	1738	1860	10
8	1842	2050	300
9	1891	2070	340
10	1996	2150	370
11	2316	2300	550
12	2359	2340	550
13	2343	2500	350
14	2515	2650	350
15	2610	2700	410

17.

	Y	*X*	*Z*
1	1715	1780	510
2	1763	1830	490
3	1746	1850	430
4	1738	1860	410
5	1842	2050	300
6	1846	2060	320
7	1908	2100	350
8	2071	2210	410
9	2316	2300	550
10	2338	2350	530
11	2405	2450	490
12	2343	2500	350
13	2515	2650	350
14	2610	2700	410
15	2829	2850	460
	2916	2900	480

18.

	Y	*X*	*Z*
1	1546	1400	630
2	1615	1500	610
3	1698	1630	570
4	1774	1810	500
5	1612	1740	420
6	1842	2050	300
7	1908	2100	350
8	2071	2210	410
9	2359	2340	550
10	2434	2600	330
11	2704	2750	440
12	3060	3000	510
13	2866	2800	550
14	2694	2650	550
15	2441	2450	520

19.

	Y	*X*	*Z*
1	1842	2050	300
2	1891	2070	340
3	1996	2150	370
4	2316	2300	550
5	2359	2340	550
6	2405	2450	490
7	2334	2600	330
8	2610	2700	410
9	2916	2900	480
10	2983	2900	550
11	2866	2800	550
12	2798	2750	540
13	2694	2650	550
14	2561	2550	530
15	2441	2450	520

20.

	Y	*X*	*Z*
1	1774	1810	500
2	1746	1850	430
3	1612	1740	420
4	1750	1910	390
5	1842	2050	300
6	1908	2100	350
7	1996	2150	370
8	2316	2300	550
9	2338	2350	530
10	2405	2450	490
11	2343	2500	350
12	2610	2700	410
13	2916	2900	480
14	2376	2400	510
15	2300	2300	540
16	2206	2250	500

3) Залежність рентабельності *Y* (%) підприємства від: 1) обсягу виробництва *X*, тис. грн.; 2) обсягу торгівлі *Z*, тис. грн., наведено статистичним розподілом:

21.

	Y	*X*	*Z*
1	35,14	1200	600
2	34,11	1300	650
3	31,30	1400	630
4	34,24	1450	620
5	30,65	1500	610
6	34,00	1550	590
7	30,46	1600	580
8	34,35	1650	560
9	29,56	1630	570
10	34,38	1680	540
11	29,73	1710	520
12	33,53	1780	510
13	30,21	1810	500
14	31,59	1830	490
15	30,58	1850	430

22.

	Y	*X*	*Z*
1	34,00	1740	420
2	30,61	1860	410
3	31,43	1910	390
4	27,58	2050	300
5	28,92	2060	320
6	27,45	2070	340
7	18,41	2100	350
8	26,25	2150	370
9	26,51	2210	410
10	23,06	2300	550
11	23,18	2350	530
12	22,51	2340	550
13	22,25	2450	490
14	21,64	2500	350
15	20,38	2600	330

23.

	Y	*X*	*Z*
1	19,28	2650	350
2	19,16	2700	410
3	17,97	2750	440
4	17,00	2850	460
5	15,91	2900	480
6	14,71	3000	510
7	15,66	2900	550
8	16,18	2850	560
9	16,89	2800	550
10	17,58	2750	540
11	18,32	2700	530
12	18,78	2650	550
13	19,80	2600	510
14	20,27	2550	530
15	20,44	2500	570

24.

	Y	*X*	*Z*
1	21,67	2450	520
2	22,47	2400	510
3	22,92	2350	530
4	23,48	2300	540
5	24,66	2250	500
6	35,14	1200	600
7	31,30	1400	630
8	34,24	1450	620
9	34,00	1550	590
10	34,35	1650	560
11	34,38	1680	540
12	33,53	1780	510
13	31,59	1830	490
14	34,00	1740	420
15	31,43	1910	390
16	28,92	2060	320

25.

	Y	*X*	*Z*
1	31,59	1830	490
2	34,00	1740	420
3	31,43	1910	390
4	28,92	2060	320
5	28,41	2100	350
6	26,51	2210	410
7	23,18	2350	530
8	22,25	2450	490
9	20,38	2600	330
10	19,16	2700	410
11	17,97	2750	440
12	15,91	2900	480
13	15,66	2900	550
14	16,89	2800	550
15	18,32	2700	530

26.

	Y	*X*	*Z*
1	30,58	1850	430
2	30,61	1860	410
3	27,58	2050	300
4	24,45	2070	340
5	26,25	2150	370
6	23,06	2300	550
7	22,51	2340	550
8	21,64	2500	350
9	19,28	2650	350
10	17,97	2750	440
11	15,91	2900	480
12	15,66	2900	550
13	16,89	2800	550
14	18,32	2700	530
15	19,80	2600	510
16	20,44	2500	570

27.

	Y	*X*	*Z*
1	23,06	2300	550
2	22,51	2340	550
3	21,64	2500	350
4	19,28	2650	350
5	17,97	2750	440
6	15,91	2900	480
7	15,66	2900	550
8	16,18	2850	560
9	17,58	2750	540
10	18,78	2650	550
11	20,27	2550	530
12	21,67	2450	520
13	22,92	2350	530
14	24,66	2250	500
15	34,24	1450	620

28.

	Y	*X*	*Z*
1	33,53	1780	510
2	31,59	1830	490
3	30,58	1850	430
4	30,61	1860	410
5	31,43	1910	390
6	27,58	2050	300
7	27,45	2070	340
8	26,25	2150	370
9	23,06	2300	550
10	22,51	2340	530
11	21,64	2500	350
12	19,28	2650	350
13	17,00	2850	460
14	14,71	3000	510
15	16,18	2850	560

29.

	Y	X	Z
1	31,30	1400	630
2	30,65	1500	610
3	30,46	1600	580
4	29,56	1630	570
5	29,73	1710	520
6	30,21	1810	500
7	31,59	1830	490
8	34,00	1749	420
9	31,43	1910	390
10	28,92	2060	320
11	28,41	2100	350
12	26,51	2210	410
13	23,18	2350	530
14	22,25	2450	490
15	20,28	2650	350

30.

	Y	X	Z
1	34,24	1450	620
2	30,65	1500	610
3	30,46	1600	580
4	34,35	1650	560
5	34,38	1680	540
6	29,73	1710	520
7	30,21	1810	500
8	31,59	1830	490
9	34,00	1740	420
10	31,43	1910	390
11	27,58	2050	300
12	27,45	2070	340
13	28,41	2100	350
14	26,51	2210	410
15	23,06	2300	550

***Потрібно*:**

1. Визначити статистичні оцінки $\beta_0^*, \beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*$ для параметрів β_0, β_1, β_2, β_3, β_4 рівняння регресії $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 z_i + \beta_4 z_i^2$.

2. З надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для нелінійної функції регресії $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 z_i + \beta_4 z_i^2$.

3. Обчислити коефіцієнт множинної регресії R.

Залежність собівартості картоплі Y, грн/ц від її врожайності X, сотні ц/га, наведено статистичними розподілами:

1.

	Y	*X*
1	11,06	1,23
2	10,36	1,45
3	10,04	1,52
4	12,28	1,07
5	11,44	1,24
6	10,06	1,35
7	10,56	1,48
8	11,24	1,29
9	10,62	1,46
10	10,82	1,29
11	10,90	1,43
12	9,16	1,52
13	11,46	1,53
14	10,42	1,46
15	10,22	1,53

2.

	Y	*X*
1	10,46	1,47
2	10,90	1,32
3	10,52	1,43
4	10,26	1,51
5	10,90	1,32
6	10,80	1,37
7	10,48	1,44
8	10,32	1,50
9	11,18	1,25
10	10,68	1,38
11	9,05	1,64
12	8,36	1,85
13	7,25	2,01
14	8,03	1,95
15	7,46	2,16

3.

	Y	*X*
1	9,26	1,72
2	10,03	1,84
3	6,35	2,34
4	7,73	1,92
5	6,95	2,07
6	6,34	2,26
7	6,58	2,04
8	9,34	1,72
9	7,58	2,03
10	9,56	1,72
11	6,22	2,35
12	7,48	2,02
13	6,93	2,27
14	7,22	2,18
15	8,83	1,94

4.

	Y	*X*
1	10,36	1,45
2	12,28	1,07
3	10,60	1,35
4	11,24	1,29
5	10,14	1,62
6	10,82	1,29
7	9,16	1,52
8	10,42	1,46
9	11,44	1,21
10	10,22	1,53
11	10,90	1,32
12	10,26	1,51
13	10,80	1,37
14	10,32	1,50

5.

	Y	*X*
1	10,68	1,38
2	8,36	1,85
3	8,03	1,95
4	9,26	1,72
5	6,35	2,34
6	6,95	2,07
7	6,58	2,04
8	7,58	2,03
9	6,22	2,35
10	6,93	2,27
11	8,83	1,94
12	9,05	1,64
13	7,25	2,01
14	7,46	2,16
15	10,03	1,84

6.

	Y	*X*
1	12,28	1,07
2	11,44	1,24
3	10,56	1,48
4	11,24	1,29
5	10,14	1,62
6	10,92	1,37
7	9,16	1,52
8	11,46	1,53
9	10,22	1,53
10	11,44	1,21
11	10,22	1,53
12	10,46	1,47
13	10,52	1,43
14	10,26	1,51
15	10,90	1,32

7.

	Y	*X*
1	11,18	1,25
2	10,68	1,38
3	8,36	1,85
4	7,25	2,01
5	7,46	2,16
6	9,26	1,72
7	6,35	2,34
8	7,73	1,92
9	6,34	2,26
10	6,58	2,04
11	7,58	2,03
12	9,56	1,72
13	7,48	2,02
14	6,93	2,27
15	8,83	1,94

8.

	Y	*X*
1	7,22	2,18
2	6,93	2,27
3	6,22	2,35
4	9,56	1,72
5	9,34	1,72
6	6,58	2,04
7	6,95	2,07
8	7,73	1,92
9	10,03	1,84
10	9,26	1,72
11	8,03	1,95
12	7,25	2,01
13	9,05	1,64
14	10,68	1,38
15	10,32	1,50

9.

	Y	*X*
1	10,42	1,46
2	11,46	1,53
3	9,16	1,52
4	10,90	1,43
5	10,82	1,29
6	10,92	1,37
7	10,14	1,62
8	10,62	1,46
9	11,24	1,29
10	10,56	1,48
11	10,60	1,35
12	11,44	1,24
13	12,28	1,07
14	10,04	1,52
15	10,36	1,45

10.

	Y	*X*
1	10,04	1,52
2	10,60	1,35
3	11,24	1,29
4	10,14	1,62
5	10,82	1,89
6	9,16	1,52
7	10,42	1,46
8	11,44	1,21
9	10,22	1,53
10	10,90	1,32
11	10,26	1,51
12	10,80	1,37
13	10,32	1,50
14	10,68	1,38
15	8,36	1,85

11.

	Y	*X*
1	12,28	1,07
2	11,44	1,24
3	10,56	1,48
4	11,24	1,29
5	10,14	1,62
6	10,92	1,37
7	10,90	1,43
8	9,16	1,52
9	10,42	1,46
10	10,22	1,53
11	10,42	1,46
12	10,22	1,53
13	10,90	1,32
14	10,52	1,43
15	10,90	1,30

12.

	Y	*X*
1	10,80	1,37
2	10,48	1,44
3	11,18	1,25
4	10,68	1,38
5	8,36	1,85
6	7,25	2,01
7	8,03	1,95
8	9,26	1,72
9	10,03	1,84
10	7,73	1,92
11	6,95	2,07
12	6,58	2,04
13	7,58	2,03
14	9,56	1,72
15	7,48	2,02

13.

	Y	*X*
1	11,44	1,21
2	10,22	1,46
3	10,90	1,32
4	10,26	1,51
5	10,80	1,37
6	10,32	1,50
7	10,68	1,38
8	8,36	1,85
9	7,25	2,01
10	7,46	2,16
11	10,30	1,84
12	7,73	1,92
13	6,34	2,26
14	9,34	1,72
15	9,56	1,72

14.

	Y	*X*
1	11,44	1,24
2	10,56	1,48
3	10,62	1,46
4	10,92	1,37
5	10,90	1,43
6	11,46	1,53
7	10,22	1,53
8	11,44	1,21
9	10,22	1,53
10	10,90	1,32
11	10,26	1,51
12	10,80	1,37
13	10,32	1,5
14	11,18	1,25
15	8,36	1,85

15.

	Y	*X*
1	9,16	1,52
2	11,46	1,53
3	10,22	1,53
4	11,44	1,21
5	10,22	1,53
6	10,46	1,47
7	10,52	1,43
8	10,26	1,51
9	10,80	1,37
10	10,48	1,44
11	11,18	1,25
12	10,68	1,38
13	8,36	1,85
14	7,25	2,01
15	7,46	2,16

16.

	Y	*X*
1	11,24	1,29
2	10,62	1,46
3	10,14	1,62
4	10,90	1,43
5	9,16	1,52
6	11,46	1,53
7	10,42	1,46
8	11,44	1,21
9	10,42	1,46
10	10,22	1,53
11	10,90	1,32
12	10,52	1,43
13	10,26	1,51
14	10,80	1,37
15	10,48	1,44

17.

	Y	*X*
1	11,06	1,23
2	12,22	1,07
3	11,44	1,24
4	10,60	1,35
5	11,24	1,29
6	10,92	1,37
7	10,82	1,29
8	10,90	1,43
9	11,46	1,53
10	11,44	1,21
11	10,42	1,46
12	10,22	1,53
13	10,90	1,32
14	10,80	1,37
15	10,48	1,44

18.

	Y	*X*
1	10,52	1,43
2	10,26	1,51
3	10,48	1,44
4	10,32	1,50
5	10,68	1,38
6	9,05	1,60
7	8,03	1,95
8	7,46	2,10
9	6,35	2,39
10	7,73	1,92
11	9,34	1,72
12	7,58	2,03
13	6,22	2,35
14	7,48	2,02
15	7,22	2,18

19.

	Y	*X*
1	10,04	1,52
2	10,28	1,07
3	10,56	1,48
4	11,24	1,29
5	10,14	1,62
6	10,92	1,37
7	9,16	1,52
8	11,46	1,53
9	11,44	1,21
10	10,42	1,46
11	10,52	1,43
12	10,26	1,51
13	10,48	1,44
14	10,32	1,50
15	7,25	2,01

20.

	Y	*X*
1	9,05	1,60
2	8,36	1,80
3	8,03	1,93
4	7,46	2,16
5	10,03	1,82
6	6,35	2,39
7	6,95	2,07
8	6,34	2,26
9	9,34	1,72
10	7,58	2,03
11	6,22	2,35
12	7,48	2,02
13	7,22	2,18
14	8,83	1,99
15	11,06	1,23

21.

	Y	*X*
1	11,24	1,29
2	10,62	1,46
3	10,14	1,62
4	10,92	1,37
5	10,82	1,29
6	10,90	1,43
7	11,46	1,53
8	10,42	1,46
9	10,22	1,53
10	10,42	1,46
11	10,22	1,53
12	10,46	1,47
13	10,52	1,43
14	10,26	1,51
15	10,90	1,32

22.

	Y	*X*
1	7,25	2,01
2	8,03	1,93
3	7,46	2,16
4	10,03	1,84
5	6,35	2,34
6	6,95	2,07
7	6,34	2,20
8	6,58	2,04
9	7,58	2,03
10	9,56	1,72
11	6,22	2,35
12	6,93	2,27
13	7,22	2,18
14	8,83	1,94
15	9,05	1,64

23.

	Y	*X*
1	12,28	1,07
2	11,44	1,24
3	10,60	1,35
4	10,62	1,46
5	10,14	1,62
6	10,92	1,37
7	10,90	1,43
8	9,16	1,52
9	11,46	1,53
10	11,44	1,21
11	10,42	1,46
12	10,22	1,53
13	10,52	1,43
14	10,26	1,51
15	10,90	1,32

24.

	Y	*X*
1	10,32	1,50
2	11,18	1,25
3	10,68	1,38
4	7,25	2,01
5	8,03	1,95
6	7,46	2,16
7	6,35	2,34
8	7,73	1,92
9	6,95	2,07
10	9,34	1,72
11	9,56	1,72
12	7,48	2,02
13	6,93	2,27
14	7,22	2,18
15	8,83	1,94

25.

	Y	*X*
1	10,14	1,62
2	10,92	1,37
3	9,16	1,52
4	11,46	1,53
5	11,44	1,21
6	10,42	1,46
7	10,90	1,32
8	10,52	1,43
9	10,90	1,32
10	10,80	1,37
11	11,18	1,25
12	10,68	1,38
13	7,25	2,01
14	8,03	1,95
15	9,26	1,72

26.

	Y	*X*
1	12,28	1,07
2	10,56	1,48
3	11,24	1,29
4	10,62	1,46
5	10,82	1,29
6	10,90	1,43
7	9,16	1,52
8	10,22	1,53
9	11,44	1,21
10	10,42	1,46
11	10,52	1,43
12	10,26	1,51
13	10,90	1,32
14	10,48	1,44
15	10,32	1,50

27.

	Y	*X*
1	11,18	1,25
2	10,68	1,38
3	7,25	2,01
4	8,03	1,95
5	9,26	1,72
6	10,03	1,84
7	7,73	1,92
8	6,95	2,07
9	9,34	1,72
10	7,58	2,03
11	9,56	1,72
12	7,48	2,02
13	6,93	2,27
14	7,22	2,18
15	8,83	1,94

28.

	Y	*X*
1	9,16	1,52
2	11,46	1,53
3	10,22	1,53
4	10,46	1,47
5	10,90	1,32
6	10,52	1,43
7	10,48	1,44
8	10,32	1,50
9	11,18	1,25
10	9,05	1,64
11	8,36	1,85
12	8,03	1,95
13	7,46	2,16
14	10,30	1,84

29.

	Y	*X*
1	11,06	1,23
2	11,44	1,24
3	10,60	1,35
4	10,56	1,48
5	10,14	1,62
6	10,92	1,37
7	10,82	1,29
8	11,46	1,53
9	10,42	1,46
10	10,22	1,53
11	10,46	1,47
12	10,90	1,32
13	10,52	1,43
14	10,32	1,50
15	11,18	1,25

30.

	Y	*X*
1	11,46	1,53
2	10,42	1,46
3	11,44	1,21
4	10,42	1,46
5	10,26	1,51
6	10,9	1,32
7	10,48	1,44
8	11,18	1,25
9	9,05	1,64
10	7,25	2,01
11	8,03	1,95
12	7,46	2,16
13	10,03	1,84
14	7,73	1,92
15	7,58	2,03

***Потрібно*:**

1. Знайти точкові статистичні оцінки β_0^*, β_1^* для параметрів β_0, β_1 функції регресії $y_i = \beta_0 + \dfrac{\beta_1}{x_i}$.

2. З надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для функції регресії $y_i = \beta_0 + \dfrac{\beta_1}{x_i}$.

3. Обчислити *R*.

4. Нелінійна модель регресії за параметрами

Обробка звітних даних по регіону дала залежності обсягу виробництва продукції *Y*, млрд грн., від основного капіталу *X*, млрд грн., робочої сили *Z*, млн осіб, і часу *t*, які наведено статистичними розподілами:

1.

	Y	*X*	*Z*	*t*
1	1,7997	3,00	2,01	1
2	1,8548	3,10	2,04	2
3	1,9640	3,15	2,06	3
4	2,0222	3,21	2,09	4
5	2,1190	3,32	2,11	5
6	2,1899	3,39	2,13	6
7	2,2490	3,43	2,15	7
8	2,3056	3,45	2,18	8
9	2,3643	3,47	2,21	9
10	2,4253	3,52	2,22	10
11	2,4943	3,57	2,24	11
12	2,5648	3,62	2,26	12
13	2,6713	3,74	2,28	13
14	2,7515	3,79	2,31	14
15	2,8121	3,81	2,33	15

2.

	Y	*X*	*Z*	*t*
1	2,9368	3,94	2,36	1
2	3,0125	3,97	2,39	2
3	3,0951	4,01	2,42	3
4	3,1459	4,05	2,40	4
5	3,2202	4,12	2,39	5
6	3,2688	4,19	2,35	6
7	3,3077	4,25	2,29	7
8	3,3606	4,32	2,25	8
9	3,3760	4,38	2,17	9
10	3,4224	4,41	2,15	10
11	3,4423	4,45	2,08	11
12	3,3735	4,52	2,03	12
13	3,4693	4,27	2,01	13
14	3,3252	4,12	1,93	14
15	3,1578	4,01	1,89	15

3.

	Y	*X*	*Z*	*t*
1	3,4138	3,95	1,84	1
2	3,0421	3,76	1,82	2
3	3,1040	3,51	1,78	3
4	2,9011	3,48	1,75	4
5	3,0358	3,43	1,72	5
6	2,7612	3,26	1,69	6
7	2,8282	3,02	1,68	7
8	2,7345	2,97	1,68	8
9	2,8554	2,93	1,69	9
10	2,8066	2,91	1,72	10
11	2,9236	2,84	1,75	11
12	2,8961	2,85	1,78	12
13	2,9823	2,74	1,81	13
14	2,9029	2,68	1,84	14

4.

	Y	*X*	*Z*	*t*
1	3,1073	2,73	1,87	1
2	3,1157	2,76	1,92	2
3	3,3254	2,81	1,95	3
4	3,3532	2,89	1,97	4
5	3,4121	2,96	2,03	5
6	1,7997	3,00	2,01	6
7	1,8548	3,10	2,04	7
8	1,9640	3,15	2,06	8
9	2,0222	3,21	2,09	9
10	2,1190	3,32	2,11	10
11	2,1899	3,39	2,13	11
12	2,2490	3,43	2,15	12
13	2,3056	3,45	2,18	13
14	2,3643	3,47	2,21	14

5.

	Y	*X*	*Z*	*t*
1	3,0421	3,76	1,82	1
2	3,1040	3,51	1,78	2
3	2,9011	3,48	1,75	3
4	3,0358	3,43	1,72	4
5	2,7612	3,26	1,69	5
6	2,8282	3,02	1,68	6
7	2,7345	2,97	1,68	7
8	2,8554	2,93	1,69	8
9	2,8066	2,91	1,72	9
10	2,9236	2,84	1,75	10
11	2,8961	2,85	1,78	11
12	2,9029	2,68	1,84	12
13	3,1073	2,73	1,87	13
14	3,1157	2,76	1,92	14
15	3,3254	2,96	2,03	15

6.

	Y	*X*	*Z*	*t*
1	3,3532	2,89	1,97	1
2	3,4121	2,96	2,03	2
3	1,7997	3,00	2,01	3
4	1,8548	3,10	2,04	4
5	1,9640	3,15	2,06	5
6	2,0222	3,21	2,09	6
7	2,1190	3,32	2,11	7
8	2,1899	3,39	2,15	8
9	2,2490	3,43	2,13	9
10	2,3056	3,45	2,18	10
11	2,3643	3,47	2,21	11
12	2,4253	3,52	2,22	12
13	2,4943	3,57	2,24	13
14	2,5648	3,62	2,26	14
15	2,6713	3,74	2,28	15

7.

	Y	*X*	*Z*	*t*
1	3,0951	4,01	2,42	1
2	3,1459	4,05	2,40	2
3	3,2202	4,12	2,39	3
4	3,2688	4,18	2,35	4
5	3,3077	4,25	2,29	5
6	3,3606	4,32	2,25	6
7	3,3760	4,38	2,17	7
8	3,4224	4,41	2,15	8
9	3,4423	4,45	2,08	9
10	3,3735	4,52	2,03	10
11	3,4693	4,27	2,01	11
12	3,3252	4,15	1,97	12
13	3,4092	4,12	1,93	13
14	3,1578	4,01	1,89	14
15	3,4138	3,95	1,84	15

8.

	Y	*X*	*Z*	*t*
1	3,0421	3,76	1,82	1
2	3,1040	3,51	1,78	2
3	2,9011	3,48	1,75	3
4	3,0358	3,43	1,72	4
5	2,7612	3,26	1,69	5
6	2,8282	3,02	1,68	6
7	2,7345	2,97	1,68	7
8	2,8554	2,93	1,69	8
9	2,8066	2,91	1,72	9
10	2,9236	2,84	1,75	10
11	2,8961	2,85	1,78	11
12	2,9823	2,74	1,81	12
13	2,9029	2,68	1,84	13
14	3,1073	2,76	1,92	14
15	3,1157	2,81	1,95	15

9.

	Y	*X*	*Z*	*t*
1	1,8548	3,10	2,04	1
2	2,0222	3,21	2,09	2
3	2,1899	3,39	2,13	3
4	2,0356	3,45	2,18	4
5	2,4253	3,52	2,22	5
6	2,5648	3,57	2,24	6
7	2,7515	3,79	2,31	7
8	2,9368	3,94	2,36	8
9	3,0951	4,01	2,42	9
10	3,2202	4,12	3,39	10
11	3,3077	4,25	2,29	11
12	3,3760	4,38	2,17	12
13	3,4423	4,45	2,08	13
14	3,4693	4,27	2,01	14
15	3,4092	4,12	1,93	15

10.

	Y	*X*	*Z*	*t*
1	3,4138	3,95	1,84	1
2	3,1040	3,51	1,78	2
3	3,0358	3,43	1,72	3
4	2,8282	3,02	1,68	4
5	2,8554	2,93	1,69	5
6	2,9236	2,84	1,75	6
7	2,9823	2,74	1,81	7
8	3,1073	2,73	1,87	8
9	3,3254	2,81	1,95	9
10	3,4121	2,96	2,03	10
11	3,0421	3,76	1,82	11
12	2,9011	3,48	1,75	12
13	2,7612	3,26	1,69	13
14	2,7345	2,97	1,68	14
15	2,8066	2,91	1,75	15

11.

	Y	*X*	*Z*	*t*
1	2,9368	3,94	2,36	1
2	3,0125	3,97	2,39	2
3	3,0951	4,01	2,42	3
4	3,1459	4,05	2,40	4
5	3,2202	4,12	2,39	5
6	3,2688	4,18	2,35	6
7	3,3077	4,25	2,29	7
8	3,3606	4,32	2,25	8
9	3,3760	4,38	2,17	9
10	3,4224	4,41	2,15	10
11	3,4423	4,45	2,08	11
12	3,3735	4,52	2,03	12
13	3,4693	4,27	2,01	13
14	3,4092	4,12	1,93	14
15	3,1578	4,01	1,89	15

12.

	Y	*X*	*Z*	*t*
1	3,1459	4,05	2,40	1
2	3,2688	4,18	2,35	2
3	3,3606	4,32	2,25	3
4	3,4224	4,41	2,15	4
5	3,3735	4,52	2,03	5
6	3,3252	4,15	1,97	6
7	3,1578	4,01	1,89	7
8	3,0421	3,76	1,82	8
9	2,9011	3,48	1,75	9
10	2,7612	3,26	1,69	10
11	2,7345	2,97	1,68	11
12	2,8066	2,91	1,72	12
13	2,8961	2,85	1,78	13
14	2,9029	2,68	1,84	14
15	3,1157	2,76	1,92	15

13.

	Y	*X*	*Z*	*t*
1	1,9640	3,15	2,06	1
2	2,1190	3,32	2,11	2
3	2,2490	3,43	2,15	3
4	2,3643	3,47	2,21	4
5	2,4943	3,57	2,24	5
6	2,6713	3,74	2,28	6
7	2,8121	3,81	2,33	7
8	3,0125	3,97	2,39	8
9	3,1459	4,05	2,40	9
10	3,2688	4,18	2,35	10
11	3,3606	4,32	2,15	11
12	3,4224	4,41	2,15	12
13	3,3735	4,52	2,03	13
14	3,3252	4,15	1,97	14
15	3,1578	4,01	1,89	15

14.

	Y	*X*	*Z*	*t*
1	3,4138	3,95	1,84	1
2	3,1040	3,51	1,78	2
3	3,0358	3,43	1,72	3
4	2,8282	3,02	1,68	4
5	2,8554	2,93	1,69	5
6	2,8066	2,91	1,72	6
7	2,9236	2,84	1,75	7
8	2,8961	2,85	1,78	8
9	2,9823	2,74	1,81	9
10	2,9029	2,68	1,84	10
11	3,1073	2,73	1,87	11
12	3,1157	2,81	1,95	12
13	3,3532	2,89	1,97	13
14	3,4121	2,96	2,03	14
15	3,0421	3,76	1,82	15

15.

	Y	*X*	*Z*	*t*
1	2,0222	3,21	2,09	1
2	2,1899	3,39	2,13	2
3	2,3056	3,45	2,18	3
4	2,4253	3,52	2,22	4
5	2,5648	3,62	2,26	5
6	2,7515	3,79	2,31	6
7	2,8121	3,81	2,33	7
8	3,0125	3,97	2,39	8
9	3,0951	4,01	2,42	9
10	3,2202	4,12	2,39	10
11	3,2688	4,18	2,35	11
12	3,3606	4,32	2,15	12
13	3,3760	4,38	2,17	13
14	3,4423	4,45	2,08	14
15	3,3735	4,52	2,03	15

16.

	Y	*X*	*Z*	*t*
1	3,4693	4,27	2,01	1
2	3,4092	4,12	1,93	2
3	3,4138	3,95	1,84	3
4	3,1040	3,51	1,78	4
5	2,9011	3,48	1,75	5
6	2,7612	3,26	1,69	6
7	2,8282	3,02	1,68	7
8	2,8554	2,93	1,69	8
9	2,8066	2,91	1,72	9
10	2,8961	2,85	1,78	10
11	2,9823	2,74	1,81	11
12	3,1073	2,73	1,87	12
13	3,3254	2,81	1,95	13
14	3,3532	2,89	1,97	14
15	3,1040	3,51	1,78	15

17.

	Y	*X*	*Z*	*t*
1	2,1899	3,39	2,13	1
2	2,2490	3,43	2,15	2
3	2,3056	3,45	2,18	3
4	2,4943	3,57	2,24	4
5	2,5648	3,62	2,26	5
6	2,6713	3,74	2,28	6
7	2,9368	3,94	2,36	7
8	3,0125	3,97	2,39	8
9	3,0951	4,01	2,42	9
10	3,2688	4,18	2,35	10
11	3,3077	4,25	2,29	11
12	3,3606	4,32	2,25	12
13	3,4224	4,41	2,15	13
14	3,4423	4,45	2,08	14
15	3,3735	4,52	2,03	15

18.

	Y	*X*	*Z*	*t*
1	3,3532	2,89	1,97	1
2	3,4121	2,96	2,03	2
3	1,7997	3,00	2,01	3
4	1,9640	3,15	2,06	4
5	2,0222	3,21	2,09	5
6	2,1899	3,39	2,13	6
7	2,2490	3,43	2,15	7
8	2,3056	3,45	2,18	8
9	2,4943	3,57	2,24	9
10	2,5648	3,62	2,26	10
11	2,6713	3,74	2,28	11
12	2,7515	3,79	2,31	12
13	2,9368	3,94	2,36	13
14	3,0125	3,97	2,39	14
15	3,0951	4,01	2,42	15

19.

	Y	*X*	*Z*	*t*
1	3,4693	4,27	2,01	1
2	3,3252	4,15	1,97	2
3	3,4092	4,12	1,93	3
4	3,4138	3,95	1,84	4
5	3,0421	3,76	1,82	5
6	3,1040	3,51	1,78	6
7	2,7612	3,26	1,69	7
8	2,8282	3,02	1,68	8
9	2,7345	2,97	1,68	9
10	2,9236	2,84	1,75	10
11	2,8961	2,85	1,78	11
12	2,9823	2,74	1,81	12
13	3,1073	2,73	1,87	13
14	3,1157	2,76	1,92	14
15	3,3254	2,81	1,95	15

20.

	Y	*X*	*Z*	*t*
1	2,9368	3,94	2,36	1
2	3,0125	3,97	2,39	2
3	3,0951	4,01	2,42	3
4	3,2688	4,18	2,35	4
5	3,3077	4,25	2,29	5
6	3,3606	4,32	2,25	6
7	3,4224	4,41	2,15	7
8	3,4423	4,45	2,08	8
9	3,3735	4,52	2,03	9
10	3,3252	4,27	2,01	10
11	3,4092	4,12	1,93	11
12	3,1578	4,01	1,89	12
13	2,7345	2,97	1,68	13
14	2,8066	2,91	1,72	14
15	2,9823	2,74	1,81	15

21.

	Y	*X*	*Z*	*t*
1	3,4224	4,41	2,15	1
2	3,4423	4,45	2,08	2
3	3,4693	4,27	2,01	3
4	3,3252	4,15	1,97	4
5	3,1578	4,01	1,89	5
6	3,4138	3,95	1,84	6
7	2,9011	3,48	1,75	7
8	3,0358	3,43	1,72	8
9	2,8282	3,02	1,68	9
10	2,7334	2,97	1,68	10
11	2,8066	2,91	1,72	11
12	2,9236	2,84	1,75	12
13	2,9823	2,74	1,81	13
14	3,1073	2,73	1,87	14
15	3,1157	2,76	1,92	15

22.

	Y	*X*	*Z*	*t*
1	3,3252	4,15	1,97	1
2	3,1578	4,01	1,89	2
3	3,4138	3,95	1,84	3
4	3,1040	3,51	1,78	4
5	3,0358	3,43	1,72	5
6	2,7612	3,26	1,69	6
7	2,7345	2,97	1,68	7
8	2,8066	2,91	1,72	8
9	2,9236	2,84	1,75	9
10	2,8961	2,85	1,78	10
11	2,9029	2,68	1,84	11
12	3,1073	2,73	1,87	12
13	3,3254	2,81	1,95	13
14	3,4121	2,96	0,03	14
15	2,3056	3,45	2,18	15

23.

	Y	*X*	*Z*	*t*
1	1,7997	3,00	2,01	1
2	1,9640	3,15	2,06	2
3	2,0222	3,21	2,09	3
4	2,1899	3,39	2,13	4
5	2,3056	3,45	2,18	5
6	2,3643	3,47	2,21	6
7	2,4943	3,57	2,24	7
8	2,5648	3,62	2,26	8
9	2,7515	3,79	2,31	9
10	2,9368	3,94	2,36	10
11	3,0125	3,97	2,39	11
12	3,1459	4,05	2,40	12
13	3,2688	4,18	2,35	13
14	3,3606	4,62	2,25	14
15	3,3760	4,38	2,17	15

24.

	Y	*X*	*Z*	*t*
1	2,4253	3,52	2,22	1
2	2,5648	3,62	2,26	2
3	2,6713	3,74	2,28	3
4	2,8121	3,81	2,33	4
5	3,0951	4,01	2,42	5
6	3,1459	4,05	2,40	6
7	3,2688	4,18	2,35	7
8	3,3606	4,32	2,25	8
9	3,3760	4,38	2,17	9
10	3,4423	4,45	2,08	10
11	3,3735	4,52	2,03	11
12	3,3252	4,15	1,97	12
13	3,1578	4,01	1,89	13
14	3,0421	3,76	1,82	14
15	3,1040	3,51	1,78	15

25.

	Y	*X*	*Z*	*t*
1	2,1190	3,32	2,11	1
2	2,1899	3,39	2,13	2
3	2,3056	3,45	2,18	3
4	2,3643	3,47	2,21	4
5	2,4943	3,57	2,24	5
6	2,5648	3,62	2,26	6
7	2,7515	3,79	2,31	7
8	2,8121	3,81	2,33	8
9	3,0125	3,97	2,39	9
10	3,0951	4,01	2,42	10
11	3,2202	4,12	2,39	11
12	3,2688	4,18	2,35	12
13	3,3606	4,32	2,25	13
14	3,3760	4,38	2,17	14
15	3,4423	4,45	2,08	15

26.

	Y	*X*	*Z*	*t*
1	2,4253	3,52	2,22	1
2	3,4943	3,57	2,24	2
3	2,5648	3,62	2,26	3
4	2,8121	3,81	2,33	4
5	2,9368	3,94	2,36	5
6	3,0125	3,97	2,39	6
7	3,2688	4,18	2,35	7
8	3,3077	4,25	2,29	8
9	3,3606	4,32	2,25	9
10	3,4224	4,41	2,15	10
11	3,4423	4,45	2,08	11
12	3,3735	4,52	2,03	12
13	3,3252	4,15	1,97	13
14	3,4092	4,12	1,93	14
15	3,1578	4,01	1,89	15

27.

	Y	*X*	*Z*	*t*
1	3,4092	4,12	1,93	1
2	3,1578	4,01	1,89	2
3	3,0421	3,76	1,82	3
4	2,9011	3,48	1,75	4
5	3,0358	3,43	1,72	5
6	2,8282	3,02	1,68	6
7	2,7345	2,97	1,68	7
8	2,8066	2,91	1,72	8
9	2,9236	2,84	1,75	9
10	2,9823	2,74	1,81	10
11	2,9029	2,68	1,84	11
12	3,1157	2,76	1,92	12
13	3,3254	2,81	1,95	13
14	3,3532	2,89	1,97	14
15	3,4121	2,96	2,03	15

28.

	Y	*X*	*Z*	*t*
1	3,4138	3,95	1,84	1
2	3,0421	3,76	1,82	2
3	2,9011	3,48	1,75	3
4	3,0358	3,43	1,72	4
5	2,8282	3,02	1,68	5
6	2,7345	2,97	1,68	6
7	2,8066	2,91	1,72	7
8	2,9236	2,84	1,75	8
9	2,9823	2,74	1,81	9
10	2,9029	2,68	1,84	10
11	3,1157	2,76	1,92	11
12	3,3254	2,81	1,95	12
13	3,4121	2,96	2,03	13
14	1,7997	3,00	2,01	14
15	1,9640	3,15	2,06	15

29.

	Y	*X*	*Z*	*t*
1	2,5648	3,62	2,26	1
2	2,6713	3,74	2,28	2
3	2,8121	3,81	2,33	3
4	2,9368	3,94	2,36	4
5	3,0951	4,01	2,42	5
6	3,1459	4,05	2,40	6
7	3,2688	4,18	2,35	7
8	3,3077	4,25	2,39	8
9	3,3760	4,38	2,17	9
10	3,4224	4,41	2,15	10
11	3,3735	4,52	2,03	11
12	3,4693	4,27	2,01	12
13	3,4092	4,12	1,93	13
14	3,1578	4,01	1,89	14
15	3,0421	3,76	1,82	15

30.

	Y	*X*	*Z*	*t*
1	2,1190	3,32	2,11	1
2	2,1899	3,39	2,13	2
3	2,4943	3,57	2,24	3
4	2,5648	3,62	2,26	4
5	2,9368	3,94	2,36	5
6	3,0125	3,97	2,39	6
7	3,2202	4,12	2,39	7
8	3,2688	4,18	2,35	8
9	2,7612	3,26	1,69	9
10	2,8282	3,02	1,68	10
11	2,9236	2,84	1,75	11
12	2,8961	2,85	1,78	12
13	3,1073	2,73	1,92	13
14	3,1157	2,76	1,92	14
15	3,3532	2,89	1,97	15

***Потрібно*:**

1. Визначити точкові статистичні оцінки β_0^*, β_1^*, β_2^*, β_3^* для параметрів β_0, β_1, β_2, β_3 функції регресії $y_i = \beta_0 x^{\beta_1} z^{\beta_2} e^{\beta_3 t}$.

2. З надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для функції регресії $y_i = \beta_0 x_i^{\beta_1} z_i^{\beta_2} e^{\beta_3 t}$.

3. Обчислити *R*.

ДОДАТКИ ДО ТЕОРІЇ ЙМОВІРНОСТЕЙ
І МАТЕМАТИЧНОЇ СТАТИСТИКИ

Додаток 1[*]

ВИПАДКОВІ ПРОЦЕСИ.
МАРКОВСЬКІ ВИПАДКОВІ ПРОЦЕСИ ТА ЕЛЕМЕНТИ ТЕОРІЇ МАСОВОГО ОБСЛУГОВУВАННЯ (ТМО)

1. Загальна інформація

Теорія випадкових процесів є математичною наукою, що вивчає закономірності випадкових подій у їх динаміці. Ця теорія (за іншою термінологією — теорія випадкових функцій) вивчає процеси, розвиток яких наперед точно неможливо передбачити. Така невизначеність (непередбачуваність) зумовлена дією випадкових факторів на розвиток процесу.

Математичною моделлю випадкового процесу є певна функція $X = X(t)$ від дійсного аргументу $t \in T$ (часу), значення якої при кожному фіксованому $T = t_i$ є випадковою величиною. Саме поняття випадкового процесу (випадкової функції) є узагальнюючим поняттям випадкової величини.

Отже, випадковим процесом $X = X(t)$ називають такий процес, коли при будь-якому можливому значенні $T = t_i$ випадкова функція $X = X(t_i)$ утворює випадкову величину.

При $t = t_i$ ми дістанемо випадкову величину $X(t) = (x_1(t_i), x_2(t_i), ..., x_k(t_i), ...)$, яку називають перерізом випадкового процесу. Чим більше перерізів буде розглянуто, тим детальніше уявлення ми будемо мати про випадковий процес.

Випадкові процеси можна класифікувати за тими чи іншими ознаками.

Елементарною класифікацією випадкових процесів є класифікація за ознаками часу та стану. Випадковий процес називають *процесом із дискретною змінною часу*. Якщо система, в якій він здійснюється, може змінювати свій стан тільки в моменти часу $t_1, t_2, ..., t_k, ...$, кількість яких є обмеженою, або зліченною.

Прикладом процесів із дискретним часом є: 1. Робота технічного пристрою, який контролюється в моменти часу $t_1, t_2, \ldots$; 2. Процес роботи ЕОМ, що може змінювати свій стан у моменти часу $t_1, t_2, ...$

Розглядаючи одновимірний випадковий процес $X(t)$ у дискретні моменти часу $t_1, t_2, ...$, дістаємо послідовність випадкових величин $X(t_1), X(t_2),...$

---

[*] Додаток до посібника: *Жлуктенко В. І., Наконечний С. І.* Теорія ймовірностей і математична статистика: Навч.-метод. посібник: У 2-х ч. — Ч. І. Теорія ймовірностей. — К.: КНЕУ, 2000. — 304 с.

Якщо експеримент здійснився, то випадковий процес позбавляється елемента випадковості і функція $X = X(t)$ набуває певного вигляду, тобто $X = X(t)$ стає невипадковою функцією від аргументу t. У цьому разі $X = X(t)$ називають *реалізацією випадкового процесу*.

Отже, реалізацією випадкового процесу $X(t)$ називається невипадкова функція $x = x(t)$, в яку реалізується випадковий процес $X = X(t)$.

Доходимо висновку, що випадковий процес є множиною реалізацій $x_i(t)$, а саме:

$$X(t) = (x_1(t), x_2(t), ..., x_k(t), ...).$$

Множина реалізацій випадкового процесу є базою експерименту, на основі якого можна дістати його характеристики. Загалом реалізацію випадкового процесу зображено на рис. 159.

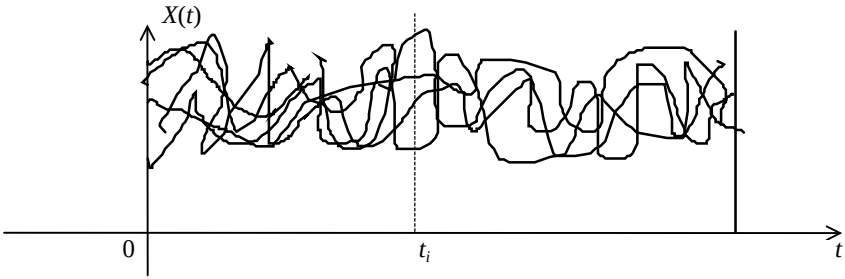


Рис. 159

Випадковий процес $X(t)$ називають *процесом із неперервним часом*, якщо перехід системи з одного стану в інший може бути здійснений у будь-який можливий момент часу t за певний спостережуваний період. У цьому разі множина можливих станів системи є незліченною.

Прикладами таких систем є:

1) $X(t)$ — число відказів технічного пристрою від початку роботи до моменту часу t;

2) число викликів, що надійшли в лікарню швидкої допомоги до моменту часу t, та ін.

Одновимірний випадковий процес $X(t)$ називають *процесом із неперервним станом*, якщо його переріз у будь-який можливий момент часу t утворює неперервну випадкову величину, а саме: множина її значень є незліченною.

Аналогічно многомірний випадковий процес називають процесом із неперервним станом, якщо в будь-який можливий мо-

мент часу t множина його можливих станів, через який проходить цей процес, є незліченною.

Прикладами випадкових процесів із неперервним станом є:

1) напруга $U(t)$ електроживлення обчислювального центру;

2) тиск газу $P(t)$ в певному резервуарі тощо.

Випадковий процес називають *процесом із дискретним станом*, якщо в будь-який можливий момент часу t множина його станів є скінченною, або зліченною, що, по суті, одне і те саме. Його переріз у будь-який можливий момент часу t утворює дискретну випадкову величину (у багатовимірному випадку — множину випадкових величин).

Прикладами таких процесів є:

1) кількість об'єктів певної технічної системи, які можуть відмовити в роботі в певний фіксований момент часу t;

2) кількість пасажирів, які проходять через станцію метро в певний момент часу t, тощо.

2. Закони розподілу та основні числові характеристики випадкових процесів

Оскільки за будь-якого можливого значення аргументу t переріз випадкового процесу $X(t)$ утворює випадкову величину, то вона має закон розподілу ймовірностей, який можна подати функцією розподілу ймовірностей, а саме:

$$F(t;x) = P(X(t) < x). \qquad (574)$$

Функція розподілу ймовірностей $F(t; x)$, як бачимо, залежить від двох аргументів — t і x.

Ця функція називається одновимірним законом розподілу випадкового процесу $X(t)$.

Якщо зафіксувати два значення аргументу t_1 і t_2, то в цьому разі два перерізи утворюють систему двох випадкових величин ($X(t_1)$, $X(t_2)$). Функція розподілу системи буде

$$F(t_1,t_2,x_1,x_2) = P(X(t_1) < x_1, X(t_2) < x_2). \qquad (575)$$

Отже, функція розподілу ймовірностей для двох перерізів уже залежатиме від чотирьох аргументів: t_1, t_2, x_1, x_2.

Теоретично можна необмежено збільшувати число перерізів і при цьому одержувати все повнішу характеристику для випадкового процесу $X(t)$. Але використовувати на практиці ці характеристики, які залежать від багатьох аргументів, дуже незручно.

Тому практично більше ніж два перерізи використовують рідко.

Для неперервного випадкового процесу $X(t)$ за фіксованого значення t_1 закон розподілу ймовірностей для неперервної випадкової величини $X(t_1)$, що утворює переріз, можна подати щільністю ймовірностей $f(t_1, x)$, а для двох перерізів закон розподілу системи можна подати щільністю ймовірностей $f(t_1, t_2; x_1, x_2)$, яку в цьому разі називають *двовимірною*.

На практиці, як правило, застосовуються числові характеристики випадкових процесів, що є аналогічними числовим характеристикам випадкових величин, а саме: математичне сподівання, дисперсія, середнє квадратичне відхилення, кореляційний момент та коефіцієнт кореляції.

На відміну від числових характеристик випадкових величин, які є сталими величинами, характеристики випадкових процесів будуть невипадковими функціями її аргументів.

Математичним сподіванням випадкового процесу $X(t)$ називається невипадкова функція $M_x(t)$, яка за будь-якого можливого значення аргументу t дорівнює математичному сподіванню, відповідному перерізу випадкового процесу.

Якщо переріз випадкового процесу при фіксованому значенні t утворює дискретну випадкову величину з рядом розподілу $X(t)$:

$x_1(t)$	$x_2(t)$	$x_3(t)$	.....	$x_k(t)$	.....
$p_1(t)$	$p_2(t)$	$p_3(t)$	.....	$p_k(t)$	.....

то

$$M_x(t) = M(X(t)) = \sum x_i(t) \cdot p_i(t) . \qquad (576)$$

Тут $x_1(t)$, $x_2(t)$, $x_3(t)$,..., $x_k(t)$ є значеннями, яких набуває випадкова величина $X(t)$ в перерізі при заданому t, а $p_1(t)$, $p_2(t)$, $p_3(t)$, ..., $p_k(t)$ — відповідні їм імовірності.

Якщо переріз випадкового процесу $X(t)$ утворює неперервну випадкову величину зі щільністю $f(t; x)$, то

$$M_x(t) = \int_{-\infty}^{\infty} x f(t; x) dx . \qquad (577)$$

Формули для обчислення інших числових характеристик для дискретного і неперервного перерізів матимуть вигляд:

$$D_x(t) = \sum x_i^2(t) p_i(t) - M_x^2(t) ; \qquad (578)$$

$$D_x(t) = \int_{-\infty}^{\infty} x^2 f(t; x) dx - M_x^2(t) ; \qquad (579)$$

$$\sigma_x(t) = \sqrt{D_x(t)}\ . \qquad (580)$$

Якщо розглянути два випадкові процеси $X_1(t)$, $X_2(t)$, реалізації яких зображено на рис. 160, 161,

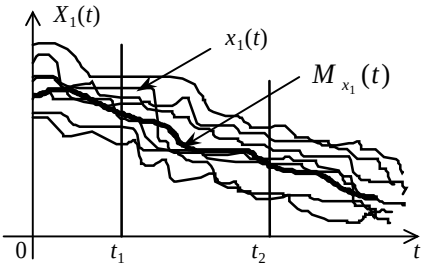


Рис. 160

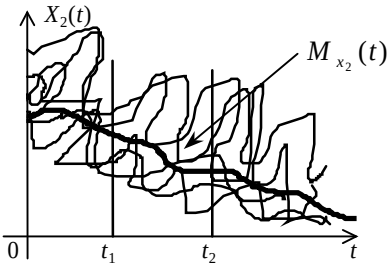


Рис. 161

то математичні сподівання і дисперсії їх майже однакові. Але якщо випадковий процес $X_1(t)$ має плавний характер змін своїх реалізацій, то $X_2(t)$ має різко виражену коливальну структуру таких реалізацій. Отже, для випадкового процесу $X_1(t)$ спостерігається більша передбачуваність реалізацій, ніж для випадкового процесу $X_2(t)$.

Так, для випадкового процесу $X_1(t)$, якщо в будь-який момент часу t певна реалізація $x_1(t)$ буде більшою ніж $M_{x_1}(t)$, то з великою ймовірністю можна очікувати, що в наступні моменти часу реалізація $x_1(t) > M_{x_1}(t)$, що не можна стверджувати для випадкового процесу $X_2(t)$.

Отже, для випадкового процесу $X_1(t)$ спостерігається тісна ймовірна залежність між двома перерізами в моменти часу t_1 і t_2, ніж для випадкового процесу $X_2(t)$.

Оскільки міра лінійної залежності між двома випадковими величинами визначається кореляційним моментом та коефіцієнтом кореляції, то аналогічно і для випадкових процесів застосовують ці самі характеристики.

Кореляційні моменти для дискретних та неперервних перерізів певного випадкового процесу $X(t)$ обчислюються за формулами:

$$K_{x'x''}(t_1,t_2) = \sum_{i=1}\sum_{j=1} x_i'(t_1)\cdot x''(t_2)\cdot p_{ij}(t_1,t_2) - M_{x'}(t_1)\cdot M_{x''}(t_2)\,, \qquad (581)$$

$$K_{x'x''}(t_1,t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_i'(t_1)\cdot x''(t_2)\, f(t_1,t_2;x',x'')dx'dx'' - $$
$$- M_{x'}(t_1)M_{x''}(t_2)\,. \qquad (582)$$

Отже, $K_{x'x''}(t_1, t_2)$ є невипадковою функцією двох аргументів t_1 і t_2.

$K_{x'x''}(t_1, t_2)$ зменшується зі збільшенням різниці $t_2 - t_1$ для випадкового процесу $X_1(t)$ повільніше, ніж для випадкового процесу $X_2(t)$.

При цьому:

1) $K_{x'x''}(t_1, t_2) = K_x(t) = D_x(t)$; при $t_1 = t_2 = t$;

2) $K_{x'x''}(t_1, t_2) = K_{x''x'}(t_2, t_1)$;

3) $K_{x'x''}(t_1, t_2) \geq 0$.

Тісноту лінійної залежності двох перерізів випадкового процесу $X(t)$ вимірюють нормованою кореляційною функцією, яка дорівнює

$$r_x(t_1; t_2) = \frac{K_x(t_1, t_2)}{\sigma_x(t_1) \cdot \sigma_x(t_2)}. \qquad (583)$$

При цьому:

1) $r_x(t, t) = 1$ при $t_1 = t_2 = t$;

2) $r_x(t_1, t_2) = r_x(t_2, t_1)$;

3) $|r_x(t_1, t_2)| \leq 1$.

Приклад 1. Випадкова функція має такий вигляд:

$$Y = X \cdot e^{-\lambda t}, t > 0,$$

де X — випадкова величина, що має закон розподілу $N(a; \sigma)$, де $a > 0$.

Визначити: $M_y(t)$, $D_y(t)$, $\sigma_y(t)$, $K_y(t_1, t_2)$, $r_y(t_1; t_2)$.

Розв'язання. Випадкова величина X має загальний нормальний закон розподілу зі значеннями параметрів $M(x) = a > 0$, $\sigma(x) = \sigma$.

$$M_y(t) = M(Xe^{-\lambda t}) = e^{-\lambda t} M(X) = ae^{-\lambda t};$$

$$D_y(t) = D(Xe^{-\lambda t}) = e^{-\lambda t} D(X) = \sigma^2 e^{-\lambda t};$$

$$\sigma_y(t) = \sqrt{D_y(t)} = \sigma \cdot e^{-\lambda t};$$

$$K_y(t_1, t_2) = M((Y' - M(Y'))(Y'' - M(Y''))) =$$

$$= M(Xe^{-\lambda t_1} - e^{-\lambda t_1} M(X))(Xe^{-\lambda t_2} - e^{-\lambda t_2} M(X)) =$$

$$= e^{-\lambda(t_1+t_2)} M(X - M(X))(X - M(X)) = e^{-\lambda(t_1+t_2)} M(X - M(X))^2 =$$

$$= \sigma^2 e^{-\lambda(t_1+t_2)}.$$

Отже, $K_y(t_1, t_2) = \sigma^2 e^{-\lambda(t_1+t_2)}$.

Оскільки $\sigma_y(t_1) = \sigma \cdot e^{-\lambda t_1}$, $\sigma_y(t_2) = e^{-\lambda t_2}$, то

$$r_y(t_1;\ t_2) = \frac{K_x(t_1;t_2)}{\sigma_x(t_1)\cdot\sigma_x(t_2)} = \frac{\sigma^2 e^{-\lambda(t_1+t_2)}}{\sigma\cdot e^{-\lambda t_1}\cdot\sigma\cdot e^{-\lambda t_2}} = 1.$$

Отже, $r_y(t_1;t_2) = 1$.

3. Марковські випадкові процеси. Ланцюги Маркова

Серед випадкових процесів, що широко застосовуються для створення стохастичних (імовірних) моделей, котрі описують процеси функціонування певних систем технічного, економічного, екологічного та соціального профілю, центральне місце належить марковським.

Випадковий процес $X(t)$ називають марковським, якщо за будь-якого можливого значення часу $t = t_1$ значення випадкової величини $x(t_1)$ не залежить від того, яких значень ця величина набувала для $t < t_1$, тобто, процес у момент часу $t = t_1$ не залежить від його поведінки в більш ранні моменти часу $t < t_1$.

Марковський процес $X(t)$ називають однорідним, якщо закономірності його поведінки на будь-якому проміжку часу ΔT не залежать від розміщення цього інтервалу на часовій осі.

Нехай $X(t)$ — однорідний марковський процес із обмеженим, або зліченним, числом можливих станів $i = 0, 1, 2, 3, ..., n, ...$

Якщо аргумент t набуває лише значення 0, 1, 2, 3, ..., то в цьому разі матимемо послідовність переходів $x(0) \to x(1) \to$ $\to x(2) \to x(3) \to .....$

Такий процес послідовностей переходів називають *ланцюгом Маркова*.

При розробленні теорії ланцюгів Маркова часто дотримуються іншої термінології, а саме: розглядається певна фізична система S, яка в кожний момент часу може перебувати в одному з несумісних станів $A_1, A_2, A_3, ..., A_k, ...$ і змінювати свій стан лише в моменти часу $t_1, t_2, t_3, ..., t_k, ...$

Процес переходу систем S утворює ланцюг Маркова, якщо ймовірність перейти в стан A_j в момент часу t ($t_k < t < t_{k+1}$) залежить лише від того, в якому стані система перебувала в момент часу t' ($t_{k-1} < t' < t_k$), і не залежить від стану системи в більш ранішні моменти часу.

Імовірності переходу зі стану A_i в стан A_j в момент часу t позначають через $p_{ij}(t)$.

Повна ймовірна картина всіх можливих переходів систем із одного стану в інший за умови, що число всіх станів дорівнює N, безпосередньо описується матрицею ймовірностей переходу

$$\pi = \begin{pmatrix} p_{11}(t) & p_{12}(t) & \dots\dots & p_{1N}(t) \\ p_{21}(t) & p_{22}(t) & \dots\dots & p_{2N}(t) \\ \dots\dots & \dots\dots & \dots\dots & \dots\dots \\ p_{N1}(t) & p_{N2}(t) & \dots\dots & p_{NN}(t) \end{pmatrix}. \qquad (584)$$

Якщо $p_{ij}(t)$ не залежить від часу, то ланцюг Маркова називають однорідним і тоді $p_{ij}(t) = p_{ij} = \text{const}$. А тому для однорідних ланцюгів Маркова матриця ймовірностей переходу набуває такого вигляду

$$\pi = \begin{pmatrix} p_{11} & p_{12} & \dots\dots & p_{1N} \\ p_{21} & p_{22} & \dots\dots & p_{2N} \\ \dots\dots & \dots\dots & \dots\dots & \dots\dots \\ p_{N1} & p_{N2} & \dots\dots & p_{NN} \end{pmatrix}. \qquad (585)$$

Для кожного рядка матриць (584), (585) виконується рівність

$$\sum_{j=1}^{N} p_{ij}(t) = \sum_{j=1}^{N} p_{ij} = 1. \qquad (586)$$

Це свідчить про те, що, перебуваючи в будь-якому можливому стані A_i у фіксований момент часу t, система обов'язково перейде з певною ймовірністю в будь-який інший можливий стан A_j або залишиться в цьому самому стані. А ці події є несумісними й утворюють повну групу.

Матрицю π (585) називають ще матрицею ймовірностей однокрокового переходу системи. Якщо позначимо множину всіх можливих станів системи через $\Omega = (\omega_1, \omega_2, \omega_3, \dots, \omega_i, \dots, \omega_j, \dots, \omega_N)$, то ймовірність переходу системи зі стану ω_i у стан ω_j за два кроки можна схематично показати на рисунку 162.

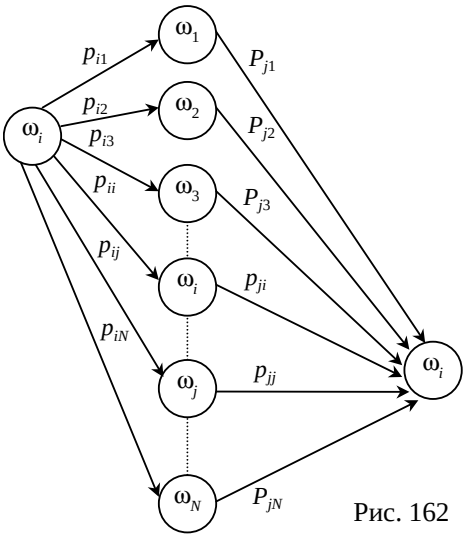


Рис. 162

Отже, ймовірність того, що система зможе перейти за два кроки зі стану ω_i до стану ω_j, дорівнюватиме:

$$P_{ij}^{(2)} = p_{i1}p_{1j} + p_{i2}p_{2j} + p_{i3}p_{3j} + \ldots + p_{ij}p_{jj} + \ldots + p_{iN}p_{Nj} =$$

$$= \sum_{k=1}^{N} p_{ik}\, p_{kj} \quad \text{для } i = \overline{1,N},\ j = \overline{1,N}. \qquad (587)$$

Вираз (587) є елементом матриці π^2, який дістанемо при множині i-го рядка матриці π на її j-й стовпчик.

Імовірність переходу системи зі стану ω_i у стан ω_j за n кроків обчислюється за формулою:

$$p_{ij}^{(n)} = \sum_{k=1}^{N} p_{ik}\, p_{kj}^{(n-1)}, \qquad (588)$$

де $p_{ij}^{(n)}$ — елемент матриці π^n.

Матрицю π^n називають n-кроковою матрицею переходу з одного стану в інший.

Імовірну картину переходу системи з одного стану в інший наочно можна зобразити з допомогою так званих *орієнтованих імовірнісних графів*, де кожний стан системи показано точкою (вершина графа), а можливі переходи з цього стану — ймовірнісними ребрами. Число станів на графі дорівнює числу вершин.

Приклад 2. За заданою матрицею ймовірностей переходу деякої системи

$$\pi = \begin{pmatrix} 0{,}1 & 0{,}2 & 0{,}7 \\ 0{,}3 & 0{,}5 & 0{,}2 \\ 0{,}4 & 0{,}1 & 0{,}5 \end{pmatrix}$$

побудувати орієнтований імовірний граф.

Розв'язання. За умовою задачі система має три стани ω_1, ω_2, ω_3, отже, граф матиме три вершини. Ймовірності переходу системи з одного стану в інший наведені в кожному рядку матриці π.

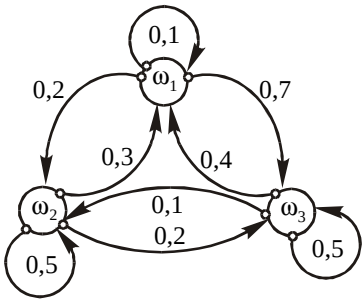


Рис. 163

Імовірний граф матиме такий вигляд (рис. 163).

Використання ймовірних графів можливе лише у тому разі, коли кількість станів систем є невеликою.

Уже за числа вершин (станів) 10 і більше рисунок набуває складної структури.

4. Класифікація ланцюгів Маркова

Якщо серед можливих станів систем, що описує ланцюг Маркова, є хоча б один, в якому система залишиться, коли перейде в нього на певному кроці функціонування, то такий ланцюг називають поглинальним ланцюгом Маркова.

Матриця однокрокового переходу для поглинаючої системи з трьох станів, наприклад, може мати такий вигляд:

$$\pi = \begin{pmatrix} 0{,}1 & 0{,}3 & 0{,}5 \\ 0{,}4 & 0{,}2 & 0{,}4 \\ 0 & 0 & 1 \end{pmatrix}.$$

Імовірний граф цієї матриці зображено на рис. 164.

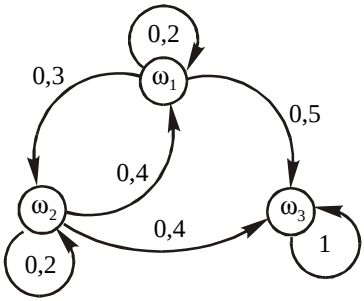


Рис. 164

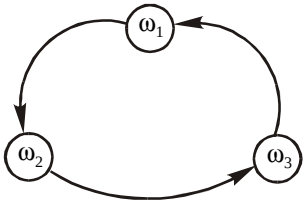


Рис. 165

Тут ω_3 є поглинаючим станом цієї системи.

Якщо в кожний можливий стан система може перейти через певний інтервал часу, то такий ланцюг Маркова, що моделює цей процес, називають *циклічним*.

Матриця циклічного ланцюга Маркова може мати для випадку трьох можливих станів системи такий вигляд:

$$\pi = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Граф цієї матриці буде таким (рис. 165).

Ланцюг Маркова називається *регулярним*, якщо існує таке ціле число k ($k > 0$), що перехід системи з будь-якого можливого стану в інший може здійснитися за k кроків.

Так, наприклад, для чотирьох можливих станів системи матриця однокрокового переходу може мати такий вигляд:

$$\pi = \begin{pmatrix} 0{,}2 & 0{,}1 & 0{,}5 & 0{,}2 \\ 0{,}3 & 0{,}2 & 0{,}1 & 0{,}4 \\ 0{,}1 & 0{,}5 & 0{,}2 & 0{,}2 \\ 0{,}4 & 0{,}1 & 0{,}4 & 0{,}1 \end{pmatrix}.$$

Графом для цієї матриці буде рис. 166.

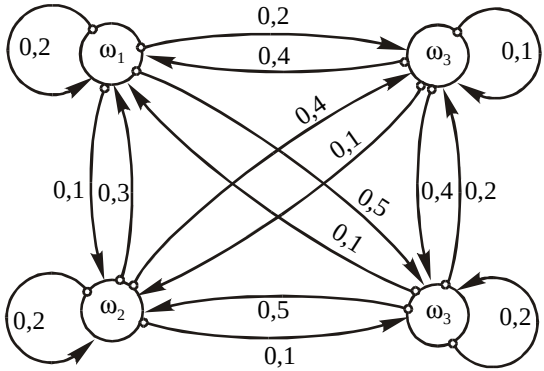


Рис. 157

Циклічні та регулярні матриці називають ергодичними.

Найширшого використання для розв’язування прикладних задач набули регулярні ланцюги Маркова.

5. Регулярні ланцюги Маркова. Стаціонарні ймовірності. Матриця вартостей

Для регулярних ланцюгів Маркова було доведено, що незалежно від початкового стану, в якому перебувала система, ймовірність перебування її в певному можливому стані прямує до деякої сталої величини при збільшенні кількості кроків k.

Тобто, при $k \to \infty$ $p_{ij}^{(n)} \to b_j = \text{const}$.

Імовірності b_j (j = 1, 2, ..., N) називають *стаціонарними*.

Отже,

$$\lim_{n\to\infty} \pi^n = B = \begin{pmatrix} b_1 & b_2 & \dots & b_N \\ b_1 & b_2 & \dots & b_N \\ \dots & \dots & \dots & \dots \\ b_1 & b_2 & \dots & b_N \end{pmatrix}.$$

Тут $\sum_{j=1}^{N} b_j = 1$, B — матриця стаціонарних імовірностей.

Таким чином, можна зробити висновок: якщо відомі π, B і при цьому $\vec{b} = (b_1, b_2, ..., b_N)$, а також задано $\vec{a} = (a_1, a_2, ..., a_N)$ — вектор початкових можливих станів системи, то виконуються рівності:

$$\lim_{n\to\infty} \vec{a}\pi^n = \vec{a} \lim_{n\to\infty} \pi^n = \vec{a}B = \vec{b}.$$

Можна довести, що коли π є регулярною матрицею і $\vec{b}$ — вектор стаціонарних імовірностей, то

$$\vec{b}\pi = \vec{b}\ . \qquad (589)$$

І справді, оскільки $\pi^n \to B$, то

$$\lim_{n\to\infty} \pi^n = \lim_{n\to\infty} \pi \cdot \pi^{n-1} = \pi \lim_{n\to\infty} \pi^{n-1} = \pi B\ ,$$

бо $\lim_{n\to\infty} \pi^{n-1} = B,\ \lim_{n\to\infty} \pi^n = B$.

Отже, маємо

$$B = \pi B. \qquad (590)$$

Таким чином, для будь-якого рядка рівняння (590) виконуватиметься рівність

$$\vec{b}\pi = \vec{b}\ . \qquad (591)$$

Отже, для визначення стаціонарних імовірностей регулярного ланцюга Маркова необхідно розв'язати систему рівнянь

$$\begin{cases} \vec{b} = \pi\vec{b} \\ \sum_{j=1}^{N} b_j = 1. \end{cases} \qquad (592)$$

Приклад. Знайти стаціонарні ймовірності для регулярної матриці

$$\pi = \begin{pmatrix} 0{,}3 & 0{,}1 & 0{,}6 \\ 0{,}2 & 0{,}5 & 0{,}3 \\ 0{,}1 & 0{,}4 & 0{,}5 \end{pmatrix}.$$

Розв'язання. Використовуючи систему (650), дістаємо:

$$\begin{cases} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 0{,}3 & 0{,}1 & 0{,}6 \\ 0{,}2 & 0{,}5 & 0{,}3 \\ 0{,}1 & 0{,}4 & 0{,}5 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \\ b_1 + b_2 + b_3 = 1 \end{cases} \to \begin{cases} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 0{,}3b_1 & 0{,}1b_2 & 0{,}6b_3 \\ 0{,}2b_1 & 0{,}5b_2 & 0{,}3b_3 \\ 0{,}1b_1 & 0{,}4b_2 & 0{,}5b_3 \end{pmatrix} \\ b_1 + b_2 + b_3 = 1 \end{cases} \to$$

$$\rightarrow \begin{cases} b_1 = 0{,}3b_1 + 0{,}1b_2 + 0{,}6b_3, \\ b_2 = 0{,}2b_1 + 0{,}5b_2 + 0{,}3b_3, \\ b_3 = 0{,}1b_1 + 0{,}4b_2 + 0{,}5b_3, \\ b_1 + b_2 + b_3 = 1 \end{cases} \rightarrow \begin{cases} -0{,}7b_1 + 0{,}1b_2 + 0{,}6b_3 = 0, \\ 0{,}2b_1 - 0{,}5b_2 + 0{,}3b_3 = 0, \\ 0{,}1b_1 + 0{,}4b_2 - 0{,}5b_3 = 0, \\ b_1 + b_2 + b_3 = 1. \end{cases}$$

Розв'язуючи одержану систему, маємо:

$$\vec{b} = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}, \text{ тобто } b_1 = \frac{1}{3},\ b_2 = \frac{1}{3},\ b_3 = \frac{1}{3}.$$

При переході системи з одного стану в інший за один крок її функціонування, а також у разі, коли система залишиться в тому самому стані, в якому вона перебувала, вводиться так званий коефіцієнт вартості r_{ij}, що інформує про витрати, які пов'язані при переході системи зі стану ω_i у стан ω_j за один крок.

Щоб визначити загальні витрати, які можна очікувати за n кроків роботи системи, використовується матриця вартостей

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \dots & r_{1N} \\ r_{21} & r_{22} & r_{23} & \dots & r_{2N} \\ \dots & \dots & \dots & \dots & \dots \\ r_{N1} & r_{N2} & r_{N3} & \dots & r_{NN} \end{pmatrix}. \qquad (593)$$

Елементи матриці R можуть набувати додатних, нульових та від'ємних числових значень. Якщо умовно позначимо через $v_i(n)$ сумарні витрати після n кроків переходів системи, починаючи із ω_i стану, тоді $\left(\vec{v}(n)\right)' = (v_1(n), v_2(n), \dots, v_N(n))$ буде вектором сумарних витрат за n кроків для кожного з ω_i станів (i = 1, 2, ..., N) системи.

Якщо система досягла ω_j стану, то очікувану винагороду після всіх переходів можна подати як $r_{ij} + v_i(n-1)$, де r_{ij} — винагорода, яку дістанемо при переході системи із ω_i стану в ω_j стан за один крок, а $v_i(n-1)$ — очікувана винагорода, яку матимемо за решту n – 1 кроків.

Ураховуючи те, що ймовірність переходу системи зі стану ω_i до стану ω_j за один крок дорівнює p_{ij}, то сумарна очікувана вартість (винагорода) за n переходів системи, починаючи зі стану ω_i, буде

$$v_i(n) = \sum_{j=1}^{N} r_{ij} p_{ij} + \sum_{j=1}^{N} v_j(n-1) p_{ij} \qquad (594)$$

або

$$v_i(n) = g_i + \sum_{j=1}^{N} v_j(n-1)\, p_{ij}\ , \qquad (595)$$

де $g_i = \sum_{j=1}^{N} r_{ij}\, p_{ij}$.

Оскільки рівняння (594), (595) виконується для всіх $i = \overline{1, N}$, то (595) можна записати у векторно-матричній формі

$$\vec{v}(n) = \vec{g} + \pi\vec{v}(n-1)\,, \qquad (596)$$

тут $\vec{g}' = (g_1, g_2, ..., g_N)$, де g_i — елементи головної діагоналі матриці $\pi \cdot R'$; R' — транспонована матриця вартостей R.

Компоненти вектора $\vec{v}(n)$ визначають із векторно-матричного рівняння

$$\vec{v}(n) = (E + \pi + \pi^2 + \pi^3 + ... + \pi^{(n-1)})\vec{g} + \pi^{(n)}\vec{v}(0)\,, \qquad (597)$$

де E — одинична матриця, $\vec{v}(0)$ — вектор винагороди за $n = 0$ кроків.

Приклад. Фермер придбав новий трактор, який може перебувати в одному з трьох несумісних станів: ω_1 — трактор працює добре, тобто не потребує жодних втручань, пов'язаних із витратами; ω_2 — трактор може бути використаний у роботі, але потребує дрібного ремонту, який, у свою чергу, вимагає певних витрат; ω_3 — трактор перебуває в неробото здатному стані, що відбивається на прибутках фермера.

Матриця ймовірностей переходу трактора з одного стану в інший за один крок (за 1 місяць) має такий вигляд:

$$\pi = \begin{pmatrix} 0{,}9 & 0{,}06 & 0{,}04 \\ 0{,}3 & 0{,}6 & 0{,}1 \\ 0{,}2 & 0{,}02 & 0{,}78 \end{pmatrix}.$$

За заданою матрицею вартостей

$$R = \begin{pmatrix} 200 & 50 & -10 \\ 150 & 40 & -20 \\ 100 & 10 & -50 \end{pmatrix}.$$

Оцінити ефективність роботи трактора за три робочих місяці. Елементи матриці вартостей r_{ij} вимірюються в гривнях.

При цьому $(\vec{v}(0))' = (000)$.

Розв'язання. З рівняння (597) для $n = 3$ дістанемо

$$\vec{v}(3) = (E + \pi + \pi^2)\vec{g} + \pi^3\vec{v}(0)\,. \qquad (598)$$

Визначимо координати вектора $\vec{g}$:

$$\pi R' = \begin{pmatrix} 0{,}9 & 0{,}06 & 0{,}04 \\ 0{,}3 & 0{,}6 & 0{,}1 \\ 0{,}2 & 0{,}02 & 0{,}78 \end{pmatrix}\begin{pmatrix} 200 & 150 & 100 \\ 50 & 40 & 10 \\ -10 & -20 & -50 \end{pmatrix} = \begin{pmatrix} 182{,}6 & 136{,}6 & 88{,}6 \\ 89 & 67 & 31 \\ 33{,}2 & 15{,}2 & -18{,}8 \end{pmatrix}.$$

Таким чином, маємо:

$$\vec{g} = \begin{pmatrix} 182{,}6 \\ 67 \\ -18{,}8 \end{pmatrix}.$$

Рівняння (598) перепишемо в такому вигляді

$$\begin{pmatrix} v_1(3) \\ v_2(3) \\ v_3(3) \end{pmatrix} = \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0{,}9 & 0{,}06 & 0{,}04 \\ 0{,}3 & 0{,}6 & 0{,}1 \\ 0{,}2 & 0{,}02 & 0{,}78 \end{pmatrix} + \begin{pmatrix} 0{,}9 & 0{,}06 & 0{,}04 \\ 0{,}3 & 0{,}6 & 0{,}1 \\ 0{,}2 & 0{,}02 & 0{,}78 \end{pmatrix}^2\right]\begin{pmatrix} 182{,}6 \\ 67 \\ -18{,}8 \end{pmatrix} +$$

$$+ \begin{pmatrix} 0{,}9 & 0{,}06 & 0{,}04 \\ 0{,}3 & 0{,}6 & 0{,}1 \\ 0{,}2 & 0{,}02 & 0{,}78 \end{pmatrix}^3 \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \to$$

$$\to \begin{pmatrix} v_1(3) \\ v_2(3) \\ v_3(3) \end{pmatrix} = \begin{pmatrix} 182{,}6 \\ 67 \\ -18{,}8 \end{pmatrix} + \begin{pmatrix} 167{,}61 \\ 93{,}1 \\ 23{,}2 \end{pmatrix} + \begin{pmatrix} 157{,}6 \\ 108{,}5 \\ 43{,}5 \end{pmatrix} = \begin{pmatrix} 507{,}81 \\ 268{,}6 \\ 47{,}9 \end{pmatrix}.$$

Таким чином, маємо:

$$\begin{pmatrix} v_1(3) \\ v_2(3) \\ v_3(3) \end{pmatrix} = \begin{pmatrix} 507{,}81 \\ 268{,}6 \\ 47{,}9 \end{pmatrix}.$$

Отже, коли трактор за три місяці перейде до стану ω_1 (справний), то прибуток фермера дорівнюватиме 507 грн. 81 коп.; для стану ω_2 цей прибуток становитиме 268 грн. 60 коп. і для стану ω_3 — лише 47 грн. 90 коп.

6. Процес народження і загибелі

Один із важливих напрямів застосування ланцюгів Маркова — моделювання процесу народження і загибелі організмів. Цей процес може бути з дискретними або з неперервними змінами часу t. Його визначальною умовою є те, що переходи можливі

лише в сусідні стани. Сутність марковського процесу в цьому разі полягає в тому, що він моделює зміни, котрі відбуваються в часі в певному об'ємі популяції, а саме — зміну числа одиниць певного виду організмів.

Такі процеси є надзвичайно зручними для математичного моделювання, що використовується для розв'язання задач теорії масового обслуговування (теорії черг).

Для процесу народження і загибелі допустимі лише переходи зі стану ω_k в стани ω_{k-1} або ω_{k+1}.

Якщо об'єм популяції дорівнює k одиниць, то процес популяції перебуває в стані ω_k.

Перехід зі стану ω_k у стан ω_{k+1} відповідає народженню одиниці виду організму, а перехід із ω_k в ω_{k-1} — загибелі одиниці організму. Події народження певного організму та його загибелі є незалежними і несумісними.

Розглянемо зміну популяції за певний проміжок часу $(t, t + \Delta t)$.

У моменті часу $t + \Delta t$ процес популяції перебуватиме у стані ω_k, якщо здійсниться одна з таких несумісних подій:

1) у момент часу t об'єм популяції дорівнював $k + 1$ одиниць і протягом часу $(t, t + \Delta t)$ загинула одиниця. Ця подія відбувається з імовірністю

$$\mu_k \Delta t + \alpha_k(\Delta t); \qquad (599)$$

2) у момент часу t об'єм популяції дорівнював $k - 1$ одиниць, і протягом часу $(t, t + \Delta t)$ народилась одиниця популяції. Ця подія відбудеться з імовірністю

$$\lambda_k \Delta t + \beta_k(\Delta t); \qquad (600)$$

3) у момент часу t об'єм популяції дорівнював k одиниць, і протягом часу $(t, t + \Delta t)$ не відбулася зміна стану, тобто жодна одиниця популяції не народилася і не загинула.

Імовірність такої події дорівнює

$$1 - \lambda_k \Delta t - \mu_k \Delta t + \gamma_k(\Delta t). \qquad (601)$$

Тут λ_k, μ_k — інтенсивність відповідно народження і загибелі одиниць популяції за певний проміжок часу, $\alpha_k(\Delta t)$, $\beta_k(\Delta t)$, $\gamma_k(\Delta t)$ — величини нескінченно малі більш високого порядку порівняно із Δt, а саме:

$$\lim_{\Delta t \to 0} \frac{\alpha_k(\Delta t)}{\Delta t} = 0, \ \lim_{\Delta t \to 0} \frac{\beta_k(\Delta t)}{\Delta t} = 0, \ \lim_{\Delta t \to 0} \frac{\gamma_k(\Delta t)}{\Delta t} = 0.$$

Вищенаведені переходи можна зобразити графом (рис. 167).

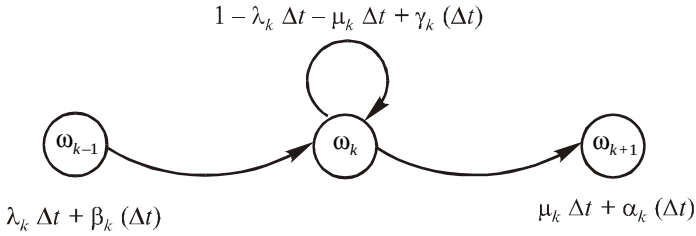


Рис. 167

Процес зміни об’єму популяції в k одиниць можна подати системою рівнянь:

$$p_k(t+\Delta t)=p_k(t)(1-\lambda_k\Delta t-\mu_k\Delta t+\gamma_k(\Delta t))+$$
$$+\ p_{k-1}(t)(\lambda_k\Delta t+\beta_k(\Delta t))+\ p_{k+1}(t)(\mu_k\Delta t+\alpha_k(\Delta t))\,; \qquad (602)$$
$$p_0(t+\Delta t)=p_0(t)(1-\lambda_0\Delta t-\mu_0\Delta t+\gamma_0(\Delta t))+\ p_1(t)(\mu_0\Delta t+\alpha_0(\Delta t))\,.$$

При цьому в будь-який момент часу t імовірності можливих станів об’єму популяції задовольняють умову нормування

$$\sum_{k=0}^{\infty} p_k(t)=1\,. \qquad (603)$$

Систему рівнянь (660) зводимо до такого вигляду:

$$\begin{cases} p_k(t+\Delta t)-p_k(t)=-(\lambda_k+\mu_k)\Delta t\cdot p_k(t)+\mu_k\Delta t\cdot p_{k+1}(t)+ \\ \qquad +\lambda_k\Delta t\cdot p_{k-1}(t)+\Theta_k(\Delta t), \\ p_0(t+\Delta t)-p_0(t)=-\lambda_0\Delta t\cdot p_0(t)+\mu_1\Delta t\cdot p_1(t)+\Theta_0(\Delta t)\,, \end{cases} \qquad (604)$$

де $\Theta_k(\Delta t)=p_k(t)\cdot\gamma_k(\Delta t)+p_{k-1}(t)\cdot\beta_k(\Delta t)+p_{k+1}(t)\cdot\alpha_k(\Delta t)$,

$\Theta_0(\Delta t)=p_1(t)\cdot\alpha_1(\Delta t)+p_0(t)\cdot\gamma_0(\Delta t)$ і при цьому $\lim\limits_{\Delta t\to 0}\dfrac{\Theta_k(\Delta t)}{\Delta t}=0$,

$\lim\limits_{\Delta t\to 0}\dfrac{\Theta_0(\Delta t)}{\Delta t}=0$.

Розділивши ліву і праву частини рівнянь системи (604) на Δt, дістанемо:

$$\begin{cases} \dfrac{p_k(t+\Delta t)-p_k(t)}{\Delta t}=-(\lambda_k+\mu_k)\,p_k(t)+\mu_k\,p_{k+1}(t)+ \\ \qquad +\lambda_k\,p_{k-1}(t)+\dfrac{\Theta_k(\Delta t)}{\Delta t}; \\ \dfrac{p_0(t+\Delta t)-p_0(t)}{\Delta t}=-\lambda_0\,p_0(t)+\mu_1\,p_1(t)+\dfrac{\Theta_0(\Delta t)}{\Delta t}. \end{cases} \qquad (605)$$

Переходимо в системі (605) до межі, при $\Delta t \to 0$ дістанемо:

$$\lim_{\Delta t \to 0} \frac{p_k(t+\Delta t) - p_k(t)}{\Delta t} = -(\lambda_k + \mu_k) p_k(t) + \mu_k p_{k+1}(t) +$$

$$+ \lambda_k p_{k-1}(t) + \lim_{\Delta t \to 0} \frac{\Theta_k(\Delta t)}{\Delta t}, \qquad (606)$$

$$\lim_{\Delta t \to 0} \frac{p_0(t+\Delta t) - p_0(t)}{\Delta t} = -\lambda_0 p_0(t) + \mu_1 p_1(t) + \lim_{\Delta t \to 0} \frac{\Theta_0(\Delta t)}{\Delta t}.$$

Із (606) дістанемо систему лінійних диференційних рівнянь

$$\begin{cases} p_k'(t) = -(\lambda_k + \mu_k) p_k(t) + \lambda_k p_{k-1}(t) + \mu_k p_{k+1}(t), \\ p_0'(t) = -\lambda_0 p_0(t) + \mu_1 p_1(t). \end{cases} \qquad (607)$$

Система (607) описує динаміку ймовірнісного процесу розмноження і загибелі популяції.

За умови $\mu_k = \mu = \text{const}$, $\lambda_k = \lambda = \text{const}$ система (607) набуває такого вигляду:

$$\begin{cases} p_0'(t) = -\lambda \cdot p_0(t) + \mu p_1(t), \\ p_k'(t) = -(\lambda + \mu) p_k(t) + \lambda p_{k-1}(t) + \mu p_{k+1}(t). \end{cases} \qquad (608)$$

Математична модель (608) застосовується в елементарній теорії масового обслуговування.

7. Елементи теорії масового обслуговування (ТМО)

7.1. Загальна інформація

Як у сфері виробництва, так і в побуті часто трапляються такі неприємні явища, як черга, що являє собою скупчення (яке буває численним) об’єктів (вимог), які чекають свого обслуговування.

Під обслуговуванням об’єкта (вимоги) розуміють виконання стосовно нього (неї) певного комплексу операцій.

Черга може мати місце у повсякденному житті: у магазинах до кас чи до продавця, у поліклініці до лікаря. У виробничій практиці вона виникає під час виробничих процесів, коли накопичуються деталі, вузли, агрегати, які потребують обслуговування робітником, верстатом-автоматом.

У більш складних системах обслуговування черги виникають під час навантажування чи розвантажування вагонів, річних та морських суден у портах, обслуговування літаків в аеропортах тощо.

Причини, які призводять до черг як у сфері виробництва, так і в сфері обслуговування, мають випадковий характер і виникають тоді, коли:

1) пропускна здатність «приладу» обслуговування (робітник, касир, верстат-автомат) не задовольняє число вимог, що надходять;

2) об'єкти (вимоги) надходять нерегулярно, тобто у випадкові моменти часу, утворюючи при цьому пуассонівський потік;

3) час обслуговування об'єкта (вимоги) не є сталою величиною, а випадковою, і припускається, що вона має експоненціальний закон розподілу ймовірностей.

7.2. Математична модель системи масового обслуговування

Математичною моделлю для найпростішої системи масового обслуговування з одним пуассонівським потоком і одним каналом обслуговування (одноканальний прилад), час якого має експоненціальний закон розподілу ймовірностей, є система диференційних рівнянь (608).

У стаціонарному режимі (при $t \to \infty$) ймовірності вже не залежать від часу

$$\lim_{t\to\infty} p_k(t) = p_k\,,\ \ k = 0,\ 1,\ 2,\ 3,\ ...$$

$$p_k'(t) = 0\,,\ k = 0,\ 1,\ 2,\ 3,\ ....$$

У цьому разі система (608) переходить в однорідну нескінченну систему алгебраїчних рівнянь відносно p_k і має вигляд:

$$\begin{cases} 0 = -\lambda p_0 + \mu p_1, \\ 0 = -(\lambda + \mu) p_k + \lambda p_{k-1} + \mu p_{k+1}. \end{cases} \qquad (609)$$

Якщо ввести величину $\rho = \dfrac{\lambda}{\mu}$, то система (609) набере вигляду:

$$\begin{cases} \rho p_0 = p_1, \\ (1+\rho) p_k = \rho p_{k-1} + p_{k+1}. \end{cases} \qquad (610)$$

При цьому $\rho < 1$.

7.3. Метод розв'язування та числові характеристики

Для розв'язування системи (610) використовують метод імовірних твірних функцій, а саме:

$$A(x) = \sum_{k=1}^{\infty} x^k p_k\ . \qquad (611)$$

Якщо помножити ліву і праву частини другого рівняння системи (610) на x^k, то дістанемо:

$$\begin{cases} \rho p_0 = p_1, \\ (1+\rho)x^k p_k = \rho x^k p_{k-1} + x^k p_{k+1}. \end{cases} \qquad (612)$$

Підсумовуючи ліву та праву частини системи (612), дістанемо функціональне рівняння щодо $A(x)$ такого вигляду:

$$(1+\rho)A(x) + \rho p_0 = \rho x p_0 + \rho x A(x) + \frac{1}{x} A(x) \to$$

$$\to \left((1-x)\rho + \left(1 - \frac{1}{x}\right)\right) A(x) = (x-1)\rho p_0 \to$$

$$\to A(x) = \frac{\rho(x-1)p_0}{\rho(1-x) + \left(1 - \frac{1}{x}\right)} = \frac{\rho \cdot x \cdot p_0}{1-\rho x}.$$

Таким чином, маємо:

$$A(x) = \frac{\rho \cdot x}{1-\rho x} p_0. \qquad (613)$$

При цьому $A(1) = \frac{\rho}{1-\rho} p_0$.

Оскільки $A(1) + p_0 = 1$, то дістанемо:

$$\frac{\rho}{1-\rho} p_0 + p_0 = 1 \to p_0 = 1 - \rho. \qquad (614)$$

ρ називають коефіцієнтом навантаження системи.

Отже, ймовірність того, що в системі будуть відсутні вимоги на обслуговування (система не працює через відсутність вимог) p_0 залежить від коефіцієнта навантаження системи. Зі збільшенням ρ зменшується p_0; при $\rho = 1$ $p_0 = 0$.

Ураховуючи (613), дістанемо:

$$A(1) = \frac{\rho}{1-\rho} p_0 = \frac{\rho}{1-\rho}(1-\rho) = \rho.$$

Отже, ймовірність того, що система зайнята обслуговуванням вимог $A(1)$, залежатиме від коефіцієнта навантаження

$$A(1) = \rho. \qquad (615)$$

Зі збільшенням ρ збільшується $A(1)$.

Для визначення математичного сподівання числа вимог у системі застосовуємо властивості $A(x)$, а саме:

$$M = A'(x)\big|_{x=1} = \left((1-\rho)\frac{\rho x}{1-\rho x}\right)'_{x=1} = (1-\rho)\frac{\rho(1-\rho x)+\rho^2 x}{(1-\rho x)^2}\Bigg|_{x=1} = \frac{\rho}{1-\rho}.$$

Отже, маємо:

$$M = \frac{\rho}{1-\rho}. \qquad (616)$$

Середнє число вимог у черзі обчислюється за формулою

$$L = M - A(1) = \frac{\rho^2}{1-\rho}. \qquad (617)$$

Середнє значення часу очікування вимоги в черзі до початку свого обслуговування приладом обчислюється так:

$$t_c = \frac{M}{\lambda}. \qquad (618)$$

Із формул (617) і (618) випливає, що для стабільності роботи системи необхідне дотримання умови $\rho < 1$.

Приклад. Комп'ютер контролює певний технологічний процес шляхом обробки інформації, яка надходить від конвеєра, що працює в автоматичному режимі. Інформація, що надходить на комп'ютер, утворює пуассонівський потік з інтенсивністю $\lambda = 0{,}1$ с$^{-1}$. Час обробки інформації є випадковою величиною, що має експоненціальний закон розподілу ймовірностей зі значенням параметра $\mu = 0{,}4$ с$^{-1}$. У процесі роботи комп'ютер може вийти з ладу. Це трапляється у випадкові моменти часу й утворює пуасонівський потік з інтенсивністю $\lambda_0 = 0{,}08$ с$^{-1}$. Відновлення роботи комп'ютера здійснюється програмним методом, і час, витрачений на відновлення, є випадковою величиною з експоненціальним законом розподілу ймовірностей зі значенням параметра $\mu_0 = 0{,}4$ с$^{-1}$.

Число інформації, що надходить для обробки, є випадковим, і при цьому інформація, що чекає на обслуговування, не залишає систему (не губиться).

Необхідно:

1) побудувати математичну модель для цієї системи в стаціонарному режимі;

2) визначити: M, L, t_c;

3) обчислити середнє число витрат при роботі заданої системи, використовуючи формулу:

$$G = (g_1 N_0 + g_2 L + g_3 M)T\ , \qquad (619)$$

де g_1 — вартість одиниці часу простою системи, зумовленого відмовою роботи комп'ютера (g_1 = 30 грн/хв);
g_2 — вартість втрат, які пов'язані з тим, що інформація очікує черги на обслуговування (g_2 = 300 грн/хв);
g_3 — вартість експлуатації одного приладу (g_3 = 100 грн/хв);
N_0 — число приладів у системі (N_0 = 1);
T — час роботи системи (T = 60 хв).

Розв'язання. Позначимо через p_k імовірність того, що в системі міститься k вимог інформації і комп'ютер обслуговує одиницю інформації, а через Q_k — імовірність того, що в системі міститься k вимог і комп'ютер перебуває в стані налагодження.

Після ліквідації несправностей система зі стану Q_k переходить у стан p_k (k = 0, 1, 2, …).

У стаціонарному режимі роботи цієї системи математична модель матиме такий вигляд:

$$\begin{cases} 1.\ (\lambda_0 + \lambda)p_0 = \mu p_1 + \mu_0 Q_0, \\ 2.\ (\lambda_0 + \lambda + \mu)p_k = \mu p_{k+1} + \lambda p_{k-1} + \mu_0 Q_k, \\ 3.\ (\lambda_0 + \mu_0)Q_0 = \lambda_0 p_0, \\ 4.\ (\lambda_0 + \mu_0)Q_k = \lambda_0 p_k + \lambda p_{k-1}. \end{cases} \qquad (620)$$

Для розв'язання системи (620) використаємо метод імовірних твірних функцій.

Упровадимо його в такому вигляді:

$$A(x) = A_1(x) + A_2(x) + p_0,$$

де $A_1(x) = \sum_{k=1}^{\infty} x^k p_k$, $A_2(x) = \sum_{k=1}^{\infty} x^k Q_k$.

У системі (620) помножимо ліву і праву частини другого та четвертого рівнянь на x^k і дістанемо:

$$\begin{cases} 1.\ (\lambda_0 + \lambda)p_0 = \mu p_1 + \mu_0 Q_0, \\ 2.\ (\lambda_0 + \lambda + \mu)p_k x^k = \mu p_{k+1} x^k + \lambda p_{k-1} x^k + \mu_0 Q_k x^k, \\ 3.\ (\lambda_0 + \mu_0)Q_0 = \lambda_0 p_0, \\ 4.\ (\lambda_0 + \mu_0)Q_k x^k = \lambda_0 p_k x^k + \lambda p_{k-1} x^k. \end{cases} \qquad (621)$$

У системі (621) просумуємо перше і друге рівняння, третє і четверте, дістанемо:

$$\begin{cases} \left(\lambda_0 + \lambda(1-x) + \mu\left(1 - \dfrac{1}{x}\right)\right)A_1(x) - \mu_0 A_2(x) = (\lambda(x-1) - \lambda_0)p_0, \\ (\lambda(1-x) + \mu_0)A_2(x) - \lambda_0 A_1(x) - \lambda_0 p_0. \end{cases} \qquad (622)$$

Розв'язуючи систему функціональних рівнянь (622) щодо $A_1(x)$, $A_2(x)$, дістанемо:

$$A_1(x) = \frac{\lambda\mu_0 + \lambda_0\lambda + \lambda^2(1-x)p_0}{\mu_0\mu\frac{1}{x} - \lambda_0\lambda - \lambda\mu_0 - \lambda^2(1-x) - \lambda\mu\left(1-\frac{1}{x}\right)}, \qquad (623)$$

$$A_2(x) = \frac{\lambda_0\mu p_0}{\mu_0\mu\frac{1}{x} - \lambda_0\lambda - \lambda\mu_0 - \lambda^2(1-x) - \lambda\mu\left(1-\frac{1}{x}\right)}. \qquad (624)$$

При $x = 1$ матимемо:

$$A_1(1) = \frac{\rho(1+\rho_0)p_0}{1-\rho(1+\rho_0)}, \quad A_2(1) = \frac{\rho_0 p_0}{1-\rho(1+\rho_0)},$$

де $\rho = \frac{\lambda}{\mu}$, $\rho_0 = \frac{\lambda_0}{\mu_0}$.

Оскільки $A(1) = A_1(1) + A_2(1) + p_0 = 1$ — умови нормування, то дістанемо:

$$\frac{\rho(1+\rho_0)}{1-\rho(1+\rho_0)}p_0 + \frac{\rho_0}{1-\rho(1+\rho_0)}p_0 + p_0 = 1 \to p_0 = \frac{1}{1+\rho_0}(1-\rho(1+\rho_0)).$$

Підставляючи p_0 у вираз для $A_1(1)$, $A_2(1)$, визначимо:

$A_1(1) = \rho$ — імовірність того, що система зайнята обслуговуванням вимог інформації;

$A_1(1) = \frac{\rho_0}{1+\rho_0}$ — імовірність того, що система перебуває у стані налагодження комп'ютера (приладу обслуговування).

Математичне сподівання кількості вимог інформації дорівнюватиме:

$$M = A'(1) = A_1'(1) + A_2'(1) = \frac{\rho + \rho_0 + \rho\cdot\rho_0(1+\alpha)}{1-\rho(1+\rho_0)}K_r,$$

де $\alpha = \frac{\mu_0}{\mu}$, $K_r = \frac{1}{1+\rho_0}$ називають коефіцієнтом готовності системи.

Беручи числові значення параметрів, дістанемо:

$$\rho = \frac{\lambda}{\mu} = \frac{0{,}1}{0{,}4} = 0{,}25, \quad \rho_0 = \frac{\lambda_0}{\mu_0} = \frac{0{,}08}{0{,}4} = 0{,}2;$$

$$\alpha = \frac{\mu_0}{\mu} = \frac{0{,}4}{0{,}4} = 1; \quad K_r = \frac{1}{1+\rho_0} = \frac{1}{1+0{,}2} = 0{,}83.$$

Тепер математичне сподівання дорівнюватиме:

$$M = \frac{0{,}25 + 0{,}2 + 0{,}25 \cdot 0{,}2(1+1)}{1 - 0{,}25(1+0{,}2)} \cdot 0{,}83 = \frac{0{,}45 + 0{,}1}{1 - 0{,}3} \cdot 0{,}83 =$$

$$= \frac{0{,}55 \cdot 0{,}83}{0{,}7} = 0{,}652 .$$

$$M = 0{,}652 .$$

Середнє число вимог у черзі дорівнюватиме:

$$L = M - A(1) = M - \left(\rho + \frac{\rho_0}{1+\rho_0}\right) = 0{,}652 - \left(0{,}25 + \frac{0{,}2}{1+0{,}2}\right) =$$

$$= 0{,}652 - (0{,}25 + 0{,}167) = 0{,}652 - 0{,}42 = 0{,}205 .$$

$$L = 0{,}205 .$$

Середнє значення часу очікування вимоги у черзі становитиме:

$$t = \frac{M}{\lambda} = \frac{0{,}652}{0{,}4} = 1{,}63 \text{ с.}$$

Середнє число витрат системи для T = 60 хв дорівнюватиме:

$$G = (g_1 N_0 + g_2 L + g_3 M)T = \left(\frac{30}{60}1 + 300\frac{1}{60}0{,}205 + 100\frac{1}{60}0{,}652\right)360 =$$

$$= (0{,}5 + 1{,}025 + 1{,}087)360 = 940 \text{ грн. 32 коп.}$$

Отже, витрати в середньому становлять 940 грн. 32 коп.

Теоретичні запитання до теми ?

1. Що називають випадковим процесом?
2. Дати класифікацію випадкових процесів.
3. Що називають реалізацією випадкового процесу?
4. Що називають перерізом випадкового процесу?
5. Що називають функцією розподілу перерізу?
6. Функція розподілу двох перерізів.
7. Математичне сподівання перерізу випадкового процесу $X = x(t)$.
8. Дисперсія та середнє квадратичне відхилення перерізу випадкового процесу $X = x(t)$.

9. Що називають кореляційним моментом для дискретних та неперервних перерізів випадкового процесу $X = x(t)$?
10. Що називають кореляційною функцією випадкового процесу $X = x(t)$? Її властивості.
11. Що називають Марковським процесом?
12. Що називають ланцюгом Маркова?
13. Матриця однокрокового переходу системи має вигляд $\pi = ...$
14. Класифікація ланцюгів Маркова.
15. Що називають поглинаючим ланцюгом Маркова?
16. Які ланцюги Маркова називають регулярними?
17. Які ланцюги Маркова називають ергодичними?
18. Який вигляд має матриця стаціонарних імовірностей?
19. Як знаходять стаціонарні ймовірності для регулярних ланцюгів Маркова?
20. Що таке матриця вартостей?
21. Як обчислюються компоненти вектора винагороди за n кроків $\vec{v}(n)$?
22. Записати математичну модель процесу розмноження та загибелі для k популяцій.
23. Які мотиви спонукають виникнення черг у сфері обслуговування?
24. Записати математичну модель для найпростішої системи масового обслуговування у стаціонарному режимі.
25. Що називають імовірною твірною функцією? Властивості цієї функції.
26. Використання ймовірної твірної функції для визначення математичного сподівання.
27. Середнє число вимог у системі обчислюється за формулою...
28. Середній час очікування вимоги до початку обслуговування обчислюється за формулою...

Задачі до теми

1. Знайти стаціонарні ймовірності для регулярних ланцюгів Маркова, що мають матриці однокрокового переходу системи:

а) $\pi = \begin{pmatrix} 0{,}5 & 0{,}4 & 0{,}1 \\ 0{,}3 & 0{,}4 & 0{,}3 \\ 0{,}2 & 0{,}1 & 0{,}7 \end{pmatrix}$; б) $\pi = \begin{pmatrix} 0{,}9 & 0{,}1 \\ 0{,}2 & 0{,}8 \end{pmatrix}$.

2. Щільність імовірностей випадкової функції $X = x(t)$ дорівнює

$$f(x,t) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x - a\sin t)^2}{2\sigma^2}},$$

де a, σ — сталі величини.

Знайти: $M(x(t))$, $D(x(t))$.

Відповідь. $M(x(t)) = a \sin t$, $D(x(t)) = \sigma^2$.

3. Двовимірна щільність випадкової функції (випадкового процесу) $X = x(t)$ дорівнює

$$f(x_1, x_2;\ t_1, t_2) = \frac{1}{2\pi\sigma^2} e^{-\frac{x_1^2 + x_2^2}{2\sigma^2}}.$$

Знайти: $M(x(t))$, $D(x(t))$, $K_x(t_1;\ t_2)$.

Відповідь. $M(x(t)) = 0$, $D(x(t)) = \sigma^2$, $K_x(t_1;\ t_2) = \sigma^2$ при $t_1 = t_2$,
$K_x(t_1;\ t_2) = 0$ при $t_1 \neq t_2$.

4. Математичне сподівання і кореляційну функцію випадкового процесу $X = x(t)$ задано $M(x(t)) = t + 4$; $K_x(t_1;\ t_2) = t_1 t_2$.

Знайти числові характеристики для $Y(t) = 5tx(t)+2$.

Відповідь. $M(y(t)) = 5t^2 + 20t + 2$; $D(y(t)) = 25t^2$; $K_x(t_1;\ t_2) = 25t_1^2\, t_2^2$.

5. Дві фірми пропонують фермерам трактори, які умовно позначимо як трактор фірми *A* і трактор фірми *B*. Ці трактори мають такі ймовірні матриці однокрокового переходу зі стану ω_1 — працює добре в стан ω_2 — трактор потребує ремонту.

Дані матриці мають такий вигляд:

$$\pi_A = \begin{pmatrix} 0{,}89 & 0{,}01 \\ 0{,}82 & 0{,}12 \end{pmatrix}, \quad \pi_B = \begin{pmatrix} 0{,}79 & 0{,}21 \\ 0{,}72 & 0{,}28 \end{pmatrix}.$$

Визначити, котрий із тракторів варто купити фермерові?

Відповідь. Трактор фірми *A*.

6. Водій таксі виявив, що коли він перебуває в аеропорту, то за одну годину простою він з імовірністю 0,6 повезе пасажира на автовокзал, з імовірністю 0,3 — на залізничний вокзал і з імовірністю 0,1 залишиться на стоянці. Коли ж він перебуває біля залізничного вокзалу, то з імовірністю 0,4 повезе пасажира до аеропорту, з імовірністю 0,3 — на автовокзал, з імовірністю 0,3 — залишиться на стоянці; коли ж водій таксі з машиною перебуває біля автовокзалу, то з імовірністю 0,7 повезе пасажира до залізничного вокзалу, з імовірністю 0,25 — до аеропорту, і з імовірністю 0,05 — залишиться на місці. Середній прибуток для кожного маршруту такий: проїзд за маршрутом аеропорт—вокзал коштує

10 грн. 30 коп.; аеропорт—залізничний вокзал — 8 грн. 20 коп. і автовокзал—залізничний вокзал — 9 грн. 50 коп. Записати однокрокову ймовірну матрицю переходів і матрицю вартостей. Визначити очікувані прибутки за чотири робочі години, якщо $\vec{v}(0) = (000)$.

Відповідь. $\pi = \begin{pmatrix} 0{,}1 & 0{,}3 & 0{,}6 \\ 0{,}4 & 0{,}3 & 0{,}3 \\ 0{,}25 & 0{,}7 & 0{,}05 \end{pmatrix}$, $R = \begin{pmatrix} 0 & 10{,}3 & 8{,}2 \\ 10{,}3 & 0 & 9{,}5 \\ 8{,}2 & 9{,}5 & 0 \end{pmatrix}$,

$$\vec{v}(3) = \begin{pmatrix} 8{,}01 \text{ грн.} \\ 6{,}47 \text{ грн.} \\ 8{,}7 \text{ грн.} \end{pmatrix}.$$

7. Потік товарних залізничних потягів, що надходить на сортувальний вузол, приймається найпростішим з інтенсивністю $\lambda = 4$ потяги за одну годину. Обчислити ймовірність того, що за 1,5 години до залізничного вузла надійде: 1) п’ять потягів; 2) не більше п’яти.

8. На станцію поточного технічного ремонту надходять автомашини, які утворюють найпростіший потік із параметром $\lambda = 0{,}8$ автомашини за годину. У ремонті бере участь одна бригада робітників. Середній час ремонту однієї автомашини є випадковою величиною, що має експоненціальний закон розподілу з параметром $\mu = 1{,}8$ автомашини за годину. На майданчику ремонтної станції одночасно може перебувати не більше як чотири автомашини.

Визначити: 1) імовірність того, що ремонтна станція простоюватиме без роботи; 2) імовірність того, що майданчик станції буде повністю заповненим; 3) середнє число автомашин на ремонтній станції.

Відповідь. $P_0 = 0{,}6399$; $P_5 = 0{,}024$; $M = 0{,}7878$.

9. Досліджується робота автозаправочної станції (АЗС), яка має три заправочні колонки. У середньому за кожну хвилину на АЗС надходять дві автомашини. Заправка однієї автомашини в середньому триває чотири хвилини. Число автомашин, що можуть перебувати в черзі для заправки, практично не обмежене і всі машини не залишають черги.

Визначити: 1) імовірність того, що на АЗС буде відсутня черга; 2) середнє число машин у черзі; 3) середній час очікування для автомашин до початку їх обслуговування.

Відповідь. $P_0 = 0{,}606$; $M = 0{,}43$; $t = 0{,}11$ хв.

Додаток 2[*]

ТАБЛИЦЯ ЗНАЧЕНЬ ФУНКЦІЇ ЛАПЛАСА $\Phi(x) = -\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{z^2}{2}}dz$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	0,26	0,1026	0,52	0,1985	0,78	0,2823
0,01	0,0040	0,27	0,1064	0,53	0,2019	0,79	0,2852
0,02	0,0080	0,28	0,1103	0,54	0,2054	0,80	0,2881
0,03	0,0120	0,29	0,1141	0,55	0,2088	0,81	0,2910
0,04	0,0160	0,30	0,1179	0,56	0,2123	0,820	0,2939
0,05	0,0199	0,31	0,1217	0,57	0,2157	0,83	0,2967
0,06	0,0239	0,32	0,1255	0,58	0,2190	0,84	0,2995
0,07	0,0279	0,33	0,1293	0,59	0,2224	0,85	0,3023
0,08	0,0319	0,34	0,1331	0,60	0,2257	0,86	0,3051
0,09	0,0359	0,35	0,1368	0,61	0,2291	0,87	0,3078
0,10	0,0398	0,36	0,1406	0,62	0,2324	0,88	0,3106
0,11	0,0438	0,37	0,1443	0,63	0,2357	0,89	0,3133
0,12	0,0478	0,38	0,1480	0,64	0,2389	0,90	0,3159
0,13	0,0517	0,39	0,1617	0,65	0,2422	0,91	0,3186
0,14	0,8557	0,40	0,1564	0,66	0,2454	0,92	0,3212
0,15	0,0596	0,41	0,1691	0,67	0,2486	0,93	0,3238
0,16	0,0636	0,42	0,1628	0,68	0,2517	0,94	0,3264
0,17	0,0675	0,43	0,1664	0,69	0,2549	0,95	0,3289
0,18	0,0714	0,44	0,1700	0,70	0,2580	0,96	0,3315
0,19	0,0753	0,45	0,1736	0,71	0,2611	0,97	0,3340
0,20	0,0793	0,46	0,1772	0,72	0,2642	0,98	0,3365
0,21	0,0832	0,47	0,1808	0,73	0,2673	0,99	0,3389
0,22	0,0871	0,48	0,1844	0,74	0,2703	1,00	0,3413
0,23	0,0910	0,49	0,1879	0,75	0,2734	1,01	0,3438
0,24	0,0948	0,50	0,1915	0,76	0,2764	1,02	0,3461
0,25	0,0987	0,51	0,1950	0,77	0,2794	1,03	0,3485

[*] Додатки 2—8 стосуються вміщеного в цій частині курсу математичної статистики.

Продовження додатка 2

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1,04	0,3508	1,33	0,4082	1,62	0,4474	1,91	0,4719
1,05	0,3531	1,34	0,4099	1,63	0,4484	1,92	0,4726
1,06	0,3554	1,35	0,4115	1,64	0,4495	1,93	0,4732
1,07	0,3577	1,36	0,4131	1,65	0,4505	1,94	0,4738
1,08	0,3599	1,37	0,4147	1,66	0,4515-	1,95	0,4744
1,09	0,3621	1,38	0,4162	1,67	0,4525	1,96	0,4750
1,10	0,3643	1,39	0,4177	1,68	0,4535	1,97	0,4756
1,11	0,3665	1,40	0,4192	1,69	0,4545	1,98	0,4761
1,12	0,3686	1,41	0,4207	1,70	0,4554	1,99	0,4767
1,13	0,3708	1,42	0,4222	1,71	0,4564	2,00	0,4772
1,14	0,3729	1,43	0,4236	1,72	0,4573	2,02	0,4783
1,15	0,3749	1,44	0,4251	1,73	0,4582	2,04	0,4793
1,16	0,3770	1,45	0,4265	1,74	0,4591	2,06	0,4803
1,17	0,3790	1,46	0,4279	1,75	0,4599	2,08	0,4812
1,18	0,3810	1,47	0,4292	1,76	0,4608	2,10	0,4821
1,19	0,3830	1,48	0,4306	1,77	0,4616	2,12	0,4830
1,20	0,3849	1,49	0,4319	1,78	0,4625	2,14	0,4838
1,21	0,3869	1,50	0,4332	1,79	0,4633	2,16	0,4846
1,22	0,3883	1,51	0,4345	1,80	0,4641	2,18	0,4854
1,23	0,3907	1,52	0,4357	1,81	0,4649	2,20	0,4861
1,24	0,3925	1,53	0,4370	1,82	0,4656	2,22	0,4868
1,25	0,3944	1,54	0,4382	1,83	0,4664	2,24	0,4875
1,26	0,3962	1,55	0,4394	1,84	0,4671	2,26	0,4881
1,27	0,3980	1,56	0,4406	1,85	0,4678	2,28	0,4887
1,28	0,3997	1,57	0,4418	1,86	0,4686	2,30	0,4893
1,29	0,4015	1,58	0,4429	1,87	0,4693	2,32	0,4898
1,30	0,4032	1,59	0,4441	1,88	0,4699	2,34	0,4904
1,31	0,4049	1,60	0,4452	1,89	0,4706	2,36	0,4909
1,32	0,4066	1,61	0,4463	1,90	0,4713	2,38	0,4913

Закінчення додатка 2

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
2,40	0,4918	2,60	0,4953	2,80	0,4974	3,20	0,49931
2,42	0,4922	2,62	0,4956	2,82	0,4976	3,40	0,49966
2,44	0,4927	2,64	0,4959	2,84	0,4977	3,60	0,49984
2,46	0,4931	2,66	0,4961	2,86	0,4979	3,80	0,499928
2,48	0,4934	2,68	0,4963	2,90	0,4981	4,00	0,499968
2,50	0,4938	2,70	0,4965	2,92	0,4982	5,00	0,499997
2,52	0,4941	2,72	0,4967	2,94	0,4984		
2,54	0,4945	2,74	0,4969	2,96	0,49846		
2,56	0,4948	2,76	0,4971	2,98	0,49856		
2,58	0,4951	2,78	0,4973	3,00	0,49865	$x > 5$	0,5

Додаток 3

ТАБЛИЦЯ ЗНАЧЕНЬ $t(j, k = n-1)$ **, ЩО ЗАДОВОЛЬНЯЮТЬ РІВНІСТЬ** $p(t) = 2\int_0^t f(x)dt = \gamma$

$k = n-1$	$p(t)$												
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	0,158	0,326	0,510	0,727	1,00	1,376	1,963	3,078	6,314	12,706	31,821	63,657	63,662
2	0,142	0,289	0,445	0,617	0,816	1,061	1,336	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	2,638	2,353	3,182	4,541	5,841	12,941
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,694	8,610
5	0,132	0,257	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,859
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,401	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,086	1,363	1,796	2,201	2,718	3,106	4,487
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073

Закінчення додатка 3

$k = n - 1$	$p(t)$												
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,103	2,552	2,872	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,859	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,857	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646

Додаток 4

ЗНАЧЕННЯ ВЕЛИЧИНИ χ_1^2 ЗАЛЕЖНО ВІД ІМОВІРНОСТІ $P(\chi^2 > \chi_1^2)$

Число ступенів свободи, *k*	$P(\chi^2 > \chi_1^2)$							
	0,2	0,10	0,05	0,02	0,01	0,005	0,002	0,001
1	1,64	2,7	3,8	5,4	6,6	7,9	9,5	10,83
2	3,22	4,6	6,0	7,8	9,2	11,6	12,4	13,8
3	4,64	6,3	7,8	9,8	11,3	12,8	14,6	16,3
4	6,0	7,8	9,5	11,7	13,3	14,9	16,9	18,5
5	7,3	9,2	11,1	13,4	15,1	16,3	18,9	20,5
6	8,6	10,6	12,6	15,0	16,8	18,6	20,7	22,5
7	9,8	12,0	14,1	16,6	18,5	20,3	22,6	24,3
8	11,0	13,4	15,5	18,2	20,1	21,9	24,3	26,1
9	12,2	14,7	16,9	19,7	21,7	23,6	26,1	27,9
10	13,4	16,0	18,3	21,2	23,2	25,2	27,7	29,6
11	14,6	17,3	19,7	22,6	24,7	26,8	29,4	31,3
12	15,8	18,5	21,0	24,1	26,2	28,3	31,0	32,9
13	17,0	19,8	22,4	25,5	27,7	29,8	32,5	34,5
14	18,2	21,1	23,7	26,9	29,1	31,0	34,0	36,1
15	19,3	22,3	25,0	28,3	30,6	32,5	35,5	37,7
16	20,5	23,5	26,3	29,6	32,0	34,0	37,0	39,2
17	21,6	24,8	27,6	31,0	33,4	35,5	38,5	40,8
18	22,8	26,0	28,9	32,3	34,8	37,0	40,0	42,3
19	23,9	27,3	30,1	33,7	36,2	38,5	41,5	43,8
20	25,0	28,4	31,4	35,0	37,6	40,0	43,0	45,3
21	26,2	29,6	32,7	36,3	38,9	41,5	44,5	46,8
22	27,3	30,8	33,9	38,7	40,3	42,5	46,0	48,3
23	28,4	32,0	35,2	39,0	41,6	44,0	47,5	49,7
24	29,6	33,2	36,4	40,3	43,0	45,5	48,5	51,2
25	30,7	34,4	37,7	41,6	44,3	47,0	50,0	52,6
26	31,8	35,6	38,9	42,9	45,6	48,0	51,5	54,1
27	32,9	36,7	40,1	44,1	47,0	49,5	53,0	55,5
28	34,0	37,9	41,3	45,4	48,3	51,0	54,5	56,9
29	35,1	39,1	42,6	46,7	49,6	52,5	56,0	58,3
30	36,3	40,3	43,8	48,0	50,9	54,0	57,5	59,7

Закінчення додатка 4

ЗНАЧЕННЯ ВЕЛИЧИНИ χ_2^2 ЗАЛЕЖНО ВІД ІМОВІРНОСТІ $P(\chi^2 > \chi_1^2)$

Число ступенів свободи, *k*	$P(\chi^2 > \chi_2^2)$							
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30
1	0,00016	0,0006	0,0039	0,016	0,064	0,148	0,455	1,07
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386	2,41
3	0,115	0,185	0,352	0,584	1,005	1,424	2,366	3,66
4	0,30	0,43	0,71	1,06	1,65	2,19	3,36	4,9
5	0,55	0,76	1,14	1,61	2,34	3,0	4,35	6,1
6	0,87	1,13	1,63	2,20	3,07	3,83	5,35	7,2
7	1,24	1,56	2,17	2,83	3,82	4,67	6,35	8,4
8	1,65	2,03	2,73	3,49	4,59	5,53	7,34	9,5
9	2,09	2,563	3,32	4,17	5,38	6,39	8,34	10,7
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,8
11	3,1	3,6	4,6	5,6	7,0	8,1	10,3	12,9
12	3,6	4,2	5,2	6,3	7,8	9,0	11,3	14,0
13	4,1	4,8	5,9	7,0	8,6	9,9	12,3	15,1
14	4,7	5,4	6,6	7,8	9,5	10,8	13,3	16,2
15	5,2	6,0	7,3	8,5	10,3	11,7	14,3	17,3
16	5,8	6,6	8,0	9,3	11,2	12,6	15,3	18,4
17	6,4	7,3	8,7	10,1	12,0	13,5	16,3	19,5
18	7,0	7,9	9,4	10,9	12,9	14,4	17,3	20,6
19	7,6	8,6	10,1	11,7	13,7	15,4	18,3	21,7
20	8,3	9,2	10,9	12,4	14,6	16,3	19,3	22,8
21	8,9	9,9	11,6	13,2	15,4	17,2	20,3	23,9
22	9,5	10,6	12,3	14,0	16,3	18,1	21,3	24,9
23	10,2	10,3	13,1	14,8	17,2	19,0	22,3	26,0
24	10,9	12,0	13,8	15,7	18,1	19,9	23,3	27,1
25	11,5	12,7	14,6	16,5	18,9	20,9	24,3	28,1
26	12,2	13,4	15,4	17,3	19,8	21,8	25,3	29,3
27	12,9	14,1	16,2	18,1	20,7	22,7	26,3	30,3
28	13,6	14,8	16,9	18,9	21,6	23,6	27,3	31,4
29	14,3	15,6	17,7	19,8	22,5	24,6	28,3	32,5
30	15,0	16,3	18,5	20,6	23,4	25,5	29,3	33,5

Додаток 5

ТАБЛИЦЯ ЗНАЧЕНЬ $q = q(\gamma, n)$

n	γ			n	γ		
	0,95	0,99	0,999		0,95	0,99	0,999
5	1,37	2,67	5,64	20	0,37	0,58	0,88
6	1,09	2,01	3,88	25	0,32	0,49	0,73
7	0,92	1,62	2,98	30	0,28	0,43	0,63
8	0,80	1,38	2,42	35	0,26	0,38	0,56
9	0,71	1,20	2,06	40	0,24	0,35	0,50
10	0,65	1,08	1,80	45	0,22	0,32	0,46
11	0,59	0,98	1,60	50	0,21	0,30	0,43
12	0,55	0,90	1,45	60	0,188	0,269	0,38
13	0,52	0,83	1,33	70	0,174	0,245	0,34
14	0,48	0,78	1,23	80	0,161	0,226	0,31
15	0,46	0,73	1,15	90	0,151	0,211	0,29
16	0,44	0,70	1,07	100	0,143	0,198	0,27
17	0,42	0,66	1,01	150	0,115	0,160	0,211
18	0,40	0,63	0,96	200	0,099	0,136	0,185
19	0,39	0,60	0,92	250	0,089	0,120	0,162

Додаток 6

КРИТИЧНІ ТОЧКИ РОЗПОДІЛУ СТЬЮДЕНТА (*t*-РОЗПОДІЛУ)

Число ступенів свободи, *k*	Рівень значущості, α						
	0,20	0,10	0,05	0,02	0,01	0,002	0,001
1	3,08	6,31	12,7	31,82	63,66	127,32	636,62
2	1,89	2,92	4,30	6,97	9,93	14,09	31,60
3	1,64	2,35	3,18	4,54	5,84	7,45	12,94
4	1,53	2,13	2,78	3,75	4,60	5,60	8,61
5	1,48	2,02	2,57	3,37	4,03	4,77	6,86
6	1,44	1,94	2,45	3,14	3,71	4,32	5,96
7	1,42	1,90	2,36	3,00	3,50	4,03	5,41
8	1,40	1,86	2,31	2,90	3,36	3,83	5,04
9	1,38	1,83	2,26	2,82	3,25	3,69	4,78
10	1,37	1,81	2,23	2,76	3,17	3,58	4,59
11	1,36	1,80	2,20	2,72	3,11	3,50	4,44
12	1,36	1,78	2,18	2,68	3,05	3,43	4,32
13	1,35	1,77	2,16	2,65	3,01	3,37	4,22
14	1,34	1,76	2,14	2,62	2,98	3,33	4,14
15	1,34	1,75	2,13	2,60	2,95	3,29	4,07
16	1,34	1,75	2,12	2,58	2,92	3,25	4,02
17	1,33	1,74	2,11	2,57	2,90	3,22	3,97
18	1,33	1,73	2,10	2,55	2,88	3,20	3,92
19	1,33	1,73	2,09	2,54	2,86	3,17	3,88
20	1,33	1,73	2,09	2,53	2,85	3,15	3,85
21	1,32	1,72	2,08	2,52	2,83	3,14	3,82
22	1,32	1,72	2,07	2,51	2,82	3,12	3,79
23	1,32	1,71	2,07	2,50	2,81	3,10	3,77
24	1,32	1,71	2,06	2,49	2,80	3,09	3,75
25	1,32	1,71	2,06	2,48	2,79	3,08	3,73
26	1,32	1,71	2,06	2,48	2,78	3,07	3,71
27	1,31	1,70	2,05	2,47	2,77	3,06	3,69
28	1,31	1,70	2,05	2,47	2,76	3,05	3,67
29	1,31	1,70	2,04	2,46	2,76	3,04	3,66
30	1,31	1,70	2,04	2,46	2,75	3,03	3,65
40	1,30	1,68	2,02	2,42	2,70	2,97	3,55
60	1,30	1,67	2,00	2,39	2,66	2,91	3,46
120	1,29	1,66	1,98	2,36	2,62	2,86	3,37
∞	1,28	1,64	1,96	2,33	2,58	2,81	3,29

Додаток 7

КРИТИЧНІ ТОЧКИ РОЗПОДІЛУ ФІШЕРА (*F*-РОЗПОДІЛУ)

Рівень значущості 0,05									
k_2 \ k_1	1	2	3	4	5	6	12	24	∞
1	164,4	199,5	215,7	224,6	230,2	234,0	244,9	249,0	254,3
2	18,5	9,2	19,2	19,3	19,3	19,3	19,4	19,5	19,5
3	10,1	9,6	9,3	9,1	9,0	8,9	8,7	8,6	8,5
4	7,7	6,9	6,6	6,4	6,3	6,2	5,9	5,8	5,6
5	6,6	5,8	5,4	5,2	5,1	5,0	4,7	4,5	4,4
6	6,0	5,1	4,8	4,5	4,4	4,3	4,0	3,8	3,7
7	5,6	4,7	4,4	4,1	4,0	3,9	3,6	3,4	3,2
8	5,3	4,5	4,1	3,8	3,7	3,6	3,3	3,1	2,9
9	5,1	4,3	3,9	3,6	3,5	3,4	3,1	2,9	2,7
10	5,0	4,1	3,7	3,5	3,3	3,2	2,9	2,7	2,5
11	4,8	4,0	3,6	3,4	3,2	3,1	2,8	2,6	2,4
12	4,8	3,9	3,5	3,3	3,1	3,0	2,7	2,5	2,3
13	4,7	3,8	3,4	3,2	3,0	2,9	2,6	2,4	2,2
14	4,6	3,7	3,3	3,1	3,0	2,9	2,5	2,3	2,1
15	4,5	3,7	3,3	3,1	2,9	2,8	2,5	2,3	2,1
16	4,5	3,6	3,2	3,0	2,9	2,7	2,4	2,2	2,0
17	4,5	3,6	3,2	3,0	2,8	2,7	2,4	2,2	2,0
18	4,4	3,6	3,2	2,9	2,8	2,7	2,3	2,1	1,9
19	4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,8
20	4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,8
22	4,3	3,4	3,1	2,8	2,7	2,6	2,2	2,0	1,8
24	4,3	3,4	3,0	2,8	2,6	2,5	2,2	2,0	1,7
26	4,2	3,4	3,0	2,7	2,6	2,4	2,1	1,9	1,7
28	4,2	3,3	2,9	2,7	2,6	2,4	2,1	1,9	1,6
30	4,2	3,3	2,9	2,7	2,5	2,4	2,1	1,9	1,6
40	4,1	3,2	2,9	2,6	2,5	2,3	2,0	1,8	1,5
60	4,0	3,2	2,8	2,5	2,4	2,3	1,9	1,7	1,4
120	3,9	3,1	2,7	2,5	2,3	2,2	1,8	1,6	1,3
∞	3,8	3,0	2,6	2,4	2,2	2,1	1,8	1,5	1,0

Рівень значущості 0,01										
k_2 \ k_1	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5859	5981	6106	6234	6366
2	98,5	99,0	99,2	99,3	99,3	99,4	99,3	99,4	99,5	99,5
3	34,1	30,8	29,5	28,7	28,2	27,9	27,5	27,1	26,6	26,1
4	21,2	18,0	16,7	16,0	15,5	15,2	14,8	14,4	13,9	13,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,3	9,9	9,5	9,0
6	13,7	10,9	9,8	9,2	8,8	8,5	8,1	7,7	7,3	6,9
7	12,3	9,6	8,5	7,9	7,5	7,2	6,8	6,5	6,1	5,7
8	11,3	8,7	7,6	7,0	6,6	6,4	6,0	5,7	5,3	4,9
9	10,6	8,0	7,0	6,4	6,1	5,8	5,5	5,1	4,7	4,3
10	10,0	7,6	6,6	6,0	5,6	5,4	5,1	4,7	4,3	3,9
11	9,7	7,2	6,2	5,7	5,3	5,1	4,7	4,4	4,0	3,6
12	9,3	6,9	6,0	5,4	5,1	4,8	4,5	4,2	3,8	3,4
13	9,1	6,7	5,7	5,2	4,9	4,6	4,3	4,0	3,6	3,2
14	8,9	6,5	5,6	5,0	4,7	4,5	4,1	3,8	3,4	3,0
15	8,7	6,4	5,4	4,9	4,6	4,3	4,0	3,7	3,3	2,9
16	8,5	6,2	5,3	4,8	4,4	4,2	3,9	3,6	3,2	2,8
17	8,4	6,1	5,2	4,7	4,3	4,1	3,8	3,5	3,1	2,7
18	8,3	6,0	5,1	4,6	4,3	4,0	3,7	3,4	3,0	2,6
19	8,2	5,9	5,0	4,5	4,2	3,9	3,6	3,3	2,9	2,4
20	8,1	5,9	4,9	4,4	4,1	3,9	3,6	3,2	2,9	2,4
22	7,9	5,7	4,8	4,3	4,0	3,8	3,5	3,1	2,8	2,3
24	7,8	5,6	4,7	4,2	3,9	3,7	3,3	3,0	2,7	2,2
26	7,7	5,5	4,6	4,1	3,8	3,6	3,3	3,0	2,6	2,1
28	7,6	5,5	4,6	4,1	3,8	3,5	3,2	2,9	2,5	2,1
30	7,6	5,4	4,5	4,0	3,7	3,5	3,2	2,8	2,5	2,0
40	7,3	5,2	4,3	3,8	3,5	3,3	3,0	2,7	2,3	1,8
60	7,1	5,0	4,1	3,7	3,3	3,1	2,8	2,5	2,1	1,6
120	6,9	4,8	4,0	3,5	3,2	3,0	2,7	2,3	2,0	1,4
∞	6,6	4,6	3,8	3,3	3,0	2,8	2,5	2,2	1,8	1,0

Закінчення додатка 7

Рівень значущості 0,001										
k_2 \ k_1	1	2	3	4	5	6	8	12	24	∞
1	Змінюється від 400 000 до 600 000									
2	998	999	999	999	999	999	999	999	999	999
3	167	148	141	137	135	133	131	128	126	123
4	74,1	61,3	56,2	53,4	51,7	50,5	49,0	47,4	45,8	44,1
5	47,0	36,6	33,2	31,1	29,8	28,8	27,6	26,4	25,1	23,8
6	35,5	27,0	23,7	21,9	20,8	20,0	19,0	18,0	16,9	15,8
7	29,2	21,7	18,8	17,2	16,2	15,5	14,6	13,7	12,7	11,7
8	25,4	18,5	15,8	14,4	13,5	12,9	12,0	11,2	10,3	9,3
9	22,9	16,4	13,9	12,6	11,7	11,1	10,4	9,6	8,7	7,8
10	21,0	14,9	12,6	11,3	10,5	9,9	9,2	8,5	7,6	6,8
11	19,7	13,8	11 ,6	10,4	9,6	9,1	8,3	7,6	6,9	6,0
12	18,6	13,0	10,8	9,6	8,9	8,4	7,7	7,0	6,3	5,4
13	17,8	12,3	10,2	9,1	8,4	7,9	7,2	6,5	5,8	5,0
14	17,1	11,8	9,7	8,6	7,9	7,4	6,8	6,1	5,4	4,6
15	16,6	11,3	9,3	8,3	7,6	7,1	6,5	5,8	5,1	4,3
16	16,1	11,0	9,0	7,9	7,3	6,8	6,2	5,6	4,9	4,1
17	15,7	10,7	8,7	7,7	7,0	6,6	6,0	5,3	4,6	3,9
18	15.4	10,4	8,5	7,5	6,8	6,4	5,8	5,1	4,5	3,7
19	15,1	10,2	8,3	7,3	6,6	6,2	5,6	5,0	4,3	3,5
20	14,8	10,0	8,1	7,1	6,5	6,0	5,4	4,8	4,2	3,4
22	14,4	9,6	7,8	6,8	6,2	5,8	5,2	4,6	3,9	3,2
24	14,0	9,3	7,6	6,6	6,0	5,6	5,0	4,4	3,7	3,0
26	13,7	9,1	7,4	6,4	5,8	5,4	4,8	4,2	3,6	2,8
28	13,5	8,9	7,2	6,3	5,7	5,2	4,7	4,1	3,5	2,7
30	13,3	8,8	7,1	6,1	5,5	5,1	4,6	4,0	3,4	2,6
40	12,6	8,2	6,6	5,7	5,1	4,7	4,2	3,6	3,0	2,2
60	12,0	7,8	6,2	5,3	4,8	4,4	3,9	3,3	2,7	1,9
120	11,4	7,3	5,8	5,0	4,4	4,0	3,5	3,0	2,4	1,6
∞	10,8	6,9	5,4	4,6	4,1	3,7	3,3	2,7	2,1	1,0

Додаток 8

КРИТИЧНІ ТОЧКИ РОЗПОДІЛУ χ^2

Число ступенів свободи, *k*	Рівень значущості, α					
	0,01	0,025	0,05	0,95	0,975	0,999
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38.9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	60,9	47,0	43,8	18,5	16,8	15,0

ЗМІСТ