МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ШОСТКИНСЬКИЙ ІНСТИТУТ Сум ДУ

ВИЩА МАТЕМАТИКА

Лекції, практичні, розв’язання прикладів, завдання для самостійної роботи

для студентів всіх спеціальностей освітньо-кваліфікаційного рівня «бакалавр» денної та заочної форм навчання

У шести частинах

Частина 6

З М І С Т

ВСТУП

Основна форма навчання студентів – самостійна робота над навчальним матеріалом, яка складається з вивчення теоретичних положень за підручником, розгляду прикладів і розв'язання задач. При вивченні матеріалу за підручником треба переходити до наступного питання тільки після правильного зрозуміння попереднього, виконуючи на папері усі обчислення, навіть і ті, які пропущені у підручнику. Розв'язання задач при вивченні дисципліни «Вища математика» часто пов'язано з багатьма складностями. Якщо складається скрутне становище при розв'язанні задачі, то треба вказати характер цього утруднення, привести припущення відносно плану розв'язку.

Відомо, що при самостійному розв'язуванні задач студентам потрібні постійні консультації щодо способів їх розв'язування, оскільки знайти шлях до розв'язування задачі без допомоги викладача або відповідного підручника студентові не під силу. Допомогти студентам технічних спеціальностей всіх форм навчання подолати ці складності, навчити їх застосовувати теоретичні знання до розв'язування задач – основне призначення цього навчального посібника.

У шостій частині навчального посібника викладено матеріал з таких розділів вищої математики: «Основні поняття математичної статистики», «Двовимірний статистичний розподіл вибірки», «Статистичні гіпотези» та «Елементи кореляційного та регресійного аналізу». Основні теоретичні положення, формули та теореми ілюструються докладним розв'язанням великої кількості задач різного рівня складності з їх повним аналізом. Для ефективності засвоєння матеріалу пропонуються завдання для самостійної роботи.

Розв'язання задач ***математичної статистики*** потребує великого обсягу обробки даних та обчислень. Відзначимо, що потужним інструментарієм статистичної обробки даних володіють електронні таблиці **Microsoft Excel.** Дуже корисним і простим є використання інших програм, наприклад: **NUMERI, MATLAB, MathCad, STATISTIKA, STATGRAPHICS.**

Автори сподіваються, що саме така побудова посібника надає студентові широкі можливості до активної самостійної роботи, яка, безумовно, сприятиме засвоєнню матеріалу при вивченні дисципліни «Вища математика».

Розділ 1
ОСНОВНІ ПОНЯТТЯ МАТЕМАТИЧОЇ СТАТИСТИКИ

1.1. Дискретний статистичний розподіл вибірки та окремі її числові характеристики

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

При реалізації вибірки кількісна ознака, наприклад X, набуває конкретних числових значень ($X = x_i$), які називають ***варіантами.***

Зростаючий числовий ряд варіант називають ***варіаційним рядом.***

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

При цьому

$$n = \sum_{i=1}^{k} n_i \ , \tag{1.1}$$

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

$$W_i = \frac{n_i}{n} . \tag{1.2}$$

Для кожної вибірки виконується рівність

$$\sum_{i=1}^{k} W_i = 1 . \tag{1.3}$$

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

$$F^*(x) = W(X < x) = \frac{n_x}{n} , \tag{1.4}$$

називається ***емпіричною функцією.***

Тут $\boldsymbol{n}$ – обсяг вибірки, $\boldsymbol{n_x}$ – кількість варіант статистичного розподілу вибірки, значення яких менше за фіксовану варіанту $\boldsymbol{x}$; $\boldsymbol{F^*(x)}$ – називають ще ***функцією розподілу відносних частот***.

Властивості $\boldsymbol{F^*(x)}$**:**

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

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

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

4) $\boldsymbol{F^*(x)}$ є неспадною функцією аргументу $\boldsymbol{x}$**,** а саме: $\boldsymbol{F(x_2) \ge F(x_1)}$ 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 $\boldsymbol{(x_i;\ n_i)}$ або $\boldsymbol{(x_i;\ W_i)}$**.**

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

***Числові характеристики*:**

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

$$\bar{x}_{\text{В}} = \frac{\sum_{i=1}^{k} x_i n_i}{n}, \tag{1.5}$$

називають ***вибірковою середньою величиною дискретного статистичного розподілу вибірки***.

Тут $\boldsymbol{x_i}$ – варіанта варіаційного ряду вибірки; $\boldsymbol{n_i}$ – частота цієї варіанти; $\boldsymbol{n}$ – обсяг вибірки (1.1);

2**) *мода*** **(Mo*)**. ***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.;

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

4**) *розмах*** **(*R*).** Для грубого оцінювання розсіювання варіант відносно $\boldsymbol{\bar{x}_{\text{В}}}$ застосовується величина, яка дорівнює різниці між найбільшою $\boldsymbol{x_{\max}}$ і найменшою $\boldsymbol{x_{\min}}$ варіантами варіаційного ряду. Ця величина називається ***розмахом***

$$\boldsymbol{R = x_{\max} - x_{\min}}. \tag{1.6}$$

Зразки розв'язування задач

Приклад 1. За заданим дискретним статистичним розподілом вибірки потрібно: побудувати $F^*(x)$ і зобразити її графічно; накреслити полігони частот і відносних частот:

$X = x_i$	– 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

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. 1.1.

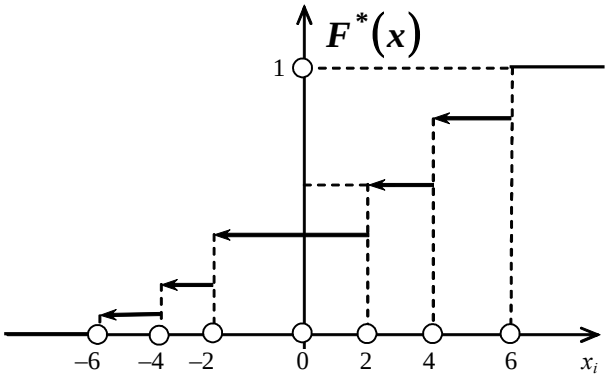


Рис. 1.1

Полігони частот та відносних частот зображено на рис. 1.2, 1.3. Таким чином графіки відрізняються тільки масштабом за віссю Y .

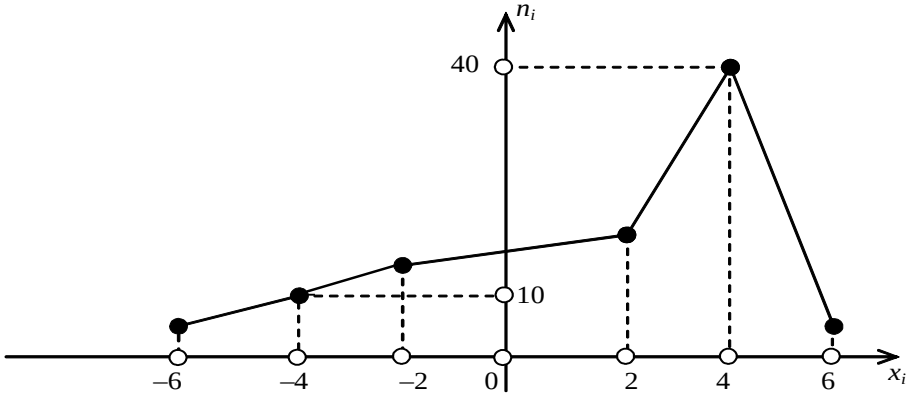


Рис. 1.2

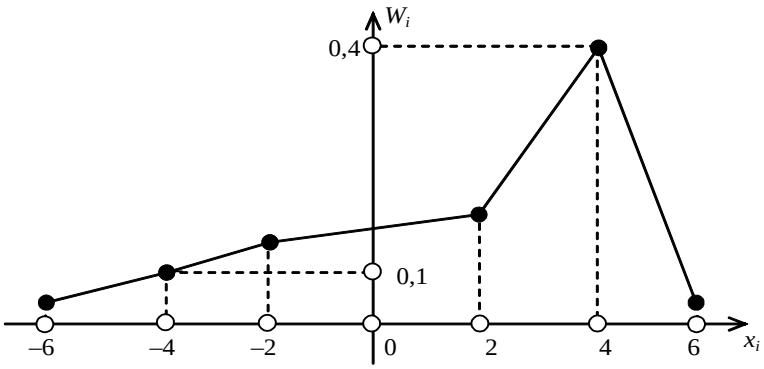


Рис. 1.3

Приклад 2. За заданим статистичним розподілом вибірки потрібно знайти **Мо***, **Ме* :**

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

Розв'язання. Обсяг вибірки $n = \sum_{i=1}^{5} n_i = 100$; **Мо*** = 6,5; 8,5.

Отже, наведений статистичний розподіл вибірки буде двомодальним. **Ме***=6,5, оскільки варіанта x_3=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.

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

Приклад 3. Вибірка, що зроблена випадковим способом, задається наведеними даними: 0,4; 0,4; 1,0; 0,3; 0,6; 0,9; 0,5; 0,4; 0,8; 0,6; 0,5; 0,5; 0,4; 0,8; 0,3; 0,3; 0,6; 0,7; 0,7; 0,3; 0,2; 0,1; 0,4; 0,5; 0,7; 0,2; 0,6; 0,5; 0,4; 0,9.

Потрібно: 1) знайти статистичний розподіл вибірки; 2) знайти **Мо***, **Ме*;** 3) побудувати полігон частот; 4) побудувати емпіричну функцію розподілу $F^*(x)$; 5) побудувати графік емпіричної функції розподілу $F^*(x)$.

Розв'язання.

1) Обсяг вибірки n = **30** ; x_{max}= **1,0;** x_{min}= **0,1 ;** розмах Δ = **1,0 – 0,1 = 0,9.** Заповнюємо стовпці **A, N** від меншого до більшого значення та в стовпці **B** робимо відмітку **(** I **)** в відповідному рядку про те, що дане значення варіанти нам зустрічалося. У стовпці **C** підводимо підсумки про кількість відміток в даному рядку, тобто вказуємо частоти цих варіант. Стовпці таблиці **A** і **C (A** і **C*** за *відносною частотою* **)** створюють шуканий статистичний розподіл.

N	A	B	C	C*
№ п/п	x_i	n_i	n_i	n_i/n
1	0.1	I	1	1/30
2	0.2	I I	2	2/30
3	0.3	I I I I	4	4/30
4	0.4	I I I I I I	6	6/30
5	0.5	I I I I I	5	5/30
6	0.6	I I I I	4	4/30
7	0.7	I I I	3	3/30
8	0.8	I I	2	2/30
9	0.9	I I	2	2/30
10	1.0	I	1	1/30
$\sum$	----	----------	30	1

2) **Mo*** =0,4 - варіанта, що має найбільшу частоту появи.

Отже, наведений статистичний розподіл вибірки буде одномодальним.

Me* = 0,5 і 0,6 - варіанти, які поділяють варіаційний ряд на дві частини, рівні за кількістю варіант.

3) Для побудови полігону частот (відносних частот) на осі абсцис відкладають варіанти x_i, на осі ординат – відповідні їм частоти n_i (відносні частоти $W_i = n_i/n$). Точки (x_i, n_i) ($(x_i, n_i/n)$ – для відносних частот) сполучають відрізками прямих і одержують полігон частот (відносних частот).

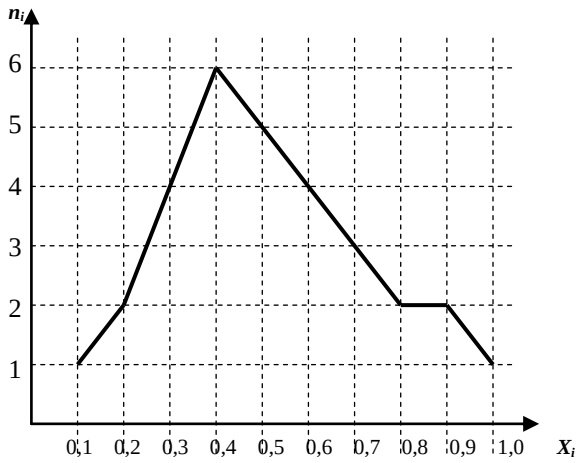


Рис. 1.4. Полігон частот

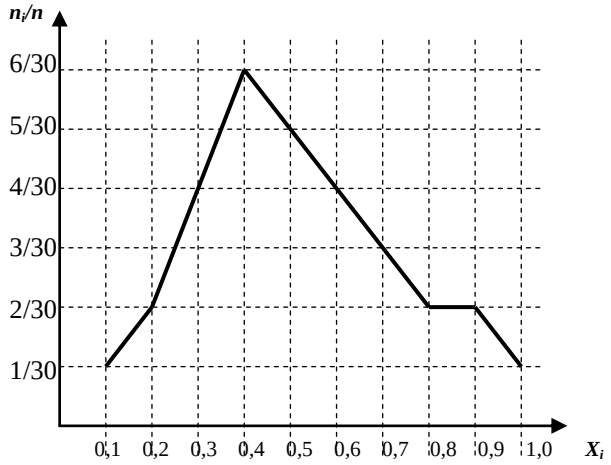


Рис. 1.5. Полігон відносних частот

Таким чином графіки відрізняються тільки масштабом за віссю ***Y*** .

4)

$$F^*(x)=\begin{cases} 0, & \textit{якщо} \quad x \le 0{,}1 \\ 1/30, & \textit{якщо} \quad 0{,}1 < x \le 0{,}2 \\ 3/30, & \textit{якщо} \quad 0{,}2 < x \le 0{,}3 \\ 7/30, & \textit{якщо} \quad 0{,}3 < x \le 0{,}4 \\ 13/30, & \textit{якщо} \quad 0{,}4 < x \le 0{,}5 \\ 18/30, & \textit{якщо} \quad 0{,}5 < x \le 0{,}6 \\ 22/30, & \textit{якщо} \quad 0{,}6 < x \le 0{,}7 \\ 25/30, & \textit{якщо} \quad 0{,}7 < x \le 0{,}8 \\ 27/30, & \textit{якщо} \quad 0{,}8 < x \le 0{,}9 \\ 29/30, & \textit{якщо} \quad 0{,}9 < x \le 1{,}0 \\ 1, & \textit{якщо} \quad x > 1{,}0 \end{cases}$$

Найменше спостережне значення варіанти дорівнює **0,1,** отже $F^*(x)=0$ при $x \le 0{,}1$.

Значення $x \le 0{,}2$, а значить $x_1=0{,}1$, спостерігалося 1 раз, отже, $F^*(x)=1/30$, якщо $0{,}1 < x \le 0{,}2$.

Значення $x \le 0.3$ спостерігалося **1+2=3** рази. Отже $F^*(x)=3/30$, якщо $0{,}2 < x \le 0{,}3$.

Значення $x \le 0{,}4$ спостерігалося **3+4=7** разів. Отже $F^*(x)=7/30$, якщо $0{,}3 < x \le 0{,}4$.

Значення $x \le 0{,}5$ спостерігалося **7+6=13** разів. Отже $F^*(x)=13/30$, якщо $0{,}4 < x \le 0{,}5$.

Значення $x \le 0{,}6$ спостерігалося **13+5=18** разів. Отже $F^*(x)=18/30$, якщо $0{,}5 < x \le 0{,}6$.

Значення $x \le 0{,}7$ спостерігалося **18+4=22** разів. Отже $F^*(x)=22/30$, якщо $0{,}6 < x \le 0{,}7$.

Значення $x \le 0{,}8$ спостерігалося **22+3=25** разів. Отже $F^*(x)= 25/30$, якщо $0{,}7 < x \le 0{,}8$.

Значення $x \le 0{,}9$ спостерігалося **25+2=27** разів. Отже $F^*(x)=27/30$, якщо $0{,}8 < x \le 0{,}9$.

Значення $x \le 1{,}0$ спостерігалося **27+2=29** разів. Отже $F^*(x)=29/30$, якщо $0{,}9 < x \le 1{,}0$.

Найбільше значення, що спостерігалося, $x=1{,}0$. Отже $F^*(x > 1{,}0)=1$.

5)

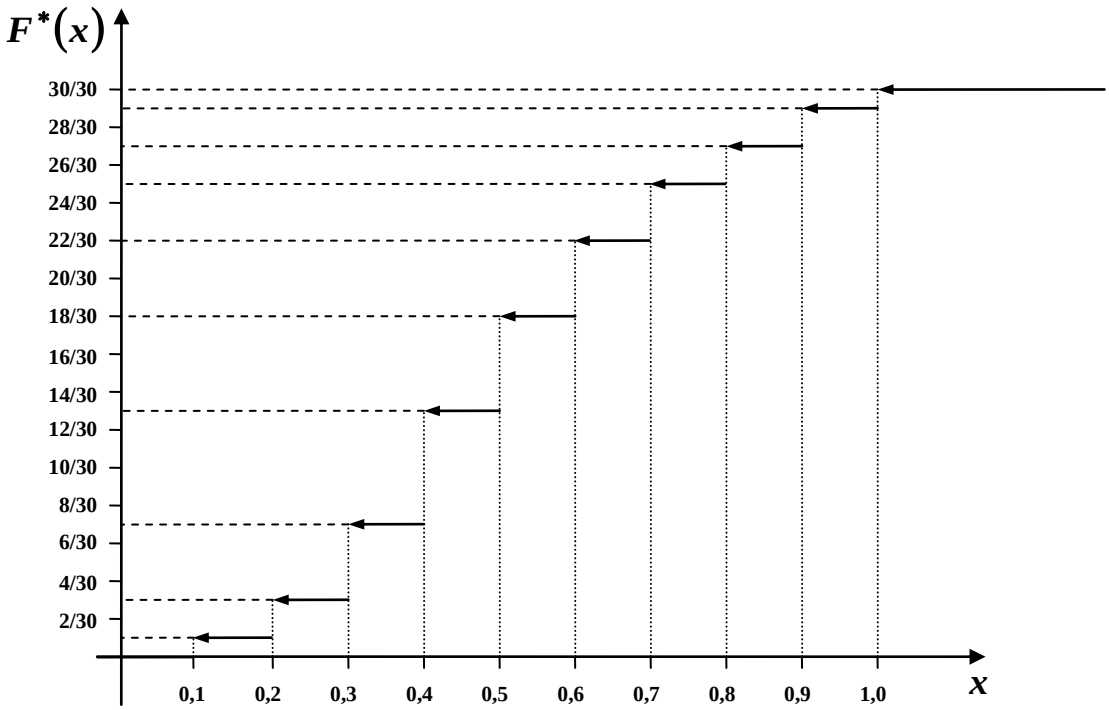


Рис. 1.6

Завдання для самостійної роботи

Вибірки, що зроблені випадковим чином, задаються наведеними даними. Виконати наступні вправи:

1) побудувати статистичний розподіл вибірки;

2) побудувати полігони частот і відносних частот;

3) знайти емпіричну функцію розподілу $F^*(x)$;

4) побудувати графік емпіричної функції розподілу $F^*(x)$;

5) знайти **Mo*, Me***.

1. Наведені дані про витрати часу на виробництво одиниці продукції: 0,09; 0,09; 0,11; 0,09; 0,08; 0,09; 0,06; 0,10; 0,09; 0,11; 0,09; 0,09; 0,11; 0,09; 0,08; 0,09; 0,06; 0,08; 0,09; 0,08; 0,10; 0,08; 0,07; 0,09; 0,07; 0,10; 0,07; 0,09; 0,10; 0,06.

2. При свердлінні 50 отворів одним і тим самим свердлом з наступними вимірюваннями діаметрів отворів отримані такі дані в мм: ; 40,29;40,32; 40,32; 40,33; 40,2; 40,29; 40,31; 40,35; 40,28; 40,29; 40,29; 40,35; 40,34; 40,33; 40,34; 40,31; 40,33; 40,34; 40,32; 40,30; 40,31; 40,30; 40,30; 40,33; 40,32; 40,33; 40,30; 40,32; 40,31; 40,32.

3. Дано статистичний розподіл:

$X = x_i$	5	7	10	13	15	18
n_i	133	48	16	6	4	3

4. Дано статистичний розподіл:

$X = x_i$	20	24	27	29	30
n_i	123	97	78	54	500

5. Дано статистичний розподіл:

$X = x_i$	10	12	13	15	18	20
n_i	365	247	152	102	72	47

Наступні завдання виконати у с н о :

1) визначити обсяг вибірки: 1,3, 5, 7,9,11,13,15,17,19, 21;
2) визначити відносні частоти (статистичні ймовірності) для варіант вибірки із завдання 1;
3) скласти статистичний ряд розподілу частот для вибірки:
   5, 5, 3, 3, 4, 4, 1, 1, 2, 2, 1, 2, 3, 4, 5;
4) знайти моду, медіану і статистичне середнє вибірки, заданої статистичним рядом:

$X = x_i$	–5	–1	0	1	5
n_i	10	8	4	8	10

5) скласти статистичний ряд розподілу відносних частот для вибірки із завдання 4.

Перевірити, чи правильні такі твердження:

1) спостережувані величини обов'язково є випадковими величинами;
2) генеральна сукупність – це множина значень спостережуваної величини;
3) вибірка спостережених значень завжди співпадає з відповідною генеральною сукупністю;
4) вибірка завжди є варіаційним рядом;
5) обсяг вибірки – це кількість спостережених значень;
6) спостережені значення є попарно різними;
7) кількість варіант у вибірці співпадає з її об'ємом;
8) за статистичним рядом розподілу частот даної вибірки можна визначити абсолютну частоту кожної варіанти вибірки;
9) за полігоном частот можна скласти відповідний статистичний ряд розподілу відносних частот;
10) за розмахом вибірки можна визначити її мінімальне та максимальне спостережене значення;
11) кожна вибірка має єдину моду;
12) кожна вибірка має єдину медіану.

1.2. Інтервальний статистичний розподіл вибірки та окремі її числові характеристики

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 з рівними між собою інтервалами.

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

У табличній формі цей розподіл має такий вигляд:

$x_i – x_{i+1}$	$x_1 - x_2$	$x_2 - x_3$	$x_3 - x_4$	…	$x_{k-1} - x_k$
n_i	n_1	n_2	n_3	…	n_k
W_i	W_1	W_2	W_3	…	W_k

Величина $\Delta x_i = x_{i+1} - x_i$ є довжиною часткового i-го інтервалу, k – кількість інтервалів. Як правило, цей інтервал береться однаковим.

Отже, для непевних величин групування вибіркових значень проводяться наступним чином: весь діапазон значень, що досліджуються, ділять на інтервали (розряди) та підраховують число значень, які належать кожному з інтервалів.

Позначивши через I_i інтервал, через x_i, x_{i+1} ***межі*** інтервалів, а через n_i – число значень величини X, що досліджувались в i – тому інтервалі, отримаємо наступний інтервальний варіаційний ряд:

I_i	$[x_1; x_2)$	$[x_2; x_3)$	…	$[x_i; x_{i+1})$	…	$[x_k; x_{k+1})$
n_i	n_1	n_2	…	n_i	…	n_k

Квадратна дужка означає, що тут включається лівий кінець інтервалу і не включається правий (хоча можна домовитись про включення правого і невключення лівого).

Для визначення раціональної величини інтервалу $\Delta x_i = h$ треба використовувати формулу Стерджеса:

$$h = (x_{max} - x_{min})/(1+3{,}322 \lg n). \qquad (1.7)$$

Якщо h буде дробовим числом, то його округляють до найближчого цілого або до найближчого нескладного дробу. За x_1 вважають значення, яке дорівнює

$$x_1 = x_{min} - h/2. \qquad (1.8)$$

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

Гістограма частот та відносних частот. Гістограма частот складається з прямокутників, кожний з яких має основу h і висоту n_i/h.

Гістограма відносних частот є фігурою, що складається з прямокутників, кожний з яких має основу завдовжки h і висоту, що дорівнює W_i/h.

Площа гістограми відносних частот $S = \sum_{i=1}^{k} h\frac{W_i}{h} = \sum_{i=1}^{k} W_i = 1$.

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

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 $[x_i–x_{i+1}]$ $F^*(x_i) < 0,5$ і $F^*(x_{i+1}) > 0,5$, то, беручи до уваги, що досліджувана ознака X є неперервною і при цьому $F^*(x)$ є неспадною функцією, всередині інтервалу $[x_i–x_{i+1}]$ неодмінно існує таке значення $X=\text{Me}^*$, де $F^*(\text{Me}^*)=0,5$.

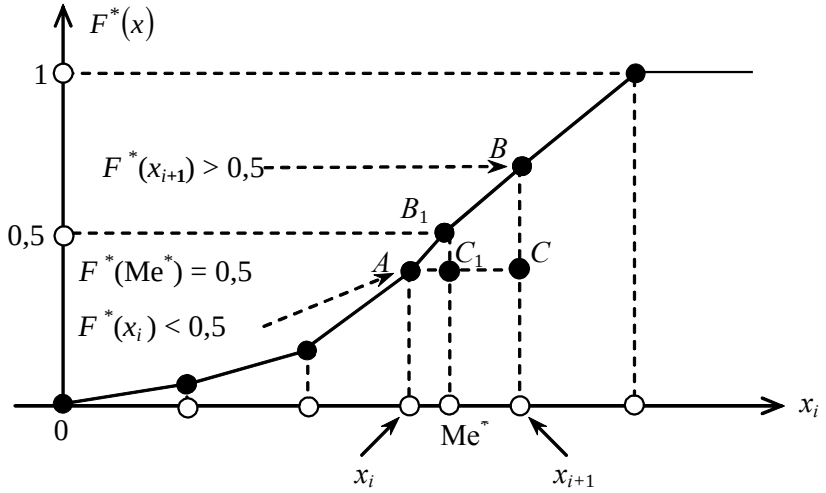


Рис. 1.7

З подібності трикутників ΔABC і ΔAB_1C_1, які зображені на рис. 1.7, маємо:

$$\frac{x_{i+1}-x_i}{\text{Me}^*-x_i} = \frac{F^*(x_{i+1})-F^*(x_i)}{0,5-F^*(x_i)} \rightarrow \text{Me}^* = x_i + \frac{0,5-F^*(x_i)}{F^*(x_{i+1})-F^*(x_i)}h, \qquad (1.9)$$

де $h = x_{i+1} - x_i$ 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:

$$\mathbf{Mo}^* = x_i + \frac{n_{\mathrm{Mo}} - n_{\mathrm{Mo}-1}}{2n_{\mathrm{Mo}} - n_{\mathrm{Mo}-1} - n_{\mathrm{Mo}+1}} h, \qquad (1.10)$$

де x_i – початок модального інтервалу; h – довжина, або крок, часткового інтервалу; n_{Mo} – частота модального інтервалу; $n_{\mathrm{Mo}-1}$ – частота домодального інтервалу; $n_{\mathrm{Mo}+1}$ — частота позамодального інтервалу.

Зразки розв'язування задач

Приклад 1. Результати вимірювань внутрішніх діаметрів шестірень: **6,75; 6,77; 6,77; 6,73; 6,76; 6,69; 6,81; 6,80; 6,72; 6,75; 6,78; 6,82; 6,74; 6,76; 6,89; 6,70; 6,71; 6,76: 6,75; 6,78.**

Знайти інтервальний варіаційний ряд.

Розв'язання. $n = 20$, $h = (6{,}89 - 6{,}69)/(1+3{,}322 \lg 20) = 0{,}037 \approx 0{,}04$. $x_1 = 6{,}69 - 0{,}02 = 6{,}67$, отримаємо наступний інтервальний варіаційний ряд

I_i	[6,67;6,71)	[6,71;6,75)	[6,75;6,79)	[6,79;6,83)	[6,83;6,87)	[6,87;6,91)
n_i	2	4	10	3	0	1

Приклад 2. За заданим інтервальним статистичним розподілом вибірки потрібно побудувати гістограму частот і відносних частот:

I_i	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

Розв'язання. Гістограми частот і відносних частот наведені на рис. 1.8, 1.9.

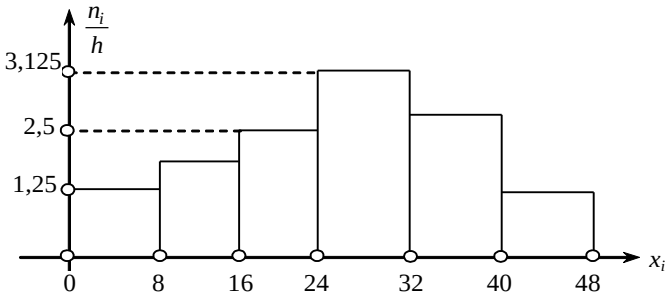


Рис. 1.8

Площа гістограми частот $S = \sum_{i=1}^{6} h \frac{n_i}{h} = \sum_{i=1}^{6} n_i = n = 100.$

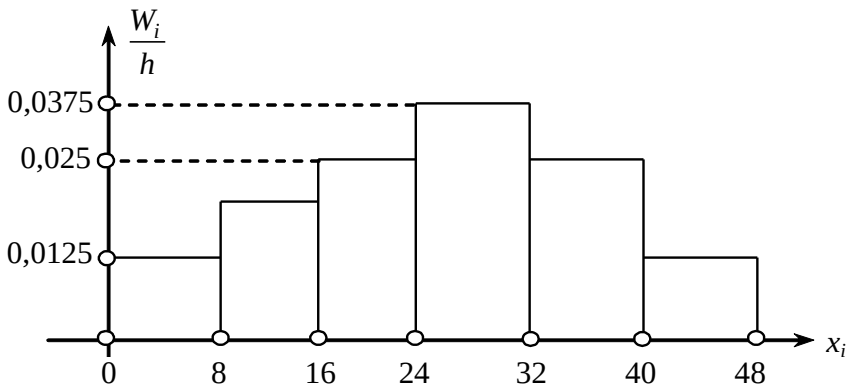


Рис. 1.9

Приклад 3. Для заданого інтервального статистичного розподілу вибірки побудувати $\boldsymbol{F^*(x)}$ і подати її графічно:

I_i	0–10	10–20	20–30	30–40	40–50	50–60
n_i	5	15	20	25	30	5

Розв'язання.

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

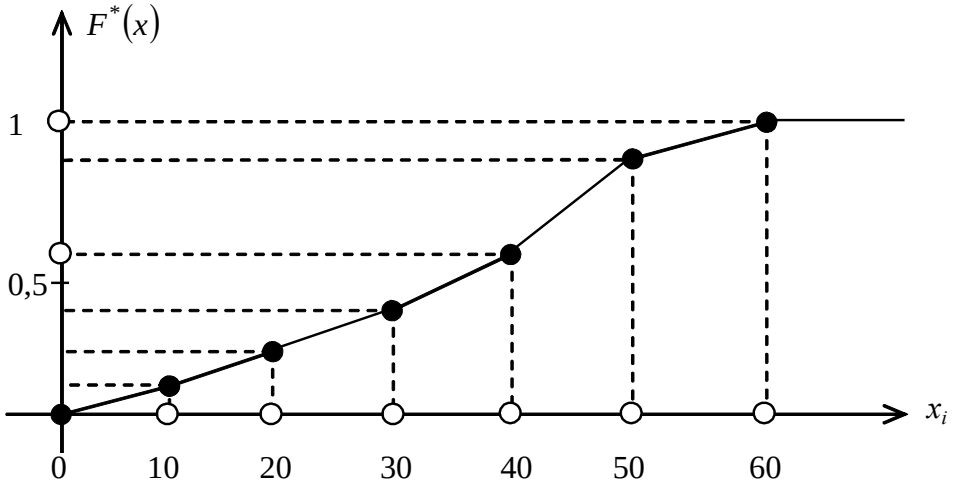


Рис. 1.10

На рис. 1.10. зображено графік $\boldsymbol{F^*(x)}$

Приклад 4. За заданим інтервальним статистичним розподілом вибірки побудувати гістограму частот і $F^*(x)$ і визначити $\mathbf{Mo}^*$, $\mathbf{Me}^*$.

I_i	0–4	4–8	8–12	12–16	16–20	20–24
n_i	6	14	20	25	30	5

Розв'язання. Гістограма частот зображена на рис. 1.11.

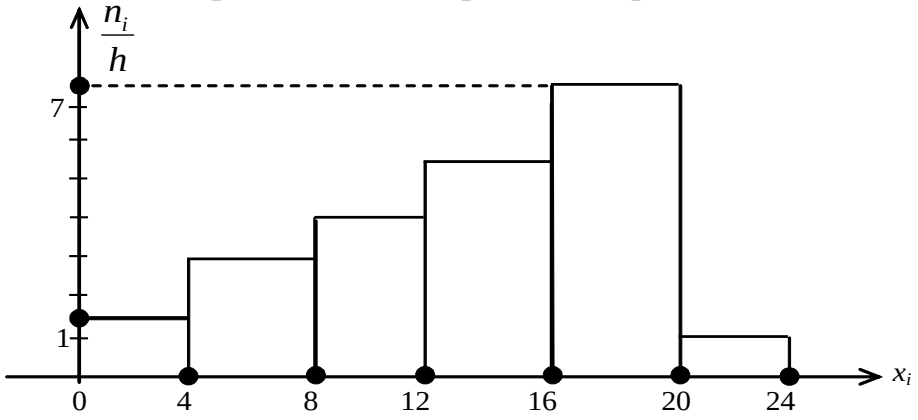


Рис. 1.11

Графік $F^*(x)$ зображено на рис. 1.12.

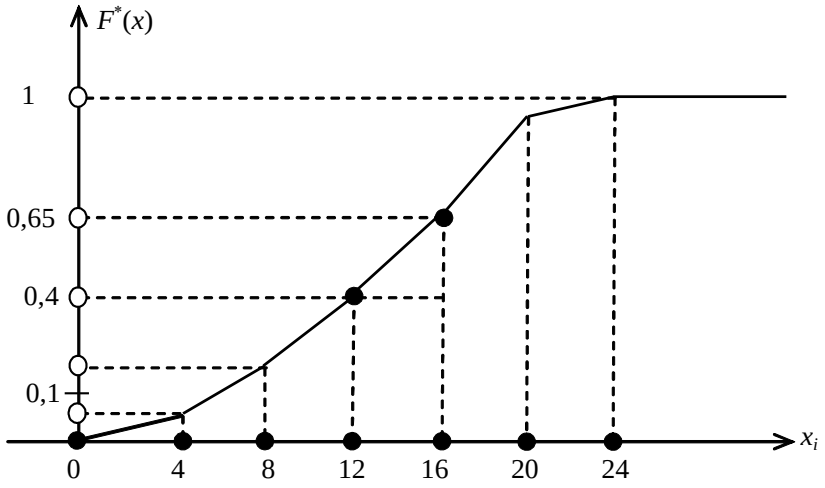


Рис. 1.12

З рис. 1.11 визначається модальний інтервал, який позначається (16–20).

Застосовуючи (1.8) і беручи до уваги, що $n_{\mathbf{Mo}} = 30$, $n_{\mathbf{Mo}-1} = 25$, $n_{\mathbf{Mo}+1} = 5$, $h = 4$, $x_i = 16$, дістанемо

$$\mathbf{Mo}^* = x_i + \frac{n_{\mathbf{Mo}} - n_{\mathbf{Mo}-1}}{2n_{\mathbf{Mo}} - n_{\mathbf{Mo}-1} - n_{\mathbf{Mo}+1}} h;$$

$$\mathbf{Mo}^* = 16 + \frac{30-25}{60-25-5} \cdot 4 = 16 + \frac{5 \cdot 4}{30} = 16{,}67.$$

Отже, $\mathbf{Mo}^* = \mathbf{16{,}67}$.

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

Беручи до уваги, що $F^*(12) = 0{,}4$, $F^*(16) = 0{,}65$, $h = 4$ і застосовуючи (1.7), дістанемо:

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

Отже, $\mathbf{Me}^* = \mathbf{13{,}6}$.

Завдання для самостійної роботи

Вибірки, що зроблені випадковим чином, задаються наведеними даними. Виконати наступні вправи:

1) побудувати інтервальний статистичний розподіл вибірки;
2) побудувати гістограму відносних частот і $F^*(x)$.
3) Визначити $\mathbf{Mo}^*, \mathbf{Me}^*$.

1. Наведені дані про витрати часу на виробництво одиниці продукції:
0,09; 0,09; 0,11; 0,09; 0,08; 0,09; 0,06; 0,10; 0,09; 0,11; 0,09; 0,09; 0,11; 0,09; 0,08; 0,09; 0,06; 0,08; 0,09; 0,08; 0,10; 0,08; 0,07; 0,09; 0,07; 0,10; 0,07; 0,09; 0,10; 0,06.

2. При свердлінні 50 отворів одним і тим самим свердлом з наступними вимірюваннями діаметрів отворів отримані такі дані в мм:
40,29; 40,32; 40,32; 40,33; 40,2; 40,29; 40,31; 40,35; 40,28; 40,29; 40,29; 40,35; 40,34; 40,33; 40,34; 40,31; 40,33; 40,34; 40,32; 40,30; 40,31; 40,30; 40,30; 40,33; 40,32; 40,33; 40,30; 40,32; 40,31; 40,32.

Для заданих статистичних розподілів побудувати гістограму відносних частот і $F^*(x)$. Визначити $\mathbf{Mo}^*, \mathbf{Me}^*$.

3. Дано статистичний розподіл:

I_i	[0; 400)	[400; 800)	[800; 1200)	[1200; 1600)
n_i	123	97	78	54

4. Дано статистичний розподіл:

I_i	[0; 5)	[5; 10)	[10; 15)	[15; 20)	[20; 25)	[25; 30)
n_i	133	48	16	6	4	3

5. Дано статистичний розподіл:

I_i	[0; 10)	[10; 20)	[20; 30)	[30; 40)	[40; 50)	[50; 60)
n_i	365	247	152	102	72	47

Перевірити, чи правильні такі твердження:

1) інтервальний статистичний ряд розподілу відносних частот застосовують лише для неперервної випадкової величини;

2) гістограма – це графік щільності розподілу статистичних ймовірностей.

Розділ 2
ТОЧКОВІ ТА ІНТЕРВАЛЬНІ ОЦІНКИ СТАТИСТИЧНИХ ХАРАКТЕРИСТИК ТА ПАРАМЕТРІВ ГЕНЕРАЛЬНОЇ СУКУПНОСТІ

2.1. Точкові оцінки статистичних характеристик та параметрів генеральної сукупності

На практиці виникає необхідність за даними вибірки зробити оцінки параметрів закону розподілу. Статистичні оцінки можна поділити на точкові та інтервальні.

Точковою називають оцінку, яка визначається одним числом. При малому об'ємі вибірки точкові оцінки можуть мати значне розходження зі значенням параметра, що оцінюється. Це, у свою чергу, може призвести до грубих помилок. Більш точними є інтервальні оцінки.

Незміщеною називають статистичну оцінку, математичне сподівання якої дорівнює параметру, що оцінюється, при будь-якому об'ємі вибірки.

***Числові характеристики*:**

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

$$\bar{x}_{\text{В}} = \frac{\sum_{i=1}^{k} x_i n_i}{n}, \tag{2.1}$$

називають ***вибірковою середньою величиною дискретного статистичного розподілу вибірки*.**

Тут x_i – варіанта варіаційного ряду вибірки; n_i – частота цієї варіанти; n – обсяг вибірки ($n = \sum_{i=1}^{k} n_i$).

Вибіркова середня величина $\bar{x}_{\text{В}}$ є незміщеною оцінкою математичного сподівання генеральної сукупності.

2) ***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

$$D_{\text{В}} = \frac{\sum_{i=1}^{k} (x_i - \bar{x}_{\text{В}})^2 n_i}{n} \tag{2.2}$$

або

$$D_{\text{В}} = \frac{\sum_{i=1}^{k} x_i^2 n_i}{n} - (\bar{x}_{\text{В}})^2. \tag{2.3}$$

Величина $D_{\text{В}}$ є зміщеною оцінкою дисперсії генеральної сукупності.

Незміщеною оцінкою дисперсії генеральної сукупності є ***виправлена дисперсія***, яка може бути розрахована за формулою

$$s^2 = \frac{n}{n-1} D_B. \qquad (2.4)$$

При великих об'ємах n вибіркова і виправлена дисперсії відрізняються мало. На практиці користуються виправленою дисперсією приблизно при $n < 30$.

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

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}}, \qquad (2.5)$$

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

4) ***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:

$$V = \frac{\sigma_{\text{В}}}{\bar{x}_{\text{В}}} 100\%. \qquad (2.6)$$

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

x_i^*	x_1^*	x_2^*	x_3^*	…	x_k^*
n_i	n_1	n_2	n_3	…	n_k

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

$$\bar{x}_{\text{В}} = \frac{\sum_{i=1}^{k} x_i^* n_i}{h}; \qquad (2.7)$$

$$D_{\text{В}} = \frac{\sum_{i=1}^{k} (x_i^*)^2 n_i}{h} - (\bar{x}_{\text{В}})^2; \qquad (2.8)$$

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}}. \qquad (2.9)$$

Зразки розв'язування задач

Приклад 1. За заданим статистичним розподілом вибірки потрібно обчислити $\bar{x}_{\text{В}}$, $D_{\text{В}}$, $\sigma_{\text{В}}$, V:

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

Розв'язання. Оскільки $n = \sum_{i=1}^{5} n_i = 100$, то згідно з формулами (2.1), (2.3), (2.4), (2.5), (2.6) дістанемо:

$$\bar{x}_{\text{В}} = \frac{\sum_{i=1}^{5} x_i n_i}{n} = \frac{2{,}5 \cdot 10 + 4{,}5 \cdot 20 + 6{,}5 \cdot 30 + 8{,}5 \cdot 30 + 10{,}5 \cdot 10}{100} = 6{,}7; \quad \bar{x}_{\text{В}} = 6{,}7.$$

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

$$\frac{\sum_{i=1}^{5} x_i^2 n_i}{n} = \frac{(2{,}5)^2 \cdot 10 + (4{,}5)^2 \cdot 20 + (6{,}5)^2 \cdot 30 + (8{,}5)^2 \cdot 30 + (10{,}5)^2 \cdot 10}{100} = 50{,}05.$$

Тоді $D_{\text{В}} = \frac{\sum_{i=1}^{5} x_i^2 n_i}{n} - (\bar{x}_{\text{В}})^2 = 50{,}05 - (6{,}7)^2 = 50{,}05 - 44{,}89 = 5{,}16$.

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

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

Приклад 2. За заданим інтервальним статистичним розподілом вибірки, в якому наведено розподіл маси новонароджених приматів 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^*	1,1	1,3	1,5	1,7	1,9	2,1	2,3	2,5	2,7	2,9	3,1
n_i	5	12	18	22	36	24	19	15	11	9	2

Беручи до уваги (2.7), (2.8), (2.9) і те, що n = 173, дістанемо:

$$\bar{x}_{\text{В}} = \frac{\sum_{i=1}^{11} x_i^* n_i}{n} = \frac{5{,}5 + 15{,}6 + 27 + 37{,}4 + 68{,}4 + 50{,}4 + 43{,}7}{173} +$$

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

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

$$\frac{\sum_{i=1}^{11}(x_i^*)^2 n_i}{n} = \frac{6{,}05 + 20{,}29 + 40{,}5 + 63{,}58 + 129{,}96 + 105{,}84 + 100{,}51}{173} +$$

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

$$D_{\text{В}} = \frac{\sum_{i=1}^{11}(x_i^*)^2 n_i}{n} - (\bar{x}_B)^2 = 4{,}251908 - (2{,}008671)^2 =$$

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

$$D_{\text{В}} = 0{,}217149.$$

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

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

Приклад 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.

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

Розв'язання. Побудуємо дискретний статистичний розподіл вибірки:

x_i	211	215	216	217	218	219	220	221	222	223	224	225	226
n_i	1	1	1	1	3	3	4	4	2	1	1	1	1

Згідно з формулами (2.1), (2.3), (2.4), (2.5), (2.6) дістанемо:

$$\bar{x}_{\text{В}} = \frac{\sum_{i=1}^{13} x_i n_i}{n} = \frac{211 \cdot 1 + 215 \cdot 1 + 216 \cdot 1 + 217 \cdot 1 + 218 \cdot 3 + 219 \cdot 3 + 220 \cdot 4}{24} +$$

$$+ \frac{221 \cdot 4 + 222 \cdot 2 + 223 + 224 + 225 + 226}{24} \approx 219{,}7.$$

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

$$\frac{\sum_{i=1}^{13} x_i^2 n_i}{n} = \frac{211^2 + 215^2 + 216^2 + 217^2 + 218^2 \cdot 3 + 219^2 \cdot 3 + 220^2 \cdot 4}{24} +$$

$$+ \frac{221^2 \cdot 4 + 222^2 \cdot 2 + 223^2 + 224^2 + 225^2 + 226^2}{24} \approx 48336{,}83.$$

Тоді

$$D_{\text{В}} = \frac{\sum_{i=1}^{13} x_i^2 n_i}{n} - (\bar{x}_{\text{В}})^2 = 48336{,}83 - (219{,}7)^2 = 48336{,}83 - 48268{,}09 = 68{,}7433.$$

$D_{\text{В}}$ =68,7433. $\sigma_{\text{В}} = \sqrt{D_{\text{В}}} = \sqrt{68{,}7433} \approx 8{,}29.$ $\sigma_{\text{В}}$ =8,29.

Оскільки $n = 24 < 30$, знайдемо

$$s^2 = \frac{n}{n-1} D_B = \frac{24}{23} \cdot 68{,}7433 = 71{,}7321, \quad s = 8{,}42.$$

$$V = \frac{\sigma_{\text{В}}}{\bar{x}_{\text{В}}} 100\% = \frac{68{,}7433}{219{,}7} 100\% = 31{,}29\%.$$

Завдання для самостійної роботи

1. 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

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

2. Із партії однотипних сталевих болтів, виготовлених заводом, була здійснена вибірка обсягом 200 шт. 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

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

3. За допомогою радіодальноміра було здійснено 16 вимірювань однієї і тієї самої відстані. Результати вимірювання в метрах наведені у вигляді статистичного ряду:

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

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

Перевірити, чи правильне таке твердження:

Статистична дисперсія може дорівнювати виправленій статистичній дисперсії.

2.2. Основні закони розподілу та вибіркові оцінки параметрів розподілу

Розглянемо приклади випадкових величин, що найчастіше використовуються у дослідженнях:

1. Неперервна випадкова величина ***розподілена нормально***, якщо її диференціальна функція розподілу має вигляд $f(x)=\frac{1}{\sigma\cdot\sqrt{2\pi}}\cdot e^{\frac{-(x-a)^2}{2\cdot\sigma^2}}$. Числа a,σ називають параметрами розподілу, їхні вибіркові оцінки знаходять за формулами $a=\bar{x}_B,\quad \sigma_B=\sqrt{s^2}=s$.

2. Випадкова величина, що має сталу диференціальну функцію $f(x)=\frac{1}{b-a}$, де $a\le x\le b$, та $f(x)=0$, якщо $x<a$ і $x>b$, розподілена ***рівномірно*** і має параметри розподілу a,b. Їхні вибіркові оцінки обчислюються за формулами $\bar{a}=\bar{x}_B-\sqrt{3}\cdot\sigma_B,\quad \bar{b}=\bar{x}_B+\sqrt{3}\cdot\sigma_B$.

3. Випадкова величина, диференціальна функція якої $f(x)=\lambda\cdot e^{-\lambda x}$ при $x\ge 0$ та $f(x)=0$, якщо $x<0$, має ***експоненціальний (показниковий) розподіл.*** Параметр розподілу λ має оцінку $\bar{\lambda}=\frac{1}{\bar{x}_B}$.

4. ***Біноміальним*** *є* закон розподілу випадкової величини X, ймовірності якої обчислюються за формулами Бернуллі:

$$P_n(k)=\frac{n!}{k!(n-k)!}\cdot p^k\cdot q^{n-k},$$

де $q=1-p;\quad n,p$ – параметри. Параметр розподілу p має оцінку $p^*=\frac{\bar{x}_B}{m}$, де m – максимальне число появи події у випробуванні.

5. У випадку, коли p є дуже малим числом, а число n – достатньо великим, замість формули Бернуллі слід використовувати формулу $P_n(k)=\frac{\lambda^k}{k!}\cdot e^{-\lambda}$ (***розподіл Пуассона).*** Параметр розподілу $\lambda=np$ має оцінку $\bar{\lambda}=\bar{x}_B$.

Зразки розв'язування задач

Приклад 1. Дано статистичний розподіл, близький до розподілу Пуассона:

$X = x_i$	0	1	2	3
n_i	116	56	22	6

Знайти оцінки параметра λ.

Розв'язання. $n = 116+56+22+6 = 200;$

$$\bar{\lambda} = \overline{x_B} = \left(\sum_{i=1}^{4} n_i \cdot x_i\right) \Big/ n = \frac{116 \cdot 0 + 56 \cdot 56 + 22 \cdot 2 + 6 \cdot 3}{200} = 0{,}6.$$

Приклад 2. Дано статистичний розподіл, близький до показникового:

$X = x_i$	[0;1)	[1;2)	[2;3)	[3;4)	[4;5)	[5;6)
n_i	259	167	109	74	70	47
x_i^*	0,5	1,5	2,5	3,5	4,5	5,5

Знайти оцінки параметра λ.

Розв'язання. $n = 259+167+109+74+70+47=726;$

$$\bar{x}_B = \frac{\sum_{i=1}^{6} n_i \cdot x_i^*}{n} = \frac{259 \cdot 0{,}5 + 167 \cdot 1{,}5 + 109 \cdot 2{,}5 + 74 \cdot 3{,}5 + 70 \cdot 4{,}5 + 47 \cdot 5{,}5}{726} =$$

$$= \frac{129{,}5 + 250{,}5 + 272{,}5 + 259 + 258{,}5}{726} = \frac{1170}{726} = 1{,}61.$$

$$\bar{\lambda} = \frac{1}{\overline{x_B}} = \frac{1}{1{,}61} = 0{,}62.$$

Приклад 3. Дано статистичний розподіл, близький до біноміального:

$X = x_i$	0	1	2	3	4	5
n_i	72	77	34	14	2	1

знайти оцінки параметра p^*.

Розв'язання. $m = 5;$ $n = 72+77+34+14+2+1 = 200;$

$$\bar{x}_B = \frac{\sum_{i=1}^{6} n_i \cdot x_i}{n} = \frac{72 \cdot 0 + 77 \cdot 1 + 34 \cdot 2 + 14 \cdot 3 + 2 \cdot 4 + 1 \cdot 5}{200} = \frac{200}{200} = 1.$$

$$p^* = \frac{\bar{x}_B}{m} = \frac{1}{5} = 0{,}2.$$

Завдання для самостійної роботи

1. Дано статистичний розподіл, близький до біноміального розподілу:

$X = x_i$	0	1	2	3	4
n_i	9	23	43	23	9

Знайти оцінки параметра p^*.

2. Дано статистичний розподіл, близький до показникового розподілу:

I_i	[0; 400)	[400; 800)	[800; 1200)	[1200; 1600)
n_i	123	97	78	54

Знайти оцінки параметра λ.

3. Дано статистичний розподіл, близький до показникового розподілу:

I_i	[0; 5)	[5; 10)	[10; 15)	[15; 20)	[20; 25)	[25; 30)
n_i	133	48	16	6	4	3

Знайти оцінки параметра λ.

4. Дано статистичний розподіл, близький до біноміального розподілу:

$X = x_i$	0	1	2	3	4
n_i	10	22	44	24	10

Знайти оцінки параметра p^*.

5. Дано статистичний розподіл, близький до розподілу Пуассона:

$X = x_i$	0	1	2	3
n_i	117	57	23	7

Знайти оцінки параметра λ.

6. Дано статистичний розподіл, близький до розподілу Пуассона:

$X = x_i$	0	1	2	3
n_i	39	38	20	8

Знайти оцінки параметра λ.

2.3. Інтервальні оцінки статистичних характеристик та параметрів генеральної сукупності

Більш точними є інтервальні оцінки. ***Інтервальною*** називають оцінку, яка визначається двома кінцями інтервалу.

Нехай за даними вибірки знайдена статистична оцінка θ^* невідомого параметра θ, яке будемо вважати постійним числом.

Надійністю оцінки θ по θ^* називають ймовірність γ, з якою виконується $P\left(\left|\theta^* - \theta\right| < \delta\right) = \gamma$. Тут число δ називають ***точністю*** оцінки.

Зазвичай, надійність оцінки задають наперед і в якості значень γ беруть число, близьке до 1, наприклад, 0,95; 0,99; 0,999.

Побудова довірчого інтервалу для $\overline{X}_\Gamma$ при відомому значенні σ_Γ із заданою надійністю γ.

Нехай ознака X генеральної сукупності має нормальний закон розподілу. Побудуємо довірчий інтервал для середнього значення генеральної сукупності $\overline{X}_\Gamma$, знаючи числове значення середнього квадратичного відхилення генеральної сукупності σ_Γ, із заданою надійністю γ. Оскільки $\overline{x}_B$ як точкова незміщена статистична оцінка для $\overline{X}_\Gamma = M(x)$ має нормальний закон розподілу з числовими характеристиками $M(\overline{x}_B) = \overline{X}_\Gamma = a,\ \sigma(\overline{x}_B) = \frac{\sigma_\Gamma}{\sqrt{n}}$, дістанемо

$$P\left(\left|\overline{x}_B - a\right| < \delta\right) = \gamma. \qquad (2.10)$$

Отже, довірчий інтервал дорівнюватиме:

$$\overline{x}_B - \frac{x \cdot \sigma_\Gamma}{\sqrt{n}} < a < \overline{x}_B + \frac{x \cdot \sigma_\Gamma}{\sqrt{n}}, \qquad (2.11)$$

Аргумент x знаходимо за значеннями функції Лапласа (додаток 1). Величина $\frac{x \cdot \sigma_\Gamma}{\sqrt{n}}$ називається ***точністю оцінки*, або *похибкою вибірки*.**

Побудова довірчого інтервалу для $\overline{X}_\Gamma$ при невідомому значенні σ_Γ із заданою надійністю γ.

У випадку малих вибірок для оцінювання $\overline{X}_\Gamma = a$ при невідомому значенні σ_Γ неможливо скористатися нормальним законом розподілу. Обчисливши за даним статистичним розподілом $\overline{x}_B$, s і визначивши за таблицею розподілу Стьюдента значення t_γ, будуємо довірчий інтервал

$$\overline{x}_B - \frac{t_\gamma \cdot s}{\sqrt{n}} < a < \overline{x}_B + \frac{t_\gamma \cdot s}{\sqrt{n}}. \qquad (2.12)$$

Тут $t_\gamma(\gamma, k = n-1)$ обчислюємо за заданою надійністю γ і числом ступенів вільності $k = n-1$ за таблицею (додаток 2).

Побудова довірчих інтервалів із заданою надійністю γ для D_Γ, σ_Γ.

У разі, коли ознака X має нормальний закон розподілу, для побудови довірчого інтервалу із заданою надійністю γ для D_Γ, σ_Γ застосовуємо випадкову величину

$$\chi^2 = \frac{n-1}{\sigma_\Gamma^2} s^2, \tag{2.13}$$

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

Для нормального розподілу справедлива формула для границь довірчого інтервалу для дисперсії при довірчій ймовірності (надійності) γ:

$$\frac{n \cdot s^2}{\chi_2^2} < D_\Gamma < \frac{n \cdot s^2}{\chi_1^2}, \tag{2.14}$$

де χ_1^2 – значення випадкової величини, що має розподіл χ^2 (хі – квадрат Пірсона) з числом ступенів вільності $f_1 = n-1$ при рівні значущості $\alpha_1 = 1 - \frac{\alpha}{2}$, де $\alpha = 1 - \gamma$; χ_2^2 – значення випадкової величини, що має розподіл χ^2 з числом ступенів вільності $f_2 = n-1$ при рівні значущості $\alpha_2 = \frac{\alpha}{2}$.

Тоді довірчий інтервал для σ_Γ випливає із (2.14) і буде таким:

$$\frac{s\sqrt{n-1}}{\chi_2} < \sigma_\Gamma < \frac{s\sqrt{n-1}}{\chi_1}. \tag{2.15}$$

Значення χ_1^2, χ_2^2 знаходимо за таблицею (додаток 3) згідно з рівностями:

$$P\left(\chi^2 > \chi_1^2\right) = 1 - \frac{\alpha}{2}; \tag{2.16}$$

$$P\left(\chi^2 > \chi_2^2\right) = \frac{\alpha}{2}, \tag{2.17}$$

де $\alpha = 1 - \gamma$.

Зразки розв'язування задач

Приклад 1. 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: $\bar{x}_{\text{В}}$, σ_{Γ}, n, x.

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

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

$$\Phi(x) = 0{,}495 \to x = 2{,}58 \text{ (додаток 1).}$$

Знайдемо числові значення кінців довірчого інтервалу:

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

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

Таким чином, маємо: $14{,}88 < \bar{X}_{\Gamma} < 15{,}12$.

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

Приклад 2. 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=2млн. грн.; з надійністю $\gamma = 0{,}999$ знайти довірчий інтервал для $\bar{X}_{\Gamma}$, якщо σ_{Γ} =5млн. грн.

Розв'язання. Інтервальний статистичний розподіл буде таким:

I_i.	2 – 4	4 – 6	6 – 8	8 – 10
n_i	9	7	10	4

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

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

$$n = \sum_{i=1}^{4} n_i = 30.$$

Тоді

$$\bar{x}_{\text{В}} = \frac{\sum_{i=1}^{4} x_i^* n_i}{n} = \frac{3\cdot 9 + 5\cdot 7 + 7\cdot 10 + 9\cdot 4}{30} = \frac{27+35+70+36}{30} = \frac{168}{30} = 5{,}6 \text{ млн. грн.}$$

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:

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

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

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

Приклад 3. Випадково вибрана партія з двадцяти приладів була випробувана щодо терміну безвідмовної роботи кожного з них $x_i = t_i$. Результати випробувань наведено у вигляді дискретного статистичного розподілу:

x_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 безвідмовної роботи приладу).

Розв'язання. Для побудови довірчого інтервалу необхідно знайти середнє вибіркове і виправлене середнє квадратичне відхилення (n = 20 < 30).

Обчислимо $\bar{x}_{\text{В}}$:

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

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

Визначимо $D_{\text{В}}$:

$$\frac{\sum_{i=1}^{5} x_i^2 n_i}{n} = \frac{100^2\cdot 2 + 170^2\cdot 5 + 240^2\cdot 10 + 310^2\cdot 2 + 380^2\cdot 1}{20} = \frac{1\ 077\ 100}{20} = 53\,855.$$

$$D_{\text{В}} = \frac{\sum_{i=1}^{5} x_i^2 n_i}{n} - (\bar{x}_{\text{В}})^2 = 53\,855 - (222{,}5)^2 = 53\,855 - 49\,506{,}25 = 4348{,}75.$$

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 розподілу Стьюдента (додаток 2) за заданою надійністю $\gamma=0{,}99$ і числом ступенів вільності $k=n-1$ = 20 – 1 = 19 знаходимо значення $t_\gamma\left(\gamma=0{,}99,\,k=19\right)=2{,}861.$

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 $\bar{X}_{\Gamma}=a$ буде міститися в інтервалі

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

Приклад 4. З генеральної сукупності, яка має нормальний закон розподілу, вилучена вибірка: 0,4; 0,4; 1,0; 0,3; 0,6; 0,9; 0,5; 0,4; 0,8; 0,6; 0,5; 0,5; 0,4; 0,8; 0,3; 0,3; 0,6; 0,7; 0,7; 0,3; 0,2; 0,1; 0,4; 0,5; 0,7; 0,2; 0,6; 0,5; 0,4; 0,9.

Визначити незміщену оцінку дисперсії. Знайти інтервал надійності для дисперсії при довірчій ймовірності γ = **0,95.**

Розв'язання. Дискретний статистичний розподіл вибірки має вигляд:

x_i	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
n_i	1	2	4	6	5	4	3	2	2	1

Оцінка для математичного сподівання (вибіркова середня) $n=30$,

$$\bar{x}_{\text{В}}=\frac{\sum_{i=1}^{10}x_i\cdot n_i}{n}=\frac{0{,}1\cdot 1+0{,}2\cdot 2+0{,}3\cdot 4+0{,}4\cdot 6+0{,}5\cdot 5}{30}+$$

$$+\frac{0{,}6\cdot 4+0{,}7\cdot 3+0{,}8\cdot 2+0{,}9\cdot 2+1{,}0\cdot 1}{30}=\frac{15{,}4}{30}=0{,}51(3)\approx 0{,}5.$$

Визначимо $D_{\text{В}}$:

$$D_{\text{В}}=\frac{\sum_{i=1}^{10}x_i^2 n_i}{n}-\left(\bar{x}_{\text{В}}\right)^2=0{,}294-0{,}25=0{,}044;$$

виправлена дисперсія може бути розрахована за формулою

$$s^2 = \frac{n}{n-1} D_B = 0{,}045. \qquad (2.18)$$

Для нормального розподілу довірчий інтервал для дисперсії має вигляд:

$$\frac{n \cdot s^2}{\chi_2^2} < D_\Gamma < \frac{n \cdot s^2}{\chi_1^2},$$

де χ_1^2 – значення випадкової величини, що має розподіл χ^2 з числом ступенів вільності $f_1 = n - 1 = 29$ при рівні значущості $\alpha_1 = 1 - \alpha/2 = 0{,}975$, де $\alpha = 1 - \gamma = 0{,}05$; χ_2^2 – значення випадкової величини, що має розподіл χ^2 з числом ступенів вільності $f_2 = n - 1 = 29$ при рівні значущості $\alpha_2 = \alpha/2 = 0{,}025$.

За таблицею розподілу «хі-квадрат» (додатки 3, 4) маємо: $\chi_1^2 = 16{,}0$; $\chi_2^2 = 45{,}7$.

Підставляємо обчислені та знайдені за таблицею значення в формулу

$$\frac{30 \cdot 0{,}045}{45{,}7} < D_\Gamma < \frac{30 \cdot 0{,}045}{16{,}0}.$$

Таким чином довірчий інтервал буде таким: $0{,}030 < D_\Gamma < 0{,}085$.

Приклад 5. Перевірена партія однотипних телевізорів x_i на чутливість до відеопрограм n_i, дані перевірки наведено як дискретний статистичний розподіл:

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

З надійністю $\gamma = 0{,}99$ побудувати довірчі інтервали для D_Γ, σ_Γ.

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

Обчислимо значення $\bar{x}_B$:

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

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

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

Обчислимо D_B:

$$\frac{\sum x_i^2 n_i}{n} = \frac{(200)^2 \cdot 2 + (250)^2 \cdot 5 + (300)^2 \cdot 6 + (350)^2 \cdot 7 + (400)^2 \cdot 5 + (450)^2 \cdot 2 +}{30}$$

$$\frac{+ (500)^2 \cdot 2 + (550)^2 \cdot 1}{30} = \frac{80\,000 + 312\,500 + 540\,000 + 857\,500 + 800\,000 +}{30}$$

$$\frac{+ 405\,000 + 500\,000 + 302\,500}{30} = \frac{3797500}{30} = 126583{,}3.$$

$$D_B = \frac{\sum x_i^2 n_i}{n} - (\bar{x}_B)^2 = 126583{,}3 - (345)^2 = 126583{,}3 - 119025 = 7558{,}3.$$

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

Виправлена дисперсія і виправлене середнє квадратичне відхилення дорівнюватимуть:

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

$$s = \sqrt{\frac{n}{n-1} D_B} = \sqrt{7818{,}9} \approx 88{,}42 \text{ мкв.}$$

Оскільки $\alpha = 1 - \gamma = 1 - 0{,}99 = 0{,}01,$ то 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 3) знаходимо:

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

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

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

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

Завдання для самостійної роботи

1. 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 $\bar{X}_\Gamma$, якщо $\sigma_\Gamma = 4$.

Відповідь: $376 < \bar{X}_\Gamma < 380{,}27.$

2. 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: $2{,}31 < \overline{X}_{\Gamma} < 3{,}33$.

3. Задана генеральна сукупність, яка розподілена нормально. Вибірка, що зроблена випадковим способом, задається наведеними даними. Виконати наступні вправи:

a) визначити границі довірчого інтервалу для математичного сподівання при довірчій ймовірності (надійності) $\gamma = 0{,}95$;

b) визначити границі довірчого інтервалу для дисперсії при довірчій ймовірності (надійності) $\gamma = 0{,}95$.

3.1. Наведені дані про витрати часу на виробництво одиниці продукції: 0,09; 0,11; 0,09; 0,08; 0,09; 0,06; 0,10; 0,09; 0,11; 0,09; 0,09; 0,11; 0,09; 0,08; 0,09; 0,06; 0,08; 0,09; 0,08; 0,10; 0,08; 0,07; 0,09; 0,07; 0,10; 0,07; 0,09; 0,10; 0,06; 0,09.

3.2. Виміряна максимальна ємність конденсаторів змінної ємності. Результати вимірювань у пФ наведені у таблиці: 437; 4,39; 4,38; 4,39; 436; 4,42; 4,36; 4,31; 4,40; 4,40; 437; 4,39; 4,38; 4,39; 436; 4,42; 4,35; 4,59; 4,40; 4,43; 4,41; 4,42; 4,40; 4,37; 4,42; 4,40; 4,35; 4,40; 4,40;4,40.

3.3. При свердлінні 50 отворів одним і тим самим свердлом з наступними вимірюваннями діаметрів отворів отримані такі дані в мм: 40,32; 40,32; 40,33; 40,2; 40,29; 40,31; 40,35; 40,28; 40,29; 40,29; 40,35; 40,34; 40,33; 40,34; 40,31; 40,33; 40,34; 40,32; 40,30; 40,31; 40,30; 40,30; 40,33; 40,32; 40,33; 40,30; 40,32; 40,31; 40,32; 40,29.

Розділ 3
ДВОВИМІРНИЙ СТАТИСТИЧНИЙ РОЗПОДІЛ ВИБІРКИ

3.1. 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}	n/n

Тут n_{ij} — частота спільної появи варіант

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

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

Загальні числові характеристики ознаки X:

загальна середня величина ознаки X

$$\bar{x}_B = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} x_j n_{ij}}{n} = \frac{\sum_{j=1}^{m} x_j n_{x_j}}{n}; \tag{3.1}$$

загальна дисперсія ознаки X

$$D_x = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} x_j^2 n_{ij}}{n} - (\bar{x}_B)^2 = \frac{\sum_{j=1}^{m} x_j^2 n_{x_j}}{n} - (\bar{x}_B)^2; \tag{3.2}$$

загальне середнє квадратичне відхилення ознаки X

$$\sigma_x = \sqrt{D_x}. \qquad (3.3)$$

Загальні числові характеристики ознаки Y:

загальна середня величина ознаки Y

$$\bar{y}_B = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} y_i\, n_{ij}}{n} = \frac{\sum_{i=1}^{k} y_i\, n_{y_i}}{n}; \qquad (3.4)$$

загальна дисперсія ознаки Y

$$D_y = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} y_i^2\, n_{ij}}{n} - (\bar{y}_B)^2 = \frac{\sum_{i=1}^{k} y_i^2\, n_{y_i}}{n} - (\bar{y}_B)^2; \qquad (3.5)$$

загальне середнє квадратичне відхилення ознаки Y

$$\sigma_y = \sqrt{D_y}. \qquad (3.6)$$

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

$$\bar{y}_{X=x_j} = \frac{\sum_{i=1}^{k} y_i\, n_{ij}}{\sum_{i=1}^{k} n_{ij}} = \frac{\sum_{i=1}^{k} y_i\, n_{ij}}{n_{x_j}}; \qquad (3.7)$$

умовна дисперсія ознаки Y

$$D\left(Y / X = x_j\right) = \frac{\sum_{i=1}^{k} y_i^2\, n_{ij}}{n_{x_j}} - (\bar{y}_{X=x_j})^2; \qquad (3.8)$$

умовне середнє квадратичне відхилення ознаки Y

$$\sigma\left(Y / X = x_j\right) = \sqrt{D(Y / X = x_j)}. \qquad (3.9)$$

$D(Y / X = x_j)$, $\sigma(Y / X = x_j)$ 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

$$\bar{x}_{Y=y_j} = \frac{\sum_{j=1}^{m} x_i \, n_{ij}}{\sum_{j=1}^{m} n_{ij}} = \frac{\sum_{j=1}^{m} x_i \, n_{ij}}{n_{y_i}}; \qquad (3.10)$$

умовна дисперсія ознаки X

$$D\left(X / Y = y_i\right) = \frac{\sum_{j=1}^{m} x_i^2 \, n_{ij}}{n_{y_i}} - (\bar{x}_{Y=y_j})^2; \qquad (3.11)$$

умовне середнє квадратичне відхилення ознаки X

$$\sigma\left(\left(X / Y = y_i\right)\right) = \sqrt{D\left(\left(X / Y = y_i\right)\right)}. \qquad (3.12)$$

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

$$\bar{y}_B = \frac{\sum_{j=1}^{n} y_{x_j} \, n_{x_j}}{n}; \qquad (3.13)$$

$$\bar{x}_B = \frac{\sum_{i=1}^{m} x_{y_i} \, n_{y_i}}{n}. \qquad (3.14)$$

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:

$$K_{xy}^{*} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{m} y_i \; x_i \; n_{ij}}{n} - \bar{x}_B \cdot \bar{y}_B . \qquad (3.15)$$

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:

$$r_{\text{В}} = \frac{K_{xy}^{*}}{\sigma_x \sigma_y} . \qquad (3.16)$$

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

Зразки розв'язування задач

Приклад 1. 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	n/n

потрібно:

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

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

Розв'язання. 1) Щоб обчислити K_{xy}^{*}, $r_{\text{В}}$, визначимо $\bar{x}_B, \sigma_x, \bar{y}_B, \sigma_y$. Оскільки $n = \sum_{i=1}^{4} \sum_{j=1}^{4} n_{ij} = 100,$ то

$$\bar{x}_B = \frac{\sum_{j=1}^{4} x_j \; n_{x_j}}{n} = \frac{10 \cdot 30 + 20 \cdot 20 + 30 \cdot 30 + 40 \cdot 20}{100} = \frac{300 + 400 + 900 + 800}{100} = \frac{2400}{100} = 24.$$

$$\frac{\sum_{j=1}^{4} x_j^2\, n_{x_j}}{n} = \frac{(10)^2 \cdot 30 + (20)^2 \cdot 20 + (30)^2 \cdot 30 + (40)^2 \cdot 20}{100} =$$

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

$$D_x = \frac{\sum_{j=1}^{4} x_j^2\, n_{x_j}}{n} - (\bar{x}_B)^2 = 700 - (24)^2 = 700 - 576 = 124.$$

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

Отже, $\sigma_x = 11{,}14.$

$$\bar{y}_B = \frac{\sum_{j=1}^{4} y_i\, n_{y_i}}{n} = \frac{2 \cdot 10 + 4 \cdot 30 + 6 \cdot 20 + 8 \cdot 40}{100} = \frac{20 + 120 + 120 + 320}{100} = 5{,}8.$$

Отже, $\bar{y}_B = 5{,}8$.

$$\frac{\sum_{j=1}^{4} y_i^2\, n_{y_i}}{n} = \frac{(2)^2 \cdot 10 + (4)^2 \cdot 30 + (6)^2 \cdot 20 + (8)^2 \cdot 40}{100} = \frac{3800}{100} = 38.$$

$$D_y = \frac{\sum_{j=1}^{4} y_i^2\, n_{y_i}}{n} - (\bar{y}_B)^2 = 38 - (5{,}8)^2 = 38 - 33{,}64 = 4{,}36,$$

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

Для визначення K_{xy}^* обчислюють:

$$\sum_{i=1}^{4}\sum_{j=1}^{4} y_i x_j n_{ij} = 2 \cdot 10 \cdot 0 + 2 \cdot 20 \cdot 2 + 2 \cdot 30 \cdot 4 + 2 \cdot 40 \cdot 4 + 4 \cdot 10 \cdot 10 + 4 \cdot 20 \cdot 8 +$$
$$+ 4 \cdot 30 \cdot 6 + 4 \cdot 40 \cdot 6 + 6 \cdot 10 \cdot 5 + 6 \cdot 20 \cdot 10 + 6 \cdot 30 \cdot 5 + 6 \cdot 40 \cdot 0 + 8 \cdot 10 \cdot 15 +$$
$$+ 8 \cdot 20 \cdot 0 + 8 \cdot 30 \cdot 15 + 8 \cdot 40 \cdot 10 = 0 + 80 + 240 + 320 + 400 + 640 + 720 +$$
$$+ 960 + 300 + 1200 + 900 + 0 + 1200 + 0 + 3600 + 3200 = 13760.$$

Тоді

$$K_{xy}^* = \frac{\sum_{i=1}^{4}\sum_{j=1}^{4} y_i x_j n_{ij}}{n} - \bar{x}_B \cdot \bar{y}_B = \frac{13760}{100} - 24 \cdot 5{,}8 = 137{,}6 - 139{,}2 = -1{,}6.$$

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 щільність кореляційного зв'язку між ознаками X та Y є слабкою.

Умовний статистичний розподіл $Y / X = 30$ 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

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

Умовна дисперсія та середнє квадратичне відхилення

$$\frac{\sum_{i=1}^{4} y_i^2 \, n_{i3}}{\sum_{i=1}^{4} n_{i3}} = \frac{(2)^2 \cdot 4 + (4)^2 \cdot 6 + (6)^2 \cdot 5 + (8)^2 \cdot 15}{30} = \frac{1252}{30} = 41{,}73;$$

$$D(Y / X = 30) = \frac{\sum_{i=1}^{4} y_i^2 \, n_{i3}}{\sum_{i=1}^{4} n_{i3}} - \left(\bar{y}_{X=30}\right)^2 = 41{,}73 - 36{,}8449 \approx 4{,}89;$$

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

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

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

$$\bar{x}_{Y=4} = \frac{\sum_{j=1}^{4} x_i \, n_{2j}}{\sum_{j=1}^{4} n_{2j}} = \frac{10 \cdot 10 + 20 \cdot 8 + 30 \cdot 6 + 40 \cdot 6}{30} = \frac{680}{30} \approx 22{,}7.$$

Отже, $\bar{x}_{Y=4} \approx 22{,}7.$

Умовна дисперсія та середнє квадратичне відхилення

$$\frac{\sum_{j=1}^{4} x_i^2 \; n_{2j}}{\sum_{j=1}^{4} n_{2j}} = \frac{(10)^2 \cdot 10 + (20)^2 \cdot 8 + (30)^2 \cdot 6 + (40)^2 \cdot 6}{30} =$$

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

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

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

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

Завдання для самостійної роботи

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

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

2. 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.**

3.2. 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_n

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

$$\bar{x}_B = \frac{\sum_{i=1}^{n} x_i}{n}; \qquad (3.17)$$

дисперсія

$$D_x = \frac{\sum_{i=1}^{n} x_i^2}{n} - (\bar{x}_B)^2; \qquad (3.18)$$

середнє квадратичне відхилення

$$\sigma_x = \sqrt{D_x}\,. \qquad (3.19)$$

***Числові характеристики ознаки Y*:**

середня величина

$$\bar{y}_B = \frac{\sum_{i=1}^{n} y_i}{n}; \qquad (3.20)$$

дисперсія

$$D_y = \frac{\sum_{i=1}^{n} y_i^2}{n} - (\bar{y}_B)^2; \qquad (3.21)$$

середнє квадратичне відхилення

$$\sigma_y = \sqrt{D_y}; \qquad (3.22)$$

емпіричний кореляційний момент

$$K_{xy}^* = \frac{\sum_{i=1}^{n} y_i \; x_i}{n} - \bar{x}_B \cdot \bar{y}_B; \qquad (3.23)$$

вибірковий коефіцієнт кореляції

$$r_B = \frac{K_{xy}^*}{\sigma_x \sigma_y}. \qquad (3.24)$$

Зразки розв'язування задач

Приклад. Залежність кількості масла y_i, що його споживає певна особа за місяць, від її прибутку 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:

$$\bar{x}_B = \frac{\sum_{i=1}^{20} x_i}{n} = \frac{70+75+82+89+95+100+105+110+115+120+125+}{20}$$
$$\frac{+130+135+140+145+150+155+160+165+170}{20} = \frac{2436}{20} = 121{,}8.$$

Отже, $\bar{x}_B = 121{,}8$.

$$\frac{\sum_{i=1}^{20} x_i^2}{n} = \frac{4900+5625+6724+7921+9025+10000+11025+}{20}$$

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

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

$$D_x = \frac{\sum_{i=1}^{20} x_i^2}{n} - (\bar{x}_B)^2 = 15728{,}5 - (121{,}8)^2 = 15728{,}5 - 14835{,}24 = 893{,}26.$$

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

$$\bar{y}_B = \frac{\sum_{i=1}^{20} y_i}{n} = \frac{10{,}5+15{,}8+17{,}8+19{,}5+20{,}4+21{,}5+22{,}2+24{,}3+25{,}3+26{,}5+}{20}$$

$$\frac{+28{,}1+30{,}1+35{,}2+36{,}4+37+38{,}5+39{,}5+40{,}5+41+42{,}5}{20} = \frac{572{,}6}{20} = 28{,}63.$$

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

$$\frac{\sum_{i=1}^{20} y_i^2}{n} = \frac{110{,}25+249{,}64+316{,}84+380{,}25+416{,}16+462{,}25+492{,}84+590{,}49+}{20}$$

$$\frac{+640{,}09+702{,}25+789{,}61+906{,}01+1239{,}04+1324{,}96+1369+1482{,}25+}{20}$$

$$\frac{+1560{,}25+1640{,}25+1681+1806{,}25}{20} = \frac{18159{,}68}{20} = 907{,}98.$$

$$D_y = \frac{\sum_{i=1}^{20} y_i^2}{n} - (\bar{y}_B)^2 = 907{,}98 - 819{,}68 = 88{,}3.$$

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

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

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

$$\sum_{i=1}^{20}\sum_{i=1}^{20} y_i x_i = 10{,}5\cdot 70 + 15{,}8\cdot 75 + 17{,}8\cdot 82 + 19{,}5\cdot 89 + 20{,}4\cdot 95 +$$

$$+21{,}5\cdot 100 + 22{,}2\cdot 105 + 24{,}3\cdot 110 + 25{,}3\cdot 115 + 26{,}5\cdot 120 + 28{,}1\cdot 125 +$$

$$+30{,}1\cdot 130 + 35{,}2\cdot 135 + 36{,}4\cdot 140 + 37\cdot 145 + 38{,}5\cdot 150 + 39{,}5\cdot 155 +$$

$$+40{,}5\cdot 160 + 41\cdot 165 + 42{,}5\cdot 170 = 735 + 1185 + 1459{,}6 + 1735{,}5 + 1938 +$$

$$+2150 + 2331 + 2673 + 2909{,}5 + 3180 + 3512{,}5 + 3913 + 4752 + 5096 +$$

$$+5365 + 5775 + 6122{,}5 + 6480 + 6765 + 7225 = 75302{,}6.$$

$$K_{xy}^{*} = \frac{\sum_{i=1}^{20} y_i x_i}{n} - \bar{x}_B \cdot \bar{y}_B = \frac{75302,6}{20} - 121,8 \cdot 28,63 = 3765,13 - 3487,13 = 278.$$

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

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

$$r_{\text{В}} \approx 0,989.$$

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

Завдання для самостійної роботи

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

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

2. 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	85	80	75	80	70	65	60	55
x_i	180	230	240	250	300	320	330	340	350	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 4
СТАТИСТИЧНІ ГІПОТЕЗИ

4.1. Загальна інформація

Правильність вибору закону розподілу чи оцінки його параметрів для генеральної сукупності перевіряється за допомогою статистичних методів перевірки статистичних гіпотез.

Під ***статистичною гіпотезою*** будемо розуміти припущення відносно закону розподілу генеральної сукупності чи оцінки його параметрів. Можливі інші гіпотези: про рівність параметрів двох чи декількох розподілів, про незалежність вибірок тощо.

Статистичною називають гіпотезу про вигляд невідомого розподілу або про параметри невідомих розподілів. Наприклад, статистичними є гіпотези:

1) генеральна сукупність, розподілена за нормальним законом;

2) дисперсії двох нормальних розподілів, рівні між собою.

Нульовою (основною) називають запропоновану гіпотезу, яку ми будемо позначати через $\boldsymbol{H_0}$.

Альтернативною (конкуруючою) називають гіпотезу $\boldsymbol{H_\alpha}$, яка суперечить основній.

Розрізняють також гіпотези за кількістю припущень.

Простою називають гіпотезу, яка має одне припущення, інакше гіпотеза є складною.

***Наприклад*:** 1. У законі Пуассона $\lambda = 3$ – проста гіпотеза;

2. Якщо $\lambda > 3$, то це складна гіпотеза.

Висунута гіпотеза може бути ***правильною або неправильною***, тому виникає необхідність її перевірки. Оскільки перевірка проводиться статистичними методами, то її називають статистичною. При прийнятті рішень за допомогою гіпотез можуть статися помилки двох родів.

Помилка першого роду полягає в тому, що буде відкинута правильна гіпотеза, тобто гіпотеза $\boldsymbol{H_0}$ є правильною, але її відхиляють на основі її перевірки.

Помилка другого роду полягає в тому, що буде прийнята неправильна гіпотеза, тобто гіпотеза $\boldsymbol{H_0}$ приймається, але в дійсності вірна конкуруюча гіпотеза $\boldsymbol{H_\alpha}$.

Імовірність здійснити помилку першого роду позначимо через α і будемо називаємо її ***рівнем значущості***.

Число α задають малим і найчастіше використовують значення α, що дорівнюють 0,05; 0,01 і т. д. Якщо, наприклад, $\alpha = 0{,}01$, то це означає, що в

одному випадку із 100 є ризик допустити помилку першого роду (відкинути гіпотезу H_0).

Для перевірки гіпотез використовуємо спеціально підібрану величину, точне чи наближене значення якої відоме.

Статистичним критерієм називають випадкову величину K, яка є основою для перевірки нульової гіпотези.

Найбільш розповсюдженим критерієм перевірки вірогідності H_0 про закон розподілу ознаки генеральної сукупності є критерій узгодженості χ^2, який визначається за формулою $\chi^2 = \sum_{i=1}^{m}(n_i - n_i')^2 / n_i'$.

Тут m – число інтервалів (часткових), на які поділяється статистичний розподіл вибірки; n_i – частота ознаки в i-у інтервалі; n_i' — теоретичні частоти, підраховані за відповідними формулами закону розподілу ймовірностей, який припускається для ознаки генеральної сукупності.

Теоретичні частоти знаходяться за формулою $n_i' = n \cdot p_i$, де n – об'єм вибірки; p_i – для дискретної випадкової величини є ймовірність події $X = x_i$, для неперервної випадкової величини p_i є ймовірність того, що ознака X попаде в i-ий інтервал.

Наприклад, для гіпотези H_0, яка припускає, що ознака генеральної сукупності має нормальний закон розподілу, імовірність p_i може бути обчислена за формулою $p_i = \Phi(x_{i+1}) - \Phi(x_i)$, де $\Phi(x)$ – функція Лапласа.

Для перевірки правильності гіпотез, як уже згадувалося, вибирається статистичний критерій, який умовно позначається через K, де K – випадкова величина, закон розподілу якої відомий. Для різних гіпотез ці критерії є різними.

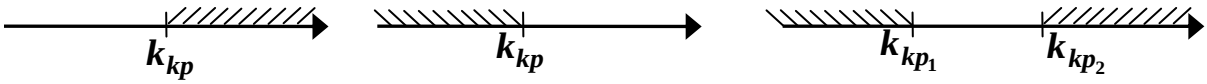
Множину R значень статистичного критерію K можна розбити на дві підмножини, що не перетинаються, A і $\overline{A}$.

Значення статистичного критерію підмножини $A \subset \Omega$, при яких нульова гіпотеза приймається, ***називається областю прийняття гіпотези***, а значення, при яких гіпотеза H_0 відхиляється, критичною областю.

За характером критичні області поділяються на односторонні та двосторонні.

Області A і $\overline{A}$ (прийняття гіпотез і критичні) між собою розділяються точками, які ми будемо називати критичними і позначати k_{kp}.

Правосторонньою критичною областю називається така область, для якої виконується нерівність $K > k_{kp}$ (рис. 4.1*а*).



а *б* *в*

Рис. 4.1

Відповідно критична область буде лівосторонньою, якщо виконується нерівність $K < k_{kp}$ (рис. 4.1*б*).

Двосторонньою будемо називати критичну область, яка задовольняє нерівності $K < k_{kp_1}$ і $K > k_{kp_2}$ (рис. 4.1*в*).

У більшості випадків для двосторонньої критичної області точки k_{kp_1} і k_{kp_2} розташовані симетрично по відношенню до нуля, тобто $k_{kp_1} = -k_{kp_2}$.

Перевірка статистичних гіпотез будь-якої природи може бути описана за допомогою такої загальної схеми.

1. Формулюється статистична гіпотеза H_0 і альтернативна H_α.

2. Вибирається статистичний критерій відповідно до сформульованої нульової гіпотези H_0.

3. Залежно від змісту нульової H_0 і альтернативної H_α гіпотез вибирається одностороння або двостороння критична область.

4. Для побудови критичної області необхідно знайти значення критичних точок.

В основі побудови критичної області покладено принцип практичної неможливості здійснитися малоймовірній випадковій події при одній спробі. За вибраним статистичним критерієм K та рівнем значущості α з допомогою спеціальних таблиць визначається критична точка k_{kp}. Згідно знайденого k_{kp} відповідно будується лівостороння, правостороння або двостороння критична область.

5. За результатами вибірки обчислюється спостережене значення критерію $K^*(K_{спост})$.

6. Приймається рішення прийняти чи відхилити нульову гіпотезу H_0 на підставі таких міркувань: якщо гіпотеза H_0 правильна, то $P(K^* \in \overline{A}) = \alpha$, тобто ймовірність того, що статистичний критерій потрапляє в критичну область $\overline{A}$, дорівнює малій ймовірності α. Якщо K^* потрапляє в критичну область $(K^* \in \overline{A})$, а ця подія малоймовірна і все ж відбулася, то в цьому разі H_0 відхилити.

7. Це твердження має наступний вигляд для різних типів областей: для лівосторонньої критичної області: $P(K^* < k_{kp}) = \alpha$, для правосторонньої

$P(K^* > k_{kp}) = \alpha$, для двосторонньої області $P(K^* < k_{kp_1}) + P(K^* > k_{kp_2}) = \alpha$.

8. Враховуючи ту обставину, що критичні точки k_{kp_1} і k_{kp_2} розташовані симетрично відносно нуля, двосторонню критичну область будують також симетричною, отже $P(K^* < k_{kp_1}) = P(K^* > k_{kp_2}) = \alpha / 2$.

9. Якщо K^* не потрапляє у критичну область, то нульова гіпотеза H_0 приймається.

4.2. 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 $k_1 = n_1 - 1$, $k_2 = n_2 - 1$ ступенями вільності, де n_1 і n_2 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 $K = s_\delta^2 / s_m^2$, яка має розподіл Фішера – Снедекора із k_1 і k_2 ступенями вільності, де s_δ^2 є більшою з виправлених дисперсій, одержаною внаслідок обробки результатів вибірок, s_m^2 є меншою з виправлених дисперсій.

Щільність ймовірностей 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 $0 \leq F < \infty$.

Спостережуване значення критерію

$$K^* = \frac{s_\delta^2}{s_m^2}. \tag{4.1}$$

Зразки розв'язування задач

Приклад 1. 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 вважати використання удосконаленого пристрою до стабілізатора температури ефективним, якщо рівень значущості $\alpha = 0{,}01$?

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:

$$\bar{y}_B = \frac{\sum y_i n_{1i}}{n_1} = \frac{21{,}2 + 21{,}4 + 21{,}0 + 21{,}3 \cdot 2 + 21{,}8}{6} = 21{,}333;$$

$$\frac{\sum y_i^2 n_{1i}}{n_1} = \frac{21{,}2^2 \cdot 1 + 21{,}4^2 \cdot 1 + 21{,}0^2 \cdot 1 + 21{,}3^2 \cdot 2 + 21{,}8^2 \cdot 1}{6} = \frac{2731{,}02}{6} = 455{,}17;$$

$$D_B = \frac{\sum y_i^2 n_{1i}}{n_1} - (\bar{y}_B)^2 = 455{,}17 - (21{,}333)^2 = 455{,}17 - 455{,}097 = 0{,}073;$$

$$s_y^2 = \frac{n_1}{n_1 - 1} D_B = \frac{6}{6-1} \cdot 0{,}073 = 0{,}0876;$$

$$\bar{x}_B = \frac{\sum x_j n_{2j}}{n_2} = \frac{37{,}7 + 37{,}6 \cdot 2 + 37{,}4}{4} = \frac{37{,}7 + 75{,}2 + 37{,}4}{4} = \frac{150{,}3}{4} = 37{,}575;$$

$$\frac{\sum x_j^2 n_{2j}}{n_2} = \frac{37{,}7^2 \cdot 1 + 37{,}6^2 \cdot 2 + 37{,}4^2 \cdot 1}{4} = \frac{5647{,}57}{4} = 1411{,}8925;$$

$$D_{\text{В}} = \frac{\sum x_j^2 n_{2j}}{n_2} - (\bar{x}_{\text{В}})^2 = 1411{,}8925 - (37{,}575)^2 =$$

$$= 1411{,}8925 - 1411{,}880625 = 0{,}011875;$$

$$s_x^2 = \frac{n_2}{n_2 - 1} D_{\text{В}} = \frac{4}{4-1} \cdot 0{,}011875 = 0{,}01583.$$

Обчислимо спостережуване значення критерію:

$$K^* = \frac{s_\delta^2}{s_m^2} = \frac{0{,}0876}{0{,}01583} = 5{,}534.$$

Число ступенів вільності для більшої виправленої дисперсії $s_\delta^2 = s_y^2$, $k_1 = n_1 - 1 = 5$, для меншої $s_m^2 = s_x^2$, $k_2 = n_2 - 1 = 3$.

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 6) відповідно до заданого рівня значущості α = 0,01 і числа ступенів вільності k_1 = 5, k_2 = 3, $k_{\text{кр}}(\alpha = 0{,}01;\ k_1 = 5;\ k_2 = 3) = 28{,}2$.

Висновок. Оскільки $K^* \in]0;\ 28{,}5]$, дані спостережень не дають підстав відхилити нульову гіпотезу, тобто вдосконалення термостабілізатора є ефективним.

Приклад 2. 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 α = 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:

$$\bar{y}_B = \frac{\sum y_i n_{1i}}{n_1} = \frac{1{,}2 \cdot 1 + 2{,}2 \cdot 2 + 3{,}2 \cdot 4 + 4{,}2 \cdot 2 + 5{,}2 \cdot 3}{12} =$$

$$= \frac{1{,}2 + 4{,}4 + 12{,}8 + 8{,}4 + 15{,}6}{12} = \frac{42{,}4}{12} \approx 3{,}53;$$

$$\frac{\sum y_i^2 n_{1i}}{n_1} = \frac{1{,}2^2 \cdot 1 + 2{,}2^2 \cdot 2 + 3{,}2^2 \cdot 4 + 4{,}2^2 \cdot 2 + 5{,}2^2 \cdot 3}{12} = \frac{168{,}48}{12} = 14{,}04;$$

$$D_{\text{В}} = \frac{\sum y_i^2 n_{1i}}{n_1} - (\bar{y}_B)^2 = 14{,}04 - (3{,}53)^2 = 14{,}04 - 12{,}4609 = 1{,}5791;$$

$$s_y^2 = \frac{n_1}{n_1 - 1} D_{\text{В}} = \frac{12}{12 - 1} \cdot 1{,}5191 = 1{,}723;$$

$$\bar{x}_B = \frac{\sum x_j n_{2j}}{n_2} = \frac{0{,}8 \cdot 2 + 1{,}6 \cdot 6 + 2{,}4 \cdot 1 + 3{,}2 \cdot 1 + 4 \cdot 2}{12} =$$

$$= \frac{1{,}6 + 9{,}6 + 2{,}4 + 3{,}2 + 8}{12} = \frac{24{,}8}{12} = 2{,}067;$$

$$\frac{\sum x_j^2 n_{2j}}{n_2} = \frac{0{,}8^2 \cdot 2 + 1{,}6^2 \cdot 6 + 2{,}4^2 \cdot 1 + 3{,}2^2 \cdot 1 + 4^2 \cdot 2}{12} = \frac{64{,}64}{12} = 5{,}39;$$

$$D_{\text{В}} = \frac{\sum x_j^2 n_{2j}}{n_2} - (\bar{x}_B)^2 = 5{,}39 - (2{,}067)^2 = 5{,}39 - 4{,}272489 = 1{,}1175;$$

$$s_x^2 = \frac{n_2}{n_2 - 1} D_{\text{В}} = \frac{12}{12 - 1} \cdot 1{,}1175 \approx 1{,}22.$$

Обчислимо спостережуване значення критерію

$$K^* = \frac{s_\delta^2}{s_m^2} = \frac{1{,}723}{1{,}22} = 1{,}41.$$

Для альтернативної гіпотези $H_\alpha : D_x > D_y$ будуємо правобічну критичну область. Знайдемо за таблицею (додаток 6) критичну точку

$$k_{\kappa\text{р}}(\alpha = 0{,}01,\ k_1 = 12 - 1 = 11,\ k_2 = 12 - 1 = 11) = k_{\kappa\text{р}}(0{,}01;\ k_1 = 11;\ k_2 = 11) = 4{,}4.$$

Висновок. Оскільки $K^* \in [0;\ 4{,}4]$, нульова гіпотеза $H_0 : D_x = D_y$ є правильною.

Завдання для самостійної роботи

1. 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: $K^* = s_\delta^2 / s_m^2 = 7{,}547$; $k_{кр}(\alpha = 0{,}001;\ k_1 = 15;\ k_2 = 19) = 5$; $K^* \overline{\in} [0; 5]$; $H_0 : D_x = D_y$ відхиляється.

2. 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: $K^* = s_\delta^2 / s_m^2 = 1{,}511$; $k_{кр}(\alpha = 0{,}001;\ k_1 = 39;\ k_2 = 39) = 2{,}2$; $K^* \in [0; 2{,}2]$; $H_0 : D_x = D_y$ не відхиляється.

4.3. Критерій перевірки гіпотези про вигляд невідомого закону розподілу

Критерієм згоди називають критерій перевірки гіпотези про вигляд невідомого закону розподілу

Є декілька критеріїв згоди: χ^2 («хі квадрат») Пірсона, Колмогорова, Смірнова і т.д. Для простоти обмежимося лише описом застосування критерію Пірсона для перевірки гіпотези про нормальний розподіл генеральної сукупності, оскільки інші закони перевіряються аналогічно.

Для перевірки критерію згоди за конкретними формулами порівнюємо емпіричні частоти (за даними вибірки) n_i і теоретичні – n'_i (обчислені в припущенні, що закон розподілу генеральної сукупності завдань, наприклад, у нашому випадку – нормальний).

В основі критерію згоди Пірсона покладена критеріальна статистика, що має вигляд:

$$\chi^2 = \sum_{i=1}^{l} (n_i - n'_i)^2 / n'_i \ . \qquad (4.2)$$

Чим менше відрізняються значення емпіричних і теоретичних частот, тим меншим буде значення $K_{спост}$ і, отже, більш точно характеризує близькість теоретичного і емпіричного розподілів.

Значення критичної точки для критерію згоди Пірсона залежить від рівня значущості α і числа ступенів вільності k.

Число ступенів вільності розподілу визначається за формулою $k = l - r - 1$, де l – число інтервалів статистичного ряду, r – число параметрів закону теоретичного розподілу, що оцінюється за даними вибірки (для нормального закону $r = 2$, оскільки цей закон виконується двома параметрами a і σ).

Зразки розв'язування задач

Приклад 1. При рівні значущості 0,01 перевірити гіпотезу про нормальний розподіл генеральної сукупності, якщо емпіричні і теоретичні частоти задаються наступною таблицею:

емпіричні частоти (n_i)	4	4	10	10	13	16
теоретичні частоти (n'_i)	6	8	11	12	10	10

Розв'язання. Складемо таблицю:

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$(n_i - n'_i)^2 / n'_i$
1	4	6	-2	4	0,67
2	4	8	-4	16	2
3	10	11	-1	1	0,09
4	10	12	-2	4	0,33
5	13	10	3	9	0,9
6	16	10	6	36	3,6
Сума	57	57			7,59

Отже, $\chi^2 = 7{,}59$.

Число ступенів вільності в нашому прикладі $\mathbf{6-3=3}$. За таблицею критичних точок розподілу $\chi^2 = \mathbf{7{,}59}$ (додаток 4) при значеннях $\boldsymbol{k=3}$ і $\boldsymbol{\alpha = 0{,}01}$ знаходимо $\chi^2_{kp} = \mathbf{11{,}3}$.

Оскільки $\chi^2_{спост} < \chi^2_{kp}$, а критична область правостороння, то немає підстави для відхилення нульової гіпотези. Іншими словами, розбіжність між теоретичними і емпіричними частотами незначна, тому дані спостережень узгоджуються з гіпотезою про нормальний закон розподілу генеральної сукупності.

Приклад 2. Вимірювання зросту юнаків віком 17 років дав такі результати:

I_i, 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 α = 0,01 перевірити правильність висунутої нульової гіпотези.

Розв'язання. Для заданого статистичного розподілу побудуємо гістограму частот (рис. 4.2).

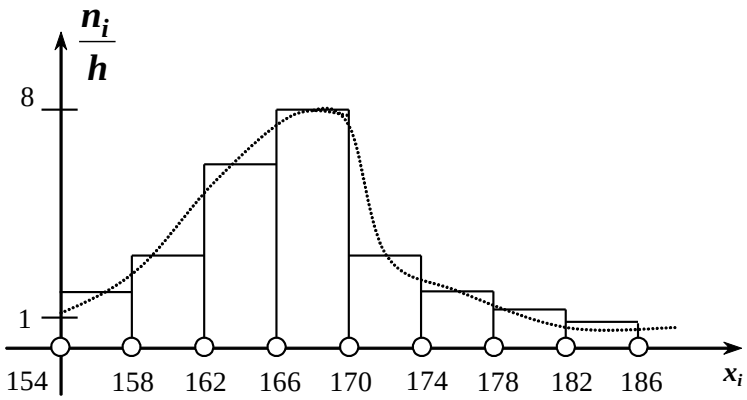


Рис. 4.2

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.

Необхідно обчислити теоретичні частоти, для цього знайдемо значення $\bar{x}_В$, $\sigma_В$, побудувавши дискретний розподіл за заданим інтервальним:

x_i	156	160	164	168	172	176	180	184
n_i	8	14	20	32	12	8	4	2

$$\bar{x}_{\text{В}} = \frac{\sum x_i\, n_i}{n} = \frac{156\cdot 8 + 160\cdot 14 + 164\cdot 20 + 168\cdot 32 +}{100}$$

$$\frac{+172\cdot 12 + 176\cdot 8 + 180\cdot 4 + 184\cdot 2}{100} = \frac{16704}{100} = 167{,}04 \text{ см};$$

$$\frac{\sum x_i^2\, n_i}{n} = \frac{156^2\cdot 8 + 160^2\cdot 14 + 164^2\cdot 20 + 168^2\cdot 32 + 172^2\cdot 12 + 176^2\cdot 8 +}{100}$$

$$\frac{+180^2\cdot 4 + 184^2\cdot 2}{100} = \frac{2794304}{100} = 27943{,}04;$$

$$D_{\text{В}} = \frac{\sum x_i^2\, n_i}{n} - (\bar{x}_{\text{В}})^2 = 27943{,}04 - (167{,}04)^2 =$$

$$= 27943{,}04 - 27902{,}3616 = 40{,}68;$$

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}} = \sqrt{40{,}68} \approx 6{,}38 \text{ см}.$$

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:

i	n_i	np_i	$n_i - np_i$	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
1	8	6	2	4	0,667
2	14	14	0	0	0
3	20	22	– 2	4	0,182
4	32	24	8	64	2,667
5	12	19	– 7	49	2,579
6	8	10	– 2	4	0,4
7	4	3	1	1	0,333
8	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 4) знаходимо значення

$$\chi^2_{\text{кр}}(\alpha = 0{,}01;\ k = 8 - 2 - 1) = \chi^2_{\text{кр}}(0{,}01; 5) = 15{,}1.$$

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 3. За заданим статистичним розподілом вибірки:

I_i, 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 α = 0,01 перевірити правильність цього припущення.

Розв'язання. Для визначення закону розподілу ознаки X побудуємо гістограму частот (рис. 4.3).

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 $\lambda = 1/\bar{x}_{\text{В}}$.

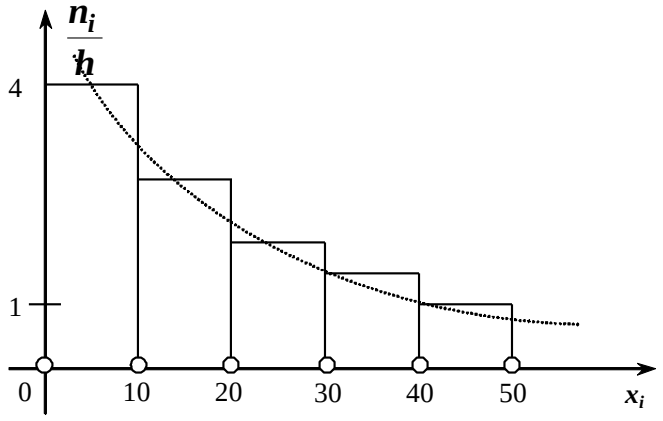


Рис. 4.3

Отже, необхідно обчислити $\bar{x}_{\text{В}}$, побудувавши дискретний статистичний розподіл за наведеним інтервальним, а саме:

x_i	5	15	25	35	45
n_i	40	30	20	6	4

Оскільки $n = \sum_{i=1}^{5} n_i = 100$, то

$$\bar{x}_{\text{В}} = \frac{\sum_{i=1}^{5} x_i\, n_i}{n} = \frac{5\cdot 40 + 15\cdot 30 + 25\cdot 20 + 35\cdot 6 + 45\cdot 4}{100} =$$

$$= \frac{200 + 450 + 500 + 910 + 180}{100} = 15{,}4. \quad \text{Тоді } \lambda = \frac{1}{\bar{x}_{\text{В}}} = \frac{1}{15{,}4} = 0{,}065.$$

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	$n_i - np_i$	$(n_i - np_i)^2$	$(n_i - np_i)^2 / np_i$
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

$$\chi^2_{\text{сп}} = \sum_{i=1}^{5} (n_i - np_i)^2 / np_i = 7{,}53.$$

За таблицею (додаток 4) знаходимо значення критичної точки

$$\chi^2_{\text{кр}}(\alpha = 0{,}01;\ k = 5 - 1 - 1 = 3) = \chi^2_{\text{кр}}(0{,}01;\ 3) = 11{,}3.$$

Висновок. Оскільки $\chi^2_{\text{сп}} \in [0;\ 11{,}3]$, нульова гіпотеза про експоненціальний закон розподілу ознаки X приймається.

Завдання для самостійної роботи

За заданими статистичними розподілами вибірки висунути H_0 про закон розподілу ознаки генеральної сукупності і при рівні значущості α = 0,01 перевірити її правильність:

1. 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	9,5–10,5	10,5–11,5	11,5–12,5	12,5–13,5
n_i	40	32	28	24	20	18	16	12	4

2. 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	0,44–0,46	0,46–0,48	0,48–0,5	0,5–0,52
n_i	10	16	24	40	32	20	16	5

4.4. Перевірка гіпотези про значущість вибіркового коефіцієнта кореляції

Нехай двомірна генеральна сукупність (X, Y) розподілена нормально. З цієї сукупності вилучена вибірка об'ємом n і обчислений вибірковий коефіцієнт кореляції r_B, який є відмінним від нуля. Так як вибірка обрана випадково, то ще неможливо стверджувати, що коефіцієнт кореляції генеральної сукупності r_Γ також відмінний від нуля. Оскільки нас цікавить саме цей коефіцієнт, виникає необхідність при заданому рівні значущості α перевірити нульову гіпотезу H_0: $r_\Gamma = 0$ – про рівність нулю генерального коефіцієнта кореляції при конкуруючій гіпотезі H_1: $r_\Gamma \neq 0$.

Якщо нульову гіпотезу буде відкинуто, то це означає, що вибірковий коефіцієнт кореляції значущо відрізняється від нуля (коротко кажучи, значно), а X та Y корельовані, тобто пов'язані лінійною залежністю.

Якщо нульову гіпотезу буде прийнято, то вибірковий коефіцієнт кореляції незначущий, а X та Y некорельовані, тобто не пов'язані лінійною залежністю.

Правило. Для того, щоб за заданим рівнем значущості α перевірити нульову гіпотезу H_0: $r_\Gamma = 0$, про рівність нулю генерального коефіцієнта кореляції нормальної двомірної випадкової величини при конкуруючій гіпотезі H_1: $r_\Gamma \neq 0$, необхідно обчислити спостережуване значення критерія:

$$T_{сп} = r_B \sqrt{n-2} \Big/ \sqrt{1 - r_B^2}\,.$$

Далі за таблицею критичних точок розподілу Стьюдента (додаток 5), за даним рівнем значущості і числом ступенів вільності $k = n - 2$ необхідно знайти критичну точку $t_{кр}(\alpha; k)$ для двосторонньої критичної області.

Якщо $|T_{сп}| < t_{кр}$ – немає сенсу відкидати нульову гіпотезу.

Якщо $|T_{сп}| > t_{кр}$ –нульову гіпотезу відкидаємо.

Приклад . За вибіркою об'ємом $n = 122$, вилученої із нормальної двовимірної сукупності, знайдено вибірковий коефіцієнт кореляції $r_B = 0{,}4$. За рівнем значущості $0{,}05$ перевіримо нульову гіпотезу про рівність нулю генерального коефіцієнта кореляції при альтернативній гіпотезі $H_1: r_\Gamma \neq 0$.

Розв'язання. Знайдемо спостережуване значення критерію:

$$T_{сп} = \frac{r_B\sqrt{n-2}}{\sqrt{1-r_B^2}} = \frac{0{,}4\sqrt{122-2}}{\sqrt{1-0{,}4^2}} = 4{,}78.$$

За умовою альтернативна гіпотеза має вигляд $r_\Gamma \neq 0$, тому критична область – двостороння.

За рівнем значущості $0{,}05$ і числом ступенів вільності $k = 122 - 2 = 120$ знаходимо за таблицею 5 для двосторонньої критичної області критичну точку $t_{кр}(0{,}05;120) = 1{,}98$.

Оскільки $T_{сп} > t_{кр}$ – нульову гіпотезу відкидаємо. Іншими словами, вибірковий коефіцієнт кореляції значно відрізняється від нуля, тобто X та Y корельовані.

Завдання для самостійної роботи

1. За вибіркою об'ємом $n = 62$, вилученої із нормальної двовимірної генеральної сукупності (X, Y), знайдено вибірковий коефіцієнт кореляції $r_B = 0{,}3$. Необхідно за рівнем значущості $0{,}01$ перевірити нульову гіпотезу про рівність нулю генерального коефіцієнта кореляції при альтернативній гіпотезі $H_1: r_\Gamma \neq 0$.

Відповідь: $k = 60$; $T_{сп} = 2{,}43$; $t_{кр}(0{,}01;60) = 2{,}66$. Немає підстав відкидати нульову гіпотезу; X та Y некорельовані випадкові величини.

2. За вибіркою об'ємом $n = 120$, вилученої із нормальної двовимірної генеральної сукупності (X, Y), знайдено вибірковий коефіцієнт кореляції $r_B = 0{,}4$. За рівнем значущості $0{,}05$ перевірити нульову гіпотезу про рівність нулю генерального коефіцієнта кореляції при конкуруючій гіпотезі $H_1: r_\Gamma \neq 0$.

Відповідь: $k = 118$; $T_{спост} = 4{,}74$; $t_{кр}(0{,}05;118) = 1{,}66$. Нульова гіпотеза відкидається; X та Y – корельовані випадкові величини.

Розділ 5
ЕЛЕМЕНТИ КОРЕЛЯЦІЙНОГО ТА РЕГРЕСІЙНОГО АНАЛІЗУ

5.1. Загальна інформація

При сумісній появі двох і більше величин у результаті проведення експерименту дослідник має підстави для встановлення певної залежності між ними, зв'язку.

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 і яка в математичній статистиці трансформується, не змінюючи своєї сутності, у статистичну залежність.

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.

За наявності кореляційного зв'язку між змінними необхідно виявити його форму функціональної залежності (лінійна чи нелінійна), а саме:

$$y = \beta_0 + \beta_1 x; \quad y = \beta_0 + \beta_1 x + \beta_2 x^2; \quad y = \beta_0 + \frac{\beta_1}{x}.$$

Наведені можливі залежності між змінними X і Y називають функціями регресії. Форму зв'язку між змінними X і Y можна встановити, застосовуючи кореляційні поля, які зображені на рисунках 5.1–5.3.

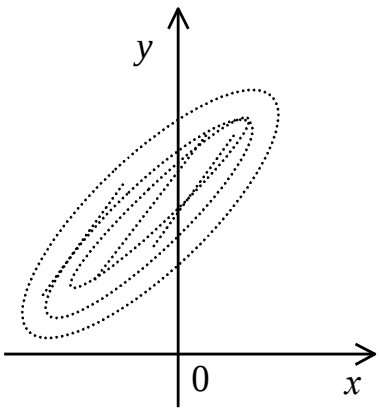


Рис. 5.1

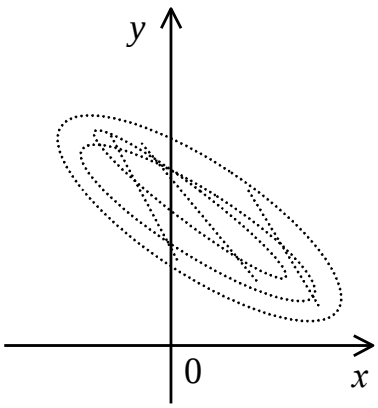


Рис. 5.2

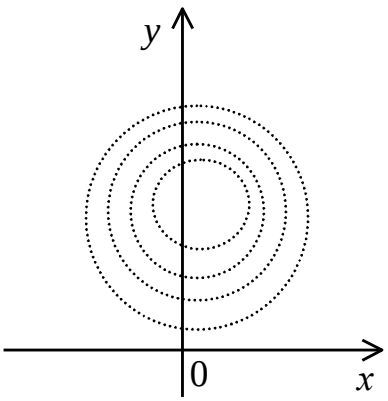


Рис. 5.3

Тут кожній точці з координатами x_i, y_i відповідає певне числове значення ознак X та Y.

Коли зв'язок між ознаками лінійний, використовують лінійний коефіцієнт кореляції. Його розраховують за формулою:

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2} \cdot \sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}} = \beta_1 \frac{\sigma_x}{\sigma_y} .$$

Він приймає значення від –1 до 1 і характеризує не тільки щільність зв'язку, а і його напрям. Чим ближче $|r|$ до 1, тим тісніше лінійний кореляційний зв'язок. Додатне значення означає прямий зв'язок між ознаками, а від'ємне – зворотний. Щільність зв'язку приблизно можна оцінити так:

– для $0{,}1 \le |r| \le 0{,}3$ – слабка,
– для $0{,}3 \le |r| \le 0{,}7$ – середня,
– для $0{,}7 \le |r| \le 0{,}99$ – цільна.

На рис. 5.1 більшість точок утворюють множину, що має тенденцію при збільшенні значень X зумовлювати збільшення значень ознаки Y (коефіцієнт кореляції більше 0). На рис. 5.2 множина точок має тенденцію при збільшенні значень X зумовлювати зменшення Y (коефіцієнт кореляції менше 0). На рис. 5.3 точки рівномірно розміщені на координатній площині xOy, що свідчить про відсутність кореляційної залежності між ознаками X і Y.

Отже, на основі розміщення точок 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 Y від X 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 $\boldsymbol{X}$ від $\boldsymbol{Y}$ називається функціональна залежність умовного середнього $\bar{x}_{y_i}$ від аргументу $\boldsymbol{y}$, що можна записати так:

$$\bar{x}_y = \beta(y).$$

Між ознаками $\boldsymbol{X}$ та $\boldsymbol{Y}$ може існувати статистична залежність і за відсутності кореляційної. Але коли існує кореляційна залежність між ознаками $\boldsymbol{X}$ та $\boldsymbol{Y}$, то обов'язково між ними існуватиме і статистична залежність.

Кореляційно - регресійний аналіз складається з таких етапів:

- вибір форми регресії;
- визначення параметрів рівняння;
- оцінка тісноти зв'язку;
- перевірка істотності зв'язку.

При виборі функції використовують графіки, аналітичні групування, теоретичне обґрунтування. Можливий перебір функцій, коли обчислюють різні рівняння регресії та обирають найкраще.

Визначення параметрів рівняння регресії проводиться методом найменших квадратів, основою якого є мінімізація суми квадратів відхилень емпіричних значень від теоретичних:

$$\sum_{i=1}^{n} (y_i - \overline{Y_i})^2 \to \min.$$

Найбільш поширена у статистичному аналізі лінійна функція.

5. 2. Рівняння лінійної парної регресії

Нехай між змінними $\boldsymbol{X}$ та $\boldsymbol{Y}$ 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. 5.4). Як бачимо, насправді між ознаками $\boldsymbol{X}$ і $\boldsymbol{Y}$ спостерігається не такий тісний зв'язок, як це передбачає функціональна залежність.

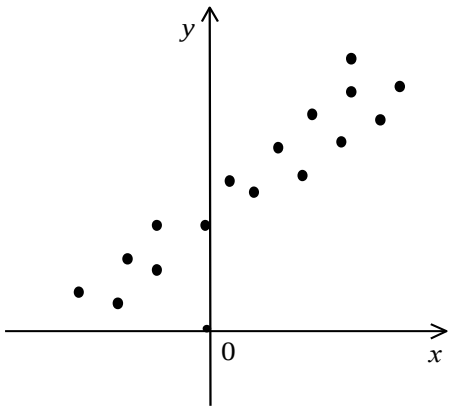


Рис. 5.4

Рівняння лінійної парної регресії набере такого вигляду:

$$y_i = r_{xy}\frac{\sigma_y}{\sigma_x}\left(x - \bar{x}\right) + \bar{y} \qquad (5.1)$$

або

$$y_i = \rho_{yx}\left(x - \bar{x}\right) + \bar{y}, \qquad (5.2)$$

де $\rho_{yx} = r_{xy}\frac{\sigma_y}{\sigma_x}$ - ***коефіцієнт регресії,*** r_{xy} —парний коефіцієнт кореляції між ознаками X і Y.

Показники кореляції часто розраховуються за даними вибірки, тому одержані показники не є точними оцінками показників кореляції в генеральній сукупності. Необхідно визначити точність оцінки показників кореляції та перевірити їх істотність.

При ***великому обсязі вибірки*** з нормально розподіленої сукупності, якщо величина лінійного коефіцієнта кореляції перевищує величину середньої квадратичної похибки $\sigma_r = \frac{1}{\sqrt{n-1}}$ більш ніж у t_α разів, а нульова гіпотеза H_0 – лінійний коефіцієнт кореляції дорівнює нулю, то вибірковий коефіцієнт кореляції можна вважати істотним (нульова гіпотеза не підтверджується). Тут α - рівень значущості, його найчастіше обирають рівним 0,05 або 0,01; r - значення лінійного коефіцієнта кореляції за даними вибірки; n – обсяг вибірки. Якщо ж відношення $\frac{|r|}{\sigma_r}$ виявиться меншим за t_α, то з імовірністю $(1-\alpha)$ треба вважати кореляційний зв'язок у генеральній сукупності відсутнім. Значення t_α в цьому випадку для γ = 0,95 дорівнює 1,96, а для γ = 0,99 дорівнює 2,58, тобто з імовірністю 0,99 можна стверджувати, що коефіцієнт кореляції генеральної сукупності знаходиться в інтервалі $\left[r - 2{,}58\frac{1-r^2}{\sqrt{n-1}}; r + 2{,}58\frac{1-r^2}{\sqrt{n-1}}\right]$.

Для **малого обсягу вибірки** розрахункове значення параметра порівнюють з табличним значенням t_α за таблицею закону розподілу Стьюдента з $(n - 2)$ ступенями вільності. Якщо розрахункове значення $t_{розр}$ перевищує табличне, то з імовірністю $(1 - \alpha)$ коефіцієнт кореляції можна вважати істотним, тобто гіпотезу про відсутність кореляційного зв'язку між випадковими величинами (X, Y) слід відкинути і прийняти альтернативну гіпотезу про наявність залежності між цими випадковими величинами.

Лінійна регресійна модель називається ***адекватною***, якщо обчислені за рівнянням регресії значення $\bar{Y}$ погоджуються з результатами спостережень.

Перевірку істотності зв'язку здійснюють за допомогою F – критерію Фішера: $F_r = \frac{r^2}{1-r^2} \cdot \frac{n-m}{m-1}$, де m – число параметрів рівняння регресії.

За рівнем значущості α, числами $k_1 = m-1 = 2-1 = 1$ та $k_2 = n-m = n-2$ за таблицями F – розподілу Фішера (додаток 6) знаходять критичне значення $F_{табл}(\alpha, k_1, k_2)$ і, якщо $F > F(\alpha, k_1, k_2)$, нульову гіпотезу спростовують, зв'язок вважають істотним, а рівняння – значущим.

Інколи використовують спрощений метод перевірки значущості коефіцієнта кореляції: якщо $|r_{xy}| \cdot \sqrt{n-1} \geq 3$, то зв'язок між випадковими величинами X та Y досить ймовірний

Зразки розв'язування задач

Приклад 1. 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) обчислити r_{xy} ;

3) побудувати графік лінії регресії.

Розв'язання. 1) Кореляційне поле залежності ознаки Y від X має такий вигляд (рис. 5.5).

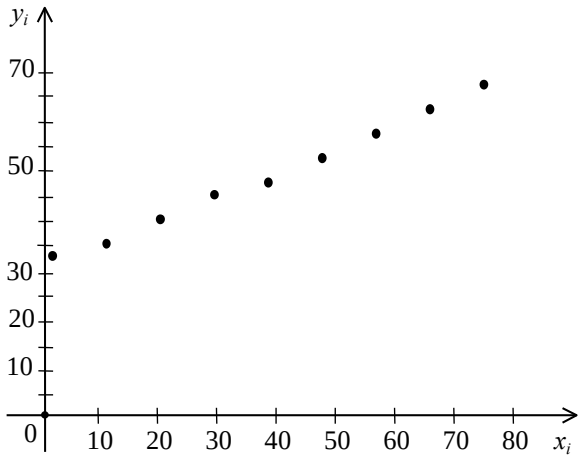


Рис. 5.5

Очевидно (рис. 5.5), що із збільшенням значень ознаки $X = x_i$ залежна зміна $Y = y_i$ має тенденцію до збільшення.

2) Для обчислення r_{xy} складемо таблицю:

№ 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

Оскільки $n = 9$, $\bar{x}_B = \dfrac{\sum_{i=1}^{9} x_i}{n} = \dfrac{360}{9} = 40$; $\bar{y}_B = \dfrac{\sum_{i=1}^{9} y_i}{n} = \dfrac{451,7}{9} = 50,19$;

$$\frac{\sum_{i=1}^{9} y_i^2}{n} = \frac{23859,05}{9} = 2651; \quad \frac{\sum_{i=1}^{9} x_i^2}{n} = \frac{20400}{9} = 2266,7;$$

$$\frac{\sum_{i=1}^{9} x_i y_i}{n} = \frac{20723}{9} = 2302,6; \quad \bar{x}_B \bar{y}_B = 40 \cdot 50,19 = 2007,6; \quad (\bar{x}_B)^2 = 1600.$$

Для обчислення r_{xy} необхідно знайти K_{xy}, σ_x, σ_y.

$$K_{xy}^* = \frac{\sum_{i=1}^{9} x_i y_i}{n} - \bar{x}_B \bar{y}_B = 2302,6 - 2007,6 = 295;$$

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{9} x_i^2}{n} - (\bar{x}_B)^2} = \sqrt{2266,7 - 40^2} = \sqrt{666,7} = 25,8;$$

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^{9} y_i^2}{n} - (\bar{y}_B)^2} = \sqrt{2651 - (50,19)^2} = \sqrt{131,96} = 11,49;$$

$$r_{xy} = \frac{K_{xy}^*}{\sigma_x \sigma_y} = \frac{295}{25,8 \cdot 11,49} = \frac{295}{296,44} = 0,995.$$

Отже, рівняння регресії буде таким:

$$y_i = 0,995 \cdot \frac{11,49}{25,8}(x_i - 40) + 50,19 \qquad y_i = 32,59 + 0,44 \cdot x_i.$$

Як бачимо, коефіцієнт кореляції близький за своїм значенням до одиниці, що свідчить про те, що залежність між X та Y є практично лінійною.

Перевіримо гіпотезу про значущість коефіцієнта регресії, яка складається у виконанні ознаки Стьюдента $|t| \geq t_{\alpha,k}$, де значення $t_{\alpha,k}$ знаходять з таблиці розподілу Стьюдента (додаток 5).

Обчислимо t за формулою: $t = r_{xy} \cdot \sqrt{\frac{n-2}{1-r_{xy}^2}} = 0{,}995 \cdot \sqrt{\frac{9-2}{1-0{,}995^2}} = 26{,}49$.

При заданій довірчій ймовірності $\gamma = 0{,}95$ і $k=7$ $t_{0,05;7} = 2{,}36$. Отже, нерівність виконана, оскільки $t = 26{,}49 > 2{,}36$. Лінійна регресія описує нашу залежність, форма зв'язку між ознаками X і Y є лінійною.

3) Графік парної лінійної функції регресії подано на рис. 5.6.

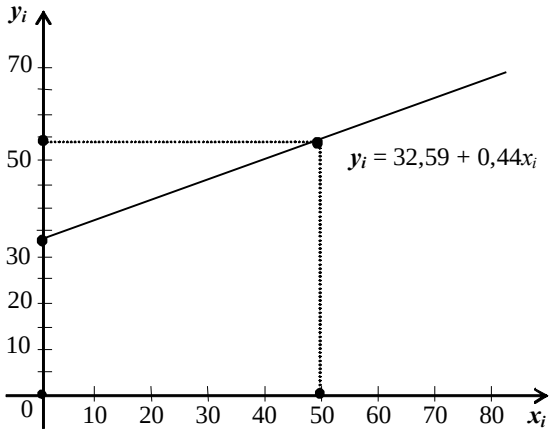


Рис. 5.6

Приклад 2. Кореляційна таблиця має вигляд:

X \ Y	4	9	14	19	24	29
130	3	3	-	-	-	-
140	-	5	4	-	-	-
150	-	-	40	2	8	-
160	-	-	5	10	6	-
170	-	-	-	4	7	3

Знайти рівняння регресії Y на X та рівняння регресії X на Y.

Розв'язання. 1. Знайдемо суми частот n_x за рядками і n_y за стовпцями.

X \ Y	4	9	14	19	24	29	n_x
130	3	3	-	-	-	-	6
140	-	5	4	-	-	-	9
150	-	-	40	2	8	-	50
160	-	-	5	10	6	-	21
170	-	-	-	4	7	3	14
n_y	3	8	49	16	21	3	100

Обчислимо середні $\bar{x}_B$ і $\bar{y}_B$:

$$\bar{x}_B = m_x = \sum_{i=1}^{5} x_i \frac{n_{x_i}}{n} =$$

$$= 130 \cdot \frac{6}{100} + 140 \cdot \frac{9}{100} + 150 \cdot \frac{50}{100} + 160 \cdot \frac{21}{100} + 170 \cdot \frac{14}{100} = 152{,}8.$$

$$\bar{y}_B = m_y = \sum_{i=1}^{5} y_i \frac{n_{y_i}}{n} =$$

$$= 4 \cdot \frac{3}{100} + 9 \cdot \frac{8}{100} + 14 \cdot \frac{49}{100} + 19 \cdot \frac{16}{100} + 24 \cdot \frac{21}{100} + 29 \cdot \frac{3}{100} = 16{,}65.$$

2. Обчислимо дисперсії:

$$\sigma_x^2 = \sum_{i=1}^{5} x_i^2 \frac{n_{x_i}}{n} - (\bar{x}_B)^2$$

$$= 130^2 \cdot \frac{6}{100} + 140^2 \cdot \frac{9}{100} + 150^2 \cdot \frac{50}{100} + 160^2 \cdot \frac{21}{100} + 170^2 \cdot \frac{14}{100} - 152{,}8^2 = 102{,}16.$$

$$\sigma_y^2 = \sum_{i=1}^{6} y_i^2 \frac{n_{y_i}}{n} - (\bar{y}_B)^2$$

$$= 4^2 \cdot \frac{3}{100} + 9^2 \cdot \frac{8}{100} + 14^2 \cdot \frac{49}{100} + 19^2 \cdot \frac{16}{100} +$$

$$+ 24^2 \cdot \frac{21}{100} + 29^2 \cdot \frac{3}{100} - 16{,}65^2 = 29{,}7275.$$

3. Щоб обчислити коефіцієнт кореляції r_{xy}, спочатку знайдемо момент кореляції:

$$K_{xy} = \sum_{i=1}^{5} \sum_{j=1}^{6} \frac{x_i y_j n_{ij}}{n} - \bar{x}_B \bar{y}_B =$$

$$= (130 \cdot (4 \cdot 3 + 9 \cdot 3) + 140 \cdot (9 \cdot 5 + 14 \cdot 4) + 150 \cdot (14 \cdot 40 + 19 \cdot 2 + 24 \cdot 8) +$$

$$+ 160 \cdot (5 \cdot 14 + 19 \cdot 10 + 24 \cdot 6) + 170 \cdot (19 \cdot 4 + 24 \cdot 7 + 29 \cdot 3)) \cdot \frac{1}{100} - 16{,}65 \cdot 152{,}8 = 42{,}08.$$

Тоді коефіцієнт кореляції:

$$r_{xy} = \frac{K_{xy}}{\sqrt{\sigma_x^2}\sqrt{\sigma_y^2}} = \frac{42{,}08}{\sqrt{29{,}7275}\sqrt{102{,}16}} = 0{,}764.$$

4. Рівняння регресії **Y** на **X**:

$$\overline{Y}_x - \bar{y}_B = r_{xy} \frac{\sigma_y}{\sigma_x}(x - \bar{x}_B), \ \overline{Y}_x - 16{,}65 = 0{,}764 \frac{5{,}45}{10{,}11}(x - 152{,}8),$$

$$\overline{Y}_x = 0{,}412\, x - 46{,}281,$$

та рівняння регресії ***X*** на ***Y***:

$$\overline{X}_x - \overline{x}_B = r_{xy}\frac{\sigma_x}{\sigma_y}(y - \overline{y}_B),\ \overline{X}_y = 1{,}417\,y + 129{,}203.$$

Перевіримо гіпотезу про значущість коефіцієнта регресії за однією з двох ознак:

а) повинна виконуватися нерівність:

$$\left|r_{xy}\right|\sqrt{n-1} \geq 3, \quad 0{,}764\sqrt{99} = 7{,}6 > 3,$$

отже, лінійна регресія описує задану залежність.

б) повинна виконуватися ознака Стьюдента:

$$|t| \geq t_{\alpha,k}, \quad t = r_{xy} \cdot \sqrt{\frac{n-2}{1-r_{xy}^2}},$$

де значення $t_{\alpha,k}$ знаходять з таблиці розподілу Стьюдента (число ступенів вільності $k = n-2$, рівень значущості $\alpha = 1-\gamma$, де γ – довірча ймовірність).

У нашому випадку $t = 11{,}72$. При заданій довірчій ймовірності $\gamma = 0{,}95$, $k = 98$, $t_{0,05;98} = 1{,}99$ виходить, що нерівність виконана, тому що $t = 11{,}72 > 1{,}99$.

Лінійна регресія описує задану залежність.

Завдання для самостійної роботи

В задачах 1–3 виконати наступні вправи:

1) побудувати кореляційне поле залежності ознаки Y від X;

2) обчислити r_{xy} ;

3) побудувати графік лінії регресії.

1. 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

2. 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
$X = x_i$, %	2,20	2,35	2,42	2,58	2,65	2,69	2,74	2,88	2,91
$Y = y_i$, %	39,1	40,5	42,4	43,8	45,6	46,9	48,5	49,4	50,0
$X = x_i$, %	2,95	2,99	3,00	3,11	3,21	3,29	3,34	3,44	3,50

3. Результати порівняння нового методу газового аналізу ***Y*** 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

.

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

4. Дана кореляційна таблиця. Написати рівняння регресії ***Y*** на ***X***, та рівняння регресії ***X*** на ***Y***.

X \ ***Y***	4	9	14	19	24	29
30	3	3	-	-	-	-
40	-	5	4	-	-	-
50	-	-	40	2	8	-
60	-	-	5	10	6	-
70	-	-	-	4	7	3

5.3. Множинна лінійна регресія

У багатьох практичних задачах залежна змінна y_i пов'язана з впливом не одного, а декількох аргументів. У цьому випадку регресію називають множинною. Якщо аргументи функції регресії мають перший степінь, то множинна регресія називається ***лінійною***, у противному разі — ***множинною нелінійною регресією.***

Лінійна множинна регресія. Лінійне рівняння регресії для ***n*** факторів має такий вигляд:

$$y = a_0 + a_1 x_1 + \ldots + a_n x_n + \varepsilon, \qquad (5.3)$$

де ε – випадкова величина, яка характеризує похибку між даними спостережень і відповідними значеннями, одержаними за допомогою теоретичних розрахунків за формулою (5.3), $a_0, a_1, \ldots, a_n$ – параметри, які знаходяться, як і у випадку парної кореляції, методом найменших квадратів.

Як і раніше , y – результуюча змінна; $x_1, \ldots, x_n$ – факторні змінні; a_0 – вільний член рівняння регресії; $a_1, \ldots, a_n$ – коефіцієнти регресії.

Будуємо квадратичну функцію:

$$F(a_0, a_1, \ldots, a_n) = \sum_{i=1}^{n} \left[y_i - \left(a_0 + a_1 x_{n1} + \ldots + a_n x_{ni} \right) \right]^2 . \qquad (5.4)$$

Ця функція набуває мінімальних значень лише тоді, коли всі частинні похідні за параметрами $a_0, a_1, \ldots, a_n$ дорівнюють нулю. Після нескладних перетворень система рівнянь $\frac{\partial F}{\partial a_i} = 0, \quad (i = 0, 1, \ldots, n)$ набуває такого вигляду:

$$\begin{cases} a_0 + a_1 \sum x_{1i} + \ldots + a_n \sum x_{ni} = \sum y_i, \\ a_0 \sum x_{1i} + a_1 \sum x_{1i}^2 + \ldots + a_n \sum x_{ni} \cdot x_{1i} = \sum y_i \cdot x_{1i}, \\ \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_0 \sum x_{ni} + a_1 \sum x_{ni} \cdot x_{1i} + \ldots + a_n \sum x_{ni}^2 = \sum y_i \cdot x_{ni} \end{cases} \tag{5.5}$$

Систему рівнянь (5.5) називають ***нормальною.***

Розв’язуючи систему рівнянь (5.5) відносно $a_0, a_1, \ldots, a_n$, одержуємо рівняння прямої лінії.

На практиці часто виникає потреба у знаходженні зв’язку між результуючою змінною і деякою фіксованою факторною змінною x_i, якщо решта факторів є сталими величинами. У загальному випадку цю задачу розв'язати неможливо, оскільки факторних змінних може бути дуже багато. При регресійній моделі з n – факторними змінними зв’язок між результуючою ознакою і факторною змінною x_i при деяких фіксованих (наприклад, середніх арифметичних) значеннях решти факторів описується рівнянням часткової регресії:

$$\bar{y}_{x/xi} = b_0 + b_i x_i,$$

де $b_0 = a_0 + a_1 \bar{x}_1 + \ldots + a_{i-1} \bar{x}_{i-1} + a_{i+1} \bar{x}_{i+1} + \ldots + a_n \bar{x}_n$; $\bar{x}_1, \ldots, \bar{x}_{i-1}, \bar{x}_{i+1}, \ldots, \bar{x}_n$ – середні значення факторних змінних.

Зауважимо, що при переході від множинного рівняння регресії до часткового рівняння регресії фактори можна фіксувати не лише на середніх, але й на інших довільних рівнях. Крім того, на відміну від множинних рівнянь рівняння часткових регресій можна зобразити графіками на площині.

Одночасно з частковою регресією розглядають також часткову кореляцію, яка характеризує зв'язок між результуючою ознакою і деякою факторною змінною x_i при фіксованих (середніх) значеннях інших факторів множинного рівняння регресії.

Приклад. За статистичними даними за 10 років має місце залежність валового випуску продукції підприємства від наявних основних виробничих фондів та оборотних коштів. Скласти рівняння регресії. Дані задаються таблицею:

Роки	1	2	3	4	5	6	7	8	9	10
Валовий випуск (тис. гри.)	425	471	510	565	592	618	615	645	641	554
Основні виробничі фонди (тис. грн.)	320	341	360	389	400	430	444	397	471	320
Оборотні кошти (тис. грн.)	120	137	140	161	180	200	222	217	195	191

Розв'язання. Запишемо рівняння кореляційної залежності у такому вигляді: $\bar{y}_x = a_0 + a_1 x_1 + a_2 x_2$, де через x_1 ми позначили вартість основних виробничих фондів, а x_2 – вартість оборотних коштів. Складаємо розрахункову таблицю:

РІК	x_1	x_2	y	x_1^2	$x_1 x_2$	x_2^2	$x_1 y$	$x_2 y$
1	320	120	425	102400	38400	14400	136000	51000
2	341	137	471	116281	46717	18769	160601	64527
3	360	140	510	129600	50400	19600	183600	71400
4	389	161	565	151321	62629	25921	219785	90965
5	400	180	592	160000	72000	32400	236800	106560
6	430	200	618	184900	86000	40000	265740	123600
7	444	222	615	197136	98568	49284	273060	136530
8	397	217	645	157609	86149	47089	256065	139965
9	471	195	641	221841	91845	38025	301911	124995
10	320	191	554	102400	61120	36481	177280	105814
Суми	3872	1763	5636	1523488	693828	321969	2210842	1015356

Підставивши обчислені суми в систему рівнянь (5.5), одержуємо систему нормальних рівнянь:

$$\begin{cases} 10a_1 + 3872a_2 + 1763a_3 = 5636, \\ 3872a_1 + 1523488a_2 + 693828a_3 = 2210842, \\ 1763a_1 + 693828a_2 + 321969a_3 = 1015356. \end{cases}$$

Знайдемо розв'язок цієї системи рівнянь, наприклад, методом Крамера: $a_0 = 110{,}535$; $a_1 = 0{,}521$; $a_3 = 1{,}425$. Тоді вибіркове рівняння множинної регресії має вигляд: $\bar{y}_x = 110{,}535 + 0{,}521x_1 + 1{,}425x_2$.

Для знаходження, наприклад, рівняння часткової регресії, яке характеризує зв'язок між валовим випуском і вартістю введених в дію виробничих фондів x_1 при деякому постійному (середньому) рівні оборотних коштів x_2, потрібно у множинне рівняння регресії $\bar{y}_x = 110{,}535 + 0{,}521x_1 + 1{,}425x_2$ замість фактора x_2, підставити його середнє значення $\bar{x}_2 = 173{,}6$. У результаті одержимо вибіркове рівняння часткової регресії між факторами y і x:

$$\bar{y}_{x_1} = 931{,}24 + 0{,}521x_1.$$

Величина $\bar{y}_{x_1}$ дорівнює теоретичному значенню результуючої ознаки з урахуванням заданої величини першої факторної змінної x_1 при закріпленій на середньому рівні факторній змінній x_2.

Коефіцієнт часткової регресії збігається з відповідним коефіцієнтом множинної регресії і має аналогічний економічний зміст. Зокрема, у наведеному прикладі частковий коефіцієнт регресії показує, що збільшення основних фондів на 1 тис. грн. при постійних (середніх) значеннях трудових ресурсів дає приріст валового випуску в середньому на 0,521 тис. грн.

Завдання для самостійної роботи

Задача . За статистичними даними за 10 років (в тис. грн.), які подані в таблиці, побудувати функцію $Y = a_0 + a_1 X_1 + a_2 X_2$; оцінити правильність вибору форми функції (валовий випуск продукції Y; основні виробничі фонди X_1; оборотні засоби X_2).

Варіанти:							
1				2			
Рік	Y	X_1	X_2	Рік	Y	X_1	X_2
1	425	305	92	1	425	435	312
2	471	326	100	2	471	491	335
3	510	348	112	3	510	520	390
4	565	400	128	4	565	565	410
5	592	440	143	5	592	592	430
6	618	520	159	6	618	622	495
7	615	100	125	7	615	615	455
8	645	262	126	8	645	645	543
9	641	236	215	9	641	741	510
10	554	179	435	10	554	554	648

5.4. Нелінійна регресія

Якщо попередній аналіз явищ, зв'язок між якими досліджується, показує, що однаковим змінам середніх значень факторної ознаки відповідають неоднакові зміни середніх значень результативної ознаки, то для вираження загального характеру зв'язку застосовують криволінійні форми кореляційних рівнянь. На практиці найчастіше використовуються такі нелінійні функції залежності: гіперболічна, параболічна другого порядку, напівлогарифмічна та деякі інші.

Статистичний зв’язок між характеристиками виділяють за допомогою такої математичної функції, яка дає найменше відхилення від отриманих зі спостережень значень характеристик. Рівняння таких функцій називаються ***рівняннями зв’язків*** між результуючими та фактичними характеристиками.

Вид функції, заданої рівнянням зв’язку, визначає і розмежовує зв’язки за видами їх прояву на ***лінійні*** **і** ***криволінійні*** (параболічні, гіперболічні, ступеневі і т.д.)

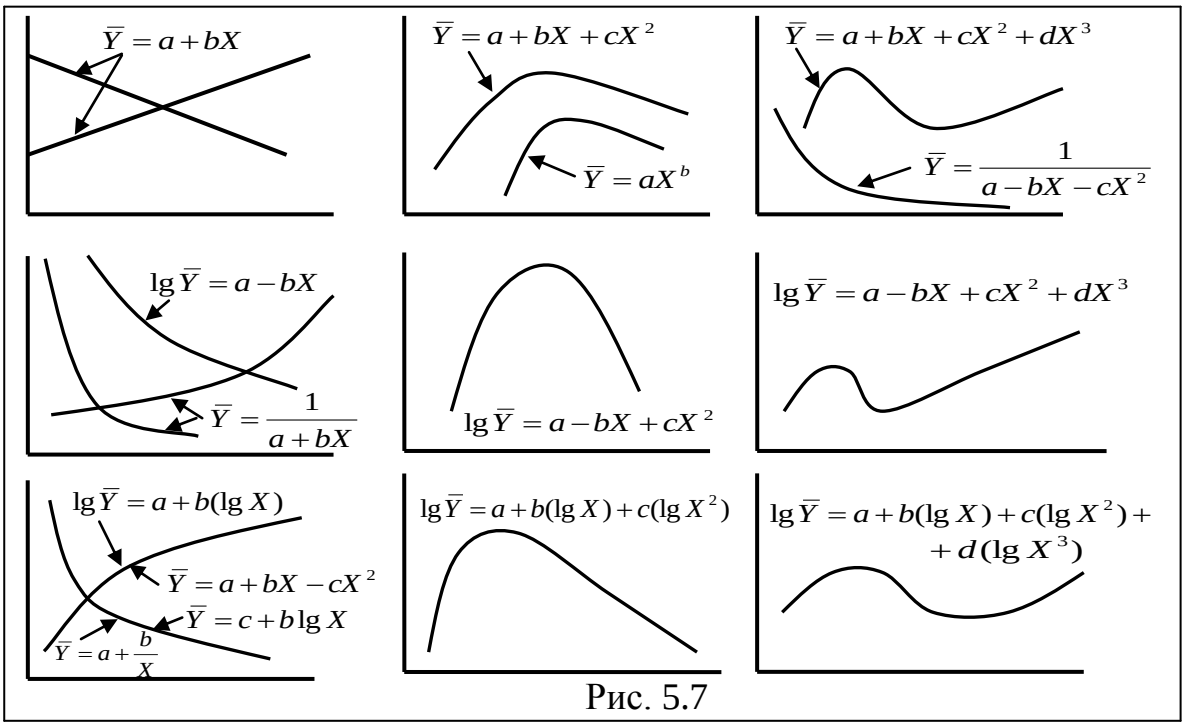


Рис. 5.7

Якщо результативна ознака при збільшенні факторної ознаки спадає, але не нескінченно, а прямує до певного рівня, то для її аналізу застосовується рівняння гіперболи $\bar{Y}_x = a_0 + a_1 \frac{1}{x}$.

Для знаходження параметрів цього рівняння методом найменших квадратів складають і розв'язують систему рівнянь з двома невідомими:

$$a_1 \sum_{i=1}^{n} \frac{1}{x_i} + n\, a_0 = \sum_{i=1}^{n} y_i \,,$$

$$a_1 \sum_{i=1}^{n} \frac{1}{x_i^2} + a_0 \sum_{i=1}^{n} \frac{1}{x_i} = \sum_{i=1}^{n} \frac{1}{x_i} y_i \ .$$

Коефіцієнт кореляції дозволяє достатньо точно оцінити щільність зв'язку у випадку лінійної залежності між ознаками. При наявності криволінійної залежності для оцінки щільності кореляційного зв'язку потрібно використовувати кореляційне відношення:

$$\eta_{x/y} = \sqrt{\sum_{i=1}^{n} \left(\bar{Y}_x - \bar{y}\right)^2 \Big/ \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2} \ . \qquad (5.6)$$

Величина $\eta_{x/y}$ задовольняє нерівності $0 \le \eta_{x/y} \le 1$. Якщо $\eta_{x/y} = 0$, то випадкова величина Y не знаходиться в кореляційній залежності від X. Тоді як по

мірі наближення $\eta_{x/y}$ до 1 щільність зв'язку Y з X зростає, і при $\eta_{x/y}$ = 1 вона стає функціональною.

Зразки розв'язування задач

Приклад 1. За даними таблиці побудувати рівняння регресії, визначити кореляційне відношення та перевірити його надійність $(\gamma = 0{,}95;\ \alpha = 0{,}05)$:

x	1,0	2,0	3,0	5,0	10,0
y	7,0	5,0	3,0	2,0	1,5

Розв'язання. Аналіз даних у таблиці показує, що результативна ознака при збільшенні факторної ознаки спадає, але не нескінченно, а прямує до певного рівня, тому для її аналізу доцільно застосовувати гіперболічну форму залежності. Розрахунки оформимо у вигляді таблиці:

Номер спостереження	x	y	$\frac{1}{x}$	$\frac{1}{x^2}$	$\frac{y}{x}$	$\bar{Y}_x = 1{,}014 + 6{,}305\frac{1}{x}$	$(y-\bar{y})^2$	$(\bar{Y}_x - \bar{y})$	$(\bar{Y}_x - \bar{y})^2$
1	1	7,0	1,00	1,00	7,00	7,22	10,89	3,52	12,39
2	2	5,0	0,50	0,25	2,50	4,17	1,69	0,47	0,22
3	3	3,0	0,33	0,11	1,00	3,09	0,49	– 0,61	0,37
4	5	2,0	0,20	0,04	0,4	2,28	2,89	– 1,42	2,02
5	10	1,5	0,10	0,01	0,15	1,64	4,84	-2,06	4,24
Разом	21	18,5	2,13	1,41	11,05	18,50	20,80	x	19,24

Отже $$a_0 = \frac{\sum_{i=1}^{n}\frac{1}{x_i^2}\sum_{i=1}^{n}y_i - \sum_{i=1}^{n}\frac{1}{x_i}\sum_{i=1}^{n}\frac{y_i}{x_i}}{n\sum_{i=1}^{n}\frac{1}{x_i^2} - \sum_{i=1}^{n}\frac{1}{x_i}\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{1{,}41\cdot 18{,}5 - 2{,}13\cdot 11{,}05}{5\cdot 1{,}41 - 2{,}13\cdot 2{,}13} = 1{,}014\ ,$$

$$a_1 = \frac{n\sum_{i=1}^{n}\frac{y_i}{x_i} - \sum_{i=1}^{n}\frac{1}{x_i}\sum_{i=1}^{n}y_i}{n\sum_{i=1}^{n}\frac{1}{x_i^2} - \sum_{i=1}^{n}\frac{1}{x_i}\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{5\cdot 11{,}05 - 2{,}13\cdot 18{,}5}{5\cdot 1{,}41 - 2{,}13\cdot 2{,}13} = 6{,}305\ .$$

Таким чином, рівняння регресії матиме вигляд: $\bar{Y}_x = 1{,}014 + 6{,}305\frac{1}{x}$.

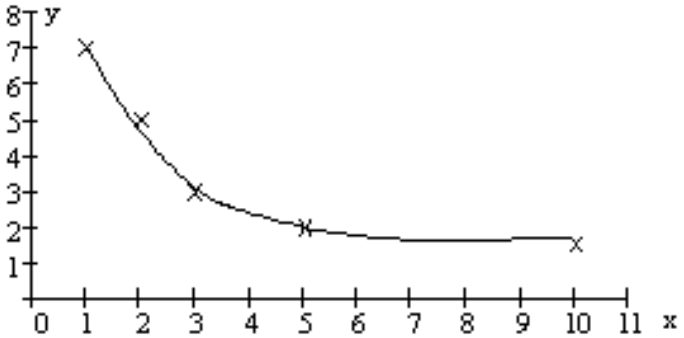


Рис. 5.8

На рис. 5.8 побудовано кореляційне поле та теоретичну лінію регресії Y по X.

Середнє значення результативної ознаки: $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{18,5}{5} = 3,7$.

Фактичні і теоретичні значення досліджуваних ознак (наведені в таблиці) не дуже відрізняються. Для визначення щільності зв'язку між результативною і факторною ознаками обчислимо кореляційне відношення:

$$\eta_{x/y} = \sqrt{\frac{\sum_{i=1}^{n} (Y_x - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}} = \sqrt{\frac{19,25}{20,8}} = \sqrt{0,925} = 0,96.$$

Значення $\eta_{x/y}$ наближене до 1, отже, кореляційне відношення показує, що між ознаками Y та X існує щільна обернена залежність. Надійність показника кореляційного відношення перевіримо за t – критерієм Стьюдента. Для цього спочатку визначимо середню похибку кореляційного відношення:

$$\mu_\eta = \frac{1-\eta^2}{\sqrt{n-1}} = \frac{1-0,96^2}{\sqrt{4}} = 0,04; \quad t = \frac{\eta}{\mu_\eta} = \frac{0,96}{0,04} = 24,$$

що свідчить про високу надійність кореляційного відношення (24 > 3,18= $t_{0,5;3}$).

Парабола другого порядку як форма математичного вираження зв'язків між Y та X застосовується у тих випадках, коли із зростанням факторної ознаки відбувається нерівномірне зростання або спадання результативної ознаки.

При знаходженні рівняння регресії застосовують тип кривої у вигляді параболи другого порядку $\bar{Y}_x = a_0 + a_1 x + a_2 x^2$. Параметри цього рівняння знаходять методом найменших квадратів шляхом складання і розв'язку системи рівнянь:

$$a_2 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i + n a_0 = \sum_{i=1}^{n} y_i ,$$

$$a_2 \sum_{i=1}^{n} x_i^3 + a_1 \sum_{i=1}^{n} x_i^2 + a_0 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i ,$$

$$a_2 \sum_{i=1}^{n} x_i^4 + a_1 \sum_{i=1}^{n} x_i^3 + a_0 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2 y_i .$$

Приклад 2. За даними таблиці побудувати рівняння регресії, визначити кореляційне відношення та перевірити його надійність $(\gamma = 0,95;\ \alpha = 0,05)$:

x	0,0	5,0	10,0	15,0	20,0	25,0	30,0	35,0	40,0	45,0
y	1,2	5,0	7,0	8,0	9,2	9,5	9,7	9,0	9,8	8,7

Розв'язання. Система рівнянь буде мати вигляд:

$$\mathbf{78 = 10a_0 + 225a_1 + 7125a_2}\ ,$$
$$\mathbf{2057{,}5 = 225a_0 + 7125a_1 + 253125a_2}\ ,$$
$$\mathbf{66697{,}5 = 7125a_0 + 253125a_1 + 9583125a_2}\ .$$

Розрахунки оформимо у вигляді таблиці. У результаті розв'язування системи рівнянь одержимо: $\boldsymbol{a_0 = 2{,}992}$; $\boldsymbol{a_1 = 0{,}372}$; $\boldsymbol{a_2 = -0{,}005.}$ Таким чином, рівняння регресії матиме вигляд: $\boldsymbol{\bar{Y}_x = 2{,}992 + 0{,}372x - 0{,}005x^2}$.

Номер спостереження	x	y	x^2	x^3	x^4	xy	x^2y	$\bar{Y}_x$
1	0	1,2	0	0	0	0,0	0,0	3,0
2	5	5,0	25	125	625	25,0	125,0	4,7
3	10	7,0	100	1000	10000	70,0	700,0	6,2
4	15	8,0	225	3375	50625	120,0	1800,0	7,4
5	20	9,2	400	8000	160000	184,0	3680,0	8,4
6	25	9,5	625	15625	390625	237,5	5937,5	9,2
7	30	9,7	900	27000	810000	291,0	8730,0	9,6
8	35	9,0	1225	42875	1500625	346,5	12127,5	10,0
9	40	9,8	1600	64000	2560000	392,0	15680,0	9,9
10	45	8,7	2025	91125	4100625	391,5	17617,5	9,6
Разом	225	78,0	7125	253125	9583125	2057,5	66697,5	78,0

Побудуємо кореляційне поле та теоретичну лінію регресії ***Y*** по ***X*** (рис. 5.9).

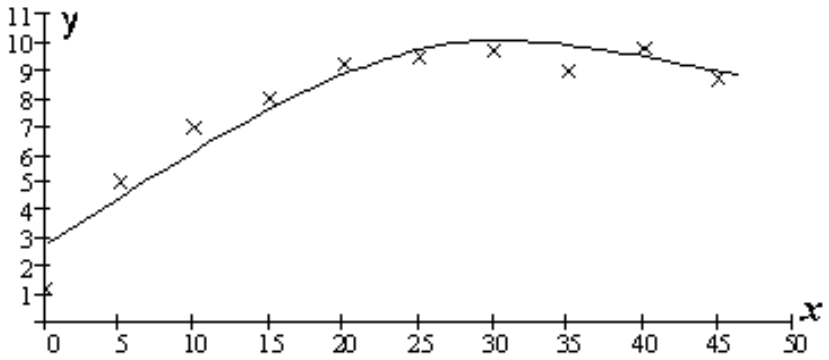


Рис. 5.9

Наведемо наступну розрахункову таблицю:

x	0,00	5,00	10,00	15,00	20,00	25,00	30,00	35	40	45	Разом
y	1,20	5,00	7,00	8,00	9,20	9,50	9,70	9,00	9,80	8,70	78,00
$\bar{Y}_x$	3,00	4,70	6,20	7,40	8,40	9,20	9,60	10,00	9,90	9,60	78,00
$(y-\bar{y})^2$	43,56	7,84	0,64	0,04	1,96	2,89	3,61	4,41	4,00	0,81	69,76
$(\bar{Y}_x-\bar{y})$	4,80	3,10	1,60	0,40	0,60	1,40	1,80	2,20	2,10	1,80	X
$(\bar{Y}_x-\bar{y})^2$	23,04	9,61	2,56	0,16	0,38	1,96	3,24	2,84	4,51	3,24	53,42

Кореляційне відношення за даними розрахунків дорівнює:

$$\eta_{x/y} = \sqrt{\frac{\sum_{i=1}^{n}(Y_x - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \sqrt{\frac{53,42}{69,76}} = \sqrt{0,766} = 0,875$$

Значення $\eta_{x/y}$ наближене до 1, отже, кореляційне відношення показує, що між ознаками ***Y*** та ***X*** існує щільна обернена залежність.

Надійність, показника кореляційного відношення перевіримо за t – критерієм Стьюдента. Для цього спочатку визначимо середню похибку кореляційного відношення:

$$\mu_\eta = \frac{1-\eta^2}{\sqrt{n-1}} = \frac{1-0,766}{3} = 0,078; \quad t = \frac{\eta}{\mu_\eta} = \frac{0,875}{0,078} = 11,22.$$

Оскільки $t_\eta > t_{табл}(11,22 > 3,18 = t_{0,5;3})$, залежність між ознаками можна вважати доведеною.

Завдання для самостійної роботи

За даними таблиць побудувати відношення та перевірити його надійність.

Варіант 1			Варіант 2		
N	***X***	***Y***	***N***	***X***	***Y***
1	3	47	1	5,5	4
2	7	35	2	6	2,8
3	11	20	3	6,5	1,8
4	15	15	4	7	1,2
5	19	12	5	7,5	0,8
6	23	10	6	8	0,5
7	27	9	7	8,5	0,3
8	31	8,2	8	9	0,2
9	35	7,7	9	9,5	0,1
10	39	7,5	10	10	0,1

Варіант 3			Варіант 4		
N	***X***	***Y***	***N***	***X***	***Y***
1	1	7	1	1	2
2	1,5	12	2	4	12
3	2	16	3	7	20
4	2,5	19	4	10	27
5	3	21	5	13	33
6	3,5	22	6	16	38
7	4	22	7	19	42
8	4,5	21	8	22	45
9	5	20	9	25	43
10	5,5	19	10	27	40

Додаток 1

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
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

Продовження додатка 1

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
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
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 2

ТАБЛИЦЯ ЗНАЧЕНЬ $t(j, k = n-1)$,

ЩО ЗАДОВОЛЬНЯЮТЬ РІВНІСТЬ $p(t) = 2\int_0^t f(x)dt = \gamma$

k	***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
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 3

ЗНАЧЕННЯ ВЕЛИЧИНИ χ_1^2 ЗАЛЕЖНО ВІД ІМОВІРНОСТІ $P(\chi^2 > \chi_1^2)$

Число ступенів вільності, 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

ЗНАЧЕННЯ ВЕЛИЧИНИ χ_2^2 ЗАЛЕЖНО ВІД ІМОВІРНОСТІ $P(\chi^2 > \chi_1^2)$

Число ступенів вільності, 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 4

КРИТИЧНІ ТОЧКИ РОЗПОДІЛУ χ^2

Число ступенів вільності, ***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

Додаток 5

КРИТИЧНІ ТОЧКИ РОЗПОДІЛУ СТЬЮДЕНТА (*t* -РОЗПОДІЛУ)

Число ступенів вільності, ***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 6

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

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 7

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

ЛІТЕРАТУРА

1. Бугір М.К. Теорія ймовірностей та математична статистика. –Тернопіль: Підручники & посібники, 1998. – 176 с.
2. Овчинников П.П. та інші. Вища математика: Підручник. У 2 ч. – К.: Техніка, 2004.
3. Жлуктенко В. І., Наконечний С. І., Савіна С. С. Теорія ймовірностей і математична статистика: Навч.-метод. посібник: У 2-х ч. –Ч. ІІ. Математична статистика. – К.: КНЕУ, 2001. – 334 с.
4. Пасічник І.В., Сясєв А.В., Маринчук Л.В. Теорія ймовірностей та випадкові процеси. Ч.1: Конспект лекцій. – Дніпропетровськ, НМетАУ. 2005. – 48 с.
5. Теорія ймовірностей та випадкові процеси. Частина ІІ: Конспект лекцій / Л.В. Маринчук, І.В. Пасічник, А.В. Сясєв, А.Г. Моня. – Дніпропетровськ: НМетАУ, 2006. – 48 с.