

ДЕРЖАВНИЙ КОМІТЕТ ЗВ'ЯЗКУ ТА ІНФОРМАТИЗАЦІЇ УКРАЇНИ
ОДЕСЬКА НАЦІОНАЛЬНА АКАДЕМІЯ ЗВ'ЯЗКУ ім. О.С. Попова

Кафедра вищої математики

**ДОСЛІДЖЕННЯ СТАТИСТИЧНИХ ХАРАКТЕРИСТИК
ВИПАДКОВИХ ВЕЛИЧИН**

Методичні вказівки та варіанти комплексного завдання
для студентів другого курсу всіх спеціальностей

ЗАТВЕРДЖЕНО
методичною радою
академії
Протокол № 10
від 11.03.2003 р.

Одеса 2004

УДК 517-519.2

План НМВ 2004 навч. р.

Рецензент – *доц. В.Е. Горбачов*

Укладачі: *доц. І.В. Стрелковська, доц. Л.І. Соколов,
доц. О.П. Яковчук, викл. Т.І. Григор'єва*

Комп'ютерна правка
та макетування *Є. С. Корнійчук*

Наведені методичні вказівки містять зразки розв'язків завдань та варіанти комплексного завдання з елементів математичної статистики, які були запропоновані проф. Г.В. Піддубним.

Методичні вказівки призначені для студентів усіх спеціальностей, що вивчають курс вищої математики.

СХВАЛЕНО
на засіданні кафедри
вищої математики і
рекомендовано до друку.
Протокол № 9
від 31.01.2003 р.

I. Стислі теоретичні відомості з обробки дослідних даних

1. Постановка задач з обробки дослідних даних

Розглянемо деяку сукупність об'єктів Ω , що мають якусь спільну ознаку. Якісна характеристика X цієї ознаки у різних об'єктів з Ω може бути різною. Для випадково обраного об'єкта з Ω кількісна характеристика ознаки X приймає деяке значення, заздалегідь невідомо яке. Якщо дослід полягає у витягу навмання якогось об'єкта з Ω , то Ω можна розглядати як простір елементарних подій. Оскільки кожному з можливих результатів такого дослідження ставиться у відповідність певне значення X , то X можна розглядати як випадкову величину, визначену на Ω .

Нехай з Ω витягаються N об'єктів (проводиться N дослідів). Позначимо через

$$x^{(1)}, x^{(2)}, \dots, x^{(N)} \quad (1)$$

значення, що приймає випадкова величина X на відповідних об'єктах.

Сукупність чисел (1) називається *вибіркою* обсягу N для випадкової величини X , а самі ці числа *варіантами* або *вибірковими значеннями*.

Вибірка (1) являє собою первинний статистичний матеріал, що підлягає обробці. При цьому виникають такі типові задачі:

а) задача визначення закону розподілу X за статистичними даними (1);

б) задача перевірки правдоподібності тієї чи іншої гіпотези відносно X , наприклад гіпотези про передбачуваний закон розподілу X ;

в) задача знаходження невідомих параметрів розподілу X , наприклад її математичного сподівання та дисперсії.

2. Статистичні розподіли

Первинна обробка даних вибірки полягає у складанні таблиці, в одному рядку якої записуються у зростаючому порядку різні значення варіант з вибірки (1), а у другому – відносні частоти, з якими ці значення зустрічаються в розглянутій вибірці.

Така таблиця називається *статистичним рядом*.

Позначимо всі різні і розташовані у зростаючому порядку значення варіант з вибірки (1) через

$$x_1, x_2, \dots, x_n; \quad x_i < x_{i+1}, \quad n \leq N \quad (2)$$

і нехай значення x_i зустрічається у вибірці (1) m_i разів. Тоді відносна частота значення x_i буде дорівнювати

$$W_i = W(X = x_i) = \frac{m_i}{N}, \quad i = 1, 2, \dots, n,$$

при цьому

$$\sum_{i=1}^n m_i = N, \quad \sum_{i=1}^n W_i = 1.$$

Тут і далі через $W(A)$ позначається відносна частота події A .

З урахуванням зроблених позначень статистичний ряд, побудований за вибіркою (1), приймає такий вигляд:

x_i	x_1	x_2	...	x_n
W_i	W_1	W_2	...	W_n

(3)

Іншою формою статистичного розподілу є *емпірична* (статистична) *функція розподілу* $F^*(x)$, що визначає для кожного значення x відносну частоту події $X < x$:

$$F^*(x) = W(X < x) = \frac{k}{N}, \quad (4)$$

де k – число тих значень варіант з вибірки (1), що менші за x .

Якщо використовувати статистичний ряд (3), то для побудови $F^*(x)$ зручною є така формула:

$$F^*(x) = \sum_{i=1}^k W_i, \quad (5)$$

де k – число тих значень x з варіаційного ряду (2), що менші за x .

Емпірична функція розподілу має такі властивості:

1) $F^*(x)$ – неспадна функція;

2) $0 \leq F^*(x) \leq 1$;

3) $F^*(x) = \begin{cases} 0, & \forall x \leq x_1, \\ 1, & \forall x > x_n, \end{cases}$

де x_1 – найменша, а x_n – найбільша варіанта.

Приклад 1. Проведено $N = 10$ дослідів над випадковою величиною X . Результати цих дослідів подані у вибірці:

i (номер досліджу)	1	2	3	4	5	6	7	8	9	10
x_i	2	7	8	5	2	7	2	7	7	8

Побудувати статистичний ряд і емпіричну функцію розподілу.

Розв'язок. На підставі дослідних даних будуємо статистичний ряд:

x_i	2	5	7	8
m_i	3	1	4	2
W_i	0,3	0,1	0,4	0,2

Оскільки найменше спостережене значення величини X дорівнює 2, то

$$F^*(x) = 0 \text{ при } x \leq 2.$$

Значення $x < 5$, а саме $x_1 = 2$ спостерігалось з частотою $W_1 = 0,3$, отже,

$$F^*(x) = W_1 = 0,3 \text{ при } 2 < x \leq 5.$$

Значення $x < 7$, а саме $x_1 = 2$ і $x_2 = 5$, спостерігалися з частотами відповідно $W_1 = 0,3$ і $W_2 = 0,1$, отже,

$$F^*(x) = W_1 + W_2 = 0,4 \text{ при } 5 < x \leq 7.$$

Аналогічно,

$$F^*(x) = W_1 + W_2 + W_3 = 0,8 \text{ при } 7 < x \leq 8,$$

$$F^*(x) = W_1 + W_2 + W_3 + W_4 = 1 \text{ при } x > 8.$$

Отже,

$$F^*(x) = \begin{cases} 0 & \text{при } x \leq 2; \\ 0,3 & \text{при } 2 < x \leq 5; \\ 0,4 & \text{при } 5 < x \leq 7; \\ 0,8 & \text{при } 7 < x \leq 8; \\ 1 & \text{при } x > 8. \end{cases}$$

Графік цієї функції зображений на рис. 1.

Якщо випадкова величина X є неперервною, то, як правило, усі її спостережені значення різні і тому, особливо при великій кількості дослідів, первинна обробка вибірки у вигляді статистичного ряду (3) позбавлена сенсу. У цьому випадку, розташовуючи спостережені значення X у неспадному порядку, розбивають проміжок $[x_1, x_{n+1}]$, де x_1 – найменше, а x_{n+1} – найбільше зі спостережених значень X , на кілька часткових інтервалів:

$$(x_1, x_2), (x_2, x_3), \dots, (x_n, x_{n+1})$$

і підраховують кількість m_i тих значень вибірки, що потрапили в i -й інтервал (x_i, x_{i+1}) . При цьому, очевидно, $\sum_{i=1}^n m_i = N$, де N – обсяг вибірки.

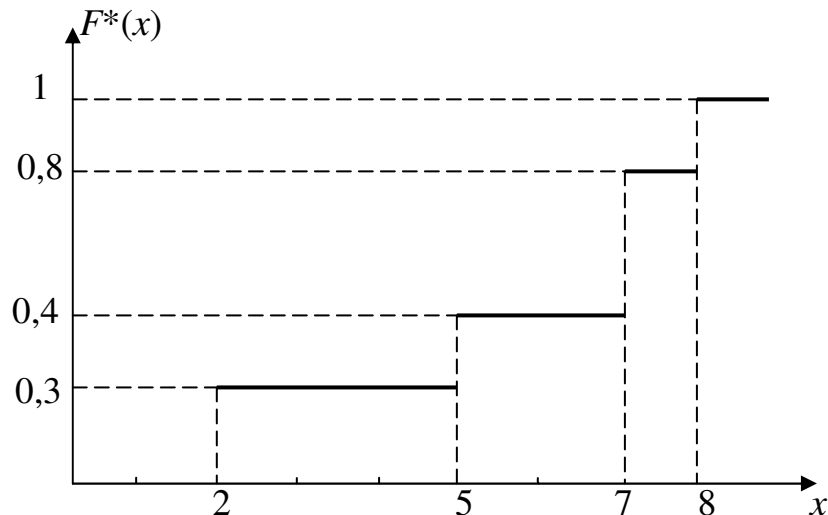


Рис. 1

Потім підраховують відносну частоту влучень спостережених значень X у кожен з інтервалів:

$$p^*_i = W(x_i < X < x_{i+1}) = \frac{m_i}{N}, \quad (6)$$

де m_i – число спостережених значень X , що потрапили в i -й частковий інтервал. Якщо будь-яке зі спостережених значень X знаходиться точно на межі двох інтервалів, то рекомендується вважати дане значення приналежним (умовно) до обох інтервалів і додавати до чисел m_i того й іншого інтервалу по 0,5.

Сума відносних частот усіх часткових інтервалів, очевидно, дорівнює одиниці. Результати такої обробки первинних даних зводяться в табл. 1. Цю таблицю будемо називати *розподілом вибірки*.

Таблиця 1

Номер інтервалу i	1	2	...	i	...	n
Межі інтервалів (x_i, x_{i+1})	x_1, x_2	x_2, x_3	...	x_i, x_{i+1}	...	x_n, x_{n+1}
Чисельність i -го інтервалу m_i	m_1	m_2	...	m_i	...	m_n
Відносні частоти p_i^*	p_1^*	p_2^*	...	p_i^*	...	p_n^*

Зауваження. 1. Число часткових інтервалів не повинно бути занадто великим. Якщо обсяг вибірки складає більш за сотню спостережень, то практика показує, що доцільно вибирати число інтервалів порядку 10 – 15.

2. Довжини часткових інтервалів можна вибирати як однаковими, так і різними. Для простоти розрахунків їх звичайно беруть однаковими. Однак, якщо число спостережень у деяких інтервалах виявиться невеликим (меншим за 5), то кожний з цих інтервалів доцільно об'єднати з якимось з найближчих.

Розподіл вибірки графічно подається у вигляді так званої *гістограми*, що будується таким чином: на осі абсцис відкладаються часткові інтервали і на кожному з них, як на основі, будується прямокутник, площа якого дорівнює відносній частоті спостережень у даному інтервалі (рис. 2). Для побудови гістограми треба відносну частоту кожного інтервалу розділити на його довжину й отримане число взяти як висоту прямокутника h_i . Таким чином, h_i обчислюється за формулою

$$h_i = \frac{p_i^*}{x_{i+1} - x_i}.$$

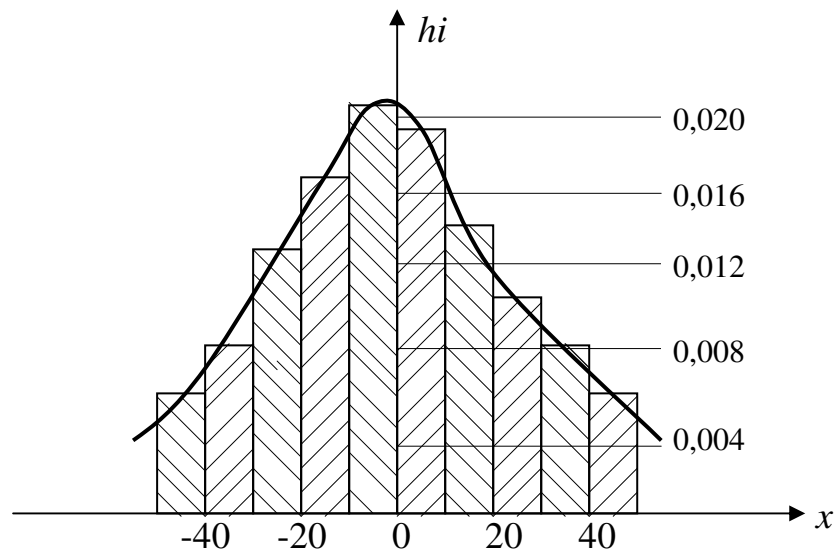


Рис. 2

Зі способу побудови гістограми видно, що повна її площа дорівнює одиниці. Крім цього, очевидно, що зі збільшенням числа дослідів можна вибирати все більш і більш дрібні часткові інтервали; при цьому гістограма буде все більш наближатися до деякої кривої, що являє собою графік щільності розподілу випадкової величини X . Таким чином, замінюючи східчасту лінію, що обмежує гістограму, неперервною кривою, можна висловити припущення про можливий закон розподілу величини X (рис. 2).

Користаючись даними розподілу вибірки, можна приблизно побудувати емпіричну функцію розподілу величини X . Побудова точної емпіричної функції розподілу з декількома сотнями стрибків у всіх спостережених значеннях X не має наочності. Звичайно, досить побудувати емпіричну функцію розподілу приблизно за декількома точками. За ці точки зручно взяти межі $x_1, x_2, \dots$ часткових інтервалів, що фігурують у розподілі вибірки. Тоді, очевидно,

$$\left. \begin{aligned} F^*(x_1) &= 0, \\ F^*(x_2) &= p_1^*, \\ F^*(x_3) &= p_1^* + p_2^*, \\ &\Lambda \\ F^*(x_i) &= \sum_{k=1}^{i-1} p_k^*, \\ &\Lambda \\ F^*(x_{n+1}) &= \sum_{k=1}^n p_k^*. \end{aligned} \right\} \quad (8)$$

З'єднуючи отримані точки плавною кривою, одержимо наближений графік емпіричної функції розподілу (рис. 3).

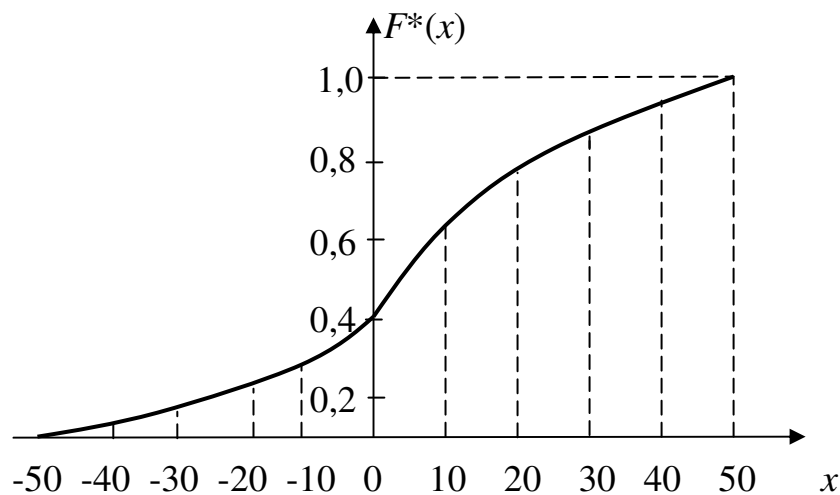


Рис. 3

3. Оцінка невідомої числової характеристики закону розподілу

Нехай над випадковою величиною X проводиться N незалежних дослідів і нехай X_k — випадкова величина, яка дорівнює значенню, що заздалегідь невідомо, прийнятому величиною X у k -му досліді. Тоді

сукупність незалежних випадкових величин

$$X_1, X_2, \dots, X_N, \quad (9)$$

що мають той самий закон розподілу, що і для X , можна розглядати як *випадкову вибірку*.

Нехай a – деяка числова характеристика величини X , що нам невідома. Тоді оцінка цієї величини $\tilde{a}$, обчислена на основі сукупності величин (9), буде являти собою деяку функцію від $X_1, X_2, \dots, X_N$,

$$\tilde{a} = f(X_1, X_2, \dots, X_N) \quad (10)$$

і, отже, сама є випадковою величиною. Закон розподілу $\tilde{a}$ залежить від обраного виду функції f , від закону розподілу X (зокрема від невідомого параметра a) і від обсягу вибірки N .

Однією з основних задач теорії оцінок параметрів розподілу є вибір виду функції (10), яку варто прийняти як оцінку параметра a . Слід вимагати від оцінки $\tilde{a}$, щоб вона при збільшенні числа дослідів наближалася до параметра a .

З закону великих чисел відомо, що середнє арифметичне спостережених значень $X_1, X_2, \dots, X_N$ випадкової величини X при необмеженому збільшенні числа іспитів N збігається за ймовірністю до математичного сподівання m_x цієї величини:

$$\frac{1}{N} \sum_{i=1}^N X_i \xrightarrow{\text{з.й.}} m_x, \quad N \rightarrow \infty.$$

Тому природно за оцінку m_x взяти функцію

$$\tilde{m}_x = \frac{1}{N} \sum_{i=1}^N X_i. \quad (11)$$

Розглянемо тепер дисперсію величини X :

$$D_x = M[(X - m_x)^2].$$

Якщо в цьому виразі математичне сподівання замінити його оцінкою, то ми одержимо оцінку дисперсії:

$$\bar{D}_x = \frac{1}{N} \sum_{i=1}^N (X_i - \tilde{m}_x)^2 \quad (12)$$

Можна показати, що $\bar{D}_x$ при $N \rightarrow \infty$ збігається за ймовірністю до оцінюваного параметра D_x :

$$\bar{D}_x \xrightarrow{\text{з.й.}} D_x, \quad N \rightarrow \infty.$$

Означення 1. Оцінка $\tilde{a}$ параметра a називається обґрунтованою,

якщо вона збігається за ймовірністю до оцінюваного параметра при необмеженому збільшенні числа дослідів.

Таким чином, оцінки (11) і (12) математичного сподівання і дисперсії є обґрунтованими.

Крім обґрунтованості, слід вимагати, щоб оцінка параметра розподілу мала ще ряд властивостей, що визначають "доброякісність" оцінки.

Означення 2. Оцінка $\tilde{a}$ параметра a називається незсуненою, якщо її математичне сподівання дорівнює оцінюваному параметру:

$$M[\tilde{a}] = a. \quad (13)$$

Неважко переконатися, що оцінка $\tilde{m}_x$ математичного сподівання є незсуненою. Оцінка ж дисперсії, що визначається за формулою (12), цієї властивості не має. Дійсно, обчислення дають

$$M[\bar{D}_x] = \frac{N-1}{N} D_x \neq D_x.$$

Користуючись оцінкою $\bar{D}_x$ для визначення дисперсії D_x , ми будемо робити деяку систематичну помилку в менший бік. Щоб ліквідувати це зміщення, необхідно ввести поправку, помноживши $\bar{D}_x$ на $\frac{N}{N-1}$. У результаті одержимо "виправлену" оцінку дисперсії:

$$\tilde{D}_x = \frac{1}{N-1} \sum_{i=1}^N (X_i - \tilde{m}_x)^2 \quad (14)$$

Оцінка $\tilde{D}_x = \frac{N}{N-1} \bar{D}_x$ вже є незсуненою, і оскільки множник $\frac{N}{N-1}$ прямує до одиниці при $N \rightarrow \infty$, то вона, як і $\bar{D}_x$, є обґрунтованою.

При великих N обидві оцінки $\tilde{D}_x$ і $\bar{D}_x$ будуть розрізнятися дуже мало, і введення поправкового множника втрачає сенс.

Крім обґрунтованості і незсуненості, бажано, щоб оцінка $\tilde{a}$ серед всіх інших можливих оцінок параметра a мала б мінімальну дисперсію:

$$D[\tilde{a}] = \min.$$

Оцінка, що має таку властивість, називається *ефективною*.

Можна показати, що оцінка $\tilde{m}_x$ є ефективною. Оцінка ж дисперсії $\tilde{D}_x$, взагалі говорячи, властивості ефективності не має.

4. Числові характеристики статистичного розподілу

Кожній числовій характеристиці випадкової величини X :

математичному сподіванню m_x , дисперсії D_x , середньому квадратичному відхиленню σ_x , відповідають їх статистичні аналогії: статистичне математичне сподівання (статистичне середнє), статистична дисперсія (вибіркова дисперсія), статистичне середньоквадратичне відхилення, що будуються на основі конкретних результатів дослідів величини X , поданих вибіркою

$$X^{(1)}, X^{(2)}, \dots, X^{(N)}.$$

Кожну зі статистичних характеристик слід припустити рівною тому значенню відповідної оцінки параметра $\tilde{a}$, що ця оцінка приймає для даної вибірки, тобто припустити

$$\tilde{a} = f(x^{(1)}, x^{(2)}, \dots, x^{(N)}).$$

Таким чином, $\tilde{m}_x$, $\tilde{D}_x$ і $\tilde{\sigma}_x$ визначаються такими виразами:

$$\tilde{m}_x = \frac{1}{N} \sum_{i=1}^N x^{(i)}, \quad (15)$$

$$\tilde{D}_x = \frac{1}{N-1} \sum_{i=1}^N (x^{(i)} - \tilde{m}_x)^2, \quad (16)$$

$$\tilde{\sigma}_x = \sqrt{\tilde{D}_x}, \quad (17)$$

які можна використовувати як наближені значення параметрів m_x , D_x і σ_x .

Зауваження. На практиці часто замість формули (16) зручно використовувати іншу, рівносильну їй формулу:

$$\tilde{D}_x = \left[\frac{1}{N} \sum_{i=1}^N (x^{(i)})^2 - \tilde{m}_x^2 \right] \frac{N}{N-1}. \quad (18)$$

Якщо вибірка (1) оброблена у вигляді статистичного ряду (3), то формули (15), (16) і (18) зводяться до такого вигляду:

$$\tilde{m}_x = \sum_{i=1}^n x_i W_i, \quad (19)$$

$$\tilde{D}_x = \frac{N}{N-1} \sum_{i=1}^n (x_i - \tilde{m}_x)^2 W_i, \quad (20)$$

$$\tilde{D}_x = \frac{N}{N-1} \left[\sum_{i=1}^n x_i^2 W_i - \tilde{m}_x^2 \right]. \quad (21)$$

При великому обсязі вибірки N обчислення числових характеристик статистичного розподілу за формулами (15), (16) чи (19), (20) стає надмірно громіздким. У цьому випадку доцільно скористатися обробкою спостережених даних у вигляді розподілу вибірки (див. табл. 1) і вважати наближене значення випадкової величини в кожному з часткових інтервалів

рівним його середині, тобто

$$\hat{x}_i = \frac{1}{2}(x_i + x_{i+1}), \quad i = 1, 2, \dots, n. \quad (22)$$

Тоді замість (15), (16) і (18) одержимо такі наближені формули:

$$\tilde{m}_x \approx \sum_{i=1}^n \hat{x}_i p_i^*, \quad (23)$$

$$\tilde{D}_x \approx \sum_{i=1}^n (\hat{x}_i - \hat{m})^2 p_i^*, \quad (24)$$

$$\tilde{D}_x \approx \sum_{i=1}^n \hat{x}_i^2 p_i^* - \tilde{m}_x^2, \quad (25)$$

які цілком аналогічні формулам, що визначають математичне сподівання m_x і дисперсію D_x дискретної випадкової величини.

5. Довірчий інтервал. Довірча ймовірність

Невідомі числові характеристики випадкової величини X ми оцінюємо за допомогою дослідних даних приблизно, вважаючи

$$m_x \approx \tilde{m}_x, D_x \approx \tilde{D}_x, \sigma_x \approx \tilde{\sigma}_x.$$

Виникають такі питання: які точність і надійність такої оцінки, тобто до яких помилок може привести заміна тієї чи іншої числової характеристики величини X відповідною їй статистичною характеристикою і з якою впевненістю можна чекати, що ці помилки не вийдуть за відомі межі?

Для відповіді на поставлене питання в математичній статистиці користуються так званими довірчими інтервалами і довірчими ймовірностями.

Означення. Нехай β – деяка ймовірність, настільки близька до одиниці (наприклад $\beta=0,9, 0,95$ чи $0,99$), що розглянуту з такою ймовірністю подію можна було б вважати практично достовірною, і нехай $\tilde{a}$ – оцінка деякого параметра a . Тоді випадковий проміжок

$$I_{\beta}^{(a)} = (\tilde{a} - \varepsilon, \tilde{a} + \varepsilon), \quad (26)$$

який з ймовірністю β накриває невідоме значення параметра a , називається довірчим інтервалом, а β – довірчою ймовірністю. Межі інтервалу $I_{\beta}^{(a)}$: $a_1 = \tilde{a} - \varepsilon$, $a_2 = \tilde{a} + \varepsilon$ називають довірчими.

Якщо β задано, то довжина довірчого інтервалу, що дорівнює 2ε , буде конкретним числом (невипадковою величиною), визначеним зі співвідношення

$$P(\tilde{a} - \varepsilon < a < \tilde{a} + \varepsilon) = \beta,$$

чи в еквівалентній формі:

$$P(|\tilde{a} - a| < \varepsilon) = \beta; \quad (27)$$

випадковість же довірчого інтервалу $I_{\beta}^{(a)}$ полягає в його випадковому розташуванні на числовій осі, тому що центр цього інтервалу $\tilde{a}$ – випадкова величина (рис. 4).

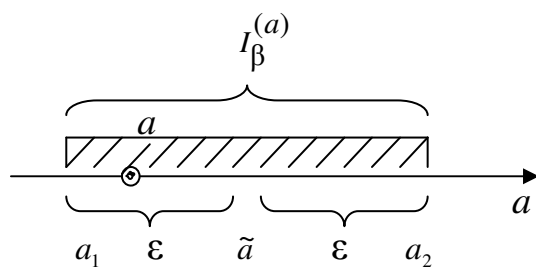


Рис. 4

Обчисливши те значення оцінки $\tilde{a}$, що вона приймає для даної вибірки, одержуємо дослідне значення довірчого інтервалу. У цьому випадку довірчий інтервал можна розглядати як діапазон тих значень оцінюваного параметра a , які не суперечать досліді,

тому що ті значення a , для яких $|\tilde{a} - a| > \varepsilon$, мають дуже малу ймовірність, що дорівнює $1 - \beta$, і тому можуть розглядатися як практично неможливі.

Розрахунок довірчого інтервалу для математичного сподівання

Нехай $X, X, \dots, X_n$ – значення випадкової величини X , які вона приймає в N незалежних дослідах, а m і D – її математичне сподівання та дисперсія, що невідомі. Тоді довжина довірчого інтервалу $2\varepsilon_\beta$ для математичного сподівання, що залежить від β , визначається відповідно до (27) з рівняння

$$P(|\tilde{m}_x - m| < \varepsilon_\beta) = \beta, \quad (28)$$

де

$$\tilde{m}_x = \frac{1}{N} \sum_{i=1}^N X_i.$$

При досить великій кількості дослідів N закон розподілу випадкової величини X можна приблизно вважати нормальним, параметри якого, як неважко переконатися, дорівнюють

$$M[\tilde{m}_x] = m_x, \quad \tilde{D}[\tilde{m}_x] = D_x / N.$$

Оскільки $D_x \approx \tilde{D}_x$, то з цим ступенем точності можна вважати

$$\sigma_{\tilde{m}_x} = \sqrt{D[\tilde{m}_x]} = \sqrt{D_x / N} \approx \sqrt{\tilde{D}_x / N}. \quad (29)$$

Використовуючи тепер відому з курсу теорії ймовірностей формулу для ймовірності влучення нормально розподіленої випадкової величини в інтервал, симетричний щодо її математичного сподівання m , маємо

$$P(|\tilde{m}_x - m_x| < \varepsilon_\beta) = 2\Phi\left(\frac{\varepsilon_\beta}{\sigma_{\tilde{m}_x}}\right) - 1. \quad (30)$$

Вважаючи тепер

$$\varepsilon_\beta = t_\beta \sigma_{\tilde{m}_x}, \quad (31)$$

для визначення t_β , як випливає з (29) – (31), одержуємо рівняння

$$\Phi(t_\beta) = (1 + \beta) / 2. \quad (32)$$

Корінь цього рівняння t_β можна знайти з таблиці Д1 для нормальної функції розподілу

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

Однак, з метою зручності в додатку наведена спеціальна таблиця

для t_β (см. табл. Д2).

Отже, обчисливши за дослідними даними $\tilde{m}_x, \tilde{D}_x$ і визначивши $\sigma_{\tilde{m}_x}$ з формули (29), а також t_β з таблиці Д2, знаходимо довірчий інтервал для математичного сподівання

$$I_\beta^{(m)} = \left(\tilde{m}_x - t_\beta \sigma_{\tilde{m}_x}, \tilde{m}_x + t_\beta \sigma_{\tilde{m}_x} \right) \quad (33)$$

Розрахунок довірчого інтервалу для дисперсії

Викладення, аналогічне тим, що були наведені нами при побудові довірчого інтервалу для математичного сподівання, дозволяють одержати довірчий інтервал для дисперсії:

$$I_\beta^{(D)} = \left(\tilde{D}_x - t_\beta \sigma_{\tilde{D}_x}, \tilde{D}_x + t_\beta \sigma_{\tilde{D}_x} \right), \quad (34)$$

де для нормального закону

$$\sigma_{\tilde{D}_x} = \sqrt{\frac{2}{N-1}} \tilde{D}_x, \quad (35)$$

для рівномірного розподілу

$$\sigma_{\tilde{D}_x} = \sqrt{\frac{0,8N+1,2}{N(N-1)}} \tilde{D}_x,$$

для показникового розподілу

$$\sigma_{\tilde{D}_x} = \sqrt{\frac{8N-6}{N(N-1)}} \tilde{D}_x,$$

а t_β , як і раніше, визначається з таблиці Д2.

Для середнього квадратичного відхилення $\sigma_x = \sqrt{D_x}$ довірчий інтервал, згідно (34), буде таким:

$$I_\beta^{(\sigma)} = \left(\sqrt{\tilde{D}_x - t_\beta \sigma_{\tilde{D}_x}}, \sqrt{\tilde{D}_x + t_\beta \sigma_{\tilde{D}_x}} \right). \quad (36)$$

Зауваження. Довірчі інтервали для математичного сподівання і дисперсії, що задаються формулами (33) і (34), є наближеними, тому що при їх одержанні невідома дисперсія D_x випадкової величини X замінювалася статистичною дисперсією $\tilde{D}_x$. Однак, якщо X розподілена нормально, то можна одержати точні співвідношення, що мають той же вигляд, що і (33), (34), однак у цих формулах параметр t_β визначається трохи інакше.

6. Вирівнювання статистичних розподілів

При обробці дослідних даних, отриманих у результаті спостережень величини X в серії дослідів, часто доводиться вирішувати питання про те, як на основі цих даних підібрати теоретичну криву розподілу, яку можна було б прийняти як передбачуваний закон розподілу величини X . Така задача називається задачею *вирівнювання (згладжування) статистичних розподілів*.

У загальному випадку для розв'язування цієї задачі, що полягає в аналітичному поданні емпіричних розподілів, користуються методом найменших квадратів, відповідно до якого вважається, що найкращим наближенням до емпіричних залежностей у даному класі функцій є таке, при якому сума квадратів відхилень перетворюється на мінімум.

У багатьох випадках теоретичний вид кривої розподілу вибирається відповідно до зовнішнього вигляду гістограми чи з фізичних міркувань. Наприклад, якщо X – це помилка вимірювального приладу або одна з координат точки влучення при стрільбі, то варто припустити, що теоретичним розподілом є нормальний закон. Якщо ж X – це час безвідмовної роботи якогось елемента системи або тривалість телефонної розмови, то варто припустити, що теоретичним розподілом цієї величини є показниковий закон. При цьому параметри передбачуваного теоретичного розподілу величини X природно вибирати так, щоб математичне сподівання і дисперсія цього розподілу були рівні відповідним статистичним характеристикам $\tilde{m}_x$, $\tilde{D}_x = \sigma_x^2$.

Так, наприклад, якщо гістограма значень величини X , що спостерігаються, має вигляд, показаний на рис. 2, то можна зробити припущення, що X розподілена за нормальним законом з щільністю

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2} = \frac{1}{\sigma} \varphi\left(\frac{x-m}{\sigma}\right),$$

де $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$. Параметри цього розподілу m і σ варто покласти рівними відповідним статистичним характеристикам:

$$m = \tilde{m}_x, \quad \sigma = \tilde{\sigma}_x.$$

Якщо гістограма має вигляд, показаний на рис. 5, то можна висловити припущення, що X розподілено з щільністю

$$f(x) = \begin{cases} \frac{1}{b-a}, & \forall a \leq x \leq b, \\ 0, & \forall x < a, \text{ або } x > b. \end{cases}$$

Оскільки для такого розподілу

$$m_x = \frac{a+b}{2}, \quad \sigma_x = \frac{b-a}{2\sqrt{3}},$$

то параметри a і b виражаються через m_x, σ_x за формулами

$$a = m_x - \sqrt{3}\sigma_x, \quad b = m_x + \sqrt{3}\sigma_x.$$

Тому для a і b варто використовувати оцінки

$$\tilde{a} = \tilde{m}_x - \sqrt{3}\tilde{\sigma}_x, \quad \tilde{b} = \tilde{m}_x + \sqrt{3}\tilde{\sigma}_x. \quad (37)$$

Якщо гістограма має вигляд, показаний на рис. 6, то

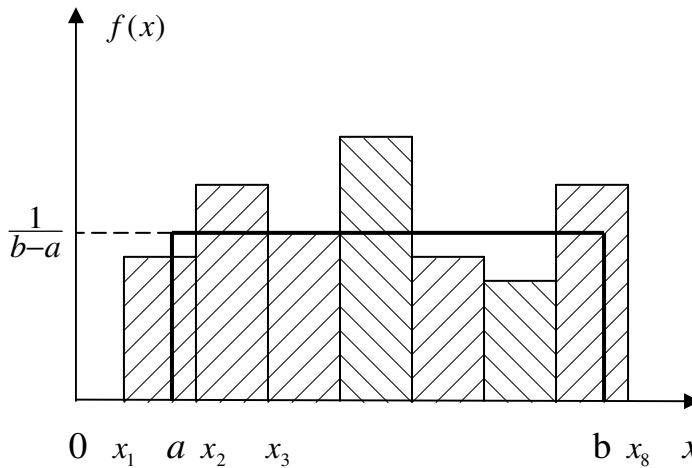


Рис. 5

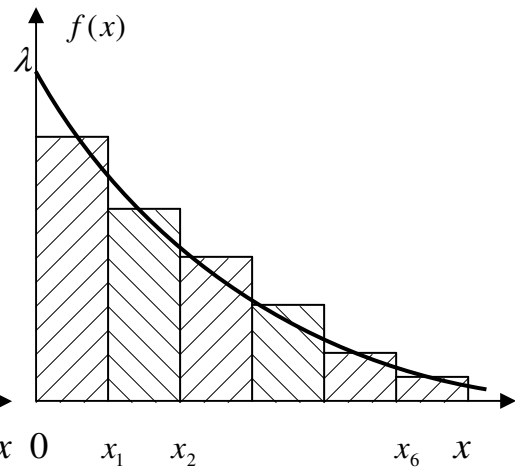


Рис. 6

можна припустити, що X розподілено за показниковим законом з щільністю

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \forall x \geq 0, \\ 0, & \forall x < 0. \end{cases}$$

Оскільки для такого розподілу

$$m_x = 1/\lambda,$$

то для параметра λ варто використовувати оцінку

$$\tilde{\lambda} = 1/\tilde{m}_x. \quad (38)$$

7. Перевірка гіпотези про закон розподілу випадкової величини

Після того, як висловлене припущення (гіпотеза) про теоретичний закон розподілу величини X , виникає питання про те, якою мірою воно погоджується зі статистичним розподілом, отриманим на основі спостережених значень цієї величини в серії незалежних дослідів. Ідея розв'язання поставленої задачі полягає в наступному.

Нехай H – це гіпотеза, яка полягає в тому, що величина X

підкоряється деякому певному закону розподілу і нехай U – деяка величина, що характеризує ступінь розбіжності теоретичного і статистичного розподілу, що виражається через результати спостережень X в серії з N дослідів і залежить від передбачуваного закону розподілу X (тобто від гіпотези H). Від істинного закону розподілу X випадкова величина U залежить побічно, через спостережені значення X в розглянутій серії дослідів.

Припустимо, що в результаті даної серії дослідів величина U прийняла значення u . Виникає питання: чи можна цю розбіжність пояснити випадковими причинами, чи вона вказує на наявність істотної різниці (непогодженості) між теоретичними і статистичними розподілами і, отже, на непридатність гіпотези H ? Для того, щоб відповісти на це питання, треба обчислити ймовірність того, що за рахунок випадкових причин, зв'язаних з недостатнім об'ємом дослідних даних, міра розбіжності U виявиться не меншою, ніж u , якщо гіпотеза H вірна, тобто обчислити ймовірність $P(U \geq u/H)$. Якщо ця ймовірність виявиться занадто малою, то гіпотезу H варто відкинути як мало правдоподібну; якщо ж ця ймовірність значна, то варто визнати, що експериментальні дані не суперечать гіпотезі H , тобто визнати правдоподібним наше припущення про теоретичний розподіл величини X .

Випадкова величина U , що грає роль міри розбіжності між теоретичним і статистичним розподілами, називається *критерієм згоди*. Існують різні критерії згоди. Одним з найбільш розповсюджених у практичних розрахунках є «критерій χ^2 » Пірсона, що має такий вигляд:

$$U = N \sum_{i=1}^n \frac{(p_i^* - p_i)^2}{p_i}, \quad (39)$$

де N – об'єм вибірки; n – число інтервалів, на які розбитий весь даний діапазон спостережених значень X ; p_i^* – відносна частота влучення спостережених значень X в i -й частковий інтервал (x_i, x_{i+1}) , а p_i – ймовірність влучення в цей інтервал величини X , якщо її істинний закон розподілу збігається з передбачуванням. Наприклад, якщо передбачуваний закон розподілу X заданий у вигляді функції розподілу $F(x)$, то

$$p_i = P(x_i < X < x_{i+1} / H) = F(x_{i+1}) - F(x_i). \quad (40)$$

Зокрема, якщо передбачуваний розподіл є нормальним, то

$$F(x) = \Phi\left(\frac{x - \tilde{m}_x}{\tilde{\sigma}_x}\right) \quad (41)$$

і формула (39) здобуває вигляд

$$p_i = \Phi(z_{i+1}) - \Phi(z_i), \quad (42)$$

де

$$z_i = (x_i - \tilde{m}_x) / \tilde{\sigma}_x. \quad (43)$$

Якщо ж передбачуваний закон розподілу є показниковим, то

$$F(x) = \begin{cases} 1 - e^{-x/\tilde{m}_x}, & \forall x \geq 0, \\ 0, & \forall x < 0 \end{cases}$$

і формула (40) здобуває вигляд

$$p_i = e^{-x_i/\tilde{m}_x} - e^{-x_{i+1}/\tilde{m}_x}. \quad (44)$$

Нарешті, якщо передбачуваний закон розподілу є рівномірним, то

$$F(x) = \begin{cases} 0, & \forall x < \tilde{a}, \\ \frac{x - \tilde{a}}{\tilde{b} - \tilde{a}}, & \forall \tilde{a} \leq x \leq \tilde{b}, \\ 1, & \forall x > \tilde{b} \end{cases}$$

і ймовірності p_i визначаються зі співвідношень

$$p_1 = \frac{x_2 - \tilde{a}}{\tilde{b} - \tilde{a}}, \quad p_i = \frac{x_{i+1} - x_i}{\tilde{b} - \tilde{a}}, \quad i=2,3,\dots,n-1, \quad p_n = \frac{\tilde{b} - x_n}{\tilde{b} - \tilde{a}}. \quad (45)$$

Випадкова величина U має такі властивості: при досить великому числі дослідів N її розподіл не залежить від функції розподілу $F(x)$ і від числа дослідів N , а залежить лише від числа *ступенів вільності* r , що визначається за формулою

$$r = n - s - 1, \quad (46)$$

де s – число параметрів розподілу $F(x)$. Зі збільшенням N розподіл величини U наближається до так званого χ^2 розподілу. У зв'язку з цим міру розбіжності теоретичного і статистичного розподілів, обумовлену співвідношенням (39), позначають χ^2 .

Якщо у формулі (39) ввести N під знак суми і врахувати, що $p_i^* = m_i / N$, де m_i – число спостережених значень величини X в i -му інтервалі, то ця формула приводиться до вигляду, більш зручному для розрахунків:

$$U = \chi^2 = \sum_{i=1}^n \frac{(m_i - Np_i)^2}{Np_i}. \quad (47)$$

Для розподілу χ^2 складена таблиця (див. табл. Д4), що дозволяє для даного числа ступенів вільності r визначити таке значення u , яке називається *критичною точкою для розподілу χ^2* , ймовірність перевищення якого дорівнює α , тобто таке u , для якого виконується умова

$$P(\chi^2 \geq u) = \alpha.$$

Ймовірність α вибирається відповідно до суті задачі і називається *рівнем значимості*.

Таким чином, обчислюючи дослідне значення χ^2 відповідно до формули (47) і визначаючи за табл. Д4 критичну точку u , що відповідає даним r і α , ми приходимо до одного з наступних висновків.

1. Якщо $\chi^2 < u$, то таке значення міри розбіжності теоретичного і статистичного розподілів варто визнати несуттєвим і, отже, на рівні значення α можна вважати, що передбачуваний закон розподілу величини X погоджується з дослідними даними.
2. Якщо $\chi^2 > u$, то таку величину міри розбіжності варто визнати значущою і тому на рівні значимості α зроблене припущення про закон розподілу величини X необхідно відкинути.

Приклад 2. Зроблено $N = 200$ спостережень випадкової величини X . Результати спостережень після первинної обробки зведені в розподіл вибірки, поданої у перших трьох рядках табл. 2.

Таблиця 2

Номер інтервалу i	1	2	3	4	5	6	7	8	9	10	11
Межі інтервалу x_i, x_{i+1}	(-50;-40)	(-40;-30)	(-30;-20)	(-20;-10)	(-10;0)	(0;10)	(10;20)	(20;30)	(30;40)	(40;50)	–
Чисельність інтервалу m_i	6	13	24	30	41	36	22	14	9	5	–
Відносні частоти p_i^*	0,030	0,065	0,120	0,150	0,205	0,180	0,110	0,070	0,045	0,025	–
Висоти h_i гістограми	0,0030	0,0065	0,0120	0,0150	0,0205	0,0180	0,0110	0,0070	0,0045	0,0025	–
$F^*(x_i)$	0	0,030	0,095	0,215	0,365	0,570	0,750	0,860	0,930	0,975	1

1. Підрахувати відносні частоти p_i^* для кожного з часткових інтервалів, побудувати гістограму й емпіричну функцію розподілу.

Висловити припущення про закон розподілу величини X .

2. Знайти статистичне математичне сподівання, дисперсію і середнє квадратичне відхилення.

3. Побудувати довірчі інтервали, що відповідають довірчій ймовірності $\beta = 0,95$ для математичного сподівання, дисперсії і середнього квадратичного відхилення величини X .

4. Вирівняти статистичний розподіл величини X . Перевірити за допомогою критерію Пірсона погодженість вирівняного (теоретичного) і статистичного розподілів на рівні значимості $\alpha = 0,3$. Визначити ймовірність $\hat{\alpha}$, з якою за рахунок чисто випадкових причин могла з'явитися розбіжність між гіпотетичним і статистичним розподілами більша, ніж отримана розбіжність χ^2 .

Розв'язок. 1. За формулою (6) підрахуємо відносні частоти p_i^* для кожного з інтервалів і заносимо в четвертий рядок табл. 2. Далі підраховуємо за формулою (7) висоти h_i прямокутників для побудови гістограми і результати розрахунку заносимо в п'ятий рядок табл. 2. За отриманими даними будуємо гістограму, що зображена на рис. 2. Проводимо приблизно неперервну криву, "що вирівнює" східчасту лінію, якою обмежена гістограма. За виглядом цієї кривої робимо припущення, що величина X розподілена за нормальним законом. Тепер за формулами (8) підраховуємо $F^*(x_i)$ і заносимо в шостий рядок табл. 2. Наближений графік емпіричної функції розподілу $F^*(x)$, побудований за результатами цього розрахунку, показаний на рис. 3.

2. Складаємо таблицю середин часткових інтервалів $\hat{x}_i$ і відповідних їм відносних частот

$$\tilde{m}_x \approx -1,8.$$

Таблиця 3

x_i	-45	-35	-25	-15	-5	5	15	25	35	45
p_i^*	0,030	0,065	0,120	0,150	0,205	0,180	0,110	0,070	0,045	0,025
x_i^2	2025	1225	625	225	25	25	225	625	1225	2025
$x_i^2 p_i$	60,75	79,62	75,00	33,75	5,125	4,500	24,75	43,75	55,13	50,63

Для обчислення $\tilde{D}_x$ зручно скористатися формулою (25). Проміжні розрахунки відповідно до цієї формули розміщуємо в табл. 3. У результаті знаходимо

$$\tilde{D}_x \approx 429,76, \quad \tilde{\sigma}_x = \sqrt{\tilde{D}_x} \approx 20,73. \quad (48)$$

3. З умови маємо: $N = 200$. Оцінки $\tilde{m}_x$, $\tilde{D}_x$ розглянутої випадкової величини були знайдені раніше:

$$\tilde{m}_x = -1,8; \quad \tilde{D}_x = 429,76.$$

Звідси за формулами (29) і (35) одержуємо: $\sigma_{\tilde{m}_x} = 1,47$; $\sigma_{\tilde{D}_x} = 43,08$.

З таблиці Д2 для $\beta = 0,95$ маємо $t_\beta = 1,96$. Таким чином, за формулами (33) і (34) знаходимо довірчі інтервали:

$$\text{для } m_x: \quad I_\beta^{(m)} = (-4,68; 1,08),$$

$$\text{для } D_x: \quad I_\beta^{(D)} = (345,32; 514,20),$$

$$\text{для } \sigma_x: \quad I_\beta^{(\sigma)} = (18,58; 22,67).$$

4. Відповідно до гістограми, отриманої вище, робимо припущення про те, що величина X розподілена нормально. Вважаючи $m_x = \tilde{m}_x = -1,8$, $\sigma_x = \tilde{\sigma}_x = 20,73$, де статистичні характеристики $\tilde{m}_x, \tilde{\sigma}_x$ знайдені в пункті 2, одержуємо теоретичну щільність розподілу

$$f(x) = \frac{1}{20,73\sqrt{2\pi}} e^{-\frac{(x+1,8)^2}{2 \cdot 20,73^2}} \quad (49)$$

$$\text{або} \quad f(x) = \frac{1}{20,73} \varphi(z), \quad z = \frac{x+1,8}{20,73},$$

де $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ – нормальна щільність розподілу, розрахунок якої даний у табл.Д3. Використовуючи табл. Д1 і Д2, підрахуємо значення $f(x) = \varphi(z)/20,73$ і $F(x) = \Phi(z)$ на кінцях інтервалів x_i і разом із проміжними обчисленнями величин $z_i = (x_i + 1,8)/20,73$ результати зводимо в табл. 4.

Результати розрахунку відкладаємо на рис.2 і проводимо через розрахункові точки плавну криву.

Обчислюємо теоретичні ймовірності p_i влучення X в часткові інтервали (x_i, x_{i+1}) відповідно до формули (42) і результати теж заносимо в табл.4.

Таблиця 4

x_i	-50	-40	-30	-20	-10	0	10	20	30	40	50
z_i	-2,33	-1,84	-1,36	-0,88	-0,39	0,09	0,57	1,05	1,53	2,02	2,5
$f(x_i) =$ $= \frac{1}{20,73} \varphi(z_i)$	0,0013	0,0036	0,0079	0,0134	0,0184	0,0198	0,0169	0,0110	0,0064	0,0025	0,0008
$F(x_i) = \Phi(z_i)$	0,0102	0,0329	0,0869	0,1894	0,3483	0,5359	0,7157	0,8531	0,9370	0,9782	0,9938
p_i	0,0227	0,0540	0,1025	0,1589	0,1876	0,1798	0,1374	0,0839	0,0512	0,0156	

Тепер розраховуємо за формулою (46) значення χ^2 , розміщуючи проміжні обчислення в табл.5. Звідси знаходимо $\chi^2 = 4,54$. сигналу 2 потім,

Таблиця 5

i	1	2	3	4	5	6	7	8	9	10
m_i	6	13	24	30	41	36	22	14	9	5
Np_i	4,54	10,8	20,5	31,78	37,52	35,96	27,18	16,78	10,24	3,12
$m_i - Np_i$	-1,46	2,20	3,50	-1,78	3,48	0,04	-5,18	-2,78	-0,76	1,88
$(m_i - Np_i)^2$	2,13	4,84	12,25	3,17	12,11	0,00	26,83	7,73	0,60	3,53
$(m_i - Np_i)^2 / Np_i$	0,77	0,45	0,60	0,10	0,32	0,00	0,95	0,46	0,06	1,13

Обчисливши тепер число ступенів вільності за формулою (46)

$$r = 10 - 2 - 1 = 7$$

і враховуючи, що $\alpha = 0,3$, за табл. Д4 знаходимо критичну точку розподілу χ^2 :

$$u = u(0,3; 7) = 8,38.$$

Порівнюючи значення χ^2 з u , $\chi^2 = 4,54 < u = 8,38$, робимо висновок, що на рівні значимості $\alpha = 0,3$ наше припущення про те, що випадкова величина X має нормальний розподіл, що характеризується щільністю (49), не суперечить дослідним даним.

Ймовірність $\hat{\alpha}$ того, що розбіжність між нормальним розподілом (49) і статистичним розподілом буде за рахунок чисто випадкових причин

більшу за $\chi^2 = 8,38$, знаходимо з табл. Д4, враховуючи $u(\alpha; 7) = 8,38$; $\hat{\alpha} = 0,7$. Результати остаточних розрахунків оформляємо у вигляді зведеної табл. 6.

Таблиця 6

i	1	2	3	4	5	6	7	8	9	10	11
Інтервали (x_i, x_{i+1})	(-50; -40)	(-40; -30)	(-30; -20)	(-20; -10)	(-10; 0)	(0; 10)	(10; 20)	(20; 30)	(30; 40)	(40; 50)	–
Численність інтервалів m_i	6	13	24	30	41	36	22	14	9	5	–
Відносна частота p_i^*	0,030	0,065	0,120	0,150	0,205	0,180	0,110	0,070	0,045	0,025	–
$F^*(x_i)$	0	0,030	0,095	0,215	0,365	0,570	0,750	0,860	0,930	0,975	1
$f(x_i)$	0,0013	0,0036	0,0079	0,0134	0,0184	0,0198	0,0169	0,0110	0,0064	0,0025	0,0008
$F(x_i)$	0,0102	0,0329	0,0829	0,1894	0,3483	0,5359	0,7157	0,8531	0,9370	0,9782	0,9938
p_i	0,0227	0,0540	0,1025	0,1589	0,1876	0,1798	0,1374	0,0839	0,0512	0,0156	–

$\tilde{m}_x$	$\tilde{D}_x$	$\tilde{\sigma}_x$	$\sigma_{\tilde{m}}$	$\sigma_{\tilde{D}}$	t_β	$I_{\mathcal{B}}^{(m)}$	$I_{\mathcal{B}}^{(D)}$	$I_{\mathcal{B}}^{(\sigma)}$	x^2	r	$u(\alpha, r)$	$\hat{\alpha}$
-1,80	429,76	20,73	1,47	43,08	1,96	(-4,68; 1,08)	(345,32; 514,20)	(18,58; 22,67)	4,54	7	8,38	0,7

$$(\alpha = 0,3; \quad \beta = 0,95)$$

II. Комплексне завдання

Варіанти 1-25

Зроблено $N = 200$ незалежних дослідів, у яких за допомогою еталонної напруги v_0 калібрувалось шкала вольтметра. Результати дослідів (показання на шкалі вольтметра), поданої в табл.7, є спостереженими значеннями випадкової величини V . Помилка вимірів

$X = V - v_0$ також випадкова величина. Її математичне сподівання m називається систематичною помилкою приладу.

Скласти таблицю спостережених значень помилки вимірів X і зробити наступну обробку цих дослідних даних.

1. Результат первинної обробки дослідних даних подати у вигляді розподілу вибірки.

2. Виконати розрахунок і графіки гістограми та емпіричної функції розподілу.

3. Знайти оцінки для математичного сподівання, дисперсії і середнього квадратичного відхилення.

4. Зробити розрахунок довірчих інтервалів для математичного сподівання, дисперсії і середнього квадратичного відхилення для довірчої ймовірності β .

5. Виписати формули для функції і щільності передбачуваного закону розподілу випадкової величини X , отримані в результаті вирівнювання статистичного розподілу.

6. Перевірити гіпотези про теоретичний розподіл випадкової величини X за допомогою критерію Пірсона на рівні значимості α .

7. Скласти зведену таблицю результатів основних розрахунків.

Значення параметрів $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6$, що фігурують у табл. 7, як і інших вихідних величин еталонної напруги v_0 , довірчої ймовірності β і рівня значимості α , подані для відповідних варіантів у табл. 8. Число часткових інтервалів n вибрати рівним 10.

Варіанти 26-50

Зроблено $N = 200$ вимірів часу роботи одного з радіотехнічних елементів. Результати вимірів представлені в табл.9.

Зробити наступну обробку цих дослідних даних.

1. Результат первинної обробки дослідних даних подати у вигляді розподілу вибірки.

2. Виконати розрахунок і графіки гістограми та емпіричної функції розподілу.

3. Знайти оцінки для математичного сподівання, дисперсії і середнього квадратичного відхилення.

4. Зробити розрахунок довірчих інтервалів для математичного сподівання, дисперсії і середнього квадратичного відхилення для довірчої ймовірності β .

5. Виписати формули для функції і щільності передбачуваного закону розподілу випадкової величини X , отримані в результаті вирівнювання статистичного розподілу.

6. Перевірити гіпотези про теоретичний розподіл випадкової величини X за допомогою критерію Пірсона на рівні значимості α .

7. Скласти зведену таблицю результатів основних розрахунків.

Значення параметрів $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7$, що фігурують у табл.9, як і інших вихідних величин – довірчої ймовірності β і рівня значимості α , подані для відповідних варіантів у табл.10.

Число часткових інтервалів вибрати рівним 10.

Варіанти 51-75

Зроблено $N = 200$ вимірів різниці фаз випадкового й опорного сигналу в різні моменти часу. Результати вимірів представлені в табл.11.

Зробити наступну обробку цих дослідних даних.

1. Результат первинної обробки дослідних даних подати у вигляді розподілу вибірки.

2. Виконати розрахунок і графіки гістограми й емпіричної функції розподілу.

3. Знайти оцінки для математичного сподівання, дисперсії і середнього квадратичного відхилення.

4. Зробити розрахунок довірчих інтервалів для математичного сподівання, дисперсії і середнього квадратичного відхилення для довірчої ймовірності β .

5. Виписати формули для функції і щільності передбачуваного закону розподілу випадкової величини X , отримані в результаті вирівнювання статистичного розподілу.

6. Перевірити гіпотези про теоретичний розподіл випадкової величини X за допомогою критерію Пірсона на рівні значимості α .

7. Скласти зведену таблицю результатів основних розрахунків.

Значення параметрів $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7$, що фігурують у табл.10, як і

інших вихідних величин – довірчої ймовірності β і рівня значимості α , подані для відповідних варіантів у табл.12. Число часткових інтервалів вибрати рівним 10.

Вказівки до виконання комплексного завдання

1. Скласти таблицю спостережених значень величини X .
2. Визначити найменше $x_{\min}$ і найбільше $x_{\max}$ спостережені значення величини X . Проміжок $(x_{\min}, x_{\max})$ розбити на $n = 10$ часткових інтервалів (x_i, x_{i+1}) , $i=1,2,\dots,10$, $x_1=x_{\min}$, $x_{11}=x_{\max}$.
3. Визначити число спостережень m_i величини X у кожному з часткових інтервалів і за формулою (6) обчислити відносні частоти p_i^* . Скласти таблицю розподілу вибірки (див.табл.1).
4. Розрахувати гістограму та обчислити значення емпіричної функції розподілу $F^*(x)$ на кінцях часткових інтервалів x_i за формулою (8). Результати звести в таблицю (див.табл.2), побудувати відповідні графіки.
5. За формулами (23), (25), і (17) обчислити оцінки математичного сподівання $\tilde{m}_x$, дисперсії $\tilde{D}_x$ і середнього квадратичного відхилення $\tilde{\sigma}_x$. Проміжні обчислення занести в таблицю (див. табл. 3).
6. За формулами (29) і (35) обчислити $\sigma_{\tilde{m}}$ і $\sigma_{\tilde{D}}$. З табл. ДЗ для даної довірчої ймовірності визначити значення параметра t_β . За формулами (33), (34) і (36) побудувати довірчі інтервали $I_\beta^{(m)}$, $I_\beta^{(D)}$, $I_\beta^{(\sigma)}$.
7. Виписати формули для щільності $f(x)$ і функції $F(x)$ передбачуваного розподілу величини X , обчисливши попередньо оцінки параметрів цього розподілу. Зокрема, якщо передбачуваний закон – нормальний, то

$$f(x) = \frac{1}{\tilde{\sigma}_x} \varphi\left(\frac{x - \tilde{m}_x}{\tilde{\sigma}_x}\right), \quad F(x) = \Phi\left(\frac{x - \tilde{m}_x}{\tilde{\sigma}_x}\right).$$

Якщо ж мова йде про рівномірний розподіл, то

$$f(x) = \begin{cases} \frac{1}{\tilde{b} - \tilde{a}}, & \forall \tilde{a} \leq x \leq \tilde{b}, \\ 0, & \forall x < \tilde{a}, x > \tilde{b}, \end{cases}, \quad F(x) = \begin{cases} 0, & \forall x < \tilde{a}, \\ \frac{x - \tilde{a}}{\tilde{b} - \tilde{a}}, & \forall \tilde{a} \leq x \leq \tilde{b}, \\ 1, & \forall x > \tilde{b}, \end{cases}$$

де параметри $\tilde{a}$ і $\tilde{b}$ визначаються через знайдені раніше оцінки $\tilde{m}_x$ і $\tilde{\sigma}_x$ за формулами (37).

Для показникового розподілу маємо:

$$f(x) = \begin{cases} \tilde{\lambda} e^{-\tilde{\lambda}x}, & \forall x \geq 0, \\ 0, & \forall x < 0. \end{cases} \quad F(x) = \begin{cases} 1 - e^{-\tilde{\lambda}x}, & \forall x \geq 0, \\ 0, & \forall x < 0. \end{cases}$$

де параметр $\tilde{\lambda}$ визначається за формулою (38).

Розрахувати значення $f(x)$ і $F(x)$ на кінцях часткових інтервалів. Графік $f(x)$ будується на рисунку, де подана гістограма, а графік $F(x)$ – на рисунку, де подана емпірична функція розподілу $F^*(x)$.

8. За формулою (40) обчислити теоретичні ймовірності p_i попадання величини X у часткові інтервали.

Результати розрахунків за пп.7 і 8 зводяться в одну таблицю (див.табл.4.)

9. За відомими m_i , p_i і N обчислити значення χ^2 (див.табл.5).

Визначити число ступенів вільності r за формулою (41) і знайти критичну точку $u = u(\alpha; r)$ розподілу χ^2 за табл. Д4.

10. Обчислити ймовірність $\hat{\alpha}$, з якою за рахунок чисто випадкових причин могла б з'явитися розбіжність $u(\hat{\alpha}, r)$ між гіпотетичним і статистичним розподілами більша, ніж обчислене значення χ^2 . З цією метою з табл. Д4 за відомими r і $u = \chi^2$ ($u = u(\hat{\alpha}; r)$) знайти відповідні значення ймовірності $\hat{\alpha}$.

11. Оформити остаточні результати розрахунків у вигляді зведеної таблиці (див.табл.6).

Примітка. Проміжні розрахунки виконуються з точністю, необхідною для отримання остаточних результатів, – до третьої значущої цифри.

Таблиця 7

Таблиця 8

Таблиця 9

Таблиця 10

Таблиця 11

Таблиця 12

Таблиця Д1

Таблиця Д1(прод)

Таблиця Д2

Значення t_β

β	0,80	0,81	0,82	0,83	0,84	0,85
t_β	1,282	1,310	1,340	1,371	1,404	1,439
β	0,86	0,87	0,88	0,89	0,90	
t_β	1,475	1,513	1,554	1,597	1,643	
β	0,91	0,92	0,93	0,94	0,95	0,96
t_β	1,694	1,750	1,810	1,880	1,960	2,053
β	0,97	0,98	0,99	0,9973	0,999	
t_β	2,169	2,325	2,576	3,000	3,290	

Таблиця Д3

Значення функції Гаусса $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

x	0	1	2	3	4	5	6	7	8	9	x
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973	0,0
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918	0,1
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825	0,2
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697	0,3
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538	0,4
0,5	0,3521	3503	3485	3467	3448	3429	3410	3391	3372	3352	0,5
0,6	3332	3312	3292	3271	3251	3230	3209	3387	3166	3144	0,6
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920	0,7
0,8	2897	2874	2850	2827	2903	2780	2757	2732	2709	2685	0,8
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444	0,9
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203	1,0
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965	1,1
1,2	1942	1919	1876	1872	1849	1826	1804	1781	1758	1736	1,2
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518	1,3
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315	1,4
1,5	0,1295	1276	1257	1238	1219	1200	1182	1163	1145	1127	1,5
1,6	1109	1092	1074	1057	1040	1023	1006	0983	0973	0957	1,6
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804	1,7
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669	1,8
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551	1,9
2,0	0,054	0519	0519	0508	0498	0488	0478	0468	0459	0449	2,0
2,1	0440	0431	0422	0413	0404	0396	0388	0379	0371	0363	2,1
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290	2,2
2,3	0283	0277	0270	0264	0258	0252	0264	0241	0235	0229	2,3
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180	2,4

Таблиця ДЗ(прод)

Таблиця Д4

Критичні точки $u = u(\alpha, r)$ розподілу χ^2

r	α						
	0,99	0,98	0,95	0,90	0,80	0,70	0,50
1	0,000	0,001	0,004	0,016	0,064	0,148	0,455
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386
3	0,115	0,185	0,352	0,584	1,005	1,424	2,37
4	0,297	0,429	0,711	1,064	1,649	2,20	3,36
5	0,554	0,752	1,145	1,61	2,34	3,00	4,35
6	0,872	1,134	1,635	2,20	3,07	3,83	5,35
7	1,239	1,564	2,17	3,83	3,82	4,67	6,35
8	1,646	2,03	2,73	3,49	4,59	5,53	7,34
9	2,09	2,53	3,32	4,17	5,38	6,39	8,34
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34
11	3,05	3,61	4,58	5,58	6,99	8,15	10,34
12	3,57	4,18	5,23	6,30	7,81	9,03	11,34
13	4,11	4,76	5,89	7,04	8,63	9,93	12,34
14	4,66	5,37	6,57	7,79	9,47	10,82	13,34
15	5,23	5,98	7,26	8,55	10,31	11,72	14,34
16	5,81	6,61	7,96	9,31	11,15	12,62	15,34
17	6,41	7,26	8,67	10,08	12,00	13,53	16,34
18	7,02	7,91	9,39	10,86	12,86	14,44	17,34
19	7,63	8,57	10,11	11,65	13,72	15,35	18,34
20	8,26	9,24	10,58	12,44	14,58	16,27	19,34
21	8,90	9,92	11,59	13,24	15,44	17,18	20,3
22	9,54	10,60	12,34	14,04	16,31	18,10	21,3
23	10,20	11,29	13,09	14,85	17,19	19,02	22,3
24	10,86	11,99	13,85	15,66	18,06	19,94	23,3
25	11,52	12,70	14,61	16,47	18,94	20,9	24,3
26	12,20	13,41	15,38	17,29	19,82	21,8	25,3
27	12,88	14,12	16,15	18,11	20,7	22,7	26,3
28	13,56	14,85	16,93	18,94	21,6	23,6	27,3
29	14,26	15,57	17,71	19,77	22,5	24,6	28,3
30	14,95	16,31	18,49	20,6	23,4	25,5	29,3

Таблиця Д4 (продовження)

r	α						
	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	1,074	1,642	2,71	3,84	5,41	6,64	10,83
2	2,41	3,22	4,60	5,99	7,82	9,21	13,82
3	3,66	4,64	6,25	7,82	9,84	11,34	16,27
4	4,88	5,99	7,78	9,49	11,67	13,28	18,46
5	6,06	7,29	9,24	11,07	13,39	15,09	20,5
6	7,23	8,56	10,64	12,59	15,03	16,81	22,5
7	8,38	9,80	12,02	14,07	16,62	18,48	24,3
8	9,52	11,03	13,36	15,51	18,17	20,1	26,1
9	10,66	12,24	14,68	16,92	19,68	21,7	27,9
10	11,78	13,44	15,99	18,31	21,2	23,2	29,6
11	12,90	14,63	17,28	19,68	22,6	24,7	31,3
12	14,01	15,81	18,55	21,0	24,1	26,2	32,9
13	15,12	16,98	19,81	22,4	25,5	27,7	34,6
14	16,22	18,15	21,1	23,7	26,9	29,1	36,1
15	17,32	19,31	22,3	25,0	28,3	30,6	37,7
16	18,42	20,5	23,5	26,3	29,6	32,0	39,3
17	19,51	21,6	24,8	27,6	31,0	33,4	40,8
18	20,6	22,8	26,0	28,9	32,3	34,8	42,3
19	21,7	23,9	27,2	30,1	33,7	36,2	43,8
20	22,8	25,0	28,4	31,4	35,0	37,6	45,3
21	23,9	26,2	29,6	32,7	36,3	38,9	46,8
22	24,9	27,3	30,8	33,9	37,7	40,9	48,3
23	26,0	28,4	32,0	35,2	39,0	41,6	49,7
24	27,1	29,6	33,2	36,4	40,3	43,0	51,2
25	28,2	30,7	34,4	37,7	41,7	44,3	52,6
26	29,2	31,8	35,6	38,9	42,9	45,6	54,1
27	30,3	32,9	36,7	40,1	44,1	47,0	55,5
28	31,4	34,0	37,9	41,3	45,4	48,3	56,9
29	32,5	35,1	39,1	42,6	46,7	49,6	58,3
30	33,5	36,2	40,3	43,8	48,0	50,9	59,7

Таблиця Д5

Значення функції e^{-x}

x	e^{-x}	x	e^{-x}	x	e^{-x}	x	e^{-x}
0,00	1,000	0,40	0,670	0,80	0,449	3,00	0,050
01	990	41	664	81	445	3,10	40
02	980	42	657	82	440	3,20	41
03	970	43	650	83	436	3,30	37
04	961	44	644	84	432	3,40	33
05	951	45	638	85	427	3,50	30
06	942	46	631	86	423	3,60	27
07	932	47	625	87	419	3,70	25
08	923	48	619	88	415	3,80	22
09	914	49	613	89	411	3,90	20
0,10	0,905	0,50	0,606	0,90	0,407	4,00	0,0183
11	896	51	600	91	403	4,10	166
12	887	52	595	92	399	4,20	150
13	878	53	589	93	395	4,30	136
14	869	54	583	94	391	4,40	123
15	861	55	577	95	387	4,50	111
16	852	56	571	96	383	4,60	101
17	844	57	565	97	379	4,70	0,0091
18	835	58	560	98	375	4,80	82
19	827	59	554	99	372	4,90	74
0,20	0,819	0,60	0,549	1,00	0,368	5,00	0,0067
21	811	61	543	1,10	333	5,10	61
22	803	62	538	1,20	302	5,20	55
23	795	63	533	1,30	273	5,30	50
24	787	64	527	1,40	247	5,40	45
25	779	65	522	1,50	223	5,50	41
26	771	66	517	1,60	202	5,60	37
27	763	67	512	1,70	183	5,70	33
28	756	68	507	1,80	165	5,80	30
29	748	69	502	1,90	150	5,90	27
0,30	0,741	0,70	0,497	2,00	0,135	6,00	0,0025
31	733	71	492	2,10	122	6,10	22
32	726	72	487	2,20	111	6,20	20
33	719	73	482	2,30	100	6,30	18
34	712	74	477	2,40	0,091	6,40	17
35	705	75	472	2,50	82	6,50	15
36	698	76	468	2,60	74	6,60	14
37	691	77	463	2,70	67	6,70	12
38	684	78	458	2,80	61	6,80	11
39	677	79	454	2,90	55	6,90	10
0,40	0,670	0,80	0,449	3,00	0,050	7,00	0,0009

Таблиця Д6

Таблиця Д6(прод)

ЗМІСТ

I. Стислі теоретичні відомості з обробки дослідних даних.....	3
1. Постановка задач з обробки дослідних даних.....	3
2. Статистичні розподіли.....	3
3. Оцінка невідомої числової характеристики закону розподілу.....	7
4. Числові характеристики статистичного розподілу.....	9
5. Довірчий інтервал. Довірча ймовірність.....	11
6. Вирівнювання статистичних розподілів.....	13
7. Перевірка гіпотези про закон розподілу випадкової величини.....	14
II. Комплексне завдання.....	20
III. Додаток. Ймовірнісні таблиці.....	29

Методичні вказівки та варіанти комплексного завдання з елементів
математичної статистики для студентів другого курсу всіх спеціальностей

Дослідження статистичних характеристик випадкових величин

доц. ***І.В. Стрелковська***, доц. ***Л.І. Соколов***,
доц. ***О.П. Яковчук***, викл. ***Т.І. Григор'єва***

Рецензент	<i>В.Є. Горбачов</i>
Редактор	<i>Л.І. Соколов</i>
Комп'ютерна правка та макетування	<i>Є. С. Корнійчук</i>