

Міністерство освіти і науки України

Одеський державний університет інтелектуальних технологій і зв'язку

Кафедра вищої математики

Вища математика та математична статистика

Навчально-наочний посібник
для студентів очної та дистанційної форм навчання

Частина 4

Одеса
2023

Рецензент: к.т.н., доц. Зборовська І.А.

Вища математика та математична статистика: Навчально-наочний посібник для студентів очної та дистанційної форм навчання. Частина 4 / Укладачі: О.В.Лінкова, А.І.Гарбуз, Одеса: ДУІТЗ, 2023. 32 с.

Навчально-наочний посібник містить теоретичні матеріали з дисципліни “Вища математика та математична статистика” та “Теорія ймовірностей та математична статистика” відповідно до навчальних програм дисциплін. Матеріали надано у вигляді таблиць, що допоможе тим, хто навчається, краще засвоїти теоретичний матеріал та наочно побачити методи розв’язання практичних задач та їх ілюстрації, краще зрозуміти алгоритми.

Навчально-наочний посібник призначений для студентів усіх спеціальностей та може використовуватися при проведенні лекційних, практичних занять з дисциплін математичного циклу, а також при виконанні студентами розрахунково-графічних робіт та при підготовці до заліків та екзаменів.

Видання містить ілюстративно-наочні матеріали, які сприяють вивченню і викладанню навчальної дисципліни математичного циклу, засвоєнню їх змісту.

Особливо корисним даний посібник може бути при підготовці студентів зі всіх видів занять та застосовуватися ними під час самостійних занять.

СХВАЛЕНО

на засіданні кафедри
Вищої математики
і рекомендовано до друку
Протокол № 6 від 27.01.2023 р.

ЗАТВЕРДЖЕНО
методичною радою
ДУІТЗ
Протокол № 2 від 08.02.2023 р.

Лінкова О.В.,
Гарбуз А.І., 2023

ДУІТЗ, 2023

ВСТУП

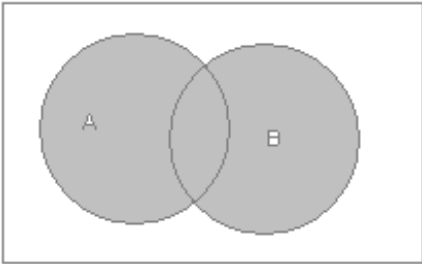
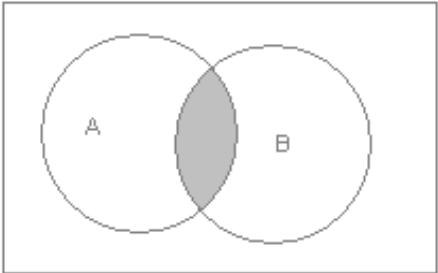
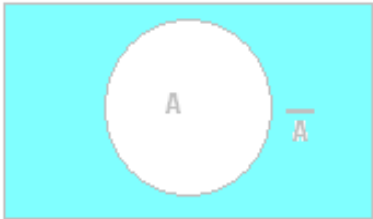
Пізнавальна діяльність студентів по оволодінню знаннями з будь-яких дисциплін полягає в вивченні основних теоретичних положень дисципліни; в можливості застосування цих положень до розв'язання практичних завдань (типових, дослідних, тощо). Таким чином вміння роз'язувати вище перелічені задачі, завдяки отриманим теоретичним знанням, може бути здійснено тільки за допомогою самостійних занять.

В сучасних реаліях життя, де дистанційна форма навчання набуває стрімке поширення, стає актуальною підтримка студентів під час їх самостійного роботи.

Тому, саме оснащена відповідними методичними матеріалами, самостійна робота в вищих навчальних закладах є ефективним напрямком підвищення якості всієї освіти.

Навчально-наочний посібник відповідає навчальній програмі з дисципліни, змісту та методиці викладання курсу і є частиною навчально-методичного забезпечення в першу чергу самостійної роботи студентів, а також і усіх видів занять як під керівництвом викладача, та й при дистанційній роботі студентів.

Навчально-наочний посібник являє собою поданий у логічному вигляді (у вигляді схем, графіків, алгоритмів розв'язування задач) матеріал курсу за окремими модулями (темами). Цей матеріал – основа знань з кожної теми, що наочно та послідовно ілюструє матеріал лекцій і допомагає студентам краще засвоїти інформацію, запам'ятати її, надає навчання творчих рис

1. ВИПАДКОВІ ПОДІЇ		
Випадкова подія – це подія, яка може відбутися або не відбутися при виконанні певного комплексу умов		
1.1 Види подій		
Достовірна – подія, яка в даному досліді і обов’язково відбудеться.		
Неможлива – подія, яка в даному досліді не може відбутися за будь-яких умов.		
Протилежні події – це події A та $\bar{A}$, настання однієї з яких в даному досліді означає ненастання іншої.		
Випадкові – це події, які можуть відбутися або не відбутися в даному досліді.		
Дії над подіями		
 $A + B$  $A \cdot B$  $\bar{A}$		
1.2 Елементи комбінаторики		
Комбінаторика – розділ математики, що вивчає задачі про розташування або вибір елементів із множини.		
Нехай задано множину $X = \{x_1; x_2; \dots; x_n\}$, що складається з n елементів.		
Види комбінацій		
Назва		Кількість
Перестановкою (множини X) з n елементів називають будь-яку впорядковану множину з усіх цих елементів, причому дві такі множини вважають різними, якщо вони відрізняються між собою порядком розташування елементів.		$P_n = n!$
Розміщенням (множини X) з n елементів по m ($m \leq n$) називають будь-яку впорядковану підмножину Y множини X , причому дві такі підмножини вважають різними, якщо вони відрізняються складом або порядком розташування елементів.		$A_n^m = \frac{n!}{(n-m)!}$
Комбінацією (множини X) з n елементів по m ($m \leq n$) називають будь-яку підмножину Y множини X , причому дві такі підмножини вважають різними, якщо вони відрізняються складом елементів		$C_n^m = \frac{n!}{m!(n-m)!}$
Приклади		
Для проведення дослідів вибрано 5 деталей. Скількома способами вони можуть бути	З 25 студентів обирають старосту та члена студентської ради. Яке число усіх можливих варіантів такого вибору?	Для участі у конференції з групи в 25 студентів обирають 3-х студентів. Яке число усіх можливих варіантів вибору такої трійки?

розподілені між п'ятьма дослідниками? Розв'язок. Число способів дорівнює числу комбінацій з 5 елементів по 5, які відрізняються тільки порядком їх розташування та є перестановками, тому $P_5 = 5! = 120$	Розв'язок. Такі комбінації з 25 елементів по 2 елементи є розміщеннями, тому що в них важливий не тільки склад елементів, що входять до комбінації, а й розташування елементів, тому $A_{25}^2 = \frac{25!}{(25-2)!} = \frac{25!}{23!} = 24 \cdot 25 = 600$	Розв'язок. Число можливих варіантів дорівнює числу комбінацій з 25 елементів по 3 елементи, причому комбінації будуть відрізнятися тільки їх елементами, бо порядок їх розташування немає значення, тому $C_{25}^3 = \frac{25!}{2!(25-3)!} = \frac{25!}{2! \cdot 22!} = 2300$
---	---	---

1.3 Означення ймовірності події

Класичне	Геометричне	Статистичне
$P(A) = \frac{m}{n}$, де m – число можливих подій, що сприяють події A ; n – число усіх можливих подій	$P(A) = \frac{g_d}{G_D}$, де g_d – площа області, об'єм, довжина (d); G_D – площа області, об'єм, довжина (D)	$P(A) = \frac{m}{n}$, де m – число дослідів, в яких з'явилася подія A ; n – загальне число дослідів; $\frac{m}{n}$ – відносна частота події A

Приклади

На деякому складі знаходяться 60 деталей, які виготовлені трьома бригадами. З них 25 деталей виготовлені 1-ю бригадою, 21 – 2-ю, 14 – 3-ю бригадою. Знайти ймовірність того, що до цеху надійшла деталь, яку виготовлено другою бригадою. Розв'язок: $n = 60$; $m = 21 \Rightarrow$ $P(A) = \frac{21}{60} = \frac{7}{20}$	На площині розташовані два концентричні кола з радіусами $r = 5$ м та $R = 10$ м. Знайти ймовірність того, що точка, яку кинуто навмання в більше коло, потрапить до кільця. Розв'язок: $P(A) = \frac{S_r}{S_R} = \frac{\pi r^2}{\pi R^2} = \frac{25}{100} = 0,25$	При випробуванні на надійність з 80 виробів 2 вироби відмовили протягом 50 годин роботи. Яка ймовірність безвідмовної роботи виробів протягом 50 годин. Розв'язок: Кількість виробів, які працювали до кінця дослідів складає $80 - 2 = 78$, $m = 78$. Загальна кількість дослідів $n = 80$. Тоді частість дорівнює $P(A) = \frac{m}{n} = \frac{78}{80} \approx 0,92$
--	---	--

1.4 Властивості ймовірності

1. Для достовірної події A : $P(A) = 1$.	2. Для неможливої події A : $P(A) = 0$.
3. $0 \leq P(A) \leq 1$.	4. $P(A) + P(\bar{A}) = 1$.
5. $q = P(\bar{A}) = 1 - P(A) = 1 - p$.	6. Якщо $\{A_1, A_2, \dots, A_n\} = E$, то $P(A_1) + P(A_2) + \dots + P(A_n) = 1$.

1.5 Основні теореми

1. Теорема множення: ймовірність добутку двох подій A та B дорівнює ймовірності однієї з них на умовну ймовірність іншої, обчислену за умови, що перша подія мала місце: $P(A \cdot B) = P(A)P(B/A) = P(B)P(A/B)$. Якщо події A та B незалежні, то $P(A/B) = P(A)$; $P(B/A) = P(B)$; $P(A \cdot B) = P(A) \cdot P(B)$.	Приклад: В партії, що містить 4% браку, 75% стандартних виробів відносяться до 1-го гатунку. Яка ймовірність того, що випадково	Розв'язок: Подія A – вибраний виріб стандартний; подія B – вибраний виріб 1-го гатунку;
---	--	---

вибраний виріб буде 1-го гатунку.	подія C – вибраний виріб стандартний та 1-го гатунку. $P(A) = 1 - P(\bar{A}) = 1 - 0,04 = 0,96$ $C = A \cdot B$ $P(C) = P(A \cdot B) = P(A)P(B/A) = 0,96 \cdot 0,75 = 0,72$
2. Теорема додавання: Ймовірність суми сумісних подій A та B дорівнює сумі їх ймовірностей мінус ймовірність їх добутку: $P(A + B) = P(A) + P(B) - P(A \cdot B)$ Якщо події A та B несумісні, то $P(A \cdot B) = 0$ та $P(A + B) = P(A) + P(B)$	
Приклад: З урни, що містить 10 куль з номерами 1,2,...,10 навмання виймається куля. Яка ймовірність того, що її номер кратний 2 або 3.	Розв'язок: Подія A_1 – номер кулі кратний 2; подія A_2 – номер кулі кратний 3; подія $A = A_1 + A_2$ – номер кулі кратний 2 або 3; подія $A_1 \cdot A_2$ – номер кулі кратний 2 та 3 одночасно. Події A_1 та A_2 сумісні, тому $P(A) = P(A_1) + P(A_2) - P(A_1 \cdot A_2)$ Оскільки $P(A_1) = \frac{4}{10}$; $P(A_2) = \frac{3}{10}$, $P(A_1 \cdot A_2) = \frac{1}{10}$, тоді $P(A) = \frac{4}{10} + \frac{3}{10} - \frac{1}{10} = \frac{6}{10}.$
Формула повної ймовірності Ймовірність події A, яка відбувається разом з однією з гіпотез $H_1, H_2, \dots, H_n$, дорівнює $P(A) = \sum_{i=1}^n P(H_i)P(A/H_i)$	
Приклад: На трьох верстатах випускаються однакові деталі. Денна потужність 1-го верстату складає 6000 шт., 2-го – 1000 шт., 3-го – 3000 шт. На 1-му верстаті брак складає 8 %, на 2-му – 2 %, на 3-му – 5 %. Деталі, що надходять до складу, змішуються. Яка ймовірність вибору з загальної партії бракованої деталі ?	Розв'язок: Позначимо гіпотези: H_1 – випуск деталі на 1-му верстаті; H_2 – випуск деталі на 2-му верстаті; H_3 – випуск деталі на 3-му верстаті. Подія A – вибір бракованої деталі. Ймовірності гіпотез дорівнюють: $P(H_1) = \frac{m_1}{n} = \frac{6000}{10000} = 0,6;$ $P(H_2) = \frac{m_2}{n} = \frac{1000}{10000} = 0,1;$ $P(H_3) = \frac{m_3}{n} = \frac{3000}{10000} = 0,3.$ Ймовірності того, що бракована деталь виготовлена на i-му верстаті ($i = 1, 2, 3$) дорівнюють: $P(A/H_1) = 0,08, P(A/H_2) = 0,02, P(A/H_3) = 0,05$ Тоді ймовірність вибору з загальної партії бракованої деталі дорівнює

	$P(A) = \sum_{i=1}^3 P(H_i)P(A/H_i) = 0,6 \cdot 0,08 +$ $+ 0,1 \cdot 0,02 + 0,3 \cdot 0,05 = 0,065$
Формула Байєса Ймовірність гіпотези після проведення дослідів дорівнює $P(H_i/A) = \frac{P(H_i) \cdot P(A/H_i)}{P(A)}$	
Приклад: Деталі попадають для перевірки на відповідність вимогам стандартів до одного з двох контролерів. Ймовірність того, що деталь попадає до 1-го контролеру $p_1 = 0,6$, до 2-го $p_2 = 0,4$. Ймовірність того, що стандартна деталь буде визнана стандартною першим контролером дорівнює 0,95; другим – 0,97. Деталь при перевірці була визнана стандартною. Знайти ймовірність того, що цю деталь перевіряв перший контролер.	Розв'язок: Позначимо подію A – деталь стандартна. Позначимо гіпотези: H_1 – деталь перевірена 1-м контролером; H_2 – деталь перевірена 2-м контролером. Тоді ймовірності гіпотез за умовою задачі дорівнюють: $P(H_1) = 0,6$; $P(H_2) = 0,4$. За умовою задачі умовні ймовірності дорівнюють: $P(A/H_1) = 0,95$; $P(A/H_2) = 0,97$. Повна ймовірність $P(A) = \sum_{i=1}^2 P(H_i)P(A/H_i) = 0,6 \cdot 0,95 + 0,4 \cdot 0,97 =$ $= 0,95865.$ За формулою Байєса ймовірність того, що цю деталь перевіряв 1-ий контролер дорівнює $P(H_1/A) = \frac{0,6 \cdot 0,95}{0,958} = 0,595.$
Формула Бернуллі Якщо відбувається n незалежних дослідів, в кожному з яких подія A може відбутися з ймовірністю p та не відбутися з ймовірністю $q = 1 - p$, то ймовірність того, що подія A відбудеться m разів в n дослідах виражається формулою $P_n(m) = C_n^m p^m q^{n-m}$	
Приклад: Для контролю якості партії деталей обирається вибірка з 10 деталей. Ймовірність попадання дефектної деталі до вибірки можна вважати сталою, що дорівнює $p = 0,03$. Визначити ймовірність того, що партія буде прийнята, якщо умовою цього є відсутність дефектних деталей у вибірці.	Розв'язок: Позначимо подію A – поява дефектної деталі у вибірці. Оскільки ймовірність події A не залежить від результатів дослідів, тобто від результатів контролю деталей, що обираються до вибірки, то досліди можна вважати незалежними. Тоді за формулою Бернуллі при $n = 10$; $m = 0$; $p = 0,03$; $q = 1 - p = 1 - 0,03 = 0,97$ шукана ймовірність дорівнюватиме $P_{10}(0) = C_{10}^0 0,03^0 0,97^{10-0} = 0,73.$

2. Випадкові величини					
Випадковою величиною (ВВ) називається величина, яка в результаті досліду набуває того чи іншого числового значення, заздалегідь невідоме.					
Дискретні ВВ (ДВВ)			Неперервні ВВ (НВВ)		
ДВВ – випадкова величина, яка набуває кінцеву множину значень або множину значень, елементи якої можна пронумерувати натуральними числами ДВВ $X = \{x_1, x_2, x_3, ..., x_n\}$			НВВ - випадкова величина, яка набуває будь-які значення на деякому інтервалі дійсної числової вісі НВВ $X = \{\alpha \leq x \leq \beta\}$		
2.1 Закон розподілу					
ДВВ			НВВ		
1) ряд розподілу			1) функція розподілу (інтегральна функція) $F(x) = P(X < x)$		
x_i	x_1	x_2	...	x_n	
p_i	p_1	p_2	...	p_n	$\sum_{i=1}^n p_i = 1$
2) багатокутник розподілу			2) щільність розподілу (диференціальна функція) $f(x) = F'(x)$		
3) функція розподілу (інтегральна функція) $F(x) = P(X < x)$					
2.2 Властивості функцій $F(x)$ та $f(x)$					
1) $F(x_2) \geq F(x_1)$ при $x_2 \geq x_1$ 2) $F(-\infty) = 0$ 3) $F(+\infty) = 1$ 4) $0 \leq F(x) \leq 1$ 5) $P(a \leq x \leq b) = F(b) - F(a)$			1) $f(x) \geq 0$ при $x \in (-\infty; +\infty)$ 2) $\int_{-\infty}^x f(x)dx = F(x)$ 3) $P(a \leq x \leq b) = \int_a^b f(x)dx$ 4) $\int_{-\infty}^{+\infty} f(x)dx = 1$ – умова нормування		
Приклади					
Для контролю стабільності процесу шліфування кілець через кожну годину обирають 5 кілець. Ймовірність того, що після обробки овальність кілець перевищить припустиму, дорівнює 0,25. Скласти ряд розподілу числа дефектних кілець в вибірці. Розв'язок: ВВ X – число дефектних кілець у вибірці є дискретною ВВ та набуває значення $ВВ \ X = \{0,1,2,3,4,5\}$. Ймовірність появи дефектного кільця є сталою, що дорівнює $p = 0,25$. За формулою Бернуллі знайдемо ймовірності, з якими ВВ X набуває усі свої значення, а саме: $p_1 = P(X = 0) = C_5^0 \cdot 0,25^0 \cdot 0,75^5 = 0,2375$; $p_2 = P(X = 1) = C_5^1 \cdot 0,25 \cdot 0,75^4 = 0,3955$; $p_3 = P(X = 2) = C_5^2 \cdot 0,25^2 \cdot 0,75^3 = 0,2637$; $p_4 = P(X = 3) = C_5^3 \cdot 0,25^3 \cdot 0,75^2 = 0,0878$;			Щільність розподілу НВВ X задана формулою $f(x) = \frac{A}{1+x^2}$. Знайти: 1) величину сталої A ; 2) функцію розподілу; 3) ймовірність попадання значень НВВ X до інтервалу $[-1;1]$. Розв'язок: 1) Коефіцієнт A знайдемо з умови нормування: $\int_{-\infty}^{+\infty} \frac{dx}{1+x^2} = 1 \Rightarrow \int_{-\infty}^{+\infty} \frac{dx}{1+x^2} = \arctg x \Big _{-\infty}^{+\infty} = \left(\frac{\pi}{2} + \frac{\pi}{2} \right) = \pi$ тоді $A = \frac{1}{\pi}$ та функція щільності має вигляд $f(x) = \frac{1}{\pi(1+x^2)}$. 2) Функцію розподілу знайдемо за		

$p_5 = P(X = 4) = C_5^4 \cdot 0,25^4 \cdot 0,75 = 0,0145;$ $p_6 = P(X = 5) = C_5^5 \cdot 0,25^5 \cdot 0,75^0 = 0,0010.$ Ряд розподілу матиме вигляд							означенням $F(x) = \int_{-\infty}^x f(x)dx = \frac{1}{\pi} \int_{-\infty}^x \frac{dx}{1+x^2} = \frac{1}{\pi} \lim_{a \rightarrow -\infty} \arctg x \Big _a^x =$ $= \frac{1}{\pi} \arctg x - \frac{1}{\pi} \cdot \left(-\frac{\pi}{2}\right) = \frac{1}{\pi} \arctg x + \frac{1}{2}.$	
x_i	0	1	2	3	4	5		
p_i	0,2375	0,3955	0,2637	0,0878	0,0145	0,0010		
Функція розподілу матиме вигляд:							3) Ймовірність попадання значень НВВ X до інтервалу $[-1;1]$ знайдемо за формулою	
$F(x) = \begin{cases} 0, & \text{якщо } x < 0; \\ 0,2375, & \text{якщо } 0 \leq x < 1; \\ 0,6330, & \text{якщо } 1 \leq x < 2; \\ 0,8967, & \text{якщо } 2 \leq x < 3; \\ 0,9845, & \text{якщо } 3 \leq x < 4; \\ 0,9990, & \text{якщо } 4 \leq x < 5; \\ 1, & \text{якщо } x \geq 5. \end{cases}$							$P(a \leq x \leq b) = \frac{1}{\pi} \int_{-1}^1 \frac{dx}{1+x^2} = \frac{1}{\pi} \cdot \arctg x \Big _{-1}^1 =$ $= \frac{1}{\pi} \arctg 1 - \frac{1}{\pi} \arctg(-1) = \frac{1}{\pi} \left(\frac{\pi}{4} + \frac{\pi}{4}\right) = \frac{1}{2}$	
2.3 Числові характеристики випадкових величин								
1. Математичне сподівання $M(X)$ характеризує положення ВВ на числовій осі, визначає деяке орієнтовне значення, біля якого зосереджені усі можливі значення ВВ.								
ДВВ X					НВВ X			
$M(X) = \sum_{i=1}^n x_i p_i$					$M(X) = \int_{-\infty}^{+\infty} x f(x) dx$			
Властивості								
1. $M(C) = C, C = const$					2. $M(C \cdot X) = C \cdot M(X), C = const$			
3. $M(X + Y) = M(X) + M(Y)$					4. $M(XY) = M(X) \cdot M(Y)$			
Мода (M_0)								
ДВВ X					НВВ X			
M_0 - це значення x_i з найбільшою ймовірністю					M_0 - це значення x_i , для якого $f(x)$ максимальне: $f(x = M_0) = f(M_0) = \max$			
Дисперсія $D(X)$ та середнє квадратичне відхилення $\sigma(X)$								
Дисперсія $D(X)$ та середнє квадратичне відхилення $\sigma(X)$ – числові характеристики розсіювання ВВ – характеризують наскільки тісно згруповані можливі значення ВВ біля центра її розсіювання (математичного сподівання)								
Дисперсією $D(X)$ називається математичне сподівання квадрата відхилення ВВ від її математичного сподівання								
ДВВ X					НВВ X			
$D(X) = \sum_{i=1}^n (x_i - M(X))^2 p_i$ $D(X) = M(X^2) - (M(X))^2$ $D(X) = \sum_{i=1}^n x_i^2 p_i - (M(X))^2$					$D(X) = \int_{-\infty}^{+\infty} (x - M(X))^2 f(x) dx$ $D(X) = \int_{-\infty}^{+\infty} x^2 f(x) dx - (M(X))^2$ $\sigma(X) = \sqrt{D(X)}$			

$\sigma(X) = \sqrt{D(X)}$					
Властивості					
1. $D(C) = 0, C = const$					
2. $D(C \cdot X) = C^2 \cdot D(X), C = const$					
3. $D(X + Y) = D(X) + D(Y)$					
4. $D(X - Y) = D(X) + D(Y)$					
5. $D(XY) = D(X) \cdot D(Y) + (M(X))^2 D(Y) + (M(Y))^2 D(X)$, якщо ВВ X та ВВ Y незалежні					
Приклади					
ДВВ X			НВВ X		
ДВВ X задана рядом розподілу			НВВ X задана функцією розподілу		
x_i	-2	-1	0	1	2
p_i	0,3	0,2	0,1	0,2	0,2
Знайти $M(X), D(X), \sigma(X)$			Знайти $M(X), D(X), \sigma(X)$		
Розв'язання					
Знайдемо математичне сподівання ВВ X та ВВ X^2 :			Знайдемо функцію щільності розподілу		
$M(X) = \sum_{i=1}^5 x_i p_i = -2 \cdot 0,3 - 1 \cdot 0,2 + 0 \cdot 0,1 + 1 \cdot 0,2 + 2 \cdot 0,2 = -0,2;$			$f(x) = F'(x) = \begin{cases} 0, & x \leq 3; \\ 1, & 3 < x \leq 4; \\ 0, & x > 4. \end{cases}$		
$M(X^2) = \sum_{i=1}^5 x_i^2 p_i = (-2)^2 \cdot 0,3 + (-1)^2 \cdot 0,2 + 0^2 \cdot 0,1 + 1^2 \cdot 0,2 + 2^2 \cdot 0,2 = 2,4;$			$M(X) = \int_3^4 x f(x) dx = \int_3^4 x dx = \frac{x^2}{2} \Big _3^4 = \frac{7}{2} = 3,5$		
$D(X) = M(X^2) - (M(X))^2 = 2,4 - (-0,2)^2 = 2,36;$			$D(X) = \int_3^4 x^2 f(x) dx - (M(X))^2 =$		
$\sigma(X) = \sqrt{D(X)} = \sqrt{2,36} \approx 1,54.$			$= \int_3^4 x^2 dx - \frac{49}{4} = \frac{x^3}{3} \Big _3^4 - \frac{49}{4} = \frac{37}{3} - \frac{49}{4} = \frac{1}{12}$		
2.4 Закони розподілу дискретних випадкових величин (ДВВ)					
2.4.1. Рівномірний закон розподілу					
Закон розподілу називається рівномірним , якщо ДВВ X набуває значення x_i з однаковою ймовірністю $p_i = \frac{1}{n}, i = 1, 2, \dots, n$.					
x_i	x_1	x_2	...	x_n	
p_i	$\frac{1}{n}$	$\frac{1}{n}$	...	$\frac{1}{n}$	
Числові характеристики					
$M(X) = \frac{1}{n} \sum_{i=1}^n x_i; \quad D(X) = \frac{1}{n} \sum_{i=1}^n x_i^2 - (M(X))^2; \quad \sigma(X) = \sqrt{D(X)}; \quad F(x_i) = \frac{1}{n} \sum_{k \leq i} k_i, k = 1, 2, \dots, n.$					
Приклад: Побудувати ряд розподілу та функцію розподілу ВВ X – цифри, що з'являється на грані при киданні кубика.					
Обчислити числові характеристики					
Розв'язання: ВВ X підпорядкована рівномірному закону розподілу, ряд розподілу має вигляд					
p_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Знайдемо функцію розподілу:

$$F(x) = \begin{cases} 0, & x < 1; \\ 1/6, & 1 \leq x < 2; \\ 2/6, & 2 \leq x < 3; \\ 3/6, & 3 \leq x < 4; \\ 4/6, & 4 \leq x < 5; \\ 5/6, & 5 \leq x < 6; \\ 1, & x \geq 6. \end{cases}$$

Математичне сподівання :

$$M(X) = \frac{1}{n} \sum_{i=1}^6 x_i = \frac{1}{6} (1 + 2 + 3 + 4 + 5 + 6) = 3,5$$

Дисперсія:

$$D(X) = \frac{1}{n} \sum_{i=1}^6 x_i^2 - (M(X))^2 = \frac{1}{6} (1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) - 3,5^2 \approx 2,92$$

2.4.2. Біноміальний закон розподілу

Дискретна ВВ X , що набуває цілочисельні значення $X = \{0, 1, 2, \dots, n\}$ розподілена за біноміальним законом, якщо ймовірності її можливих значень визначаються за формулою $P(X = m) = C_n^m p^m q^{n-m}$, де n – число незалежних дослідів, в яких може відбутися подія A ; p – ймовірність події A ; $q = 1 - p$

Числові характеристики

$$M(X) = n \cdot p \quad D(X) = n \cdot p \cdot q \quad \sigma(X) = \sqrt{n \cdot p \cdot q}$$

Приклад:

Є три верстати, коефіцієнт використання яких за часом складає 0,8. Знайти ймовірність того, що у середині зміни за нормальних умов виробництва з трьох верстатів будуть працювати не більше ніж два.

Розв'язання:

Випадкова величина X – число одночасно працюючих верстатів – набуває значення $X = \{0, 1, 2, 3\}$ та розподілена за біноміальним законом.

Ймовірності ВВ X обчислюються за формулою

$$P(X = m) = C_n^m p^m q^{n-m}, \text{ де } n = 3, p = 0,8, q = 1 - p = 1 - 0,8 = 0,2.$$

Необхідно знайти ймовірність

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2);$$

$$P(X = 0) = C_3^0 p^0 q^{3-0} = \frac{3!}{0!3!} 0,8^0 0,2^3 = 0,008;$$

$$P(X = 1) = C_3^1 p^1 q^{3-1} = \frac{3!}{1!2!} 0,8^1 0,2^2 = 0,096;$$

$$P(X = 2) = C_3^2 p^2 q^{3-2} = \frac{3!}{2!1!} 0,8^2 0,2 = 0,384.$$

Отже,

$$P(X \leq 2) = 0,008 + 0,096 + 0,384 = 0,488.$$

2.4.3. Гіпергеометричний закон розподілу

Дискретна ВВ X , що набуває цілочисельні значення $X = \{0, 1, 2, \dots, n\}$, розподілена за **гіпергеометричним законом**, якщо ймовірності її значень задаються формулою

$$P(X = d) = \frac{C_D^d C_{N-D}^{n-d}}{C_N^n}$$

Застосовується у статистичному контролі якості при складанні планів приймального контролю якості.

Параметри гіпергеометричного закону розподілу:

N – об'єм партії; D – число дефектних виробів в партії; n – об'єм вибірки.

Числові характеристики

$$M(X) = n \cdot \frac{D}{N} = n \cdot p$$

$$D(X) = n \cdot \frac{N-n}{N-1} \cdot p(1-p)$$

Приклад:

В партії зі 100 виробів міститься 10% браку. Навмання беруть 5 виробів для перевірки їх якості. Необхідно визначити ймовірність того, що у вибірці буде менш ніж 2 дефектних вироби, функцію розподілу, математичне сподівання та дисперсію числа дефектних виробів у вибірці

Розв'язання:

ДВВ X – число дефектних виробів у вибірці – набуває цілочисельні значення $X = \{0, 1, 2, 3, 4, 5\}$ та підпорядкована гіпергеометричному закону з параметрами $N = 100, D = 100 \cdot 0,1 = 10, n = 5$. Ймовірність того, що у вибірці буде рівно d дефектних виробів, знаходиться за формулою

$$P(X = d) = \frac{C_{10}^d C_{100-10}^{5-d}}{C_{100}^5}$$

Ряд розподілу матиме вигляд

x_i	0	1	2	3	4	5
p_i	0,583	0,340	0,070	0,007	0,0000	0,0000

Функція розподілу матиме вигляд:

$$F(x) = \begin{cases} 0, & x < 0 \\ 0,583, & 0 \leq x < 1 \\ 0,923, & 1 \leq x < 2 \\ 0,993, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

Числові характеристики:

$$M(X) = n \cdot \frac{D}{N} = n \cdot p = 5 \cdot 0,1 = 0,5$$

$$D(X) = n \cdot \frac{N-n}{N-1} \cdot p(1-p) = 5 \cdot \frac{100-5}{100-1} \cdot 0,1(1-0,1) = 0,4318$$

2.4.4. Розподіл Пуассона

Дискретна ВВ X , що набуває множини значень $X = \{0, 1, 2, \dots\}$, підпорядкована **закону Пуассона**, якщо ймовірності її можливих значень задаються формулою

$$P(X = m) = \frac{\lambda^m \cdot e^{-\lambda}}{m!}, \text{ де } \lambda = n \cdot p \text{ – параметр закону Пуассона.}$$

Закон Пуассона описує розподіл числа подій, які відбуваються в процесі випробування незалежно одна від одної, причому число випробувань є великим, а ймовірність появи події в кожному

випробуванні є малою.	
Числові характеристики	
$M(X) = D(X) = \lambda$	
Приклад: Завод відвантажив до бази 5000 виробів. Ймовірність пошкодження виробу в дорозі складає 0,0002. Знайти ймовірність того, що до бази надійде не більш ніж 3 пошкоджених вироби.	Розв'язання: ВВ X – число пошкоджених в дорозі виробів набуває значення $X = \{0, 1, 2, \dots, 5000\}$, ймовірність пошкодження виробу $p = 0,0002$. Таким чином, ВВ X підпорядкована закону Пуассона з параметром $\lambda = n \cdot p = 5000 \cdot 0,0002 = 1$. Тоді $P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) =$ $= e^{-1} \left(\frac{1^0}{0!} + \frac{1^1}{1!} + \frac{1^2}{2!} + \frac{1^3}{3!} \right) = 0,981.$ При обчисленнях можна користатися таблицями розподілу Пуассона.
2.5 Закони розподілу неперервних випадкових величин (НВВ)	
2.5.1. Рівномірний закон розподілу	
НВВ X розподілена за рівномірним законом розподілу на $[a; b]$, якщо її щільність розподілу має вигляд	
$f(x) = \begin{cases} 0, & x < a, x > b; \\ \frac{1}{b-a}, & a \leq x \leq b. \end{cases}$	
та функція розподілу має вигляд	
$F(x) = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & a \leq x \leq b; \\ 1, & x > b. \end{cases}$	
Рівномірно розподілена НВВ X зустрічається при знятті показань вимірювальних пристроїв. Похибка при заокругленні відліку до найближчого цілого ділення шкали – є випадковою величиною, яка може зі сталою щільністю ймовірності набуває будь-які значення між двома сусідніми діленнями.	
Числові характеристики	
1. $M(X) = \frac{a+b}{2}$ 2. $D(X) = \frac{(a+b)^2}{12}$ 3. $\sigma(x) = \frac{b-a}{2\sqrt{3}}$ 4. $P(a \leq x \leq b) = \frac{b-a}{b-a}$	
Приклад: Ціна ділення шкали амперметра дорівнює 0,1 ампера. Показники амперметра заокруглюють до найближчого цілого ділення. Знайти ймовірність того, що при відліку буде зроблена помилка, яка не перевищує 0,02 ампера.	Розв'язання: Помилку заокруглення відліку можна розглядати як випадкову величину X , яка розподілена за рівномірним законом в інтервалі між двома сусідніми цілими діленнями. Довжина інтервалу, в якому знаходяться можливі значення НВ X , дорівнює $b-a = 0,1$, тому функція щільності дорівнює $f(x) = \frac{1}{b-a}$, де $b-a$ – довжина інтервалу, в якому знаходяться можливі значення ВВ X та $f(x) = 0$ поза цього інтервалу. Тоді

	$f(x) = \frac{1}{b-a} = \frac{1}{0,1} = 10$ Помилка відліку перевищить 0,02, якщо вона буде знаходитися в інтервалі (0,02;0,08) . За формулою $P(a < X < b) = \int_a^b f(x)dx$ отримаємо $P(0,02 < X < 0,08) = \int_{0,02}^{0,08} 10dx = 0,6$
2.5.2. Експоненціальний розподіл	
Неперервна випадкова величина X розподілена за експоненціальним розподілом, якщо її щільність розподілу має вигляд $f(x) = \begin{cases} 0, & x < 0; \\ e^{-\lambda x}, & x \geq 0. \end{cases}$ та функція розподілу має вигляд $F(x) = \begin{cases} 0, & x < 0; \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$ де λ – параметр розподілу	
Експоненціальний розподіл застосовується в теорії надійності для опису часу безвідмовної роботи пристроїв. Так, якщо НВВ T – час безвідмовної роботи пристроїв, t ($0 \leq t < \infty$)- значення НВВ $T=(t, 0 \leq t < \infty)$, то функція розподілу визначає ймовірність відмови пристрою до настання моменту часу t. Тоді ймовірність безвідмовної роботи пристрою, яку можна обчислити за формулою $P(t) = 1 - F(t) = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t}$, називається функцією надійності, а параметр λ – інтенсивністю відмов.	
Числові характеристики	
1. $M(X) = \frac{1}{\lambda}$ 2. $D(X) = \frac{1}{\lambda^2}$ 3. $\sigma(x) = \frac{1}{\lambda}$	
Приклад: Час, необхідний для ремонту верстатів, задовольняє експоненціальному розподілу з параметром $\lambda = 0,25 \frac{1}{год}$. Знайти ймовірність того, що час ремонту одного верстата менш ніж 6 год. Знайти середній час, що витрачається на ремонт одного верстата.	Розв’язання: Позначимо через T час ремонту верстата. T – неперервна випадкова величина, яка набуває дійсні додатні значення на числовій осі: $T = \{0 \leq t < \infty\}$. За умовою, T підкорена експоненціальному закону, тому ймовірність того, що час ремонту менш ніж 6 годин, дорівнює $P(T \leq 6) = F(6) = 1 - e^{-0,25 \cdot 6} = 1 - 0,223 = 0,777.$ Середній час, що витрачається на ремонт одного верстата, є математичним сподівання часу ремонту T та дорівнює $m_n = \frac{1}{\lambda} = \frac{1}{0,25} = 4 \text{ год.}$

2.5.3. Розподіл Вейбула

Неперервна випадкова величини X підпорядкована **закону Вейбула**, якщо її функція щільності розподілу має вигляд

$$f(x) = \begin{cases} \frac{b}{a} \left(\frac{x-c}{a} \right)^{b-1} \cdot e^{-\left(\frac{x-c}{a} \right)^b}, & x \geq c; \\ 0, & x < c. \end{cases}$$

та функція розподілу має вигляд

$$F(x) = \begin{cases} 1 - e^{-\left(\frac{x-c}{a} \right)^b}, & x \geq c; \\ 0, & x < c. \end{cases}$$

де a – параметр масштабу, b – параметр форми, c – параметр зсуву.

При $b=1$ розподіл Вейбула перетворюється на експоненціальний розподіл

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

де $y = x - c$; $\lambda = \frac{1}{a}$.

Закон Вейбула застосовується в теорії надійності для опису строків служби багатьох виробів.

Числові характеристики

$$1. M(X) = a \cdot K_b + c$$

$$2. \sigma(X) = a q_b$$

де коефіцієнти K_b, q_b залежать від параметра форми розподілу b . Таблиці коефіцієнтів розподілу Вейбула надані в ДСУ.

Приклад:

Дослідження закону розподілу овальності в партії валиків показало, що овальність підкорена закону Вейбула з параметрами $c = 0$, $b = 2$, $a = 14$ мкм. Визначити середній відсоток браку, якщо допуск на овальність складає 28 мкм.

Розв'язання:

Запишемо функцію розподілу овальності, підставляючи відомі нам значення параметрів:

$$F(x) = 1 - e^{-\frac{x^2}{14^2}} \text{ при } x \geq 0.$$

Ймовірність того, що значення овальності вийде за межі допуску, дорівнює

$$\begin{aligned} P(x \geq 28) &= 1 - P(0 \leq X \leq 280) = \\ &= 1 - (F(28) - F(0)) = 1 - F(28) = \\ &= 1 - \left(1 - e^{-\frac{28^2}{14^2}} \right) = e^{-4} = 0,018. \end{aligned}$$

Отже, середній відсоток браку складе $q = P(X > 28) \cdot 100\% = 0,018 \cdot 100\% = 1,8\%$.

2.5.4 Нормальний закон розподілу (закон Гаусса)

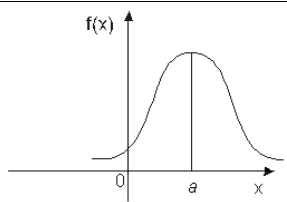
Неперервна випадкова величини X , що набуває значення на дійсній числовій осі $-\infty < x < +\infty$, має **нормальний закон розподілу**, якщо її щільність розподілу має вигляд

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}.$$

Функція розподілу такої ВВ має вигляд

$$F(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-m_x)^2}{2\sigma_x^2}} dx,$$

де σ_x, m_x - параметри розподілу.

Нормальний закон розподілу проявляється в тих випадках, коли ВВ X утворюється в результаті дії великого числа взаємно незалежних факторів, причому кожний фактор впливає на НВВ X несуттєво та жоден фактор не домінує над іншим за ступенем свого впливу на ВВ X.		
При $\sigma_x = 1, m_x = 0$ щільність розподілу та функція розподілу називають нормованими, а самі функції мають вигляд: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ та $F(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{x^2}{2}} dx$ – функція Лапласа. Значення $f(x)$ та $\Phi(x)$ знаходяться у відповідних таблицях.		
Властивості		
1. Графік $f(x)$ має назву кривої Гауса: – при $x \rightarrow \pm\infty$ $f(x) \rightarrow 0$, крива Гауса асимптотично наближується до осі ОХ; – крива Гауса симетрична відносно прямої $x = m_x = a$		
2. Максимальне значення функція щільності розподілу набуває в точці $x = m_x = a$, яке дорівнює $f_{\max}(a) = \frac{1}{\sqrt{2\pi}}$.		
3. При зміні параметра m_x крива Гауса зміщується вздовж осі ОХ, зберігаючи свій вигляд. При зміні параметрів σ_x вигляд кривої змінюється: – при зростанні σ_x крива притискається до осі ОУ; – при спаданні σ_x крива витягується вздовж осі ОУ.		
4. Функція розподілу $F(x)$ виражається через функцію Лапласа $\Phi(x)$ наступним чином $F(x) = \Phi(y), \text{ де } y = \frac{x - m_x}{\sigma_x}.$		
5. $F(-x) = \Phi(-y) = 1 - \Phi(y)$.		
6. Ймовірність попадання випадкової величини X в інтервал $[a; b]$ обчислюється за формулою $P(a \leq x \leq b) = \Phi\left(\frac{b - m_x}{\sigma_x}\right) - \Phi\left(\frac{a - m_x}{\sigma_x}\right)$		
7. Ймовірність попадання випадкової величини X в інтервал, симетричний відносно центра розсіювання $[m_x - d; m_x + d]$, обчислюється за формулою $P(m_x - d \leq x \leq m_x + d) = P(X - m_x \leq d) = 2\Phi\left(\frac{d}{\sigma_x}\right) - 1$		
Числові характеристики		
1. $M(X) = m_x$	2. $D(X) = \sigma_x^2$	3. $\sigma(X) = \sigma_x$
Правило «трьох сигм»		
У випадку, коли інтервал $[\alpha; \beta]$ довжиною l є симетричним відносно m_x, то ймовірність влучання в цей інтервал дорівнює $P(X - m_x \leq l) = 2\Phi\left(\frac{l}{\sigma_x}\right) - 1, \text{ де } l = \frac{\beta - \alpha}{2}.$ Якщо задати $l = 3\sigma_x$, то ймовірність влучення в інтервал $[m_x - 3\sigma; m_x + 3\sigma]$ буде дорівнювати $P(X - m_x \leq 3\sigma) = 2\Phi(3) - 1 \approx 2 \cdot 0,9986 - 1 \approx 0,9973 \approx 1.$		
Правило «трьох сигм»: всі значення випадкової величини X розподіляються навколо центра		

розсіювання m_x (математичного сподівання) в межах $\pm 3\sigma$.	
Приклад: Міцність сталі є нормально розподіленою випадковою величиною з наступними характеристиками: $\sigma_x = 3 \text{ кг/мм}^2, m_x = 31 \text{ кг/мм}^2$. Визначити: 1) ймовірність отримання сталі з границею міцності, що лежить в межах $31 \text{ кг/мм}^2 \leq X \leq 35 \text{ кг/мм}^2$; 2) ймовірність отримання сталі з границею міцності меншою за 36 кг/мм^2; 3) ймовірність отримання сталі з границею міцності більшою за 39 кг/мм^2; 4) границі допуску, ймовірність виходу за які для границі міцності дорівнює 0,005.	Розв'язання: Шукані ймовірності дорівнюють: 1) $P(31 \leq x \leq 35) = \Phi\left(\frac{35-31}{3}\right) - \Phi\left(\frac{31-31}{3}\right) = \Phi(1,3) - \Phi(0) = 0,9030$; 2) $P(x < 36) = F(36) = \Phi\left(\frac{36-31}{3}\right) = \Phi(1,66) = 0,9515$; 3) $P(x > 39) = 1 - P(x < 39) = 1 - F(39) = 1 - \Phi\left(\frac{39-31}{3}\right) = 1 - \Phi(2,66) = 0,0040$; 4) з умови відомо, що $P\left(X - m_x < \frac{\delta}{2}\right) = 0,995$, де δ – область допустимих значень границі міцності, яка дорівнює різниці верхньої та нижньої границі допуску $\delta = T_B - T_H; T_{B,H} = m_x \pm \frac{\delta}{2}$ У відповідності з формулою $P(m_x - d \leq x \leq m_x + d) = P(X - m_x \leq d) = 2\Phi\left(\frac{d}{\sigma_x}\right) - 1,$ отримаємо: $P(X - m_x \leq d) = 2\Phi\left(\frac{\delta}{2\sigma_x}\right) - 1, \text{ звідки}$ $\Phi\left(\frac{\delta}{2\sigma_x}\right) = \frac{1}{2}(1 + 0,995) = 0,997.$ За таблицею значень функції Лапласа знаходимо значення аргументу, що відповідає цій ймовірності 0,997: $\frac{\delta}{2\sigma_x} = 2,8$ Звідки $\frac{\delta}{2} = 8,4$ та $T_B = 39,1 \text{ кг/мм}^2$; $T_H = 22,6 \text{ кг/мм}^2$.
Приклад: Здійснюється вимірювання діаметра валу без систематичних помилок. Випадкові помилки вимірювання X підпорядковані нормальному закону з середнім квадратичним відхиленням $\sigma = 10 \text{ мм}$. Знайти ймовірність того, що вимірювання буде проведено з помилкою, яка не перевищує за абсолютною величиною 15 мм	Розв'язання: Тому що математичне сподівання випадкових помилок дорівнює нулю, то можна застосувати формулу $P(X < \delta) = 2\Phi\left(\frac{\delta}{\sigma_x}\right).$ При $\delta = 15$; $\sigma_x = 10$ отримаємо $P(X < 15) = 2\Phi\left(\frac{15}{10}\right) = 2\Phi(1,5).$ За таблицею значень функції Лапласа знаходимо $\Phi(1,5) = 0,4332$. Тоді шукана ймовірність дорівнює $P(X < 15) = 2 \cdot 0,4332 = 0,8664.$

3. ЕЛЕМЕНТИ МАТЕМАТИЧНОЇ СТАТИСТИКИ

Математична статистика вивчає масові випадкові явища в природі, суспільстві та техніці методами теорії ймовірностей. Найбільш важливими задачами математичної статистики є:

- 1) визначення способів отримання та групування статистичних даних, об'єму необхідних дослідів до початку та у ході дослідження;
- 2) розробка методів обробки на аналізу статистичних даних, що дозволяють:
 - а) перевірити правильність статистичних гіпотез про невідомий закон розподілу та його параметрів;
 - б) оцінити невідому функцію розподілу та числові характеристики випадкових величин;
 - в) провести згладжування експериментальних залежностей.

3.1 Статистична сукупність. Генеральна сукупність та вибірка

У математичній статистиці прийнято оперувати такими поняттями, як **генеральна сукупність (ГС) та вибіркова сукупність (ВС) – вибірка**.

Означення. Сукупність однорідних об'єктів, що вивчається вибіркоvim методом, називається **генеральною сукупністю (ГС)**.

Висновки про генеральну сукупність (ГС) визначаються на основі вивчення вибірки.

Означення. **Вибіркою** називається сукупність випадково вибраних об'єктів з генеральної сукупності (ГС).

Число об'єктів у генеральній сукупності N та вибірці n називають їх **об'ємами**.

Відбір об'єктів до вибірки

а) випадковим повторним або неповторним	N об'єктів генеральної сукупності записують на картках та вибирають одну картку з усіх.
б) механічним	Генеральна сукупність ділиться на групи, щоб в кожну групу до вибірки увійшов один об'єкт.
в) типовим	Генеральну сукупність ділять на частини та з кожної частини обирають об'єкт до вибірки випадковим способом.
г) серійним	Генеральну сукупність розбивають на серії, їх нумерують та потім користуючись випадковим відбором, утворюють вибірку, в яку відібрані серії входять цілком.

3.2 Статистичний ряд. Статистична функція розподілу. Гістограма та полігон

Для вивчення результатів вибірових спостережень їх треба, по-перше, впорядкувати. Розташуємо значення, що спостерігаються у наступному порядку:

$$x_1 < x_2 < x_3 < \dots < x_n \quad \text{або} \quad x_1 > x_2 > x_3 > \dots > x_n.$$

Отримані ряди називаються **ранжованими**.

У ранжованому ряді різні значення X_i можуть зустрічатися декілька разів.

Позначимо через m_i число, що показує, скільки разів зустрічається варіанта у вибірці, тоді m_i називається **частотою варіанти**.

Відношення частоти варіанти m_i до об'єму вибірки називається **відносною частотою** p_i^* або **частістю** $p_i^* = \frac{m_i}{n}$. Частість p_i^* характеризує частку кожного значення в загальній кількості спостережень і є статистичною ймовірністю.

Варіанти X_i та відповідні їм частоти m_i або частоти (відносні частоти) p_i^* утворюють статистичний ряд вибірки :

x_i	x_1	x_2	...	x_n
m_i	m_1	m_2	...	m_n

x_i	x_1	x_2	...	x_n
p_i^*	p_1^*	p_2^*	...	p_n^*

Приклад 1. Для оцінки граничної похибки вимірювання мікрометра проведені 8 повторних вимірювань діаметра в заданому місці (з відліком десятої долі ціни ділення шкали). В результаті отримано ряд значень діаметра: 19,990; 19,998; 19,995; 19,997; 20,007; 19,996; 19,993; 19,994 мм. Скласти статистичний ряд частот та відносних частот.

Розв'язання: статистичний ряд частот та відносних частот має вигляд

x_i	19,990	19,993	19,994	19,995	19,996	19,997	19,998	20,007
m_i	1	1	1	1	1	1	1	1

$$\sum_{i=1}^8 m_i = 8$$

x_i	19,990	19,993	19,994	19,995	19,996	19,997	19,998	20,007
p_i^*	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

$$\sum_{i=1}^8 p_i^* = 1$$

Інтервальный ряд

Якщо число членів варіаційного ряду велике, то для зручності його вивчення утворюють інтервальный ряд, групуючи варіанти у інтервали.

Для інтервального ряду частота m_i дорівнює числу значень варіантів, що потрапляють до заданого інтервалу та відповідний статистичний ряд складає перелік інтервалів та відповідні їм частоти або частоти.

x_i	$[x_1; x_2]$	$[x_2; x_3]$	...	$[x_{n-1}; x_n]$
m_i	m_1	m_2	...	m_{n-1}

3.3 СТАТИСТИЧНА ФУНКЦІЯ РОЗПОДІЛУ

Функцію розподілу генеральної сукупності будемо називати теоретичною функцією розподілу F_T . Її статистичним аналогом є статистична (емпірична) функція розподілу $F_n^*(x)$.

Означення. Статистична функція розподілу визначає для кожного значення x_i частість або статистичну ймовірність події, яка полягає в тому, що випадкова величина X набуде значення, менше за x_i , тобто $F_n^*(x_i) = P(X < x_i)$.

Значення $F_n^*(x)$ обчислюється за формулою :

$$F_n^*(x_i) = \frac{\sum_{k < i} m_k}{n}.$$

Для графічного зображення статистичного ряду застосовують полігон або гістограму. Полігон служить для зображення дискретного статистичного ряду, а гістограма для зображення інтервального статистичного ряду.

Для графічного зображення статистичної функції розподілу $F_n^*(x)$ застосовують **кумулятивну криву**.

Практичні рекомендації з побудови гістограми та статистичної функції розподілу

1. При побудові гістограм складається таблиця, у першому рядку якої вказуються межі інтервалів (x_k, x_{k+1}) ; у другому – середина $\bar{x}_k$ цього інтервалу, у третьому – кількість варіант m_k або частота варіант p_k^*

(x_k, x_{k+1})	(x_1, x_2)	(x_2, x_3)	...	(x_{n-1}, x_n)
$\bar{x}_k$	$\bar{x}_1$	$\bar{x}_2$	...	$\bar{x}_n$
m_k	m_1	m_2	...	m_n
p_k^*	p_1^*	p_2^*	...	p_n^*

2. Для побудови статистичної функції розподілу $F_n^*(x)$ також треба скласти таблицю:

x_k	x_1	x_2	...	x_n
m_k	m_1	m_2	...	m_n
p_k^*	p_1^*	p_2^*	...	p_n^*

3. При великому об'ємі вибірки статистичний варіаційний ряд замінюють на інтервальний, який отримується групуванням вибірки.

Таке групування полягає у тому, що весь інтервал $(x_{\min}, x_{\max})$ значень ГС розбивається на N груп довжиною :

$$d = \frac{R}{N} = \frac{x_{\max} - x_{\min}}{N},$$

де R – розмах варіювання. На практиці рекомендується обирати величину N рівною $N = 1 + 3,322 \lg n$. При групуванні вибірки для кожного конкретного розряду визначаються межі (x_k, x_{k+1}) , $k = 1, \dots, N$ і підраховується частота m_i та відносна частота $p_i^* = \frac{m_i}{n}$. За такого групування інтервальний ряд замінюється дискретним, для чого всі значення X у межах розряду (x_k, x_{k+1}) замінюються середнім значенням :

$$\bar{x}_k = \frac{x_k + x_{k+1}}{2}$$

Для графічного зображення $F_n^*(x)$ та гістограми утворюють таблицю:

(x_k, x_{k+1})	$(x_1, x_1 + d)$	$(x_1 + d, x_1 + 2d)$		$(x_1 + (N-1)d, x_n)$
m_k	m_1	m_2	...	m_n
$\bar{x}_k$	$\bar{x}_1$	$\bar{x}_2$	...	$\bar{x}_n$
p_k^*	p_1^*	p_2^*	...	p_n^*

3.4 СТАТИСТИЧНІ ОЦІНКИ ПАРАМЕТРІВ РОЗПОДІЛУ

Числові характеристики статистичного розподілу

Середнє арифметичне значень випадкової величини у випадку, коли варіанти не повторюються:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}.$$

Середнє арифметичне значень випадкової величини у випадку, коли варіанти повторюються:

$$\bar{x} = \frac{\sum_{i=1}^n m_i x_i}{n}.$$

Вибіркова середня при групуванні вибірки на N розрядів: $\bar{x} = \sum_{k=1}^n \bar{x}_k \cdot p_k^*$.

Вибіркова дисперсія: $S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$.

Вибіркова дисперсія для групуваної вибірки: $S^2 = \sum_{k=1}^n (x_k - \bar{x})^2 \cdot p_k^*$.

- ✓ **мода** вибірки – найбільше значення частоти;
- ✓ **вибіркова медіана** – число у ранжованому ряді спостережень, яке стоїть у його середині, або середнє арифметичне сусідніх, розташованих у середині значень при парній кількості чисел ряду;
- ✓ **коефіцієнт варіацій** : $V = \frac{S}{\bar{x}} \cdot 100\%$, де $S = \sqrt{S^2}$ - середнє квадратичне відхилення.

Приклад. Контрольні обміри діаметрів болтів дали наступні результати: 2,31; 2,28; 2,29; 2,28; 2,32; 2,28; 2,32; 2,29; 2,31; 2,32.

Знайти точкові оцінки процесу обмірювання.

Розв'язання: Випадкова величина X має об'єм $n=10$, вибіркве середнє :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{10} \sum_{i=1}^{10} x_i = \frac{1}{10} (2,31 + 2,28 + 2,29 + 2,28 + 2,32 + 2,28 + 2,32 + 2,29 + 2,31 + 2,32) = 2,3$$

Вибіркова дисперсія :

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{10} \sum_{i=1}^{10} (x_i - \bar{x})^2 = \frac{1}{10} ((2,31 - 2,3)^2 + (2,28 - 2,3)^2 + (2,29 - 2,3)^2 + (2,28 - 2,3)^2 + (2,32 - 2,3)^2 + (2,28 - 2,3)^2 + (2,29 - 2,3)^2 + (2,31 - 2,3)^2 + (2,32 - 2,3)^2) = 0,0003.$$

Середнє квадратичне відхилення :

$$S = \sqrt{S^2} = \sqrt{0,0003} \approx 0,017; \text{ мода дорівнює } 3.$$

Ранжований ряд має вигляд:

x_i	2,28	2,29	2,31	2,32
m_i	3	2	2	3

	Медіана дорівнює $\frac{2,29 + 2,31}{2} = 2,3$; коефіцієнт варіації $V = \frac{S}{\bar{x}} \cdot 100\% = \frac{0,017}{23} \cdot 100\% \approx 0,74\% .$
3.4.1 Точкові оцінки	
Задача оцінювання у загальному вигляді формується так: нехай X – генеральна сукупність (або вибірка сукупність), що має довільний закон розділу з щільністю розподілу $f(x, \theta)$, де θ – параметр розподілу, значення якого невідоме. Необхідно зробити висновки про істинне значення параметра θ^0 на основі обробки вибірки $X_n = (x_1, x_2, \dots, x_n)$ з генеральної сукупності. Означення. Оцінкою θ^* параметра розподілу θ називається функція (статистика) $\theta^* = \theta^*(\bar{x}_n)$, що залежить від вибірки X_n, яка у певному статистичному розумінні близька до істинного значення параметра θ^0	
3.4.2 Способи оцінювання параметрів	
Точковий	При точковому оцінюванні шукані значення θ^0 вважається рівним його вибірковій оцінці θ^* , тобто залишається точка, біля якої знаходиться невідомий параметр.
Інтервальний	При інтервальній оцінці знаходять інтервал (θ_1, θ_2) у якому з зазначеною ймовірністю знаходиться точне значення θ^0 .
Оцінки параметрів розподілу – математичного сподівання та дисперсії	
Точкові	
Статистичною оцінкою математичного сподівання є середня вибірка або вибірка середня	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Статистичною оцінкою дисперсії є вибірка дисперсія та виправлена вибірка дисперсія	$D(x) = \frac{1}{n} \sum_{i=1}^n n_i (x_i - \bar{x})^2; S_0^2 = \frac{1}{n-1} \sum_{i=1}^n n_i (x_i - \bar{x})^2$
Незміщена, обґрунтована, ефективна оцінка дисперсії	$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - M(x))^2$
У випадку коли $M(x)$ невідоме, для оцінки дисперсії використовують оцінку	$S_0^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \text{ або } S_0^2 = \frac{n}{n-1} S^2$
Інтервальні	
Інтервал $((\theta_1, \theta_2))$, в якому з заданою ймовірністю γ буде знаходитися шуканий параметр, називається довірчим інтервалом	
– довірчий інтервал для математичного сподівання $M(x)$ при відомій дисперсії	При відомій дисперсії $D(x) = \sigma^2$ $\left(\bar{x} - t_\gamma \cdot \frac{\sigma}{\sqrt{n}}; \bar{x} + t_\gamma \cdot \frac{\sigma}{\sqrt{n}} \right),$ де $\bar{x} \approx a, \sigma = \sqrt{D(x)}, n$ – об'єм вибірки. Значення t_γ знаходяться з рівняння $\Phi(t_\gamma) = \frac{\gamma}{2}$, або $F(t_\gamma) = \frac{\gamma+1}{2}$, де $\Phi(t_\gamma)$ – функція Лапласа, значення якої визначаються за відповідною таблицею.

	Точність оцінки або середня помилка вибірки, з якою довірчий інтервал покриває невідомий параметр, визначається за формулою $\delta = \frac{t_\gamma \sigma}{\sqrt{n}}$, де t_γ значення аргументу функції Лапласа $\Phi(t_\gamma) = \frac{\gamma}{2}$. Значення визначається за таблицею функції Лапласа за заданою надійністю γ.
Приклад. Генеральна сукупність X має нормальний закон розподілу з $\sigma=2$. З вибірки об'єму $n=25$ обчислено вибіркове середнє $\bar{x}=11$. Обчислити довірчий інтервал для математичного сподівання при довірчій ймовірності $\gamma=0,95$.	Розв'язання. За таблицею функції Лапласа з рівняння $F(t_\gamma) = \frac{\gamma+1}{2} = \frac{0,95+1}{2} = 0,975$ знайдемо значення $t_\gamma=1,96$. Тоді довірчий інтервал матиме вигляд $\left(11 - 1,96 \cdot \frac{2}{\sqrt{25}}; 11 + 1,96 \cdot \frac{2}{\sqrt{25}} \right) = (10,216; 11,784).$
– довірчий інтервал для математичного сподівання при невідомій дисперсії (при невеликому об'ємі вибірки).	Оберемо за точкову оцінку математичного сподівання вибіркове середнє $\bar{x}$, а за точкову оцінку дисперсії – виправлену вибірккову дисперсію $S_0^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \text{ тоді}$ $\left(\bar{x} - t_{\gamma,k} \cdot \frac{S_0}{\sqrt{n}}; \bar{x} + t_{\gamma,k} \cdot \frac{S_0}{\sqrt{n}} \right),$ де $t_{\gamma,k}$ визначається за таблицею розподілу Стюдента за даними значеннями γ та $k=n-1$.
Приклад. З генеральної сукупності здійснено вибірку об'єму $n=9$, знайдено вибіркове середнє $\bar{x}=30,1$ та вибіркове середнє квадратичне відхилення $S_0=6$. Знайти довірчий інтервал для математичного сподівання при довірчій ймовірності $\gamma=0,99$.	Розв'язання. За довірчою ймовірністю $\gamma=0,99$ та рівнем значущості $k=n-1=9-1=8$ за таблицею розподілу Стюдента знаходимо значення статистики $t_{\gamma,k}=t_{0,99,8}=3,5$. Тоді довірчий інтервал матиме вигляд $(30,1 - 3,5 \cdot \frac{6}{\sqrt{25}}; 30,1 + 3,5 \cdot \frac{6}{\sqrt{25}}) =$ $= (30,1 - 7; 30,1 + 7) = (23,1; 37,1).$
– довірчий інтервал для дисперсії при відомому математичному сподіванні	При невідомому математичному сподіванні довірчий інтервал має вигляд $\left(\frac{(n-1)S_0^2}{\chi_2^2}; \frac{(n-1)S_0^2}{\chi_1^2} \right),$ де $S_0^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$, а значення статистик $\chi_1^2; \chi_2^2$ визначаються за таблицею значень χ^2 відповідно значенням параметрів k та довірчій ймовірності α наступним чином: χ_1^2 при $k=n-1$ та $\beta_1 = \frac{1+\gamma}{2}$; χ_2^2 при $k=n-1$ та $\beta_2 = \frac{1-\gamma}{2}$.

4. СТАТИСТИЧНІ ГІПОТЕЗИ	
Означення: Статистичною гіпотезою називається припущення про форму невідомого закону розподілу або про параметри відомого закону, що перевіряється на основі обробки даних вибіркового спостереження. Статистичні гіпотези відносяться до  законів розподілу  параметрів розподілу	
ГІПОТЕЗИ Висунута гіпотеза називається основною (нульовою) – H_0. Основній гіпотезі H_0 протиставляється альтернативна (конкуруюча) гіпотеза H_1, яка відкидає гіпотезу H_0. Простою називають гіпотезу, якій відповідає один розподіл або один параметр. Складеною називається гіпотеза, яка зводиться до вибору деякого розподілу з їх множини, або декількох параметрів.	
Помилки при перевірці гіпотез	
Помилки 1-го типу	основна гіпотеза H_0 відхиляється, хоча вона є вірною (помилка відхилю вірної гіпотези)
Помилка 2 –го типу	основна гіпотеза H_0 приймається, хоча є хибною (помилка прийняття хибної гіпотези)
РІВЕНЬ ЗНАЧУЩОСТІ	
Ймовірність припустити помилку 1-го типу називається рівнем значущості (α). Стандартні критичні значення α дорівнюють 0,1 ; 0,05 ; 0,01 ; 0,001.	
4.1. Критерії та загальний алгоритм перевірки статистичної гіпотези Для перевірки основної гіпотези H_0 користуються спеціально підбраною випадковою величиною ξ, розподіл якої відомий. Серед таких відомих законів розподілу виділяються: нормальний закон розподілу; розподіл Стюдента; розподіл Фішера; χ^2-розподіл Пірсона.	
4.1.1. Розподіл Стюдента	
Застосовується для оцінки генеральної середньої a однієї сукупності та для перевірки гіпотез відносно генеральних середніх декількох сукупностей при малих ($n \leq 30$) об'ємах вибірок. Для розподілу Стюдента розглядається випадкова величина t: $t_k = \frac{x_0}{\sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}}$ Критичні значення $t_{\gamma,k}$ знаходяться з умови $P(t_k \leq t_{\gamma,k}) = \gamma$, де γ - довірча ймовірність.	
4.1.2. Розподіл Фішера	
Застосовується для перевірки гіпотез про рівність генеральних дисперсій S_x^2 та S_y^2. Для розподілу Фішера розглядається випадкова величина F_{k_1,k_2}: $F_{k_1,k_2} = \frac{\frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - a_x)^2}{\frac{1}{n_2} \sum_{j=1}^{n_2} (y_j - a_y)^2} = \frac{S_x^2}{S_y^2}, \text{ де } k_1, k_2 - \text{число ступенів вільності } (k_1 = n_1, k_2 = n_2).$ Значення статистики F_{k_1,k_2} визначається за таблицею розподілу Фішера.	

4.1.3. Критерій згоди Пірсона (критерій χ^2)

Застосовується для знаходження невідомого закону розподілу генеральної сукупності X за вибіркою об'єму n .

Передбачає виконання трьох етапів:

1) обробки вибірки з метою побудови гістограми (полігона) частот та статистичної функції розподілу $F^*(x)$;

2) формування гіпотези H_0 відносно теоретичної функції розподілу $F_T(x)$;

3) перевірка згоди розподілів $F^*(x)$ та $F_T(x)$ за χ^2 -критерієм.

Мірою різниці вибірових та теоретичних частот за критерієм Пірсона є величина:

$$\chi_{k_0}^2 = \sum_{k=1}^n \frac{(m_k - m^T)^2}{m^T}.$$

Алгоритм

перевірки гіпотези про вид функції розподілу $F_T(x)$ генеральної сукупності X на основі вибірки за допомогою χ^2 – критерію:

1. за вибіркою побудувати гістограму (полігон) частот;

2. на основі гістограми (полігона) висунути гіпотезу H_0 про теоретичний закон розподілу $F_T(x)$ генеральної сукупності X ;

3. для нормального закону розподілу обчислити його параметри $\bar{x}$ та $\sigma = S^2$;

4. обчислити теоретичні частоти m^T за формулою

$$m^T = n \cdot p^T;$$

5. знайти величину $\chi_{k_0}^2$ за формулою

$$\chi_{k_0}^2 = \sum_{k=1}^n \frac{(m_k - m^T)^2}{m^T};$$

6. визначити число ступенів вільностей k_0 за формулою $k_0 = N - r - 1$, де N – число інтервалів групування; r – число параметрів теоретичного розподілу.

Якщо висунута основна гіпотеза H_0 полягає в тому, що генеральна сукупність підпорядкована нормальному закону розподілу з $r=2$ (параметри $M(x), D(x)$), то $k_0 = N - r - 1 = N - 3$;

7. за рівнем значущості α за таблицею χ^2 знайти критичне значення χ_{α, k_0}^2 ; порівняти $\chi_{k_0}^2$ та χ_{α, k_0}^2 і зробити висновок про висунуту гіпотезу H_0 .

5. КОРЕЛЯЦІЙНА ЗАЛЕЖНІСТЬ

Статистичним (ймовірнісним) називається зв'язок між величинами, коли значенню однієї з них відповідає декілька значень іншої, причому, яке саме значення вона приймає заздалегідь невідомо.

Статистичною називається така залежність між величинами, коли зміна однієї з них викликає зміну іншої. Якщо ця зміна призводить до зміни середнього значення іншої величини, то така статистична залежність називається **кореляційною**.

Вичерпною характеристикою статистичного зв'язку величин є їх спільний закон розподілу, який є невідомим, але в принципі може бути одержаний на основі вибірки.

На практиці для дослідження статистичної залежності між $ВВ X$ та $ВВ Y$ обмежуються визначенням залежності між однією з них та умовним математичним сподіванням іншої; тобто між X та $M(Y/X) = x$ або між Y та $M(X/Y) = y$.

Умовне математичне сподівання $M((Y/X) = x) = y(x)$ є функцією від x та називається **функцією регресії Y на X** , $M((X/Y) = y) = x(y)$ є функцією від Y та називається **функцією регресії X на Y** .

Залежність, що описується функцією регресії називається **кореляційною**.

Дві **основні задачі**, які розв'язуються методом кореляційного аналізу:

- 1) визначення форми і параметрів рівняння зв'язку;
- 2) оцінювання щільності зв'язку.

5.1. Точкові оцінки лінії регресії та коефіцієнта кореляції

Вигляд рівняння регресії може бути відомий заздалегідь з теоретичних положень про зв'язок досліджуваних величин, з міркувань аналогії або порівняння експериментальних даних з відомими функціями.

Для такого порівняння зручно використовувати лінію регресії, яка будується за кореляційним полем вибірки.

Кореляційним полем називається графічне зображення на площі $ХОУ$ усіх точок вибірки $(x_1; y_1), (x_2; y_2), (x_3; y_3), \dots, (x_n; y_n)$.

Параметри рівняння ліній регресії $\tilde{y}(x)$ виражаються через вибіркові величини:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i; \quad S_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2; \quad S_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

Для лінійної регресії $\tilde{y}(x)$ показником щільності зв'язку (близькості точок $(x_i; y_i)$ до ліній регресії $\tilde{y}(x)$) є коефіцієнт кореляції

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{S_x - S_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{S_x - S_y}$$

Для лінійної регресії вводиться коефіцієнт регресії

$$1) \text{ для } x \text{ на } y: \rho_{xy} = r \cdot \frac{\sigma_x}{\sigma_y};$$

$$2) \text{ для } y \text{ на } x: \rho_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}.$$

Ці коефіцієнти вказують середню зміну залежної змінної при зміні незалежної на одиницю. Величини ρ_{xy} та ρ_{yx} мають певний знак:

– якщо $\rho_{xy}, \rho_{yx} > 0$, то зв'язок називається прямим;

– якщо $\rho_{xy}, \rho_{yx} < 0$ – оберненим.

При $\rho_{xy} = 0$ ($\rho_{yx} = 0$) зв'язок між X та Y відсутній, і величини X та Y є некорельовані.

З порівняння оцінки лінійної регресії $\tilde{y} = a + bx$ і теоретичної лінії регресії $y(x)$ випливає, що параметр b є вибіркоvim коефіцієнтом регресії ρ_{yx} :

$$\rho_{yx} = \frac{\sum_{i=1}^n \Delta x_i \Delta y_i}{\sum_{i=1}^n \Delta x_i^2}.$$

5.2. Побудова лінійної регресії за вибіркою за методом найменших квадратів

Складемо суму $S = \sum_{i=1}^n (y_i - a - bx)^2$, що є функцією двох змінних a та b . За необхідною умовою існування екстремуму маємо:

$$\begin{cases} S'_a = 0 \\ S'_b = 0 \end{cases} \Rightarrow \begin{cases} 2 \sum_{i=1}^n (y_i - a - bx_i) \cdot (-1) = 0 \\ 2 \sum_{i=1}^n (y_i - a - bx_i) \cdot (-x_i) = 0 \end{cases} \Rightarrow \begin{cases} \sum_{i=1}^n y_i = an + b \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 \end{cases}.$$

Розв'язком цієї системи є:

$$a = \bar{y} - b\bar{x}; \quad b = \frac{\sum_{i=1}^n \Delta x_i \Delta y_i}{\sum_{i=1}^n \Delta x_i^2},$$

де $\Delta x_i = x_i - \bar{x}$; $\Delta y_i = y_i - \bar{y}$; $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$; $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$.

Лінійна регресія буде мати вигляд: $\tilde{y} = a + bx$.

Приклад

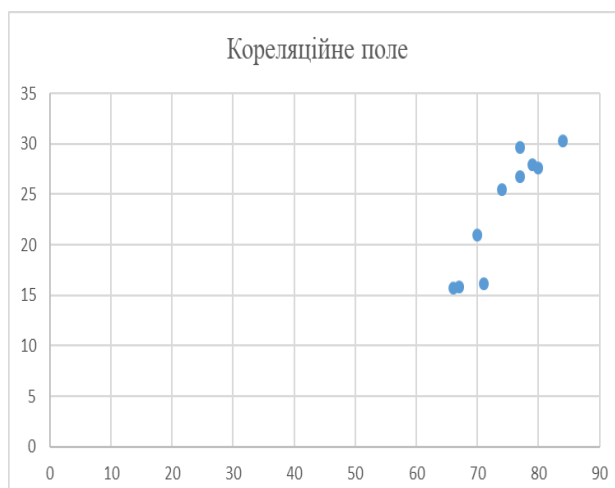
Експериментальні дані залежності X від Y надані в таблиці:

№	1	2	3	4	5	6	7	8	9	10
x_i	79	70	80	71	77	77	84	66	74	67
y_i	28	21	27,6	16,2	29,7	26,8	30,3	15,7	25,5	15,8

Встановити вид залежності між цими величинами, визначити параметри рівняння регресії Y на X та щільність зв'язку між ними.

Розв'язок

Для встановлення виду залежності між величинами X та Y побудуємо кореляційне поле результатів експерименту.



Аналіз кореляційного поля свідчить про те, що залежність близька до лінійної, а її вид – пряма лінія: $\bar{y}(x) = \bar{y} + \rho_{yx}(x - \bar{x})$. Визначимо рівняння цієї регресії, отримані розрахункові величини зведемо до таблиці:

<i>Вхідні дані</i>			<i>Розрахункові величини</i>					
i	x_i	y_i	Δx_i	Δy_i	$\Delta x_i \Delta y_i$	Δx_i^2	Δy_i^2	$\bar{y}_i = \bar{y}(x)$
1	79	28	4,5	4,34	19,53	20,25	18,84	28,20
2	70	21	-4,5	-2,66	11,97	20,25	7,08	19,79
3	80	27,6	5,5	3,94	21,67	30,25	15,52	29,13
4	71	16,2	-3,5	-7,46	26,11	12,25	55,65	20,72
5	77	29,7	2,5	6,04	15,1	6,25	36,48	26,33
6	77	26,8	2,5	3,14	7,85	6,25	9,86	26,33
7	84	30,3	9,5	6,64	62,08	90,25	44,09	32,87
8	66	15,7	-8,5	-7,96	67,66	72,25	63,36	16,05
9	74	25,5	-0,5	1,84	-0,92	0,25	3,39	23,53
10	67	15,8	-7,5	-7,86	58,95	56,25	61,78	16,99
$\sum$	745	236,6	0	0	293,76	314,5	316,05	239,9
$\frac{1}{n} \sum$	$\bar{x} = 74,5$	$\bar{y} = 23,66$						23,99

Коефіцієнт регресії за вибіркою: $\bar{\rho}_{yx} = \frac{293,76}{314,5} = 0,934$.

Рівняння регресії матиме вигляд:

$$\bar{y}(x) = 23,66 + 0,934(x - 74,5)$$

$$\bar{y}(x) = -45,59 + 0,934x$$

Визначимо щільність зв'язку, для цього оцінимо вибірковий коефіцієнт кореляції:

$$r_{xy} = \frac{\sum_{i=1}^n \Delta x_i \Delta y_i}{\sqrt{\sum_{i=1}^n \Delta x_i^2} \sqrt{\sum_{i=1}^n \Delta y_i^2}} = \frac{293,76}{\sqrt{314,5} \sqrt{316,05}} = 0,93.$$

Оскільки $r_{xy} \approx 1$, то зв'язок між X та Y є щільним.

Довірчий інтервал для лінії регресії

Границі довірчого інтервалу для $M(Y) = a_y$	$\bar{y}', \bar{y}'' = \bar{y} \pm t_{\alpha,k} \cdot \sigma_{\bar{y}}$, де $\sigma_{\bar{y}} = \frac{\tilde{\sigma}_{yp}}{\sqrt{n}}$ – середнє квадратичне відхилення загальної вибіркової середньої; $t_{\alpha,k}$ – величина, яка визначається за таблицею розподілу Стюдента для заданої довірчої ймовірності γ та числом ступенів вільності $k = n - 2$.
Границі довірчого інтервалу для коефіцієнта регресії ρ_{yx}	$\tilde{\rho}'_{yx}, \tilde{\rho}''_{yx} = \bar{\rho} \pm t_{\alpha,k} \cdot \sigma_{\tilde{\rho}_{yx}}$, де $\tilde{\rho}_{yx}$ – точкова оцінка ρ_{yx} , $\sigma_{\tilde{\rho}_{yx}}$ – середнє квадратичне значення коефіцієнту $\tilde{\rho}_{yx}$, причому $\sigma_{\tilde{\rho}_{yx}} = \frac{\tilde{\sigma}_{yp}}{\sqrt{\sum \Delta x_i^2}}$.

5.3. Нелінійна кореляційна залежність (Параболічна регресія та знаходження її рівняння за вибіркою)

Припустимо, що вид теоретичної функції регресії $\bar{y}(x)$ відомий і є многочленом другого порядку $\bar{y} = a_0x^2 + b_0x + c_0$.

При нелінійній кореляційній залежності величин коефіцієнт кореляції не може бути мірою щільності зв'язку. Як показник тісноти зв'язку в загальному випадку застосовується кореляційне відношення h_{yx} або його квадрат $h^2_{yx} = d_{yx}$ – **коефіцієнт детермінації**, причому $0 \leq h_{yx} \leq 1$; $0 \leq d_{yx} \leq 1$.

Рівності $h_{yx} = 0$ або $d_{yx} = 0$ свідчать про відсутність зв'язку між Y та X ; близькість h_{yx} та d_{yx} до 1 вказує, що цей зв'язок є тісним.

Коефіцієнт детермінації обчислюється за формулами:

$$d_{yx} = h^2_{yx} = 1 - \frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{або} \quad d_{yx} = h^2_{yx} = 1 - \frac{\sum_{i=1}^n (\Delta \tilde{y}_i)^2}{\sum_{i=1}^n (\Delta y_i)^2}.$$

Якщо d_{yx} виражається у відсотках, то він показує, на скільки відсотків зміна Y обумовлена зміною X ; величина $100\% - d_{yx}$ показує на скільки відсотків зміна Y обумовлена випадковими факторами.

5.4. Побудова нелінійної регресії за вибіркою за методом найменших квадратів

Система лінійних алгебраїчних рівнянь для отримання коефіцієнтів ліній регресії, а саме a, b, c має вигляд:

$$\begin{cases} a \sum_{i=1}^n x_i^4 + b \sum_{i=1}^n x_i^3 + c \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i^2 y_i \\ a \sum_{i=1}^n x_i^3 + b \sum_{i=1}^n x_i^2 + c \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \\ a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i + cn = \sum_{i=1}^n y_i \end{cases}$$

Розв'язок цієї системи спрощується, якщо ввести підстановку:

$$\Delta x_i = x_i - \bar{x}, \quad \text{де} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

За такої підстановки спочатку визначається параболічне рівняння регресії у вигляді

$$\tilde{y}(x) = a'(x - \bar{x})^2 + b'(x - \bar{x}) + c',$$

де a', b', c' - параметри цього проміжного рівняння, причому

$$a' = \frac{n \sum_{i=1}^n \Delta x_i^2 \Delta y_i}{n \sum_{i=1}^n \Delta x_i^4 - \left(\sum_{i=1}^n \Delta x_i \right)^2}; \quad b' = \frac{\sum_{i=1}^n \Delta x_i \Delta y_i}{\sum_{i=1}^n \Delta x_i^2}; \quad c' = \frac{n \bar{y} \sum_{i=1}^n \Delta x_i^4 - \sum_{i=1}^n \Delta x_i^2 \sum_{i=1}^n \Delta x_i^2 \Delta y_i - \left(\sum_{i=1}^n \Delta x_i^2 \right)^2 \bar{y}}{n \sum_{i=1}^n \Delta x_i^4 - \left(\sum_{i=1}^n \Delta x_i \right)^2}.$$

Проміжне рівняння зводиться до вигляду $\tilde{y}(x) = ax^2 + bx + c$ шляхом обчислення його параметрів за формулами: $a = a'; b = b' + 2a'\bar{x}; c = c' + a'\bar{x}^2 - b'\bar{x}$.

Приклад

Експериментально отримана залежність однієї величини (Y) від іншої величини (X)

N_i	1	2	3	4	5	6	7
x_i	0	1	2	3	4	5	6
y_i	2.89	1.41	0.29	-0.41	-0.69	-0.61	-0.11

Встановити вигляд залежності між величинами, визначити параметри рівняння регресії Y на X та щільність зв'язку між ними.

Розв'язання

1) Для встановлення виду залежності Y від X побудуємо кореляційне поле:

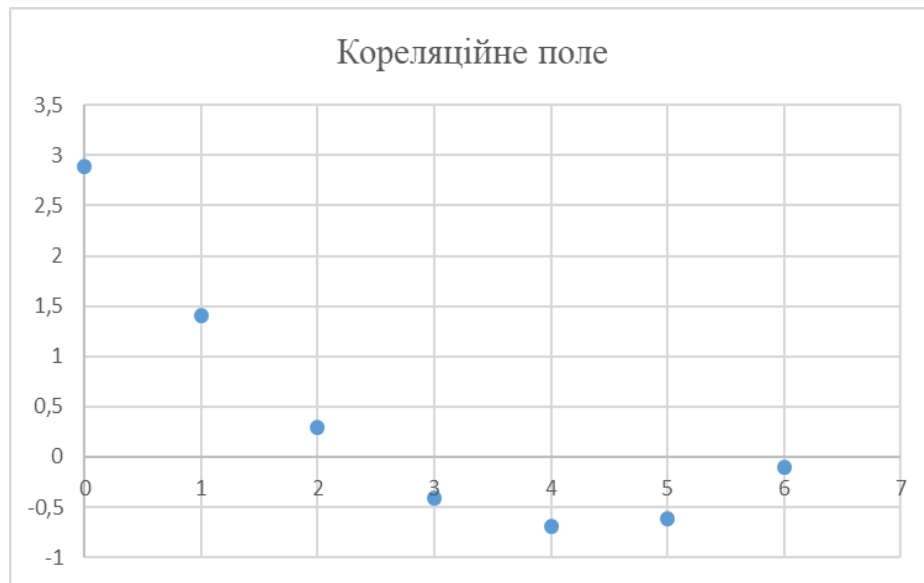


Рисунок 1

Аналіз кореляційного поля свідчить про те, що залежність між Y та X є параболічною та може бути описана рівнянням: $y(x) = ax^2 + bx + c$.

2) Визначимо параметри цього рівняння, результати обчислень зведемо до таблиці:

Вхідні дані			Розрахункові дані						
i	x_i	y_i	Δx_i	Δy_i	$\Delta x_i \Delta y_i$	Δx_i^2	$\Delta x_i^2 \Delta y_i$	Δx_i^4	$\tilde{y}_i = \tilde{y}(x_i)$
1	0	2,89	-3	2,19	-7,47	9	22,41	81	2,9
2	1	1,41	-2	1,01	-2,02	4	4,04	16	1,4
3	2	0,29	-1	-0,11	0,11	1	-0,11	11	0,3
4	3	-0,41	0	-0,81	0	0	0	0	0,4
5	4	-0,69	1	-1,09	-1,09	1	1,09	1	-0,7
6	5	-0,61	2	-1,01	-2,02	4	-4,04	16	-0,16
7	6	-0,11	3	-0,51	-1,53	9	-4,56	81	-0,1
Σ	21	2,77	0	-0,03	-14,02	28	16,62	19,6	2,8
$\frac{1}{n} \Sigma$	$\bar{x} = 3$	$\bar{y} = 0,4$							

Обчислимо проміжні параметри:

$$a' = \frac{n \sum_{i=1}^n \Delta x_i^2 \Delta y_i}{n \sum_{i=1}^n \Delta x_i^4 - \left(\sum_{i=1}^n \Delta x_i^2 \right)^2} = \frac{7 \cdot 16,62}{7 \cdot 196 - 28^2} = 0,2,$$

$$b' = \frac{\sum_{i=1}^n \Delta x_i \Delta y_i}{\sum_{i=1}^n \Delta x_i^2} = -\frac{14,02}{28} = -0,5,$$

$$c' = \frac{n \bar{y} \sum_{i=1}^n \Delta x_i^4 - \sum_{i=1}^n \Delta x_i^2 \sum_{i=1}^n \Delta x_i^2 \Delta y_i - \left(\sum_{i=1}^n \Delta x_i^2 \right)^2 \bar{y}}{n \sum_{i=1}^n \Delta x_i^4 - \left(\sum_{i=1}^n \Delta x_i^2 \right)^2} = \frac{7 \cdot 0,4 \cdot 196 - 28 \cdot 16,02 - 28^2 \cdot 0,4}{7 \cdot 196 - 28^2} = -0,4.$$

Тоді

$$a = a' = 0,2;$$

$$b = b' + 2a'\bar{x} = -0,5 - 2 \cdot 0,2 \cdot 3 = -1,7;$$

$$c = c' + a'\bar{x}^2 - b'\bar{x} = -0,4 + 0,2 \cdot 3^2 - (-0,5) = 2,9$$

Рівняння регресії набуде вигляду:

$$\tilde{y}(x) = ax^2 + bx + c = 0,2x^2 - 1,7x + 2,9.$$

3) Визначимо щільність зв'язку між величинами. Для цього обчислимо коефіцієнт детермінації за формулою

$$d_{yx} = h_{yx}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (\Delta \tilde{y}_i)^2}{\sum_{i=1}^n (\Delta y_i)^2}$$

Результати обчислень зведемо до таблиці:

№	y_i	Δy_i	$(\Delta y_i)^2$	$\tilde{y}_i$	$\Delta \tilde{y}_i$	$(\Delta \tilde{y}_i)^2$
1	2,82	2,49	6,2	2,9	-0,01	10^{-4}
2	1,41	1,01	1,02	1,4	0,01	10^{-4}
3	0,29	-0,11	0,01	0,3	-0,01	10^{-4}
4	-0,41	-0,81	0,66	-0,4	-0,01	10^{-4}
5	-0,69	-1,09	1,19	-0,7	0,01	10^{-4}
6	-0,61	-1,01	1,02	-0,6	-0,01	10^{-4}
7	-0,11	-0,51	0,26	-0,1	-0,01	10^{-4}

Σ	2,77	≈ 0	10,36	2,8	$=0$	$7 \cdot 10^{-4}$
----------	------	-------------	-------	-----	------	-------------------

$$d_{yx} = h_{yx}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (\Delta \tilde{y}_i)^2}{\sum_{i=1}^n (\Delta y_i)^2} = 1 - \frac{7 \cdot 10^{-4}}{10,36} = 1 - 0,7 \cdot 10^{-4} \approx 1.$$

Оскільки коефіцієнт детермінації близький до одиниці, то можна вважати, що між величинами Y та X існує щільний зв'язок.

Зміст

1.	Випадкові події.....	3
1.1.	Види подій.....	3
1.2.	Елементи комбінаторики.....	3
1.3.	Означення ймовірності події.....	4
1.4.	Властивості ймовірності.....	4
1.5.	Основні теореми.....	4
2.	Випадкові величини.....	7
2.1.	Закон розподілу.....	7
2.2.	Властивості функцій $F(x)$ та $f(x)$	7
2.3.	Числові характеристики випадкових величин.....	8
2.4.	Закони розподілу дискретних випадкових величин (ДВВ).....	9
2.4.1.	Рівномірний закон розподілу.....	9
2.4.2.	Біноміальний закон розподілу.....	10
2.4.3.	Гіпергеометричний закон розподілу.....	11
2.4.4.	Розподіл Пуассона.....	11
2.5.	Закони розподілу неперервних випадкових величин (НВВ).....	12
2.5.1.	Рівномірний закон розподілу.....	12
2.5.2.	Експоненціальний розподіл.....	13
2.5.3.	Розподіл Вейбула.....	14
2.5.4.	Нормальний закон розподілу (закон Гауса).....	14
3.	Елементи математичної статистики.....	17
3.1.	Статистична сукупність. Генеральна сукупність та вибірка.....	17
3.2.	Статистичний ряд. Статистична функція розподілу. Гістограма та полігон.....	17
3.3.	Статистична функція розподілу.....	18
3.4.	Статистичні оцінки параметрів розподілу.....	20
3.4.1.	Точкові оцінки.....	21
3.4.2.	Способи оцінювання параметрів.....	21
4.	Статистичні гіпотези.....	23
4.1.	Критерії та загальний алгоритм перевірки статистичної гіпотези.....	23
4.1.1.	Розподіл Стьюдента.....	23
4.1.2.	Розподіл Фішера.....	23
4.1.3.	Критерій згоди Пірсона (критерій χ^2).....	24
5.	Кореляційна залежність.....	25
5.1.	Точкові оцінки лінії регресії та коефіцієнта кореляції.....	25
5.2.	Побудова лінійної регресії за вибіркою за методом найменших квадратів...	26
5.3.	Нелінійна кореляційна залежність.....	28
5.4.	Побудова нелінійної регресії за вибіркою за методом найменших квадратів	28
	Література	34

Література

1. Барковський В. , Барковська Н., Лепатін О. Теорія ймовірностей та математична статистика. Київ: Центр навчальної літератури, 2019. 424 с.
2. Кармелюк Г. Теорія ймовірностей та математична статистика. Київ: Центр навчальної літератури, 2019. 576 с.
3. Огірко О.Ш., Галайко Н.В. Львів: ЛьвДУВС, 2017. 292 с.
4. Драгомирецька Х.Т. Теорія ймовірностей та математична статистика. Львів: Львівська політехніка, 2012. 352 с.
5. Руденко В.М. Математична статистика. Київ: Центр навчальної літератури, 2019, 336с.

Інформаційні ресурси в Інтернеті:

1. <http://ir.nmu.org.ua/handle/123456789/2238>
2. <http://maup.com.ua/assets/files/lib/book/nw22.pdf>
3. http://eprints.kname.edu.ua/14499/1/%D0%BF%D0%B5%D1%87._%D0%9B%D0%B5%D0%BA-%D0%9C%D0%9F-%5B2%2B%5D.pdf
4. <http://www.tsatu.edu.ua/vmf/wp-content/uploads/sites/17/konspekt-lekcij-halanchuk-na-sajt-s-oblozhkoj.pdf>

Навчальне видання

Лінкова Олена Володимирівна
Гарбуз Артем Ігорович

Вища математика та математична статистика

Навчально-наочний посібник
для студентів очної та дистанційної форми навчання.
частина 4

Видається в авторській редакції

ДУІТЗ
65023, Одеса, вул. Кузнечна, 1
Обсяг 2,0 д.а., електронна версія