Міністерство освіти і науки України
Сумський державний університет

О. А. Гончаров, І. О. Князь, О. В. Хоменко

ТЕОРІЯ ЙМОВІРНОСТЕЙ

І МАТЕМАТИЧНА

СТАТИСТИКА

Навчальний посібник

Рекомендовано вченою радою
Сумського державного університету

Суми
Сумський державний університет
2022

УДК 519.2 (075.8)
Г 65

Рецензенти:
Корніч Г. В., доктор фізико-математичних наук, професор, завідувач кафедри системного аналізу та обчислювальної математики НУ «Запорізька політехніка»;
Шуда І. О. , доктор фізико-математичних наук, професор, зав. Кафедри математичного аналізу та методів оптимізації.

Рекомендовано до видання
вченою радою Сумського державного університету
як навчальний посібник
(протокол № 2 від08 вересня 2022 року)

Гончаров О. А., Князь І. О., Хоменко О. В.
Г 65 Теорія ймовірностей і математична статистика : навчальний посібник / Гончаров О. А., Князь І. О.,Хоменко О. В. – Суми : СумДУ, 2022. – 174 с.

ISBN

Посібник містить основні теоретичні положення та практичні вказівки з курсу теорії ймовірностей та математичної статистики, варіанти контрольних і тестових завдань, методичні рекомендації до їх виконання, а також перелік необхідної літератури.

УДК 519.2 (075. 8)

ISBN

ЗМІСТ

ВСТУП

Теорія ймовірностей - математична наука, що вивчає кількісні закономірності випадкових масових явищ. Теорія ймовірностей є також теоретичною основою математичної та прикладної статистики. У зв’язку з розвитком масових процесів у виробництві та економіці, а також через необхідність більш тонкого аналізу результатів експерименту, математичного моделювання та розв’язання оптимізаційних задач, імовірнісні й статистичні методи широко проникли в економічні, технічні й технологічні науки.

У нашому посібнику наведено основні теоретичні матеріали й задачі для організації самостійної роботи студентів з основ теорії ймовірностей і математичної статистики. Кількість варіантів кожного виду задач подана з надлишком: 30 варіантів, (а для контрольних робіт студентів-заочників потрібно 25). Додаткові варіанти можна використовувати як демонстраційні на лекціях і практичних заняттях. Крім того, в академічних групах стаціонару може бути більше 25 студентів.

Об’єм необхідних знань міститься в типовій програмі курсу.

1 Випадкові події та їх імовірності:

а) простір елементарних подій; класифікація подій; статистичне, класичне й геометричне визначення ймовірності випадкової події;

б) основні теореми теорії ймовірностей: теорема додавання ймовірностей, умовна ймовірність, теореми множення ймовірностей; формула повної ймовірності, формула Байеса;

в) формула Бернуллі для схеми повторних незалежних випробувань, локальна та інтегральна формули Лапласа, асимптотична формула Пуассона.

2 Випадкові величини:

а) випадкові величини та їх числові характеристики; математичне сподівання, дисперсія й середнє квадратичне відхилення;

б) закони розподілення дискретних випадкових величин;

в) закони розподілення неперервних випадкових величин; диференціальна та інтегральна функції розподілення, їх властивості та ймовірнісний зміст;

г) нормальний закон розподілення, його числові характеристики; графіки та властивості диференціальної й інтегральної функцій розподілення;

д) закон великих чисел; нерівність Чебишева; теорема Чебишева.

3 Системи двох випадкових величин:

а) функція розподілення, її властивості; густина розподілення, властивості; числові характеристики системи випадкових величин; кореляційний момент, коефіцієнт кореляції; найпростіші випадки нелінійної кореляції;

б) рівняння регресії, залежність між коефіцієнтами регресії.

4 Основи математичної статистики:

а) генеральна сукупність і вибірка, вибіркові числові характеристики; перевірка гіпотези про закон розподілення; критерій Пірсона;

б) інтервал надійності та ймовірність надійності; точкові та інтервальні оцінки параметрів розподілення; розподілення Стьюдента та χ^2;

в) випадкові процеси, класифікація; ланцюги Маркова, їх властивості; марковські процеси.

5 Статистичне дослідження закономірностей:

а) основні гіпотези регресійного аналізу; метод найменших квадратів; кореляційне відношення, його властивості, кореляційне відношення за умов лінійної та нелінійної регресій;

б) перевірка гіпотези про значущість вибіркового коефіцієнта кореляції, побудова інтервалу надійності для коефіцієнта кореляції генеральної сукупності, область надійності для рівняння регресії;

в) перевірка гіпотези про лінійність моделі; критерій Фішера;

г) множинна кореляційна залежність; звідний коефіцієнт множинної кореляції;

д) стаціонарні випадкові процеси, числові характеристики за результатами досліду; кореляційна функція випадкового процесу.

Номер варіанта для контрольних робіт студентів заочної форми навчання визначаю за двома останніми цифрами в номері залікової книжки (табл. 1).

Таблиця 1 Вибір варіанту студентом

Дві останні цифри залікової книжки	01	02	03	….	23	24	25
	26	27	28	....	48	49	50
	51	52	53	….	73	74	75
	76	77	78	….	98	99	00
Номер варіанта	1	2	3	….	23	24	25

ЧАСТИНА І. ОСНОВНІ ПОНЯТТЯ ТЕОРІЇ ЙМОВІРНОСТЕЙ

Розділ 1. Основні поняття теорії ймовірностей та комбінаторики

1.1 Елементи комбінаторики. Розміщення, перестановки, комбінації,

Комбінаторикою називають розділ математики, що вивчає правила групування елементів множини. Ров’язання задач ґрунтується на основному правилі та принципах комбінаторики, конкретних способах підрахунку кількості підмножин, які можна утворити з певної множини елементів.

Основне правило комбінаторики. Між скінченними множинами A та B можна встановити взаємно однозначну відповідність тоді й лише тоді, коли вони мають однакову кількість елементів.

Основне правило комбінаторики дозволяє виконувати обчислення кількості елементів множини фактично зведенням до множини з відомою кількістю елементів або до множини, для якої кількість можна підрахувати дещо простіше. Здебільшого його використання не акцентуються, а приймається як дещо зрозуміле.

Значна кількість формул і теорем комбінаторики ґрунтується на двох основних елементарних принципах, що називають принципами суми й добутку.

Принцип суми. Нехай $n(A)$, $n(B)$ – кількості елементів скінченних неперетинних множин A та B відповідно. Тоді для множини $C = A \cup B$ кількість елементів

$$n(C) = n(A) + n(B). \tag{1.1}$$

Принцип добутку. Нехай потрібно послідовно виконати k дій, причому першу дію можна виконати n_1 способом, після чого другу дію – n_2 способами й так далі до k-ї дії, яку можна виконати n_k способами. Тоді всі k дій можна виконати $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ способами.

Розміщенням з n елементів за k називають довільну впорядковану підмножину з k елементів із множини M (множина з n елементів).

Два розміщення вважають різними не лише тоді, коли вони складаються з різних елементів, а й тоді, коли вони складаються з однакових елементів, але відрізняються порядком їх розташування.

Кількість розміщень A_n^k з n елементів за k обчислюють за формулою

$$A_n^k = \frac{n!}{(n-k)!} = n(n-1)(n-2)\cdot\ldots\cdot(n-k+1)\,. \tag{1.2}$$

Приклад. Скласти всі двозначні числа з трьох цифр 1, 2, 3 (числа з однаковими цифрами не враховувати).

Розв’язання. Загальну кількість різних чисел можна підрахувати за формулою (1.2): $A_3^2 = 3\cdot 2 = 6$. Маємо 12, 13, 21, 23, 31, 32.

Розміщенням з повтореннями з n елементів за k називають довільну впорядковану підмножину з k елементів із множини M. Водночас елементи не обов’язково повинні бути різними. Кількість розміщень із повтореннями з n елементів за k обчислюють за формулою

$$\overline{A_n^k} = n^k\,. \tag{1.3}$$

Приклад. Скількома способами можна розкласти 3 предмети в 4 ящики.

Розв'язання. Кількість таких способів дорівнює

$$\overline{A_4^3} = 4^3 = 64\,.$$

Перестановками називають розміщення з n елементів за n.

Різні перестановки відрізняються лише порядком розташування елементів. Кількість перестановок P_n з n елементів дорівнює добутку всіх натуральних чисел від 1 до n

$$P_n = n! \tag{1.4}$$

Приклад. Скількома способами можна розмістити 4 особи за столом?

Розв'язання. За формулою (1.4) одержимо: $P_4 = 4! = 24$.

Перестановками з повтореннями називають розміщення з n елементів, що є однотипними.

Кількість різних перестановок із повторенням, що можна утворити з n елементів, серед яких є k_1 елементів першого типу, k_2 елементів другого типу, …, k_m елементів m-го типу

$$P_n = \frac{n!}{k_1!k_2!\ldots k_m!}\,. \tag{1.5}$$

Приклад. Скільки різних п'ятибуквених слів можна скласти з букв слова «манна»?

Розв'язання. У слові букви «а» та «н» повторюються 2 рази, а буква «м» один раз. За формулою (1.5) одержуємо

$$P_5 = \frac{5!}{2!2!1!} = 30\,.$$

Комбінацією з n елементів за k називають довільну підмножину з k елементів із множини M, що відрізняється від неї принаймні одним елементом.

Порядок розташування елементів у комбінаціях не має значення. Кількість комбінацій з n елементів за k ($k \le n$) позначають C_n^k та обчислюють за формулою

$$C_n^k = \frac{n!}{k!(n-k)!}\,. \tag{1.6}$$

Приклад. Скількома способами можна обрати представників колективу з 30 студентів у складі трьох осіб?

Розв'язання. Кількість способів дорівнює кількості комбінацій із 30 елементів за 3

$$C_{30}^3 = \frac{30!}{3!\,27!} = 4060\,.$$

Комбінацією з повтореннями з n елементів за k називають довільну множину з k елементів із множини M (елементи не обов'язково різні).

Кількість комбінацій із повтореннями з n елементів за k обчислюють за формулою

$$N_n^k = \frac{(n+k-1)!}{k!(n-1)!}\,. \tag{1.7}$$

1.2 Простір елементарних подій. Випадкові події.

Експерименти, які в незмінних умовах можна повторити будь-яку кількість разів, але результати яких непередбачувані, у теорії ймовірностей називають **випробуваннями**.

Елементарною подією називають найпростіший результат випробування в певному експерименті.

Сукупність усіх можливих елементарних подій випробування називають **простором елементарних** подій і позначають знаком Ω.

Підмножина A простору елементарних подій називається **випадковою подією**.

Достовірною назвемо таку подію, що обов'язково відбудеться за умови виконання сукупності умов досліду. **Неможливою** назвемо подію, яка при таких умовах завідомо не може відбутися. Випадкова подія може відбутися, а може не відбутися за одних і тих самих умов.

Добутком двох подій A та B є така подія $C = A \cap B$ ($C = AB$), що полягає в одночасному виконанні обох подій A та B. Події A та B несумісні, коли їхній добуток дорівнює нулю, тобто є неможливим.

Сумою двох подій A та B є така подія $C = A \cup B$ (або $C = A+B$), що полягає в наставанні хоча б однієї з подій A і B.

Дві події A та B є **протилежними** ($A = \overline{B}$), коли їх добуток є неможливою подією, а сума, – достовірною подією.

Різницею подій A та B називають подію $C = A \backslash B$ (або $C = A - B$), що настає тоді, коли настає подія A й не настає подія B.

Події A_1, $A_2 \dots A_n$ утворюють повну групу, якщо вони попарно несумісні, а в сумі дають достовірну подію.

Властивості операцій над подіями

Додавання подій:

1) $A + B = B + A$;

2) $(A+B)+C = A+(B+C) = A+B+C$;

3) $A+B = A, \quad B \subset A$;

4) $A+B = B, \quad A \subset B$;

5) $A+A = A$;

6) $A+\overline{A} = E$**,** E *– достовірна подія;*

7) $A+E = E$;

8) $A+\varnothing = A$.

Множення подій:

1) $AB = BA$;

2) $(AB)C = A(BC) = ABC$;

3) $AB = A, \quad A \subset B$;

4) $AB = B, \quad B \subset A$;

5) $AA = A$;

6) $A\overline{A} = \varnothing$;

7) $AE = A$;

8) $A\varnothing = \varnothing$

1.3 Ймовірності подій

Для кількісного оцінювання можливості появи випадкової події A вводять поняття ймовірності. Ймовірністю P даної події A назвемо відношення числа випадків m, які сприяють появі даної події, до загальної кількості n несумісних, рівно можливих і єдино можливих результатів випробувань, тобто:

$$P = \frac{m}{n}. \tag{1.8}$$

Приклад. До супермаркету надійшло 15 холодильників, 10 із них виготовлені донецьким заводом, а інші – дніпропетровським заводом. Знайти ймовірність того, що серед п'яти навмання вибраних холодильників виявляться три, виготовлених донецьким заводом.

Розв’язання. Нехай A – подія: серед п’яти навмання вибраних холодильників виявиться три, виготовлених донецьким заводом. Скористуємось формулою (1.8)

$$P(A)=\frac{m}{n},$$

де m – число випадків, що сприяють появі даної події, n – загальна кількість випадків.

Тут n – кількість способів вибрати 5 холодильників із 15

$$n=C_{15}^{5}=\frac{15!}{5!\cdot 10!}=\frac{15\cdot 14\cdot 13\cdot 12\cdot 11}{5\cdot 4\cdot 3\cdot 2}=3\cdot 7\cdot 13\cdot 11=3003,$$

а m – кількість способів вибрати 2 холодильники з п’яти, виготовлених донецьким заводом і 3 холодильники з 10, виготовлених дніпропетровським заводом

$$m=m_1\cdot m_2.$$

Тут m_1 – кількість способів вибрати 2 холодильники з п'яти

$$m_1=C_5^2=\frac{5!}{2!\cdot 3!}=\frac{5\cdot 4}{2}=10,$$

m_2 – кількість способів вибрати 3 холодильники з 10

$$m_2=C_{10}^{3}=\frac{10!}{3!\cdot 7!}=\frac{10\cdot 9\cdot 8}{3\cdot 2}=5\cdot 3\cdot 8=120.$$

Отже, $m=10\cdot 40=400$. Тоді

$$P(A)=\frac{10\cdot 5\cdot 3\cdot 8}{3\cdot 7\cdot 13\cdot 11}=\frac{1200}{3003}\approx 0,4.$$

До основних понять теорії ймовірностей, разом з імовірністю, належить відносна частота.

Відносною частотою події називають відношення кількості випробувань, у яких подія настала m разів, до загальної кількості n фактично проведених випробувань: m/n.

Хоча, за зовнішнім виглядом, формули для обчислення ймовірності й відносної частоти однакові, самі поняття не є тотожними. Пояснимо це твердження «словесними формулами»

$$P=\frac{m}{n}:$$

$$\text{ймовірність події}=\frac{\text{кількість сприятливих випадків}}{\text{загальна кількість випадків}},$$

$$p=\frac{m}{n}:$$

$$\text{відносна частота події}=\frac{\text{кількість появ події}}{\text{загальна кількість випробувань}}.$$

Різниця між імовірністю й відносною частотою полягає в тому, що ймовірність обчислюють до експерименту, а відносна частота може бути обчислена лише після проведення досліду. Важливою властивістю відносної частоти є її стійкість. Ця властивість є теоретичною основою математичної статистики.

Статистичною ймовірністю події A називають межу, до якої наближається відносна частота m/n події A за умови необмеженого збільшення кількості всіх випробувань n, тобто

$$P(A)=\lim_{n\to\infty}\frac{m}{n}.$$

Статистичний спосіб завдання ймовірності має перевагу. Він грунтується на реальних випробовуваннях, але має й недоліки, бо для надійного визначення ймовірності необхідно провести велику кількість випробувань. Статистичне означення ймовірності події хоча й досить повно відображає зміст цього поняття, але не дає можливості фактичного обчислення ймовірності, тобто не є «робочим» означенням.

Розділ 2. Основні теореми теорії ймовірностей

2.1 Теореми додавання для сумісних і несумісних ймовірностей

Дві події називають **несумісними**, якщо вони не можуть одночасно настати в одному досліді, іншими словами, настання однієї події виключає можливість настання другої.

Теорема додавання для несумісних подій. *Якщо події A та B несумісні* ($A \cap B = \varnothing$)*, причому відомі їх імовірності P(A) і P(B), то ймовірність суми цих подій дорівнює сумі їх імовірностей*

$$P(A \cup B) = P(A) + P(B). \tag{2.1}$$

Доведення. Дійсно, нехай n – кількість усіх елементарних подій у певному досліді; m_1 – кількість елементарних подій, сприятливих події A; m_2 – кількість елементарних подій, сприятливих події B. Тоді настанню події $P(A \cup B)$ сприяють m_1+m_2 елементарних подій. Отже, за класичним означенням імовірності

$$P(A\cup B)=\frac{m_1+m_2}{n}=\frac{m_1}{n}+\frac{m_2}{n}=P(A)+P(B).$$

Теорема допускає узагальнення в разі суми скінченої (або зліченої) кількості попарно несумісних подій

$$P(A_1\cup A_2\cup A_3\cup\ldots\cup A_n) = P(A_1) + P(A_2) + P(A_3) +\ldots + P(A_n).$$

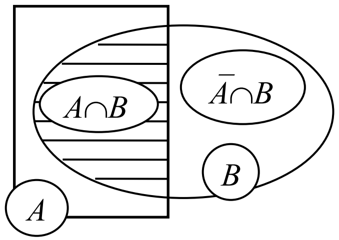
Звідси випливають такі властивості ймовірностей:

$$P(\bar{A}) = 1-P(A);$$

$$0\le P(A)\le 1.$$

Теорема додавання для сумісних імовірностей. *Для будь-яких подій А та В справедливе співвідношення*

$$P(A\cup B)=P(A)+P(B)-P(A\cap B). \qquad (2.2)$$



Доведення. Подамо події $A \cup B$ та B у вигляді суми попарно несумісних подій $B=A \cap B + \overline{A} \cap B$, $A\cup B=A+\overline{A} \cap B$. Тоді на підставі (2.1) одержимо $P(A\cup B)=P(A)+P(\overline{A} \cap B)$ і $P(B)=P(A\cap B)+P(\overline{A} \cap B)$. Підставляючи вираз $P(\overline{A} \cap B)$ із другого співвідношення в перше, приходимо до рівності (2.2).

2.2 Умовні ймовірності та незалежні події

Як зазначалося, в основі означення ймовірності випадкової події лежить сукупність певних умов. Якщо ж ніяких інших обмежень, окрім цих умов, у разі обчислення ймовірності $P(A)$ не накладається, така ймовірність називають ***безумовною***. Якщо ж подія A відбувається за умови, що відбулась інша подія B, то ймовірність настання події A називають ***умовною*** (позначають $P(A\,|\,B)$ або $P_B(A)$) та обчислюють за формулою

$$P(A\,|\,B)=\frac{P(A\cap B)}{P(B)},\quad P(B)>0. \qquad (2.3)$$

Події A та B $(P(A)>0,\ P(B)>0)$ називають **незалежними**, якщо $P(A\,|\,B)=P(A)$, $P(A\,|\,B)=P(B)$, тобто, якщо настання однієї з них не змінює ймовірності настання другої. У протилежному разі події A і B називають залежними.

Зі співвідношення (2.3) випливає **теорема множення ймовірностей:** імовірність сумісної появи двох залежних подій A та B дорівнює

$$P(AB)=P(A)P(B/A)=P(B)P(A/B). \qquad (2.4)$$

Теорема. Події A та B незалежні тоді і тільки тоді, коли

$$P(AB)=P(A)\,P(B). \qquad (2.5)$$

Твердження теореми випливає з формули (2.3) та з означення незалежності подій.

Приклад. В урні 5 білих і 7 чорних кульок. Із неї почергово випадково виймають дві кулі двома способами:

а) навмання виймають кулю, фіксують її колір і куля повертається в урну, після чого всі кулі в урні перемішують.

б) вийняту кулю в урну не повертають.

Яка ймовірність того, що обидві вийняті кулі білі?

Розв'язання. У способі реєстрації а) можна вважати обидві спроби незалежними (за умов ретельного перемішування) й потрібна ймовірність дорівнює (2.5)

$$P=\frac{5}{12}\cdot\frac{5}{12}=\frac{25}{144}. \qquad .$$

У способі реєстрації б) ймовірність білого кольору другої кулі залежить від кольору першої витягнутої кулі і результат розрахунку інший (формула (2.4)):

$$P = \frac{5}{12} \cdot \frac{4}{11} = \frac{5}{33}.$$

Звичайно, у разі б) можливий розрахунок імовірності й комбінаторним способом:

$$P = \frac{m}{n} = \frac{C_5^2}{C_{12}^2} = \frac{10}{66} = \frac{5}{33}.$$

Приклад. Знайти ймовірність того, що під час кидання трьох гральних костей цифра «шість» з'явиться хоча б на одній із них.

Розв'язання. Імовірність, що випаде 6 в одному кидку кості дорівнює 1/6. Імовірність того, що не випаде 6 дорівнює 5/6. Імовірність того, що під час кидання трьох гральних костей не випаде 6 ні одного разу, дорівнює $P = \left(\frac{5}{6}\right)^3 = \frac{125}{216}$. Тоді ймовірність того, що хоча б один раз випаде 6 дорівнює: $1 - \frac{125}{216} = \frac{91}{216}$.

2.2 Формули повної ймовірності та Байєса

Теорема. Нехай $H_i \ (i = \overline{1, n})$ – повна група несумісних між собою подій ($P(H_i) > 0 \ \ \forall i$), нехай випадкова подія A настає лише після настання однієї з подій $H_i \ (i = \overline{1, n})$, тоді актуальна рівність

$$P(A) = \sum_{i=1}^{n} P(H_i) P(A / H_i). \qquad (2.6)$$

Доведення.

$$P(A) = P(EA) = P\left(\left(\bigcup_{i=1}^{n} H_i\right) A\right) = P\left(\bigcup_{i=1}^{n} (H_i A)\right) = \sum_{i=1}^{n} P(H_i A) =$$
$$= \sum_{i=1}^{n} P(H_i) P(A / H_i).$$

Формулу (2.6) називають формулою повної ймовірності. Її можна узагальнити й на зліченне число подій (гіпотез), якщо останні утворюють повну групу.

Теорема. Нехай $H_i \ (i = \overline{1, n})$ – повна група подій ($P(H_i) > 0 \ \ \forall i$). Тоді для будь-якої випадкової події A, що вже настала після настання однієї з гіпотез $H_i \ (i = \overline{1, n})$, актуальні рівності

$$P(H_i / A) = \frac{P(H_i) P(A / H_i)}{\sum_{i=1}^{n} P(H_i) P(A / H_i)} \quad (i = \overline{1, n}). \qquad (2.7)$$

Доведення. В умовах теореми згідно з теоремою множення ймовірностей

$$P(H_i A) = P(H_i) P(A / H_i) = P(A) P(H_i / A).$$

Звідки з урахуванням формули повної ймовірності маємо

$$P(H_i / A) = \frac{P(AH_i)}{P(A)} = \frac{P(H_i) P(A / H_i)}{\sum_{i=1}^{n} P(H_i) P(A / H_i)}.$$

Формулу (2.7) називають теоремою гіпотез, або формулою Байєса.

Приклад. Імовірність того, що під час свердлування будуть знайдені ґрунтові води дорівнює 0,25. З імовірністю 0,5 разом із ґрунтовими водами трапляються тверді породи. Якщо ґрунтових вод немає, то тверді породи трапляються з імовірністю 0,8. Знайти ймовірність того, що під час свердлування будуть знайдені тверді породи.

Розв'язання. Нехай H_1 – подія, яка полягає в тому, що під час свердлування будуть знайдені ґрунтові води; H_2 подія, яка полягає в тому, що під час свердлування не будуть знайдені ґрунтові води.

Події H_1, H_2 – несумісні та утворюють повну групу подій.

Подія A – полягає в тому, що під час свердлування будуть знайдені тверді породи. Імовірність події A обчислимо за формулою повної ймовірності (2.6)

$$P(A)=P(H_1)P_{H_1}(A)+P(H_2)P_{H_2}(A)$$

Враховуючи, що

$$P(H_1)=0,25\,,\;P(H_2)=0,75\,,\;P_{H_1}(A)=0,5\,,\;P_{H_2}(A)=0,8\,,$$

знайдемо ймовірність події А:

$$P(A)=0,25\cdot 0,5+0,75\cdot 0,8=0,125+0,6=0,725\,.$$

Приклад. Імовірність того, що вироби задовольняють вимоги стандарту, дорівнює 0,95. Пропонують спрощену систему контролю, яка пропускає з ймовірністю 0,98 вироби, що задовольняють стандарт, з імовірністю 0,05 вироби, які стандарту не задовольняють.

Знайти ймовірність того, що виріб, який пройшов цей контроль, задовольняє стандарт.

Розв'язання. Нехай A – подія, яка полягає в тому, що виріб пройшов контроль; H_1 – подія, яка полягає в тому, що виріб задовольняє стандарт; H_2 – подія, яка полягає в тому, що виріб не задовольняє стандарту. Події H_1 і H_2 утворюють повну групу несумісних подій

Подія A уже відбулась, потрібно знайти ймовірність події H_1.

За умовами

$$P(H_1)=0,95\,,\;P(H_2)=1-0,95=0,05\,,$$

$$P_{H_1}(A)=0,98\,,\;\;P_{H_2}(A)=0,05\,.$$

Імовірність події A (знаменник формули Байеса)

$$P(A)=P(H_1)P_{H_1}(A)+P(H_2)P_{H_2}(A)=0,95\cdot 0,98+0,05\cdot 0,05=$$
$$=0,931+0,0025=0,933.$$

Тоді ймовірність того, що виріб, який пройшов онтроль, задовольняє стандарту (2.7):

$$P_A(H_1)=\frac{0,95\cdot 0,98}{0,9335}=0,9973\,.$$

2.3 Формули Бернуллі, Лапласа й Пуассона

Якщо n випробувань проводити за однакових умов і ймовірність настання події A в усіх випробуваннях однакова та не залежить від настання або ненастання A в інших випробуваннях, таку послідовність незалежних випробувань називають ***схемою Бернуллі.***

Розглянемо серію з n незалежних випробувань із двома можливими наслідками, в кожному з яких подія A може відбуватися з імовірністю p. Нехай A_j ($j = 1,2,...,n$) позначає подію, що означає наставання події A у j-му випробуванні. Тоді кожну з 2^n елементарних подій серії можна зобразити у вигляді добутку n множників, кожен із яких дорівнює A_j або $\overline{A}_j$.

Теорема. Ймовірність $P_n(k)$ того, що в серії з n випробовувань подія настає k раз, задається рівністю

$$P_n(k) = C_n^k p^k q^{n-k}, \quad q = 1-p, \quad k = 0,1,2,... \tag{2.8}$$

Доведення. Подіям, що нас цікавлять, сприяють ті елементарні події, у яких події A_j спостерігаються k разів, а події $\overline{A}_j$ – $(n-k)$ разів (наприклад, $A_1 \cdot A_2 \cdot ... \cdot A_k \cdot \overline{A}_{k+1} \cdot ... \cdot \overline{A}_n$, $A_1 \cdot A_2 \cdot ... \cdot A_{k-1} \cdot \overline{A}_k \cdot A_{k+1} \cdot \overline{A}_{k+2} \cdot ... \cdot \overline{A}_n$ тощо.). З огляду незалежність подій A_j імовірність кожної такої елементарної події на підставі теореми множення ймовірностей дорівнює $p^k q^{n-k}$. Оскільки подібних елементарних подій буде C_n^k, то з урахуванням їх несумісності й теореми додавання ймовірностей остаточно одержимо формулу (2.8).

Приклад. На кожному із двох крил літака установлені по два двигуни, кожен із яких може вийти з ладу під час польоту незалежно один від одного з імовірністю $P = 0,1$. Яка ймовірність того, що політ закінчиться нормально, якщо: а) літак може летіти на будь-яких двох двигунах;

б) літак може летіти за умови, що на кожному крилі працює хоча б один двигун.

Розв'язання. Позначимо через A подію, яка полягає в тому, що політ закінчиться нормально.
1) Нехай D_k – подія, яка полягає у тому, що під час польоту вийдуть із ладу лише k ($k = 0,1,2,3,4$) двигуни. Тоді $\overline{A} = D_3 \cup D_4$, де події D_3 і D_4 несумісні. Отже, $P(\overline{A}) = P(D_3) + P(D_4)$. Під час обчислення $P(D_3)$ і $P(D_4)$ скористаємося схемою Бернуллі за умови $n = 4$ і $p = 0.1$:

$$P(D_3) = p_4(3) = C_4^3 \cdot p^3 \cdot (1-p) = 4 \cdot (0,1)^3 \cdot 0.9 = 0,0036;$$

$$P(D_4) = p_4(4) = C_4^4 \cdot p^4 \cdot (1-p)^0 = 4 \cdot (0,1)^4 = 0,0001.$$

Тому $P(A) = 1 - P(\overline{A}) = 1 - 0,0036 - 0,0001 = 0,996$.

2) Нехай $П_k$ ($Л_k$) – подія, яка полягає в тому, що під час польоту на правому (лівому) крилі вийдуть із ладу лише k ($k = 0,1,2$) двигуни.

Тоді подію A можна подати у вигляді суми несумісних подій: $A = П_1 \cdot Л_0 + П_0 \cdot Л_1 + П_1 \cdot Л_1 + П_0 \cdot Л_0$. Під час обчислення $P(П_0) = P(Л_0)$ і $P(П_1) = P(Л_1)$ скористаємося схемою Бернуллі при $n = 2$ і $p = 0.1$:

$$P(П_0) = p_2(0) = C_2^0 \cdot p^0 \cdot (1-p)^2 = (0,9)^2 = 0,81;$$

$$P(П_1) = p_2(1) = C_2^1 \cdot p \cdot (1-p) = 2 \cdot 0,1 \cdot 0,9 = 0,18.$$

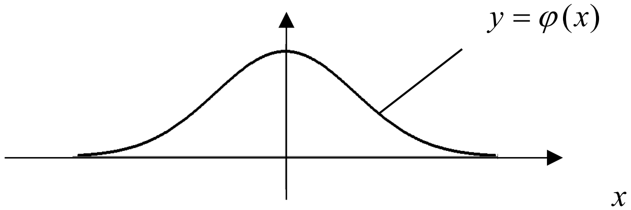
Отже, з урахуванням незалежності подій $П_k$ і $Л_j$ одержимо

$$P(A) = P(П_1) \cdot P(Л_0) + P(П_0) \cdot P(Л_1) + P(П_1) \cdot P(Л_1) + P(П_0) \cdot P(Л_0) =$$
$$= 0,18 \cdot 0,81 + 0,81 \cdot 0,18 + 0,18 \cdot 0,18 + 0,81 \cdot 0,81 = 0,98.$$

Якщо n велике, то формула Бернуллі потребує значних обчислень, тому користуються асимптотичною формулою Лапласа

$$P_n(k)=\frac{1}{\sqrt{npq}}\phi(t),\quad \phi(t)=\frac{1}{\sqrt{2\pi}}e^{\frac{-t^2}{2}},\quad t=\frac{k-np}{\sqrt{npq}}\ . \qquad (2.9)$$

Таблиця функції $\varphi(t)$ (функція Гаусса) наведена в додатку А (табл. А.1). Її графік має такий вигляд:



Зауважимо, що формула Лапласа дає непоганий результат за умови $p\geq 0,5$; $npq>10$.

Приклад. Імовірність виготовлення деталі першого сорту на певному верстаті дорівнює 0,6. Знайти ймовірність того, що серед навмання взятих 100 деталей буде 55 деталей першого сорту.

Розв'язання.

$$p=0,6>0,5;\ npq=100\cdot 0,6\cdot 0,4=24>10.$$

Отже, використаємо формулу Лапласа:

$$P_{100}(55)=\frac{1}{\sqrt{24}}\phi(x)=\left\{x=\frac{m-np}{\sqrt{24}}=\frac{55-100\cdot 0,6}{\sqrt{24}}\right\}=$$

$$=\frac{1}{\sqrt{24}}\phi\left(\frac{55-100\cdot 0,6}{\sqrt{24}}\right)=\frac{0,2371}{\sqrt{24}}\approx 0,04835.$$

Точна формула Бернулі

$$P_{100}(55)=0,04781.$$

Похибка менше 1 %.

Якщо ж n – дуже велике число, а ймовірність появи події дуже мала ($p<0,1$), то вдаються до асимптотичної формули Пуассона

$$P_n(k)=\frac{\lambda^k e^{-\lambda}}{k!}. \qquad (2.10)$$

Приклад. В упаковці є 10 % нестандартних деталей. Навмання відібрано 4 деталі. Знайти ймовірності того, що серед відібраних деталей немає нестандартних, одна, дві, три, чотири нестандартних.

Розв'язання. Імовірність появи нестандартної деталі завжди дорівнює 0,1.

Знайдемо потрібні ймовірності за формулою Бернуллі (2.8):

1) серед відібраних деталей узагалі немає нестандартних:
$$P_4(0)=\frac{4!}{0!\cdot 4!}0,1^0\cdot 0,9^4=0,6561;$$
2) серед відібраних деталей одна нестандартна:
$$P_4(1)=\frac{4!}{1!\cdot 3!}0,1^1\cdot 0,9^3=0,2916;$$
3) серед відібраних деталей дві нестандартні:
$$P_4(2)=\frac{4!}{2!\cdot 2!}0,1^2\cdot 0,9^2=0,0486;$$
4) серед відібраних деталей три нестандартні:
$$P_4(3)=\frac{4!}{3!\cdot 1!}0,1^3\cdot 0,9^1=0,0036;$$
5) серед відібраних деталей чотири нестандартні:
$$P_4(4)=\frac{4!}{4!\cdot 0!}0,1^4\cdot 0,9^0=0,0001.$$

Приклад. Найти ймовірність того, що з $n=100$ зернин зійде рівно $k=80$, якщо їх схожість $p=0,8$.

Розв'язання. За формулою (1.9):

$$t = \frac{80 - 100 \cdot 0,8}{\sqrt{100 \cdot 0,8 \cdot 0,2}} = 0, \quad \phi(0) = 0,3989.$$

$$P_{100}(80) = \frac{0,3989}{4} = 0,0997.$$

Приклад. Книга має 1000 сторінок. Відомо, що в ній міститься 100 помилок. Яка ймовірність того, що на випадково відібраній сторінці немає помилок, одна помилка, дві помилки?

Розв'язання. Середня кількість помилок на одну сторінку $a = \frac{100}{1000} = 0,1.$

Застосуємо формулу Пуассона (2.10).

1) на сторінці немає помилок:

$$P(0) = \frac{(0,1)^0}{0!} e^{-0,1} = 0,3679;$$

2) на сторінці одна помилка:

$$P(0) = \frac{(0,1)^1}{1!} e^{-0,1} = 0,03679;$$

3) на сторінці дві помилки:

$$P(0) = \frac{(0,1)^2}{2!} e^{-0,1} = 0,0018.$$

Імовірність того, що в n випробуваннях подія відбудеться від k_1 до k_2 разів обчислюють за формулою (інтегральною формулою Лапласа).

$$P_n(k_1, k_2) = \Phi(x_1) - \Phi(x_2), \qquad (2.11)$$

де $\Phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ – функція Лапласа;

$$x_1 = \frac{k_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{k_2 - np}{\sqrt{npq}}.$$

Таблиця функції Лапласа наведена в додатку А (табл. А. 2).

Приклад. Імовірність появи події за час випробувань $p = 0,8$. Зроблено сто випробувань. Знайти ймовірність того, що подія відбудеться не менше 75 разів і не більше 90 разів.

Розв'язання. Скористаємося інтегральною формулою Лапласа (2.11).

$$x_1 = \frac{75 - 100 \cdot 0,8}{\sqrt{100 \cdot 0,8 \cdot 0,2}} = -1,25; \quad x_2 = \frac{90 - 100 \cdot 0,8}{\sqrt{100 \cdot 0,8 \cdot 0,2}} = 2,5\,.$$

Ураховуючи, що функція Лапласа непарна, тобто $\Phi(-x) = \Phi(x)$, одержимо:

$$P_{100}(75,90) = \Phi(2,5) - \Phi(-1,25) = \Phi(2,5) + \Phi(1,25) =$$

$$= 0,4938 + 0,3944 = 0,8882.$$

ЧАСТИНА ІІ. ВИПАДКОВІ ВЕЛИЧИНИ ТА ПРОЦЕСИ. РОЗПОДІЛЕННЯ ВИПАДКОВИХ ВЕЛИЧИН

Розділ 3. Одновимірні випадкові величини та основні закони розподілення дискретних і неперервних величин

Одновимірною ***випадковою величиною*** називається величина, значення якої залежить від наслідку експерименту (інакше кажучи, випадковою величиною називають числова функція, визначена на просторі елементарних подій).

Між випадковими подіями й випадковими величинами є тісний зв'язок. Випадкова подія – це якісна характеристика випадкового результату досліду, а випадкова величина – його кількісна характеристика. Випадкові величини за певними властивостями поділяють на ***дискретні*** та ***неперервні***.

3.1 Дискретні випадкові величини та їх числові характеристики

Дискретною називають випадкову величину, що може набувати скінченну чи нескінченну множину значень із певними ймовірностями. Задають дискретні випадкові величини за допомогою закону розподілення, коли задають імовірності їх можливих випадкових значень.

Для дискретної випадкової величини X закон розподілення може бути заданий таблично або графічно. У першій ситуації закон розподілення називають ***рядом розподілення ймовірностей*** випадкової величини X

X	x_1	x_2	$\dots$	x_n	$\dots$
P	p_1	p_2	$\dots$	p_n	$\dots$

У першому рядку таблиці записують усі можливі значення випадкової величини, а в другому − відповідні їм імовірності. Оскільки події становлять повну групу несумісних подій, то за теоремою додавання ймовірностей маємо

$$\sum_{k=1}^{\infty} p_k = 1,$$

тобто сума ймовірностей усіх можливих значень випадкової величини дорівнює одиниці.

До основних характеристик випадкових величин відносять математичне сподівання й дисперсію.

Математичним сподіванням дискретної випадкової величини є сума попарних добутків усіх можливих значень випадкової величини на їх імовірності

$$m_x = M(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i . \qquad (3.1)$$

Дисперсія дорівнює різниці між математичним сподіванням квадрата випадкової величини й квадратом її математичного сподівання

$$D(X) = M(X^2) - \left[M(X)\right]^2 . \qquad (3.2)$$

Важливо зрозуміти їх визначення та зміст: математичне сподівання характеризує середню величину можливих значень випадкової величини, а дисперсія (середнє квадратичне відхилення) – розсіяння можливих значень навкруги математичного сподівання.

Розмірність дисперсії дорівнює квадрату розмірності випадкової величини. Тому вводять також величину $\sigma_X = \sqrt{DX}$, що називають ***середнім квадратичним*** (стандартним) відхиленням, вимірність якого співпадає з вимірністю випадкової величини. Саме це є причиною його широкого застосування для характеристики відхилень та ймовірного оцінювання поведінки випадкової величини. Зокрема, середнє квадратичне відхилення має надзвичайно важливе значення для критеріальної характеристики так званого принципу практичної впевненості. Принцип практичної впевненості можна сформулювати так: якщо ймовірність деякої події в певному досліді досить мала, то можна бути практично впевненим у тому, що за одноразового проведення досліду подія A не настане.

Чим менше $\sigma_X(DX)$, тим тісніше групуються значення випадкової величини навколо її математичного сподівання.

Приклад. Закон розподілення випадкової величини має вигляд (табл. 3.1).

Таблиця 3.1

X	0	1	2
p	0,0625	0,375	0,5625

Знайти математичне сподівання й дисперсію.

Розв'язання

1) математичне сподівання:

$$M(X) = 0 \cdot 0,0625 + 1 \cdot 0,375 + 2 \cdot 0,5625 = 1,5\,;$$

2) дисперсія: знайдемо математичне сподівання квадрата випадкової величини, для чого побудуємо допоміжну таблицю 3.2.

Таблиця 3.2

X	0	1	2
X^2	0	1	4
p	0,0625	0,375	0,5625

$$M(X^2) = 0 \cdot 0,0625 + 1 \cdot 0,375 + 4 \cdot 0,5625 = 2,625\,.$$

Знайдемо власне дисперсію

$$D(X) = 2,625 - [1,5]^2 = 0,375\,.$$

Властивості математичного сподівання:

1.Математичне сподівання сталої величини – стала $M(C) = C$.

2. $M(C \cdot X) = C \cdot M(X)$;

3. $M(X + Y) = M(X) + M(Y)$;

4. $M(X - Y) = M(X) - M(Y)$;

5.Математичне сподівання добутку двох незалежних випадкових величин
$M(X \cdot Y) = M(X) \cdot M(Y)$.

Властивості дисперсії:

1. $D(C) = 0$, C – стала величина;

2. $D(C \cdot X) = C^2 \cdot D(X)$;

3.Дисперсія суми незалежних випадкових величин дорівнює сумі їх дисперсій:
$D(X + Y) = D(X) + D(Y)$;

4. Дисперсія різниці двох незалежних випадкових величин дорівнює **сумі** їх дисперсій:
$D(X - Y) = D(X) + D(Y)$.

Приклад. Дискретна випадкова величина X може набувати лише два значення: x_1 і x_2, причому $x_1 < x_2$. Відомі: ймовірність p_1 значення x_1, математичне сподівання $M(X)$ і дисперсія $D(X)$: $p_1 = 0,2$; $M(X) = 3,8$; $D(X) = 0,16$. Знайти закон розподілення випадкової величини.

Розв'язання. Сума ймовірностей $p_1 + p_2$ дорівнює 1, тому

$$p_2 = 1 - 0,2 = 0,8.$$

Запишемо закон розподілення випадкової величини X

X	x_1	x_2
p	0,2	0,8

Запишемо математичне сподівання й дисперсію через x_1 і x_2 відомі ймовірності

$$M(X) = 0,2x_1 + 0,8x_2,$$

$$D(X) = M(X^2) - M^2(X) = 0,2x_1^2 + 0,8x_2^2 - 3,8^2 = 0,16.$$

Ці два рівняння утворюють систему, що має два розв'язки

$$x_1 = 3;\ x_2 = 4 \text{ і } x_1 = 4,6;\ x_2 = 3,6.$$

За умовою $x_1 < x_2$, тому розв'язком буде пара:

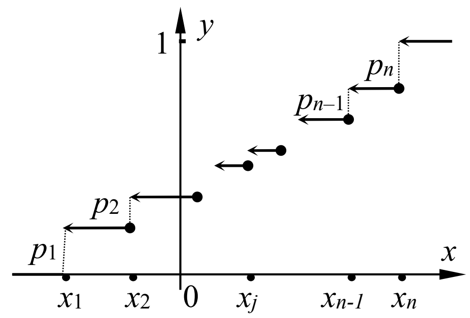
$$x_1 = 3;\ x_2 = 4.$$

Отже, маємо такий закон розподілу випадкової величини X:

X	3	4
p	0,2	0,8

Для кількісного оцінювання закону розподілення випадкової величини (дискретної або неперервної) задають ***функцію розподілу ймовірностей випадкової величини,***

котру визначають як імовірність того, що випадкова величина X набуде значення, меншого від певного фіксованого числа x і позначають $y = F(x) = P(X < x)$ або $F(x) = P(-\infty < X < x)$. Приблизний графік функції розподілення дискретної випадкової величини



Як бачимо, $F_X(x) = \sum p_k$, де знак суми стосується лише тих значень k, для яких $x_k < x$. Звідси випливає, що значення функції розподілення дискретної випадкової величини в проміжку $(x_k; x_{k+1}]$ дорівнює $p_1 + p_2 + ... + p_k$.

Функція розподілення є неспадною функцією, такою, що $F_X(-\infty) = 0$ (ймовірність неможливої події), $F_X(+\infty) = 1$ (імовірність вірогідної події). Знаючи функцію розподілення, можна знайти ймовірність того, що випадкова величина попадає в проміжок $[a; b)$:

$$P\{X \in [a; b)\} = P\{X < b\} - P\{X < a\} = F_X(b) - F_X(a).$$

Приклад. Побудувати інтегральну функцію розподілення для дискретної випадкової величини, знайденої в попередньому прикладі

X	3	4
P	0,2	0,8

Розв'язання. Для $-\infty < x < 3$ маємо $F(x) = 0$ (зліва точки $x = 3$ значень немає); для $3 < x < 4$ маємо $F(x) = 0{,}2$ (у цьому інтервалі є лише одне значення X з імовірністю 0,2). Нарешті, за умови $x \geq 4$ маємо $F(x) = 0{,}2 + 0{,}8 = 1$. Отже, функція $F(x)$ має такий вигляд:

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

3.2 Неперервні випадкові величини та їх числові характеристики

Неперервна випадкова величина має значення, що заповнюють скінчений або нескінченний проміжки числової осі.

Закон розподілення неперервної випадкової величини зручно задавати за допомогою функції густини ймовірності, або диференціальної функції розподілення $f(x)$. Зв'язок між інтегральною $F(x)$ і диференціальною функціями такий:

$$f(x) = F'(x), \qquad F(x) = \int_{-\infty}^{x} f(x)dx. \qquad (3.3)$$

Властивості функції $f(x)$

$$P(a < X < b) = \int_{a}^{b} f(x)dx;$$

$$f(x) \geq 0; \quad \int_{-\infty}^{\infty} f(x)dx = 1.$$

Числові характеристики неперервної випадкової величини:

- математичне сподівання $M(X) = \int_{-\infty}^{\infty} xf(x)dx$; (3.4)
- дисперсія $D(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - [M(X)]^2$. (3.5)

Під час вивчення неперервних випадкових величин необхідно звернути увагу на те, що інтегральну функцію розподілення використовується для завдання як неперервної, так і дискретної випадкових величин. Диференціальна функція розподілення використовується лише для величин неперервних. Власне кажучи, випадкову величину будемо називати неперервною, якщо її інтегральна функція неперервно диференційована.

Розглядаючи закони розподілення випадкових неперервних величин, необхідно вивчити властивості функцій $f(x)$ і $F(x)$, їх взаємозв'язок, імовірнісний зміст і числові характеристики.

Приклад. Задана неперервна випадкова величина X своєю диференціальною функцією розподілення $f(x)$

$$f(x) = \begin{cases} A\cos 2x, & -\frac{\pi}{4} \leq x \leq \frac{\pi}{4} \\ 0, & |x| > \frac{\pi}{4} \end{cases}.$$

Потрібно знайти коефіцієнт A, інтегральну функцію розподілення, побудувати графіки обох функцій, знайти числові характеристики випадкової величини, знайти ймовірність того, що випадкова величина x попадає в інтервал $\left(\frac{\pi}{6}; 2\right)$.

Розв'язання

1) Знайдемо коефіцієнт A:

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{-\pi/4} 0dx + \int_{-\pi/4}^{\pi/4} A\cos 2x dx + \int_{\pi/4}^{\infty} 0dx = \frac{A\sin 2x}{2}\Big|_{-\pi/4}^{\pi/4} = A = 1.$$

2) Знайдемо інтегральну функцію розподілення:

– якщо $x < -\frac{\pi}{4}$: $F(x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} 0dx = 0;$

– якщо $-\frac{\pi}{4} \le x \le \frac{\pi}{4}$:

$$F(x) = \int_{-\infty}^{-\pi/4} 0dx + \int_{-\pi/4}^{x} \cos 2x dx = \frac{\sin 2x}{2}\Big|_{-\pi/4}^{x} = \frac{\sin 2x}{2} + \frac{1}{2};$$

– якщо $x > \frac{\pi}{4}$:

$$F(x) = \int_{-\infty}^{-\pi/4} 0dx + \int_{-\pi/4}^{\pi/4} \cos 2x dx + \int_{\pi/4}^{x} 0dx = \frac{\sin 2x}{2}\Big|_{-\pi/4}^{\pi/4} = 1.$$

Отже,

$$f(x) = \begin{cases} \cos 2x, & \text{при} \quad -\frac{\pi}{4} \le x \le \frac{\pi}{4} \\ 0, & \text{при} \quad |x| > \frac{\pi}{4} \end{cases};$$

$$F(x) = \begin{cases} 0, & \text{при} \quad x < -\frac{\pi}{4} \\ \frac{\sin 2x + 1}{2}, & \text{при} \quad -\frac{\pi}{4} \le x \le \frac{\pi}{4} \\ 1, & \text{при} \quad x > \frac{\pi}{4} \end{cases}.$$

Побудуємо графіки (рис. 3.1, 3.2).

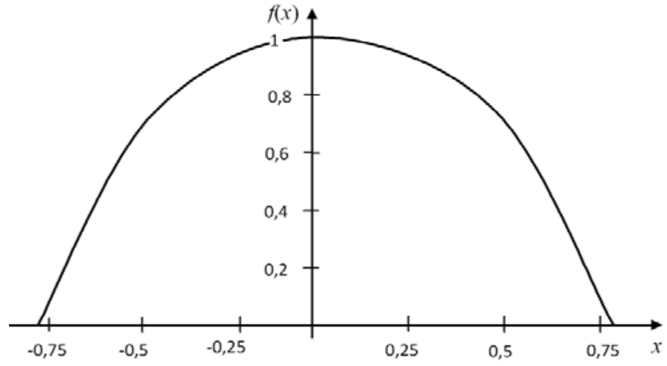


Рисунок 3.1 – Диференціальна функція розподілення

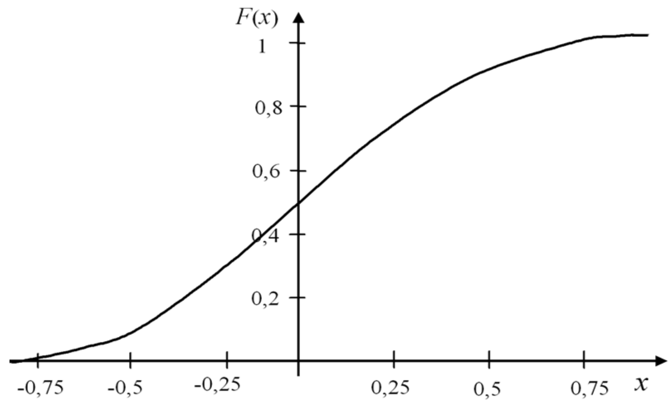


Рисунок 3.2 – Інтегральна функція розподілення

3) Числові характеристики випадкової величини X:

$$M(X)=\int_{-\infty}^{\infty}xf(x)dx=\int_{-\infty}^{-\pi/4}0dx+\int_{-\pi/4}^{\pi/4}x\cos 2xdx+\int_{\pi/4}^{\infty}0dx=$$

$$=\int_{-\pi/4}^{\pi/4}x\cos 2xdx=\frac{x\sin 2x}{2}\Big|_{-\pi/4}^{\pi/4}-\int_{-\pi/4}^{\pi/4}\frac{\sin 2x}{2}dx=\frac{\cos 2x}{4}\Big|_{-\pi/4}^{\pi/4}=0.$$

$$M(X^2)=\int_{-\infty}^{\infty}x^2f(x)dx=\int_{-\infty}^{-\pi/4}0dx+\int_{-\pi/4}^{\pi/4}x^2\cos 2xdx+\int_{\pi/4}^{\infty}0dx=$$

$$=\frac{x^2\sin 2x}{2}\Big|_{-\pi/4}^{\pi/4}-\int_{-\pi/4}^{\pi/4}x\sin 2xdx=\frac{\pi^2}{16}+\frac{x\cos 2x}{2}\Big|_{-\pi/4}^{\pi/4}-$$

$$-\int_{-\pi/4}^{\pi/4}\frac{\cos 2x}{2}dx=\frac{\pi^2}{16}-\frac{\sin 2x}{4}\Big|_{-\pi/4}^{\pi/4}=\frac{\pi^2}{16}-\frac{1}{2}=0{,}1163.$$

$$D(X)=M(X^2)-\left[M(X)\right]^2=0{,}1163-0=0{,}1163.$$

4) Імовірність попадання випадкової величини в заданий інтервал $\left(\frac{\pi}{6};2\right)$:

$$P\left(\frac{\pi}{6}<x<2\right)=\int_{\pi/6}^{2}f(x)dx=\int_{\pi/6}^{\pi/4}\cos 2xdx+\int_{\pi/4}^{2}0dx=\frac{\sin 2x}{2}\Big|_{\pi/6}^{\pi/4}=$$

$$=\frac{1}{2}-\frac{\sqrt{3}}{4}=0{,}067;$$

Цю ймовірність можна знайти в інший спосіб:

$$P\left(\frac{\pi}{6}<x<2\right)=F(2)-F(\pi/6)=1-\frac{\sin(\pi/3)+1}{2}=$$

$$=\frac{1}{2}-\frac{\sqrt{3}}{4}=0{,}067.$$

Приклад. Яким може бути математичне сподівання неперервної випадкової величини, якщо вона визначена на інтервалі [2; 5] і її інтегральна функція розподілення на цьому інтервалі є квадратичною (поліном другого степеня).

Розв'язання. Ураховуючи основні властивості інтегральної функції розподілення, запишемо її формулу в загальному вигляді:

$$F(x)=\begin{cases}0, & x<2,\\ ax^2+bx+c, & 2\le x\le 5,\\ 1, & 5<x.\end{cases}$$

З умов неперервності запишемо:

$$F(2)=4a+2b+c=0;$$
$$F(5)=25a+5b+c=1.$$

Один із коефіцієнтів, наприклад a, візьмемо за параметр, тоді:

$$b=\frac{1}{3}-7a,\qquad c=10a-\frac{2}{3},$$

$$F(x)=a(x^2-7x+10)+\frac{x-2}{3},\quad x\in[2;5].$$

Для оцінювання цього параметра врахуємо, що інтегральна функція розподілення будь-якої випадкової величини є неспадною

$$f(x) = F'(x) = a(2x-7) + \frac{1}{3} \geq 0, \quad x \in [2;5].$$

Тому $|a| \leq \frac{1}{9}$, (лінійна функція своїх екстремальних значень набуває на кінцях інтервалу).

Виразимо тепер математичне сподівання через параметр a:

$$M(X) = \int_2^5 xf(x)dx = \frac{9}{2}a + \frac{7}{2}.$$

Беручи до уваги область допустимих значень параметра a, одержимо інтервал можливих значень математичного сподівання

$$3 \leq M(X) \leq 4.$$

Функція розподілення для екстремальних значень математичного сподівання показана на рисунку 3.3. На осі абсцис відмічено інтервал можливих значень $M(X)$. Не випадково трапилось, що він повністю належить інтервалу [2; 5], на якому розподілена випадкова величина. Математичне сподівання прогнозує середнє значення випадкової величини й тому не повинно виходити за область її допустимих значень.

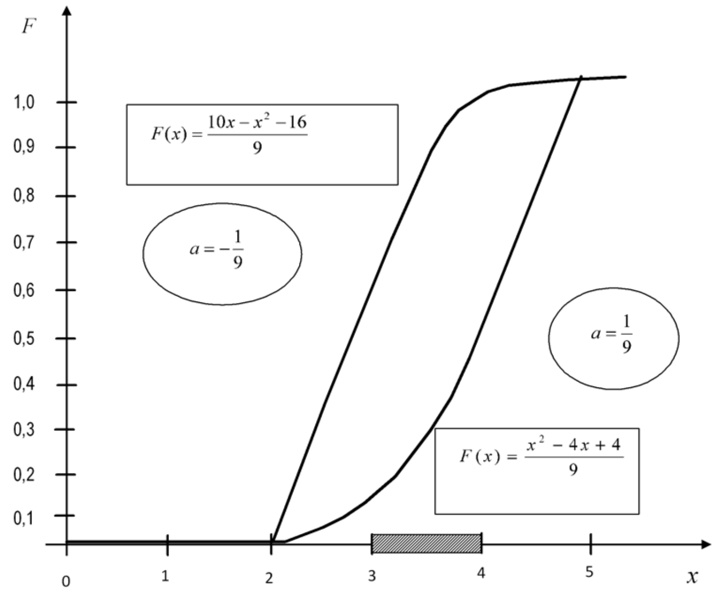


Рисунок 3.3 – Можливі межі інтегральної функції розподілення.

3.3 Основні закони розподілу дискретних і неперервних випадкових величин

Розглянемо кілька найбільш важливих дискретних розподілень.

Біноміальне розподілення (розподілення Я. Бернуллі). Випадкова величина X набуває значення $0,1,...,n$ і водночас

$$p_k = C_n^k p^k (1-p)^{n-k}, \quad k = 0,1,...,n, \quad p \in (0;1). \quad (3.6)$$

Біноміальне розподілення має випадкова величина, що дорівнює кількості появ події A в серії з n незалежних випробувань за ймовірності p її появи в одному експерименті. На рисунку 3.4а наведено графік біноміального розподілення за умови $n = 20$, $p = 0.5$.

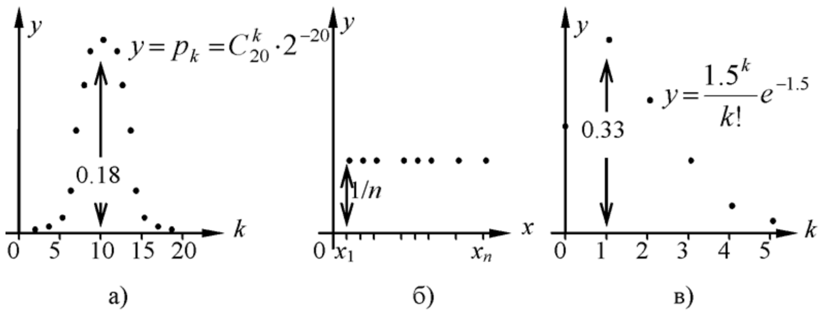


Рисунок 3.4 – (а) Графік біноміального розподілення за умови n = 20, p = 0.5; (б) рівномірне дискретне розподілення; (*в*) *графік розподілення Пуассона із параметром* $\lambda = 1.5$

Рівномірне дискретне розподілення. Випадкова величина X набуває n значень із відповідними ймовірностями, що дорівнюють $1/n$ (рис. 3.4б).

Розподілення Пуассона. Випадкова величина X набуває значення 0,1,... з імовірностями

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0,1,2,\ldots, \quad \lambda > 0. \qquad (3.7)$$

На рисунку 3.4в наведено графік розподілення Пуассона з параметром λ=1,5.

Розглянемо деякі важливі **неперервні розподілення**. **Рівномірне (прямокутне) розподілення**. Випадкова величина X рівномірно розподілена в проміжку $[a\,;\,b]$, якщо її густина ймовірності має вигляд (рис. 3.5):

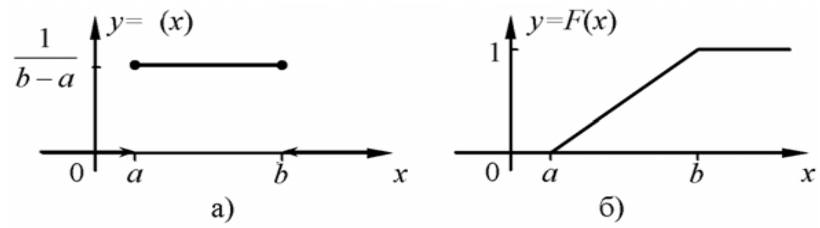


Рисунок 3.5 – Графіки $f(x)$ і $F(x)$ *рівномірного розподілення на проміжку* $[a\,;\,b]$

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{якщо } x \in [a;b]; \\ 0, & \text{якщо } x \notin [a;b]. \end{cases} \qquad (3.8)$$

Показникове розподілення. Випадкова величина X має показникове розподілення з параметром $\lambda > 0$, якщо її густина розподілення

$$f(x) = \begin{cases} 0, & x \le 0; \\ \lambda e^{-\lambda x}, & x > 0. \end{cases}$$

Відповідна функція розподілення має вигляд

$$F(x) = \begin{cases} 0, & x \le 0; \\ 1 - e^{-\lambda t}, & x > 0. \end{cases} \qquad (3.10)$$

Графіки густини ймовірності та функції розподілення приведені на рисунку 3.6.

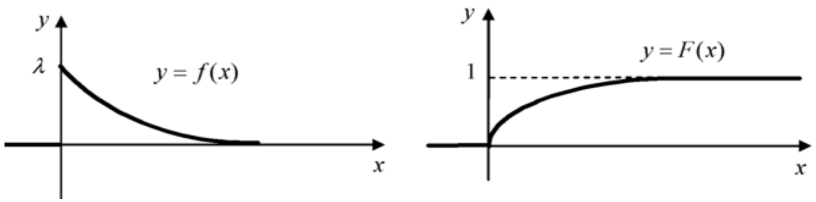


Рисунок 3.6 – Графіки $f(x)$ і $F(x)$ показникового розподілення

Показникове розподілення (й лише воно серед неперервних розподілень) має властивість «відсутності післядії»:

$$P\{X > x_1 + x_2 / X > x_1\} = P\{X > x_2\} \quad (x_1,\ x_2 > 0).$$

Дійсно,

$$P\left\{X > x_1 + x_2 / X > x_1\right\} = \frac{P\left\{X > x_1 + x_2, X > x_1\right\}}{P\left\{X > x_1\right\}} =$$

$$= \frac{P\left\{X > x_1 + x_2\right\}}{P\left\{X > x_1\right\}} == \frac{1 - F(x_1 + x_2)}{1 - F(x_1)} = \frac{e^{-\lambda(x_1 + x_2)}}{e^{-\lambda x_1}} = e^{-\lambda x_2} =$$

$$= 1 - F_X(x_2) = P\left\{X > x_2\right\}.$$

Знайдемо числові характеристики показникового розподілення:

$$MX = \int_{-\infty}^{+\infty} x\, f(x)\, dx = \int_{0}^{+\infty} x\, \lambda\, e^{-\lambda x} dx = -\int_{0}^{+\infty} x\, de^{-\lambda x} = -\left(0 - \int_{0}^{+\infty} e^{-\lambda x} dx\right) =$$

$$= -\frac{1}{\lambda} e^{-\lambda x}\Big|_0^{+\infty} = -\frac{1}{\lambda}(0 - 1) = \frac{1}{\lambda}.$$

$$DX = MX^2 - (MX)^2;$$

$$MX^2 = \int_{-\infty}^{+\infty} x^2 f(x)\, dx = \int_{0}^{+\infty} x^2 \lambda e^{-\lambda x} dx = -\int_{0}^{+\infty} x^2 de^{-\lambda x} =$$

$$= -\left(0 - 2\int_{0}^{+\infty} x e^{-\lambda x} dx\right) = \frac{2}{\lambda}\int_{0}^{+\infty} x\, \lambda e^{-\lambda x} dx = \frac{2}{\lambda} MX = \frac{2}{\lambda^2},$$

$$DX = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

Отже, для показникового розподілення

$$MX = \frac{1}{\lambda}, \quad DX = \frac{1}{\lambda^2}, \quad \sigma X = \frac{1}{\lambda}.$$

Розділ 4. Нормальний закон розподілення. Граничні теореми теорії ймовірностей.

4.1 Нормальний закон розподілення. Його основні характеристики

Нормальний закон розподілення (або закон Гауса) відіграє винятково важливу роль у теорії ймовірностей і посідає серед інших законів розподілення особливе місце. Саме тому ми винесли його в окремий розділ цього посібника. Це закон, що найчастіше трапляється на практиці. Основна особливість, яка виділяє його серед інших , полягає в тому, що він є граничним законом, до якого наближаються інші закони розподілення.

Так, наприклад, велика кількість гарматних пострілів, здійснених у різних умовах, показує, що розсіювання снарядів на площині під час пострілу з однієї гармати за умови встановленого прицілу підлягає нормальному закону.

«Універсальність» нормального закону пояснюється тим, що будь-яка випадкова величина, яка є сумою великої кількості окремих числових значень, кожне з яких підпорядковане різним законам розподілення й неістотно впливає на суму, розподілену майже за нормальним законом.

Більшість випадкових величин, таких, наприклад, як похибки вимірів, похибки гарматних стрільб тощо можуть бути подані як суми великої кількості малих доданків – елементарних похибок, кожна з яких визначається дією окремої причини, що не залежить від інших. Яким би законам розподілення не підпорядковувались окремі елементарні похибки, особливості цих розподілень у сумі великої кількості доданків нівелюються й сума підпорядковується закону, близькому до нормального.

Підсумовані похибки в загальній сумі повинні відігравати порівняно малу роль.

Випадкова величина X має нормальне розподілення з параметрами a та σ^2, якщо її густина розподілення має вигляд (рис. 4.1а)

$$f\left(x\right)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}. \qquad (4.1)$$

У подальшому запис $N(a;\sigma^2)$ означатиме, що випадкова величина X має розподілення Гауса з параметрами a та σ^2. Графік розподілення Гауса є симетричним стосовно прямої $x = a$. Єдиний максимум досягається при $x=a$ і дорівнює $\frac{1}{\sigma\sqrt{2\pi}}$. Оскільки площа під графіком дорівнює 1, то під час зменшення σ графік стає більш «високим» та «вузьким».

Функція розподілення випадкової величини X виражена через функцію Лапласа $\Phi(x)$. Дійсно,

$$F\left(x\right)=\int_{-\infty}^{x}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t-a)^2}{2\sigma^2}}dt=\int_{-\infty}^{a}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t-a)^2}{2\sigma^2}}dt+$$

$$+\int_{a}^{x}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t-a)^2}{2\sigma^2}}dt=\left\{\left(t-a\right)/\sigma=k\right\}=\frac{1}{2}+\frac{1}{\sqrt{2\pi}}\int_{a}^{\frac{x-a}{\sigma}}e^{-\frac{k^2}{2}}dk=$$

$$=\frac{1}{2}+\Phi\left(\frac{x-a}{\sigma}\right)$$

Графіки густини розподілення ймовірностей та їх функції наведено на рисунку 4.1.

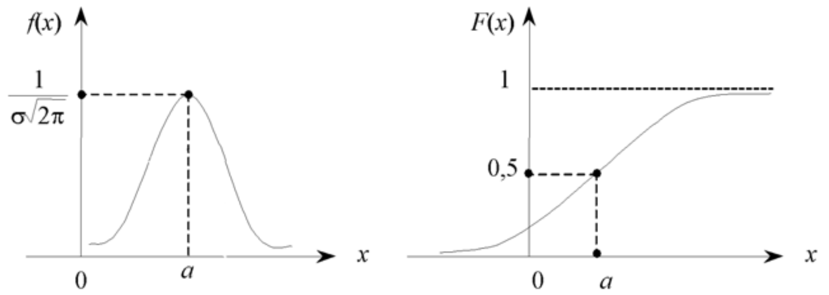


Рисунок 4.1 – Графіки щільності розподієнняу ймовірностей та їх функції

Розподілення Гауса відіграє фундаментальну роль у застосуваннях теорії ймовірності.

Знайдемо числові характеристики нормального розподілення:

$$MX=\int\limits_{-\infty}^{+\infty} x\,f(x)dx=\frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{+\infty} x\,e^{-\frac{(x-a)^2}{2\sigma^2}}dx=\left\{\frac{x-a}{\sigma}=t\right\}=$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{+\infty}(a+\sigma\,t)e^{-t^2/2}\,\sigma\,dt=\frac{1}{\sqrt{2\pi}}\left[a\int\limits_{-\infty}^{+\infty}e^{-t^2/2}dt+\sigma\int\limits_{-\infty}^{+\infty}t\,e^{-t^2/2}dt\right]=a$$

,

оскільки $\int\limits_{-\infty}^{+\infty}e^{-t^2/2}dt=\sqrt{2\pi}$, $\int\limits_{-\infty}^{+\infty}t\,e^{-t^2/2}dt=0$.

Обчислимо дисперсію

$$DX=\int\limits_{-\infty}^{\infty}(x-MX)^2\,p(x)dx=\frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{\infty}(x-a)^2e^{-\frac{(x-a)^2}{2\sigma^2}}dx=$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\sigma\int\limits_{-\infty}^{\infty}(\sigma t)^2e^{-\frac{t^2}{2}}dt=\frac{1}{\sqrt{2\pi}}\sigma^2\int\limits_{-\infty}^{\infty}t^2e^{-\frac{t^2}{2}}dt=$$

$$=-te^{-\frac{t^2}{2}}\Bigg|_{-\infty}^{\infty}+\frac{1}{\sqrt{2\pi}}\sigma^2\int\limits_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt=\sigma^2.$$

Отже, для нормального розподілення $MX=a$, $DX=\sigma^2$, $\sigma X=\sigma$.

4.2 Правило трьох сигм

Оскільки ймовірність попадання випадкової величини в проміжок дорівнює різниці значень функції розподілення на кінцях проміжку, то

$$P\{X\in[p;m]\}=\Phi\left(\frac{m-a}{\sigma}\right)-\Phi\left(\frac{p-a}{\sigma}\right). \qquad (4.2)$$

Якщо проміжок $[p;\,m]$ довжиною $2k\sigma$ розташований симетрично точці $x=a$, то формула (4.2) набуває особливо простого вигляду

$$P\{X\in[a-k\sigma\,;\,a+k\sigma]\}=\Phi(k)-\Phi(-k)=2\Phi(k). \qquad (4.3)$$

Імовірність попадання в проміжок $[a–3\sigma;\,a+3\sigma]$ дорівнює 0,9973. Отже, можна стверджувати, що подія $\{X\notin[a\text{-}3\sigma;\,a+3\sigma]\}$ є практично неможливою. У цьому полягає відоме правило «трьох сигм».

Приклад. Відхилення розміру деталі від стандартного розподілено за законом $N(0;16\,\text{мм}^2)$. Деталь вважають придатною, якщо відхилення від стандарту не перевищує 6 мм. Який відсоток випуску непридатних деталей?

Розв'язання. Нехай випадкова величина X – відхилення розміру деталі від номінального. Знайдемо ймовірність того, що деталь буде забраковано

$$P\left\{\,|X|>6\right\}=1-P\left\{\,|X|\le 6\right\}=1-\left(\Phi\left(\frac{6}{4}\right)-\Phi\left(-\frac{6}{4}\right)\right)=$$

$$=1-2\Phi\left(\frac{3}{2}\right)=1-2\cdot 0.4332=0.134\,.$$

Отже, брак становить майже 13,5 %.

4.3 Розподілення χ^2 (Пірсона) та Стьюдента

Нехай випадкові величини $X_0, X_1, X_2, ..., X_n$ є незалежні та розподілені за законом Гауса $N(0;1)$. Розподілення випадкової величини $\chi_n^2=X_1^2+X_2^2+...+X_n^2$ називається **χ^2-*розподіленням*** (хі-квадрат розподіленням) з n ступенями свободи. Густина ймовірності випадкової величини χ_n^2 задається співвідношенням

$$p_{\chi_n^2}\left(x\right)=K_n\cdot x^{\frac{n}{2}-1}\cdot e^{-\frac{x}{2}},\quad x>0\,, \tag{4.4}$$

де стала K_n визначається умовою нормування. Зокрема, випадкова величина $\chi_1^2 = X_1^2$ має щільність ймовірності

$$p_{\chi_1^2}\left(x\right)=\frac{1}{\sqrt{2\pi x}}\cdot e^{-\frac{x}{2}}$$

Графіки густини розподілення χ^2 наведено на рисунку 4.2. Як бачимо з рисунка, це розподілення визначається параметром n. Якщо n зростає, розподілення повільно наближається до нормального.

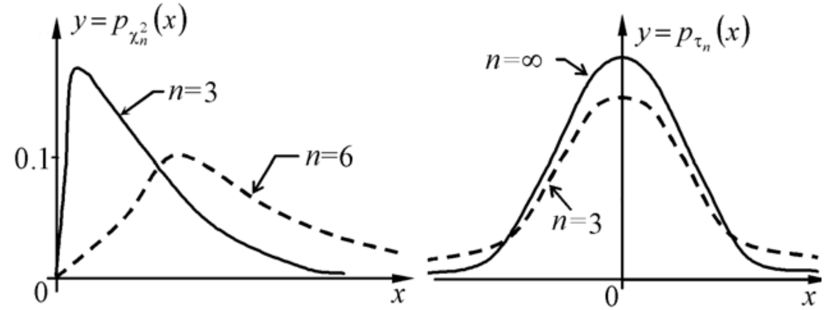


Рисунок 4.2 – Графіки густини розподієнняу χ^2 (ліворуч) та Стьюдента (праворуч)

Розподілення випадкової величини $\tau_n = X_0\big/\sqrt{\chi_n^2/n}$ називається **τ-*розподіленням*** (розподіленням Стьюдента) з n ступенями свободи. Густина ймовірності випадкової величини τ_n задається співвідношенням (графік густини приведено на мал. 2.25):

$$p_{\tau_n}\left(x\right)=L_n\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \tag{4.5}$$

де стала L_n визначається умовою нормування. Графік функції $p_{\tau_n}\left(x\right)$ наведено на рисунку 4.2.

Графік густини є симетричним відносно прямої x=0 і подібний до графіка густини нормального розподілення $N(0;1)$. Зі зростанням кількості ступенів свободи n розподілення Стьюдента наближається до нормального з параметрами a = 0, σ^2 = 1.

Зазначимо, що розподілення Стьюдента − один із найбільш відомих серед використовуваних під час аналізу реальних даних. Його застосовують для оцінювання математичного сподіванння, прогнозного значення та інших характеристик за допомогою надійних інтервалів, під час перевірки гіпотез про значення математичних сподівань, коефіцієнтів регресійної залежності, гіпотез однорідності виборок.

4.4 Закон великих чисел та центральна гранична теорема

Граничні теореми теорії ймовірностей установлюють відповідність між теоретичними й дослідними характеристиками випадкових величин або випадкових подій за умови великої кількості випробувань. Граничні теореми описують також граничні закони розподілення.

Граничні теореми, що встановлюють відповідність між теоретичними та дослідними характеристиками випадкових подій, об'єднують загальною назвою **закон великих чисел**. Закон великих чисел відіграє важливу роль у різних процесах, пов'язаних із масовим виробництвом.

Граничні теореми, що встановлюють граничні закони розподілення випадкових величин, об'єднують загальною назвою **центральна гранична теорема**.

Необхідність граничних теорем обумовлена потребою розв'язування, наприклад, таких задач:

1) коли сума багатьох випадкових величин мало відрізняється від постійної величини, тобто майже перестає бути випадковою величиною, і тому її поведінка може прогнозуватися із значною імовірністю;

2) За яких умов можна зі значною ймовірністю прогнозувати число появи деякої випадкової події в разі великій кількості незалежних випробувань;

3) у разі обмежень за якими сума багатьох випадкових величин буде розподілена за нормальним законом.

Перед тим, як перейти до центральної граничної теореми, розглянемо, спочатку, нерівність Чебишева. Нерівність П. Чебишева встановлює верхню межу для ймовірності відхилення випадкової величини від її математичного сподівання.

*Теорема***.** Якщо X – невід'ємна випадкова величина, для якої існує MX, то $\forall \varepsilon > 0$

$$P(X \geq \varepsilon) \leq \frac{MX}{\varepsilon}. \qquad (4.6)$$

Доведення. Якщо $P(X \geq 0) = 1$, то

$$MX = \int_0^{+\infty} x\, f(x)dx \geq \int_\varepsilon^{+\infty} x\, f(x)dx \geq \varepsilon \int_\varepsilon^{+\infty} f(x)dx = \varepsilon\, P(X \geq \varepsilon)$$

звідки і випливає нерівність (4.6).

Нерівність (4.6) називається допоміжною або першою нерівністю Чебишева.

Теорема. Якщо X – випадкова величина, для якої існує DX, то $\forall \varepsilon > 0$

$$P\left(\left|X - MX\right| \geq \varepsilon\right) \leq \frac{DX}{\varepsilon^2}. \qquad (4.7)$$

Доведення. Оскільки $\left|X - MX\right|$ – невід'ємна випадкова величина, тому відповідно до допоміжної нерівності Чебишева маємо

$$P\left(\left|X - MX\right| \geq \varepsilon\right) = P\left(\left(X - MX\right)^2 \geq \varepsilon^2\right) \leq \frac{M(X - MX)^2}{\varepsilon^2} = \frac{DX}{\varepsilon^2}$$

Нерівність (4.7) називають основною, або другою нерівністю Чебишева.

Зазначимо, що основна нерівність Чебишева універсальна, оскільки не висуває ніяких вимог до розподілення X, але й саме внаслідок своєї універсальності, малоінформативна кількісно – для конкретних значень ε оцінки ймовірностей $P\left(|X-MX|\geq\varepsilon\right)$ і $P\left(|X-MX|<\varepsilon\right)$ вкрай грубі. Незважаючи на сказане, покращити її неможливо, оскільки (це можна показати на прикладах) вона є точною.

Теорема Чебишева $D(X_n)<L$ дає змогу обґрунтувати правило середнього, яким широко користуються в практиці вимірювань: за достатньо великої кількості вимірювань з імовірністю, як завгодно близькою до одиниці, можна вважати, що середнє арифметичне результатів вимірювань буде як завгодно мало відрізнятися від вимірюваної величини, тобто точно характеризує вимірювану величину.

Теорема Чебишева має велике теоретичне, практичне й методологічне значення. На цій теоремі ґрунтується вибірковий метод, що застосовують у статистиці. Його суть полягає в тому, що за порівняно невеликою випадковою вибіркою роблять висновок про всю сукупність досліджуваних об’єктів.

Вважають, що для послідовності випадкових величин X_n $(n=\overline{1,\infty})$ для яких існують математичні сподівання, актуальний **закон великих чисел**, якщо

$$\frac{1}{n}\sum_{i=1}^{n}X_i \xrightarrow{P} \frac{1}{n}\sum_{i=1}^{n}MX_i\,. \qquad (4.8)$$

Тобто, для послідовності випадкових величин виконується закон великих чисел, якщо середнє арифметичне випадкових величин сходиться до середнього

арифметичного математичних очікувань цих випадкових величин.

Зазначимо, що співвідношення (4.8) виражає той факт, що в разі деяких додаткових припущень, поведінка середнього арифметичного достатньо великої кількості випадкових величин, за необмеженого зростання цієї кількості, втрачає випадковий характер і стає закономірною.

Теорема А. А. Маркова. Позначимо через S_n суму $\sum_1^n X_i$. Якщо $\frac{D(S_n)}{n^2}\to 0,\ n\to\infty$, то закон великих чисел справедливий для введеної нами в розгляд послідовності випадкових величин.

Теорема Чебишева. Якщо випадкові величини $X_1, X_2, ..., X_n...$ попарно незалежні, мають кінцеві математичні очікування й рівномірно обмежені дисперсії, тобто $D(X_n)<L$, де L - задане число, що не залежить від n, то

$$\frac{D(S_n)}{n^2}=\frac{D(\sum_1^n X_i)}{n^2}=\frac{\sum_1^n D(X_i)}{n^2}\leq\frac{nL}{n^2}=\frac{L}{n}\to 0$$

Зі свого боку, теорема Маркова виходить із нерівності Чебишева. Нехай умови теореми Маркова виконані, тобто $\frac{D(S_n)}{n^2}\to 0, n\to\infty$. Покладемо $Z_n=\frac{S_n}{n}$, тоді $D(Z_n)=\frac{1}{n^2}D(S_n)\to 0$. Але оскільки нерівність Чебишева

$$MX_n \ (n = \overline{1, \infty})$$

$$p(\left|\frac{1}{n}\sum_1^n X_i - \frac{1}{n}\sum_1^n M(X_i)\right| \geq \varepsilon) =$$

$$= p(\left|Z_n - M(Z_n)\right| \geq \varepsilon) \leq \frac{D(Z_n)}{\varepsilon^2} \to 0$$

У центральній граничній теоремі описуються умови, за яких виникає нормальний закон розподілення. Виявляється, що він виникає щоразу, коли випадкова величина може бути репрезентована у вигляді суми досить великого числа попарно незалежних випадкових величин, кожна з яких порівняно мало впливає на всю суму.

Нехай випадкові величини $X_1, X_2, ..., X_n ...$ попарно незалежні й кожна з них має математичне очікування й дисперсію: $M(X_i) = a_i$, $D(X_i) = \sigma_i^2$. Позначимо через S_n суму $\sum_{i=1}^n X_i$, через A_n суму $\sum_{i=1}^n a_i$, а через B_n суму $\sum_{i=1}^n \sigma_i^2$.

Будемо говорити, що до послідовності застосовна *центральна гранична теорема*, якщо для будь-яких чисел t_1, t_2, $t_1 < t_2$ в разі $n \to \infty$ справедливо граничне співвідношення

$$p(t_1 \leq \frac{S_n - A_n}{\sqrt{B_n}} \leq t_2) \to \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{t^2}{2}} dt$$

Іншими словами, у разі $n \to \infty$ випадкова величина S_n має наближено нормальне розподілення з параметрами $a = A_n$ та $\sigma = \sqrt{B_n}$. Можна показати, що якщо випадкові числа однаково розподілені та є незалежними випадковими величинами, то для послідовності справедлива центральна гранична теорема.

Розділ 5. Системи випадкових величин

Вивчені нами в попередніх розділах випадкові величини є одномірними, тобто значення їх визначається одним числом. Крім них, трапляються випадкові величини, значення яких визначають двома, трьома, …, n числами. Такі величини називають, відповідно, двовимірними, тривимірними, …, n-вимірними.

Для наочності дослідження n-вимірних випадкових величин зручно користуватися геометричною інтерпретацією. Так, систему двох випадкових величин (X, Y) можна зобразити випадковою точкою на площині xy з координатами X та Y, або, що рівносильно, випадковим вектором, спрямованим з початку координат у точку (X, Y). Аналогічно, систему трьох випадкових величин (X, Y, Z) можна зобразити в тривимірному просторі як випадковий вектор, спрямованим із початку координат у точку (X, Y, Z).

Якщо вимірність величини більше від 3, то геометрична інтерпретація втрачає наочність, але користуватися геометричною термінологією зручно. Так, про систему випадкових величин будемо казати як про випадкову точку в n-вимірному просторі R_n або як про випадковий вектор, спрямований із початку координат у точку $(X_1, X_2, …, X_n)$. На n-вимірні випадкові величини поширюються майже без змін основні означення, які було розглянуто для одновимірної випадкової величини. Очевидно, що властивості n-вимірної випадкової величини не обмежуються властивостями окремих складових випадкових величин; істотні також зв'язки між складовими величинами.

Повна характеристика n-вимірної випадкової величини – її закон розподілення, що як і для окремих випадкових величин може мати різні форми: функцію розподілення, щільність розподілення, таблицю ймовірностей окремих значень випадкового вектора тощо. Для практики важливий випадок $n = 2$, яким ми обмежимося в подальшому.

5.1 Функція розподілення двовимірної випадкової величини

Функцією розподілення ймовірностей (або сумісною функцією розподілєня) двовимірної випадкової величини (X, Y) називається така функція двох аргументів x, y, що визначає ймовірність сумісного настання подій

$$F(x, y) = P(X < x,\ Y < y).$$

Подія в дужках означає добуток подій:

$$(X < x,\ Y < y) = (X < x) \cap (Y < y).$$

Властивості функції розподілення двовимірної випадкової величини (X, Y) аналогічні властивостям одновимірної випадкової величини:

1. $0 \le F(x,y) \le 1$;
2. неперервність зліва за кожним аргументом;
3. функція $F(x,y)$ – неспадна функція за кожним аргументом: $F(x_1, y) \le F(x_2, y)$, якщо $x_1 < x_2$ та $F(x, y_1) \le F(x, y_2)$, якщо $y_1 < y_2$;
4. $F(x, +\infty) = F(x_1)$ – функція розподілення компонента x, $F(+\infty, y) = F(y_2)$ – функція розподілу компонента Y (властивість установлює природний зв'язок між функцією розподілення двовимірної випадкової величини та функціями розподілення $F(x_1)$ та $F(y_2)$, що також називають частинними функціями одновимірних випадкових величин X і Y.
5. $F(-\infty, y) = 0$, $F(x, -\infty) = 0$, $F(-\infty, -\infty) = 0$, $F(+\infty, +\infty) = 1$;

6. імовірність потрапляння випадкової точки до прямокутника ($x_1 \le X \le x_2$; $y_1 \le Y \le y_2$) обчислюють за формулою

$$P(x_1 < X < x_2;\ y_1 < Y < y_2) = (F(x_2, y_2) - F(x_1, y_2)) - (F(x_2, y_1) - F(x_1, y_1)).$$

5.2 Двовимірні випадкові величини

Розглянемо двовимірну дискретну випадкову величину (X, Y). Кожна з величин X, Y, що її складають, є звичайною дискретною випадковою величиною. Тому двовимірну дискретну випадкову величину (X, Y) можна назвати системою двох випадкових величин.

Закон розподілення випадкової величини задають таблицею (табл. 5.1).

Таблиця 5.1Закон розподілення двовимірної величини

	x_1	x_2	$\dots$	x_n
y_1	$p(x_1, y_1)$	$p(x_2, y_1)$	$\dots$	$p(x_n, y_1)$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
y_j	$p(x_1, y_j)$	$p(x_2, y_j)$	$\dots$	$p(x_n, y_j)$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
y_m	$p(x_1, y_m)$	$p(x_2, y_m)$	$\dots$	$p(x_n, y_m)$

У цій таблиці в клітинках із координатами (x_i, y_j) знаходяться події ймовірності $\{X = x_i, Y = y_j\}$. Вважаємо, що всі можливі комбінації подій утворюють повну групу.

Іншими словами,

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p\left(x_i, y_j\right) = 1. \qquad (5.1)$$

Знаючи закон розподілення двовимірної дискретної випадкової величини, можна знайти закони розподілення окремих її складових. Наприклад, за теоремою додавання несумісних подій

$$p\left(x_i\right) = p(x_i, y_1) + p(x_i, y_2) + ... + p(x_i, y_m) \qquad (5.2)$$

одержали ймовірність того, що X прийме x_i. Отже, одержали закон розподілення випадкової величини X (табл. 5.2).

Таблиця 5.2 – Закон розподілу X

X	x_1	$\cdots$	x_i	$\cdots$	x_n
p	$p(x_1)$	$\cdots$	$p(x_i)$	$\cdots$	$p(x_n)$

Так само можна знайти закон розподілення випадкової величини Y (табл. 5.3)

Таблиця 5.3 – Закон розподілу Y

Y	y_1	$\cdots$	y_j	$\cdots$	y_m
p	$p(x_1)$	$\cdots$	$p(x_i)$	$\cdots$	$p(y_m)$

Перед тим, як ми перейдемо до розглядання числових характеристик двовимірної випадкової величини, розглянемо поняття числових моментів випадкової величини.

5.3 Початкові й центральні моменти

Початковим моментом (або просто моментом) випадкової величини X порядку k називають число

$$\alpha_k = MX^k . \qquad (5.3)$$

Центральним моментом випадкової величини X порядку k називають число

$$\mu_k = M\,(\,X - MX\,)^k . \qquad (5.4)$$

Між центральним і початковим моментами існує простий зв'язок. Дійсно,

$$\mu_k = M\,(\,X - MX\,)^k = M\sum_{i=0}^{k} C_k^i X^i (-MX)^{k-i} = \sum_{i=0}^{k} C_k^i MX^i (-MX)^{k-i} =$$

$$= \sum_{i=0}^{k} (-1)^{k-i} C_k^i \alpha_i \alpha_1^{k-i} = \sum_{i=2}^{k} (-1)^{k-i} C_k^i \alpha_i \alpha_1^{k-i} + (-1)^k \alpha_1^k + (-1)^{k-1} k\alpha_1^k =$$

$$= \sum_{i=2}^{k} (-1)^{k-i} C_k^i \alpha_i \alpha_1^{k-i} + (-1)^{k-1}(k-1)\alpha_1^k .$$

Випишемо формули, що встановлюють цей зв'язок, для перших чотирьох значень k :

$\mu_1 = 0$;

$\mu_2 = \alpha_2 - \alpha_1^2 = DX$;

$\mu_3 = \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3$;

$\mu_4 = \alpha_4 - 4\alpha_3\alpha_1 + 6\alpha_2\alpha_1^2 - 3\alpha_1^4$.

Зі старшими моментами $k > 2$ пов’язані дві характеристики випадкової величини X (або розподілення

X), які часто використовують у застосуваннях – асиметрії та ексцесу.

Асиметрією випадкової величини X називають число

$$AX = \frac{\mu_3}{(\sigma X)^3} = \frac{M(X - MX)^3}{(\sigma X)^3}. \qquad (5.5)$$

Зміст асиметрії випливає з формули (5.5). Якщо $AX < 0$, розподілення має праву асиметрію, якщо $AX > 0$ – ліву, якщо $AX = 0$, розподілення симетричний відносно прямої $X = MX$ (рис. 5.1).

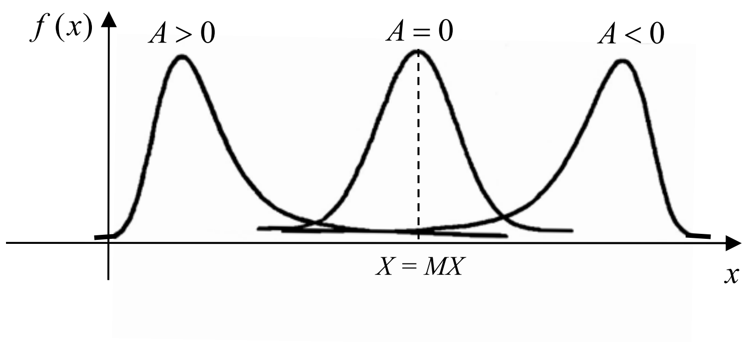


Рисунок 5.1 – Асиметрія розподілення випадкової величини X

Ексцесом випадкової величини X називають число

$$EX = \frac{\mu_4}{(\sigma X)^4} - 3 = \frac{M(X - MX)^4}{(\sigma X)^4} - 3. \qquad (5.6)$$

Ексцес характеризує згладженість кривої густини розподілення порівняно зі стандартним нормальним розподіленням, для якого $EX = 0$, відповідно $\mu_4 / (\sigma X)^4 = 3$ (рис. 5.2).

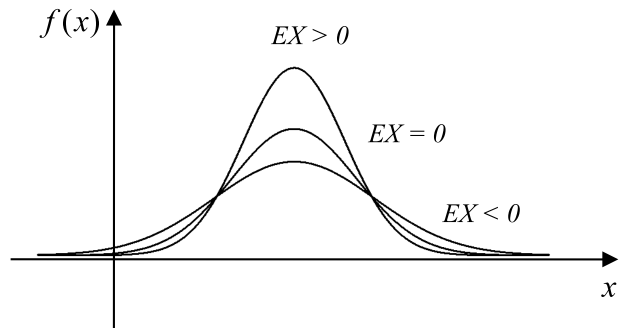


Рисунок 5.2 – Ексцес розподілення випадкової величини X

Отже, AX і EX – безрозмірні випадкові величини, що характеризують ступінь відмінності густини розподілення $f(x)$ від густини стандартного нормального розподілення.

5.4 Числові характеристики двовимірної випадкової величини. Коефіцієнт кореляції та його властивості.

Розглянемо числові характеристики системи двох випадкових величин. Вважаємо заданим закон розподілення дискретної двовимірної випадкової величини (X, Y). Тоді можна обчислити математичні сподівання складових X та Y

$$M(X) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i p(x_i, y_j), \qquad (5.7)$$
$$M(Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} y_i p(x_i, y_j).$$

Кореляційним моментом μ_{xy} випадкових величин X та Y називають математичне сподівання добутку відхилень цих величин від відповідного математичного сподівання

$$\mu_{xy} = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(x_i - M\left(X\right)\right)\left(y_j - M\left(Y\right)\right)p\left(x_i, y_j\right) . \quad (5.8)$$

Приклад. Знайти закони розподілення випадкових величин X та Y, якщо їх система цих випадкових величин $\left(X,Y\right)$ має такий закон розподілення:

	x_1	x_2
y_1	0,1	0,55
y_2	0,2	0,15

Розв'язання. Знайдемо

$$p\left(x_1\right) = p\left(x_1, y_1\right) + p\left(x_1, y_2\right) = 0,1 + 0,2 = 0,3 ;$$
$$p\left(x_2\right) = p\left(x_2, y_1\right) + p\left(x_2, y_2\right) = 0,55 + 0,15 = 0,7 .$$

Це означає, що закон розподілення складової X має вигляд:

X	x_1	x_2
p	0,3	0,7

Аналогічно

$$p\left(y_1\right) = p\left(x_1, y_1\right) + p\left(x_2, y_1\right) = 0,1 + 0,55 = 0,65 ;$$

$$p\left(y_2\right) = p\left(x_1, y_{21}\right) + p\left(x_2, y_2\right) = 0,2 + 0,15 = 0,35 .$$

Закон розподілення складової Y має вигляд:

Y	y_1	y_2
p	0,65	0,35

Найважливішою числовою характеристикою багатовимірних випадкових величин, крім математичного сподівання й дисперсії кожної компоненти, є кореляційний момент, що характеризує взаємозв'язок її одномірних складових. Якщо, наприклад, розглянути систему двох дискретних випадкових величин (табл. 5.4), то кореляційний момент (5.8) має вигляд

$$\mu_{xy} = \sum_{i=1}^{2}\sum_{j=1}^{2}(x_i - M\left(X\right))(y_j - M\left(Y\right))p_{ij}$$

і може набувати значення з інтервалу

$$[-\alpha;\alpha], \quad \text{де} \quad \alpha = \frac{\left|x_2 - x_1\right|\left|y_2 - y_1\right|}{4} .$$

Водночас крайні значення μ_{xy} відповідають найсильнішому взаємозв'язку складових X і Y, а середнє значення кореляційного моменту $\left(\mu_{xy} = 0\right)$ вказує на відсутність зв'язку між X і Y (тобто їх незалежність).

Так, за умови $\mu_{xy} = \alpha$ закон розподілення двовимірної випадкової величини (X,Y) зображують у вигляді такої таблиці (табл. 5.5).

Таблиця 5.4

	x_1	x_2
y_1	p_{11}	p_{12}
y_2	p_{21}	p_{22}

Таблиця 5.5

	x_1	x_2
y_1	½	0
y_2	0	½

стверджувати, що $Y = y_1$. Такий взаємний зв'язок випадкових величин називають функціональним, і він є найсильнішим з усіх можливих тому, що елемент випадковості повністю зникає. З іншого боку, рівність $\mu_{xy} = 0$ у нашому випадку приводить до пропорційності рядків і стовпців таблиці закону розподілення:

$$\frac{p_{11}}{p_{12}} = \frac{p_{21}}{p_{22}}.$$

Отже, закон розподілення випадкової величини Y один і той самий, як за умови $X = x_1$, так і при $X = x_2$, а також і за умови невідомого значення X. Це й визначає незалежність випадкових величин X і Y. Частіше найзручнішою числовою характеристикою, ніж кореляційний момент, для оцінки тісноти зв'язку є безрозмірний коефіцієнт кореляції:

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \cdot \sigma_y}, \quad |r_{xy}| \le 1. \tag{5.9}$$

У розглянутому прикладі коли $\mu_{xy} = \pm\alpha$, одержимо $r_{xy} = \pm 1$. Зазначимо, що для більш складних систем випадкових величин (наприклад, із більшою кількістю можливих значень X, Y) зв'язок між компонентами не так легко оцінюється, але в будь-якому разі зберігаються такі властивості:

а) якщо X і Y незалежні, то $r_{xy} = 0$;

б) якщо $Y = aX + b$ або $X = a,\ Y = b$, то $r_{xy} = \pm 1$.

Геометрична інтерпретація різних видів кореляційного зв'язку наведена на рисунку 5.3.

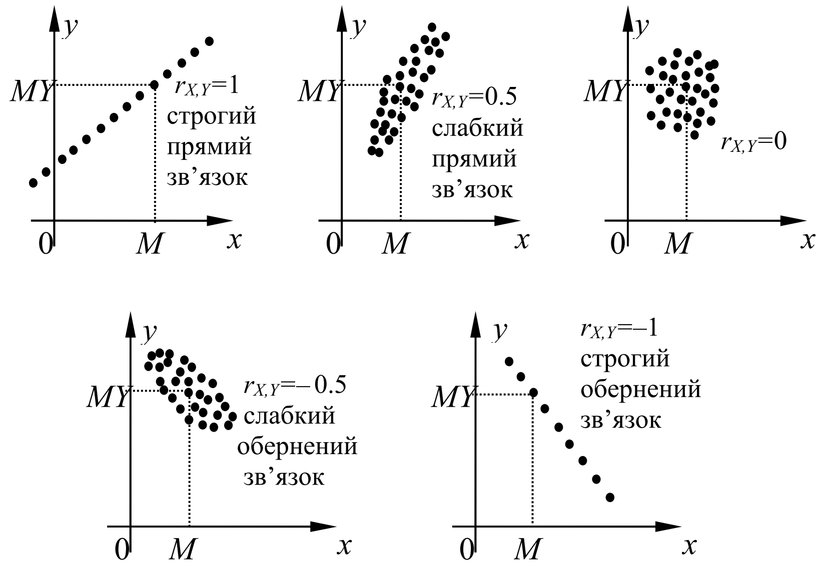


Рисунок 5.3 – Різні види кореляційного зв'язку згідно з формулою (5.9)

5.5 Лінійна регресія

Розглянемо поняття умовного математичного сподівання. Якщо ми знаємо розподілення однієї координати випадкового вектора за умови, що інша координата набуває певного значення, то можна ввести поняття умовного математичного сподівання.

Умовним математичним сподіванням $M(Y/X = x)$ випадкової величини Y за умови, що випадкова величина $X = x$, називають число, яке знаходиться за формулою

$$M\left(Y \big/ X = x\right) = \sum_{j=1}^{n(+\infty)} y_j \cdot P\left\{Y = y_j \big/ X = x_k\right\} \qquad (5.10)$$

для дискретного випадкового вектора, та

$$M\left(Y \big/ X = x\right) = \int_{-\infty}^{+\infty} y \cdot p\left(y \big/ X = x\right) dy \qquad (5.11)$$

для неперервного випадкового вектора.

Аналогічно визначається умовне математичне сподівання $M(X/Y = y)$.

У разі зміни x, взагалі кажучи, змінюється умовне математичне сподівання $M(Y/X = x)$, яке можна розглядати у цьому разі як функцію x:

$$M(Y/X = x) = g(x).$$

Цю функцію називають ***регресією*** Y на X (Y відносно X), а її графік $y = g(x)$ – ***лінією регресії*** Y на X.

Аналогічно визначають регресію X на Y та лінію регресії X на Y:

$$M(X/Y=y)=h(y), \quad x=h(y).$$

Випадкові величини X та Y називаються ***лінійно корельованими***, якщо лінії регресії є прямими. Рівняння цих прямих такі:

$$\begin{aligned} y &= MY + \beta_{Y/X}(x - MX), \\ x &= MX + \beta_{X/Y}(y - MY). \end{aligned} \qquad (5.12)$$

Зміст регресії Y на X полягає в тому, що функція $g(X)$ є найкращим наближенням до випадкової величини Y. Це означає, що для довільної функції $v(X)$ виконується співвідношення

$$M\,[Y - v(X)]^2 \geq M\,[Y - g(X)]^2.$$

Якщо лінія регресії Y на X (X на Y) не є прямою, можна використати одну з прямих регресії (5.12) в якості наближення до істинної лінії регресії. У цьому разі цю пряму називають прямою ***наближеної*** регресії, водночас (5.12) є найкращим наближенням до Y (до X) серед усіх лінійних функцій випадкової величини X (або Y).

Кутові коефіцієнти $\beta_{Y/X}$ і $\beta_{X/Y}$ називаються відповідно ***коефіцієнтами регресії*** Y на X та X на Y. Водночас

$$\beta_{Y/X} = r_{X,Y}\frac{\sigma_Y}{\sigma_X}, \quad \beta_{X/Y} = r_{X,Y}\frac{\sigma_X}{\sigma_Y}. \qquad (5.13)$$

Прямі регресії (5.12) проходять через точку з координатами (MX; MY). У разі $|r_{X,Y}|=1$ прямі регресії співпадають, а в разі $r_{X,Y}=0$ – паралельні осям координат.

Як міру розсіювання випадкової величини Y відносно регресії $g(X)$ (Y на X) розглядають ***кореляційне відношення***

$$\rho^2_{Y/X} = \frac{M[g(X) - MY]^2}{\sigma^2_Y}. \qquad (5.14)$$

Із (5.14) випливають такі властивості кореляційного відношення:

- $0 \leq \rho^2_{Y/X} \leq 1$;
- $\rho^2_{Y/X} = 1$ тоді й лише тоді, коли між випадковими

величинами X та Y є функціональна залежність $Y = g(X)$;

- $\rho^2_{Y/X} = 0$ тоді й лише тоді, коли $g(X)=MY$, тобто лінія регресії є горизонтальною прямою і, тим самим, випадкові величини X та Y є некорельованими.

Узагалі кажучи, $\rho^2_{Y/X} \neq \rho^2_{X/Y}$.

Приклад. Випадковий вектор має такі числові характеристики: $MX = 1$, $DX = 4$, $MY = 3$, $DY = 9$, $r_{X,Y} = 0.7$. У цьому випробуванні випадкова величина мала значення $X = 1$. Яке математичне сподівання випадкової величини Y?

Розв'язання. Скористаємося рівнянням прямої наближеної регресії Y на X (5.12):

$$y = 3 + \frac{3}{2} \cdot 0.7\left(x - 1\right).$$

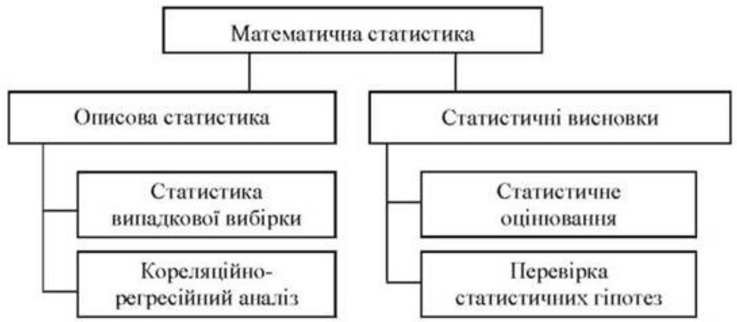
Підставляючи в це рівняння $x = 1$, одержимо $y = 3$. Отже, $M(Y/_{X=1}) = 3$.

ЧАСТИНА III. МАТЕМАТИЧНА СТАТИСТИКА

Розділ 6. Елементи математичної статистики

6.1 Основні задачі математичної статистики

Метою математичної статистики є розроблення методів оброблення статистичних даних для одержання наукових і практичних висновків. Теоретичною основою математичної статистики є теорія ймовірностей. В структурі математичної статистики традиційно виділяють два основні розділи: *описова статистика* й статистичні висновки



Описову статистику використовують для:

узагальнення показників однієї змінної (статистика випадкової вибірки);

виявлення взаємозв'язків між двома та більше змінними (кореляційно-регресійний аналіз).

Описова статистика дає можливість одержати нову інформацію, швидше зрозуміти та всебічно оцінити її, тобто виконує наукову функцію опису об'єктів дослідження, чим і виправдовує свою назву. Методи описової статистики покликані перетворити сукупність окремих емпіричних даних на систему наочних для сприйняття форм і чисел: розподілення частот; показники тенденцій, варіативності, зв'язку. Цими методами розраховують статистики випадкової вибірки, що служать підставою для здійснення статистичних висновків.

Статистичні висновки надають можливість:

- оцінити точність, надійність та ефективність вибіркових статистик, виявити похибки, що виникають у процесі статистичних досліджень (статистичне оцінювання);
- узагальнити параметри генеральної сукупності, одержані на підставі вибіркових статистик (перевірка статистичних гіпотез).

6.2 Генеральна та вибіркова сукупності

Групу об’єктів, поєднаних за якоюсь якісною чи кількісною ознакою, називають статистичною сукупністю. Розрізнюють генеральну й вибіркову сукупності. Вибірковою сукупністю або вибіркою називають сукупність випадково відібраних об’єктів. Генеральною сукупністю називають сукупність об’єктів, із яких виконують вибірку. Об’ємом сукупності називають кількість об’єктів, що до неї входить.

Термін «генеральна сукупність» використовують тоді, коли йдеться про велику, але кінцеву сукупність досліджуваних об'єктів. Генеральні сукупності можуть бути значними, скінченними й нескінченними. На практиці зазвичай мають справу зі скінченними сукупностями. І якщо відношення обсягу генеральної сукупності до обсягу вибірки становить більш, ніж 100, то, за словами Гласса й Стенлі, методи оцінювання для скінченних і нескінченних сукупностей дають по суті однакові результати. Генеральною сукупністю можна називати й повну сукупність значень якоїсь ознаки. Приналежність вибірки до генеральної сукупності є

основною підставою для оцінювання характеристик генеральної сукупності за характеристиками вибірки.

Основна *ідея* математичної статистики грунтується на переконанні про те, що повне вивчення всіх об'єктів генеральної сукупності в більшості наукових завдань або практично неможливе, або економічно недоцільне, оскільки вимагає багато часу та значних матеріальних витрат. Тому в математичній статистиці застосовують *вибірковий підхід*.

Просту випадкову вибірку здійснюють за допомогою простого випадкового відбору (кожний елемент із однаковою ймовірністю може потрапити до вибірки).

Вибірка повинна бути **репрезентативною**, тобто представницькою – її дані повинні правильно відображати відповідну ознаку генеральної сукупності, яку вивчають.

Способи відбору:

1). вибір, що не потребує розділення генеральної сукупності на частини: простий випадковий неповторюваний відбір та простий випадковий повторюваний відбір;

2). вибір, за якого генеральну сукупність поділяють на частини: типовий, механічний та серійний відбори.

Найкращий спосіб здійснення простої вибірки – використання випадкових вибіркових чисел. Ці числа складаються із цифр від 0 до 9, генеруються випадково (ймовірність розміщення від будь-якої цифри в будь-якому місці таблиці не більше й не менше ймовірності розміщення будь-якої іншої цифри з десяти наведених цифр), їх записуються до спеціальної таблиці. Використання таблиць випадкових чисел гарантує, що не буде скоєно систематичної помилки, тобто помилки, яка робить дані не репрезентативними.

6.3 Первинне оброблення даних

Статистичне дослідження проводять за таким планом:

1. формулюють завдання дослідження й визначають об'єм, місце та час потрібної вибірки.
2. збирають необхідні дані та наочно подаються (аналітично, таблично, графічно).
3. проводять оброблення зібраних статистичних даних та формулюють висновок.

Наприклад, ми маємо вибірку: $x_1, x_2, x_3, \ldots, x_n$. Об'єм вибірки дорівнює n. Припустимо, що ці числа одержані в результаті вимірювання маси деталей тощо.

Подальші наші дії залежать від кількості в цій вибірці різних чисел.

Якщо різних чисел небагато, то маємо справу з дискретною величиною. Якщо багато – з неперервною. Відповідно до цього розглядаємо два випадки: дискретний і неперервний.

Дискретний випадок

Перший етап оброблення вибірки – це складання варіаційного ряду. Його одержують так: серед усіх значень x_i $(i = 1,2,3,\ldots,n)$ відбирають усі різні і розміщують у порядку зростання, одержимо

$a_1, a_2, a_3, \ldots, a_m$ де $m \le n$, $m=1,2,3, \ldots$

$a_1 < a_2 < a_3 < \ldots < a_m$

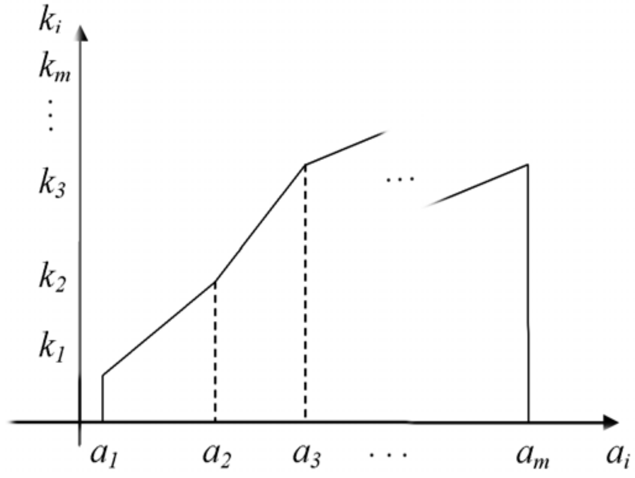
Другий етап оброблення вибірки – це складання дискретної таблиці частоти й відносної частоти.

a_1	a_2	a_3	…	a_m
k_1	k_2	k_3	…	k_m
$n_1=\frac{k_1}{n}$	$n_2=\frac{k_2}{n}$	$n_3=\frac{k_3}{n}$	…	$n_m=\frac{k_m}{n}$

де $\sum k_i = n,\ \sum n_i = 1$,

k_i – число вимірювань, у яких спостерігалася ознака a_i або частота, $n_i=\frac{k_i}{n}$ – відносні частоти.

Третій етап – графічне зображення дискретної таблиці. Це можна зробити за допомогою стовпцевої діаграми та полігону частот. Для зображення діаграми будують систему координат, на осі абсцис відкладають значення a_i, а на осі ординат – відповідні частоти. Потім будуємо точки з координатами (a_i, k_i) а з них опускаємо перпендикуляри на вісь абсцис. Одержуємо:



Полігоном частот (відносних частот) називають ламану, відрізки якої з'єднують точки, абсцисами яких є значення варіант a_i, а ординатами – відповідні їм частоти k_i (відносні частоти n_i).

Неперервний випадок

Якщо різних значень у вибірці буде багато, або всі вони будуть різними, то складена таблиця частот не демонструє особливостей вибірки. У цьому разі її здійснюють поетапно:

Перший етап. Увесь проміжок зміни значень вибірки від найменшого до найбільшого розбивають на інтервали або на класи. Важливе значення має вибір оптимальної величини інтервалу та правильне включення варіант у відповідний інтервал. Для цього в кожному інтервалі потрібно розрізняти верхню й нижню межу. Розмах всієї вибірки дорівнює $R = x_{\max} - x_{\min}$. Оптимальна кількість інтервалів (m) зазвичай лежить у межах від 5 до 15. Нижню межу i-го інтервалу позначають $x_i(min)$, верхню – $x_i(man)$, де i змінюється від 1 до m.

Другий етап. Розбивши ряд на інтервали підраховують число значень із вибірки (частоти), що потрапили в кожний інтервал $[p_i, p_{i+1}]$, а потім відносні частоти. У результаті одержуємо інтервальну таблицю частот

$[p_1, p_2]$	$(p_2, p_3]$	...	$(p_m, p_{m+1}]$
k_1	k_2	...	k_m
$N_1 = \frac{k_1}{n}$	$N_2 = \frac{k_2}{n}$	...	$N_m = \frac{k_m}{n}$

де n – об’єм вибірки, m – число інтервалів, k_i – кількість значень, що потрапила в i-й інтервал.

Третій етап. Графічною ілюстрацією таблиці частоти є гістограма та полігон.

Гістограмою частот (відносних частот) називають ступінчасту фігуру, що складається з прямокутників основами яких є довжини інтервалів значень вибірки, а висоти дорівнюють $\frac{k_i}{h}$.

Площа i-го стовпчика дорівнює n_i , а площа всієї гістограми відносних частот дорівнює одиниці.

Полігон для інтервальної таблиці частоти легко дістати з гістограми. Для цього досить сполучити відрізками середини верхньої сторони прямокутників.

На практиці часто замість повного вивчення даних

вибірки буває достатньо обмежитися знаходженням їх числових характеристик. За аналогією з числовими характеристиками дискретних випадкових величин визначають вибіркові числові характеристики, водночас замінюючи ймовірності відносними частотами.

6.4 Числові характеристики статистичного розподілення

Статистичним розподіленням вибірки називають перелік варіантів та відповідних частот або відносних частот. Статистичний закон розподілення зручно задавати таблицею, що встановлює зв’язок між значеннями випадкової величини та їх частотами.

Числові характеристики, обчислені за вибіркою називають ***статистиками***.

Числові характеристики, обчислені за генеральною сукупністю називають ***параметрами***.

Наведемо основні статистики.

Мода – це значення, яке в статистичному ряді трапляється найчастіше (M_0).

Медіана – це значення, що посідає центральне місце у впорядкованому ряді розподілення (M_e). Для дискретного впорядкованого варіаційного ряду з непарним числом елементів медіану знаходять як варіанту x із порядковим номером $\frac{n+1}{2}$, тобто $M_e = x_{\frac{n+1}{2}}$.

Для ряду з парним числом елементів медіану розраховують як середню арифметичну двох варіант із порядковими номерами $\frac{n}{2}$ та $\frac{n}{2}+1$: $M_e = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$.

Для інтервального ряду розподілення медіану обчислюють за формулою

$$M_e = x_0 + h\frac{\frac{1}{2}\sum n_i - S_{Me-1}}{n_{Me}},$$

де x_0 – нижня межа медіанного інтервалу, h – довжина медіанного інтервалу, S_{Me-1} – сума частот накопичених перед медіанним інтервалом, $\frac{1}{2}\sum n_i$ – половина суми частот, n_{Me} – частота медіанного інтервалу.

Вибірковою середньою статистичного розподілення вибірки $\overline{x}$ називають середню арифметичну значень її варіант x_i з урахуванням їхніх частот k_i, тобто

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{m} x_i k_i = \frac{k_1x_1 + k_2x_2 + \ldots + k_mx_m}{n}. \qquad (6.1)$$

Вибіркова середня є основною характеристикою статистичного розподілення вибірки та аналогом математичного сподівання.

Вибірковою дисперсією D статистичного розподілення вибірки називають середню арифметичну квадратів відхилень варіант від вибіркової середньої, тобто

$$D = \frac{1}{n}\sum k_i(x_i - \overline{x_i})^2. \qquad (6.2)$$

Якщо вибірка складається з двох (або кількох) груп, то внутрішньогрупову дисперсію обчислюється за формулою

$$D_{\text{вн. гр.}} = \frac{D_A \cdot n_A + D_B \cdot n_B}{n_A + n_B}, \qquad (6.3)$$

де D_A, D_B – дисперсії груп A і B; n_A, n_B – об’єми груп A та B.

Міжгрупова дисперсія

$$D_{\text{між}} = \frac{\left(\overline{x}_A - \overline{x}\right)^2 \cdot n_A + \left(\overline{x}_B - \overline{x}\right)^2 \cdot n_B}{n_A + n_B}. \qquad (6.4)$$

Вибіркова дисперсія дорівнює сумі внутрішньогрупової та міжгрупової дисперсій.

Асиметрією $\boldsymbol{A_B}$ називають відношення центрального емпіричного моменту третього порядку до куба середнього квадратичного відхилення

$$A_B = \frac{m_3}{\sigma_B^{\,3}} \qquad (6.5)$$

де m_3 – центральний емпіричний момент 3-го порядку, σ_B – середнє квадратичне відхилення статистичного розподілення вибірки.

Ексцесом $\boldsymbol{E_B}$ називають зменшене на три одиниці відношення центрального моменту четвертого порядку до четвертого степеня середнього квадратичного відхилення:

$$E_B = \frac{m_4}{\sigma_B^{\,4}} - 3, \qquad (6.6)$$

де m_4 – центральний емпіричний момент 4-го порядку, σ_B – середнє квадратичне відхилення статистичного розподілення вибірки.

Усі вищезазначені формули можуть бути використані під час обчислення числових характеристик вибірки для випадку, коли емпіричні дані згруповані за допомогою інтервального варіаційного ряду, зокрема, якщо вважати, що x_i – середини інтервалів.

6.5 Точкові оцінювання невідомих параметрів розподілення

Повною характеристикою випадкової величини є її закон розподілення. Для його встановлення дослідними методами потрібні певні затрати ресурсів і часу. Однак часто є підстави вважати, що той чи інший закон розподілення є відомим (наприклад: пуассонівський, нормальний, показниковий тощо). Але щоб конкретизувати цей закон, потрібно знати його параметри, що називають ***параметрами розподіленя***. Зокрема для нормального закону розподілення параметрами є a і σ, для пуассонівського – λ, для показникового – α тощо.

Отже, вивчаючи певну ознаку X генеральної сукупності, ми можемо знати характер закону розподілення випадкової величини X, але параметри цього закону залишаються невідомими. Тоді виникає завдання визначити наближені числові значення невідомих параметрів розподілення. Такі наближені числові значення параметрів розподілення називають їх точковими **статистичними оцінками**, або просто ***точковими оцінками***.

Щоб статистичні оцінки давали найкращі наближення параметрів, вони повинні задовольняти певні вимоги. Розглянемо їх.

Нехай θ^* є статистична оцінка невідомого параметра θ теоретичного розподілення. Припустимо, що за вибіркою об'єму n знайдена оцінка θ_1^*. При інших вибірках того ж об'єму одержимо деякі інші оцінки $\theta_2^*, \theta_3^*, ..., \theta_m^*$. Саме оцінку θ^* можна розглядати, як випадкову величину, а числа $\theta_2^*, \theta_3^*, ..., \theta_m^*$, як її можливі значення.

Якщо числа θ_k^* (k=1,2, …,m) будуть більші значення θ, тоді оцінка θ^* дає наближене значення θ з надлишком. У цьому разі математичне сподівання випадкової величини θ^* буде більше θ, $M(\theta^*) > \theta$.

Якщо θ^* дає оцінку θ з недостачею, тоді $M(\theta^*) < \theta$.

Отже, використання статистичної оцінки, математичне сподівання якої не дорівнює параметра θ, приводить до систематичних похибок.

Вимога $M(\theta^*) = \theta$ застерігає від систематичних похибок.

Статистичні оцінки класифікують за такими ознаками:

- статистичну оцінку θ^* параметра θ називають ***незміщеною***, якщо $M(\theta^*)=\theta$;
- ***зміщеною*** будемо називати статистичну оцінку, для якої $M(\theta^*)\neq\theta$;
- ***ефективною*** називають статистичну θ^* оцінку, що при заданому обсязі вибірки n має найменшу дисперсію;
- ***змістовною*** називають статистичну оцінку, яка в разі $n\to\infty$ прямує за ймовірністю до параметра, що оцінюють. Прикладом змістовної оцінки можуть слугувати закони великих чисел, наприклад, теорема Бернуллі. Вимогу змістовності висувають до статистичних оцінок під час розглядання вибірки великого обсягу;
- ***точковою*** називають оцінку, визначену одним числом. Наприклад, вибіркова середня, вибіркова дисперсія – точкові оцінки відповідних числових характеристик генеральної сукупності.

Розглянемо точкову оцінку **математичного сподівання**. Нехай $x_1, x_2, x_3, ..., x_n$ – вибірка одержана в результаті n незалежних випробувань над випадковою величиною X – деякою ознакою генеральної сукупності, що має математичне сподівання $M(X)=a$. За точкову оцінку математичного сподівання a беруть вибіркове середнє $a_n^* = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Легко довести, що $a_n^* = \overline{x}$ є незміщеною для $M(X) = a$, тобто $M(\overline{x}) = a$.

Якщо додатково припустити, що випадкова величина X має скінчену дисперсію $D(X)=\sigma^2$, тоді можна стверджувати, що оцінка $a_n^* = \overline{x}$ є змістовною. Якщо обчислити дисперсію вибіркової середньої $\overline{x}$, то одержимо

$$D(X) = D\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(x_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

Оскільки $\lim\limits_{n\to\infty} D(\overline{x}) = \lim\limits_{n\to\infty} \frac{\sigma^2}{n} = 0$, то це означає, що оцінка $\overline{x}$ є змістовною для параметра a.

Необхідно зазначити, що, якщо випадкова величина X нормально розподілена з параметрами $M(X)=a$ й $D(X)=\sigma^2$, то оцінка $\overline{x}$ має в класі всіх незміщених оцінок математичного сподівання a мінімальну дисперсію, що дорівнює $\frac{\sigma^2}{n}$. Тому $\overline{x}$ є ефективною оцінкою параметра a.

Розглянемо точкову оцінку ***дисперсії.*** За точкову оцінку дисперсії беруть вибіркову дисперсію $D = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$, що є зміщеною оцінкою параметра $D(X)=\sigma^2$. Цей факт випливає з рівності $M(D) = \frac{n}{n-1}\sigma^2$, яку неважко встановити за допомогою безпосередніх обчислень. Тому вибіркову дисперсію доцільно виправити так, щоб вона стала незміщеною оцінкою. Для цього достатньо D помножити на дріб $\frac{n}{n-1}$.

Виправлену вибіркову дисперсію позначають $S^2 = \frac{n}{n-1} D$.

Тоді ***виправленим середньоквадратичним відхиленням*** вибірки буде $s = \sqrt{S^2}$.

Можна показати, що оцінки D і S^2 є змістовними і не є ефективними.

У випадку, коли математичне сподівання a відоме й випадкова величина X нормально розподілена, то незміщеною, змістовною та ефективною оцінкою дисперсії $D(X)=\sigma^2$ є оцінка $\frac{1}{n}\sum_{i=1}^{n}\left(x_i - a\right)^2$

Точкові оцінки параметрів розподілення є випадковими величинами, їх можна вважати первинними результатами обробки вибірки тому, що невідомо, з якою точністю кожна з них оцінює відповідну числову характеристику генеральної сукупності.

Однак, у разі малого об’єму вибірки точкові оцінки можуть мати значні розходження зі значенням оцінюваного параметра. Це призводить до грубих помилок.

6.6 Метод моментів

Метод моментів ґрунтується на тому, що невідомі параметри теоретичного розподілення (розподілення генеральної сукупності) визначають із рівнянь, які добуваються прирівнюванням важливіших числових характеристик (моментів) теоретичного розподілення відповідним числовим характеристикам емпіричного розподілення. Так, нехай заданий, наприклад, вид теоретичного розподілення $F(x,\theta)$, що визначається невідомим параметром θ. Для знаходження одного параметра необхідне одне рівняння відносно даного параметра. Для цього використовують момент 1-го порядку (математичне сподівання) теоретичного розподілення й відповідна числова характеристика емпіричного розподілення – вибіркове середнє. Знаходимо математичне сподівання як функцію від θ

$$M(X) = \int_{-\infty}^{\infty} x\,\mathrm{d}F(x,\theta) = m(\theta)$$

і функцію вибірки

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

Порівнюючи їх, одержуємо рівняння для визначення оцінки невідомого параметра θ

$$m(\theta) = \overline{x}.$$

Для знаходження оцінок двох невідомих параметрів θ_1, θ_2 звичайно беруть математичне сподівання й дисперсію теоретичного розподілу та відповідні їм числові характеристики емпіричного розподілення – вибіркове середнє $\overline{x}$ і вибіркову дисперсію s^2. Одержують два рівняння

$$m(\theta_1,\theta_2) = \overline{x}, \quad \sigma^2(\theta_1,\theta_2) = s^2 .$$

Розв'язуючи цю систему, знаходять відповідні оцінки $\tilde{\theta}_1, \tilde{\theta}_2$.

Оцінки методу моментів звичайно є *слушними*, однак за ефективністю вони не є найкращими. Тим не менш, метод моментів часто використовують на практиці, оскільки приводить до порівняно простих обчислень.

Приклад. Знайдемо методом моментів оцінювання

невідомого параметра λ експоненціального розподілення.

Генеральна сукупність має експоненціальне розподілення із густини розподілення $f(x,\lambda) = \lambda e^{-\lambda x}$ $(x > 0)$, де λ – невідомий параметр. Вибіркові значення $x_1,...,x_n$ узяті з однієї й тієї самої генеральної сукупності. Знайдемо методом моментів оцінку $\tilde{\lambda}$ невідомого параметра розподілення λ.

Розв'язання. Визначимо математичне сподівання експоненціального розподілення:

$$m(\lambda) = \lambda\int_{0}^{\infty} xe^{-\lambda x}\mathrm{d}x = \frac{1}{\lambda}.$$

Далі за вибіркою знаходимо вибіркове середнє емпіричного розподілення $\overline{x}$:

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

Із рівняння

$$m(\lambda) = \frac{1}{\lambda} = \overline{x}.$$

знаходимо оцінку $\tilde{\lambda}$ параметра λ:

$$\tilde{\lambda} = \frac{1}{\overline{x}}.$$

Особливо зручне застосування методу моментів, коли шукані параметри розподілення самі є деякими числовими характеристиками. Знайдемо, наприклад, методом моментів оцінювання невідомих параметрів a та

σ нормального розподілення.

Приклад. Вибірка $x_1,...,x_n$ одержана з генеральної сукупності з нормальним розподіленням зі щільністю розподілення:

$$f(x,a,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}},\ (-\infty<x<\infty),$$

де a і σ – невідомі параметри розподілення. Знайдемо методом моментів оцінювання $\tilde{a}$ та $\tilde{\sigma}$ невідомих параметрів a і σ.

Розв'язання. Визначимо математичне сподівання й дисперсію розподілення:

$$a=\int_{-\infty}^{\infty}x\,f(x,a,\sigma)dx\,,\quad \sigma^2=\int_{-\infty}^{\infty}(x-a)^2\,f(x,a,\sigma)dx.$$

За даними вибірки знаходимо вибіркові числові характеристики m i s і прирівнюємо їх до відповідних числових характеристик теоретичного розподілення: $a=m,\ \ \sigma=s$.

У результаті одержуємо шукані оцінки параметрів: $\tilde{a}=m,\ \tilde{\sigma}=s$.

6.7 Метод максимальної правдоподібності

Одним із найбільш універсальних методів одержання оцінок параметрів розподіленя генеральної сукупності є метод *максимальної правдоподібності.*

Основу його становить *функція правдоподібності*, що виражає густину ймовірності (ймовірність) сумісної появи результатів вибірки $x_1,...,x_n$:

$$L(x_1,...,x_n,\theta)=f(x_1,\theta)\,f(x_2,\theta)...\,f(x_n,\theta)=\prod_{i=1}^{n}f(x_i,\theta)\,.$$

Згідно з методом максимальної правдоподібності за оцінку невідомого параметра θ приймають таке його значення $\tilde{\theta}$, що максимізує функцію L. Величина $\tilde{\theta}$, за якої функція правдоподібності досягає максимального значення, називається *оцінкою максимальної правдоподібності*.

Природність такого підходу до визначення статистичних оцінок випливає зі смислу функції правдоподібності, яка при кожному фіксованому значенні параметра θ є *мірою правдоподібності* одержання вибірки $x_1,...,x_n$. Оцінка $\tilde{\theta}$ є такою, що вибірка $x_1,...,x_n$, яка одержана у результаті спостережень, є найбільш вірогідною. Знаходження оцінки $\tilde{\theta}$ спрощується, якщо максимізувати не саму функцію L, а $\ln(L)$, оскільки максимум обох функцій досягається в разі одного й того самого значення θ. Для одержання оцінки максимальної правдоподібності треба розв'язати рівняння

$$\frac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}=0\,.$$

Якщо потрібно оцінити не один, а декілька параметрів $\theta_1,...,\theta_m$, то оцінку максимальної правдоподібності цих параметрів знаходять із системи рівнянь

$$\frac{\partial\ln L(\theta_1,...,\theta_m)}{\partial\theta_i}=0\quad(i=1,...,m)\,.$$

Переваги методу максимальної правдоподібності полягає у тому, що для широкого класу розподілень він приводить до оцінок, які є слушними, асимптотично ефективними, мають асимптотично нормальне розподілення і, якщо для параметра θ існує ефективна оцінка, то рівняння правдоподібності має єдиний розв'язок, що співпадає з нею. Однак оцінка максимальної правдоподібності може виявитися зсуненою.

Приклад. Методом максимальної правдоподібності оцінимо параметр λ розподілення Пуассона.

Розв'язання. Генеральна сукупність має розподілення Пуассона

$$P_n(X = x_i) = \frac{\lambda^{x_i}}{x_i!} e^{-\lambda},$$

де n – кількість випробувань у кожній серії, x_i – кількість появ події у i-й серії $(i = 1,...,n)$. Знайдемо оцінку $\tilde{\lambda}$ невідомого параметра λ за вибіркою $x_1,...,x_n$.

Складемо функцію правдоподібності:

$$L(x_1,...,x_n) = p(x_1,\lambda)\,p(x_2,\lambda)...p(x_n,\lambda)$$

Визначимо логарифм цієї функції

$$\ln L(\lambda) = \sum x_i \ln\lambda - n\lambda - \ln(x_1!...x_2!) =$$

$$= \frac{\lambda^{x_1} e^{-\lambda}}{x_1!} \cdot \frac{\lambda^{x_2} e^{-\lambda}}{x_2!} ... \frac{\lambda^{x_n} e^{-\lambda}}{x_n!} = \frac{\lambda^{\sum x_i} e^{-n\lambda}}{x_1!...x_n!}.$$

Прирівнюючи похідну цієї функції за λ до нуля,

одержуємо рівняння правдоподібності

$$\frac{d\ln L(\lambda)}{d\lambda} = \frac{\sum x_i}{\lambda} - n = 0.$$

Розв'язуючи це рівняння відносно λ, знаходимо:

$$\tilde{\lambda} = \frac{1}{n}\sum x_i = \bar{x}.$$

Оскільки

$$\frac{d^2 \ln L(\lambda)}{d\lambda^2} = -\frac{\sum x_i}{\lambda^2} < 0,$$

то $\tilde{\lambda} = \bar{x}$ є точкою максимуму функції $\ln L(\lambda)$.

Із вищенаведеного одержуємо, що $\tilde{\lambda} = \bar{x}$ є оцінка максимальної правдоподібності параметра λ розподілення Пуассона.

Приклад. Методом максимальної правдоподібності знайдемо оцінку параметрів a і σ нормального розподілення.

Розв'язання. Записуємо функцію правдоподібності:

$$\ln L(a,\sigma) = \frac{n}{2}(\ln\ 2n + \ln\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2.$$

Диференціюючи $\ln L(a,\sigma)$ за a і σ^2, одержуємо систему рівнянь

$$\sum_{i=1}^{n}(x_i - a) = 0,$$

$$-\frac{n}{2} + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2 = 0.$$

Звідки знаходимо оцінки

$$\tilde{a} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}, \quad \tilde{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2} = s.$$

Ці оцінки співпадають з оцінками методу моментів. Вони слушні, причому $\tilde{a}$ є незсуненою, а $\tilde{\sigma}$ – зсуненою і, як було сказано раніше, $\tilde{a}$ – ефективна, $\tilde{\sigma}$ – асимптотично ефективна.

6.8 Метод найменших квадратів

Сутність методу найменших квадратів полягає в тому, що оцінки невідомих параметрів розподілення $\theta = (\theta_1, ..., \theta_m)$ визначають з умови мінімізації суми квадратів відхилень вибіркових даних x_i $(i = 1, ..., n)$ від оцінки, що визначається. Наприклад, знайдемо оцінку за методом найменших квадратів $\tilde{\theta}$ для генеральної середньої $\overline{x}_0$. Згідно з цим методом, оцінку $\tilde{\theta}$ знаходимо з умови

$$S(\tilde{\theta}) = \sum_{i=1}^{n} (x_i - \theta)^2 \to \min.$$

Використовуючи необхідну умову екстремуму функції, прирівнюємо до нуля похідну

$$\frac{dS}{d\theta} = -2\sum_{i=1}^{n} (x_i - \theta) = 0 \quad \text{або} \quad \sum_{i=1}^{n} x_i - \theta n = 0.$$

Звідки

$$\tilde{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}.$$

Отже, оцінка генеральної середньої є вибіркова середня $\overline{x}$.

Метод найменших квадратів має широке застосування в практиці статистичних досліджень, оскільки, по-перше, не потребує знання закону розподілення вибіркових даних; по-друге, для нього досить добре розроблений математичний апарат числової реалізації.

Метод найменших квадратів застосовують у моделях кореляційного й регресійного аналізу.

Розділ 7. Статистичні оцінки параметрів розподілення

7.1 Надійні інтервали

У попередньому розділі були розглянуті методи оцінки параметрів генеральної сукупності одним числом. Такі оцінки параметрів називаються *точковими.* Однак точкова оцінка $\tilde{\theta}$ є лише наближеним значенням невідомого параметра θ навіть у тому разі, якщо вона незсунена (у середньому співпадає з θ), слушна (прямує до θ із зростанням n) та ефективна (має найменше відхилення від θ), і для вибірок малого об'єму може істотно відрізнятися від оцінюваного параметра θ і бути ненадійною.

Щоб одержати уявлення про точність і надійність оцінки $\tilde{\theta}$ параметра θ, використовують інтервальні оцінки.

Оцінку називають *інтервальною*, якщо вона визначається двома числами – кінцями інтервалу, що накриває оцінюваний параметр. Інтервальну оцінку визначають за допомогою надійного інтервалу.

Надійним інтервалом для параметра θ називають числовий інтервал $(\tilde{\theta}^{(1)}, \tilde{\theta}^{(2)})$ який із заданою ймовірністю γ накриває невідоме значення параметра θ, тобто

$$P\left(\tilde{\theta}^{(1)} < \theta < \tilde{\theta}^{(2)}\right) = \gamma \ .$$

Імовірність γ, із якою виконується нерівність $(\tilde{\theta}^{(1)} < \theta < \tilde{\theta}^{(2)})$, називають *надійною ймовірністю* або *надійністю оцінки*, а ймовірність $\alpha = 1 - \gamma$ – *рівнем значущості* оцінки.

Оскільки межі інтервалу знаходяться за вибірковими даними, то вони є випадковими величинами на відміну від оцінюваного параметра θ – величини невипадкової. Величина надійного інтервалу істотно залежить від об'єму вибірки n (зменшується зі зростанням n) і від значення надійної ймовірності γ (збільшується з наближенням γ до одиниці). Дуже часто надійний інтервал вибирається симетричним відносно параметра θ, тобто $(\theta - \delta; \theta + \delta)$, де величину δ називають *точністю оцінки*.

Найбільше відхилення δ вибіркової середньої (або частки) від генеральної середньої (або частки), яке можливе із заданою надійною ймовірністю γ, називають *граничною похибкою вибірки*. Похибка δ є *похибкою репрезентативності* (*представництва*) вибірки. Вона виникає в наслідок того, що досліджується не вся сукупність, а лише її частина (вибірка), відібрана випадково. Цю похибку часто називають *випадковою похибкою репрезентативності*. Вона відрізняється від систематичної похибки репрезентативності, що з'являється в результаті порушення принципу випадковості під час відбору елементів сукупності до вибірки.

Зазвичай надійність оцінки γ задають наперед і беруть близькою до одиниці: 0,9; 0,95; 0,99. Чим менша для вибраної ймовірності γ величина $\left|\tilde{\theta}^{(1)} < \theta < \tilde{\theta}^{(2)}\right|$, тим точніша оцінка невідомого параметра θ. Вибір надійної ймовірності не є математичною задачею, а визначається конкретно вирішуваною проблемою. Наприклад, нехай на двох підприємствах імовірність випуску придатної продукції дорівнює $\gamma = 1 - \alpha = 0{,}99$, тобто ймовірність випуску браку $\alpha = 0{,}01$. Перше підприємство випускає парасольки, а друге парашути. Викинути $1\,\%$ парасольок дешевше, ніж змінити технологічний процес. Якщо ж на

100 парашутів трапиться один бракований, то це матиме серйозні наслідки. Отже, у першому випадку ймовірність браку α допускається, а в другому – ні.

Для побудови надійних інтервалів для параметрів генеральних сукупностей можуть бути реалізовані два підходи, що ґрунтуються на знанні точного (за умови даного об'єму вибірки n), або асимптотичного (при $n \to \infty$) розподілення вибіркових характеристик (статистик). Перший підхід реалізується в разі побудови інтервальних оцінок параметрів для малих вибірок, другий – для великих вибірок. Розглянемо спочатку другий підхід, що застосовують для великих вибірок.

7.2 Надійний інтервал для математичного сподівання нормального розподілення *a* при відомому σ

За великого об'єму вибірки ($n \to \infty$) розподілення вибіркових характеристик (статистик) необмежено наближається до нормального (практично за умови $n > 30-40$ розподілення вибіркової середньої $\bar{x}$ можна вважати приблизно нормальним).

Нехай генеральна сукупність (випадкова величина X) розподілена нормально з невідомим математичним сподіванням a й відомим середнім квадратичним відхиленням σ. Треба оцінити невідоме математичне сподівання a (генеральне середнє арифметичне) за вибірковим середнім $\bar{x}$ і знайти надійний інтервал із довірчою ймовірністю γ.

Для *повторної вибірки* вибіркові значення $X_1,...,X_n$ є незалежні випадкові величини, розподілені як і величина X (генеральна сукупність) за нормальним законом. Відносно випадкових величин X_i і $\bar{x}$ відомо:

- математичне сподівання X_i дорівнює a, дисперсія σ^2: $M(X_i)=a$, $D(X_i)=\sigma^2$;
- оскільки величини X_i розподілені нормально, то і їх сума (середнє арифметичне) $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} X_i$ також розподілена за нормальним законом з параметрами
$$M(\bar{x}) = M\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = a,$$
$$\sigma(\bar{x}) = \sqrt{D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)} = \sqrt{\frac{1}{n^2}n\sigma^2} = \frac{\sigma}{\sqrt{n}};$$
- випадкова величина $\dfrac{\bar{x}-a}{\sigma/\sqrt{n}}$ розподілена за нормальним законом $N(0, 1)$ із параметрами 0 і 1 і розподіленням $N(0, 1)$ не залежить від оцінюваного параметра a.

Задамо надійну ймовірність γ та визначимо величину $\delta > 0$ із рівняння
$$\mathrm{P}\left(\left|\bar{x}-a\right| < \delta\right) = \gamma .$$

Оскільки $\bar{x}$ нормально розподілена величина з параметрами a і $\sigma/\sqrt{n}$, то ймовірність того, що $\left|\bar{x}-a\right| < \delta$ буде дорівнювати
$$P\left(\left|\bar{x}-a\right| < \delta\right) = 2\Phi\left(\frac{\delta}{\sigma(\bar{x})}\right) = 2\Phi\left(\frac{\delta}{\sigma/\sqrt{n}}\right) = 2\Phi(t_\alpha),$$
де $\Phi(t)$ – функція Лапласа (інтеграл імовірностей), а t_α дорівнює:
$$t_\alpha = \frac{\delta\sqrt{n}}{\sigma}.$$

Величину t_α визначаємо із рівності

$$2\Phi(t_\alpha)=\gamma \text{ або } \Phi\left(t_\alpha\right)=\frac{\gamma}{2}=\frac{1-\alpha}{2},$$

де $\alpha=1-\gamma$ – рівень значущості. Знайшовши t_α із таблиць значень функції Лапласа, визначаємо граничну похибку оцінки δ:

$$\delta=t_\alpha\sigma\left(\overline{x}\right)=\frac{t_\alpha\sigma}{\sqrt{n}}. \qquad (7.1)$$

Обчисливши δ, одержуємо, що із імовірністю γ виконана нерівність

$$\overline{x}-\delta<a<\overline{x}+\delta. \qquad (7.2)$$

На рисунку 7.1 показано, що випадкова величина $\dfrac{\overline{x}-a}{\sigma/\sqrt{n}}$, яка має стандартне нормальне розподілення, з імовірністю $1-\alpha$ набуває значень, що трапляють в інтервал $(-t_\alpha, t_\alpha)$, отже, з імовірністю $1-\alpha$ виконується нерівність (7.2).

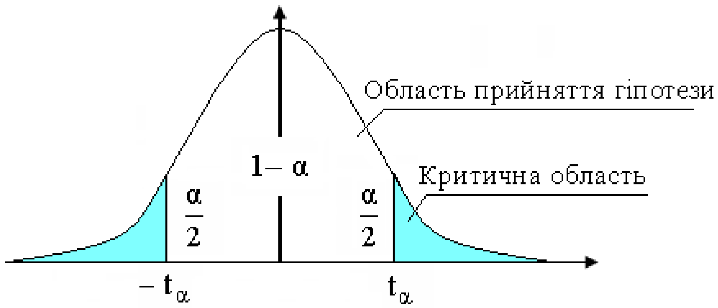


Рисунок 7.1 – Надійний інтервал для математичного сподівання

Отже, інтервал $(\overline{x}-\delta;\overline{x}+\delta)$ є надійним інтервалом для математичного сподівання нормального розподілення. Сенс цього співвідношення такий: упевнено можна стверджувати, що надійний інтервал $(\overline{x}-\delta;\ \overline{x}+\delta)$ покриває невідоме середнє арифметичне генеральної сукупності.

Приклад. Випадкова величина розподілена за нормальним законом із параметром σ. Знайти мінімальний об’єм вибірки, щоб із надійністю γ та точністю Δ виконувалась рівність $\overline{x}=a$, якщо $\sigma=0,5;\ \gamma=0,95;\ \Delta=0,1$.

Розв’язання.

Для $\gamma=0,95$ маємо

$$2\Phi(t)=\gamma \quad\Rightarrow\quad \Phi(t)=\frac{\gamma}{2}=\frac{0,95}{2}=0,475 \quad\Rightarrow\quad t=1,96.$$

Використовуючи формулу знаходження об’єму вибірки, знайдене t, та задані Δ, σ, одержуємо

$$n=\left(\frac{1,96\cdot0,5}{0,1}\right)^2=96,04.$$

Отже, мінімальний об’єм вибірки $n=96$.

Для *безповторної вибірки* об’єму n із генеральної сукупності об’єму N, $\overline{x}$ є сумою залежних випадкових величин. Однак і в цьому разі при $n\to\infty$ закон розподілення $\overline{x}$ як завгодно близько наближається до нормального. Водночас середнє квадратичне відхилення дорівнює

$$\sigma=\left(\overline{x}\right)=\frac{\sigma}{\sqrt{n}}\sqrt{1-\frac{n}{N}},$$

значення якого підставляють у (7.1). Отже, і для безповторної вибірки надійний інтервал для a має вигляд (7.2).

Формула (7.1) пов'язує між собою три величини: надійну ймовірність γ, граничну похибку вибірки δ та об'єм вибірки n. У кожній конкретній задачі дві із цих величин задають і визначають третю. У результаті одержуємо три типи задач:

1. відомі n і γ, визначають δ;
2. відомі n і δ, визначають γ;
3. відомі γ і δ, визначають n.

Перші дві задачі пов'язані з аналізом результатів уже зробленої вибірки даного об'єму n, отже, і із заданою точковою оцінкою $\overline{x}$.

Метою розв'язання задач третього типу є розрахунок необхідного об'єму вибірки n, що забезпечить задану граничну похибку вибірки δ за умови вибраної величини надійної ймовірності γ.

Перша задача – визначення точності оцінки δ – є задача побудови надійних інтервалів для оцінювання параметрів розподілень, що розглянутих вище. Розглянемо тепер другу й третю задачі.

Алгоритм визначення надійної ймовірності γ:

Якщо заданий об'єм вибірки n і гранична похибка вибірки δ і необхідно визначити надійну ймовірність γ, то алгоритм розв'язку задачі такий:

- визначаємо $t_\alpha = \dfrac{\delta\sqrt{n}}{\sigma}$;
- імовірність γ знаходимо за формулою $\gamma = 2\Phi\left(t_\alpha\right)$.

Алгоритм визначення об'єму вибірки:

Установлення об'єму вибірки n для проведення вибіркових спостережень є важливим, оскільки це визначає необхідні при цьому часові, трудові й вартісні витрати. Для визначення n необхідно задати: надійну ймовірність γ (надійність оцінки) та граничну похибку оцінки δ (точність оцінки). Із формули (7.1) одержуємо

$$n = \frac{t^2\sigma^2}{\delta^2}.$$

Якщо вибірка безповторна, то об'єм вибірки з генеральної сукупності об'єму N визначають за формулою

$$n' = \frac{N\,t^2\sigma^2}{t^2\sigma^2 + N\delta^2}.$$

Якщо знайдений об'єм повторної вибірки n, то об'єм безповторної вибірки можна визначити за формулою

$$n' = \frac{nN}{n+N}.$$

Оскільки $\dfrac{N}{n+N} < 1$, то за умов однієї й тієї самої точності й надійності оцінок, об'єм безповторної вибірки n' завжди менший за об'єм повторної вибірки n. Цим пояснюється той факт, що на практиці здебільшого використовують безповторну вибірку.

Припускають, що математичне сподівання a розподілення генеральної сукупності невідоме, але відома її дисперсія σ^2.

Коментар. Цей метод стійкий у разі помірних відхилень розподілення від нормальності.

7.3 Надійний інтервал для математичного сподівання нормального розподілення *a* в разі невідомого *σ*

Нехай випадкова величина X розподілена нормально. Необхідно знайти надійний інтервал для невідомого математичного сподівання a в разі невідомого середнього квадратичного відхилення σ.

Визначимо за даними вибірки $X_1,...,X_n$ об'єму n із генеральної сукупності X вибіркове середнє та виправлену випадкову дисперсію s^2 (незсунену оцінку для σ^2):

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \overline{x})^2.$$

Випадкова величина $\overline{x}$ розподілена нормально з параметрами a і $\sigma / \sqrt{n}$.

Оскільки величина σ невідома, у разі побудови надійного інтервалу для a не можна користуватись нормальністю розподілення $\overline{x}$. У цьому разі розглянемо статистику

$$T = \frac{\overline{x} - a}{s / \sqrt{n}}.$$

Доведено, що випадкова величина T розподілена за законом Стьюдента з $k = n-1$ ступенями свободи, який не

залежить від розподілення X.

Нехай задана надійність оцінки γ. Знайдемо таке число t_γ, щоб виконувалась рівність

$$P\left(\left|\frac{\overline{x} - a}{s / \sqrt{n}}\right| < t_\gamma\right) = \gamma.$$

Замінивши нерівність у дужках рівносильною їй подвійною нерівністю, одержимо

$$P\left(\overline{x} - \frac{t_\gamma s}{\sqrt{n}} < a < \overline{x} + \frac{t_\gamma s}{\sqrt{n}}\right) = \gamma$$

Величину $t_\gamma = t(\gamma, n-1)$ знаходимо з розподілення Стьюдента (t-розподілення). Підставляючи знайдене значення t_γ в нерівність, одержимо надійний інтервал

$$\left(\overline{x} - \frac{t_\gamma s}{\sqrt{n}}; \overline{x} + \frac{t_\gamma s}{\sqrt{n}}\right)$$

що накриває невідомий параметр a з надійністю γ.

7.4 Обробка вибірки методом найменших квадратів

Під час експериментального вивчення функціональної залежності однієї величини Y від іншої величини X роблять ряд вимірів величини y за різних значень x. Результати можуть бути репрезентовані у вигляді таблиці

x	x_1	x_2	…	x_n
y	y_1	y_2	…	y_n

За допомогою методу найменших квадратів (МНК) зображують статистичну функціональну залежність у вигляді аналітичної залежності й виражають такі оцінки параметрів рівняння регресії, що зводять до мінімуму обрану міру розкиду. У результаті відбувається вирівнювання емпіричних значень в одну лінію регресії.

Водночас, для однозначного визначення як міру розкиду використовують один із показників розсіювання випадкової величини – дисперсію.

Припустимо, що діаграма розсіювання така, що між величинами x та y існує лінійна залежність

$$\hat{y} = ax + b,$$

де параметри a та b невідомі.

Це означає, що відхилення фактичних значень функції від «підібраної прямої» $\delta_i = y - \overline{y_i}$ повинні бути мінімальними, тобто пряму підбирають так, щоб сума квадратів відхилень була мінімальною

$$\delta_1^2 + \delta_2^2 + \ldots + \delta_n^2 \Rightarrow \min .$$

Тоді повинна виконуватися рівність

$$\delta = \sum_{i=1}^{n} (y_i - ax_i - b)^2 = \min .$$

Потрібно визначити параметри a та b так, щоб z досягло мінімуму.

Відомо, що необхідна умова існування мінімуму полягає в тому, що

$$\frac{\partial \delta}{\partial a} = 0, \quad \frac{\partial \delta}{\partial b} = 0.$$

Після диференціювання та спрощень, одержимо систему рівнянь

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i; \\ a\sum_{i=1}^{n} x_i + b \cdot n = \sum_{i=1}^{n} y_i. \end{cases}$$

що називають системою нормальних рівнянь у разі вибору емпіричної функції у вигляді лінійної залежності.

Приклад. Методом найменших квадратів знайти значення параметрів емпіричної функції, якщо дослідницькі дані про значення x та y репрезентовані в таблиці

x	6	8	9	9	10	11	12	13	14	15
y	4	4	5	7	5	6	8	7	9	10

Розв'язання. За вибіркою спостережень побудуємо в системі координат *x0y* діаграму розсіювання:

Як видно з діаграми, за емпіричну функцію можна взяти лінійну функцію

$$\overline{y} = ax + b\,.$$

Для знаходження параметрів a та b застосуємо МНК.

Тоді для визначення параметрів a та b будемо мати систему нормальних рівнянь

$$\begin{cases} a\sum_{i=1}^{n} x_i^{\,2} + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i; \\ a\sum_{i=1}^{n} x_i + b\cdot n = \sum_{i=1}^{n} y_i. \end{cases}$$

Формуючи систему рівнянь, маємо

$$\begin{cases} 1217a + 107b = 743; \\ 107a + \ \ 10b = 65. \end{cases}$$

Розв'язавши систему, одержимо

$$a = 0{,}659;\ \ b = -0{,}55\,.$$

Підставляючи ці значення параметрів, одержимо емпіричну функцію

$$\overline{y} = 0{,}659x - 0{,}55\,,$$

яка описує залежність між випадковими величинами x та y.

7.5 Основні характеристики зв'язку

Залежність величини y або x називають функціональною, якщо кожному значенню x відповідає єдине значення величини y. Якщо x – детермінована величина, тобто та, що набуває цілком певного значення, то й y є детермінована, якщо ж x – випадкова величина, то й y також випадкова величина.

Однак частіше має місце не функціональна, а імовірнісна залежність, коли кожному фіксованому значенню незалежної змінної x відповідає не одне, а множина значень змінних y, причому сказати заздалегідь, якого саме значення набуде величина y не можна.

Припустимо, що існує імовірнісна залежність випадкової змінної y від x. Зафіксуємо деяке значення x змінної X. У разі $X = x$ змінна y може набути будь-якого значення з деякої множини, причому яке саме, заздалегідь невідомо. Середнє цієї множини називають груповим генеральним середнім змінної Y у разі $X = x$ або математичним сподіванням випадкової величини Y, обчисленої за умови, що $X = x$, це умовне математичне сподівання позначають як $M(Y\,/\,X = x)$. Якщо існує стохастична залежність Y від X, то насамперед намагаються з'ясувати, змінюються чи ні під час зміни X умовні математичні сподівання $M(Y\,/\,X = x)$.

Статистичною називають залежність, за якої зміна однієї з величин спричиняє зміну розподілення іншої. Зокрема, статистична залежність проявляється в тому, що під час зміни однієї з величин змінюється середнє значення іншої; у цьому разі статистичну залежність називають кореляційною. Говорять, що має місце кореляційна залежність Y від X, якщо під час зміни X змінюються умовні математичні сподівання $M(Y\,/\,X = x)$, якщо ж умовні математичні сподівання залишаються незмінними, то говорять, що кореляційна залежність Y від X відсутня.

Функцію $\varphi(x) = M(Y / X = x)$, що описує зміну умовного математичного сподівання випадкової змінної Y у разі зміни значень x змінної X, називаються функцією регресії.

Умовним середнім $\overline{y}_x$ називають середнє арифметичне значень y, що спостерігалися у разі відповідних $X = x$.

Умовним середнім $\overline{x}_y$ називають середнє арифметичне значень x, що спостерігалися за відповідних $Y = y$.

За умови великої кількості спостережень одне й те саме значення x може трапитися n_x разів, одне й те саме значення y – n_y разів, одна й та сама пара чисел $(x; y)$ може трапитися n_{xy} раз. Тому дані спостережень групують, підраховуючи частоти n_x, n_y, n_{xy}. Усі згруповані дані записують у вигляді таблиці, що називають кореляційною:

$Y \backslash X$	x_1	x_2	…	x_1	$\sum$
y_1	n_{11}	n_{21}	…	n_{l1}	n_1
y_2	n_{12}	n_{22}	…	n_{l2}	n_2
…	…	…	…	…	…
y_k	n_{1k}	n_{2k}	…	n_{lk}	n_k
$\sum$	m_1	m_2	…	m_l	n

Якщо обидві лінії регресії Y на X та X на Y – прямі, то кореляцію називають лінійною. Для оцінювання сили лінійного кореляційного зв'язку служить вибірковий коефіцієнт кореляції $r_в$. Вибірковий коефіцієнт кореляції визначають рівністю

$$r_в = \frac{\sum n_{xy} \cdot x \cdot y - n \cdot \overline{x} \cdot \overline{y}}{n \cdot \sigma_x \cdot \sigma_y},$$

де x, y – варіанти ознак X та Y; n_{xy} – частота пари (x, y); n – об'єм вибірки; σ_x, σ_y – вибіркові середні квадратичні відхилення; $\overline{x}$, $\overline{y}$ – вибіркові середні ознак X та Y.

Нехай $x_{i+1} - x_i = h_1$, $y_{i+1} - y_i = h_2$ – кроки таблиці (варіанти рівновіддалені). Якщо перейти до умовних варінт

$$u_i = \frac{x_i - c_1}{h_1}, \quad v_i = \frac{x_i - c_2}{h_2},$$

де c_1, c_2 – умовні нулі, то вибірковий коефіцієнт кореляції обчислюють за формулою

$$r_в = \frac{\sum n_{uv} \cdot u \cdot v - n \cdot \overline{u} \cdot \overline{v}}{n \cdot \sigma_u \cdot \sigma_v}.$$

Рівняння прямих регресії Y на X та X на Y мають вигляд

$$\overline{y_x} - \overline{y} = r_в \frac{\sigma_y}{\sigma_x}(x - \overline{x}), \qquad \overline{x_y} - \overline{x} = r_в \frac{\sigma_x}{\sigma_y}(y - \overline{y}).$$

де $\overline{x_y}, \overline{y_x}$ – умовні середні, тобто середні значення однієї змінної, що відповідають певному значенню іншої.

Розділ 8. Статистична перевірка гіпотез

На практиці часто необхідно знати закон розподілення генеральної сукупності. Якщо він невідомий, але є припущення його певного вигляду, тоді висувають гіпотезу, що випадкова величина має це розподілення і виникає потреба в перевірці гіпотези. Розглянемо одну з основних задач математичної статистики – перевірку правдоподібності гіпотез. Для цього потрібно використовувати так звані критерії узгодження, серед яких зупинимося на критеріях Пірсона та Романовського.

Статистичною гіпотезою називають будь-яке твердження про вигляд або властивості розподілення випадкових величин, що спостерігаються в експерименті.

Нульовою (основною) називають висунуту гіпотезу H_0.

Конкуруючою (альтернативною) називають гіпотезу H_1, що суперечить нульовій.

Простою називають гіпотезу, що містить лише одне твердження.

Складною називають гіпотезу, що складається зі скінченної або нескінченної кількості простих.

Висунута статистична гіпотеза підлягає перевірці, що здійснюють за даними вибірки.

Помилка першого роду полягає в прийнятті альтернативи (відхилення нульової гіпотези) в умовах, коли справджується нульова гіпотеза.

Помилка другого роду полягає в прийнятті нульової гіпотези (відхилення альтернативи) в умовах, коли справджується альтернативна гіпотеза.

Для перевірки нульової гіпотези вибирають деяку випадкову величину *К*, розподілення якої відоме, і її називають статистичним критерієм перевірки нульової гіпотези.

Критичною областю називають множину значень критерію, за яких нульову гіпотезу відхиляють. Областю прийняття гіпотези називають множину значень критерію, за яких нульову гіпотезу приймають.

Критичними точками називають точки, що відокремлюють критичну область від області прийняття гіпотези.

Розрізняють однобічну (правобічну та лівобічну) та двобічну критичні області. Правобічною називають критичну область, що визначається нерівністю $K > k_{кр}$, де $k_{кр}$ – додатне число. Лівобічною називають критичну область, що визначається нерівністю $K < k_{кр}$, де $k_{кр}$– від'ємне число.

Щоб знайти однобічну критичну область, треба знайти критичну точку $k_{кр}$. Для цього задають достатньо малу ймовірність – рівень значущості α, а потім шукають критичну точку з урахуванням вимоги

$$P(K > k_{кр}) = \alpha$$

у разі правобічної критичної області, або

$$P(K < k_{кр}) = \alpha$$

у разі лівобічної критичної області.

У випадку двобічної критичної області повинна виконуватись тотожність

$$P(K < k_1) + P(K > k_2) = \alpha .$$

Для кожного критерію узгодження є відповідні таблиці (наприклад, таблиці додатка), що дозволяють знайти таку точку $k_{кр}$, яка задовольняє потрібну умову.

Для перевірки правильності основної статистичної гіпотези H_0 потрібно:

1) визначити гіпотезу H_1 альтернативну до гіпотези H_0;

2) обрати статистичну характеристику перевірки;

3) визначити допустиму ймовірність похибки першого роду, тобто рівень значущості α;

4) знайти за відповідною таблицею критичну область (критичну точку) для обраної статистичної характеристики.

До критичної області належать такі значення статистичної характеристики, за яких гіпотеза H_0 відхиляється на користь альтернативної гіпотези.

8.1 Критерій узгодження Пірсона про вигляд розподілення

Критерій узгодження Пірсона (χ-квадрат) ефективно використовують для перевірки гіпотези про розподілення генеральної сукупності, тобто що розподілення випадкової величини має певний функціональний вираз. Обмежимось застосуванням цього критерію для перевірки гіпотези про нормальне розподілення генеральної сукупності. Нехай вибірка має таке розподілення об'єму n

варіанти x_k	x_1	x_2	…	x_m
частоти n_k	n_1	n_2	…	n_m

або

варіанти l_k	(x_1, x_2)	(x_2, x_3)	…	(x_{m-1}, x_m)
частоти n_k	n_1	n_2	…	n_m

Потрібно з рівнем значущості α перевірити основну гіпотезу H_0: генеральна сукупність розподілена нормально. Критерієм перевірки цієї гіпотези беруть випадкову величину χ^2, що в різних випробуваннях набуває різних, наперед невідомих значень Критичне значення цієї випадкової величини залежить від рівня значущості α та степенів вільності її розподілення k

$$\chi^2_{kp} = \chi^2(\alpha, k).$$

Для розподілення генеральної сукупності за нормальним законом степінь вільності буде $k = m - 3$, де m – кількість варіантів вибірки або часткових інтервалів варіант.

Правило Пірсона. Щоб під час заданого рівня значущості α перевірити основну гіпотезу H_0, що генеральна сукупність розподілена нормально, необхідно:

1) обчислити теоретичні частоти n'_k для варіант вибірки;

2) обчислити спостережне значення критерія χ^2 за формулою

$$\chi^2_{cn} = \sum_{k=1}^{m} \frac{(n_k - n'_k)^2}{n'_k};$$

3) знайти степінь вільності $k = m - 3$;

4) знайти з таблиці значень критичних точок розподілення Пірсона (таблиця А.5) критичну точку χ^2_{kp}, що відповідає заданому рівню значущості α та степені вільності k;

5) порівняти χ^2_{cn} та χ^2_{kp} і зробити висновок:

якщо $\chi^2_{cn} < \chi^2_{kp}$, то гіпотезу H_0 треба прийняти;

якщо $\chi^2_{cn} > \chi^2_{kp}$, то гіпотезу H_0 треба відхилити.

8.2 Перевірка гіпотези про рівність математичних очікувань (про рівність генеральних середніх) двох нормально розподілених генеральних сукупностей у разі великих вибірок або відомих дисперсій

Нехай є дві нормально розподілені генеральні сукупності, причому в першій сукупності досліджувана ознака *X*, у другій сукупності досліджувана ознака *Y*, тобто кожна ознака має нормальний закон розподілення зі своїми параметрами.

Ми надалі будемо розглядати ситуації, що стосуються великих вибірок із цих двох генеральних сукупностей: $n > 30,\ \ m > 30$. Випадки малих вибірок аналізуються у відповідних розділах підручників; тут такі ситуації не розглядають.

Постановка задачі:

$$H_0 : M(X) = M(Y);$$
$$H_1 : M(X) \neq M(Y);\ \ \text{або}\ \ M(X) > M(Y),\ M(X) < M(Y).$$

Оскільки тут розглянутий приклад великих вибірок (або приклад, коли дисперсії відомі), то можна припустити, що невідомі дисперсії генеральних сукупностей досить добре апроксимуються вибірковими дисперсіями, тобто можна вважати, що

$$\sigma_X^2 \approx S_X^2;\quad \sigma_Y^2 \approx S_Y^2;$$

де

$$S_X^2 = \frac{\sum_{i=1}^{n_1}\left(x_i - \overline{x}\right)^2}{n-1};\quad S_Y^2 = \frac{\sum_{i=1}^{n_2}\left(y_i - \overline{y}\right)^2}{m-1}$$

Для вирішення завдання використовують критерій

$$K = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{S_X^{\ 2}}{n} + \frac{S_Y^{\ 2}}{m}}}.$$

Ця випадкова величина в разі справедливості нульової гіпотези наближено має стандартний нормальний закон розподілення. Розв’яжемо конкретну задачу, у якій реалізується описаний вище підхід.

Приклад.

Проводиться порівняння зростання 20-річних юнаків, які проживають у Сумах та Харкові. На основі двох випадкових вибірок, виконаних у двох містах, були одержані певні дані. У Сумах відібрали 75 юнаків. Було виміряно зростання кожного юнака. На основі вимірів були обчислені дві характеристики: середнє зростання юнаків, що дорівнювало 179 см, і стандартне відхилення, яке дорівнювало 8 см. У Харкові були випадково відібрані 57 юнаків, їх середній зріст дорівнював 176 см зі стандартним відхиленням 10 см. На основі цих експериментальних даних необхідно перевірити гіпотезу про рівність середніх величин зростання 20-річних юнаків (про рівність генеральних середніх) у цих містах. Прийняти довірчу ймовірність дорівнює 90 %. Передбачається, що зростання юнаків підпорядковується нормальному закону розподілення.

Розв’язання.

Потрібно з’ясувати, чи значуще відрізняються один від одного вибіркові середні значення зросту юнаків. Якщо буде показано, що вибіркові середні відрізняються незначно, то звідси можна буде зробити висновок про справедливість нульової гіпотези, тобто висновок про приблизний рівність генеральних середніх значень зростання юнаків, які проживають у різних містах. В іншому разі буде зроблено висновок про істотне розходження генеральних середніх значень зростання юнаків, які проживають у цих містах.

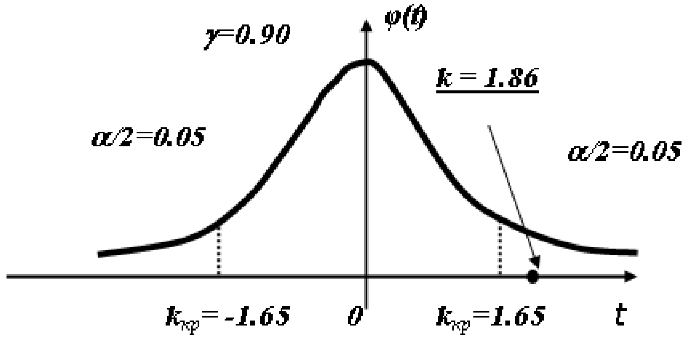
У разі такої постановки завдання потрібно будувати двосторонню критичну область. Визначимо її межі на основі табличного розв'язання рівняння

$$\Phi(k_{кр}) = \frac{\gamma}{2} \rightarrow \Phi(k_{кр}) = \frac{0.90}{2} = 0.45 \rightarrow k_{кр} \approx 1.65.$$

Обчислимо на основі експериментальної інформації спостережуване значення критерію, що в разі справедливості нульової гіпотези наближено має стандартний нормальний закон розподілення

$$K = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} = \frac{179 - 176}{\sqrt{\frac{8^2}{75} + \frac{10^2}{57}}} = 1.86\ .$$

Зобразимо результати графічно на графіку густини стандартного нормального закону розподілу:



Бачимо, що значення критерію потрапило в критичну область значень параметра. Це означає, що потрібно відкинути основну гіпотезу на користь альтернативної

гіпотези й сказати, що середній зріст сумських та харківських 20-річних юнаків відрізняється значуще.

8.3 Перевірка гіпотези щодо рівності дисперсій двох нормально розподілених генеральних сукупностей

Нехай є дві випадкові величини X та Y, що мають нормальне розподілення. Сукупність значень кожної з випадкових величин будемо розглядати як генеральну. Отже, маємо дві нормально розподілених генеральних сукупності X та Y, із яких вилучені дві незалежні вибіркові сукупності обсягами n і m. Для кожної з вибіркових сукупностей обчислили виправлені дисперсії S_X^2 і S_Y^2. Необхідно в разі заданого рівня значущості α перевірити нульову гіпотезу про рівність дисперсій цих випадкових величин, тобто $H_0 : D(X) = D(Y)$. Як статистичний критерій для перевірки цієї гіпотези розглядають критерій Фішера, тобто випадкову величину F, яку визначають як відношення більшої дисперсії до меншої

$$F = \frac{S_б^2}{S_м^2},$$

де $S_б^2$ – більша з виправлених дисперсій; $S_м^2$ – менша з виправлених дисперсій.

Випадкова величина F розподілена за статистикою Фішера. Якщо альтернативною є гіпотеза $H_1 : D(X) > D(Y)$, то критична область є односторонньою, і її межу (критичну точку) визначають співвідношенням

$$P\left(F > F_\alpha(k_1; k_2)\right) = \alpha\ ,$$

тобто за умови, що основна гіпотеза не є хибною, ймовірність того, що значення випадкової величини F

належатиме критичні області, дорівнює прийнятому заздалегідь рівню значущості α. Величину $F_\alpha(k_1;k_2)$ визначають за статистикою Фішера в разі заданого рівня значущості α та кількості степенів свободи k_1 більшої з дисперсій, а також кількості степенів свободи k_2 меншої з дисперсій.

Якщо $F < F_\alpha(k_1;k_2)$, то з надійністю $\gamma = 1-\alpha$ нульову гіпотезу нема підстав відхилити. Якщо $F > F_\alpha(k_1;k_2)$, то нульова гіпотеза відхиляється.

Якщо альтернативною є гіпотеза $H_1: D(X) \neq D(Y)$, то критична область є двосторонньою, тоді права критична точка, із якою порівнюють емпіричне значення критерію, визначають співвідношенням

$$P\left(F > F_\alpha(k_1;k_2)\right) = 0,5 \cdot \alpha .$$

Приклад.

За двома методиками проведені вимірювання однієї й тієї самої фізичної величини. Одержані такі результати:

а) для першої методики:

$x_1 = 9,6;\ x_2 = 10,0;\ x_3 = 9,8;\ x_4 = 10,2;\ x_5 = 10,6$;

б) для другої методики:

$y_1 = 10,4;\ y_2 = 9,7;\ y_3 = 10,0;\ y_4 = 10,3$.

Перевірити, чи можна вважати, що обидві методики забезпечують однакову точність вимірювань, якщо прийняти рівень значущості $\alpha = 0,1$?

Розв'язання.

Вважатимемо, що результати вимірювань розподілені нормально й вибірки є незалежними. Оскільки точність вимірювання визначається величиною дисперсій, то задача полягає в перевірці нульової гіпотези $H_0: D(X) = D(Y)$ у разі конкуруючої гіпотези $H_1: D(X) \neq D(Y)$. Розрахунки

основних числових характеристик випадкових величин X та Y дали такі результати: $\overline{x} = 10,04$ і $S_X^2 = 0,148$, $\overline{y} = 10,1$ і $S_Y^2 = 0,1$. Отже,

$$F = S_X^2 \,/\, S_Y^2 = 0,148 / 0,1 = 1,48$$

Конкуруючою є гіпотеза $H_1: D(X) \neq D(Y)$, тобто критична область двостороння, тому критичну точку розподілення Фішера визначаємо для $0,5 \cdot \alpha = 0,05$. Кількість ступенів вільності становить $k_1 = n-1 = 5-1 = 4$ та $k_2 = m-1 = 4-1 = 3$. Відповідно $F_{0,05}(4;\,3) = 9,12$. Оскільки $F < F_{0,05}$, то гіпотезу про рівність генеральних дисперсій немає підстав відхилити. Іншими словами, виправлені дисперсії розрізняються неістотно, отже, обидві методики забезпечують однакову точність вимірювань.

8.4 Перевірка гіпотези щодо рівності двох середніх нормальних генеральних сукупностей за невідомих дисперсій

Нехай є дві вибіркові сукупності обсягами n і m $(n < 30,\ m < 30)$, що містять значення випадкових величин X та Y, дисперсії яких невідомі. Необхідно при заданому рівні значущості α перевірити нульову гіпотезу про рівність математичних сподівань цих випадкових величин, тобто $H_0: M(X) = M(Y)$. В якості альтернативної вибирають гіпотезу $H_1: M(X) \neq M(Y)$.

Розглянемо приклад, коли дисперсії невідомі та однакові (малі незалежні вибірки). Визначимо вибіркові середні $\overline{x}$ і $\overline{y}$ та виправлені вибіркові дисперсії S_X^2 і S_Y^2

цих випадкових величин. Їх генеральні дисперсії хоча й невідомі, але вважаються однаковими.

Перевірку основної гіпотези здійснюють за критерієм Стьюдента. Спостережуване значення критерію обчислюють співвідношенням

$$T = \frac{\overline{x} - \overline{y}}{\sqrt{(n-1)S_X^2 + (m-1)S_Y^2}} \cdot \sqrt{\frac{mn \cdot (m+n-2)}{m+n}}.$$

Його порівнюють із критичною точкою розподілення Стьюдента для заданого рівня значущості α та кількості ступенів свободи $k = n + m - 2$. Згідно з альтернативною гіпотезою критична область є двосторонньою. Отже, за таблицею визначаємо значення $t_\alpha(k)$. Якщо $|T| < t_\alpha(k)$, то нульову гіпотезу немає підстав відхилити, отже, за даним рівнем значущості $M(X) = M(Y)$; якщо $|T| > t_\alpha(k)$, то нульову гіпотезу відхиляють.

Приклад.

За двома незалежними вибірками обсягами $n = 12$ і $m = 18$, вилучених із нормально розподілених генеральних сукупностей, знайдені вибіркові середні: $\overline{x} = 31,2$, $\overline{y} = 29,2$ та виправлені дисперсії: $S_X^2 = 0,84$ і $S_Y^2 = 0,40$. За умови рівня значущості $\alpha = 0,05$ перевірити нульову гіпотезу $H_0 : M(X) = M(Y)$ при конкуруючій гіпотезі $H_1 : M(X) \neq M(Y)$.

Розв'язання.

Застосування критерію Стьюдента передбачає, що обидві генеральні сукупності мають однакові дисперсії, отже, спочатку за критерієм Фішера необхідно перевірити гіпотезу про рівність генеральних дисперсій $H_0 : D(X) = D(Y)$. Оскільки величина S_X^2 значно більша за

S_Y^2, то конкуруючою вважатимемо гіпотезу $H_1 : D(X) > D(Y)$, тому критична область є правобічною. У разі рівня значущості $\alpha = 0,05$ і кількості ступенів вільності $k_1 = n - 1 = 12 - 1 = 11$ і $k_2 = m - 1 = 18 - 1 = 17$ знаходимо критичну точку $F_{0,05}(11;17) = 2,41$. Оскільки $F < F_{0,05}$, то нульову гіпотезу про рівність генеральних дисперсій немає підстав відхилити.

Оскільки припущення про рівність генеральних дисперсій виконується, то можна порівнювати середні за критерієм Стьюдента. Обчислимо спостережуване значення критерію

$$T = \frac{31,2 - 29,2}{\sqrt{(12-1) \cdot 0,84 + (18-1) \cdot 0,4}} \sqrt{\frac{12 \cdot 18 \cdot (12+18-2)}{12+18}} = 7,1.$$

За умовою альтернативна гіпотеза має вигляд $H_1 : M(X) \neq M(Y)$, тому критична область є двосторонньою. За рівнем значущості 0,05 і кількістю ступенів свободи $k = n + m - 2 = 12 + 18 - 2 = 28$ знаходимо точку $t_{0,05}(28) = 2,05$. Оскільки $T > t_{0,05}$, то нульову гіпотезу про рівність середніх потрібно відхилити, оскільки розбіжність між вибірковими середніми є статистично значущою.

8.5 Приклад статистичного дослідження вибірки

Нехай на основі деяких вимірювань одержана така вибірка: 2,01 3,89 2,95 1,54 3,42 2,48 2,95 3,42 2,95 3,89 2,48 2,01 4,83 3,42 2,48 3,42 3,42 2,48 3,89 2,48 2,01 4,36 1,54 2,95 2,95 3,42 2,95 2,95 3,89 3,42 2,95 2,48 3,42 2,48 3,42 2,95 2,95 3,89 3,89 2,01 3,42 2,48 2,48 2,95 3,42 2,95 2,48 4,83 2,95 3,42 2,48 2,48 1,54 2,48 2,95 2,95 3,42 3,42 4,36

2,95 2,01 3,42 2,95 3,42 3,42 2,95 2,95 2,95 3,42 3,42 2,48 4,36 3,89 3,42 3,42 3,42 1,54 2,95 2,48 2,48 2,95 2,48 2,95 2,95 3,89 2,95 3,42 2,95 1,54 2,48 2,48 2,48 2,95 3,42 2,95 3,42 2,01 2,01 2,48 2,95

1) Побудуємо варіаційний ряд. Розташувавши числа вибірки в порядку зростання та вказавши їх частоти, складемо варіаційний ряд (табл. 8.1).

Таблиця 8.1 – Варіаційний ряд.

1,54	2,01	2,48	2,95	3,42	3,89	4,36	4,83
5	7	21	29	25	8	3	2

2) Обчислимо розмах вибірки $x_{\max} - x_{\min} = 4{,}83 - 1{,}54 = 3{,}29$. Розрахуємо моду та медіану.

Оскільки медіана – це варіанта, що ділить вибірку на дві рівні за об'ємом частини й, зважаючи на те, що в даному прикладі об'єм вибірки – число парне, маємо

$$Me = (X_{50} + X_{51}) / 2 = (2{,}95 + 2{,}95) / 2 = 2{,}95.$$

Мода в цьому прикладі: $Mo = 2{,}95$.

3) Побудуємо полігон частот та гістограму. Як зазначалося, полігон частот – це ламана, що з'єднує точки з координатами (x_i, n_i) (рис. 8.1).

Для побудови гістограми потрібно вибрати на осі $0x$ основи прямокутників. У разі варіаційного ряду з рівновіддаленими варіантами x_i, зручно розбивку $[\alpha_i ; \beta_i]$ вибрати за таким правилом:

$$\alpha_i + \beta_i = 2x_i, \quad \alpha_{i+1} = \beta_i , \quad i = 1, \ldots, k.$$

Під час побудови гістограми інтервали, що відповідають малим частотам n_i, називають непоказними та об'єднують із сусідніми. Якщо сусідніх два, то вибирають той, у якого частота n_i вища. Рівень

зображуваності, взагалі кажучи, величина інтуїтивна. Приймемо, що об'єднанню підлягають інтервали, у яких $n_i < 5$. Інколи навіть після об'єднання двох інтервалів утворюється непоказний інтервал, тоді процес об'єднання продовжується на наступний сусідній інтервал.

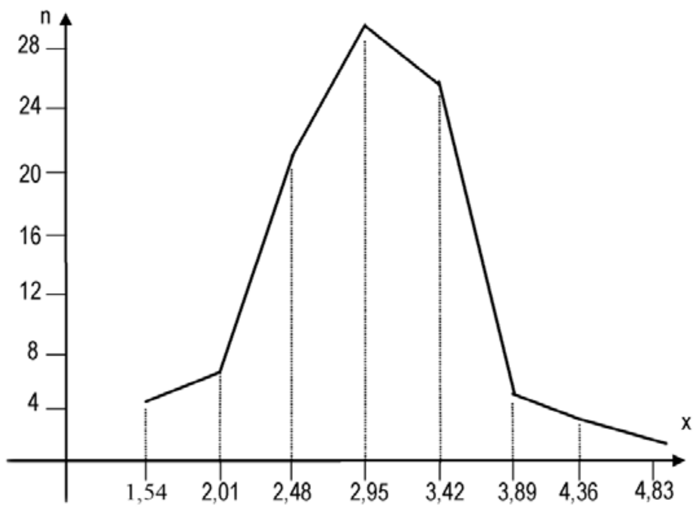


Рисунок 8.1 – Полігон частот

Об'єднання двох інтервалів описують співвідношенням

$$[\alpha_j ; \beta_j] \cup [\alpha_{j+1} ; \beta_{j+1}] \to [\alpha_j ; \beta_{j+1}].$$

Після об'єднання інтервалів виконується їх перенумерація, щоб індекс «i» набував всіх значень підряд без пропусків. У результаті зміни інтервалів варіаційний ряд дещо видозмінюється: частоти стосуються не чисел, а до інтервалів – показують, скільки елементів із вибірки попадає в цей інтервал. Запишемо інтервальний варіаційний ряд у таблицю 8.2 (стовпці «i», «інтервал», «δ_i» – довжина інтервалу, $\delta_i = \beta_i - \alpha_i$).

Для обчислення висот прямокутників треба вибрати масштаб, це можна зробити способом,описаним нижче. Прямокутник найбільшої висоти буде відповідати моді варіаційного ряду. Тому, для того щоб рисунок гістограми ефективно використовував площу рисунка висотою H, масштабний множник μ повинен задовольняти співвідношення

$$S = H \cdot \delta = \mu \cdot n_M\,, \quad \mu = \frac{H \cdot \delta}{n_M},$$

де δ і S – довжина основи та площа прямокутника, що відповідає моді варіаційного ряду, n_M – частота моди (найбільша частота у варіаційному ряді).

Тепер висота будь-якого прямокутника гістограми потрібно обчислювати за формулою

$$h_i = \mu \frac{n_i}{\delta_i}.$$

У цьому прикладі, узявши висоту 70 мм, одержимо (рис. 8.2)

$$\mu = \frac{70 \cdot 0.47}{29} = 1{,}13.$$

Вісь $0x$ розмітимо так, щоб і ширина рисунка використовувалась ефективно. Для цього крайні відмітки 1,305 і 5,065 розташуємо біля лівої та правої меж рисунка**.** Результати розрахунків занесемо в стовпець «h» таблиці 8.2.

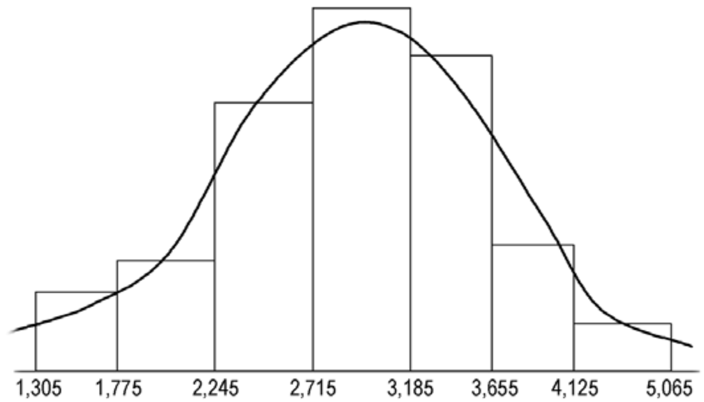


Рисунок 8.2 – Гістограма

4) Обчислимо характеристики вибірки:

$$n = 5 + 7 + 21 + 29 + 25 + 8 + 5 = 100,$$

$$\bar{x} = \frac{5 \cdot 1{,}54 + 7 \cdot 2{,}01 + 21 \cdot 2{,}48 + 29 \cdot 2{,}95 + 25 \cdot 3{,}42}{100} +$$

$$+ \frac{8 \cdot 3{,}89 + 3 \cdot 4{,}36 + 2 \cdot 4{,}83}{100} = 2{,}988,$$

$$D = \frac{5 \cdot 1{,}54^2 + 7 \cdot 2{,}01^2 + 21 \cdot 2{,}48^2 + 29 \cdot 2{,}95^2 + 25 \cdot 3{,}42^2}{100} +$$

$$+ \frac{8 \cdot 3{,}89^2 + 3 \cdot 4{,}36^2 + 2 \cdot 4{,}83^2}{100} - 2{,}988^2 = 0{,}4624,$$

$$s^2 = \frac{n}{n-1} D = \frac{100}{99} \cdot 0{,}4624 = 0{,}4670,$$

$$s = \sqrt{0{,}4670} = 0{,}6833.$$

Таблиця 8.2 Характеристики вибірки

i	***інтерв ал***	n_i	δ_i	h_i	***x***	***Ф***	p_i	n'_i
1	1,305 1,775	5	0,47	10	–2,463 –1,775	–0,4931 –0,4620	0,0311	3,11
2	1,775 2,245	7	0,47	14	–1,775 –1,087	–0,4620 –0,3614	0,1006	10,06
3	2,245 2,715	21	0,47	42	–1,087 –0,400	–0,3614 –0,1554	0,2060	20,60
4	2,715 3,185	29	0,47	58	–0,400 0,288	–0,1554 0,1130	0,2684	26,84
5	3,185 3,655	25	0,47	50	0,288 0,976	0,1330 0,3355	0,2225	22,25
6	3,655 4,125	8	0,47	16	0,976 1,664	0,3355 0,4519	0,1164	11,64
7	4,125 5,065	5	0,94	5	1,664 3,040	0,4519 0,4988	0,0469	4,69

5) З огляду на вигляд з вигляду гістограми (близькі до центра рисунка прямокутники високі, а до периферії знижуються), висунемо гіпотезу про нормальний закон розподілу генеральної сукупності. За параметри нормального закону візьмемо:

$$a = \bar{x} = 2{,}988, \quad \sigma = s = 0{,}6833.$$

Заповнимо стовпець «*x*» таблиці 8.2, у який заносять аргументи інтеграла ймовірностей

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp(-\frac{t^2}{2})dt ,$$

на «*i*»-ому інтервалі $x = \frac{r-a}{s}$, де r – число зі стовпця «інтервал».

З стовпця «*x*» таблиці 8.2 можна обчислити значення інтеграла ймовірностей, використовуючи Додаток А

(табл. А.2). Треба мати на увазі, що, обчислюючи ймовірності $P = \Phi(x'') - \Phi(x')$, ми віднімаємо близькі між собою числа, а це, як відомо, приводить до значного зросту відносної похибки [9]. Тому значення функції *Ф(x)* треба одержувати з максимальною точністю. Аргумент «*x*» у Додатку А заданий з двома знаками після коми, отже, якщо в стовпці «*x*» таблиці 8.2 залишати три знаки після коми, то за допомогою лінійної інтерполяції можна знайти потрібні значення *Ф*(*x*) достатньо точно.

Приклад лінійної інтерполяції. Нехай треба знайти *Ф*(1,577). У Додатку А (табл. А.2) знаходимо *Ф*(1,57) = =0,4418, *Ф*(1,58) = 0,4429. Можна обчислити різницю 0,4429 – 0,4418 = 0,0011. Отже, на одну соту частину аргументу припадає 0,0011, а на одну тисячну 0,00011. Помножимо на однозначне число 7 ($7 \cdot 0{,}00011 = 0{,}00077$) та округлимо до чотирьох знаків після коми – 0,0008. Отже,

$$\Phi(1{,}577) = 0{,}4418 + 0{,}0008 = 0{,}4426.$$

Заповнимо стовпець *Ф*(*x*) таблиці 8.2. Тепер можна обчислити теоретичні ймовірності попадання в інтервали:

$$P_i = \Phi(x''_i) - \Phi(x'_i)$$

(від нижнього числа клітини стовпця *Ф*(*x*) віднімаємо верхнє число тієї самої клітини). Нарешті, останній стовпець таблиці 8.2 – теоретичні частоти:

$$n'_i = P_i \cdot n ,$$

де n – об'єм вибірки.

Зауважимо, що, на відміну від фактичних, теоретичні частоти не округляють до цілого значення.

6) За стовпцем «n_i'» на рисунок гістограми наносимо точки теоретичної кривої за правилом:

- абсциса «i-ї» точки – це середина «i-го» інтервалу,
- ордината «i-ї» точки = $\mu \cdot n_i' / \Delta_i$.

З'єднуємо точки плавною лінією (вершина кривої не зобов'язана співпадати з якою-небудь точкою), вона повинна бути симетричною відносно прямої $x = \overline{x}$.

7) Обчислимо суму Пірсона:

$$\chi^2_{спост} = \frac{(5-3,11)^2}{3,11} + \frac{(7-10,06)^2}{10,06} + \frac{(21-20,06)^2}{20,06} + \frac{(29-26,84)^2}{26,84} +$$

$$+ \frac{(25-22,25)^2}{22,25} + \frac{(8-11,64)^2}{11,64} + \frac{(5-4,69)^2}{4,69} = 3,759 ;$$

8) Для обчислення кількості ступенів вільності треба від числа інтервалів відняти кількість зв'язків, що накладає підсумовування. У разі нормального закону зв'язків 3, це підсумовування під час обчислення $\overline{x}$, D, χ^2 . В даному прикладі ступенів вільності 7 – 3 = 4. Нехай в даному прикладі заданий рівень значущості α = 0,05. Використовуючи таблицю А.5, знайдемо критичне значення розподілення χ^2: $\chi^2_{крит} = 9,5$, отже $\chi^2_{спост} < \chi^2_{крит}$, тому немає підстав відкинути гіпотезу про нормальне розподілення.

к) Визначимо ймовірність P_{ab} попадання в інтервал [2,48; 3,65] двома способами – за відносною частотою та за теоретичною функцією розподілення.

За відносною частотою

$$P_{ab} = \frac{0,5 \cdot 21 + 29 + 25}{100} = 0,645$$

(у чисельнику враховуємо частоти тих варіант варіаційного ряду, що належать інтервалу [a; b], для точок, які знаходяться на межі частота зменшується вдвічі).

За теоретичним розподілення:

$$P_{ab} = P(a \le X \le b) = \Phi(x'') - \Phi(x'),$$

$$x'' = \frac{3,65 - 2,988}{0,6833} = 0,969 ,$$

$$x' = \frac{2,48 - 2,988}{0,6833} = -0,743 ,$$

$P_{ab} = \Phi$ (0,969) – Φ (–0,743) = 0,3338 + 0,2731 = 0,6069.

10) Знайдемо інтервали довіри для параметрів нормального розподілення; нехай їх треба обчислити з надійністю γ = 0,95. У таблиці А.3 для n = 100 і $\gamma = 0,95$ знаходимо $t_\gamma = 1,984$ та обчислюємо відхилення:

$$\delta = \frac{1,984 \cdot 0,6833}{\sqrt{100}} = 0,136 ,$$

$$a - \delta = 2,852,$$

$$a + \delta = 3,124.$$

Із надійністю γ = 0,95 одержимо 2,852 < a < 3,124.

Додаток А (табл. А.4) дозволяє знайти q_γ = 0,143. Тому з надійністю γ = 0,95 отримаємо інтервал довіри

$$\sigma\,(1\text{-}q) < \sigma < (1{+}q),$$
$$0{,}586 < \sigma < 0{,}781.$$

Якщо задана двовимірна вибірка, то її дослідження виконують за схемою, наведеною в наступному прикладі.

Приклад. Нехай заданий початковий масив (табл. 8.3)

Таблиця 8.3 Початковий масив

x	57	25	51	29	43	32	61	18	45	21	36	61	28	14	53	16	14	44	32	60
y	47	56	37	59	52	51	49	74	40	66	65	27	72	79	56	82	70	52	61	33

і задана надійність $\gamma = 0{,}999$. Підрахунок об'єму дає $n = 20$.

1) Знайдемо числові характеристики:

$$S_x = 57 + 25 + 51 + 29 + 43 + 32 + 61 + 18 + 45 + 21 + 36 + 61 + 28 + 14 + 53 + 16 + 14 + 44 + 32 + 60 = 740,$$
$$S_y = 47 + 56 + 37 + 59 + 52 + 51 + 49 + 74 + 40 + 66 + 65 + 27 + 72 + 79 + 56 + 82 + 70 + 52 + 61 + 33 = 1128,$$
$$S_{xx} = 57^2 + 25^2 + 51^2 + 29^2 + 43^2 + 32^2 + 61^2 + 18^2 + 45^2 + 21^2 + 36^2 + 61^2 + 28^2 + 14^2 + 53^2 + 16^2 + 14^2 + 44^2 + 32^2 + 60^2 = 32518,$$
$$S_{yy} = 47^2 + 56^2 + 37^2 + 59^2 + 52^2 + 51^2 + 49^2 + 74^2 + 40^2 + 66^2 + 65^2 + 27^2 + 72^2 + 79^2 + 56^2 + 82^2 + 70^2 + 52^2 + 61^2 + 33^2 = 1128,$$
$$S_{xy} = 57\cdot 47 + 25\cdot 56 + 51\cdot 37 + 29\cdot 59 + 43\cdot 52 + 32\cdot 51 + 61\cdot 49 + 18\cdot 74 + 45\cdot 40 + 21\cdot 66 + 36\cdot 65 + 61\cdot 27 + 28\cdot 72 + 14\cdot 790 + 53\cdot 56 + 16\cdot 82 + 14\cdot 70 + 44\cdot 52 + 32\cdot 61 + 60\cdot 33 = 37641$$

$$\overline{x} = S_x / n = 740 / 20 = 37{,}00,$$
$$\overline{y} = S_y / n = 1128 / 20 = 56{,}40,$$
$$D_x = S_{xx} / n - \left(\overline{x}\right)^2 = 32518 / 20 - 37{,}00^2 = 256{,}90,$$
$$D_y = S_{yy} / n - \left(\overline{y}\right)^2 = 67986 / 20 - 56{,}40^2 = 218{,}34.$$

2) знайдемо коефіцієнти рівняння лінійної регресії:

$$k = \frac{\frac{1}{n}\cdot S_{xy} - \overline{x}\cdot\overline{y}}{D_x} = \frac{37641/20 - 37{,}00\cdot 56{,}40}{256{,}90} = -0{,}7970;$$

$$b = \overline{y} - k\cdot\overline{x} = 56{,}40 + 0{,}797\cdot 37{,}00 = 85{,}89.$$

3) знайдемо вибірковий коефіцієнт кореляції:

$$r = k\cdot\sqrt{D_x / D_y} = -0{,}8645.$$

Можна оцінити тісноту лінійного зв'язку між факторами за шкалою

$$|r| < 0{,}6 \text{ – слабка},$$
$$0{,}6 \le |r| \le 0{,}9 \text{ – середня},$$
$$|r| > 0{,}9 \text{ – сильна}.$$

Тоді, застосовуючи одержаний результат, можна зробити висновок: між факторами «*x*» і «*y*» спостерігається лінійний зв'язок із середнім рівнем тісноти.

4) зробимо передбачення в заданих точках, вважаючи цими точками найменше, середнє й найбільше значення x.

$$y(x) = k \cdot x + b,$$
$$y(x_{\min}) = y(14,00) = -0,7970 \cdot 14 + 85,89 = 74,73,$$
$$y(x) = y(37,00) = -0,7970 \cdot 37,00 + 85,89 = 56,40,$$
$$y(x_{\max}) = y(82,00) = -0,7970 \cdot 82 + 85,89 = 20,54;$$

5) знайдемо інтервали довіри в заданих точках.

$$s_y = \sqrt{\frac{n}{n-1} D_y} = \sqrt{\frac{20}{19} \cdot 218,34} = 15,16\,.$$

Використовуючи таблицю А.3, знаходимо t_γ = 3,883. Тепер можна записати формулу для обчислення ширини інтервалу довіри для будь-якого «x» [10]:

$$\delta(x) = \frac{t_y \cdot s_y}{\sqrt{n}} \sqrt{\left(x - \overline{x}\right)^2 / D};$$
$$\frac{t_\gamma \cdot s_y}{\sqrt{n}} = \frac{3,883 \cdot 15,16}{4,472} = 13,163\,;$$
$$\delta(14) = 13,163 \cdot \sqrt{1 + (14 - 37,00)^2 / 256,9} = 23,02\,;$$
$$\delta(37,00) = 13,16\,;$$
$$\delta(82) = 13,163 \cdot \sqrt{1 + (82 - 37,00)^2 / 256,9} = 39,23\,.$$

Отже, одержимо інтервали довіри для передбачень «y» у заданих точках:

$$y_x - \delta_x \le y \le y_x + \delta_x;$$
$$x = 14:\ 51,71 \le y \le 97,75,$$
$$x = 37,00:\ 43,241 \le y \le 69,56,$$
$$x = 82:\ -18,98 \le y \le 59,48.$$

6) результати розрахунків і заданий числовий масив подамо на одному рисунку (рис. 8.3). Область довіри виявилась широкою тому, що задана дуже висока надійність γ = 0,999 (загалом усі точки не зобов'язані попадати в область довіри).

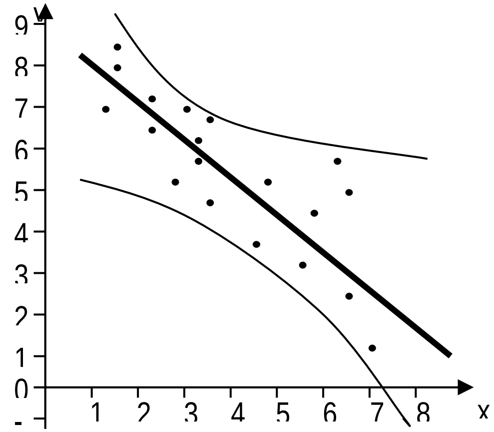


Рисунок 8.3 – Результати розрахунків і заданий числовий масив

ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ

Завдання 1 Розв'язати задачу

1.1 Серед 50 деталей 3 нестандартні. Навмання взяли 2 деталі. Знайти ймовірність того, що серед відібраних деталей одна стандартна, а інша нестандартна.

1.2 Є п'ять букв: р, е, г, й, о. Яка ймовірність того, що довільне розташування їх одна за іншою дає слово «герой»?

1.3 Серед 40 деталей 5 нестандартних. Узяті навмання 2 деталі. Знайти ймовірність того, що обидві деталі будуть нестандартними.

1.4 На клумбі 20 червоних айстр, 10 синіх і 30 білих. Яка ймовірність того, що зірвана в темряві айстра буде червоною або синьою (тобто не білою)?

1.5 З 60 питань, які входять до екзаменаційних білетів, студент вивчив 50. Яка ймовірність того, що взятий навмання студентом білет, який містить два питання, буде складатися з вивчених ним питань?

1.6 Телефонна лінія, що з'єднує районний центр із селом має довжину 12 км. Під час грози сталося пошкодження цієї лінії. Знайти ймовірність того, що пошкодження сталося на перших трьох кілометрах від районного центру.

1.7 У складальника є 10 деталей, що мало відрізняються одна від іншої. Із них 4 першого, по 2 другого, третього й четвертого різновидів. Яка ймовірність того, що з шести взятих одночасно деталей три будуть першого різновиду, дві – другого та одна – третього?

1.8 З урни, яка містить 6 пронумерованих куль, навмання витягли одну за одною всі кулі, що там знаходились. Знайти ймовірність того, що номери вийнятих куль будуть йти по черзі?

1.9 На кожній з шести однакових карток надрукована одна з таких букв: А, Т, М, Р, С, О. Картки перемішані. Знайти ймовірність того, що на чотирьох витягнутих за чергою та розкладених в один ряд картках можна буде прочитати слово «трос».

1.10 Із 10 хлопчиків і 8 дівчаток потрібно виділити для участі в туристичному поході 5 осіб. Знайти ймовірність того, що будуть виділені 2 хлопчики й 3 дівчинки (на умовах випадкового відбору).

1.11 Бібліотека складається з десяти різних книг, причому п'ять коштують по 4 грн. кожна, три книги – по 1 грн. і дві книги – по 3 грн. Знайти ймовірність того, що взяті навмання дві книги коштують 5 грн.

1.12 Куб, усі грані якого пофарбовані, розпиляний на 125 кубиків однакового розміру, які потім ретельно змішали. Знайти ймовірність того, що навмання витягнутий кубик буде мати пофарбованих граней: а) три; б) дві; в) одну.

1.13 У ящику 10 деталей, серед яких 2 нестандартні. Знайти ймовірність того, що у відібраних навмання 6 деталях опиниться не більше однієї нестандартної.

1.14 З урни, що містить 5 куль із номерами 1, 2, ..., 5 витягли навмання 4 кулі. Яка ймовірність того, що всі номери добутих куль непарні?

1.15 Набираючи номер телефону, абонент забув останні дві цифри і, лише пам'ятаючи, що ці цифри різні і не нулі, набрав їх навмання. Знайти ймовірність того, що він набрав потрібні цифри.

1.16 У майстерню для ремонту надійшло 10 годинників. Відомо, що 6 шт. з них мають потребу в загальному чищенні механізму. Майстер бере перші, які трапляться, 5 годинників. Знайти ймовірність того, що два з них потребують загального чищення механізму.

1.17 У коробці лежать 10 темних і 5 світлих краваток. Продавець навмання витяг 2 краватки. Знайти ймовірність того, що обидві краватки будуть темними.

1.18 До кінця дня в ятці залишилось 60 кавунів, із яких 50 стиглих. Покупець вибирає 2 кавуни. Яка ймовірність, що обидва кавуни стиглі?

1.19 Є два «секретних» цифрових замка, що відмикаються лише в разі певного набору цифр. Один замок має на осі шість дисків, поділених на п'ять секторів кожний, другий – п'ять дисків, поділених на шість секторів. Який замок кращий? Для якого замка ймовірність відімкнути його випадковим набором цифр менша?

1.20 Із 30 деталей, серед яких 10 вищої якості, випадковим чином вибирають на складання 20. Яка ймовірність того, що серед них опиниться 7 деталей вищої якості?

1.21 На однакових картках написані всі натуральні числа від 1 до 25 включно. Випадко беруть дві картки. Знайти ймовірність того, що на цих картках будуть написані прості числа.

1.22 Чотири білети в театр розігрують випадково чином серед п'яти юнаків і семи дівчат. Знайти ймовірність того, що білети отримають два юнаки та дві дівчини.

1.23 На складі телеательє є п'ятнадцять кінескопів, до того ж, десять із них виготовлені московським, а решта- львівським заводами. Знайти ймовірність того, що серед п'яти навмання взятих кінескопів опиняться три кінескопи, виготовлені московським заводом.

1.24 Букви, що входять до слова «ремонт», виписані кожна на окремій картці. Картки старанно перемішують, після чого виймають чотири по черзі. Яка ймовірність отримати при цьому скласти слово «море»?

1.25 У студентській групі 18 юнаків і 12 дівчат. За списком випадковим чином вибирають делегацію з двох

осіб. Визначити ймовірність того, що вибрані дівчина та юнак.

1.26 Упаковка містить 20 плиток, причому 3 мають дефекти. Контролер витягає навмання 4 плитки. Знайти ймовірність того, що упаковка буде прийнята контролером, якщо для цього потрібно, щоб він не знайшов ні однієї бракованої плитки.

1.27 Склад отримав 20 контрольно-вимірювальних приладів, але лише 12 із них відтаровані. Знайти ймовірність того, що з п'яти взятих приладів чотири відтаровані.

1.28 У фізкультурній групі 11 спортсменів, серед них 6 першорозрядників. Знайти ймовірність того, що серед 5 випадково вибраних спортсменів опиниться три першорозрядники.

1.29 Є 6 деталей першого ґатунку, 5 – другого ґатунку, 4 – третього ґатунку. Яка ймовірність того, що серед 3 випадково відібраних деталей опиняться деталі всіх ґатунків?

1.30 Із десяти білетів виграють два. Знайти ймовірність того, що серед узятих навмання п'яти білетів опиниться один виграшний.

Завдання 2 Розв'язати задачу

2.1 В електричний ланцюг послідовно ввімкнуті 3 елементи, що працюють незалежно один від одного. Ймовірності відмови 1-го, 2-го й 3-го елементів відповідно дорівнюють 0,1, 0,15 і 0,2. Знайти ймовірність того, що струму в ланцюзі не буде.

2.2 У ящику 8 деталей, серед яких 2 нестандартні. Знайти ймовірність того, що серед 5 навмання відібраних деталей опиниться не більше однієї нестандартної деталі.

2.3 Пристрій містить два незалежно працюючих елементи. Ймовірність відмови елементів відповідно дорівнюють 0,05 і 0,08. Знайти ймовірність відмови пристрою, якщо для цього достатньо, щоб відмовив хоча б один елемент.

2.4 Один з хлопчиків народився в березні, а другий – в квітні. Яка ймовірність того, що обидва вони народились в перші 10 днів місяця?

2.5 Робітник обслуговує три верстати. Ймовірність безперебійної роботи продовж однієї години після налаштування для 1-го верстата дорівнює 0,8, для 2-го – 0,9, для 3-го – 0,75. Знайти ймовірність того, що продовж години лише один верстат буде потребувати втручання робітника.

2.6 Імовірності того, що кожен із трьох друзів прийде в умовлене місце, дорівнюють: 0,8, 0,4, 0,7. Визначити ймовірність того, що зустріч відбудеться, якщо для цього достатньо з'явитися двом із трьох друзів?

2.7 ВТК перевіряє деякі вироби на стандартність. Імовірність того, що виріб нестандартний, дорівнює 0,1. Знайти ймовірність того, що нестандартним виявиться лише четвертий за ліком перевірений виріб (на цьому перевірка закінчується).

2.8 По лінії зв'язку, що має чотири приймально-передавальних пункти, передається повідомлення. Імовірності того, що повідомлення буде спотворене на першому, другому, третьому й четвертому пунктах відповідно дорівнюють 0,1, 0,15, 0,2 і 0,25. Яка ймовірність одержання неспотвореного повідомлення?

2.9 Для сигналізації про пожежу встановлені два незалежно працюючих датчики. Імовірності того, що під час пожежі датчик спрацює, для першого й другого відповідно дорівнюють 0,9 і 0,95. Визначити ймовірність того, що під час пожежі спрацює хоча б один датчик.

2.10 Абонент забув останню цифру номера телефону й набирає її навмання. Знайти ймовірність того, що йому доведеться набирати номер не більше трьох разів.

2.11 Робітник обслуговує чотири верстати, кожен із яких може вийти з ладу на упродовж зміни з імовірністю 0,02. Визначити ймовірність того, що з ладу вийдуть не більше 2 верстатів?

2.12 У партії з десяти деталей вісім стандартних. Визначити ймовірність того, що серед двох навмання вибраних із партії деталей є хоча б одна стандартна.

2.13 У двох ящиках знаходяться деталі: в першому – 10, з них 3 стандартні, у другому – 15, із них 6 стандартних. З кожного ящика навмання виймають по одній деталі. Знайти ймовірність того, що обидві деталі виявляться стандартними.

2.14 У студії телебачення 3 телевізійні камери. Для кожної камери ймовірність того, що вона ввімкнена в поточний момент, дорівнює 0,6. Знайти ймовірність того, що в поточний момент увімкнена хоча б одна камера.

2.15 Підприємство виготовляє 95 % виробів стандартних, причому з них 86 % – першого ґатунку. Знайти ймовірність того, що взятий навмання виріб, виготовлений на цьому підприємстві, виявиться виробом першого ґатунку.

2.16 Відділ технічного контролю перевіряє вироби на стандартність. Імовірність того, що виріб нестандартний, дорівнює 0,1. Знайти ймовірність того, що нестандартним виявиться лише четвертий із п'яти перевірених виробів.

2.17 У ящику лежать 15 плавких запобіжників, що відрізняються лише силою струму, на який вони розраховані. Із них 7 розраховані на 10А, 5 – на 8А і 3 – на 5А. Навмання беруть два запобіжники, знайти ймовірність того, що вони розраховані на максимальний струм.

2.18 Лабораторія одержала два прилади, причому ймовірність того, що прилад полагоджений, становить для першого приладу 0,75, а для другого – 0,8. Знайти ймовірність того, що під час перевірки хоча б один прилад виявиться полагодженим.

2.19 Є три деталі, причому ймовірність того, що деталь придатна, для першої становить 0,8, для другої – 0,75, для третьої – 0,6. Найти ймовірність того, що при перевірці тільки дві деталі виявляться придатними.

2.20 Ймовірність того, що прилад на протязі гарантійного терміну не буде вимагати ремонту для першого приладу складає 0,75, для другого – 0,6, для третього – 0,8. Знайти ймовірність того, що впродовж гарантійного терміну лише один із приладів не буде потребувати ремонту.

2.21 Імовірність того, що прилад упродовж гарантійного терміну не буде потребувати ремонту становить: для першого приладу 0,7, для другого – 0,8, для третього – 0,9. Знайти ймовірність того, що впродовж гарантійного терміну лише два прилади не будуть вимагати ремонту.

2.22 Для сигналізації про аварію встановлені два незалежно працюючих сигналізатори. Імовірність того, що під час аварії сигналізатор спрацює, дорівнює 0,95 для першого сигналізатора і 0,9 для другого. Знайти ймовірність того, що в разі аварії спрацює лише один сигналізатор.

2.23 Із партії виробів товарознавець відбирає вироби вищого ґатунку. Імовірність того, що навмання вибраний виріб буде мати вищий ґатунок, дорівнює 0,8. Знайти ймовірність того, що з трьох перевірених виробів лише 2 вироби мають вищий ґатунок.

2.24 Студент шукає потрібну йому формулу в трьох довідниках. Імовірність того, що формула міститься в першому, другому, третьому довіднику, відповідно дорівнює 0,6, 0,7, 0,8. Знайти ймовірність того, що формула міститься лише у двох довідниках.

2.25 Потрібний складальникові тип деталі знаходиться в кожному з чотирьох ящиків з імовірностями 0,6, 0,7, 0,8,

0,9. Знайти ймовірність того, що потрібний тип деталі міститься не менше ніж у двох ящиках.

2.26 Три дослідники незалежно один від одного виконують вимірювання деякої фізичної величини. Імовірність того, що перший дослідник зробить помилку під час зчитування показань приладу, дорівнює 0,1. Для другого й третього дослідників ця ймовірність відповідно дорівнює 0,15 і 0,2. Знайти ймовірність того, що під час одного вимірювання хоча б один із дослідників зробить помилку.

2.27 Фірма купила 3 комп’ютери, причому ймовірності того, що комп’ютери справні, відповідно дорівнюють 0,8, 0,9, 0,95. Знайти ймовірність того, що під час перевірки хоча б два комп’ютери будуть справними.

2.28 Із 25 питань студент знає 15. Йому ставлять 3 питання. Яка ймовірність того, що студент знає не менше двох питань?

2.29 Є два сигналізатори на випадок аварії. Імовірність спрацьовування кожного дорівнює 0,9. Знайти ймовірність, що в разі аварії спрацює один із сигналізаторів, якщо другий із них вмикається тільки тоді, коли перший сигналізатор не спрацював.

2.30 Деталь обробляють трьома автоматами, що працюють незалежно один від одного. Імовірності браку для кожного з автоматів дорівнюють 0,95, 0,95, 0,9. Деталь вважається 2-го ґатунку, якщо брак допущений лише одним з автоматів. Знайти ймовірність того, що виготовлена деталь 2-го ґатунку.

Завдання 3 Розв’язати задачу

3.1. Під час транспортування та вантажно-розвантажувальних роботах 6 % цегли, що надходить, виявляється битою. Яка ймовірність того, що з партії в 10 000 цеглин битими опиняться не більше 700 штук.

3.2. При масовому виробництві інтегральних схем ймовірність появи браку дорівнює 0,005. Визначити ймовірність того, що в партії з 600 виробів бракованими будуть: а) не більше ніж три вироби, б) рівно чотири вироби.

3.3. Гравець накидає кільця на кілочок, імовірність удачі при цьому дорівнює 0,1. Визначити ймовірність того, що з шести кілець на кілочок попадуть хоча б два.

3.4. В місті N у середньому за один рік народжується 12 300 дітей. Імовірність народження хлопчика дорівнює 0,515. Визначити ймовірність того, що в цьому місті за рік хлопчиків народиться менше ніж дівчаток.

3.5. Металеві труби, кожна довжиною вісім метрів, мають середню концентрацію мікродефектів у 0,375 мікродефектів на один погонний метр. Визначити ймовірність того, що: а) ця труба буде бракованою, якщо за технічними умовами допустимо не більше п’яти мікродефектів на кожну трубу; б) ця труба має рівно чотири мікродефекти.

3.6. Будівельна організація має п’ять бульдозерів, для кожного з них імовірність безвідмовної роботи упродовж деякого часу T дорівнює 0,9. Визначити такі ймовірності: а) жоден із п’яти бульдозерів не буде потребувати ремонту за час T; б) два бульдозери будуть мати потребу в проведенні ремонту.

3.7. На диспетчерський пункт у середньому надходить три замовлення на таксі за хвилину. Визначити ймовірність того, що за дві хвилини надійдуть: а) не більше п’яти викликів; б) рівно п’ять.

3.8. У механічному цеху працюють 120 токарів. Імовірність того, що кожному токарю в поточний момент часу буде потрібен різець певного типу, дорівнює 0,2. Скільки різців даного типу повинна мати інструментальна комора, щоб забезпечити з імовірністю 0,95 потребу в них.

3.9. Читальний зал інституту розрахований на 300 студентів, кожен із яких з імовірністю, що дорівнює 0,2, бере англо-український словник. Скільки таких словників повинно бути в читальному залі, щоб з імовірністю 0,85 можна було забезпечити всіх охочих?

3.10. Станок-автомат виробляє 70 % усіх виробів першим ґатунком, а ті, що залишились – другим. Потрібно з’ясувати, що є більш імовірним – отримати два вироби першого ґатунку з п’яти навмання вибраних, або п’ять першого ґатунку з десяти.

3.11. До технічного водогону підключені 160 підприємств, кожне з яких з імовірністю 0,7 у поточний момент часу виконує відбір води з магістралі. Визначити ймовірність того, що в цей момент відбір води виконують не менше 80 й не більше 120 підприємств.

3.12. Відвальний щит бульдозера захоплює смугу ґрунту завширшки 5 метрів. Середня концентрація великих каменів на одному квадратному метрі площі дорівнює 0,05 каменя. До скидання ґрунту бульдозер кожен раз проходить 20 метрів. Яка ймовірність захоплення: а) не більше 4 каменів, б) рівно 6 каменів?

3.13. Трос складається з 200 окремих стальних жил (дротів). Імовірність того, що одна жила не задовольняє технічним умовам, дорівнює 0,015. Трос бракується, коли в ньому більше чотирьох дефектних жил. Визначити ймовірність появи браку.

3.14. Упродовж години комутатор одержує в середньому шістдесят викликів. Телефоністка вийшла на дві хвилини. Яка ймовірність, що за цей час: а) не надійде ні одного виклику, б) надійде не більше двох викликів?

3.15. У камері Вільсона в середньому реєструють 15 елементарних частинок на годину. Визначити ймовірність того, що упродовж двадцяти хвилин буде зареєстровано:

а) дві частинки; б) більше двох елементарних частинок.

3.16 У телевізійній студії знаходяться чотири телевізійні камери. Імовірність того, що одна камера в поточний момент часу ввімкнена, дорівнює 0,6. Визначити імовірність того, що в поточний момент ввімкнені: а) рівно дві камери; б) хоча б одна камера.

3.17 За деякий проміжок часу середня кількість помилкових з'єднань, що припадають на одного телефонного абонента, дорівнює 5. Яка ймовірність того, що за цей час для певного абонента кількість помилкових з'єднань буде більшою ніж два?

3.18 У магазин надійшло 150 телевізорів. Імовірність того, що кожен окремий телевізор вимагає регулювання перед продажем, дорівнює 0,4. Визначити ймовірність того, що не менше ніж 50 і не більше ніж 80 телевізорів будуть потребувати додаткового регулювання.

3.19 Імовірність того, що після одного навчального року підручник буде мати потребу в новій обкладинці, дорівнює 0,25. Визначити ймовірність того, що не менше ніж 800 і не більше ніж 1100 підручників треба буде оправити заново, якщо фонд навчальної бібліотеки становить 4000 книг.

3.20 Зразок радіоактивної речовини в середньому за 10 секунд випромінює чотири заряджені частинки. Визначити ймовірність того, що за одну секунду зразок випроменить: а) хоча б одну частинку; б) рівно одну частинку.

3.21 Упродовж години комутатор одержує в середньому 60 викликів. Знайти ймовірність того, що за дві хвилини: а) не буде ні одного виклику; б) буде не більше одного виклику.

3.22 Студент здає екзамен автоматичному екзаменатору. На кожне питання відповідь «так» чи «ні». Яка ймовірність здати екзамен навмання, якщо для цього потрібно дати правильні відповіді не менше ніж на сім питань із десяти?

3.23 Керамічні труби довжиною 6 метрів мають випадкове розподілення мікродефектів, середня концентрація яких дорівнює 0,1 мікродефектів на один погонний метр труби. Визначити ймовірність того, що дана труба має: а) не більше 5 мікродефектів, б) рівно 6 мікродефектів.

3.24 Стінові блоки площею в шість квадратних метрів мають випадкове розподілення мікротріщин із середньою концентрацією 0,1 мікротріщин на один квадратній метр. Визначити ймовірність того, що даний блок: а) не має ні однієї тріщини; б) має не більше чотирьох мікротріщин.

3.25 У середньому на 1 кв. м поверхні штучного супутника потрапляє за час його роботи на орбіті 400 мікрометеоритів. Визначити ймовірність попадання більше п'яти мікрометеоритів у скло ілюмінатора, якщо його площа дорівнює 100 кв. см.

3.26 Робітник обслуговує 5 верстатів, кожен із яких може вийти з ладу впродовж зміни з імовірністю 0,01. Знайти ймовірність того, що принаймні чотири верстати пропрацюють усю зміну.

3.27 Імовірність вийти з ладу за деякий час T для одного конденсатора дорівнює 0,1. Знайти ймовірність того, що з 100 конденсаторів упродовж часу T з ладу вийдуть: а) рівно 16 конденсаторів; б) від 4 до 19 конденсаторів.

3.28 Знайти ймовірність того, що серед навмання взятих 100 деталей 55 виявляться відполірованими, якщо в загальній масі деталей є порівну відполірованих і невідполірованих. Яка ймовірність того, що кількість відполірованих деталей буде не меншою від 45 й не більшою від 50?

3.29 Під час масового виробництва діодів імовірність браку в процесі формування дорівнює 0,1. Яка ймовірність того, що з 400 навмання взятих діодів 50 будуть бракованими? Знайти ймовірність того, що бракованих буде від 25 до 55 діодів.

3.30 У сім'ї 4 дітей. Вважаючи ймовірність народження хлопчика й дівчинки однаковою, знайти ймовірність того, що в цій сім'ї 3 хлопчики й 1 дівчинка.

Завдання 4 Розв'язати задачу

4.1 ВТК проводить контроль виготовлених приладів. Прилади мають приховані дефекти з ймовірністю, що дорівнює 0,15. Під час перевірки наявність дефекту виявляється з імовірністю 0,9. Крім того, з ймовірністю 0,05 справний прилад може бути помилково признаний дефектним. У разі виявлення дефекту прилади бракуються. Визначити ймовірність того, що забракований прилад має дефект.

4.2 На деякому заводі перший верстат виробляє 40 % усієї продукції, а другий – 60 %. У середньому 9 з 1000 деталей, що виробляє перший верстат, виявляються бракованими, у другого – одна бракована деталь із 250. Випадково відібрана з усієї денної продукції деталь. Вона виявилась за результатами перевірки бракованою. Яка ймовірність того, що вона виготовлена на першому верстаті?

4.3 На інструментальний склад надійшли десять нових інструментів. Кожній із трьох змін видають випадковим чином два інструменти, які після закінчення роботи повертають знову на склад. Визначити ймовірність того, що третя зміна отримає обидва нові інструменти.

4.4 У ящику лежать 20 тенісних м'ячів, із них 12 нових і 8 вже використовуваних. Із ящика навмання дістають для гри два м'ячі й після гри повертають у ящик. Після цього виймають два м'ячі для наступної гри. Визначити ймовірність того, що обидва ці м'ячі виявляться новими.

4.5 На станцію очищення стічних вод 30 % стоку надходить із першого підприємства, 40 % – із другого й залишок – із третього. Імовірність появи в стічних водах солей важких металів для першого, другого й третього підприємств відповідно дорівнює 0,01, 0,02 і 0,04. Визначити ймовірність появи солей важких металів у всьому стоці.

4.6 На деякій фабриці 30 % продукції виготовляє перша машина, 25 % – друга, а решту продукції – третя. Перша машина дає 1 % браку, друга – 2 % і третя – 3 %. Випадково вибрана одиниця продукції виявилася бракованою. Визначити ймовірність того, що вона виготовлена на першій машині.

4.7 На складі знаходяться електролампи, виготовлені двома заводами. Серед них 70 % виготовлені першим заводом, а решта – другим. Відомо, що з кожних 100 лампочок, виготовлених першим заводом, 90 відповідають стандарту, а зі 100 ламп виготовлених другим, 80 відповідають стандарту. Визначити ймовірність того, що взята навмання лампочка буде відповідати вимогам стандарту.

4.8 На конвеєр надходять однотипні вироби, виготовлені двома робітниками. Водночас перший робітник постачає 60 %, а другий – 40 % загальної кількості виробів. Імовірність того, що виріб, виготовлений першим робітником, виявиться нестандартним, дорівнює 0,005, а другим робітником – 0,01. Узятий навмання з конвеєра виріб виявився нестандартним. Визначити ймовірність того, що він виготовлений першим робітником.

4.9 На складання надходять однотипні деталі з трьох підприємств, причому перше постачає 50 %, друге – 30 % і третє – 20 % усієї кількості. Імовірності появи браку для першого, другого й третього підприємств відповідно дорівнюють 0,05, 0,1 і 0,15. Вибірковий контроль виявив браковану деталь. Яка ймовірність того, що брак відбувається з вини другого підприємства?

4.10 Є десять однакових урн, у дев'яти з яких знаходяться по дві білих і дві чорних кулі, а в одній – п'ять білих куль та одна чорна. З урни, взятої навмання, витягли білу кулю. Яка ймовірність того, що кулю витягли з урни, яка містить п'ять білих куль?

4.11 На склад надійшло сім нових інструментів. Перша зміна отримала три інструменти й, після закінчення роботи повернула їх на склад. Друга зміна отримала два інструменти й також потім повернула їх на склад. Третя зміна отримала два інструменти. Яка ймовірність того, що всі вони нові, якщо видача інструментів випадкова?

4.12 Імовірність того, що вироби деякого виробництва відповідають стандарту, дорівнює 0,96. Пропонується спрощена система контролю, яка пропускає з ймовірністю 0,98 вироби, що відповідають стандарту, і з імовірністю 0,05 вироби, що не відповідають стандарту. Яка ймовірність того, що виріб, який пройшов такий контроль, відповідає стандарту?

4.13 У цеху три типи автоматичних верстатів виробляють одні й ті самі деталі. Продуктивність їх однакова, але якість роботи різна: верстати першого типу виробляють 90 % продукції відмінної якості, другого – 85 % і третього – 80 %. Усі виготовлені за зміну деталі надходять на склад в одну ємкість. Визначити ймовірність того, що навмання вибрана деталь буде мати відмінну якість, якщо верстатів першого типу є 10 штук, другого – 6 і третього – 4.

4.14 Об'єкт будують три бригади монтажників. Імовірності того, що дана бригада допустить порушення технології під час монтажу одного блоку дорівнюють відповідно 0,01, 0,015 і 0,02. Перша бригада виконала 50 % усього об'єму робіт, друга – 30 %, третя – 20 %. Яка

ймовірність того, що вибраний випадково блок змонтований із порушенням технології?

4.15 На деякій фабриці 30 % усієї продукції виробляє перша машина, 25 % – другою, а решта продукції – третя. Перша машина дає 1 % браку, друга – 1,5 % і третя – 2 %. Визначити ймовірність того, що випадково вибрана одиниця продукції виявиться бракованою.

4.16 40 % усіх приладів, що випускає завод, збирають з високоякісних деталей, а решта – із деталей звичайної якості. Імовірність безвідмовної роботи за час T приладу, зібраного з деталей звичайної якості, дорівнює 0,7, а з високоякісних – 0,95. Прилад випробували упродовж часу T і він працював безвідмовно. Знайти ймовірність того, що прилад був зібраний із високоякісних деталей.

4.17 У партії з 600 радіоламп 200 виготовлені на першому заводі, 250 – на другому і решта – на третьому. Імовірність того, що лампа, виготовлена на першому заводі, виявиться стандартною, дорівнює 0,97, на другому заводі – 0,91 і на третьому – 0,95. Навмання взята лампа виявилася стандартною. Визначити ймовірність того, що вона виготовлена на першому заводі.

4.18 У лікарню в середньому надходить 50 % хворих із хворобою A, 30 % – із хворобою B і 20 % – із хворобою C. Імовірності повного видужання після хвороб A, B і C відповідно дорівнюють 0,7, 0,8 і 0,9. Визначити ймовірність того, що хворий, який опинився в лікарні, буде виписаний здоровим.

4.19 На двох верстатах виробляють однакову продукцію. Продуктивність першого верстата вдвічі більша за продуктивність другого. Імовірність появи браку на першому верстаті 0,1, на другому – 0,15. Виготовлені за зміну деталі складають у контейнер. Знайти ймовірність того, що випадково вибраний із контейнера виріб не виявиться бракованим.

4.20 Деякі вироби перевіряють на стандартність два контролери, причому перший перевіряє 60 %, а другий – 40 % усієї продукції. Імовірність того, що стандартний виріб буде визнаний стандартним під час перевірки першим контролером, дорівнює 0,95, другим контролером – 0,9. (Якість контролю залежить від кваліфікації контролера). Визначити ймовірність того, що стандартний виріб буде визнаний стандартним.

4.21 Є дві партії однотипних виробів з 12 і 10 штук, причому в кожній партії є по одному бракованому. Виріб, узятий навмання з першої партії, переклали в другу, після чого вибрали навмання один виріб із другої партії. Визначити ймовірність того, що другий раз не буде витягнутий бракований виріб.

4.22 Перевірка агрегатів машини під час технічному обслуговування дозволяє визначити несправність з імовірністю 0,8. Ймовірність помилкового виявлення «несправності» дорівнює 0,01. Несправні машини складають 20 % усіх машин, що надходять на техобслуговування. Визначити ймовірність того, що машина справна, якщо вона була визнана несправною.

4.23 Імовірність того, що під час буріння свердловини будуть знайдені ґрунтові води, дорівнює 0,3. Ґрунтові води супроводжуються твердими породами з імовірністю 0,6. Там, де ґрунтових вод немає, тверді породи трапляються з імовірністю 0,8. Знайти ймовірність того, що під час буріння будуть знайдені тверді породи.

4.24 У цеху перший, другий і третій верстати виготовляють відповідно 25 %, 35 % і 40 % усіх болтів, що виробляються. Брак під час їх виготовлення становить відповідно 5 %, 4 % і 2 %. Випадково взятий зі складу болт виявився дефектним. Визначити ймовірність того, що він виготовлений на другому станку.

4.25 Є два набори деталей. Імовірність того, що деталь першого набору є стандартною, дорівнює 0,8, а другого –

0,9. Знайти ймовірність того, що випадкова деталь із навмання взятого набору нестандартна.

4.26 По лінії зв’язку передаються два сигнали: А і В відповідно з імовірностями 0,84 і 0,16. Через перешкоди 1/6 частина сигналів А спотворюється та приймається як В-сигнали, а 1/8 частина переданих В-сигналів приймається як А-сигнали. Знайти ймовірність того, що на пункті прийому з’явиться А-сигнал.

4.27 Два автомати виробляють деталі, що надходять на спільний конвеєр. Імовірність отримання нестандартної деталі на першому автоматі дорівнює 0,06, а на другому – 0,3. Продуктивність першого автомата вдвоє більша ніж продуктивність другого. Знайти ймовірність того, що навмання взята з конвеєра деталь є нестандартною.

4.28 Прилад може працювати у двох режимах: нормальному й ненормальному. Нормальний режим спостерігається у 80 % всіх випадків роботи приладу, ненормальний – в 20 %. Імовірність виходу з ладу приладу за час Т в нормальному режимі дорівнює 0,1, у ненормальному – 0,7. Знайти ймовірність виходу приладу з ладу за час Т.

4.29 Пасажир може звернутися для купівлі квитка в одну з двох кас. Імовірність звернення в кожну касу залежить від їх розташування й дорівнює відповідно 0,7 і 0,3. Імовірність того, що до моменту появи пасажира білети будуть розпродані, дорівнює для першої каси 0,8, для другої – 0,4. Знайти ймовірність того, що, вибравши навмання касу, пасажир придбає квиток.

4.30 У двох ящиках знаходяться радіолампи. У першому ящику міститься 12 ламп, із них одна нестандартна; у другому 10 ламп, із них 2 нестандартні. З першого ящику навмання взята лампа й перекладена в другий. Знайти ймовірність того, що навмання витягнута з другого ящика лампа буде нестандартною.

Завдання 5 У схемі повторних випробувань точне значення ймовірності можна одержати за формулою Бернуллі**.** У разі заданих «*n*» і «*m*» для вказаних значень $p_i = \dfrac{A_i}{1000}$, (i = 1**,** 2, 3, 4**,** 5) обчислити $P_n(m)$ за трьома формулами: Бернуллі**,** Лапласа й Пуассона**.** Для кожного значення p_i знайти відносну похибку формул Лапласа й Пуассона. Побудувати на одному рисунку графіки відносних похибок цих формул як функцій *p*. Показати області переважного використання формул Лапласа й Пуассона. (Числові параметри варіантів наведено в таблиці 1).

Таблиця 1 Числові параметри

Варіант	*n*	*m*	A_1	A_2	A_3	A_4	A_5	x_1	x_2	x_3	*M*	*D*
1	12	1	207	233	250	270	306	-3	-2	2	-1	4,0
2	12	1	203	240	254	272	304	-2	1	2	0	2,8
3	12	1	202	231	259	274	300	-1	0	5	1	7,0
4	12	1	209	238	255	278	308	0	1	4	2	2,8
5	12	1	200	236	253	277	301	1	2	6	3	4,0
6	12	2	257	282	299	320	356	2	5	6	4	2,8
7	12	2	253	289	303	322	354	3	4	9	5	7,0
8	12	2	252	280	308	324	350	4	5	8	6	2,8
9	12	2	259	287	304	328	358	-4	-3	1	-2	4,0
10	12	2	250	285	302	327	351	-3	0	1	-1	2,8
11	13	1	177	203	220	240	277	-2	-1	4	0	7,0
12	13	1	173	210	224	242	275	-1	0	3	1	2,8
13	13	1	172	201	229	244	271	0	1	5	2	4,0
14	13	1	179	208	225	248	279	1	4	5	3	2,8
15	13	1	170	206	223	247	272	2	3	8	4	7,0
16	13	2	246	272	289	309	346	3	4	7	5	2,8
17	13	2	242	279	293	311	344	4	5	9	6	4,0
18	13	2	241	270	298	314	340	-4	-1	0	-2	2,8
19	13	2	248	277	294	318	348	-3	-2	3	-1	7,0
20	13	2	239	275	292	317	341	-2	-1	2	0	2,8
21	14	1	147	173	190	210	247	-1	0	4	1	4,0
22	14	1	143	180	194	212	245	0	3	4	2	2,8
23	14	1	142	171	199	214	241	1	2	7	3	7,0
24	14	1	149	178	195	218	249	2	3	6	4	2,8
25	14	1	140	176	193	217	242	3	4	8	5	4,0
26	14	2	246	272	289	309	346	4	7	8	6	2,8
27	14	2	242	279	293	311	344	-4	-3	2	-2	7,0
28	14	2	241	270	298	314	340	-3	-2	1	-1	2,8
29	14	2	248	277	294	318	348	-2	-1	3	0	4,0
30	14	2	239	275	292	317	341	-1	2	3	1	2,8

Завдання 6 Випадкова величина може набувати значень x_1, x_2, x_3 з імовірностями p_1, p_2, p_3 (див. табл. 1). За заданих значеннь випадкової величини та її математичного сподівання *M* імовірності p_i визначаються неоднозначно. Виконати такі дослідження:

а) визначити області можливих значень для кожної з імовірностей p_1, p_2, p_3;

б) знайти в параметричному вигляді $p_1 = p_1(t)$, $p_2 = p_2(t)$, $p_3 = p_3(t)$ взаємозв'язок між імовірностями (параметр *t* вибрати так**,** щоб функції $p_i(t)$ були лінійними). Побудувати графіки цих функцій на одному рисунку;

в) використовуючи графіки пункту б), знайти закон розподілення випадкової величини *X*, що найменше відрізняється від рівномірного, тобто таке розподілення, при якому величина $R = \max p_i - \min p_i$ набуває найменших значень.

г) знайти область можливих значень $D(X)$.

Завдання 7 Випадкова величина *X* може набувати значень x_1, x_2, x_3 з математичним сподіванням *M* і дисперсією *D* (див. табл. 1). Знайти закон розподілення *X*.

Завдання 8 Випадкова величина X розподілена за законом

$$f(x)=\begin{cases} 0, & x \le a; \\ C_1 x + C_0, & a < x < b; \\ 0, & x > b. \end{cases} \qquad (9.1)$$

із математичним сподіванням M_1 (табл. 2). Виконати такі завдання:

а) визначити коефіцієнти C_1, C_0 і побудувати графік функції густини ймовірності (функції $f(x)$);

б) знайти інтегральну функцію розподілення (функцію $F(x)$) і побудувати її графік;

в) знайти дисперсію та середнє квадратичне відхилення X.

Завдання 9 Виконати такі завдання для випадкової величини X, розподіленої за законом (9.1) (табл. 2):

а) знайти діапазон можливих значень математичного сподівання M;

б) побудувати графіки $f(x)$ при $M = M_{\min}$ і $M = M_{\max}$ (на одному рисунку);

в) знайти залежність $D = D(M)$, побудувати її графік;

г) знайти діапазон можливих значень дисперсії D;

д) знайти діапазон можливих значень $P\{c \le X \le d\}$.

Завдання 10 Випадкова величина X розподілена за законом (9.1). Відома ймовірність p попадання X в інтервал [c; d] (див. табл. 2). Знайти закон розподілення X та обчислити її числові характеристики.

Завдання 11 Виконати такі завдання (див. табл. 2):

а) обчислити числові характеристики випадкової величини X;

Таблиця 2 Дані для завдання

Варіант	a	b	c	d	P	M_1	M_2	D	u_1	u_2
1	-5	1	0	2	17/72	-3	3/2	3/20	1/10	1/2
2	-4	4	3	6	1/6	1	4	1	1/14	1/4
3	-3	2	0	3	13/25	-1	3/4	27/80	1/6	1/2
4	-2	4	2	5	4/9	2	4	3/4	1/8	1/2
5	-1	7	5	9	1/8	2	8	3/5	1/12	1/4
6	0	5	2	8	18/25	3	6	4	1/4	1/6
7	-6	0	-4	2	7/9	-4	-5/2	27/20	1/4	1/4
8	-5	3	2	4	1/6	0	7/2	1/60	1/14	1/2
9	-4	1	0	2	7/25	-2	1/2	3/20	1/8	1/2
10	-3	3	1	4	4/9	1	3	1	1/8	1/2
11	-2	6	5	8	1/6	1	29/4	27/80	1/14	1/4
12	-1	4	3	6	7/25	2	4	3/4	1/8	1/4
13	0	6	3	7	5/8	2	4	3/5	1/6	1/2
14	-6	2	-2	4	3/8	-1	0	4	1/8	1/4
15	-5	0	-3	3	18/25	-3	3/2	27/20	1/4	1/6
16	-4	2	1	3	17/72	0	3/2	1/60	1/10	1/2
17	-3	5	4	6	1/6	0	11/2	3/20	1/14	1/2
18	-2	3	2	5	7/25	1	3	1	1/8	1/4
19	-1	5	3	6	4/9	1	15/4	27/80	1/8	1/2
20	0	8	6	9	1/8	5	8	3/4	1/12	1/2
21	-6	-1	-3	1	13/25	-4	0	3/5	1/6	1/4
22	-5	1	-2	4	5/8	-1	2	4	1/6	1/6
23	-4	4	0	6	3/8	-1	3/2	27/20	1/8	1/4
24	-3	2	1	3	7/25	0	5/2	1/60	1/8	1/2
25	-2	4	3	5	17/72	0	7/2	3/20	1/10	1/2
26	-1	7	5	8	1/8	4	7	1	1/12	1/2
27	0	5	4	7	7/25	2	25/4	27/80	1/8	1/4
28	-6	0	-1	2	17/72	-2	0	3/4	1/10	1/4

Варіант	*a*	*b*	*c*	*d*	*P*	M_1	M_2	*D*	u_1	u_2
29	-5	3	0	4	1/4	-2	1	3/5	1/10	1/2
30	-4	1	-3	3	22/25	-1	-1	4	1/2	1/4

б) знайти інтегральну функцію розподілення (функцію $F(x)$) і побудувати її графік, якщо випадкова величина X розподілена за законом

$$f(x)=\begin{cases} 0, & x<a; \\ u_1, & a\le x\le b; \\ 0, & b\le x\le c; \\ u_2, & c\le x\le d; \\ 0, & d<x. \end{cases}$$

Завдання 12 Неперервна випадкова величина X розподілена за законом

$$f\left(x\right)=\begin{cases} 0, & x<c, \\ c_2x^2+c_1x+c_0, & c\le x\le d, \\ 0, & x>d. \end{cases}$$

із математичним сподіванням M_2 і дисперсією D (див. табл. 2).

Виконати такі завдання:

а) визначити коефіцієнти c_0, c_1, c_2 і побудувати графік густини ймовірності (функції $f(x)$);

б) на одному рисунку побудувати графіки $f(x)$ і функції густини рівномірного розподілення на тому самому інтервалі $[c; d]$;

в) знайти найбільше й найменше значення $f(x)$, дати їм імовірнісну інтерпретацію;

г) знайти найбільше відхилення $f(x)$ від рівномірного розподілення на тому самому інтервалі.

Завдання 13 Обчислити ймовірність попадання нормально розподіленої випадкової величини X з параметрами $a = M_2$, $\sigma=\sqrt{D}$ в інтервал $[c; d]$ (див. табл. 2).

Завдання 14 Двовимірна випадкова величина розподілена за законом (табл. 5.4) з відомими x_i, y_j, p_{ij} (табл. 3)

Обчислити її числові характеристики:

а) математичне сподівання й дисперсії компонентів випадкової величини $M[X]$, $M[Y]$, $D[X]$, $D[Y]$;

б) кореляційний момент μ_{xy} і коефіцієнт кореляції r_{xy}.

Завдання 15 Двовимірна випадкова величина розподілена за законом (табл. 5.4), з відомими значеннями x_i, y_j і числовими характеристиками $M[X]$, $M[Y]$, μ_{xy} (кореляційний момент) (див. табл. 3). Знайти інші параметри закону розподілення (тобто ймовірності p_{ij}).

Завдання 16 Двовимірна випадкова величина розподілена за законом (табл. 5.4) з відомими значеннями x_i, y_j і числовими характеристиками $M[X]$, $M[Y]$, μ_{xy} (див. табл. 3). Знайти інші параметри закону розподілення, тобто ймовірності, якщо X та Y незалежні.

Таблиця 3 Задані параметри

Варіант	x_1	x_2	y_1	y_2	p_{11}	p_{12}	p_{22}	*M[X]*	*M[Y]*	μ_{xy}
1	-2	-1	1	3	0,29	0,41	0,06	-1,2	1,6	-0,2

Варіант	x_1	x_2	y_1	y_2	p_{11}	p_{12}	p_{22}	$M[X]$	$M[Y]$	μ_{xy}
2	1	4	-1	1	0,20	0,30	0,41	2,9	0,0	-0,7
3	-1	0	0	3	0,05	0,59	0,06	-0,4	1,0	-0,4
4	1	3	2	3	0,58	0,25	0,11	2,6	2,1	-0,1
5	0	4	0	1	0,69	0,10	0,08	3,1	0,2	-0,1
6	-1	0	0	1	0,37	0,22	0,15	-0,3	0,4	0,1
7	0	1	0	3	0,37	0,08	0,30	0,5	1,6	-0,4
8	1	3	1	2	0,13	0,36	0,07	1,8	1,5	0,0
9	-1	3	0	4	0,15	0,69	0,10	0,5	0,6	0,3
10	2	5	-1	3	0,08	0,59	0,23	3,2	0,3	-0,3
11	-1	0	1	2	0,20	0,49	0,23	-0,1	1,3	-0,0
12	2	6	0	3	0,27	0,15	0,15	4,3	1,7	-0,6
13	-1	3	2	3	0,12	0,19	0,25	0,9	2,6	0,3
14	0	2	0	2	0,33	0,16	0,32	1,0	1,0	-0,1
15	-1	0	0	3	0,11	0,68	0,06	-0,8	0,6	0,1
16	2	5	1	4	0,43	0,10	0,33	3,9	2,4	0,2
17	-1	2	1	3	0,23	0,33	0,22	1,1	1,8	-0,6
18	0	1	2	5	0,21	0,50	0,11	0,7	2,8	-0,3
19	0	2	2	4	0,06	0,09	0,55	1,1	3,7	0,1
20	0	3	0	1	0,11	0,06	0,42	2,4	0,8	0,1
21	2	4	0	1	0,26	0,21	0,47	2,8	0,5	-0,3
22	0	4	2	5	0,13	0,27	0,47	0,9	3,8	-0,6
23	-1	2	-1	3	0,13	0,36	0,26	0,0	1,0	-0,9
24	0	3	1	5	0,43	0,40	0,10	1,2	1,6	0,5
25	0	1	0	1	0,09	0,10	0,46	0,5	0,8	-0,0
26	2	4	-1	1	0,26	0,14	0,38	3,3	0,2	0,2
27	1	5	1	4	0,42	0,18	0,15	2,7	2,2	-1,2
28	-1	3	0	3	0,41	0,08	0,42	0,7	1,5	-1,7
29	1	5	1	5	0,11	0,26	0,09	3,8	3,5	1,1
30	-1	2	-1	1	0,40	0,30	0,14	1,2	-0,3	-0,3

Завдання 17. Двовимірна випадкова величина розподілена за законом (табл. 4.4) із заданими значеннями x_i, y_j і числовими характеристиками $M[X]$, $M[Y]$, (див. табл. 3). Знайти область можливих значень кореляційного моменту μ_{xy}.

Для розв'язання наступних трьох завдань студенту потрібно згідно зі своїм варіантом з таблиці 4 вибрати так званий *набір*.

Таблиця 5.4 Задані параметри

Варіант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
набір	3 4 5 6 8	1 2 3 6 8	2 3 5 7 8	2 3 4 6 8	1 2 5 6 8	1 4 5 6 7	1 3 5 6 8	3 5 6 7 8	1 2 3 4 7	1 3 4 7 8	1 3 4 6 7	2 3 6 7 8	2 3 4 5 6	1 4 5 7 8	1 2 3 4 5	2 5 6 7 8	1 2 4 5 8

Продовження таблиці 5.4

Варіант	**19**	**20**	**21**	**22**	**23**	**25**	**26**	**27**	**28**	**29**
набір	1 4 5 6 8	1 3 5 7 8	1 2 4 5 7	2 3 4 5 7	1 2 5 6 7	2 3 4 5 8	1 2 3 4 6	1 2 4 6 8	1 3 4 5 7	1 4 6 7 8

Завдання 18 Із групи *A* таблиці 5 вибрати числа, які записані в клітинах із номерами з *набору* (див. табл. 4) для даного варіанта. Порожня клітина означає відсутність числа. Виписані такий спосіб числа назвемо вибіркою *A*. Аналогічно за допомогою рядка *B* обираємо вибірку *B*. Позначимо $C = A \cup B$ (об'єднана вибірка). Виконати такі завдання:

а) знайти вибіркові середні й дисперсії для *A* і *B*;

б) обчислити внутрішньо й міжгрупові дисперсії;

в) визначити середню дисперсію вибірки *C* за обчисленими числовими характеристиками *A* та *B*;

г) обчислити виправлену вибіркову дисперсію та середнє квадратичне відхилення для «*C*».

Таблиця 5 Набір даних

Набір	1	2	3	4	5	6	7	8
Група А		-2,4	3,7		0,8	5,8	-1,1	-3,1
Група В	-2,9	8,3	1,8	-3,7	4,2		-0,5	6,1

Завдання 19 Із рядків таблиці 6 згідно зі своїм набором (див. табл. 4) скласти вибірку, тобто початковий числовий масив. Інші потрібні дані наведені в таблиці 7.

Пояснення. Числа набору вказують, які рядки належать до даного варіанта. Так, для варіанта 8 у початковий числовий масив з таблиці 6 треба включити елементи рядків 3, 5, 6, 7, 8; тобто одержимо набір із 26 + 21 + 24 + 25 + 23 = 119 елементів. Для спрощення запису в таблиці 6 дев'ятьма буквами: A, B, C, D, E, F, G, H, K позначені числа, що утворюють арифметичну прогресію. Два перших члени цієї прогресії вказані в таблиці 7. Наприклад, для варіанта 3 з таблиці 7 виписуємо A = 1,0, B = 1,4. Отже, крок прогресії 0,4й наступні букви позначають числа C = 1,8, D = 2,2 і т. д.

Виконати такі завдання:

а) скласти варіаційний ряд;

б) знайти розмах, медіану й моду вибірки;

в) побудувати полігон частот;

г) побудувати гістограму вибірки. Основи прямокутників гістограми вибрати так, щоб точки, які відповідають числам «А», «В» тощо опинились на

серединах основ. Інтервали, частота яких виявиться меншою 5, об'єднати з більш показними сусідніми інтервалами;

д) обчислити вибіркове середнє, дисперсію, виправлену дисперсію та середнє квадратичне відхилення;

е) обчислити теоретичні частоти для нормального закону, крайні інтервали брати напівнескінченними (сума всіх частот водночас буде дорівнювати об'єму вибірки);

ж) накласти теоретичну криву на гістограму;

з) обчислити суму Пірсона;

и) визначити кількість ступенів вільності, з рівнем значущості α перевірити гіпотезу про нормальне розподілення;

к) визначити ймовірність попадання в інтервал [*a*; *b*] двома способами – за відносною частотою й за теоретичною функцією розподілу;

л) знайти інтервали довіри для числових параметрів *a* та σ нормально розподіленої генеральної сукупності.

Таблиця 6 Набір даних

Набір	*Елементи вибірки*	*шт*
1	F E E F C E D E B E F D G D H C G B E B C C D D	24
2	B D E D G K G F D E D C H E D C G E K C	20
3	C E F F C D F G G D G C E K E H D E E F D D C G K A	26
4	E F D E D E K F E F A G E C E G G D D B D F H E G H	26
5	E F G F H E B G E G E K F B E B D G E F K	21
6	B D F F F F H D C E D A G K D E D B E F F K D G	24
7	D F H F D C D H G G E H B K E G D B B A D E G D E	25
8	H E H E E F H G E D D C K A E D B F D C D G C	23

Таблиця 7 Набір даних

Варіант	*A*	*B*	*a*	*b*	γ	α
1	0,6	1,5	3,3	4,8	0,999	0,025
2	0,8	1,2	1,2	1,7	0,999	0,025
3	1,0	1,4	1,4	1,9	0,95	0,05
4	2,8	3,6	3,6	4,9	0,99	0,05
5	0,5	1,0	1,5	2,2	0,99	0,05
6	0,5	1,0	1,0	1,7	0,999	0,05
7	1,2	1,6	2,4	2,9	0,999	0,01
8	1,9	2,7	3,5	4,8	0,999	0,05
9	0,5	1,2	1,2	2,3	0,99	0,01
10	1,7	2,3	2,9	3,8	0,95	0,05
11	1,1	1,9	2,7	4,0	0,95	0,01
12	2,1	3,0	3,9	5,4	0,99	0,02
13	1,2	1,9	3,3	4,4	0,999	0,01
14	0,7	1,2	2,2	2,9	0,99	0,05
15	0,9	1,4	1,9	2,6	0,999	0,01
16	2,9	3,6	4,3	5,4	0,95	0,025
17	2,6	3,2	4,4	5,3	0,999	0,01
18	1,2	2,1	3,9	5,4	0,99	0,05
19	1,9	2,6	4,0	5,1	0,99	0,05
20	1,7	2,6	4,4	5,9	0,99	0,025
21	1,7	2,2	3,2	3,9	0,95	0,05
22	1,5	2,2	2,9	4,0	0,99	0,05
23	1,0	1,9	2,8	4,3	0,99	0,01
24	0,7	1,2	1,7	2,4	0,95	0,05
25	0,7	1,3	1,3	2,2	0,999	0,01
26	1,8	2,4	3,6	4,5	0,99	0,01
27	1,7	2,3	2,9	3,8	0,95	0,01
28	1,6	2,4	4,0	5,3	0,999	0,025
29	2,9	3,3	3,7	4,2	0,99	0,05
30	0,7	1,1	1,1	1,6	0,999	0,05

Завдання 20 (таблиця 8). Для заданої двовимірної вибірки виконати такі завдання:

а) обчислити числові характеристики S_x, S_y, S_{xx}, S_{yy}, S_{xy}, $\bar{x}$, $\bar{y}$, D_x, D_y;

б) обчислити коефіцієнти рівняння лінійної регресії;

в) обчислити вибірковий коефіцієнт кореляції, оцінити тісноту лінійного зв'язку між факторами;

г) для X_{min}, X, X_{max} знайти передбачення «y»;

д) для X_{min}, X, X_{max} знайти інтервали довіри лінії регресії з надійністю γ;

е) побудувати на одному рисунку кореляційне поле, лінію регресії та область довіри.

У таблиці 8 число «k» дорівнює остачі від ділення номера варіанта на 5, а число «n» набуває значення з набору для цього варіанта (табл. 5.4). Отже, у початковий числовий масив треба включити пари чисел «x» і «y», записані в стовпці «k» в тих клітинах (відокремлених товстими горизонтальними лініями), номер «n» яких належить набору даного варіанта.

Таблиця 8 Набір даних

$k \rightarrow$	0		1		2		3		4	
$n \downarrow$	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*
1	35	30	58	13	83	28	65	15	75	28
	31	38	29	32	86	29	86	39	77	34
	1	73	2	54	80	25	76	28	99	63
	36	28	29	34	90	34	86	37	96	58
2	6	63	31	31	61	14	86	37	65	18
	43	25	37	25	75	24	81	32	95	58
	27	42	21	39	76	23	63	15	65	21

$k \rightarrow$ / $n \downarrow$	0		1		2		3		4	
	x	y	x	y	x	y	x	y	x	y
	3	69	55	11	70	19	96	44	62	16
3	1	70	0	61	86	33	99	50	92	53
	6	63	49	23	92	38	84	37	77	33
	33	34	56	20	98	41	77	27	78	35
	46	47	26	34	79	23	62	19	96	55
4	10	58	38	26	78	25	76	27	82	40
	49	16	14	40	65	18	78	29	71	23
	45	24	33	31	71	17	87	40	82	38
	46	21	21	36	64	16	72	24	75	28
5	17	52	40	27	63	16	92	46	85	40
	4	71	7	52	96	45	80	31	64	17
	17	50	4	53	66	17	85	36	92	51
	34	32	24	38	91	36	91	48	60	16
6	2	76	57	17	74	22	61	19	66	24
	4	70	43	24	95	43	60	18	96	61
	39	28	51	17	85	29	92	42	91	54
	4	70	7	49	85	31	89	43	68	24
7	9	61	55	22	70	20	66	23	88	49
	16	54	45	25	61	12	71	26	64	21
	35	28	46	21	82	26	64	18	88	45
	9	65	35	26	83	26	74	24	87	45
8	3	73	34	30	93	39	68	20	96	59
	18	52	28	31	86	29	87	42	81	33
	45	15	17	40	86	29	75	27	65	20
	3	74	16	38	88	38	73	25	69	25

Список літератури

1. Чистяков В. П. Курс теории вероятностей / В. П. Чистяков. – Москва : Наука, 1982. – 224 с.

2. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике / В. Е. Гмурман. – Москва : Высш. шк., 1979. – 400 с.

3. Гмурман В. Е. Теория вероятностей и математическая статистика / В. Е. Гмурман. – Москва: . Высш. шк., 1979. – 479 с.

4. Вентцель Е. С. Теория вероятностей. / Е. С Вентцель, Л. А. Овчаров. – Москва : Наука, 1969. – 432 с.

5. Пискунов Н. С. Дифференциальное и интегральное исчисления : учебник для вузов: в 2 т. / Н. С. Пискунов. – М. : Интеграл-пресс, 2001. – Т 2. – 544 с. – ISBN 5-89602-013-9.

6. Кремер Н. Ш. Теория вероятностей и математическая статистика : учебник для вузов / Н. Ш. Кремер. – М. : ЮНИТИ-ДАНА, 2003. – 543 с. – ISBN 5-238-00141-X.

7. Карасёв А. И. Теория вероятностей и математическая статистика / А. И. Карасёв. – Москва : Статистика, 1970. – 368 с.

8. Венецкий И. Г., Теория вероятностей и математическая статистика / И. Г. Венецкий, Г. С. Кильдишев. – Москва : Статистика. 1975. – 346 с.

9. Демидович Б. П. Основы вычислительной математики / Б. П. Демидович, И. А. Марон. – Москва : Наука, 1970. – 668 с.

10. Дрейпер Н. Прикладной регрессионный анализ / Н. Дрейпер, Г. Смит. – Москва : Финансы и статистика, 1986. – Том 1. – 432 с.

11. Бугір М. К. Теорія ймовірності та математична статистика : посібник для студентів економічних

спеціальностей вузів / М. К. Бугір. – Тернопіль : Підручники і посібники, 1998. – 176 с. – ISBN 966-562-175-0.

ДОДАТОК А.
ТАБЛИЦІ СПЕЦІАЛЬНИХ ФУНКЦІЙ

Таблиця А.1 – Значення функції ϕ(x)

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3725	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1899	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013

	0	1	2	3	4	5	6	7	8	9
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Таблиця А.2. – Таблиця значень функції Ф(х)

X	*Ф*(*X*)	*X*	*Ф*(*X*)	*X*	*Ф*(*X*)	*X*	*Ф*(*X*)
0,00	0,0000	0,41	0,1591	0,82	0,2939	1,23	0,3907
0,01	0,0040	0,42	0,1627	0,83	0,2967	1,24	0,3925
0,02	0,0080	0,43	0,16664	0,84	0,2995	1,25	0,3944
0,03	0,0120	0,44	0,1700	0,85	0,3023	1,26	0,3962
0,04	0,0160	0,45	0,1736	0,86	0,3051	1,27	0,3980
0,05	0,0199	0,46	0,1772	0,87	0,3079	1,28	0,3997
0,06	0,0239	0,47	0,1808	0,88	0,3106	1,29	0,4015
0,07	0,0279	0,48	0,1844	0,89	0,3133	1,30	0,4032
0,08	0,0319	0,49	0,1879	0,90	0,3159	1,31	0,4049
0,09	0,0359	0,50	0,1915	0,91	0,3186	1,35	0,4115
0,10	0,0398	0,51	0,1950	0,92	0,3212	1,36	0,4131
0,11	0,0438	0,52	0,1985	0,93	0,3238	1,37	0,4147
0,12	0,0478	0,53	0,2019	0,94	0,3264	1,38	0,4162
0,13	0,0517	0,54	0,2054	0,95	0,3289	1,39	0,4177
0,14	0,0557	0,55	0,2088	0,96	0,3315	1,40	0,4193
0,15	0,0596	0,56	0,2122	0,97	0,3340	1,41	0,4207
0,16	0,0636	0,57	0,2157	0,98	0,3365	1,42	0,4222
0,17	0,0675	0,58	0,2190	0,99	0,3389	1,43	0,4236
0,18	0,0714	0,59	0,2224	1,00	0,3414	1,44	0,4251
0,19	0,0754	0,60	0,2257	1,01	0,3438	1,45	0,4265
0,20	0,0793	0,61	0,2291	1,02	0,3461	1,46	0,4279
0,21	0,0832	0,62	0,2324	1,03	0,3485	1,47	0,4292
0,22	0,0871	0,63	0,2356	1,04	0,3508	1,48	0,4306
0,23	0,0910	0,64	0,2389	1,05	0,3531	1,49	0,4319
0,24	0,0948	0,65	0,2421	1,06	0,3554	1,50	0,4332
0,25	0,0987	0,66	0,2454	1,07	0,357	1,51	0,4345
0,26	0,1026	0,67	0,2486	1,08	0,3599	1,52	0,4347
0,27	0,1064	0,68	0,2517	1,09	0,3622	1,53	0,4370
0,28	0,1103	0,69	0,2549	1,10	0,3634	1,54	0,4382
0,29	0,1141	0,70	0,2580	1,11	0,3665	1,55	0,4394

X	*Ф*(*X*)	*X*	*Ф*(*X*)	*X*	*Ф*(*X*)	*X*	*Ф*(*X*)
0,30	0,1179	0,71	0,2611	1,12	0,3687	1,56	0,4406
0,31	0,1217	0,72	0,2642	1,13	0,3708	1,57	0,4418
0,32	0,1255	0,73	0,2673	1,14	0,3729	1,58	0,4429
0,33	0,1293	0,74	0,2703	1,15	0,3749	1,59	0,4441
0,34	0,1331	0,75	0,2734	1,16	0,3770	1,60	0,4452
0,35	0,1368	0,76	0,2764	1,17	0,3790	1,61	0,4463
0,36	0,1406	0,77	0,2793	1,18	0,3810	1,62	0,4474
0,37	0,1443	0,78	0,2823	1,19	0,3830	1,63	0,4485
0,38	0,1480	079	0,2852	1,20	0,3849	1,64	0,4495
0,39	0,1517	0,80	0,2881	1,21	0,3869	1,65	0,4505
0,40	0,1554	0,81	0,2910	1,22	0,3888	1,66	0,4515
						1,67	0,4525

Продовження таблиці А.2

X	*Ф*(*X*)	*X*	*Ф*(*X*)	*X*	*Ф*(*X*)	*X*	*Ф*(*X*)
1,68	0,4535	1,92	0,4726	2,32	0,4898	2,78	0,4973
1,69	0,4545	1,93	0,4732	2,34	0,4903	2,80	0,4974
1,70	0,4554	1,94	0,4738	2,36	0,4909	2,82	0,4976
1,71	0,4564	1,95	0,4744	2,38	0,4913	2,84	0,4977
1,72	0,4573	1,96	0,4750	2,40	0,4918	2,86	0,4979
1,73	0,4582	1,97	0,4756	2,42	0,4922	2,88	0,4980
1,74	0,4591	1,98	0,4761	2,44	0,4926	2,90	0,4981
1,75	0,4599	1,99	0,4767	2,46	0,4930	2,92	0,4982
1,76	0,4608	2,00	0,4772	2,48	0,4934	2,94	0,4984
1,77	0,4616	2,02	0,4783	2,50	0,4938	2,96	0,4985
1,78	0,4625	2,04	0,4793	2,52	0,4941	2,98	0,4986
1,79	0,4633	2,06	0,4803	2,54	0,4944	3,00	0,4986
1,80	0,4641	2,08	0,4812	2,56	0,4948	3,20	0,4993
1,81	0,4648	2,10	0,4821	2,58	0,4951	3,40	0,4996
1,82	0,4656	2,12	0,4830	2,60	0,4953	3,60	0,4998
1,83	0,4664	2,14	0,4838	2,62	0,4956	3,80	0,4999
1,84	0,4671	2,16	0,4846	2,64	0,4958	4,00	0,4999
1,85	0,4678	2,18	0,4854	2,66	0,4961	4,25	0,4999
1,86	0,4686	2,20	0,4861	2,68	0,4963	4,50	0,4999
1,87	0,4693	2,22	0,4868	2,70	0,4965	5,00	0,4999
1,88	0,4699	2,24	0,4874	2,72	0,4967	∞	0,5
1,89	0,4706	2,26	0,4881	2,74	0,4969		
1,90	0,4713	2,28	0,4887	2,76	0,4971		
1,91	0,4719	2,30	0,4893	2,77	0,4972		

Таблиця А.3. – Таблиця значень функції $t_\gamma = t(\gamma, n)$,

n \ γ	0,95	0,99	0,999
5	2,78	4,60	8,61
6	2,57	4,03	6,86
7	2,45	3,71	5,96
8	2,37	3,50	5,41
9	2,31	3,36	5,04
10	2,26	3,25	4,78
11	2,23	3,17	4,59
12	2,20	3,11	4,44
13	2,18	3,06	4,32
14	2,16	3,01	4,22
15	2,15	2,98	4,14
16	2,13	2,95	4,07
17	2,12	2,92	4,02
18	2,11	2,90	3,97
19	2,10	2,88	3,92

n \ γ	0,95	0,99	0,999
20	2,093	2,861	3,883
25	2,064	2,797	3,745
30	2,045	2,756	3,659
35	2,032	2,720	3,600
40	2,023	2,708	3,558
45	2,016	2,692	3,527
50	2,009	2,679	3,502
60	2,001	2,662	3,464
70	1,996	2,649	3,439
80	1,991	2,640	3,418
90	1,987	2,633	3,403
100	1,984	2,627	3,392
120	1,980	2,617	3,374
∞	1,960	2,576	3,291

Таблиця А.4 – Таблиця значень функції $q_\gamma = q(\gamma, n)$

n \ γ	0,95	0,99	0,999
5	1,37	2,67	5,64
6	1,09	2,01	3,88
7	0,92	1,62	2,98
8	0,80	1,38	2,42
9	0,71	1,20	2,06
10	0,65	1,08	1,80
11	0,59	0,98	1,60
12	0,55	0,90	1,45
13	0,52	0,83	1,33

n \ γ	0,95	0,99	0,999
20	0,37	0,58	0,88
25	0,32	0,49	0,73
30	0,28	0,43	0,63
35	0,26	0,38	0,56
40	0,24	0,35	0,50
45	0,22	0,32	0,46
50	0,21	0,30	0,43
60	0,188	0,269	0,38
70	0,174	0,245	0,34

n \ γ	0,95	0,99	0,999
14	0,48	0,78	1,23
15	0,46	0,73	1,15
16	0,44	0,70	1,07
17	0,42	0,66	1,01
18	0,40	0,63	0,96
19	0,39	0,60	0,92

n \ γ	0,95	0,99	0,999
80	0,161	0,226	0,31
90	0,151	0,211	0,29
100	0,143	0,198	0,27
150	0,115	0,160	0,211
200	0,099	0,136	0,185
250	0,089	0,120	0,162

Таблиця А.5 – Критичні точки розподілення χ^2

Кількість ступенів вільності	Рівень значущості α		
	0,01	0,025	0,05
1	6,6	5	3,8
2	9,2	7,4	6
3	11,3	9,3	7,8
4	13,3	11,1	9,5
5	15,1	12,8	11,1
6	16,8	14,4	12,6
7	18,5	16	14,1
8	20,1	17,5	15,5
9	21,7	19	16,9
10	23,2	20,5	18,3
11	24,7	21,9	19,7
12	26,2	23,3	21
13	27,7	24,7	22,4
14	29,1	26,1	23,7
15	30,6	27,5	25
16	32	28,8	26,3
17	33,4	30,2	27,6
18	34,8	31,5	28,9
19	36,2	32,9	30,1

Кількість ступенів вільності	Рівень значущості α		
	0,01	0,025	0,05
20	37,6	34,2	31,4
21	38,9	35,5	32,7
22	40,3	36,8	33,9
23	41,6	38,1	35,2
24	43	39,4	36,4
25	44,3	40,6	37,7
26	45,6	41,9	38,9
27	47	43,2	40,1
28	48,3	44,5	41,3
29	49,6	45,7	42,6
30	50,9	47	43,8

Навчальне видання

Гончаров Олександр Андрійович
Князь Ігор Олександрович
Хоменко Олексій Віталійович

ТЕОРІЯ ЙМОВІРНОСТЕЙ
І МАТЕМАТИЧНА СТАТИСТИКА

Навчальний посібник

За загальною редакцією О. А. Гончаров

Художнє оформлення обкладинки О. Ф. Дубровіна

Редактор

Комп’ютерна верстання

. Формат 60 х 84/16.
Ум. друк. арк. . Обл.-вид. арк. .
Тираж прим. Зам. № .
Свідоцтво про внесення суб’єкта видавничої справи
до Державного реєстру
серія ДК №1633 від 24.12.03.