

Міністерство освіти та науки України
Луцький національний технічний університет



ТЕХНІЧНИЙ ФАХОВИЙ КОЛЕДЖ

Луцького національного технічного університету

**ОСНОВИ ТЕОРІЇ ЙМОВІРНОСТЕЙ
ТА СТАТИСТИКИ**

Конспект лекцій
для здобувачів фахової передвищої освіти
освітньо-професійної програми «Транспортні технології
(на автомобільному транспорті)»
галузі знань 27 Транспорт
спеціальності 275.03 Транспортні технології (на автомобільному транспорті)
та освітньо-професійної програми «Автомобільний транспорт»
галузі знань 27 Транспорт
спеціальності 274 Автомобільний транспорт
денної форми навчання

Луцьк 2024

УДК 519.21 (042.4)
Б 83

До друку

Електронна копія друкованого видання передана для внесення в репозитарій
ТФК ЛНТУ

Бібліотекар _____ Оксана РОМАНЮК

Рекомендовано до видання Навчально-методичною радою ТФК ЛНТУ,
протокол

№ ___ від «___» _____ 2024 року.

Голова НМР _____ Світлана БУСНЮК

Розглянуто і схвалено на засіданні циклової комісії природничо-математичних
дисциплін ТФК ЛНТУ, протокол № ___ від «___» _____ 2024 року.

Голова ЦК _____ Неля Стефанська

Укладач: _____ Юлія Боровська, викладач ТФК ЛНТУ.

Рецензент: _____ Юрій Харкевич, кандидат фізико-математичних
наук, професор факультету інформаційних систем, фізики та математики ВНУ
імені Лесі Українки.

Б 83	Основи теорії ймовірностей та статистики [Текст]: конспект лекцій для здобувачів фахової передвищої освіти освітньо-професійної програми «Транспортні технології (на автомобільному транспорті)» галузі знань 27 Транспорт спеціальності 275.03 Транспортні технології (на автомобільному транспорті) та освітньо-професійної програми «Автомобільний транспорт» галузі знань 27 Транспорт спеціальності 274 Автомобільний транспорт денної форми навчання / уклад. Ю. В. Боровська. – Луцьк : ТФК Луцького НТУ, 2023. – 88 с.
------	---

Методичне видання складене відповідно до робочої програми з дисципліни «Основи теорії ймовірностей та статистики», вміщує теоретичні відомості з кожної теми та приклади розв'язань типових задач. Мета цієї розробки – допомогти здобувачам освіти зрозуміти новий матеріал на лекціях, при підготовці до практичних занять, краще засвоїти основні поняття та методи розв'язання задач з основ теорії ймовірностей та статистики при самостійній підготовці.

© Ю.В.Боровська, 2024

ЗМІСТ

РОЗДІЛ I. ТЕОРІЯ ЙМОВІРНОСТЕЙ.....	5
ТЕМА 1. ВИПАДКОВІ ПОДІЇ.....	5
Лекція 1. Основні поняття теорії ймовірностей. Простір елементарних подій	5
Лекція 2. Алгебра подій. Аксиоми теорії ймовірностей та їх наслідки. Класичне означення ймовірності.....	6
Формула Байєсса.....	11
Лекція 4. Повторювання випробувань. Формула Бернуллі.....	14
Лекція 5. Поліноміальна формула. Граничні теореми в схемі Бернуллі.....	16
ТЕМА 2. ВИПАДКОВІ ВЕЛИЧИНИ.....	20
Лекція 6. Основні поняття випадкових величин. Дискретні і неперервні випадкові величини. Щільність ймовірностей.....	20
Лекція 7. Числові характеристики випадкових величин та їх властивості.....	25
Лекція 8. Числові характеристики основних розподілів.....	29
Лекція 9. Багатовимірні випадкові величини. Система двох випадкових величин та її числові характеристики.....	32
ТЕМА 3. СТАТИСТИЧНІ РОЗПОДІЛИ ВИБІРОК.....	36
ТА ЇХ ЧИСЛОВІ ХАРАКТЕРИСТИКИ.....	36
Лекція 10. Предмет і методи статистики.....	36
Лекція 12. Вибіркові числові характеристики.....	43
Лекція 13. Інтервальний статистичний розподіл вибірки та його числові характеристики.....	47
Лекція 14. Двовимірний статистичний розподіл вибірки та його числові характеристики. Статистичне вивчення кореляційного зв'язку випадкових величин.....	51
Лекція 15. Точкові оцінки параметрів розподілу. Методи знаходження точкових оцінок.....	55
Лекція 16. Інтервальні оцінки параметрів розподілу. Довірчі інтервали для математичного сподівання при відомій дисперсії.....	58
Лекція 17. Довірчі інтервали для дисперсії.....	62
ТЕМА 5. СТАТИСТИЧНІ ГІПОТЕЗИ.....	64
Лекція 18. Статистичні гіпотези.....	64
Лекція 19. Перевірка правильності нульової гіпотези про рівність математичних сподівань.....	68
ДОДАТКИ.....	76
СПИСОК ЛІТЕРАТУРИ.....	86

ВСТУП

Теорія ймовірностей та статистика дедалі ширше застосовуються в багатьох галузях науки і техніки і є важливими складовими фундаментальної підготовки сучасних висококваліфікованих фахівців.

Конспект лекцій містить теоретичні відомості та приклади розв'язання задач до кожної з тем програми та ставить за мету допомогти здобувачу освіти оволодіти основними методами кількісної оцінки дії випадкових факторів, що впливають на безліч процесів; показати, як використовувати ці знання при плануванні, організації та управлінні виробництвом, аналізі технологічних процесів та прогнозуванні; виробити вміння самостійно вивчати навчальну та наукову літературу, навички моделювання випадкових явищ, процесів. Вивчення курсу дає можливість зрозуміти новий аспект явищ, що відбуваються навколо, причини рішень, які люди приймають щодо цих явищ, вже на новому, більш чіткому рівні.

Для реалізації цієї мети здобувач освіти повинен опанувати основні методи знаходження ймовірностей випадкових величин та їх числових характеристик; вміти застосовувати статистичні методи до обробки й аналізу даних і приймати на основі цього обґрунтовані рішення. Ці вміння та навички дадуть можливість добре засвоїти теоретичні і практичні основи теорії ймовірностей та статистики, що сприятиме формуванню професійних компетентностей у здобувачів освіти щодо ефективного розв'язання різноманітних завдань майбутньої професійної діяльності в умовах інформаційного суспільства.

РОЗДІЛ I. ТЕОРІЯ ЙМОВІРНОСТЕЙ

ТЕМА 1. ВИПАДКОВІ ПОДІЇ

Лекція 1. Основні поняття теорії ймовірностей. Простір елементарних подій

План

1. Класифікація подій
2. Простір елементарних подій

1. Класифікація подій

Математична наука, що вивчає закономірності масових випадкових явищ, називається *теорією ймовірностей*. Мова йде не про будь-які явища, а саме ті, що можуть бути багаторазово повторені при незмінних умовах і демонструють при цьому так звану статистичну стійкість, тобто відносна кількість появ даного явища прямує до сталого значення із збільшенням кількості спостережень.

Математична статистика – це розділ математики, в якому вивчаються методи обробки і аналізу експериментальних даних, одержаних в результаті досліджень масових випадкових явищ.

Послідовність операцій, що виконуються з додержанням певного комплексу умов, називають *експериментом* (дослідом, випробуванням). Наслідком будь-якого експерименту є *подія*. Події поділяються на *вірогідні*, *неможливі* та *випадкові*.

Якщо в результаті експерименту, проведеного з додержанням певного комплексу умов, деяка подія обов'язково настає, то вона називається *вірогідною*. Вірогідна подія позначається символом Ω («омега»).

Подія називається *неможливою*, якщо в результаті експерименту, проведеного з додержанням певного комплексу умов, вона не настає ніколи. Неможлива подія позначається символом \emptyset (порожня множина).

Подія називається *випадковою*, якщо за певного комплексу умов у результаті експерименту вона може настати або не настати залежно від дії численних несуттєвих факторів, врахувати які дослідник не в змозі.

Випадкові події позначають символами A, B, C, \dots або $A_1, A_2, A_3, \dots, A_k; B_1, B_2, \dots, B_n$.

Подія, що може відбутися внаслідок проведення однієї і лише однієї спроби (експерименту), називається *простою (елементарною) випадковою подією*.

Елементарні події позначаються ω_i ($i = 1, 2, 3, \dots$) і в теорії ймовірностей, так само як, скажімо, точка в геометрії, не поділяються на простіші складові.

2. Простір елементарних подій

Відправним поняттям теорії ймовірностей є *стохастичний експеримент* – дія, яку можна повторювати багато разів, спостерігаючи за її випадковими наслідками. Прикладами найпростішого стохастичного експерименту можуть бути: підкидання монети або грального кубика, витягання карти з колоди, дані

про народжуваність та смертність, влучання у мішень, тощо. Кожному експерименту з випадковими результатами (наслідками) відповідає певна множина Ω елементарних подій ω_i , кожна з яких може відбутися внаслідок його проведення: $\omega_i \in \Omega$. Множину всіх можливих результатів експерименту, які не можна розкласти на простіші, називають *простором елементарних подій*.

Приклад 1. Монету підкидають чотири рази. Побудувати простір елементарних подій для цього експерименту і такі випадкові події:

- 1) A — герб випаде двічі; 2) B — герб випаде не менш як тричі.

Розв'язання. Шуканий простір елементарних подій:

$\Omega = \{гггг, гггц, ггцг, гцгг, цггг, ггцц, гццг, гцгц, цггц, цггг, цгцц, цгцг, цгцг, ццгг, ццгц, ццгг, цццг, цццц\};$

- 1) $A = \{ггцц, цггг, гцгг, цггг, гцгг, цггг\};$

- 2) $B = \{гггг, гггц, гггг, гггг, цггг\}.$

Простір елементарних подій може бути як дискретним, так і неперервним. Якщо множина є зчисленною, тобто всі її елементи можна перелічити або принаймні пронумерувати (кожній елементарній події поставити у відповідність один і тільки один елемент нескінченної послідовності натуральних чисел $1, 2, 3, \dots$), то простір елементарних подій називають *дискретним*. Він може бути обмеженим і необмеженим. У протилежному випадку простір елементарних подій називають *неперервним*.

Простір елементарних подій є математичною моделлю певного ідеалізованого експерименту в тому розумінні, що будь-який можливий його наслідок описується однією і лише однією елементарною подією – наслідком експерименту.

Лекція 2. Алгебра подій. Аксиоми теорії ймовірностей та їх наслідки. Класичне означення ймовірності.

План

1. **Алгебра подій.**
2. **Повна група подій. Протилежні події.**
3. **Аксиоми теорії ймовірностей та їх наслідки.**
4. **Класичне означення ймовірності.**

1. Алгебра подій.

Між реальними подіями та підмножинами простору елементарних подій можна встановити взаємнооднозначну відповідність. Тому над подіями виконують ті ж операції, що й над множинами.

Додавання. Сумою двох подій A і B називається така подія $A \cup B$ ($A + B$), складається з елементарних подій, що належать A або B (тобто відбувається хоча б одна з подій A чи B). Подію $A \cup B$ схематично зображено на рис. 1 заштрихованою областю.

Операція $A \cup B$ називається *об'єднанням* цих подій.

Множення. Добутком двох подій A і B називається така подія $A \cap B$ (AB), яка складається з елементарних подій, що належать A і B (події A і B відбуваються одночасно).

Операція $A \cap B$ називається *перерізом* цих подій (рис. 2).

Віднімання. Різницею двох подій A і B називається така подія $A \setminus B$ ($A - B$), яка складається з елементарних подій, що належать події A і не належать події B (рис. 3).

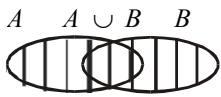


рис.1

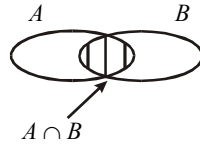


рис.2

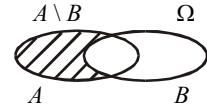


рис.3

Якщо $A \cap B \neq \emptyset$, то випадкові події A і B називають *сумісними*. Якщо $A \cap B = \emptyset$, то такі випадкові події A і B називають *несумісними*.

2. Повна група подій. Протилежні події

Якщо $A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n = \bigcup_{i=1}^n A_i = \Omega$, то такі випадкові події утворюють *повну групу*, а саме: внаслідок експерименту якась із подій A_i обов'язково настане.

Дві несумісні випадкові події, що утворюють повну групу, називають *протилежними*.

Подія, яка протилежна події A , позначається \bar{A} і складається з елементарних подій, що не належать A . Справедливі наступні співвідношення: $A \cup \bar{A} = \Omega$, $A \cap \bar{A} = \emptyset$.

Випадкові події A, B, C ($A \subset \Omega, B \subset \Omega, C \subset \Omega$), для яких визначено операції додавання, множення та віднімання, підлягають таким законам:

1. $A \cup A = A, A \cap A = A$.

2. $A \cup B = B \cup A$. | Комутативний закон для операцій

3. $A \cap B = B \cap A$. | додавання та множення.

4. $(A \cup B) \cup C = A \cup (B \cup C)$. | Асоціативний закон для

5. $(A \cap B) \cap C = A \cap (B \cap C)$. | операцій додавання та множення.

6. $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$. | Перший дистрибутивний закон.

7. $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$. | Другий дистрибутивний закон.

8. $A \cup \Omega = \Omega$.

13. $\bar{\Omega} = \emptyset$.

9. $A \cap \Omega = A$.

14. $\bar{\emptyset} = \Omega$.

10. $A \cup \emptyset = A$.

15. $A \cup (A \cap \bar{B}) = A; B = B \cup (B \cap \bar{A})$.

11. $A \cap \emptyset = \emptyset$.

16. $\overline{A \cup B} = \bar{A} \cap \bar{B}$.

12. $\bar{\bar{A}} = A$.

17. $\overline{A \cap B} = \bar{A} \cup \bar{B}$.

Елементарні випадкові події задовольняють наступні твердження: 1) між собою несумісні; 2) утворюють повну групу; 3) є рівноможливими, а саме: усі елементарні події мають однакові можливості відбутися внаслідок проведення одного експерименту.

Для кількісного вимірювання появи випадкових подій і їх комбінацій уводиться поняття ймовірності події, що є числом такої ж природи, як і відстань у геометрії або маса в теоретичній механіці.

3. Аксиоми теорії ймовірностей та їх наслідки.

Нехай Ω – множина елементів ω_i , які є елементарними подіями, а F – множина підмножин з Ω . Елементами множини F є випадкові події. Тоді F є алгеброю множин, що означає, що якщо $\Omega \in F$, то об'єднання, перетин чи різниця двох множин системи знов належить до цієї системи.

Множина F називається *алгеброю подій*, якщо:

1. $\Omega \in F$, тобто алгебра подій містить вірогідну подію.
2. Разом з будь-якою подією, алгебра подій містить протилежну їй подію, тобто якщо $A \in F$, то $\bar{A} \in F$.
3. Із того, що $A \in F, B \in F$, випливає: що $A \cap B \in F, A \cup B \in F, A \setminus B \in F$.

Якщо множина містить n елементів, то кількість усіх підмножин буде 2^n .

Числова функція P , що визначена на множині F , називається *ймовірністю*, якщо виконуються аксіоми:

I. F є алгеброю подій.

II. Кожній події A із множини F ($A \subset F$) поставлено у відповідність невід'ємне дійсне число $P(A)$, $P(A) \geq 0$.

III. Ймовірність вірогідної події дорівнює одиниці: $P(\Omega) = 1$.

IV. Якщо випадкові події A і B є несумісними ($A \cap B = \emptyset$), то ймовірність суми подій дорівнює сумі їх ймовірностей:

$$P(A \cup B) = P(A) + P(B) \quad (2.1)$$

Трійка (F, Ω, P) , де F є алгеброю подій і P задовольняє аксіоми, називається *простором ймовірностей*.

Наслідки аксіом

1. Аксіома IV поширюється на будь-яку скінченну кількість попарно несумісних випадкових подій $A_1, A_2, A_3, \dots, A_n$:

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) \text{ або } P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n).$$

2. Якщо випадкові події $A_1, A_2, A_3, \dots, A_n$ утворюють повну групу, $P\left(\bigcup_{i=1}^n A_i\right) = 1$, то формула додавання двох сумісних випадкових подій буде:

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

3. Формула додавання для трьох сумісних випадкових подій має такий вигляд:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) +$$

$$+ P(A \cap B \cap C).$$

4. Якщо випадкова подія A сприяє появі B ($A \subset B$), то $P(A) \leq P(B)$.

4. Класичне означення ймовірності

Ймовірністю випадкової події A називається невід'ємне число $P(A)$, що дорівнює відношенню числа елементарних подій m ($0 \leq m \leq n$), які сприяють появі A , до кількості всіх елементарних подій n простору Ω :

$$P(A) = \frac{m}{n} \quad (2.3)$$

Властивості ймовірності:

1. Ймовірність неможливої події дорівнює нулю: $P(\emptyset) = 0$.
2. Сума ймовірностей протилежних подій дорівнює 1, тому ймовірність протилежної події шукаємо за формулою:

$$P(\bar{A}) = 1 - P(A). \quad (2.4)$$

Для довільної випадкової події $0 < P(A) < 1$.

Приклад 1. Гральний кубик підкидають один раз. Яка ймовірність того, що на грані кубика з'явиться число, кратне 3?

Розв'язання. Число всіх елементарних подій для цього експерименту $n = 6$. Нехай B — поява на грані числа, кратного 3. Число елементарних подій, що сприяють появі B , дорівнює двом ($m = 2$).

$$\text{Отже, } P(B) = \frac{m}{n} = \frac{2}{6} = \frac{1}{3}.$$

Приклад 2. В урні білі, червоні і сині кульки. Ймовірність того, навмання взята кулька червона дорівнює 0,3; синя – 0,5; біла – 0,2. Знайти ймовірність того, що навмання взята кулька – кольорова (не біла).

Розв'язання. Подія A – навмання взята кулька червона; подія B – синя, C – біла. Тоді $P(A) = 0,3$; $P(B) = 0,5$; $P(C) = 0,2$.

Ймовірність появи кольорової кульки дорівнює ймовірності появи червоної чи синьої кульок. За теоремою додавання ймовірностей несумісних подій маємо:

$$P(A + B) = P(A) + P(B) = 0,3 + 0,5 = 0,8.$$

Приклад 3. В лотереї випущено 100000 квитків і встановлено 100 виграшів по 20 тис.грн., 1000 – по 10тис.грн., 5000 – по 2,5тис.грн і 10000 – по 500грн. Яка ймовірність того, що придбавши один квиток, можна виграти не менше 2,5тис.грн?

Розв'язання. Нехай подія A – виграш не менше 2,5 тис.грн., A_1 – виграш дорівнює 2,5 тис.грн., A_2 – виграш 10 тис.грн., A_3 – виграш 20 тис.грн. Оскільки куплено один білет і події попарно несумісні, то

$$P(A) = P(A_1) + P(A_2) + P(A_3) = \frac{5000}{100000} + \frac{1000}{100000} + \frac{100}{100000} = 0,061.$$

Приклад 4. Придбали касовий апарат. Ймовірність того, що придбали апарат, виробництва Японії – 0,4; виробництва США – 0,3; виробництва України – 0,1; іншого іноземного виробника – 0,2. Яка ймовірність того, що придбаний апарат іноземного виробництва?

Розв'язання. Подія A – придбали апарат іноземного виробництва; A_1 – виробництва Японії; A_2 – виробництва США; A_3 – іншого іноземного виробництва; A_4 – виробництва України. Тоді маємо:

$$P(A_1) = 0,4; P(A_2) = 0,3; P(A_3) = 0,2; P(A_4) = 0,1.$$

Події $A_1; A_2; A_3; A_4$ утворюють повну групу, отже $P(A_1+A_2+A_3) = 1 - P(A_4) = 0,9$.
Відповідь. 0,9.

Комбінаціями з n елементів по m ($0 \leq m \leq n$) називаються такі множини з m елементів, які різняться між собою хоча б одним елементом.

Кількість таких множин:

$$C_n^m = \frac{A_n^m}{P_m} = \frac{n!}{m!(n-m)} \quad (2.5)$$

Приклад 5. У цеху працює 10 верстатів-автоматів, кожен із яких може з певною ймовірністю перебувати в роботоздатному стані або в стані поломки. Яка ймовірність того, що під час роботи верстатів-автоматів із ладу вийдуть три з них?

Розв'язання. Оскільки кожен верстат-автомат може перебувати у двох несумісних станах — роботоздатному або нероботоздатному, то кількість усіх елементарних подій множини Ω буде $n = 2^{10}$.

Позначимо через A випадкову подію – із ладу вийде три верстати з десяти. Тоді кількість елементарних подій, що сприяють появі A , буде

$$m = C_{10}^3 = \frac{10!}{3!7!} = \frac{10 \cdot 9 \cdot 8 \cdot 7!}{1 \cdot 2 \cdot 3 \cdot 7!} = 120.$$

$$\text{Отже, } P(A) = \frac{m}{n} = \frac{C_{10}^3}{2^{10}} = \frac{120}{2^{10}}.$$

Лекція 3. Умовна ймовірність. Формула повної ймовірності.

Формула Байєсса

План

1. Умовна ймовірність та її властивості
2. Формули множення ймовірностей
3. Формула повної ймовірності
4. Формула Байєсса

1. Умовна ймовірність та її властивості

Випадкові події A і B називають *залежними*, якщо поява однієї з них (A або B) впливає на ймовірність появи іншої.

У протилежному випадку випадкові події A і B називаються *незалежними*.

Якщо ймовірність випадкової події A обчислюється за умови, що подія B відбулася, то така ймовірність називається *умовною*. Ця ймовірність обчислюється за формулою

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, P(B) \neq 0 \quad (3.1)$$

Аналогічно

$$P(B/A) = \frac{P(A \cap B)}{P(A)}, P(A) \neq 0 \quad (3.2)$$

Властивості умовної ймовірності:

1. $P(A/B) = 0$, якщо $A \cap B = \emptyset$.
2. $P(A/B) = 1$, якщо $A \cap B = B$.
3. У решті випадків $0 < P(A/B) < 1$.

Приклад 1. Задана множина цілих чисел. $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. Навмання беруть одне число. Яка ймовірність того, що це число виявиться кратним 3, коли відомо, що воно є непарним?

Розв'язання. Нехай подія A — поява числа кратного 3, B — непарного числа. Тоді маємо:

$A = (3, 6, 9, 12)$, $m_1 = 4$; $B = (1, 3, 5, 7, 9, 11)$, $m_2 = 6$; $A \cap B = (3, 9)$, $m_3 = 2$;

$$P(A) = \frac{m_1}{n} = \frac{4}{12} = \frac{1}{3}; P(B) = \frac{6}{12} = \frac{1}{2}; P(A \cap B) = \frac{m_3}{n} = \frac{2}{12} = \frac{1}{6};$$

$$P(A/B) = P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

Оскільки $P(A) \neq P(A/B)$, то події A і B є залежними

2. Формули множення ймовірностей

Згідно із (3.1) і (3.2) для двох залежних випадкових подій маємо:

$$P(A \cap B) = P(B) P(A/B) = P(A) P(B/A). \quad (3.3)$$

Формула множення для n залежних випадкових подій A_1, A_2, \dots, A_n :

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1) P(A_2 / A_1) P(A_3 / A_1 A_2) \dots P(A_n / A_1 A_2 \dots A_{n-1}) \quad (3.4)$$

Приклад 2. У ящику міститься 15 однотипних деталей. Із них 9 стандартні, решта — браковані. Деталі виймають по одній без повернення. Так було вийнято три деталі. Обчислити ймовірності таких випадкових подій:

- 1) A — три деталі виявляться стандартними;
- 2) B — усі три виявляться бракованими;
- 3) C — дві стандартні й одна бракована.

Розв'язання. Нехай A_i — поява стандартної, \bar{A}_i — бракованої деталі при i -му вийманні.

$$\text{Подія } A = A_1 \cap A_2 \cap A_3, \quad B = \bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3,$$

$$C = (A_1 \cap A_2 \cap \bar{A}_3) \cup (A_1 \cap \bar{A}_2 \cap A_3) \cup (\bar{A}_1 \cap A_2 \cap A_3).$$

Оскільки випадкові події A_i, \bar{A}_i є залежними, то:

$$P(A) = P(A_1 \cap A_2 \cap A_3) = P(A_1) P(A_2 / A_1) P(A_3 / A_1 A_2) = \frac{9}{15} \frac{8}{14} \frac{7}{13} = \frac{12}{65};$$

$$P(B) = P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) = P(\bar{A}_1) P(\bar{A}_2 / \bar{A}_1) P(\bar{A}_3 / \bar{A}_1 \bar{A}_2) = \frac{6}{15} \frac{5}{14} \frac{4}{13} = \frac{4}{91}.$$

$$\begin{aligned} P(C) &= P\left((A_1 \cap A_2 \cap \bar{A}_3) \cup (A_1 \cap \bar{A}_2 \cap A_3) \cup (\bar{A}_1 \cap A_2 \cap A_3)\right) = \\ &= P(A_1 \cap A_2 \cap \bar{A}_3) + P(A_1 \cap \bar{A}_2 \cap A_3) + P(\bar{A}_1 \cap A_2 \cap A_3) = \\ &= P(A_1) P(A_2 / A_1) P(\bar{A}_3 / A_1 A_2) + P(A_1) P(\bar{A}_2 / A_1) P(A_3 / A_1 \bar{A}_2) + \\ &+ P(\bar{A}_1) P(A_2 / \bar{A}_1) P(A_3 / \bar{A}_1 A_2) = \frac{9}{15} \frac{8}{14} \frac{6}{13} + \frac{9}{15} \frac{6}{14} \frac{8}{13} + \frac{6}{15} \frac{9}{14} \frac{8}{13} = \frac{216}{455}. \end{aligned}$$

Якщо випадкові події A і B є незалежними, то $P(A / B) = P(A)$, $P(B / A) = P(B)$.

Формули (3.3), (3.4) для незалежних випадкових подій наберуть такого вигляду:

$$P(A \cap B) = P(A) P(B); \quad P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i).$$

Приклад 3. Гральний кубик і монету підкидають по одному разу. Яка ймовірність того, що при цьому на грані кубика випаде число, кратне 3, а на монеті герб?

Розв'язання. Нехай поява числа, кратного трьом — подія A , а поява герба — подія B . Випадкові події A і B є між собою незалежними. Отже,

$$P(A) = \frac{2}{6} = \frac{1}{3}; \quad P(B) = \frac{1}{2}; \quad P(A \cap B) = P(A) P(B) = \frac{1}{3} \frac{1}{2} = \frac{1}{6}.$$

3. Формула повної ймовірності

У разі, коли випадкова подія A може відбутися лише за умови, що відбудеться одна з несумісних випадкових подій B_i , які утворюють повну групу і між собою є попарно несумісними $\left(B_i \cap B_j = \emptyset, i \neq j, i, j = 1, \dots, n, \bigcup_{i=1}^n B_i = \Omega \right)$, імовірність події A обчислюється за формулою

$$P(A) = \sum_{i=1}^n P(B_i) P(A / B_i) \quad (3.5)$$

яка називається *формулою повної ймовірності*.

Випадкові події B_1, B_2, \dots, B_n називають *гіпотезами*.

Приклад 4. До складального цеху надходять деталі від трьох інших цехів. Від першого надходить 45% усіх деталей, від другого — 35% і від третього — 20%. Перший цех допускає в середньому 6% браку, другий — 2% і третій — 8%.

Яка ймовірність того, що до складального цеху надійде стандартна деталь?

Розв'язання. Позначимо через A появу стандартної деталі, подія B_1 — деталь надійде від першого цеху, B_2 — від другого, B_3 — від третього. За умовою задачі:

$$P(B_1) = 0,45, \quad P(A / B_1) = 0,94;$$

$$P(B_2) = 0,35, \quad P(A / B_2) = 0,98;$$

$$P(B_3) = 0,2, \quad P(A / B_3) = 0,92.$$

Згідно з (3.5) маємо:

$$P(A) = P(B_1) P(A/B_1) + P(B_2) P(A/B_2) + P(B_3) P(A/B_3) = 0,45 \cdot 0,94 + 0,35 \cdot 0,98 + 0,2 \cdot 0,92 = 0,423 + 0,343 + 0,184 = 0,95.$$

4. Формула Байєса

Застосовуючи формулу множення ймовірностей для залежних випадкових подій $A, B_i (i = \overline{1, n})$, дістаємо $P(A) P(B_i / A) = P(B_i) P(A / B_i)$

$$P(B_i / A) = \frac{P(B_i) P(A / B_i)}{P(A)} = \frac{P(B_i) P(A / B_i)}{\sum_{i=1}^n P(B_i) P(A / B_i)} \quad (3.6)$$

Залежність (5.6) називається *формулою Байєса*. Її використовують для переоцінювання ймовірностей гіпотез B_i за умови, що випадкова подія A здійсниться.

Після переоцінювання всіх гіпотез B_i маємо:

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

Згідно з формулою Байєса можна прийняти рішення, провівши експеримент. Але для цього необхідно, аби вибір тієї чи іншої гіпотези мав

ґрунтовні підстави, тобто щоб унаслідок проведення експерименту ймовірність $P(B_i/A)$ була близька до одиниці.

Приклад 5. В магазині три холодильники, в яких закінчується морозиво. У першому – 4 білих морозива і 6 шоколадних, у другому – 2 білих і 8 шоколадних, у третьому – 3 білих і 7 шоколадних. Навмання вибирають холодильник і виймають з нього біле морозиво. Визначити ймовірність того, що морозиво витягнули з другого холодильника.

Розв'язання. Позначимо через подію A вибір білого морозива, події B_i — вибір i -го холодильника. Оскільки нас цікавить, щоб біле морозиво було саме з другого холодильника, то шукана ймовірність $P(B_2/A)$ за формулою (3.6) матиме вигляд:

$$P(B_2/A) = \frac{P(B_2)P(A/B_2)}{P(A)}.$$

Тоді ймовірності витягнути морозиво з кожного з трьох холодильників однакові і рівні $P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}$.

Ймовірності витягнути біле морозиво будуть рівні кількості морозиват в кожному з холодильників, розділену на загальну кількість морозива:

$$P(A/B_1) = \frac{4}{4+6} = 0,4; \quad P(A/B_2) = \frac{2}{2+8} = 0,2, \quad P(A/B_3) = \frac{3}{3+7} = 0,3.$$

Знайшовши повну ймовірність:

$$P(A) = P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + P(B_3)P(A/B_3) = \\ \frac{1}{3} \cdot 0,4 + \frac{1}{3} \cdot 0,2 + \frac{1}{3} \cdot 0,3 = 0,3,$$

маємо:

$$P(B_2/A) = \frac{\frac{1}{3} \cdot 0,2}{0,3} = \frac{2}{9}.$$

Лекція 4. Повторювання випробувань. Формула Бернуллі.

План

1. Формула Бернуллі

2. Найімовірніше число появи випадкової події

Якщо кожний експеримент має лише два несумісні наслідки (події) зі сталими ймовірностями p і q , то їх називають експериментами за схемою Бернуллі. У кожному експерименті випадкова подія з імовірністю p відбувається, а з імовірністю q – не відбувається, тобто $p + q = 1$.

Простір елементарних подій для одного експерименту містить дві елементарні події, а для n експериментів за схемою Бернуллі – 2^n елементарних подій.

1. Формула Бернуллі

Імовірність того, що в результаті n незалежних експериментів за схемою Бернуллі подія A з'явиться m раз, подається у вигляді

$$P_n(m) = C_n^m p^m q^{n-m} = \frac{n!}{m!(n-m)!} p^m q^{n-m} \quad (4.1)$$

Імовірність того, що в результаті n незалежних експериментів подія A з'явиться від m_i до m_j раз, обчислюється за формулою:

$$P_n(m_i \leq m \leq m_j) = \sum_{m=m_i}^{m_j} C_n^m p^m q^{n-m} = \sum_{m=m_i}^{m_j} \frac{n!}{m!(n-m)!} p^m q^{n-m} \quad (4.2)$$

Оскільки

$$P_n(0 \leq m \leq n) = \sum_{m=0}^n C_n^m p^m q^{n-m} = 1 \quad (4.3)$$

Дістанемо

$$P(0 \leq m \leq m_i) = 1 - \sum_{m=m_i+1}^n C_n^m p^m q^{n-m} \quad (4.4)$$

$$P(m_i \leq m \leq n) = 1 - \sum_{m=0}^{m_i-1} C_n^m p^m q^{n-m} \quad (4.5)$$

Приклад 1. Імовірність того, що електролампочка не перегорить при ввімкненні її в електромережу, є величиною сталою і дорівнює 0,9.

Обчислити ймовірність того, що з п'яти електролампочок, увімкнених у електромережу за певною схемою, не перегорять: 1) дві; 2) не більш як дві; 3) не менш як дві.

Розв'язання. За умовою задачі маємо: $p = 0,9$; $q = 0,1$; $n = 5$; $m = 2$. Згідно з (4.1), (4.4), (4.5) дістанемо:

$$1) P_5(2) = C_5^2 p^2 q^3 = \frac{5!}{2! 3!} (0,9)^2 (0,1)^3 = 10 \cdot 0,81 \cdot 0,001 = 0,0081;$$

$$\begin{aligned} 2) P_5(0 \leq m \leq 2) &= \sum_{m=0}^2 C_5^m p^m q^{5-m} = C_5^0 p^0 q^5 + C_5^1 p q^4 + C_5^2 p^2 q^3 = \\ &= q^5 + 5p q^4 + 10p^2 q^3 = (0,1)^5 + 5 \cdot 0,9 (0,1)^4 + 10 (0,9)^2 (0,1)^3 = \\ &= 0,00001 + 5 \cdot 0,9 \cdot 0,0001 + 10 \cdot 0,81 \cdot 0,001 = \\ &= 0,00001 + 0,00045 + 0,0081 = 0,00856; \end{aligned}$$

$$\begin{aligned} 3) P_5(2 \leq m \leq 5) &= \sum_{m=2}^5 C_5^m p^m q^{5-m} = 1 - \sum_{m=0}^1 C_5^m p^m q^{5-m} = \\ &= 1 - C_5^0 p^0 q^5 - C_5^1 p q^4 = 1 - (0,00001 + 0,00045) = 1 - 0,00046 = 0,99954. \end{aligned}$$

2. Найімовірніше число появи випадкової події

Найімовірнішим числом появи випадкової події A в результаті n незалежних експериментів за схемою Бернуллі називається таке число m_0 , для якого ймовірність $P_n(m_0)$ перевищує або в усякому разі є не меншою за ймовірність кожного з решти можливих наслідків експериментів.

$$np - q \leq m_0 \leq np + p \quad (4.6)$$

Число m_0 називають також *модю*.

Приклад 2. Імовірність того, що студент складе іспит з математики, є величиною сталою і дорівнює в середньому 0,8. Нехай є група з восьми студентів. Знайти найімовірнішу кількість членів цієї групи котрі складуть іспит з математики, і обчислити відповідну ймовірність.

Розв'язання. За умовою задачі $n = 8$; $p = 0,8$; $q = 0,2$.

$$np - q \leq m_0 \leq np + p \rightarrow$$

$$\rightarrow 8 \cdot 0,8 - 0,2 \leq m_0 \leq 8 \cdot 0,8 + 0,8 \rightarrow$$

$$\rightarrow 6,2 \leq m_0 \leq 7,2.$$

Отже, $m_0 = 7$; $P_8(7) = C_8^7 p^7 q = 8 (0,8)^7 0,2 = 1,6 (0,8)^7 = 0,524288$.

Висновок: найімовірніша кількість студентів, які складуть екзамен, $m_0 = 7$. Відповідна ймовірність дорівнює 0,524288.

Приклад 3. Яку найменшу кількість разів потрібно підкинути гральний кубик, щоб найімовірніше число випадання п'яти очок було рівне 6?

Розв'язання. Ймовірність випадання п'яти очок вважаємо незалежною подією з ймовірністю появи $p = \frac{1}{6}$, оскільки граней 6 і тільки на одній п'ятірці.

Протилежна подія матиме ймовірність $q = 1 - p = 1 - \frac{1}{6} = \frac{5}{6}$. За формулою (4.6)

отримаємо подвійну нерівність:

$$n \cdot \frac{1}{6} - \frac{5}{6} \leq 6 \leq n \cdot \frac{1}{6} + \frac{1}{6}.$$

Запишемо її у вигляді системи двох нерівностей:

$$\left\{ \begin{array}{l} \frac{1}{6}n - \frac{5}{6} \leq 6; \\ \frac{1}{6}n + \frac{1}{6} \geq 6; \end{array} \right. \left\{ \begin{array}{l} n \leq \frac{6 + \frac{5}{6}}{\frac{1}{6}}; \\ n \geq \frac{6 - \frac{1}{6}}{\frac{1}{6}}; \end{array} \right. \left\{ \begin{array}{l} n \leq 41; \\ n \geq 35. \end{array} \right.$$

Найменшим числом із заданого проміжку є 35. Отже, гральний кубик потрібно підкинути мінімум 35 разів.

Лекція 5. Поліноміальна формула. Граничні теореми в схемі Бернуллі.

План

1. Поліноміальна формула
2. Локальна теорема
3. Інтегральна теорема

1. Поліноміальна формула

Як відмічалось раніше, схема Бернуллі є послідовністю незалежних випробувань з двома можливими результатами. При цьому в кожному випробуванні подія A може з'явитись з однією і тією ж ймовірністю p , а подія \bar{A} – з ймовірністю $q = 1 - p$.

За поліноміальною схемою здійснюється перехід від послідовності незалежних випробувань з двома можливими результатами (A та \bar{A}) до послідовності незалежних випробувань з k взаємовиключними результатами A_1, A_2, \dots, A_k . При цьому в кожному випробуванні події A_1, A_2, \dots, A_k з'являються відповідно з ймовірностями p_1, p_2, \dots, p_k . Тоді ймовірність $P_n(m_1, m_2, \dots, m_k)$ того, що в n незалежних випробуваннях подія A_1 відбудеться m_1 разів, A_2 – m_2 раз і т.д., подія A_k – m_k разів ($m_1 + m_2 + \dots + m_k = n$), визначиться за формулою:

$$P_n(m_1, m_2, \dots, m_k) = \frac{n!}{m_1! \cdot m_2! \cdot \dots \cdot m_k!} \cdot p_1^{m_1} \cdot p_2^{m_2} \cdot \dots \cdot p_k^{m_k}. \quad (7.1)$$

2. Локальна теорема

Якщо ймовірність появи випадкової події в кожному з n незалежних експериментів є величиною сталою і дорівнює p ($0 < p < 1$), то для великих значень n і m ймовірність того, що випадкова подія A настане m раз, подається такою асимптотичною формулою:

$$P_n(m) \approx \frac{\varphi(x)}{\sqrt{npq}} \quad (7.2)$$

де $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ називається *функцією Гаусса*. Функція Гаусса протабульована, і її значення наведено в дод. 1, де

$$x = \frac{m - np}{\sqrt{npq}} \quad (7.3)$$

Тут x є рівномірно обмеженою величиною відносно n і m .

Властивості функції Гаусса:

- 1) $\varphi(x)$ визначена на всій осі абсцис; $\varphi(x) > 0$;
- 2) $\varphi(x)$ є функцією парною: $\varphi(-x) = \varphi(x)$;
- 3) $\lim_{x \rightarrow \pm\infty} \varphi(x) = 0$;

- 4) $\varphi'(x) = -x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$; $\varphi'(0) = 0$;

$\varphi'(x)|_{x < 0} > 0$; $\varphi'(x)|_{x > 0} < 0$; отже, $\varphi(0) = \frac{1}{\sqrt{2\pi}}$ — максимум функції Гаусса;

- 5) $\varphi''(x) = (x^2 - 1) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ $\varphi''(x)|_{x=\pm 1} = 0$.

Графік функції Гаусса зображено на рис. 4.

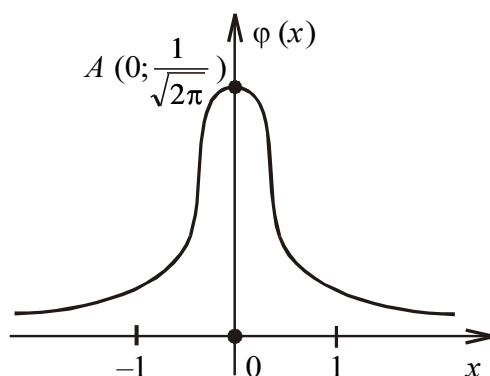


Рис. 4

Приклад 1. Фабрика випускає 75% виробів 1-го сорту. Із партії готових виробів навмання беруть 400 деталей. Обчислити ймовірності таких випадкових подій:

- 1) виробів 1-го сорту виявиться 290 шт.;
- 2) 300 шт.;
- 3) 320 шт.

Розв'язання. За умовою задачі маємо:

$$n = 400; p = 0,75; q = 0,25; m = 290; 300; 320.$$

$$1) \sqrt{npq} = \sqrt{400 \cdot 0,75 \cdot 0,25} = \sqrt{75} \approx 8,7; np = 400 \cdot 0,75 = 300;$$

$$x = \frac{m - np}{\sqrt{npq}} = \frac{290 - 300}{8,7} = -1,15;$$

$$P_{400}(290) \approx \frac{\varphi(-1,15)}{8,7} = \frac{\varphi(1,15)}{8,7} = \frac{0,2059}{8,7} \approx 0,0237;$$

$$2) x = \frac{m - np}{\sqrt{npq}} = \frac{300 - 300}{8,7} = 0;$$

$$P_{400}(300) \approx \frac{\varphi(0)}{8,7} = \frac{0,3989}{8,7} \approx 0,046;$$

$$3) x = \frac{m - np}{\sqrt{npq}} = \frac{320 - 300}{8,7} = \frac{20}{8,7} \approx 2,3;$$

$$P_{400}(320) \approx \frac{\varphi(2,3)}{8,7} \approx \frac{0,0283}{8,7} \approx 0,0033.$$

3. Інтегральна теорема

Якщо ймовірність появи випадкової події в кожному з n незалежних експериментів є величиною сталою і дорівнює p ($0 < p < 1$), то для великих значень n імовірність появи випадкової події від m_i до m_j раз обчислюється за такою асимптотичною формулою:

$$P_n(m_i \leq m \leq m_j) \approx \Phi(x_j) - \Phi(x_i) \quad (5.1)$$

де $x_j = \frac{m_j - np}{\sqrt{npq}}$, $x_i = \frac{m_i - np}{\sqrt{npq}}$, а $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{x^2}{2}} dx$ є функцією Лапласа, значення якої наведено в дод. 2.

Властивості функції Лапласа:

1. $\Phi(x)$ визначена на всій осі абсцис.
2. $\Phi(-x) = -\Phi(x)$, отже, $\Phi(x)$ є непарною функцією.
3. $\Phi(0) = 0$.

4. $\Phi(\infty) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{x^2}{2}} dx = 0,5$, оскільки $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$ є інтегралом

Пуассона.

5. $\Phi(-\infty) = -0,5$, як непарна функція.
6. $\Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} > 0$, отже, $\Phi(x)$ є функцією неспадною.
7. $\Phi''(0) = 0$; $\Phi''(x)|_{x<0} > 0$; $\Phi''(x)|_{x>0} < 0$.

Таким чином, $x = 0$ є точкою перегину.

Графік функції $\Phi(x)$ зображено на рис. 5

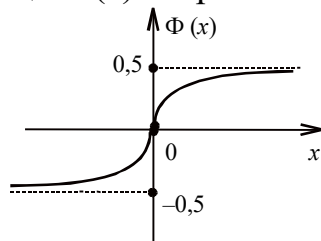


Рис. 5

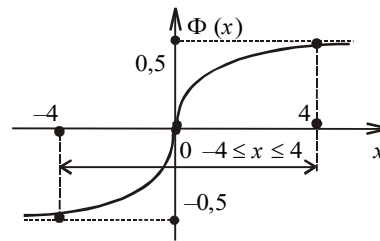


Рис. 6

Розв'язуючи задачі, додержують такого правила:

$$\Phi(x)|_{x \geq 4} \approx 0,5, \quad \Phi(x)|_{x \leq -4} \approx -0,5.$$

Отже, практично функція Лапласа застосовується для значень $x \in [-4; 4]$, що ілюструє рис. 6.

Приклад 2. Верстат-автомат виготовляє однотипні деталі. Імовірність того, що виготовлена одна деталь виявиться стандартною, є величиною сталою і дорівнює 0,95. За зміну верстатом було виготовлено 800 деталей. Яка ймовірність того, що стандартних деталей серед них буде: 1) від 720 до 780 шт.; 2) від 740 до 790 шт.?

Розв'язання. За умовою задачі:

$$n = 800; \quad p = 0,95; \quad q = 0,05; \quad 720 \leq m \leq 780; \quad 740 \leq m \leq 780;$$

$$\sqrt{npq} = \sqrt{800 \cdot 0,95 \cdot 0,05} = \sqrt{38} \approx 6,2, \quad np = 800 \cdot 0,95 = 760.$$

$$1) \quad x_j = \frac{m_j - np}{\sqrt{npq}} = \frac{780 - 760}{6,2} = \frac{20}{6,2} \approx 3,23;$$

$$x_i = \frac{m_i - np}{\sqrt{npq}} = \frac{720 - 760}{6,2} = -\frac{40}{6,2} \approx -6,5;$$

$$P_{800}(720 \leq m \leq 780) \approx \Phi(3,23) - \Phi(-6,5) = \Phi(3,23) + \Phi(6,5) = 0,49931 + 0,5 = 0,99931;$$

$$2) x_j = \frac{m_j - np}{\sqrt{npq}} = \frac{790 - 760}{6,2} \approx 4,84;$$

$$x_i = \frac{m_i - np}{\sqrt{npq}} = \frac{740 - 760}{6,2} = -\frac{20}{6,2} \approx -3,23;$$

$$P_{800}(740 \leq m \leq 790) \approx \Phi(4,84) - \Phi(-3,23) = \Phi(4,84) + \Phi(3,23) = 0,5 + 0,499 \cdot 31 = 0,99931.$$

ТЕМА 2. ВИПАДКОВІ ВЕЛИЧИНИ

Лекція 6. Основні поняття випадкових величин. Дискретні і неперервні випадкові величини. Щільність ймовірностей

План

1. Дискретні та неперервні випадкові величини
2. Закон розподілу випадкової величини
3. Функція розподілу ймовірностей
4. Щільність ймовірностей

1. Дискретні та неперервні випадкові величини

Розглянемо такий простір елементарних подій, в якому кожній елементарній події $\omega_i \in \Omega$ відповідає одне і лише одне число x або набір чисел (x_1, x_2, \dots, x_k) , тобто на множині Ω визначена певна функція $\alpha(\omega_i)$, яка кожній елементарній події ω_i ставить у відповідність певний елемент одновимірного простору R_1 або n -вимірного простору R_n .

Цю функцію називають *випадковою величиною*. У разі, коли $\alpha(\omega_i)$ відображає множину Ω на одновимірний простір R_1 , випадкову величину називають *одновимірною*. Якщо відображення здійснюється на R_n , то випадкову величину називають *n -вимірною* (системою n випадкових величин або n -вимірним випадковим вектором).

Схематично одновимірну випадкову величину унаочнює рис. 7.

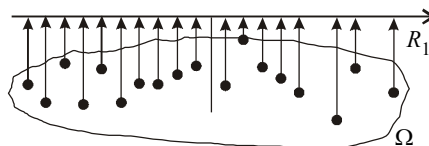


Рис. 7

Отже, величина називається *випадковою*, якщо внаслідок проведення експерименту під впливом випадкових факторів вона набуває того чи іншого можливого числового значення з певною ймовірністю.

Якщо множина можливих значень випадкової величини є зчисленною то таку величину називають *дискретною*. У протилежному разі її називають *неперервною*.

Випадкові величини позначають великими літерами латинського алфавіту X, Y, Z, \dots , а їх можливі значення — малими $x; y; z, \dots$.

2. Закон розподілу випадкової величини

Для опису випадкової величини необхідно навести не лише множину можливих її значень, а й указати, з якими ймовірностями ця величина набуває того чи іншого можливого значення.

З цією метою вводять поняття закону розподілу ймовірностей.

Співвідношення, що встановлює зв'язок між можливими значеннями випадкової величини та відповідними їм імовірностями, називають *законом розподілу випадкової величини*.

Закон розподілу дискретної випадкової величини X можна задати в табличній формі або за допомогою ймовірнісного многокутника.

У разі табличної форми запису закону подається послідовність можливих значень випадкової величини X , розміщених у порядку зростання, та відповідних їм імовірностей:

$X = x_i$	x_1	x_2	x_3	x_k
$P(X = x_i) = p_i$	p_1	p_2	p_3	p_k

Оскільки випадкові події є між собою несумісними і утворюють повну групу, то необхідною є *умовою нормування*:

$$\sum_{j=1}^k P(X = x_j) = \sum_{j=1}^k p_j = 1. \quad (6.1)$$

Закон розподілу ймовірностей можна уявити графічно. Для цього візьмемо систему координат, відклавши на осі абсцис можливі значення випадкової величини x_i , а на осі ординат — імовірності p_i цих можливих значень. Точки з координатами $(x_i; p_i)$ послідовно сполучимо відрізками прямої. Утворену при цьому фігуру називають імовірнісним многокутником.

Приклад 1. За заданим у табличній формі законом розподілу дискретної випадкової величини X :

$X = x_i$	-2,5	1	3,5	5	6,5	8
$P(X = x_i) = p_i$	0,1	0,2	0,1	0,3	0,2	0,1

побудувати ймовірнісний многокутник.

Розв'язання. Імовірнісний многокутник зображено на рис. 8.

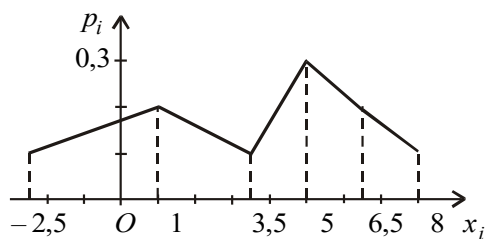


Рис. 8

Сума ординат імовірнісного багатокутника завжди дорівнює одиниці.

3. Функція розподілу

Закон розподілу ймовірностей можна подати ще в одній формі, яка придатна і для дискретних, і для неперервних випадкових величин, а саме: як функцію розподілу ймовірностей випадкової величини $F(x)$, так звану інтегральну функцію.

Функцію аргументу x , що визначає ймовірність випадкової події $X < x$, називають *функцією розподілу ймовірностей*:

$$F(x) = P(X < x) \quad (6.2)$$

Цю функцію можна тлумачити так: унаслідок експерименту випадкова величина може набути значення, меншого за x .

Наприклад, $F(5) = P(X < 5)$ означає, що в результаті експерименту випадкова величина X (дискретна чи неперервна) може набути значення, яке міститься ліворуч від $x = 5$, що ілюструє рис. 9.

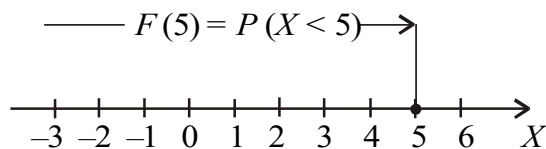


Рис. 9

Розглянемо властивості $F(x)$:

1. $0 \leq F(x) \leq 1$.
2. $F(x)$ є неспадною функцією, а саме $F(x_2) \geq F(x_1)$, якщо $x_2 > x_1$.

Приклад 1. Закон розподілу дискретної випадкової величини X задано таблицею:

$X = x_i$	-4	-1	2	6	9	13
$P(X=x_i)=p_i$	0,1	0,2	0,1	0,3	0,1	0,2

Побудувати $F(x)$ та її графік.

Розв'язання. Згідно з властивостями $F(x)$, дістаємо наведені далі співвідношення.

- 1) $F(-4) = P(X < -4) = 0$;
- 2) $F(-1) = P(X < -1) = P(X = -4) = 0,1$;
- 3) $F(2) = P(X < 2) = P(X = -4) + P(X = -1) = 0,1 + 0,2 = 0,3$;
- 4) $F(6) = P(X < 6) = P(X = -4) + P(X = -1) + P(X = 2) = 0,1 + 0,2 + 0,1 = 0,4$;

- 5) $F(9) = P(X < 9) = P(X = -4) + P(X = -1) + P(X = 2) + P(X = 6) = 0,1 + 0,2 + 0,1 + 0,3 = 0,7$;
- 6) $F(12) = P(X < 13) = P(X = -4) + P(X = -1) + P(X = 2) + P(X = 9) = 0,1 + 0,2 + 0,1 + 0,3 + 0,1 = 0,8$;
- 7) $F(x)|_{x > 13} = P(X > 13) = P(X = -4) + P(X = -1) + P(X = 2) + P(X = 9) + P(X = 13) = 0,1 + 0,2 + 0,1 + 0,1 + 0,3 + 0,1 + 0,2 = 1$.

Компактно $F(x)$ можна записати в такій формі:

$$F(x) = P(X < x) = \begin{cases} 0, & x \leq -4; \\ 0,1, & -4 < x \leq -1; \\ 0,3, & -1 < x \leq 2; \\ 0,4, & 2 < x \leq 6; \\ 0,7, & 6 < x \leq 9; \\ 0,8, & 9 < x \leq 12; \\ 1, & x > 12. \end{cases}$$

Графік функції $F(x)$ зображено на рис. 10.

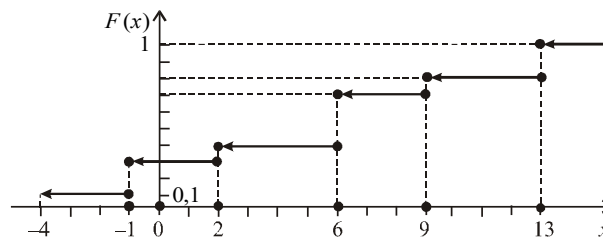


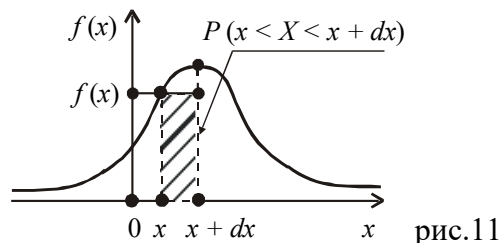
Рис. 10

4. Щільність ймовірностей

Для неперервних випадкових величин закон розподілу ймовірностей зручно описувати з допомогою щільності ймовірностей, яку позначають $p(x)$.

Щільністю ймовірностей неперервної випадкової величини X називається перша похідна від інтегральної функції $F(x)$.

Геометрично на графіку щільності ймовірності $p(x)dx$ відповідає площа прямокутника з основою dx і висотою $p(x)$ (рис. 11).



Властивості щільності:

1. $p(x) \geq 0$. Ця властивість впливає з означення щільності ймовірності як першої похідної від $F(x)$ за умови, що $F(x)$ є неспадною функцією.
2. Умова нормування неперервної випадкової величини X :

$$\int_{-\infty}^{\infty} p(x) dx = 1. \quad (6.3)$$

3. Імовірність попадання неперервної випадкової величини в інтервалі $[\alpha; \beta]$ обчислюється за формулою

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} p(x) dx. \quad (6.4)$$

4. Функція розподілу ймовірностей неперервної випадкової величини має вигляд

$$F(x) = \int_{-\infty}^x p(x) dx. \quad (6.5)$$

Приклад 2. Закон неперервної випадкової величини X задано у вигляді:

$$p(x) = \begin{cases} 0, & x \leq 0; \\ \frac{1}{2} \sin x, & 0 < x \leq \pi; \\ 0, & x > \pi. \end{cases}$$

Знайти $F(x)$ і побудувати графіки функцій $p(x)$, $F(x)$. Обчислити $P\left(\frac{\pi}{6} < X < \frac{\pi}{4}\right)$.

Розв'язання. Згідно 4 властивості маємо:

$$\begin{aligned} F(x) &= \int_0^x f(x) dx = \int_0^x \frac{1}{2} \sin dx = \frac{1}{2} \int_0^x \sin dx = \frac{1}{2} (-\cos x \Big|_0^x) = \frac{1}{2} (-\cos x + 1) = \\ &= \frac{1 - \cos x}{2}. \end{aligned}$$

Отже, функція розподілу ймовірностей буде така:

$$F(x) = \begin{cases} 0, & x \leq 0; \\ \frac{1}{2}(1 - \cos x), & 0 < x \leq \pi; \\ 1, & x > \pi. \end{cases}$$

Графіки функцій $p(x)$, $F(x)$ зображені відповідно на рис. 12 і 13.

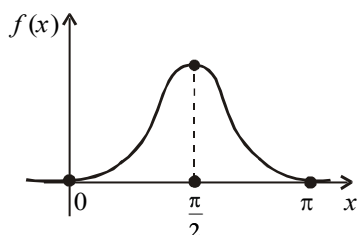


Рис. 12

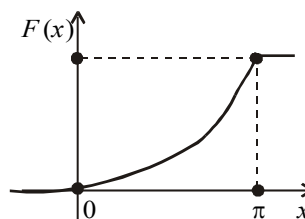


Рис. 13

Імовірність події $\frac{\pi}{6} < X < \frac{\pi}{4}$ можна обчислити згідно 3 властивості:

$$P\left(\frac{\pi}{6} < X < \frac{\pi}{4}\right) = \int_{\frac{\pi}{6}}^{\frac{\pi}{4}} \sin x \, dx = -\cos x \Big|_{\frac{\pi}{6}}^{\frac{\pi}{4}} = -\cos \frac{\pi}{4} + \cos \frac{\pi}{6} = -\frac{\sqrt{2}}{2} + \frac{\sqrt{3}}{2} = \frac{\sqrt{3} - \sqrt{2}}{2}.$$

Лекція 7. Числові характеристики випадкових величин та їх властивості

План

1. Математичне сподівання
2. Мода та медіана випадкової величини
3. Дисперсія та середнє квадратичне відхилення
4. Початкові та центральні моменти
5. Асиметрія і ексцес

1. Математичне сподівання

Термін «математичне сподівання» випадкової величини X є синонімом терміна «середнє значення» випадкової величини X .

Математичним сподіванням випадкової величини X , визначеної на дискретному просторі Ω (Ω – обмежена множина), називається величина

$$M(X) = \sum_{i=1}^n x_i p_i = x_1 p_1 + x_2 p_2 + \dots + x_n p_n \quad (7.1)$$

Якщо простір Ω є неперервним, то математичним сподіванням неперервної випадкової величини X називається величина

$$M(X) = \int_{\Omega} x f(x) dx$$

Якщо $\Omega = (-\infty; \infty)$, то

$$M(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Якщо $\Omega = [a; b]$, то

$$M(X) = \int_a^b x f(x) dx.$$

Властивості математичного сподівання

1. Математичне сподівання від сталої величини C дорівнює самій сталій:

$$M(C) = C.$$

- 2.

$$M(CX) = CM(X).$$

3. Якщо A і B є сталими величинами, то

$$M(AX + B) = AM(X) + B.$$

Приклад 1. Випадкові величини X та Y незалежні і розподілені таким чином

$X=x_i$	2	7	9		$Y=y_j$	4	6
p_i	0,1	0,4	0,5		p_j	0,7	0,3

Знайти $M(Z)$, якщо $Z = 2XY + X^2 + 3Y - 71$.

Розв'язання. Скориставшись властивостями математичного сподівання, маємо:

$$\begin{aligned} M(Z) &= M(2XY + X^2 + 3Y - 71) = M(2XY) + M(X^2) + M(3Y) + M(-71) = \\ &= 2M(XY) + M(X^2) + 3M(Y) - 71 = 2M(X)M(Y) + M(X^2) + 3M(Y) - 71. \end{aligned}$$

Для розрахунку $M(Z)$ необхідно знайти $M(X)$, $M(Y)$, $M(X^2)$.

Згідно з формулою (7.1):

$$M(X) = \sum_{i=1}^3 x_i p_i = x_1 p_1 + x_2 p_2 + x_3 p_3 = 2 \cdot 0,1 + 7 \cdot 0,4 + 9 \cdot 0,5 = 7,5;$$

$$M(X^2) = \sum_{i=1}^3 x_i^2 p_i = x_1^2 p_1 + x_2^2 p_2 + x_3^2 p_3 = 2^2 \cdot 0,1 + 7^2 \cdot 0,4 + 9^2 \cdot 0,5 = 60,5;$$

$$M(Y) = \sum_{j=1}^2 y_j p_j = y_1 p_1 + y_2 p_2 = 4 \cdot 0,7 + 6 \cdot 0,3 = 4,6;$$

$$\text{Маємо } M(Z) = 2 \cdot 7,5 \cdot 4,6 + 60,5 + 3 \cdot 4,6 - 71 = 72,3.$$

2. Мода та медіана випадкової величини

Моду (M_o) дискретної випадкової величини X називають те її можливе значення, якому відповідає найбільша ймовірність появи.

Моду для неперервної випадкової величини X називають те її можливе значення, якому відповідає максимальне значення щільності ймовірності.

Якщо випадкова величина має одну моду, то такий розподіл імовірностей називають *одномодальним*; якщо розподіл має дві моди — *двомодальним* і т. ін. Існують і такі розподіли, які не мають моди. Їх називають *антимодальними*.

Медіаною (M_e) неперервної випадкової величини X називають те її значення, для якого виконуються рівність імовірностей подій: $F(M_e) = 0,5$.

Отже, M_e — можливе значення випадкової величини X , причому таке, що пряма, проведена перпендикулярно до відповідної точки на площині $X = M_e$, поділяє площу фігури, яка обмежена функцією $f(x)$, на дві рівні частини.

3. Дисперсія та середнє квадратичне відхилення

Математичне сподівання не дає достатньо повної інформації про випадкову величину, оскільки одному й тому самому значенню $M(X)$ може відповідати безліч випадкових величин, які будуть різнитися не лише можливими значеннями, а й характером розподілу і самою природою можливих значень.

Дисперсією випадкової величини X називається математичне сподівання квадрата відхилення цієї величини $D(X) = M(X - M(X))^2$.

Властивості дисперсії:

1. Якщо C — стала величина, то $D(C) = 0$.

$$2. D(CX) = C^2 D(X).$$

$$3. \text{ Якщо } A \text{ і } B \text{ — сталі величини, то } D(AX + B) = A^2 D(X).$$

$$4. D(X) = M(X^2) - M^2(X).$$

Для дискретної випадкової величини X зручно шукати дисперсію саме за останньою властивістю, а не за означенням, враховуючи, що

$$M(X^2) = \sum_{i=1}^n x_i^2 p_i \quad (7.2)$$

$$\text{для неперервної } D(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - M^2(X).$$

Отже, дисперсія характеризує розсіювання випадкової величини відносно свого математичного сподівання.

Середнім квадратичним відхиленням випадкової величини X називають корінь квадратний із дисперсії:

$$\sigma(X) = \sqrt{D(X)} \quad (7.3)$$

Приклад 2. Випадкові величини X та Y незалежні і відомо, що $D(X)=2$; $D(Y)=3$. Знайти $D(Z)$ та $\sigma(Z)$, якщо $Z=3X-Y$.

Розв'язання. Скористаємось властивостями дисперсії:

$$D(Z) = D(3X - Y) = D(3X) + D(Y) = 3^2 D(X) + D(Y) = 9D(X) + D(Y) = 9 \cdot 2 + 3 = 21.$$

$$\sigma(Z) = \sqrt{21}.$$

4. Початкові та центральні моменти

Узагальненими числовими характеристиками випадкових величин є початкові та центральні моменти.

Початковим моментом k -го порядку дискретної випадкової величини X називають математичне сподівання величини X^k :

$$\nu_k = M(X^k) = \sum_{i=1}^n x_i^k p_i \quad (k=1, 2, 3, \dots) \quad (7.4)$$

$$\text{Коли } k=1, \quad \nu_1 = M(X) = \sum_{i=1}^n x_i p_i; \quad \text{коли } k=2, \quad \nu_2 = M(X^2) = \sum_{i=1}^n x_i^2 p_i \quad \text{і т. д.}$$

Центральним моментом k -го порядку називається математичне сподівання від $(X - M(X))^k$:

$$\mu_k = M(X - M(X))^k \quad (k=1, 2, 3, \dots). \quad (7.5)$$

Центральні моменти виражаються через початкові:

$$\text{при } k=1, \quad \mu_1 = 0;$$

$$\text{коли } k=2, \quad \mu_2 = \nu_2 - \nu_1^2 = D(X);$$

$$\text{коли } k=3, \quad \mu_3 = \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3;$$

$$\text{коли } k=4, \quad \mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_1^2\nu_2 - 3\nu_1^4.$$

5. Асиметрія і ексцес

Третій центральний момент характеризує асиметрію закону розподілу випадкової величини. Якщо $\mu_3 = 0$, то випадкова величина X симетрично розподілена відносно $M(X)$. Оскільки μ_3 має розмірність випадкової величини в кубі, то вводять безрозмірну величину — коефіцієнт асиметрії:

$$A_s = \frac{\mu_3}{\sigma^3} \quad (7.6)$$

Центральний момент четвертого порядку використовується для визначення ексцесу, що характеризує плосковершинність, або гостровершинність щільності ймовірності $f(x)$. Ексцес обчислюється за формулою

$$E_s = \frac{\mu_4}{\sigma^4} - 3. \quad (7.7)$$

Зауважимо, що число 3 віднімається ось чому. Для центрального закону розподілу, так званого нормального закону, виконується рівність: $\frac{\mu_4}{\sigma^4} = 3$. Отже,

$$E_s = 0.$$

Приклад 3. Закон розподілу дискретної випадкової величини X задано таблицею:

x_i	2	3	10
p_i	0,1	0,4	0,5

Обчислити математичне сподівання $M(X)$; дисперсію $D(X)$; середнє квадратичне відхилення $\sigma(X)$; моду M_o ; початкові та центральні моменти до третього порядку включно: $\nu_1, \nu_2, \nu_3, \mu_1, \mu_2, \mu_3$; коефіцієнт асиметрії A_s ; ексцес.

Розв'язання. Згідно формул (7.1) – (7.7) маємо:

$$M(X) = \sum_{i=1}^3 x_i p_i = x_1 p_1 + x_2 p_2 + x_3 p_3 = 2 \cdot 0,1 - 3 \cdot 0,4 + 10 \cdot 0,5 = 6,4;$$

$$M(X^2) = \sum_{i=1}^3 x_i^2 p_i = x_1^2 p_1 + x_2^2 p_2 + x_3^2 p_3 = 2^2 \cdot 0,1 - 3^2 \cdot 0,4 + 10^2 \cdot 0,5 = 54;$$

$$D(X) = M(X^2) - M^2(X) = 54 - 6,4^2 = 13,04;$$

$$\sigma(X) = \sqrt{D(X)} = \sqrt{13,04} \approx 3,61;$$

$M_o = 10$, оскільки саме значення $x_3 = 10$ має найбільшу ймовірність $p_3 = 0,5$.

$$\nu_1 = M(X) = 6,4; \quad \nu_2 = M(X^2) = 54;$$

$$\nu_3 = M(X^3) = \sum_{i=1}^3 x_i^3 p_i = x_1^3 p_1 + x_2^3 p_2 + x_3^3 p_3 = 2^3 \cdot 0,1 - 3^3 \cdot 0,4 + 10^3 \cdot 0,5 = 511,6;$$

$$\nu_4 = M(X^4) = \sum_{i=1}^4 x_i^4 p_i = x_1^4 p_1 + x_2^4 p_2 + x_3^4 p_3 = 2^4 \cdot 0,1 - 3^4 \cdot 0,4 + 10^4 \cdot 0,5 = 5034;$$

$$\mu_1 = 0; \quad \mu_2 = D(X) = 13,04;$$

$$\mu_3 = \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3 = 511,6 - 3 \cdot 6,4 \cdot 54 + 2 \cdot 6,4^3 = 511,6 - 1036,8 + 524,288 = -0,912$$

$$\mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_1^2\nu_2 - 3\nu_1^4 = 5034 - 4 \cdot 511,6 \cdot 6,4 + 6 \cdot 6,4^2 \cdot 54 - 3 \cdot 6,4^4 = 174,9152;$$

$$As = \frac{\mu_3}{\sigma^3} = \frac{-0,912}{3,61^3} = -\frac{0,912}{47,045881} \approx -0,0194;$$

$$Es = \frac{\mu_4}{\sigma^4} - 3 = \frac{174,9152}{3,61^4} - 3 \approx -1,9701.$$

Лекція 8. Числові характеристики основних розподілів

План

1. Закони розподілу дискретної випадкової величини
2. Закони розподілу неперервної випадкової величини

1. Закони розподілу дискретної випадкової величини

До найбільш поширених законів розподілу ДВВ належать: біномний; Пуассона; геометричний.

Біномний закон розподілу ДВВ задається таблицею, в якій ймовірності p_k розраховуються за формулою Бернуллі:

X	0	1	2	...	n-1	n
p	q^n	$C_n^1 p q^{n-1}$	$C_n^2 p^2 q^{n-2}$...	$C_n^{n-1} p^{n-1} q$	p^n

Сталі параметри n та p називають параметрами розподілу.

Числові характеристики ДВВ X при біномному розподілі:

$$M(X) = np \quad (8.1)$$

$$D(X) = npq \quad (8.2)$$

$$\sigma = \sqrt{npq} \quad (8.3)$$

Якщо в схемі повторних незалежних випробувань $n \rightarrow \infty$, а число p близьке до 0, і крім того $np \rightarrow \lambda$ при $n \rightarrow \infty$, тоді ДВВ X, що визначає кількість появ певної події в схемі Бернуллі, має розподіл Пуассона, який задається наступною таблицею.

X	0	1	2	...	n-1	n
P	$l^{-\lambda}$	$\lambda l^{-\lambda}$	$\frac{\lambda^2}{2!} l^{-\lambda}$...	$\frac{\lambda^{n-1}}{(n-1)!} l^{-\lambda}$	$\frac{\lambda^n}{n!} l^{-\lambda}$

Числові характеристики ДВВ X , що має розподіл Пуассона:

$$M(X) = \lambda \quad (8.4)$$

$$D(X) = \lambda \quad (8.5)$$

Геометричний розподіл задається формулою

$$P(X=m) = pq^{m-1}, \quad (8.6)$$

де $p = P(A)$ – ймовірність появи події A в кожному з незалежних випробувань; $q = 1-p$; ДВВ X – це кількість випробувань у схемі Бернуллі до першої появи події A .

Числові характеристики ДВВ X , що має геометричний закон розподілу:

$$M(X) = \frac{1}{p} \quad (8.7)$$

$$D(X) = \frac{q}{p^2} \quad (8.8)$$

Приклад 1. Знайти $M(X)$, $D(X)$, $\sigma(X)$, якщо ДВВ X – кількість виграшних лотерейних квитків, якщо куплено 40 квитків і ймовірність виграшу будь-якого з них – 0,02.

Розв’язання. ДВВ X має біномний розподіл з параметрами $n = 40$, $p = 0,02$. Згідно формул (6) – (8) маємо $M(X) = np = 40 \cdot 0,02 = 0,8$;

$$D(X) = npq = 40 \cdot 0,02 \cdot 0,98 = 0,784; \quad \sigma(X) = \sqrt{npq} = \sqrt{0,784} \approx 0,885.$$

2. Закони розподілу неперервної випадкової величини

Основними законами розподілу НВВ є рівномірний, показниковий, нормальний та розподіл Стьюдента.

Випадкова величина X розподілена рівномірно у проміжку $[a;b]$, якщо її щільність ймовірності має вигляд:

$$f(x) = \begin{cases} c = \frac{1}{b-a}, & \text{при } x \in [a;b] \\ 0, & \text{при } x \notin [a;b] \end{cases} \quad (8.9)$$

Числові характеристики НВВ X , що рівномірно розподілена:

$$M(X) = \frac{a+b}{2}; \quad D(X) = \frac{(b-a)^2}{12} \quad (8.10)$$

Випадкова величина X розподілена за показниковим законом з параметром λ , якщо щільність її ймовірності має вигляд:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{при } x \geq 0; \\ 0, & \text{при } x < 0. \end{cases} \quad (8.11)$$

Числові характеристики НВВ X , що має показниковий розподіл, визначаються:

$$M(X) = \frac{1}{\lambda}; \quad D(X) = \frac{1}{\lambda^2}$$

Інтегральна функція розподілу для ВВ X , що має показниковий розподіл, задається формулою:

$$F(x) = \begin{cases} 1 - \lambda^{-\lambda x}, & \text{при } x \geq 0; \\ 0, & \text{при } x < 0. \end{cases} \quad (8.12)$$

НВВ X розподілена за нормальним законом, якщо щільність її ймовірності має вигляд:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \quad (8.13)$$

де a та σ – параметри розподілу. Графік цієї функції називається нормальною кривою або *кривою Гаусса*.

Інтегральна функція нормального розподілу має вигляд:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-a)^2}{2\sigma^2}} dt \quad (8.14)$$

При $a=0$, $\sigma=1$ нормальна крива називається нормованою та НВВ X має нормований нормальний розподіл.

Числові характеристики НВВ X , що розподілена за нормальним законом:

$$M(X) = a; D(X) = \sigma^2. \quad (8.15)$$

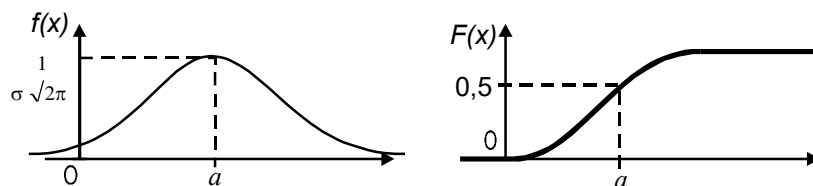
Ймовірність того, що нормально розподілена величина потрапить в проміжок $(c;d)$:

$$P(c < X < d) = \Phi\left(\frac{d-a}{\sigma}\right) - \Phi\left(\frac{c-a}{\sigma}\right) \quad (8.16)$$

де $\Phi(x)$ – функція Лапласа, що задається формулою

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz \quad (8.17)$$

Графіки функцій $f(x)$ та $F(x)$, що задані формулами (8.13), (8.14) мають вигляд:



Приклад 5. Знайти числові характеристики ВВ X , що задана функцією розподілу:

$$F(x) = \begin{cases} 0, & \text{якщо } x \leq 0; \\ \frac{x^3}{8}, & \text{якщо } 0 < x \leq 2; \\ 1, & \text{якщо } x > 2 \end{cases}$$

Розв'язання. Знайдемо щільність ймовірності за формулою $f(x) = F'(x)$:

$$f(x) = \begin{cases} \frac{3x^2}{8}, & x \in (0;2]; \\ 0, & x \notin (0;2]. \end{cases}$$

Знайдемо математичне сподівання:

$$M(X) = \int_0^2 x \frac{3x^2}{8} dx = \frac{3}{8} \cdot \frac{x^4}{4} \Big|_0^2 = \frac{3}{2}.$$

Знайдемо дисперсію:

$$D(X) = \int_0^2 x^2 \frac{3x^2}{8} dx - \left(\frac{3}{2}\right)^2 = \frac{3}{40} x^5 \Big|_0^2 - \frac{9}{4} = \frac{12}{5} - \frac{9}{4} = \frac{3}{20}.$$

Середнє квадратичне відхилення:

$$\sigma(X) = \sqrt{\frac{3}{20}}.$$

Лекція 9. Багатовимірні випадкові величини. Система двох випадкових величин та її числові характеристики

План

1. Основні поняття
2. Основні числові характеристики
3. Кореляційний момент. Коефіцієнт кореляції та його властивості

1. Основні поняття

На одному й тому самому просторі елементарних подій Ω можна визначити не одну, а кілька випадкових величин. Така потреба постає, наприклад, коли досліджуваний об'єкт характеризується кількома випадковими параметрами. Так, у разі виготовлення валів такі їх параметри, як діаметр, довжина, овальність є випадковими величинами, значення яких наперед не можна передбачити. Або, скажімо, структура витрат випадково взятої окремої сім'ї на їжу, одяг, взуття, транспорт, задоволення духовних потреб також є випадковими величинами, визначеними на одному й тому самому просторі елементарних подій.

На багатовимірні випадкові величини поширюються майже без змін основні означення, які були розглянуті для одновимірної випадкової величини.

Одночасна поява внаслідок проведення експерименту n випадкових величин (X_1, X_2, \dots, X_n) з певною ймовірністю являє собою n -вимірну випадкову величину, яку називають також *системою n випадкових величин*, або *n -вимірним випадковим вектором*.

Законом розподілу двох дискретних випадкових величин називають перелік можливих значень $Y = y_i, X = x_j$ та відповідних їм імовірностей спільної появи.

У табличній формі цей закон має такий вигляд:

$X = x_j$ $Y = y_i$	x_1	x_2	x_3	\dots	x_m	p_{yi}
y_1	p_{11}	p_{12}	p_{13}		p_{1m}	p_{y1}
y_2	p_{21}	p_{22}	p_{23}		p_{2m}	p_{y2}
y_3	p_{31}	p_{32}	p_{33}		p_{3m}	p_{y3}
\dots	\dots	\dots	\dots	\dots	\dots	\dots
y_k	p_{k1}	p_{k2}	p_{k3}	\dots	p_{km}	p_{ym}
p_{xj}	p_{x1}	p_{x2}	p_{x3}	\dots	p_{xm}	

Тут використано такі позначення

$$p_{ij} = P((Y = y_i) \cap (X = x_j)); \quad p_{y_i} = \sum_{j=1}^m p_{ij}; \quad p_{x_j} = \sum_{i=1}^k p_{ij}.$$

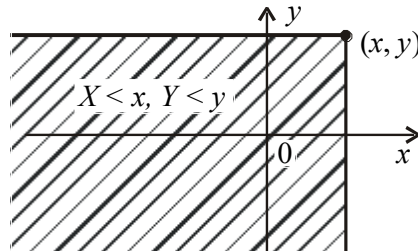
Умова нормування має вигляд:

$$\sum_{i=1}^k \sum_{j=1}^m p_{ij} = \sum_{i=1}^k p_{y_i} = \sum_{j=1}^m p_{x_j} = 1. \quad (9.1)$$

Функцією розподілу ймовірностей системи двох випадкових величин (X, Y) називають таку функцію двох аргументів x, y , яка визначає ймовірність спільної появи подій $(X < x) \cap (Y < y)$:

$$F(x, y) = P((X < x) \cap (Y < y)). \quad (9.2)$$

Геометрично ця функція зображена на рис.:



Для функції розподілу виконується нерівність:

$$0 \leq F(x, y) \leq 1,$$

оскільки $0 \leq P((X < x) \cap (Y < y)) \leq 1$.

2. Основні числові характеристики для випадкових величин X, Y , що утворюють систему (X, Y)

$$M(X) = \sum_{i=1}^k \sum_{j=1}^m x_j p_{ij} = \sum_{j=1}^m x_j p_{x_j}. \quad (9.3)$$

$$D(X) = M(X^2) - M^2(X) = \sum_{i=1}^k \sum_{j=1}^m x_j^2 p_{ij} - M^2(X) = \sum_{j=1}^m x_j^2 p_{x_j} - M^2(X). \quad (9.4)$$

$$\sigma(X) = \sigma_x = \sqrt{D(X)}. \quad (9.5)$$

$$M(Y) = \sum_{i=1}^k \sum_{j=1}^m y_i p_{ij} = \sum_{i=1}^k y_i p_{y_i}. \quad (9.6)$$

$$D(Y) = M(Y^2) - M^2(Y) = \sum_{i=1}^k \sum_{j=1}^m y_i^2 p_{ij} - M^2(Y) = \sum_{i=1}^k y_i^2 p_{y_i} - M^2(Y) \quad (9.7)$$

$$\sigma(Y) = \sigma_y = \sqrt{D(Y)}. \quad (9.8)$$

3. Кореляційний момент. Коефіцієнт кореляції та його властивості

Під час вивчення системи двох і більше випадкових величин доводиться з'ясувати наявність зв'язку між цими величинами та його характер. З відповідною метою застосовують так званий *кореляційний момент*:

$$K_{xy} = M(XY) - M(X)M(Y) = \sum_{i=1}^k \sum_{j=1}^m y_i x_j p_{ij} - M(X)M(Y). \quad (9.9)$$

У разі $K_{xy} = 0$ зв'язок між величинами X та Y , що належать системі (X, Y) , відсутній. Коли $K_{xy} \neq 0$, то між відповідними X і Y кореляційний зв'язок існує.

Тісноту кореляційного зв'язку характеризує коефіцієнт кореляції:

$$r_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y}, \quad (9.10)$$

$$|r_{xy}| \leq 1, \text{ або } -1 \leq r_{xy} \leq 1.$$

Отже, якщо випадкові величини X та Y є незалежними, то $K_{xy} = 0$ і $r_{xy} = 0$. Рівність нулеві r_{xy} є необхідною, але не достатньою умовою незалежності випадкових величин.

Справді, може існувати система залежних випадкових величин, в якій коефіцієнт кореляції дорівнює нулю. Прикладом такої системи є система двох випадкових величин (X, Y) , яка рівномірно розподілена всередині кола радіусом R із центром у початку координат. Дві випадкові величини X і Y називають *некорельованими*, якщо $r_{xy} = 0$, і *корельованими*, якщо $r_{xy} \neq 0$.

Отже, якщо X і Y незалежні, то вони будуть і некорельованими. Але з некорельованості випадкових величин у загальному випадку не випливає їх незалежність.

Приклад 1. Задано закон розподілу системи двох дискретних випадкових величин (X, Y) :

$X = x_j$ $Y = y_i$	5,2	10,2	15,2	P_{y_i}
2,4	0,1a	2a	0,9a	
4,4	2a	0,2a	1,8a	
6,4	1,9a	0,8a	0,3a	
P_{x_j}				

Знайти a . Обчислити $M(X)$; $D(X)$; $\sigma(X)$; $M(Y)$; $D(Y)$; $\sigma(Y)$; K_{xy} ; r_{xy} ; $P(2,4 \leq Y < 6,4; 5,2 < X \leq 15,2)$.

Розв'язання. Скориставшись умовою нормування (13.1), дістанемо:

$$\sum_{i=1}^3 \sum_{j=1}^3 P_{ij} = 0,1a + 2a + 0,9a + 2a + 0,2a + 1,8a + 1,9a + 0,8a + 0,3a = 1 \Rightarrow a = 0,1.$$

Зі знайденим a закон системи набирає такого вигляду:

$X = x_j$ $Y = y_i$	5,2	10,2	15,2	P_{y_i}
2,4	0,01	0,2	0,09	0,3
4,4	0,2	0,02	0,18	0,4
6,4	0,19	0,08	0,03	0,3
P_{x_j}	0,4	0,3	0,3	

Основні числові характеристики обчислюємо за формулами (9.3) - (9.10):

$$M(X) = \sum_{j=1}^3 x_j p_{x_j} = 5,2 \cdot 0,4 + 10,2 \cdot 0,3 + 15,2 \cdot 0,3 = 2,08 + 3,06 + 4,56 = 9,7;$$

$$M(X^2) = \sum_{j=1}^3 x_j^2 p_{x_j} = (5,2)^2 \cdot 0,4 + (10,2)^2 \cdot 0,3 + (15,2)^2 \cdot 0,3 = \\ = 10,816 + 31,212 + 69,312 = 111,34;$$

$$D(X) = M(X^2) - M^2(X) = 111,34 - 94,09 = 17,25;$$

$$\sigma_x = \sqrt{D(X)} = 4,15;$$

$$M(Y) = \sum_{i=1}^3 y_i p_{y_i} = 2,4 \cdot 0,3 + 4,4 \cdot 0,4 + 6,4 \cdot 0,3 = 0,72 + 1,76 + 1,92 = 4,4;$$

$$M(Y^2) = \sum_{i=1}^3 y_i^2 p_{y_i} = (2,4)^2 \cdot 0,3 + (4,4)^2 \cdot 0,4 + (6,4)^2 \cdot 0,3 = \\ = 1,728 + 7,744 + 12,288 = 21,76;$$

$$D(Y) = M(Y^2) - M^2(Y) = 21,76 - (4,4)^2 = 21,76 - 19,36 = 2,4;$$

$$\sigma_y = \sqrt{D(Y)} = 1,55;$$

$$M(XY) = \sum_{i=1}^3 \sum_{j=1}^3 y_i x_j p_{ij} = 2,4 \cdot 5,2 \cdot 0,01 + 2,4 \cdot 10,2 \cdot 0,2 + 2,4 \cdot 15,2 \cdot 0,09 + \\ + 4,4 \cdot 5,2 \cdot 0,2 + 4,4 \cdot 10,2 \cdot 0,02 + 4,4 \cdot 15,2 \cdot 0,18 + 6,4 \cdot 5,2 \cdot 0,19 + \\ + 6,4 \cdot 10,2 \cdot 0,08 + 6,4 \cdot 15,2 \cdot 0,03 = 0,1248 + 4,896 + 3,2832 + 4,576 + \\ + 0,8976 + 12,0384 + 6,3232 + 5,2224 + 2,9184 = 40,28;$$

$$K_{xy} = M(XY) - M(X) M(Y) = 40,28 - 9,7 \cdot 4,4 = 40,28 - 42,68 = -2,4.$$

Оскільки $K_{xy} < 0$, то між відповідними величинами існує кореляційний зв'язок. Для вимірювання тісноти кореляційного зв'язку обчислимо коефіцієнт кореляції

$$r_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y} = \frac{-2,4}{4,15 \cdot 1,55} \approx 0,37.$$

Тому маємо: $p(2,4 \leq Y < 6,4; 5,2 < X \leq 15,2) = 0,2 + 0,02 + 0,09 + 0,18 = 0,31$.

РОЗДІЛ II. СТАТИСТИКА

ТЕМА 3. СТАТИСТИЧНІ РОЗПОДІЛИ ВИБІРОК ТА ЇХ ЧИСЛОВІ ХАРАКТЕРИСТИКИ

Лекція 10. Предмет і методи статистики

План

1. *Статистика як наука.*
2. *Предмет статистики.*
3. *Загальні поняття про статистичну методологію.*

1. *Статистика як наука*

З давніх часів людство здійснювало облік багатьох явищ і предметів, які виникали в процесі його життєдіяльності. Це і чисельність чоловічого і жіночого населення країни, і прибуток скарбниці країни, і земельні угіддя і їх кількість. При подальшому поглибленні суспільного поділу праці, збільшення її продуктивності розвитку суспільних відносин відбувалося зростання кількості даних факторів у виробничій та соціальних сферах, встановилися їх зв'язки на господарському, регіональному та загальнодержавному рівнях.

Всю перераховану інформацію надає суспільству статистика. Термін статистика визначається сукупністю латинських та італійських слів: «status» (становище, положення), «stato» (держава), «statista» (знавець держав, політик). В наукове використання термін статистика було введено німецьким вченим Г.Ахенвалем у 1743 р. для визначення сукупності знань, які характеризують державний устрій, визначні пам'ятки країни, її добробут. Однак таке визначення далеко від сучасного розуміння статистики як науки.

Сьогодні термін „статистика” має декілька значень. Так, в широкому загалі під статистикою розуміють практичну діяльність зі збирання й узагальнення даних про різні явища і процеси суспільного життя. Наприклад, облік лікарняних листків у сфері охорони здоров'я, облік сільськогосподарських робіт у сільськогосподарських формуваннях, збирання відомостей про підписку на газети й журнали, про діяльність тих чи інших господарських одиниць.

Під статистикою розуміють також масив цифрових показників, що характеризують певні статистичні сукупності чи суспільство в цілому. До таких масивів цифрових показників, які трактують як статистику, можна віднести відомості про населення, чисельність спеціалістів, школярів і студентів, інформацію про виробничу діяльність окремих галузей матеріального і нематеріального виробництва.

В той же час під статистикою розуміють окрему галузь науки, яка має свій предмет і метод дослідження.

Отже, *статистика* – це галузь практичної діяльності, економічної науки та навчальної дисципліни з вивчення способів збирання, обробки та аналізу даних про масові соціально-економічні явища та предмети.

Об'єктами статистичного аналізу можуть бути найрізноманітніші явища і процеси суспільного життя.

Статистика як самостійна наука користується специфічним категорійним апаратом, який властивий лише їй. До таких категорійних понять статистики належать:

- статистична сукупність;
- одиниця сукупності;
- ознака сукупності;
- варіюючі ознаки;
- закономірність масових суспільних явищ.

Масові суспільні явища, які досліджуються статистикою, складаються з однорідних в певному відношенні одиниць або елементів. Чисельність одиниць або елементів (одиниць, фактів, випадків) одного й того ж виду, які підлягають статистичному дослідженню, називають *статистичною сукупністю*, а окремі її елементи – *одиницями сукупності*.

Одиниці сукупності мають багато ознак з різним значенням, різним виміром. *Ознака сукупності* – це статистичний еквівалент властивостей одиниць сукупності. Так, для одиниці статистичної сукупності «підприємство» ознаками можуть бути: обсяг виготовленої продукції співвідношення власних та позикових коштів.

Ознаки, які приймають різні значення або відрізняються від окремих одиниць сукупності, називаються *варіюючими* (змінними, мінливими) (наприклад вік, стать людини, для підприємства форма власності, спеціалізація), або *варіаційними*, а окремі їх значення або відмінності – *варіантами*.

Варіюючі ознаки діляться на *атрибутивні* (якісні) та *кількісні*. *Атрибутивними* називаються такі ознаки, окремі значення яких не виражаються числами, наприклад, стать (чоловіча, жіноча), соціальна приналежність, партійність та ін. *Кількісними* називаються такі ознаки, окремі значення (варіанти) яких виражаються числами (вік людини, продуктивність праці, число народжених і померлих, наявність тракторів в окремих господарствах). Залежно від характеру зміни кількісні ознаки діляться на *неперервні* і *дискретні* (*перервні*). *Неперервними* називаються ознаки, які в певних межах можуть приймати будь-які значення (наприклад, вік, врожайність певної культури, зарплата, собівартість одиниці продукції, ціна одиниці товару тощо). *Дискретними* (*перервними*) називаються такі ознаки, які можуть приймати тільки деякі певні значення, між якими не може бути проміжних (наприклад, чисельність сім'ї, кількість поверхів у будинку, кількість коліс в автомашині). Варіанти іскретних ознак виражаються цілими числами.

Ознаки, що характеризують статистичну сукупність взаємопов'язані між собою, тому розрізняють також *факторні* та *результативні* ознаки. *Факторні* – це незалежні ознаки, які впливають на інші ознаки і є причиною їх зміни. *Результативними* ознаками називають залежні ознаки, які змінюються під впливом факторних ознак. Так кваліфікація, стаж роботи – це факторні ознаки, а продуктивність праці – це результативна ознака.

Спостереження за значною кількістю випадків (масове спостереження) та наступне об'єднання цих випадків в однорідну сукупність і розрахунок за цими

сукупностями узагальнюючих показників дає статистиці можливість звільнитися від впливу випадковостей і встановити характерні риси досліджуваних явищ, істотні взаємозв'язки і тенденції їх розвитку, закономірності їх зміни в часі та просторі.

2. Предмет статистики.

Предметом статистики є процеси суспільного життя, які мають масовий прояв. Це означає, що основою дослідження статистики є не індивідуальні (одиначні), а масові суспільні явища. В окремих випадках із загальної маси суспільних явищ може бути виділений окремий об'єкт для поглибленого вивчення, але таке дослідження не може проводитись відірвано від масових явищ, а лише у логічному їх взаємозв'язку, в порівнянні з масовими явищами

Слід виділити три основні риси предмету соціально-економічної статистики:

- вона вивчає суспільні явища;
- статистика вивчає кількісний бік суспільних явищ у нерозривному зв'язку з їх якісним змістом;
- вона вивчає масові суспільні явища, за наслідками яких можливий аналіз та оцінка одиничних явищ на основі застосування монографічного методу.

Кількісну сторону масових суспільних явищ і процесів статистика виражає у вигляді *статистичних показників*. *Статистичним показником* називають узагальнену числову характеристику будь-якого масового явища (процесу) з його якісною визначеністю в конкретних умовах місця та часу. Наприклад, кількість працюючих на підприємстві, обсяги виробленої та реалізованої продукції.

Статистичні показники можуть бути виражені у вигляді *абсолютних та відносних величин*. Якщо статистичний показник стосується окремого явища (наприклад конкретного підприємства), то його називають *індивідуальним*, якщо ж сукупності явищ (наприклад однотипні підприємства регіону), то *узагальненими або зведеними*.

3. Загальні поняття про статистичну методологію.

Для вивчення свого предмету (тобто процесів суспільного життя, які мають суспільний прояв) статистика використовує різні методи, сукупність яких утворює статистичну методологію. Під *методом* будь-якої науки розуміють спосіб дослідження наукою явищ природи і процесів суспільного життя.

У будь-якому статистичному дослідженні можна виділити три послідовні стадії: статистичне спостереження; статистичне зведення і групування; аналіз зведеного та опрацьованого матеріалу, виявлення закономірностей і зв'язків суспільних явищ.

Статистичне спостереження – перша стадія статистичного дослідження. На цій стадії перед статистикою стоїть завдання обліку кожної одиниці сукупності та індивідуальних значень властивих їй ознак. Характерною особливістю цієї стадії є метод масових спостережень. Важливим прийомом статистики є абстрагування під випадкового і виявлення типового,

закономірного. Статистичні закономірності виявляються тільки в масовій кількості одиниць спостереження, тобто у великих сукупностях. Статистична закономірність проявляється лише на підставі дослідження достатньо великої кількості одиниць спостереження. Лише за цієї умови проявляється закон великих чисел. Щоб знайти правильну характеристику явищам в цілому, не можна опиратись на окремі, одиничні спостереження, потрібно взяти й узагальнити всю сукупність фактів або достатньо велике їх число. Прояв закону великих чисел полягає в тому, що зі збільшенням кількості спостережень вплив випадкових причин, що визначають величину ознаки окремих одиниць сукупності, в цілому взаємно погашається і у зведених характеристиках виражається дія основних причин, тобто проявляється закономірність. Сам закон великих чисел не є регулятором масових процесів і явищ, а лише необхідною передумовою їх прояву.

Другим етапом статистичного дослідження є *статистичне зведення і групування* матеріалів статистичного спостереження. Зібрані факти класифікують і систематизують. Їх ділять за ознаками відмінності і об'єднують за ознаками подібності. Найважливішим специфічним прийомом на цьому етапі дослідження є метод групування. За допомогою методу групувань статистики ділять досліджувані явища на типи, характерні групи і підгрупи. Так формують статистичні сукупності, обмежують якісно однорідні в істотному відношенні сукупності, що є необхідною передумовою для застосування узагальнюючих показників. На другому етапі статистичного дослідження переходять від характеристики одиничних фактів до характеристики їх сукупності, від вивчення індивідуальних значень ознаки до їх узагальнення у всій масі одиниць спостереження.

Третім етапом статистичного дослідження є *аналіз зведеного та опрацьованого матеріалу, виявлення закономірностей і зв'язків суспільних явищ*. Специфічним для статистичних методів на цьому етапі є застосування узагальнюючих показників – підсумкових, відносних і середніх величин, статистичних коефіцієнтів. Важливим прийомом є метод порівняння. Порівняння результатів спостереження дає змогу виявити закономірності розвитку явищ у часі і просторі. Для оцінки взаємозв'язків масових явищ статистика використовує індексний, балансовий, кореляційний та інші методи. Все це доповнюється табличним методом, найбільш раціональним викладом цифрового матеріалу, і графічним методом – методом наочного висвітлення статистичних даних.

Таким чином, статистика – це наука, яка за допомогою своїх специфічних методів досліджує кількісний бік масових суспільних явищ і закономірності, які проявляються в них.

Лекція 11. Основні поняття математичної статистики. Статистичний розподіл. Полігон частот. Емпірична функція розподілу

План

1. Генеральна і вибіркова сукупності
2. Статистичний розподіл вибірки
3. Емпірична функція розподілу
4. Полігон частот

1. Генеральна і вибіркова сукупності

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

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

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

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

Математична статистика розв'язує дві категорії задач:

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

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

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

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

Зростаючий числовий ряд варіант називають *варіаційним*.

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

$$n = \sum_{i=1}^k n_i \quad (11.1)$$

де k — кількість варіант, що різняться числовим значенням;
 n — обсяг вибірки.

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

$$w_i = \frac{n_i}{n} \quad (11.2)$$

Для кожної вибірки виконується рівність $\sum_{i=1}^k w_i = 1$.

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

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

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

2. Статистичний розподіл вибірки

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

У табличній формі він має такий вигляд:

$X = x_i$	x_1	x_2	x_3	...	x_k
n_i	n_1	n_2	n_3	...	n_k
w_i	w_1	w_2	w_3	...	w_k

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

3. Емпірична функція розподілу

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

$$F^*(x) = W(X < x) = \frac{n_x}{n} \quad (11.3)$$

називається *емпіричною*, або *кумулятою*. Тут n — обсяг вибірки; n_x — кількість варіант статистичного розподілу вибірки, значення яких менше за фіксовану варіанту x ; $F^*(x)$ — називають ще *функцією нагромадження відносних частот*.

Властивості емпіричної функції розподілу:

- 1) $0 \leq F^*(x) \leq 1$;
- 2) $F(x_{\min}) = 0$, де x_{\min} є найменшою варіантою варіаційного ряду;
- 3) $F(x) \Big|_{x > x_{\max}} = 1$, де x_{\max} є найбільшою варіантою варіаційного ряду;
- 4) $F(x)$ є неспадною функцією аргументу x , а саме: $F(x_2) \geq F(x_1)$ при $x_2 \geq x_1$.

4. Полігон частот і відносних частот.

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

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

Приклад. За заданим дискретним статистичним розподілом вибірки

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

потрібно:

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

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

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

Графічне зображення $F^*(x)$ подано на рис. 14.

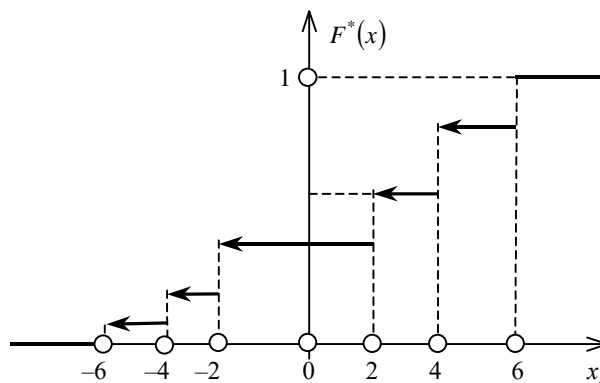


Рис. 14

Полігони частот та відносних частот зображено на рис.15, 16.

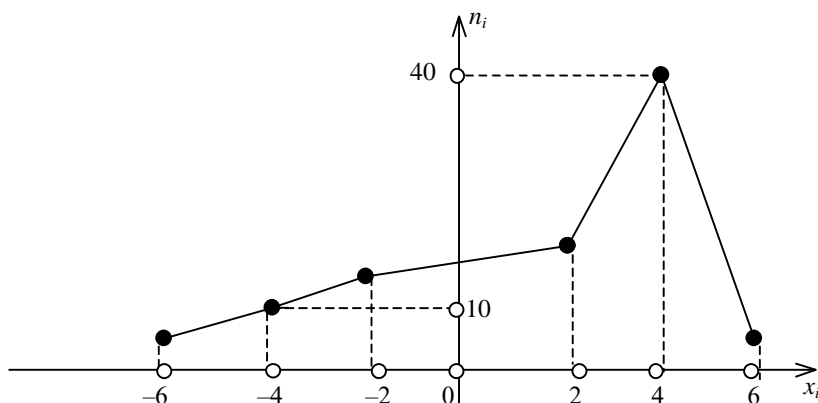


Рис. 15

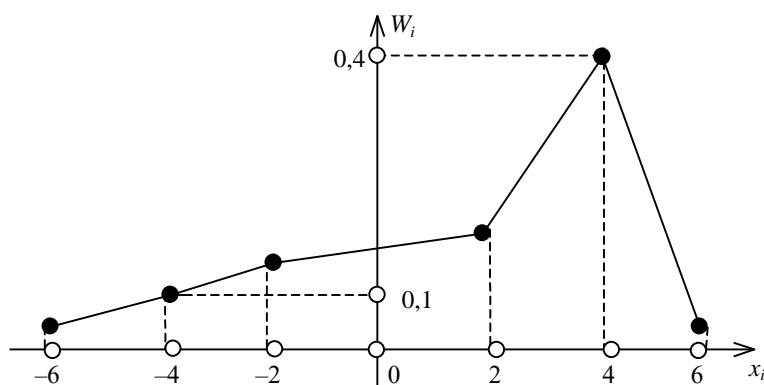


Рис. 16

Лекція 12. Вибіркові числові характеристики

План

1. Вибіркові числові характеристики
2. Емпіричні моменти
3. Коефіцієнт асиметрії. Ексцес.

1. Вибіркові числові характеристики

1) *вибіркова середня величина* \bar{x}_B . Величину, яка визначається формулою

$$\bar{x}_B = \frac{\sum x_i n_i}{n} \quad (12.1)$$

називають *вибірковою середньою величиною дискретного статистичного розподілу вибірки*.

Тут x_i — варіанта варіаційного ряду вибірки;

n_i — частота цієї варіанти;

n — обсяг вибірки ($n = \sum n_i$).

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

то

$$\bar{x}_B = \frac{\sum x_i}{n} \quad (12.2)$$

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

При цьому

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

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

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

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

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

5) *дисперсія*. Для вимірювання розсіювання варіант вибірки відносно \bar{x}_B вибирається дисперсія.

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

$$D_B = \frac{\sum (x_i - \bar{x}_B)^2 n_i}{n}$$

або

$$D_B = \frac{\sum x_i^2 n_i}{n} - (\bar{x}_B)^2 \quad (12.3)$$

6) *середнє квадратичне відхилення вибірки σ_B* . При обчисленні D_B відхилення підноситься до квадрата, а отже, змінюється одиниця виміру ознаки X , тому на основі дисперсії вводиться середнє квадратичне відхилення

$$\sigma_B = \sqrt{D_B} \quad (12.4)$$

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

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

$$R = x_{\max} - x_{\min} \quad (12.6)$$

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

$$V = \frac{\sigma_B}{\bar{x}_B} 100\% \quad (12.7)$$

2. Емпіричні моменти

Середнє зважене значення варіант у степені k ($k = 1, 2, 3, \dots$) називають *початковим емпіричним моментом k -го порядку v_k* , який обчислюється за формулою

$$v_k = \frac{\sum x_k n_i}{n} \quad (12.8)$$

При $k = 1$ дістанемо початковий момент першого порядку:

$$v_1 = \frac{\sum x_i n_i}{n} = \bar{x}_B \quad (12.9)$$

При $k = 2$ обчислимо початковий момент другого порядку:

$$v_2 = \frac{\sum x_i^2 n_i}{n} \quad (12.10)$$

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

$$D_B = v_2 - (v_1)^2 \quad (12.11)$$

Середнє зважене відхилення варіант у степені k ($k = 1, 2, 3, \dots$) називають *центральним емпіричним моментом k -го порядку*

$$\mu_k = \frac{\sum (x_i - \bar{x}_B)^k n_i}{n} \quad (12.12)$$

При $k = 1$ дістанемо:

$$\mu_1 = \frac{\sum (x_i - \bar{x}_B) n_i}{n} = \frac{\sum x_i n_i}{n} - \bar{x}_B \cdot \frac{\sum n_i}{n} = \bar{x}_B - \bar{x}_B = 0.$$

При $k = 2$ маємо:

$$\mu_2 = \frac{\sum x_i^2 n_i}{n} - (\bar{x}_B)^2 = D_B.$$

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

$$\mu_3 = v_3 - 3v_2 \cdot v_1 + 2(v_1)^3, \quad (12.13)$$

$$\mu_4 = v_4 - 4v_3 \cdot v_1 + 6v_2(v_1)^2 - 3(v_1)^4. \quad (12.14)$$

3. Коефіцієнт асиметрії. Ексцес.

Центральний емпіричний момент третього порядку застосовується для обчислення коефіцієнта асиметрії:

$$A_s = \frac{\mu_3}{\sigma_B^3} \quad (12.15)$$

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

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

Центральний емпіричний момент четвертого порядку застосовується для обчислення ексцесу:

$$E_s = \frac{\mu_4}{\sigma_B^4} - 3. \quad (12.16)$$

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

Приклад 1. За заданим статистичним розподілом вибірки

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

потрібно:

- 1) обчислити \bar{x}_B , D_B , σ_B ;
- 2) знайти Mo^* , Me^* ;
- 3) обчислити R , V .

Розв'язання. Оскільки $n = \sum n_i = 100$, то дістанемо:

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

$$\bar{x}_B = 6,7.$$

Для обчислення D_B визначається

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

Тоді

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

$$D_B = 5,16.$$

$$\sigma_B = \sqrt{D_B} = \sqrt{5,16} \approx 2,27.$$

$$\sigma_B = 2,27.$$

$$Mo^* = 6,5; 8,5.$$

Отже, наведений статистичний розподіл вибірки буде двомодальним. $Me^* = 6,5$, оскільки варіанта $x = 6,5$ поділяє варіаційний ряд 2,5; 4,5; **6,5**; 8,5; 10,5 на дві частини: 2,5; 4,5 і 8,5; 10,5, які мають однакову кількість варіант.

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

$$V = \frac{\sigma_B}{\bar{x}_B} 100\% = \frac{2,27}{6,7} 100\% = 33,88\%.$$

Лекція 13. Інтервальний статистичний розподіл вибірки та його числові характеристики

План

1. Інтервальний статистичний розподіл вибірки
2. Гістограма частот та відносних частот.
3. Емпірична функція $F^*(x)$
4. Медіана та мода вибірки
5. \bar{x}_B, D_B, σ_B для інтервального статистичного розподілу вибірки

1. Інтервальний статистичний розподіл вибірки

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

У табличній формі цей розподіл має такий вигляд:

h	$x_1 - x_2$	$x_2 - x_3$	$x_3 - x_4$...	$x_{k-1} - x_k$
n_i	n_1	n_2	n_3	...	n_k
w_i	w_1	w_2	w_3	...	w_k

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

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

2. Гістограма частот та відносних частот.

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

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

Приклад 1. За заданим інтервальним статистичним розподілом вибірки

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

потрібно побудувати гістограму частот і відносних частот.

Розв'язання. Гістограми частот і відносних частот наведені на рис. 17, 18.

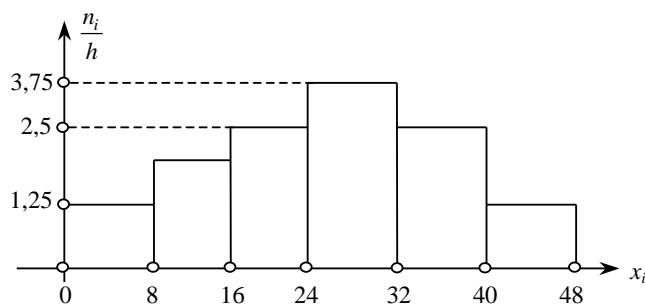


Рис. 17

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



Рис. 18

Площа гістограми відносних частот

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

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

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

4. Медіана та мода вибірки

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

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

Використовуючи лінійну інтерполяцію, моду обчислимо за формулою

$$Mo^* = x_{i-1} + \frac{n_{Mo} - n_{Mo-1}}{2n_{Mo} - n_{Mo-1} - n_{Mo+1}} h \quad (13.1)$$

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

Приклад 2. За заданим інтервальним статистичним розподілом вибірки

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

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

Розв'язання. Гістограма частот зображена на рис. 19.

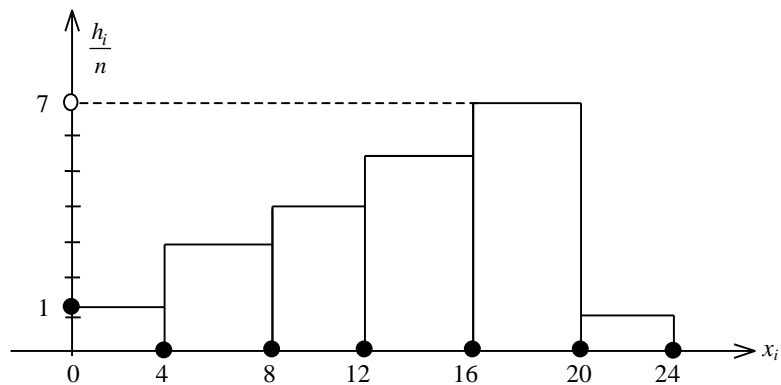


Рис. 19

Графік $F^*(x)$ зображено на рис. 20.

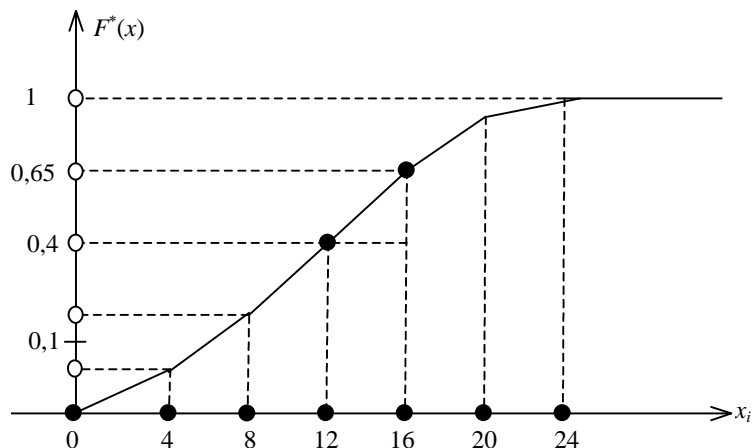


Рис. 20

З рис. 19 визначається модальний інтервал, який дорівнює 16-20.

Застосовуючи (13.1) і беручи до уваги, що $n_{Mo} = 30$, $n_{Mo-1} = 25$, $n_{Mo+1} = 5$, $h = 4$, $x_{i-1} = 16$, дістанемо

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

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

Отже, $Mo^* = 16,17$.

З графіка $F^*(x)$ визначається медіанний інтервал, який дорівнює 12-16. Беручи до уваги, що $F(12) = 0,4$, $F(16) = 0,65$, $h = 4$ дістанемо:

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

Отже, $Me^* = 13,6$.

5. \bar{x}_B, D_B, σ_B для інтервального статистичного розподілу вибірки

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

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

Тоді \bar{x}_B, D_B, σ_B обчислюються за формулами:

$$\bar{x}_B = \frac{\sum x_i^* n_i}{h} \quad (13.2)$$

$$D_B = \frac{\sum (x_i^*)^2 n_i}{h} - (\bar{x}_B)^2; \quad (13.3)$$

$$\sigma_B = \sqrt{D_B} \quad (13.4)$$

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

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

обчислити \bar{x}_B, D_B, σ_B .

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

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

Беручи до уваги те, що $n = 173$, дістанемо:

$$\bar{x}_B = \frac{\sum x_i^* n_i}{n} = \frac{5,5 + 15,6 + 27 + 37,4 + 68,4 + 50,4 + 43,7}{173} + \frac{37,5 + 29,7 + 26,1 + 6,2}{173} = \frac{347,5}{173} \approx 2,008671 \text{ кг.}$$

Отже, $\bar{x}_B = 2,008671$ кг.

$$\frac{\sum (x_i^*)^2 n_i}{n} = \frac{6,05 + 20,29 + 40,5 + 63,58 + 129,96 + 105,84 + 100,51}{173} + \frac{93,75 + 80,19 + 75,69 + 19,22}{173} = \frac{735,58}{173} = 4,251908.$$

$$D_B = \frac{\sum (x_i^*)^2 n_i}{n} - (\bar{x}_B)^2 = 4,251908 - (2,008671)^2 = 4,251908 - 4,034759 = 0,217149.$$

$$D_B = 0,217149.$$

$$\sigma_B = \sqrt{D_B} = \sqrt{0,217149} \approx 0,466.$$

Отже, $\sigma_B = 0,466$ кг.

Лекція 14. Двовимірний статистичний розподіл вибірки та його числові характеристики. Статистичне вивчення кореляційного зв'язку випадкових величин

План

1. Двовимірний статистичний розподіл вибірки
2. Числові характеристики ознак X та Y
3. Кореляційний момент, вибірковий коефіцієнт кореляції

1. Двовимірний статистичний розподіл вибірки

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

У табличній формі цей розподіл має такий вигляд:

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

Тут n_{ij} — частота спільної появи варіант

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

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

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

Загальні числові характеристики ознаки X:

загальна середня величина ознаки X

$$\bar{x} = \frac{\sum_{i=1}^k \sum_{j=1}^m x_j n_{ij}}{n} = \frac{\sum_{j=1}^m x_j n_{x_j}}{n}; \quad (14.1)$$

загальна дисперсія ознаки X

$$D_x = \frac{\sum_{i=1}^k \sum_{j=1}^m x_j^2 n_{ij}}{n} - (\bar{x})^2 = \frac{\sum_{j=1}^m x_j^2 n_{x_j}}{n} - (\bar{x})^2; \quad (14.2)$$

загальне середнє квадратичне відхилення ознаки X

$$\sigma_x = \sqrt{D_x}. \quad (14.3)$$

Загальні числові характеристики ознаки Y:

загальна середня величина ознаки Y

$$\bar{y} = \frac{\sum_{i=1}^k \sum_{j=1}^m y_i n_{ij}}{n} = \frac{\sum_{i=1}^k y_i n_{y_i}}{n}; \quad (14.4)$$

загальна дисперсія ознаки Y

$$D_y = \frac{\sum_{i=1}^k \sum_{j=1}^m y_i^2 n_{ij}}{n} - (\bar{y})^2 = \frac{\sum_{i=1}^k y_i^2 n_{y_i}}{n} - (\bar{y})^2; \quad (14.5)$$

загальне середнє квадратичне відхилення ознаки Y

$$\sigma_y = \sqrt{D_y}. \quad (14.6)$$

3. Кореляційний момент, вибіркового коефіцієнта кореляції

Під час дослідження двовимірного статистичного розподілу вибірки постає потреба з'ясувати наявність зв'язку між ознаками X і Y , який у статистиці називають кореляційним. Для цього обчислюється емпіричний кореляційний момент K_{xy}^* за формулою

$$K_{xy}^* = \frac{\sum_{i=1}^k \sum_{j=1}^m y_i x_j n_{ij}}{n} - \bar{x} \cdot \bar{y} \quad (14.7)$$

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

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

Для вимірювання тісноти кореляційного зв'язку обчислюється вибіркового коефіцієнта кореляції r_B за формулою

$$r_B = \frac{K_{xy}^*}{\sigma_x \sigma_y} \quad (14.8)$$

Як і в теорії ймовірностей, $|r_B| \leq 1$, $-1 \leq r_B \leq 1$.

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

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

потрібно: обчислити K_{xy}^* , r_B .

Розв'язання. Щоб обчислити K_{xy}^* , r_B визначимо \bar{x} , σ_x , \bar{y} , σ_y . Оскільки $n = \sum \sum n_{ij} = 100$, то

$$\begin{aligned} \bar{x} &= \frac{\sum x_j n_{x_j}}{n} = \frac{10 \cdot 30 + 20 \cdot 20 + 30 \cdot 30 + 40 \cdot 20}{100} = \\ &= \frac{300 + 400 + 900 + 800}{100} = \frac{2400}{100} = 24. \\ \bar{x} &= 24. \end{aligned}$$

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

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

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

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

Отже, $\sigma_x = 11,14$.

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

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

Отже, $\bar{y} = 5,8$.

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

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

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

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

Для визначення K_{xy}^* обчислюють

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

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

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

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

Тоді

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

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

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

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

Отже, $r_B = -0,068$, тобто тіснота кореляційного зв'язку між ознаками X та Y є слабкою.

ТЕМА 4. СТАТИСТИЧНІ ОЦІНКИ ПАРАМЕТРІВ ГЕНЕРАЛЬНОЇ СУКУПНОСТІ

Лекція 15. Точкові оцінки параметрів розподілу. Методи знаходження точкових оцінок

План

1. Основні поняття
2. Загальні вимоги до точкових оцінок.
3. Методи знаходження точкових оцінок
4. Виправлена дисперсія, виправлене середнє квадратичне відхилення.

1. Основні поняття

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

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

Параметри генеральної сукупності $M(x) = \bar{X}_\Gamma, D_\Gamma, \sigma_\Gamma, Mo, Me, r_{xy}$ є величинами сталими, але їх числове значення невідоме. Ці параметри оцінюються параметрами вибірки: $\bar{x}_B, D_B, \sigma_B, Mo^*, Me^*, r_B$, які дістають при обробці вибірки. Вони є величинами непередбачуваними, тобто випадковими. Схематично це можна показати так (рис. 21).

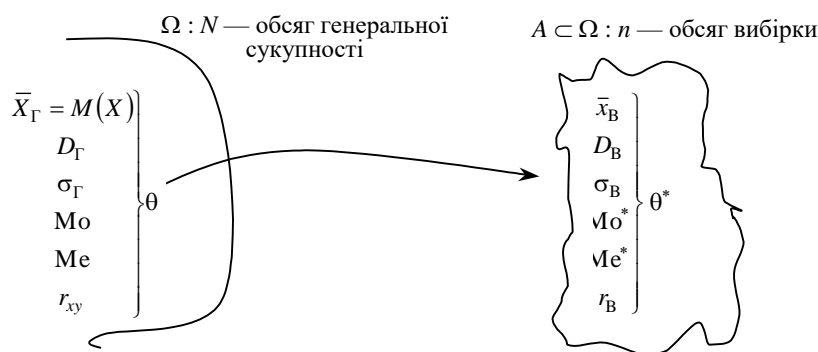


Рис. 21

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

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

2. Загальні вимоги до точкових оцінок

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

$$M(\theta^*) = \theta,$$

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

$$M(\theta^*) \neq \theta,$$

точкова статистична оцінка θ^* називається *зміщеною відносно параметра генеральної сукупності* θ . Різниця $\theta^* - \theta = \delta$ називається *зміщенням статистичної оцінки* θ^* .

Оцінювальний параметр може мати кілька точкових незміщених статистичних оцінок, що можна зобразити так (рис. 22):

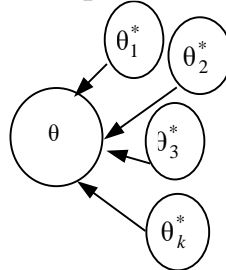


Рис. 22

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

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

$$\lim_{n \rightarrow \infty} P(|\theta^* - \theta| < \delta) = 1.$$

3. Методи визначення точкових статистичних оцінок

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

Метод аналогій. Цей метод базується на тому, що для параметрів генеральної сукупності вибирають такі самі параметри вибірки, тобто для оцінки $\bar{X}_T = M(X)$, D_T вибирають аналогічні статистики — \bar{x}_B , D_B .

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

Метод максимальної правдоподібності.

Цей метод посідає центральне місце в теорії статистичного оцінювання параметрів. Нехай ознака генеральної сукупності X визначається лише одним

параметром θ і має щільність імовірностей $f(x; \theta)$. У разі реалізації вибірки з варіантами x_1, x_2, \dots, x_n щільність імовірностей вибірки буде такою:

$$f(x_1, x_2, \dots, x_n, \theta^*) = f(x_1, \theta^*) \cdot f(x_2, \theta^*) \cdot \dots \cdot f(x_n, \theta^*). \quad (15.1)$$

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

Суть цього методу полягає в тому, що, фіксуючи значення варіант x_1, x_2, \dots, x_n , визначають таке значення параметра θ^* , при якому функція (15.1) максимізується. Вона називається *функцією максимальної правдоподібності* і позначається так: $L = L(\theta^*)$.

При цьому за статистичні оцінки вибирають ті їх значення, за яких задана вибірка буде найімовірнішою, тобто функція досягає максимуму.

4. *Виправлена дисперсія, виправлене середнє квадратичне відхилення.*

Точковою незміщеною статистичною оцінкою для $M(X) \in \bar{x}_B$.

D_B є точковою зміщеною статистичною оцінкою для D_Γ , де $\frac{n-1}{n}$ — коефіцієнт зміщення, який зменшується зі збільшенням обсягу вибірки n .

Отже, $\frac{n}{n-1} D_B$ буде точковою незміщеною статистичною оцінкою для D_Γ .

Її назвали *виправленою дисперсією* і позначили через S^2 .

Звідси точковою незміщеною статистичною оцінкою для D_Γ є виправлена дисперсія $S^2 = \frac{n}{n-1} D_B$ або

$$S^2 = \frac{n}{n-1} \cdot \frac{\sum_{i=1}^n (x_i - \bar{x}_B)^2}{n} = \frac{\sum_{i=1}^n (x_i - \bar{x}_B)^2}{n-1}. \quad (15.2)$$

Величину

$$S = \sqrt{\frac{n}{n-1} D_B} \quad (15.3)$$

називають *виправленим середнім квадратичним відхиленням*.

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

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

Знайти точкові незміщені статистичні оцінки для $M(x)$, D_Γ .

Розв'язання. Оскільки точковою незміщеною оцінкою для $M(x) \in \bar{x}_B$, то обчислимо

$$\begin{aligned}\bar{x}_B &= \frac{\sum x_i n_i}{n} = \\ &= \frac{3,7 \cdot 1 + 3,8 \cdot 22 + 3,9 \cdot 40 + 4,0 \cdot 79 + 4,1 \cdot 27 + 4,2 \cdot 26 + 4,3 \cdot 4 + 4,4 \cdot 1}{200} = \\ &= \frac{3,7 + 83,6 + 156 + 316 + 110,7 + 109,2 + 17,2 + 4,4}{200} = \frac{808,8}{200} = 4,004 \text{ мм.}\end{aligned}$$

Для визначення точкової незміщеної статистичної оцінки для D_T обчислимо D_B :

$$\begin{aligned}\frac{\sum x_i^2 n_i}{n} &= \frac{(3,7)^2 \cdot 1 + (3,8)^2 \cdot 22 + (3,9)^2 \cdot 40 + (4,0)^2 \cdot 79 + \\ &\quad + (4,1)^2 \cdot 27 + (4,2)^2 \cdot 26 + (4,3)^2 \cdot 4 + (4,4)^2 \cdot 1}{200} = \\ &= \frac{13,69 + 317,68 + 608,4 + 1264 + 453,87 + 458,64 + 73,96 + 19,36}{200} = \\ &= \frac{3209,6}{200} = 16,048.\end{aligned}$$

$$\begin{aligned}D_B &= \frac{\sum x_i^2 n_i}{n} - (\bar{x}_B)^2 = 16,048 - (4,004)^2 = \\ &= 16,048 - 16,032016 = 0,015984.\end{aligned}$$

Тоді точкова незміщена статистична оцінка для D_T дорівнюватиме:

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

Лекція 16. Інтервальні оцінки параметрів розподілу. Довірчі інтервали для математичного сподівання при відомій дисперсії.

План

- 1. Довірчі інтервали*
- 2. Довірчі інтервали для математичного сподівання при відомій дисперсії*
- 3. Довірчі інтервали для математичного сподівання при невідомій дисперсії*

1. Довірчі інтервали

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

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

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

$$|\theta^* - \theta| < \delta, \quad (16.1)$$

де δ є точністю оцінки.

Оскільки θ^* є випадковою величиною, то і δ буде випадковою, тому нерівність (16.1) справджуватиметься з певною ймовірністю.

Ймовірність, з якою береться нерівність (16.1), тобто

$$P(|\theta^* - \theta| < \delta) = \gamma \quad (16.2)$$

називають *надійністю*.

Рівність (16.2) можна записати так:

$$P(\theta^* - \delta < \theta < \theta^* + \delta) = \gamma \quad (16.3)$$

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

2. Довірчі інтервали для математичного сподівання при відомій дисперсії

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

$\sigma(\bar{x}_B) = \frac{\sigma_\Gamma}{\sqrt{n}}$, то, скориставшись (16.3), дістанемо

$$P(|\bar{x}_B - a| < \delta) = \gamma \quad (16.4)$$

Аргумент x знаходимо з формули за значенням функції Лапласа, яка дорівнює $0,5\gamma$ за таблицею (додаток 2):

$$2\Phi(x) = \gamma \rightarrow \Phi(x) = 0,5\gamma.$$

Отже, довірчий інтервал дорівнюватиме:

$$\bar{x}_B - \frac{x \cdot \sigma_\Gamma}{\sqrt{n}} < a < \bar{x}_B + \frac{x \cdot \sigma_\Gamma}{\sqrt{n}} \quad (16.5)$$

що можна зобразити умовно на рис. 23.

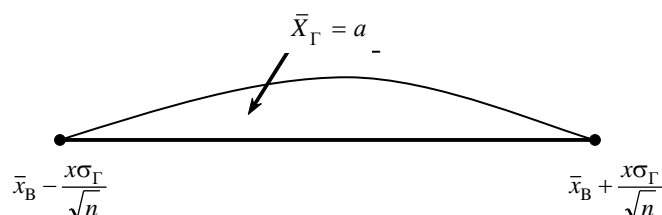


Рис. 23

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

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

Розв'язання. Для побудови довірчого інтервалу необхідно знати: \bar{x}_B , σ_{Γ} , n , x .

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

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

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

Знайдемо числові значення кінців довірчого інтервалу:

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

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

Таким чином, маємо:

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

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

3. Довірчі інтервали для математичного сподівання при невідомій дисперсії

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

$$t = \frac{\bar{x}_B - a}{\frac{S}{\sqrt{n}}}, \quad (16.6)$$

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

Обчисливши за даним статистичним розподілом \bar{x}_B , S і визначивши за таблицею розподілу Стьюдента значення t_{γ} , будемо довірчий інтервал

$$\bar{x}_B - \frac{t_{\gamma} \cdot S}{\sqrt{n}} < a < \bar{x}_B + \frac{t_{\gamma} S}{\sqrt{n}}. \quad (16.7)$$

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

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

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

З надійністю $\gamma = 0,99$ побудувати довірчий інтервал для «а» (середнього часу безвідмовної роботи приладу).

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

Обчислимо \bar{x}_B :

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

Отже, дістали $\bar{x}_B = 222,5$ год.

Визначимо D_B :

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

$$D_B = \frac{\sum t_i^2 n_i}{n} - (\bar{x}_B)^2 = 53\,855 - (222,5)^2 = 53\,855 - 49\,506,25 = 4348,75.$$

Отже, $D_B = 4348,75$.

Виправлене середнє квадратичне відхилення дорівнюватиме:

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

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

Обчислимо кінці довірчого інтервалу:

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

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

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

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

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

Лекція 17. Довірчі інтервали для дисперсії.

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

$$\chi^2 = \frac{n-1}{\sigma_\Gamma^2} S^2, \quad (17.1)$$

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

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

$$\frac{(n-1)S^2}{\chi_2^2} < D_\Gamma < \frac{(n-1)S^2}{\chi_1^2} \quad (17.2)$$

Тоді довірчий інтервал для σ_Γ буде таким:

$$\frac{S\sqrt{n-1}}{\chi_2} < \sigma_\Gamma < \frac{S\sqrt{n-1}}{\chi_1} \quad (17.3)$$

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

$$P(\chi^2 > \chi_1^2) = 1 - \frac{\alpha}{2}; \quad (17.4)$$

$$P(\chi^2 > \chi_2^2) = \frac{\alpha}{2}, \quad (17.5)$$

де $\alpha = 1 - \gamma$.

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

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

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

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

Обчислимо значення \bar{x}_B :

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

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

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

Обчислимо D_B :

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

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

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

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

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

Отже, $D_B = 7558,3 [\text{мкВ}]^2$.

Виправлена дисперсія і виправлене середнє квадратичне відхилення дорівнюватимуть:

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

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

Оскільки $\alpha = 1 - \gamma = 1 - 0,99 = 0,01$, то згідно з (17.4), (17.5) знаходимо значення χ_1^2 , χ_2^2 , а саме:

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

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

За таблицею (додаток 4) знаходимо:

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

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

Обчислимо кінці довірчого інтервалу для D_T :

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

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

Отже, довірчий інтервал для D_T буде таким:

$$4319,0 < D_T < 15856,5.$$

Довірчий інтервал для σ_T становить

$$68,3 < \sigma_T < 130,83.$$

ТЕМА 5. СТАТИСТИЧНІ ГІПОТЕЗИ

Лекція 18. Статистичні гіпотези.

План

1. Загальна інформація
2. Статистичний критерій
3. Область прийняття гіпотези. Критична область
4. Загальний алгоритм перевірки правильності нульової гіпотези
5. Помилки першого та другого роду. Потужність критерію

1. Загальна інформація

Інформація, яку дістають на підставі вибірки, реалізованої із генеральної сукупності, може бути використана для формулювання певних суджень про всю генеральну сукупність. Наприклад, розпочавши виготовляти покриття нового типу для автомобілів, відбирають певну кількість цих покриттів і піддають їх певним тестам.

За результатами тестів можна зробити висновок про те, чи кращі нові покриття від покриттів старого типу, чи ні. А це, у свою чергу, дає підставу для прийняття рішення: виготовляти їх чи ні.

Такі рішення називають *статистичними*.

Статистичні рішення мають імовірнісний характер, тобто завжди існує ймовірність того, що прийняті рішення будуть помилковими.

Головна цінність прийняття статистичних рішень полягає в тому, що в межах імовірнісних категорій можна об'єктивно виміряти ступінь ризику, що відповідає тому чи іншому рішення.

Будь-які статистичні висновки, здобуті на підставі обробки вибірки, називають *статистичними гіпотезами*.

Гіпотезу, що підлягає перевірці, називають *основною*. Оскільки ця гіпотеза припускає відсутність систематичних розбіжностей (нульові розбіжності) між невідомим параметром генеральної сукупності і величиною, що одержана внаслідок обробки вибірки, то її називають *нульовою гіпотезою* і позначають H_0 .

Зміст нульової гіпотези записується так:

$$H_0 : \bar{x}_T = a ;$$

$$H_0 : \sigma_T = 2 ;$$

$$H_0 : r_{xy} = 0,95 .$$

Кожній нульовій гіпотезі можна протиставити кілька альтернативних (конкуруючих) гіпотез, які позначають символом H_α , що заперечують

твердження нульової. Так, наприклад, нульова гіпотеза стверджує: $H_0 : \bar{x}_G = a$, а альтернативна гіпотеза — $H_a : \bar{x}_G > a$, тобто заперечує твердження нульової.

Проста гіпотеза, як правило, належить до параметра ознак генеральної сукупності і є однозначною.

Наприклад, згідно з простою гіпотезою параметр генеральної сукупності дорівнює конкретному числу, а саме:

$$H_0 : \bar{x}_G = 4;$$

$$H_0 : \sigma_G = 4.$$

Складна статистична гіпотеза є неоднозначною. Вона може стверджувати, що значення параметра генеральної сукупності належить певній області ймовірних значень, яка може бути дискретною і неперервною.

Наприклад:

$$H_0 : \bar{x}_G \in [2; 2,1; 2,2] \quad \text{або} \quad H_0 : \bar{x}_G \in [5, 2 \div 6, 5].$$

Нульова гіпотеза може стверджувати як про значення одного параметра генеральної сукупності, так і про значення кількох параметрів, а також про закон розподілу ознаки генеральної сукупності.

2. Статистичний критерій. Емпіричне значення критерію

Для перевірки правильності висунутої статистичної гіпотези вибирають такзваний статистичний критерій, керуючись яким відхиляють або не відхиляють нульову гіпотезу. Статистичний критерій, котрий умовно позначають через K , є випадковою величиною, закон розподілу ймовірностей якої нам заздалегідь відомий. Так, наприклад, для перевірки правильності $H_0 : \bar{X}_G = a$ як статистичний критерій K можна взяти випадкову величину, яку позначають через $K = Z$, що дорівнює

$$Z = \frac{\bar{x}_G - a}{\sigma(\bar{x}_G)} \quad (18.1)$$

і яка має нормований нормальний закон розподілу ймовірностей. При великих обсягах вибірки ($n > 30$) закони розподілу статистичних критеріїв наближатимуться до нормального.

Спостережуване значення критерію, який позначають через K^* , обчислюють за результатом вибірки.

3. Область прийняття гіпотези. Критична область

Множину Ω всіх можливих значень статистичного критерію K можна поділити на дві підмножини A і \bar{A} , які не перетинаються.

$$(A \cup \bar{A} = \Omega, \quad A \cap \bar{A} = \emptyset).$$

Сукупність значень статистичного критерію $K \in A$, за яких нульова гіпотеза не відхиляється, називають *областю прийняття нульової гіпотези*.

Сукупність значень статистичного критерію $K \in \bar{A}$, за яких нульова гіпотеза не приймається, називають *критичною областю*.

Отже, A — область прийняття H_0 ,

\bar{A} — критична область, де H_0 відхиляється.

Точку або кілька точок, що поділяють множину Ω на підмножини A і \bar{A} , називають *критичними* і позначають через $K_{кр}$.

Існують три види критичних областей:

Якщо при $K < K_{кр}$ нульова гіпотеза відхиляється, то в цьому разі ми маємо лівобічну критичну область, яку умовно можна зобразити (рис. 24).

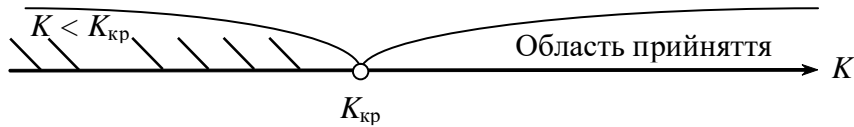


Рис. 24

Якщо при $K > K_{кр}$ нульова гіпотеза відхиляється, то в цьому разі маємо правобічну критичну область (рис. 25).

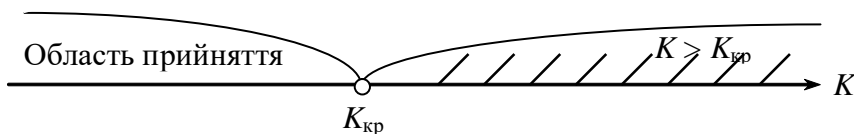


Рис. 25

Якщо ж при $K < K'_{кр}$ і при $K > K''_{кр}$ нульова гіпотеза відхиляється, то маємо двобічну критичну область (рис. 26).

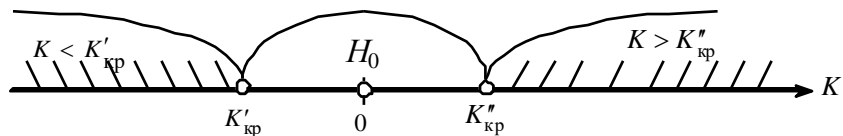


Рис. 26

Лівобічна і правобічна області визначаються однією критичною точкою, двобічна критична область — двома критичними точками, симетричними відносно нуля.

4. Загальний алгоритм перевірки правильності нульової гіпотези

Для перевірки правильності H_0 задається так званий *рівень значущості* α .

α — це мала ймовірність, якою наперед задаються. Вона може набувати значення $\alpha = 0,005; 0,01; 0,001$.

В основу перевірки H_0 покладено принцип $P(K \in \bar{A}) = \alpha$, тобто ймовірність того, що статистичний критерій потрапляє в критичну область \bar{A} , дорівнює малій ймовірності α . Якщо ж виявиться, що $K \in \bar{A}$, а ця подія малоїмовірна і все ж відбулася, то немає підстав приймати нульову гіпотезу.

Пропонується такий алгоритм перевірки правильності H_0 :

1. Сформулювати H_0 й одночасно альтернативну гіпотезу H_α .
2. Вибрати статистичний критерій, який відповідав би сформульованій нульовій гіпотезі.
3. Залежно від змісту нульової та альтернативної гіпотез будується правобічна, лівобічна або двобічна критична область, а саме:

нехай $H_0 : \bar{x}_r = a$, тоді, якщо

$H_\alpha : \bar{x}_r > a$, то вибирається правобічна критична область, якщо

$H_\alpha : \bar{x}_r < a$, то вибирається лівобічна критична область і коли

$H_\alpha : \bar{x}_r \neq a$, то вибирається двобічна критична область.

4. Для побудови критичної області (лівобічної, правобічної чи двобічної) необхідно знайти критичні точки. За вибраним статистичним критерієм та рівнем значущості α знаходяться критичні точки.

5. За результатами вибірки обчислюється спостережуване значення критерію $K_{\text{сп}}^*$.

6. Відхиляють чи приймають нульову гіпотезу на підставі таких міркувань: у разі, коли $K^* \in \bar{A}$, а це є малоїмовірною випадковою подією, $P(K^* \in \bar{A}) = \alpha$ і, незважаючи на це, вона відбулася, то в цьому разі H_0 відхиляється:

для лівобічної критичної області

$$P(K_{\text{сп}}^* < K_{\text{кр}}) = \alpha \quad (18.2)$$

для правобічної критичної області

$$P(K_{\text{сп}}^* > K_{\text{кр}}) = \alpha \quad (18.3)$$

для двобічної критичної області

$$P(K_{\text{сп}}^* < K'_{\text{кр}}) + P(K_{\text{сп}}^* > K''_{\text{кр}}) = \alpha \quad (18.4)$$

або

$$P(K_{\text{сп}}^* < K'_{\text{кр}}) = P(K_{\text{сп}}^* > K''_{\text{кр}}) = \frac{\alpha}{2} \quad (18.5)$$

ураховуючи ту обставину, що критичні точки $K'_{\text{кр}}$ і $K''_{\text{кр}}$ симетрично розташовані відносно нуля.

4. Помилки першого та другого роду. Потужність критерію

Якою б не була малою величиною α , потрапляння спостережуваного значення $K_{\text{сп}}^*$ у критичну область ($K_{\text{сп}}^* \in \bar{A}$) ніколи не буде подією абсолютно неможливою. Тому не виключається той випадок, коли H_0 буде правильною, а $K_{\text{сп}}^* \in \bar{A}$, а тому нульову гіпотезу буде відхилено.

Отже, при перевірці правильності H_0 можуть бути допущені помилки. Розрізняють при цьому помилки першого і другого роду.

Якщо H_0 є правильною, але її відхиляють на основі її перевірки, то буде допущена помилка першого роду.

Якщо H_0 є неправильною, але її приймають, то в цьому разі буде допущена помилка другого роду.

Імовірність помилки першого роду: α . Імовірність помилки другого роду позначають символом β . Різницю $\pi = 1 - \beta$ називають *імовірністю обґрунтованого відхилення H_0* , або *потужністю критерію*.

Лекція 19. Перевірка правильності нульової гіпотези про рівність математичних сподівань

План

1. Перевірка правильності нульової гіпотези про рівність математичних сподівань при великому обсязі вибірки і відомих значеннях дисперсій

2. Малий обсяг вибірки ($n' < 40$, $n'' < 40$) і невідомі значення дисперсій генеральної сукупності

1. Перевірка правильності нульової гіпотези про рівність математичних сподівань при великому обсязі вибірки і відомих значеннях дисперсій

Нехай задано дві генеральні сукупності, ознаки яких X і Y мають нормальний закон розподілу і при цьому незалежні одна від одної. Необхідно перевірити правдивість гіпотези $H_0: M(X) = M(Y)$.

Тут можуть спостерігатися два випадки:

Випадок 1. Обсяг вибірки великий ($n > 40$) і відомі значення D_x , D_y ознак генеральних сукупностей.

З кожної генеральної сукупності здійснюють вибірку відповідно з обсягами n' і n'' і будують статистичні розподіли:

x_i	x_1	x_2	x_3	...	x_k	y_j	y_1	y_2	y_3	...	y_m
n'_i	n'_1	n'_2	n'_3	...	n'_k	n''_j	n''_1	n''_2	n''_3	...	n''_m

Тут $n' = \sum n'_i$, $n'' = \sum n''_j$.

Обчислюються значення

$$\bar{x}_\Gamma = \frac{\sum x_i n'_i}{n'}, \quad \bar{y}_\Gamma = \frac{\sum y_j n''_j}{n''}.$$

Оскільки $D(\bar{x}_\Gamma - \bar{y}_\Gamma) = \frac{D_x}{n'} + \frac{D_y}{n''}$, за статистичний критерій береться

випадкова величина

$$Z = \frac{\bar{x}_B - \bar{y}_B}{\sqrt{\frac{D_x}{n'} + \frac{D_y}{n''}}} \quad (19.1)$$

що має закон розподілу $N(0; 1)$.

Коли $D_x = D_y = D$, дістанемо:

$$Z = \frac{\bar{x}_B - \bar{y}_B}{\sigma_\Gamma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (19.2)$$

Залежно від формулювання альтернативної гіпотези H_α будуються відповідно правобічна, лівобічна та двобічна критичні області.

Спостережуване значення критерію відповідно обчислюється:

$$Z^* = \frac{\bar{x}_B - \bar{y}_B}{\sqrt{\frac{D_x}{n'} + \frac{D_y}{n''}}} \quad (19.3)$$

або

$$Z^* = \frac{\bar{x}_B - \bar{y}_B}{\sigma_{\Gamma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (19.4)$$

Приклад. За заданими статистичними розподілами двох вибірок, реалізованих із двох генеральних сукупностей, ознаки яких мають нормальний закон розподілу зі значенням дисперсій генеральних сукупностей $D_x = 10$; $D_y = 15$,

x_i	12,2	13,2	14,2	15,2	16,2
n'_i	5	15	40	30	10

y_j	8,4	12,4	16,4	20,4	24,4
n''_j	10	15	35	20	20

при рівні значущості $\alpha = 0,01$ перевірити правдивість нульової гіпотези

$H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза

$H_{\alpha} : M(X) > M(Y)$.

Розв'язання. Оскільки $n' = \sum n'_i = 100$; $n'' = \sum n''_j = 100$, обчислимо \bar{x}_B , \bar{y}_B :

$$\begin{aligned} \bar{x}_B &= \frac{\sum x_i n'_i}{n'} = \frac{12,2 \cdot 5 + 13,2 \cdot 15 + 14,2 \cdot 40 + 15,2 \cdot 30 + 16,2 \cdot 10}{100} \\ &= \frac{61 + 198 + 568 + 456 + 162}{100} = \frac{1446,5}{100} = 14,45. \end{aligned}$$

$$\begin{aligned} \bar{y}_B &= \frac{\sum y_j n''_j}{n''} = \frac{8,4 \cdot 10 + 12,4 \cdot 15 + 16,4 \cdot 35 + 20,4 \cdot 20 + 24,4 \cdot 20}{100} = \\ &= \frac{84 + 186 + 574 + 408 + 488}{100} = \frac{1740}{100} = 17,4. \end{aligned}$$

Для альтернативної гіпотези $H_{\alpha} : M(X) > M(Y)$ будується правобічна критична область. Критичну точку $z_{\text{кр}}$ знаходимо з рівності

$$\Phi(z_{\text{кр}}) = \frac{1 - 2\alpha}{2} = \frac{1 - 2 \cdot 0,01}{2} = \frac{0,98}{2} = 0,49 \rightarrow z_{\text{кр}} = 2,34.$$

Правобічна критична область зображена на рис. 27.

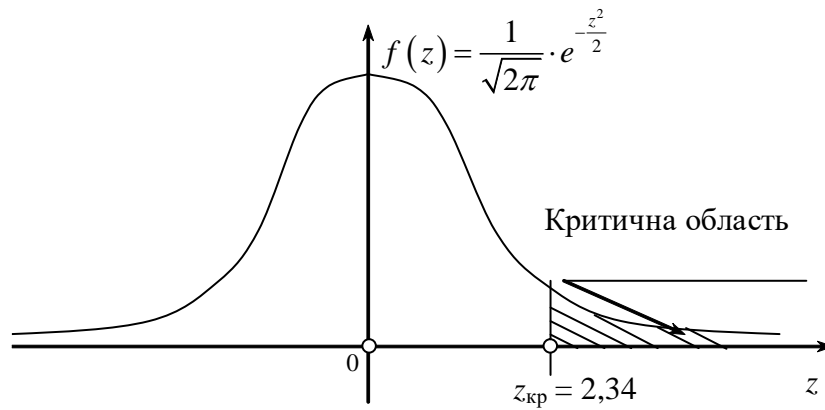


Рис. 27

Обчислимо спостережуване значення критерію

$$Z^* = \frac{\bar{x}_B - \bar{y}_B}{\sqrt{\frac{D_x}{n'} + \frac{D_y}{n''}}} = \frac{14,45 - 17,4}{\sqrt{\frac{10}{100} + \frac{15}{100}}} = -\frac{2,95}{\sqrt{0,1 + 0,15}} = -\frac{2,95}{\sqrt{0,25}} = -\frac{2,95}{0,5} = -5,9.$$

Висновок. Оскільки $Z^* \in]-\infty; 2,34]$, то $H_0 : M(X) = M(Y)$ не відхиляється.

Випадок 2. Якщо обсяг вибірки великий ($n > 40$), але невідомі значення генеральних дисперсій D_x, D_y , то у цьому випадку застосовують їх точкові незміщені статистичні оцінки, а саме:

$$D(\bar{x}_B - \bar{y}_B) \rightarrow S^2 = \frac{\sum (x_j - \bar{x}_B) \cdot n_j'' + \sum (y_i - \bar{y}_B) \cdot n_i'}{n' + n'' - 2} = \frac{(n' - 1)S_x^2 + (n'' - 1)S_y^2}{n' + n'' - 2} \quad (19.5)$$

При великих обсягах вибірок n', n'' статистичний критерій

$$Z = \frac{\bar{x}_B - \bar{y}_B}{\sqrt{\frac{(n' - 1)S_x^2 + (n'' - 1)S_y^2}{n' + n'' - 2}} \cdot \sqrt{\frac{1}{n'} + \frac{1}{n''}}} \quad (19.6)$$

асимптотично наближається до закону розподілу $N(0; 1)$. Тому для визначення критичних точок застосовується функція Лапласа.

Приклад. З допомогою двох радіовимірних приладів вимірювалась відстань до певного об'єкта. Результати вимірювання наведені у вигляді двох статистичних розподілів ознак: Y — відстань, виміряна першим радіоприладом, та X — другим. При цьому Y і X є незалежними між собою і підпорядковані нормальному закону розподілу. Статистичні розподіли мають такий вигляд:

y_i , км	195	198	201	204	207	210
n'_i	10	20	30	20	15	5

x_j , км	184	188	192	196	200	204
n_j''	5	15	30	40	6	4

При рівні значущості $\alpha = 0,01$ перевірити правильність нульової гіпотези

$H_0: M(X) = M(Y)$, якщо альтернативна гіпотеза

$H_\alpha: M(Y) > M(X)$.

Розв'язання. Значення дисперсій генеральних сукупностей невідомі. Необхідно обчислити \bar{x}_B , \bar{y}_B , S_x^2 , S_y^2 .

Оскільки $n' = \sum n_i' = n'' = \sum n_j'' = 100$, то

$$\begin{aligned}\bar{x}_B &= \frac{\sum x_j n_j''}{n''} = \frac{184 \cdot 5 + 188 \cdot 15 + 192 \cdot 30 + 196 \cdot 40 + 200 \cdot 6 + 204 \cdot 4}{100} \\ &= \frac{920 + 2820 + 5760 + 7840 + 1200 + 816}{100} = \frac{19356}{100} = 193,56 \text{ км.}\end{aligned}$$

$$\begin{aligned}\frac{\sum x_j^2 n_j''}{n''} &= \frac{184^2 \cdot 5 + 188^2 \cdot 15 + 192^2 \cdot 30 + 196^2 \cdot 40 + 200^2 \cdot 6 + 204^2 \cdot 4}{100} = \\ &= \frac{3748464}{100} = 37484,64.\end{aligned}$$

$$\begin{aligned}D_B &= \frac{\sum x_j^2 n_j''}{n'} - (\bar{x}_B)^2 = 37484,64 - (193,56)^2 = ; \\ &= 37484,64 - 37465,47 = 19,17;\end{aligned}$$

$$S_x^2 = \frac{n''}{n'' - 1} D_B = \frac{100}{100 - 1} \cdot 19,17 = 19,36;$$

$$S_x = \sqrt{19,36} \approx 4,4.$$

$$\begin{aligned}\bar{y}_B &= \frac{\sum y_i n_i'}{n'} = \frac{195 \cdot 10 + 198 \cdot 20 + 201 \cdot 30 + 204 \cdot 20 + 207 \cdot 15 + 210 \cdot 5}{100} = \\ &= \frac{1950 + 3960 + 6030 + 4080 + 3105 + 1050}{100} = \frac{20175}{100} = 201,75 \text{ км.}\end{aligned}$$

$$\begin{aligned}\frac{\sum y_i^2 n_i'}{n'} &= \frac{195^2 \cdot 10 + 198^2 \cdot 20 + 201^2 \cdot 30 + 204^2 \cdot 20 + 207^2 \cdot 15 + 210^2 \cdot 5}{100} = \\ &= \frac{4071915}{100} = 40719,15;\end{aligned}$$

$$D_B = \frac{\sum y_i^2 n_i'}{n'} - (\bar{y}_B)^2 = 40719,15 - (201,75)^2 = = 40719,15 - 40703,0625 = 16,0875;$$

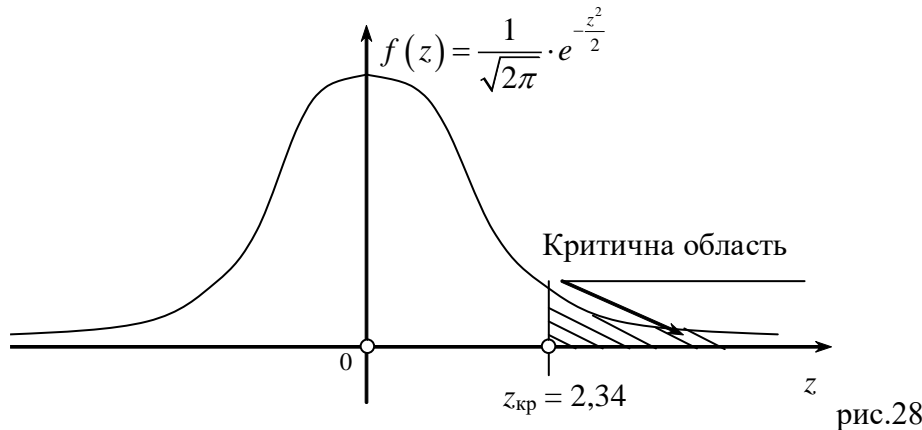
$$S_y^2 = \frac{n'}{n' - 1} D_B = \frac{100}{100 - 1} 16,0875 = 16,25;$$

$$S_y = \sqrt{16,25} \approx 4,03.$$

При альтернативній гіпотезі $H_\alpha: M(X) > M(Y)$ будемо правобічну критичну область, критична точка якої, ураховуючи те, що обсяг вибірки великий, знаходиться з рівності

$$\Phi(z_{\text{кр}}) = \frac{1-2\alpha}{2} = \frac{1-2 \cdot 0,01}{2} = \frac{0,98}{2} = 0,49 \rightarrow z_{\text{кр}} = 2,34.$$

Критична область зображена на рис. 28.



Спостережуване значення критерію обчислюється так:

$$\begin{aligned} Z^* &= \frac{\bar{x}_B - \bar{y}_B}{\sqrt{\frac{(n' - 1)S_x^2 + (n'' - 1)S_y^2}{n' + n'' - 2}} \cdot \sqrt{\frac{1}{n'} + \frac{1}{n''}}} = \\ &= \frac{193,56 - 201,75}{\sqrt{\frac{99 \cdot 19,36 + 99 \cdot 16,25}{100 + 100 - 2}} \cdot \sqrt{\frac{1}{100} + \frac{1}{100}}} = \frac{8,19}{\sqrt{4,215 \cdot 0,02}} = \frac{8,19}{0,29} = -28,24. \end{aligned}$$

Висновок. Оскільки $Z^* \in]-\infty; 2,34]$, то відсутні підстави для відхилення $H_0: M(X) = M(Y)$.

2. Малий обсяг вибірки ($n' < 40, n'' < 40$) і невідомі значення дисперсій генеральної сукупності

При малих обсягах вибірок статистичний критерій

$$z = \frac{\bar{x}_B - \bar{y}_B}{\sqrt{\frac{(n' - 1)S_x^2 + (n'' - 1)S_y^2}{n' + n'' - 2}} \cdot \sqrt{\frac{1}{n'} + \frac{1}{n''}}} \quad (19.7)$$

матиме розподіл Стьюдента з $k = n' + n'' - 2$ ступенями свободи. У цьому разі для побудови критичних областей критичні точки знаходять за таблицею (додаток 3).

Приклад. Протягом доби двома приладами вимірювали напругу в електромережі. Результати вимірювання наведено у вигляді статистичних розподілів

y_i	223	227	229	230	235
n'_i	1	2	6	2	1

x_j	216	217	219	228	236
n''_j	2	3	5	1	1

Припускаючи, що випадкові величини X і Y (напруга у вольтах) є незалежними і мають нормальний закон розподілу ймовірностей, за рівня значущості $\alpha = 0,001$ перевірити правильність нульової гіпотези $H_0 : M(X) = M(Y)$ при альтернативних гіпотезах:

- 1) $H_\alpha : M(X) > M(Y)$;
- 2) $H_\alpha : M(X) \neq M(Y)$.

Розв'язання. Обсяги вибірок відповідно дорівнюють $n' = \sum n'_i = 12$, $n'' = \sum n''_j = 12$.

Обчислимо значення $\bar{x}_B, \bar{y}_B, S_x^2, S_y^2$:

$$\begin{aligned}\bar{y}_B &= \frac{\sum y_i n'_i}{n'} = \frac{223 \cdot 1 + 227 \cdot 2 + 229 \cdot 6 + 230 \cdot 2 + 235 \cdot 1}{12} = \\ &= \frac{223 + 454 + 1374 + 460 + 235}{12} = \frac{2746}{12} = 228,83;\end{aligned}$$

$$\begin{aligned}\frac{\sum y_i^2 n'_i}{n'} &= \frac{223^2 \cdot 1 + 227^2 \cdot 2 + 229^2 \cdot 6 + 230^2 \cdot 2 + 235^2 \cdot 1}{12} = \\ &= \frac{628458}{12} \approx 52371,5;\end{aligned}$$

$$D_B = \frac{\sum y_i^2 n'_i}{n'} - (\bar{y}_B)^2 = 52371,5 - (228,8)^2 = 52371,5 - 52349,44 = 22,06;$$

$$S_y^2 = \frac{n'}{n'-1} D_B = \frac{12}{12-1} \cdot 22,06 \approx 24,1;$$

$$\begin{aligned}\bar{x}_B &= \frac{\sum x_j n''_j}{n''} = \frac{216 \cdot 2 + 217 \cdot 3 + 219 \cdot 5 + 228 \cdot 1 + 236 \cdot 1}{12} = \\ &= \frac{432 + 651 + 1095 + 228 + 236}{12} = \frac{2642}{12} \approx 220,17;\end{aligned}$$

$$\begin{aligned}\frac{\sum x_j^2 n''_j}{n''} &= \frac{216^2 \cdot 2 + 217^2 \cdot 3 + 219^2 \cdot 5 + 228^2 \cdot 1 + 236^2 \cdot 1}{12} = \\ &= \frac{582064}{12} \approx 48505,3;\end{aligned}$$

$$D_B = \frac{\sum x_j^2 n_j''}{n''} - (\bar{x}_B)^2 = 48505,3 - (220,17)^2 = 48505,3 - 48474,83 \approx 30,47;$$

$$S_x^2 = \frac{n''}{n''-1} D_B = \frac{12}{12-1} \cdot 30,47 \approx 33,24.$$

1) Для перевірки правильності нульової гіпотези

$H_0 : M(X) = M(Y)$ при альтернативній гіпотезі

$H_\alpha : M(X) > M(Y)$ будемо правобічну критичну область. Враховуючи, що статистичний критерій має розподіл Стюдента з $k = n' + n'' - 2 = 12 + 12 - 2 = 22$ та рівнем значущості $\alpha = 0,001$, за таблицею (додаток 3) знаходимо критичну точку $z_{кр}(\alpha = 0,001; k = 22) = 3,79$.

Правобічна критична область зображена на рис. 29.

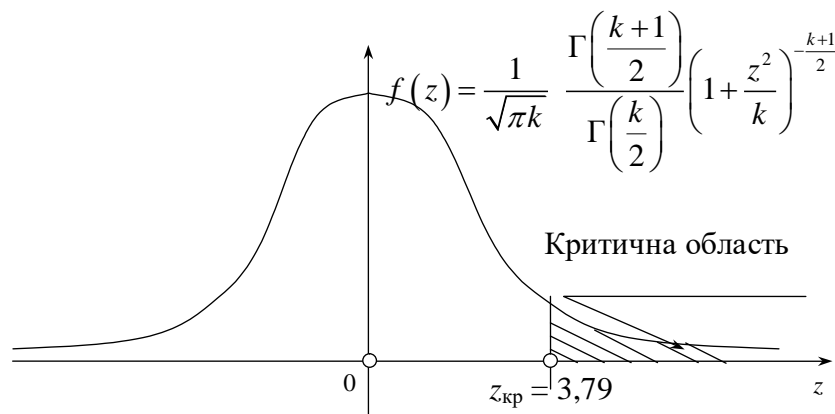


Рис. 29

За формулою (30.7) обчислюємо спостережуване значення критерію

$$\begin{aligned} z^* &= \frac{\bar{x}_B - \bar{y}_B}{\sqrt{\frac{(n'-1)S_x^2 + (n''-1)S_y^2}{n'+n''-2} \cdot \left(\frac{1}{n'} + \frac{1}{n''}\right)}} = \\ &= \frac{220,17 - 228,8}{\sqrt{\frac{11 \cdot 33,24 + 11 \cdot 24,1}{12+12-2} \cdot \left(\frac{1}{12} + \frac{1}{12}\right)}} = \frac{8,63}{\sqrt{\frac{365,64 + 265,1}{22} \cdot 0,17}} = \\ &= -\frac{8,63}{\sqrt{28,67 \cdot 0,17}} = -\frac{8,63}{\sqrt{4,8739}} = -\frac{8,63}{2,21} \approx -3,91. \end{aligned}$$

Висновок. Оскільки $z^* \in [-\infty; 3,79]$, то $H_0 : M(X) = M(Y)$ приймається.

2) Для альтернативної гіпотези $H_\alpha : M(X) \neq M(Y)$ будується двобічна критична область. Беручи до уваги, що $z'_{кр} = -z''_{кр}$, а $z''_{кр} = 3,79$, тоді $z'_{кр} = -3,79$. Двобічна критична область зображена на рис. 30.

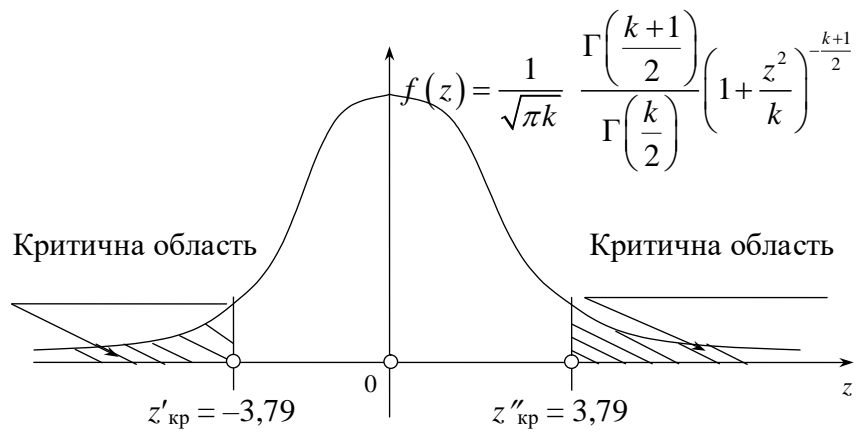


Рис. 30

З попередніх обчислень маємо $z^* = -3,91$.

Висновок. Оскільки $z^* \in]-3,79; 3,79]$, то в цьому разі немає підстав для прийняття $H_0 : M(X) = M(Y)$.

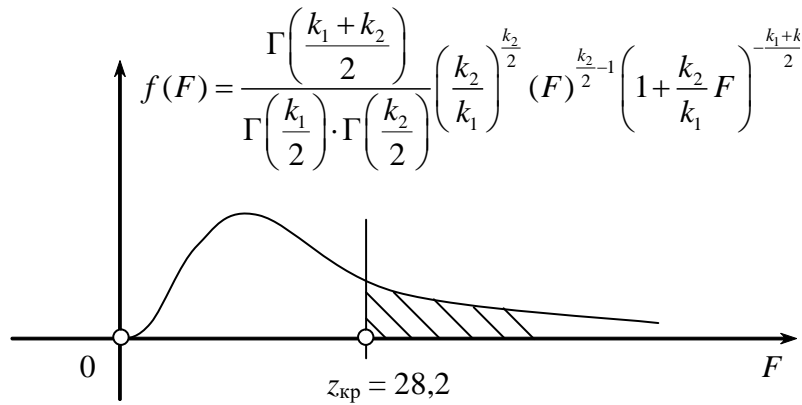


Рис. 31

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

ДОДАТКИ

Додаток 1

ТАБЛИЦЯ ЗНАЧЕНЬ ФУНКЦІЇ ГАУССА $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

	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	3726	3712	3697
0.4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0.5	3521	3503	3485	3467	3478	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	2813	2589	2565	2541	2516	2492	2468	2444
1.0	0.2420	2396	2371	2347	2323	2293	2275	2251	2227	2203
1.1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1.2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1.3	1714	1691	1669	1647	1646	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	1107
1.6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1.7	0940	0925	0909	0893	0978	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	0525	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	0279	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	0164	0163	0158	0154	0151	0147	0143	0139
2.6	0136	0132	0129	0126	0122	0118	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	0040	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	0014	0014	0013	0013	0013
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	0003	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	0002	000

ТАБЛИЦЯ ЗНАЧЕНЬ ФУНКЦІЇ ЛАПЛАСА $\Phi(x) = -\frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	0,26	0,1026	0,52	0,1985	0,78	0,2823
0,01	0,0040	0,27	0,1064	0,53	0,2019	0,79	0,2852
0,02	0,0080	0,28	0,1103	0,54	0,2054	0,80	0,2881
0,03	0,0120	0,29	0,1141	0,55	0,2088	0,81	0,2910
0,04	0,0160	0,30	0,1179	0,56	0,2123	0,820	0,2939
0,05	0,0199	0,31	0,1217	0,57	0,2157	0,83	0,2967
0,06	0,0239	0,32	0,1255	0,58	0,2190	0,84	0,2995
0,07	0,0279	0,33	0,1293	0,59	0,2224	0,85	0,3023
0,08	0,0319	0,34	0,1331	0,60	0,2257	0,86	0,3051
0,09	0,0359	0,35	0,1368	0,61	0,2291	0,87	0,3078
0,10	0,0398	0,36	0,1406	0,62	0,2324	0,88	0,3106
0,11	0,0438	0,37	0,1443	0,63	0,2357	0,89	0,3133
0,12	0,0478	0,38	0,1480	0,64	0,2389	0,90	0,3159
0,13	0,0517	0,39	0,1617	0,65	0,2422	0,91	0,3186
0,14	0,8557	0,40	0,1564	0,66	0,2454	0,92	0,3212
0,15	0,0596	0,41	0,1691	0,67	0,2486	0,93	0,3238
0,16	0,0636	0,42	0,1628	0,68	0,2517	0,94	0,3264
0,17	0,0675	0,43	0,1664	0,69	0,2549	0,95	0,3289
0,18	0,0714	0,44	0,1700	0,70	0,2580	0,96	0,3315
0,19	0,0753	0,45	0,1736	0,71	0,2611	0,97	0,3340
0,20	0,0793	0,46	0,1772	0,72	0,2642	0,98	0,3365
0,21	0,0832	0,47	0,1808	0,73	0,2673	0,99	0,3389
0,22	0,0871	0,48	0,1844	0,74	0,2703	1,00	0,3413
0,23	0,0910	0,49	0,1879	0,75	0,2734	1,01	0,3438
0,24	0,0948	0,50	0,1915	0,76	0,2764	1,02	0,3461
0,25	0,0987	0,51	0,1950	0,77	0,2794	1,03	0,3485

Продовження додатка 2

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1,04	0,3508	1,33	0,4082	1,62	0,4474	1,91	0,4719
1,05	0,3531	1,34	0,4099	1,63	0,4484	1,92	0,4726
1,06	0,3554	1,35	0,4115	1,64	0,4495	1,93	0,4732
1,07	0,3577	1,36	0,4131	1,65	0,4505	1,94	0,4738
1,08	0,3599	1,37	0,4147	1,66	0,4515	1,95	0,4744
1,09	0,3621	1,38	0,4162	1,67	0,4525	1,96	0,4750
1,10	0,3643	1,39	0,4177	1,68	0,4535	1,97	0,4756
1,11	0,3665	1,40	0,4192	1,69	0,4545	1,98	0,4761
1,12	0,3686	1,41	0,4207	1,70	0,4554	1,99	0,4767
1,13	0,3708	1,42	0,4222	1,71	0,4564	2,00	0,4772
1,14	0,3729	1,43	0,4236	1,72	0,4573	2,02	0,4783
1,15	0,3749	1,44	0,4251	1,73	0,4582	2,04	0,4793
1,16	0,3770	1,45	0,4265	1,74	0,4591	2,06	0,4803
1,17	0,3790	1,46	0,4279	1,75	0,4599	2,08	0,4812
1,18	0,3810	1,47	0,4292	1,76	0,4608	2,10	0,4821
1,19	0,3830	1,48	0,4306	1,77	0,4616	2,12	0,4830
1,20	0,3849	1,49	0,4319	1,78	0,4625	2,14	0,4838
1,21	0,3869	1,50	0,4332	1,79	0,4633	2,16	0,4846
1,22	0,3883	1,51	0,4345	1,80	0,4641	2,18	0,4854
1,23	0,3907	1,52	0,4357	1,81	0,4649	2,20	0,4861
1,24	0,3925	1,53	0,4370	1,82	0,4656	2,22	0,4868
1,25	0,3944	1,54	0,4382	1,83	0,4664	2,24	0,4875
1,26	0,3962	1,55	0,4394	1,84	0,4671	2,26	0,4881
1,27	0,3980	1,56	0,4406	1,85	0,4678	2,28	0,4887
1,28	0,3997	1,57	0,4418	1,86	0,4686	2,30	0,4893
1,29	0,4015	1,58	0,4429	1,87	0,4693	2,32	0,4898
1,30	0,4032	1,59	0,4441	1,88	0,4699	2,34	0,4904
1,31	0,4049	1,60	0,4452	1,89	0,4706	2,36	0,4909
1,32	0,4066	1,61	0,4463	1,90	0,4713	2,38	0,4913

Закінчення додатка 2

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
2,40	0,4918	2,60	0,4953	2,80	0,4974	3,20	0,49931
2,42	0,4922	2,62	0,4956	2,82	0,4976	3,40	0,49966
2,44	0,4927	2,64	0,4959	2,84	0,4977	3,60	0,49984
2,46	0,4931	2,66	0,4961	2,86	0,4979	3,80	0,499928
2,48	0,4934	2,68	0,4963	2,90	0,4981	4,00	0,499968
2,50	0,4938	2,70	0,4965	2,92	0,4982	5,00	0,499997
2,52	0,4941	2,72	0,4967	2,94	0,4984		
2,54	0,4945	2,74	0,4969	2,96	0,49846		
2,56	0,4948	2,76	0,4971	2,98	0,49856		
2,58	0,4951	2,78	0,4973	3,00	0,49865	$x > 5$	0,5

КРИТИЧНІ ТОЧКИ РОЗПОДІЛУ СТЬЮДЕНТА (*t*-РОЗПОДІЛУ)

Число ступенів свободи, <i>k</i>	Рівень значущості, α						
	0,20	0,10	0,05	0,02	0,01	0,002	0,001
1	3,08	6,31	12,7	31,82	63,66	127,32	636,62
2	1,89	2,92	4,30	6,97	9,93	14,09	31,60
3	1,64	2,35	3,18	4,54	5,84	7,45	12,94
4	1,53	2,13	2,78	3,75	4,60	5,60	8,61
5	1,48	2,02	2,57	3,37	4,03	4,77	6,86
6	1,44	1,94	2,45	3,14	3,71	4,32	5,96
7	1,42	1,90	2,36	3,00	3,50	4,03	5,41
8	1,40	1,86	2,31	2,90	3,36	3,83	5,04
9	1,38	1,83	2,26	2,82	3,25	3,69	4,78
10	1,37	1,81	2,23	2,76	3,17	3,58	4,59
11	1,36	1,80	2,20	2,72	3,11	3,50	4,44
12	1,36	1,78	2,18	2,68	3,05	3,43	4,32
13	1,35	1,77	2,16	2,65	3,01	3,37	4,22
14	1,34	1,76	2,14	2,62	2,98	3,33	4,14
15	1,34	1,75	2,13	2,60	2,95	3,29	4,07
16	1,34	1,75	2,12	2,58	2,92	3,25	4,02
17	1,33	1,74	2,11	2,57	2,90	3,22	3,97
18	1,33	1,73	2,10	2,55	2,88	3,20	3,92
19	1,33	1,73	2,09	2,54	2,86	3,17	3,88
20	1,33	1,73	2,09	2,53	2,85	3,15	3,85
21	1,32	1,72	2,08	2,52	2,83	3,14	3,82
22	1,32	1,72	2,07	2,51	2,82	3,12	3,79
23	1,32	1,71	2,07	2,50	2,81	3,10	3,77
24	1,32	1,71	2,06	2,49	2,80	3,09	3,75
25	1,32	1,71	2,06	2,48	2,79	3,08	3,73
26	1,32	1,71	2,06	2,48	2,78	3,07	3,71
27	1,31	1,70	2,05	2,47	2,77	3,06	3,69
28	1,31	1,70	2,05	2,47	2,76	3,05	3,67
29	1,31	1,70	2,04	2,46	2,76	3,04	3,66
30	1,31	1,70	2,04	2,46	2,75	3,03	3,65
40	1,30	1,68	2,02	2,42	2,70	2,97	3,55
60	1,30	1,67	2,00	2,39	2,66	2,91	3,46
120	1,29	1,66	1,98	2,36	2,62	2,86	3,37
∞	1,28	1,64	1,96	2,33	2,58	2,81	3,29

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

Число ступенів в свободі, k	$P(\chi^2 > \chi_1^2)$							
	0,2	0,10	0,05	0,02	0,01	0,005	0,002	0,001
1	1,64	2,7	3,8	5,4	6,6	7,9	9,5	10,83
2	3,22	4,6	6,0	7,8	9,2	11,6	12,4	13,8
3	4,64	6,3	7,8	9,8	11,3	12,8	14,6	16,3
4	6,0	7,8	9,5	11,7	13,3	14,9	16,9	18,5
5	7,3	9,2	11,1	13,4	15,1	16,3	18,9	20,5
6	8,6	10,6	12,6	15,0	16,8	18,6	20,7	22,5
7	9,8	12,0	14,1	16,6	18,5	20,3	22,6	24,3
8	11,0	13,4	15,5	18,2	20,1	21,9	24,3	26,1
9	12,2	14,7	16,9	19,7	21,7	23,6	26,1	27,9
10	13,4	16,0	18,3	21,2	23,2	25,2	27,7	29,6
11	14,6	17,3	19,7	22,6	24,7	26,8	29,4	31,3
12	15,8	18,5	21,0	24,1	26,2	28,3	31,0	32,9
13	17,0	19,8	22,4	25,5	27,7	29,8	32,5	34,5
14	18,2	21,1	23,7	26,9	29,1	31,0	34,0	36,1
15	19,3	22,3	25,0	28,3	30,6	32,5	35,5	37,7
16	20,5	23,5	26,3	29,6	32,0	34,0	37,0	39,2
17	21,6	24,8	27,6	31,0	33,4	35,5	38,5	40,8
18	22,8	26,0	28,9	32,3	34,8	37,0	40,0	42,3
19	23,9	27,3	30,1	33,7	36,2	38,5	41,5	43,8
20	25,0	28,4	31,4	35,0	37,6	40,0	43,0	45,3
21	26,2	29,6	32,7	36,3	38,9	41,5	44,5	46,8
22	27,3	30,8	33,9	38,7	40,3	42,5	46,0	48,3
23	28,4	32,0	35,2	39,0	41,6	44,0	47,5	49,7
24	29,6	33,2	36,4	40,3	43,0	45,5	48,5	51,2
25	30,7	34,4	37,7	41,6	44,3	47,0	50,0	52,6
26	31,8	35,6	38,9	42,9	45,6	48,0	51,5	54,1
27	32,9	36,7	40,1	44,1	47,0	49,5	53,0	55,5
28	34,0	37,9	41,3	45,4	48,3	51,0	54,5	56,9
29	35,1	39,1	42,6	46,7	49,6	52,5	56,0	58,3
30	36,3	40,3	43,8	48,0	50,9	54,0	57,5	59,7

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

Число ступенів свободи, k	$P(\chi^2 > \chi_1^2)$							
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30
1	0,00016	0,0006	0,0039	0,016	0,064	0,148	0,455	1,07
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386	2,41
3	0,115	0,185	0,352	0,584	1,005	1,424	2,366	3,66
4	0,30	0,43	0,71	1,06	1,65	2,19	3,36	4,9
5	0,55	0,76	1,14	1,61	2,34	3,0	4,35	6,1
6	0,87	1,13	1,63	2,20	3,07	3,83	5,35	7,2
7	1,24	1,56	2,17	2,83	3,82	4,67	6,35	8,4
8	1,65	2,03	2,73	3,49	4,59	5,53	7,34	9,5
9	2,09	2,563	3,32	4,17	5,38	6,39	8,34	10,7
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,8
11	3,1	3,6	4,6	5,6	7,0	8,1	10,3	12,9
12	3,6	4,2	5,2	6,3	7,8	9,0	11,3	14,0
13	4,1	4,8	5,9	7,0	8,6	9,9	12,3	15,1
14	4,7	5,4	6,6	7,8	9,5	10,8	13,3	16,2
15	5,2	6,0	7,3	8,5	10,3	11,7	14,3	17,3
16	5,8	6,6	8,0	9,3	11,2	12,6	15,3	18,4
17	6,4	7,3	8,7	10,1	12,0	13,5	16,3	19,5
18	7,0	7,9	9,4	10,9	12,9	14,4	17,3	20,6
19	7,6	8,6	10,1	11,7	13,7	15,4	18,3	21,7
20	8,3	9,2	10,9	12,4	14,6	16,3	19,3	22,8
21	8,9	9,9	11,6	13,2	15,4	17,2	20,3	23,9
22	9,5	10,6	12,3	14,0	16,3	18,1	21,3	24,9
23	10,2	10,3	13,1	14,8	17,2	19,0	22,3	26,0
24	10,9	12,0	13,8	15,7	18,1	19,9	23,3	27,1
25	11,5	12,7	14,6	16,5	18,9	20,9	24,3	28,1
26	12,2	13,4	15,4	17,3	19,8	21,8	25,3	29,3
27	12,9	14,1	16,2	18,1	20,7	22,7	26,3	30,3
28	13,6	14,8	16,9	18,9	21,6	23,6	27,3	31,4
29	14,3	15,6	17,7	19,8	22,5	24,6	28,3	32,5
30	15,0	16,3	18,5	20,6	23,4	25,5	29,3	33,5

КРИТИЧНІ ТОЧКИ РОЗПОДІЛУ ФІШЕРА (*F*-РОЗПОДІЛУ)

Рівень значущості 0,05										
k_2	k_1	1	2	3	4	5	6	12	24	∞
1		164,4	199,5	215,7	224,6	230,2	234,0	244,9	249,0	254,3
2		18,5	9,2	19,2	19,3	19,3	19,3	19,4	19,5	19,5
3		10,1	9,6	9,3	9,1	9,0	8,9	8,7	8,6	8,5
4		7,7	6,9	6,6	6,4	6,3	6,2	5,9	5,8	5,6
5		6,6	5,8	5,4	5,2	5,1	5,0	4,7	4,5	4,4
6		6,0	5,1	4,8	4,5	4,4	4,3	4,0	3,8	3,7
7		5,6	4,7	4,4	4,1	4,0	3,9	3,6	3,4	3,2
8		5,3	4,5	4,1	3,8	3,7	3,6	3,3	3,1	2,9
9		5,1	4,3	3,9	3,6	3,5	3,4	3,1	2,9	2,7
10		5,0	4,1	3,7	3,5	3,3	3,2	2,9	2,7	2,5
11		4,8	4,0	3,6	3,4	3,2	3,1	2,8	2,6	2,4
12		4,8	3,9	3,5	3,3	3,1	3,0	2,7	2,5	2,3
13		4,7	3,8	3,4	3,2	3,0	2,9	2,6	2,4	2,2
14		4,6	3,7	3,3	3,1	3,0	2,9	2,5	2,3	2,1
15		4,5	3,7	3,3	3,1	2,9	2,8	2,5	2,3	2,1
16		4,5	3,6	3,2	3,0	2,9	2,7	2,4	2,2	2,0
17		4,5	3,6	3,2	3,0	2,8	2,7	2,4	2,2	2,0
18		4,4	3,6	3,2	2,9	2,8	2,7	2,3	2,1	1,9
19		4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,8
20		4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,8
22		4,3	3,4	3,1	2,8	2,7	2,6	2,2	2,0	1,8
24		4,3	3,4	3,0	2,8	2,6	2,5	2,2	2,0	1,7
26		4,2	3,4	3,0	2,7	2,6	2,4	2,1	1,9	1,7
28		4,2	3,3	2,9	2,7	2,6	2,4	2,1	1,9	1,6
30		4,2	3,3	2,9	2,7	2,5	2,4	2,1	1,9	1,6
40		4,1	3,2	2,9	2,6	2,5	2,3	2,0	1,8	1,5
60		4,0	3,2	2,8	2,5	2,4	2,3	1,9	1,7	1,4
120		3,9	3,1	2,7	2,5	2,3	2,2	1,8	1,6	1,3
∞		3,8	3,0	2,6	2,4	2,2	2,1	1,8	1,5	1,0

Рівень значущості 0,01											
k_2	k_1	1	2	3	4	5	6	8	12	24	∞
1		4052	4999	5403	5625	5764	5859	5981	6106	6234	6366
2		98,5	99,0	99,2	99,3	99,3	99,4	99,3	99,4	99,5	99,5
3		34,1	30,8	29,5	28,7	28,2	27,9	27,5	27,1	26,6	26,1
4		21,2	18,0	16,7	16,0	15,5	15,2	14,8	14,4	13,9	13,5
5		16,3	13,3	12,1	11,4	11,0	10,7	10,3	9,9	9,5	9,0
6		13,7	10,9	9,8	9,2	8,8	8,5	8,1	7,7	7,3	6,9
7		12,3	9,6	8,5	7,9	7,5	7,2	6,8	6,5	6,1	5,7
8		11,3	8,7	7,6	7,0	6,6	6,4	6,0	5,7	5,3	4,9
9		10,6	8,0	7,0	6,4	6,1	5,8	5,5	5,1	4,7	4,3
10		10,0	7,6	6,6	6,0	5,6	5,4	5,1	4,7	4,3	3,9
11		9,7	7,2	6,2	5,7	5,3	5,1	4,7	4,4	4,0	3,6
12		9,3	6,9	6,0	5,4	5,1	4,8	4,5	4,2	3,8	3,4
13		9,1	6,7	5,7	5,2	4,9	4,6	4,3	4,0	3,6	3,2
14		8,9	6,5	5,6	5,0	4,7	4,5	4,1	3,8	3,4	3,0
15		8,7	6,4	5,4	4,9	4,6	4,3	4,0	3,7	3,3	2,9
16		8,5	6,2	5,3	4,8	4,4	4,2	3,9	3,6	3,2	2,8
17		8,4	6,1	5,2	4,7	4,3	4,1	3,8	3,5	3,1	2,7
18		8,3	6,0	5,1	4,6	4,3	4,0	3,7	3,4	3,0	2,6
19		8,2	5,9	5,0	4,5	4,2	3,9	3,6	3,3	2,9	2,4
20		8,1	5,9	4,9	4,4	4,1	3,9	3,6	3,2	2,9	2,4
22		7,9	5,7	4,8	4,3	4,0	3,8	3,5	3,1	2,8	2,3
24		7,8	5,6	4,7	4,2	3,9	3,7	3,3	3,0	2,7	2,2
26		7,7	5,5	4,6	4,1	3,8	3,6	3,3	3,0	2,6	2,1
28		7,6	5,5	4,6	4,1	3,8	3,5	3,2	2,9	2,5	2,1
30		7,6	5,4	4,5	4,0	3,7	3,5	3,2	2,8	2,5	2,0
40		7,3	5,2	4,3	3,8	3,5	3,3	3,0	2,7	2,3	1,8
60		7,1	5,0	4,1	3,7	3,3	3,1	2,8	2,5	2,1	1,6
120		6,9	4,8	4,0	3,5	3,2	3,0	2,7	2,3	2,0	1,4
∞		6,6	4,6	3,8	3,3	3,0	2,8	2,5	2,2	1,8	1,0

Рівень значущості 0,001											
k_2	k_1	1	2	3	4	5	6	8	12	24	∞
1		Змінюється від 400 000 до 600 000									
2		998	999	999	999	999	999	999	999	999	999
3		167	148	141	137	135	133	131	128	126	123
4		74,1	61,3	56,2	53,4	51,7	50,5	49,0	47,4	45,8	44,1
5		47,0	36,6	33,2	31,1	29,8	28,8	27,6	26,4	25,1	23,8
6		35,5	27,0	23,7	21,9	20,8	20,0	19,0	18,0	16,9	15,8
7		29,2	21,7	18,8	17,2	16,2	15,5	14,6	13,7	12,7	11,7
8		25,4	18,5	15,8	14,4	13,5	12,9	12,0	11,2	10,3	9,3
9		22,9	16,4	13,9	12,6	11,7	11,1	10,4	9,6	8,7	7,8
10		21,0	14,9	12,6	11,3	10,5	9,9	9,2	8,5	7,6	6,8
11		19,7	13,8	11,6	10,4	9,6	9,1	8,3	7,6	6,9	6,0
12		18,6	13,0	10,8	9,6	8,9	8,4	7,7	7,0	6,3	5,4
13		17,8	12,3	10,2	9,1	8,4	7,9	7,2	6,5	5,8	5,0
14		17,1	11,8	9,7	8,6	7,9	7,4	6,8	6,1	5,4	4,6
15		16,6	11,3	9,3	8,3	7,6	7,1	6,5	5,8	5,1	4,3
16		16,1	11,0	9,0	7,9	7,3	6,8	6,2	5,6	4,9	4,1
17		15,7	10,7	8,7	7,7	7,0	6,6	6,0	5,3	4,6	3,9
18		15,4	10,4	8,5	7,5	6,8	6,4	5,8	5,1	4,5	3,7
19		15,1	10,2	8,3	7,3	6,6	6,2	5,6	5,0	4,3	3,5
20		14,8	10,0	8,1	7,1	6,5	6,0	5,4	4,8	4,2	3,4
22		14,4	9,6	7,8	6,8	6,2	5,8	5,2	4,6	3,9	3,2
24		14,0	9,3	7,6	6,6	6,0	5,6	5,0	4,4	3,7	3,0
26		13,7	9,1	7,4	6,4	5,8	5,4	4,8	4,2	3,6	2,8
28		13,5	8,9	7,2	6,3	5,7	5,2	4,7	4,1	3,5	2,7
30		13,3	8,8	7,1	6,1	5,5	5,1	4,6	4,0	3,4	2,6
40		12,6	8,2	6,6	5,7	5,1	4,7	4,2	3,6	3,0	2,2
60		12,0	7,8	6,2	5,3	4,8	4,4	3,9	3,3	2,7	1,9
120		11,4	7,3	5,8	5,0	4,4	4,0	3,5	3,0	2,4	1,6
∞		10,8	6,9	5,4	4,6	4,1	3,7	3,3	2,7	2,1	1,0

КРИТИЧНІ ТОЧКИ РОЗПОДІЛУ χ^2

Число ступенів свободи, k	Рівень значущості, α					
	0,01	0,025	0,05	0,95	0,975	0,999
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	60,9	47,0	43,8	18,5	16,8	15,0

СПИСОК ЛІТЕРАТУРИ

1. Статистика: підручник / С.І. Пирожков, В.В. Рязанцева, Р.М. Моторин та ін. Київ: Київ. нац. торг.-екон. ун-т, 2020. 328 с.
2. Статистика: навч. посіб. / Л.М. Карпенко. Одеса: ОРІДУ НАДУ, 2019. 184 с.
3. Теорія ймовірностей і математична статистика: Навч.-метод. посібник. у 2 ч. Ч. І. Теорія ймовірностей / В. І. Жлуктенко, С. І. Наконечний – К.: КНЕУ, 2020. — 304 с.
4. Теорія ймовірностей і математична статистика: Навч.-метод. посібник. у 2 ч. — Ч. II. Математична статистика / В. І. Жлуктенко, С. І. Наконечний – К.: КНЕУ, 2019. — 364 с.
5. Теорія ймовірностей і математична статистика: Навчальний посібник / М. К. Бугір – Тернопіль, «Підручники й посібники», 2020 р. — 404 с.
6. Теорія ймовірностей і математична статистика: Навчальний посібник / Барковський В.В., Барковська Н.В., Лопатін О.К. – Київ: Центр навчальної літератури, 2021. — 424 с.
7. Теорія ймовірностей та математична статистика у прикладах і задачах: навч. посіб. для студ. вищ. навч. закл. / О.Б. Жильцов; за ред. Г.О. Михаліна. — К.: Київ. ун-т ім. Б. Грінченка, 2022. — 336 с.
8. Теорія ймовірностей. [Електронний ресурс]: навч. посіб. для студ. вищ. навч. закл. / КПІ ім. Ігоря Сікорського; уклад.: О.В. Барабаш, А.П. Мусієнко, О.В. Свинчук. – Київ: КПІ ім. Ігоря Сікорського, 2021. – 193 с. URL: https://ela.kpi.ua/bitstream/123456789/42046/1/Navch_Posib_Teor_Ymovirn_BarabashO_MusienkoA_SvynchukO.pdf (дата звернення 23.01.2024)
9. Теорія ймовірностей та математична статистика. [Електронний ресурс]: навчальний посібник / О. І. Огірко, Н. В. Галайко. – Львів: ЛьвДУВС, 2017. – 292 с. URL: <https://dspace.lvduvs.edu.ua/bitstream/1234567890/629/1/%D1%82%D0%B5%D0%BE%D1%80%D1%96%D1%8F%20%D0%B9%D0%BC%D0%BE%D0%B2%D1%96%D1%80%D0%BD%D0%BE%D1%81%D1%82%D0%B5%D0%B9%20%D0%BF%D1%96%D0%B4%D1%80%D1%83%D1%87%D0%BD%D0%B8%D0%BA.pdf> (дата звернення 23.01.2024)
10. Теорія ймовірностей і математична статистика. [Електронний ресурс]: Підручник для студентів фізико-математичних та інформатичних спеціальностей педагогічних університетів. Видання четверте, доповнене / М.І. Жалдак, Н.М. Кузьміна, Г.О. Михалін. – Київ. НПУ імені М.П. Драгоманова, 2020 – 750 с. URL: <https://enpuir.npu.edu.ua/bitstream/handle/123456789/35207/Zhaldak-Kuzmina-Mykhalin.pdf?sequenc> (дата звернення 23.01.2024)

Б 83	<p>Основи теорії ймовірностей та статистики [Текст]: конспект лекцій для здобувачів фахової передвищої освіти освітньо-професійної програми «Транспортні технології (на автомобільному транспорті)» галузі знань 27 Транспорт спеціальності 275.03 Транспортні технології (на автомобільному транспорті) та освітньо-професійної програми «Автомобільний транспорт» галузі знань 27 Транспорт спеціальності 274 Автомобільний транспорт денної форми навчання / уклад. Ю. В. Боровська. – Луцьк : ТФК Луцького НТУ, 2024. – 88 с.</p>
------	--

Методичне видання складене відповідно до робочої програми з дисципліни «Основи теорії ймовірностей та статистики», вміщує теоретичні відомості з кожної теми та приклади розв’язань типових задач. Мета цієї розробки – допомогти здобувачам освіти зрозуміти новий матеріал на лекціях, при підготовці до практичних занять, краще засвоїти основні поняття та методи розв’язання задач з основ теорії ймовірностей та статистики при самостійній підготовці.

Комп’ютерний набір
Редактор

Ю.В.Боровська
Ю.В.Боровська

Підп. до друку «__»_____202__р.
Формат 60x84/16. Папір офс. Гарнітура Таймс.
Ум. друк. арк. ____. Тираж ____ прим.

Відокремлений структурний підрозділ
«Технічний фаховий коледж
Луцького національного технічного університету»
43023 м. Луцьк, вул. Конякіна, 5

