Львівський державний університет внутрішніх справ

О. І. ОГІРКО,
Н. В. ГАЛАЙКО

ТЕОРІЯ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНА СТАТИСТИКА

Навчальний посібник

Львів
2017

УДК 519.2(075.8)
О-36

Рекомендовано до друку та поширення через мережу Інтернет
Вченою радою Львівського державного університету внутрішніх справ
(протокол від 29 листопада 2017 р. № 4)

Рецензенти:
В. М. Юзевич*, доктор фізико-математичних наук, професор;*
Н. М. Пирич*, кандидат фізико-математичних наук, доцент;*
О. Р. Волошин*, кандидат фізико-математичних наук, доцент*

Огірко О. І., Галайко Н. В.
О-36 Теорія ймовірностей та математична статистика: навчальний посібник / О. І. Огірко, Н. В. Галайко. – Львів: ЛьвДУВС, 2017. – 292 с.

Навчальний посібник підготовлений відповідно до навчальної програми дисципліни «Теорія ймовірностей та математична статистика» і «Вища та прикладна математика» для здобувачів освітнього ступеня «бакалавр» галузі знань 07 «Управління та адміністрування».

Посібник складається з дванадцяти розділів. У ньому викладено теоретичні відомості з комбінаторики, основ теорій ймовірностей, теорії оцінювання невідомих параметрів, перевірки статистичних гіпотез, елементів кореляційно-регресійного аналізу. Теоретичний матеріал супроводжується великою кількістю розв'язань типових прикладів і задач. В кінці кожного розділу подано питання для самоконтролю та задачі для самостійного розв'язування.

Для здобувачів вищої освіти економічних спеціальностей закладів вищої освіти, аспірантів та фахівців у сфері економіки.

УДК 519.2(075.8)

ЗМІСТ

Частина II

МАТЕМАТИЧНА СТАТИСТИКА

ВСТУП

Навчальна дисципліна «Теорія ймовірностей і математична статистика» займає важливе місце у навчальному процесі, оскільки формує базові знання у сфері застосування ймовірнісно-статистичного апарата, вивчення закономірностей у масових випадкових явищах, визначення їх ймовірнісних характеристик з метою застосування до аналізу економічних явищ та прогнозування.

Теорія ймовірностей і математична статистика є основою для побудови кількісних моделей керування економічними системами. Ймовірнісно-статистичні методи є базовими для теорії ухвалення рішень як важливої складової сучасного менеджменту. Статистичні показники застосовують під час оцінки ризиків в інвестиційній діяльності, в діяльності страхових компаній, а також у багатьох галузях економіки.

Метою навчального посібника є ознайомлення здобувачів вищої освіти (а також усіх зацікавлених фахівців) з основними поняттями комбінаторики, основ теорій ймовірностей, теорії оцінювання невідомих параметрів, перевірки статистичних гіпотез, елементів кореляційно-регресійного аналізу, методами, теоремами та формулами теорії ймовірностей та математичної статистики.

У посібнику визначені основні засади математичної статистики, яка використовується під час планування, організації та управління виробництвом, оцінювання якості продукції, системного аналізу економічних структур та технологічних процесів, застосування математичних методів у економіці.

Навчальний посібник складений відповідно до програми навчальної дисципліни «Теорія ймовірностей та математична статистика» і «Вища та прикладна математика» для здобувачів освітнього ступеня «бакалавр» галузі знань 07 «Управління та адміністрування. Складається з двох частин – теорії ймовірностей та математичної статистики і розділений на 12 розділів. Розподіл матеріалу на частини, розділи та підрозділи дає змогу виокремлювати головне і зосереджує на ньому увагу читача. Для зручності читача нумерацію формул, теорем, прикладів виконано окремо для кожного розділу.

Навчальний посібник містить теоретичний матеріал, який супроводжується великою кількістю розв'язань типових прикладів і задач, питання для самоконтролю, задачі для самостійного розв'язку, додатки, тестові завдання та список використаних та рекомендованих джерел.

Частина I

ТЕОРІЯ ЙМОВІРНОСТЕЙ

Розділ 1

ВИПАДКОВІ ПОДІЇ. ОЗНАЧЕННЯ ЙМОВІРНОСТІ

Теорія ймовірностей – це розділ математики, що вивчає закономірності випадкових явищ: випадкові події, випадкові величини, їхні функції, властивості й операції над ними.

Теорія ймовірностей виникла в середині XVII ст. через намагання створити теорію азартних ігор. Перші праці, в яких зароджувались основні поняття теорії ймовірностей, належать таким видатним ученим свого часу, як П. Ферма (1601–1665), Б. Паскаль (1623–1662), Х. Гюйгенс (1629–1695).

Визначна роль у розвитку теорії ймовірності належить знаменитому математику Лапласу (1749–1827). Він вперше систематично виклав основи теорії ймовірностей, довів одну з форм центральної граничної теореми (теорема Муавра-Лапласа) і показав практичне застосування теорії ймовірностей до конкретних практичних задач, зокрема, до аналізу помилок спостережень при вимірюванні.

Вагоме значення для розвитку теорії ймовірностей мали і праці видатних математиків XX століття – С. Н. Бернштейна, О. Я. Хінчина, А. М. Колмогорова, Б. В. Гнєденко, Р. А. Фішера, Р. Е. Мізеса, К. Пірсона та ін.

Нині теорію ймовірностей будують на аксіоматичній основі, яку подав А. М. Колмогоров у 30-тих роках минулого століття.

Предметом теорії ймовірностей є вивчення математичних моделей реальних випадкових явищ (подій), які називають ймовірнісними моделями. Такі моделі дозволяють зрозуміти математичну сутність реальних випадкових подій та надають можливість прогнозувати перебіг досліджуваних випадкових подій. Як приклад можна привести зростання чи спадання курсів валют, темпів виробництва, попиту на товари та послуги, прогнозування результатів виборів, результатів спортивних змагань тощо.

1.1. Випадкові події. Класифікація подій

Основні поняття теорії ймовірностей випливають із призначення теорії ймовірностей як науки. Теорія ймовірностей займається виявленням закономірностей в масових випадкових явищах природи. Тому основним операційним матеріалом для теорії ймовірностей є події. Внаслідок проведеного дослідження, яке ми будемо називати випробуванням, настає подія.

Експериментом (або випадковим експериментом) називають певний комплекс умов, що забезпечують спостереження за певним реальним випадковим явищем (певною реальною випадковою подією). Кожне окреме проведення експерименту (тобто забезпечення певних умов) називають **випробуванням**, а відповідний результат випробування називають **наслідком, або елементарним наслідком або елементарною подією**.

Сукупність усіх елементарних подій, пов'язаних з конкретним експериментом, позначають $\boldsymbol{\Omega}$ і називають **множиною (або простором) елементарних подій**. Простір елементарних подій може містити скінчену, злічену або незлічену множину елементів.

Кожну реальну випадкову подію, пов'язану з цим експериментом, ототожнюють з її математичною моделлю – певною сукупністю результатів цього експерименту. Такі моделі називають **подіями**. Події позначають великими літерами латинського алфавіту $A, B, C,$ *або* $A_1, A_2, \ldots, A_n$.

Наприклад: підкидання монети – випробування, випав герб – подія; здача екзамену – випробування, отримана оцінка за екзамен – подія.

Подія A, яка під час виконання комплексу умов може відбутися, або не відбутися, називається **випадковою**.

Наприклад, якщо кинути монету, то вона може впасти так, що зверху буде або герб, або напис. Тому подія «при киданні монети випав герб» є випадковою.

Подія, яка обов'язково відбувається під час кожного виконання комплексу умов, називається **вірогідною (достовірною)**, якщо подія A напевно не може відбутися при виконанні комплексу умов, то вона навивається **неможливою**.

Наприклад, якщо в ящику є лише білі кулі, то те, що витягнута куля, буде білою – достовірна подія, а те, що витягнута куля буде іншого кольору – неможлива подія.

Достовірну подію позначають *U*, а неможливу подію *V*.

Події *A, B, C* ... називаються **несумісними**, якщо поява однієї із цих подій унеможливлює появу іншої події в одному і тому ж випробуванні.

Наприклад, кинули монету. Поява герба виключає появу напису. Подія «з'явився герб» і « з'явився напис» – несумісні.

Якщо поява однієї події не виключає появи іншої, то такі події називаються **сумісними**.

Наприклад, якщо маємо дві коробки деталей. Із кожної коробки вибирають навмання по одній деталі. В цьому разі вибір стандартної деталі і з першої коробки, і з другої – будуть подіями сумісними.

Події називаються **протилежними**, якщо в умовах випробування вони, будучи єдиними його наслідками, і є не сумісні. Подію, протилежну події *A*, позначають через $\bar{A}$.

Прикладами протилежних подій можуть бути влучення і промах при пострілі.

Події A_1, A_2, ..., A_n називаються **рівноможливими**, якщо жодна з цих подій не має «переваги» над іншою.

Наприклад, поява герба і поява напису при киданні монети є події рівноможливі. Або при киданні грального кубика ні одна із граней не має переваги перед іншою.

Події A_1, A_2, ..., A_n., попарно несумісні та рівноможливі, утворюють **повну групу подій**, з яких хоча б одна неминуче відбудеться, тобто сума цих подій є подією достовірною.

Наприклад, підкидають шестигранний кубик. Позначимо події: A_1 – випала грань з цифрою 1; A_2 – випала грань з цифрою 2; A_3 – випала грань з цифрою 3; A_4 – випала грань з цифрою 4; A_5 – випала грань з цифрою 5; A_6 – випала грань з цифрою 6. Тоді події A_1, A_2, A_3, A_4, A_5, A_6 утворюють повну групу подій.

1.2. Класичне означення ймовірності

Якщо подія *A* складається з m частинних випадків, що входять до складу повної групи з *n* попарно несумісних і рівноможливих випадків, то ймовірність *P(A)* події *A* дорівнює

$$P(A) = \frac{m}{n}, \qquad (1.1)$$

тобто ймовірність *P(A)* події *A* дорівнює відношенню числа результатів випробувань, сприятливих події *A*, до числа всіх можливих результатів випробувань.

Із означення ймовірності випливають такі властивості:

1. Ймовірність достовірної події дорівнює одиниці, тобто:

$$P(U) = 1.$$

Якщо подія достовірна, то число сприятливих випадків дорівнює числу всіх можливих, тобто $m = n$ і тому:

$$P(U) = \frac{m}{n} = \frac{n}{n} = 1.$$

2. Ймовірність неможливої події *V* дорівнює нулю, тобто:

$$P(V) = 0.$$

Якщо подія неможлива, то ні один із результатів випробувань не сприяє появі події, тобто $m = 0$ і, отже:

$$P(V) = \frac{m}{n} = \frac{0}{n} = 0.$$

3. Ймовірність випадкової події *A* задовольняє таку нерівність:

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

Випадковій події *A* сприяє лише частина із загального числа результатів випробувань. У цьому разі $0 \leq m \leq n$, а отже,

$$\frac{0}{n} \leq \frac{m}{n} \leq \frac{n}{n}, \text{ тобто } 0 \leq P(A) \leq 1.$$

Приклад 1.1. В ящику 7 однакових за розміром куль: 1 червоні, 2 сині, 4 білих. Знайти ймовірність появи синьої кулі, якщо беруть одну кулю з ящика навмання.

Розв'язання. Нехай подія *A* – навмання взята синя куля.

З ящика можна взяти будь-яку кулю із семи, тому усіх можливих наслідків 7, тому $n = 7$.

Для появи синьої кулі сприяти будуть лише 2 кулі, тому $m = 2$. За формулою (1.1) одержуємо:

$$P(A) \cdot= \frac{2}{7}.$$

Приклад 1.2. Знайти ймовірність того, що вибране випадковим чином двозначне число ділиться на 5.

Розв'язання. Експеримент полягає в тому, що випадково вибирається двозначне число. Наслідком такого випробування є одне з чисел від 10 до 99. Оскільки двозначних чисел 90, то $n = 90$, причому вибір кожного числа можна вважати рівноможливим.

Нехай подія A – вибране двозначне число ділиться на 5.

Загальна кількість наслідків експерименту $n = 90$. Кількість чисел, які діляться на 5 є 18, тому $m = 18$.

Тоді за формулою (1.1) ймовірність події *A:*

$$P(A) = \frac{18}{90} = \frac{1}{35}.$$

1.3. Елементи комбінаторики та їх застосування під час обчислення ймовірностей

Комбінаторика – це розділ математики, який вивчає розміщення об'єктів відповідно до спеціальних правил і методів підрахунку числа всіх можливих способів, якими ці розміщення можуть бути зроблені.

1. Перестановки. Усякий встановлений у скінченній множині порядок називається **перестановкою** його елементів.

Число всіх перестановок з n елементів позначають P_n. Воно дорівнює добутку послідовних натуральних чисел від 1 до n.

$$P_n = n! = 1 \cdot 2 \cdot 3 \dots \cdot (n-1) \cdot n. \qquad (1.2)$$

Приклад 1.3. Президент компанії планує відвідати 7 філій, які знаходяться в різних містах. Скільки є різних маршрутів поїздок?

Розв'язання. Кількість різних маршрутів буде рівною кількістю перестановок з 7 елементів.

Тоді за формулою (1.2) одержимо:

$$P_7 = 7! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 = 5040 \text{ маршрутів поїздок.}$$

2. Розміщення. Множина, в якій задано порядок розміщення її елементів, називається **впорядкованою**. Нехай задано скінченну множину, яка складається з n елементів. Усяка її упорядкована k-елементна підмножина ($k < n$) називається **розміщенням** з n елементів по k.

Число розміщень з n елементів по k позначають A_n^k. Це число знаходять за формулою:

$$A_n^k = \frac{n!}{(n-k)!} \tag{1.3}$$

Приклад 1.4. Керівництво компанії, яке складається з голови, заступника, та головного бухгалтера, обирають з 10 претендентів. Скільки може бути варіантів вибору керівництва компанії?

Розв'язання. Оскільки порядок вибору членів керівництва важливий, то кількість варіантів вибору буде розміщення з 10 по 3.

Тоді за формулою (1.3) одержимо:

$$A_{10}^3 = \frac{10!}{(10-3)!} = 8 \cdot 9 \cdot 10 = 720$$

різних варіантів вибору керівництва компанії.

3. Сполучення. Нехай дано скінченну множину, яка складається з n елементів. Усяка її k-елементна підмножина ($k < n$) називається **сполученням** з n елементів по k.

Число сполучень з n елементів по k позначають C_n^k. Це число знаходиться за формулою:

$$C_n^k = \frac{n!}{m!(n-m)!}\,. \tag{1.4}$$

Приклад 1.5. Приватне підприємство має ліцензію на проведення десяти видів комерційної діяльності. На початку роботи підприємство планує займатися п'ятьма видами діяльності. Скільки є способів вибору цих п'яти видів комерційної діяльності?

Розв'язання. Оскільки порядок вибору видів діяльності не є важливим, то число способів вибору п'яти видів комерційної діяльності буде рівне числу комбінацій з 10 елементів по 5.

Тоді за формулою (1.4) одержимо:

$$C_{10}^5 = \frac{10!}{5!(10-5)!} = \frac{6 \cdot 7 \cdot 8 \cdot 9 \cdot 10}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 252 \text{ способи.}$$

Під час знаходження ймовірностей подій доволі часто для обчислення кількості елементарних подій (сприятливих деякій події або всіх можливих подій) використовують елементи комбінаторики (розміщення, перестановки, сполучення).

Приклад 1.6. Чотирнадцять варіантів контрольної роботи написані на окремих картках і випадково розподіляються серед 12 здобувачів освіти, які сидять в одному ряду. Кожний здобувач освіти отримує одну картку. Знайти ймовірність того, що варіанти 1 і 2 не будуть використані.

Розв'язання. Маємо експеримент з розподілу 14 карток серед 12 здобувачів освіти. Результатами експерименту є впорядковані (за здобувачами освіти) набори розданих 12 з 14 варіантів контрольних робіт. У цьому разі елементарні події відрізняються одна від одної не лише номерами варіантів, що розподіляються серед здобувачів освіти, а й порядком розподілу. Тому такі елементарні події є розміщеннями, а кількість усіх розміщень (елементарних подій) обчислюється за формулою (1.3):

$$n = A_{14}^{12} = \frac{14!}{(14-12)!}\,.$$

Вважаємо, що всі елементарні події рівноможливі.

Нехай подія A – варіанти 1 і 2 залишаться нерозподіленими.

Тоді інші 12 карток розподіляться серед 12 здобувачів освіти.

Такі розподіли є перестановками, а їх кількість обчислюється за формулою (1.2):

$$m = P_{12} = 12!$$

– кількість елементарних подій, що сприяють події A.

Отже, за формулою (1.1) маємо:

$$P(A) = \frac{12!}{A_{14}^{12}} = \frac{1 \cdot 2}{13 \cdot 14} = \frac{1}{91} \approx 0{,}01\,.$$

Приклад 1.7. У групі 10 хлопців і 5 дівчат, серед яких вибирають дві особи для участі у конференції. Яка ймовірність того, що:

А. Виберуть двох хлопців;

Б. Виберуть хлопця й дівчину?

Розв'язання. У групі 15 осіб, серед яких 10 хлопців і 5 дівчат. Вибір двох осіб із 15 є сполученням, кількість яких:

$$C_{15}^{2} = \frac{15!}{2!\,(15-2)!} = \frac{14 \cdot 15}{1 \cdot 2 \cdot} = 105.$$

Це є загальна кількість випадків $n = 15$.

А. Нехай подія A – вибрали двох хлопців.

Кількість випадків, що сприяють події A, визначається кількістю виборів 2 хлопців із 10:

$$m = C_{10}^{2} = \frac{10!}{2!(10-2)!} = \frac{9 \cdot 10}{1 \cdot 2 \cdot} = 45.$$

Тоді за формулою (1.1) одержимо:

$$P(A) = \frac{45}{105} = 0{,}43.$$

Б. Нехай подія B – вибрали хлопця й дівчину.

Кількість випадків, що сприяють події B, визначається кількістю виборів 1 хлопця із 10 і 1 дівчини з 5.

Це можливо зробити

$$m = C_{10}^{1} \cdot C_{5}^{1} = 10 \cdot 5 = 50 \text{ способами.}$$

Отже,

$$P(B) = \frac{50}{105} = 0{,}48.$$

Приклад 1.8. (Задача про вибірку). У партії з N деталей – M бракованих. Навмання беруть n деталей. Знайти ймовірність того, що з n деталей буде m бракованих.

Розв'язання. Елементарним наслідком є вибір будь-яких деталей з їх загального числа N. Число всіх таких наслідків дорівнює числу комбінацій з N по n, тобто C_N^n.

Подія A є виймання n деталей, з яких m бракованих. Наслідком, який сприяє настанню події A, є поява групи з n деталей, в яких n-m якісних деталей і m бракованих. Число таких груп дорівнює $C_{N-M}^{n-m} \cdot C_M^m$, бо групу з m бракованих деталей можна утворити C_M^m способами, а групу n-m якісних деталей – C_{N-M}^{n-m} способами. Водночас будь-яка група якісних деталей може комбінуватись з будь-якою групою несправних деталей.

Шукана ймовірність події A дорівнює відношенню числа наслідків, які сприяють цій події, до числа всіх елементарних наслідків:

$$P(A) = \frac{C_{N-M}^{n-m} \cdot C_M^m}{C_N^n}.$$

1.4. Статистичне означення ймовірності

Використання класичного означення до задач природничо-наукового або економічного характеру не завжди можливо з різних причин. Зокрема, часто неможливо подати результат експерименту як сукупність подій, які можна було б вважати рівноможливими. Наприклад, з міркувань симетрії, на яких ґрунтуються міркування про рівноймовірність подій, вивести ймовірність того, що народжена дитина була хлопчиком, неможливо.

З цієї причини поряд з класичним визначенням ймовірності користуються також **статистичним визначенням ймовірності**, приймаючи за ймовірність події її відносну частоту.

Відносна частота, поряд з ймовірністю, належить до основних понять теорії ймовірностей. Якщо позначити через μ – число появ події в n – незалежних випробуваннях, то відношення числа появ події до загального числа проведених випробувань називають **відносною частотою появи даної події**, тобто

$$W = \frac{\mu}{n}, \tag{1.5}$$

де μ – число появ події, n – загальне число випробувань, W – відносна частота.

Порівнюючи класичне визначення ймовірностей і визначення відносної частоти, можна зробити такий висновок: визначення ймовірностей не вимагає, щоб проводились випробування в дійсності; визначення відносної частоти вимагає фактичного проведення випробувань. Тобто ймовірність визначають до випробування, а відносну частоту – після випробування.

Приклад 1.9. Відділ технічного контролю виявив у партії з 90 деталей 3 браковані деталі. Знайти відносну частоту бракованих деталей.

Розв'язання. Позначимо через A таку подію, як поява бракованої деталі. Загальне число деталей $n = 90$. Бракованих деталей є $\mu = 3$.

Тоді за формулою (1.5) маємо

$$W(A) = \frac{3}{90} = \frac{1}{30} \approx 0{,}03\,.$$

Якщо дослідним шляхом встановлена відносна частота, то одержане число можна прийняти за наближене значення ймовірності.

Наприклад: за даними статистики, відносна частота народження дівчат на 1000 дітей характеризується такими числами за кожен місяць:

Місяці	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Частота w	0,486	0,489	0,490	0,471	0,478	0,482	0,462	0,484	0,485	0,491	0,482	0,473

Відносна частота коливається близько 0,482, яке можна прийняти за наближене значення ймовірності народжування дівчат.

1.5. Геометричне означення ймовірності

Класичне означення ймовірностей використовується лише для експериментів з обмеженим числом рівномірних елементарних подій, тобто коли простір елементарних подій обмежений.

Для математичного опису досліду з нескінченним числом рівноможливих результатів використовують **геометричне означення ймовірності**.

Ймовірність випадкової події A дорівнює

$$P(A) = \frac{m(A)}{m(\Omega)}, \tag{1.6}$$

де $m(A)$ – довжина, площа чи об'єм підмножини A; $m(\Omega)$ – довжина, площа чи об'єм області Ω.

Якщо множина Ω вимірюється в лінійних одиницях, то $P(A)$ дорівнюватиме відношенню довжини, якщо Ω вимірюється в квадратних одиницях, то $P(A)$ дорівнюватиме відношенню площ.

Таке означення ймовірності має просту геометричну інтерптитацію: це частина, яку складає довжина, площа чи об’єм множини *A* від довжини, площі чи об’єму усієї області Ω. Випадкові події в такому разі реалізуються як певні геометричні об’єкти.

Приклад 1.10. (Задача про зустріч). Два студенти *A* і *B* домовились зустрітися в певному місці між 19 і 20 год. Той, хто приходить першим, чекає на другого протягом 20 хв., і не дочекавшись, відходить. Чому дорівнює ймовірність зустрічі студентів *A* і *B,* якщо прихід кожного з них протягом вказаної години може відбутись випадково в незалежні моменти?

Розв’язання. Позначимо момент приходу студента *A* через *x*, а студента *B* – через *y.* Щоб зустріч відбулась, необхідно, щоб виконувалась умова:

$$|x-y| \le 20$$
$$-20 \le x-y \le 20$$

Маємо $x-20 \le y$, $x+20 \ge y$.

Будемо зображати *x* і *y* як координати точки на площині. За одиницю масштабу виберемо хвилину. Всі можливі результати позначимо точками квадрата зі сторонами 60, сприятливі випадки зображуємо нерівностями:

$$x-20 \le y$$
$$x+20 \ge y$$

у заштрихованій області (рис. 1.1):

$m(\Omega)$ – площа квадрату зі стороною 60, тому $m(\Omega) = 60^2$;

$m(A)$ – площа заштрихованої фігури, яка шукається як різниця площ квадратів зі сторонами 60 та 40 відповідно.

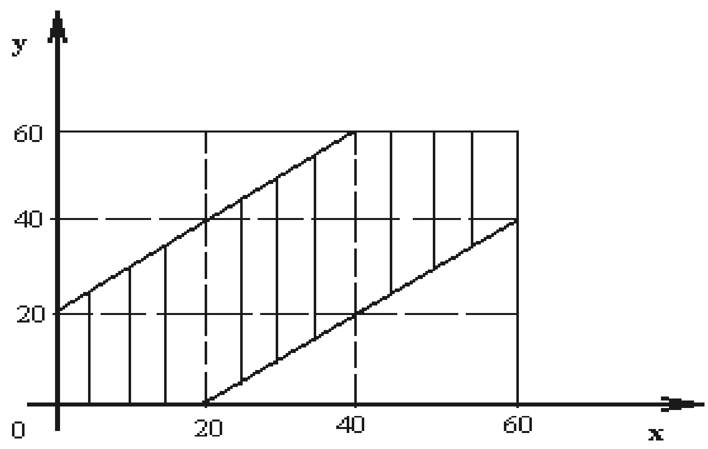

Рис. 1.1. Можливість зустрічі студентів

Тоді за формулою (1.6) одержимо:

$$P(A) = \frac{60^2 - 40^2}{60^2} = \frac{5}{9}.$$

Запитання для самоконтролю

1. Що є предметом теорії ймовірностей?
2. Що таке подія?
3. Що називають множиною (або простором) елементарних подій?
4. Що таке випробування?
5. Які події називаються достовірними, неможливими, випадковими?
6. Які події називаються сумісними, несумісних? Наведіть приклад.
7. Які події називаються протилежними?
8. Які події називаються рівноможливими?
9. Сформулюйте класичне означення ймовірності.
10. Яку нерівність задовольняє ймовірність випадкової події?
11. Які ймовірності достовірної та неможливої подій?
12. Назвіть основні елементи комбінаторики?
13. Як обчислити кількість перестановок з n елементів?
14. Як обчислити кількість розміщень з n по m елементів?
15. Як обчислити кількість сполучень з n по m елементів?
16. Що називається відносною частотою події?
17. Яку властивість відносної частоти спостерігають в практичній діяльності?
18. Що називається статистичною ймовірністю події?
19. Сформулюйте геометричне означення ймовірності.

Задачі для самостійного розв'язування

1. У столі знаходиться 12 олівців однакової форми і розмірів, з них 4 олівці кольорові, а решта – прості. Яка ймовірність взяття навмання простого олівця?

2. В ящику 7 червоних кульок і 3 зелених. Виймають навмання одну кульку. Яка ймовірність, що вона червона?

3. Абонент чекає телефонний дзвінок протягом однієї години. Яка ймовірність того, що йому зателефонують в останні 15 хвилин цієї години?

4. Абонент чекає телефонний дзвінок протягом однієї години. Яка ймовірність того, що йому зателефонують в перші 30 хвилин цієї години?

5. З карток, які складають слово «ціноутворення», навмання вибирають одну. Визначити ймовірність того, що на ній буде написана буква: а) «н»; б) «у».

6. В ящику 40 куль: 28 червоних і 12 синіх. З ящика навмання виймають одну кулю. Визначити ймовірність того, що ця куля: а) червона; б) синя.

7. Із 40 доброякісних і 4 бракованих деталей для контролю взято навмання вісім, які виявились доброякісними деталями. Визначити ймовірність того, що взята наступна деталь буде доброякісною.

8. В групі є 25 студентів. Скількома способами можна вибрати серед них старосту та заступника?

9. Шість викладачів мають розподілити між собою 6 розділів книги так, щоб кожен викладач писав один розділ. Скількома способами можна це зробити?

10. На екзамені з вищої математики має бути комісія з 3 викладачів. Скількома способами можна утворити комісію з 12 осіб?

11. У групі налічується 20 студентів. Скількома способами можна обрати:

а) студентську раду із трьох студентів;

б) голову, заступника та секретаря студентської ради?

12. В ящику 7 червоних кульок і 3 зелених. Виймають навмання дві кульки. Яка ймовірність, що вони обидві червоні?

13. В групі 15 студентів, серед яких 7 відмінників. За списком навмання вибирають 6 студентів. Знайти ймовірність того, що серед вибраних студентів будуть 4 відмінники.

14. В цеху працюють 6 мужчин і 4 жінки. По табельних номерах навмання вибирають сім людей. Знайти ймовірність того, що серед відібраних людей є три жінки.

15. В урні міститься 10 кульок, серед них 4 сині і 6 червоних. Із урни навмання беруть 5 кульок. Знайти ймовірність того, що серед вибраних кульок є 2 сині та 3 червоні.

16. На книжковій полиці в довільному порядку розміщені книги з п'ятитомного зібрання творів. Яка ймовірність того, що всі томи розміщені в порядку зростання номерів?

17. На книжковій полиці в довільному порядку розміщені книги з десятитомного зібрання творів. Яка ймовірність того, що третій, четвертий, п'ятий томи після переставляння стоятимуть поруч в порядку зростання номерів?

18. На книжковій полиці в довільному порядку розміщені книги з восьмитомного зібрання творів. Яка ймовірність того, що п'ятий, шостий, сьомий томи після переставляння стоятимуть поруч в порядку зростання номерів?

19. Серед 17 студентів групи, в якій є 7 дівчат розігруються 5 квитків на концерт. Знайти ймовірність того, що квитки виграють 4 дівчини.

20. Шістнадцять варіантів контрольної роботи написані на окремих картках і розподіляються випадково серед 14 студентів, які сидять в одному ряду. Кожний студент отримує одну картку. Знайти ймовірність того, що:

а) варіанти 1 і 2 не будуть використані;

б) варіанти 1 і 2 одержать студенти, які сидять поруч.

21. Студент знає відповіді до 45 із 60 питань програми. Кожен екзаменаційний білет містить 2 питання. Яка ймовірність того, що студент відповість на обидва питання?

22. Серед 4000 перших простих чисел натурального ряду є 551 просте число. Знайдіть частість появи простого числа.

23. Проведено 100 пострілів у мішень. Виявлено, що відносна частота влучень дорівнює 0,85. Скільки пострілів були влучними?

24. Під час перевірки готової продукції було виявлено 5 бракованих одиниць товару з 200 перевірених. Знайти відносну частоту бракованих одиниць товару.

25. На проміжку [0; 5] випадково вибирають два дійсних числа. Яка ймовірність того, що:

а) сума чисел менша за 4;

б) добуток вибраних чисел більший за 5;

в) модуль різниці вибраних чисел менший за 2, а їх добуток більший за 3.

26. На відрізку [0, l] навмання вибирають два числа. Знайти ймовірність того, що сума цих чисел не перевищує 1.

27. На площині проведено паралельні прямі на відстані 15 см одна від одної. На площину кидають монету діаметром 5 см. Знайти ймовірність того, що монета перетне одну із паралельних прямих.

28. На відрізку одиничної довжини навмання вибирається точка. Знайти ймовірність того, що відстань від цієї точки до кожного з кінців відрізка буде більша, ніж k, де k > 2.

Розділ 2

ОСНОВНІ ТЕОРЕМИ ТЕОРІЇ ЙМОВІРНОСТЕЙ

2.1. Умовна ймовірність. Залежні та незалежні події

Якщо при обчисленні ймовірності події A не вказуються ніякі умови, крім комплексу умов, то такі ймовірності називаються **безумовними** і позначаються $P(A)$.

Ймовірність настання події A, обчислена в припущені, що подія B уже відбулася, називається **умовною ймовірністю** події A при умові B і позначається

$$P(A/B) \text{ або } P_B(A).$$

Словами це означає: ймовірність події A при умові, що подія B уже відбулась.

Приклад 2.1. В ящику міститься 5 червоних і 7 білих куль. Із ящика два рази виймають по одній кулі. Знайти ймовірність того, що за другим випробуванням буде вийнята біла куля (подія A), якщо першою була вийнята також біла куля (подія B).

Розв'язання. Всього в ящику є 12 $(5 + 7 = 12)$ куль. Після першого випробування в урні залишилось 11 куль, із них 6 білих.

Отже, шукана ймовірність, на основі класичного означення ймовірностей, дорівнює:

$$P_B(A) = \frac{6}{11}.$$

Дві події A і B називаються **незалежними**, якщо ймовірність появи однієї із них не залежить від появи або непояви іншої події.

Події $A_1, A_2, \ldots A_n$ називаються **незалежними в сукупності**, якщо ймовірність кожної з них не змінюється через настання або не настання інших подій окремо або в будь якій комбінації.

2.2. Теорема додавання ймовірностей несумісних подій

Сумою двох подій $A + B$ називають подію, що полягає у появі події A або події B, або обох цих подій.

Наприклад, якщо зроблено два постріли A – подія, що означає влучання при першому пострілі, а B – при другому пострілі, то $A+B$ означає влучання при першому або другому, або в обох пострілах.

Зокрема, якщо дві події A і B несумісні, то $A+B$ – подія, що полягає в появі однієї із цих подій, або події A, або події B.

Сумою $\sum_{\text{i}} A_{\,i}$ подій $A_1, A_2, \dots A_n$ називають таку подію, поява якої рівносильна появі принаймні однієї з подій A_i , $i=1, 2, \dots , n$.

Теорема 2.1. Ймовірність появи однієї з кількох несумісних подій, не важливо якої, дорівнює сумі ймовірностей цих подій.

$$P(A+B) = P(A) + P(B). \qquad (2.1)$$

$$P(\textstyle\sum_i^n A_i) = P(A_1 + A_2 + \ \dots + A_n) = P(A_1) + P(A_2) + \ \dots + P(A_{1n}). \qquad (2.2)$$

Приклад 2.2. В ящику знаходиться 6 білих і 4 чорних кулі. З ящика виймають навмання 2 кулі. Яка ймовірність того, що вони однокольорові?

Розв'язання. Нехай подія A – поява двох білих куль, а подія B – поява двох чорних куль. Події A та B несумісні.

Число всіх єдиноможливих, рівноможливих і несумісних випадків дорівнює числу пар, які можна утворити з десяти різних куль.

$$n = C_{10}^3 = \frac{9 \cdot 10}{2} = 45$$

– число загальних випадків.

Число випадків, сприятливих для появи двох білих куль, дорівнює $C_6^2 = 15$, двох чорних – $C_4^2 = 6$.

Тоді за формулою (2.1) маємо:

$$P(A+B) = P(A) + P(B) = \frac{15}{45} + \frac{6}{45} = \frac{7}{15}.$$

2.3. Теорема множення залежних і незалежних подій

Добутком двох подій *A і B* називають таку третю подію *C*, що полягає в одночасній появі цих подій.

Добутком $\prod_i A_i$ подій A_1, A_2, ...A_n називають таку подію, поява якої рівносильна появі кожної з цих подій A_i, $i=1, 2, \ldots ,n$.

Якщо події *A* і *B* залежні між собою, причому відомі $P(A)$, $P_B(A)$, то ймовірність добутку цих подій, тобто ймовірність того, що з'явилась подія *A* і *B* шукається за допомогою теореми.

Теорема 2.2. Якщо події *A* і *B* незалежні між собою, то ймовірність добутку цих подій дорівнює добутку ймовірностей цих подій, тобто:

$$P(A \cdot B) = P(A) \cdot P(B). \qquad (2.3)$$

Ймовірність появи кількох подій, незалежних сукупно, обчислюється за формулою:

$$P\left(\prod_i A_i\right) = P(A_1 \cdot \mathrm{A}_2 \cdot \ldots \cdot A_n) = P(A_1) \cdot P(\mathrm{A}_2) \cdot \ldots \cdot P(A_1). \qquad (2.4)$$

Приклад 2.3. У трьох коробках знаходиться по 10 деталей. У першій коробці – 8, в другій – 7 і в третій – 9 стандартних деталей. Із кожної коробки навмання виймають по одній деталі. Знайти ймовірність того, що всі три вийняті деталі будуть стандартними.

Розв'язання. Ймовірність того, що з першої коробки вийнята стандартна деталь (подія *A*), дорівнює:

$$P(A) = \frac{8}{10} = 0{,}8.$$

Ймовірність того, що з другої коробки вийнята стандартна деталь (подія *B*), дорівнює:

$$P(B) = \frac{7}{10} = 0{,}7.$$

Ймовірність того, що із третьої коробки вийнята стандартна деталь (подія С), дорівнює:

$$P(C) = \frac{9}{10} = 0{,}9.$$

Оскільки події *A, B, C* незалежні, то шукана ймовірність (за теоремою множення – формула (2.4) дорівнює:

$$P(A \cdot B \cdot C) = P(A) \cdot P(B) \cdot P(C) = 0{,}8 \cdot 0{,}7 \cdot 0{,}9 = 0{,}504.$$

Теорема 2.3. Ймовірність добутку двох залежних подій дорівнює добутку ймовірності однієї із цих подій на умовну ймовірність другої, обчислену за умови, що відбулась перша подія.

$$P(A \cdot B) = P(A) \cdot P_A(B). \qquad (2.5)$$

Якщо маємо три події *A, B, C,* залежні між собою, то теорема множення ймовірностей матиме вигляд:

$$P(A \cdot B \cdot \mathrm{C}) = P(A) \cdot P_A(B) \cdot P_{AB}(C). \qquad (2.6)$$

Оскільки $A \cdot B = B \cdot A$, то і для множення ймовірностей має місце закон комутативності:

$$P(A \cdot B) = P(A) \cdot P_A(B) = P(B) \cdot P_B(A). \qquad (2.7)$$

Приклад 2.4. Студент знає 20 із 25 питань програми. Знайти ймовірність того, що студент знає три питання, які були в екзаменаційному білеті.

Розв'язання. Ймовірність того, що студент знає перше питання (подія *A*), дорівнює:

$$P(A) = \frac{20}{25}.$$

Ймовірність того, що студент знає друге питання (подія *B*), дорівнює:

$$P(B) = \frac{19}{24}$$

(оскільки події *A, B* залежні між собою).

Ймовірність того, студент знає третє питання (подія *C*), дорівнює:

$$P(C) = \frac{18}{23}.$$

(оскільки події *A, B, C* залежні між собою).

На основі формули (2.6) маємо:

$$P(A \cdot B \cdot \mathrm{C}) = \frac{20}{25} \cdot \frac{19}{24} \cdot \frac{18}{23} = \frac{57}{115} \approx 0{,}5\ .$$

2.4. Теорема додавання ймовірностей сумісних подій

Суму двох сумісних подій можна подати у вигляді:

$$\mathrm{A} + \mathrm{B} = \mathrm{A} \cdot \overline{\mathrm{B}} + \mathrm{B} \cdot \overline{\mathrm{A}} + \mathrm{A} \cdot \mathrm{B}.$$

Геометрично суму двох сумісних подій *A* і *B* показано на рис 2.1.

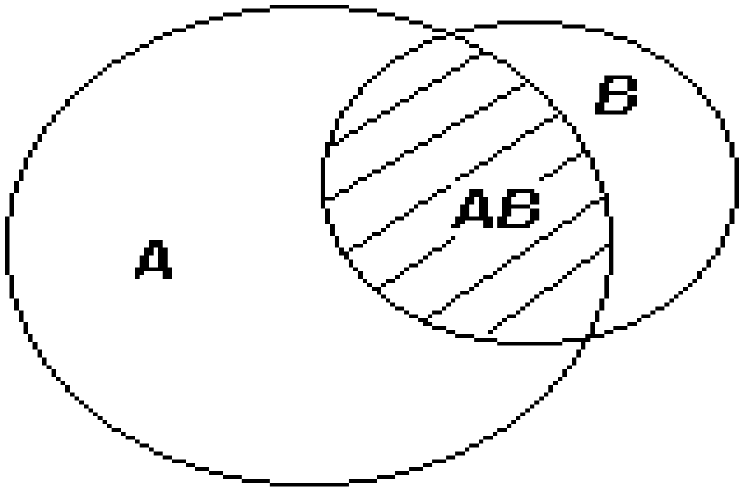

Рис. 2.1. Сума сумісних подій

Теорема 2.4. Ймовірність появи хоча б однієї із двох сумісних подій дорівнює сумі ймовірностей цих подій без імовірності добутку цих подій, тобто:

$$P(A + B) = P(A) + P(B) - P(A \cdot B) \qquad (2.8)$$

Зауваження: при використанні формули (2.8) треба мати на увазі, що події *A* і *B* можуть бути і залежними, і незалежними.

Для сумісних і незалежних подій теорема додавання запишеться формулою:

$$P(A + B) = P(A) + P(B) - P(A) \cdot P(B). \qquad (2.9)$$

Для сумісних і залежних подій теорема додавання запишеться формулою:

$$P(A + B) = P(A) + P(B) - P(A) \cdot P_A(B). \qquad (2.10)$$

Приклад 2.5. Для виконання завдання замовник звернувся до двох виконавців. Ймовірність того, що перший виконавець надасть згоду виконати замовлення дорівнює 0,8, а другий – 0,9. Знайти ймовірність того, що замовлення буде прийнято принаймні одним з цих виконавців.

Розв'язання. Позначимо події:

A – завдання прийнято для виконання;

A_1 – перший виконавець дав згоду виконати завдання;

A_2 – другий виконавець дав згоду виконати завдання.

Тоді $A = A_1 + A_2$. Оскільки згода виконати замовлення першим виконавцем не виключає згоди виконання замовлення іншим, події сумісні. Тоді:

$$P(A) = P(A_1) + P(A_2) - P(A_1 \cdot A_2).$$

Події A_1 та A_2 незалежні, тому що ймовірність згоди кожного виконавця не залежить від того, чи дав згоду інший виконавець. Тому

$$P(A_1 \cdot A_2) = P(A_1) \cdot P(A_2).$$

Отже:

$$P(A) = 0{,}8 + 0{,}9 - 0{,}8 \cdot 0{,}9 = 0{,}98.$$

2.5. Протилежні події. Ймовірність протилежної події

Протилежними навивають дві єдиноможливі події, які утворюють повну групу подій. Подію, протилежну події A (тобто ненастання події A), позначають через $\bar{A}$.

Наприклад: влучання і промах при пострілі в задану ціль – протилежні події. Складання іспиту і провал іспиту – протилежні події. Якщо подія A – влучання в ціль, то $\bar{A}$ – промах. Якщо A – подія складання іспиту, то $\bar{A}$ – подія провалу іспиту.

Теорема 2.5. Ймовірність суми протилежних подій дорівнює одиниці, тобто:

$$\mathrm{P(A)} + \mathrm{P}(\bar{\mathrm{A}}) = 1. \quad (2.11)$$

Приклад 2.6. Ймовірність того, що банк надасть кредит, дорівнює 0,7. Знайти ймовірність того, що банк відмовить у кредитуванні.

Розв'язання. Події A – банк надасть кредит та $\bar{A}$ – банк відмовить у кредиті – протилежні. Використовуючи формулу (2.11) ймовірність буде дорівнювати:

$$P(\bar{A}) = 1 - P(A) = 1 - 0{,}7 = 0{,}3.$$

Приклад 2.7. В ящику в довільному порядку розкладено 20 деталей, причому 5 з них стандартні. Навмання беруть три деталі. Знайти ймовірність того, що принаймні одна з них буде стандартною (подія A).

Розв'язання. Задачу можна розв'язати двома способами.

Перший спосіб. Принаймні одна з взятих деталей буде стандартною, якщо відбудеться одна з трьох несумісних подій:

A_1 – одна деталь стандартна і дві нестандартні,

A_2 – дві деталь стандартна і одна нестандартні,

A_3 – три деталі стандартні.

Подію A можна записати у вигляді суми цих трьох подій:

$$A = A_1 + A_2 + A_3.$$

За теоремою додавання формула (2.2) маємо:

$$P(A) = P(A_1) + P(A_2) + P(A_3).$$

Знаходимо ймовірність кожної з цих подій (використовуємо задачу про вибірку приклад 1.8):

$$P(A_1) = \frac{C_5^1 \cdot C_{15}^2}{C_{20}^3} = \frac{5}{1} \cdot \frac{14 \cdot 15}{1 \cdot 2} \cdot \frac{1 \cdot 2 \cdot 3}{18 \cdot 19 \cdot 20} = \frac{35}{76};$$

$$P(A_2) = \frac{C_5^2 \cdot C_{15}^1}{C_{20}^3} = \frac{4 \cdot 5}{1 \cdot 2} \cdot \frac{15}{1} \cdot \frac{1 \cdot 2 \cdot 3}{18 \cdot 19 \cdot 20} = \frac{5}{38};$$

$$P(A_3) = \frac{C_5^3}{C_{20}^3} = \frac{3 \cdot 4 \cdot 5}{1 \cdot 2 \cdot 3} \cdot \frac{1 \cdot 2 \cdot 3}{18 \cdot 19 \cdot 20} = \frac{1}{114}.$$

Додавши знайдені величини, отримаємо:

$$P(A) = \frac{35}{76} + \frac{5}{38} + \frac{1}{114} = \frac{137}{228}.$$

Другий спосіб. Події *A* (принаймні одна з взятих деталей буде стандартною) і Ā (жодна зі взятих деталей не стандартна) протилежні, тому за формулою:

$$P(A) + P(\bar{A}) = 1, \text{ або } P(A) = 1 - P(\bar{A}). \qquad (2.11)$$

Ймовірність появи події Ā дорівнює:

$$P(\bar{A}) = \frac{C_{15}^3}{C_{20}^3} = \frac{13 \cdot 14 \cdot 15}{1 \cdot 2 \cdot 3} \cdot \frac{1 \cdot 2 \cdot 3}{18 \cdot 19 \cdot 20} = \frac{91}{228}.$$

Отже, шукана ймовірність:

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{91}{228} = \frac{137}{228}.$$

2.6. Ймовірність появи хоча б однієї події

Нехай в результаті випробовування може з'явитися n подій, причому відомі ймовірності цих подій. Ймовірність того, що з'явиться хоча б одна із цих подій шукається за теоремою.

Теорема 2.6. Ймовірність появи хоча б однієї із подій A_1, A_2, ... A_n незалежних сукупно, дорівнює різниці між одиницею і добутком ймовірностей протилежних подій $\overline{A_1}, \overline{A_2}, \ldots, \overline{A_n}$.

$$P(A) = 1 - q_1 \cdot q_2 \cdot \ldots \cdot q_n, \qquad (2.12)$$

де $q_{\text{i}} = 1 - p_i$ – ймовірності відповідних протилежних подій.

Приклад 2.8. Ймовірність влучання в ціль із трьох рушниць, відповідно, дорівнює $p_1=0{,}7$, $p_2=0{,}8$, $p_3=0{,}9$. Знайти ймовірність хоч би одного влучання (подія *A*) при одному залпі із трьох рушниць.

Розв'язання. Події A_1, A_2, A_3 – влучання в ціль із трьох рушниць незалежні сукупно. Ймовірність промахів, відповідно, будуть рівні:

$$q_1 = 1 - p_1 = 1 - 0{,}7 = 0{,}3 = P(\overline{A_1});$$

$$q_2 = 1 - p_2 = 1 - 0{,}8 = 0{,}2 = P(\overline{A_2});$$

$$q_3 = 1 - p_3 = 1 - 0{,}9 = 0{,}1 = P(\overline{A_3}).$$

Шукана ймовірність за формулою (2.12) буде дорівнювати:

$$P(A) = 1 - 0{,}3 \cdot 0{,}2 \cdot 0{,}1 = 0{,}994.$$

Зокрема, якщо події A_1, A_2, ... A_n мають однакову ймовірність, рівну p, тобто $P(A_1) = P(A_2) = \cdots = P(A_n) = p$, то ймовірність появи – хоч би однієї із цих подій дорівнює:

$$P(A) = 1 - q^n \,. \qquad (2.13)$$

Приклад 2.9. Укладено 4 однотипних договори страхування терміном на 1 рік. Ймовірність того, що за будь-яким договором протягом року надійде запит на відшкодування, дорівнює 0,1. Припускаючи, що надходження запитів незалежні між собою, знайти ймовірність того, що

А. Запити надійдуть за чотирма договорами.

Б. Запити не надійдуть за жодним договором.

В. Запит надійде принаймні за одним договором.

Розв'язання. Нехай A_i , *i=1, 2, 3, 4* – подія, яка полягає в тому, що запит надійде за *i*-тим договором. Тоді $\overline{\mathrm{A}_i}$ протилежна подія до події $\mathrm{A}i$. Тому:

$$P(\mathrm{A}_i) = p = 0{,}1,\ P(\overline{\mathrm{A}_i}) = q = 1 - p = 1 - 0{,}1 = 0{,}9.$$

А. Оскільки події A_i незалежні, то за формулою незалежних подій:

$$P(\mathrm{A}_1 \cdot \mathrm{A}_2 \cdot \mathrm{A}_3 \cdot \mathrm{A}_4) = P(\mathrm{A}_1) \cdot P(\mathrm{A}_2) \cdot P(\mathrm{A}_3) \cdot P(\mathrm{A}_4) = 0{,}1^4 = 0{,}0001.$$

Б.

$$P(\overline{\mathrm{A}_1} \cdot \overline{\mathrm{A}_2} \cdot \overline{\mathrm{A}_3} \cdot \overline{\mathrm{A}_4}) = P(\overline{\mathrm{A}_1}) \cdot P(\overline{\mathrm{A}_2}) \cdot P(\overline{\mathrm{A}_3}) \cdot P(\overline{\mathrm{A}}_4) = 0{,}9^4 = 0{,}6561.$$

В. Нехай подія *A* – запит надійде принаймні за одним договором є протилежною подією до події $\overline{\mathrm{A}}$ – запит не надійде за жодним договором. Скориставшись формулою (2.13) одержимо:

$$P(A) = 1 - 0{,}9^4 = 0{,}3439.$$

2.7. Формула повної ймовірності

Нехай подія B може відбутися з однією і лише однією із n-попарно несумісних подій $H_1, H_2, \ldots H_n$, тобто

$$A = A \cdot H_1 + A \cdot H_2 + \cdots + A \cdot H_n.$$

Потрібно знайти ймовірність події A. На основі теореми додавання несумісних подій маємо:

$$P(A) = P(A \cdot H_1) + P(A \cdot H_2) + \cdots + P(A \cdot H_n).$$

Далі, використавши теорему множення залежних подій, отримаємо:

$$P(A) = P(H_1) \cdot P_{H_1}(A) + P(H_2) \cdot P_{H_2}(A) + \cdots + P(H_n) \cdot P_{H_n}(A)$$

або

$$P(A) = \sum_{i=1}^{n} P(H_i) \cdot P_{H_i}(A). \qquad (2.14)$$

Формулу (2.14) називають **формулою повної ймовірності**.

Подій $H_1, H_2, \ldots H_n$, які утворюють повну груп, тобто:

$$P(H_1) + P(H_2) + \cdots + P(H_n) = 1. \qquad (2.15)$$

Приклад 2.10. Маємо дві коробки деталей. Ймовірність того, що в першій коробці деталь стандартна, дорівнює 0,8, а в другій – 0,9. Знайти ймовірність того, що взята навмання деталь із будь-якої коробки буде стандартна.

Розв'язання. Нехай A – подія – вийнята стандартна деталь.

H_1 – деталь може бути вийнята із першої коробки;

H_2 – деталь може бути вийнята з другої коробки, тобто:

$$A = A \cdot H_1 + A \cdot H_2.$$

На основі формули повної імовірності (2.14), маємо:

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

Ймовірність вибору коробки:

$$P(H_1) = \frac{1}{2}; \qquad P(H_2) = \frac{1}{2}.$$

Ймовірність стандартної деталі з відповідної коробки:

$$P_{H_1}(A) = 0{,}8\ ; \qquad P_{H_2}(A) = 0{,}9.$$

Підставивши значення, одержимо:

$$P(A) = \frac{1}{2} \cdot 0{,}8 + \frac{1}{2} \cdot 0{,}9 = 0{,}85.$$

2.8. Ймовірність гіпотез. Формула Байєса

Нехай знову подія A може відбутися з однією і лише однією із n-попарно незалежних подій H_1, H_2, ... H_n, тобто:

$$A = A \cdot H_1 + A \cdot H_2 + \cdots + A \cdot H_n.$$

Події H_1, H_2, ...H_n, називають **гіпотезами**.

Нехай подія A – відбулася. Треба знайти ймовірність події H_i, із якою відбулася подія A, тобто $P_A(H_i)$. На основі теореми множення ймовірностей, маємо:

$$P(H_i A) = P_{H_i}(A) \cdot P(H_i) = P(A) \cdot P_A(H_i),$$

звідси:

$$P_A(H_i) = \frac{P(H_i) P_{H_i}(A)}{P(A)}.$$

Використавши формулу повної імовірності (2.14), одержимо:

$$P_A(H_i) = \frac{P(H_i) P_{H_i}(A)}{\sum_{i=1}^{n} P(H_i) \cdot P_{H_i}(A)}. \qquad (2.16)$$

Формулу (2.15) називають **формулою Байєса** (за ім'ям англійського математика, який опублікував в 1764 р. цю формулу) або **формулами гіпотез**.

Формула Байєса дає можливість переоцінити ймовірність гіпотез після того, як стає відомим результат випробування, внаслідок якого з'явилась подія *A*.

Приклад 2.11. Маємо 6 ящиків: 2 ящики із складу H_1, в яких є 3 білі та 2 червоні кулі, 1 ящик із складу H_2, в якому є 4 білі та 1 червона кулі, 3 ящики зі складу H_3, в яких є 2 білі і 3 червоні кулі. Навмання вибирають ящик і навмання вибирають кулю. Знайти ймовірність того, що

А. Біла куля вийнята із складу H_1.

Б. Біла куля вийнята із складу H_2.

В. Біла куля вийнята із складу H_3.

Розв'язання.

Ймовірність вибору ящиків з першого складу: $P(H_1) = \frac{2}{6}$.

Ймовірність вибору ящиків з другого складу: $P(H_2) = \frac{1}{6}$.

Ймовірність вибору ящиків з третього складу: $P(H_3) = \frac{3}{6}$.

Подія *A* – вийнята біла куля.

Ймовірність білої кулі при умові вибору першого складу: $P_{H_1}(A) = \frac{3}{5}$.

Ймовірність білої кулі при умові вибору другого складу: $P_{H_2}(A) = \frac{4}{5}$.

Ймовірність білої кулі при умові вибору третього складу: $P_{H_3}(A) = \frac{2}{5}$.

Знайдемо за формулою повної ймовірності (2.14):

$$P(A) = \sum_{i=1}^{3} P(H_i) \cdot P_{H_i}(A) = \frac{2}{6} \cdot \frac{3}{5} + \frac{1}{6} \cdot \frac{4}{5} + \frac{3}{6} \cdot \frac{2}{5} = \frac{16}{30} = \frac{8}{15}.$$

За формулою Байєса (2.16), маємо:

А. $P_A(H_1) = \frac{P(H_1)P_{H_1}(A)}{\sum_{i=1}^{n} P(H_i) \cdot P_{H_i}(A)} = \frac{\frac{2}{6} \cdot \frac{3}{5}}{\frac{16}{30}} = \frac{3}{8} \approx 0{,}38.$

Б. $P_A(H_2) = \frac{P(H_2)P_{H_2}(A)}{\sum_{i=1}^{n} P(H_i) \cdot P_{H_i}(A)} = \frac{\frac{1}{6} \cdot \frac{4}{5}}{\frac{16}{30}} = \frac{4}{16} = 0{,}25.$

В. $P_A(H_3) = \frac{P(H_3)P_{H_3}(A)}{\sum_{i=1}^{n} P(H_i) \cdot P_{H_i}(A)} = \frac{\frac{3}{6} \cdot \frac{2}{5}}{\frac{16}{30}} = \frac{3}{8} \approx 0{,}38.$

Порівнюючи вказані ймовірності, можна зробити висновок, що біла куля була вийнята із коробок першого або третього складів, оскільки ймовірність цих подій найбільша.

Запитання для самоконтролю

1. Які події називаються незалежними?
2. Що називають умовною ймовірністю події A за умови появи події B?
3. Ймовірності яких подій називаються безумовними.
4. Що називають сумою подій?
5. Сформулюйте теорему додавання ймовірностей несумісних подій.
6. Які події називаються протилежними?
7. Чому рівна сума протилежних подій?
8. За якою формулою шукається ймовірність появи хоча б однієї події?
9. Запишіть формулу для знаходження ймовірність появи хоча б однієї подій, якщо виконується рівність $P(A_1) = P(A_2) = \cdots = P(A_n) = p$.
10. Що називають добутком подій?
11. Сформулюйте теорему множення ймовірностей залежних подій.
12. Сформулюйте теорему множення ймовірностей незалежних подій.
13. Запишіть формулу додавання сумісних і залежних подій.
14. Запишіть формулу додавання сумісних і незалежних подій.
15. Виведіть формулу повної ймовірності.
16. Чому рівна сума подій, які утворюють повну групу подій?
17. Виведіть формулу Байєса.

Задачі для самостійного розв'язування

1. В урні знаходиться 6 білих і 4 чорних кулі. З урни виймають навмання 2 кулі. Яка ймовірність того, що вони однокольорові?

2. У коробці знаходиться 30 куль. Із них 15 червоних, 10 синіх і 5 білих. Із коробки навмання виймають кулю. Знайти ймовірність того, що вийнята куля – кольорова.

3. В ящику в довільному порядку розкладено 20 деталей, причому 5 з них стандартних. Беруть три деталі. Знайти ймовірність того, що принаймні одна з узятих деталей буде стандартною (подія A).

4. Під час складання іспиту студентом ймовірність одержання п'ятірки дорівнює 0,3; четвірки – 0,45; двійки – 0,1; не з'явитись на екзамен – 0,05. Яка ймовірність того, що студент одержить позитивну оцінку?

5. У відділі технічного контролю швейної фабрики представлено на огляд 80 жіночих костюмів, з них 50 одного фасону і 30 – іншого. Визначити ймовірність того, що взяті навмання для огляду два костюми виявляться різних фасонів.

6. Стан використання підприємствами наданих їм кредитів контролюється центральним та обласним банками. Центральний банк перевіряє вибірково 10%, а обласний – 20% підприємств. Яка ймовірність того, навмання вибране підприємство буде перевірене центральним банком і не перевірене обласним, якщо рішення щодо перевірки на різних рівнях ухвалюється незалежно?

7. Продавець обслуговує в магазині два відділи. Ймовірність того, що протягом певного часу йому доведеться відпустити товар з I відділу, дорівнює 0,8; з II відділу – 0,7. Яка ймовірність того, що протягом певного часу продавець не відпускатиме товару?

8. Три студенти одночасно складають іспит. Ймовірність складання іспиту на «відмінно» першим студентом – 0,2; другим – 0,5; третім – 0,3. Знайти ймовірність того, що три студенти складуть іспит на «відмінно».

9. Прилад складається з трьох вузлів, кожний з яких незалежно від інших може протягом часу t відмовити у роботі. Вихід з ладу хоча б одного вузла призводить до зупинки приладу загалом. За час t ймовірність безвідмовної роботи першого вузла дорівнює 0,8; другого – 0,9; третім – 0,7. Знайти надійність приладу загалом.

10. Серед 50 електроламп є три нестандартні. Знайти ймовірність того, що дві взяті одночасно електролампи виявляються нестандартними.

11. Ймовірність того, що подія відбудеться хоча б один раз у трьох незалежних випробуваннях, дорівнює 0,916. Знайти ймовірність появи події в одному випробуванні.

12. Для отримання кредиту підприємець звертається до двох банків. Ймовірність того, що перший банк не відмовить в наданні кредиту становить 0,75, інший – 0,8. Знайти ймовірність того, що:

а) хоча б один банк дасть згоду на кредитування;

б) обидва банки відмовляться надавати кредит;

в) один дасть згоду на кредитування.

13. У першій урні 10 куль, з них 8 білих. У другій урні 20 куль, з них 4 білих. Із кожної урни навмання взято по одній кулі, а потім із двох обраних навмання взято одну. Знайти ймовірність того, що остання куля буде білою.

14. Студент розшукує потрібну йому формулу у трьох довідниках. Ймовірність того, що формула міститься у першому довіднику дорівнює 0,6, у другому – 0,7, в третьому – 0,8. Знайти ймовірність того, що формула міститься:

а) тільки в одному довіднику;

б) тільки в двох довідниках;

в) в усіх трьох довідниках;

г) хоча б в одному із довідників.

15. Бізнесмен має контракти з трьома банками і може брати кредити в кожному з них. Протягом 5 попередніх років перший банк погодився надати кредит 6 разів, другий банк – 7 разів, третій – 9 разів при 10 зверненнях до

кожного з них. Яка ймовірність того, що нині принаймні один з банків видасть бізнесменові кредит?

16. Ймовірність того, що під час роботи комп'ютера станеться збій в арифметичному пристрої, у оперативній пам'яті або в пристрої введення, співвідносяться як 2:1:3. Ймовірність зафіксувати збій у цих пристроях дорівнюють, відповідно, 0,9; 0,75; 0,7. Знайти ймовірність фіксування збою в роботі комп'ютера.

17. На складання надходять деталі з трьох автоматів. Перший автомат дає 20%, другий – 30%, третій – 50% деталей такого типу, які надходять на складання. Перший автомат дає 0,2% браку, другий – 0,3%, третій – 0,1%. Знайти ймовірність того, що на складання надійшла бракована деталь.

18. У змаганнях беруть участь п'ять команд по 5 чоловік. У перших двох командах є по два майстри спорту, у третій і четвертій – по 1, а в п'ятій команді – 4 майстри. Один майстер спорту завоював «золото». Знайти ймовірність того, що «золото» завоював спортсмен:

1) з перших двох команд;
2) з третьої і четвертої;
3) з п'ятої команди.

19. На склад надходять однакові електричні лампи. Перший завод постачає 70%, другий – 30% усієї кількості ламп. Відомо, що перший завод випускає 95% продукції, яка може відпрацювати встановлений термін, а другий – 92%. Взята навмання електролампа відпрацювала встановлений термін. Знайти ймовірність того, що ця лампа поступила з першого заводу, з другого заводу?

20. Два економісти заповнюють документи, які складають у спільну папку. Ймовірність помилки для першого економіста дорівнює 0,1, для другого – 0,2. Перший економіст заповнив 40 документів, другий – 60. Під час перевірки у навмання взятому з папки документі виявили помилку. Знайти за класичним способом ймовірність того, що документ складав перший економіст.

21. До центру статистичних досліджень надходять данні з трьох пунктів: з першого – 50%, з другого – 30%, з третього – 20% всіх даних. Ймовірність припущення помилки під час обробки статистичних даних у першому пункті дорівнює 0,1, у другому – 0,05, у третьому – 0,15. Яка ймовірність того, що отримані центром дані достовірні?

22. Виробництво певної продукції може проводитись у трьох температурних режимах з ймовірностями 0,45; 0,25; 0,3 відповідно. Залежно від температурного режиму ймовірність отримання продукції вищої якості становить 0,8; 0,85; 0,9. Яка ймовірність того, що: а) виготовлена продукція вищої якості?

б) продукцію вищої якості виготовлено при третьому температурному режимі?

23. До каси підприємства надійшли банкноти у пачках від двох банків: 50 пачок від першого і 70 – від другого. Ймовірність помилки касирів першого банку становить 0,15, другого – 0,2. Яка ймовірність того, що:

а) навмання вибрану пачку сформовано без помилок?

б) пачку без помилок було сформовано касирами другого банку?

24. У рекламному агентстві працює три групи дизайнерів: перша обслуговує 25 фірм, друга – 45, третя – 40. Протягом одного місяця кошти, витрачені на рекламу дизайнерами першої групи, повертаються до 40% фірм, другої – до 45%, третьої – до 35%. Яка ймовірність того, що:

а) навмання вибрана фірма окупила витрачені на рекламу кошти протягом місяця?

б) фірма, що окупила протягом місяця витрачені на рекламу кошти, обслуговувалась третьою групою дизайнерів?

25. Магазин отримує продукцію від двох виробників: перший постачає 2/5 усіх виробів, другий – 3/5. Ймовірність продажу виробів першого постачальника становить 0,95, другого – 0,8. Яка ймовірність того, що:

а) навмання вибраний виріб не буде реалізовано?

б) нереалізований виріб отримано від першого виробника?

26. Справи клієнтів банку зберігаються у 8 сейфах: у трьох по 150 справ, у п'яти – по 250. Ймовірність вчасного повернення кредиту клієнтами, справи яких лежать у перших трьох сейфах, становить 0,96, в останніх п'яти – 0,95. Яка ймовірність того, що:

а) навмання вибрано справу клієнта, який вчасно поверне кредит?

б) справа клієнта, який своєчасно повернув кредит, лежала в одному з перших трьох сейфів?

27. Виробник комп'ютерів отримує комплектуючі деталі від трьох постачальників, частки яких становлять 20%, 45% та 35%. Деталі першого постачальника мають 2% браку, другого 1,5%, третього – 1,7%. Яка ймовірність того, що:

а) навмання вибрана деталь буде з браком?

б) браковану деталь отримано від другого постачальника?

28. Із 30 студентів групи, що прийшли на екзамен, 8 підготовлені відмінно, 10 – добре, 8 – задовільно, а решта – незадовільно. Програма екзамену включає 50 питань. Білет містить 3 питання. Студент, підготовлений відмінно, знає всі питання; добре – 40 питань; задовільно – 25 питань і незадовільно – 10 питань. Знайти ймовірність того, що навмання викликаний студент відповість на всі 3 питання білета.

29. У крамницю для продажу надійшла продукція з трьох фабрик, відносними частками яких є: 1 – 50%; 2 – 30%; 3 – 20%. Брак продукції цих фабрик становить 2%; 3% та 5% відповідно. Знайти ймовірність того, що навмання куплений у магазині виріб цих фабрик буде якісним.

Розділ 3

ПОВТОРНІ ВИПРОБУВАННЯ

3.1. Послідовність незалежних випробувань. Формула Бернуллі

На практиці часто трапляються випадкові експерименти, які полягають в тому, що одне й те ж випробування проводять декілька разів поспіль. Водночас у кожному випробуванні деякі події можуть відбуватись, а можуть і не відбуватись. Наприклад, підкидання кубика чи монети декілька раз, перевірка на якість виробів, які взяті по одному із певної кількості.

Послідовні випробування називаються незалежними, якщо здійснення будь-якого результату в n-му за рахунком випробуванні не залежить від результатів у попередніх випробуваннях.

Нехай проводиться n послідовних випробувань, ймовірність події A в кожному випробуванні одна й та сама, а випробування незалежні, тобто поява події A в i-му випробуванні не залежить від її появи в інших випробуваннях, то таку серію випробувань називають **схемою Бернуллі**.

Якщо випадкова подія А відбувається в кожному випробуванні з ймовірністю $P(A) = p$, тоді вона не відбувається з ймовірністю $q = 1 - p$. Ймовірність того, що при n випробуваннях за схемою Бернуллі подія А відбудеться k разів визначається за **формулою Бернуллі**:

$$P_n(k) = C_n^k p^k q^{n-k}. \qquad (3.1)$$

Приклад 3.1. Укладено чотири однотипних договори страхування терміном на один рік. Ймовірність того, що за будь-яким договором протягом року надійде запит на відшкодування, дорівнює 0,1. Знайти ймовірність того, що надійде рівно два запити.

Розв'язання. За умовою $n = 4$, $k = 2$, $p = 0{,}1$, тоді

$$q = 1 - 0{,}1 = 0{,}9.$$

Скориставшись формулою Бернуллі (3.1) матимемо, що

$$P_4(2) = C_4^2 0{,}1^2 0{,}9^{4-2} = 0{,}0486.$$

Приклад 3.2. Статистикою встановлено, що із кожної тисячі дітей, що народились, в середньому народжується 485 дівчат і 515 хлопчиків. У сім'ї є 5 дітей. Знайти ймовірність того, що серед тих дітей:

А. 3 дівчинки.

Б. Не більше трьох дівчат.

Розв'язання.

А. Згідно з умовою задачі, ймовірність народження дівчинки $p = 0{,}485$; а ймовірність народження хлопчика $q = 0{,}515$; $n = 5$; $k = 3$.

За формулою Бернуллі маємо:

$$P_5(3) = C_5^3 (0{,}485)^3 (0{,}515)^2 = 0{,}31.$$

Б. Вимога, щоб в сім'ї було не більше трьох дівчат, буде виконуватися, коли серед 5-ти дітей в сім'ї:

1) немає ні однієї дівчинки – $P_5(0)$,

$$P_5(0) = C_5^0 (0{,}485)^5 (0{,}515)^0 = 0{,}02;$$

2) є одна дівчинка і 4 хлопчики – $P_5(1)$,

$$P_5(1) = C_5^1 (0{,}485)^1 (0{,}515)^4 = 0{,}17;$$

3) є дві дівчинки і 3 хлопчики – $P_5(2)$,

$$P_5(2) = C_5^2 (0{,}485)^2 (0{,}515)^3 = 0{,}33;$$

4) є три дівчинки і 2 хлопчики – $P_5(3)$,

$$P_5(3) = C_5^3 (0{,}485)^3 (0{,}515)^2 = 0{,}31.$$

Всі ці події несумісні, тому шукана ймовірність, за теоремою додавання несумісних подій, дорівнює сумі ймовірностей кожної із цих подій:

$$P_5(k \leq 3) = P_5(0) + P_5(1) + P_5(2) + P_5(3) = 0{,}83.$$

3.2. Найімовірніше число настання події під час повторних випробувань

Під час обчислення ймовірностей $P_n(k)$ для різних значень k будуть різні значення імовірності. Число k_0 (появи події в n незалежних випробуваннях) називається **найімовірнішим**, якщо ймовірність $P_n(k_0)$ – найбільша.

Для визначення найімовірнішого числа появи використовуємо нерівність:

$$np - q \leq k_0 \leq np + p. \qquad (3.2)$$

Можливі такі випадки:

1. Якщо $np - q$ – дробове число, то маємо єдине найімовірніше число.
2. Якщо $np - q$ – ціле число, то маємо два найімовірніших числа k_0 і k_0+1.
3. Якщо np – ціле число, то найімовірніше число $k_0 = np$.

Приклад 3.3. Укладено чотири однотипних договори страхування терміном на один рік. Ймовірність того, що за будь-яким договором протягом року надійде запит на відшкодування, дорівнює 0,1. Знайти найімовірніше число запитів.

Розв'язання. За умовою $n = 4$, $k = 1$, $p = 0{,}1$, тоді

$$q = 1 - 0{,}1 = 0{,}9.$$

Скориставшись нерівністю (3.2) матимемо:

$$\mathbf{4 \cdot 0,1 - 0,9 \leq \boldsymbol{k_0} \leq 4 \cdot 0,1 + 0,1}$$

або

$$\mathbf{-0,5 < \boldsymbol{k_0} < 0,5}.$$

Отже, найімовірніше число $\mathrm{k}_0 = 0$.

3.3. Локальна теорема Муавра–Лапласа

Формула Бернуллі дає можливість обчислити ймовірність того, що подія *A* з'явиться k раз в n незалежних випробуваннях, якщо ймовірність появи події в кожному випробуванні є сталою.

Але користування формулою Бернуллі при великих n призводить до труднощів, пов'язаних з громіздкими обчисленнями. Локальна теорема Муавра–Лапласа дає так звану асимптотичну формулу, яка дає можливість наближено знайти ймовірність $P_n(k)$, якщо число випробувань доволі велике.

Зауважимо, що для частинного випадку, коли $p=0{,}5$, асимптотична формула була знайдена ще в 1730 році Муавром. У 1783 році Лаплас узагальнив цю формулу для довільного значення p, $0 < p < 1$. Тому цю теорему часто називають теоремою Муавра-Лапласа.

Локальна теорема Лапласа: якщо ймовірність p настання події *A* в n незалежних випробуваннях стала і задовольняє нерівність $0 < p < 1$, то ймовірність того, що подія з'явиться k раз в n незалежних випробуваннях, наближено дорівнює:

$$P_n(k) \approx \frac{1}{\sqrt{npq}}\varphi(x), \qquad (3.3)$$

де

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x}{2}}, \quad x = \frac{k-np}{\sqrt{npq}}.$$

Функція $\varphi(x)$ – є парна, тобто виконується рівність

$$\varphi(-x) = \varphi(x).$$

Для функції $\varphi(x)$ складені таблиці, в яких за заданими значеннями x знаходять значення функції $\varphi(x)$ (див. – додатки, табл. 1).

Приклад 3.4. Записи страхової компанії показали, що 30% власників договорів страхування подали запити на відшкодування. Для перевірки було відібрано 15 чоловік, які мають договори. Знайти ймовірність того, що протягом року надійде п'ять запитів.

Розв'язання. За умовою задачі $n = 15, k = 5, p = 0{,}3$, тоді

$$q = 1 - 0{,}3 = 0{,}7.$$

Скористаємось локальною теоремою Лапласа, тому спочатку обчислимо

$$x = \frac{5 - 15 \cdot 0{,}3}{\sqrt{15 \cdot 0{,}3 \cdot 0{,}7}} \approx 0{,}28.$$

За таблицею 1 у додатку знайдемо значення функції $\varphi(x)$ при $x = 0{,}28$,

$$\varphi(0{,}28) = 0{,}3836.$$

Тоді за формулою (3.3) одержимо

$$P_{15}(5) = \frac{0{,}3836}{\sqrt{15 \cdot 0{,}3 \cdot 0{,}7}} = \frac{0{,}3836}{1{,}77} = 0{,}217.$$

3.4. Інтегральна теорема Лапласа

Часто потрібно зайти, при n-незалежних випробувань, в кожному із яких ймовірність появи події *A* стала і дорівнює p $(0 < p < 1)$, ймовірність того, що подія *A* відбудеться не менше k_1 і не більше k_2 раз $k_1 < k_2$. Цю ймовірність позначають: $P_n(k_1, k_2)$.

Отримати точний результат ймовірностей у цьому разі можна за допомогою інтегральної теореми Лапласа.

Інтегральна теорема Лапласа. Якщо ймовірність *P* настання події *A* в кожному випробуванні стала $(0 < p < 1)$, то ймовірність $P_n(k_1, k_2)$ того, що подія *A* відбудеться в «*n*» випробуваннях від k_1 до k_2 разів визначається наближено такою формулою:

$$P_n(k) = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{x}{2}} dx, \qquad (3.4)$$

де

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

Якщо ввести позначення: $\Phi(\mathrm{x}) = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{x}{2}} dx$, то формула (3.4) буде мати вигляд:

$$P_n(k_1, k_2) = \Phi(x_2) - \Phi(x_1). \qquad (3.5)$$

Функцію $\Phi(x)$ називають **функцією Лапласа**.

$\Phi(x)$ – є не парною функцією, тобто виконується рівність

$$\Phi(-x) = \Phi(x).$$

Функція $\Phi(x)$ табульована. В таблиці 2 Додатків наведені значення лише до x = 5, тому для $x > 5$ можна вважати, що $\Phi(x) = 0{,}5$.

Приклад 3.5. Записи страхової компанії показали, що 30% власників договорів страхування подали запити на відшкодування. Для перевірки було відібрано 15 чоловік, які мають договори. Знайти ймовірність того, що протягом року надійде принаймні 10 запитів.

Розв'язання. За умовою задачі $n = 15$, $k_1 = 10$, $k_2 = 15$

$$p = 0{,}3, \text{ тоді } q = 1 - 0{,}3 = 0{,}7.$$

Скористаємось інтегральною теоремою Лапласа, тому спочатку обчислимо:

$$x_1 = \frac{10 - 15\cdot 0{,}3}{\sqrt{15\cdot 0{,}3\cdot 0{,}7}} \approx 3{,}10;$$

$$x_2 = \frac{15 - 15\cdot 0{,}3}{\sqrt{15\cdot 0{,}3\cdot 0{,}7}} \approx 5{,}92.$$

За таблицею 2 у Додатках знайдемо значення функції $\Phi(x)$,

$$\Phi(3{,}10) = 0{,}49903\ , \Phi(5{,}92) = 0{,}5.$$

Тоді за формулою (3.5):

$$P_{15}(10{,}15) = P(10 \le k \le 15) = \Phi(5{,}92) - \Phi(3{,}10)$$

$$P(10 \le k \le 15) = 0{,}5 - 0{,}49903 = 0{,}00097.$$

Із теореми Лапласа випливає така теорема.

Теорема Бернуллі. Якщо p $(0 < p < 1)$ – постійне, при $n \to \infty$

$$P_n\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) \to 2\Phi\left(\varepsilon \cdot \sqrt{\frac{n}{pq}}\right). \tag{3.6}$$

Теорема Бернуллі при доволі великих n, p, q дає змогу оцінити ймовірність того, що в n незалежних випробуваннях абсолютна величина відхилення відносної частки появи події A від її ймовірності p не перевищить заданого додатного числа ε;

$$P_n\left(\left|\frac{m}{n}-p\right| \leq \varepsilon\right) \approx 2\Phi\left(\varepsilon \cdot \sqrt{\frac{n}{pq}}\right). \qquad (3.7)$$

За допомогою формули (3.7) можна знайти:

1. *Число випробувань* необхідне для того, щоб зі заданою ймовірністю γ відхилення відносної частоти настання події A від ймовірності p настання її у кожному випробуванні за абсолютною величиною не перевищувало це додатне число ε.

$$n = \frac{t_\gamma^2}{\varepsilon^2} pq. \qquad (3.8)$$

$2\Phi(t_\gamma) = \gamma$, t_γ знаходимо за таблицею 2 у Додатках.

2. *Межі*, в яких із заданою ймовірністю γ знаходиться частота появи події A, якщо відомо число випробувань n і ймовірність p настання події в кожному випробуванні.

$$np - t_\gamma\sqrt{npq} \leq k \leq np + t_\gamma\sqrt{npq}. \qquad (3.9)$$

Приклад 3.6. Оцінити мінімальне число договорів, яке потрібно укласти, щоб з ймовірністю $\gamma=0,95$ можна було стверджувати, що частка страхових випадків відхиляється від ймовірності $p=0,1$ (за абсолютною величиною) не більше ніж на $\varepsilon=0,02$.

Розв'язання. За умовою задачі $\gamma=0,95,\ p=0,1,\ q=0,9,\ \varepsilon=0,02$

$$2\Phi(t_\gamma) = 0,95 \Rightarrow \Phi(t_\gamma) = 0,475.$$

Скориставшись таблицею 2 з Додатків отримаємо, що $t_\gamma = 1,96$.

Підставивши дані у формулу (3.8), отримаємо:

$$n = \frac{(1,96)^2}{(0,02)^2} \cdot 0,1 \cdot 0,9 = 864,36.$$

Отже, потрібно укласти не менше 865 договорів страхування.

Приклад 3.7. Укладено 120 договорів страхування з ймовірністю $p=0,4$ настання події A за одним договором страхування. Знайти межі, в яких міститься частота появи події A з ймовірністю $\gamma=0,975$.

Розв'язання. За умовою задачі $n = 120$, $p=0,4$, $q=0,6$.

$$2\Phi(t_\gamma) = 0{,}975 \Rightarrow \Phi(t_\gamma) = 0{,}4875.$$

Скориставшись таблицею 2 з додатків, отримаємо $t_\gamma = 2{,}66$.

Підставивши дані у нерівність (3.9) отримаємо:

$$120 \cdot 0{,}4 - 2{,}66\sqrt{120 \cdot 0{,}4 \cdot 0{,}6} \le k \le 120 \cdot 0{,}4 + 2{,}66\sqrt{120 \cdot 0{,}4 \cdot 0{,}6}$$
$$33{,}7 \le k \le 62{,}3.$$

Отже, потрібно укласти від 34 до 62 договорів страхування.

3.5. Теорема Пуассона

За теоремою Пуассона вираховують імовірності $P_n(k)$ в тих випадках, коли ймовірність настання події стала і дуже мала, а число випробувань доволі велике.

Теорема Пуасонна. Якщо ймовірність p настання події A мала, то для великого числа випробувань ймовірність того, що подія відбудеться k раз, асимптотично виражається формулою:

$$P_n(k) = \frac{a^k e^{-a}}{k!}, \tag{3.10}$$

де $a = np$.

Приклад 3.8. Телевізійний завод відвантажив на базу 5000 доброякісних кольорових телевізорів. Ймовірність того, що під час перевезення продукції якість втратиться, дорівнює 0,0002. Знайти ймовірність того, що на базу потрапить три бракованих телевізори.

Розв'язання. За умовою задачі n = 500; p = 0,0002; k = 3.

Знайдемо $a = np = 5000 \cdot 0{,}0002 = 1$.

Шукана ймовірність за теоремою Пуассона (3.10) приблизно дорівнює:

$$P_{5000}(3) = \frac{a^k e^{-a}}{k!} = \frac{e^{-1}}{3!} \approx 0{,}06.$$

Потоком подій називають послідовність подій, які настають у випадковий момент. Прикладом потоку подій може бути надходження викликів на пункт швидкої допомоги, прибуття покупців у магазин та ін.

Серед властивостей, якими володіють потоки, виокремимо властивості стаціонарності, відсутності післядії та ординарності.

Властивість стаціонарності характеризується тим, що ймовірність появи k подій за довільний проміжок залежить лише від числа k і від довжини t проміжку і не залежить від початку його відліку; водночас вважаємо, що різні проміжки не перетинаються.

Отже, якщо потік володіє властивістю стаціонарності, то ймовірність появи k подій за проміжок t є функцією, яка залежить лише від k і t.

Властивість «відсутності післядії» характеризується тим, що ймовірність появи k подій за довільний проміжок не залежить від того, відбувались чи не відбувались події у моменти часу, які передували початку проміжку, що розглядається.

Отже, якщо потік володіє властивістю післядії, то є взаємна незалежність появи тієї чи іншої кількості подій в неперетинних проміжках.

Властивість ординарності характеризується тим, що поява двох чи більше подій за малий проміжок часу практично неможлива, тобто ймовірність появи більше, ніж одної, події за невеликий проміжок настільки мала порівняно з імовірністю появи лише однієї події, що нею можна знехтувати.

Відтак, якщо потік володіє властивістю ординарності, то за нескінченно малий проміжок може відбутись не більше однієї події.

Будемо розглядати такі потоки, які мають властивість стаціонарності.

Інтенсивністю потоку λ називають середнє число подій, які з'являються за одиницю часу.

Якщо постійна інтенсивність потоку відома, то ймовірність появи k подій простого потоку за певний час t визначається формулою Пуассона:

$$P_t(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}. \tag{3.11}$$

Приклад 3.8. Середнє число викликів, що надходять на АТС за одну хвилину дорівнює двом. Знайти ймовірність того, що за 5 хв надійде на АТС два виклики.

Розв'язання. За умовою задачі λ =2; *t*=5; *k*=2.

Використовуючи формулу Пуассона **(**3.11**)**, одержимо:

$$P_5(2) = \frac{(2\cdot 5)^2 \cdot e^{-2\cdot 5}}{2!} = \frac{100\cdot 0{,}000045}{2} \approx 0{,}000025.$$

Ця подія практично неможлива, бо ймовірність настання події дуже мала.

Запитання для самоконтролю

1. Напишіть формулу Бернуллі.
2. У яких випадках використовують формулу Бернуллі?
3. Як знаходять найімовірніше число настання події при послідовності незалежних випробувань?
4. Сформулюйте локальну теорему Муавра–Лапласа.
5. У якому випадку її застосовують теорему Лапласа?
6. Напишіть формулу Пуассона.
7. У яких випадках використовують формулу Пуасонна?
8. Що називають потоком подій?
9. Чим характеризується властивість стаціонарності?
10. Чим характеризується властивість «відсутності післядії?
11. Чим характеризується властивість ординарності?
12. Що називається інтенсивністю потоку?
13. У чому полягає інтегральна теорема Лапласа?
14. Яка функція називається функцією Лапласа?

Задачі для самостійного розв'язування

1. Серед 5-ти студентів проводиться психологічний тест на визначення типу характеру людини. Ймовірність того, що за результатами тестування буде правильно визначено тип характеру кожної людини дорівнює 0,9. Знайти ймовірність того, що буде правильно визначено тип характеру лише трьох протестованих студентів.

2. Ймовірність того, що витрати електроенергії протягом доби не перевищили норму дорівнює 0,75. Знайти ймовірність того, що у найближчі 6 діб витрати електроенергії впродовж 4-ох днів не перевищать норму.

3. Автопарк налічує 12 автомашин. Ймовірність виходу на лінію кожної з них дорівнює p = 0,8. Знайти ймовірність нормальної роботи автопарку, якщо для цього на лінії треба мати не менше 8-ми автомашин?

4. У цеху є 10 верстатів. Ймовірність бути ввімкнутим для кожного верстата дорівнює 0,9. Яка ймовірність

а) одночасної роботи 5 верстатів?

б) не менше 4 верстатів?

5. Ймовірність виграшу по білету грошово-речової лотереї дорівнює 0,1. Яка ймовірність, маючи 8 білетів,

а) виграти на 3 білети;

б) виграти принаймні не менше ніж на 4 білетах?

6. Ймовірність того, що взята навмання деталь є нестандартною, становить 0,2. Знайти ймовірність того, що серед 6 навмання взятих деталей:

а) 4 деталі нестандартні;

б) не менше як 5 нестандартні деталі.

7. Стрілець зробив 4 постріли в ціль. Ймовірність влучання при кожному пострілі є сталою p = 0,6. Треба визначити ймовірність того, що буде хоча б одне влучання в ціль.

8. Банк видає кредитні картки VISA. Було встановлено, що 40% усіх рахунків оплачуються повністю за їх допомогою. З попереднього року вибрали навмання 6 рахунків. Яка ймовірність, що

а) 4 з них оплачені за допомогою карток VISA?

б) не більше чотирьох?

9. Служба податків визначила, що 50% всіх особистих декларацій про прибуток містить принаймні одну помилку. Якщо випадково відібрати десять декларацій, то яка ймовірність того, що рівно 6 із них будуть містити принаймні одну помилку?

10. Митниця дає офіційну оцінку того, що 20% усіх осіб, що повертаються з-за кордону, не декларує весь товар, на який накладається податок. Якщо випадково відібрати 6 осіб, які повертаються з-за кордону, то яка ймовірність того, що не менше трьох з них не задекларує товар, який обкладається податком?

11. Ймовірність одягу з дефектами, який шиє швейний цех, дорівнює 0,05. Скільки костюмів повинно бути у партії, щоб найімовірніше число костюмів з дефектом дорівнювало 20?

12. Знайти найімовірніше число зіпсованих деталей, якщо випробується 100 деталей, а ймовірність того, що деталь зіпсована, дорівнює 0,1.

13. Скільки потрібно провести незалежних випробувань, щоб найімовірніше число настання певної події дорівнювало 50, якщо ймовірність настання цієї події в окремому випробуванні $P = 0,9$?

14. Ймовірність виготовлення бракованої деталі становить 0,01. Знайти найімовірніше число бракованих деталей серед 500 деталей і ймовірність такої кількості їх у партії.

15. Ймовірність несплати податків для кожного з 500 підприємців регіону рівна 0,1. Знайти найімовірнішу кількість підприємців регіону, які не сплачують податків.

16. В одержаній партії текстильних виробів 0,6% браку. Яка ймовірність при випадковому відборі 1000 виробів виявити:

а) шість бракованих виробів;

б) хоча б один бракований виріб?

17. Скільки треба взяти деталей, щоб найімовірнішим числом стандартних деталей було число 50, якщо ймовірність того, що взята навмання деталь буде стандартною, дорівнює 0,81.

18. При якій кількості лотерейних білетів найімовірніше число виграшів дорівнюватиме 16, якщо ймовірність виграшу для кожного білета дорівнює 0,01.

19. В інкубатор закладено 750 яєць. Ймовірність того, що з яйця вилупиться курочка дорівнює 0,5. Яка ймовірність того, що вилупиться рівно 428 курок?

20. Ймовірність несплати податку для кожного з 400 підприємців дорівнює 0,1. Яка ймовірність того, що податки не сплатять не більше 37 підприємців.

21. Ймовірність своєчасної реалізації одиниці продукції дорівнює 0,8. Визначити ймовірність своєчасної реалізації не менше ніж 312 одиниць продукції з 400, що поступили на реалізацію.

22. Ймовірність того, що власник квартири не має заборгованості в оплаті за використання електроенергії дорівнює 0,6. Яка ймовірність того, що з 2400 власників квартир:

а) 1400 не мають названої заборгованості;

б) більше 1400 не мають названої заборгованості?

23. Ймовірність браку виробництва складає 15%. Яке буде найімовірніше значення браку для 500 виготовлених деталей? Яка ймовірність того, що бракованих деталей буде:

а) від 150 до 300?

б) рівно 220?

24. Визначити ймовірність того, що з 400 виготовлених на фабриці виробів 80 – вищого сорту. Ймовірність того, що кожний виріб вищого сорту дорівнює 0,2.

25. Ймовірність виявлення бракованого виробу дорівнює 0,005. Чому дорівнює ймовірність того, що з 10000 взятих навмання виробів бракованих буде 40?

26. Прядильниця обслуговує 1000 веретен. Ймовірність розриву нитки на одному веретені за хвилину дорівнює 0,004. Знайти ймовірність того, що протягом однієї хвилини нитка обірветься на п'яти веретенах.

27. Гральний кубик кидають 800 разів. Яка ймовірність того, що кількість очок, кратна трьом, з'явиться не менше 260 разів і не більше 274 разів?

28. Зі статистичних даних відомо, що ймовірність захворіти грипом під час епідемії для кожної особи дорівнює 0,1. Яка ймовірність того, що із 100 перевірених осіб хворими виявляться від 20 до 50 осіб?

29. Ймовірність випуску радіоламп з дефектом дорівнює 0,08. Визначити ймовірність того, що серед 1000 радіоламп відхилення від цього відсотка браку не перевищує 0,01.

30. Книгу надруковано тиражем 90000 примірників. Ймовірність неправильного брошурування книги дорівнює 0,0001. Знайти ймовірність того, що тираж має 5 бракованих книг.

31. Ймовірність появи події в кожному з незалежних випробуваннях – 0,5. Знайти число випробувань n, при якому з ймовірністю 0,7698 можна очікувати, що відносна частота появи події відхилиться від його ймовірності за абсолютною величиною не більш як на 0,02.

32. Ймовірність виробництва бракованого виробу дорівнює 0,008. Знайти ймовірність того, що серед 1000 виробів бракованих буде не більше ніж: а) 3; б) 4; в) 5; г) 10.

Розділ 4

ВИПАДКОВІ ВЕЛИЧИНИ І ФУНКЦІЇ РОЗПОДІЛУ

4.1. Дискретна випадкова величина

Випадковою величиною X називають таку величину, яка в результаті випробування прийме лише одне з можливих значень наперед невідоме і залежне від випадкових причин, які попередньо не можуть бути враховані.

Дискретною випадковою величиною називають таку випадкову величину, яка приймає ізольовані числові значення із усіх можливих.

Для будь-якої випадкової величини недостатньо відомостей про їх можливі значення. Потрібно знати ще ймовірність появи цих значень. Перелік всіх випадкових значень дискретної випадкової величини та їх відповідних ймовірностей називається **закон розподілу дискретних випадкових величин**. Причому повинна виконуватись рівність

$$p_1 + p_2 + \ ... + p_n = 1.$$

Закон розподілу дискретних випадкових величин може бути заданий такими способами.

1. Табличний. Цей спосіб полягає в тому, що дається таблиця, в якій записані окремі значення випадкової величини *X* та відповідні їм ймовірності.

X	x_1	x_2	. . .	x_n
P	p_1	p_2	. . . ..	p_n

Якщо випадкова величина X приймає скінчене число різних значень $x_1,\ x_2,\ \dots,\ x_n$, то випадкові величини утворюють повну групу попарно несумісних подій, а отже, сума їх ймовірностей повинна бути рівна 1.

2. Графічний. Під час цього способу функцію задають графіком зазвичай у прямокутній системі координат. На осі абсцис відкладають значення випадкової величини, а на осі ординат – ймовірності. Якщо на площині *XOY* сполучити послідовно точки (x_1, p_1), (x_2, p_2), ... , (x_n, p_n), то дістанемо ламану, яку називають **многокутником розподілу випадкової величини *X***. Прикладом цього способу задання закону розподілу є полігон розподілу ймовірностей.

3. Аналітичний закон розподілу полягає в тому, що дається математична формула, за якою можна обчислити ймовірності того, що випадкова величина прийме те чи інше значення.

Найважливішими є такі види дискретних розподілів: рівномірний, біноміальний, показниковий, геометричний.

Рівномірний розподіл – це розподіл випадкової величини, яка набуває n різних значень x_i з однаковими ймовірностями

$$p_i = \frac{1}{n}\text{, де } i=1, 2, 3, ..., n.$$

Наприклад, число очок, що випадають на верхній грані грального кубика, є дискретна випадкова величина, яка має такий рівномірний закон розподілу:

X	1	2	3	4	5	6
P	1/6	1/6	1/6	1/6	1/6	1/6

Приклад 4.1. Один з п’яти ключів відмикає замок. Скласти закон розподілу кількості випробувань при відмиканні замка, якщо випробуваний ключ у дальших випробуваннях не використовується.

Розв’язання. Ймовірність відмикання замка кожним ключем

$$p = \frac{1}{5}\cdot$$

Випадкова величина *X* (кількість випробуваних ключів під час відмикання замка) може набувати значень: 1, 2, 3, 4, 5, оскільки маємо п’ять ключів.

При *X*=1 $p_1 = \frac{1}{5}$ (ймовірність відкрити першим ключем з п’яти).

При *X*=2 $p_2 = \frac{1}{5}$ (ймовірність відкрити другим ключем з п’яти).

При *X*=3 $p_3 = \frac{1}{5}$ (ймовірність відкрити третім ключем з п’яти).

При X=4 $p_4 = \frac{1}{5}$ (ймовірність відкрити четвертим ключем з п'яти).

При X=5 $p_5 = \frac{1}{5}$ (ймовірність відкрити п'ятим ключем з п'яти).

Закон розподілу цієї випадкової величини буде мати вигляд:

X	1	2	3	4	5
P	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

Контроль $\frac{1}{5}+\frac{1}{5}+\frac{1}{5}+\frac{1}{5}+\frac{1}{5} = 1.$

Біноміальний розподіл (розподіл Бернуллі) – це розподіл випадкової величини, яка набуває значення *i = 1, 2, 3, ..., n*, а ймовірності знаходяться з допомогою формули $p_i = C_n^i p^i q^{n-i}$.

Біноміальний розподіл має випадкова величина, яка дорівнює кількості «успіхів» у серії з n незалежних випробувань, у кожному з яких «успіх» відбувається з ймовірністю p.

Приклад 4.2. Ймовірність складання іспиту на «відмінно» для кожного з шести студентів дорівнює 0,4. Скласти закон розподілу кількості п'ятірок, які студенти одержали на екзамені.

Розв'язання. Позначимо випадкову величину кількості п'ятірок, одержаних студентами на екзамені, через *X*. Вона може набувати значень: 0, 1, 2, 3, 4, 5, 6.

За умовою задачі p = 0,4; q = 0,6.

Знайдемо ймовірності, з якими *X* набуває відповідних значень.

При *X* = 0 (на іспиті жоден із 6 студентів не одержав п'ятірки)

$$p(0)= C_6^0(0{,}4)^0(0{,}6)^6 \approx 0{,}047.$$

При *X* = 1 (на іспиті один із 6 студентів одержав п'ятірку)

$$p(1)= C_6^1(0{,}4)^1(0{,}6)^5 \approx 0{,}187.$$

При *X* = 2 (на іспиті двоє із 6 студентів одержали п'ятірки)

$$p(2)= C_6^2(0{,}4)^2(0{,}6)^4 \approx 0{,}311.$$

При *X* = 3 (на іспиті троє із 6 студентів одержали п'ятірки)

$$p(3)= C_6^3(0{,}4)^3(0{,}6)^3 \approx 0{,}276.$$

При X = 4 (на іспиті четверо із 6 студентів одержали п'ятірки)

$$p(4) = C_6^4(0{,}4)^4(0{,}6)^2 \approx 0{,}138.$$

При X = 5 (на іспиті п'ятеро із 6 студентів одержали п'ятірки)

$$p(5) = C_6^5(0{,}4)^5(0{,}6)^1 \approx 0{,}037.$$

При X = 6 (на іспиті 6 студентів одержали п'ятірки)

$$p(6) = C_6^6(0{,}4)^6(0{,}6)^0 \approx 0{,}004.$$

Закон розподілу для цієї випадкової величини буде мати вигляд:

x_i	0	1	2	3	4	5	6
p_i	0,047	0,187	0,311	0,276	0,138	0,037	0,004

Многокутник розподілу для цієї випадкової величини подано на рисунку 4.1.

Показниковий розподіл (розподіл Пуассона) – це розподіл випадкової величини, яка набуває значень k = *1, 2, 3, ...,* n, … а ймовірності знаходяться з допомогою формули $P_n(k) = \frac{a^k e^{-a}}{k!}$, де $a = n \cdot p$.

З допомогою цього розподілу будуються математичні моделі роботи об'єктів обслуговування населення.

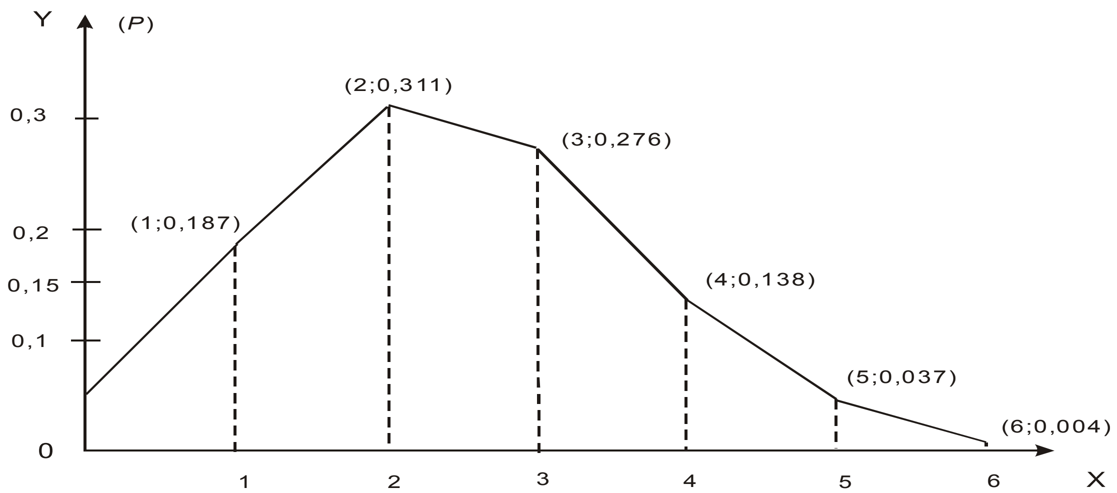

Рис. 4.1. Многокутник розподілу випадкової величини

Приклад 4.3. Магазин отримав 100 пляшок мінеральної води. Ймовірність того, що під час перевезення пляшка розіб'ється, дорівнює 0,01. Знайти ряд розподілу випадкової величини X, що характеризує кількість розбитих пляшок ($X \in [1, 5]$).

Розв'язування. Ймовірності появи окремих значень випадкової величини обчислюють за формулою:

$$P(k) = \frac{\mathrm{a}^k e^{-\mathrm{a}}}{k!},$$

де за умовою задачі $a = n \cdot p = 100 \cdot 0{,}01 = 1$.

При $X = 1$ (під час перевезення розбилась одна пляшка мінеральної води)

$$P(1) = \frac{1^1 e^{-1}}{1!} \approx 0{,}367884;$$

при $X = 2$ (під час перевезення розбилось дві пляшки мінеральної води)

$$P(2) = \frac{1^2 e^{-1}}{2!} \approx 0{,}18394;$$

при $X = 3$ (під час перевезення розбилось три пляшки мінеральної води)

$$P(3) = \frac{1^3 e^{-1}}{3!} \approx 0{,}06131;$$

при $X = 4$ (під час перевезення розбилось чотири пляшки мінеральної води)

$$P(4) = \frac{1^4 e^{-1}}{4!} \approx 0{,}01533;$$

при $X = 5$ (під час перевезення розбилось п'ять пляшок мінеральної води)

$$P(5) = \frac{1^5 e^{-1}}{5!} \approx 0{,}00306.$$

Закон розподілу для цієї випадкової величини буде мати вигляд (див. рис. 4.2):

x_i	1	2	3	4	5
p_i	0,36788	0,18394	0,06131	0,01533	0,00306

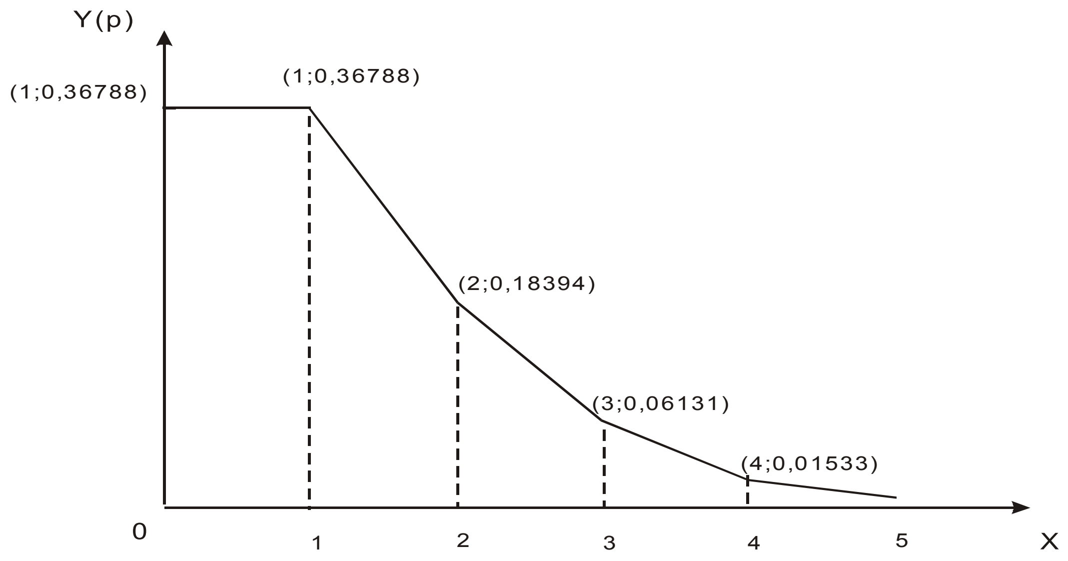

Рис. 4.2. Многокутник розподілу цієї випадкової величини

Геометричний розподіл – це розподіл випадкової величини, що набуває значень $k \in \boldsymbol{N}$ з ймовірностями $p_i = p \cdot q^{k-1}$, де *k=1, 2, 3, ..., n, …* число p – фіксоване, а q*=1 –* p.

Геометричний розподіл має випадкова величина, яка дорівнює кількості спроб до першого «успіху» в серії незалежних випробувань, у кожному з яких ймовірність успіху дорівнює p.

Приклад 4.4. Проводяться послідовні випробування п'яти приладів на надійність. Кожний наступний прилад випробують тільки тоді, коли попередній виявиться надійним. Скласти закон розподілу випадкової величини кількості випробувань приладів, якщо дорівнює 0,9.

Розв'язання. Число приладів, що проходить випробування, позначимо через *X.* Тоді *X* може дорівнювати 1, 2, 3, 4, 5. Якщо перший взятий для випробування прилад виявиться ненадійним, то випробування на цьому закінчаться, тобто буде тільки одне випробування. Ймовірність того, що перший прилад виявиться ненадійним, дорівнює $q = 1 - p = 1 - 0{,}9 = 0{,}1$. Тому ймовірність величини *X* буде дорівнювати одиниці: $p_1 = 0{,}1$.

Якщо перший прилад виявиться надійним, то аби було зроблено тільки два випробування, треба, щоб другий прилад був ненадійним. Ймовірність ненадійності другого приладу дорівнює 0,1, тому за теоремою множення ймовірностей (випадкова величина *X* має геометричний розподіл):

для $X = 2$, $p_2 = 0{,}9 \cdot 0{,}1 = 0{,}09$;
для $X = 3$, $p_3 = 0{,}9 \cdot 0{,}9 \cdot 0{,}1 = 0{,}081$;
для $X = 4$, $p_4 = 0{,}9 \cdot 0{,}9 \cdot 0{,}9 \cdot 0{,}1 = 0{,}079$;
для $X = 5$, $p_5 = 0{,}9 \cdot 0{,}9 \cdot 0{,}9 \cdot 0{,}9 = 0{,}6561$.

Закон розподілу для цієї випадкової величини буде мати вигляд:

x_i	1	2	3	4	5
p_i	0,1	0,09	0,081	0,079	0,6561

Використовуючи ці дані, будуємо многокутник розподілу випадкової величини (рис. 4.3).

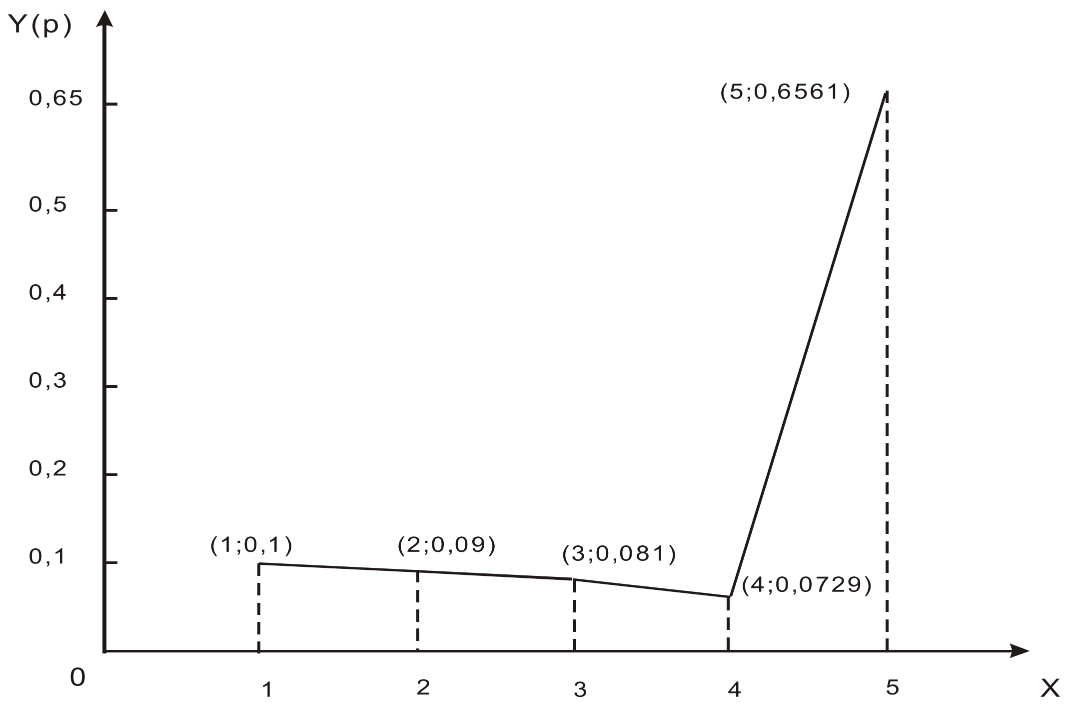

Рис. 4.3. Многокутник розподілу випадкової величини

Гіпергеометричний розподіл – це розподіл випадкової величини, яка набуває значень $k = 1,\ 2,\ 3,\ \ldots,\ n$, а ймовірності знаходяться з допомогою формули

$$P_k = \frac{C_n^k C_{N-n}^{m-k}}{C_N^k},$$

де $n \geq m,\ N \geq n$.

Цей розподіл має широке використання. Він виникає під час перевірки виробленої продукції, вивчення громадської думки в основі цього знаходиться експериментальна схема, яка полягає у витяганні (опитуванні) об’єктів без повернення у вихідну сукупність.

Приклад 4.5. Із картотеки служби зайнятості навмання вибирають 8 карток, серед яких три відповідають особам працевлаштованим, а решта безробітним. Скласти закон розподілу випадкової величини *X*-число карток серед відібраних, які відповідають працевлаштованим особам.

Розв'язання. Випадкова величина може набувати значення 0, 1, 2, 3. Оскільки ми маємо трьох працевлаштованих осіб. Потрібно знайти ймовірності цих значень. Для цього використаємо класичне означення ймовірності.

Ймовірність того, що маємо 4 безробітних і 0 працевлаштованих:

$$P_0 = \frac{C_5^4 C_3^0}{C_8^4} = \frac{\frac{5!}{4!\cdot 1!}\cdot\frac{3!}{0!3!}\cdot}{\frac{8!}{4!\cdot 4!}} = \frac{1}{14}.$$

Ймовірність того, що маємо 3 безробітних і 1 працевлаштованого:

$$P_1 = \frac{C_5^3 C_3^1}{C_8^4} = \frac{6}{14}.$$

Ймовірність того, що маємо 2 безробітних і 2 працевлаштованих:

$$P_2 = \frac{C_5^2 C_3^2}{C_8^4} = \frac{6}{14}.$$

Ймовірність того, що маємо 1 безробітного і 3 працевлаштованих:

$$P_3 = \frac{C_5^1 C_3^3}{C_8^4} = \frac{1}{14}.$$

Випадкова величина *X*-число карток серед відібраних, які відповідають працевлаштованим особам, буде мати такий закон розподілу:

X	0	1	2	3
P	$\frac{1}{14}$	$\frac{6}{14}$	$\frac{6}{14}$	$\frac{1}{14}$

Контроль $\frac{1}{14} + \frac{6}{14} + \frac{6}{14} + \frac{1}{14} = 1.$

4.2. Неперервна випадкова величина

Неперервною випадковою величиною називають таку величину, яка набуває всі свої можливі значення з деякого проміжку.

На відміну від дискретної випадкової величини розподіл неперервної випадкової величини неможливо задати, зазначивши певні значення, яких вона набуває, та відповідні їм ймовірності. Множина можливих значень неперервної випадкової величини незлічення. Тому виникає питання про новий спосіб задання випадкової величини. Він полягає у заданні функції розподілу.

Законом розподілу випадкової величини називають таке співвідношення, яке встановлює зв'язок між можливими значеннями випадкової величини і відповідними їм ймовірностями.

4.2.1. Інтегральна функція розподілу та її основні властивості

Інтегральною функцією розподілу називають функцію *F(x),* яка визначає для кожного значення *x* ймовірність того, що випадкова величина *X* набуває значення, менше дійсного числа *x*, тобто:

$$F(x) = P(X < x). \tag{4.1}$$

Геометрично ця рівність пояснюється так: *F(x)* є ймовірність того, що випадкова величина набуде значення, яке зображується на числовій вісі точкою, що знаходиться лівіше значення *x*.

Випадкову величину будемо називати неперервною, якщо її інтегральна функція розподілу *F(x)* є неперервно диференційована.

Властивості інтегральної функції розподілу:

1. Інтегральна функція розподілу *F(x)* задовольняє таку нерівність:

$$0 \le F(x) \ \le 1.$$

Ця властивість випливає з означення інтегральної функції як імовірності події.

2. Інтегральна функція $F(x)$ не є спадною, тобто, $F(x_2) \geq F(x_1)$ якщо $x_2 > x_1$.

3. Ймовірність того, що випадкова величина X набуде значення, що належить інтервалу (a, b), дорівнює приросту інтегральної функції, тобто

$$P(a < X < b) = F(b) - F(a). \tag{4.2}$$

Приклад 4.5. Дискретна випадкова величина X задана табличним розподілом:

X	1	4	8
P	0,3	0,1	0,6

Знайти інтегральну функцію і побудувати її графік.

Розв'язання.

Випадкова величина не може бути меншою за одиницю. Отже, для всіх $x \leq 1$ події неможливі, то $F(x) = 0$.

Якщо $1 < x \leq 4$, то $F(x) = 0{,}3$.

Справді, нехай, наприклад, $x = 1{,}4$. Тоді $F(1{,}4)$ є ймовірністю події $X < 1{,}4$. Але випадкова величина X тільки в одному випадку набуває значень, меншим від 1,4, а саме, з ймовірністю 0,3.

Якщо $4 < x \leq 8$, то $F(x) = 0{,}3 + 0{,}1 = 0{,}4$.

Бо якщо $x = 5$. Тоді $F(5)$ виражає ймовірність події $X < 5$. Значення величини будуть менші за 5, тобто 4 або 3. Застосовуючи теорему додавання ймовірностей, одержимо $F(x)$=0,91.

Якщо $x > 8$, то $F(x) = 0{,}4 + 0{,}6 = 1$.

Отже, інтегральна функція може бути записана так:

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

Графік *F(x)* матиме вигляд (рис. 4.4):

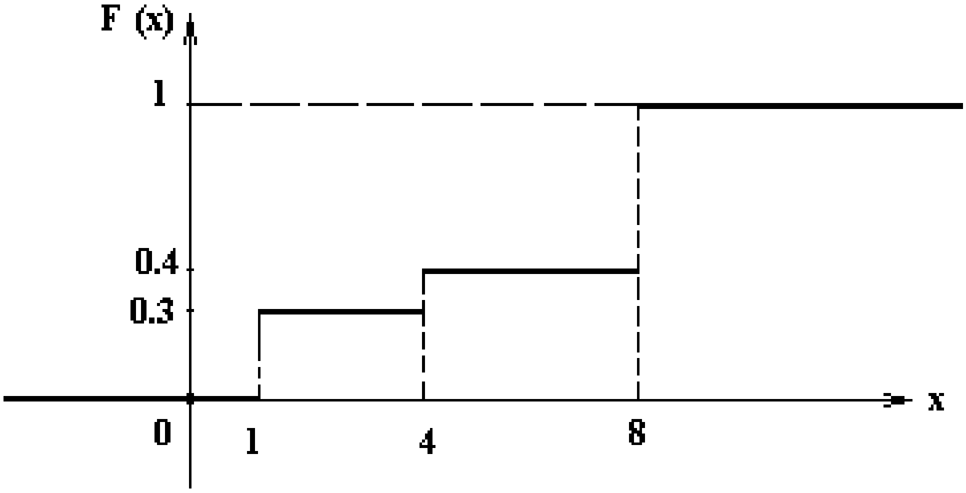

Рис. 4.4. Інтегральна функція

Приклад 4.6. Випадкова величина *X* задана інтегральною функцією розподілу

$$F(x) = \begin{cases} 0 & \text{при } x \leq 0; \\ x - 4 & \text{при } 0 < x \leq 1; \\ 1 & \text{при } x > 1. \end{cases}$$

Знайти ймовірність того, що в результаті випробування випадкова величина набуде значення, що належить інтервалу $(-\frac{1}{2}, \frac{1}{2})$.

Розв'язування. За формулою (4.2) маємо:

$$P\left(-\frac{1}{2} < X < \frac{1}{2}\right) = F\left(\frac{1}{2}\right) - F\left(-\frac{1}{2}\right) = \left(\frac{1}{2} - 4\right) - 0 = -\frac{7}{2} = -3{,}5.$$

4.2.2. Диференціальна функція розподілу та її основні властивості

Диференціальною функцією розподілу *f(x)* або **щільністю ймовірностей** неперервної випадкової величини *X* називають першу похідну від інтегральної функції:

$$f(x) = F'(x). \tag{4.3}$$

З цього означення випливає, що інтегральна функція розподілу неперервної випадкової величини *X* є первісною для диференціальної функції.

Знаючи інтегральну функцію розподілу *F(x)*, легко знайти диференціальну функцію *f(x)*, бо за означенням $f(x) = F'(x)$.

Знаючи диференціальну функцію *f(x)*, інтегральну функцію *F(x)* можна знайти за формулою:

$$F(x) = \int_{-\infty}^{\mathrm{x}} f(x)\,\mathrm{dx}. \qquad (4.4)$$

Приклад 4.7. Заданий рівномірний закон розподілу неперервної випадкової величини *X*:

$$f(x) = \begin{cases} 0 & \text{при } x \le a; \\ \dfrac{1}{b-a} & \text{при } a < x \le b; \\ 1 & \text{при } x > b. \end{cases}$$

Знайти інтегральну функцію рівномірного закону розподілу.

Розв'язання. Скористаємось формулою (4.4)

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

Якщо $x \le a$, то $f(x) = 0$ і, отже $F(x) = 0$; якщо $a < x \le b$, то $f(x) = \frac{1}{b-a}$, отже,

$$F(x) = \int_{-\infty}^{\mathrm{x}} f(x)\,\mathrm{dx} = \int_{-\infty}^{\mathrm{a}} 0 \cdot \mathrm{dx} + \int_{\mathrm{a}}^{\mathrm{x}} \frac{1}{b-a} dx = \frac{x-a}{b-a}.$$

Якщо $x > b$, то

$$F(x) = \int_{-\infty}^{\mathrm{a}} 0 \cdot \mathrm{dx} + \int_{\mathrm{a}}^{\mathrm{b}} \frac{1}{x-a} dx + \int_{b}^{x} 0 dx = \frac{b-a}{b-a} = 1\ .$$

Отже, шукана інтегральна функція рівномірного закону розподілу матиме вигляд:

$$F(x) = \begin{cases} 0 & \text{при } x \le a; \\ \dfrac{x-a}{b-a} & \text{при } a < x \le b; \\ 1 & \text{при } x > b. \end{cases}$$

Властивості диференціальної функції розподілу

1. Диференціальна функція додатна, тобто $f(x) \geq 0$. Ця властивість випливає з того, що інтегральна функція не є спадною і $f(x) = F'(x) \geq 0$.

2. Невластивий інтеграл від диференціальної функції дорівнює одиниці, тобто:

$$\int_{+\infty}^{-\infty} f(x)dx = 1.$$

Ця властивість випливає і з того, що ймовірність випадкової величини попадає в інтервал $(+\infty, -\infty)$, є подія достовірна.

Зокрема, якщо всі можливі значення випадкової величини належать інтервалу *(a, b)*, то $\int_a^b f(x)dx = 1$.

Знаючи диференціальну функцію, можна обчислити ймовірність того, що неперервна випадкова величина набуває значення, що належить інтервалу *(a, b)*.

3. Ймовірність того, що неперервна випадкова величина набуде значення, що належить інтервалу *(a, b)*, дорівнює визначеному інтегралу від диференціальної функції, в межах від a до b:

$$P(a < X < b) = \int_a^b f(x)dx. \quad (4.5)$$

Геометрично формула (4.5) визначає площу криволінійної трапеції, обмеженою віссю абсцис, кривою-диференціальною функцією розподілу *f(x)* і прямими $x = a,\ x = b$. Необхідність в формулі (4.5) пов'язана з тим, що розподіл неперервних випадкових величин, як правило, задається не інтегральною, а диференціальною функцією розподілу.

Приклад 4.8. Неперервна випадкова величина *X* задана інтегральною функцією розподілу:

$$F(x) = \begin{cases} 0 & \text{при } x \leq 0; \\ x^2 & \text{при } 0 < x \leq 1; \\ 1 & \text{при } x > 1. \end{cases}$$

Знайти:

А. Диференціальну функцію розподілу;

Б. Ймовірність того, що випадкова величина *X* набуде значення з інтервалу $(0, \frac{1}{2})$.

В. Побудувати графіки інтегральної та диференціальної функції розподілу.

Розв'язання.

А. За означенням диференціальної функції, маємо:

$$f(x) = F'(x) = \begin{cases} 0 & \text{при } x \leq 0; \\ 2x & \text{при } 0 < x \leq 1; \\ 1 & \text{при } x > 1. \end{cases}$$

Б. Ймовірність того, що випадкова величина *X* набуде значення з інтервалу *(a, b)*, визначається формулою (4.5):

$$P(a < X < b) = \int_a^b f(x)dx.$$

Підставивши значення, одержимо:

$$P\left(0 < X < \frac{1}{2}\right) = \int_0^{\frac{1}{2}} 2xdx = x^2\Big|_0^{1/2} = \frac{1}{4}.$$

В. Схематичні графіки функцій $F(x)$ і $f(x)$ мають вигляд:

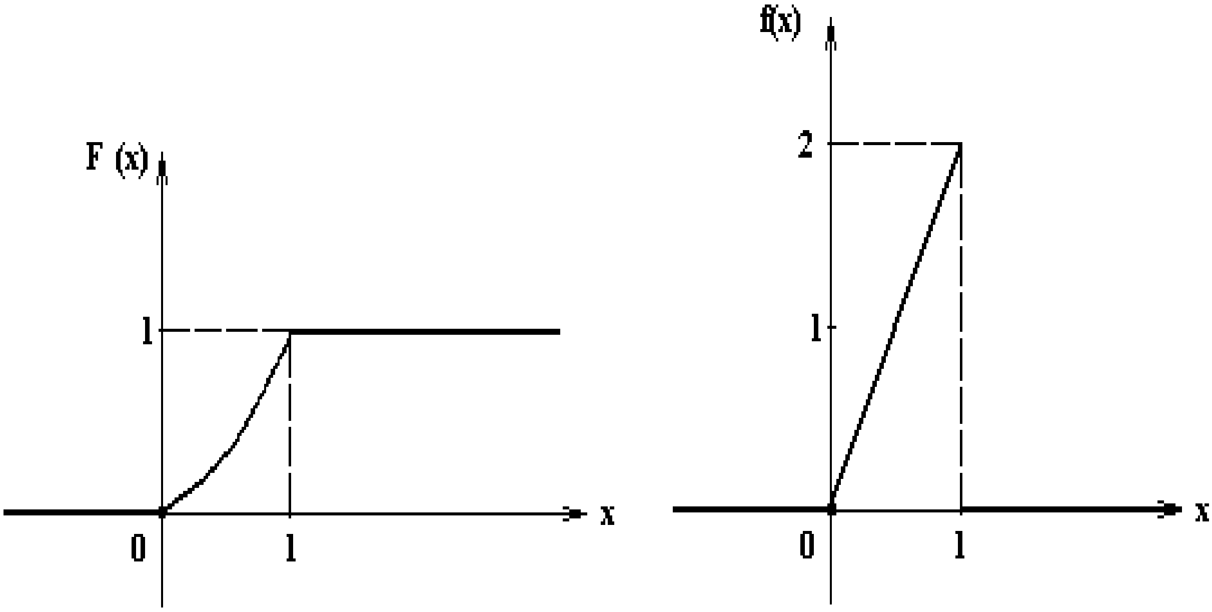

Рис. 4.5. Графіки інтегральної та диференціальної функції розподілу

У багатьох випадках диференціальну функцію розподілу неперервної випадкової величини часто називають **густиною ймовірності в точці *x***.

4.2.3. Нормальний закон розподілу

Нормальний закон розподілу (який, часто називають законом Гауса) виконує важливу роль в теорії ймовірностей і має серед інших законів розподілу особливе значення. Цей закон найчастіше трапляється на практиці. Важливою особливістю цього закону є те, що він є граничним законом, до якого наближаються інші закони розподілу, які почасти трапляються в типових умовах.

Нормальним законом розподілу випадкової величини X називають закон, диференціальна функція якого задається формулою:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi\cdot}} e^{\frac{-(x-a)^2}{2\cdot\sigma^2}}. \tag{4.6}$$

Нормальний закон розподілу визначається двома параметрами: a іσ. Достатньо задати ці параметри, щоб задати нормальний закон розподілу.

Графік диференціальної функції нормального закону розподілу часто називають **нормальною кривою або кривою Гауса**.

Дослідивши диференціальну функції $f(x) = \frac{1}{\sigma\sqrt{2\pi\cdot}} e^{\frac{-(x-a)^2}{2\cdot\sigma^2}}$ методом диференціального числення, отримаємо:

1. Функція $f(x)$ визначена для всіх значень x.
2. При всіх значеннях x функція $f(x)$ набуває додатні значення, тобто графік нормальної кривої розміщений над віссю OX.
3. Границя функції за необмеженого зростання x (за абсолютною величиною) дорівнює нулю, тобто:

$$\lim_{x\to\pm\infty} \frac{1}{\sigma\sqrt{2\pi\cdot}} e^{\frac{-(x-a)^2}{2\cdot\sigma^2}} = 0.$$

Це означає, що вісь OX є горизонтальною асимптотою графіка функції.

4. Дослідивши функцію на екстремум, знайдемо функцію, що має максимум, рівний $f(a) = \frac{1}{\sigma\sqrt{2\pi\cdot}}$.

5. Функція $f(x)$ має дві точки перегину:

$$\left(a-\sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right) i \ \left(a+\sigma, \frac{1}{\sigma\sqrt{2\pi e}}\right).$$

Схематичний графік нормальної кривої $f(x) = \frac{1}{\sigma\sqrt{2\pi\cdot}} e^{\frac{-(x-a)^2}{2\cdot\sigma^2}}$ наведений на рисунку 4.6.

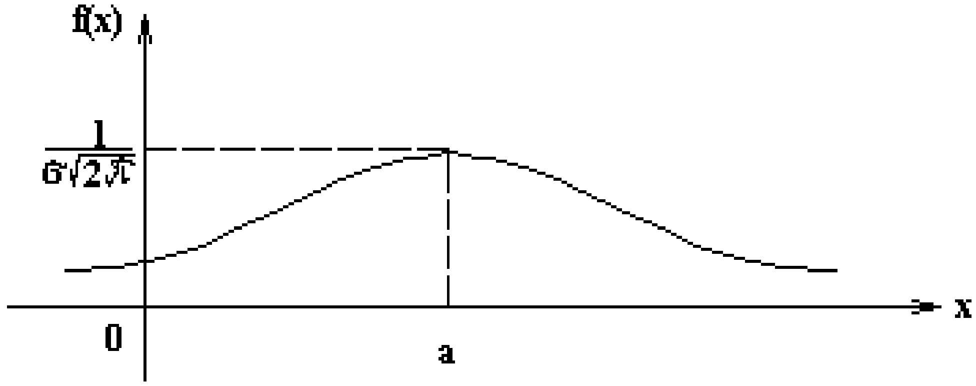

Рис. 4.6. Графік нормальної кривої

При а = 0 і $\sigma = 1$ нормальну криву називають **нормованою** і позначають:

$$\varphi(x) = \frac{1}{\sqrt{2\pi\cdot}} e^{\frac{-(x)^2}{2}}.$$

Ймовірність попадання в заданий інтервал (α, β) випадкової величини, розподіленої за нормальним законом, визначається за формулою:

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right). \qquad (4.7)$$

Приклад 4.9. Випадкова величина *X* розподілена за нормальним законом. Дано $a = 30$ і $\sigma = 10$. Знайти ймовірність того, що *X* набуде значення, що належить інтервалу (10, 50).

Розв'язання. За умовою задачі $a = 30$; $\sigma = 10$; $\alpha = 10$; $\beta = 50$. Використовуючи формулу (4.7), отримаємо:

$$P(10 < X < 50) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right) =$$
$$= \Phi\left(\frac{50 - 30}{10}\right) - \Phi\left(\frac{10 - 30}{10}\right) = 2\Phi(2).$$

За таблицею 2 Додатків знаходимо: $\Phi(2) = 0{,}4772$.
Звідси: $P(10 < X < 50) = 2 \cdot 0{,}4772 = 0{,}9544$.

Правило трьох сигм

Якщо випадкова величина *X* розподілена нормально, то $P(|X - a| \geq$ $\geq 3\sigma) \to 0$, тобто ймовірність того, що абсолютна величина відхилення *X* від її математичного сподівання більше 3σ прямує до 0, то це означає, що $\left|X - a\right| < 3\sigma$ – практично достовірна подія.

У практиці це правило використовують так:

Якщо закон розподілу випадкової величини X невідомий, але $\left|X - a\right| < 3\sigma$, тоді можна припустити, що *X* розподілена нормально.

Нормальний закон розподілу випадкової величини широко застосовується на практиці. Це пов'язано з тим, що коли на випадкову величину впливає сума багатьох факторів, кожний із яких впливає як завгодно, мало, то випадкова величина має розподіл близький до нормального. На практиці розглядають ще інші закони розподілу, такі, наприклад, як розподіл χ^2, розподіл Стьюдента, розподіл *F* – Фішера–Снеддона, гама-розподіл та інші.

Розглянемо основні з них.

Розподіл χ^2 («хі-квадрат»)

Нехай *Xi* (*i=1, 2, n)* – нормальні, нормовані незалежні величини. Тоді сума квадратів цих величин

$$\chi^2 = \sum_{i=1}^{n} X_i^2$$

розподілена по закону χ^2 з $k = n$ степенями вільності.

Якщо величини *Xi* зв'язані одним лінійним співвідношенням, наприклад, $\sum_{i=1}^{n} X_i = \bar{X}$, то число степеней вільності буде $k = n + 1$. Диференціальна функція χ^2 розподілу має вигляд

$$f(x) = \begin{cases} 0 & \text{при } x < 0, \\ \dfrac{1}{2^{\frac{k}{2}} \cdot \Gamma(\frac{k}{2})} \cdot e^{-\frac{x}{2}-1} \cdot e^{\frac{x}{2}-1} & \text{при } x \geq 0, \end{cases}$$

де $\Gamma(x) = \int_0^{\infty} t^{x-1} \cdot e^{-t}$ гама-функція, $\Gamma(n+1) = n!$

Розподіл χ^2 визначається параметром – числом степеней вільності *k*. Коли *k* зростає, розподіл χ^2 прямує до нормального розподілу дуже повільно.

Розподіл Стьюдента

Нехай *X* – нормальна нормована випадкова величина, а *Y* – незалежна від *X* величина, яка розподілена за законом хі-квадрат з *k* степенями вільності.

Тоді величина

$$T = \frac{X}{\sqrt{\frac{y}{k}}}$$

має розподіл, який називають **t-розподілом** або **розподілом Стьюдента** з k степенями вільності.

При зростанні k розподіл Стьюдента швидко наближається до нормального розподілу.

4.2.4. Показниковий закон розподілу

Показниковим (експотенціальним) законом розподілу випадкової величини X називають розподіл йімовірностей, який описується такою диференціальною функцією:

$$f(x) = \begin{cases} 0 & \text{при } x < 0, \\ \lambda e^{-\lambda x} & \text{при } x \geq 0, \end{cases} \qquad (4.8)$$

де λ – стала величина $\lambda > 0$.

Прикладом неперервної випадкової величини, яка має показниковий закон розподілу, може бути випадковий час між двома послідовними подіями простого потоку (простий потік Пуассона).

Скориставшись формулу (4.4) можна знайти інтегральну функцію показникового закону розподілу:

$$F(x) = \int_{-\infty}^{\mathrm{x}} f(x)\,\mathrm{dx} = \int\limits_{-\infty}^{0} 0 \cdot dx + \lambda \int\limits_{0}^{x} e^{-\lambda x}\,\mathrm{dx} = 1 - \mathrm{e}^{-\lambda x}$$

тобто,

$$F(x) = 1 - \mathrm{e}^{-\lambda x}\,. \qquad (4.9)$$

Графіки диференціальної та інтегральної функцій показникового розподілу мають вигляд (див. рис. 4.6):

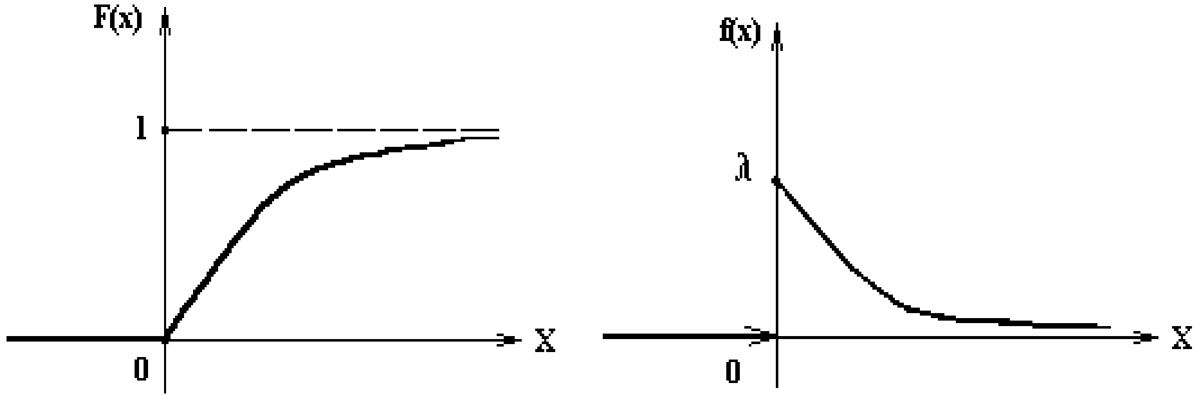

Рис. 4.6. Графіки диференціальної та інтегральної функцій показникового розподілу

Приклад 4.10. Диференціальна функція має вигляд:

$$f(x) = \begin{cases} 0 & \text{при } x < 0, \\ 2e^{-2x} & \text{при } x \geq 0. \end{cases}$$

Знайти ймовірність того, що в результаті випробування випадкова величина *X* набуде значення з інтервалу (0,3; 1).

Розв'язання. Ймовірність того, що в результаті випробування випадкова величина *X* набуде значення з інтервалу (0,3; 1) будемо шукати за формулою $P(a < X < b) = F(b) – F(a)$. Скориставшись формулою (4.9), за умовою $\lambda = 2$ маємо:

$$P(0{,}3 < X < 1) = 1 – \mathrm{e}^{-2\cdot 0{,}3} – (1 – \mathrm{e}^{-2\cdot 1}) = 0{,}54881 – 0{,}13534.$$

4.3. Багатовимірні випадкові величини

Крім одномірних випадкових величин, вивчаються величини, можливі значення яких визначаються двома, трьома, … , *n* числами. Такі величини називаються відповідно **двовимірними, трьохвимірними, … , n-вимірними**.

Будемо позначати через (*X, У*) двовимірну величину. Обидві величини *X* і *У* розглядаються одночасно, утворюючи систему двох випадкових величин. Двовимірну випадкову величину (*X, У*) геометрично можна інтерпретувати як випадкову точку $M(X, У)$ на площині (тобто як точку з випадковими координатами).

Наприклад, верстат-автомат штампує сталеві плитки. Якщо контролюють розміри – довжину X і ширину $У$, то маємо двовимірну випадкову величину.

Законом розподілу дискретної двовимірної величини (X і У) називають перелік всіх можливих значень цієї величини (тобто пар чисел) x_i, y_i і їх ймовірності $p(x_i, y_i)$ $i=1, 2, ..., n, j=1, 2, ..., m$. Закон розподілу подають у вигляді таблиці з подвійним входом.

Перший рядок таблиці містить всі можливі значення складової X, а перший стовпчик – всі можливі значення складової $У$. В клітинах, що на перетині стовпчика x_i і рядка y_i, вказані імовірності $p(x_i, y_i)$ того, що двовимірна випадкова величина набуде значення (x_i, y_i).

X У	x_1	x_2	.	.	.	x_n
y_1	$p(x_1, y_1)$	$p(x_2, y_1)$	.	.	.	$p(x_n, y_1)$
y_2	$p(x_1, y_2)$	$p(x_2, y_2)$	.	.	.	$p(x_n, y_2)$
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
y_m	$p(x_1, y_m)$	$p(x_2, y_m)$	.	.	.	$p(x_n, y_m)$

Оскільки події X, $У$, утворюють повну групу подій, то сума ймовірностей, розміщених в усіх клітках таблиці, дорівнює одиниці.

Знаючи закон розподілу двовимірної випадкової величини, можна знайти закон розподілу кожної зі складових.

Приклад 4.11. Знайти закон розподілу складових двовимірної випадкової величини, заданої законом розподілу:

X У	x_1	x_2	x_3
y_1	0,30	0,10	0,20
y_2	0,18	0,16	0,06

Розв'язання. Розклавши ймовірності за стовпчиками, одержимо ймовірності можливих значень

$$X\colon P(x_1) = 0{,}30 + 0{,}18 = 0{,}48;\ P(x_2) = 0{,}10 + 0{,}16 = 0{,}26;$$
$$P(x_3) = 0{,}20 + 0{,}06 = 0{,}26.$$

Закон розподілу складової випадкової величини *X* матиме вигляд:

X	x_2	x_2	x_2
P	0,48	0,26	0,26

Контроль 0,48 + 0,26 + 0,26 = 1.

Аналогічно, склавши ймовірності по рядках, отримаємо ймовірності можливих значень *У*: $P(y_1) = 0{,}60$; $P(y_2) = 0{,}40$.

Закон розподілу складової випадкової величини *X* матиме вигляд:

Y	y_2	y_2
P	0,60	0,40

Запитання для самоконтролю

1. Що називають випадковою величиною?
2. Що називають функцією розподілу ймовірності випадкової величини?
3. Яка випадкова величина називається дискретною?
4. Як можна задати дискретний розподіл випадкової величини?
5. Як побудувати многокутник розподілу випадкової величини?
6. Чому має дорівнювати сума всіх значень ймовірностей p_i дискретного розподілу випадкової величини?
7. Які є основні види дискретних розподілів ймовірностей?
8. Який закон розподілу називають рівномірним?
9. Який закон розподілу називають біноміальним?
10. Якою формулою визначається закон розподілу Пуассона? Який зміст параметра *a* в цій формулі?
11. За якою формулою обчислюється ймовірність p_i геометричного розподілу ймовірностей випадкової величини?
12. За якою формулою обчислюється ймовірність p_i гіпергеометричного розподілу ймовірностей випадкової величини?
13. Яка функція називається інтегральною функцією розподілу випадкової величини?
14. Назвіть властивості інтегральної функції розподілу випадкової величини?
15. Що називається щільністю розподілу ймовірностей або диференціальною функцією розподілу випадкової величини?

16. Назвіть властивості диференціальної функції розподілу.

17. Якою диференціальною функцією розподілу випадкової величини визначається нормальний закон розподілу?

18. Напишіть формулу для обчислення ймовірності того, що випадкова величина, яка підлягає нормальному розподілу, набуває якогось значення з інтервалу (α, β)?

19. Намалюйте схематичний графік нормальної кривої.

20. Графік диференціальної функції нормального закону розподілу називають?

21. У чому полягає зміст правила трьох сигм.

22. З чим пов'язане широке використання нормального закон розподілу?

23. Який закон розподілу називають розподілом χ^2?

24. Який закон розподілу називають розподілом Стьюдента?

25. Якою дифиренціальною функцією описується показниковий закон розподілу?

26. Запишіть інтегральну функцію показникового закону розподілу.

27. Намалюйте графіки диференціальної та інтегральної функцій показникового розподілу.

28. Які випадкові величини називають багатовимірними?

29. Що називається законом розподілу дискретної двовимірної величини?

30. Як подають закон розподілу дискретної двовимірної величини?

Задачі для самостійного розв'язування

1. У лотереї випущено 1000 білетів. Розігруються: 1 виграш по 1000 грн, 4 – 500 грн, 5 – 400 грн, 10 – 1000 грн. Знайти закон розподілу вартості виграшів для володаря 1 лотерейного білету.

2. Монету підкидують 4 рази. Необхідно знайти закон розподілу ймовірності випадкової величини числа випадання герба.

3. Студент знає відповідь на 20 питань з 25. Навмання вибирається 5 питань. Випадкова величина X – кількість питань, на які студент знає відповідь. Знайти ряд розподілу випадкової величини X.

4. Нехай X – випадкова величина, що дорівнює кількості хлопців у навмання вибраній сім'ї з трьома дітьми. Вважаючи народження хлопця й дівчини рівноймовірними подіями, побудувати ряд розподілу та многокутник розподілу ймовірності випадкової величини X, а також обчислити ймовірність того, що в сім'ї буде більше хлопчиків, ніж дівчаток.

5. Нехай випадкова величина X дорівнює кількості номерів, угаданих гравцем у лотереї «6 із 39». Побудувати ряд розподілу випадкової величини X. Знайти значення функції розподілу в точці $x = 3$.

6. Два стрільці роблять по одному незалежному пострілу в одну мішень. Ймовірність влучення для першого стрільця дорівнює 0,8, а для другого – 0,5. Знайти ряд розподілу і многокутник розподілу випадкової величини X – кількості влучень у мішень, а також ймовірність того, що кількість влучень дорівнюватиме кількості промахів.

7. Екзаменатор задає студентам додаткові питання про ймовірність того, що студенти знатимуть відповідь на будь-яке питання, дорівнює 0,9. Використовуючи припущення, що іспит закінчиться, якщо студент не знає запропоноване питання, знайти закон розподілу числа заданих додаткових запитань екзаменатором.

8. Комутатор деякого підприємства обслуговує 100 клієнтів, ймовірність дзвінка за 1 хвилину на комутатор дорівнює 0,01. Записати закон розподілу кількості дзвінків на комутатор протягом 1 хвилини.

9. Автоматична телефонна станція обслуговує 10000 телефонних номерів. Ймовірність того, що впродовж 1 хвилини до АТС надійде виклик абонента, дорівнює 0,0004. Знайти ряд розподілу випадкової величини X, яка дорівнює кількості викликів, які надійшли до АТС впродовж 1 хв, та ймовірність того, що за цей час надійде хоча б один виклик.

10. Ймовірність влучення у мішені, що рухаються, з пістолета дорівнює 0,001. Виконується 2000 пострілів. Випадкова величина X дорівнює кількості влучень. Побудувати ряд розподілу випадкової величини X.

11. Обчислити ймовірності появи деяких значень випадкової величини X, розподіленої за законом Пуассона, при p = 0,01; n = 100.

12. Підручник видано тиражем 10000 примірників. Ймовірність того, що зброшурований екземпляр виявиться неякісним, дорівнює 0,001. Знайти ряд розподілу випадкової величини X, що характеризує кількість неякісно зброшурованих екземплярів.

13. Магазин отримав 1000 пляшок мінеральної води. Ймовірність того, що під час перевезення пляшка розіб’ється, дорівнює 0,0004. Знайти ряд розподілу випадкової величини X, що характеризує кількість розбитих пляшок.

14. Один з п’яти ключів відмикає замок. Скласти закон розподілу кількості випробувань при відмиканні замка, якщо випробуваний ключ у дальших випробуваннях використовується.

15. На шляху руху автомобіля 6 світлофорів, кожен з яких дозволяє або забороняє рух автомобіля з ймовірністю 0,5. Скласти ряд розподілу і побудувати функцію розподілу кількості світлофорів, які автомобіль минув без зупинки.

16. Монету кидають 6 разів. Скласти ряд розподілу та побудувати функцію розподілу відношення частоти випадіння герба до числа випадіння надпису.

17. Ймовірність можливості будь-якого абонента подзвонити на комутатор протягом години дорівнює 0,01. Телефонна станція обслуговує 300 абонентів. Яка ймовірність того, що протягом години подзвонять 4 абоненти?

18. Нехай X – кількість появ числа 5 при двох киданнях грального кубика. Знайти закон розподілу величини X.

19. Виконується 4 незалежних постріла по мішені. Ймовірність влучення при одному пострілі дорівнює 0,25. Знайти закон розподілу кількості влучень X, найімовірніше число влучень.

20. Закон розподілу випадкової величини набуває вигляду. Знайти інтегральну функцію розподілу випадкової величини X.

1.

X	1	2	3	4
P	0,7	0,21	0,063	0,027

2.

X	4	5	6	8	10
P	0,21	0,17	0,18	0,23	0,21

3.

X	3	6	7	8	9
P	0,15	0,18	0,36	0,21	0,10

4.

X	6	8	9	10	12
P	0,17	0,14	0,27	0,23	0,19

5.

X					
P	,31	,08	,15	,21	,25

21. У задачах випадкову величину X задано інтегральною функцією розподілу $F(x)$. Знайти:

а) диференціальну функцію розподілу $f(x)$,

б) ймовірність того, що X набуває значення, що міститься в інтервалі $(\frac{1}{2},\ 1)$;

в) побудувати графіки функцій $F(x)$ і $f(x)$.

1) $F(x)=\begin{cases} 0 & \text{при } x\le 0; \\ \frac{1}{3}x & \text{при } 0<x\le 2; \\ 1 & \text{при } x>2; \end{cases}$

2) $F(x)=\begin{cases} 0 & \text{при } x\le -1; \\ \frac{x}{5} & \text{при} -1<x\le 2; \\ 1 & \text{при } x>2; \end{cases}$

3) $F(x)=\begin{cases} 0 & \text{при } x\le 0; \\ x^2-4x & \text{при } 0<x\le 1; \\ 1 & \text{при } x>1; \end{cases}$

4) $F(x)=\begin{cases} 0 & \text{при } x\le -2; \\ 3x^2-2x-1 & \text{при} -2<x\le 1; \\ 1 & \text{при } x>1; \end{cases}$

5) $F(x)=\begin{cases} 0 & \text{при } x\le 0; \\ \sin x & \text{при } 0<x\le \frac{\pi}{2}; \\ 1 & \text{при } x>\frac{\pi}{2}; \end{cases}$

6) $F(x)=\begin{cases} 0 & \text{при } x\le 0; \\ \frac{1}{2}(1-\cos x) & \text{при } 0<x\le \frac{\pi}{2}; \\ 1 & \text{при } x>\frac{\pi}{2}; \end{cases}$

7) $F(x)=\begin{cases} 0 & \text{при } x\le 0; \\ \operatorname{tg} x & \text{при } 0<x\le \frac{\pi}{4}; \\ 1 & \text{при } x>\frac{\pi}{4}. \end{cases}$

21. Випадкова величина X має нормальний розподіл ймовірностей з параметрами $a = 7$ *і* $\sigma = 2$. Порівняти ймовірності влучення значень випадкової величини X у проміжки [3; 7] і [–100; 1]. Зазначити інтервал, у який випадкова величина X влучає з практичною достовірністю.

22. Випадкова величина X має нормальний розподіл із параметрами а, σ. Записати вираз для щільності та функції розподілу випадкової величини X і знайти ймовірність того, що випадкова величина X набуде значення з проміжку (α; β), якщо:

а) a=0, σ=1, α=0, β=5;

б) a=5, σ=3, α=2, β=8;

в) a=9, σ=2, α=3, β=15.

23. Випадкова величина *X* має нормальний розподіл із параметрами a, σ. У кожному з наступних пунктів знайти інтервал, у який значення випадкової величини *X* влучають з практичною достовірністю (із ймовірністю 0,9973):

а) a = 0, σ = 1;

б) a = 5, σ = 2;

в) a = 5, σ = 100.

24. Випадкова величина *X* має показниковий розподіл. Ймовірність того, що ця випадкова величина набуде значення з проміжку [0; 5], дорівнює 0,7. Знайти ймовірність того, що ця випадкова величина набуде значення з проміжку [7; 9].

25. Неперервна випадкова величина *X* розподілена за законом, заданим диференціальною функцією

$$f(x) = \begin{cases} 0 & \text{при } x < 0; \\ 3e^{-3x} & \text{при } x \geq 0. \end{cases}$$

Знайти ймовірність того, що *X* потрапить в інтервал:

а) (0,13; 0,7);

б) (0,8; 1,2);

в) $(\frac{1}{2}, \frac{3}{2})$.

24. Незалежні випадкові величини *X* і *У* задані такими таблицями розподілу. Скласти таблицю розподілу добутку випадкових величин *X* і Y

X	0	2	4	6
P (X)	0,1	0,2	0,3	0,4

У	1	3	5
P(Y)	0,5	0,3	0,2

Розділ 5

ЧИСЛОВІ ХАРАКТЕРИСТИКИ ВИПАДКОВИХ ВЕЛИЧИН

До основних числових характеристик, які описують розподіл випадкової величини, належать математичне сподівання, дисперсія, середнє квадратичне відхилення, мода, медіана.

5.1. Числові характеристики дискретних випадкових величин

Математичним сподіванням дискретної випадкової величини називають суму добутків всіх можливих значень цієї величини на відповідні імовірності та позначають ***M(X)***.

Нехай випадкова величина X задана табличним законом розподілу,

X	x_1	x_2	. . .	x_n
P	p_1	p_2	. . . ..	p_n

то математичне сподівання визначається формулою:

$$M(X) = x_1 \cdot p_1 + x_2 \cdot p_2 + \ ... + x_n \cdot p_n$$

або

$$M(X) = \sum_{i=1}^{n} x_n \cdot p_n \,. \tag{5.1}$$

Приклад 5.1. У лотереї з кожних 100 білетів виграють 15. За розміром виграші розподілені так:

Розмір виграшу в грн.	200	100	20
Кількість виграшів	1	4	10

Скласти закон розподілу випадкової величини – розміру виграшу у лотереї та визначити її математичне сподівання.

Розв'язання. Щоб скласти закон розподілу випадкової величини *X* у вигляді таблиці, треба визначити всі можливі її значення. У нашому прикладі, крім записаних в таблиці значень 200, 50 та 10, є ще одне значення – 0 (програш). Відповідні імовірності дорівнюють $\frac{1}{100} = 0{,}01$; $\frac{4}{100} = 0{,}01$; $\frac{10}{100} = 0{,}1$ та $\frac{85}{100} = 0{,}850$ (кількість білетів, що відповідають варіантові 0, дорівнює 85).

Закон розподілу випадкової величини *X* має вигляд:

X	200	100	20	0
P	0,01	0,04	0,10	0,85

Математичне сподівання *X* знаходило за формулою (5.1):

$$M(X) = 200 \cdot 0{,}01 + 100 \cdot 0{,}04 + 20 \cdot 0{,}10 + 0 \cdot 0{,}85 = 8.$$

Математичне сподівання має ту ж розмірність, що й випадкова величина.

Ймовірнісний зміст математичного сподівання полягає в тому, що математичне сподівання наближено дорівнює середньому арифметичному значень випадкової величини.

Приклад 5.2. Знайти математичне сподівання числа появи події *A* в одному випробуванні, якщо ймовірність настання події *A* дорівнює p.

Розв'язання. Випадкова величина *X*-числа появи події *A* в одному випробуванні може набувати тільки двох значень:

$x_1 = 1$ – подія *A* відбулась, з ймовірністю p;

$x_2 = 0$ – подія *A* не відбулась, з ймовірністю q=1–*p*.

Математичне сподівання за формулою (5.1) буде дорівнювати

$$M(X) = 1 \cdot p + 0 \cdot q = p.$$

Отже, математичне сподівання числа появи події в одному випробуванні дорівнює ймовірності цієї події.

Властивості математичного сподівання

1. Математичне сподівання сталої величини дорівнює сталій величині:

$$M(C) = C.$$

2. Сталу величину можна виносити за знак математичного сподівання:

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

3. Математичне сподівання добутку кількох незалежних випадкових величин дорівнює добутку їх математичних сподівань:

$$M(X \cdot Y) = M(X) \cdot M(Y).$$

4. Математичне сподівання суми двох випадкових величин дорівнює сумі їх математичних сподівань:

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

5. Математичне сподівання числа появи події в n незалежних випробуваннях дорівнює добутку числа випробувань на ймовірність події у кожному випробуванні:

$$M(X) = n \cdot p.$$

Приклад 5.3. Знайти математичне сподівання суми числа очок, які можуть випасти при киданні двох гральних кубиків.

Розв'язання. Позначимо число очок, що можуть випасти на першому кубику, через X, а на другому – через Y. Вони набуватимуть однакових значень 1, 2, 3, 4, 5, 6 з імовірностями $p = \frac{1}{6}$.

Маємо:

$$M(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = \frac{7}{3},$$
$$M(Y) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{3}.$$

Математичне сподівання суми числа очок дорівнює:

$$M(X + Y) = M(X) + M(Y) = \frac{7}{2} + \frac{7}{2} = 7.$$

Приклад 5.4. Ймовірність влучання в мішень при одному пострілі дорівнює $p = 0{,}7$. Знайти математичне сподівання числа влучань при десяти пострілах.

Розв'язання. Влучання в мішень при кожному пострілі не залежить від результату інших пострілів, тому вказані події є незалежні і шукане математичне сподівання дорівнює:

$$M(X) = n \cdot p = 10 \cdot 0{,}7 = 7 \text{ (влучань)}.$$

Математичне сподівання не дає повної характеристики випадкової величини, бо для різних законів розподілу можуть бути однакові математичні сподівання.

Для обчислення відхилення випадкової величини від її математичного сподівання використовують іншу числову характеристику, дисперсію.

Дисперсією (розсіюванням) дискретної випадкової величини називають математичне сподівання квадрата відхилення випадкової величини від її математичного сподівання і позначають ***D(X)***:

$$D(X) = M(X - M(X))^2. \tag{5.2}$$

Для обчислення дисперсії застосовують формулу, яку отримують, розписавши формулу (5.2).

$$D(X) = M\big(X - M(X)\big)^2 = M\left(X^2 - 2 \cdot X \cdot M(X) + \big(M(X)\big)^2\right) =$$
$$= M(X^2) - 2 \cdot M(X) \cdot M(X) + \big(M(X)\big)^2 =$$
$$= M(X^2) - 2 \cdot (M(X))^2 + (M(X))^2 = M(X^2) - (M(X))^2$$

Отже,

$$D(X) = M(X^2) - \big(M(X)\big)^2. \tag{5.3}$$

Приклад 5.5. Знайти дисперсію випадкової величини *X,* яка має такий закон розподілу

X	2	3	5
P	0,1	0,6	0,3

Розв'язання. Знайдемо математичне сподівання *M(X):*

$$M(X) = 2 \cdot 0{,}1 + 3 \cdot 0{,}6 + \ 5 \cdot 0{,}3 = 3{,}5;$$

$$M(X^2) = 2^2 \cdot 0{,}1 + 3^2 \cdot 0{,}6 + \ 5^2 \cdot 0{,}3 = 13{,}3.$$

Дисперсію шукаємо за формулою (5.3):

$$D(X) = \ M(X^2) – \left(M(X)\right)^2 = 13.3 – 3{,}5^2 = 1{,}05.$$

Властивості дисперсії

1. Дисперсія сталої величини *C* дорівнює нулю:

$$D(\mathrm{C}) = \ 0.$$

2. Сталий множник можна виносити за знак дисперсії, піднісши його до квадрату:

$$D(\mathrm{C} \cdot X) = \ \mathrm{C}^2 \cdot D(X).$$

3. Дисперсія суми двох незалежних випадкових величин дорівнює сумі дисперсії цих величин:

$$D(X + Y) = D(X) + D(Y).$$

4. Дисперсія різниці двох незалежних випадкових величин дорівнює сумі дисперсій:

$$D\big(X – Y\big) = D(X) + D(Y).$$

5. Дисперсія числа появи події *A* в n незалежних випробуваннях, в кожному із яких ймовірність *p* появи події стала, дорівнює добутку числа випробувань на ймовірність появи і непояви події в одному випробуванні:

$$D(X) = n \cdot p \cdot q.$$

6. Дисперсія добутку двох незалежних випадкових величин *X, У* дорівнює:

$$D(X \cdot Y) = D(X) \cdot D(Y) + n^2 D(X) + m^2 D(Y),$$

де

$$m = M(X), n = M(Y).$$

Приклад 5.6. Кубик підкидується 10 разів. Знайти математичне сподівання та дисперсію появи 6 очок на випавшій грані.

Розв'язання:

З умови задачі маємо $n = 10, p = \frac{1}{6}, q = 1 - p = 1 - \frac{1}{6} = \frac{5}{6}$.

Для того, щоб знайти математичне сподівання, скористаємося п'ятою властивістю математичного сподівання:

$$M(X) = n \cdot p = 10 \cdot \frac{1}{6} = \frac{5}{6}.$$

Для того, щоб знайти дисперсію, використаємо п'яту властивістю дисперсії:

$$D(X) = n \cdot p \cdot q = 10 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{25}{18}.$$

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

$$\sigma(X) = \sqrt{D(X)}. \tag{5.4}$$

Середнє квадратичне відхилення має ту розмірність, що і випадкова величина *X*. Тому в тих випадках, коли оцінка розсіювання повинна мати ту ж розмірність, що і випадкова величина, замість дисперсії розглядають середнє квадратичне відхилення.

Властивості середнього квадратичного відхилення

1. Середнє квадратичне відхилення сталої величини *C* дорівнює нулю:

$$\sigma(\mathrm{C}) = 0.$$

2. Сталий множник можна виносити за знак середнього квадратичного відхилення, взявши по модулю:

$$\sigma(\mathrm{C} \cdot X) = |\mathrm{C}| \cdot \sigma(X).$$

3. Середнє квадратичне відхилення суми двох незалежних випадкових величин дорівнює:

$$\sigma(X + Y) = \sqrt{\sigma(X)^2 + \sigma(Y)^2}.$$

Модою $\boldsymbol{Mo(X)}$ дискретної випадкової величини називають найімовірніше її значення в деякому околі цього значення.

Розподіл називається **унімодальним**, якщо він має єдину моду (біноміальний, нормальний, показниковий тощо).

Розподіл називається **полімодальним**, якщо він має більше однієї моди (рівномірний тощо).

Модою абсолютно неперервної випадкової величини є точка максимуму щільності розподілу.

Медіаною випадкової величини називається таке число $\boldsymbol{Me(X)}$, для якого виконується умова:

$$P\big(X < Me(X)\big) = P\big(X > Me(X)\big).$$

Медіана неперерваної випадкової величини Х завжди існує, а якщо *X* має дискретний розподіл, то $Me(X)$) може не існувати.

Приклад 5.7. Випадкова величина *X* має ряд розподілу, наведений у таблиці. Знайти числові характеристики випадкової величини *X*.

X	1	2	3
P	0,2	0,6	0,2

Розв'язання.

Математичне сподівання випадкової величини *X* обчислимо за формулою (5.1):

$$M(X) = \sum_{i=1}^{n} x_n \cdot p_n = 1 \cdot 0{,}2 + 2 \cdot 0{,}6 + 3 \cdot 0{,}2 = 2.$$

Дисперсію випадкової величини Х знайдемо за формулою (5.3):

$$D(X) = \ M(X^2) – (M(X))^2 = x_1^2 \cdot p_1 + x_2^2 \cdot p_2 + x_3^2 \cdot p_3 – (M(X))^2;$$

$$D(X) == 1^2 \cdot 0{,}2 + 2^2 \cdot 0{,}6 + 3^2 \cdot 0{,}2 – 2^2 = 4{,}4 – 4 = 0{,}4.$$

Середнє квадратичне відхилення випадкової величини X вирахуємо за формулою (5.4):

$$\sigma(X) = \sqrt{D(X)} = \sqrt{0{,}4} = 0{,}633.$$

Мода і медіана дорівнюють

$$Mo(X) = Me(X) = 2.$$

Числові характеристики основних законів розподілу дискретних випадкових величин, мають такий вигляд.

1. Рівномірний розподіл з параметром n

$$M(X) = \frac{n+1}{2}, D(X) = \frac{n^2-1}{12}, \sigma(X) = \sqrt{\frac{n^2-1}{12}}.$$

2. Біномінальний розподіл з параметром n і p

$$M(X) = n \cdot p, D(X) = n \cdot p \cdot (1-p), \sigma(X) = \sqrt{n \cdot p \cdot (1-p)}.$$

3. Показниковий з параметром λ

$$M(X) = \lambda, D(X) = \lambda, \sigma(X) = \sqrt{\lambda}.$$

4. Геометричний з параметром p

$$M(X) = \frac{1}{p}, D(X) = \frac{1-p}{p^2}, \sigma(X) = \frac{\sqrt{1-p}}{p}.$$

Приклад 5.8. Ймовірність влучення у мішень, що рухається, з пістолета дорівнює 0,001. Відбувається 2000 пострілів. Знайти числові характеристики випадкової величини *X*, яка дорівнює кількості влучень у мішень.

Розв'язання. Випадкової величини X має показниковий розподіл з параметром λ, який дорівнює $\lambda = n \cdot p$.

За умовою задачі $n = 2000,\ p = 0{,}001,$

тому

$$\lambda = 2000 \cdot 0{,}001 = 2.$$

Скористаємось формулами для знаходження числових характеристик основних законів розподілу дискретних випадкових величин. При показниковому розподілі:

$$M(X) = \lambda, D(X) = \lambda, \sigma(X) = \sqrt{\lambda}.$$

Тому

$$M(X) = 2, D(X) = 2, \sigma(X) = \sqrt{2}.$$

5.2. Числові характеристики середнього арифметичного однаково розподілених взаємно незалежних випадкових величин

Нехай $X_1, X_2, \dots, X_n$ – взаємно незалежні випадкові величини, які мають однаковий розподіл і, отже, однакові характеристики (математичне сподівання, дисперсію).

Середнє арифметичне цих величин:

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Математичне сподівання середнього арифметичного однаково розподілених взаємно незалежних випадкових величин дорівнює математичному сподіванню кожної з цих величин:

$$M(\bar{X}) = a.$$

Дисперсія середнього арифметичного однаково розподілених взаємно незалежних випадкових величин в n-разів менша дисперсії D – кожної із цих величин:

$$D(\bar{X}) = \frac{D}{n}.$$

Середнє квадратичне відхилення середнього арифметичного n однаково розподілених взаємно незалежних випадкових величин в $\sqrt{n}$ раз менше середнього квадратичного відхилення σ кожної із величин:

$$\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

Наприклад, якщо середнє квадратичне відхилення окремого виміру дорівнює $\sigma = 4$ м, м, а всього зроблено n = 16 вимірів, то середнє квадратичне відхилення середнього арифметичного тих вимірів дорівнює лише 1 м:

$$\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{16}} = 1.$$

Середнє арифметичне вимірів виявилось більш близьким до істинного значення величини, яку вимірюють, ніж результат окремого виміру.

При $n \to \infty$ математичне сподівання середнього арифметичного n незалежних однаково розподілених випадкових величин не змінюється і дорівнює середньому випадкових величин, а дисперсія прямує до нуля. Це означає, що середнє арифметичне випадкових величин при $n \to \infty$ перестає бути випадковою величиною і прямує до сталої a.

Для встановлення певної ознаки деякої величини проводять n вимірювань $X_1, X_2, \ldots, X_n$, які залежать від багатьох випадкових факторів та мають однаків розподілі і є взаємно незалежними. За характеристику досліджувальної ознаки доцільніше обирати середнє арифметичне отриманих результатів вимірювання $X_1, X_2, \ldots, X_n$, оскільки розсіювання їх середнього арифметичного є менше ніж розсіювання кожного результату вимірювання і воно прямує до нуля при $n \to \infty$.

Значення середнього арифметичного результатів вимірювань ознаки випадкової величини є надійнішим і ближчим до справжньої характеристики цієї ознаки, ніж окремий результат.

5.3. Числові характеристики неперервних випадкових величин

Математичним сподіванням неперервної випадкової величини X, можливі значення якої належать відрізку $[a, b]$, називають визначений інтеграл:

$$M(X) = \int_a^b x \cdot f(x) dx. \tag{5.5}$$

Якщо можливі значення випадкової величини X належать всій дійсній осі, тобто $x \in (-\infty; +\infty)$, то

$$M(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx. \tag{5.6}$$

Дисперсія неперервної випадкової величини, тобто якщо можливі значення належать відрізку $[a, b]$, то дисперсія визначається формулою:

$$D(X) = \int_a^b (X - M(X))^2 \cdot f(x)dx, \qquad (5.7)$$

де

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

Якщо можливі значення випадкової величини X належать всій дійсній осі, тобто $x \in (-\infty;\ +\infty)$, то дисперсія визначається формулою:

$$D(X) = \int_{-\infty}^{+\infty} (X - M(X))^2 \cdot f(x)dx, \qquad (5.8)$$

де

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

Формули (5.7), (5.8), як у випадку дискретної випадкової величини, можна записати у вигляді:

$$D(X) = \int_a^b x^2 \cdot f(x)dx - (M(X))^2\ ; \qquad (5.9)$$

$$D(X) = \int_{-\infty}^{+\infty} x^2 \cdot f(x)dx - (M(X))^2. \qquad (5.10)$$

Середнє квадратичне відхилення, аналогічно як і для дискретної випадкової величини, визначається за формулою:

$$\sigma(X) = \sqrt{D(X)}. \qquad (5.11)$$

Приклад 5.9. Знайти математичне сподівання, дисперсію та середнє квадратичне відхилення випадкової величини *X*, якщо

$$f(x) = \begin{cases} 0 & \text{при } x \le 0; \\ 2x & \text{при } 0 < x \le 1\ ; \\ 0 & \text{при } x > 1. \end{cases}$$

Розв'язання. Математичне сподівання знаходимо за формулою (5.6):

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

Підставивши значення, одержимо:

$$M(X) = \int_0^1 x \cdot 2xdx = 2\int_0^1 x^2dx = 2\frac{x^3}{3}\Big|_0^1 = \frac{2}{3}.$$

Дисперсію визначаємо за формулою (5.9).
Підставивши значення, одержимо:

$$D(X) = \int_a^b x^2 \cdot f(x)dx - (M(X))^2 = \int_0^1 x^2 \cdot 2xdx - \left(\frac{2}{3}\right)^2 = \frac{1}{18}.$$

Середнє квадратичне відхилення визначається формулою (5.11):

$$\sigma(X) = \sqrt{D(X)} = \sqrt{\frac{1}{18}}.$$

Приклад 5.10. Знайти математичне сподівання і дисперсію випадкової величини *X*, якщо на інтервалі $(-\frac{\pi}{2}, \frac{\pi}{2})$ диференціальна функція $f(x) = \frac{2}{\pi}\cos^2 x$ і дорівнює нулю при $|x| > \frac{\pi}{2}$.

Розв'язання.
Математичне сподівання для неперервної величини *X*:

$$M(X) = \int_{-\infty}^{+\infty} x \cdot \frac{2}{\pi}\cos^2 x\,dx = \int_{-\infty}^{-\frac{\pi}{2}} x \cdot 0dx + \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} x \cdot \frac{2}{\pi}\cos^2 x\,dx =$$

$$= \int_{-\infty}^{-\frac{\pi}{2}} x \cdot 0dx = 0.$$

Оскільки $\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} x \cdot \frac{2}{\pi}\cos^2 x\,dx = 0$, як інтеграл від непарної функції.

Дисперсія випадкової величини X:

$$D(X) = \int_{-\infty}^{+\infty} x^2 \cdot \frac{2}{\pi} \cos^2 x \, dx - (M(X))^2 = \int_{-\infty}^{+\infty} x^2 \cdot \frac{2}{\pi} \cos^2 x \, dx,$$

бо числове значення математичного сподівання $M(X) = 0$

$$D(X) = \int_{-\infty}^{+\infty} x^2 \cdot \frac{2}{\pi} \cos^2 x \, dx = \frac{1}{\pi} \int_{-\infty}^{+\infty} x^2 \cdot (1 + \cos\ 2x) \, dx$$

$$= \frac{2}{\pi} \left(\frac{x^3}{3} + \frac{x^3 \cdot sin2x}{2} + \frac{x^2 \cdot cos2x}{2} + \frac{sin2x}{4} \right) \Big| \begin{matrix} \frac{\pi}{2} \\ 0 \end{matrix} = \frac{\pi^2}{12} - \frac{1}{12}.$$

5.4. Числові характеристики нормального закону розподілу

Нехай неперервна випадкова величина X розподілена за нормальним законом:

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

За означенням математичного сподівання для неперервної випадкової величини маємо:

$$M(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx = \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi \cdot}} e^{\frac{-(x-a)^2}{2 \cdot \sigma^2}} dx$$

$$= \frac{1}{\sigma\sqrt{2\pi \cdot}} \int_{-\infty}^{+\infty} x \cdot e^{\frac{-(x-a)^2}{2 \cdot \sigma^2}} dx = a.$$

Інтеграл шукається за допомогою заміни $z = \frac{x-a}{\sigma}$, $x = a - \sigma z$, $dx = \sigma dz$.

Отже, **математичне сподівання нормального закону розподілу** дорівнює параметру a диференціальної функції розподілу

$$M(X) = a. \qquad (5.12)$$

За означенням дисперсії неперервної випадкової величини, враховуючи, що $M(X) = a$, маємо:

$$D(X) = \int_{-\infty}^{+\infty} x^2 \cdot f(x)dx - (M(X))^2 = \int_{-\infty}^{+\infty} x^2 \cdot \frac{1}{\sigma\sqrt{2\pi\cdot}} e^{\frac{-(x-a)^2}{2\cdot\sigma^2}} dx - (a)^2 \quad =$$

$$\frac{1}{\sigma\sqrt{2\pi\cdot}} \int_{-\infty}^{+\infty} (x-a)^2 \cdot e^{\frac{-(x-a)^2}{2\cdot\sigma^2}} dx.$$

Інтеграл шукається за допомогою заміни $z = \frac{x-a}{\sigma}$, $x = a - \sigma z$, $dx = \sigma dz$.

$$D(X) = \sigma^2 \tag{5.13}$$

Звідси:

$$\sigma = \sqrt{D(X)} \tag{5.14}$$

Відтак **середнє квадратичне відхилення нормально розподіленої випадкової величини** *X* дорівнює параметру σ.

Ймовірність відхилення випадкової величини, розподіленої за нормальним законом, від свого математичного сподівання шукається за формулою:

$$P(|X - a| < \alpha) = 2\Phi\left(\frac{\alpha}{\sigma}\right). \tag{5.15}$$

5.5. Числові характеристики показникового закону розподілу

Нехай неперервна випадкова величина *X* розподілена за показниковим законом:

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

де λ – стала величина $\lambda > 0$.

Використовуючи формулу (5.6), одержимо:

$$M(X) = \int_0^{+\infty} x \cdot f(x)dx = \int_0^{+\infty} x \cdot \lambda e^{-\lambda x} dx = \lambda \int_0^{+\infty} x \cdot e^{-\lambda x} dx.$$

Інтегруючи частинами, одержимо:

$$M(X) = \frac{1}{\lambda} \tag{5.16}$$

Відтак **математичне сподівання показникового розподілу дорівнює оберненій величині параметра $\boldsymbol{\lambda}$**.

Використовуючи формулу (1.6) для обчислення дисперсії, одержимо:

$$D(X) = \int_0^{\infty} x^2 \cdot f(x)dx - (M(X))^2 = \lambda \int_0^{\infty} x^2 \cdot e^{-\lambda x} dx - \frac{1}{\lambda^2};$$

$$D(X) = \int_0^{\infty} x^2 f(x)dx - [M(X)]^2 = \lambda \int_0^{\infty} x^2 e^{-\lambda x} dx - \frac{1}{\lambda^2}.$$

Інтегруючи частинками, одержимо:

$$D(X) = \frac{1}{\lambda^2}. \tag{5.17}$$

Звідси середнє квадратичне відхилення, як корінь квадратний із дисперсії, $\sigma(X) = \sqrt{D(X)}$ дорівнює:

$$\sigma(X) = \sqrt{\frac{1}{\lambda^2}} = \frac{1}{\lambda}. \tag{5.18}$$

Відтак математичне сподівання і середнє квадратичне відхилення показникового закону розподілу рівні між собою, тобто:

$$M(X) = \sigma(X) = \frac{1}{\lambda}.$$

Приклад 5.11. Знайти числові характеристики випадкової величини, розподіленої за законом:

$$f(x) = \begin{cases} 0 & \text{при } x < 0; \\ 4e^{-4x} & \text{при } x \geq 0. \end{cases}$$

Розв'язання. У цьому разі випадкова величина X розподілена за показниковим законом із параметром $\lambda = 4$.

Згідно з формулами (5.16), (5.17) та (5.18) маємо:

$$M(X) = \frac{1}{\lambda} = \frac{1}{4};$$

$$D(X) = \frac{1}{\lambda^2} = \frac{1}{4^2} = \frac{1}{16};$$

$$\sigma(X) = \sqrt{\frac{1}{\lambda^2}} = \frac{1}{\lambda} = \frac{1}{4}.$$

Приклад 5.12. Знайти числові характеристики випадкової величини, розподіленої за законом:

$$f(x) = \begin{cases} 0 & \text{при } x < 0; \\ \dfrac{2}{\mathrm{e}^{2x}} & \text{при } x \geq 0. \end{cases}$$

Розв'язання. Запишемо випадкову величину *X* у вигляді

$$f(x) = \begin{cases} 0 & \text{при } x < 0; \\ 2 \cdot \mathrm{e}^{-2x} & \text{при } x \geq 0. \end{cases}$$

Отримаємо випадкову величину, розподілену за показниковим законом із параметром $\lambda = 2$.

Згідно з формулами (5.16), (5.17) та (5.18) маємо:

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

$$D(X) = \frac{1}{\lambda^2} = \frac{1}{2^2} = \frac{1}{4};$$

$$\sigma(X) = \sqrt{\frac{1}{\lambda^2}} = \frac{1}{\lambda} = \frac{1}{2}.$$

5.6. Кореляційний момент. Коефіцієнт кореляції

Для характеристики зв'язку між випадковими величинами *X* і *У* використовується кореляційний момент.

Кореляційним моментом $\mathbf{M_{xy}}$ випадкових величин *X* і *У* називають математичне сподівання добутку відхилень цих величин від своїх математичних сподівань:

$$M_{xy} = M\left[\left(X - M(X)\right) \cdot (Y - M(Y))\right]. \qquad (5.19)$$

Для обчислення кореляційних моментів дискретних випадкових величин використовують формулу:

$$M_{xy} = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\left(x_i - M(x)\right) \cdot \left(y_j - M(y)\right) p(x_i, y_j)\right). \qquad (5.20)$$

$$M_{xy} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(x - M(x)\right) \cdot \left(y - M(y)\right) f(x, y) dx dy. \qquad (5.21)$$

Із визначення кореляційного моменту випливає, що він має розмірність, рівну добутку розмірностей випадкових величин *X* і *У*. Для того, щоб мати однакову розмірність, або безрозмірну величину, яка характеризує випадкові величини *X* і *У*, вводять коефіцієнт кореляції r_{xy}.

Коефіцієнтом кореляції r_{xy} – випадкових величин *X* і *У* називають відношення кореляційного моменту до добутку середніх квадратичних відхилень цих величин, тобто:

$$r_{xy} = \frac{M_{xy}}{\sigma_x \cdot \sigma_y}. \qquad (5.19)$$

Властивості коефіцієнта кореляції:

1) $|r_{xy}| \le 1$;

2) якщо *X* та *Y* незалежні, то $r_{xy} = 0$;

3) якщо між *X* та *Y* є лінійна залежність *Y* = *aX* + *b*, де a та b – постійні, то $|r_{xy}| = 1$.

Дві випадкові величини *X* і *У* називаються **корельованими**, якщо їх кореляційний момент (або те саме – коефіцієнт кореляції) відмінний від нуля.

Дві випадкові величини *X* і *У* називаються **некорельованими величинами**, якщо їх кореляційний момент дорівнює нулю.

Із означення корельованості величин *X* і *У* випливає їх залежність. Дійсно, припустимо протилежне: ми повинні вважати, що $M_{xy} = 0$, а це суперечить умові. Оскільки для корельованих величин $M_{xy} \neq 0$. Обернене твердження не завжди можливе, тобто якщо дві величини залежні, то вони можуть бути корельованими і некорельованими. Тобто, кореляційний момент двох залежних величин може не дорівнювати нулю, але і може дорівнювати.

Запитання для самоконтролю

1. Які числові характеристики існують для неперервних випадкових величин та що характеризує кожна з них?
2. Що називають математичним сподіванням дискретної випадкової величини?
3. В яких одиницях вимірюється математичне сподівання щодо одиниць вимірювання випадкової величини?
4. Вказати основні властивості математичного сподівання.
5. Що називають дисперсією дискретної випадкової величини?
6. Сформулюйте властивості дисперсії.
7. Що називають середнім квадратичним відхиленням дискретної випадкової величини?
8. Сформулюйте властивості середнього квадратичного відхилення.
9. Що таке мода?
10. Що таке медіана?
11. Запишіть числові характеристики основних законів розподілу дискретних випадкових величин.
12. Що називають математичним сподіванням середнього арифметичного однаково розподілених взаємонезалежних випадкових величин?
13. Що називають дисперсією середнього арифметичного однаково розподілених взаємно незалежних випадкових величин?
14. Що називають середнім квадратичним відхиленням середнього арифметичного n однаково розподілених взаємно незалежних випадкових величин?
15. За якими формулами обчислюють числові характеристики неперервних випадкових величин?
16. Чому дорівнює математичне сподівання нормального закону розподілу?
17. Чому дорівнює середнє квадратичне відхилення нормально розподіленої випадкової величини?
18. Чому дорівнює математичне сподівання показникового розподілу дорівнює оберненій величині?
19. Чому дорівнює середнє квадратичне відхилення показникового закону розподілу?
20. Що називають кореляційним моментом M_{xy}.
21. Для чого використовують кореляційний момент?

22. Що називають коефіцієнтом кореляції r_{xy}.

23. Сформулюйте властивості коефіцієнта кореляції

24. Коли дві випадкові величини *X* і *У* називаються корельованими?

25. Коли дві випадкові величини *X* і *У* називаються некорельованими величинами?

Задачі для самостійного розв'язування

1. Знайти за даним законом її розподілу:

а) математичне сподівання;

б) дисперсію;

в) середнє квадратичне відхилення дискретної випадкової величини.

1.

X	1	2	3	4	5
P	0,05	0,18	0,23	0,41	0,13

2.

X	2	3	4	5	8
P	0,25	0,15	0,27	0,08	0,25

3.

X	4	5	6	7	8
P	0,15	0,25	0,27	0,25	0,08

4.

X	4	5	6	8	10
P	0,21	0,17	0,18	0,23	0,21

5.

X	3	6	7	8	9
P	0,15	0,18	0,36	0,21	0,10

6.

X	6	8	9	10	12
P	0,17	0,14	0,27	0,23	0,19

7.

X	2	3	5	6	9
P	0,31	0,08	0,15	0,21	0,25

8.

X	1	3	6	8	11
P	0,07	0,16	0,05	0,23	0,49

9.

X	3	5	8	10	12
P	0,41	0,18	0,06	0,21	0,14

10.

X	2	6	9	12	15
P	0,27	0,33	0,13	0,11	0,16

2. Знайти числові характеристики випадкової величини *X*, яка дорівнює кількості хлопчиків у сім'ї з трьома дітьми (вважаючи народження хлопчика і дівчинки рівноймовірними подіями).

3. Стрілець робить 4 постріли по мішені. Знайти числові характеристики випадкової величини *X*, яка дорівнює кількості влучень, вважаючи, що ймовірність влучення при одному пострілі 0,7.

4. У ящику лежить 5 пронумерованих кульок (номери від 1 до 5). Навмання виймається кулька. Випадкова величина *X* – номер кульки. Вважаючи, що ймовірність дістати будь-яку кульку однакова, знайти математичне сподівання, дисперсію та середнє квадратичне відхилення цієї випадкової величини, а також моду і медіану.

5. Нехай *X* – кількість появ числа 5 при двох киданнях грального кубика. Знайти закон розподілу величини X та її математичне сподівання.

6. Технологічний процес є таким, що брак складає 5% усіх виробів. Навмання взято 20 виробів. Знайти математичне сподівання та дисперсію кількості бракованих виробів.

7. На шляху руху автомобіля 6 світлофорів, кожен з яких дозволяє або забороняє рух автомобіля з ймовірністю 0,5. Скласти ряд розподілу. Чому дорівнює математичне сподівання та дисперсія цієї випадкової величини?

8. Студент знає відповідь на 20 питань з 25. Навмання і за умови рівноможливості виймається 5 питань. Знайти числові характеристики випадкової величини *X*, яка дорівнює кількості питань, на які студент знає відповідь.

9. У групі з 20 студентів троє відмінників. Випадково і за умови рівноможливості вибирається 4 студенти. Знайти числові характеристики випадкової величини *X*, яка дорівнює кількості відмінників серед вибраних студентів

10. Серед 10 деталей є 2 браковані. Для контролю навмання і за умови рівноможливості вибирають 3 деталі. Знайти числові характеристики випадкової величини *X*, яка дорівнює кількості бракованих деталей серед вибраних.

11. Знайти числові характеристики випадкової величини *X*, яка дорівнює кількості вгаданих номерів у лотереї «5 із 30» за умови рівноможливості вибору.

12. Магазин отримав 10000 пляшок мінеральної води. Ймовірність того, що при перевезенні пляшка розіб'ється, дорівнює 0,0004. Знайти числові характеристики випадкової величини, яка дорівнює кількості розбитих пляшок.

13. Підручник видано тиражем 10000 примірників. Ймовірність того, що зброшурований екземпляр виявиться неякісним, дорівнює 0,001. Знайти числові характеристики випадкової величини, яка дорівнює кількості неякісних екземплярів.

14. Автоматична телефонна станція обслуговує 10000 телефонних номерів. Ймовірність того, що протягом 1 хв на АТС надійде виклик від абонента, дорівнює 0,0004. Знайти числові характеристики випадкової величини *X*, яка дорівнює кількості викликів, що надійшли на АТС протягом 1 хв.

15. Ймовірність влучення у мішень, що рухається, з пістолета дорівнює 0,001. Відбувається 2000 пострілів. Знайти числові характеристики випадкової величини *X*, яка дорівнює кількості влучень у мішень.

16. Випадкова величина *X* рівномірно розподілена на відрізку [1; 10]. Знайти числові характеристики випадкової величини X а також моду та медіану.

17. Випадкову величину *X* задано інтегральною функцією розподілу $F(x)$. Знайти:

а) диференціальну функцію розподілу $f(x)$,

б) математичне сподівання і дисперсію *X*,

$$1.\ F(x)=\begin{cases} 0 & \text{при } x\le 0; \\ \frac{1}{3}x & \text{при } 0<x\le 2; \\ 1 & \text{при } x>2. \end{cases}$$

$$2.\ F(x)=\begin{cases} 0 & \text{при } x\le -1; \\ \frac{x}{5} & \text{при} -1<x\le 2; \\ 1 & \text{при } x>2. \end{cases}$$

$$3.\ F(x)=\begin{cases}0 & \text{при } x\le 0;\\ x^2-4x & \text{при } 0<x\le 1;\\ 1 & \text{при } x>1.\end{cases}$$

$$4.\ F(x)=\begin{cases}0 & \text{при } x\le -2;\\ 3x^2-2x-1 & \text{при} -2<x\le 1;\\ 1 & \text{при } x>1.\end{cases}$$

$$5.\ F(x)=\begin{cases}0 & \text{при } x\le 0;\\ \sin x & \text{при } 0<x\le \dfrac{\pi}{2}\ ;\\ 1 & \text{при } x>\dfrac{\pi}{2}.\end{cases}$$

$$6.\ F(x)=\begin{cases}0 & \text{при } x\le 0;\\ \dfrac{1}{2}(1-\cos x) & \text{при } 0<x\le \dfrac{\pi}{2}\ ;\\ 1 & \text{при } x>\dfrac{\pi}{2}.\end{cases}$$

$$7.\ F(x)=\begin{cases}0 & \text{при } x\le 0;\\ \operatorname{tg} x & \text{при } 0<x\le \dfrac{\pi}{4};\\ 1 & \text{при } x>\dfrac{\pi}{4}.\end{cases}$$

$$8.\ F(x)=\begin{cases}0 & \text{при } x\le -1;\\ x^2-5x-7 & \text{при} -1<x\le 2;\\ 1 & \text{при } x>2.\end{cases}$$

$$9.\ F(x)=\begin{cases}0 & \text{при } x\le 0;\\ \dfrac{1}{2}(1-\sin x) & \text{при } 0<x\le \dfrac{\pi}{2};\\ 1 & \text{при } x>\dfrac{\pi}{2}.\end{cases}$$

18. Знайти числові характеристики випадкової величини, розподіленої за законом:

$$1.\ f(x)=\begin{cases}0 & \text{при } x<0;\\ \frac{4}{e^{4x}} & \text{при } x\ge 0.\end{cases}$$

$$2.\ f(x) = \begin{cases} 0 & \text{при } x < 0; \\ 5e^{-5x} & \text{при } x \geq 0. \end{cases}$$

19. Всередині круга радіуса R випадково вибирається точка. Ймовірність того, що точка потрапить до будь-якої області, що є в крузі, пропорційна площі цієї області. Знайти функцію розподілу та дисперсію випадкової величини X, яка дорівнює відстані від точки до центра круга.

20. Випадкова величина доходу підприємства X має диференціальну функцію розподілу

$$f(x) = \begin{cases} 0 & \text{при } x < -2; \\ 0{,}2 & \text{при} -2 \leq x \leq 3; \\ 0 & \text{при } x > 3. \end{cases}$$

Знайти математичне сподівання, дисперсію та ймовірність одержання доходу $X \in (1, 5]$.

21. Диференціальна функція розподілу прибутку X підприємства відома

$$f(x) = \begin{cases} 0 & \text{при } x < 0; \\ \frac{2}{25}x & \text{при } 0 \leq x \leq 5; \\ 0 & \text{при } x > 5. \end{cases}$$

Знайти математичне сподівання прибутку, дисперсію та ймовірність одержання прибутку.

22. Диференціальна функція розподілу прибутку X підприємства відома

$$0.F.F(x) = \begin{cases} 0 & \text{при } x \leq 0; \\ x^5 - 4x^4 + x^3 & \text{при } 0 < x \leq 3; \\ 1 & \text{при } x > 3. \end{cases}$$

Знайти математичне сподівання прибутку, дисперсію та ймовірність одержання прибутку .

23. На дослідному полі площею 1000 га урожайність певного сорту пшениці має такий розподіл:

Урожайність (ц/га)	17	18	19	20	21	22	23
Площа (га)	40	100	150	360	210	80	60

Розглядаючи врожайність як випадкову величину, визначити її математичне сподівання, дисперсію і середнє квадратичне відхилення.

24. Нормально розподілена випадкова величина має математичне сподівання $M(X)$ та дисперсію $D(X)$. Записати диференціальну функцію її розподілу, якщо

а) $M(X) = 3, D(X) = 4$;
б) $M(X) = -1, D(X) = 9$;
в) $M(X) = 0, D(X) = 1$.

25. Випадкова величина X розподілена за законом Пуассона з параметром a, випадкова величина Y розподілена за тим же законом з параметром b. Ці випадкові величини незалежні.

Знайти закон розподілу, математичне сподівання та дисперсію їх суми, якщо

1) a = 0,3, b = 1,2;
2) a = 1,7, b = 2,1;
3) a = 5, b = 2;
4) a = 0,2, b = 1,5;

26. Знайти ймовірність того, що нормально розподілена випадкова величина X набуває значень, що містяться в інтервалі (α, β), якщо математичне сподівання величини X дорівнює a, а середнє квадратичне відхилення – σ.

1) $a = 4, \sigma = 2, \alpha = 8, \beta = 10$;
2) $a = 5, \sigma = 4, \alpha = 10, \beta = 12$;
3) $a = 3, \sigma = 2, \alpha = 5, \beta = 8$;
4) $a = 2, \sigma = 3, \alpha = 3, \beta = 6$;
5) $a = 3, \sigma = 4, \alpha = 8, \beta = 9$;
6) $a = 4, \sigma = 5, \alpha = 10, \beta = 16$;
7) $a = 5, \sigma = 8, \alpha = 18, \beta = 21$;
8) $a = 8, \sigma = 3, \alpha = 14, \beta = 17$;
9) $a = 2, \sigma = 8, \alpha = 12, \beta = 18$.

Розділ 6

ЗАКОН ВЕЛИКИХ ЧИСЕЛ. ЦЕНТРАЛЬНА ГРАНИЧНА ТЕОРЕМА

Теорія ймовірностей вивчає закономірності масових випадкових подій. Закономірності виявляються при великій кількості випадкових явищ, що відбуваються при однакових умовах.

Це означає, що характеристики випадкових подій і випадкових величин в цих умовах стають **стійкими**: середній їх результат (наприклад частота події, середні значення випадкових величин) перестає бути випадковою і може бути передбачений з великою часткою вірогідності.

Саме в стійкості середніх величин і полягає фізичний зміст «**закону великих чисел**»: за дуже великої кількості випадкових експериментів середній їх результат практично перестає бути випадковим і може бути передбаченим майже вірогідно.

У теорії ймовірностей під законом великих чисел мають на увазі групу математичних теорем, у кожній з яких за тих чи інших умов установлюється факт наближення середніх характеристик великої кількості випадкових експериментів до деяких невипадкових сталих.

Закон великих чисел – це свого роду зв’язок між теорією ймовірностей як математичною наукою і закономірностями випадкових явищ у разі масових спостережень над ними.

Другу групу граничних теорем можна об’єднують під назвою – **центральна гранична терема**. У всіх її формах визначаються умови виникнення нормального розподілу випадкової величини. На основі тверджень центральної граничної теореми випливає, що нормальний розподіл виникає тоді, коли підсумовується багато незалежних випадкових величин, які є порівняльними щодо порядку свого впливу на розсіювання суми.

Різні форми закону великих чисел і центральної граничної теореми відрізняються умовами, які накладаються на випадкові величини і їх суми.

6.1. Нерівність Чебишева

Для доведення граничних теорем, а також під час розв'язування задач важливу роль виконує нерівність Чебишева.

Розглянемо закон розподілу дискретної величини X:

X	x_1	x_2	...	x_n
p	p_1	p_2	…..	p_n

Знайдемо ймовірність відхилення випадкової величини від його математичного сподівання.

Перша нерівність Чебишева: для довільної випадкової величини X, яка набуває невід'ємні значення та має скінченне математичне сподівання

$$P(X \geq 1) \leq M(X). \tag{6.1}$$

Якщо X – дискретна випадкова величина, то

$$P(X \geq 1) = \sum_i p(x_i) \leq \sum_i x_i \cdot p(x_i) = M(X).$$

Якщо X – неперервна випадкова величина, $f(x)$ – щільність її ймовірностей, то

$$P(X \geq 1) = \int_1^\infty f(x)dx \leq \int_1^\infty xf(x)dx = M(X).$$

Якщо X – набуває лише невід'ємні значення

$$M(X) < \infty, \alpha > 0,$$

то

$$P(X \geq \alpha) \leq \frac{M(X)}{\alpha}. \tag{6.2}$$

Нерівність (6.1) називають **нерівністю Маркова**.

Друга нерівність Чебишева: ймовірність того, що відхилення випадкової величини X від його математичного сподівання (за абсолютною величиною менше додатного числа ε) задовольняє таку нерівність:

$$P\big(\big|X - M(X)\big| < \varepsilon\big) \geq 1 - \frac{D(X)}{\varepsilon^2}. \qquad (6.3)$$

Приклад 6.1. Випадкова величина X має характеристики $M(X)=1$ і $\sigma=0,2$. Скориставшись нерівністю Чебишева, оцінити ймовірності таких подій:

А. $A = \{0,5 < X < 1,5\}$.

Б. $B = \{0,75 < X < 1,35\}$.

В. $C = \{X < 2\}$.

Розв'язання: Використовуючи формулу (6.3), одержимо:

$$\text{А. } P(A) = P\{0,5 < X < 1,5\} = P\{\big|X - 1\big| < 0,5\} =$$
$$= 1 - P\{|X - 1| \geq 0,5\} \geq 1 - \frac{0,2^2}{0,5^2} = 0,84.$$

Отже, $P(A) \geq 0,84$.

Б. Використаємо те, що звуження інтервалу не може призвести до збільшення ймовірності:

$$P(B) = P\{0,75 < X < 1,35\} \geq P\{0,75 < X < 1,25\} =$$
$$= P\{|X - 1| < 0,25\} = 1 - P\{|X - 1| \geq 0,25\} \geq 1 - \frac{0,2^2}{0,25^2} = 0,36.$$

Отже, $P(B) \geq 0,36$.

$$\text{В. } P(C) = P\{X < 2\} = 1 - P\{X - 1 \geq 1\} \geq$$
$$\geq 1 - P\{|X - 1| \geq 1\} \geq 1 - \frac{0,2^2}{1^2} = 0,96.$$

Отже, $P(C) \geq 0,96$.

Додаткова інформація про розподіл випадкової величини дає змогу уточнити оцінку ймовірності на основі нерівності Чебишева.

Приклад 6.2. Дискретна випадкова величина задана законом розподілу.

X	-2	-1	2	3	4
p	0,1	0,2	0,3	0,15	0,25

Знайти ймовірність того, що $\left|X - M(X)\right| \leq 3$ та оцінити її, користуючись нерівністю Чебишева.

Розв'язання:

Знайдемо спочатку математичне сподівання та дисперсія:

$$\mathrm{M(X)} = (-2)\cdot 0{,}1 + (-1)\cdot 0{,}2 + 3\cdot 0{,}15 + 4\cdot 0{,}25 = 1{,}65;$$

$$D(X) = (-2)^2\cdot 0{,}1 + (-1)^2\cdot 0{,}2 + 3^2\cdot 0{,}15 + 4^2\cdot 0{,}25 - (1{,}65)^2 = 4{,}4275.$$

Тоді нерівність $\left|X - M(X)\right| \leq 3$ розпишеться:

$$\left|X - 1{,}65\right| \leq 3 \Rightarrow -3 + 1{,}65 < X < 3 + 1{,}65 \Rightarrow -1{,}35 < X < 4{,}65;$$

$$P(|X - 1{,}65| \leq 3) = 0{,}2 + 0{,}3 + 0{,}15 + 0{,}25 = 0{,}9$$

Оцінимо цю ймовірність за допомогою нерівності Чебишева:

$$P\left(\left|X - 1{,}65\right| \leq 3\right) \geq 1 - \tfrac{4{,}4275}{9} \approx 0{,}51.$$

Якщо, $\left(\left|X - M(X)\right| \geq \varepsilon\right)$, це еквівалентне $\left(\left|X - M(X)\right|^2 \geq \varepsilon^2\right)$ застосувавши першу нерівність Чебишева

$$P(|X - M(X)| \geq \varepsilon) = P(\frac{1}{\varepsilon^2}|X - M(X)|^2) \geq 1) \leq \frac{M((X - M(X)^2)}{\varepsilon^2}$$

$$P\left(\left|X - M(X)\right| \geq \varepsilon\right) = \tfrac{D(X)}{\varepsilon^2}. \qquad (6.4)$$

Приклад 6.3. Дисперсія випадкової величини X рівна 0,001. Знайти ймовірність того, що випадкова величина X відрізняється від її математичного сподівання більше ніж на 0,1.

Розв'язання. Для знаходження ймовірності використаємо нерівність (6.3)

$$P\left(\left|X - M(X)\right| > 0{,}1\right) = \tfrac{0{,}001}{0{,}1^2} = 0{,}1.$$

Наслідок з нерівності Чебишева: якщо всі послідовності незалежних випадкових величин *(X_1, X_2, ..., X_n)* мають скінченні математичні сподівання $M(X_i) = a$ і дисперсії $D(X_i) = \sigma$, то для будь-якого числа ε > 0

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - a\right| < \varepsilon\right) \geq 1-\frac{\sigma^2}{n\varepsilon^2}. \qquad (6.5)$$

З допомогою цього наслідку можна визначити, скільки необхідно провести вимірювань, щоб похибка вимірювання зі заданою ймовірністю не перевищувала заданого значення ε.

Приклад 6.4. Скільки потрібно взяти хлібин з середньою масою a кожної, щоб з ймовірністю, не меншою ніж 0,9, можна було стверджувати, що середнє арифметичне значення мас цих хлібин відрізняється від а за абсолютною величиною менше, ніж на 0,05 кг, якщо середнє квадратичне відхилення маси хлібини 0,1 кг.

Розв'язання:

Скористаємось формулою (6.4), яка для цієї задачі буде мати такий вигляд:

$$1-\frac{\sigma^2}{n\varepsilon^2} \geq 0{,}9 \Rightarrow \frac{0{,}01}{n0{,}0025} \leq 0{,}1 \Rightarrow n \geq \frac{0{,}01}{0{,}1 \cdot 0{,}0025} = 40.$$

Тобто для досягнення заданого відхилення потрібно взяти не менше 40 хлібин.

6.2. Закон великих чисел Чебишева

Теорема Чебишева. Якщо X_1, X_2, ..., X_n, попарно незалежні випадкові величини, причому дисперсії їх рівномірно обмежені, тобто $D(X_1) \leq C$, $D(X_2) \leq C$, ..., $D(X_n) \leq C$, де C=const, то для довільного ε>0 є така рівність:

$$\lim_{n\to\infty} P\left(\left|\bar{X} - M(\bar{X})\right| < \varepsilon\right) = 1, \qquad (6.6)$$

де

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Зміст теореми Чебишева (так званого *закону великих чисел*) полягає в тому, що середнє арифметичне доволі великого числа незалежних випадкових величин (дисперсії яких рівномірно обмежені) перестають бути випадковими величинами. Цей факт пояснюється тим, що при вимірюванні чи проведенні експерименту за істинну величину беруть середнє значення цих величин.

Середнє арифметичне великого числа незалежних випадкових величин втрачає випадковий характер і має **властивість стійкості**.

На теоремі Чебишева ґрунтується вибірковий метод, який використовують у статистиці. Суть методу полягає в тому, що на підставі вивчення певної ознаки для достатньо великої випадкової вибірки об'єктів роблять висновок про генеральну сукупність.

6.3. Теореми Бернуллі та Пуассона

Важливим частковим випадком теореми Чебишева є теорема Бернуллі.

Теорема Бернуллі. Нехай m – число появ події A в n незалежних випробуваннях, а p – ймовірність появи події в кожному випробуванні. Тоді з ймовірністю, близькою до одиниці, можна стверджувати, що при доволі великому n, частота події A не надто відрізнятиметься від ймовірності p.

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n}-p\right| < \varepsilon\right) = 1$$

Правильність теореми Бернуллі випливає з нерівності Чебишева (6.2), застосованої до випадкової величини $\frac{m}{n}$, тобто:

$$P\left(\left|\frac{m}{n}-p\right| < \varepsilon\right) \geq 1-\frac{p\cdot q}{n\cdot\varepsilon^2}\,. \tag{6.7}$$

Приклад 6.5. Ймовірність того, що автомат, який продає квитки, при опусканні монети працюватиме безпомилково дорівнює 0,96. За теоремою Бернуллі треба визначити, скільки слід зробити випробувань правильності роботи автомату, щоб з ймовірністю не меншою 0,99 можна було стверджувати, що відхилення частоти правильної роботи автомату від ймовірності правильної роботи за абсолютною величиною не перевищить 0,03.

Розв'язання.

Використаємо теорему Бернуллі:

$$P\left(\left|\frac{m}{n}-p\right| < \varepsilon\right) \geq 1-\frac{p\cdot q}{n\cdot\varepsilon^2}\,.$$

За умовою ε > 0,03, p = 0,96.

Отже, $q = 1 - p = 1 - 0{,}96 = 0{,}04$.

Вимога задачі $p \geq 0{,}99$ справджуватиметься, коли

$1 - \frac{p \cdot q}{n \cdot \varepsilon^2} \geq 0{,}99$ або при $\frac{p \cdot q}{n \cdot \varepsilon^2} \leq 0{,}01$, звідки $n \geq \frac{0{,}96 \cdot 0{,}04}{0{,}01 \cdot 0{,}03^2} = \frac{12800}{3}$;

n може бути лише цілим числом, заокруглення треба виконувати в подібних випадках лише у бік збільшення n.

Отже, $n \geq 4267$. При таких значеннях n ймовірність P буде більша за 0,99.

Другим, більш загальним, окремим випадком теореми Чебишева, є теорема Пуасона.

Теорема Пуасона. При збільшенні кількості n – незалежних випробувань ймовірність того, що частота $\frac{m}{n}$ появ деякої події A не надто відрізняється від середньої величини ймовірностей появ події A в окремих випробуваннях та прямує до 1, якщо $n \to \infty$.

Тобто

$$\lim_{n \to \infty} P\left(\left|\frac{m}{n} - \bar{p})\right| < \varepsilon\right) = 1, \qquad (6.8)$$

де

$$\bar{p} = \frac{p_1 + p_2 + \ldots + p_n}{n}.$$

Крім законів великих чисел, описаних у теоремах, спостерігається ще одне доволі цікаве явище, яке полягає в тому, що за великої кількості випадкових доданків, кожний з яких робить лише невеликий внесок у загальну суму, розподіл кожного з випадкових доданків не впливає на сумарний результат.

6.4. Центральна гранична теорема

Центральні граничні теореми використовують під час розв’язування таких задач.

1. Коли сума багатьох випадкових величин не надто відрізняється від постійної величини, тобто майже перестає бути випадковою величиною і тому її поведінка може прогнозуватись із значною ймовірністю?

2. За яких умов можна із значною ймовірністю прогнозувати число появ деякої випадкової події при великій кількості незалежних випробувань?

3. При яких обмеженнях сума багатьох випадкових величин буде розподілена за нормальним законом?

Теорема (центральна гранична теорема). Нехай Xi – незалежні однаково розподілені випадкові величини з математичним сподіванням $M\ (Xi\ ) = a$ і дисперсією $D\ (Xi\ ) =\sigma^2$. Тоді при великому n розподіл суми $X_1 + X_2 + \ldots + X_n = X$ близький до нормального розподілу.

При цьому

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - na}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{\alpha - na}{\sigma\sqrt{n}}\right). \qquad (6.9)$$

Отже,

$$P\left(\left|x - na\right| \leq \delta\right) = P\left(na - \delta < x < na + \delta\right) \approx 2\Phi\left(\frac{\delta}{\sigma\sqrt{n}}\right). \qquad (6.10)$$

Із останніх рівностей випливає, що ймовірність p події можна оцінити, використовуючи експериментальні результати.

Справді, якщо в кожному випробуванні подія A відбувається з ймовірністю p, і в серії з n незалежних випробувань досліджувана подія A відбувається k разів, то статистична частота появи події A дорівнює $\frac{k}{n}$. Ця частота є випадковою величиною X, яка є сумою незалежних випадкових величин X_k, що набувають значень $\frac{1}{n}$ з ймовірністю р, коли подія A відбувається у k-му випробуванні, і 0 з ймовірністю $q=1-p$ в іншому разі, $k=1, 2, \ldots, n$. Для таких випадкових величин X_k математичне сподівання:

$$M(X_k) = \frac{1}{n} \cdot p + 0 \cdot \left(1 - p\right) = \frac{p}{n},$$

а дисперсія $\sigma^2 = D(X_k) = \frac{p \cdot (1-p)}{n^2}$, $k=1, 2, \ldots, n$.

Тоді при доволі великих значеннях n $(n \geq 30)$ виконується рівність

$$P\left(\left|\frac{k}{n} - p\right| \leq \delta\right) = 2\Phi\left(\frac{\delta}{\sigma\sqrt{n}}\right) = 2\Phi\left(\delta \cdot \sqrt{\frac{n}{p \cdot (1-p)}}\right). \qquad (6.11)$$

Приклад 6.6. При виробництві цифрових дисків брак становить 1%. Скільки дисків потрібно відібрати для перевірки якості, щоб з ймовірністю 0,95 можна було стверджувати, що у випадковій вибірці дисків відсоток бракованих відрізняється від 1% не більше як на 0,5%?

Розв'язання. Кількість бракованих дисків є випадковою величиною $X_1+X_2+\ldots+X_n=X$, де $X_1, X_2, \ldots, X_n$ – незалежні однаково розподілені випадкові величини; X_i – випадкова величина, яка дорівнює кількості бракованих дисків при виготовленні одного диска, тобто X_i може набувати значення або 0, або 1 з ймовірністю відповідно $p=0{,}99$ і $q=0{,}01$.

Якщо n – доволі велике число, то за центральною граничною теоремою розподіл випадкової величини X близький до нормального, причому

$$P\left(\left|\frac{k}{n}-p\right| \le \delta\right) = 2\Phi(\delta \cdot \sqrt{\frac{n}{p\cdot(1-p)}}),$$

де $\frac{k}{n}$ – частота браку; $p=0{,}01$; $\delta = \frac{0{,}5}{100} = 0{,}005$; n – невідома кількість дискет.

Число n потрібно вибрати таким, щоб

$$P\left(\left|\frac{k}{n}-0{,}01\right| \le 0{,}005\right) = 2\Phi\left(0{,}005 \cdot \sqrt{\frac{n}{0{,}01\cdot(1-0{,}01)}}\right) = 0{,}95.$$

Тобто,

$$2\Phi\left(0{,}005 \cdot \sqrt{\frac{n}{0{,}01\cdot 0{,}99}}\right) = 0{,}475.$$

За таблицею значень функції Лапласа (Додатки, таблиця 2) знаходимо значення аргументу x таке, що $\Phi\ (x) = 0{,}475$ тоді $x = 1{,}96$.

Розв'язавши рівняння

$$\left(0{,}005 \cdot \sqrt{\frac{n}{0{,}01\cdot 0{,}99}}\right) = 1{,}96,$$

отримаємо $n \ge 1522$.

Приклад 6.7. Ймовірність того, що деталь не є стандартною дорівнює 0,1. Знайти ймовірність того, що серед виготовлених 400 деталей відносна частота появи нестандартної деталі відхиляється від ймовірності 0,1 не більше, ніж на 0,03.

Розв'язання. Для знаходження ймовірності скористаємось формулою (6.10).

За умовою задачі маємо $n = 400, p = 0{,}1, \delta = 0{,}03$.

$$P\left(\left|\frac{k}{400}-0{,}1\right| \leq 0{,}03\right) = 2\Phi(0{,}03 \cdot \sqrt{\frac{400}{0{,}1\cdot(1-0{,}01)}});$$

$$P\left(\left|\frac{k}{400}-0{,}1\right| \leq 0{,}03\right) = 2\Phi(2).$$

За таблицею значень функції Лапласа (див. Додатки, таблиця 2) знаходимо значення *Ф (2)=0,4772.*

$$P\left(\left|\frac{k}{400}-0{,}1\right| \leq 0{,}03\right) = 2\Phi(2) = 2 \cdot 0{,}4772 = 0{,}9544.$$

Запитання для самоконтролю

1. Що називають законом великих чисел?
2. Що розуміють під граничними теоремами теорії ймовірності?
3. У чому полягає фізичний зміст «закону великих чисел»?
4. Чим відрізняються різні форми закону великих чисел і центральної граничної теореми?
5. Яку нерівність називають нерівністю Маркова?
6. Запишіть нерівністі Чебишева.
7. Сформулюйте наслідок з нерівності Чебишева.
8. Що можна визначити за допомогою наслідку з нерівності Чебишева?
9. Сформулюйте теорему Чебишева.
10. У чому полягає зміст теореми Чебишева (так званого закону великих чисел).
11. Який метод, що використовується в статистиці, ґрунтується на теоремі Чебишева.
12. Середнє арифметичне великого числа незалежних випадкових величин втрачає випадковий характер і має … ?
13. Сформулюйте теорему Бернуллі.
14. Запишіть теорему Пуасонна.
15. Наслідком яких теорем є інтегральна теорема Лапласа?
16. Сформулюйте центральну граничну теорему.
17. Під час розв'язування яких задач використовують центральні граничні теореми?
18. При яких значеннях *n* виконується рівність

$$P\left(\left|\frac{k}{n}-p\right| \leq \delta\right) = 2\Phi\left(\frac{\delta}{\sigma\sqrt{n}}\right) = 2\Phi(\delta \cdot \sqrt{\frac{n}{p\cdot(1-p)}})?$$

Задачі для самостійного розв'язування

1. Використовуючи нерівність Чебишева, оцінити ймовірність того, що $|X - M(X)| \leq 3$, якщо $D(X) = 0{,}0025$. Оцінити ймовірність тієї самої події, якщо відомо, що X має нормальний розподіл.

2. Використовуючи нерівність Чебишева, оцінити ймовірність того, що $|X - M(X)| > 0{,}1$, якщо $\sigma(X) = 0{,}4$. Оцінити ймовірність тієї самої події, якщо відомо, що X має нормальний розподіл.

3. Сума всіх вкладів в банку становить 20 млн грн, а ймовірність того, що випадково вибраний вклад не перевищує 10000 грн, дорівнює 0,8. Оцініть кількість вкладників.

4. Маса деталей, що виготовляються на верстаті, є випадковою величиною, середнє значення якої (математичне сподівання) дорівнює 1,2 кг. Дисперсія цієї величини дорівнює 0,012. Використовуючи нерівність Чебишева, оцінити ймовірність того, що:

а) відхилення маси деталі від її середнього значення за абсолютною величиною не перевищить 0,2;

б) маса деталі набуде значення від 1,18 до 1,22.

5. Використовуючи нерівність Чебишева, оцінити ймовірність появи події A, яка полягає у тому, що випадкова величина X набуде значення, яке відрізнятиметься від математичного сподівання $M(X)$ на величину, що не перевищує потроєного середнього квадратичного відхилення. Чи зміниться відповідь, якщо відомо, що випадкова величина X має нормальний розподіл?

6. Річна виручка авіакомпанії від перевезення пасажирів – випадкова величина з середнім значенням 200 млн грн і стандартним (середнім квадратичним) відхиленням 20 млн грн. Знайти:

а) оцінку ймовірності того, що в наступному році авіакомпанія матиме виручку не менше 220 млн грн;

б) оцінку ймовірності того, що виручка міститиметься в межах від 180 до 220 млн грн;

в) в який межах з імовірністю не меншою за 0,95 можна очікувати виручку в наступному році.

7. Скільки потрібно провести незалежних випробувань, щоб ймовірність нерівності $\left|\frac{m}{n} - p\right| \leq 0{,}06$ перевищувала б 0,78, якщо ймовірність появи події в окремому випробуванні $p=0{,}7$.

8. Ймовірність того, що покупцю взуттєвого магазину потрібні черевики 40-го розміру дорівнює 0,15. Оцініть межі відсотка покупців серед 2000 відвідувачів магазину, яким потрібні черевики 40-го розміру, якщо ці межі потрібно гарантувати з ймовірність 0,98.

9. Скільки виробів необхідно відібрати для перевірки якості продукції, щоб з імовірністю 0,95 можна було стверджувати, що частка бракованих деталей відрізняється від імовірності $p=0,1$ випуску бракованої деталі не більше ніж на 0,02?

10. При виробництві поліетиленових пакетів брак становить 5%. Скільки виробів потрібно відібрати для перевірки якості продукції, щоб з ймовірністю 0,9 можна було стверджувати, що частка бракованих пакетів становить від 4 до 6%?

11. Із 1000 виробів навмання відібрано 200. Після перевірки серед них виявлено 15 бракованих. Прийнявши частку бракованих виробів за ймовірність виготовлення браку, оцінити ймовірність того, що в усій партії бракованих виробів буде не більше 10% і не менше 5%.

12. Скільки необхідно провести експериментів n, щоб ймовірність відхилення відносної частоти появи випадкової події W(A) від імовірності $p = 0,85$, взяте за абсолютною величиною, на $e = 0,001$, була б не меншою за 0,99.

13. Ймовірність виготовити стандартну деталь робітником дорівнює 0,95. Контролю підлягає 400 деталей. Оцінити ймовірність відхилення відносної частоти появи стандартної деталі *W(A)* від імовірності 0,95 не більше ніж на величину 0,02.

14. Унаслідок медичного огляду 900 допризовників було виявлено, що середня маса кожного з них на 1,2 кг більша від середньої маси попереднього призову. Чи можна це констатувати як випадковість, якщо середнє відхилення маси допризовника дорівнює 8 кг?

15. Дисперсія кожної із 4500 незалежних випадкових величин, що мають один і той самий закон розподілу ймовірностей, дорівнює 5. Оцінити ймовірність того, що відхилення середнього арифметичного цих величин від середнього арифметичного їх математичних сподівань, взяте за абсолютною величиною, не перевищить 0,4.

16. Додають 24 незалежні випадкові величини, кожна з яких рівномірно розподілена на (0;1). Написати наближений вираз для щільності розподілу суми цих випадкових величин і обчислити ймовірність того, що їх сума лежить в межах від 10 до 14.

17. Кожна з 14 незалежних випадкових величин розподілена за рівномірним законом на інтервалі (0,1). Записати наближений вигляд для густини об’єднання цих випадкових величин. Знайти ймовірність того, що сума буде в межах від 6 до 8.

18. У певній місцевості середня річна кількість сонячних днів дорівнює 100. Оцінити ймовірність того, що протягом року в цій місцевості буде не більше ніж 125 сонячних днів.

19. В деякій місцевості середня швидкість вітру на певній висоті дорівнює 20 км/год, а середнє квадратичне відхилення 4 км/год. Оцінити швидкість вітру на цій висоті з ймовірністю не меншою, ніж 0,9.

20. Похибка спостереження X при обчислюванні довжини розподілена нормально з параметрами а = 5 мм і σ = 4 мм. Знайти ймовірність того, що значення, що вимірюється, відхиляється від дійсного більше ніж на 10 мм.

21. Довести, що при центруванні та нормуванні нормально розподіленої випадкової величини X одержимо нормально розподілену випадкову величину з математичним сподіванням 0 та дисперсією 1.

22. Відхилення розміру деталі від стандарту є випадкова величина, яка розподілена нормально з $M(X) = 0$ та $D(X) = 36$. Скільки потрібно виготовити деталей, щоб з імовірністю не менше 0,95 можна було стверджувати, що серед них буде хоч би одна стандартна, якщо дозволено відхилення розміру деталі від стандарту в інтервалі (–2, 4).

23. Ймовірність банкрутства фірми (подія A) р = 0,2. Яке число фірм потрібно відібрати, щоб ймовірність відхилення не більше ніж на 0,1 відносної частоти від ймовірності події була не меншою за 0,95.

24. Для визначення середньої маси хлібини в партії зі 100 однакових кареток взяли по хлібині з кожної. Оцінити знизу ймовірність того, що середня маса відібраних 100 хлібин відрізняється від середньої маси усієї партії за абсолютною величиною менше ніж на 0,05 кг, якщо відомо, що середнє квадратичне відхилення маси хлібини в кожній каретці менше, ніж 0,1 кг.

25. Зріст дорослих чоловіків є випадковою величиною, розподіленою за нормальним законом. Нехай математичне сподівання її дорівнює 170 см, а дисперсія – 36 см. Знайти щільність розподілу ймовірностей і функцію розподілу цієї випадкової величини. Обчислити ймовірність того, що:

а) хоча б один із навмання вибраних чоловіків матиме зріст 168–172 см;

б) хоча б один із чотирьох навмання вибраних чоловіків матиме зріст 168–172 см.

26. Дисперсія кожної з 2500 незалежних величин не перевищує 5. Оцінити ймовірність того, що абсолютна величина відхилення середнього арифметичного цих випадкових величин від середнього арифметичного їх математичних сподівань не перевищить 0,4.

27. Середнє квадратичне відхилення кожної з 3000 незалежних величин не перевищує 4. Оцінити ймовірність того, що абсолютна величина відхилення середньої арифметичної цих випадкових величин від середньої арифметичної їх математичних сподівань не перевищить 0,4.

28. Добова потреба електроенергії в населеному пункті є випадковою величиною, математичне сподівання якої дорівнює 2000 квт/год., а дисперсія – 20000. Оцінити ймовірність того, що в найближчий день витрата електроенергії в цьому населеному пункті становитиме 1500-25 – квт/год.

Частина II

МАТЕМАТИЧНА СТАТИСТИКА

Розділ 7

СТАТИСТИЧНІ РОЗПОДІЛИ ВИБІРКИ

Статистика (від італійського слова *stato* – держава) вивчає кількісний бік суспільних явищ і процесів у нерозривному зв'язку з їх якісним змістом.

Статистику поділяють на описову та пояснювальну. Прикладом описової статистики є книга рекордів Гіннеса. Пояснювальна статистика формулює певні висновки про ті чи інші явища, робить прогноз.

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

Математична статистика – розділ математики, в якому на основі дослідних даних вивчаються імовірнісні закономірності масових явищ.

Основними завданнями математичної статистики є статистична перевірка гіпотез, оцінка розподілу статистичних ймовірностей та його параметрів, вивчення статистичної залежності, визначення основних числових характеристик випадкових вибірок, якими є: вибіркове середнє, вибіркові дисперсії, стандартне відхилення.

Математична статистика виникла (XVII ст.) та почала розвиватись одночасно з теорією ймовірностей. Подальшим розвитком (кінець XIX – початок XX ст.) математичної статистики займалися П. Чебишов, А. Марков, О. Ляпунов, а також К. Гаусс, Ф. Гальтон, К. Пірсон та інші. У XX ст. найбільший внесок у математичну статистику зробили В. Романовський, Е. Слуцький, А. Колмогоров, Стьюдент (псевдонім У. Госсета), Е. Пірсон, Ю. Нейман, А. Вальд, А. Скороход, В. Королюк та інші вчені.

Прикладом перевірки статистичних гіпотез є з’ясування питання про те, змінюється чи не змінюється виробничий процес з часом. Прикладом оцінки параметрів є оцінка середнього значення статистичної змінної за дослідними даними. Для вивчення статистичної залежності використовують методи теорії кореляції. Загальні методи математичної статистики є основою теорії похибок.

Предмет математичної статистики полягає в розробці методів збору та обробки статистичних даних для одержання наукових та практичних висновків.

7.1. Генеральна та вибіркова сукупність

Основні поняття математичної статистики:

– **Генеральна сукупність** – це вся сукупність об’єктів, які досліджуються.

– **Вибірка або вибіркова сукупність** – це об’єкти, довільно або випадково відібрані з генеральної сукупності для дослідження.

– **Обсяг (об’єм) сукупності** – це кількість об’єктів цієї сукупності.

Приклад 7.1. Плоди дерева (200 шт.) обстежують на наявність специфічного смаку цього сорту. Для цього відбирають 10 штук. Тут 200 – обсяг генеральної сукупності, 10 – обсяг вибірки.

Вибірка повинна бути репрезентативною (представницькою), тобто правильно відображати ті властивості генеральної сукупності, які вивчаються. Досліджувані явища мають бути масовими. Лише тоді статистичні дані будуть достовірними.

Вибірки бувають повторні та безповторні. *Повторною* називають вибірку, за якої відібраний об’єкт повертається до генеральної сукупності перед відбором іншого об’єкта. Вибірку називають *безповторною*, якщо взятий об’єкт до генеральної сукупності не повертається. Найчастіше використовують безповторні вибірки.

План проведення статистичних досліджень:

1) формулюють завдання дослідження та визначають обсяг вибірки (мета, об’єкти вивчення, їх кількість, певні ознаки, характеристики);

2) збирають потрібні дані, вибирають доцільну форму їх подання для подальшого дослідження (водночас застосовують такі методи, як спостереження, порівняння, усне та письмове анкетування);

3) проводять остаточну обробку статистичного матеріалу та його вивчення;

4) за результатами формулюють певні висновки.

Компактною та наочною формою подання даних, отриманих у результаті статистичного дослідження, є: таблиці, графіки, стовпчасті та кругові діаграми.

7.2. Способи відбору статистичного матеріалу

У практичній діяльності використовують різноманітні способи відбору об'єктів із генеральної сукупності.

1. *Простий випадковий відбір.*

Навмання вибирається певна кількість об'єктів з усієї генеральної сукупності. Наприклад, вибираємо m об'єктів із n об'єктів генеральної сукупності. Для цього нумеруємо картки від 1 до n на об'єктах і вибираємо по одній картці і т. д. Простий відбір може бути повторним або безповторним. Для великого обсягу вибірки використовують готові таблиці «випадкових чисел».

2. *Типовий відбір.*

Об'єкти вибирають не з усієї сукупності, а з певної її типової частини. Наприклад, соціологи вивчають різні групи людей. Типовими ознаками можуть бути: місце проживання, вік, професія…

3. *Механічний відбір.*

Генеральну сукупність поділяють на стільки груп, частин, скільки планується об'єктів у вибірку. З кожної частини вибирають один об'єкт. Наприклад, якщо обстежується 20% телевізорів з партії, то беруть кожен п'ятий, якщо 10% – кожен десятий. Щоб механічний відбір був репрезентативним, треба враховувати специфіку технологічного процесу.

4. *Серійний відбір.*

Об'єкти з генеральної сукупності вибирають не по одному, а серіями, партіями. Серійний відбір використовують тоді, коли ознака, яку досліджують, мало змінюється в різних серіях.

На практиці розглянуті способи відбору поєднують і комбінують довільним способом.

7.3. Статистичний розподіл вибірки

Об'єкти обстежують за їх певними характеристиками або ознаками (вік, професія, …), які називають **варіантами**.

Сукупність значень ознаки (статистичної змінної), записаних у порядку їх зростання, називається **варіаційним рядом**.

Нехай X – статистична змінна, x_i $(i = 1, 2, \dots, n)$ – її значення, то тоді $x_1', x_2', \dots, x_n'$ – варіаційний ряд, якщо $x_{i+1}' \geq x_i'$.

Якщо кількість варіант велика, то сукупність їх значень для зручності розбивають на інтервали. Кожен інтервал характеризується одним числом – його серединою. Величину інтервалу ще називають інтервальною різницею. Частинні інтервали варіант, які розміщені у зростаючій послідовності, утворюють **інтервальний варіаційний ряд**.

Важлива характеристика – частота появи цієї варіанти.

Додатне число n_i, яке показує, скільки разів варіанта x_i $(i = 1, 2, \dots, m)$ трапляється в таблиці даних, називається **частотою** варіанти x_i.

Ряд $n_1, n_2, \dots, n_m$ називається **рядом частот**. Сума усіх частот повинна дорівнювати об'єму вибірки:

$$\sum_{i=1}^{m} n_i = n. \tag{7.1}$$

Статистичний розподіл вибірки встановлює зв'язок між рядом варіант, що зростає або спадає, і відповідними частотами. Він може бути зображений у вигляді таблиці

x_i	x_1	x_2	…	x_m
n_i	n_1	n_2	…	n_m

Відношення частоти n_i варіанти x_i до об'єму вибірки n називають **відносною частотою** і позначають

$$W_i = \frac{n_i}{n}. \tag{7.2}$$

Сума усіх відносних частот

$$\sum_{i=1}^{m} W_i = 1. \tag{7.3}$$

Залежність між упорядкованим рядом варіант і відповідними їм відносними частотами називається **статистичним розподілом відносних частот вибірки**:

x_i	x_1	x_2	…	x_m
W_i	$\frac{n_1}{n}$	$\frac{n_2}{n}$	…	$\frac{n_m}{n}$

Приклад 7.2. Під час дослідження кількісної ознаки X із генеральної сукупності було отримано вибірку

4; 3; 6; 4; 7; 2; 5; 1; 2; 5; 4; 4; 3; 5; 6; 3; 4; 1; 3; 4.

Знайти обсяг вибірки, побудувати її варіаційний ряд, статистичні розподіли частот та відносних частот.

Розв'язання. Оскільки вибірка містить 20 значень, то об'єм вибірки $n = 20$.

Побудуємо варіаційний ряд вибірки, тобто запишемо всі її значення у порядку зростання:

1; 1; 2; 2; 3; 3; 3; 3; 4; 4; 4; 4; 4; 4; 5; 5; 5; 6; 6;7.

У цій вибірці лише сім різних значень, тобто варіант:

1; 2; 3; 4; 5; 6; 7.

Знайдемо їх частоти:

$$n_1 = 2; \quad n_2 = 2; \quad n_3 = 4; \quad n_4 = 6; \quad n_5 = 3; \quad n_6 = 2; \quad n_7 = 1.$$

Запишемо шуканий статистичний розподіл частот вибірки:

x_i	1	2	3	4	5	6	7
n_i	2	2	4	6	3	2	1

Знайдемо відносні частоти варіант вибірки:

$$W_1 = \frac{2}{20} = 0{,}1; \; W_2 = \frac{2}{20} = 0{,}1; \; W_3 = \frac{4}{20} = 0{,}2; \; W_4 = \frac{6}{20} = 0{,}3;$$

$$W_5 = \frac{3}{20} = 0{,}15; \quad W_6 = \frac{2}{20} = 0{,}1; \quad W_7 = \frac{1}{20} = 0{,}05.$$

Отже, шуканий розподіл відносних частот має такий вигляд:

x_i	1	2	3	4	5	6	7
W_i	0,1	0,1	0,2	0,3	0,15	0,1	0,05

7.4. Емпірична функція розподілу та її властивості

За даними статистичного розподілу вибірки будується емпірична функція розподілу.

Емпіричною функцією розподілу (або функцією розподілу вибірки) називають функцію $F^*(x)$, яка визначає для кожного значення x відносну частоту події $X < x$.

Математично це означення записується:

$$F^*(x) = \frac{n_x}{n}, \tag{7.4}$$

де n_x – кількість варіант, які менші від x, n – об’єм вибірки.

Емпірична функція розподілу $F^*(x)$ має такі властивості:

1. $0 \le F^*(x) \le 1$;
2. $F^*(x)$– зростаюча функція;
3. $F^*(x) = \begin{cases} 0, & x \le x_1; \\ 1, & x > x_m, \end{cases}$

де x_1 – найменше значення варіанти, x_m – найбільше значення варіанти.

Приклад 7.3. Нехай маємо таку частотну таблицю (розподіл частот):

x_i	3	5	7	10	15
n_i	2	4	7	4	3

Записати емпіричну функцію розподілу та побудувати її графік.

Розв’язання. Об’єм вибірки:

$$n = \sum_{i=1}^{m} n_i = 2 + 4 + 7 + 4 + 3 = 20.$$

Оскільки найменше значення варіанти $x_1 = 3$, то

$$F^*(x) = 0$$

для всіх $x \le 3$.

Значення $x < 5$, а саме: $x_1 = 3$, спостерігається двічі, тому

$$F^*(x) = \frac{2}{20} = 0{,}1,$$

при $3 < x \le 5$.

Значення $x < 7$, а саме: $x_1 = 3$, $x_2 = 5$, спостерігалися $2 + 4 = 6$ разів, тому

$$F^*(x) = \frac{6}{20} = 0{,}3,$$

при $5 < x \le 7$.

Значення $x < 10$, а саме, $x_1 = 3$, $x_2 = 5$, $x_3 = 7$, спостерігалися $2 + 4 + 7 = 13$ разів, тому

$$F^*(x) = \frac{13}{20} = 0{,}65,$$

при $7 < x \leq 10$.

Значення $x < 15$, а саме, $x_1 = 3$, $x_2 = 5$, $x_3 = 7$, $x_4 = 10$, спостерігалися $2 + 4 + 7 + 4 = 17$ разів, тому

$$F^*(x) = \frac{17}{20} = 0{,}85,$$

при $10 < x \leq 15$.

Оскільки $x_5 = 15$ – найбільше значення варіанти, то

$$F^*(x) = 1$$

при $x > 15$.

Отже, запишемо шукану емпіричну функцію розподілу:

$$F^*(x) = \begin{cases} 0, \text{при } x \leq 3; \\ 0{,}1, \text{при } 3 < x \leq 5; \\ 0{,}3, \text{при } 5 < x \leq 7; \\ 0{,}65, \text{при } 7 < x \leq 10; \\ 0{,}85, \text{при } 10 < x \leq 15; \\ 1, \text{при } x > 15. \end{cases}$$

Графік цієї функції зображено на рисунку 7.1.

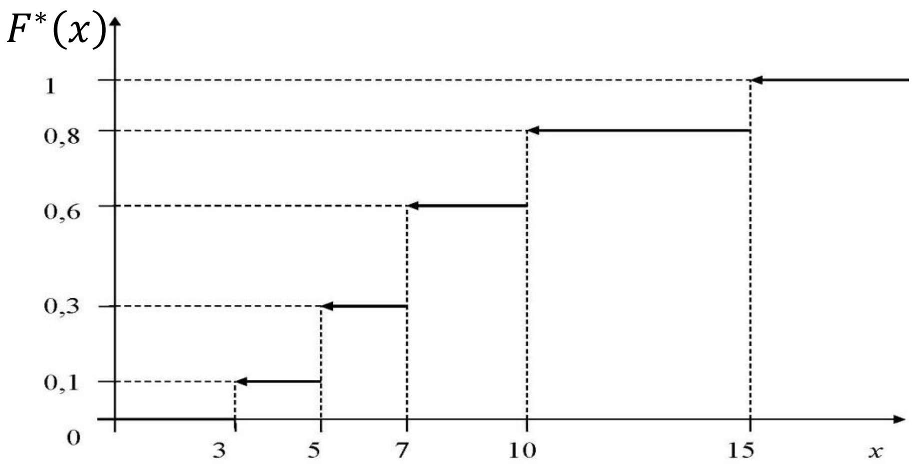

Рис. 7.1. Графік емпіричної функції розподілу

7.5. Згруповані розподіли вибірки

Часто для спрощення статистичного дослідження використовують згрупований розподіл вибірки.

Загальна схема побудови згрупованого розподілу частот

1. Визначити найбільше x_{max} та найменше x_{min} значення варіанти x_i і обчислити розмах варіант $R = x_{max} - x_{min}$.
2. Задати певне непарне число класів k. При загальному числі замірів $n \geq 100$ доцільно брати $9 \leq k \leq 15$, а при $n \leq 100$ можна вибрати $k = 7$.
3. Визначити ширину класу $h = \frac{R}{k}$.
4. Встановити границі класів і підрахувати кількість варіант у кожному класі. При підрахунку числа варіант значення x_i, що знаходиться на границі класів, слід відносити завжди до одного й того ж класу, а саме там, де це число трапилося вперше. Відтак воно стає нижньою границею класу.
5. Визначити частоту для кожного класу і записати ряд розподілу.

Згрупований розподіл накопиченої частоти та відносної частоти

Часто поряд з розподілом частот варіанти необхідно мати розподіл накопичених частот.

Розподіл накопиченої частоти одержують послідовним додаванням частот чергового інтервалу, починаючи з першого і закінчуючи останнім.

Розподіл накопиченої частоти (позначається F_i) дозволяє відповісти на питання: «Скільки існує варіант, які менші, ніж варіанта x_i ?».

Розподіл накопиченої відносної частоти одержують послідовним додаванням відносних частот чергового інтервалу, починаючи з першого і закінчуючи останнім.

Розподіл накопиченої відносної частоти (позначається $\frac{F_i}{n}$) дозволяє відповісти на питання: «Яка пропорція варіант, які менші, ніж x_i ?»

Згрупований розподіл щільності частот і щільності відносних частот

Якщо поділити всі частоти на ширину інтервалу h, то отримаємо **розподіл щільності частот вибірки** $\frac{n_i}{h}$.

Якщо поділити всі відносні частоти на ширину інтервалу h, то отримаємо **розподіл щільності відносних частот вибірки** $\frac{W_i}{h}$.

Приклад 7.4. У таблиці 1 наведена вибірка середньомісячної платні 100 співробітників фірми. Впорядкувати вибірку. Записати розподіл частот та відносних частот середньомісячної платні співробітників фірми; згрупований розподіл частот та відносних частот середньомісячної платні співробітників фірми; згрупований розподіл накопичених частот та накопичених відносних частот середньомісячної платні співробітників фірми; згрупований розподіл щільності частот та щільності відносних частот.

Таблиця 7.1

Вибірка середньомісячної платні 100 співробітників фірми									
338	348	304	314	326	314	324	304	342	308
336	304	302	338	314	304	320	321	322	321
312	323	336	324	312	312	364	356	362	302
322	310	334	292	362	381	304	366	298	304
381	368	304	298	368	290	340	328	316	322
302	314	292	342	321	322	290	332	298	296
296	298	324	338	352	326	318	304	332	322
360	312	331	331	304	316	332	282	342	338
342	322	324	325	302	328	354	330	316	324
334	350	334	324	332	340	324	314	326	323

Розв'язання. Запишемо розподіл частот та відносних частот цієї вибірки:

Таблиця 7.2

Розподіл частот та відносних частот середньомісячної платні співробітників фірми											
x_i	n_i	W_i	x_i	n_i	W_i	x_i	n_i	W_i	x_i	n_i	W_i
282	1	0,01	314	5	0,05	328	2	0,02	350	1	0,01
290	2	0,02	316	3	0,03	330	1	0,01	352	1	0,01
292	2	0,02	318	1	0,01	331	2	0,02	354	1	0,01
296	2	0,02	320	1	0,01	332	4	0,04	356	1	0,01
298	4	0,04	321	3	0,03	334	3	0,03	360	1	0,01
302	4	0,04	322	6	0,06	336	2	0,02	362	2	0,02
304	9	0,09	323	2	0,02	338	4	0,04	364	1	0,01
308	1	0,01	324	7	0,07	340	2	0,02	366	1	0,01
310	1	0,01	325	1	0,01	342	4	0,04	368	2	0,02
312	4	0,04	326	3	0,03	348	1	0,01	381	2	0,02

Наступним кроком в обробці даних, що веде до суттєвого спрощення досліджень, є групування.

Знаходимо максимальне та мінімальне значення варіанти вибірки:

$$x_{max} = 381, x_{min} = 282.$$

Розмах

$$R = x_{max} - x_{min} = 381 - 282 = 99.$$

Задамо число класів k =11.

Визначимо ширину класу $h = \frac{R}{k} = \frac{99}{11} = 9.$

Просумуємо частоти для кожного класу інтервалів, значення x_i, що знаходиться на границі класів, заносимо до того класу, де це число трапилось вперше. Результат запишемо у вигляді таблиці.

Таблиця 7.3

Згрупований розподіл частот середньомісячної платні співробітників фірми		
Інтервали платні	Частоти n_i	
282 – 291	3	=(1+2)
291 – 300	8	=(2+2+4)
300 – 309	14	=(4+9+1)
309 – 318	14	=(1+4+5+3+1)
318 – 327	23	=(1+3+6+2+7+1+3)
327 – 336	14	=(2+1+2+4+3+2)
336 – 345	10	=(4+2+4)
345 – 354	4	=(1+1+1+1)
354 – 363	4	=(1+1+2)
363 – 272	4	=(1+1+2)
372 – 381	2	=(2)
Разом:	100	

Запишемо розподіл накопичених частот, який одержимо послідовним додаванням частот чергового інтервалу.

Таблиця 7.4

Згрупований розподіл накопичених частот середньомісячної платні співробітників фірми		
Платня	Накопичені частоти F_i	
менше ніж 291	3	(=3)
менше ніж 300	11	(=3+8)
менше ніж 309	25	(=11+14)
менше ніж 318	39	(=25+14)
менше ніж 327	62	(=39+23)
менше ніж 336	76	(=62+14)
менше ніж 345	86	(=76+10)
менше ніж 354	90	(=86+4)
менше ніж 363	94	(=90+4)
менше ніж 372	98	(=94+4)
менше ніж 381	100	(=98+2)

Якщо поділити частоти та накопичені частоти на об'єм вибірки $n = 100$, то отримаємо, відповідно, розподіл відносних частот W_i та накопичених відносних частот $\frac{F_i}{n}$.

Запишемо розподіл відносних частот W_i та накопичених відносних частот $\frac{F_i}{n}$ (таблиця 7.5).

Таблиця 7.5

Згрупований розподіл відносних та накопичених відносних частот середньомісячної платні співробітників фірми			
Інтервали платні	Відносні частоти W_i	Платня	Накопичені відносні частоти $\frac{F_i}{n}$
282 – 291	0,03	менше ніж 291	0,03
291 – 300	0,08	менше ніж 300	0,11
300 – 309	0,14	менше ніж 309	0,25
309 – 318	0,14	менше ніж 318	0,39
318 – 327	0,23	менше ніж 327	0,62
327 – 336	0,14	менше ніж 336	0,76
336 – 345	0,1	менше ніж 345	0,86
345 – 354	0,04	менше ніж 354	0,9
354 – 363	0,04	менше ніж 363	0,94
363 – 272	0,04	менше ніж 372	0,98
372 – 381	0,02	менше ніж 381	1

Запишемо згрупований розподіл щільності частот та щільності відносних частот.

Для цього поділимо частоти та відносні частоти на ширину інтервалу $h = 10$. Щоб підсумувати результати, які одержали в *Прикладі 7.4*, зведемо разом в одну таблицю розглянуті розподіли *Таблиця 7.6*.

Ця таблиця містить усі важливі статистичні розподіли вибірки.

Таблиця 7.6

Статистичні розподіли вибірки							
x_i	n_i	W_i	$\frac{n_i}{h}$	$\frac{W_i}{h}$	x_i	F_i	$\frac{F_i}{n}$
282 – 291	3	0,03	$\frac{1}{3}$	$\frac{1}{300}$	менше, ніж 291	3	0,03
291 – 300	8	0,08	$\frac{8}{9}$	$\frac{2}{225}$	менше, ніж 300	11	0,11
300 – 309	14	0,14	$\frac{14}{9}$	$\frac{7}{450}$	менше, ніж 309	25	0,25
309 – 318	14	0,14	$\frac{14}{9}$	$\frac{7}{450}$	менше, ніж 318	39	0,39
318 – 327	23	0,23	$\frac{23}{9}$	$\frac{23}{900}$	менше, ніж 327	62	0,62
327 – 336	14	0,14	$\frac{14}{9}$	$\frac{7}{450}$	менше, ніж 336	76	0,76
336 – 345	10	0,1	$\frac{10}{9}$	$\frac{1}{90}$	менше, ніж 345	86	0,86
345 – 354	4	0,04	$\frac{4}{9}$	$\frac{1}{225}$	менше, ніж 354	90	0,9
354 – 363	4	0,04	$\frac{4}{9}$	$\frac{1}{225}$	менше, ніж 363	94	0,94
363 – 272	4	0,04	$\frac{4}{9}$	$\frac{1}{225}$	менше, ніж 372	98	0,98
372 – 381	2	0,02	$\frac{2}{9}$	$\frac{1}{450}$	менше, ніж 381	100	1

7.6. Полігон частот та відносних частот

Статистичні розподіли вибірки можуть бути представлені також графічно. Завдяки цьому ми можемо побачити характерні зміни ряду розподілу, не користуючись аналізом цифрових даних.

Якщо в результаті вибірки ми одержали статистичний розподіл ознаки X, яку треба дослідити, то будемо мати перелік варіант ознаки

$$x_1, x_2, x_3, \dots, x_m$$

та відповідних їм частот

$$n_1, n_2, n_3, \dots, n_m$$

або відносних частот

$$W_1, W_2, W_3, \dots, W_m.$$

Значення варіант та частот або відносних частот можна розглядати як координати точок

$$K_1(x_1;\ n_1),\ K_2(x_2;\ n_2),\ K_3(x_3;\ n_3),\ \dots,\ K_m(x_m;\ n_m)$$

або

$$L_1(x_1;\ W_1),\ L(x_2;\ W_2),\ L_3(x_3;\ W_3),\ \dots,\ L_m(x_m;\ W_m).$$

Полігоном частот називають ламану, відрізки якої з'єднують точки з координатами

$$K_1(x_1;\ n_1),\ K_2(x_2;\ n_2),\ K_3(x_3;\ n_3),\ \dots,\ K_m(x_m;\ n_m).$$

Полігоном відносних частот називають ламану, відрізки якої з'єднують точки з координатами

$$L_1(x_1;\ W_1),\ L(x_2;\ W_2),\ L_3(x_3;\ W_3),\ \dots,\ L_m(x_m;\ W_m).$$

Для побудови полігону частот на осі абсцис відкладають варіанти x_i ознаки X, а на осі ординат – відповідні їм частоти n_i. Точки $K_i(x_i;\ n_i)$ $(i = 1, 2, 3,\ \dots,\ m)$ з'єднують відрізками прямих і одержують полігон частот.

Для побудови полігону відносних частот на осі абсцис відкладають варіанти x_i ознаки X, а на осі ординат – відповідні їм відносні частоти W_i $(i = 1, 2, 3, \dots, m)$, а потім точки $L_i(x_i;\ W_i)$ з'єднують відрізками прямих.

При неперервному розподілі ознаки X у разі великої кількості спостережень увесь інтервал, в якому розміщені спостережні значення ознаки, як правило, розбивають на кілька частинних інтервалів однакової довжини h і знаходять

n_i – суму частот варіант, що потрапили в i-й інтервал. Для вибору оптимальної величини інтервалу рекомендовано використовувати формулу:

$$h \approx \frac{x_{max} - x_{min}}{1 + 3{,}2 \cdot \lg n},$$

де x_{max}, x_{min} – відповідно найбільше і найменше значення вибірки, n – об'єм вибірки.

Якщо задано інтервальний статистичний розподіл вибірки, то для побудови полігону частот за даними вибірки з'єднують точки, абсциси яких є значення середин частинних інтервалів, а ординатами – відповідні їм значення частот.

Приклад 7.5. Вибірку задано інтервальним розподілом частот:

$(x_i\,; x_{i+1})$	(1; 3)	(3; 5)	(5; 7)	(7; 9)	(9; 11)
n_i	13	9	5	16	7

Побудувати полігон відносних частот.

Розв'язання. Знайдемо об'єм вибірки:

$$n = 13 + 9 + 5 + 16 + 7 + 7 = 50.$$

Для побудови інтервального статистичного розподілу відносних частот визначимо відносні частоти:

$$W_1 = \frac{n_1}{n} = \frac{13}{50} = 0{,}26; \qquad W_2 = \frac{n_2}{n} = \frac{9}{50} = 0{,}18;$$

$$W_3 = \frac{n_3}{n} = \frac{5}{50} = 0{,}1; \qquad W_4 = \frac{n_4}{n} = \frac{16}{50} = 0{,}32;$$

$$W_5 = \frac{n_5}{n} = \frac{7}{50} = 0{,}14.$$

Побудуємо інтервальний статистичний розподіл відносних частот:

$(x_i\,; x_{i+1})$	(1; 3)	(3; 5)	(5; 7)	(7; 9)	(9; 11)
W_i	0,26	0,18	0,1	0,32	0,14

Знайдемо середини частинних інтервалів:

$$x_1^{\text{сер}} = \frac{x_1+x_2}{2} = \frac{1+3}{2} = 2; \qquad x_2^{\text{сер}} = \frac{x_2+x_3}{2} = \frac{3+5}{2} = 4;$$

$$x_3^{\text{сер}} = \frac{x_3+x_4}{2} = \frac{5+7}{2} = 6; \qquad x_4^{\text{сер}} = \frac{x_4+x_5}{2} = \frac{7+9}{2} = 8;$$

$$x_5^{\text{сер}} = \frac{x_5+x_6}{2} = \frac{9+11}{2} = 10.$$

Відкладемо на осі абсцис значення середин частинних інтервалів $x_i^{\text{сер}}$, а на осі ординат – значення відповідних їм відносних частот W_i. Послідовно з’єднуючи між собою точки ($x_i^{\text{сер}}$; W_i) відрізками, отримаємо полігон відносних частот (рис. 7.2).

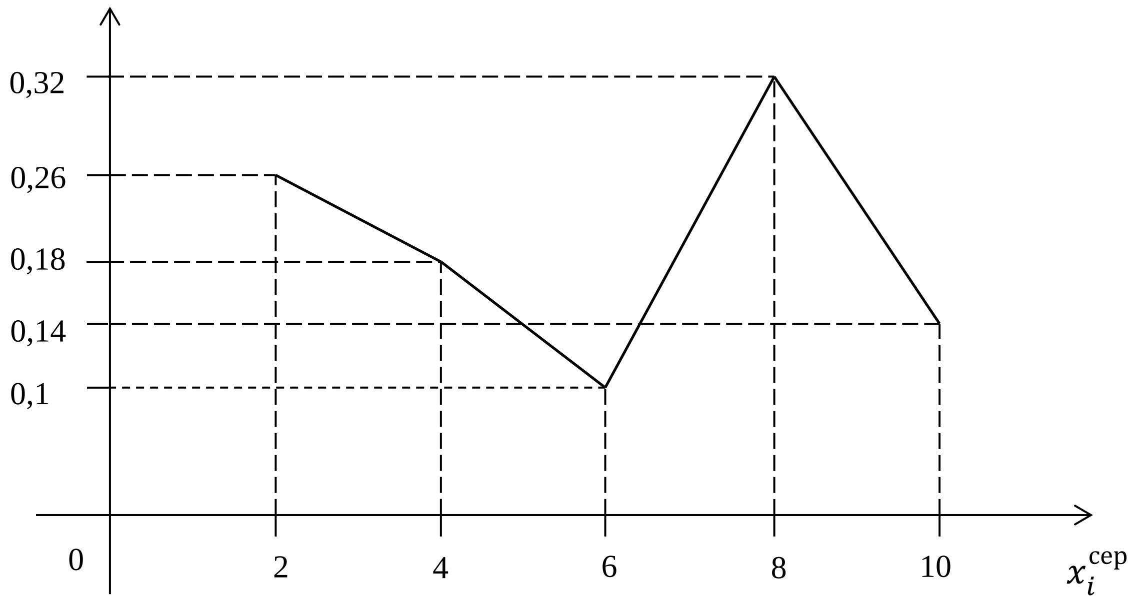

Рис. 7.2. Графік побудови полігону відносних частот

7.7. Гістограма частот та відносних частот

Гістограмою частот називають ступінчасту фігуру, яка складається з прямокутників, основами яких є частинні інтервали варіант довжиною $h = x_i - x_{i-1}$, а висоти дорівнюють $\frac{n_i}{h}$ (щільність частоти).

Гістограмою відносних частот називають ступінчасту фігуру, яка складається з прямокутників, основами яких є частинні інтервали варіант довжиною $h = x_i - x_{i-1}$, а висоти дорівнюють $\frac{W_i}{h}$ (щільність відносної частоти).

Площа гістограми частот дорівнює об'єму вибірки, а площа гістограми відносних частот – одиниці.

Для побудови гістограми частот (відносних частот) проміжок $[x_{min}\,;\,x_{max}]$, тобто від найменшого значення x_{min}, що спостерігали, до найбільшого значення x_{max}, розбивають на декілька відрізків рівної довжини h. Потім підраховують суму частот (відносних частот) тих значень варіант, які належать кожному із одержаних відрізків.

Якщо в i-му відрізку ($i = 1, 2, 3, \dots, m$) кількість варіант, що спостерігали, з врахуванням їх частот дорівнює n_i, то будують прямокутник Π_i, основою якого буде i-ий відрізок довжиною h, а висотою буде відрізок довжиною $\frac{n_i}{h}$ (для відносних частот висота – $\frac{W_i}{h}$). Площа такого прямокутника дорівнює $h \cdot \frac{n_i}{h} = n_i$ (у випадку відносних частот площа дорівнює $h \cdot \frac{W_i}{h} = W_i$). Тому площа усіх прямокутників буде дорівнювати сумі n_i (W_i), тобто $S = \sum_{i=1}^{m} n_i = n$ об'єму вибірки.

У випадку гістограми відносних частот площа прямокутників буде дорівнювати сумі усіх відносних частот

$$S = \sum_{i=1}^{m} W_i = 1.$$

Приклад 7.6. Вибірку задано розподілом частот:

x_i	2	4	6	8	10	12	14
n_i	2	3	5	1	4	2	3

Побудувати гістограму відносних частот.

Розв'язання. Щоб побудувати гістограму відносних частот на основі цього статистичного розподілу частот, складемо інтервальний статистичний розподіл відносних частот і знайдемо їх щільність.

Обсяг вибірки:

$$n = 2 + 3 + 5 + 1 + 4 + 2 + 3 = 20.$$

Відносні частоти:

$$W_1 = \frac{n_1}{n} = \frac{2}{20} = 0{,}1; \qquad W_2 = \frac{n_2}{n} = \frac{3}{20} = 0{,}15;$$

$$W_3 = \frac{n_3}{n} = \frac{5}{20} = 0{,}25; \qquad W_4 = \frac{n_4}{n} = \frac{1}{20} = 0{,}05;$$

$$W_5 = \frac{n_5}{n} = \frac{34}{20} = 0{,}2; \qquad W_6 = \frac{n_6}{n} = \frac{2}{20} = 0{,}1;$$

$$W_7 = \frac{n_7}{n} = \frac{1}{20} = 0{,}05.$$

Знайдемо частинні інтервали, їх довжину і щільність відносних частот. Частинні інтервали визначимо з умови: задані в дискретному статистичному розподілі варіанти мають бути серединами частинних інтервалів.

Отже, довжина частинних інтервалів

$$h = x_i - x_{i-1} = 2.$$

Шуканий інтервальний статистичний розподіл відносних частот має вигляд:

$(x_i; x_{i+1})$	(1; 3)	(3; 5)	(5; 7)	(7; 9)	(9; 11)	(11; 13)	(13; 15)
W_i	0,1	0,15	0,25	0,05	0,2	0,1	0,15

Щільності відносних частот:

$$\frac{W_1}{h} = \frac{0{,}1}{2} = 0{,}05; \qquad \frac{W_2}{h} = \frac{0{,}15}{2} = 0{,}075;$$

$$\frac{W_3}{h} = \frac{0{,}25}{2} = 0{,}125; \qquad \frac{W_4}{h} = \frac{0{,}05}{2} = 0{,}025;$$

$$\frac{W_5}{h} = \frac{0{,}2}{2} = 0{,}1; \qquad \frac{W_6}{h} = \frac{0{,}1}{2} = 0{,}05;$$

$$\frac{W_7}{h} = \frac{0{,}15}{2} = 0{,}075.$$

Побудуємо гістограму відносних частот (рис. 7.3).

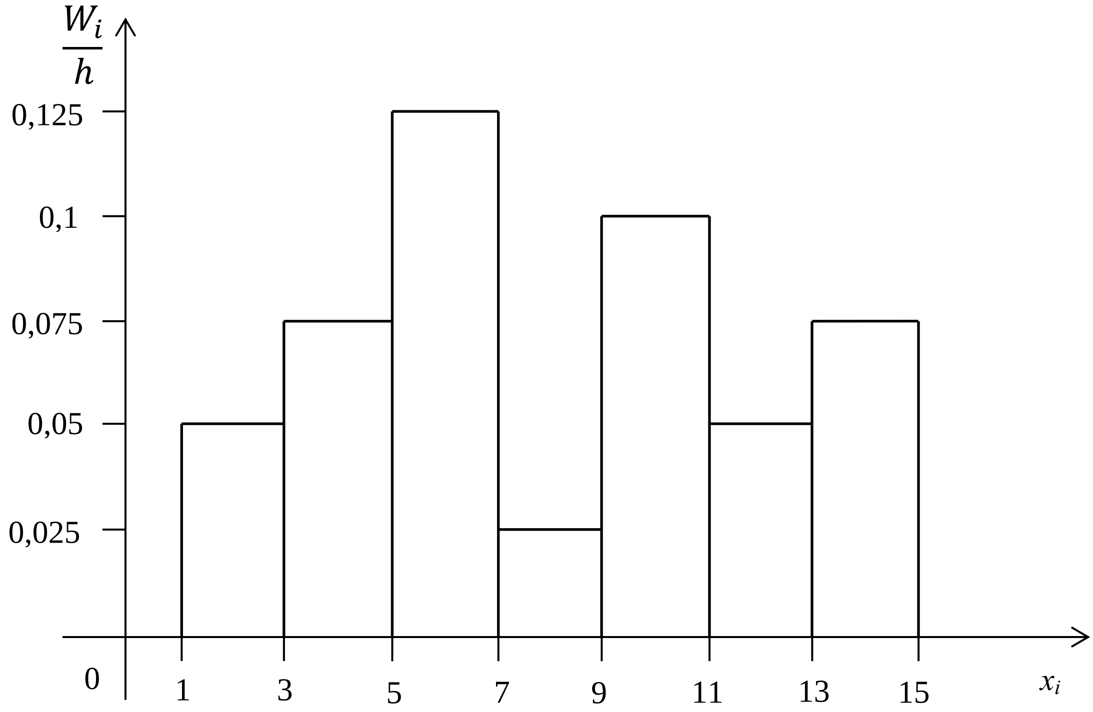

Рис. 7.3. Гістограма відносних частот

Запитання для самоконтролю

1. Що є предметом математичної статистики?
2. Вказати основні завдання математичної статистики.
3. Що називають генеральною та вибірковою сукупністю, об'ємом цих сукупностей?
4. Навести приклади генеральної сукупності та вибірки.
5. Способи відбору статистичного матеріалу (проілюструвати на прикладах).
6. Що називається варіантою, варіаційним рядом?
7. Що таке частота, відносна частота варіант?
8. Що називається емпіричною функцією розподілу?
9. Властивості емпіричної функція розподілу та її графік.
10. Статистичний розподіл частот вибірки.
11. Статистичний розподіл відносних частот вибірки.
12. Згрупований розподіл вибірки, загальна схема його побудови.
13. Що називають згрупованим розподілом накопичених частот вибірки; накопичених відносних частот вибірки?
14. Що називають згрупованим розподілом щільності частот; щільності відносних частот?

15. Що називається інтервальним статистичним розподілом вибірки? Які рекомендації до вибору числа класів інтервалів?

16. Що називається полігоном частот і відносних частот?

17. Що називається гістограмою частот і відносних частот?

Задачі для самостійного розв'язування

1. Для заданої вибірки із генеральної сукупності побудувати варіаційний ряд, скласти розподіли частот та відносних частот.

а) 7, 4, 4, 8, 12, 12, 12, 7, 8, 12, 8, 12, 4, 12, 12, 4, 12, 4,12, 12, 6, 9, 5, 3,6, 6, 9, 3, 5, 6, 9, 5, 6, 6, 9, 6, 9, 6, 6, 6.

б) 0,45; 0,21; 0,14; 0,15; 1,52; 0,1; 0,52; 1,59; 3,38; 2,25; 0,8; 1,26; 2,31; 0,84; 3,72; 2,11; 1,02; 4,2; 2,53; 0,78; 2,92; 0,71; 4,7; 3,02; 1,58; 4,12; 2,59; 0,88; 0,96; 1,76; 1,93; 4,9; 2,82; 1,14; 5,7; 1,21; 1,47; 3,52; 0,36; 0,64.

в) 4; 4,3; 5,68; 6,2; 5,64; 5,8; 4,25; 5,4; 5,3; 5,2; 4,55; 5,32; 6; 6,15; 4,56; 6,64; 6,5; 4,7; 6,8; 6,15; 5,6; 5,1; 4,2; 4,8; 6,9; 7; 4,9; 5; 5,25; 6,2.

г) 40,25; 40,29; 40,46; 40,33; 40,37; 40,27; 40,39; 40,34; 40,33; 40,35; 40,38; 40,32; 40,28; 40,41; 40,45; 40,39; 40,29;40,3; 40,44; 40,37; 40,41; 40,33; 40,35; 40,35; 40,35; 40, 40; 40, 40; 40,3; 40,28; 40,34; 40,45; 40,44.

2. Знайти емпіричну функцію заданого розподілу вибірки та побудувати її графік.

а)

x_i	2	5	7	8
n_i	1	3	2	4

б)

x_i	4	8	12	16
n_i	15	10	25	30

в)

x_i	2	3	5	6	6
n_i	6	4	5	5	5

г)

x_i	-2	-1	0	1	2
n_i	7	13	14	15	11

3. П'ятдесят абітурієнтів на вступних іспитах з інформатики отримали таку кількість балів:

12, 14, 19, 15, 14, 18, 13, 16, 17, 12, 20, 17, 15, 13, 17, 16, 20, 14, 14, 13, 17, 16, 15, 19, 16, 15, 18, 17, 15, 14, 16, 15, 15, 18, 15, 15, 19, 14, 16, 18, 18, 15, 15, 17, 15, 16, 16, 14, 14, 17.

Побудувати дискретний статистичний розподіл частот та відносних частот. Знайти емпіричну функцію розподілу та побудувати її графік.

4. 200 однотипних деталей x_i після шліфування були піддані контрольним вимірюванням, результати яких наведено в таблиці:

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

Побудувати полігон відносних частот, емпіричну функцію розподілу.

5. Побудувати полігон частот та відносних частот заданого розподілу вибірки.

а)

x_i	-2	-1	0	1
n_i	1	3	2	4

б)

x_i	2	3	5	6
n_i	15	10	25	30

в)

x_i	2	5	7	8	10
n_i	6	4	5	5	5

г)

x_i	-2	0	3	5	7
n_i	7	13	14	15	11

Побудувати гістограму частот та відносних частот.

6. У таблиці наведена вибірка оцінок за екзамен з дисципліни «Теорія ймовірностей та математична статистика» 100 студентів економічного факультету. Впорядкувати вибірку. Записати розподіл частот та відносних частот оцінок за екзамен; згрупований розподіл частот та відносних частот оцінок за екзамен; згрупований розподіл накопичених частот та накопичених відносних частот оцінок за екзамен; згрупований розподіл щільності частот та щільності відносних частот.

Вибірка оцінок за екзамен 100 студентів									
52	54	70	73	75	52	82	93	95	82
60	62	50	79	92	91	81	53	80	70
52	82	93	95	82	55	60	62	50	90
52	54	70	73	75	52	82	93	95	82
82	82	93	55	60	62	50	79	28	100
70	73	75	52	58	73	75	52	82	84
52	54	70	73	75	52	82	93	95	82
100	66	84	93	95	82	82	93	95	82
84	60	62	50	79	52	54	70	73	73
73	75	52	82	66	70	73	75	52	54

7. Із генеральної сукупності зроблено вибірку обсягом $n = 32$. Здобуто такі реалізації випадкової величини: 2,2; 7,1; 6,3; 3,9; 5,9; 5,6; 5,6; 4,7; 7,9; 3,2; 6,1; 5,5; 6,4; 6,0; 6,9; 4,7; 6,4; 6,9; 6,7; 7,9; 4,2; 6,7; 6,0; 9,2; 5,5; 6,5; 3,5; 4,9; 7,2; 4,9; 8,9; 5,7. Скласти інтервальний ряд і побудувати гістограму.

8. Було виміряно глибину виробу. Результати вимірювання, мм: 2,41; 2,62; 2,73; 2,52; 2,54; 2,41; 2,81; 2,53; 2,64; 2,61; 2,72; 2,45; 2,52; 2,1; 2,64; 2,52; 2,5; 2,33; 2,24; 2,4; 2,72; 2,51; 2,4; 2,61; 2,42; 2,43; 2,65; 2,54; 2,35; 2,54; 2,62; 2,9; 2,75; 2,24; 2,65; 2,45; 2,53; 2,32; 2,24; 2,55. Скласти інтервальний ряд і побудувати гістограму.

9. Для дослідження розподілу маси новонароджених x_i була зібрана інформація про 100 дітей. Ця інформація подана як інтервальний статистичний розподіл, що має такий вигляд у табличній формі:

$(x_i; x_{i+1})$	(1; 1,5)	(1,5; 2)	(2; 2,5)	(2,5; 3)	(3; 3,5)	(3,5; 4)	(4; 4,5)	(4,5; 5)
n_i	2	8	10	30	40	6	3	1

10. Побудувати емпіричну функцію розподілу, гістограму та полігон частот вибірки, поданої у вигляді таблиці частот:

а)

$[x_i; x_{i+1})$	[-3;-2)	[-2; -1)	[-1; 0)	[0;1)	[1;2)	[2;3)	[3;4)	[4;5)
n_i	3	10	15	24	25	13	7	3

б)

$[x_i; x_{i+1})$	[0,2; 2,2)	[2,2; 4,2)	[4,2; 6,2)	[6,2; 8,2)	[8,2; 10,2)
n_i	70	20	4	3	3

11. Із партії однотипних сталевих болтів, виготовлених заводом, була здійснена вибірка обсягом 200 шт. і результати вимірювання їх діаметрів x_i наведено у вигляді інтервального статистичного розподілу:

x_i, мм h = 2 мм	14,40–14,42	14,42–14,44	14,44–14,46	14,46–14,48	14,48–14,50	14,50–14,52	14,52–14,54	14,54–14,56	14,56–14,58	14,58–14,60	14,60–14,62	14,62–14,64
n_i	2	2	8	9	9	14	41	76	21	11	4	3

Побудувати гістограму частот і емпіричну функцію розподілу.

12. Відсоток виконання плану підприємства за рік та кількість підприємств, що виконують цей план, наведено у вигляді інтервального статистичного розподілу:

x_i,% h = 10%	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100	100–110	110–120
n_i	2	6	13	16	25	12	10	8	5	3	1

Побудувати гістограму частот і емпіричну функцію розподілу.

Розділ 8

ЧИСЛОВІ ХАРАКТЕРИСТИКИ СТАТИСТИЧНОГО МАТЕРІАЛУ

8.1. Числові характеристики вибірки

Як доповнення до табличних та графічних методів представлення даних, ще одним найважливішим засобом обробки даних є обчислення їх числових характеристик. Найважливіші з них: *вибіркове середнє, дисперсія, середньоквадратичне відхилення.*

Ці характеристики можуть бути обчислені за даними, що знаходяться у вибірці, або за даними, що входять у кінцеву генеральну сукупність.

Числові характеристики, обчислені за вибіркою або ті, що використовуються для опису даних вибірки, називають **статистиками**.

Вибірковим середнім статистичного матеріалу називають суму всіх значень вибірки, поділену на обсяг вибірки *n*:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{8.1}$$

Якщо маємо частотну таблицю, то вибіркове середнє вираховуємо за формулою:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i, \tag{8.2}$$

де n – об'єм вибірки, m – число різних варіант, $n_1,\ n_2,\ \dots,\ n_m$ – частоти варіант ($n = n_1 + n_2 + \cdots + n_m$), x_i – значення i-ої варіанти.

Вибіркове середнє для інтервального статистичного ряду знаходиться за формулою

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{s} n_i z_i, \tag{8.3}$$

де n – об'єм вибірки, s – кількість інтервалів, z_i – середина i-ого інтервалу, $n_1, n_2, \dots, n_s$ – число елементів вибірки, які потрапили в i-ий інтервал.

Вибіркове середнє є аналогом математичного сподівання і використовується дуже часто. Воно може набувати різних числових значень при різних вибірках однакового об'єму.

Тому можна розглядати розподіли (теоретичний та емпіричний) вибіркового середнього та числові характеристики цього розподілу (цей розподіл називають *вибірковим).*

Основні властивості вибіркового середнього

1. При множенні усіх варіант вибірки на однаковий множник вибіркове середнє також множиться на цей множник

$$\frac{1}{n}\sum_{i=1}^{m} n_i(cx_i) = \frac{c}{n}\sum_{i=1}^{m} n_i x_i = c\bar{x}.$$

2. Якщо додати (відняти) до всіх варіант вибірки однакове число, то вибіркове середнє зростає (зменшується) на це число

$$\frac{1}{n}\sum_{i=1}^{m} n_i(x_i \pm c) = \frac{1}{n}\sum_{i=1}^{m} n_i x_i \pm \frac{1}{n} c \sum_{i=1}^{m} n_i = \bar{x} \pm c.$$

Степеневим середнім вибірки називають середнє значення, яке знаходять за формулою

$$\overline{x_c} = \left(\sum_{i=1}^{m} \frac{n_i}{n} x_i^{\propto}\right)^{\frac{1}{\alpha}}. \qquad (8.4)$$

При $\alpha = 1$ одержимо вибіркове середнє.

При $\alpha = 2$ одержимо середньоквадратичне значення вибірки

$$\overline{x_2} = \left(\sum_{i=1}^{m} \frac{n_i}{n} x_i^2\right)^{\frac{1}{2}}.$$

При $\alpha = -1$ одержимо середнє гармонічне

$$\bar{x}_{-1} = \bar{x}_{\text{гарм}} = \frac{n}{\sum_{i=1}^{m} \frac{n_i}{x_i}}.$$

Середнім геометричним вибірки називають середнє значення, яке обчислюється за формулою

$$\bar{x}_{\Gamma} = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \ldots \cdot x_m^{n_m}}. \qquad (8.5)$$

Це середнє обчислюється лише за умови, що всі значення варіанти є додатні, тобто $x_i >, i = 1, 2, \ldots, m$.

Середнє геометричне застосовується у статистиці для визначення темпу зростання під час дослідження змін ознаки з часом.

Зауваження 1. Обрання того чи іншого середнього для характеристики розподілу пов'язане з якісним аналізом цього розподілу.

Зауваження 2. Крім вказаних степеневих середніх, у статистиці застосовують ще структурні середні, які не залежать від значень варіант, які розташовані на краях розподілу, а пов'язані з рядом частот.

До структурних середніх відносять моду та медіану.

Мода M_o – це елемент, який найчастіше трапляється у вибірці.

Для дискретних статистичних рядів

$$M_o = x_i, \text{ якщо } n_i = \max_i n_i$$

Для інтервальних статистичних рядів

$$M_o = x_i + h \cdot \frac{n_i - n_{i-1}}{2n_i - n_{i-1} - n_{i+1}}, \tag{8.6}$$

де x_i – початок інтервалу з найбільшою частотою, n_i – частота i-го інтервалу.

Медіаною M_e називають варіанту, яка ділить варіаційний ряд на дві частини, які рівні за числом варіант.

Для дискретних статистичних рядів

$$M_e = \begin{cases} x_m, & \text{при } n = 2m - 1, \\ \frac{x_m + x_{m+1}}{2}, & \text{при } n = 2m, \end{cases} \tag{8.7}$$

де x_m – середина варіаційного ряду.

Для інтервальних статистичних рядів

$$M_e = x_i + h \cdot \frac{\frac{n}{2} - \sum_{j=1}^{i-1} n_j}{n_i}, \tag{8.8}$$

де x_i – початок медіанного інтервалу, тобто такого, якому відповідає перша з нагромаджених частот, що перевищує половину всіх спостережень, h – довжина i-ого інтервалу, n_i – частота медіанного інтервалу.

Вибірковою дисперсією $D_в$ називають середнє значення квадратів відхилення варіант від вибіркового середнього з урахуванням відповідних частостей.

$$D_в = \frac{1}{n}\sum_{i=1}^{m} n_i\left(x_i - \bar{x}\right)^2. \quad (8.9)$$

Зауваження. Обчислення вибіркової дисперсії спрощується, якщо її знаходити за формулою:

$$D_в = \frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{m} n_i x_i\right)^2 = \frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - (\bar{x})^2. \quad (8.10)$$

Виправленою вибірковою дисперсією (варіансою) називають суму квадратів відхилень елементів від вибіркового середнього, поділену на $(n-1)$:

$$S^2 = \frac{n}{n-1}D_в = \frac{1}{n-1}\sum_{i=1}^{m} n_i\left(x_i - \bar{x}\right)^2. \quad (8.11)$$

Вибіркову дисперсію та виправлену вибіркову дисперсію для інтервального розподілу знаходять, відповідно, за формулами:

$$D_в = \frac{1}{n}\sum_{i=1}^{s} n_i\left(z_i - \bar{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{s} n_i z_i^2 - (\bar{x})^2, \quad (8.12)$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{s} n_i\left(z_i - \bar{x}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{s} n_i z_i^2 - (\bar{x})^2, \quad (8.13)$$

$$\text{або } S^2 = \frac{n}{n-1}D_в. \quad (8.14)$$

Вибірковим середньоквадратичним відхиленням називають квадратний корінь із вибіркової дисперсії:

$$\sigma_в = \sqrt{D_в}. \quad (8.15)$$

Стандарт – це арифметичний квадратний корінь із варіанси:

$$S = \sqrt{S^2}. \quad (8.16)$$

Розмах – це різниця між найбільшим і найменшим значенням варіаційного ряду (між крайніми елементами) і позначається

$$R = x_{max} - x_{min}. \quad (8.17)$$

Приклад 8.1. Обчислити вибіркове середнє, вибіркову дисперсію, вибіркове середньоквадратичне відхилення, варіанту, стандарт, моду, медіану та розмах вибірки, заданої статистичним розподілом частот:

x_i	-1	2	5	8	10
n_i	5	10	20	5	10

Розв'язання. Об'єм вибірки $n = 5 + 10 + 20 + 5 + 10 = 50.$

Знайдемо, відповідно, вибіркове середнє, вибіркову дисперсію, вибіркове середньоквадратичне відхилення, варіансу, стандарт, моду та медіану вибірки:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i = \frac{1}{50}\big(-1\cdot 5 + 2\cdot 10 + 5\cdot 20 + 8\cdot 5 + 10\cdot 10\big) = \frac{255}{50} = 5{,}1;$$

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - (\bar{x})^2 = \frac{1}{50}(1\cdot 5 + 4\cdot 10 + 25\cdot 20 + 64\cdot 5 + +100\cdot$$
$$\cdot 10) - 5{,}1^2 = \frac{1}{50}\cdot 1865 - 26{,}01 = 37{,}3 - 26{,}01 = 11{,}29;$$

$$\sigma_{\text{в}} = \sqrt{D_{\text{в}}} = \sqrt{11{,}29} \approx 3{,}36;$$

$$S^2 = \frac{n}{n-1}D_{\text{в}} = \frac{50}{49}\cdot 11{,}29 = 11{,}52;$$

$$S = \sqrt{S^2} = \sqrt{11{,}52} \approx 3{,}39;$$

$$M_o = 5;$$

$$M_e = 5;$$

$$R = x_{max} - x_{min} = 10 - (-1) = 11.$$

Приклад 8.2. Обчислити вибіркове середнє, вибіркову дисперсію, вибіркове середньоквадратичне відхилення, варіанту, стандарт, моду, медіану та розмах вибірки, заданої інтервальним розподілом:

інтервал	[2; 4)	[4; 6)	[6; 8)	[8; 10)	[10; 12)
n_i	2	8	35	40	15

Розв'язання. Об'єм вибірки $n = 2 + 8 + 35 + 40 + 15 = 100.$

Перетворимо цей інтервальний статистичний розподіл вибірки на точковий, знайшовши середини частинних інтервалів:

$$z_1 = \frac{x_1+x_2}{2} = \frac{2+4}{2} = 3; \qquad z_2 = \frac{x_2+x_3}{2} = \frac{4+6}{2} = 5;$$

$$z_3 = \frac{x_3+x_4}{2} = \frac{6+8}{2} = 7; \qquad z_4 = \frac{x_4+x_5}{2} = \frac{8+10}{2} = 9;$$

$$z_5 = \frac{x_5+x_6}{2} = \frac{10+12}{2} = 11.$$

Отже, ми отримали такий статистичний розподіл вибірки:

z_i	3	5	7	9	11
n_i	2	8	35	40	15

Знайдемо, відповідно, вибіркове середнє, вибіркову дисперсію, вибіркове середньоквадратичне відхилення, варіансу, стандарт, моду та медіану вибірки:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{s} n_i z_i = \frac{1}{100}(3\cdot 2 + 5\cdot 8 + 7\cdot 35 + 9\cdot 40 + 11\cdot 15) =$$
$$= \frac{1}{100}\cdot 816 = 8{,}16;$$

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{s} n_i z_i^2 - (\bar{x})^2 = \frac{1}{100}(9\cdot 2 + 25\cdot 8 + 49\cdot 35 + 81\cdot 40 + +121\cdot$$
$$15) - -8{,}16^2 = \frac{1}{100}\cdot 6988 - 66{,}5856 = 69{,}88 - 66{,}59 = 3{,}29;$$

$$\sigma_{\text{в}} = \sqrt{D_{\text{в}}} = \sqrt{3{,}29} \approx 1{,}81;$$

$$S^2 = \frac{n}{n-1}D_{\text{в}} = \frac{100}{99}\cdot 3{,}29 = 3{,}32;$$

$$S = \sqrt{S^2} = \sqrt{3{,}32} \approx 1{,}82;$$

$$M_o = 8 + 2\cdot\frac{40-35}{2\cdot 40-35-15} = 8 + 2\cdot\frac{5}{30} = 8{,}33;$$

$$M_e = 8 + 2\cdot\frac{\frac{100}{2}-(8+2+35)}{40} = 8 + 2\cdot\frac{50-45}{40} = 8{,}25;$$

$$R = x_{max} - x_{min} = 12 - 2 = 10.$$

8.2. Метод добутків обчислення вибіркового середнього та вибіркової дисперсії

Нехай вибірку задано у вигляді розподілу рівновіддалених варіант і відповідних їм частот. У цьому разі вибіркове середнє та вибіркову дисперсію зручно знаходити методом добутків.

Алгоритм методу добутків

1. У перший стовпчик таблиці записують рівновіддалені варіанти x_i вибірки, розміщуючи їх у порядку зростання.
2. У другий стовпчик таблиці записують відповідні частоти n_i відповідних варіант. Суму елементів цього стовпчика (об'єм вибірки n) записують в останню його клітинку.
3. Третій стовпчик містить умовні варіанти u_i вибірки. Для знаходження умовних варіант вибірки треба:

а) значення варіанти вибірки, яка знаходиться приблизно посередині варіаційного ряду С обрати за умовний нуль;

б) знайти різницю h між будь-якими двома сусідніми варіантами;

в) обчислити умовні варіанти $u_i = \frac{x_i - C}{h}$. Зауважимо, що умовні варіанти завжди будуть цілими числами.

4. У четвертий стовпчик записують добутки частот та відповідних їм умовних варіант $n_i \cdot u_i$. Суму елементів стовпчика $\sum_{i=1}^{m} n_i \cdot u_i$ записують в його останню клітинку.
5. Знаходять добутки частот та квадратів умовних варіант $n_i \cdot u_i^2$ і записують їх у п'ятий стовпчик. Суму елементів стовпчика $\sum_{i=1}^{m} n_i \cdot u_i^2$ записують в його останню клітинку.
6. Знаходять добутки частот та квадратів умовних варіант, збільшених на одиницю, $n_i \cdot (u_i + 1)^2$ і записують їх у шостий стовпчик. Суму елементів стовпчика $\sum_{i=1}^{m} n_i \cdot (u_i + 1)^2$ записують в його останню клітинку.
7. Перевіряють обчислення так: сума елементів шостого стовпчика повинна задовольняти тотожність:

$$\sum_{i=1}^{m} n_i \cdot (u_i + 1)^2 = \sum_{i=1}^{m} n_i \cdot u_i^2 + 2 \cdot \sum_{i=1}^{m} n_i \cdot u_i + n.$$

8. Обчислюють умовні моменти за формулами:

$$M_1^* = \frac{1}{n} \sum_{i=1}^{m} n_i \cdot u_i; \qquad M_2^* = \frac{1}{n} \sum_{i=1}^{m} n_i \cdot u_i^2.$$

9. Обчислюють вибіркове середнє та вибіркову дисперсію за формулами:

$$\bar{x} = M_1^* \cdot h + C, \qquad D_{\text{в}} = \left(M_2^* - (M_1^*)^2 \right) \cdot h^2.$$

Зауваження. Якщо варіанти вибірки не є рівновіддаленими, то інтервал, у якому розміщені всі варіанти вибірки, ділять на кілька однакових за довжиною частинних інтервалів, кожен з яких має містити не менше 8–10 варіант. Потім знаходять середини частинних інтервалів, які й утворюють послідовність рівновіддалених варіант. За частоту кожної середини частинного інтервалу беруть суму частот варіант, які потрапили у відповідний частинний інтервал.

Під час обчислення вибіркової дисперсії з метою зменшення похибки, зумовленої групуванням (особливо при малій кількості інтервалів), роблять поправку Шеппарда, а саме: обчислюють дисперсію за формулою:

$$D'_{\text{в}} = D_{\text{в}} - \frac{1}{12} \cdot h^2.$$

Приклад 8.3. Знайти методом добутків вибіркове середнє та вибіркову дисперсію для статистичного розподілу вибірки об’єму $n = 100$.

x_i	2	8	9	13	15	18	20	21	24	27
n_i	8	1	15	13	9	3	4	10	22	15

Розв’язання. Розіб’ємо інтервал $[2;\ 27]$ на п’ять частинних інтервалів довжиною $h = 5$:

$$[2;\ 7], \quad (7;\ 12], \quad (12;\ 17], \quad (17;\ 22], \quad (22;\ 27].$$

Новими варіантами будуть середини цих частинних інтервалів:

$$y_1 = \frac{2+7}{2} = 4{,}5; \qquad y_2 = \frac{7+12}{2} = 9{,}5; \qquad y_3 = \frac{12+17}{2} = 14{,}5;$$

$$y_4 = \frac{17+22}{2} = 19{,}5; \qquad y_5 = \frac{22+27}{2} = 24{,}5.$$

За частоти m_i варіант y_i візьмемо суму частот варіант, які потрапили у відповідний i-ий інтервал:

$$m_1 = 8; \qquad m_2 = 1 + 15 = 16; \qquad m_3 = 13 + 9 = 22;$$

$$m_4 = 3 + 4 + 10 = 17; \qquad m_5 = 22 + 15 = 37.$$

Отже, ми отримали такий статистичний розподіл рівновіддалених варіант:

y_i	4,5	9,5	14,5	19,5	24,5
m_i	8	16	22	17	37

Складемо розрахункову таблицю:

y_i	m_i	u_i	$m_i u_i$	$m_i u_i^2$	$m_i(u_i+1)^2$
4,5	8	-2	-16	32	8
9,5	16	-1	-16	16	0
14,5	22	0	0	0	22
19,5	17	1	17	17	68
24,5	37	2	74	148	333
	$n=100$		$\sum_{i=1}^{5} m_i u_i = 59$	$\sum_{i=1}^{5} m_i u_i^2 = 213$	$\sum_{i=1}^{5} m_i(u_i+1)^2 = 431$

Перевірка:

$$\sum_{i=1}^{5} m_i(u_i+1)^2 = 431;$$

$$\sum_{i=1}^{5} m_i u_i^2 + 2\cdot\sum_{i=1}^{5} m_i u_i + n = 213 + 2\cdot 59 + 100 = 431.$$

Обчислимо умовні моменти першого і другого порядків:

$$M_1^* = \frac{1}{n}\sum_{i=1}^{m} n_i \cdot u_i = \frac{59}{100} = 0{,}59; \quad M_2^* = \frac{1}{n}\sum_{i=1}^{m} n_i \cdot u_i^2 = \frac{213}{100} = 2{,}13.$$

Методом добутків знайдемо вибіркове середнє й дисперсію, враховуючи, що хибний нуль $\text{C} = 14{,}5$:

$$\bar{x} = M_1^* \cdot h + C = 0{,}59 \cdot 5 + 14{,}5 = 17{,}45,$$

$$D_{\text{в}} = \left(M_2^* – (M_1^*)^2 \right) \cdot h^2 = \left(2{,}13 – 0{,}59^2\right) \cdot 5^2 = 44{,}5475.$$

Оскільки кількість частинних інтервалів мала, то скористаємося поправкою Шеппарда:

$$D'_{\text{в}} = D_{\text{в}} – \frac{1}{12} \cdot h^2 = 44{,}5475 – \frac{1}{12} \cdot 5^2 \approx 42{,}46.$$

8.3. Метод сум обчислення вибіркового середнього та вибіркової дисперсії

Нехай вибірку задано у вигляді розподілу рівновіддалених варіант і відповідних їм частот. Тоді вибіркове середнє та вибіркову дисперсію зручно знаходити методом сум.

Алгоритм методу сум

1. У перший стовпчик таблиці записують рівновіддалені варіанти x_i вибірки, розміщуючи їх у порядку зростання.
2. У другий стовпчик таблиці записують відповідні частоти n_i варіант. Суму елементів цього стовпчика (об'єм вибірки n) записують в його останню клітинку.
3. Третій стовпчик містить умовні варіанти u_i вибірки. Для знаходження умовних варіант вибірки треба:

   а) значення варіанти вибірки з найбільшою частотою C обрати за умовний нуль, це значення називають модою;

   б) у порожніх клітинках третього стовпця, які розміщені над нулем (крім верхньої), зверху вниз записують послідовно нагромаджені частоти. Додавши всі нагромаджені частоти зверху до нуля, отримують число b_1, яке записують у верхній клітинці;

   в) у порожніх клітинках третього стовпця, які розміщені під нулем (крім нижньої), знизу вверх записують послідовно нагромаджені частоти. Додавши всі нагромаджені частоти знизу до нуля, отримують число a_1, яке записують у його нижню клітинку.
4. У четвертому стовпці знову записують хибний нуль, над і під ним записують ще по одному нулю. Аналогічно до третього, заповнюють четвертий стовбець, причому нагромаджуються частоти третього стовпця. Суму нагромаджених частот, розміщених над нулем, позначають b_2 і записують у верхній клітинці четвертого стовпця. Суму нагромаджених частот, розміщених під нулем, позначають a_2 і записують у нижню клітинку четвертого стовпця.
5. Обчислюють d_1, s_1, s_2:

$$d_1 = a_1 - b_1; \qquad s_1 = a_1 + b_1; \qquad s_2 = a_2 + b_2.$$

6. Обчислюють умовні моменти за формулами:

$$M_1^* = \frac{d_1}{n}; \qquad M_2^* = \frac{s_1 + 2s_2}{n}.$$

7. Обчислюють вибіркове середнє та вибіркову дисперсію за формулами:

$$\bar{x} = M_1^* \cdot h + C,$$

$$D_{\text{в}} = \left(M_2^* - (M_1^*)^2 \right) \cdot h^2.$$

Приклад 8.4. Знайти методом сум вибіркове середнє та вибіркову дисперсію для статистичного розподілу вибірки об’єму $n = 100$.

x_i	52	58	64	70	76	82	88	94	100	106
n_i	4	8	10	13	18	21	14	9	2	1

Розв’язання. Складемо розрахункову таблицю

1	2	3	4
x_i	n_i	$b_1 = 126$	$b_2 = 131$
52	4	4	4
58	8	12	16
64	10	22	38
70	13	35	73
76	18	53	0
82	21	0	0
88	14	26	0
94	9	12	16
100	2	3	4
106	1	1	1
		$a_1 = 42$	$a_2 = 21$

Знайдемо d_1, s_1, s_2:

$$d_1 = a_1 - b_1 = 42 - 126 = -84;$$

$$s_1 = a_1 + b_1 = 42 + 126 = 168;$$

$$s_2 = a_2 + b_2 = 21 + 131 = 152.$$

Обчислимо умовні моменти першого і другого порядку:

$$M_1^* = \frac{d_1}{n} = \frac{-84}{100} =- 0,84;$$

$$M_2^* = \frac{s_1+2s_2}{n} = \frac{168+2\cdot152}{100} = 4,72.$$

Знайдемо вибіркове середнє та дисперсію, врахувавши, що хибний нуль $\mathrm{C} = 82$, а крок (відстань між двома сусідніми варіантами) $h = 6$, отримаємо:

$$\bar{x} = M_1^* \cdot h + C =- 0,84 \cdot 6 + 82 = 76,96;$$

$$D_{\text{в}} = \left(M_2^* - (M_1^*)^2\right) \cdot h^2 = \left(4,72 - \left(- 0,84\right)^2\right) \cdot 6^2 = 144,5184.$$

Запитання для самоконтролю

1. Що відноситься до числових характеристики вибірки?
2. Які числові характеристики називаються статистичними?
3. Як визначається вибіркова середня для статистичного розподілу вибірки?
4. Як визначається вибіркова середня для інтервального розподілу вибірки?
5. Які властивості має вибіркова середня?
6. Що таке степеневе середнє; середнє геометричне?
7. Як визначається дисперсія для статистичного розподілу вибірки?
8. Як визначається дисперсія для інтервального розподілу вибірки?
9. Як визначається середньоквадратичне відхилення?
10. Що таке мода і як її знайти?
11. Що таке медіана і як її знайти?
12. Що таке варіанса, стандарт, розмах і як їх обчислити?
13. Метод добутків для знаходження вибіркового середнього та вибіркової дисперсії, його алгоритм.
14. Метод сум для знаходження вибіркового середнього та вибіркової дисперсії, його алгоритм.
15. У яких випадках використовують виправлену вибіркову дисперсію і як вона пов’язана із вибірковою дисперсією?
16. В яких випадках обчислюють характеристики вибірки методом добутків; методом сум?

Задачі для самостійного розв'язування

1. Знайти вибіркове середнє, дисперсію, моду і медіану для вибірки. Побудувати емпіричну функцію розподілу та полігон частот.

а)

x_i	-2	5	7	8
n_i	12	12	16	20

б)

x_i	4	8	12	16
n_i	5	10	5	20

в)

x_i	1	2	3	4	5
n_i	5	3	2	8	2

г)

x_i	-2	-1	0	1	2
n_i	7	15	23	31	39

г)

x_i	0	1	2	3	4	5	6	7	8	9	10	11
n_i	45	41	34	54	39	43	41	33	37	41	47	39

2. Через кожну годину вимірювалась напруга в електромережі. Результати вимірювання напруги у вольтах наведені у вигляді статистичного ряду:

222, 219, 224, 220, 218, 217, 221, 220, 215, 218, 223, 225, 220, 226, 221, 216, 211, 219, 220, 221, 222, 218, 221, 219.

Побудувати статистичний розподіл, полігон частот. Обчислити числові характеристики вибірки.

3. Знайти вибіркове середнє, дисперсію, середньоквадратичне відхилення, варіансу, стандарт, моду, медіану і розмах для вибірки, заданої інтервальним розподілом. Побудувати гістограму та полігон частот.

а)

$[x_i; x_{i+1})$	[4; 6)	[6; 8)	[8; 10)	[10; 12)	[12; 14)
n_i	5	3	2	8	2

б)

$[x_i; x_{i+1})$	[3; 5)	[5; 7)	[7; 11)	[11; 13)	[13; 15)
n_i	4	3	6	3	4

в)

$[x_i; x_{i+1})$	[-3; -1)	[-1;0)	[0; 1)	[1; 2)	[2; 3)	[3; 5)
n_i	13	15	24	25	13	10

г)

$[x_i; x_{i+1})$	[0; 1)	[1;2)	[2; 3)	[3; 4)	[4; 5)	[5; 6)	[6; 7)	[7; 8)	[8; 9)	[9; 10)
n_i	35	16	15	17	17	19	11	16	30	24

4. Рівень води x_i в річці щодо номіналу вимірювався протягом 50-ти років навесні. Результати вимірювання наведено у формі інтервального статистичного розподілу.

x_i, см $h = 24$	0–24	24–48	48–72	72–96	96–120	120–144	144–168	168–192	192–216
n_i	0	2	4	6	12	16	6	3	1

Побудувати гістограму частот. Обчислити числові характеристики вибірки.

5. Методом добутків знайти вибіркове середнє та вибіркову дисперсію для статистичного розподілу вибірки.

а)

x_i	12	14	16	18	20	22
n_i	5	15	50	16	10	4

б)

x_i	65	70	75	80	85
n_i	2	5	25	15	3

6. Методом сум знайти вибіркове середнє та вибіркову дисперсію для статистичного розподілу вибірки.

а)

x_i	40	46	52	68	74	80
n_i	10	14	12	5	11	8

б)

x_i	20	40	60	80	100
n_i	5	10	15	20	10

7. Середню температуру повітря у червні в Києві та Львові вимірювали протягом 40 років. Дані вимірювання наведено в таблиці (X – середня температура повітря у червні в Києві; Y– середня температура повітря у червні у Львові):

X	Y	X	Y	X	Y	X	Y
12	10,8	14	10	16	13	18	14
12	11,3	14	12	15,9	15	18	14
12	12	13,9	12,4	16	16	18,1	14,8
12,8	11	15	11	16,9	12,9	18,4	17,8
12	12	14,9	13	17,2	13,9	19,2	15
13,8	9,8	14,9	14,2	16,9	14,8	19,3	16
13,1	11,5	14,2	13,8	16,9	15	20	17
13	13	15	16	17	16	20	17,7
13,9	10,1	15,5	13,9	16,8	17	14	14,8
14,2	10	15,9	14,7	17,5	16	14	15,2

Знайти числові характеристики вибірки X та Y. Зробити висновки про те, як змінювалася середня температура повітря у червні в Києві та Львові.

8. Записати цю вибірку у вигляді варіаційного ряду; скласти статистичний ряд частот; скласти статистичний ряд відносних частот; згрупований розподіл частот та відносних частот; згрупований розподіл накопичених частот та накопичених відносних частот. Знайти емпіричну функцію розподілу та побудувати її графік. Побудувати гістограму та полігон частот та відносних частот. Обчислити вибіркові числові характеристики (вибіркове середнє, варіансу, стандарт, моду, медіану, розмах).

2, 5, 8, 6, 4, 5, 8, 6, 8, 4, 8, 5, 8, 10, 2, 5, 8, 6, 8, 6, 8, 4, 8, 5, 8, 6, 8, 4, 8, 5, 8, 10, 5, 8, 6, 8, 6, 5, 8, 6, 8, 4, 8, 5, 8, 10, 2, 5, 8,10, 4, 8, 5, 8, 10, 5, 8, 6, 8, 6, 5, 8, 6, 8, 4, 8, 5, 8, 10, 2, 5, 8,10, 2, 5, 8, 6, 4, 5, 8, 6, 8, 4, 8, 5, 8, 10, 2, 5, 8, 6, 8, 6, 8, 4, 8, 5, 8, 6, 8.

Розділ 9

СТАТИСТИЧНІ ОЦІНКИ ПАРАМЕТРІВ РОЗПОДІЛУ

9.1. Точкові оцінки

Статистичною оцінкою θ^* невідомого параметра θ теоретичного розподілу називають функцію $f(X_1, X_2, \dots, X_n)$, де $X_1,\ X_2,\ \dots\ ,\ X_n$ – спостережувальні випадкові величини.

Точковою оцінкою називають статистичну оцінку, яка визначається одним числом $\theta^* = f(x_1, x_2, \dots, x_n)$, де $x_1, x_2, \dots, x_n$ – результати n спостережень над кількісною ознакою випадкової величини *X*.

Наприклад, вибіркова середня $\bar{x}$, вибіркова дисперсія $D_{\text{в}}$ та вибіркове середньоквадратичне $\sigma_{\text{в}}$ – точкові оцінки відповідних числових характеристик генеральної сукупності.

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

Якщо об'єм вибірки доволі великий, то точкові оцінки задовольняють практичні потреби точності.

Щоб статистичні оцінки давали найкращі наближення, вони повинні задовольняти певним вимогам. Розглянемо ці вимоги.

Нехай θ^* – статистична оцінка невідомого параметра θ теоретичного розподілу. Припустимо, що за вибіркою об'єму n знайдена оцінка θ_1^*. Для інших вибірок того ж об'єму одержано оцінки $\theta_2^*,\ \theta_3^*,\ \dots,\ \theta_m^*$. Оцінку θ^* можна розглядати як випадкову величину, а числа $\theta_1^*,\ \theta_2^*,\ \dots,\ \theta_m^*$ – як її можливі значення.

Якщо θ_k^* $(k = 1, 2, \dots, m)$ будуть більші від значення θ, то оцінка θ^* дає наближене значення θ з надлишком. У цьому випадку математичне сподівання випадкової величини θ^* є більшим від θ $(\mathrm{M}(\theta^*) > \theta)$.

Якщо θ^* дає оцінку з недостачею, математичне сподівання випадкової величини θ^* є меншим від θ $(\mathrm{M}(\theta^*) < \theta)$.

Відтак використання статистичної оцінки, математичне сподівання якої не дорівнює параметру θ, спричиняє систематичні похибки. Вимога $\mathrm{M}(\theta^*) = \theta$ застерігає від таких похибок.

Незміщеною називають точкову оцінку θ^*, математичне сподівання якої дорівнює оцінюваному параметру θ для будь-якого обсягу вибірки:

$$\mathrm{M}(\theta^*) = \theta.$$

Зміщеною називають точкову оцінку θ^*, математичне сподівання якої відмінне від оцінюваного параметра θ.

Якщо $\mathrm{M}(\theta^*) = \theta + b(\theta)$, то $b(\theta)$ називають **зсувом оцінки $\boldsymbol{\theta}^*$**.

Вимога про незміщеність оцінки θ^* є недостатньою, оскільки можливі значення θ^* можуть бути дуже розсіяні від середнього значення, дисперсія $D(\theta^*)$ може бути великою. Тоді, знайдена за даними однієї вибірки оцінка θ_k^* може значно відрізнятися від середнього значення θ^*, а отже, і від параметра θ.

Якщо дисперсія $\mathrm{D}(\theta^*)$ буде незначною, то можливість допустити помилку виключена. Тому для статистичної оцінки виникає вимога про її ефективність.

Ефективною називають таку оцінку θ^*, яка при заданому об'ємі n має найменшу можливу дисперсію.

Обґрунтованою називають статистичну оцінку, яка при $n \to \infty$ прямує за ймовірністю до оцінюваного параметра.

Послідовність оцінок θ_k^* $(k = 1, 2, \ldots, m)$ параметра θ називається ***спроможною***, якщо $\theta_k^* \xrightarrow{P} \theta, n \to \infty$.

Послідовність оцінок θ_k^* $(k = 1, 2, \ldots, m)$ параметра θ називається **сильно спроможною**, якщо $\theta_k^* \to \theta$ з ймовірністю одиниця при $n \to \infty$.

Незміщеною оцінкою генерального середнього (математичного сподівання) є вибіркове середнє

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i,$$

де n_i – частота варіанти x_i; x_i – варіанта вибірки; $n = n_1 + n_2 + \cdots + n_m$ – об'єм вибірки.

Зауваження 1. Якщо варіанти x_i вибірки є дуже великими або дуже малими (близькими до нуля) числами, то для спрощення розрахунків доцільно відняти (у разі великих від'ємних чисел – додати) від кожної варіанти одне й те саме число С (як С можна вибрати будь – яке число, розміщене приблизно посередині варіаційного ряду), потім поділити (у разі близьких до нуля чисел – помножити) на одне й те саме число b (як b можна вибрати найбільше спільне кратне), тобто перейти до умовних варіант

$$u_i = \frac{x_i \pm C}{b} \qquad (u_i = x_i b). \tag{9.1}$$

Тоді

$$\bar{x} = \frac{b}{n}\sum_{i=1}^{m} n_i u_i \mp C = b\bar{u}_b \mp C \quad \left(\bar{x} = \frac{\sum_{i=1}^{m} n_i u_i}{bn} = \frac{\bar{u}_b}{b}\right). \tag{9.2}$$

Зміщеною оцінкою генеральної дисперсії є вибіркова дисперсія

$$D_в = \frac{1}{n}\sum_{i=1}^{m} n_i \left(x_i - \bar{x}\right)^2.$$

Ця оцінка зміщена, оскільки

$$M(D_в) = \frac{n-1}{n} D_Г.$$

Для обчислення вибіркової дисперсії можна скористатися зручнішою формулою

$$D_в = \frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{m} n_i x_i\right)^2 = \overline{x^2} - (\bar{x})^2.$$

Зауваження 2. Якщо варіанти x_i вибірки є дуже великими або дуже малими (близькими до нуля) числами, то для спрощення розрахунків доцільно відняти (у разі великих від'ємних чисел – додати) від кожної варіанти одне й те саме число C (як C можна вибрати будь-яке число, розміщене приблизно посередині варіаційного ряду), потім поділити (у разі близьких до нуля чисел – помножити) на одне й те саме число b (як b можна вибрати найбільше спільне кратне), тобто перейти до умовних варіант

$$u_i = \frac{x_i \pm C}{b} \qquad (u_i = x_i b).$$

Тоді

$$D_в = b^2\left(\frac{1}{n}\sum_{i=1}^{m} n_i u_i^2 - \left(\frac{1}{n}\sum_{i=1}^{m} n_i u_i\right)^2\right) = b^2\left(\overline{u^2} - (\bar{u})^2\right). \tag{9.3}$$

$$\left(D_в = \left(\frac{1}{b^2 n}\sum_{i=1}^{m} n_i u_i^2 - \left(\frac{1}{bn}\sum_{i=1}^{m} n_i u_i\right)^2\right) = \frac{\left(\overline{u^2} - (\bar{u})^2\right)}{b^2}\right). \tag{9.4}$$

Незміщеною оцінкою генеральної дисперсії є виправлена вибіркова дисперсія

$$S^2 = \frac{n}{n-1} D_{\text{в}}.$$

Поправку $\frac{n}{n-1}$ називають ***поправкою Бесселя***.

Приклад 9.1. З генеральної сукупності отримано деяку вибірку об'єму $n = 100$:

x_i	0,002	0,004	0,006	0,008	0,01	0,012	0,014
n_i	7	29	35	12	9	5	3

Знайти зміщену оцінку генеральної дисперсії.

Розв'язання. Зміщеною оцінкою генеральної дисперсії D_{Γ} є вибіркова дисперсія $D_{\text{в}}$. Оскільки варіанти вибірки є малими, близькими до нуля числами, перейдемо до умовних варіант

$$u_i = x_i \cdot 1000$$

(як b вибрано число 1000, оскільки в такому разі ми отримаємо цілі числа):

$$u_1 = x_1 \cdot 1000 = 0{,}002 \cdot 1000 = 2;$$
$$u_2 = x_2 \cdot 1000 = 0{,}004 \cdot 1000 = 4;$$
$$u_3 = x_3 \cdot 1000 = 0{,}006 \cdot 1000 = 6;$$
$$u_4 = x_4 \cdot 1000 = 0{,}008 \cdot 1000 = 8;$$
$$u_5 = x_5 \cdot 1000 = 0{,}01 \cdot 1000 = 10;$$
$$u_6 = x_6 \cdot 1000 = 0{,}012 \cdot 1000 = 12;$$
$$u_7 = x_7 \cdot 1000 = 0{,}014 \cdot 1000 = 14.$$

Тепер можна знайти вибіркову дисперсію:

$$D_{\text{в}} = \left(\frac{1}{b^2 n} \sum_{i=1}^{7} n_i u_i^2 - \left(\frac{1}{bn} \sum_{i=1}^{7} n_i u_i\right)^2\right) =$$

$$= \frac{7 \cdot 2^2 + 29 \cdot 4^2 + 35 \cdot 6^2 + 12 \cdot 8^2 + 9 \cdot 10^2 + 5 \cdot 12^2 + 3 \cdot 14^2}{1000^2 \cdot 100} -$$

$$-\left[\frac{7\cdot 2+29\cdot 4+35\cdot 6+12\cdot 8+9\cdot 10+5\cdot 12+3\cdot 14}{1000\cdot 100}\right]^2=$$

$$=\frac{28+464+1260+768+900+720+588}{100000000}-$$

$$-\left[\frac{14+116+210+96+90+60+42}{100000}\right]^2=$$

$$=4728\cdot 10^{-8}-394384\cdot 10^{-10}=78416\cdot 10^{-10}=0{,}0000078416.$$

Приклад 9.2. У результаті статистичних досліджень випадкової величини X отримано таку вибірку: 47, 45, 46, 46, 46, 45, 47, 44, 46, 45, 45, 46, 46, 44, 46, 48, 46, 46, 45, 46, 44, 46, 45, 47, 46, 46, 47, 46, 46, 48, 44, 46, 45, 46, 45, 46, 44, 47, 46, 46, 45, 47, 48, 44, 46, 46, 45, 46, 47, 45.

Знайти незміщені оцінки генерального середнього та генеральної дисперсії.

Розв'язання. Об'єм вибірки $n=50$. Побудуємо статистичний розподіл вибірки:

x_i	44	45	46	47	48
n_i	6	11	23	7	3

Перевірка: $n = 6 + 11 + 23 + 7 + 3 = 50.$

Незміщеною оцінкою генерального середнього є вибіркове середнє.

Оскільки варіанти вибірки є великими числами, перейдемо до умовних варіант

$$u_i=x_i-46.$$

Як С вибрано число 46:

$$u_1=x_1-46=44-46=-2; \qquad u_2=x_2-46=45-46=-1;$$

$$u_3=x_3-46=46-46=0; \qquad u_4=x_4-46=47-46=1;$$

$$u_5=x_5-46=48-46=5.$$

Тепер можна знайти вибіркове середнє:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i u_i + C = \frac{-2\cdot 6+(-1)\cdot 11+0\cdot 23+1\cdot 7+2\cdot 3}{50} + 46 =$$

$$= \frac{-10}{50} + 46 = 45{,}8.$$

Щоб знайти незміщену оцінку генеральної дисперсії (виправлену вибіркову дисперсію), визначимо вибіркову дисперсію й помножимо її на поправку Бесселя:

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{m} n_i x_i\right)^2 =$$

$$= \frac{(-2)^2\cdot 6 + (-1)^2\cdot 11 + 0^2\cdot 23 + 1^2\cdot 7 + 2^2\cdot 3}{50} -$$

$$- \left[\frac{-2\cdot 6 + (-1)\cdot 11 + 0\cdot 23 + 1\cdot 7 + 2\cdot 3}{50}\right]^2 =$$

$$= \frac{54}{50} - \left[\frac{-10}{50}\right]^2 = 1{,}04.$$

$$S^2 = \frac{n}{n-1} D_{\text{в}} = \frac{50}{49}\cdot 1{,}04 \approx 1{,}06.$$

Отже, незміщеними оцінками генерального середнього та генеральної дисперсії є

$$\bar{x}_{\Gamma}^{*} = 45{,}8; \qquad \bar{D}_{\Gamma}^{*} \approx 1{,}06.$$

9.2. Методи визначення точкових статистичних оцінок

9.2.1. Метод моментів

Методом моментів знаходження точкових оцінок називають метод, за яким для обчислення невідомих параметрів заданого розподілу прирівнюють теоретичні та емпіричні моменти.

Якщо розподіл визначається одним параметром, то для знаходження оцінки прирівнюють математичне сподівання до вибіркового середнього

$$M(X) = \bar{x},$$

а потім із цього рівняння визначають шукану точкову оцінку невідомого параметра.

Якщо розподіл визначається двома параметрами, то їх точкові оцінки знаходять із системи рівнянь:

$$\begin{cases} M(X) = \bar{x}; \\ D(X) = D_{\text{в}}. \end{cases} \tag{9.5}$$

Лівими частинами цих рівнянь є математичне сподівання та дисперсія, які дорівнюють відповідно вибірковому середнього та вибірковій дисперсії.

Приклад 9.3. Випадкова величина *X* (зріст дорослої людини) розподілена за нормальним розподілом з параметрами *a, σ.* У результаті статистичних досліджень отримано такий статистичний розподіл зросту дорослих людей для $n = 1000$ осіб:

Зріст, см	(145; 155]	(155; 165]	(165; 175]	(175; 185]	(185; 195]	(195; 205]	(205; 215]
Кількість осіб	24	112	263	322	202	66	11

Знайти методом моментів точкову оцінку невідомих параметрів *a, σ* нормального розподілу.

Розв'язання. Перетворимо інтервальний статистичний розподіл на точковий, вибравши як варіанти середини частинних інтервалів:

Зріст, см	150	160	170	180	190	200	210
Кількість осіб	24	112	263	322	202	66	11

Оскільки нормальний закон розподілу залежить від двох параметрів, треба знайти вибіркове середнє та вибіркову дисперсію, а потім прирівняти їх відповідно до математичного сподівання та дисперсії.

Перейдемо до умовних варіант

$$u_i = \frac{x_i - 180}{10}$$

(x_i – зріст людини):

$$u_1 = \frac{150-180}{10} = -3; \quad u_2 = \frac{160-180}{10} = -2; \quad u_3 = \frac{170-180}{10} = -1;$$

$$u_4 = \frac{180-180}{10} = 0; \quad u_5 = \frac{190-180}{10} 1; \quad u_6 = \frac{200-180}{10} = 2;$$

$$u_7 = \frac{210-180}{10} = 3.$$

Знайдемо вибіркове середнє та вибіркову дисперсію:

$$\bar{x} = \frac{10}{n}\sum_{i=1}^{m} n_i u_i + 180 =$$

$$= \frac{10\cdot(-3\cdot24+(-2)\cdot112+(-1)\cdot263+0\cdot322+1\cdot202+2\cdot66+3\cdot11)}{1000} + 180 =$$

$$= 0{,}01\cdot(-72-224-263+0+202+132+33) + 180 =$$

$$= 0{,}01\cdot(-192) + 180 = 178{,}08;$$

$$D_{\text{в}} = b^2\left(\frac{1}{n}\sum_{i=1}^{m} n_i u_i^2 - \left(\frac{1}{n}\sum_{i=1}^{m} n_i u_i\right)^2\right) =$$

$$= 10^2\cdot\left(\frac{(-3)^2\cdot24+(-2)^2\cdot112+(-1)^2\cdot263+0^2\cdot322+1^2\cdot202+2^2\cdot66+3^2\cdot11}{1000}\right) -$$

$$-\left[\frac{(-3)\cdot24+(-2)\cdot112+(-1)\cdot263+0\cdot322+1\cdot202+2\cdot66+3\cdot11}{1000}\right]^2 =$$

$$= 100\cdot\left(\frac{1492}{1000} - \left[\frac{-192}{1000}\right]^2\right) = 100 * 1{,}455136 = 145{,}5136.$$

Оскільки параметр a нормального закону розподілу є математичним сподіванням, а параметр σ – середнім квадратичним відхиленням, то оцінками цих параметрів є

$$a^* = \bar{x} = 178{,}08; \qquad \sigma^* = \sqrt{D_{\text{в}}} = \sqrt{145{,}5136} \approx 12{,}0629.$$

9.2.2. Метод найменших квадратів

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

Отож, використовуючи метод найменших квадратів, можна, наприклад, визначити статистичну оцінку для $\overline{X_\Gamma} = M(X)$. Для цього скористаємося функцією $u = \sum_{i=1}^{n}\left(x_i - \theta^*\right)^2 n_i$. Використовуючи умову екстремуму, отримаємо:

$$\frac{\partial u}{\partial \theta^*} = -2\sum_{i=1}^{n}\left(x_i - \theta^*\right)n_i = 0 \to$$

$$\to \sum_{i=1}^{n} x_i n_i - \sum_{i=1}^{n} \theta^* n_i = 0 \to$$

$$\theta^* = \frac{\sum_{i=1}^{n} x_i n_i}{n} = \overline{x_{\text{В}}}.$$

Звідси для $\theta = \overline{X_\Gamma}$ точковою статистичною оцінкою буде $\theta^* = \bar{x}$ – вибіркова середня.

9.2.3. Метод максимальної правдоподібності

Метод максимальної правдоподібності займає головне місце в теорії статистичного оцінювання параметрів θ. На нього свого часу зауважував К. Гаусс, а розробив його Р. Фішер. Цей метод розглянемо детальніше.

Метод максимальної правдоподібності полягає у знаходженні максимуму функції одного або кількох оцінюваних параметрів.

Припустимо, що X – дискретна випадкова величина з відомим законом розподілу, який визначається невідомим параметром θ. За даними вибірки $x_1, x_2, \ldots, x_n$, отриманої в результаті спостережень над випадковою величиною X, необхідно знайти точкову оцінку $\theta^* = \theta^*(x_1, x_2, \ldots, x_n)$ параметра θ.

Функцією правдоподібності L дискретної випадкової величини називають функцію аргумента θ

$$L(x_1, x_2, \ldots, x_n; \theta) = p(x_1; \theta) \cdot p(x_2; \theta) \cdot \ldots \cdot p(x_n; \theta), \qquad (9.6)$$

де $p(x_i; \theta)$ – ймовірність того, що в результаті випробування випадкова величина X набуде значення x_i.

Оцінкою найбільшої або максимальної правдоподібності параметра θ називають таке його значення θ^*, за якого функція правдоподібності досягає свого максимуму.

Логарифмічною функцією правдоподібності називають функцію $\ln L$.

Оскільки функції L і $\ln L$ досягають свого максимуму при одному і тому самому значенні θ, здебільшого зручніше знаходити максимум функції $\ln L$, а не L.

Якщо випадкова величина *X* неперервна, то відомою вважається щільність розподілу ймовірностей $f(x)$, а невідомим – параметр, від якого залежить ця щільність.

Функцією правдоподібності L неперервної випадкової величини називають функцію аргумента θ

$$L(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_n; \theta). \tag{9.7}$$

Оцінку найбільшої правдоподібності невідомого параметра розподілу неперервної випадкової величини шукають так само, як і у випадку дискретної випадкової величини, а саме:

1) визначають похідну $\frac{\partial \ln L}{\partial \theta}$ (або $\frac{\partial L}{\partial \theta}$);

2) знаходять корені θ_i^* рівняння $\frac{\partial \ln L}{\partial \theta} = 0$ (або $\frac{\partial L}{\partial \theta} = 0$). Ці рівняння називають ***рівняннями правдоподібності***;

3) визначають другу похідну $\frac{\partial^2 \ln L}{\partial \theta^2}$ (або $\frac{\partial^2 L}{\partial \theta^2}$). Корінь θ_i^* рівняння правдоподібності, для якого друга похідна від'ємна, беруть за оцінку θ^* найбільшої правдоподібності параметра θ.

Якщо параметр θ двовимірний, тобто $\theta = (\theta_1, \theta_2)$, то для знаходження максимуму функції правдоподібності складають і розв'язують систему рівнянь:

$$\begin{cases} \frac{\partial \ln L}{\partial \theta_1} = 0; \\ \frac{\partial \ln L}{\partial \theta_2} = 0. \end{cases} \quad \text{або} \quad \begin{cases} \frac{\partial L}{\partial \theta_1} = 0; \\ \frac{\partial L}{\partial \theta_2} = 0. \end{cases} \tag{9.8}$$

Наприклад, коли ознака генеральної сукупності *X* має нормальний закон розподілу, то функція максимальної правдоподібності набере такого вигляду:

$$L(x_1, x_2, \dots, x_n; \theta_1^*, \theta_2^*) = \frac{1}{(2\pi\theta_2)^{\frac{n}{2}}} \cdot e^{-\frac{\sum_{i=1}^{n}(x_i - \theta_1^*)^2}{2\theta_2^*}}. \tag{9.9}$$

Водночас за статистичні оцінки θ_1^*, θ_2^* вибирають ті їх значення, за яких задана вибірка буде найімовірнішою, тобто функція $L(x_1, x_2, \dots, x_n; \theta_1^*, \theta_2^*)$ досягає максимуму.

На практиці зручно від функції (9.9) перейти до її логарифма, а саме:

$$\ln L(x_1, x_2, \dots, x_n; \theta_1^*, \theta_2^*) = -\frac{n}{2}(\ln \pi + \ln \theta_2^*) - \frac{\sum_{i=1}^{n}(x_i - \theta_1^*)^2}{2\theta_2^*}. \tag{9.10}$$

Згідно з необхідною умовою екстремуму для цієї функції дістанемо:

$$\begin{cases} \frac{\partial \ln L}{\partial \theta_1^*} = - \frac{\sum_{i=1}^{n}(x_i - \theta_1^*)}{\theta_2^*} = 0; \\ \frac{\partial \ln L}{\partial \theta_2^*} = - \frac{n}{2\theta_2^*} + \frac{\sum_{i=1}^{n}(x_i - \theta_1^*)^2}{2(\theta_2^*)^2} = 0. \end{cases} \quad (9.11)$$

З першого рівняння системи дістанемо:

$$\theta_1^* = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$$

З другого рівняння системи маємо:

$$\theta_2^* = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = D_{\text{в}}.$$

Отже, для $\overline{X_\Gamma} = M(X)$ точковою статистичною оцінкою є $\bar{x}$ для $D_\Gamma - D_{\text{в}}$.

Приклад 9.4. Випадкова величина *X* (кількість битого скляного посуду в одній упаковці) розподілена за законом Пуассона з невідомим параметром λ. У результаті статистичних досліджень отримано такий емпіричний розподіл кількості битого скляного посуду в $n = 1000$ упаковках:

x_i	0	1	2	3	4	5	6
n_i	554	324	98	19	3	1	1

Знайти методом максимальної правдоподібності точкову оцінку невідомого параметра λ розподілу Пуассона.

Розв'язання. Оскільки випадкова величина *X* розподілена за законом Пуассона, то функція правдоподібності має вигляд:

$$L = \left(\frac{\lambda^0 e^{-\lambda}}{0!}\right)^{554} \cdot \left(\frac{\lambda^1 e^{-\lambda}}{1!}\right)^{324} \cdot \left(\frac{\lambda^2 e^{-\lambda}}{2!}\right)^{98} \cdot \left(\frac{\lambda^3 e^{-\lambda}}{3!}\right)^{19} \cdot \left(\frac{\lambda^4 e^{-\lambda}}{4!}\right)^{3} \cdot \left(\frac{\lambda^5 e^{-\lambda}}{5!}\right)^{1} \cdot \left(\frac{\lambda^6 e^{-\lambda}}{6!}\right)^{1}.$$

Запишемо логарифмічну функцію правдоподібності:

$$\ln L = 600 \ln \lambda - 1000\lambda - \ln(2^{133} \cdot 3^{25} \cdot 5^2).$$

Визначимо похідну логарифмічної функції розподілу за λ:

$$\frac{\partial \ln L}{\partial \lambda} = \frac{600}{\lambda} - 1000.$$

Прирівнявши її до нуля, знайдемо єдиний корінь рівняння правдоподібності:

$$\lambda^* = 0{,}6.$$

Оскільки друга похідна логарифмічної функції правдоподібності

$$\frac{\partial^2 \ln L}{\partial \lambda^2} = -\frac{600}{\lambda^2} < 0$$

завжди від'ємна, точковою оцінкою максимальної правдоподібності параметра λ розподілу Пуассона буде

$$\lambda^* = 0{,}6.$$

9.3. Інтервальні оцінки

Якщо об'єм вибірки малий, то точкові оцінки можуть давати значні похибки, тому питання точності оцінок у цьому випадку дуже важливе і використовують інтервальні оцінки.

Інтервальною називають статистичну оцінку, що визначається двома числами, кінцями інтервалів.

Нехай знайдена за даними вибірки статистична оцінка θ^* буде оцінкою невідомого параметра θ.

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

$$\left|\theta^* - \theta\right| < \delta, \tag{9.12}$$

де δ є точністю оцінки.

Оскільки θ^* є випадковою величиною, то і δ буде випадковою, тому нерівність (9.12) справджуватиметься з певною ймовірністю.

Імовірність, з якою береться нерівність (9.12), тобто

$$P\left(\left|\theta^* - \theta\right| < \delta\right) = \gamma \tag{9.13}$$

називають *надійністю*.

Рівність (9.13) можна записати так:

$$P\left(\theta^* - \delta < \theta < \theta^* + \delta\right) = \gamma \tag{9.14}$$

Найчастіше число γ задається наперед і, залежно від обставин, воно дорівнює 0,95 або 0,99 або 0,999.

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

Зауваження. Кінці довірчого інтервалу є випадковими величинами.

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

Нехай ознака X генеральної сукупності має нормальний закон розподілу. Побудуємо довірчий інтервал для $\overline{X_\Gamma}$, знаючи числове значення середнього квадратичного відхилення генеральної сукупності σ_Γ; із заданою надійністю γ. Оскільки $\bar{x}$ як точкова незміщена статистична оцінка для $\overline{X_\Gamma} = M(X)$ має нормальний закон розподілу з числовими характеристиками $M(\bar{x}) = \alpha$, $\sigma(\bar{x}) = \frac{\sigma_\Gamma}{\sqrt{n}}$, то, скориставшись (9.14), дістанемо

$$P\left(\left|\bar{x} - \alpha\right| < \delta\right) = \gamma. \tag{9.15}$$

Випадкова величина $\bar{x} - \alpha$ має нормальний закон розподілу з числовими характеристиками

$$M\left(\bar{x} - \alpha\right) = M(\bar{x}) - \alpha = \alpha - \alpha = 0,$$

$$D\left(\bar{x} - \alpha\right) = D(\bar{x}) = \frac{D_\Gamma}{n},$$

$$\sigma(\bar{x}) = \frac{\sigma_\Gamma}{\sqrt{n}}.$$

Тому $\frac{\bar{x}-\alpha}{\frac{\sigma_\Gamma}{\sqrt{n}}}$ матиме нормований нормальний закон розподілу $N(0;\ 1)$.

Звідси рівність (9.15) можна записати, назначивши $\frac{\delta}{\frac{\sigma_\Gamma}{\sqrt{n}}} = x$ так:

$$P\left(\left|\frac{\bar{x}-\alpha}{\frac{\sigma_\Gamma}{\sqrt{n}}}\right| < x\right) = \gamma \tag{9.16}$$

або

$$P\left(\bar{x}-\frac{x\cdot\sigma_{\Gamma}}{\sqrt{n}}<a<\bar{x}+\frac{x\cdot\sigma_{\Gamma}}{\sqrt{n}}\right)=\gamma.$$

Згідно з формулою нормованого нормального закону

$$P(|X-a|<\delta)=2\Phi(\delta)$$

для (9.16) вона набирає такого вигляду:

$$P\left(\left|\frac{\bar{x}-a}{\frac{\sigma_{\Gamma}}{\sqrt{n}}}\right|<x\right)=2\Phi(x)=\gamma. \qquad (9.17)$$

З рівності (9.17) знаходимо аргументи *x*, а саме:

$$2\Phi(x)=\gamma\to\Phi(x)=\frac{\gamma}{2}.$$

Аргумент *x* знаходимо за значенням функції Лапласа, яка дорівнює $\frac{1}{2}\gamma$ за таблицею.

Отже, довірчий інтервал дорівнюватиме:

$$\bar{x}-\frac{x\cdot\sigma_{\Gamma}}{\sqrt{n}}<a<\bar{x}+\frac{x\cdot\sigma_{\Gamma}}{\sqrt{n}}. \qquad (9.18)$$

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

Приклад 9.5. Вимірявши 40 випадково відібраних після виготовлення деталей, знайшли вибіркову середню, що дорівнює 15 см. Із надійністю $\gamma=0{,}99$ побудувати довірчий інтервал для середньої величини всієї партії деталей, якщо генеральна дисперсія дорівнює 0,99 см^2.

Розв'язання. Для побудови довірчого інтервалу необхідно знати: $\bar{x},\ \sigma_{\Gamma},\ n,\ x.$

З умови задачі маємо: $\bar{x}=15$ см, $\sigma_{\Gamma}=\sqrt{D_{\Gamma}}=\sqrt{0{,}99}=0{,}3$ см, $n=40$. Величина *x* обчислюється з рівняння:

$$\Phi(x)=\frac{\gamma}{2}=\frac{0{,}99}{2}=0{,}495;$$

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

Знайдемо числові значення кінців довірчого інтервалу:

$$\bar{x} - \frac{x \cdot \sigma_{\Gamma}}{\sqrt{n}} = 15 - \frac{2{,}58 \cdot 0{,}3}{\sqrt{40}} = 15 - 0{,}12 = 14{,}88 \text{ см};$$

$$\bar{x} + \frac{x \cdot \sigma_{\Gamma}}{\sqrt{n}} = 15 + \frac{2{,}58 \cdot 0{,}3}{\sqrt{40}} = 15 + 0{,}12 = 15{,}12 \text{ см}.$$

Відтак маємо:

$$14{,}88 < \overline{X_{\Gamma}} < 15{,}12.$$

Отже, з надійністю 0,99 (99% гарантії) оцінюваний параметр $\overline{X_{\Gamma}}$ перебуває всередині інтервалу $[14{,}87;\ 15{,}13]$.

Приклад 9.6. Маємо такі дані про розміри основних фондів (у млн грн) на 30-ти випадково вибраних підприємствах:

4,2; 2,4; 4,9; 6,7; 4,5; 2,7; 3,9; 2,1; 5,8; 4,0;
2,8; 7,8; 4,4; 6,6; 2,0; 6,2; 7,0; 8,1; 0,7; 6,8;
9,4; 7,6; 6,3; 8,8; 6,5; 1,4; 4,6; 2,0; 7,2; 9,1.

Побудувати інтервальний статистичний розподіл із довжиною кроку h = 2 млн грн.

З надійністю $\gamma = 0{,}999$ знайти довірчий інтервал для $\overline{X_{\Gamma}}$, якщо σ_{Γ} = 5 млн грн.

Розв'язання. Інтервальний статистичний розподіл буде таким:

h = 2 млн грн.	2 – 4	4 – 6	6 – 8	8 – 10
n_i	9	7	10	4

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

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

$$n = \sum n_i = 30.$$

Тоді

$$\bar{x} = \frac{\sum n_i x_i^*}{n} = \frac{3\cdot 9+5\cdot 7+7\cdot 10+9\cdot 4}{30} = \frac{168}{30} = 5{,}6 \text{ млн грн.}$$

Для побудови довірчого інтервалу із заданою надійністю $\gamma = 0{,}999$ необхідно знайти x:

$$\Phi(x) = \frac{\gamma}{2} = \frac{0{,}999}{2} = 0{,}4995;$$

$\Phi(x) = 0{,}4995 \ \rightarrow x = 3{,}4$ (за таблицею значень функції Лапласа).

Обчислюємо кінці інтервалу:

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

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

Отже, довірчий інтервал для $\overline{X_\Gamma}$ буде $2{,}5 < \overline{X_\Gamma} < 8{,}7$.

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

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

$$t = \frac{\bar{x} - \alpha}{\frac{S}{\sqrt{n}}}, \qquad (9.19)$$

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

Тоді (9.19) набуває такого вигляду:

$$P\left(\left|\frac{\bar{x}-\alpha}{\frac{S}{\sqrt{n}}}\right| < t_\gamma\right) = P\left(\bar{x} - \frac{t_\gamma\cdot S}{\sqrt{n}} < a < \bar{x} + \frac{t_\gamma\cdot S}{\sqrt{n}}\right) = 2\int_0^{t_\gamma} f(t)dt = \gamma,$$

оскільки $f(t)$ для розподілу Стьюдента є функцією парною.

Обчисливши за цим статистичним розподілом $\bar{x}$, S і визначивши за таблицею розподілу Стьюдента значення t_γ, будуємо довірчий інтервал

$$\bar{x} - \frac{t_\gamma \cdot S}{\sqrt{n}} < a < \bar{x} + \frac{t_\gamma \cdot S}{\sqrt{n}}. \tag{9.20}$$

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

Приклад 9.7. Випадково вибрана партія з двадцяти приладів була випробувана щодо терміну безвідказної 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 «a» (середнього часу безвідказної роботи приладу).

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

Обчислимо $\bar{x}$:

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

Визначимо D_B:

$$D_B = \frac{100^2 \cdot 2 + 170^2 \cdot 5 + 240^2 \cdot 10 + 310^2 \cdot 2 + 380^2 \cdot 1}{20} - 222{,}5^2 = \frac{1077100}{20} - 49506{,}25 = 53855 - 49506{,}25 = 4348{,}75.$$

Виправлене середнє квадратичне відхилення дорівнюватиме:

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

За таблицею значень $\int_0^t f(t)dt = \gamma = 0{,}99$ розподілу Стьюдента за заданою надійністю $\gamma = 0{,}99$ і числом ступенів свободи $k = n-1 = 20-1 = 19$ знаходимо значення $t_\gamma(\gamma = 0{,}99; k = 19) = 2{,}86$.

Обчислимо кінці довірчого інтервалу:

$$\bar{x} - \frac{t_\gamma \cdot S}{\sqrt{n}} = 222{,}5 - \frac{2{,}86 \cdot 67{,}66}{\sqrt{0}} = 222{,}5 - 43{,}27 = 179{,}23 \text{ год.};$$

$$\bar{x} + \frac{t_\gamma \cdot S}{\sqrt{n}} = 222{,}5 + \frac{2{,}86 \cdot 67{,}66}{\sqrt{0}} = 222{,}5 + 43{,}27 = 265{,}77 \text{ год.}$$

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

$$179{,}23 < a < 265{,}77.$$

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

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

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

$$\chi^2 = \frac{n-1}{\sigma_\Gamma^2} S^2, \tag{9.21}$$

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

Оскільки випадкові події

$$A(\chi_1^2 < \chi^2 < \chi_2^2) \text{ і } B\left(\frac{1}{\chi_2^2} < \frac{1}{\chi^2} < \frac{1}{\chi_1^2}\right)$$

є рівноймовірними, тобто їх імовірності рівні $(P(A) = P(B))$, то маємо:

$$P(\chi_1^2 < \chi^2 < \chi_2^2) = P\left(\frac{1}{\chi_2^2} < \frac{1}{\chi^2} < \frac{1}{\chi_1^2}\right). \tag{9.22}$$

Підставляючи в (9.22) $\chi^2 = \frac{n-1}{\sigma_\Gamma^2} S^2$ дістанемо

$$P\left(\frac{1}{\chi_2^2} < \frac{1}{\chi^2} < \frac{1}{\chi_1^2}\right) = P\left(\frac{1}{\chi_2^2} < \frac{1}{\frac{n-1}{\sigma_\Gamma^2} S^2} < \frac{1}{\chi_1^2}\right) = P\left(\frac{1}{\chi_2^2} < \frac{\sigma_\Gamma^2}{(n-1)S^2} < \frac{1}{\chi_1^2}\right) =$$

$$= P\left(\frac{(n-1)S^2}{\chi_2^2} < \sigma_\Gamma^2 < \frac{(n-1)S^2}{\chi_1^2}\right) = \gamma.$$

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

$$\frac{(n-1)S^2}{\chi_2^2} < D_\Gamma < \frac{(n-1)S^2}{\chi_1^2}. \tag{9.23}$$

Тоді довірчий інтервал для σ_Γ випливає із (9.23) і буде таким:

$$\frac{S\sqrt{n-1}}{\chi_2} < \sigma_\Gamma < \frac{S\sqrt{n-1}}{\chi_1}. \tag{9.24}$$

Значення χ_1^2, χ_2^2 знаходимо за таблицею згідно з рівностями:

$$P(\chi^2 > \chi_1^2) = 1 - \frac{\alpha}{2}; \tag{9.25}$$

$$P(\chi^2 > \chi_2^2) = \frac{\alpha}{2}; \tag{9.26}$$

$$\alpha = 1 - \gamma.$$

Приклад 9.8. У результаті статистичних досліджень випадкової величини X отримано вибірку об'єму $n = 25$ із таким статистичним розподілом:

x_i	0	1	2	3	4	5	6
n_i	1	3	4	6	5	4	2

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 $\bar{x}$:

$$\bar{x} = \frac{\sum x_i n_i}{n} = \frac{0\cdot1+1\cdot3+2\cdot4+3\cdot6+4\cdot5+5\cdot4+6\cdot2}{25} = \frac{81}{25} = 3,24.$$

Обчислимо D_B:

$$D_B = \frac{0^2 \cdot 1 + 1^2 \cdot 3 + 2^2 \cdot 4 + 3^2 \cdot 6 + 4^2 \cdot 5 + 5^2 \cdot 4 + 6^2 \cdot 2}{25} - 3{,}24^2 = \frac{325}{25} - 10{,}498 =$$
$$= 13 - 10{,}498 = 2{,}502.$$

Виправлена дисперсія і виправлене середнє квадратичне відхилення дорівнюватимуть:

$$S^2 = \frac{n}{n-1} \cdot D_B = \frac{25}{25-1} \cdot 2{,}502 = 2{,}634;$$

$$S = \sqrt{S^2} = \sqrt{2{,}634} \approx 1{,}623.$$

Оскільки $\alpha = 1 - \gamma = 1 - 0{,}99 = 0{,}01$, то згідно з (9.25), (9.26) знаходимо значення χ_1^2, χ_2^2, а саме:

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

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

За таблицею знаходимо:

$$\chi_1^2(0{,}995; 24) = 10{,}9;$$

$$\chi_2^2(0{,}005; 24) = 45{,}5.$$

Обчислимо кінці довірчого інтервалу для $D_Г$:

$$\frac{(n-1)S^2}{\chi_2^2} = \frac{25-1}{45{,}5} \cdot 2{,}634 = 1{,}389;$$
$$\frac{(n-1)S^2}{\chi_1^2} = \frac{25-1}{10{,}9} \cdot 2{,}634 = 5{,}799.$$

Отже, довірчий інтервал для $D_Г$ буде таким:

$$1{,}389 < D_Г < 5{,}799.$$

Довірчий інтервал для $\sigma_Г$ становить

$$1{,}179 < \sigma_Г < 2{,}408.$$

Запитання для самоконтролю

1. Що називається статистичною оцінкою?
2. Що називається точковою статистичною оцінкою?
3. Що таке незміщена точкова статистична оцінка?
4. Що таке зміщена точкова статистична оцінка?
5. Що називають ефективною точковою статистичною оцінкою?
6. Що називають обґрунтованою точковою статистичною оцінкою?
7. Яку статистичну оцінку називають спроможною?
8. Яку статистичну оцінку називають сильно спроможною?
9. У чому сутність методу моментів?
10. У чому сутність методу найменших квадратів?
11. У чому сутність методу максимальної правдоподібності?
12. Що називається функцією правдоподібності дискретної випадкової?
13. Що називається логарифмічною функцією правдоподібності?
14. Що називається функцією правдоподібності неперервної випадкової?
15. Який вигляд мають рівняння правдоподібності.
16. Що є точковою незміщеною статистичною оцінкою для $\bar{X}_\Gamma$?
17. Що означає точкова незміщена статистична оцінка для D_Γ?
18. Що називається виправленою дисперсією?
19. Що називається виправленим середнім квадратичним відхиленням?
20. Який закон розподілу ймовірностей має випадкова величина $\bar{x}$?
21. Який закон розподілу має випадкова величина $\frac{n-1}{\sigma_\Gamma^2}S^2$?
22. Який закон розподілу ймовірностей має випадкова величина $\frac{S\sqrt{n-1}}{\sigma_\Gamma}$?
23. Визначення інтервальної статистичної оцінки для параметрів генеральної сукупності.
24. Що називають точністю оцінки?
25. Що називають надійністю оцінки?
26. Що називають довірчим інтервалом?
27. Який закон розподілу ймовірностей має випадкова величина $\frac{\bar{x}-\alpha}{\frac{\sigma_\Gamma}{\sqrt{n}}}$?
28. Як побудувати довірчий інтервал із заданою надійністю γ при відомому значенні σ_Γ?
29. Який закон розподілу ймовірностей має випадкова величина $\frac{\bar{x}-\alpha}{\frac{S}{\sqrt{n}}}$?
30. Як побудувати довірчий інтервал для $\bar{X}_\Gamma$ із заданою надійністю γ при невідомому значенні σ_Γ?
31. Як побудувати довірчий інтервал для $\bar{X}_\Gamma$ із заданою надійністю γ при невідомому значенні σ_Γ і при обсягах вибірки $n > 30$?
32. Як побудувати довірчий інтервал із надійністю γ для D_Γ, σ_Γ, використовуючи розподіл χ^2?

Задачі для самостійного розв'язування

1. Із нормально розподіленої сукупності з $D(X) = \sigma^2$ зроблено вибірку обсягом n. Знайти оцінку для $M(X)$. Перевірити її на незміщеність, ефективність та обґрунтованість.

2. Із нормально розподіленої сукупності з $M(X) = a$ зроблено вибірку обсягом n. Знайти оцінку для $D(X)$. Перевірити її на незміщеність, ефективність та обґрунтованість.

3. З генеральної сукупності отримано деяку вибірку:

x_i	0,001	0,003	0,005	0,007	0,009	0,011	0,013	0,015
n_i	5	30	34	13	8	5	3	2

Знайти зміщену оцінку генеральної дисперсії.

4. Випадкова величина X розподілена за нормальним розподілом з параметрами *a, σ.* Таблиця містить дані про зріст 1000 учнів старших класів:

Зріст, см	143-146	146-149	149-152	152-155	155-158	158-161	161-164	164-167	167-170	170-173	173-176	176-179	179-182	182-185	185-188
Число учнів	1	2	8	26	65	120	181	201	170	120	64	27	10	3	1

Знайти методом моментів точкову оцінку невідомих параметрів *a, σ* нормального розподілу.

5. Випадкова величина X розподілена за нормальним розподілом. Знайти методом моментів точкову оцінку невідомих параметрів.

$[x_i;\ x_{i+1})$	$[-1;\ 0)$	$[0;\ 1)$	$[1;\ 2)$	$[2;\ 3)$	$[3;\ 4)$	$[4;\ 5)$	$[5;\ 6)$
n_i	1	2	5	11	6	3	2

6. Випадкова величина X розподілена за розподілом Пуассона. Знайти методом моментів точкову оцінку невідомих параметрів.

x_i	0	1	2	3	4	5	6	7	8	9
n_i	4	11	17	27	16	15	5	2	2	1

7. Випадкова величина X розподілена за біноміальним розподілом. Знайти методом моментів точкову оцінку невідомих параметрів.

x_i	0	1	2	3	4	5	6	7	8
n_i	3	10	16	34	17	11	5	3	1

8. Визначити мінімальний обсяг вибірки n для того, щоб із надійністю $\gamma = 0{,}98$ можна було дістати оцінку математичного сподівання нормально розподіленої сукупності $M(X) = 0{,}2$, якщо середнє квадратичне відхилення в генеральній сукупності $\sigma = 1{,}5$ і оцінка знаходиться за допомогою вибіркової середньої величини.

9. Знайти з надійністю 0,95 довірчий інтервал оцінки математичного сподівання a нормально розподіленої ознаки X генеральної сукупності, якщо відомі вибіркова середня $\bar{x}$, об'єм вибірки n та середнє квадратичне відхилення σ генеральної сукупності.

а) $\bar{x} = 14;\ \ n = 25,\ \ \sigma = 5;$ б) $\bar{x} = 10{,}2;\ \ n = 16,\ \ \sigma = 4;$

в) $\bar{x} = 16{,}8;\ \ n = 25,\ \ \sigma = 5;$ г) $\bar{x} = 2000;\ \ n = 1600,\ \ \sigma = 40.$

10. У будинку відпочинку випадковим способом було відібрано 20 осіб і виміряно їх зріст x_i. Здобуті результати наведено у вигляді інтервального статистичного розподілу:

x_i, см	165,5 – 170,5	170,5 – 175,5	175,5 – 180,5	180,5 – 185,5
n_i	4	6	8	2

Із надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для $\bar{X}_\Gamma$, якщо $\sigma_\Gamma = 2$.

11. 28 однотипних приладів були випробувані щодо їх безвідмовної роботи x_i. Результати вимірювання наведено в таблиці:

x_i, год	100	110	120	130	140	150
n_i	10	6	5	4	2	1

З надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для $\bar{X}_\Gamma$.

12. За вибіркою з нормального розподілу з надійністю $\gamma = 0{,}95$ знайти інтервальні оцінки для $\bar{X}_{\Gamma}$, якщо

а) $\sigma_{\Gamma} = 4$; б) σ – невідоме.

x_i	-2	1	2	3	4	5
n_i	4	3	4	6	2	1

Знайти також довірчий інтервал для D_{Γ}, σ_{Γ} якщо $\gamma = 0{,}9$.

13. Під час вибіркового аналізу максимального завантаження літака при певній кількості пального, дістали такі дані, т: 78; 80; 85; 88; 89; 90; 92; 93; 92; 85; 93; 89; 80; 90; 85; 89.

Припускаючи, що максимальне завантаження літака має нормальний розподіл, знайти довірчий інтервал для дисперсії та середнього квадратичного відхилення з надійністю γ = 0,95.

14. Побудувати довірчий інтервал з надійністю γ = 0,96 для невідомої дисперсії генеральної сукупності спостережень випадкової величини X, розподіленої нормально, за вибіркою із обсягом $n = 20$ та відомою виправленою вибірковою дисперсією $S^2 = 10$.

Розділ 10

ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ

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

За результатами тестів можна зробити висновок про те, чи кращі нові покришки від покришок старого типу, чи ні. А це, своєю чергою, дає підставу для ухвалення рішення: виготовляти їх чи ні.

Такі рішення називають *статистичними*.

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

Головна цінність ухвалення статистичних рішень полягає в тому, що в межах імовірнісних категорій можна об’єктивно виміряти ступінь ризику, що відповідає тому чи іншому рішенню.

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

Статистичні гіпотези про значення параметрів ознак генеральної сукупності називають *параметричними*.

Наприклад, висувається статистична гіпотеза про числові значення генеральної середньої $\bar{x}_\Gamma$, генеральної дисперсії D_Γ, генерального середнього квадратичного відхилення σ_Γ та ін.

Статистичні гіпотези, що висуваються на підставі обробки вибірки про закон розподілу ознаки генеральної сукупності, називаються *непараметричними*. Наприклад, на підставі обробки вибірки може бути висунута гіпотеза, що ознака генеральної сукупності має нормальний закон розподілу, експоненціальний закон та ін.

10.1. Статистична перевірка гіпотез

Статистичною називається гіпотеза про вигляд невідомого розподілу випадкової величини або про параметри відомих розподілів.

Нульовою (основною) називають висунуту гіпотезу H_0.

Зміст нульової гіпотези записується так:

$$H_0: \ \bar{x}_\Gamma = a;$$

Конкуруючою (альтернативною) називають гіпотезу H_a, яка суперечить нульовій.

Кожній нульовій гіпотезі можна протиставити кілька альтернативних (конкуруючих) гіпотез.

Приміром, нульова гіпотеза стверджує: H_0: $\bar{x}_\Gamma = a$, а альтернативна гіпотеза – H_a: $\bar{x}_\Gamma > a$, тобто заперечує твердження нульової.

Простою називають гіпотезу, яка містить лише одне твердження.

Складною називають гіпотезу, яка складається зі скінченного або нескінченного числа простих гіпотез.

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:

$$H_0: \ \bar{x}_\Gamma \in [2; \ 2{,}1; \ 2{,}2].$$

Перевірка гіпотези здійснюється за даними вибірки, тобто статистичними методами. Тому перевірку гіпотези за даними вибірки називають статистичною.

У результаті статистичної перевірки правильності основної гіпотези H_0 за результатами вибірки у двох випадках може бути ухвалено неправильне рішення, тобто можуть бути допущені помилки двох видів.

Помилка першого роду полягає в тому, що буде відхилена правильна гіпотеза.

Помилка другого роду полягає в тому, що буде прийнята неправильна гіпотеза.

Ймовірність здійснити помилку першого роду позначають α і називають **рівнем значущості.**

Найчастіше рівень значущості приймають рівним 0,05 або 0,01. Якщо прийнято рівень значущості рівним 0,05, то це означає, що в п’яти випадках із 100 ми ризикуємо одержати похибку першого роду (відкинути правильну гіпотезу).

Ймовірність здійснити помилку другого роду позначають β.

Для перевірки нульової гіпотези вибирається деяка випадкова величина *К,* розподіл якої відомий, і яка називається **статистичним критерієм** (або просто критерієм) перевірки нульової гіпотези.

Спостережним значенням критерію узгодження називають значення відповідного критерію, обчислене за даними вибірки.

Критичною областю називають множину значень критерію, при яких нульову гіпотезу відхиляють.

Областю прийняття гіпотези (областю допустимих значень) називають множину значень критерію, при яких нульову гіпотезу приймають.

Основний принцип перевірки статистичних гіпотез такий: якщо спостережуване значення критерію належить критичній області, нульову гіпотезу відхиляють; якщо спостережуване значення критерію належить області прийняття гіпотези, гіпотезу приймають.

Критичними точками $k_{\text{кр}}$ називають точки, які відокремлюють критичну область від області прийняття гіпотези.

Правосторонньою називають критичну область, яка визначається нерівністю

$$К > k_{\text{кр}} \ (k_{\text{кр}} > 0). \tag{10.1}$$

Лівосторонньою називають критичну область, яка визначається нерівністю

$$К < k_{\text{кр}} \ (k_{\text{кр}} < 0). \tag{10.2}$$

Двосторонньою називають критичну область, яка визначається нерівностями

$$К < k_1, К > k_2 \ (k_2 > k_1). \tag{10.3}$$

Зокрема, якщо критичні точки симетричні відносно нуля, то двостороння критична область визначається нерівностями

$$К < -k_{\text{кр}}, \ К > k_{\text{кр}} \ (k_{\text{кр}} > 0), \tag{10.4}$$

або рівносильною їй нерівністю

$$|K| > k_{\text{кр}}. \tag{10.5}$$

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

Для знаходження правосторонньої критичної області достатньо знайти критичну точку $k_{\text{кр}}$. Для цього задають малу ймовірність α; число α називається рівнем значущості критерію. Далі знаходять критичну точку, виходячи з умови, що при справедливості нульової гіпотези виконується рівність

$$P\{ К > k_{кр}\} = \alpha.$$

Для кожного критерію є відповідні таблиці, з яких знаходять критичну точку, що задовольняє цю формулу.

Аналогічно знаходять лівосторонню критичну область.

Двосторонню критичну область знаходять за умови справедливості нульової гіпотези з рівності

$$P\{ К < k_1 \} + P\{ К > k_2\} = \alpha. \quad (10.6)$$

Якщо розподіл критерію симетричний щодо нуля, то

$$P\{ К < k_1\} = P\{ К > k_2 \}.$$

Врахувавши рівність (10.6), одержимо

$$P\{ К > k_{кр} \} = \frac{\alpha}{2}. \quad (10.7)$$

Тоді двосторонню критичну область знаходять з рівності (10.7). Критичні точки знаходять з відповідних таблиць.

При знаходженні критичної області доцільно враховувати потужність критерію.

Потужністю критерію називають імовірність належності критерію критичній області за умови, що правильна альтернативна гіпотеза.

Тобто потужність критерію є імовірність того, що основна гіпотеза буде відхилена, якщо альтернативна гіпотеза правильна.

Якщо рівень значущості вже обрано, то критичну область доцільно будувати так, щоб потужність критерію була максимальною. Виконання цієї вимоги забезпечує мінімальну імовірність похибки другого роду.

Зауваження. Єдиний спосіб одночасного зменшення ймовірностей похибок першого та другого роду це є збільшення об’єму вибірки.

Загальний алгоритм перевірки правильності нульової гіпотези

Для перевірки правильності H_0 задається так званий рівень значущості α.

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. Визначити допустиму імовірність похибки першого роду, тобто рівень значущості α;

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

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

Наголосимо, що між рівнем значущості (та критичною областю існує такий зв'язок: якщо гіпотеза H_0 правильна, то з імовірністю значення вибіркової функції будуть належати критичній області.

10.2. Перевірка рівності вибіркового середнього гіпотетичному генеральному середньому

Нехай із **генеральної сукупності з відомою дисперсією $\boldsymbol{\sigma^2}$ отримано вибірку об'єму $\boldsymbol{n}$**.

Правило 1. Для того, щоб при заданому рівні значущості α перевірити нульову гіпотезу $H_0\colon a = a_0$ (про рівність генерального середнього a нормальної сукупності з відомою дисперсією σ^2 гіпотетичному (прогнозованому) значенню a_0) при конкуруючій гіпотезі $H_a\colon a \neq a_0$, потрібно обчислити спостережуване значення критерію

$$U_{\text{спост}} = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma} \tag{10.8}$$

і за таблицею функції Лапласа знайти критичну точку $u_{\text{кр}}$ двосторонньої критичної області з рівності

$$\Phi(u_{\text{кр}}) = \frac{1-\alpha}{2}.$$

Якщо $|U_{\text{спост}}| < u_{\text{кр}}$, немає підстав відхиляти нульову гіпотезу. Якщо $|U_{\text{спост}}| > u_{\text{кр}}$, нульову гіпотезу відхиляють.

Правило 2. При конкуруючій гіпотезі $H_a\colon a > a_0$ критичну точку правосторонньої критичної області знаходять з рівності

$$\Phi(u_{\text{кр}}) = \frac{1}{2} - \alpha.$$

Якщо $U_{\text{спост}} < u_{\text{кр}}$, немає підстав відхиляти нульову гіпотезу. Якщо $U_{\text{спост}} > u_{\text{кр}}$, то нульову гіпотезу відхиляють.

Правило 3. При конкуруючій гіпотезі $H_a: a < a_0$ спочатку знаходять допоміжну критичну точку $u_{\text{кр}}$ за правилом 2, а потім вважають межею лівосторонньої критичної області

$$u'_{\text{кр}} = -u_{\text{кр}}.$$

Якщо $U_{\text{спост}} > -u_{\text{кр}}$, то немає підстав відхиляти нульову гіпотезу. Якщо $U_{\text{спост}} < u_{\text{кр}}$, то нульову гіпотезу відхиляють.

Потужність критерію перевірки основної гіпотези $H_0 : a = a_0$ про рівність генерального середнього a гіпотетичному значенню a_0 при відомому середньому квадратичному відхиленні σ знаходять залежно від вигляду альтернативної гіпотези.

При альтернативній гіпотезі $H_a : a > a_0$ для гіпотетичного значення генерального середнього $a = a_1 > a_0$ потужність правостороннього критерію

$$1 - \beta = \frac{1}{2} - \Phi(u_{\text{кр}} - \lambda),$$

де $u_{\text{кр}}$ знаходять з рівності $\Phi\left(u_{\text{кр}}\right) = \frac{1}{2} - \alpha,\ \lambda = \frac{(a_1 - a_0)\sqrt{n}}{\sigma}$. При різних значеннях a_1 функція потужності одностороннього критерію

$$\pi_1(a_1) = \frac{1}{2} - \Phi(u_{\text{кр}} - \lambda).$$

При альтернативній гіпотезі $H_a: a \neq a_0$ для гіпотетичного значення генерального середнього $a = a_1$ потужність двостороннього критерію

$$1 - \beta = 1 - \left[\Phi(u_{\text{кр}} - \lambda) + \Phi(u_{\text{кр}} + \lambda)\right],$$

де $u_{\text{кр}}$ знаходять з рівності $\Phi\left(u_{\text{кр}}\right) = \frac{1-\alpha}{2}$, $\lambda = \frac{(a_1 - a_0)\sqrt{n}}{\sigma}$. При різних значеннях a_1 функція потужності одностороннього критерію

$$\pi_1(a_1) = 1 - \left[\Phi(u_{\text{кр}} - \lambda) + \Phi(u_{\text{кр}} + \lambda)\right].$$

Приклад 10.1. Із нормальної генеральної сукупності з відомим середнім квадратичним відхиленням $\sigma = 4{,}8$ отримано вибірку об'єму $n = 144$, за якою знайдено вибіркове середнє $\bar{x} = 16$. Потрібно при рівні значущості $\alpha = 0{,}05$ перевірити нульову гіпотезу $H_0 : a = a_0 = 15$ при конкуруючій гіпотезі:

А. $H_a: a \neq 15$.

Б. $H_a: a > 15$.

В. $H_a: a < 15$.

Крім того, необхідно знайти потужність правостороннього та двохстороннього критеріїв.

Розв'язання. Обчислимо спочатку спостережуване значення критерію

$$U_{\text{спост}} = \frac{(\bar{x}-a_0)\sqrt{n}}{\sigma} = \frac{(16-15)\sqrt{144}}{4.8} = 2{,}5.$$

А. Скористаємося правилом 1. Знайдемо критичну точку з рівності

$$\Phi(u_{\text{кр}}) = \frac{1-\alpha}{2} = \frac{1-0{,}05}{2} = 0{,}475.$$

За таблицею функції Лапласа визначимо критичну точку:

$$u_{\text{кр}} \approx 1{,}96.$$

Оскільки $|U_{\text{спост}}| > u_{\text{кр}}$, основна гіпотеза відхиляється. Тобто, вибіркове та генеральні середні суттєво відрізняються.

Б. Скористаємося правилом 2. Знайдемо критичну точку з рівності:

$$\Phi(u_{\text{кр}}) = \frac{1}{2} - \alpha = \frac{1}{2} - 0{,}05 = 0{,}45.$$

За таблицею функції Лапласа визначимо критичну точку:

$$u_{\text{кр}} \approx 1{,}64.$$

Оскільки $U_{\text{спост}} > u_{\text{кр}}$, основна гіпотеза відхиляється. Інакше кажучи, вибіркове та гіпотетичне генеральне середні суттєво відрізняються.

В. Скористаємося правилом 3. Критична точка буде такою ж, як і в пункті б), але з протилежним знаком:

$$u_{\text{кр}} \approx -1{,}64.$$

Оскільки $U_{\text{спост}} > u_{\text{кр}}$, то немає підстав відхиляти нульову гіпотезу. Тобто вибіркове та гіпотетичне генеральне середні несуттєво відрізняються.

Тепер знайдемо потужності правостороннього та двохстороннього критеріїв. Нагадаємо, що критичні точки у цих випадках різні та дорівнюють $1{,}64$ і $1{,}96$ відповідно .

Знайдемо параметр λ, який входить в обидва рівняння для визначення потужності критеріїв:

$$\lambda = \frac{(\bar{x}-a_0)\sqrt{n}}{\sigma} = \frac{(16-15)\sqrt{144}}{4{,}8} = 2{,}5.$$

Отже, потужності відповідно правостороннього та двостороннього критеріїв будуть такі:

$$\pi_1^{\text{правост}}(16) = \frac{1}{2} - \Phi(u_{\text{кр}} - \lambda) = \frac{1}{2} - \Phi(1{,}64 - 2{,}5) = \frac{1}{2} + \Phi(0{,}86) =$$
$$= 0{,}5 + 0{,}3051 = 0{,}8051.$$

$$\pi_1^{\text{двост}}(16) = 1 - \left[\Phi(u_{\text{кр}} - \lambda) + \Phi(u_{\text{кр}} + \lambda)\right] =$$
$$= 1 - [\Phi(1{,}96 - 2{,}5) + \Phi(1{,}96 - 2{,}5)] =$$
$$= 1 - \left[-\Phi(0{,}54) + \Phi(4{,}46)\right] == 1 + 0{,}2054 - 0{,}5 = 0{,}7054.$$

Тобто ймовірності того, що нульова гіпотеза буде відхилена, якщо правильною є конкуруюча гіпотеза, дорівнюють $0{,}8051$ і $0{,}7054$ відповідно для правостороннього та двостороннього критеріїв.

Нехай із **генеральної сукупності отримано вибірку об'єму *n* з невідомою дисперсією**.

Якщо дисперсія генеральної сукупності невідома, то як критерій перевірки основної гіпотези беруть випадкову величину

$$T = \frac{(\bar{x}-a_0)\sqrt{n}}{s}, \tag{10.9}$$

де $s = \sqrt{\frac{n\sum_i n_i x_i^2 - \left(\sum_i n_i x_i\right)^2}{n(n-1)}}$ – виправлене вибіркове середнє квадратичне відхилення. Величина T має розподіл Стьюдента з $k = n - 1$ ступенями вільності.

Правило 1. Для того, щоб при заданому рівні значущості α перевірити нульову гіпотезу $H_0 : a = a_0$ про рівність невідомого генерального середнього a нормальної сукупності з невідомою дисперсією гіпотетичному (прогнозованому) значенню a_0 при конкуруючій гіпотезі $H_a: a \neq a_0$ потрібно обчислити спостережуване значення критерію

$$T_{\text{спост}} = \frac{(\bar{x} - a_0)\sqrt{n}}{s}$$

і за таблицею критичних точок розподілу Стьюдента при заданому рівні значущості α, розміщеному у верхній частині таблиці, і кількості ступенів вільності $k = n - 1$ знайти критичну точку $t_{\text{двост кр}}(\alpha; k)$.

Якщо $|T_{\text{спост}}| < t_{\text{двост кр}}$, то немає підстав відхиляти нульову гіпотезу. Якщо $|T_{\text{спост}}| > t_{\text{двост кр}}$ – нульову гіпотезу відхиляють.

Правило 2. При конкуруючій гіпотезі $H_a: a > a_0$ за рівнем значущості α, розміщеним у нижній частині таблиці розподілу Стьюдента, і при кількості ступенів вільності $k = n - 1$ знаходять критичну точку правосторонньої критичної області $t_{\text{правост кр}}(\alpha; k)$.

Якщо $T_{\text{спост}} < t_{\text{правост кр}}$, немає підстав відхиляти нульову гіпотезу. Якщо $T_{\text{спост}} > t_{\text{правост кр}}$, нульову гіпотезу відхиляють.

Правило 3. При конкуруючій гіпотезі $H_a: a < a_0$ спочатку знаходять допоміжну критичну точку (за правилом 2) $t_{\text{правост кр}}(\alpha; k)$ і вважають межею лівосторонньої критичної області $t_{\text{лівост кр}}(\alpha; k) = - t_{\text{правост кр}}(\alpha; k)$.

Якщо $T_{\text{спост}} > - t_{\text{правост кр}}$, немає підстав відхиляти нульову гіпотезу. Якщо $T_{\text{спост}} < - t_{\text{правост кр}}$, нульову гіпотезу відхиляють.

Приклад 10.2. Проведено 10 незалежних експериментів над випадковою величиною *X*, що має нормальний закон розподілу з невідомими значеннями a, σ. Наслідки експериментів подано у вигляді статистичного ряду:

x_i	2,5	2	–2,3	1,9	–2,1	2,4	2,3	–2,5	1,5	–1,7
n_i	1	1	1	1	1	1	1	1	1	1

При рівні значущості α = 0,001 перевірити правильність нульової гіпотези $H_0: a = 0{,}9$, при альтернативній гіпотезі $H_a: a < 0{,}9$.

Розв'язання. Запишемо статистичний ряд у вигляді статистичного розподілу й обчислимо $\bar{x}$, s:

x_i	–2,5	–2,3	–2,1	–1,7	1,5	1,9	2	2,3	2,4	2,5
n_i	1	1	1	1	1	1	1	1	1	1

$$n = 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 = 10;$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i = \frac{1}{10}\Big(-2{,}5 \cdot 1 + (-2{,}3) \cdot 1 + (-2{,}1) \cdot 1 + (-1{,}7) \cdot 1 + 1{,}5 \cdot 1 + \\ +1{,}9 \cdot 1 + 2 \cdot 1 + 2{,}3 \cdot 1 + 2{,}4 \cdot 1 + 2{,}5 \cdot 1\Big) = 0{,}4;$$

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{m} n_i x_i\right)^2 = \frac{1}{10}\Big((-2{,}5)^2 \cdot 1 + (-2{,}3)^2 \cdot 1 + (-2{,}1)^2 \cdot 1 + \\ +(-1{,}7)^2 \cdot 1 + 1{,}5^2 \cdot 1 + 1{,}9^2 \cdot 1 + 2^2 \cdot 1 + 2{,}3^2 \cdot 1 + +2{,}4^2 \cdot 1 + 2{,}5^2 \cdot \\ 1\Big) - -0{,}4^2 = 4{,}6 - 0{,}16 = 4{,}44;$$

$$S^2 = \frac{n}{n-1} D_{\text{в}} = \frac{10}{9} \cdot 4{,}44 = 4{,}933;$$

$$s = \sqrt{S^2} = \sqrt{4{,}933} \approx 2{,}22.$$

При альтернативній гіпотезі $H_a\colon a < 0{,}9$ будується лівобічна критична область. Для цього необхідно знайти критичну точку

$$t_{\text{лівост кр}}(\alpha; k) = -\ t_{\text{правост кр}}(\alpha; k);$$

$$t_{\text{правост кр}}(\alpha; k) = t(0{,}001; 9) = 4{,}78;$$

$$t_{\text{лівост кр}}(\alpha; k) = -4{,}78.$$

Обчислимо спостережуване значення критерію:

$$T_{\text{спост}} = \frac{(\bar{x} - a_0)\sqrt{n}}{s} = \frac{(0{,}4 - 0{,}9) \cdot \sqrt{10}}{2{,}22} = -\frac{0{,}5 \cdot \sqrt{10}}{2{,}22} \approx -0{,}712.$$

Оскільки $T_{\text{спост}} > -\ t_{\text{правост кр}}$, то немає підстав відхилити $H_0\colon a = 0{,}9$.

Приклад 10.3. Реалізувавши вибірку з генеральної сукупності, ознака якої *X* має нормальний закон розподілу, дістали статистичний розподіл:

x_i	6	7	8	9	10	11	12	13	14
n_i	1	3	6	8	6	6	5	3	2

При рівні значущості α = 0,01 перевірити правильність нульової гіпотези

$H_0\colon a = 8$, якщо альтернативна гіпотеза $H_a\colon a \neq 8$.

Розв'язання. Обчислимо значення $\bar{x}$, s:

$$n = 1 + 3 + 6 + 8 + 6 + 6 + 5 + 3 + 2 = 40;$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i = \frac{1}{40}(6 \cdot 1 + 7 \cdot 3 + 8 \cdot 6 + 9 \cdot 8 + 10 \cdot 6 + 11 \cdot 6 + +12 \cdot 5 + 13 \cdot \cdot 3 + 14 \cdot 2) = \frac{6+21+48+72+60+60+60+39+28}{40} = 10;$$

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{m} n_i x_i\right)^2 = \frac{1}{40}(6^2 \cdot 1 + 7^2 \cdot 3 + 8^2 \cdot 6 + 9^2 \cdot 8 + 10^2 \cdot 6 + +11^2 \cdot 6 + 12^2 \cdot 5 + 13^2 \cdot 3 + 1{,}4^2 \cdot 2) - 10^2 = = \frac{36+147+384+648+600+726+720+507+392}{100} - 100 = 104 - 100 = 4;$$

$$S^2 = \frac{n}{n-1} D_{\text{в}} = \frac{40}{39} \cdot 4 = 4{,}103;$$

$$s = \sqrt{S^2} = \sqrt{4{,}103} \approx 2{,}03.$$

При альтернативній гіпотезі $H_a: a \neq 8$ будуємо двобічну критичну область. Враховуючи, що σ_{Γ} є невідомою величиною, для побудови цієї області беремо статистичний критерій.

$$t_{\text{двост кр}}(\alpha; k) = t_{\text{двост кр}}(0{,}01; 39) = 2{,}7.$$

Обчислимо спостережуване значення критерію:

$$T_{\text{спост}} = \frac{(\bar{x} - a_0)\sqrt{n}}{s} = \frac{(10-8)\sqrt{40}}{2{,}03} = \frac{2 \cdot 2\sqrt{10}}{2{,}03} \approx 6{,}23.$$

Оскільки $|T_{\text{спост}}| > t_{\text{двост кр}}$, то немає підстав приймати нульову гіпотезу, тобто гіпотеза $H_0: a = 8$ відхиляється.

10.3. Перевірка рівності виправленої вибіркової дисперсії генеральній дисперсії

Нехай з нормальної генеральної сукупності отримано вибірку об'єму n, для якої знайдено виправлену вибіркову дисперсію S^2.

Правило 1. Для того, щоб при заданому рівні значущості α перевірити нульову гіпотезу $H_0: \sigma^2 = \sigma_0^2$ про рівність невідомої генеральної дисперсії σ^2

гіпотетичному (прогнозованому) значенню σ_0^2 при конкуруючій гіпотезі $H_a : \sigma^2 > \sigma_0^2$, потрібно обчислити спостережуване значення критерію

$$\chi^2_{\text{спост}} = \frac{(n-1)S^2}{\sigma_0^2} \qquad (10.10)$$

і за таблицею критичних точок розподілу χ^2 при заданому рівні значущості α і кількості ступенів вільності $k = n - 1$ знайти критичну точку $\chi^2_{\text{крит}}(\alpha; k)$.

Якщо $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, то немає підстав відхиляти нульову гіпотезу. Якщо $\chi^2_{\text{спост}} > \chi^2_{\text{крит}}$ – нульову гіпотезу відхиляють.

Правило 2. При конкуруючій гіпотезі $H_a: \sigma^2 \neq \sigma_0^2$ знаходять ліву $\chi^2_{\text{лів крит}}(1-\frac{\alpha}{2}; k)$ і праву $\chi^2_{\text{прав крит}}(\frac{\alpha}{2}; k)$ критичні точки.

Якщо $\chi^2_{\text{лів крит}} < \chi^2_{\text{спост}} < \chi^2_{\text{прав крит}}$, то немає підстав відхиляти нульову гіпотезу.

Якщо $\chi^2_{\text{спост}} < \chi^2_{\text{лів крит}}$ або $\chi^2_{\text{спост}} > \chi^2_{\text{прав крит}}$ – нульову гіпотезу відхиляють.

Правило 3. При конкуруючій гіпотезі $H_a: \sigma^2 < \sigma_0^2$ знаходять критичну точку $\chi^2_{\text{крит}}(1-\alpha; k)$.

Якщо $\chi^2_{\text{спост}} > \chi^2_{\text{крит}}$, то немає підстав відхиляти нульову гіпотезу. Якщо $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$ – нульову гіпотезу відхиляють.

Приклад 10.4. Із нормальної генеральної сукупності отримано вибірку об'єму $n = 20$, для якої знайдено вибіркову виправлену дисперсію $S^2 = 2{,}7$. Перевірити при рівні значущості $\alpha = 0{,}05$ основну гіпотезу $H_0: \sigma^2 = \sigma_0^2 = 3$, якщо альтернативна гіпотеза:

А. $H_a: \sigma^2 > 3$;

Б. $H_a: \sigma^2 \neq 3$;

В. $H_a: \sigma^2 < 3$.

Розв'язання. Знайдемо спостережуване значення критерію:

$$\chi^2_{\text{спост}} = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(20-1)\cdot 2{,}7}{3} = 17{,}1.$$

А. Скористаємося правилом 1. За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0{,}05$ і кількості ступенів вільності

$$k = n - 1 = 20 - 1 = 19$$

знаходимо критичну точку:

$$\chi^2_{\text{крит}}(0{,}05; 19) = 30{,}14.$$

Оскільки $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, то немає підстав відхиляти основну гіпотезу про рівність генеральної дисперсії гіпотетичному значенню. Інакше кажучи, різниця між виправленою вибірковою дисперсією $S^2 = 2{,}7$ і гіпотетичною генеральною дисперсією $\sigma_0^2 \leq 3$ незначуща.

Б. Скористаємося правилом 2. За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0{,}05$ і кількості ступенів вільності

$$k = n - 1 = 20 - 1 = 19$$

знаходимо критичні точки:

$$\chi^2_{\text{лів крит}}\left(1 - \frac{\alpha}{2}; k\right) = \chi^2_{\text{лів крит}}(0{,}975; 19) = 8{,}91;$$

$$\chi^2_{\text{прав крит}}\left(\frac{\alpha}{2}; k\right) = \chi^2_{\text{прав крит}}(0{,}025; 19) = 32{,}85.$$

Оскільки $\chi^2_{\text{лів крит}} < \chi^2_{\text{спост}} < \chi^2_{\text{прав крит}}$, то немає підстав відхиляти основну гіпотезу про рівність генеральної дисперсії гіпотетичному значенню.

В. Скористаємося правилом 3. За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0{,}05$ і кількості ступенів вільності

$$k = n - 1 = 20 - 1 = 19$$

знаходимо критичну точку:

$$\chi^2_{\text{крит}}(1 - \alpha; k) = \chi^2_{\text{крит}}(0{,}95; 19) = 10{,}12.$$

Оскільки $\chi^2_{\text{спост}} > \chi^2_{\text{крит}}$, то немає підстав відхиляти основну гіпотезу про рівність генеральної дисперсії гіпотетичному значенню.

10.4. Перевірка правильності нульової гіпотези про рівність двох генеральних середніх (M(X) = M(Y))

Нехай задано дві генеральні сукупності, ознаки яких X і Y мають нормальний закон розподілу і водночас незалежні одна від одної. Необхідно перевірити правдивість $H_0: M(X) = M(Y)$ $(\bar{x}_Г = \bar{y}_Г)$ при відомих значеннях 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_m
n'_i	n'_1	n'_2	n'_3	.....	n'_m

y_j	y_1	y_2	y_3	.....	y_m
n''_j	n''_1	n''_2	n''_3	.....	n''_m

Тут $n' = \sum n'_i,\ n'' = \sum n''_j$.

Обчислюються значення

$$\bar{x} = \frac{\sum_{i=1}^{m} x_i n'_i}{n'}; \quad \bar{y} = \frac{\sum_{j=1}^{m} y_j n''_j}{n''}.$$

За статистичний критерій береться випадкова величина

$$U_{\text{спост}} = \frac{(\bar{x} - \bar{y})}{\sigma(\bar{x} - \bar{y})}, \tag{10.11}$$

що має нормальний закон розподілу.

Оскільки $D(\bar{x} - \bar{y}) = \frac{D_X}{n'} + \frac{D_Y}{n''}$, то статистичний критерій набере такого вигляду:

$$U_{\text{спост}} = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{D_X}{n'} + \frac{D_Y}{n''}}}. \tag{10.12}$$

Залежно від формулювання альтернативної гіпотези H_a будуються відповідно правобічна, лівобічна та двобічна критичні області.

Спостережуване значення критерію, відповідно, обчислюється:

$$U_{\text{спост}} = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{D_X}{n'} + \frac{D_Y}{n''}}}.$$

Приклад 10.5. 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

$H_0\colon M(X) = M(Y)$, якщо альтернативна гіпотеза $H_a : M(X) > M(Y)$.

Розв'язання. Оскільки $n' = \sum n'_i = 100$, $n'' = \sum n''_j = 100$, обчислимо $\bar{x}$ та $\bar{y}$:

$$\bar{x} = \frac{\sum_{i=1}^{m} x_i n'_i}{n'} = \frac{12{,}2 \cdot 5 + 13{,}2 \cdot 15 + 14{,}2 \cdot 40 + 15{,}2 \cdot 30 + 16{,}2 \cdot 10}{100} =$$
$$= \frac{62{,}5 + 198 + 568 + 456 + 162}{100} = \frac{1446{,}5}{100} = 14{,}465;$$

$$\bar{y} = \frac{\sum_{j=1}^{m} y_j n''_j}{n''} = \frac{8{,}4 \cdot 10 + 12{,}4 \cdot 15 + 16{,}4 \cdot 35 + 20{,}4 \cdot 20 + 24{,}4 \cdot 20}{100} =$$
$$= \frac{84 + 186 + 574 + 408 + 488}{100} = \frac{1740}{100} = 17{,}4.$$

Для альтернативної гіпотези $H_a\colon M(X) > M(Y)$ будується правобічна критична область. Критичну точку $u_{\text{кр}}$ знаходимо з рівності

$$\Phi(u_{\text{кр}}) = \frac{1-2\alpha}{2} = \frac{1-2\cdot 0{,}01}{2} = \frac{0{,}98}{2} = 0{,}49.$$

За таблицею визначимо критичну точку:

$$u_{\text{кр}} \approx 2{,}34.$$

Обчислимо спостережуване значення критерію

$$U_{\text{спост}} = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{D_X}{n'} + \frac{D_Y}{n''}}} = \frac{14{,}465 - 17{,}4}{\sqrt{\frac{10}{100} + \frac{15}{100}}} = -\frac{2{,}935}{\sqrt{0{,}1 + 0.15}} = -\frac{2{,}935}{0{,}5} = -5{,}87.$$

Оскільки $U_{\text{спост}} < u_{\text{кр}}$, то $H_0\colon M(X) = M(Y)$ не відхиляється.

10.5. Перевірка правильності нульової гіпотези про рівність двох дисперсій

Одним із важливих завдань математичної статистики є порівняння двох або кількох вибіркових дисперсій. Таке порівняння дає можливість визначити, чи можна вважати вибіркові дисперсії статистичними оцінками однієї і тієї ж дисперсії генеральної сукупності. Воно застосовується насамперед під час обчислення дисперсій за результатами технологічних вимірювань.

Порівняння дисперсій D_X , D_Y здійснюється зіставленням виправлених дисперсій S_x^2, S_y^2, які, відповідно, мають закон розподілу χ^2 із $k_1 = n' - 1$, $k_2 = n'' - 1$ ступенями свободи, де n' і n'' є обсяги першої і другої вибірок.

Нехай перша вибірка здійснена з генеральної сукупності з ознакою *Y*, дисперсія якої дорівнює D_Y, друга – з генеральної сукупності з ознакою *X*, дисперсія якої дорівнює D_X. Необхідно перевірити правильність нульової гіпотези

$$H_0: D_X = D_Y.$$

За статистичний критерій береться випадкова величина

$$F = \frac{S_\delta^2}{S_m^2}, \tag{10.13}$$

яка має розподіл Фішера–Снедекора із k_1 і k_2 ступенями свободи, де S_δ^2 є більшою з виправлених дисперсій, одержаною внаслідок обробки результатів вибірок, S_m^2 є меншою з виправлених дисперсій.

Приклад 10.6. Під час дослідження стабільності температури в термостаті дістали такі результати: 21,2; 21,8; 21,3; 21,0; 21,4; 21,3.

З метою стабілізації температури було використано удосконалений пристрій, після цього заміри температури показали такі результати: 37,7; 37,6; 37,6; 37,4. Чи можна за рівня значущості α = 0,01 вважати використання удосконаленого пристрою до стабілізатора температури ефективним?

Розв'язання. Очевидно, що ефективність стабілізаторів без удосконаленого пристрою і з ним залежить від дисперсій вимірюваних ними температур. Отже, задача звелась до порівняння двох дисперсій.

Обчислимо виправлені вибіркові дисперсії:

$$\bar{y} = \frac{\sum_{j=1}^{m} y_j n_j''}{n''} = \frac{21{,}2+21{,}8+21{,}3+21{,}0+21{,}4+21{,}3}{6} = 21{,}333;$$

$$D_{\text{в}}(Y) = \frac{1}{n''}\sum_{j=1}^{m} n_j'' y_j^2 - \left(\frac{1}{n''}\sum_{j=1}^{m} n_j'' y_j\right)^2 =$$
$$= \frac{21{,}2^2+21{,}8^2+21{,}3^2+21{,}0^2+21{,}4^2+21{,}3^2}{6} - 21{,}333^2 =$$
$$= \frac{2731{,}02}{6} - 455{,}097 = 455{,}17 - 455{,}097 = 0{,}073;$$

$$S_y^2 = \frac{n''}{n''-1} D_{\text{в}}(Y) = \frac{6}{6-1} \cdot 0{,}073 = 0{,}0876;$$

$$\bar{x} = \frac{\sum_{i=1}^{m} x_i n_i'}{n'} = \frac{37{,}7+37{,}6+37{,}6+37{,}4}{4} = \frac{150{,}3}{4} = 37{,}575;$$

$$D_{\text{в}}(X) = \frac{1}{n'}\sum_{i=1}^{m} n_i' x_i^2 - \left(\frac{1}{n'}\sum_{i=1}^{m} n_i' x_i\right)^2 =$$
$$= \frac{37{,}7^2+37{,}6^2+37{,}6^2+37{,}4^2}{4} - 37{,}575^2 =$$
$$= \frac{5647{,}57}{4} - 1411{,}880625 = 1411{,}8925 - 1411{,}880625 = 0{,}011875;$$
$$S_y^2 = \frac{n'}{n'-1} D_{\text{в}}(X) = \frac{4}{4-1} \cdot 0{,}011875 = 0{,}01583.$$

Обчислимо спостережуване значення критерію:

$$F^* = \frac{S_\delta^2}{S_m^2} = \frac{0{,}0876}{0{,}01583} = 5{,}534.$$

Число ступенів свободи:

для більшої виправленої дисперсії $S_\delta^2 = S_y^2,\ k_1 = n'' - 1 = 6 - 1 = 5,$

для меншої $S_m^2 = S_x^2, k_2 = n' - 1 = 4 - 1 = 3.$

Оскільки удосконалення стабілізатора температур може тільки зменшити дисперсію, то будуємо правобічну критичну область. Отже,

$$H_a: S_y^2 > S_x^2.$$

Критичну точку знаходимо за таблицею відповідно до заданого рівня значущості $\alpha = 0{,}01$ і числа ступенів свободи $k_1 = 5$, $k_2 = 3$,

$$F_{\text{кр}}(\alpha;\ k_1, k_2) = F_{\text{кр}}(0{,}01;\ 5, 3) = 28{,}2.$$

Оскільки $F^* < F_{\text{кр}}$, то дані спостережень не дають підстав відхилити нульову гіпотезу, тобто вдосконалення термостабілізатора є ефективним.

Запитання для самоконтролю

1. Статистична гіпотеза – це … ?
2. Які статистичні гіпотези називають параметричними?
3. Які статистичні гіпотези називають непараметричними?
4. Нульовою (основною) називають гіпотезу …?
5. Конкуруючою (альтернативною) називають гіпотезу …?
6. Простою називають гіпотезу …?
7. Складною називають гіпотезу…?
8. Коли виникає помилка першого роду?
9. Коли виникає помилка другого роду?
10. Що таке рівень значущості?
11. Що називають статистичним критерієм перевірки?
12. Що називають спостережним значенням критерію узгодження?
13. Що називають критичною областю?
14. Областю прийняття гіпотези (областю допустимих значень) називають …?
15. Основний принцип перевірки статистичних гіпотез.
16. Критичними точками $k_{кр}$ називають …
17. Правосторонньою називають критичну область …
18. Лівосторонньою називають критичну область …
19. Двосторонньою називають критичну область …
20. Потужністю критерію називають …
21. Загальний алгоритм перевірки правильності нульової гіпотези.
22. Алгоритм перевірки правильності нульової гіпотези про рівність генерального середнього a нормальної сукупності з відомою дисперсією σ^2 гіпотетичному значенню a_0 при конкуруючій гіпотезі $H_a\colon a \neq a_0$; $H_a\colon a < a_0$; $H_a\colon a > a_0$.
23. Алгоритм перевірки правильності нульової гіпотези про рівність генерального середнього a нормальної сукупності з невідомою дисперсією гіпотетичному значенню a_0 при конкуруючій гіпотезі $H_a\colon a \neq a_0$; $H_a\colon a < a_0$; $H_a\colon a > a_0$.
24. Який статистичний критерій вибирається для перевірки основної гіпотези $H_0\colon a = a_0$ при невідомій дисперсії?
25. Який закон розподілу ймовірностей має статистичний критерій $T = \frac{(\bar{x}-a_0)\sqrt{n}}{s}$?
26. Алгоритм перевірки правильності нульової гіпотези про рівність невідомої генеральної дисперсії σ^2 гіпотетичному (прогнозованому) значенню σ_0^2 при конкуруючій гіпотезі $H_a\colon \sigma^2 > \sigma_0^2$; $H_a\colon \sigma^2 \neq \sigma_0^2$; $H_a\colon \sigma^2 < \sigma_0^2$.

27. Алгоритм перевірки правильності гіпотези про рівність двох генеральних середніх при відомих значеннях D_X, D_Y.

28. Який статистичний критерій вибирається для перевірки правильності: $M(X) = M(Y)$?

29. Який закон розподілу ймовірностей має статистичний критерій $U_{\text{спост}} = \frac{(\bar{x}-\bar{y})}{\sigma(\bar{x}-\bar{y})}$?

30. Алгоритм перевірки правильності нульової гіпотези про рівність двох дисперсій.

31. Який статистичний критерій вибирається для перевірки правильності: $H_0: D_X = D_Y$;

32. Який закон розподілу ймовірностей $F = \frac{S_\delta^2}{S_m^2}$?

Задачі для самостійного розв'язування

1. За заданим статистичним розподілом вибірки, реалізованим із генеральної сукупності, ознака *X* якої має нормальний закон розподілу

x_i	4,2	6,2	8,2	10,2	12,2
n_i	6	8	12	8	2

При рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези $H_0: a = 10$, якщо альтернативна гіпотеза $H_a: a > 10$, $\sigma_\Gamma^2 = 16$.

2. Проведено 25 незалежних вимірювань випадкової величини *X*, що має нормальний закон розподілу зі значенням $\sigma_\Gamma^2 = 4$:

x_i	2,4	5,4	8,4	11,4	14,4	17,4
n_i	2	3	10	6	3	1

При рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези $H_0: a = 10{,}5$, якщо альтернативна гіпотеза $H_a: a < 10{,}5$.

3. Маємо дані про розподіл підприємств певної області за зростанням виробітку на одного працівника у відсотках до наступного року:

x_i,%	75	85	95	105	115	125
n_i	5	8	10	5	2	1

Враховуючи, що ознака має нормальний закон розподілу зі значенням $\sigma_{\Gamma}^2 = 36$, перевірити правильність нульової гіпотези при $\alpha = 0{,}01$: $H_0: a = 90$, якщо $H_a: a \neq 90$.

4. У результаті двадцяти незалежних вимірювань певної величини *X* дістали статистичний розподіл:

x_i	3,4	6,4	9,4	12,4	15,4	18,4
n_i	2	4	8	3	2	1

Припускаючи, що випадкова величина *X* має нормальний закон розподілу, при рівні значущості α = 0,01 перевірити правильність $H_0: a = 10$, якщо $H_a: a > 10$.

5. Результати вимірювання зросту дівчаток віком 16 років дали такі показники:

h = 4, см	160–164	164–168	168–172	172–176	176–180
n_i	4	6	20	4	2

Вважаючи, що випадкова величина *X* – зріст дівчаток – має нормальний закон розподілу, при рівні значущості $\alpha = 0{,}001$ перевірити правильність нульової гіпотези $H_0: a = 180$, якщо альтернативна гіпотеза $H_a: a \neq 180$.

6. Вимірювання барометром атмосферного тиску протягом 100 діб дали такі результати:

x_i, мм рт. ст.	744,4	746,4	748,4	750,4	752,4	754,4
n_i	10	20	30	20	15	5

Вважаючи, що *X* – атмосферний тиск – є випадковою величиною, яка має нормальний закон розподілу, при рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези $H_0: a = 749{,}2$, якщо альтернативна гіпотеза $H_a: a < 749{,}2$.

7. Із нормальної генеральної сукупності отримано вибірку об’єму $n = 15$, для якої знайдено вибіркову виправлену дисперсію $S^2 = 14{,}9$. Переві-

рити при рівні значущості $\alpha = 0{,}01$ основну гіпотезу $H_0: \sigma^2 = \sigma_0^2 = 14{,}1$, якщо альтернативна гіпотеза:

а) $H_a: \sigma^2 > 14{,}1$;

б) $H_a: \sigma^2 \neq 14{,}1$;

в) $H_a: \sigma^2 < 14{,}1$.

8. Із нормальної генеральної сукупності отримано вибірку об'єму $n = 17$, для якої знайдено вибіркову виправлену дисперсію $S^2 = 0{,}24$. Перевірити при рівні значущості $\alpha = 0{,}05$ основну гіпотезу $H_0: \sigma^2 = \sigma_0^2 = 0{,}18$, якщо альтернативна гіпотеза:

а) $H_a: \sigma^2 > 0{,}18$;

б) $H_a: \sigma^2 \neq 0{,}18$;

в) $H_a: \sigma^2 < 0{,}18$.

9. Точність роботи верстата-автомата перевіряється по дисперсії контрольованого розміру виробів, яка не повинна перевищувати 0,15. Вибірковому контролю було піддано 25 виробів і за результатами визначена оцінка дисперсії $S^2 = 0{,}25$. Припускається, що розмір виробів – нормально розподілена випадкова величина. Перевірити гіпотезу, що верстат забезпечує необхідну точність.

10. Із нормальної генеральної сукупності отримано вибірку об'єму $n = 31$.

x_i	10,1	10,3	10,6	11,2	11,5	11,8	12,0
n_i	1	3	7	10	6	3	1

Перевірити при рівні значущості $\alpha = 0{,}05$ основну гіпотезу $H_0: \sigma^2 = \sigma_0^2 = 0{,}18$, якщо альтернативна гіпотеза:

а) $H_a: \sigma^2 > 0{,}18$;

б) $H_a: \sigma^2 \neq 0{,}18$;

в) $H_a: \sigma^2 < 0{,}18$.

11. Електролампочки на 220 В виготовлялися двома заводами. З першої партії, виготовленої заводом № 1, здійснили вибірку обсягом $n''_j = 25$, а з другої партії – обсягом $n'_i = 36$. Першу і другу партії електролампочок перевірили на тривалість роботи. Результати перевірки наведено у вигляді статистичних розподілів такого вигляду:

y_j	48	50	52	54	56
n''_j	2	3	14	5	1

x_i	53	56	59	62	65
n'_i	4	6	10	12	4

Відомо, що ознаки Y – тривалість роботи електролампочки першого заводу і X – тривалість роботи електролампочки другого заводу є випадковими величинами, які незалежні між собою і мають нормальний закон розподілу зі значеннями $\sigma_Y = 50$, $\sigma_X = 72$. При рівні значущості $\alpha = 0{,}01$ перевірити правильність $H_0: M(X) = M(Y)$, якщо альтернативна гіпотеза $H_a: M(X) > M(Y)$.

12. Досліджувався місячний прибуток робітників у гривнях двох заводів однієї і тієї ж галузі виробництва. Результати досліджень подано двома статистичними розподілами:

y_j	150,6	160,6	170,6	180,6	190,6
n''_j	12	28	40	18	2

x_i	140,8	160,8	180,8	200,8	220,8
n'_i	2	6	32	8	2

Ознаки X і Y є незалежними і мають нормальний закон розподілу. При рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези $H_0: M(X) = M(Y)$, якщо альтернативна гіпотеза $H_a: M(X) < M(Y)$.

13. Визначався обсяг валової продукції на підприємствах однієї і тієї ж галузі у двох районах України. Результати розрахунків подано двома статистичними розподілами:

y_j, млн грн	380	400	420	440	460
n''_j	5	15	30	40	10

x_i, млн грн	360	400	440	480	500	540
n'_i	10	20	30	20	15	5

Враховуючи, що ознаки X і Y є незалежними і мають нормальний закон розподілу, при рівні значущості $\alpha = 0{,}01$ перевірити правильність $H_0 : M(X) = M(Y)$, якщо альтернативна гіпотеза $H_a : M(X) \neq M(Y)$.

14. Вимірювалось споживання масла за одну добу одним мешканцем у двох регіонах країни. Результати вимірювання подано двома статистичними розподілами:

y_j, мг	15,99	18,99	21,99	21,99	24,99
n''_j	4	6	20	10	5

x_i, мг	14,55	20,55	26,55	30,55	36,55
n'_i	6	14	16	6	4

Ознаки X і Y (добове споживання масла в мг) є незалежними випадковими величинами, які мають нормальний закон розподілу ймовірностей. При рівні значущості $\alpha = 0{,}01$ перевірити правильність нульової гіпотези $H_0: D_X = D_Y$, якщо альтернативна гіпотеза $H_a: D_X > D_Y$.

15. Визначалися річні середні витрати електроенергії на комунально-побутові вимоги для одного мешканця у двох містах. Результати розрахунків подано двома статистичними розподілами для першого і другого міст:

y_j, Вт/м.	700	708	716	724	732	740
n''_j	5	6	9	6	3	1

x_i, Вт/м.	706	710	714	718	722	726	730
n'_i	8	10	12	5	2	2	1

Ознаки X і Y (річні витрати в кВт/особу) є незалежними між собою і мають нормальний закон розподілу. При рівні значущості $\alpha = 0{,}0\,01$ перевірити правильність нульової гіпотези $H_0 : D_X = D_Y$, якщо альтернативна гіпотеза $H_a : D_X < D_Y$.

Розділ 11

СТАТИСТИЧНА ПЕРЕВІРКА ГІПОТЕЗИ ПРО ЗАКОН РОЗПОДІЛУ ГЕНЕРАЛЬНОЇ СУКУПНОСТІ

11.1. Перевірка гіпотези про нормальний розподіл

Нехай статистичний розподіл вибірки задано у вигляді послідовності рівновіддалених варіант і відповідних їм частот:

x_i	x_1	x_2	....	x_m
n_i	n_1	n_2	....	n_m

За допомогою критерію Пірсона перевірити гіпотезу про нормальний розподіл генеральної сукупності.

Правило 1. Для того, щоб при заданому рівні значущості α перевірити гіпотезу про нормальний розподіл генеральної сукупності, необхідно:

1) обчислити вибіркове середнє $\bar{x}$ і вибіркове середнє квадратичне відхилення $\sigma_в$;

2) визначити теоретичні частоти

$$n_i' = \frac{n \cdot h}{\sigma_в} \cdot \varphi(u_i), \qquad (11.1)$$

де n – об'єм вибірки; h – крок (різниця між двома сусідніми варіантами); $\varphi(u)$ – диференціальна функція Лапласа; $u_i = \frac{x_i - \bar{x}}{\sigma_в}$;

3) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона. Для цього:

а) складають розрахункову таблицю

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
1.					

і знаходять спостережуване значення критерію Пірсона

$$\chi^2_{\text{спост}} = \sum_{i=1}^{m} \frac{\left(n_i - n'_i\right)^2}{n'_i}; \tag{11.2}$$

б) за таблицею критичних точок розподілу χ^2 при заданому рівні значущості α і кількості ступенів вільності $k = s - 3$ (s – кількість варіант вибірки) знаходять критичну точку $\chi^2_{\text{крит}}(\alpha; k)$ правосторонньої критичної області. Якщо $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, немає підстав відхиляти гіпотезу про нормальний розподіл генеральної сукупності. Іншими словами, емпіричні та теоретичні частоти відрізняються несуттєво (випадково). Якщо $\chi^2_{\text{спост}} > \chi^2_{\text{крит}}$, гіпотезу відхиляють. Іншими словами, емпіричні та теоретичні частоти відрізняються суттєво.

Нехай статистичний розподіл вибірки задано у вигляді послідовності інтервалів і відповідних їм частот

$(x_i;\ x_{i+1}]$	$(x_1;\ x_2]$	$(x_2;\ x_3]$	.	$(x_m;\ x_{m+1}]$
n_i	n_1	n_2	. . . .	n_m

За допомогою критерію Пірсона перевірити гіпотезу про нормальний розподіл генеральної сукупності.

Правило 2. Для того, щоб при заданому рівні значущості α перевірити гіпотезу про нормальний розподіл генеральної сукупності, необхідно:

1) обчислити вибіркове середнє $\overline{x^*}$ і вибіркове середнє квадратичне відхилення σ^*, причому як варіанти x_i^* беруть середнє арифметичне кінців інтервалу:

$$x_i^* = \frac{x_i + x_{i+1}}{2};$$

2) пронумерувати досліджувану випадкову величину X, тобто перейти до випадкової величини $Z = \frac{X - \overline{x^*}}{\sigma^*}$ і обчислити кінці інтервалів: $z_i = \frac{x_i - \overline{x^*}}{\sigma^*}$,

причому найменше значення Z, тобто z_1, вважають рівним $-\infty$, а найбільше, тобто z_{m+1} – рівним ∞;

3) обчислити теоретичні частоти

$$n'_i = n \cdot P_i, \tag{11.3}$$

де n – об'єм вибірки; $P_i = \Phi(z_{i+1}) - \Phi(z_i)$ – ймовірності потрапляння X в інтервали $(x_i;\ x_{i+1}]$; $\Phi(z)$ – функція Лапласа;

4) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона. Для цього:

А. Складають розрахункову таблицю, за якою знаходять спостережуване значення критерію Пірсона

$$\chi^2_{\text{спост}} = \sum_{i=1}^{m} \frac{\left(n_i - n'_i\right)^2}{n'_i}.$$

Б. За таблицею критичних точок розподілу χ^2 при заданому рівні значущості α і кількості ступенів вільності $k = s - 3$ (s – кількість інтервалів вибірки) знаходять критичну точку $\chi^2_{\text{крит}}(\alpha; k)$ правосторонньої критичної області. Якщо $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, немає підстав відхиляти гіпотезу про нормальний розподіл генеральної сукупності. Тобто, емпіричні та теоретичні частоти відрізняються несуттєво (випадково). Якщо $\chi^2_{\text{спост}} > \chi^2_{\text{крит}}$, гіпотезу відхиляють. Інакше кажучи, емпіричні та теоретичні частоти відрізняються суттєво.

Приклад 11.1. Використовуючи критерій Пірсона, при рівні значущості $\alpha = 0{,}05$ перевірити, чи справджується гіпотеза про нормальний розподіл генеральної сукупності X, якщо з цієї сукупності отримано таку вибірку об'ємом $n = 100$:

$(x_i;\ x_{i+1}]$	(3; 7]	(7; 11]	(11; 15]	(15; 19]	(19; 23]	(23; 27]	(27; 31]
n_i	6	16	19	17	15	14	13

Розв'язання. Спочатку обчислимо вибіркове середнє $\overline{x^*}$ і середнє квадратичне відхилення вибірки σ^*. Для цього перетворимо інтервальний статистичний розподіл вибірки на точковий:

x_i	5	9	13	17	21	25	29
n_i	6	16	19	17	15	14	13

Отже,

$$\overline{x^*} = \frac{5\cdot6+9\cdot16+13\cdot19+17\cdot17+21\cdot15+25\cdot14+29\cdot13}{100} = 17{,}52;$$

$$\sigma^* = \sqrt{\frac{5^2\cdot6+9^2\cdot16+13^2\cdot19+17^2\cdot17+21^2\cdot15+25^2\cdot14+29^2\cdot13}{100} - 17{,}52^2} =$$

$$= \sqrt{\frac{35868}{100} - 306{,}95} = \sqrt{51{,}73} \approx 7{,}19;$$

Знайдемо нормовані інтервали $(z_i; z_{i+1}]$, враховуючи, що вибіркове середнє $\overline{x^*} = 17{,}52$ і середнє квадратичне відхилення вибірки $\sigma^* = 7{,}19$. Для цього складемо розрахункову таблицю (лівий кінець першого інтервалу вважаємо рівним $-\infty$, а правий кінець останнього інтервалу – рівним ∞.

i	x_i	x_{i+1}	$x_i - \overline{x^*}$	$x_{i+1} - \overline{x^*}$	$z_i = \frac{x_i - \overline{x^*}}{\sigma^*}$	$z_{i+1} = \frac{x_{i+1} - \overline{x^*}}{\sigma^*}$
1	3	7	-14,52	-10,52	$-\infty$	-1,46
2	7	11	-10,52	-6,52	-1,46	-0,91
3	11	15	-6,52	-2,5	-0,91	-0,35
4	15	19	-2,5	1,48	-0,35	0,21
5	19	23	1,48	5,48	0,21	0,76
6	23	27	5,48	9,48	0,76	1,32
7	27	31	9,48	13,48	1,32	∞

Визначимо теоретичні ймовірності P_i і теоретичні частоти n_i':

$$n_i' = n \cdot P_i$$

Для цього складемо розрахункову таблицю:

i	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	$P_i = \Phi(z_{i+1}) - \Phi(z_i)$	$n_i' = n \cdot P_i$
1	$-\infty$	-1,46	-0,5000	-0,4279	0,0721	7,21
2	-1,46	-0,91	-0,4279	-0,3186	0,1093	10,93
3	-0,91	-0,35	-0,3186	-0,1368	0,1818	18,18
4	-0,35	0,21	-0,1368	0,0832	0,22	22
5	0,21	0,76	0,0832	0,2764	0,1932	19,32
6	0,76	1,32	0,2764	0,4066	0,1302	13,02
7	1,32	∞	0,4066	0,5000	0,0934	9,34
$\sum$	–	–	–	–	1	100

Порівняємо емпіричні та теоретичні частоти, використовуючи критерій Пірсона. Для цього обчислимо спочатку спостережуване значення критерію Пірсона за допомогою розрахункової таблиці.

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\dfrac{(n_i - n'_i)^2}{n'_i}$	n_i^2	$\dfrac{n_i^2}{n'_i}$
1	6	7,21	-1,21	1,4641	0,2031	36	4,9931
2	16	10,93	5,07	25,7049	2,3518	256	23,4218
3	19	18,18	0,82	0,6724	0,0370	361	19,8570
4	17	22	-5	25	1,1364	289	13,1364
5	15	19,32	-4,32	18,6624	0,9660	225	11,6460
6	14	13,02	0,98	0,9604	0,0738	196	15,0538
7	13	9,34	3,66	13,3956	1,4342	169	18,0942
$\sum$	100	100			$\chi^2_{\text{спост}} = 6{,}2023$	6,2023	106,2023

Перевірка:

$$\chi^2_{\text{спост}} = \sum_{i=1}^{m} \frac{n_i^2}{n'_i} - 100.$$

Оскільки $\sum_{i=1}^{m} \frac{n_i^2}{n'_i} - 100 = 106{,}2023 - 100 = 6{,}2023 = \chi^2_{\text{спост}}$, то обчислення проведено правильно.

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0{,}05$ і кількості ступенів вільності

$$k = s - 3 = 7 - 3 = 4$$

знайдемо критичну точку правосторонньої критичної області:

$$\chi^2_{\text{крит}}(0{,}05; 4) = 9{,}49.$$

Оскільки $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, гіпотезу про нормальний розподіл генеральної сукупності не відхиляємо. Тобто, емпіричні та теоретичні частоти відрізняються несуттєво.

11.2. Перевірка гіпотези про рівномірний розподіл

Нехай задано інтервальний статистичний розподіл вибірки:

$(x_i;\ x_{i+1}]$	$(x_1;\ x_2]$	$(x_2;\ x_3]$		$(x_m;\ x_{m+1}]$
n_i	n_1	n_2	....	n_m

n – об'єм вибірки.

За допомогою критерію Пірсона перевірити гіпотезу про рівномірний розподіл генеральної сукупності.

Правило 1. Для того, щоб перевірити гіпотезу про рівномірний розподіл досліджуваної випадкової величини X, тобто про щільність X:

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a;b]; \\ 0, & x \notin [a;b], \end{cases} \tag{11.4}$$

необхідно:

1) оцінити параметри a і b – кінці інтервалу, у якому спостерігалися можливі значення X, за формулами:

$$a^* = \bar{x} - \sqrt{3}\sigma_{\text{в}}, \qquad b^* = \bar{x} + \sqrt{3}\sigma_{\text{в}}, \tag{11.5}$$

де a^* і b^* – оцінки параметрів);

2) знайти щільність ймовірності передбачуваного розподілу:

$$f(x) = \begin{cases} \frac{1}{b^*-a^*}, & x \in [a^*;b^*]; \\ \\ 0, & x \notin [a^*;b^*], \end{cases} \tag{11.6}$$

3) визначити теоретичні частоти:

$$n'_1 = n \cdot P_1 = n \cdot \frac{1}{b^*-a^*}(x_1 - a^*); \tag{11.7}$$

$$n'_i = n \cdot P_i = n \cdot \frac{1}{b^*-a^*}(x_i - x_{i-1});$$

$$n'_m = n \cdot P_m = n \cdot \frac{1}{b^*-a^*}(b^* - x_m).$$

4) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона, взявши кількість ступенів вільності $k = s - 3$, де s – кількість інтервалів вибірки.

Приклад 11.2. Передбачалося, що про стабільність економічної ситуації в країні (відсутність воєн, стихійних лих і т. д.) за останні 50 років можна судити за характером розподілу населення за віком (n_i – млн чол.): при спокійній обстановці розподіл повинен бути рівномірним. В результаті проведеного дослідження, для однієї з країн були отримані такі дані.

$(x_i;\ x_{i+1}]$	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
n_i	14	9	10	8	16	13	12	18

Чи є підстави вважати, що в країні була нестабільна ситуація? При рівні значущості $\alpha = 0{,}01$ перевірити гіпотезу про рівномірний розподіл та зробити відповідні висновки.

Розв'язання. Спочатку обчислимо вибіркове середнє $\overline{x^*}$ і середнє квадратичне відхилення вибірки σ^*. Для цього перетворимо інтервальний статистичний розподіл вибірки на точковий:

x_i	5	15	25	35	45	55	65	75
n_i	14	9	10	8	16	13	12	18

Отже,

$$\overline{x^*} = \frac{5\cdot14+15\cdot9+25\cdot10+35\cdot8+45\cdot16+55\cdot13+65\cdot12+75\cdot18}{100} = \frac{4300}{100} = 43;$$

$$\sigma^* = \sqrt{\frac{5^2\cdot14+15^2\cdot9+25^2\cdot10+35^2\cdot8+45^2\cdot16+55^2\cdot13+65^2\cdot12+75^2\cdot18}{100} - 43^2} =$$

$$= \sqrt{\frac{242100}{100} - 1849} = \sqrt{2421 - 1849} = \sqrt{572} \approx 23{,}92.$$

Оцінимо параметри a і b – кінців інтервалу, у якому спостерігалися можливі значення X:

$$a^* = \bar{x} - \sqrt{3}\sigma_{\text{в}} = 43 - \sqrt{3} \cdot 23{,}92 = 1{,}57;$$

$$b^* = \bar{x} + \sqrt{3}\sigma_{\text{в}} = 43 + \sqrt{3} \cdot 23{,}92 = 84{,}43.$$

Знайдемо щільність ймовірності передбачуваного розподілу:

$$f(x) = \begin{cases} 0{,}0121, & x \in [a^*; b^*]; \\ 0, & x \notin [a^*; b^*]. \end{cases}$$

Визначимо теоретичні частоти:

$$n_1' = n \cdot P_1 = 100 \cdot 0{,}0121 \cdot (10 - 1{,}57) = 10{,}1;$$
$$n_2' = n \cdot P_2 = 100 \cdot 0{,}0121(20 - 10) = 12{,}1;$$
$$n_3' = n \cdot P_3 = 100 \cdot 0{,}0121(30 - 20) = 12{,}1;$$
$$n_4' = n \cdot P_4 = 100 \cdot 0{,}0121(40 - 30) = 12{,}1;$$
$$n_5' = n \cdot P_5 = 100 \cdot 0{,}0121(50 - 40) = 12{,}1;$$
$$n_6' = n \cdot P_6 = 100 \cdot 0{,}0121(60 - 50) = 12{,}1;$$
$$n_7' = n \cdot P_7 = 100 \cdot 0{,}0121(70 - 60) = 12{,}1;$$
$$n_8' = n \cdot P_8 = 100 \cdot 0{,}0121(84{,}43 - 70) = 17{,}3.$$

Порівняємо емпіричні та теоретичні частоти за допомогою критерію Пірсона, взявши кількість ступенів вільності $k = s - 3$, де s – кількість інтервалів вибірки.

Для цього обчислимо спочатку спостережуване значення критерію Пірсона за допомогою розрахункової таблиці.

i	n_i	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$	n_i^2	$\frac{n_i^2}{n_i'}$
1	14	10,1	3,9	15,21	1,5059	196	19,4059
2	9	12,1	-3,1	9,61	0,7942	81	6,6942
3	10	12,1	-2,1	4,41	0,3645	100	8,2645
4	8	12,1	-4,1	16,81	1,3893	64	5,2893
5	16	12,1	3,9	15,21	1,2570	256	21,157
6	13	12,1	0,9	0,81	0,0669	169	13,9669
7	12	12,1	-0,1	0,01	0,0008	144	11,9008
8	18	17,3	0,7	0,49	0,0283	324	18,7283
$\sum$	100	100			$\chi^2_{\text{спост}} = 5{,}4069$		105,4069

Перевірка:

$$\chi^2_{\text{спост}} = \sum_{i=1}^{m} \frac{n_i^2}{n_i'} - 100.$$

Оскільки $\sum_{i=1}^{m} \frac{n_i^2}{n_i'} - 100 = 105{,}4069 - 100 = 5{,}4069 = \chi^2_{\text{спост}}$, то обчислення проведено правильно.

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0{,}01$ і кількості ступенів вільності

$$k = s - 3 = 8 - 3 = 5$$

знайдемо критичну точку правосторонньої критичної області:

$$\chi^2_{\text{крит}}(0{,}01; 5) = 15{,}09.$$

Оскільки $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, гіпотезу про рівномірний розподіл генеральної сукупності не відхиляємо. Тобто, емпіричні та теоретичні частоти відрізняються несуттєво. Отже, підстави вважати, що в країні була нестабільна ситуація, немає.

11.3. Перевірка гіпотези про показниковий розподіл

Нехай задано інтервальний статистичний розподіл вибірки

$(x_i;\ x_{i+1}]$	$(x_1;\ x_2]$	$(x_2;\ x_3]$		$(x_m;\ x_{m+1}]$
n_i	n_1	n_2	....	n_m

n – об'єм вибірки.

За допомогою критерію Пірсона перевірити гіпотезу про показниковий розподіл генеральної сукупності.

Правило 1. Для того, щоб при рівні значущості α перевірити гіпотезу про показниковий розподіл досліджуваної випадкової величини X, необхідно:

1) знайти вибіркове середнє $\bar{x}$, узявши як варіанти середини інтервалів:

$$x_i^* = \frac{x_i + x_{i+1}}{2};$$

2) за оцінку параметра λ показникового розподілу взяти величину, обернену до вибіркового середнього,

$$\lambda^* = \frac{1}{\bar{x}}, \tag{11.8}$$

3) знайти ймовірності потрапляння випадкової величини X у частинні інтервали $(x_i;\ x_{i+1}]$ за формулою

$$P_i = P(x_i < X < x_{i+1}) = e^{-\lambda^* x_i} - e^{-\lambda^* x_{i+1}}; \tag{11.9}$$

4) обчислити теоретичні частоти

$$n_i' = n \cdot P_i;$$

5) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона, узявши кількість ступенів вільності $k = s - 2$, де s – кількість інтервалів вибірки.

Приклад 11.3. В результаті випробування 200 елементів на тривалість роботи отримано емпіричне розподіл, наведене в таблиці (у першому рядку вказані інтервали часу в годинах, в другому рядку – частоти, тобто кількість елементів, які працювали в межах відповідного інтервалу). Використовуючи критерій Пірсона, при рівні значущості $\alpha = 0{,}05$ перевірити, чи узгоджується гіпотеза про показниковий розподіл генеральної сукупності X з емпіричним розподілом вибірки.

$(x_i;\ x_{i+1}]$	$(0;\ 5]$	$(5;\ 10]$	$(10;\ 15]$	$(15;\ 20]$	$(20;\ 25]$	$(25;\ 30]$
n_i	133	45	15	4	2	1

Розв'язання. Спочатку обчислимо вибіркове середнє $\bar{x}$. Для цього перетворимо інтервальний статистичний розподіл вибірки на точковий:

x_i	2,5	7,5	12,5	17,5	22,5	27,5
n_i	133	45	15	4	2	1

Отже,

$$\bar{x} = \frac{2{,}5\cdot 133 + 7{,}5\cdot 45 + 12{,}5\cdot 15 + 17{,}5\cdot 4 + 22{,}5\cdot 2 + 27{,}5\cdot 1}{200} = \frac{1000}{200} = 5.$$

Оцінимо параметр λ показникового розподілу взявши величину, обернену до вибіркового середнього,

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

Знайдемо ймовірності потрапляння випадкової величини X у частинні інтервали $(x_i;\ x_{i+1}]$ за формулою

$$P_1 = P(0 < X < 5) = e^{-0,2\cdot 0} – e^{-0,2\cdot 5} =$$
$$= 1 – e^{-1} = 1 – 0{,}3679 = 0{,}6321;$$

$$P_2 = P(5 < X < 10) = e^{-0,2\cdot 5} – e^{-0,2\cdot 10} =$$
$$= e^{-1} – e^{-2} = 0{,}3679 – – 0{,}1353 = 0{,}2326;$$

$$P_3 = P(10 < X < 15) = e^{-0,2\cdot 10} – e^{-0,2\cdot 15} =$$
$$= e^{-2} – e^{-3} = 0{,}1353 – – 0{,}0498 = 0{,}0855;$$

$$P_4 = P(15 < X < 20) = e^{-0,2\cdot 15} – e^{-0,2\cdot 20} =$$
$$= e^{-3} – e^{-4} = 0{,}0498 – – 0{,}0183 = 0{,}0315;$$

$$P_5 = P(20 < X < 25) = e^{-0,2\cdot 20} – e^{-0,2\cdot 25} =$$
$$= e^{-4} – e^{-5} = 0{,}0183 – – 0{,}0067 = 0{,}0116;$$

$$P_6 = P(25 < X < 30) = e^{-0,2\cdot 25} – e^{-0,2\cdot 30} =$$
$$= e^{-5} – e^{-6} = 0{,}0067 – – 0{,}0025 = 0{,}0042.$$

Визначимо теоретичні частоти:

$n'_1 = n \cdot P_1 = 200 \cdot 0{,}6321 = 126{,}42;$

$n'_2 = n \cdot P_2 = 200 \cdot 0{,}2326 = 46{,}52;$

$n'_3 = n \cdot P_3 = 200 \cdot 0{,}0855 = 17{,}1;$

$n'_4 = n \cdot P_4 = 200 \cdot 0{,}0315 = 6{,}3;$

$n'_5 = n \cdot P_5 = 200 \cdot 0{,}0116 = 2{,}32;$

$n'_6 = n \cdot P_6 = 200 \cdot 0{,}0042 = 0{,}84.$

Порівняємо емпіричні та теоретичні частоти за допомогою критерію Пірсона, взявши кількість ступенів вільності $k = s - 2$, де s – кількість інтервалів вибірки.

Для цього обчислимо спочатку спостережуване значення критерію Пірсона за допомогою розрахункової таблиці.

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
1	133	126,42	6,58	43,2964	0,3425
2	45	46,52	-1,52	2,3104	0,0497
3	15	17,1	-2,1	4,41	0,2579
4	4	6,3	-2,3	5,29	0,8397
5	2	2,32	-0,32	0,1024	0,0441
6	1	0,84	0,16	0,0256	0,0305
$\sum$					$\chi^2_{\text{спост}} = 1{,}5644$

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0{,}05$ і кількості ступенів вільності

$$k = s - 3 = 6 - 2 = 1$$

знайдемо критичну точку правосторонньої критичної області:

$$\chi^2_{\text{крит}}(0{,}05; 3) = 9{,}49.$$

Оскільки $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, гіпотезу про показниковий розподіл генеральної сукупності не відхиляємо. Інакше кажучи, емпіричні та теоретичні частоти відрізняються несуттєво.

11.4. Перевірка гіпотези про біноміальний розподіл

Нехай проведено n дослідів. Кожний дослід складається із N незалежних випробувань, для яких ймовірність появи деякої події A одна й та ж. Реєструється кількість появ події A в кожному досліді. У підсумку отримаємо статистичний розподіл дискретної випадкової величини X, яка характеризує кількість появ події A (у першому рядку наведено кількість появ події A в одному досліді, а в другому рядку – частоту n_i, тобто кількість дослідів, у яких зареєстровано x_i появ події A):

x_i	0	1	....	N
n_i	n_1	n_2		n_N

За допомогою критерію Пірсона перевірити гіпотезу про біноміальний закон розподілу дискретної випадкової величини X.

Правило 1. Для того, щоб при рівні значущості α перевірити гіпотезу про біноміальний розподіл дискретної випадкової величини X, необхідно:

1) знайти за формулою Бернуллі

$$P_N(i) = C_N^i p^i (1-p)^{N-i} \tag{11.10}$$

ймовірності $P_N(i)$ появи події A рівно i разів у N випробуваннях $(i = 0,1,2,\dots,s)$, де s – максимальна кількість спостережуваних появ події A в одному досліді, тобто $s \le N$);

2) знайти теоретичні частоти

$$n_i' = n \cdot P_N(i);$$

де n – кількість дослідів;

3) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона, поклавши кількість ступенів вільності $k = s - 1$ (вважається, що ймовірність p появи події A задана, тобто вона не оцінювалась за вибіркою і не об'єднувалися малочисельні частоти).

Якщо ймовірність p була оцінена за вибіркою, то $k = s - 2$. Якщо, крім того, було об'єднано малочисельні частоти, то s – кількість варіант вибірки, які залишилися після об'єднання частот.

Приклад 11.4. Чотири монети були підкинуті 20160 разів, водночас комбінації: чотири «герби», три «герби» і «цифра», два «герби» і дві «цифри», один «герб» і три «цифри», чотири «цифри» з'явилися таку відповідну кількість разів:

1181, 4909, 7583, 5085, 1402.

Чи свідчать ці дані про те, що кількість «гербів», яка з'явилась на чотирьох монетах має біноміальний розподіл. Застосувати критерій χ^2 , при рівні значущості $\alpha = 0{,}05$.

Розв'язання. Спочатку знайдемо за формулою Бернуллі (11.10) ймовірності $P_N(i)$ появи події A – кількість випавши гербів. Дані запишемо у таблицю.

$$P_4(4) = C_4^4 \cdot 0{,}5^4 \cdot (1\text{–}0{,}5)^{4-4} = 1 \cdot 0{,}0625 \cdot 1 = 0{,}0625;$$

$$P_4(3) = C_4^3 \cdot 0{,}5^3 \cdot \left(1\text{–}0{,}5\right)^{4-3} = 4 \cdot 0{,}125 \cdot 0{,}5 = 0{,}25;$$

$$P_4(2) = C_4^2 \cdot 0{,}5^2 \cdot (1\text{–}0{,}5)^{4-2} = 6 \cdot 0{,}25 \cdot 0{,}25 = 0{,}375;$$

$$P_4(1) = C_4^1 \cdot 0{,}5^1 \cdot (1\text{–}0{,}5)^{4-1} = 4 \cdot 0{,}5 \cdot 0{,}125 = 0{,}25;$$

$$P_4(0) = C_4^0 \cdot 0{,}5^0 \cdot (1\text{–}0{,}5)^{4-0} = 1 \cdot 1 \cdot 0{,}0625 = 0{,}0625.$$

X	4	3	2	1	0
$P(X)$	0,0625	0,25	0,375	0,25	0,0625

Знайдемо теоретичні частоти, використовуючи формулу:

$$n'_i = n \cdot P_N(i).$$

$n'_1 = 20160 * 0{,}0625 = 1260;$

$n'_2 = 20160 * 0{,}25 = 5040;$

$n'_3 = 20160 * 0{,}375 = 7560;$

$n'_4 = 20160 * 0{,}25 = 5040;$

$n'_5 = 20160 * 0{,}0625 = 1260.$

Порівняємо емпіричні та теоретичні частоти за допомогою критерію Пірсона. Для цього складемо розрахункову таблицю:

i	n_i	n'_i	$n_i\text{–}n'_i$	$(n_i\text{–}n'_i)^2$	$\frac{(n_i\text{–}n'_i)^2}{n'_i}$
0	1181	1260	-79	6241	4,9532
1	4909	5040	-131	17161	3,405
2	7583	7560	23	529	0,0699
3	5085	5040	45	2025	0,4018

4	1402	1260	142	20164	16,0032
5	1181	1260	-79	6241	4,9532
$\sum$					$\chi^2_{\text{спост}} = 24{,}8331$

Із розрахункової таблиці отримаємо

$$\chi^2_{\text{спост}} = 24{,}8331.$$

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0{,}05$ і кількості ступенів вільності

$$k = s - 2 = 5 - 2 = 3$$

знаходимо критичну точку правосторонньої критичної області:

$$\chi^2_{\text{крит}}(0{,}05; 3) = 7{,}82.$$

Оскільки $\chi^2_{\text{спост}} > \chi^2_{\text{крит}}$, є підстави відхиляти гіпотезу про розподіл випадкової величини X (кількості «гербів» за біноміальним законом).

11.5. Перевірка гіпотези про розподіл Пуассона

Нехай задано точковий статистичний розподіл вибірки. За допомогою критерію Пірсона перевірити гіпотезу про розподіл генеральної сукупності за законом Пуассона.

Правило 1. Для того, щоб при рівні значущості α перевірити гіпотезу про те, що досліджувана випадкова величина X розподілена за законом Пуассона, необхідно:

1) знайти за цим статистичним розподілом вибіркове середнє $\bar{x}$;

2) взяти за оцінку параметра λ розподілу Пуассона вибіркове середнє

$$\lambda^* = \bar{x}; \tag{11.11}$$

3) знайти за формулою Пуассона (або за таблицею значень функції Пуассона) ймовірності P_i появи рівно i подій у n випробуваннях

($i = 1, 2, 3, \ldots, r$), де r – максимальна кількість спостережуваних подій; n – об'єм вибірки;

4) визначити теоретичні частоти

$$n_i' = n \cdot P_i;$$

5) порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона, узявши кількість ступенів вільності $k = s - 2$, де s – кількість варіант вибірки (якщо проводилося об'єднання малочисельних частот в одну групу, то s – кількість варіант, які залишилися після об'єднання).

Приклад 11.4. У $n = 1000$ перевірках партій товару реєструвалася кількість x_i неякісної продукції, внаслідок чого було отримано такий статистичний розподіл кількості браку x_i в n_i партіях товару:

x_i	0	1	2	3	4	5	6	7
n_i	242	349	234	107	43	21	3	1

При рівні значущості $\alpha = 0{,}05$ перевірити гіпотезу про те, що кількість неякісної продукції X розподілена за законом Пуассона.

Розв'язання. Спочатку знайдемо вибіркове середнє:

$$\bar{x} = \frac{0\cdot242+1\cdot349+2\cdot234+3\cdot107+4\cdot43+5\cdot21+6\cdot3+7\cdot1}{1000} = \frac{1440}{1000} = 1{,}44.$$

Візьмемо за оцінку параметра λ розподілу Пуассона вибіркове середнє

$$\lambda^* = \bar{x} = 1{,}44.$$

Припускаємо, що закон Пуассона має вигляд:

$$P_{1000}(i) = 1{,}44^i \cdot \frac{e^{-1{,}44}}{i!}.$$

Поклавши $i = 0, 1, 2, \ldots, 7$, обчислимо ймовірності $P_i = P_{1000}(i)$:

$$P_0 = 1{,}44^0 \cdot \frac{e^{-1{,}44}}{0!} \approx 0{,}2369; \qquad P_1 = 1{,}44^1 \cdot \frac{e^{-1{,}44}}{1!} \approx 0{,}3412;$$

$$P_2 = 1{,}44^2 \cdot \frac{e^{-1{,}44}}{2!} \approx 0{,}2456; \qquad P_3 = 1{,}44^3 \cdot \frac{e^{-1{,}44}}{3!} \approx 0{,}1179;$$

$P_4 = 1{,}44^4 \cdot \frac{e^{-1,44}}{4!} \approx 0{,}0424;$ $P_5 = 1{,}44^5 \cdot \frac{e^{-1,44}}{5!} \approx 0{,}0122;$

$P_6 = 1{,}44^6 \cdot \frac{e^{-1,44}}{6!} \approx 0{,}0029;$ $P_7 = 1{,}44^7 \cdot \frac{e^{-1,44}}{7!} \approx 0{,}0006;$

Знайдемо теоретичні частоти $n'_i = n \cdot P_i$:

$n'_0 = 1000 * 0{,}2369 = 236{,}9;$ $n'_1 = 1000 * 0{,}3112 = 311{,}2;$

$n'_2 = 1000 * 0{,}2369 = 236{,}9;$ $n'_3 = 1000 * 0{,}1179 = 117{,}9;$
$n'_4 = 1000 * 0{,}0424 = 42{,}4;$ $n'_5 = 1000 * 0{,}0122 = 12{,}2;$

$n'_6 = 1000 * 0{,}0029 = 2{,}9;$ $n'_7 = 1000 * 0{,}0006 = 0{,}6.$

Оскільки частоти $n_6 = 3$ і $n_7 = 1$ – малочисельні (менше п'яти), об'єднаємо їх із частотою $n_5 = 21$, а саме

$$n_5 = 21 + 3 + 1 = 25.$$

Як теоретичну частоту, що відповідає об'єднаній частоті, візьмемо суму відповідних теоретичних частот:

$$n'_5 = 12{,}2 + 2{,}9 + 0{,}6 = 15{,}7.$$

Порівняємо емпіричні та теоретичні частоти за допомогою критерію Пірсона. Для цього складемо розрахункову таблицю:

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
0	242	236,9	5,1	26,01	0,1098
1	349	341,2	7,8	60,84	0,1784
2	234	245,6	-11,6	134,56	0,5479
3	107	117,9	-10,9	118,81	1,0077
4	43	42,4	0,6	0,36	0,0085
5	25	15,7	9,3	86,49	5,5089
$\sum$	1000				$\chi^2_{\text{спост}} = 7{,}3612$

Із розрахункової таблиці отримаємо

$$\chi^2_{\text{спост}} = 7{,}3612.$$

За таблицею критичних точок розподілу χ^2 при рівні значущості $\alpha = 0{,}05$ і кількості ступенів вільності

$$k = s - 1 = 6 - 2 = 4$$

знаходимо критичну точку правосторонньої критичної області:

$$\chi^2_{\text{крит}}(0{,}05; 4) = 9{,}49.$$

Оскільки $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, то немає підстав відхиляти гіпотезу про розподіл випадкової величини X (кількості неякісного товару в партії) за законом Пуассона.

Запитання для самоконтролю

1. Диференціальна функція Лапласа має вигляд …?
2. Як знайти спостережуване значення критерію Пірсона?
3. Як знайти критичне значення критерію Пірсона $\chi^2_{\text{крит}}$?
4. Якщо $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$, то гіпотеза про нормальний розподіл … ?
5. Якщо $\chi^2_{\text{спост}} > \chi^2_{\text{крит}}$, то гіпотеза про нормальний розподіл … ?
6. $\Phi(z)$ – функція Лапласа має вигляд … ?
7. Як за допомогою критерію Пірсона перевірити гіпотезу про нормальний розподіл генеральної сукупності?
8. Як визначити теоретичні частоти при перевірці гіпотези про нормальний розподіл генеральної сукупності?
9. Як за допомогою критерію Пірсона перевірити гіпотезу про нормальний розподіл, якщо вибірка задана інтервальним розподілом?
10. Як визначити кількість ступенів вільності при перевірці гіпотези про нормальний розподіл за критерієм Пірсона?
11. Як за допомогою критерію перевірити гіпотезу про рівномірний розподіл генеральної сукупності?
12. Як оцінити параметри a і b – кінці інтервалу, у якому спостерігалися можливі значення X (рівномірний розподіл)?
13. Як визначити теоретичні частоти при перевірці гіпотези про рівномірний розподіл?
14. Як визначити кількості ступенів вільності при перевірці гіпотези про рівномірний розподіл за критерієм Пірсона?
15. Як за допомогою критерію Пірсона перевірити гіпотезу про показниковий розподіл генеральної сукупності?
16. Яку величину беруть за оцінку параметра λ показникового розподілу?

17. Як знайти ймовірності потрапляння випадкової величини X у частинні інтервали $(x_i;\ x_{i+1}]$ (показників розподіл)?

18. Як визначити теоретичні частоти при перевірці гіпотези про показниковий розподіл?

19. Як визначити кількість ступенів вільності при перевірці гіпотези про показниковий розподіл за критерієм Пірсона?

20. Як знайти ймовірності $P_N(i)$ появи події A рівно i разів у N випробуваннях?

21. Як визначити теоретичні частоти при перевірці гіпотези про біноміальний розподіл?

22. Як визначити кількість ступенів вільності при перевірці гіпотези про біноміальний розподіл за критерієм Пірсона?

23. Як за допомогою критерію Пірсона перевірити гіпотезу про біноміальний розподіл генеральної сукупності?

24. Яку величину беруть за оцінку параметра λ розподілу Пуассона?

25. Як знайти ймовірності P_i появи рівно i подій у n випробуваннях (розподіл Пуассона)?

26. Як визначити кількість ступенів вільності під час перевірки гіпотези про розподіл Пуассона за критерієм Пірсона?

27. Як за допомогою критерію Пірсона перевірити гіпотезу про розподіл Пуассона генеральної сукупності?

Задачі для самостійного розв’язування

1. Використовуючи критерій Пірсона (χ^2 – квадрат) з рівнем значущості α, перевірити гіпотезу про нормальний розподіл генеральної сукупності X, якщо відомі емпіричні n_i та теоретичні n_i' частоти.

а)	n_i	15	26	25	30	26	21	24	20	13
$\alpha = 0{,}05$	n_i'	9,1	16,5	25,3	32	33,9	29,8	22	13,5	7

б)	n_i	15	26	25	30	26	21	24	20	13
$\alpha = 0{,}01$	n_i'	9,1	16,5	25,3	32	33,9	29,8	22	13,5	7

в) $\alpha = 0{,}05$

n_i	8	16	40	72	36	18	10
n_i'	6	18	36	76	39	18	7

г) $\alpha = 0{,}05$

n_i	5	10	20	8	7
n_i'	6	14	18	7	5

2. Застосовуючи критерій Пірсона, при рівні значимості 0,05 перевірити узгодженість з нормальним розподілом генеральної сукупності X емпіричного розподілу вибірки об'єму $n = 200$:

x_i	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,1	2,3
n_i	6	9	26	25	30	26	21	24	20	8	5

3. За даним інтервальним розподілом вибірки об'єму $n = 100$ при рівні значущості $\alpha = 0{,}05$ за критерієм згоди Пірсона перевірити гіпотезу про нормальний розподіл генеральної сукупності.

$(x_i;\ x_{i+1}]$	(3,8]	(8,13]	(13,18]	(18,23]	(23,28]	(28,33]	(33,38]
n_i	6	8	15	40	16	8	7

4. Таблиця містить дані про зріст 1000 учнів старших класів. Перевірити, чи не протирічать статистичні дані нормальному законові розподілу ($\alpha = 0{,}01$).

Зріст, см	143-146	146-149	149-152	152-155	155-158	158-161	161-164	164-167	167-170	170-173	173-176	176-179	179-182	182-185	185-188
Число учнів	1	2	8	26	65	120	181	201	170	120	64	27	10	3	1

5. За допомогою критерію χ^2 – квадрат перевірити гіпотезу про нормальний закон розподілу у сукупності – розміри деталей після шліфування, на підставі даних, які наводяться в таблиці ($\alpha = 0{,}01$).

$(x_i - x_{i+1}]$	3,6–3,7	3,7–3,8	3,8–3,9	3,9–4,0	4,0–4,1	4,1–4,2	4,2–4,3	4,3–4,4
n_i	1	22	40	79	27	26	4	1

6. Перевірити з допомогою критерію χ^2 – квадрат узгодженість статистичного розподілу з нормальним розподілом при рівні значимості $\alpha = 0{,}01$:

$x_i - x_{i+1}$	0–3	3–6	6–9	9–12	12–15	15–18	18–21	21–24	24–27	27–30
n_i	1	3	4	6	11	10	7	5	2	1

7. Застосовуючи критерій Пірсона χ^2 при рівні значимості $\alpha = 0{,}05$ перевірити, чи узгоджується гіпотеза про нормальний розподіл генеральної сукупності з цим емпіричним розподілом.

$x_i - x_{i+1}$	–20 – –10	– 10 – 0	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
n_i	20	47	80	89	40	16	8

8. Перевірити з допомогою критерію χ^2 узгодженість статистичних даних з рівномірним законом розподілу, рівень значимості $\alpha = 0{,}05$:

$(x_i - x_{i+1}]$	(–1 ; 1]	(1 ; 3]	(3 ; 5]	(5 ; 7]	(7 ; 9]
n_i	10	11	8	9	12

9. Проведено $n = 200$ випробувань, в результаті яких подія A виникла в різні моменти. Отримали емпіричний розподіл. За допомогою критерію Пірсона при $\alpha = 0{,}05$ перевірити гіпотезу про те, що час появи події розподілений рівномірно.

$(x_i - x_{i+1}]$	2 – 4	4 – 6	6 – 8	8 – 10	10 – 12	12 – 14	14 – 16	16 – 18	18 – 20	20 – 22
n_i	21	16	15	26	22	14	21	22	18	25

10. За спостереженнями, наведеними в таблиці, за допомогою критерію χ^2 перевірити згоду з рівномірним розподілом. У першому рядку таблиці вказано ліву границю інтервалу (i – номер інтервалу (i; i + 1)).

а)	i	0	1	2	3	4	5	6	7	8	9	10	11
$\alpha = 0{,}05$	n_i	45	41	34	54	39	43	41	33	37	41	47	39

б) $\alpha = 0{,}1$

i	0	1	2	3	4	5	6	7	8	9
n_i	74	92	83	79	80	73	77	75	76	91

в) $\alpha = 0{,}05$

i	0	1	2	3	4	5	6	7	8	9
n_i	16	15	19	13	14	19	14	11	13	16

11. За спостереженнями, наведеними в таблиці, за допомогою критерію χ^2 перевірити гіпотезу, що випадкова величина має пуассонівський розподіл:

а) $\alpha = 0{,}05$

x_i	0	1	2	3	4
n_i	110	65	21	3	1

б) $\alpha = 0{,}05$

x_i	0	1	2	3	4	5	6	7
n_i	112	168	130	68	32	5	1	1

в) $\alpha = 0{,}01$

x_i	0	1	2	3	4	5	6	7
n_i	8	17	16	10	6	2	0	1

12. В результаті випробування 200 елементів на тривалість роботи отримали емпіричний розподіл. Треба при рівні значимості $\alpha = 0{,}05$ перевірити гіпотезу про те, що час роботи елементів розподілено за показниковим законом.

$x_i - x_{i+1}$	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30
n_i	133	45	15	4	2	1

13. В результаті реєстрації часу приходу 800 відвідувачів виставки отримали статистичний розподіл. При рівні значимості $\alpha = 0{,}01$ перевірити за критерієм χ^2 гіпотезу про те, що час приходу відвідувачів виставки розподілений за показниковим законом.

$x_i - x_{i+1}$	0 – 1	1 – 2	2 – 3	3 – 4	4 – 5	5 – 6	6 – 7	7 – 8
n_i	259	167	109	74	70	47	40	34

14. Дослід, який полягає в одночасному підкиданні чотирьох монет, повторили 100 раз. Емпіричний розподіл дискретної випадкової величини X – число випавших «гербів» вийшов таким:

x_i	0	1	2	3	4
n_i	8	20	42	22	8

При рівні значимості $\alpha = 0{,}05$ перевірити гіпотезу про біноміальний закон за допомогою критерію χ^2 .

15. За заданими статистичними розподілами вибірки висунути H_0 про закон розподілу ознаки генеральної сукупності і при рівні значущості $\alpha = 0{,}01$ перевірити її правильність:

А. Вимірювалась маса новонароджених дітей у пологовому будинку. Результати вимірювань подано інтервальним статистичним розподілом:

x_i, кг, h = 0,2	1 – 1,2	1,2 – 1,4	1,4 – 1,6	1,6 – 1,8	1,8 – 2	2 – 2,2	2,2 – 2,4	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

Б. Протягом 50 років вимірювався рівень води навесні під час повені відносно умовного нуля. Результати вимірювання наведено у вигляді інтервального статистичного розподілу:

x_i, см, h = 24	0–24	24–48	48–72	72–96	96–120	120–146	146–170	170–196	196–220
n_i	1	2	4	6	12	16	6	3	1

В. Вимірювання показників виконання річного плану підприємствами певної галузі наведено у вигляді інтервального статистичного розподілу:

x_i,%, h = 10	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100	100–120
n_i	2	5	13	16	25	12	10	5	3	1

Г. Вимірювався місячний дохід робітника певної галузі. Результати вимірювання подано інтервальним статистичним розподілом:

x_i, грн., h = 20	100–120	120–140	140–160	160–180	180–200	200–220	220–240	240–260	260–280
n_i	10	15	20	25	30	40	10	4	2

16. За наведеними результатами 50-ти вимірів значень деякої неперервної випадкової величини потрібно:

1) згрупувати результати спостережень (побудувати інтервальний статистичний ряд);
2) побудувати гістограму частот та емпіричну функцію розподілу;
3) знайти точкові оцінки числових характеристик;
4) вважаючи, що досліджувана величина має нормальний закон розподілу, обчислити теоретичні частоти та побудувати теоретичну криву розподілу;
5) за критеріями Пірсона та Колмогорова при заданому рівні значущості перевірити гіпотезу про узгодженість емпіричних даних з теоретичними;
6) побудувати довірчі інтервали з певною надійністю для невідомих числових характеристик розподілу.

24	23	20	28	32	28	12	19	18	40
17	39	38	22	24	23	31	22	30	23
15	14	24	30	23	12	34	27	28	22
13	11	24	8	24	26	19	7	24	27
22	29	24	18	26	25	24	15	20	25

$$\gamma = 0{,}95;\ \alpha = 0{,}05.$$

Розділ 12

ЕЛЕМЕНТИ КОРЕЛЯЦІЙНОГО ТА РЕГРЕСІЙНОГО АНАЛІЗУ

Кожній величині, яку отримують у результаті проведення експерименту, притаманний елемент випадковості, що виявляється більшою чи меншою мірою залежно від її природи.

У разі сумісної появи двох і більше величин у результаті проведення експерименту дослідник має підстави для встановлення певної залежності між ними, зв'язку.

Строгої функціональної залежності між змінними, у буквальному розумінні цього слова, у реальному світі не існує, бо вони перебувають під впливом випадкових факторів, наслідки якого передбачити практично неможливо. Тому між змінними існує особлива форма зв'язку, яку називають стохастичною і яка в математичній статистиці трансформується, не змінюючи своєї сутності, у статистичну залежність.

Наприклад, при дослідженні двох змінних X та Y зміна значень $X = x_i$ призводить до такої зміни значень Y, яку можна розбити на два компоненти: систематичну, що пов'язана із залежністю, котра існує між X та Y, і випадкову, яка зазнає впливу випадкових факторів.

Показником, що вимірює стохастичний зв'язок між змінними, є *коефіцієнт кореляції*, який свідчить, з певною мірою ймовірності, наскільки зв'язок між змінними близький.

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

$$y = a_0 + a_1 x; \tag{12.1}$$

$$y = a_0 + a_1 x + a_2 x^2; \tag{12.2}$$

$$y = a_0 + \frac{a_1}{x}. \tag{12.3}$$

Наведені можливі залежності між змінними X і Y (12.1), (12.2), (12.3) називають *функціями регресії*. Форму зв'язку між змінними X і Y можна встановити, застосовуючи кореляційні поля. Кожній точці з координатами x_i, y_i відповідає певне числове значення ознак X та Y.

Отже, на основі розміщення точок кореляційного поля дослідник має підстави для гіпотетичного припущення про лінійні чи нелінійні залежності між ознаками X і Y.

Припустимо, що нам відома функціональна залежність між випадковими величинами Y та X вигляду

$$Y = f(X; a_1; a_2; \ldots; a_m) \tag{12.4}$$

з невідомими параметрами $a_1; a_2; \ldots; a_m$.

Нехай внаслідок n незалежних випробувань одержані варіанти ознак Y та X, які оформлені у статистичній таблиці вигляду:

X	x_1	x_2	…	x_k	…	x_n
Y	y_1	y_2	…	y_k	…	y_n

Для знаходження оцінок параметрів функціональної залежності $a_1; a_2; \ldots; a_m$ за даними вибірки застосуємо метод найменших квадратів. Цей метод ґрунтується на тому, що найімовірніші значення параметрів $a_1; a_2; \ldots; a_m$ повинні давати мінімум функції

$$S = \sum_{k=1}^{n}\left[y_k - f(x_k; a_1; a_2; \ldots; a_m)\right]^2 \tag{12.5}$$

Якщо функція $f(x_k; a_1; a_2; \ldots; a_m)$ має неперервні частинні похідні відносно невідомих параметрів $a_1; a_2; \ldots; a_m$, то необхідною умовою існування мінімуму функції S буде система m рівнянь з m невідомими

$$\frac{\partial S}{\partial a_k} = 0, \quad k = 1, 2, \ldots, m. \tag{12.6}$$

Знаходження функціональної залежності між випадковими величинами X та Y з використанням даних випробувань (або вибірки) називають вирівнюванням емпіричних даних вздовж кривої $y = f(x; a_1; a_2; \ldots; a_m)$.

12.1. Рівняння лінійної парної регресії

Нехай між змінними X та Y теоретично існує певна лінійна залежність. Це твердження може ґрунтуватися на тій підставі, наприклад, що кореляційне поле для пар $(x_i; y_i)$ має такий вигляд (рис. 12.1).

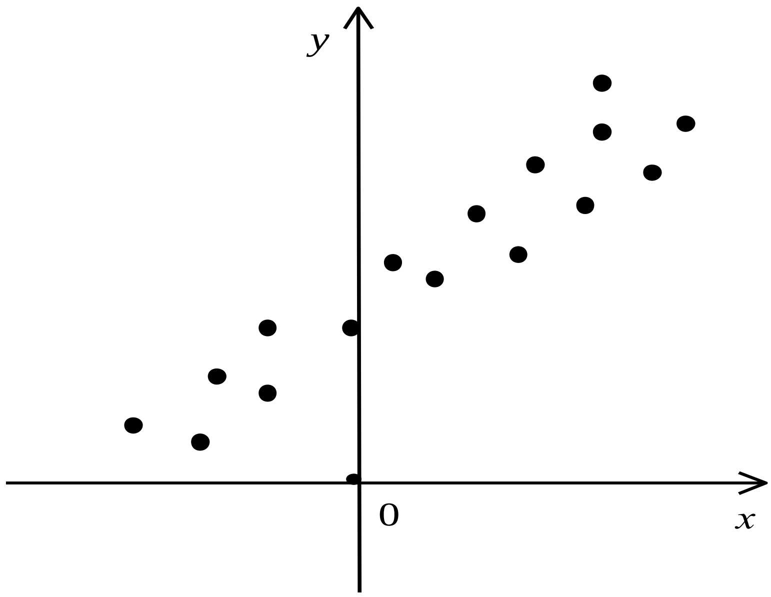

Рис. 12.1. Кореляційне поле для пар

Як бачимо, насправді між ознаками X та Y спостерігається не такий тісний зв'язок, як це передбачає функціональна залежність.

Окремі спостережувані значення y, як правило, відхилятимуться від передбаченої лінійної залежності під впливом випадкових збудників, які здебільшого є невідомими. Відхилення від передбаченої лінійної форми зв'язку можуть статися внаслідок неправильної специфікації рівняння, тобто ще з початку неправильно вибране рівняння, що описує залежність між X і Y.

Будемо вважати, що специфікація рівняння вибрана правильно. Враховуючи вплив на значення Y збурювальних випадкових факторів, лінійне рівняння зв'язку X і Y можна подати в такому вигляді:

$$y = a_0 + a_1 x + \varepsilon, \qquad (12.7)$$

де a_0, a_1 є невідомі параметри регресії, ε є випадковою змінною, що характеризує відхилення y від гіпотетичної теоретичної регресії.

Отже, в рівнянні (12.7) значення «y» подається у вигляді суми двох частин: систематичної $a_0 + a_1 x$ і випадкової ε. Параметри a_0, a_1 є невідомими величинами, а ε є випадковою величиною, що має нормальний закон розподілу з числовими характеристиками: $M(\varepsilon) = 0; D(\varepsilon) = const$.

У результаті статистичних спостережень дослідник дістає характеристики для незалежної змінної x і відповідні значення залежної змінної y.

Отже, необхідно визначити параметри a_0, a_1. Але істинні значення цих параметрів дістати неможливо, оскільки ми користуємося інформацією, здобутою від вибірки обмеженого обсягу. Тому знайдені значення параметрів будуть лише статистичними оцінками істинних (невідомих нам) параметрів a_0, a_1.

На практиці найчастіше оцінки невідомих параметрів a_0, a_1 визначаються за методом найменших квадратів, розробка якого належить К. Гауссу і П. Лапласу. Цей метод почали широко застосовувати в економіко-статистичних обчисленнях, відколи була створена теорія регресії.

В основі методу найменших квадратів є принцип мінімізації суми квадратів залишків моделі.

Згідно з формулою (12.5) маємо

$$S = \sum_{k=1}^{n}\left[y_k - (a_0 + a_1 x_k)\right]^2 \qquad (12.8)$$

Ця функція S неперервно диференційовна, тому згідно з необхідними умовами існування мінімуму S повинні виконуватись рівності $\frac{\partial S}{\partial a_0} = 0$ та $\frac{\partial S}{\partial a_1} = 0$.

У нашому випадку ці рівності мають вигляд:

$$\begin{cases} \sum_{k=1}^{n}\left[y_k - (a_0 + a_1 x_k)\right] = 0; \\ \sum_{k=1}^{n}\left[y_k - (a_0 + a_1 x_k)\right] \cdot x_k = 0; \end{cases}$$

Реалізація цього принципу дає можливість отримати систему нормальних рівнянь:

$$\begin{cases} n a_0 + a_1 \sum_{k=1}^{n} x_k = \sum_{k=1}^{n} y_k; \\ a_0 \sum_{k=1}^{n} x_k + a_1 \sum_{k=1}^{n} x_k^2 = \sum_{k=1}^{n} x_k y_k . \end{cases} \qquad (12.9)$$

У цій системі n – кількість спостережень, $\sum_{k=1}^{n} x_k$, $\sum_{k=1}^{n} y_k, \sum_{k=1}^{n} x_k^2$, $\sum_{k=1}^{n} x_k y_k$ – величини, які можна розрахувати на основі вихідних спостережень над змінними Y і X.

Розв’язавши систему нормальних рівнянь, одержимо оцінки невідомих параметрів моделі $\hat{a}_0$ і $\hat{a}_1$:

$$\hat{a}_1 = \frac{\left(\sum_{k=1}^{n} x_k\right)\cdot\left(\sum_{k=1}^{n} y_k\right) - n\cdot\sum_{k=1}^{n} x_k y_k}{\left(\sum_{k=1}^{n} x_k\right)^2 - n\cdot\sum_{k=1}^{n} x_k^2}; \qquad (12.10)$$

$$\hat{a}_0 = \frac{\left(\sum_{k=1}^{n} x_k\right)\cdot\left(\sum_{k=1}^{n} x_k y_k\right) - \left(\sum_{k=1}^{n} x_k^2\right)\cdot\left(\sum_{k=1}^{n} y_k\right)}{\left(\sum_{k=1}^{n} x_k\right)^2 - n\cdot\sum_{k=1}^{n} x_k^2}. \qquad (12.11)$$

Тоді рівняння прямої регресії матиме вигляд:

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 x. \qquad (12.12)$$

Якщо кількість значень x_k та y_k велика, то обчислення параметрів $\hat{a}_0$ та $\hat{a}_1$ за формулами (12.10; 12.11) ускладнюється. Для спрощення обчислень поділимо два рівняння системи (12.9) на n. Отримаємо:

$$\begin{cases} a_0 + a_1 \dfrac{\sum_{k=1}^{n} x_k}{n} = \dfrac{\sum_{k=1}^{n} y_k}{n}; \\ a_0 \dfrac{\sum_{k=1}^{n} x_k}{n} + a_1 \dfrac{\sum_{k=1}^{n} x_k^2}{n} = \dfrac{\sum_{k=1}^{n} x_k y_k}{n}; \end{cases}$$

або

$$\begin{cases} a_0 + a_1 \bar{x} = \bar{y}; \\ a_0 \bar{x} + a_1 \overline{x^2} = \overline{xy}. \end{cases} \qquad (12.13)$$

Розв'язавши систему рівнянь (12.13), отримали:

$$\hat{a}_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2}; \qquad (12.14)$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}. \qquad (12.15)$$

Коваріацією вибірки називається величина:

$$K(X;Y) = \overline{xy} - \bar{x} \cdot \bar{y}. \qquad (12.16)$$

Коефіцієнтом кореляції називають величину

$$r = \frac{K(X;Y)}{\sigma_x \cdot \sigma_y}, \qquad (12.17)$$

де $\sigma_x = \overline{x^2} - (\bar{x})^2$ – середньоквадратичне відхилення змінної *X*, $\sigma_y = \overline{y^2} - (\bar{y})^2$ – середньоквадратичне відхилення змінної Y.

Коефіцієнт кореляції оцінює залежність між *X* та *Y* і має такі властивості:

1. $-1 \leq r \leq 1$ – залежність між *X* та *Y* тим сильніша, чим значення r ближче до 1. Якщо $r > 0$, то із зростанням *X* зростає і *Y* (кореляція додатна). Якщо $r < 0$, то при зростанні *X* величина *Y* у середньому спадає (кореляція від'ємна).

2. $r = \pm 1$ тоді і лише тоді, коли *Y* є лінійною функцією від *X* і навпаки.

3. Якщо статистичні змінні X та Y незалежні, то $r = 0$ ($K(X;Y) = 0$). Такі змінні X та Y, для яких $r = 0$, називають некорельованими, а для яких $r \neq 0$ – корельованими.

Коефіцієнт кореляції служить для оцінки тісноти лінійного зв'язку між випадковими змінними X та Y: чим $|r|$ ближче до 1, тим зв'язок сильніший, чим ближче $|r|$ до нуля – слабший. Від'ємний знак свідчить про обернений зв'язок, додатній – про прямий.

Коефіцієнтом детермінації називається величина R^2, яка визначається за формулою:

$$R^2 = r^2. \tag{12.18}$$

На основі коефіцієнта детермінації R^2 можна зробити висновок про ступінь значущості вимірюваного зв'язку на основі лінійної регресії. $R^2 \epsilon [0;1]$.

Оскільки коефіцієнт детермінації R^2 характеризує, якою мірою варіація залежної змінної визначається варіацією незалежної змінної, то що ближче R^2 до одиниці, то суттєвішим є зв'язок між цими змінними.

Приклад 12.1. Залежність обсягу отриманого прибутку деяким умовним підприємством регіону від вартості основних виробничих фондів наведено парним статистичним розподілом вибірки:

Основні фонди, млн грн, x_k	2,5	2,8	3	3,2	3,5	4,2	4,5	5	5,3	6
Прибуток, млн грн, y_k	1,2	1,5	1,7	2,2	2,6	3,1	3,4	4,2	4,7	5,4

Методом найменших квадратів визначити оцінки невідомих параметрів лінійної парної регресії. Обчислити коефіцієнт кореляції та детермінації, зробити висновки.

Розв'язання. З таблиці бачимо, що зі збільшенням значень ознаки X залежна зміна Y має тенденцію до збільшення.

Тому припускаємо, що між ознаками X та Y існує лінійна функціональна залежність

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 x.$$

Для визначення параметрів $\hat{a}_0$ та $\hat{a}_1$ скористаємося розрахунковою таблицею, що має такий вигляд:

№ з/п	x_k	y_k	x_k^2	$x_k y_k$	y_k^2
1.	2,5	1,2	6,25	3,0	1,44
2.	2,8	1,5	7,84	4,2	2,25
3.	3	1,7	9	5,1	2,89
4.	3,2	2,2	10,24	7,0	4,84
5.	3,5	2,6	12,25	9,1	6,76
6.	4,2	3,1	17,64	13,0	9,61
7.	4,5	3,4	20,25	15,3	11,56
8.	5	4,2	25	21,0	17,64
9.	5,3	4,7	28,09	24,9	22,09
10	6	5,4	36	32,4	29,16
$\sum$	**40**	**30**	**172,56**	**135,07**	**108,24**

Скориставшись формулами (12.10; 12.11), де $n = 10$, отримаємо:

$$\hat{a}_1 = \frac{(\sum_{k=1}^{n} x_k)\cdot(\sum_{k=1}^{n} y_k) - n\cdot\sum_{k=1}^{n} x_k y_k}{(\sum_{k=1}^{n} x_k)^2 - n\cdot\sum_{k=1}^{n} x_k^2} = \frac{40\cdot 30 - 10\cdot 135,07}{40^2 - 10\cdot 172,56} = 1,12;$$

$$\hat{a}_0 = \frac{(\sum_{k=1}^{n} x_k)\cdot(\sum_{k=1}^{n} x_k y_k) - (\sum_{k=1}^{n} x_k^2)\cdot(\sum_{k=1}^{n} y_k)}{(\sum_{k=1}^{n} x_k)^2 - n\cdot\sum_{k=1}^{n} x_k^2} = \frac{40\cdot 135,07 - 172,56\cdot 30}{40^2 - 10\cdot 172,56} = -1,799.$$

Отже, рівняння регресії буде таким:

$$\hat{Y} = -1,799 + 1,12x.$$

Для обчислення r необхідно знайти $K(X;Y)$, σ_x, σ_y:

$$K(X;Y) = \overline{xy} - \bar{x}\cdot\bar{y} = 13,507 - 4\cdot 3 = 1,507;$$

$$\sigma_x = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{17,256 - 4^2} = \sqrt{1,256} = 1,12;$$

$$\sigma_y = \sqrt{\overline{y^2} - (\bar{y})^2} = \sqrt{10,824 - 3^2} = \sqrt{1,824} = 1,35;$$

$$r = \frac{K(X;Y)}{\sigma_x\cdot\sigma_y} = \frac{1,507}{1,12\cdot 1,35} = 0,996;$$

$$R^2 = r^2 = 0,996^2 = 0,992.$$

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

Коефіцієнт детермінації $R^2 = 0{,}992$. Це означає, що зміна обсягу прибутку підприємства на 99,2% визначається варіацією вартості основних фондів, і 0,8% – іншими випадковими факторами.

12.2. Парна нелінійна регресія

Якщо відображені на площині XOY групи точок $(x_i;\ y_i)$, розміщуються, нагадуючи деякі криві, то доцільно вважати, що між досліджуваними величинами існує нелінійна залежність. Тепер знову виникло завдання підібрати таку криву, яка б на основі методу найменших квадратів мала найменші відхилення від точок, здобутих при спостереженні, знайти її рівняння і визначити тісноту зв'язку.

Розглянемо деякі найпростіші види нелінійної кореляційної залежності.

Нехай зі зростанням однієї випадкової величини умовні середні другої зростають (спадають), досягають максимуму (мінімуму), а потім спадають (зростають). Тоді можна вважати, що між ними існує ***параболічна залежність*** виду:

$$y = a_0 + a_1 x + a_2 x^2 + \varepsilon. \tag{12.19}$$

Методом найменших квадратів на основі даних випробувань необхідно оцінити невідомі параметри a_0, a_1, a_2. Для цієї залежності формула (12.5) матиме вигляд:

$$S = \sum_{k=1}^{n} \left[y_k - (a_0 + a_1 x_k + a_2 x_k^2) \right]^2. \tag{12.20}$$

Необхідні умови існування мінімуму функції S є рівності нулю частинних похідних першого порядку: $\frac{\partial S}{\partial a_0} = 0; \frac{\partial S}{\partial a_1} = 0; \frac{\partial S}{\partial a_2} = 0.$

У нашому випадку:

$$\begin{cases} -2\sum_{k=1}^{n} \left[y_k - (a_0 + a_1 x_k + a_2 x_k^2) \right] x_k^2 = 0; \\ -2\sum_{k=1}^{n} \left[y_k - (a_0 + a_1 x_k + a_2 x_k^2) \right] x_k = 0; \\ -2\sum_{k=1}^{n} \left[y_k - (a_0 + a_1 x_k + a_2 x_k^2) \right] = 0. \end{cases}$$

Реалізація цього принципу дає можливість отримати систему нормальних рівнянь:

$$\begin{cases} a_2 \sum_{k=1}^{n} x_k^4 + a_1 \sum_{k=1}^{n} x_k^3 + a_0 \sum_{k=1}^{n} x_k^2 = \sum_{k=1}^{n} x_k^2 y_k\,; \\ a_2 \sum_{k=1}^{n} x_k^3 + a_1 \sum_{k=1}^{n} x_k^2 + a_0 \sum_{k=1}^{n} x_k = \sum_{k=1}^{n} x_k y_k; \\ a_2 \sum_{k=1}^{n} x_k^2 + a_1 \sum_{k=1}^{n} x_k + n \cdot a_0 = \sum_{k=1}^{n} y_k. \end{cases} \tag{12.21}$$

Система (12.21) є неоднорідною лінійною системою трьох рівнянь з невідомими a_0, a_1, a_2. Розв'язок цієї системи можна знайти різними методами (матричним, за правилом Крамера, методом Гаусса, а його вигляд буде громіздкий при доволі великій кількості випробувань n).

Розв'язавши систему нормальних рівнянь, одержимо оцінки невідомих параметрів моделі $\hat{a}_0$; $\hat{a}_1$; $\hat{a}_2$.

Тоді рівняння регресії матиме вигляд:

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 x^2. \tag{12.22}$$

Приклад 12.2. За даними (див. табл.) про господарську діяльність десяти підприємств побудувати парну кореляційно-регресійну модель залежності обсягів виробництва (y – показник) від основних виробничих фондів (x – фактор). Встановити форму зв'язку та математичне рівняння зв'язку шляхом побудови графіка кореляційної залежності. Знайти оцінки параметрів рівняння парної параболічної регресії $\hat{a}_0$; $\hat{a}_1$; $\hat{a}_2$.

x	2,8	4,2	5	5,4	5,8	7	8,2	9	10	10,2
y	7	9	9,8	11	10	13	12	11,4	8	6,6

Розв'язання. З таблиці бачимо, що зі збільшенням значень ознаки X залежна зміна Y має тенденцію до зростання, досягає максимуму, а потім спадає.

Тому припускаємо, що між ознаками X та Y існує параболічна функціональна залежність:

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 x^2.$$

Для визначення параметрів $\hat{a}_0$, $\hat{a}_1$ та $\hat{a}_2$ скористаємося розрахунковою таблицею, що має такий вигляд:

№ з/п	x_k	y_k	x_k^2	$x_k y_k$	y_k^2	x_k^3	x_k^4	$x_k^2 y_k$
1.	2,8	7	7,84	19,6	49	21,952	61,4656	54,88
2.	4,2	9	17,64	37,8	81	74,088	311,17	158,76
3.	5	9,8	25	49	96,04	125	625	245
4.	5,4	11	29,16	59,4	121	157,464	850,306	320,76
5.	5,8	10	33,64	58	100	195,112	1131,65	336,4
6.	7	13	49	91	169	343	2401	637
7.	8,2	12	67,24	98,4	144	551,368	4521,22	806,88
8.	9	11,4	81	102,6	129,96	729	6561	923,4
9.	10	8	100	80	64	1000	10000	800
10.	10,2	6,6	104,04	67,32	43,56	1061,21	10824,3	686,664
$\sum$	**67,6**	**97,8**	**514,56**	**663,12**	**997,56**	**4258,19**	**37287,1**	**4969,74**

Підставивши у систему (12.21) дані з розрахункової таблиці, отримаємо:

$$\begin{cases} 37287{,}1a_2 + 4258{,}19a_1 + 514{,}56a_0 = 4969{,}74; \\ 4258{,}19a_2 + 514{,}56a_1 + 67{,}6a_0 = 663{,}12; \\ 514{,}56a_2 + 67{,}6a_1 + 10a_0 = 97{,}8. \end{cases}$$

Розв'язавши систему нормальних рівнянь, одержимо оцінки невідомих параметрів моделі $\hat{a}_0$; $\hat{a}_1$; $\hat{a}_2$.

$$\hat{a}_0 = -4{,}78;$$

$$\hat{a}_1 = 4{,}87;$$

$$\hat{a}_2 = -0{,}36.$$

Отже, рівняння регресії буде таким:

$$\hat{Y} = -4{,}78 + 4{,}87x - 0{,}36x^2.$$

Нехай зі зростанням однієї випадкової величини умовні середні другої спадають, але не на ту ж величину, як це буває в разі лінійної залежності, а розмір зміни ніби згасає. У такому разі можна вважати, що залежність ***гіперболічна***:

$$y = a_0 + \frac{a_1}{x} + \varepsilon. \tag{12.23}$$

Методом найменших квадратів на основі даних випробувань необхідно оцінити невідомі параметри a_0, a_1. Для цієї залежності формула (12.5) матиме вигляд:

$$S = \sum_{k=1}^{n} \left[y_k - \left(a_0 + \frac{a_1}{x_k} \right) \right]^2. \tag{12.24}$$

Необхідні умови існування мінімуму функції S є рівності нулю частинних похідних першого порядку: $\frac{\partial S}{\partial a_0} = 0; \frac{\partial S}{\partial a_1} = 0.$

У нашому випадку:

$$\begin{cases} \sum\limits_{k=1}^{n} \left[y_k - \left(a_0 + \dfrac{a_1}{x_k} \right) \right] = 0; \\ \sum\limits_{k=1}^{n} \left[y_k - (a_0 + a_1 \dfrac{1}{x_k}) \right] \cdot \dfrac{1}{x_k} = 0. \end{cases}$$

Реалізація цього принципу дає можливість отримати систему нормальних рівнянь:

$$\begin{cases} n a_0 + a_1 \sum_{k=1}^{n} \frac{1}{x_k} = \sum_{k=1}^{n} y_k; \\ a_0 \sum_{k=1}^{n} \frac{1}{x_k} + a_1 \sum_{k=1}^{n} \frac{1}{x_k^2} = \sum_{k=1}^{n} \frac{y_k}{x_k}. \end{cases} \tag{12.25}$$

Система (12.25) є неоднорідною лінійною системою двох рівнянь з невідомими a_0, a_1.

Розв'язавши систему нормальних рівнянь, одержимо оцінки невідомих параметрів моделі $\hat{a}_0$; $\hat{a}_1$.

Тоді рівняння регресії матиме вигляд:

$$\hat{Y} = \hat{a}_0 + \frac{\hat{a}_1}{x}. \tag{12.26}$$

Приклад 12.3. Підприємства легкої промисловості регіону отримали інформацію, що характеризує залежність обсягу випуску продукції (Y, млн грн) від обсягу капіталовкладень (X, млн грн). Установити форму залежності між X і Y, знайти рівняння регресії і оцінити тісноту зв'язку.

x_k	3	7	7	10	12	14	17	20	21	22
y_k	13	15	19	22	21	20	26	30	26	27

Розв'язання. Припускаємо, що між ознаками X та Y існує гіперболічна функціональна залежність

$$\hat{Y} = \hat{a}_0 + \frac{\hat{a}_1}{x}.$$

Для визначення параметрів $\hat{a}_0$ та $\hat{a}_1$ скористаємося розрахунковою таблицею, що має такий вигляд:

№ з/п	x_k	y_k	$\frac{1}{x_k}$	$\frac{1}{x_k^2}$	$\frac{y_k}{x_k}$	y_k^2
1.	3	13	0,333333	0,111111	4,333333	169
2.	7	15	0,142857	0,020408	2,142857	225
3.	7	19	0,142857	0,020408	2,714286	361
4.	10	22	0,1	0,01	2,2	484
5.	12	21	0,083333	0,006944	1,75	441
6.	14	20	0,071429	0,005102	1,428571	400
7.	17	26	0,058824	0,00346	1,529412	676
8.	20	30	0,05	0,0025	1,5	900
9.	21	26	0,047619	0,002268	1,238095	676
10.	22	27	0,045455	0,002066	1,227273	729
$\sum$	133	219	1,075707	0,184268	20,06383	5061

Скориставшись формулами (12.25), отримаємо систему нормальних рівнянь:

$$\begin{cases} 10a_0 + 1{,}075707a_1 = 219; \\ 1{,}075707a_0 + 0{,}184268a_1 = 20{,}06383. \end{cases}$$

Розв'язавши цю систему нормальних рівнянь, одержимо оцінки невідомих параметрів моделі $\hat{a}_0$; $\hat{a}_1$.

$$\hat{a}_0 = 27{,}38;$$

$$\hat{a}_1 =- 50{,}97.$$

Отже, рівняння регресії буде таким:

$$\hat{Y} = 27{,}38 - \frac{50{,}97}{x.}.$$

Для обчислення r необхідно знайти $K(X;Y)$, σ_x, σ_y:

$$K(X;Y) = \overline{\left(\frac{y}{x}\right)} - \frac{1}{\bar{x}} \cdot \bar{y} = 2{,}006383 - 0{,}1075707 \cdot 21{,}9 =- 0{,}3494;$$

$$\sigma_x = \sqrt{\overline{\left(\frac{1}{x^2}\right)} - \overline{\left(\frac{1}{x}\right)}^2} = \sqrt{0{,}184268 - 0{,}1075707^2} = \sqrt{0{,}0068553} = 0{,}082797;$$

$$\sigma_y = \sqrt{\overline{y^2} - (\bar{y})^2} = \sqrt{506{,}1 - 21{,}9^2} = \sqrt{26{,}49} = 5{,}146844;$$

$$r = \frac{K(X;Y)}{\sigma_x \cdot \sigma_y} = \frac{-0{,}3494}{0{,}082797 \cdot 5{,}146844} \approx - 0{,}82.$$

Як бачимо, коефіцієнт кореляції $|r|$ близький до одиниці, що свідчить про те, що зв'язок між Xта Y є тісним.

Нехай зі зростанням однієї випадкової величини умовні середні другої зростають (спадають), але не на ту ж величину, як це буває в разі лінійної залежності, а розмір зміни ніби прискорюється (згасає). У такому разі можна вважати, що залежність ***показникова***:

$$y = a_0 \cdot a_1{}^x + \varepsilon. \tag{12.27}$$

Для оцінки невідомих параметрів a_0, a_1 рівняння показникової регресії, насамперед зведемо її до лінійного вигляду прологарифмувавши ліву і праву частину рівняння (12.27).

$$\ln y = \ln(a_0 \cdot a_1{}^x) \quad \Rightarrow \qquad \ln y = \ln a_0 + x \ln a_1.$$

Введемо нові змінні $y' = \ln y$, $b_0 = \ln a_0$; $b_1 = \ln a_1$. Тоді рівняння регресії матиме вигляд:

$$y' = b_0 + b_1 \cdot x. \tag{12.28}$$

Методом найменших квадратів оцінимо невідомі параметри b_0, b_1, розв'язавши систему нормальних рівнянь:

$$\begin{cases} nb_0 + b_1 \sum_{k=1}^{n} x_k = \sum_{k=1}^{n} y'_k; \\ b_0 \sum_{k=1}^{n} x_k + b_1 \sum_{k=1}^{n} x_k^2 = \sum_{k=1}^{n} x_k y'_k. \end{cases} \tag{12.29}$$

Розв'язавши систему нормальних рівнянь (12.29), одержимо оцінки невідомих параметрів моделі $\hat{b}_0$; $\hat{b}_1$.

Оцінки $\hat{a}_0$ та $\hat{a}_1$ невідомих параметрів здійснимо за формулами:

$$\hat{a}_0 = e^{\hat{b}_0}; \quad \hat{a}_1 = e^{\hat{b}_1}.$$

Тоді рівняння регресії матиме вигляд:

$$\hat{Y} = \hat{a}_0 \cdot \hat{a}_1{}^x. \tag{12.30}$$

Приклад 12.4. Дослідити виробничий процес у регіоні (табл.) за допомогою класичної виробничої моделі, що описує залежність між обсягом валової продукції (y) та обсягом основного капіталу (x). Установити форму залежності між X і Y, знайти рівняння регресії і оцінити тісноту зв'язку.

x_k	60	64	72	76	78	82	84	92	94	98
y_k	6,4	6,8	7,8	10,4	15,4	25,4	30,4	40,8	50,4	56,2

Розв'язання. Припускаємо, що між ознаками X та Y існує показникові функціональна залежність

$$\hat{Y} = \hat{a}_0 \cdot \hat{a}_1{}^x.$$

Прологарифмуємо ліву і праву частину рівняння та введемо додаткові позначення.

$$\ln y = \ln(a_0 \cdot a_1{}^x) \quad \Rightarrow \quad \ln y = \ln a_0 + x \ln a_1$$

$y' = \ln y, \quad b_0 = \ln a_0; \;\; b_1 = \ln a_1.$

Тоді рівняння регресії матиме вигляд:

$$y' = b_0 + b_1 \cdot x.$$

Для визначення параметрів b_0 та b_1 скористаємося розрахунковою таблицею, що має такий вигляд:

№ з/п	$\boldsymbol{x_k}$	$\boldsymbol{y_k}$	$\boldsymbol{y'_k = \ln y_k}$	$\boldsymbol{x_k^2}$	$\boldsymbol{y'_k \cdot x_k}$	$\boldsymbol{(y'_k)^2}$
1.	60	6,4	1,856298	3600	111,3779	3,445842
2.	64	6,8	1,916923	4096	122,683	3,674592
3.	72	7,8	2,054124	5184	147,8969	4,219424
4.	76	10,4	2,341806	5776	177,9772	5,484054
5.	78	15,4	2,734368	6084	213,2807	7,476766
6.	82	25,4	3,234749	6724	265,2494	10,4636
7.	84	30,4	3,414443	7056	286,8132	11,65842
8.	92	40,8	3,708682	8464	341,1988	13,75432
9.	94	50,4	3,919991	8836	368,4792	15,36633
10.	98	56,2	4,028917	9604	394,8338	16,23217
$\sum$	800	250	29,2103	65424	2429,79	91,77552

$$b_1 = \frac{\left(\sum_{k=1}^{n} x_k\right)\cdot\left(\sum_{k=1}^{n} y'_k\right) - n\cdot\sum_{k=1}^{n} x_k y'_k}{\left(\sum_{k=1}^{n} x_k\right)^2 - n\cdot\sum_{k=1}^{n} x_k^2} = \frac{800\cdot 29{,}2103 - 10\cdot 2429{,}79}{800^2 - 10\cdot 65424} = 0{,}065285;$$

$$b_0 = \frac{\left(\sum_{k=1}^{n} x_k\right)\cdot\left(\sum_{k=1}^{n} x_k y'_k\right) - \left(\sum_{k=1}^{n} x_k^2\right)\cdot\left(\sum_{k=1}^{n} y'_k\right)}{\left(\sum_{k=1}^{n} x_k\right)^2 - n\cdot\sum_{k=1}^{n} x_k^2} = \frac{800\cdot 2429{,}79 - 65424\cdot 29{,}2103}{800^2 - 10\cdot 65424} =- 2{,}301779.$$

Знайдемо оцінки $\hat{a}_0$ та $\hat{a}_1$ невідомих параметрів, використовуючи такі формули:

$$\hat{a}_0 = e^{b_0} = e^{-2{,}301779} = 0{,}1001;$$

$$\hat{a}_1 = e^{b_1} = e^{0{,}065285} = 1{,}0675.$$

Тоді рівняння регресії матиме вигляд:

$$\hat{Y} = 0{,}1001 \cdot 1{,}0675^{x}.$$

Для оцінки тісноти зв'язку використовуємо коефіцієнт кореляції:

$$r = \frac{K(X;Y)}{\sigma_x \cdot \sigma_{y'}} = \frac{9{,}2966}{11{,}93314711 \cdot 0{,}803203423} = 0{,}97;$$

$$K(X;Y') = \overline{xy'} - \bar{x} \cdot \overline{y'} = 242{,}979 - 80 \cdot 2{,}92103 = 9{,}2966;$$

$$\sigma_x = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{6542{,}4 - 80^2} = 11{,}93314711;$$

$$\sigma_y = \sqrt{\overline{(y')^2} - (\overline{y'})^2} = \sqrt{9{,}177552 - 2{,}92103^2} = 0{,}803203423.$$

Як бачимо, коефіцієнт кореляції близький за своїм значенням до одиниці, що свідчить про те, що зв'язок між Xта Y є тісним.

12.3. Множинна лінійна регресія

На практиці здебільшого залежна змінна Y пов'язана з впливом не одного, а кількох аргументів. У цьому разі регресію називають множинною. Водночас якщо аргументи в функції регресії в першій степені, то множинна регресія називається *лінійною*, в іншому разі – *множинною нелінійною регресією*.

Лінійна множинна регресія

Визначення статистичних точкових оцінок

Розглянемо лінійну залежність Y від m аргументів $(X_1, X_2, \dots, X_m)$.

Лінійна модель у цьому разі набуває такого вигляду:

$$y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + \dots + a_m x_{im}. \quad (12.31)$$

Для вибірки обсягу n матимемо систему лінійних рівнянь

$$\begin{aligned} y_1 &= a_0 + a_1x_{11} + a_2x_{12} + \cdots + a_mx_{1m} + \varepsilon_1; \\ y_2 &= a_0 + a_1x_{21} + a_2x_{22} + \cdots + a_mx_{2m} + \varepsilon_2; \\ y_3 &= a_0 + a_1x_{31} + a_2x_{32} + \cdots + a_mx_{3m} + \varepsilon_3; \\ y_n &= a_0 + a_1x_{n1} + a_2x_{n2} + \cdots + a_mx_{nm} + \varepsilon_n. \end{aligned} \quad (12.32)$$

Параметри рівняння (12.31) є величинами сталими, але невідомими. Ці параметри оцінюють статистичними точковими оцінками $\hat{a}_0$; $\hat{a}_1$; $\hat{a}_2$; …;$\hat{a}_m$, які отримують шляхом обробки результатів вибірки, і є величинами випадковими.

Для визначення компонентів $\hat{a}_0$; $\hat{a}_1$; $\hat{a}_2$; …;$\hat{a}_m$ (статистичних точкових оцінок) застосовується метод найменших квадратів. Необхідно, щоб сума квадратів відхилень фактичних даних від теоретичних була мінімальною. Цю вимогу можна представити так:

$$S = \sum_{i=1}^{n}\left[y_i - (a_0 + a_1x_{i1} + a_2x_{i2} + \cdots + a_mx_{im})\right]^2. \quad (12.33)$$

Ця функція S неперервно диференційовна, тому згідно з необхідними умовами існування мінімуму S повинні виконуватись рівності частинних похідних нулю.

У нашому випадку ці рівності мають вигляд:

$$\begin{cases} \sum_{i=1}^{n}\left[y_i - (a_0 + a_1x_{i1} + a_2x_{i2} + \cdots + a_mx_{im})\right] = 0; \\ \sum_{i=1}^{n}\left[y_i - (a_0 + a_1x_{i1} + a_2x_{i2} + \cdots + a_mx_{im})\right] \cdot x_{i1} = 0; \\ \sum_{i=1}^{n}\left[y_i - (a_0 + a_1x_{i1} + a_2x_{i2} + \cdots + a_mx_{im})\right] \cdot x_{i2} = 0; \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \sum_{i=1}^{n}\left[y_i - (a_0 + a_1x_{i1} + a_2x_{i2} + \cdots + a_mx_{im})\right] \cdot x_{im} = 0. \end{cases}$$

Реалізація цього принципу дає можливість отримати систему нормальних рівнянь:

$$\begin{cases} na_0 + a_1 \sum_{i=1}^{n} x_{i1} + a_2 \sum_{i=1}^{n} x_{i2} + \dots + a_m \sum_{i=1}^{n} x_{im} = \sum_{i=1}^{n} y_i; \\ a_0 \sum_{i=1}^{n} x_{i1} + a_1 \sum_{i=1}^{n} x_{i1}^2 + a_2 \sum_{i=1}^{n} x_{i1}x_{i2} + \dots + a_m \sum_{i=1}^{n} x_{i1}x_{im} = \sum_{i=1}^{n} x_{i1}y_i; \\ a_0 \sum_{i=1}^{n} x_{i2} + a_1 \sum_{i=1}^{n} x_{i1}x_{i2} + a_2 \sum_{i=1}^{n} x_{i2}^2 + \dots + a_m \sum_{i=1}^{n} x_{i2}x_{im} = \sum_{i=1}^{n} x_{i2}y_i; \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_0 \sum_{i=1}^{n} x_{im} + a_1 \sum_{i=1}^{n} x_{i1}x_{im} + a_2 \sum_{i=1}^{n} x_{i2}x_{im} + \dots + a_m \sum_{i=1}^{n} x_{im}^2 = \sum_{i=1}^{n} x_{im}y_i. \end{cases}$$

Розв'язавши систему нормальних рівнянь, одержимо оцінки невідомих параметрів моделі $\hat{a}_0$ $\hat{a}_1$; $\hat{a}_2$; …;$\hat{a}_m$.

Вираження для a_0 , a_1 , a_2…, a_m стає доволі складним, тому доцільно це зробити з допомогою матричної алгебри.

У векторно-матричній формі система (12.32) набуває такого вигляду:

$$Y = XA + \varepsilon, \tag{12.34}$$

де

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}; \quad A = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \dots \\ a_m \end{pmatrix}; \; X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1m} \\ 1 & x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nm} \end{pmatrix}; \; \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \\ \varepsilon_n \end{pmatrix}.$$

Матрицю X розміром $n \times (m + 1)$ називають регресійною, а елементи x_{ij} цієї матриці – регресорами. Статистична оцінка буде визначатись вектором

$$\hat{A} = \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \\ \dots \\ \hat{a}_m \end{pmatrix}.$$

Рівняння (24) представимо у вигляді:

$$S = (Y - XA)' \cdot (Y - XA) = Y'Y - 2A'X'Y + A'X'XA, \tag{12.35}$$

де символ штрих (′) означає операцію транспонування.

Знайдемо частинну похідну окресленого виразу за компонентами вектора A і прирівняємо до нуля:

$$\frac{\partial S}{\partial A} = -2X'Y + 2X'XA = 0.$$

Звідси, отримаємо систему рівнянь у матричній формі, якій повинен задовольняти вектор A при дотриманні вимоги:

$$X'XA = X'Y. \tag{12.36}$$

Якщо до матриці $X'X$ існує обернена матриця $(X'X)^{-1}$, то отримаємо розв'язком системи нормальних рівнянь вектор-стовпець шуканих оцінок параметрів регресії:

$$A = (X'X)^{-1} \cdot (X'Y)\,. \tag{12.37}$$

На відміну від простої моделі регресії алгоритм визначення параметрів багатофакторної моделі є більш складним та трудомістким.

Приклад 12.5. На основі наведених у таблиці даних залежність обсягу отриманого прибутку підприємствами регіону від розміру основних виробничих фондів та затрат праці побудувати множинну лінійну регресійну модель.

Таблиця

Номер підприємства	Прибуток, млн грн, y	Основні фонди, млн грн, x_1	Затрати праці, млн днів, x_2
1	1,2	2,5	4,0
2	1,5	2,8	4,2
3	1,9	3,0	3,6
4	2,2	3,6	4,6
5	2,8	3,9	4,3
6	3,1	4,2	5,1
7	3,4	4,5	5,3
8	4,5	5,0	4,8
9	4,8	5,6	5,4
10	5,4	6,0	5,8

Розв'язування. Попередній аналіз вхідної інформації дає можливість зробити висновок про наявність лінійної форми зв'язку між вибраними економічними показниками:

$$\hat{y} = a_0 + a_1 x_1 + a_2 x_2,$$

де y – прибуток, млн грн; x_1 – вартість основних виробничих фондів, млн грн; x_2 – затрати праці, млн днів.

Для знаходження оцінок параметрів моделі використаємо математичний апарат матричної алгебри.

Введемо позначення:

$$A = \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \end{pmatrix}; \qquad Y = \begin{pmatrix} 1{,}2 \\ 1{,}5 \\ 1{,}9 \\ 2{,}2 \\ 2{,}8 \\ 3{,}1 \\ 3{,}4 \\ 4{,}5 \\ 4{,}8 \\ 5{,}4 \end{pmatrix}; \qquad X = \begin{pmatrix} 1 & 2{,}5 & 4{,}0 \\ 1 & 2{,}8 & 4{,}2 \\ 1 & 3{,}0 & 3{,}6 \\ 1 & 3{,}6 & 4{,}6 \\ 1 & 3{,}9 & 4{,}3 \\ 1 & 4{,}2 & 5{,}1 \\ 1 & 4{,}5 & 5{,}3 \\ 1 & 5{,}0 & 4{,}8 \\ 1 & 5{,}6 & 5{,}4 \\ 1 & 6{,}0 & 5{,}8 \end{pmatrix}.$$

Вектор-стовпець шуканих оцінок параметрів регресії знайдемо, користуючись формулою:

$$A = (X'X)^{-1} \cdot (X'Y),$$

де X' – матриця, транспонована до матриці X.

1. Знаходимо добуток двох матриць:

$$X'X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2{,}5 & 2{,}8 & 3{,}0 & 3{,}6 & 3{,}9 & 4{,}2 & 4{,}5 & 5{,}0 & 5{,}6 & 6{,}0 \\ 4{,}0 & 4{,}2 & 3{,}6 & 4{,}6 & 4{,}3 & 5{,}1 & 5{,}3 & 4{,}8 & 5{,}4 & 5{,}8 \end{pmatrix} \cdot$$

$$\cdot \begin{pmatrix} 1 & 2{,}5 & 4{,}0 \\ 1 & 2{,}8 & 4{,}2 \\ 1 & 3{,}0 & 3{,}6 \\ 1 & 3{,}6 & 4{,}6 \\ 1 & 3{,}9 & 4{,}3 \\ 1 & 4{,}2 & 5{,}1 \\ 1 & 4{,}5 & 5{,}3 \\ 1 & 5{,}0 & 4{,}8 \\ 1 & 5{,}6 & 5{,}4 \\ 1 & 6{,}0 & 5{,}8 \end{pmatrix} = \begin{pmatrix} 10 & 14{,}1 & 47{,}1 \\ 41{,}1 & 181{,}5 & 200{,}2 \\ 47{,}1 & 200{,}2 & 226{,}2 \end{pmatrix}.$$

2. Знаходимо обернену матрицю до матриці $(X'X)$:

$$(X'X)^{-1} = \begin{pmatrix} 8,92 & 1,22 & -2,93 \\ 1,22 & 0,4 & -0,61 \\ -2,93 & -0,61 & 1,51 \end{pmatrix}.$$

3. Знаходимо добуток матриць X' та Y:

$$X'Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2,5 & 2,8 & 3,0 & 3,6 & 3,9 & 4,2 & 4,5 & 5,0 & 5,6 & 6,0 \\ 4,0 & 4,2 & 3,6 & 4,6 & 4,3 & 5,1 & 5,3 & 4,8 & 5,4 & 5,8 \end{pmatrix} \cdot \begin{pmatrix} 1,2 \\ 1,5 \\ 1,9 \\ 2,2 \\ 2,8 \\ 3,1 \\ 3,4 \\ 4,5 \\ 4,8 \\ 5,4 \end{pmatrix} =$$

$$= \begin{pmatrix} 30,8 \\ 141,84 \\ 152,77 \end{pmatrix}.$$

4. Знаходимо вектор-стовпець шуканих оцінок параметрів регресії:

$$A = (X'X)^{-1} \cdot (X'Y) = \begin{pmatrix} 8,92 & 1,22 & -2,93 \\ 1,22 & 0,4 & -0,61 \\ -2,93 & -0,61 & 1,51 \end{pmatrix} \cdot \begin{pmatrix} 30,8 \\ 141,84 \\ 152,77 \end{pmatrix} =$$
$$= \begin{pmatrix} -0,97 \\ 1,4 \\ -0,37 \end{pmatrix}.$$

Отже, нами отримано таку множинну лінійну регресійну модель:

$$\hat{y} = -0,97 + 1,4x_1 - 0,37x_2.$$

Запитання для самоконтролю

1. Дати визначення статистичної залежності між ознаками X та Y.
2. Що означає кореляційна залежність між ознаками X та Y?
3. Що називають функціями регресії?
4. Лінійне рівняння зв'язку X і Y.
5. Яким методом найчастіше визначають оцінки невідомих параметрів?
6. Який принцип лежить в основі методу найменших квадратів?
7. Який вигляд має функція S у випадку парної лінійної регресії?
8. Необхідними умовами існування мінімуму S?
9. У якому випадку можна вважати, що між X та Y існує лінійна залежність?
10. Записати модель парної лінійної регресії.
11. Коваріацією вибірки називається?
12. Коефіцієнтом кореляції називають?
13. Залежність між X та Y тим сильніша, чим значення … ?
14. Який кореляційний зв'язок називається прямим; оберненим?
15. У яких межах знаходиться коефіцієнт кореляції?
16. Якщо статистичні змінні X та Y незалежні, то r дорівнює?
17. Коефіцієнтом детермінації називається величина R^2 …
18. У яких межах находиться коефіцієнт детермінації?
19. У якому випадку можна вважати, що між X та Y існує параболічна залежність?
20. Записати модель парної регресії при параболічній залежності.
21. Який вигляд має функція S у випадку параболічної залежності?
22. У якому випадку можна вважати, що між X та Y існує гіперболічна залежність?
23. Записати модель парної регресії при гіперболічній залежності.
24. Який вигляд має функція S у випадку гіперболічної залежності?
25. У якому випадку можна вважати, що між X та Y існує показникова залежність?
26. Записати модель парної регресії при показниковій залежності.
27. Який вигляд має функція S у випадку показникової залежності?
28. Якщо залежна змінна Y пов'язана з впливом не одного, а кількох аргументів, то регресію називають …?
29. Записати модель множинної лінійної регресії.
30. Записати у векторно-матричній формі модель лінійної множинної регресії.
31. Який вигляд має функція S у випадку множинної лінійної регресії?
32. Вектор-стовпець оцінок параметрів множинної лінійної регресії знаходять за формулою.

Задачі для самостійного розв'язування

1. За дослідними даними методом найменших квадратів визначити оцінки невідомих параметрів лінійної парної регресії. Обчислити коефіцієнт кореляції та детермінації, зробити висновки.

а)

x_k	0,5	1,5	2,5	3,5
y_k	1,32	0,81	0,18	-0,46

б)

x_k	1	2	3	4
y_k	1,1	0,48	-0,1	-0,8

в)

x_k	2	4	6	8
y_k	0,81	2,3	3,4	4,54

г)

x_k	0	0,5	1	1,5	2	2,5	3	3,5	4	4,5
y_k	1,67	1,32	1,1	0,81	0,48	0,18	-0,1	-0,46	-0,8	-1,15

2. За дослідними даними методом найменших квадратів визначити оцінки невідомих параметрів рівняння регресії при параболічній залежності.

а)

x_k	0	2	4	6	8	10
y_k	5	-1	0,5	1,5	1,5	8,5

б)

x_k	-4	-3	-2	-1	0	1	2	3	4
y_k	6	3	1	0,3	-0,1	-0,2	0	0,2	1

в)

x_k	-3	-2	-1	0	1	2	3
y_k	-0,7	0	0,5	0,4	0,9	0,8	0,5

г)

x_k	0	1	2	3	4	5	6	7	8
y_k	0,1	0,48	0,81	1,26	2,3	2,85	3,4	3,96	4,54

3. За дослідними даними методом найменших квадратів визначити оцінки невідомих параметрів рівняння регресії при гіперболічній залежності.

а)

x_k	0,25	0,5	1	2
y_k	6	4	3	3

б)

x_k	1	2	5	10
y_k	10	5	2	1

в)

x_k	8	9	10	11	12	15	18	20	22	24
y_k	50	45	42	36	30	15	12	8	7	5

4. За дослідними даними методом найменших квадратів визначити оцінки невідомих параметрів рівняння регресії при показниковій залежності.

а)

$\boldsymbol{x_k}$	48	50	54	58	60	62	64	66	68	70
$\boldsymbol{y_k}$	18	18	19	26	25	26	30	28	26	34

б)

$\boldsymbol{x_k}$	1	1,2	1,3	1,7	2,5	2,8	3	3,5	3,8	4,2
$\boldsymbol{y_k}$	10	15	20	25	30	36	40	42	40	42

5. Маючи статистичні дані про результати господарської діяльності підприємств побудувати багатофакторну лінійну модель залежності показника y (чистий прибуток підприємств, млн грн) від двох факторів: x_1 (основні фонди, млн грн), x_2 (оборотні фонди, млн грн).

$\boldsymbol{y}$	13	8	15	7	18	21	25	26	28	30	33	38
$\boldsymbol{x_1}$	21	19	18	15	24	23	21	24	25	26	25	29
$\boldsymbol{x_2}$	16	7	11	9	15	17	19	21	20	22	21	20

6. Підприємство, що складається з багатьох філій, досліджує залежність свого річного товарообігу y (млн у. о.) від торгової площі своїх філій x_1 (тис. кв. м) і середньоденної інтенсивності потоку покупців x_2 (тис. чол./день). Побудувати багатофакторну лінійну модель залежності. Дані за філіями наведено в таблиці.

Номер філії	Річний товарообіг, (млн у.о.)	Торгова площа філії, тис. кв. м	Середньоденна інтенсивність потоку покупців, тис. чол./день
1.	2,93	0,31	10,24
2.	5,27	0,98	7,15
3.	6,85	1,21	10,81
4.	7,01	1,29	9,89
5.	7,02	1,12	13,72
6.	8,35	1,49	13,92
7.	4,33	0,78	8,54
8.	5,77	0,94	12,36
9.	7,68	1,29	12,27
10.	3,16	0,48	11,01
11.	1,52	0,24	8,25
12.	3,15	0,55	9,31

7. Маючи статистичні дані про результати господарської діяльності підприємств, побудувати багатофакторну лінійну модель залежності прибутку підприємства y від інвестицій x_1, витрат на рекламу x_2 та витрат на заробітну плату x_3.

Номер підприємства	Прибуток підприємства	Розмір інвестицій	Витрати на рекламу	Витрати на заробітну плату
1.	15,70	17,37	5,28	1,42
2.	17,34	18,24	6,47	1,58
3.	21,57	22,47	6,98	1,98
4.	33,50	18,47	7,05	2,04
5.	32,30	16,82	7,94	2,38
6.	37,90	17,60	8,12	3,48
7.	40,78	17,12	8,69	3,07
8.	48,02	19,81	9,31	3,84
9.	43,30	18,67	10,45	4,28

ТЕСТОВІ ЗАВДАННЯ

1. Що вивчає теорія ймовірностей?
а) закономірності масових подій;
б) закономірності руху планет;
в) закономірності випадання опадів на певній території;
г) правильної відповіді немає.

2. Випадкова подія – це подія, яка:
а) внаслідок експерименту не настає ніколи;
б) внаслідок експерименту обов’язково настане;
в) подія, яка у результаті експерименту може настати або не настати;
г) множина несумісних подій.

3. Вірогідна подія – це подія, яка:
а) внаслідок експерименту не настає ніколи;
б) внаслідок експерименту обов’язково настане;
в) подія, яка у результаті експерименту може настати або не настати;
г) множина несумісних подій.

4. Неможлива подія – це подія, яка:
а) внаслідок експерименту не настає ніколи;
б) внаслідок експерименту обов’язково настане;
в) подія, яка у результаті експерименту може настати або не настати;
г) множина несумісних подій.

5. Залежні події – це події:
а) ймовірність появи однієї з них залежить від появи або не появи іншої;
б) множина подій, які неможливо перелічити;
в) внаслідок експерименту одна із подій обов’язково настане;
г) ймовірність появи однієї з них не залежить від появи або не появи іншої.

6. Незалежні події – це події:
а) ймовірність появи однієї з них залежить від появи або не появи іншої;
б) множина подій, які неможливо перелічити;
в) внаслідок експерименту одна із подій обов’язково настане;
г) ймовірність появи однієї з них не залежить від появи або не появи іншої.

7. Сумісні події – це події:

а) поява однієї з них виключає появу інших в одному і тому ж випробуванні;

б) подія, яка розкладається на елементарні події;

в) поява однієї з них не виключає появу інших в одному і тому ж випробуванні;

г) множина подій, які неможливо перелічити.

8. Рівноможливі – це такі події:

а) поява однієї з яких унеможливлює появу інших;

б) які мають однакові можливості появи;

в) несумісні події;

г) залежні події.

9. Число всіх перестановок з *n* елементів знаходять за формулою:

а) $A_n^k = \frac{n!}{(n-k)!}$;

б) $P_{\ n} = n! = 1 \cdot 2 \cdot 3 \ldots \cdot (n-1) \cdot n$;

в) $C_n^k = \frac{n!}{m!(n-m)!}$;

г) правильної відповіді немає.

10. Число розміщень з *n* елементів по *k* знаходять за формулою:

а) $A_n^k = \frac{n!}{(n-k)!}$;

б) $P_{\ n} = n! = 1 \cdot 2 \cdot 3 \ldots \cdot (n-1) \cdot n$;

в) $C_n^k = \frac{n!}{m!(n-m)!}$;

г) правильної відповіді немає.

11. Число сполучень з *n* елементів по *k* знаходиться за формулою:

а) $A_n^k = \frac{n!}{(n-k)!}$;

б) $P_{\ n} = n! = 1 \cdot 2 \cdot 3 \ldots \cdot (n-1) \cdot n$;

в) $C_n^k = \frac{n!}{m!(n-m)!}$;

г) правильної відповіді немає.

12. Скількома способами 5 осіб можуть утворити чергу?

а) 5;

б) 120;

в) 1;

г) правильної відповіді немає.

13. Обчислити A_5^3 (кількість розміщень з 5 елементів по 3).

а) 60;

б) 20;

в) 100;

г) правильної відповіді немає.

14. Обчислити C_7^2.

а) 5040;

б) 15;

в) 10;

г) правильної відповіді немає.

15. Якщо $\sum_{i=1}^{n} P(A_i) = 1$, тоді події A_i:

а) протилежні;

б) утворюють повну групу;

в) несумісні;

г) незалежні.

16. Для довільної випадкової події:

а) $0 < P(A) < 1$;

б) $0 \leq P(A) \leq 1$;

в) $P(A) \leq 1$;

г) $P(A) \geq 0$.

17. Для математичного опису досліду з нескінченним числом рівно-можливих результатів використовують:

а) геометричне означення ймовірності;

б) статистичне означення ймовірності;

в) класичне означення ймовірності;

г) правильної відповіді немає.

18. Якщо події A і B незалежні між собою, то ймовірність добутку цих подій дорівнює:

а) $P(A \cdot B) = P(A) \cdot P(B)$;

б) $P(A \cdot B) = P(A) \cdot P_A(B)$;

в) $P(A \cdot B) = P(A) + P(B)$;

г) правильної відповіді немає.

19. У класі навчається 18 дівчаток і 12 хлопців. Навмання вибирають одного учня для участі в шкільних зборах. Яка ймовірність того, що буде вибрано хлопця?

а) $2/_3$;

б) $1/_2$;

в) $3/_5$;

г) $2/_5$.

20. Ймовірність добутку двох залежних подій дорівнює:
а) $P(A \cdot B) = P(A) \cdot P(B)$;
б) $P(A \cdot B) = P(A) \cdot P_A(B)$;
в) $P(A \cdot B) = P(A) + P(B)$;
г) правильної відповіді немає.

21. Ймовірність появи хоча б однієї із двох сумісних подій дорівнює:
а) $P(A + B) = P(A) + P(B) - P(A \cdot B)$;
б) $P(A + B) = P(A) + P(B) + P(A \cdot B)$;
в) $P(A + B) = P(A) \cdot P(B)$;
г) правильної відповіді немає.

22. Ймовірність суми протилежних подій дорівнює:
а) $P(A) + P(\bar{A}) = 0$;
б) $P(A) + P(\bar{A}) = \infty$;
в) $P(A) + P(\bar{A}) = 1$;
г) правильної відповіді немає.

23. Події A і B незалежні $P(A) = 0,4$, $P(B) = 0,7$. Обчислити $P(A \cdot B)$.
а) 1,1;
б) 0,3;
в) 0,28;
г) правильної відповіді немає.

24. Події A і B сумісні $P(A) = 0,4$, $P(B) = 0,7$. Обчислити $P(A + B)$.
а) 0,3;
б) 0,28;
в) 0,82;
г) правильної відповіді немає.

25. Яку формулу називають формулою повної ймовірності?

а) $P(A) = \sum_{i=1}^{n} P(H_i) \cdot P_{H_i}(A)$;

б) $P(H_1) + P(H_2) + \cdots + P(H_n) = 1$;

в) $P_A(H_i) = \frac{P(H_i)P_{H_i}(A)}{\sum_{i=1}^{n} P(H_i) \cdot P_{H_i}(A)}$;

г) правильної відповіді немає.

26. Яку формулу називають формулою повної ймовірності?

а) $P(A) = \sum_{i=1}^{n} P(H_i) \cdot P_{H_i}(A)$;

б) $P_n(k) = C_n^k p^k q^{n-k}$;

в) $P_A(H_i) = \frac{P(H_i)P_{H_i}(A)}{\sum_{i=1}^{n} P(H_i) \cdot P_{H_i}(A)}$;

г) правильної відповіді немає.

27. Яку формулу називають формулою Бернуллі?

а) $P(A) = \sum_{i=1}^{n} P(H_i) \cdot P_{H_i}(A)$;

б) $P_n(k) = C_n^k p^k q^{n-k}$;

в) $P_A(H_i) = \frac{P(H_i)P_{H_i}(A)}{\sum_{i=1}^{n} P(H_i) \cdot P_{H_i}(A)}$;

г) правильної відповіді немає.

28. Гіпотези, які висуваються в задачі, що передбачає використання формули повної ймовірності, обов'язково мають задовольняти такі умови:

а) їх має бути не менше п'яти;

б) всі гіпотези попарно незалежні та в сумі становлять простір елементарних подій;

в) всі гіпотези попарно незалежні;

г) правильної відповіді немає.

29. Функцію $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{x}{2}} dx$ називають:

а) функцією Лапласа;

б) функцією Муавра;

в) показниковою функцією;

г) правильної відповіді немає.

30. Досліджується група із 150 осіб на наявність певної ознаки. Ймовірність того, що особа має цю ознаку, становить 0,6. Обчислюється ймовірність того, що в цій групі досліджувану ознаку мають від 120 до 140 осіб. Для обчислення зазначеної ймовірності доцільно застосувати:

а) формулу Пуассона;

б) формулу Бернуллі;

в) локальну теорему Муавра-Лапласа;

г) інтегральну теорему Лапласа.

31. Якщо ймовірність *p* настання події *A* мала, то для великого числа випробувань ймовірність того, що подія відбудеться *k* раз, асимптотично виражається формулою:

а) $P_n(k) = C_n^k p^k q^{n-k}$;

б) $P_n(k_1, k_2) = \Phi(x_2) - \Phi(x_1)$;

в) $P_n(k) = \frac{a^k e^{-a}}{k!}$;

г) правильної відповіді немає.

32. Дискретна випадкова величина – це величина, яка приймає:

а) відокремлені ізольовані одне від одного числові значення;

б) значення рівні 1 або 0;

в) будь-яке числове значення з деякого скінченого або нескінченого інтервалу;

г) будь-яке числове значення з інтервалу (0; 1).

33. Неперервна випадкова величина – це величина, яка приймає:

а) відокремлені ізольовані одне від одного числові значення;

б) значення рівні 1 або 0;

в) будь-яке числове значення з деякого скінченого або нескінченого інтервалу;

г) будь-яке числове значення з інтервалу (0; 1).

34. Яким способом може бути заданий закон розподілу дискретних випадкових величин?

а) табличним;

б) графічним;

в) аналітичним;

г) всі відповіді правильні.

35. Розподіл випадкової величини, яка набуває *n* різних значень x_i з однаковими ймовірностями, називається:

а) рівномірний розподіл;

б) біноміальний розподіл;

в) показниковий розподіл;

г) правильної відповіді немає.

36. Який з таких законів розподілів стосується до закону розподілу дискретної випадкової величини:

а) нормальний розподіл;
б) біноміальний розподіл;
в) показниковий розподіл;
г) правильної відповіді немає.

37. Інтегральною функцією розподілу називають функцію *F(x)*, яка визначає для кожного значення *x* ймовірність того, що:

а) $F(x) = P(X < x)$;
б) $F(x) = P(X \leq x)$;
в) $F(x) = P(X < x)$;
г) правильної відповіді немає.

38. Деяка випадкова величина *X* задана рядом розподілу ймовірностей. Знайти ймовірність $\boldsymbol{p_4 = P(X < x_4)}$

x_i	1	2	3	4	5
p_i	0,1	0,09	0,081	p_4	0,6561

а) 0,1;
б) 0,079;
в) 0,023;
г) правильної відповіді немає.

39. Деяка задана таблиця не може бути рядом розподілу ймовірностей деякої випадкової величини *X*, оскільки:

а) значення x_i не рівновіддалені одне від одного;
б) є від'ємні значення випадкової величини;
в) сума p_i $i = 1, 2, \ldots, 6$ перевищує 1;
г) правильної відповіді немає.

40. Нормальним законом розподілу випадкової величини *X* називають закон, диференціальна функція якого задається формулою:

а) $f(x) = \frac{1}{\sigma\sqrt{2\pi}\cdot} e^{\frac{-(x-a)^2}{2\cdot\sigma^2}}$;

б) $f(x) = \begin{cases} 0 & \text{при } x < 0, \\ \frac{1}{2^{\frac{k}{2}}\cdot\Gamma(\frac{k}{2})} \cdot e^{-\frac{x}{2}-1} \cdot e^{\frac{x}{2}-1} & \text{при } x \geq 0; \end{cases}$

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

г) правильної відповіді немає.

41. Показниковим (експотенціальним) законом розподілу випадкової величини X називають розподіл ймовірностей, диференціальна функція якого задається формулою:

а) $f(x) = \frac{1}{\sigma\sqrt{2\pi}\cdot} e^{\frac{-(x-a)^2}{2\cdot\sigma^2}}$;

б) $f(x) = \begin{cases} 0 & \text{при } x < 0, \\ \frac{1}{2^{\frac{k}{2}}\cdot\Gamma(\frac{k}{2})} \cdot e^{-\frac{x}{2}-1} \cdot e^{\frac{x}{2}-1} & \text{при } x \geq 0; \end{cases}$

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

г) правильної відповіді немає.

42. Ймовірність того, що неперервна випадкова величина прийме значення, що належить інтервалу *(a, b)*, дорівнює:

а) $P(a < X < b) = \int_a^b f(x)dx$;

б) $P(a < X < b) = \int_b^a f(x)dx$;

в) $P(a < X < b) = \int_a^b x \cdot f(x)dx$;

г) правильної відповіді немає.

43. Математичне сподівання дискретної випадкової величини Х називається число, яке дорівнює:

а) сумі всіх можливих значень Х;

б) сумі ймовірностей усіх можливих значень Х;

в) різниці добутків усіх можливих значень Х на відповідні їх ймовірності;

г) сумі добутків усіх можливих значень Х на відповідні їх ймовірності.

44. Дисперсією дискретної випадкової величини Х називається число, яке дорівнює:

а) сумі добутків усіх можливих значень Х на відповідні їх ймовірності;

б) математичному сподіванню квадрата відхилення ДВВ від її математичного сподівання;

в) квадрату відхилення ДВВ від її математичного сподівання;

г) математичному сподіванню квадрата суми ДВВ та її математичного сподівання.

45. Математичне сподівання позначають:

а) $M(X)$;

б) $D(X)$;

в) $\sigma(X)$;

г) правильної відповіді немає.

46. Модою $Mo(X)$ дискретної випадкової величини називають:

а) найімовірніше значення дискретної випадкової величини в деякому околі цього значення;

б) точку максимуму щільності розподілу;

в) всі відповіді правильні;

г) правильної відповіді немає.

47. Виберіть правильно записану теорему Я. Бернуллі.

а) $\lim_{n\to\infty} P\left(\left|\frac{m}{n}-p\right| < \varepsilon\right) = 1$;

б) $\lim_{n\to\infty} P\left(\left|\frac{m}{n}-p\right| < \varepsilon\right) = 0$;

в) $\lim_{n\to\infty} P\left(\left|\frac{m}{n}-p\right| \geq \varepsilon\right) = 1$;

г) правильної відповіді немає.

48. Випадкова величина X вимірюється в кілограмах. В яких одиницях вимірюється $M(X)$?

а) кг^2;

б) кг;

в) безрозмірна;

г) правильної відповіді немає.

49. Випадкова величина X вимірюється в сантиметрах. В яких одиницях вимірюється $\sigma(X)$?

а) см;

б) см^2;

в) безрозмірна;

г) правильної відповіді немає.

50. Випадкова величина X вимірюється в метрах. В яких одиницях вимірюються $D(X)$?

а) м^2;

б) м;

в) безрозмірна;

г) правильної відповіді немає.

51. Вставити пропущений символ у нерівності Чебишева $P\left(\left|X - ?\right| < \varepsilon\right) \geq 1 - \frac{D(X)}{\varepsilon^2}$.

а) $M(X)$;

б) $D(X)$;

в) $\sigma(X)$;

г) правильної відповіді немає.

52. Для того, щоб оцінити ймовірність відхилення значення випадкової величини від її математичного сподівання, потрібно використати:

а) теорему Бернуллі;

б) теорему Чебишева;

в) нерівність Чебишева;

г) центральну граничну теорему.

53. Випадкова величина називається n-мірною:

а) якщо елементарній події відповідає один елемент x;

б) якщо елементарній події відповідає набір чисел (X_1, $X_{2, \ldots}$ X_n), які одночасно появляються під час експерименту з певною ймовірністю;

в) якщо множина можливих значень випадкової величини приймає окремі ізольовані значення з визначеною ймовірністю;

г) правильної відповіді немає.

54. Що дають змогу встановити закони великих чисел?

а) умови, при яких можна мати надприбутки;

б) всі відповіді правильні;

в) умови, при яких сумарна поведінка досить великого числа випадкових величин втрачає випадковий характер і стає закономірною;

г) правильної відповіді немає.

55. Випадкові величини X та Y називаються некорельованими, якщо:

а) коефіцієнт кореляції дорівнює 0;

б) коефіцієнт кореляції дорівнює 1;

в) коефіцієнт кореляції дорівнює –1;

г) коефіцієнт кореляції більший 0.

56. Математична статистика – розділ математики:

а) в якому на основі дослідних даних вивчаються імовірнісні закономірності масових явищ;

б) що вивчає закономірності випадкових явищ: випадкові події, випадкові величини, їхні функції, властивості й операції над ними;

в) всі відповіді правильні;
г) правильної відповіді немає.

57. Які способи відбору об’єктів із генеральної сукупності використовують у практичній діяльності?
а) простий випадковий відбір;
б) типовий та механічний відбір;
в) серійний відбір;
г) всі відповіді правильні.

58. Предмет математичної статистики полягає в розробці:
а) методів збору статистичних даних для одержання наукових та практичних висновків;
б) методів обробки статистичних даних для одержання наукових та практичних висновків;
в) методів збору та обробки статистичних даних;
г) всі відповіді правильні.

59. Дати визначення вибіркової сукупності (вибірки).
а) це сукупність об’єктів, з якої зроблено вибірку;
б) сукупність випадково взятих об’єктів;
в) множина однорідних об’єктів;
г) правильної відповіді немає.

60. Що називається обсягом (об’ємом) сукупності?
а) це кількість об’єктів цієї сукупності;
б) це зростаючий числовий ряд варіант;
в) множина однорідних об’єктів;
г) правильної відповіді немає.

61. Що називається варіаційним рядом?
а) це спадний числовий ряд варіант;
б) це довільний числовий ряд варіант;
в) це зростаючий числовий ряд варіант;
г) правильної відповіді немає.

62. Гістограма частот – це:
а) ступінчаста фігура, яка складається з прямокутників;
б) ламана, відрізки якої з’єднують точки $(x_1, n_1), (x_2, n_2), \ldots, (x_k, n_k)$;
в) ознака випадкової величини;
г) правильної відповіді немає.

63. Полігон частот – це:

а) ступінчаста фігура, яка складається з прямокутників;

б) ламана, відрізки якої з'єднують точки *(x_1, n_1), (x_2, n_2), …, (x_k, n_k)*;

в) ознака випадкової величини;

г) правильної відповіді немає.

64. Частота – це:

а) ознака випадкової величини

б) додатне число, що вказує, скільки раз варіанта трапляється в таблиці даних;

в) ступінчаста фігура, яка складається з прямокутників;

г) правильної відповіді немає.

65. Просту середньоарифметичну вибірки обчислюють за формулою:

а) $\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i$;

б) $\bar{x} = \frac{n}{\sum_{i=1}^{m}\frac{n_i}{x_i}}$;

в) $\bar{x} = \frac{1}{n}\sum_{i=1}^{m} x_i$;

г) $\bar{x} = \left(\sum_{i=1}^{m}\frac{n_i}{n} x_i^{\alpha}\right)^{\frac{1}{\alpha}}$.

66. Гіпотезу називають складною, якщо

а) вона складається із нескінченної кількості простих гіпотез;

б) містить лише одне припущення;

в) вона суперечить основні гіпотезі;

г) вона є основною гіпотезою.

67. Якщо за висновком буде відкинута правильна гіпотеза, то кажуть, що це:

а) похибка нульового роду;

б) похибка другого роду;

в) похибка першого роду;

г) висновок не є хибним.

68. Якщо за висновком буде прийнята неправильна гіпотеза, то кажуть, що це:

а) похибка нульового роду;

б) похибка другого роду;

в) похибка першого роду;

г) висновок не є хибним.

69. Для перевірки гіпотези про розподіл генеральної сукупності, тобто про розподіл випадкової величини, використовують:

а) критерій дисперсійного аналізу;

б) критерій Стьюдента;

в) критерій Фішера;

г) критерій Пірсона.

70. Сума площ всіх прямокутників, із яких складається гістограма частот статистичного розподілу дорівнює:

а) одиниці;

б) об'єму вибірки;

в) найбільшій частоті;

г) правильної відповіді немає.

71. Сума відносних частот усіх варіант дорівнює:

а) об'єму вибірки;

б) одиниці;

в) кількості різних варіант вибірки;

г) правильної відповіді немає.

72. Задано вибірку:

2, 3, 7, 2, 7, 3, 3, 2, 7, 1, 5, 9, 8, 3, 1, 9, 8, 5, 7, 2. 1.

Визначити об'єм вибірки:

а) 7;

б) 4;

в) 20;

г) правильної відповіді немає.

73. Для того, щоб задану вибірку перетворити на варіаційний ряд, потрібно:

а) виписати один раз ті варіанти, які трапляються у виборці;

б) нічого не змінювати;

в) записати всі варіанти у порядку зростання;

г) правильної відповіді немає.

74. Вибірковою дисперсією $D_{в}$ називають:

а) середнє значення квадратів відхилення варіант від вибіркового середнього з урахуванням відповідних частостей;

б) варіанту, яка ділить варіаційний ряд на дві частини, які рівні за числом варіант;

в) початок медіанного інтервалу, тобто такого, якому відповідає перша з нагромаджених частот;

г) правильної відповіді немає.

75. Вибіркове середньоквадратичне відхилення обчислюють за формулою:

а) $\sigma_в = \sqrt{D_в}$;

б) $S = \sqrt{S^2}$;

в) $R = x_{max} - x_{min}$;

г) правильної відповіді немає.

76. Стандарт обчислюють за формулою:

а) $\sigma_в = \sqrt{D_в}$;

б) $\sigma_в = \sqrt{S^2}$;

в) $\sigma_в = \sqrt{x_{max} - x_{min}}$;

г) правильної відповіді немає.

77. Точкову оцінку θ^*, математичне сподівання якої відмінне від оцінюваного параметра θ називають:

а) обґрунтованою;

б) ефективною;

в) зміщеною;

г) правильної відповіді немає.

78. Оцінку θ^*, яка при заданому об'ємі n має найменшу можливу дисперсію називають:

а) обґрунтованою;

б) ефективною;

в) зміщеною;

г) правильної відповіді немає.

79. Статистичну оцінку, яка при $n \to \infty$ прямує за ймовірністю до оцінюваного параметра, називають:

а) обґрунтованою;

б) ефективною;

в) зміщеною;

г) правильної відповіді немає.

80. Метод найбільшої правдоподібності полягає:

а) у знаходженні критичних точок функції розподілу ймовірностей;

б) у знаходженні екстремумів будь-якої функції, що пов’язує оцінювані параметри;

в) у відшуканні максимуму функції правдоподібності від одного або кількох оцінюваннях параметрів;

г) залежно від виду випадкової величини змінюється і те, в чому полягає метод найбільшої правдоподібності.

81. Якщо деякий розподіл визначається одним параметром, то за методом моментів для оцінки параметру математичне сподівання прирівнюється до:

а) будь-якого числа;

б) до вибіркового середнього;

в) до нуля;

г) правильної відповіді немає.

82. Відхилення правильної нульової гіпотези внаслідок перевірки називають:

а) похибкою першого року;

б) похибкою другого року;

в) похибкою третього року;

г) критичною помилкою;

д) немає правильної відповіді.

83. Виберіть термін, який не стосується перевірки статистичних гіпотез:

а) критична область;

б) потужність критерію;

в) умовна ймовірність;

г) рівень значущості.

84. Незміщеною оцінкою генерального середнього є:

а) $\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i$;

б) $M(D_{\text{в}}) = \frac{n-1}{n} D_{\Gamma}$;

в) $u_i = \frac{x_i \pm C}{b}$;

г) правильної відповіді немає.

85. Функцією правдоподібності L дискретної випадкової величини називають функцію:

а) $L(x_1, x_2, \dots, x_n; \theta) = p(x_1; \theta) \cdot p(x_2; \theta) \cdot \dots \cdot p(x_n; \theta)$;

б) $L(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_n; \theta)$;

в)$\ln L = f(x_1; \theta) \cdot p(x_1; \theta)$;

г) правильної відповіді немає.

86. Функцією правдоподібності L неперервної випадкової величини називають функцію:

а) $L(x_1, x_2, \dots, x_n; \theta) = p(x_1; \theta) \cdot p(x_2; \theta) \cdot \dots \cdot p(x_n; \theta)$;

б) $L(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_n; \theta)$;

в)$\ln L = f(x_1; \theta) \cdot p(x_1; \theta)$;

г) правильної відповіді немає.

87. Простою називають гіпотезу:

а) яка містить лише одне твердження;

б) яка складається зі скінченного або нескінченного числа простих гіпотез;

в) яка складається з нескінченного числа простих гіпотез;

г) правильної відповіді немає.

88. Складною називають гіпотезу:

а) яка містить лише одне твердження;

б) яка складається зі скінченного або нескінченного числа простих гіпотез;

в) яка складається зі нескінченного числа простих гіпотез;

г) правильної відповіді немає.

89. Виберіть формулу, за якою визначаються теоретичні частоти:

а) $n_i' = \frac{n \cdot h}{\sigma_{\text{в}}} \cdot \varphi(u_i)$;

б) $n_i' = \frac{n \cdot h}{\sigma_{\text{в}}} \cdot \Phi(u_i)$;

в) всі відповіді правильні;

г) правильної відповіді немає.

90. Рівняння прямої регресії має вигляд:

а) $\hat{Y} = \hat{a}_0 + \hat{a}_1 x$;

б) $y = a_0 + a_1 x + a_2 x^2 + \varepsilon$;

в) $\hat{Y} = \hat{a}_0 + \frac{\hat{a}_1}{x}$;

г) правильної відповіді немає.

91. Якою формулою визначається коваріація?

а) $K(X;Y) = \overline{xy} - \bar{x} \cdot \bar{y}$;

б) $r = \frac{K(X;Y)}{\sigma_x \cdot \sigma_y}$;

в) $R^2 = r^2$;

г) правильної відповіді немає.

92. Якою формулою визначається коефіцієнт кореляції?

а) $K(X;Y) = \overline{xy} - \bar{x} \cdot \bar{y}$;

б) $r = \frac{K(X;Y)}{\sigma_x \cdot \sigma_y}$;

в) $R^2 = r^2$;

г) правильної відповіді немає.

93. Якою формулою визначається коефіцієнт детермінації:

а) $K(X;Y) = \overline{xy} - \bar{x} \cdot \bar{y}$;

б) $r = \frac{K(X;Y)}{\sigma_x \cdot \sigma_y}$;

в) $R^2 = r^2$;

г) правильної відповіді немає.

94. Параболічна залежність – це коли:

а) зі зростанням однієї випадкової величини умовні середні другої зростають (спадають), досягають максимуму (мінімуму), а потім спадають (зростають);

б) зі зростанням однієї випадкової величини умовні середні другої спадають, але не на ту ж величину, як це буває в разі лінійної залежності, а розмір зміни ніби згасає;

в) зі зростанням однієї випадкової величини умовні середні другої зростають (спадають), але не на ту ж величину, як це буває в разі лінійної залежності, а розмір зміни ніби прискорюється (згасає);

г) правильної відповіді немає.

95. Лінійна модель множинної регресії має вигляд:

а) $y_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + \cdots + a_m x_{im},\ i = 1, 2, \ldots, n$;

б) $y = a_0 + a_1 x$;

в) $y = a_0 + a_1 x + a_2 x^2$;

г) $y = a_0 + \frac{a_1}{x}$.

96. Якщо між ознаками *X* та *Y* існує параболічна функціональна залежність, то рівняння регресії має вигляд:

а) $\hat{Y} = \hat{a}_0 + \hat{a}_1 x$;

б) $\hat{Y} = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 x^2$;

в) $\hat{Y} = \hat{a}_0 + \frac{\hat{a}_1}{x}$;

г) правильної відповіді немає.

97. Якщо між ознаками *X* та *Y* існує гіперболічна функціональна залежність, то рівняння регресії має вигляд:

а) $\hat{Y} = \hat{a}_0 + \hat{a}_1 x$;

б) $\hat{Y} = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 x^2$;

в) $\hat{Y} = \hat{a}_0 + \frac{\hat{a}_1}{x}$;

г) правильної відповіді немає.

98. Якщо між ознаками *X* та *Y* існує лінійна функціональна залежність, то рівняння регресії має вигляд:

а) $\hat{Y} = \hat{a}_0 + \hat{a}_1 x$;

б) $\hat{Y} = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 x^2$;

в) $\hat{Y} = \hat{a}_0 + \frac{\hat{a}_1}{x}$;

г) правильної відповіді немає.

99. Якщо між ознаками *X* та *Y* існує показникова функціональна залежність, то рівняння регресії має вигляд:

а) $\hat{Y} = \hat{a}_0 \cdot \hat{a}_1{}^x$;

б) $\hat{Y} = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 x^2$;

в) $\hat{Y} = \hat{a}_0 + \frac{\hat{a}_1}{x}$;

г) правильної відповіді немає.

100. Гіперболічна залежність – це коли:

а) зі зростанням однієї випадкової величини умовні середні другої зростають (спадають), досягають максимуму (мінімуму), а потім спадають (зростають);

б) зі зростанням однієї випадкової величини умовні середні другої спадають, але не на ту ж величину, як це буває в разі лінійної залежності, а розмір зміни ніби згасає;

в) зі зростанням однієї випадкової величини умовні середні другої зростають (спадають), але не на ту ж величину, як це буває в разі лінійної залежності, а розмір зміни ніби прискорюється (згасає);

г) правильної відповіді немає.

101. Показникова залежність – це коли:

а) зі зростанням однієї випадкової величини умовні середні другої зростають (спадають), досягають максимуму (мінімуму), а потім спадають (зростають);

б) зі зростанням однієї випадкової величини умовні середні другої спадають, але не на ту ж величину, як це буває в разі лінійної залежності, а розмір зміни ніби згасає;

в) зі зростанням однієї випадкової величини умовні середні другої зростають (спадають), але не на ту ж величину, як це буває в разі лінійної залежності, а розмір зміни ніби прискорюється (згасає);

г) правильної відповіді немає.

102. Під час перевірки гіпотези про розподіл генеральної сукупності за законом Пуассона за оцінку параметра λ беремо:

а) вибіркове середнє: $\lambda^* = \bar{x}$;

б) вибіркову дисперсію: $\lambda^* = D_{\text{в}}$;

в) вибіркове середньоквадратичне відхилення: $\lambda^* = \sigma$;

г) правильної відповіді немає.

103. При перевірці гіпотези про біноміальний розподіл теоретичні частоти обчислюються за формулою:

а) $n_i' = C_N^i p^i (1-p)^{N-i}$;

б) $n_i' = n \cdot C_N^i p^i (1-p)^{N-i}$;

в) $n_i' = n$;

г) правильної відповіді немає.

104. За оцінку параметра λ показникового розподілу береться величина:

а) рівна вибірковому середньому $\lambda^* = \bar{x}$;

б) обернена до вибіркового середнього $\lambda^* = \frac{1}{\bar{x}}$;

в) рівна вибірковому середньоквадратичному відхиленню $\lambda^* = \sigma$;

г) правильної відповіді немає.

105. При перевірці за критерієм Пірсона гіпотеза про нормальний розподіл приймається, якщо:

а) $\chi^2_{\text{спост}} < \chi^2_{\text{крит}}$;

б) $\chi^2_{\text{спост}} > \chi^2_{\text{крит}}$;

в) $\chi^2_{\text{спост}} = \chi^2_{\text{крит}}$;

г) $\chi^2_{\text{спост}} \neq \chi^2_{\text{крит}}$.

ТЕРМІНОЛОГІЧНИЙ СЛОВНИК

Б

Багатовимірні випадкові величини – це такі випадкові величини, можливі значення яких визначаються двома, трьома, … , n числами.

Безповторна вибірка – це вибірка, в якій взятий об'єкт до генеральної сукупності не повертається.

Безумовна ймовірність – це ймовірності події, при обчисленні якої не вказуються жодні додаткові умови.

В

Варіанти – це об'єкти вибірки, які обстежують за їх певними характеристиками або ознаками.

Варіаційний ряд – це сукупність значень ознаки (статистичної змінної), записаних у порядку їх зростання.

Вибірка, або вибіркова сукупність – це об'єкти, довільно або випадково відібрані з генеральної сукупності для дослідження.

Вибіркове середнє статистичного матеріалу – це сума всіх значень вибірки, поділена на її обсяг n.

Вибіркове середньоквадратичне відхилення – це квадратний корінь із вибіркової дисперсії.

Вибіркова дисперсія $D_в$ – це середнє значення квадратів відхилення варіант від вибіркового середнього з урахуванням відповідних частостей.

Випадкова величина X – це величина, яка в результаті випробування прийме одне і тільки одне з можливих значень наперед невідоме і залежне від випадкових причин, які попередньо не можуть бути враховані.

Випадкова подія – це така подія, яка при виконанні комплексу умов може відбутися, або не відбутися.

Виправлена вибіркова дисперсія (варіанса) – це сума квадратів відхилень елементів від вибіркового середнього, поділена на $(n-1)$.

Відносна частота появи події – це відношення числа появ події до загального числа проведених випробувань.

Вірогідна (достовірна) подія – це подія, яка обов'язково відбувається при кожному виконанні комплексу умов.

Впорядкована множина – це множина, в якій задано порядок розміщення її елементів.

Г

Генеральна сукупність – це вся сукупність об'єктів, які досліджуються.

Гіпотези – це такі попарно незалежні події H_1, H_2, ..., H_n, якщо подія A може відбутися тільки при реалізації однією з них.

Гістограма відносних частот – це ступінчаста фігура, яка складається з прямокутників, основами яких є частинні інтервали варіант довжиною $h = x_i - x_{i-1}$, а висоти дорівнюють $\frac{W_i}{h}$ (щільність відносної частоти).

Гістограма частот – це ступінчаста фігура, яка складається з прямокутників, основами яких є частинні інтервали варіант довжиною $h = x_i - x_{i-1}$, а висоти дорівнюють $\frac{n_i}{h}$ (щільність частоти).

Д

Дискретна випадкова величина – це випадкова величина, яка приймає ізольовані числові значення із усіх можливих.

Дисперсія (розсіювання) дискретної випадкової величини – це математичне сподівання квадрата відхилення випадкової величини від її математичного сподівання.

Диференціальна функція Лапласа – це функція вигляду $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}}$ де $x = \frac{k-np}{\sqrt{npq}}$.

Диференціальна функція розподілу *f(x)* або щільність ймовірностей неперервної випадкової величини *X* – це перша похідна від інтегральної функції, тобто: $f(x) = F'(x)$.

Добуток двох подій *A і B* – це подія *C*, що полягає в одночасній появі цих подій.

Е

Експеримент (або випадковий експеримент) – це певний комплекс умов, що забезпечує спостереження за певним реальним випадковим явищем (певною реальною випадковою подією).

Емпірична функція розподілу (або функцією розподілу вибірки) – це функція $F^*(x)$, яка визначає для кожного значення x відносну частоту події $X < x$.

Ефективна оцінка – це така оцінка θ^*, яка при заданому об’ємі n має найменшу можливу дисперсію.

З

Закон великих чисел – це свого роду зв’язок між теорією ймовірностей як математичною наукою і закономірностями випадкових явищ у разі масових спостережень над ними.

Закон розподілу дискретної випадкової величини – це перелік всіх її можливих значень та відповідних їм ймовірностей.

Залежні події – це випадкові події А і В, якщо поява однієї з них (А або В) впливає на ймовірність появи іншої.

Зміщена точкова оцінка θ* – це оцінка θ^*, математичне сподівання якої відмінне від оцінюваного параметра θ.

Зсув оцінки $\boldsymbol{\theta}^*$ – це величина $b(\theta) = \mathrm{M}(\theta^*) - \theta$.

Й

Ймовірність *P(A)* події *A* – це відношення числа результатів випробувань, сприятливих події *A,* до числа всіх можливих результатів випробувань.

І

Інтегральна функція Лапласа – це функція вигляду $\Phi(\mathrm{x}) = \frac{1}{\sqrt{2\pi}}\int_{x_1}^{x_2} e^{-\frac{x}{2}}\,dx$.

Інтегральна функція розподілу – це функція *F(x),* яка визначає для кожного значення *x* ймовірність того, що випадкова величина *X* приймає значення, менше дійсного числа *x*, тобто: $F(x) = P(X < x)$.

Інтервальний варіаційний ряд – це частинні інтервали варіант, які розміщені у зростаючій послідовності.

Інтервальна оцінка – це статистична оцінка, що визначається двома числами, кінцями інтервалів.

К

Коваріація вибірки – це величина $K(X;Y) = \overline{xy} - \bar{x}\cdot\bar{y}$.

Коефіцієнт детермінації – це величина $R^2 = r^2$.

Коефіцієнт кореляції – це показник, що вимірює тісноту стохастичного зв’язку між змінними.

Комбінаторика – це розділ математики, який вивчає розміщення об'єктів відповідно до спеціальних правил і методів підрахунку числа всіх можливих способів, якими ці розміщення можуть бути зроблені.

Конкуруюча (альтернативна) гіпотеза – це гіпотеза H_a, яка суперечить нульовій.

Кореляційний момент $\mathbf{M_{xy}}$ випадкових величин $\boldsymbol{X}$ і $\boldsymbol{У}$ – це математичне сподівання добутку відхилень цих величин від своїх математичних сподівань.

Корельовані випадкові величини $\boldsymbol{X}$ і $\boldsymbol{У}$ – це такі випадкові величини, для яких кореляційний момент відмінний від нуля.

Критичниі точки $\boldsymbol{k}_{\text{кр}}$ – це точки, які відокремлюють критичну область від області прийняття гіпотези.

Критична область – це множина значень критерію, при яких нульову гіпотезу відхиляють.

Л

Логарифмічна функція правдоподібності – це функція $\ln L$, де L – функція правдоподібності.

М

Математична статистика – розділ математики, в якому на основі дослідних даних вивчаються імовірнісні закономірності масових явищ.

Математичне сподівання дискретної випадкової величини – це сума добутків всіх можливих значень цієї величини на відповідні їм імовірності.

Медіана $\mathbf{M_e}$ в статистиці – це варіанта, яка ділить варіаційний ряд на дві частини, які рівні за числом варіант.

Медіана випадкової величини – це таке число $Me(X)$, для якого виконується умова $P\big(X < Me(X)\big) = P\big(X > Me(X)\big)$.

Метод максимальної правдоподібності – це метод, який полягає у знаходженні максимуму функції одного або кількох оцінюваних параметрів.

Метод моментів знаходження точкових оцінок – це метод, згідно з яким статистичні оцінки визначаються з умови мінімізації суми квадратів відхилень варіант вибірки від статистичної оцінки θ^*.

Множинна лінійна регресія – це регресія, коли змінна Y пов'язана з впливом не одного, а кількох аргументів в першій степені.

Мода $\mathbf{M_o}$ в статистиці – це елемент, який найчастіше трапляється у вибірці.

Мода абсолютно неперервної випадкової величини – це точка максимуму щільності розподілу.

Мода $Mo(X)$ дискретної випадкової величини – це найімовірніше її значення в деякому колі цього значення.

Н

Незалежні події – це події А і В, для яких поява однієї з них (А або В) не впливає на ймовірність появи іншої.

Незалежні послідовні випробування – це такі випробування, якщо здійснення будь-якого результату в n-му за рахунком випробуванні не залежить від результатів у попередніх.

Незміщена точкова оцінка $\boldsymbol{\theta}^*$ – це оцінка θ^*, математичне сподівання якої дорівнює оцінюваному параметру θ для будь-якого обсягу вибірки.

Некорельовані випадкові величини *X* і *У* – це такі величини, для яких кореляційний момент дорівнює нулю.

Неможлива подія – це подія, яка напевно не може відбутися при виконанні комплексу умов.

Неперервна випадкова величина – це величина, яка приймає всі свої можливі значення з деякого проміжку.

Нормальний закон розподілу випадкової величини *X* – це закон, диференціальна функція якого задається формулою $f(x)=\frac{1}{\sigma\sqrt{2\pi\cdot}}e^{\frac{-(x-a)^2}{2\cdot\sigma^2}}$.

Нормальна крива або крива Гауса – це графік диференціальної функції величини, розподіленої за нормальним законом.

Нульова (основна) гіпотеза – це висунута гіпотеза H_0.

О

Обґрунтована статистична оцінка – це оцінка, яка при $n\to\infty$ прямує за ймовірністю до оцінюваного параметра.

Область прийняття гіпотези (область допустимих значень) – це множина значень критерію, за яких нульову гіпотезу приймають.

Обсяг (об'єм) сукупності – це кількість об'єктів цієї сукупності.

Основний принцип перевірки статистичних гіпотез такий: якщо спостережуване значення критерію належить критичній області, нульову гіпотезу відхиляють; якщо спостережуване значення критерію належить області прийняття гіпотези, гіпотезу приймають.

Оцінка найбільшої або максимальної правдоподібності параметра θ – це таке його значення θ^*, при якому функція правдоподібності досягає свого максимуму.

П

Параметричні гіпотези – це статистичні гіпотези про значення параметрів ознак генеральної сукупності.

Перестановка – це всякий встановлений у скінченній множині порядок його елементів.

Повна група подій – це попарно несумісні та рівноможливі події A_1, A_2, ...,A_n.

Повторна вибірка – це вибірка, при якій відібраний об'єкт повертається до генеральної сукупності перед відбором іншого об'єкта.

Показниковий закон розподілу випадкової величини *X* – це розподіл ймовірностей, який описується диференціальною функцією: $f(x) = \begin{cases} 0 & \text{при } x < 0 \\ \lambda e^{-\lambda x} & \text{при } x \geq 0 \end{cases}$, де λ – стала величина $\lambda > 0$.

Полігон відносних частот – це ламана, відрізки якої з'єднують точки, координати яких є значення варіант та відповідних їм відносних частот.

Полігон частот – це ламана, відрізки якої з'єднують точки, координати яких є значення варіант та відповідних їм частот.

Помилка другого роду полягає в тому, що буде прийнята неправильна гіпотеза.

Помилка першого роду полягає в тому, що буде відхилена правильна гіпотеза.

Потужність критерію – це ймовірність належності критерію критичній області при умові, що правильна альтернативна гіпотеза.

Проста гіпотеза – це гіпотеза, яка містить лише одне твердження.

Протилежні події – це дві єдиноможливі події, які утворюють повну групу подій.

Р

Репрезентативна (представницька) вибірка – це вибірка, яка правильно відображає ті властивості генеральної сукупності, які вивчаються.

Рівень значущості – це ймовірність здійснити помилку першого роду.

Рівноможливі події – це події A_1, A_2, ...,A_n, якщо ні одна з цих подій не має «переваги» над іншою.

Розмах – це різниця між найбільшим і найменшим значенням варіаційного ряду (між крайніми елементами).

Розміщення з *n* елементів по *k* – це всяка упорядкована k-елементна підмножина ($k < n$).

С

Середнє геометричне вибірки – це середнє значення, яке обчислюється за формулою $\bar{x}_{\Gamma} = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \ldots \cdot x_m^{n_m}}$.

Середнє квадратичне відхилення дискретної випадкової величини *X* – це квадратний корінь з дисперсії.

Складна гіпотеза – це гіпотеза, яка складається зі скінченного або нескінченного числа простих гіпотез.

Сполучення з *n* елементів по *k* – це будь-яка її k- елементна підмножина ($k < n$).

Спостережне значення критерію узгодження – це значення відповідного критерію, обчислене за даними вибірки.

Спроможна послідовність оцінок – це послідовність θ_k^* $(k = 1, 2, \ldots, m)$ параметра θ, якщо $\theta_k^* \xrightarrow{P} \theta,\ n \to \infty$.

Стандарт – це арифметичний квадратний корінь із варіанси.

Статистики – це числові характеристики, обчислені по вибірці або ті, що використовуються для опису даних вибірки.

Статистичний критерій перевірки нульової гіпотези – це деяка випадкова величина, розподіл якої відомий.

Статистичні гіпотези – це будь-які статистичні висновки, здобуті на підставі обробки вибірки.

Статистична оцінка $\boldsymbol{\theta^*}$ невідомого параметра $\boldsymbol{\theta}$ теоретичного розподілу – це функція $f(X_1, X_2, \ldots, X_n)$, де $X_1,\ X_2,\ \ldots,\ X_n$ – спостережувальні випадкові величини.

Степеневе середнє вибірки – це середнє значення, яке знаходять за формулою $\overline{x_c} = \left(\sum_{i=1}^{m} \frac{n_i}{n} x_i^{\alpha}\right)^{\frac{1}{\alpha}}$.

Сума двох подій *А і В* – це подія, що полягає у появі події *А* або події *В*, або обох цих подій.

Т

Теорія ймовірностей – розділ математики, що вивчає закономірності випадкових явищ: випадкові події, випадкові величини, їхні функції, властивості й операції над ними.

Точкова оцінка – це статистична оцінка, яка визначається одним єдиним числом $\theta^* = f(x_1, x_2, \ldots, x_n)$, де $x_1, x_2, \ldots, x_n$ – результати n спостережень над кількісною ознакою випадкової величини *X*.

Точність оцінки – це різниця між статистичною оцінкою θ^* та її оцінювальним параметром θ, взята за абсолютним значенням.

У

Умовна ймовірність події *A* при умові *B* – це ймовірність настання події *A,* обчислена в припущені, що подія *B* уже відбулася.

Ф

Функція правдоподібності L дискретної випадкової величини – це функція аргумента θ $L(x_1, x_2, \dots, x_n; \theta) = p(x_1; \theta) \cdot p(x_2; \theta) \cdot \dots \cdot p(x_n; \theta)$.

Функція правдоподібності L неперервної випадкової величини – це функція аргумента θ $L(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_n; \theta)$.

Функція регресії – це можливі залежності між змінними X і Y.

Ч

Частота варіанти – це додатне число, яке показує, скільки разів варіанта трапляється в таблиці даних.

Щ

Щільність відносної частоти – це величина $\frac{W_i}{h}$, утворена діленням відносної частоти на ширину інтервалу.

Щільність частоти вибірки – це величина $\frac{n_i}{h}$, утворена діленням частоти на ширину інтервалу.

СПИСОК ВИКОРИСТАНИХ ТА РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

1. Барковський В. В., Н. В. Барковська, О. К. Лопатін. Теорія ймовірностей та математична статистика: навчальний посібник. К.: ЦНЛ, 2006. 424 с.

2. Валь О. Д., Мельничук О. Д., Королюк С. Л. Теорія ймовірностей від найпростішого: навчальний посібник. Чернівці: Книги-ХХІ, 2004. 160 с.

3. Волошин О. Р., Галайко Н. В. Математична статистика: курс лекцій. Львів: ЛьвДУВС, 2010. 88 с.

4. Волощенко А. Б., Джалладова І. Б. Теорія ймовірностей та математична статистика: навчально-методичний посібник для самост. вивч. дисципліни. К.: КНЕУ, 2003. 356 с.

5. Горбань С. Ф., Снижко Н. В. Теория вероятностей и математическая статистика: курс лекций. К.: МАУП, 1999. 168 с.

6. Донченко В. С., Сидоров М. В., Шарапов М. М. Теорія ймовірності та математична статистика: навчальний посібник. К.: Академія, 2009. 288 с.

7. Єрьоменко В. О., Шинкарик М. І. Теорія імовірностей: навчальний посібник для студентів економічних спеціальностей. Тернопіль: Економічна думка, 2000. 176 с.

8. Жалдак М. І., Михалін Г. О Елементи стохастики з комп’ютерною підтримкою: посібник для вчителів. К.: Шкільний світ, 2002. 128 с.

9. Жерновий Ю. В. Збірник задач з теорії ймовірностей та математичної статистики для студентів нематематичних спеціальностей. Львів, 2009. 18 с. URL: http://zyurvas.narod.ru/Lekcyi_z_TIMS/zbirn_zadach.pdf.

10. Жерновий Ю.В. Теорія ймовірностей та математична статистика: тексти лекцій для студентів нематематичних спеціальностей. Львів, 2008. 101 с. URL: http://zyurvas.narod.ru/Lekcyi_z_TIMS/Lekcii_z_TIMS.pdf.

11. Жлуктенко В. І., Наконечний С. І. Теорія ймовірностей і математична статистика: навчально-методичний посібник у 2-х частинах. Ч. 1. Теорія ймовірностей. К.: КНЕУ, 2000. 304 с.

12. Жлуктенко В. І., Наконечний С. І. Теорія ймовірностей і математична статистика: навчально-методичний посібник у 2-х частинах. Ч. 2. Математична статистика. К.: КНЕУ, 2001. 336 с.

13. Конет І. М. Теорія ймовірностей та математична статистика в прикладах і задачах: навчально-методичний посібник. Кам’янець-Подільський: Абетка, 2001. 218 с.

14. Медведєв М. Г., Пащенко І. О. Теорія ймовірностей та математична статистика: підручник. К.: Кондор, 2008. 536 с.

15. Практикум з теорії ймовірностей та математичної статистики: навчальний посібник / за ред. Р. К. Чорнея. Київ: МАУП, 2003. 328 с.

16. Пушак Я. С., Лозовий Б. Л. Теорія імовірностей і елементи математичної статистики: навчальний посібник. Львів: УАД, 2006. 428 с.

17. Рудомино-Дусятская И. А., Кепич Е. И. Теория вероятностей и математическая статистика: учебное пособие. К.: Университет экономики и права «КРОК», 2007. 116 с.

18. Турчин В. М. Теорія ймовірностей: Основні поняття, приклади, задачі: навчальний посібник. К.: А.С.К., 2004. 476 с.

19. Черняк О. І., Обушна О.М., Ставицький А.В. Теорія ймовірностей та математична статистика. Збірник задач: навчальний посібник. 2-ге видання, виправлене. К.: Т-во «Знання», КОО, 2002. 199 с.

ДОДАТКИ

Таблиця значень функції $\varphi(x) = \frac{1}{2\pi} e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3988	3986	3989	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3952	3845	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3725	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0794	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0011	0011	0010	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001
4,0	0001	0001	0001	0001	0001	0001	0001	0001	0001	0001
4,1	0001	0001	0001	0001	0001	0001	0001	0001	0001	0001
4,2	0001	0001	0001	0001	0000	0000	0000	0000	0000	0000

Таблиця значень функції $\Phi(x) = \frac{1}{2\pi}\int_0^x e^{-\frac{z^2}{2}}\,dz$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	0,50	0,1915	1,00	0,3413	1,50	0,4332	2,00	0,4772	2,50	0,4938
0,01	0,0040	0,51	0,1950	1,01	0,3438	1,51	0,4345	2,01	0,4778	2,52	0,4941
0,02	0,0080	0,52	0,1985	1,02	0,3461	1,52	0,4357	2,02	0,4783	2,54	0,4945
0,03	0,0120	0,53	0,2019	1,03	0,3485	1,53	0,4370	2,03	0,4788	2,56	0,4948
0,04	0,0160	0,54	0,2054	1,04	0,3508	1,54	0,4382	2,04	0,4793	2,58	0,4951
0,05	0,0199	0,55	0,2088	1,05	0,3531	1,55	0,4394	2,05	0,4798	2,60	0,4953
0,06	0,0239	0,56	0,2123	1,06	0,3554	1,56	0,4406	2,06	0,4803	2,62	0,4956
0,07	0,0279	0,57	0,2157	1,07	0,3577	1,57	0,4418	2,07	0,4808	2,64	0,4959
0,08	0,0319	0,58	0,2190	1,08	0,3599	1,58	0,4429	2,08	0,4812	2,66	0,4961
0,09	0,0359	0,59	0,2224	1,09	0,3621	1,59	0,4441	2,09	0,4817	2,68	0,4963
0,10	0,0398	0,60	0,2257	1,10	0,3643	1,60	0,4452	2,10	0,4821	2,70	0,4965
0,11	0,0438	0,61	0,2291	1,11	0,3665	1,61	0,4463	2,11	0,4826	2,72	0,4967
0,12	0,0478	0,62	0,2324	1,12	0,3686	1,62	0,4474	2,12	0,4830	2,74	0,4969
0,13	0,0517	0,63	0,2357	1,13	0,3708	1,63	0,4484	2,13	0,4834	2,76	0,4971
0,14	0,0557	0,64	0,2389	1,14	0,3729	1,64	0,4495	2,14	0,4838	2,78	0,4973
0,15	0,0596	0,65	0,2422	1,15	0,3749	1,65	0,4505	2,15	0,4842	2,80	0,4974
0,16	0,0636	0,66	0,2454	1,16	0,3770	1,66	0,4515	2,16	0,4846	2,82	0,4976
0,17	0,0675	0,67	0,2486	1,17	0,3790	1,67	0,4525	2,17	0,4850	2,84	0,4977
0,18	0,0714	0,68	0,2517	1,18	0,3810	1,68	0,4535	2,18	0,4854	2,86	0,4979
0,19	0,0753	0,69	0,2549	1,19	0,3830	1,69	0,4545	2,19	0,4857	2,88	0,4980
0,20	0,0793	0,70	0,2580	1,20	0,3849	1,70	0,4554	2,20	0,4861	2,90	0,4981
0,21	0,0832	0,71	0,2611	1,21	0,3869	1,71	0,4564	2,21	0,4864	2,92	0,4982
0,22	0,0871	0,72	0,2642	1,22	0,3888	1,72	0,4573	2,22	0,4868	2,94	0,4984
0,23	0,0910	0,73	0,2673	1,23	0,3907	1,73	0,4582	2,23	0,4871	2,96	0,4985
0,24	0,0948	0,74	0,2704	1,24	0,3925	1,74	0,4591	2,24	0,4875	2,98	0,4986
0,25	0,0987	0,75	0,2734	1,25	0,3944	1,75	0,4599	2,25	0,4878	3,00	0,4987
0,26	0,1026	0,76	0,2764	1,26	0,3962	1,76	0,4608	2,26	0,4881	3,05	0,4989
0,27	0,1064	0,77	0,2794	1,27	0,3980	1,77	0,4616	2,27	0,4884	3,10	0,49903
0,28	0,1103	0,78	0,2823	1,28	0,3997	1,78	0,4625	2,28	0,4887	3,15	0,49918
0,29	0,1141	0,79	0,2852	1,29	0,4015	1,79	0,4633	2,29	0,4890	3,20	0,49931
0,30	0,1179	0,80	0,2881	1,30	0,4032	1,80	0,4641	2,30	0,4893	3,25	0,49942
0,31	0,1217	0,81	0,2910	1,31	0,4049	1,81	0,4649	2,31	0,4896	3,30	0,49952
0,32	0,1255	0,82	0,2939	1,32	0,4066	1,82	0,4656	2,32	0,4898	3,35	0,49960
0,33	0,1293	0,83	0,2967	1,33	0,4082	1,83	0,4664	2,33	0,4901	3,40	0,49966
0,34	0,1331	0,84	0,2995	1,34	0,4099	1,84	0,4671	2,34	0,4904	3,50	0,49977
0,35	0,1368	0,85	0,3023	1,35	0,4115	1,85	0,4678	2,35	0,4906	3,60	0,49984
0,36	0,1406	0,86	0,3051	1,36	0,4131	1,86	0,4686	2,36	0,4909	3,70	0,49989
0,37	0,1443	0,87	0,3078	1,37	0,4147	1,87	0,4693	2,37	0,4911	3,80	0,499928
0,38	0,1480	0,88	0,3106	1,38	0,4162	1,88	0,4699	2,38	0,4913	3,90	0,499952
0,39	0,1517	0,89	0,3133	1,39	0,4177	1,89	0,4706	2,39	0,4916	4,00	0,499968
0,40	0,1554	0,90	0,3159	1,40	0,4192	1,90	0,4713	2,40	0,4918	4,20	0,499987
0,41	0,1591	0,91	0,3186	1,41	0,4207	1,91	0,4719	2,41	0,4920	4,40	0,4999946
0,42	0,1628	0,92	0,3212	1,42	0,4222	1,92	0,4726	2,42	0,4922	4,60	0,4999979
0,43	0,1664	0,93	0,3238	1,43	0,4236	1,93	0,4732	2,43	0,4925	4,80	0,4999992
0,44	0,1700	0,94	0,3264	1,44	0,4251	1,94	0,4738	2,44	0,4927	5,00	0,4999997
0,45	0,1736	0,95	0,3289	1,45	0,4265	1,95	0,4744	2,45	0,4929		
0,46	0,1772	0,96	0,3315	1,46	0,4279	1,96	0,4750	2,46	0,4931		
0,47	0,1808	0,97	0,3340	1,47	0,4292	1,97	0,4756	2,47	0,4932		
0,48	0,1844	0,98	0,3365	1,48	0,4306	1,98	0,4761	2,48	0,4934		
0,49	0,1879	0,99	0,3389	1,49	0,4319	1,99	0,4767	2,49	0,4936		

Таблиця значень функції Пуассона $P_n(m) = \frac{\lambda^m}{m!} e^{-\lambda}$

m \ λ	**0,1**	**0,2**	**0,3**	**0,4**	**0,5**	**0,6**	**0,7**	**0,8**	**0,9**	**1**
0	0,9048	8187	7408	6703	6065	5488	4966	4493	4066	3679
1	0905	1637	2222	2681	3033	3293	3476	3595	3659	3679
2	0045	0164	0333	0536	0758	0988	1217	1438	1647	1839
3	0002	0011	0033	0072	0126	0198	0284	0383	0494	0613
4	0000	0001	0003	0007	0016	0030	0050	0077	0111	0153
5	0000	0000	0000	0001	0002	0004	0007	0012	0020	0031
6	0000	0000	0000	0000	0000	0000	0001	0002	0003	0005

m \ λ	**1,1**	**1,2**	**1,3**	**1,4**	**1,5**	**1,6**	**1,7**	**1,8**	**1,9**	**2**
0	0,3329	3012	2725	2466	2231	2019	1827	1653	1496	1353
1	3662	3614	3543	3452	3347	3230	3106	2975	2842	2707
2	2014	2169	2303	2417	2510	2584	2640	2678	2700	2707
3	0738	0867	0998	1128	1255	1378	1496	1607	1710	1804
4	0203	0260	0324	0395	0471	0551	0636	0723	0812	0902
5	0045	0062	0084	0111	0141	0176	0216	0260	0309	0361
6	0008	0012	0018	0026	0035	0047	0061	0078	0098	0120
7	0001	0002	0003	0005	0008	0011	0015	0020	0027	0034
8	0000	0000	0001	0001	0001	0002	0003	0005	0006	0009
9	0000	0000	0000	0000	0000	0000	0001	0001	0001	0002

m \ λ	**2,1**	**2,2**	**2,3**	**2,4**	**2,5**	**2,6**	**2,7**	**2,8**	**2,9**	**3**
0	0,1225	1108	1003	0907	0821	0743	0672	0608	0550	0498
1	2572	2438	2306	2177	2052	1931	1815	1703	1596	1494
2	2700	2681	2652	2613	2565	2510	2450	2384	2314	2240
3	1890	1966	2033	2090	2138	2176	2205	2225	2237	2240
4	0992	1082	1169	1254	1336	1414	1488	1557	1622	1680
5	0417	0476	0538	0602	0668	0735	0804	0872	0940	1008
6	0146	0174	0206	0241	0278	0319	0362	0407	0455	0504
7	0044	0055	0068	0083	0099	0118	0139	0163	0188	0216
8	0011	0015	0019	0025	0031	0038	0047	0057	0068	0081
9	0003	0004	0005	0007	0009	0011	0014	0018	0022	0027
10	0001	0001	0001	0002	0002	0003	0004	0005	0006	0008
11	0000	0000	0000	0000	0000	0001	0001	0001	0002	0002

m \ λ	**3**	**3,5**	**4**	**4,5**	**5**	**6**	**7**	**8**	**9**	**10**
0	0,0498	0302	0183	0111	0067	0025	0009	0003	0001	0000
1	1494	1057	0733	0500	0337	0149	0064	0027	0011	0005
2	2240	1850	1465	1125	0842	0446	0223	0107	0050	0023
3	2240	2158	1954	1687	1404	0892	0521	0286	0150	0076
4	1680	1888	1954	1898	1755	1339	0912	0573	0337	0189
5	1008	1322	1563	1708	1755	1606	1277	0916	0607	0378
6	0504	0771	1042	1281	1462	1606	1490	1221	0911	0631
7	0216	0385	0595	0824	1044	1377	1490	1396	1171	0901
8	0081	0169	0298	0463	0653	1033	1304	1396	1318	1126
9	0027	0066	0132	0232	0363	0688	1014	1241	1318	1251
10	0008	0023	0053	0104	0181	0413	0710	0993	1186	1251
11	0002	0007	0019	0043	0082	0225	0452	0722	0970	1137
12	0001	0002	0006	0016	0034	0113	0263	0481	0728	0948
13	0000	0001	0002	0006	0013	0052	0142	0296	0504	0729
14	0000	0000	0001	0002	0005	0022	0071	0169	0324	0521
15	0000	0000	0000	0001	0002	0009	0033	0090	0194	0347
16	0000	0000	0000	0000	0000	0003	0014	0045	0109	0217
17	0000	0000	0000	0000	0000	0001	0006	0021	0058	0128
18	0000	0000	0000	0000	0000	0000	0002	0009	0029	0071
19	0000	0000	0000	0000	0000	0000	0001	0004	0014	0037
20	0000	0000	0000	0000	0000	0000	0000	0002	0006	0019
21	0000	0000	0000	0000	0000	0000	0000	0001	0003	0009
22	0000	0000	0000	0000	0000	0000	0000	0000	0001	0004

Таблиця значень функції $t_\alpha = t(\alpha; n)$

n \ α	0,95	0,99	0,999	n \ α	0,95	0,99	0,999
5	2,78	4,60	8,61	**20**	2,093	2,861	3,883
6	2,57	4,03	6,86	**25**	2,064	2,797	3,745
7	2,45	3,71	5,96	**30**	2,045	2,756	3,659
8	2,37	3,50	5,41	**35**	2,032	2,720	3,600
9	2,31	3,36	5.04	**40**	2,023	2,708	3,558
10	2,26	3,25	4,78	**45**	2,016	2,692	3,527
11	2,23	3,17	4,59	**50**	2,009	2,679	3,502
12	2,20	3,11	4,44	**60**	2,001	2,662	3,464
13	2,18	3,06	4,32	**70**	1,996	2,649	3,439
14	2,16	3,01	4,22	**80**	1,991	2,640	3,418
15	2,15	2,98	4,14	**90**	1,987	2,633	3,403
16	2,13	2,95	4,07	**100**	1,984	2,627	3,392
17	2,12	2,92	4,02	**120**	1,980	2,617	3,374
18	2,11	2,90	3,97	∞	1,960	2,576	3,291
19	2,10	2,88	3,92				

Таблиця значень функції $q = q(\alpha; n)$

n \ α	0,95	0,99	0,999	n \ α	0,95	0,99	0,999
5	1,37	2,67	5,64	**20**	0,37	0,58	0,88
6	1,09	2,01	3,88	**25**	0,32	0,49	0,73
7	0,92	1,62	2,98	**30**	0,28	0,43	0,63
8	0,80	1,38	2,42	**35**	0,26	0,38	0,56
9	0,71	1,20	2,06	**40**	0,24	0,35	0,50
10	0,65	1,08	1,80	**45**	0,22	0,32	0,46
11	0,59	0,98	1,60	**50**	0,21	0,30	0,43
12	0,55	0,90	1,45	**60**	0,188	0,269	0,38
13	0,52	0,83	1,33	**70**	0,174	0,245	0,34
14	0,48	0,78	1,23	**80**	0,161	0,226	0,31
15	0,46	0,73	1,15	**90**	0,151	0,211	0,29
16	0,44	0,70	1,07	**100**	0,143	0,198	0,27
17	0,42	0,66	1,01	**150**	0,115	0,160	0,211
18	0,40	0,63	0,96	**200**	0,099	0,136	0,185
19	0,39	0,60	0,92	**250**	0,089	0,120	0,162

Критичні точки розподілу Стьюдента

Кількість ступенів вільності	Рівень значущості α (двостороння критична область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,71	31,82	63,66	318,29	636,58
2	2,92	4,30	6,96	9,92	22,33	31,60
3	2,35	3,18	4,54	5,84	1021	12,92
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,02	2,57	3,36	4,03	5,89	6,87
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,41
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,02	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,97
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,72	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,50	3,79
24	1,71	2,06	2,49	2,80	3,47	3,75
25	1,71	2,06	249	2,79	3,45	3,73
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	202	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29
	0,05	0,025	0,01	0,005	0,001	0,0005
	Рівень значущості α (одностороння критична область)					

Критичні точки розподілу χ^2

Кількість ступенів вільності k	Рівень значущості α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,63	5,02	3,84	0,0039	0,00098	0,00016
2	9,21	7,38	5,99	0,103	0,051	0,020
3	11,34	9,35	7,81	0,352	0,216	0,115
4	13,28	11,14	9,49	0,711	0,484	0,297
5	15,09	12,83	11,07	1,15	0,831	0,554
6	16,81	14,45	12,59	1,64	1,24	0,872
7	18,48	16,01	14,07	2,17	1,69	1,24
8	20,09	17,53	15,51	2,73	2,18	1,65
9	21,67	19,02	16,92	3,33	2,70	2,09
10	23,21	20,48	18,31	3,94	3,25	2,56
11	24,73	21,92	19,68	4,57	3,82	3,05
12	26,22	23,34	21,03	5,23	4,40	3,57
13	27,69	24,74	22,36	5,89	5,01	4,11
14	29,14	26,12	23,68	6,57	5,63	4,66
15	30,58	27,49	25,00	7,26	6,26	5,23
16	32,00	28,85	26,30	7,96	6,91	5,81
17	33,41	30,19	27,59	8,67	7,56	6,41
18	34,81	31,53	28,87	9,39	8,23	7,01
19	36,19	32,85	30,14	10,12	8,91	7,63
20	37,57	34,17	31,41	10,85	9,59	8,26
21	38,93	35,48	32,67	11,59	10,28	8,90
22	40,29	36,78	33,92	12,34	10,98	9,54
23	41,64	38,08	35,17	13,09	11,69	10,20
24	42,98	39,36	36,42	13,85	12,40	10,86
25	44,31	40,65	37,65	14,61	13,12	11,52
26	45,64	41,92	38,89	15,38	13,84	12,20
27	46,96	43,19	40,11	16,15	14,57	12,88
28	48,28	44,46	41,34	16,93	15,31	13,56
29	49,59	45,72	42,56	17,71	16,05	14,26
30	50,89	46,98	43,77	18,49	16,79	14,95

Критичні точки розподілу *F* Фішера–Снедекора

Рівень значущості α = 0,01												
k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5404	5624	5764	5859	5928	5981	6022	6056	6083	6107
2	98,50	99,00	99,16	99,25	99,30	99,33	99,36	99,38	99,39	99,40	99,41	99,42
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,96	9,89
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77	4,71
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62	3,55
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,43	3,37
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29	3,23
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,24	3,17
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,09	3,03
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	3,06	2,99
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	3,02	2,96
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,99	2,93
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,96	2,90
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,93	2,87
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,91	2,84
31	7,53	5,36	4,48	3,99	3,67	3,45	3,28	3,15	3,04	2,96	2,88	2,82
32	7,50	5,34	4,46	3,97	3,65	3,43	3,26	3,13	3,02	2,93	2,86	2,80
33	7,47	5,31	4,44	3,95	3,63	3,41	3,24	3,11	3,00	2,91	2,84	2,78
34	7,44	5,29	4,42	3,93	3,61	3,39	3,22	3,09	2,98	2,89	2,82	2,76
35	7,42	5,27	4,40	3,91	3,59	3,37	3,20	3,07	2,96	2,88	2,80	2,74
36	7,40	5,25	4,38	3,89	3,57	3,35	3,18	3,05	2,95	2,86	2,79	2,72
37	7,37	5,23	4,36	3,87	3,56	3,33	3,17	3,04	2,93	2,84	2,77	2,71
38	7,35	5,21	4,34	3,86	3,54	3,32	3,15	3,02	2,92	2,83	2,75	2,69
39	7,33	5,19	4,33	3,84	3,53	3,30	3,14	3,01	2,90	2,81	2,74	2,68
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,73	2,66
41	7,30	5,16	4,30	3,81	3,50	3,28	3,11	2,98	2,87	2,79	2,71	2,65

Рівень значущості α = 0,01												
k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
42	7,28	5,15	4,29	3,80	3,49	3,27	3,10	2,97	2,86	2,78	2,70	2,64
43	7,26	5,14	4,27	3,79	3,48	3,25	3,09	2,96	2,85	2,76	2,69	2,63
44	7,25	5,12	4,26	3,78	3,47	3,24	3,08	2,95	2,84	2,75	2,68	2,62
45	7,23	5,11	4,25	3,77	3,45	3,23	3,07	2,94	2,83	2,74	2,67	2,61
46	7,22	5,10	4,24	3,76	3,44	3,22	3,06	2,93	2,82	2,73	2,66	2,60
47	7,21	5,09	4,23	3,75	3,43	3,21	3,05	2,92	2,81	2,72	2,65	2,59
48	7,19	5,08	4,22	3,74	3,43	3,20	3,04	2,91	2,80	2,71	2,64	2,58
49	7,18	5,07	4,21	3,73	3,42	3,19	3,03	2,90	2,79	2,71	2,63	2,57
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70	2,63	2,56
55	7,12	5,01	4,16	3,68	3,37	3,15	2,98	2,85	2,75	2,66	2,59	2,53
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,50
65	7,04	4,95	4,10	3,62	3,31	3,09	2,93	2,80	2,69	2,61	3,53	2,47
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,51	2,45
75	6,99	4,90	4,05	3,58	3,27	3,05	2,89	2,76	2,65	2,57	2,49	2,43
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55	2,48	2,42
85	6,94	4,86	4,02	3,55	3,24	3,02	2,86	2,73	2,62	2,54	2,46	2,40
90	6,93	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52	2,45	2,39
95	6,91	4,84	3,99	3,52	3,22	3,00	2,83	2,70	2,60	2,51	2,44	2,38
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,43	2,37
125	6,84	4,78	3,94	3,47	3,17	2,95	2,79	2,66	2.55	2,47	2,39	2,33
150	6,81	4,75	3,91	3,45	3,14	2,92	2,76	2,63	2,53	2,44	2,37	2,31
175	6,78	4,73	3,90	3,43	3,12	2,91	2,74	2,61	2,51	2,42	2,35	2,29
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,34	2,27
300	6,72	4,68	3,85	3,38	3,08	2,86	2,70	2,57	2,47	2,38	2,31	2,24
400	6,70	4,66	3,83	3,37	3,06	2,85	2,68	2,56	245	2,37	2,29	2,23
500	6,69	4,65	3,82	3,36	3,05	2,84	2,68	2,55	2,44	2,36	2,28	2,22
700	6,67	4,64	3,81	3,35	3,04	2,83	2,66	2,54	2,43	2,35	2,27	2,21
1000	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34	2,27	2.20
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,25	2,18

Рівень значущості α = 0,01												
k_2 \ k_1	14	16	20	24	30	40	50	75	100	200	500	∞
1	6143	6170	6209	6234	6260	6286	6302	6324	6334	6350	6360	6366
2	99,43	99,44	99,45	99,46	99,47	99,48	99,48	99,48	99,49	99,49	99,50	99,50
3	26,92	26,83	26,69	26,60	26,50	26,41	26,35	26,28	26,24	26,18	26,15	26,13
4	14,25	14,15	14,02	13,93	13,84	13,75	13,69	13,61	13,58	13,52	13,49	13,46
5	9,77	9,68	9,55	9,47	9,38	9,24	9,29	9,17	9,13	9,08	9,04	9,02
6	7,60	7,52	7,40	7,31	7,23	7,14	7,09	7,02	6,99	6,93	6,90	6,88
7	6,36	6,28	6,16	6,07	5,99	5,91	5,86	5,79	5,75	5,70	5,67	5,65
8	5,56	5,48	5,36	5,28	5,20	5,12	5,07	5,00	4,96	4,91	4,88	4,86
9	5,01	4,92	4,81	4,73	4,65	4,57	4,52	4,45	4,41	4,36	4,33	4,31
10	4,60	4,52	4,41	4,33	4,25	4,17	4,12	4,05	4,01	3,96	3,93	3,91
11	4,29	4,21	4,10	4,02	3,94	3,86	3,81	3,74	3,71	3,66	3,62	3,60
12	4,05	3,97	3,86	3,78	3,70	3,62	3,57	3,50	3,47	3,41	3,38	3,36
13	3,86	3,78	3,66	3,59	3,51	3,43	3,38	3,31	3,27	3,22	3,19	3,17
14	3,70	3,62	3,51	3,43	3,35	3,27	3,22	3,14	3,11	3,06	3,03	3,00
15	3,56	3,49	3,37	3,29	3,21	3,13	3,08	3,01	2,98	2,92	2,89	2,87
16	3,45	3,37	3,26	3,18	3,10	3,02	2,97	2,90	2,86	2,81	2,78	2,75
17	3,35	3,27	3,16	3,08	3,00	2,92	2,87	2,80	2,76	2,71	2,68	2,65
18	3,27	3,19	3,08	3,00	2,92	2,84	2,78	2,71	2,68	2,62	2,59	2,57
19	3,19	3,12	3,00	2,92	2,84	2,76	2,71	2,64	2,60	2,55	2,51	2,49
20	3,13	3,05	2,94	2,86	2,78	2,69	2,64	2,57	2,54	2,48	2,44	2,42
21	3,07	2,99	2,88	2,80	2,72	2,64	2,58	2,51	2,48	2,42	2,38	2,36
22	3,02	2,94	2,83	2,75	2,67	2,58	2,53	2,46	2,42	2,36	2,33	2,31
23	2,97	2,89	2,78	2,70	2,62	2,54	2,48	2,41	2,37	2,32	2,28	2,26
24	2,93	2,85	2,74	2,66	2,58	2,49	2,44	2,37	2,33	2,27	2,24	2,21
25	2,89	2,81	2,70	2,62	2,54	2,45	2,40	2,33	2,29	2,23	2,19	2,17
26	2,86	2,78	2,66	2,58	2,50	2,42	2,36	2,29	2,25	2,19	2,16	2,13
27	2,82	2,75	2,63	2,55	2,47	2,38	2,33	2,26	2,22	2,16	2,12	2,10
28	2,79	2,72	2,60	2,52	2,44	2,35	2,30	2,23	2,19	2,13	2,09	2,06
29	2,77	2,69	2,57	2,49	2,41	2,33	2,27	2,20	2,16	2,10	2,06	2,03
30	2,74	2,66	2,55	2,47	2,39	2,30	2,25	2,17	2,13	2,07	2,03	2,01
31	2,72	2,64	2,52	2,45	2,36	2,27	2,22	2,14	2,11	2,04	2,01	1,98
32	2,70	2,62	2,50	2,42	2,34	2,25	2,20	2,12	2,08	2,02	1,98	1,96
33	2,68	2,60	2,48	2,40	2,32	2,23	2,18	2,10	2,06	2,00	1,96	1,93
34	2,66	2,58	2,46	2,38	2,30	2,21	2,16	2,08	2,04	1,98	1,94	1,91
35	2,64	2,56	2,44	2,36	2,28	2,19	2,14	2,06	2,02	1,96	1,92	1,89
36	2,62	2,54	2,43	2,35	2,26	2,18	2,12	2,04	2,00	1,94	1,90	1,87
37	2,61	2,53	2,41	2,33	2,25	2,16	2,10	2,03	1,98	1,92	1,88	1,85
38	2,59	2,51	2,40	2,32	2,23	2,14	2,09	2,01	1,97	1,90	1,86	1,84
39	2,58	2,50	2,38	2,30	2,22	2,13	2,07	1,99	1,95	1,89	1,85	1,82
40	2,56	2,48	2,37	2,29	2,20	2,11	2,06	1,98	1,94	1,87	1,83	1,80
41	2,55	2,47	2,36	2,28	2,19	2,10	2,04	1,97	1,92	1,86	1,82	1,79

Рівень значущості α = 0,01												
k_2 \ k_1	14	16	20	24	30	40	50	75	100	200	500	∞
42	2,54	2,46	2,34	2,26	2,18	2,09	2,03	1,95	1,91	1,85	1,80	1,78
43	2,53	2,45	2,33	2,25	2,17	2,08	2,02	1,94	1,90	1,83	1,79	1,76
44	2,52	2,44	2,32	2,24	2,15	2,07	2,01	1,93	1,89	1,82	1,78	1,75
45	2,51	2,43	2,31	2,23	2,14	2,05	2,00	1,92	1,88	1,81	1,77	1,74
46	2,50	2,42	2,30	2,22	2,13	2,04	1,99	1,91	1,86	1,80	1,76	1,73
47	2,49	2,41	2,29	2,21	2,12	2,03	1,98	1,90	1,85	1,79	1,74	1,71
48	2,48	2,40	2,28	2,20	2,12	2,02	1,97	1,89	1,84	1,78	1,73	1,70
49	2,47	2,39	2,27	2,19	2,11	2,02	1,96	1,88	1,83	1,77	1,72	1,69
50	2,46	2,38	2,27	2,18	2,10	2,01	1,95	1,87	1,82	1,76	1,71	1,68
55	2,42	2,34	2,23	2,15	2,06	1,97	1,91	1,83	1,78	1,71	1,67	1,64
60	2,39	2,31	2,20	2,12	2,03	1,94	1,88	1,79	1,75	1,68	1,63	1,60
65	2,37	2,29	2,17	2,09	2,00	1,91	1,85	1,77	1,72	1,65	1,60	1,57
70	2,35	2,27	2,15	2,07	1,98	1,89	1,83	1,74	1,70	1,62	1,57	1,54
75	2,33	2,25	2,13	2,05	1,96	1,87	1,81	1,72	1,67	1,60	1,55	1,52
80	2,31	2,23	2,12	2,03	1,94	1,85	1,79	1,70	1,65	1,58	1,53	1,49
85	2,30	2,22	2,10	2,02	1,93	1,83	1,77	7,69	1,64	1,56	1,51	1,47
90	2,29	2,21	2,09	2,00	1,92	1,82	1,76	1,67	1,62	1,55	1,49	1,46
95	2,28	2,20	2,08	1,99	1,90	1,81	1,75	1,66	1,61	1,53	1,48	1,44
100	2,27	2,19	2,07	1,98	1,89	1,80	1,74	1,65	1,60	1,52	1,47	1,43
125	2,23	2,15	2,03	1,94	1,85	1,76	1,69	1,60	1,55	1,47	1,41	1,37
150	2,20	2,12	2,00	1,92	1,83	1,73	1,66	1,57	1,52	1,43	1,38	1,33
175	2,19	2,10	1,98	1,90	1,81	1,71	1,64	1,55	1,50	1,41	1,35	1,30
200	2,17	2,09	1,97	1,89	1,79	1,69	1,63	1,53	1,48	1,39	1,33	1,28
300	2,14	2,06	1,94	1,85	1,76	1,66	1,59	1,50	1,44	1,35	1,28	1,22
400	2,13	2,05	1,92	1,84	1,75	1,64	1,58	1,48	1,42	1,32	1,25	1,19
500	2,12	2,04	1,92	1,83	1,74	1,63	1,57	1,47	1,41	1,31	1,23	1,16
600	2,11	2,03	1,91	1,82	1,73	1,63	1,56	1,46	1,40	1,30	1,22	1,15
700	2,11	2,03	1,90	1,82	1,72	1,62	1,55	1,45	1,39	1,29	1,21	1,14
800	2,10	2,02	1,90	1,81	1,72	1,62	1,55	1,45	1,39	1,29	1,20	1,13
1000	2,10	2,02	1,90	1,81	1,72	1,61	1,54	1,44	1,38	1,28	1,19	1,11
∞	2,08	2,00	1,88	1,79	1,70	1,59	1,52	1,42	1,36	1,25	1,15	1,00

Рівень значущості α = 0,05												
k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
1	161	199	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38
18	4,41	3,55	316	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,24	2,20
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,22	2,18
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,17	2,13
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,14	2,10
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09
31	4,16	3,30	2,91	2,68	2,52	2,41	2,32	2,25	2,20	2,15	2,11	2,08
32	4,15	3,29	2,90	2,67	2,51	2,40	2,31	2,24	2,19	2,14	2,10	2,07
33	4,14	3,28	2,89	2,66	2,50	2,39	2,30	2,23	2,18	2,13	2,09	2,06
34	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23	2,17	2,12	2,08	2,05
35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11	2,07	2,04
36	4,11	3,26	2,87	2,63	2,48	2,36	2,28	2,21	2,15	2,11	2,07	2,03
37	4,11	3,25	2,86	2,63	2,47	2,36	2,27	2,20	2,14	2,10	2,06	2,02
38	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,05	2,02
39	4,09	3,24	2,85	2,61	2,46	2,34	2,26	2,19	2,13	2,08	2,04	2,01
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00
41	4,08	3,23	2,83	2,60	2,44	2,33	2,24	2,17	2,12	2,07	2,03	2,00

Рівень значущості α = 0,05												
k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
42	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17	2,11	2,06	2,03	1,99
43	4,07	3,21	2,82	2,59	2,43	2,32	2,23	2,16	2,11	2,06	2,02	1,99
44	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10	2,05	2,01	1,98
45	4,06	3,20	2,81	2,58	2,42	2,31	2,22	2,15	2,10	2,05	2,01	1,97
46	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,15	2,09	2,04	2,00	1,97
47	4,05	3,20	2,80	2,57	2,41	2,30	2,21	2,14	2,09	2,04	2,00	1,96
48	4,04	3,19	2,80	2,57	2,41	2,29	2,21	2,14	2,08	2,03	1,99	1,96
49	4,04	3,19	2,79	2,56	2,40	2,29	2,20	2,13	2,08	2,03	1,99	1,96
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95
55	4,02	3,16	2,77	2,54	2,38	2,27	2,18	2,11	2,06	2,01	1,97	1,93
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92
65	3,99	3,14	2,75	2,51	2,36	2,24	2,15	2,08	2,03	1,98	1,94	1,90
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89
75	3,97	3,12	2,73	2,49	2,34	2,22	2,13	2,06	2,01	1,96	1,92	1,88
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88
85	3,95	3,10	2,71	2,48	2,32	2,21	2,12	2,05	1,99	1,94	1,90	1,87
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,98	1,93	1,89	1,86
95	3,94	3,09	2,70	2,47	2,31	2,20	2,11	2,04	1,98	1,93	1,89	1,86
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85
125	3,92	3,07	2,68	2,44	2,29	2,17	2,08	2,01	1,96	1,91	1,87	1,83
150	3,90	3,06	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,85	1,82
175	3,90	3,05	2,66	2,42	2,27	2,15	2,06	1,99	1,93	1,89	1,84	1,81
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,84	1,80
300	3,87	3,03	2,63	2,40	2,24	2,13	2,04	1,97	1,91	1,86	1,82	1,78
400	3,86	3,02	2,63	2,39	2,24	2,12	2,03	1,96	1,90	1,85	1,81	1,78
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,81	1,77
600	3,86	3,01	2,62	2,39	2,23	2,11	2,02	1,95	1,90	1,85	1,80	1,77
700	3,85	3,01	2,62	2,38	2,23	2,11	2,02	1,95	1,89	1,84	1,80	1,77
800	3,85	3,01	2,62	2,38	2,23	2,11	2,02	1,95	1,89	1,84	1,80	1,76
1000	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84	1,80	1,76
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,79	1,75

Рівень значущості α = 0,05												
k_2 \ k_1	14	16	20	24	30	40	50	75	100	200	500	∞
1	245	246	248	249	250	251	252	253	253	254	254	254
2	19,42	19,43	19,45	19,45	19,46	19,47	19,48	19,48	19,49	19,49	19,49	19,50
3	8,71	8,69	8,66	8,64	8,62	8,59	8,58	8,56	8,55	8,54	8,53	8,53
4	5,87	5,84	5,80	5,77	5,75	5,72	5,70	5,68	5,66	5,65	5,64	5,63
5	4,64	4,60	4,56	4,53	4,50	4,46	4,44	4,42	4,41	4,39	4,37	4,37
6	3,96	3,92	3,87	3,84	3,81	3,77	3,75	3,73	3,71	3,69	3,68	3,67
7	3,53	3,49	3,44	3,41	3,38	3,34	3,32	3,29	3,27	3,25	3,24	3,23
8	3,24	3,20	3,15	3,12	3,08	3,04	3,02	2,99	2,97	2,95	2,94	2,93
9	3,03	2,99	2,94	2,90	2,86	2,83	2,80	2,77	2,76	2,73	2,72	2,71
10	2,86	2,83	2,77	2,74	2,70	2,66	2,64	2,60	2,59	2,56	2,55	2,54
11	2,74	2,70	2,65	2,61	2,57	2,53	2,51	2,47	2,46	2,43	2,42	2,40
12	2,64	2,60	2,54	2,51	2,47	2,43	2,40	2,37	2,35	2,32	2,31	2,30
13	2,55	2,51	2,46	2,42	2,38	2,34	2,31	2,28	2,26	2,23	2,22	2,21
14	2,48	2,44	2,39	2,35	2,31	2,27	2,24	2,21	2,19	2,16	2,14	2,13
15	2,42	2,38	2,33	2,29	2,25	2,20	2,18	2,14	2,12	2,10	2,08	2,07
16	2,37	2,33	2,28	2,24	2,19	2,15	2,12	2,09	2,07	2,04	2,02	2,01
17	2,33	2,29	2,23	2,19	2,15	2,10	2,08	2,04	2,02	1,99	1,97	1,96
18	2,29	2,25	2,19	2,15	2,11	2,06	2,04	2,00	1,98	1,95	1,93	1,92
19	2,26	2,21	2,16	2,11	2,07	2,03	2,00	1,96	1,94	1,91	1,89	1,88
20	2,22	2,18	2,12	2,08	2,04	1,99	1,97	1,93	1,91	1,88	1,86	1,84
21	2,20	2,16	2,10	2,05	2,01	1,96	1,94	1,90	1,88	1,84	1,83	1,81
22	2,17	2,13	2,07	2,03	1,98	1,94	1,91	1,87	1,85	1,82	1,80	1,78
23	2,15	2,11	2,05	2,01	1,96	1,91	1,88	1,84	1,82	1,79	1,77	1,76
24	2,13	2,09	2,03	1,98	1,94	1,89	1,86	1,82	1,80	1,77	1,75	1,73
25	2,11	2,07	2,01	1,96	1,92	1,87	1,84	1,80	1,78	1,75	1,73	1,71
26	2,09	2,05	1,99	1,95	1,90	1,85	1,82	1,78	1,76	1,73	1,71	1,69
27	2,08	2,04	1,97	1,93	1,88	1,84	1,81	1,76	1,74	1,71	1,69	1,67
28	2,06	2,02	1,96	1,91	1,87	1,82	1,79	1,75	1,73	1,69	1,67	1,65
29	2,05	2,01	1,94	1,90	1,85	1,81	1,77	1,73	1,71	1,67	1,65	1,64
30	2,04	1,99	1,93	1,89	1,84	1,79	1,76	1,72	1,70	1,66	1,64	1,62
31	2,03	1,98	1,92	1,88	1,83	1,78	1,75	1,70	1,68	1,65	1,62	1,61
32	2,01	1,97	1,91	1,86	1,82	1,77	1,74	1,69	1,67	1,63	1,61	1,59
33	2,00	1,96	1,90	1,85	1,81	1,76	1,72	1,68	1,66	1,62	1,60	1,58
34	1,99	1,95	1,89	1,84	1,80	1,75	1,71	1,67	1,65	1,61	1,59	1,57
35	1,99	1,94	1,88	1,83	1,79	1,74	1,70	1,66	1,63	1,60	1,57	1,56
36	1,98	1,93	1,87	1,82	1,78	1,73	1,69	1,65	1,62	1,59	1,56	1,55
37	1,97	1,93	1,86	1,82	1,77	1,72	1,68	1,64	1,62	1,58	1,55	1,54
38	1,96	1,92	1,85	1,81	1,76	1,71	1,68	1,63	1,61	1,57	1,54	1,53
39	1,95	1,91	1,85	1,80	1,75	1,70	1,67	1,62	1,60	1,56	1,53	1,52
40	1,95	1,90	1,84	1,79	1,74	1,69	1,66	1,61	1,59	1,55	1,53	1,51
41	1,94	1,90	1,83	1,79	1,74	1,69	1,65	1,61	1,58	1,54	1,52	1,50

Рівень значущості α = 0,05												
k_2 \ k_1	14	16	20	24	30	40	50	75	100	200	500	∞
42	1,94	1,89	1,83	1,78	1,73	1,68	1,65	1,60	1,57	1,53	1,51	1,49
43	1,93	1,89	1,82	1,77	1,72	1,67	1,64	1,59	1,57	1,53	1,50	1,48
44	1,92	1,88	1,81	1,77	1,72	1,67	1,63	1,59	1,56	1,52	1,49	1,48
45	1,92	1,87	1,81	1,76	1,71	1,66	1,63	1,58	1,55	1,51	1,49	1,47
46	1,91	1,87	1,80	1,76	1,71	1,65	1,62	1,57	1,55	1,51	1,48	1,46
47	1,91	1,86	1,80	1,75	1,70	1,65	1,61	1,57	1,54	1,50	1,47	1,46
48	1,90	1,86	1,79	1,75	1,70	1,64	1,61	1,56	1,54	1,49	1,47	1,45
49	1,90	1,85	1,79	1,74	1,69	1,64	1,60	1,56	1,53	1,49	1,46	1,44
50	1,89	1,85	1,78	1,74	1,69	1,63	1,60	1,55	1,52	1,48	1,46	1,44
55	1,88	1,83	1,76	1,72	1,67	1,61	1,58	1,53	1,50	1,46	1,43	1,41
60	1,86	1,82	1,75	1,70	1,65	1,59	1,56	1,51	1,48	1,44	1,41	1,39
65	1,85	1,80	1,73	1,69	1,63	1,58	1,54	1,49	1,46	1,42	1,39	1,37
70	1,84	1,79	1,72	1,67	1,62	1,57	1,53	1,48	1,45	1,40	1,37	1,35
75	1,83	1,78	1,71	1,66	1,61	1,55	1,52	1,47	1,44	1,39	1,36	1,34
80	1,82	1,77	1,70	1,65	1,60	1,54	1,51	1,45	1,43	1,38	1,35	1,32
85	1,81	1,76	1,70	1,65	1,59	1,54	1,50	1,45	1,42	1,37	1,34	1,31
90	1,80	1,75	1,68	1,64	1,59	1,53	1,49	1,44	1,41	1,36	1,33	1,30
95	1,80	1,75	1,68	1,63	1,58	1,52	1,48	1,43	1,40	1,35	1,32	1,29
100	1,79	1,75	1,68	1,63	1,57	1,52	1,48	1,42	1,39	1,34	1,31	1,28
125	1,77	1,73	1,66	1,60	1,55	1,49	1,45	1,40	1,36	1,31	1,27	1,25
150	1,76	1,71	1,64	1,59	1,54	1,48	1,4	1,38	1,34	1,29	1,25	1,22
175	1,75	1,70	1,63	1,58	1,52	1,46	1,42	1,36	1,33	1,27	1,23	1,20
200	1,74	1,69	1,62	1,57	1,52	1,46	1,41	1,35	1,32	1,26	1,22	1,19
300	1,72	1,68	1,61	1,5	1,50	1,43	1,39	1,33	1,30	1,23	1,19	1,15
400	1,72	1,67	1,60	1,54	1,49	1,42	1,38	1,32	1,28	1,22	1,17	1,13
500	1,71	1,6	1,59	1,54	1,48	1,42	1,38	1,31	1,28	1,21	1,16	1,11
600	1,71	1,6	1,59	1,54	1,48	1,41	1,37	1,31	1,27	1,20	1,15	1,10
700	1,71	1,66	1,59	1,53	1,48	1,41	1,37	1,30	1,27	1,20	1,15	1,09
800	1,70	1,6	1,58	1,53	1,47	1,41	1,37	1,30	1,26	1,20	1,14	1,09
1000	1,70	1,65	1,58	1,53	1,47	1,41	1,36	1,30	1,26	1,19	1,13	1,08
∞	1,69	1,64	1,57	1,52	1,46	1,39	1,35	1,28	1,24	1,17	1,11	1,00

НАВЧАЛЬНЕ ВИДАННЯ

ОГІРКО Ольга Ігорівна,
кандидат технічних наук
ГАЛАЙКО Наталія Володимирівна

ТЕОРІЯ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНА СТАТИСТИКА

Навчальний посібник

Редагування *І. Б. Попик*

Макетування *Н. М. Лесь*

Друк *А. М. Радченко*

Підписано до друку 28.12.2017.
Формат 60×84/8. Папір офсетний. Умовн. друк. арк. 33,94.
Тираж 100 прим. Зам. № 145-17.

Львівський державний університет внутрішніх справ
Україна, 79007, м. Львів, вул. Городоцька, 26.

Свідоцтво про внесення суб’єкта видавничої справи до державного реєстру
видавців, виготівників і розповсюджувачів видавничої продукції
ДК № 2541 від 26 червня 2006 р.