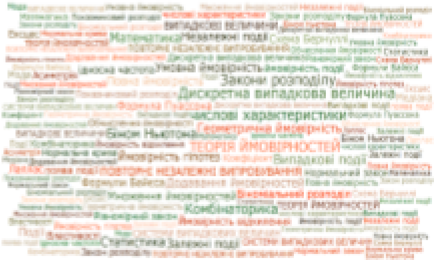
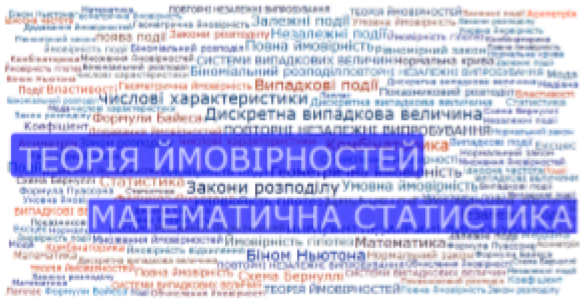
Савченко О.Г. Кавун Г.М. Валько Н.В. Кузьмич Л.В.





Савченко О.Г. Кавун Г.М.
Валько Н.В. Кузьмич Л.В.

Савченко О.Г.
Кавун Г.М.
Валько Н.В.
Кузьмич Л.В.

Теорія ймовірностей та математична статистика

Базовий курс з прикладами і задачами

Херсон – Айлант – 2017

УДК 519.21 : 519.22

Рекомендовано до друку рішенням вченої ради Херсонського державного аграрного університету, протокол №11 від 29 травня 2017 р.

Рецензенти:

Львов М.С. – доктор фізико-математичних наук, професор кафедри інформатики, програмної інженерії та економічної кібернетики Херсонського державного університету.

Котова О.В. – кандидат фізико-математичних наук, доцент кафедри алгебри, геометрії та математичного аналізу Херсонського державного університету.

Савченко О.Г., Валько Н.В., Кавун Г.М., Кузьмич Л.В.

С13 Теорія ймовірностей та математична статистика: [базовий курс з прикладами і задачами] – Херсон: Айлант, 2017. – 400 с.

ISBN 978-966-630-165-2

Базовий курс дисципліни «Теорія ймовірностей та математична статистика» містить загальний інформативний блок, методичні рекомендації до вивчення курсу, питання для самоконтролю з усіх тем навчального посібника, зразки розв'язування задач, завдання для аудиторної та індивідуальної роботи з усіх тем, методичні поради щодо їх виконання, питання для самоперевірки, зразки типових котрольних робіт, термінологічний словник та список рекомендованої літератури.

Для підготовки бакалаврів та магістрів з економіки. Рекомендований для студентів економічних спеціальностей вищих навчальних закладів та коледжів, які вивчають курс «Теорія ймовірностей та математична статистика». Навчальний посібник може бути корисним працівникам інших спеціальностей як довідник з теорії ймовірностей та математичної статистики.

ISBN 978-966-630-165-2

ЗМІСТ

ЧАСТИНА 1
ТЕОРІЯ ЙМОВІРНОСТЕЙ

Теорія ймовірностей - це математична наука, яка вивчає закономірності масових випадкових явищ. Під *випадковими явищами* розуміють явища з невизначеним наслідком, які відбуваються при повторному відтворенні одного й того ж випробування (досліду). Як би точно та детально не були фіксовані умови досліду, неможливо досягнути того, щоб при повторенні досліду результати повністю та в точності співпадали. Наприклад: при підкиданні монети неможливо вгадати, якою стороною вона впаде; результати декількох вимірів однієї й тієї ж величини одним приладом при одних і тих самих умовах різні; практично нереально врахувати залежність об'єму продаж товару від попиту і т. д. Неоднозначність результату в наведених прикладах пояснюється тим, що при зберіганні основних умов досліду не враховується зв'язок з другорядними факторами, які змінюються від одного досліду до іншого і вносять випадкові відмінності (елемент невизначеності) в їхні результати. Елемент невизначеності, складності, багатопричинності, що притаманний випадковим явищам, вимагає створення спеціальних методів для вивчення цих явищ. Розробкою таких методів і можливістю прогнозувати появу випадкових явищ займається теорія ймовірностей. Її *предметом* є специфічні закономірності, які спостерігаються у випадкових явищах; основною *задачею*, яка повинна бути вирішена в процесі вивчення дисципліни, - навчити студентів виконувати якісний і кількісний математичний аналіз випадкових явищ, випадкових величин і систем таких величин. Тому методи теорії ймовірностей все ширше й ширше проникають у різні області науки та техніки, сприяючи їхньому прогресу.

РОЗДІЛ 1
ОСНОВНІ ПОНЯТТЯ ТЕОРІЇ ЙМОВІРНОСТЕЙ

1.1. Передмова

Сучасний розвиток суспільства характеризується високим розвитком науки й техніки, які не можливі без використання математичних досліджень і розрахунків. Тому спеціалісти різних областей виробництва повинні володіти математичним апаратом, що необхідний для розв'язання багатьох важливих задач економічної науки. Особливо важливе місце в їх роботі повинні займати задачі планування та раціонального використання ресурсів, а також розробка принципів ефективного розміщення продуктивних сил.

При узагальненні явищ, які спостерігаються в економічному житті суспільства, часто приходиться будувати математичне вираження закономірностей, які характеризують деякі процеси та явища. При науковому дослідженні фізичних і технічних задач часто приходиться зустрічатися з особливого типу явищами, які називають *випадковими.* Випадкові відхилення невід'ємно супроводжують будь-яке закономірне явище. При цьому з нескінченної множини факторів, які впливають на дане явище, виділяються основні, а впливом другорядних факторів просто нехтують. Така схема вивчення випадкових явищ постійно використовується у фізиці, механіці, техніці. Очевидно, що повинна існувати принципова різниця в методах обліку основних, вирішальних факторів і другорядних. Теорія ймовірностей - це математична наука, яка вивчає ці явища. Предметом теорії ймовірностей є вивчення ймовірнісних закономірностей масових, однорідних, випадкових подій стійкої частоти.

Науку, що використовує теорію ймовірностей для обробки численних одиниць інформації як наслідків експерименту, називають *математичною статистикою*. Статистичні закономірності спостерігаються завжди, коли ми маємо справу з масою випадкових однорідних явищ. Чим більша кількість випадкових однорідних явищ, тим значніше та відчутніше проявляються специфічні закони, тим з більшою ймовірністю та точністю можна здійснити науковий прогноз.

Ймовірнісний або статистичний метод у науці є доповненням, який дає можливість глибше аналізувати явище з врахуванням притаманних йому елементів випадковості.

Характерним для сучасного етапу розвитку природничих і технічних наук є широке та плідне використання статистичних методів у всіх областях знань. Це очевидно, оскільки при глибшому вивченні довільної кількості явищ, обов'язково наступає етап, коли потрібно виявити не тільки основні закономірності, але й проаналізувати можливі відхилення від них. У теперішній час немає майже жодної науки, в якій так чи інакше не використовувалися б імовірнісні методи. Цілі розділи сучасної фізики базуються на методах теорії ймовірностей. Все ширше використовуються ймовірнісні методи в сучасній електротехніці та радіотехніці, метереології та астрономії, теорії автоматичного управління. Методи теорії ймовірностей і математичної статистики знайшли також широке застосування в різних галузях виробництва та господарювання: теорії надійності, теорії масового обслуговування, теорії інформації, економетричному моделюванні, теорії помилок спостереження; при плануванні та організації виробництва, аналізі технологічних процесів, контролю якості продукції та ін. У свою чергу, теорія ймовірностей і математична статистика опираються на інші математичні дисципліни: функціональний аналіз, алгебру, аналітичну геометрію, теорію функцій комплексної змінної.

Математичні закони теорії ймовірностей – це відображення реальних статистичних законів, що об'єктивно існують у масових випадкових явищах природи та суспільства. Для вивчення цих явищ теорія ймовірностей використовує математичний метод, тому є одним із розділів математики, таким же логічно точним і строгим, як й інші математичні науки.

У статистичних законах необхідно враховувати не лише основні фактори, але і множину другорядних, які приводять до випадкових похибок результату, тобто вносять в нього елемент невизначеності. У такому випадку неможливо однозначно прогнозувати результат при заданих умовах. Але якщо випадкові явища повторяють багаторазово при впливі певного комплексу умов, то середній результат вийде практично уже не випадковим. Отже, багаторазово підтверджена досвідом стійкість масових випадкових явищ і служить базою для використання ймовірнісних (статистичних) методів дослідження.

На теорію ймовірностей опирається математична статистика, задача якої полягає в тому, щоб за обмеженими даними (вибіркою) встановити з визначеним степенем достовірності характеристики, які притаманні генеральній сукупності, тобто всьому набору даних, які описують явище, що вивчається. Зв'язок між математичною статистикою та теорією ймовірностей обумовлений тим, що методи, які розробляються в математичній статистиці, зумовлюють отримання даних у ході вибіркових спостережень або дослідів. Використовуючи результати, отримані в теорії ймовірностей, математична статистика не тільки визначає значення шуканих характеристик, але й дає можливість оцінити ступінь точності отриманих висновків при обробці вибіркових даних.

1.2. Історична довідка

Теорія ймовірностей, подібно до інших математичних наук, виникла та розвивалась на основі потреб практики, в процесі розв'язування ряду окремих задач ігрового та прикладного характеру. Перші поняття, які дійшли до нас, відносяться до XVI - XVII ст. і пов'язані з розв'язуванням задач, що виникали в азартних іграх (слово "азарт" в перекладі з французької мови означає випадок). Вони належали Б. Паскалю (1623–1662), П.Ферма (1601–1665), Д. Кардано (1501–1576) та іншим видатним математикам. Початок систематичного дослідження задач, які відносяться до масових випадкових явищ, і поява відповідного математичного апарату відносяться до XVII ст.

Вирішальним кроком вперед у розвитку теорії ймовірностей послугували роботи Я. Бернуллі (1654–1705), який довів одне з найважливіших положень теорії ймовірностей – так званий закон великих чисел.

Другий важливий етап у розвитку теорії ймовірностей пов'язаний з ім'ям А. Муавра (1667–1754). Цей учений вперше ввів і обґрунтував закон, який дуже часто спостерігається у випадкових явищах - нормальний закон розподілу (закон Гаусса). Теореми, які обґрунтовують цей закон для тих чи інших умов, носять в теорії ймовірностей загальну назву "центральної граничної теореми".

Визначна роль у розвитку теорії ймовірностей належить знаменитому математику П. Лапласу (1749–1827). Він уперше розробив чітке та систематичне положення основ теорії ймовірностей, дав доведення одній із форм центральної граничної теореми (теореми Муавра-Лапласа) та розвинув ряд важливих додатків теорії ймовірностей до питань практики, частково до аналізу помилок спостережень і вимірів.

Значний крок вперед у розвитку теорії ймовірностей пов'язаний з іменем К. Гаусса (1777–1855), який дав ще більш загальне обґрунтування нормальному закону та розробив метод обробки експериментальних даних, що відомий під назвою "метод найменших квадратів". Слід відзначити роботи С. Пуассона (1781–1840), який довів більш загальну, ніж у Я. Бернуллі, форму закону великих чисел, а також закон розподілу дискретних випадкових величин.

Третій період розвитку теорії ймовірностей (друга половина XIX ст.) збігається зі створенням у Росії Петербурзької математичної школи. Її членами були В.Я. Буняковський (1804–1889) – автор першого курсу теорії ймовірностей російською мовою, та його учень П.Л. Чебишов (1821–1894), якому належить подальше розширення та узагальнення закону великих чисел (нерівність та теорема Чебишова). Учнями Чебишова були А.А. Марков і О.М. Ляпунов, які розширили сферу застосування закону великих чисел і центральної граничної теореми.

Петербурзька математична школа послугувала основою для подальшого розвитку теорії ймовірностей радянськими вченими: С.Н. Бернштейном, О.Я. Хінчіним, А.М. Колмогоровим.

У XIX ст. теорію ймовірностей почали успішно застосовувати в страховій справі, артилерії, статистиці.

Як самостійний розділ математики, теорія ймовірностей сформувалися в кінці XIX – на початку XX ст., причому, її бурний розвиток почався лише в XX ст. Остаточне становлення теорії ймовірностей пов'язане з ім’ям А.М. Колмогорова (1903–1987), який дав найдосконалішу аксіоматичну побудову теорії ймовірностей. Її аксіоми та теореми в абстрактній формі виражають закономірності, що притаманні випадковим явищам. У теперішній час аксіоматичний підхід є загальноприйнятим.

Значний внесок у подальший розвиток і практичне застосування теорії ймовірностей належить українським вченим: Б.В. Гнєденку, А.В. Скороходу, В.С. Королюку, І.І. Гіхману, І.М. Коваленку та іншим.

Розвиток зарубіжної теорії ймовірностей у теперішній час також іде збільшеними темпами у зв'язку з нагальними вимогами практики. Визначні роботи в цій області належать Н. Венеру, В. Феллеру, Д. Дубу, Р. Фішеру.

Сучасний розвиток теорії ймовірностей характеризується загальною зацікавленістю нею, а також широким її практичним застосуванням. У США, Франції, Україні, Швеції, Японії, Великобританії, Польщі, Італії, Угорщині та інших країнах світу є чимало вчених, які збагачують теорію ймовірностей важливими результатами.

За останні роки ми є свідками зародження нових і своєрідних методів прикладної теорії ймовірностей, поява яких пов'язана зі специфікою досліджуваних технічних проблем. Мова йде, зокрема, про такі нові дисципліни, як "теорія інформації" та "теорія масового обслуговування". Ці розділи теорії ймовірностей виникли із безпосередніх потреб практики та набувають загального технічного значення, а сфера їхнього використання швидко збільшується.

1.3. Елементи комбінаторики

Комбінаторика - один з розділів дискретної математики, який набув важливого значення у зв’язку з використанням його в теорії ймовірностей, математичній логіці, теорії чисел, обчислювальній техніці, кібернетиці. Комбінаторика має опосередковане відношення до теорії ймовірностей, перш за все завдяки класичному способу обчислення ймовірностей, при якому потрібно знаходити число всіх елементарних результатів появи події та число результатів, які сприяють її появі; при обчисленні ймовірностей різних подій, пов'язаних з експериментами, які мають скінчене число наслідків. Крім теорії ймовірностей, комбінаторика використовується в теорії обчислювальних машин,теорії автоматів, у деяких задачах економіки, природознавства і т.д.

Часто при вивченні закономірностей деяких об'єктів (предметів, явищ) приходиться підраховувати число всіх можливих способів розташування певних предметів або число всіх можливих сpособів здійснення дії; вибирати об'єкти з деякої множини та поєднувати у визначені групи за якимись заздалегідь заданими ознаками, розміщувати елементи однієї або декількох груп у вказаному порядку. Будь-які групи об'єктів (предметів, явищ), об'єднаних за якоюсь ознакою, називаються комбінаціями або сполуками. Об'єкти, з яких складаються сполуки, називаються їх елементами. Прикладами сполук можуть служити: будинки, вулиці у місті, автоколони машин (вантажні, таксопарк, автобуси), групи студентів і т.д.

Множину, що складається з n елементів, називають *упорядкованою*, якщо кожному її елементу можна поставити у відповідність число (номер елемента) від 1 до n так, що різним елементам відповідають різні числа (номери).

Розглянемо найпростіші види комбінаторних сполук, якими є розміщення, перестановки, комбінації (сполучення).

Розміщення. Нехай задано деяку множину, яка складається з n елементів $a_1, a_2, \dots, a_n$. Із цієї множини можна утворити підмножини з m елементів ($0 \leq m \leq n$), що обрані з даних n елементів, які відрізняються одна від одної або самими елементами, або порядком їх розташування. Такі множини називаються розміщеннями з n елементів по m і позначаються A_n^m.

Отже, *розміщеннями з n елементів по m елементів* називають різні впорядковані m-елементні підмножини, що вибрані із n-елементної множини ($m \le n$).

Виведемо формулу для підрахунку кількості розміщень без повторень. Для цього спочатку складемо всілякі розміщення цих n елементів по два елементи в кожному. Візьмемо елемент a_i і приєднаємо до кожного з інших (n-1) елементів. Потім приєднаємо елемент a_2до всіх інших (n-1) елементів і т.д. У результаті отримаємо такі множини з n елементів по два: $a_1a_2,\ a_1a_3, \dots, a_1a_n, a_2a_1, a_2a_3, \dots, a_2a_n, \dots, a_na_1, \dots, a_na_{n-1}$. У кожному рядку міститься (n-1) розміщень, а всього рядків n.

Отже, загальна кількість розміщень з n елементів по 2 обчислюється за формулою $A_n^2 = n(n-1)$. Аналогічно,

$A_n^3 = n(n-1).(n-2)$.

У загальному вигляді формула числа розміщень із n елементів по m має вигляд:

$$A_n^m = \frac{n!}{(n-m)!} = n(n-1).(n-2)...(n-m+1), \tag{1.3.1}$$

де $m \le n$. Зазначимо, що $A_n^n = n!$

Символ ***n*!** (ен-факторіал) позначає добуток усіх натуральних чисел від 1 до даного числа n включно: $n!=1\cdot 2\cdot 3\cdot \ldots \cdot n$.

За означенням **0! = 1**.

Наприклад, 4! = 1·2·3·4 = 24.

Приклад 1.3.1. Скільки двозначних чисел можна скласти із цифр 1, 2, 3, 4, 5, 6; якщо цифри в запису двозначного числа не повторюються?

Розв'язання. За умовою з 6 цифр треба скласти двозначні числа . Числа відрізняються не тільки складом, а й порядком цифр їх запису. Так як у запису двозначного числа цифри не повторюються, то їх кількість буде дорівнювати числу розміщень із шести елементів по два елементи:

$$N = A_6^2 = \frac{6!}{(6-2)!} = \frac{6!}{4!} = \frac{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6}{1\cdot 2\cdot 3\cdot 4} = 5\cdot 6 = 30.$$

Перестановки. Впорядковані множини, які утворені з усіх елементів множини, що містить n елементів, називають *перестановками*.

Можна дати й інше означення: *перестановками із n елементів* називаються розміщення, які складені по n елементів із даних n елементів, і відмінні одне від одного лише порядком розташування елементів.

Отже, перестановки є частинним випадком розміщень, коли в кожні сполуки входять всі елементи. Число перестановок із n елементів позначають $\boldsymbol{P_n}$ і обчислюють за формулою:

$$P_n = 1\cdot 2\cdot 3\cdot \ldots \cdot n = n! \qquad (1.3.2)$$

Приклад 1.3.2. Скількома способами можна розмістити на полиці 6 різних книжок?

Розв'язання. Одне розміщення книг на полиці буде відрізнятися від іншого лише порядком їх розташування, тому шукане число способів дорівнює числу способів впорядкування множини, яка містить 6 елементів, тобто числу перестановок із 6 елементів:

$$P_6 = 6! = 1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6 = 720.$$

Комбінації (сполуки). Різні невпорядковані m-елементні підмножини множини, що містить n елементів ($m \le n$), називають комбінаціями з n елементів по m. Дві комбінації вважаються різними, якщо вони відрізняються одна від одної хоча б одним елементом (порядок неістотний!). Прикладом сполук є різні варіанти груп у складі m осіб, відібраних із групи n осіб, що відрізняються один від одного хоча б однією людиною. Кількість сполук із n елементів по m дорівнює

$$C_n^m = \frac{n!}{m!(n-m)!}; \qquad C_n^m = \frac{n(n-1)(n-2)...(n-m+1)}{m!} \tag{1.3.3}$$

де $m \le n$.

Справедливі такі співвідношення:

$$C_n^m = C_n^{n-m}; \qquad C_n^0 = C_n^n = 1;$$
$$C_m^k + C_m^{k+1} = C_{m+1}^{k+1}; \qquad C_n^0 + C_n^1 + C_n^2 + ... + C_n^n = 2^n. \tag{1.3.4}$$

Приклад 1.3.3. Бригадир повинен відправити на роботу бригаду із 5 робітників. Скільки бригад по 5 робітників в кожній можна сформувати із 12 робітників?

Розв'язання. З 12 робітників виділяються бригади по 5 робітників. У якому порядку будуть перераховані прізвища 5 робітників - немає різниці. Бригади вважаються різними, якщо вони відрізняються хоча б однією людиною. Множина із 5 робітників не є впорядкованою, отже, число способів сформувати бригаду дорівнює числу сполук із 12 елементів по 5 елементів:

$$C_{12}^5 = \frac{12!}{(12-5)!5!} = \frac{12!}{7!5!} = \frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 11 \cdot 9 \cdot 8 = 792.$$

Перестановки з повтореннями. Множину, яка складається з n елементів, завжди можна подати у вигляді різних m об'єднань множин $B_1, B_2, \ldots, B_m$, які попарно не перетинаються та містять відповідно $k_1, k_2, \ldots, k_m$ елементів ($k_1+k_2+\ldots+k_m = n$). Такі множини називаються *перестановками з повтореннями.* Число всіх різних перестановок з повтореннями вказаного складу обчислюється за формулою

$$P_n(k_1; k_2; \ldots; k_m) = \frac{\mathrm{n}!}{k^1!\, k^2! \ldots k_m!} \tag{1.3.5}$$

Приклад 1.3.4. Скількома способами можна розселити 8 студентів у три кімнати: одномісну, тримісну, чотиримісну?

Розв’язання. Очевидно, що будь-якому розселенню однозначно відповідає вибірка заданого складу (1; 3; 4) об'ємом 1+3+4=8. Отже, шукане число способів рівне

$$P_8(1; 3; 4) = \frac{8!}{1!3!4!} = 280.$$

Розміщення з повтореннями*.* Якщо в розміщеннях з *n* елементів по *m* елементів деякі (або всі) елементи можуть бути однакові, то такі розміщення називають розміщеннями з повтореннями з *n* елементів по *m* елементів*.*

Кількість усіх розміщень з повтореннями позначають $\overline{A_n^m}$ і обчислюють за формулою

$$\overline{A_n^m} = n^m. \tag{1.3.6}$$

Комбінації з повтореннями. Якщо з множини, яка складається з *n* елементів, вибираються *m*-елементні підмножини з повтореннями, які відрізняються між собою лише складом елементів (хоча б одним), то їх називають комбінаціями з повтореннями з *n* елементів по *m* (*m*$\geq n$)*.* Або: якщо в комбінаціях з *n* елементів по *m* елементів деякі (або всі) елементи можуть бути однакові, то такі комбінації називають комбінаціями з повтореннями з *n* елементів по *m* елементів*.*

Наприклад, із трьох елементів a_1, a_2, a_3 по два можна створити 6 сполучень з повтореннями : a_1a_1; a_1a_2; a_1a_3; a_2a_2; a_2a_3; a_3a_3.

Число всіх сполучень з повтореннями позначається $\overline{\mathrm{C}_n^m}$ і обчислюється за формулою:

$$\overline{\mathrm{C}_n^m} = \mathrm{C}_{n+m-1}^m = \frac{(n+m-1)!}{m!\,(n-1)!} \qquad (1.3.7)$$

Приклад 1.3.5. Скількома способами можна відібрати 3 із 12 букв*: В, В, В, Д, Д, Д, Г, Г, Г, М, М, М ?*

Розв'язання. У даному випадку *n*= 4 (є 4 різних видів букв *В, Д, Г, М),* а *m*=3.

Тому шукане число знайдемо за формулою

$$\overline{C_4^3} = C_6^3 = \frac{6!}{3!3!} = \frac{4\cdot5\cdot6}{1\cdot2\cdot3} = 4\cdot5 = 20\,.$$

Тепер ознайомимося з *основними правилами (принципами)* комбінаторики. Ними є правило суми та правило добутку.

Правило суми. Якщо першу дію можна виконати n_1 способами, другу - n_2 способами і т. д., *k*-ту дію - n_k способами, то хоча б одну з цих дій (або 1-шу дію, або 2-гу дію, …, або *k*-ту дію) можна виконати $n_1 + n_2 + \cdots + n_k$ способами.

Правило суми можна формулювати й інакше: якщо деякий об'єкт N можна вибрати n різними способами, а об'єкт M – *m* різними способами, причому будь-який спосіб вибору об'єкта N відрізняється від будь-якого способу вибору об'єкта M, то вибір «*N або M*» можна зробити *n+m* способами.

Правило добутку. Нехай необхідно виконати одну за одною *k* дій. Якщо першу дію можна виконати n_1 способами, другу - n_2 способами і т. д., *k*-ту дію - n_k способами, то *k* дій разом (одночасно і 1-шу дію, і 2-гу дію, …, і *k*-ту дію) можна виконати $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ способами.

Правило добутку можна сформулювати й інакше: якщо деякий об’єкт N можна вибрати n різними способами, а об’єкт M – m різними способами, причому будь-який спосіб вибору об’єкта N відрізняється від будь-якого способу вибору об’єкта M, то вибір пари об’єктів «*N і M*» можна зробити $n \cdot m$ способами.

Правило суми і правило добутку можна узагальнити для будь-якого числа *k* об’єктів.

Приклад 1.3.6. У квітковій вазі є 2 червоних троянди і три білих. Скількома способами можна відібрати: а) хоча б одну троянду, б) і білу, і червону троянди одночасно (тобто пару квіток (біла, червона))?

Розв’язання. Червону троянду можна вибрати двома різними способами, а білу – трьома. а) Хоча б одну квітку (білу або червону) можна, за правилом суми, вибрати 2+3=5 способами;
б) пару квіток (біла, червона), за правилом добутку, можна вибрати 2*3=6 способами, оскільки до кожного способу вибору червоної троянди можна «приєднати» кожен із способів вибору білої.

Правило-орієнтир для розв’язування задач на комбінаторику

Задачі на комбінаторику рекомендується розв’язувати за такою схемою:

1. Підрахувати число елементів *n* основної множини.
2. Підрахувати число елементів *m*, що входять у вибірку (підмножину).
3. Вияснити, чи впорядковані підмножини та чи повторюються в них елементи.
4. При підрахуванні числа сполук використати відповідні формули.

1.4. Біном Ньютона

Біномом зазвичай називають двочлен $a+b$, який є окремим випадком многочлена.

Зі шкільного курсу математики відомі такі рівності:

$$(a+b)^2 = a^2 + 2ab + b^2, \qquad (a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3,$$

які можна записати у вигляді:

$$(a+b)^2 = C_n^0 a^2 + C_n^1 a^1 b^1 + C_n^2 b^2,$$

$$(a+b)^3 = C_n^0 a^3 + C_n^1 a^2 b^1 + C_n^2 a^1 b^2 + C_n^3 b^3.$$

Біномом Ньютона називається формула

$$(a+b)^n = \boldsymbol{C}_n^0 a^n b^0 + C_n^1 a^{n-1} b^1 + C_n^2 a^{n-2} b^2 + \cdots + C_n^{n-1} a^1 b^{n-1} +$$

$$+C_n^n a^0 b^n \qquad (1.4.1)$$

Запишемо ліву частину бінома в вигляді добутку *n* однакових множників $(a+b)^n = (a+b)(a+b)\ldots(a+b)$. Після розкриття дужок (до зведення подібних доданків) отримаємо суму різних добутків, складених із *n* букв *a* або *b*, по одній із кожного множника. Число доданків, які містять, наприклад, рівно *k* разів букву *b*, і *(n-k)* разів букву *a*, рівне числу C_n^k. Отже, сума всіх членів, які містять букву *b* множником рівно *k* раз, дорівнює $C_n^k a^{n-k} b^k$. Оскільки *k* може приймати значення 0, 1, 2, ..., *n-1, n,* то, згрупувавши доданки з однаковими степенями *a* та *b,* і звівши подібні доданки, отримаємо праву частину розкладу бінома Ньютона.

Числа C_n^0, C_n^1, … , C_n^n називаються *біноміальними коефіцієнтами.*

На підставі основних властивостей сполучень можна відмітити деякі *властивості* розкладу бінома.

1. Кількість членів розкладу бінома на одиницю більша від показника степеня бінома.

2. Сума показників *a* та b кожного члена розкладу рівна показнику степеня бінома, причому показник першого члена (елемента) *a* спадає від *n* до 0, а показник другого елемента b зростає від 0 до *n.*

3. Біноміальні коефіцієнти, що рівновіддалені від кінців розкладу, рівні. Тому вони складаються до половини, а потім симетрично виставляються.

4. Сума всіх біноміальних коефіцієнтів дорівнює 2^n.

5.Сума біноміальних коефіцієнтів, які стоять на парних місцях, дорівнює сумі коефіцієнтів, які стоять на непарних місцях, і рівна 2^{n-1},тобто:

$$C1 + C_n^2 + C_n^4 + C_n^6 + \cdots = C_n^1 + C_n^3 + C_n^5 + \cdots = 2^{n-1}. \qquad (1.4.2)$$

*6. Д*ля будь-якого *k* (1 ≤ *k* ≤ *n*) має місце рівність

$$C_n^k = C_{n-1}^k + C_{n-1}^{k-1}, \qquad (1.4.3)$$

яку легко отримати з формули (1.3.3).

7. Рівність (1.4.3) дозволяє обчислювати значення C_n^k, якщо відомі C_{n-1}^k і C_{n-1}^{k-1}. Інакше, за допомогою цієї рівності можна послідовно обчислити C_n^k спочатку для *n=1*, потім для *n=2*, *n=3* і т.д. Отримуємо так званий *трикутник Паскаля*, який зручно записувати у вигляді трикутної таблиці, в (*n+1*) рядку якої по порядку стоять числа C_n^0; C_n^1; C_n^2; …; C_n^n:

$$\begin{array}{c} 1 \\ 1 \quad 1 \\ 1 \quad 2 \quad 1 \\ 1 \quad 3 \quad 3 \quad 1 \\ 1 \quad 4 \quad 6 \quad 4 \quad 1 \\ 1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1 \\ 1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1 \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \end{array} \qquad (1.4.4)$$

Тут крайні числа рядка, тобто C_n^0 і C_n^n, дорівнюють 1, а решта чисел знаходяться за формулою (1.4.3), тобто треба додати числа, що розташовані зліва і справа від шуканого числа, з попереднього рядка. Наприклад, число 20 у сьомому рядку ми отримуємо, додавши 10 і 10 із шостого рядка.

8. Щоб записати у загальному вигляді доданки у формулі (1.4.1), доцільно (*k*+*1*)-й доданок вважати *k*-тим членом і позначити його через T_k. Тоді

$$T_k = C_n^k a^{n-k} b^k. \qquad (1.4.5)$$

Формула (1.4.5) називається *формулою загального члена* розкладу бінома. Так, $T_0 = C_n^0 a^n b^0 = C_n^0 a^n = a^n$ - перший доданок; $T_1 = C_n^1 a^{n-1} b^1 = n a^{n-1} b$ - другий доданок;
$T_2 = C_n^2 a^{n-2} b^2 = \frac{n(n-1)}{1 \cdot 2} a^{n-2} b^2$ – третій доданок і т. д.

Примітка. Не слід забувати про різницю між коефіцієнтом доданка розкладу бінома і біноміальним коефіцієнтом того ж доданка. Наприклад, у розкладі

$$(2a - 3b)^4 = 16a^4 - 96a^3 b + 216a^2 b^2 - 216ab^3 + 81b^4$$

коефіцієнт третього члена розкладу дорівнює 216, а його біноміальний коефіцієнт дорівнює $C_4^2 = 6$, для четвертого – відповідно 216 і $C_4^3 = 4$.

Узагальненням формули бінома Ньютона є формула

$$(a_1 + a_2 + \ldots + a_m)^n = \sum \frac{n!}{k_1!\, k_2! \ldots k_m!} a_1^{k_1} a_2^{k_2} \ldots a_m^{k_m}, \qquad (1.4.6)$$

яка називається *поліноміальною* (у правій частині стоїть поліном (многочлен) від $(a_1, a_2, \ldots, a_m)$, де підсумовування відбувається за різними наборами цілих невід'ємних чисел $k_1, k_2, \ldots, k_m$, таких що $k_1 + k_2 + \ldots + k_m = n$). Коефіцієнти $\frac{n!}{k_1!\, k_2! \ldots k_m!}$, з якими члени $a_1^{k_1} a_2^{k_2} \ldots a_m^{k_m}$ входять у розклад $(a_1 + a_2 + \ldots + a_m)^n$ називають *поліноміальними коефіцієнтами*.

Частинним випадком формули полінома є формула

$$(a + b)^n = \sum \frac{n!}{k_1!\, k_2!} a^{k_1} b^{k_2} \text{ (де } k_1 + k_2 = n), \qquad (1.4.7)$$

яка лише формою запису відрізняється від формули бінома Ньютона.

Приклад 1.4.1. Піднести до степеня $(a + b + c)^3$.

Розв'язання. Користуючись формулою розкладу полінома, отримаємо $(a + b + c)^3 = \sum \frac{3!}{k_1! k_2! k_3!} a^{k_1} b^{k_2} c^{k_3}$, де підсумовування проводиться за всіма наборами невід'ємних цілих чисел k_1, k_2, k_3, таких що $k_1 + k_2 + k_3 = 3$. Число 3 у сумі дають такі числа: (0; 1; 2), (0; 2; 1), (0; 3; 0), (0; 0; 3), (1; 0; 2), (1; 2; 0), (1; 1;1), (2; 0; 1), (2; 1; 0), (3; 0; 0). У відповідності до цього маємо суму десяти доданків:

$$(a + b + c)^3 = \sum \frac{3!}{0!\,1!\,2!} a^0 b^1 c^3 + \sum \frac{3!}{0!\,2!\,1!} a^0 b^2 c^1 + \sum \frac{3!}{0!\,3!\,0!} a^0 b^3 c^0 +$$

$$+ \sum \frac{3!}{0!\,0!\,3!} a^0 b^0 c^3 + \sum \frac{3!}{1!\,0!\,2!} a^1 b^0 c^2 + \sum \frac{3!}{1!\,2!\,0!} a^1 b^2 c^0 +$$

$$+ \sum \frac{3!}{1!1!1!} a^1 b^1 c^1 + \sum \frac{3!}{2!0!1!} a^2 b^0 c^1 + \sum \frac{3!}{2!1!0!} a^2 b^1 c^0 + \sum \frac{3!}{3!0!0!} a^3 b^0 c^0 =$$

$$= a^3 + b^3 + \mathrm{c}^3 + 3a^2 b + 3ab^2 + 3a^2 \mathrm{c} + 3b^2 \mathrm{c} + 3\mathrm{c}^2 a + 3c^2 b + 6abc.$$

Приклад 1.4.2. Знайти номер члена розкладу $(a + a^{-2})^{12}$, що не містить *a*.

Розв'язання. Якщо доданок не містить *a*, це означає що він містить a^0. За формулою (1.4.5) маємо

$$T_k = C_{12}^k a^{12-k} (a^{-2})^k = C_{12}^k a^{12-3k}.$$

За умовою, $12 - 3k = 0$, звідки $k = 4$. Отже, шуканим є четвертий член розкладу.

1.5. Випадкові події та операції над ними

Подією називається будь-яке явище природи або суспільства, яке при заданому комплексі умов може настати або не настати. Подія настає в результаті реалізації комплексу умов, тобто в результаті випробування (досліду, експерименту, спостереження).

Наприклад, при стрільбі по мішені в комплекс умов входить наявність зброї та патронів до неї, наявність стрільця, мішені, відстані до мішені повинні бути зазначені, спосіб стрільби (лежачи, з руки). Реалізацією комплексу цих умов є в даному випадку постріл. Влучення в мішень (або промах) - подія.

Розрізняють *елементарні події (наслідки)*, які не можна розкласти на більш прості, та *складні події*, які можна розкласти на елементарні. Множину усіх можливих елементарних подій називають *простором елементарних подій,* а самі елементарні події – точками цього простору. Простір елементарних подій може містити скінчену, злічену або незлічену множину елементів.

У ролі елементарних подій можна розглядати точку *n*-вимірного простору, відрізок деякої лінії, точки поверхні S або об'єму V тривимірного простору, функцію однієї або багатьох змінних.

Розрізняють вірогідні (достовірні), неможливі і випадкові події.

Подію називають *вірогідною,* якщо при заданому комплексі умов вона обов'язково настане. Прикладом вірогідної події є дістання білої кулі з урни, в якій містяться тільки білі кулі.

Подія називається *неможливою*, якщо при заданому комплексі умов вона відбутись не може. Дістання чорної кулі з урни, у якій знаходяться тільки білі кулі, є неможливою подією.

Подія називається *випадковою*, якщо при заданому комплексі умов вона може настати і може не настати. Дістання білої кулі з урни, в якій знаходяться білі і чорні кулі, є випадковою подією.

Події позначаються $A, B, C, \ldots$ (букви латинського алфавіту) або $A_1, A_2, A_3, \ldots, A_n$.

Наприклад, нехай в урні є білі, червоні і зелені кулі. Навмання дістається одна куля. Через A позначимо подію, яка полягає в тому, що витягнута куля виявилася білою, B - червоною і C - зеленою.

У взаємозв'язку між собою події поділяються на несумісні (неспільні) або сумісні, єдиноможливі, рівноможливі, протилежні.

Дві або кілька подій називаються *несумісними*, якщо настання однієї з них виключає можливість настання кожної з інших у даному випробуванні. У противному випадку події називаються *сумісними*.

Дві або кілька подій називаються *єдиноможливими*, якщо яка-небудь із них обов'язково настане в даному випробуванні. Події A, B, C з попереднього прикладу є несумісними і єдиноможливими.

Дві або кілька подій називаються *рівноможливими*, якщо немає підстав вважати, що яка-небудь із них настане частіше, ніж будь-яка інша.

При підкиданні монети поява герба або цифри є рівноможливими подіями.

Дві події називаються *протилежними*, якщо вони єдиноможливі та несумісні. При одному пострілі по мішені влучення в мішень або промах - протилежні події, тому що вони єдиноможливі і несумісні. Якщо подія позначена буквою A, то протилежна подія позначається $\overline{A}$ (не *A*). Прикладом протилежних подій є народження дівчинки або хлопчика, зійшло або не зійшло посаджене зерно.

Сукупність подій $A_1, A_2, A_3, \ldots, A_n$ складає *повну групу подій*, якщо ці події єдиноможливі та несумісні. Так, у попередньому прикладі події *A, B*, *C* складають повну групу подій при одному випробуванні.

Оскільки події – це деякі підмножини множини всіх елементарних подій, то над ними можна виконувати операції, як і над множинами.

Об'єднанням (сумою) випадкових подій $A_1, A_2, A_3, \ldots, A_n$ називають таку випадкову подію, яка полягає у появі хоча б однієї з цих подій, і позначають $A_1 \cup A_2 \cup \ldots \cup A_n = \bigcup_{i=1}^{n} A_i$, зокрема, якщо події несумісні, то їх сума полягає в тому, що має з'явитись подія A_1, або A_2, або A_3, ..., A_n, тобто $A_1 + A_2 + \ldots + A_n = \sum_{i=1}^{n} A_i$.

Добутком (суміщенням) $A \cap B$ (або $A \cdot B$) випадкових подій *A*, *B* називають таку випадкову подію, яка полягає у появі подій *A, B* одночасно.

Якщо *A, B* несумісні, то їх добутком є порожня (пуста) множина $\emptyset$, яка не містить жодного елемента, тобто $A \cdot B = \emptyset$.

Різницею $B - A$ (або $B \setminus A$) двох випадкових подій *B*, *A* називають усі наслідки, які полягають у тому, що подія А не з'являється при появі події *B*.

Операції додавання, множення, заперечення (протилежної події) ілюструють за допомогою діаграм Ейлера-Венна (рис.1.5.1).

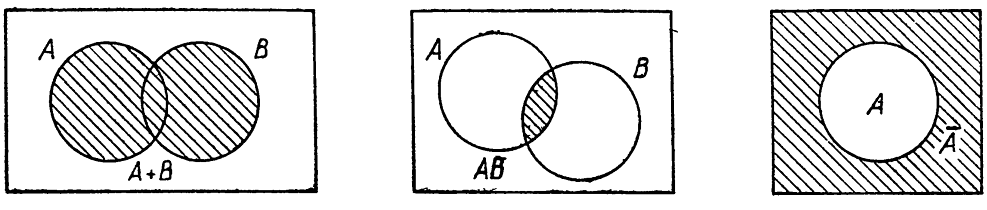
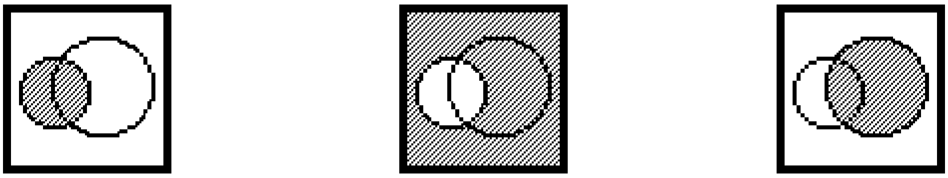
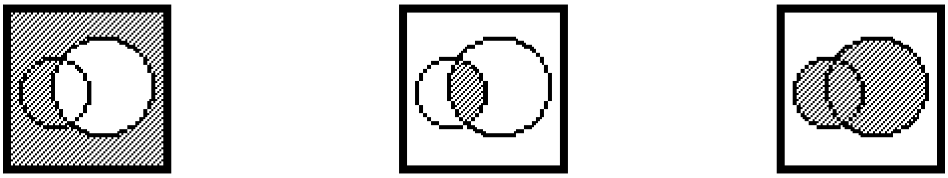


Рис.1.5.1

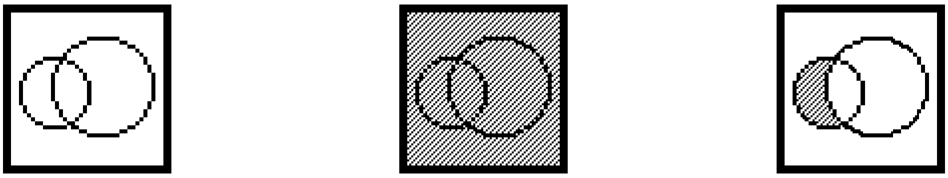
Поняття протилежної події, добутку, суми, різниці подій проілюстровані також на рис. 1.5.2. На квадрат навмання кидають точку. Попадання точки у заштриховану область означає появу відповідної події.



Подія A Подія $\overline{A}$ Подія B



Подія $\overline{B}$ Подія AB Подія $A+B$



Подія $A\overline{A}=B\overline{B}=\emptyset$ Подія $A+\overline{A}=B+\overline{B}=U$ Подія $A-B=A\overline{B}$

Рис. 1.5.2

1.6. Математичне, статистичне і аксіоматичне означення ймовірності події. Геометрична ймовірність

Перейдемо до основного поняття теорії ймовірностей – поняття ймовірності події. У методологічних термінах говорять, що ймовірність події є *мірою можливості* її *появи*. Але це означення не дає числа, з яким можна було б порівняти випадкові події за мірою можливості їх появи, яке повинно бути тим більше, чим більш можлива подія. Таким числом є p – ймовірність події. Існує декілька означень імовірності.

Розглянемо приклад. Нехай в урні знаходиться 12 білих і 5 чорних куль (усього 17). Навмання дістається одна куля. При цьому є об'єктивна можливість витягнення як білої, так і чорної кулі. Однак, можливість витягнення білої кулі більша, ніж чорної. Тому виникає необхідність оцінити об'єктивну можливість настання випадкової події числом. Зручніше за все об'єктивну можливість витягнення білої кулі (подія *A*) оцінити числом $\frac{12}{17}$, тому що з 17 результатів випробування події *A* сприяє 12. Аналогічно, об'єктивна можливість витягнення чорної кулі (подія *B)* оцінюється числом $\frac{5}{17}$. Отже, $\mathrm{P(A)} = \frac{12}{17}$; $\mathrm{P(B)} = \frac{5}{17}$.

Математичною ймовірністю випадкової події *A* називається відношення *m* - числа результатів випробування, сприятливих цій події *A*, до *n* - числа всіх єдиноможливих, рівноможливих і попарно несумісних результатів випробування. Позначається дана ймовірність *P(A)* та згідно означення

$$P(A) = \frac{m}{n}. \qquad (1.6.1)$$

Приклад 1.6.1. У коробці є 3 червоних, 4 зелених і 8 синіх кульок. Навмання послідовно дістають три кульки. Яка ймовірність того, третя витягнута кулька буде зеленою, якщо перші дві червоні?

Розв'язання*.* Після того, як із коробки дістають 2 червоні кульки, в ній залишається 1 червона кулька, 4 зелених і 8 синіх кульок, всього 13 кульок. Отже: *m* = 4; *n* = 13 і $\mathrm{P(A)} = \frac{4}{13}$**.**

Приклад 1.6.2. У партії із 50 бетонних плит є 5 нестандартних. На машину вантажать навмання 4 плити. Яка ймовірність того, що серед них 3 стандартні?

Розв'язання. Так як кожний набір із 4 плит відрізняється хоча б однією плитою, то число можливих наборів дорівнює $n = C_{50}^{4}$. Кількість наборів із трьох стандартних і одної нестандартної плити дорівнює $m = C_{45}^{3} \cdot C_{5}^{1}$.

Нехай *A* – шукана подія (на машину вантажать три стандартні та одну нестандартну плиту), то

$$P(A) = \frac{m}{n} = \frac{C_{45}^{3} \cdot C_{5}^{1}}{C_{50}^{4}} = \frac{1419}{4606} = 0{,}31.$$

При розв'язуванні багатьох задач знаходження чисел m та n може мати певні трудності. Тому при їх обчисленні часто використовують формули та правила комбінаторики.

Формулу (1.6.1) використовують для безпосереднього обчислення ймовірностей подій лише тоді, коли дослід зводиться до схеми випадків, коли m та n скінчені, усі наслідки рівноможливі. У противному випадку цією формулою користуватись не можна, і доцільно використовувати статистичне або геометричне означення ймовірності.

Статистична ймовірність. Відносною частотою події *A* (частковою частотою, частістю) називають відношення числа випробувань *m*, в яких подія *A* з'являється, до загального числа *n* проведених випробувань. Відносна частота події обчислюється за формулою

$$W(A) = \frac{m}{n}. \qquad (1.6.2)$$

Я. Бернуллі довів, що якщо зробити кілька серій по *n* випробувань у кожній, то частота у кожній серії буде коливатися навколо деякого числа, що є ймовірністю випадкової події. Тому *W(A)* при досить великій кількості випробувань називають *статистичною ймовірністю*.

Отже, *статистичною ймовірністю* випадкової події називається відношення числа випробувань, у яких подія настала, до числа всіх випробувань, якщо їх кількість достатньо велика.

Приклад 1.6.3. Для визначення схожості насіння пшениці висівається 500 насінин у лабораторних умовах. Отримано 492 результати. Знайти ймовірність того, що проросте одне посаджене зерно.

Розв'язання. Випробування – висівання зерна. Подія *A* - посаджене зерно проросте. Так як *n* = 500, *m*= 492, то шукана ймовірність дорівнює $P(A) = \frac{492}{500} = 0{,}984$. Сходи насіння складають 98,4% (з кожних 1000 насінин проросте 984).

Відзначимо, що ймовірність $\mathrm{P(A)} = \frac{\mathrm{m}}{\mathrm{n}}$ появи події A обчислюють до випробувань, а відносну частоту $W(A) = \frac{m}{n}$ - після випробування.

Геометрична ймовірність. Нехай множина усіх елементарних наслідків нескінчена, і, як наслідок, займає деяку область *G*. Події *A* сприяє лише частина $g \in G$. Тоді ймовірність випадкової події *A* дорівнює відношенню міри g до міри *G*:

$$P(A) = \frac{m(g)}{m(G)}. \tag{1.6.3}$$

Зокрема, ймовірність попадання точки на відрізок *l*, що є частиною проміжку *L,* і не залежить від його розміщення відносно *L,* обчислюється за формулою

P = Довжина l / Довжина L.

Аналогічно означається ймовірність попадання точки на фігуру *g,* яка складає частину фігури *G* на площині*;* в просторову фігуру *v*, яка складає частину просторової фігури *V*:

P = Площа g/ Площа G.

P = Об'єм v / Об'єм V.

Приклад 1.6.4. Яка ймовірність того, що навмання кинута в даний круг точка опиниться в площині вписаного у нього квадрата?

Розв'язання. Шукана ймовірність дорівнює відношенню площі квадрата до площі круга, в який вписано квадрат. Площа круга дорівнює $S\text{кр.} = \pi r^2$, а площа квадрата $S_{\text{кв.}} = 2r^2$. Отже,

$$\text{P} = \frac{S\text{кв.}}{S\text{кр.}} = \frac{2r^2}{\pi r^2} = \frac{2}{\pi} = 0{,}63.$$

Аксіометричне означення ймовірності. Для кількісного описання ступеня об'єктивної можливості появм тієї чи іншої спостережуваної в експерименті події вводяться спеціальні числові функції *P(A)*, що називаються ймовірністю події А.

Нехай $\mathbb{F}$ – поле подій для деякого експеримента. *Ймовірністю P(A)* називається числова функція, що визначена для всіх $A \in \mathbb{F}$ і задовольняє наступним трьом умовам (аксіомам ймовірностей):

1. $P(A) \geq 0.$
2. $P(\Omega) = 1.$

3. Для будь-якої скінченної або нескінченної послідовності спостережуваних подій $A_1, A_2, \dots, A_n, \dots$ таких, що $A_i\, A_j = \emptyset$ при $i \neq j$,

$$P(\textstyle\sum_k A_k) = \sum_k P(A_k).$$

Оскільки подія є множина, то ймовірність є також функцією множини. Наведені аксіоми виділяють спеціальний клас функцій, що є ймовірнісними. У відповідності до змісту цих аксіом ймовірність є невід'ємна, нормована і адитивна функція множини, що належить алгебрі $\mathbb{F}$.

Трійку $\{\Omega,\ \mathrm{F}, P\}$, де $\mathbb{F}$ – алгебра підмножин множини Ω, P – числова функція, що задовольняє аксіомам 1-3, називають ймовірнісним простором випадкового експеримента.

Аксіоматична теорія ймовірностей у її сучасному вигляді була створена радянським ученим А.М.Колмогоровим у 1933 р.

Приклад 1.6.5. Довести, що якщо для деякого експерименту $A \subset B$, то $P(A) \le P(B)$.

Розв'язання. Оскільки $A \subset B$, то $AB = A$ (рис. 1.6.1), тому $B = B\left(A + \overline{A}\right) = BA + B\overline{A} = A + B\overline{A}$, причому обидва доданки несумісні. В силу властивості адитивності $P(B) = P(A) + P(B\overline{A})$, а в силу аксіоми 1 $P(B\overline{A}) \ge 0$. Отже, $P(B) \ge P(A)$.

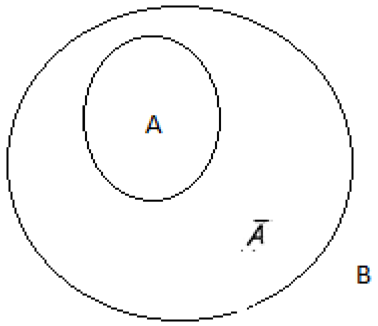


Рис. 1.6.1

Використовують також інші практичні прийоми для визначення ймовірностей подій:

- обчислення ймовірностей подій за допомогою дерев (графів) можливих результатів. Тут випробування може бути представлене як багатокроковий процес, у якому кожен попередній результат має декілька наступних результатів. У загальному випадку результати кожного кроку випробування різноймовірні. Подія, яка нас цікавить, може відбутися після будь-якого одного або декількох кроків (рис. 1.6.2).
- обчислення ймовірностей складних подій, що представлені у вигляді комбінацій елементарних подій. У цьому випадку використовуються формули ймовірностей суми і добутку подій (Розділ 2 даного підручника). Тоді алгоритм обчислення ймовірностей буде наступним:

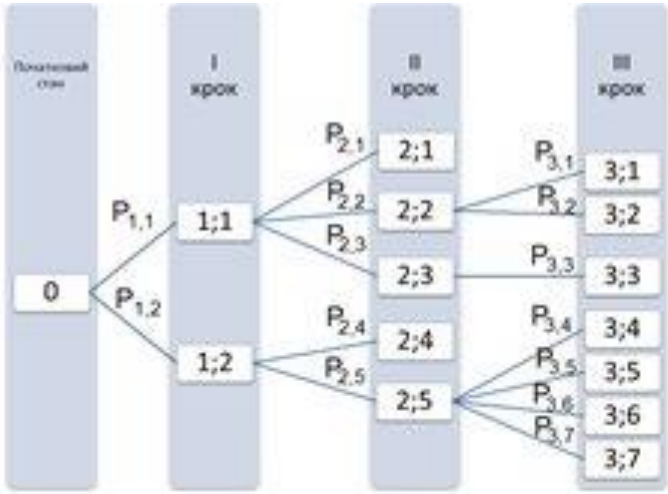


Рис. 1.6.2

- ✓ записати досліджувану подію у вигляді комбінації елементарних подій, що пов'язані знаками суми або добутку;
- ✓ застосувати формули ймовірностей суми або добутку подій;
- ✓ обчислити ймовірності елементарних подій, якщо вони неввідомі;
- ✓ обчислити шукану ймовірність.

Основні властивості ймовірностей

1. Ймовірність достовірної події дорівнює 1.

Якщо подія *U* - достовірна подія, то в n випробуваннях вона настане n раз і тому $P(U)=\frac{n}{n}=1$.

2. Ймовірність неможливої події дорівнює 0.

Якщо *V* - неможлива подія, то в n випробуваннях вона настане 0 раз і тому $P(V)=\frac{0}{n}=0.$

3. Ймовірність випадкової події є число, що знаходиться між 0 і 1.

Дійсно, якщо *A* - випадкова подія, то в n випробуваннях вона настане m раз, причому 0< m< n . Розділивши останню нерівність на n, одержимо $0<\frac{m}{n}<1$. Отже 0< *P (A)*< *1.*

Отже, ймовірність будь-якої випадкової події - це невід'ємне число, яке не перевищує 1.

Правило-орієнтир для обчислення ймовірностей

1. Уяснити суть випробування, що розглядається в задачі.
2. Встановити події та позначити їх літерами.
3. За допомогою введених позначень виразити подію, ймовірність появи якої потрібно знайти.
4. Якщо потрібно знайти ймовірність суми подій, треба вияснити, сумісні чи несумісні ці події. Якщо ж потрібно знайти ймовірність добутку подій, то необхідно встановити, залежні чи незалежні ці події.
5. Вибрати відповідну до умови задачі формулу і виконати необхідні обчислення.

Питання для самоперевірки

1. Що є предметом комбінаторики?
2. Які сполуки називають перестановками, розміщеннями, комбінаціями (сполученнями)? Як позначають та обчислюють кількість цих сполук?
3. Сформулюйте основні властивості числа сполучень.
4. Сформулюйте основні властивості розкладу бінома.
5. Що означає в теорії ймовірностей слово "подія"?
6. Яка подія називається випадковою? Наведіть приклади.
7. Яка подія називається неможливою, достовірною ? Наведіть приклади.
8. Які події називаються несумісними, рівноможливими, єдиноможливими? Наведіть приклади.
9. Які події називаються протилежними? Наведіть приклади.

10. Що називається повною групою подій? Наведіть приклади.

11. Сформулюйте математичне, статистичне та геометричне означення ймовірності. Наведіть приклади їх використання.

12. Сформулюйте основні властивості ймовірностей.

Задачі та вправи

1.1. Керуючий транспортною фірмою повинен скласти маршрути для своїх водіїв. Є шість покупців, куди повинен бути доставлений вантаж. Скільки можна скласти маршрутів почергового відвідування покупців?

1.2. Аудиторська фірма має 4 різних вакантних посади у своїх філіях. Є 10 претендентів на ці посади. Скільки існує способів заповнювання цих посад ?

1.3. Скількома способами можна 15 шахістів поділити на три команди по 5 чоловік?

1.4. Скільки різних бригад для обслуговування гідротехнічного спорудження в складі 3-х чоловіків і 2-х жінок можна скласти із 18 робітників, серед яких 6 жінок?

1.5. Скільки тризначних чисел можна скласти із цифр 0, 1, 2,3, 4, 5, 6 якщо: а) цифри не повторюються; б) цифри можуть повторюватися?

1.6. При маркетинговому дослідженні просять зробити дегустацію 5 різних сортів чаю та назвати три кращі в порядку спадання. Скільки існує способів вибору трьох кращих сортів?

1.7. Абонент забув дві останні цифри номера телефону, пам'ятаючи лише, що вони різні. Скільки максимально прийдеться абоненту набирати номер телефону, щоб додзвонитися?

1.8. У хірургічному відділенні працює 40 лікарів. Скількома способами можна організувати бригаду в складі хірурга і чотирьох його асистентів?

1.9. Група із 18 випускників проводить ювілейну зустріч. Скільки було рукостискань, якщо всі обміняються рукостисканнями по одному разу?

1.10. Піднести до степеня $(3x + 2)^6$.

1.11. Вісім студентів подорожували в двох човнах, у меншому з яких могло вміститись не більше 4 осіб, а в більшому – не більше 6 осіб. Скількома способами вони можуть розміститись у цих човнах?

1.12. Кожного із семи студентів можна направити для проходження практики на одне з двох підприємств. Скількома різними способами це можна зробити ?

1.13. Скількома різними способами можна розподілити 6 різних предметів між трьома особами так, щоб кожна з них отримала по два предмети ?

1.14. Скільки треба мати словників, щоб можна було безпосередньо робити переклади з будь-якої з п'яти мов: української, російської, французької, німецької, англійської на будь-яку іншу мову?

1.15**.** На збірку поступило 3000 деталей з першого верстата і 2000 з другого. Перший дає 0,2% браку, а другий – 0,3%. Знайти ймовірність того, що взята навмання деталь із нерозібраної продукції верстатів буде бракованою**.**

1.16. Відносна частота бракованої продукції складає 0,05%. Знайти число бракованих деталей в партії із 140 штук.

1.17. Цифри 1, 2, 3, 4, 5 написані на 5 картках. Навмання послідовно беруть три картки і розмішують зліва на право. Чому дорівнює ймовірність того, що отримане таким чином тризначне число буде парним?

1.18.Транспортна фірма повинна доставити вантаж 4 покупцям. Фірма має 11 автомобілів, з яких 5 – нові. Знайти ймовірність того, що весь вантаж буде перевезено новими автомобілями.

1.19. У партії із 20 деталей є 8 нестандартних. Для перевірки на якість беруть навмання 5 деталей. Яка ймовірність того, що в цій вибірці 2 нестандартні деталі?

1.20. Серед 25 студентів, з яких 15 дівчат, розігруються 4 білети, причому кожна особа може виграти тільки один білет. Яка ймовірність того, що серед чотирьох володарів білетів виявляться: а) чотири дівчини; б) чотири юнаки; в) три юнаки та одна дівчина?

1.21. У конкурсі газети *N* бере участь 12 чоловіків та 8 жінок. Є два призових місця. Визначити ймовірність подій: а) *A* (обидва призових місця займуть жінки); б) *B* (місця розділять між собою чоловік і жінка); в) *C* (обидва місця займуть чоловіки).

1.22. Вісім різних книжок розставляють на одній полиці. Знайти ймовірність того, що дві визначені книги будуть поставлені поруч на полиці.

1.23. Кинуто гральний кубик. Знайти ймовірність того, що випаде парна кількість очок.

1.24. Серед 25 фірм, з яких 10 українських, а інші польські, розігрується 5 урядових контрактів. Вважається, що кожна фірма має рівні шанси на отримання контракту. Знайти ймовірність того, що контракт виграють дві українські фірми.

1.25. У групі із 12 економістів тільки 8 мають досвід роботи у запропонованій новій галузі. Для проекту треба відібрати 4 особи. В припущені, що відбір претендентів ведеться навмання, визначити ймовірність того, що в команду з чотирьох осіб потраплять ті, що: а) мають досвід роботи; б) не мають досвіду роботи; в) дві особи мають досвід роботи,а дві - не мають досвіду роботи.

1.26. У слюсаря 10 деталей, які мало відрізняються одна від одної. Із них 4 першого, по дві другого, третього та четвертого видів. Яка ймовірність того, що серед взятих одночасно шести деталей три деталі будуть першого виду, дві – другого виду і одна третього виду?

1.27.З урни, в якій лежать 6 білих і 4 чорних кулі, навмання взяли 3 кулі. Яка ймовірність того, що серед взятих куль будуть дві білі?

1.28. Є 25 екзаменаційних білетів, кожний з них має по 2 питання. Студент знає відповіді тільки на 30 питань з 50.Знайти ймовірність того, що: а) студент здасть випробування, якщо для цього достатньо відповісти на 2 питання з білета, або на одне питання з білета і одне додаткове питання; б) йому попав білет з питаннями, які він знав.

1.29. З двадцяти студентів, що прийшли на випробування, 8 підготувались відмінно, 6 - добре, 4 - посередньо і 2 - погано. В екзаменаційних білетах є 40 питань. Студент, що підготувався відмінно, знає відповіді на всі питання; що підготувався добре - на 35 питань, посередньо - на 15 питань і погано - на 10 питань. Студент, якого викликали, відповів на три питання. Знайти ймовірність того, що цей студент: а) підготувався добре; б) погано.

1.30. Серед 50 лотерейних білетів є 4 виграшних. Яка ймовірність того, що серед взятих навмання будь-яких двох білетів хоча б один виявиться виграшним?

1.31. Контрольне завдання містить п'ять питань, на кожне з яких є чотири варіанти відповідей, причому, одна відповідь правильна, а три - неправильні. Знайти ймовірність того, що студент дасть три правильних відповіді.

1.32. На електропоїзд, що має шість вагонів, чекають дванадцять пасажирів, причому вибір вагону кожним пасажиром рівноможливий. Знайти ймовірність того, що: а) у кожний вагон ввійшли по два пасажири; б) в один вагон ніхто не ввійшов, у другий ввійшов один пасажир, у два вагони - по два пасажири і в останні два вагони ввійшли відповідно три і чотири пасажири.

1.33. Знайти ймовірність того, що номер першого зустрічного автомобіля не має: а) цифри 5; б) двох і більше цифр 5 (номер має чотири цифри).

1.34. У круг радіуса R вписаний рівносторонній трикутник. Яка ймовірність того, що навмання кинута в даний круг точка, опиниться всередині трикутника?

1.35. У круг радіуса R кидають точку. Знайти ймовірність того, що відстань від цієї точки до центра круга не перевищує r.

1.36. У групі з 15 студентів, серед яких шестеро дівчат; по списку навмання відібрано вісім осіб. Знайти ймовірність того, що серед відібраних студентів три дівчини.

1.37. На відрізку довжиною *l* навмання вибирають дві точки. Знайти ймовірність того, що відстань між цими точками буде не менша від 3.

1.38. У секретному замку чотири диски, кожний з яких установлюється в одному із 10 положень, що відповідають цифрам від 0 до 9. Замок відкривається лише в тому випадку, коли диски встановлені так, що цифри на них складатимуть певне чотиризначне число. Знайти ймовірність того, що при довільній установці дисків замок буде відкритий.

1.39. На площині зображено два концентричних кола, радіуси яких 5 і 10 см відповідно. Знайти ймовірність того, що точка кинута навмання у великий круг, попаде також і в кільце, яке утворене побудованими колами.

1.40. На площині проведено дві паралельні прямі, відстань між якими 8 см. Визначити ймовірність того, що навмання кинутий на цю площину круг радіуса 3 см, не буде перетинати жодну пряму.

1.41. Знайти х, якщо п'ятий член розкладу $(\sqrt[4]{x}+x^{-1})^{6}$ дорівнює $\frac{5}{9}$.

1.42. Третій член розкладу $(2x+\frac{1}{x^{2}})^{m}$ не містить *x.* При яких дійсних значеннях *x* цей член дорівнює другому члену розкладу $(1+x^{3})^{30}$?

РОЗДІЛ 2
ОСНОВНІ ТЕОРЕМИ ТЕОРІЇ ЙМОВІРНОСТЕЙ

2.1.Теореми додавання ймовірностей

Сумою двох випадкових подій *A* та *B* (позначається *A+B*) називається така подія, яка відбувається, коли відбувається принаймні одна з цих подій: *A* або *B*. Це співвідношення між подіями є звичайним співвідношенням між множинами, яке можна представити графічно за допомогою кругів Ейлера-Венна (рис.2.1.1):

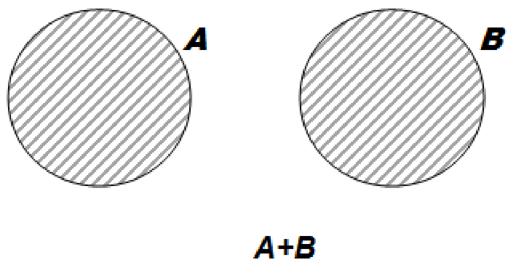


Рис. 2.1.1

Теорема 2.1.1. Ймовірність появи однієї із двох несумісних подій, дорівнює сумі ймовірностей цих подій

$$P(A + B) = P(A) + P(B). \qquad (2.1.1)$$

Доведення. Нехай повна сукупність подій містить n рівноможливих подій, причому k з них сприятливі появі події *A,* а m – події *B*. Так як *A* і *B* несумісні, то немає подій сприятливих і події *A* і події *B*. Тому з означення суми подій випливає, що для події *A+B* сприятливими є $k + m$ подій. Отже, маємо

$$P(A + B) = \frac{k+m}{n} = \frac{k}{n} + \frac{m}{n} = P(A) + (B).$$

Очевидно, що аналогічна формула має місце і для будь-якої кількості попарно несумісних подій $A_1, A_2, \ldots, A_m$:

$$P(A_1 + A_2 + \ldots + A_m) = P(A_1) + P(A_2) + \ldots + P(A_m). \qquad (2.1.2)$$

Теорема 2.1.2. Ймовірність появи хоча б однієї із двох сумісних подій дорівнює сумі ймовірностей цих подій без ймовірності їх сумісної появи:

$$P(A+B) = P(A) + P(B) - P(A\,B). \qquad (2.1.3)$$

Доведення. Подамо події *A, B* та *A+B* у вигляді суми відповідних двох несумісних подій $A = \overline{B}A + AB$; $B = \overline{A} + AB$; $A + B = A + \overline{A}B$ (рис. 2.1.2). Тоді, використавши три рази формулу суми ймовірностей двох подій послідовно, отримаємо

$$P(A)= P(\overline{B}A) + P(AB); \;\; P(B) = P(\overline{A}B)+ P(AB);$$

$$P(A + B) = P(A) + P(\overline{A}B) = P(A) + P(B) - P(A\,B).$$

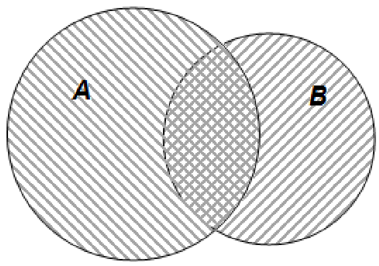


Рис. 2.1.2

Ймовірність появи хоча б однієї із трьох сумісних подій дорівнює

$$P(A+B+C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC).$$

Приклад 2.1.1. Стрілок стріляє по мішені, яка поділена на дві області. Ймовірність попадання в першу область дорівнює 0,45; у другу - 0,35. Знайти ймовірність того, що стрілок при одному пострілі влучить в одну з областей.

Розв'язання. Подія *A* – стрілок влучив у першу область і подія *B* – стрілок влучив у другу область є несумісними (при одному пострілі попадання в одну область виключає попадання в іншу). За умовою задачі: *P(A)* = 0,45; *P(B)* = 0,35. Для знаходження ймовірності використаємо теорему додавання для несумісних подій

$$P(A + B) = P(A) + P(B) = 0{,}45 + 0{,}35 = 0{,}80.$$

Декілька подій в умовах даного випробування утворюють *повну групу подій*, якщо в результаті даного випробування обов'язково відбудеться хоча б одна з них. Наприклад, в урні знаходяться сині, білі та червоні кульки. В результаті випробування (дістати з урни кульку) можливі такі події: *A* – дістали синю кульку, *B* - дістали білу кульку, *C* - дістали червону кульку. Події *A, B, C* утворюють повну групу подій.

Теорема 2.1.3. Сума ймовірностей повної групи випадкових подій дорівнює одиниці:

$$P(A_1) + P(A_2) + \cdots + P(A_n) = 1. \qquad (2.1.4)$$

Доведення. Якщо випадкові події $A_1, A_2, \dots, A_n$ утворюють повну групу, то вони попарно несумісні, а їх об'єднання буде достовірною подією. Маємо

$$P\left(\sum_{i=1}^{n} A_i\right) = P(A_1 + A_2 + \cdots + A_n) = 1 \qquad (2.1.5)$$

Ймовірність достовірної події також дорівнює одиниці, тому

$$P(A_1) + P(A_2) + \cdots + P(A_n) = 1. \qquad (2.1.6)$$

Праві частини рівностей (2.1.5) та (2.1.6) однакові, тому ліві частини будуть рівними, тобто має місце рівність (2.1.4). Теорема доведена.

Частинним випадком подій, які утворюють повну групу, є протилежні події.

Подією, що *протилежна* до *A,* називається подія $\overline{A}$ (не *A*), яка настає тоді і тільки тоді, коли не настає подія *A* (рис. 2.1.3).

$$P(\overline{A}) = 1 - P(A). \qquad (2.1.7)$$

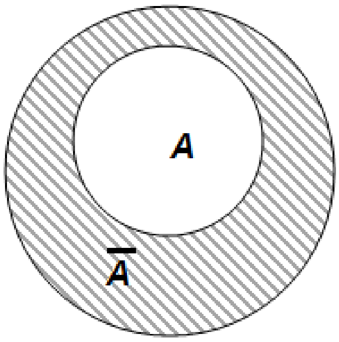


Рис. 2.1.3

Наприклад, виграш і програш, влучення в мішень та промах є протилежними подіями.

Дві протилежні події *A* та $\overline{A}$ утворюють повну групу, тому має місце рівність $P(A)+P(\overline{A}) = 1$, з якої одержуємо формулу знаходження ймовірності протилежної події:

Приклад 2.1.2. Ймовірність отримати повідомлення від певної особи на протязі доби дорівнює 0,25. Знайти ймовірність того, що повідомлення на протязі доби від цієї особи не буде отримане.

Розв'язання. Позначимо як подію *A* - повідомлення від особи на протязі доби поступить. За умовою задачі має місце співвідношення $P(A) = 0{,}25$. Протилежна подія $\overline{A}$означає, що на протязі доби від цієї особи повідомлення не поступить. Імовірність цієї події обчислимо за формулою

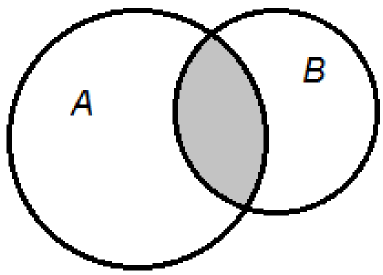
$$P(\overline{A})=1-P(A) =1 – 0{,}25 = 0{,}75.$$

2.2. Залежні та незалежні події. Умовна ймовірність. Теореми множення ймовірностей

Випадкові події *A* та *B* називають *залежними*, якщо ймовірність появи однієї з них залежить від появи або не появи іншої події. У противному випадку події називають *незалежними.*

Ймовірність події *B*, обчислена при умові появи події *A*, називають *умовною ймовірністю* події *B* і позначають *P(B/A)* або $P_A(B)$. Аналогічно, ймовірність події *A*, обчислена при умові появи події *B*, називають *умовною ймовірністю* події *A* і позначають *P(A/B)* або $P_B(A)$.

Добутком двох випадкових подій *A та B* (позначається $A{\cdot}B$) називається така випадкова подія, яка полягає в тому, що відбуваються одночасно обидві події *A* та *B* (рис. 2.2.1).



$A \cdot B$

Рис. 2.2.1

Якщо події *A* та *B* несумісні, то вони не можуть відбуватись одночасно,отже, $A{\cdot}B = 0$.

Приклад 2.2.1. В урні 20 куль, з них 13 білих та 7 чорних. Навмання беруть дві кулі. Нехай подія *A* - взята біла куля; подія *B* - взята чорна куля. Якщо кулю, що взяли першою, повертають до урни, то ймовірність появи другої кулі не залежить від того, яку кулю взяли першою. Якщо перша куля не повертається до урни, то ймовірність другої події залежить від результату першого випробування.

Якщо першою взяли білу кулю, то в урні залишилося 12 білих куль та 7 чорних, всього 19 куль, тому $P_A(B) = \frac{7}{19}$. Якщо першою взяли чорну кулю, то в урні залишилося 13 білих куль та 6 чорних куль, всього 19, тому $P_A(B) = \frac{6}{19}$.Отже, ймовірність події *B* залежить від появи або не появи події *A.*

Зауваження. Якщо події A та B незалежні,то їх умовна ймовірність дорівнює безумовній імовірності, тобто

$$P_A(B) = P(B) \text{ або } P_B(A) = P(A). \qquad (2.2.1)$$

Теорема 2.2.1. Ймовірність сумісної появи двох залежних випадкових подій A та B дорівнює добутку ймовірностей однієї з цих подій і умовної ймовірності другої події при умові, що перша подія з'явилась, тобто

$$P(A \cdot B) = P(A) \cdot P_A(B) = P(B) \cdot P_B(A). \qquad (2.2.2)$$

Доведення. Нехай появі події A сприяють m наслідків, а появі події B – s наслідків. Усіх можливих наслідків n, а події A B будуть сприяти r наслідків.

Оскільки $P(A\,B) = \frac{r}{n}$, $P(A) = \frac{m}{n}$, $\mathrm{P(B)} = \frac{s}{n}$, $P_A(B) = \frac{r}{m}$, $P_B(A) = \frac{r}{s}$.

$P(AB) = \frac{r}{n} = \frac{m}{n} \cdot \frac{r}{m} = P(A) \cdot P_A(B) = \frac{s}{n} \cdot \frac{r}{s} = P(B)\, P_B(A).$

Дане співвідношення називають формулою *множення ймовірностей залежних випадкових подій*.

Приклад 2.2.2. У коробці є 6 білих і 4 чорних кульки. Яка ймовірність того, що обидві кульки, які достають із коробки, будуть чорними?

Розв'язання. Ймовірність події A (перша кулька чорна) дорівнює $\mathrm{P(A)} = \frac{4}{10}$, а ймовірність події B (друга кулька чорна) - $P_A(B) = \frac{3}{9}$. Дані події залежні, оскільки після того, як відбулася подія A, кількість кульок зменшилася. Шукана подія C (обидві кульки є чорними) дорівнює добутку подій A та B. Для обчислення ймовірності суміщення подій використаємо теорему 2.2.1 множення ймовірностей залежних подій:

$$P(C) = P(A \cdot B) = P(A) \cdot P_A(B) = \frac{4}{10} \cdot \frac{3}{9} = \frac{2}{15}.$$

Наслідок. У випадку незалежних випадкових подій A і B ймовірність сумісної появи двох подій дорівнює добутку ймовірностей цих подій

$$P(A \cdot B) = P(A) \cdot P(B). \qquad (2.2.3)$$

Приклад 2.2.3. Прилад складається з двох вузлів. Імовірність безвідмовної роботи за деякий період цих вузлів відповідно дорівнює 0,8 і 0,9. Знайти ймовірність безвідмовної роботи приладу за вказаний період.

Розв'язання. Подія A – перший вузол працює безвідмовно, подія B – другий вузол працює безвідмовно за вказаний період. Прилад працює безвідмовно (шукана подія C) означає, що безвідмовно працюють обидва вузла, тобто $C = A\ B$. Так як події A та B незалежні, то шукана ймовірність обчислюється за теоремою множення незалежних подій

$$P(C) = P(A \cdot B) = P(A) \cdot P(B) = 0{,}8 \cdot 0{,}9 = 0{,}72.$$

Розглянемо приклад, де використовуються як теорема додавання, так і теорема множення подій.

Приклад 2.2.4. Коефіцієнти використання робочого часу двох комбайнів відповідно дорівнюють 0,8 і 0,6. Враховуючи, що зупинки в роботі кожного комбайна виникають випадково та незалежно одна від одної, знайти ймовірність: а) сумісної роботи комбайнів; б) роботи тільки одного комбайна; в) простою обох комбайнів.

Розв'язання. Позначимо події: A_1 – робота першого комбайна, $P(A_1) = 0{,}8$; A_2 – робота другого комбайна, $P(A_2) = 0{,}6$.

а) подія A (сумісна робота обох комбайнів) – це суміщення подій A_1 і A_2 ; тобто $A = A_1 \cdot A_2$, так як A_1, A_2 – події незалежні, то

$$P(A) = P(A_1 \cdot A_2) = P(A_1) \cdot P(A_2) = 0{,}8 \cdot 0{,}6 = 0{,}48.$$

б) подія *B* (робота тільки одного комбайну) має місце тоді, коли перший комбайн працює, а другий не працює або навпаки: $B = A_1 \cdot \overline{A_2} + \overline{A_1} \cdot A_2$, де $\overline{A}_1$ – простій першого комбайну, $\overline{A_2}$ – простій другого комбайну. $P(\overline{A_1})=0{,}2$; $P(\overline{A_2})=0{,}4$.

Використовуючи теореми додавання несумісних і множення незалежних подій, обчислимо

$P(B) = P(A_1)\cdot P(\overline{A_2}) + P(\overline{A}_1)\cdot P(A_2) = 0{,}8 \cdot 0{,}4 + 0{,}2 \cdot 0{,}6 = 0{,}44$.

в) подія С (простій обох комбайнів) - це суміщення подій $\overline{A_1}$, $\overline{A_2}$. Тому $P(C) = P(\overline{A_1} \cdot \overline{A_2}) = P(\overline{A_1}) \cdot P(\overline{A_2}) = 0{,}2 \cdot 0{,}4 = 0{,}08$.

2.3. Ймовірність появи хоча б однієї випадкової події

Нехай є *n* сумісних випадкових подій $A_1, A_2, \ldots, A_n$. Позначимо через *A* подію, яка полягає в тому, що з'явиться хоча б одна з цих подій. Тоді подія $\overline{A}$ полягає в тому, що події $A_1, A_2, \ldots, A_n$ не з'являться, тобто $\overline{A} = \overline{A}_1 \cdot \overline{A}_2 \cdot \ldots \cdot \overline{A}_n$. Події *A* *і* $\overline{A}$ утворюють повну групу подій: $P(A)+P\left(\overline{A}\right) = 1$. Звідси отримуємо формулу знаходження ймовірності появи хоча б однієї випадкової події:

$$P(A) = 1 - P(\overline{A}) = 1 - \mathrm{P}(\overline{A}_1 \cdot \overline{A}_2 \cdot \ldots \cdot \overline{A}_n = 1 - \mathrm{P}(\overline{A}_1) \cdot \mathrm{P}(\overline{A}_2) \cdot \ldots \cdot \mathrm{P}(\overline{A}_n). \quad (2.2.4)$$

Приклад 2.3.1. Імовірність того, що потрібна працівнику деталь знаходиться в першому та другому ящику, відповідно дорівнює 0,7 і 0,8. Яка ймовірність того, що деталь міститься хоча б в одному ящику?

Розв'язання. Позначимо події: A_1 – деталь знаходиться в першому ящику, $P(A_1) = 0{,}7$; A_2 – деталь знаходиться в другому ящику, $P(A_2) = 0{,}8$.

Ймовірність події Д (деталь міститься хоча б в одному ящику) можна знаходити за різними формулами:

$$a) P(Д) = 1 - P(\overline{A_1}) \cdot P(\overline{A_2}) = 1 - 0{,}3 \cdot 0{,}2 = 1 - 0{,}06 = 0{,}94;$$

де $P(\overline{A_1})=0,3$; $P(\overline{A_2}) = 0,2$.

б) $P(Д) = P(A_1 + A_2) = P(A_1) + P(A_2) – P(A_1)\cdot P(A_2) = 0,8 + 0,7 - 0,7\cdot 0,8 = 1,5 – 0,56 = 0,94$.

в) $Д = A_1\cdot \overline{A_2} + \overline{A_1}\cdot A_2 + A_1\cdot A_2$,

$P(Д) = P(A_1)\cdot P(\overline{A_2}) + P(\overline{A_1})\cdot P(A_2) + P(A_1)\cdot P(A_2) = 0,7\cdot 0,2 + 0,3\cdot 0,8 + 0,7\cdot 0,8 = 0,94$.

2.4. Визначення повної ймовірності події. Ймовірність гіпотез. Формула Байєса

Наслідком теорем додавання та множення ймовірностей є так звана *формула повної ймовірності.*

Нехай треба визначити ймовірність деякої події *A*, яка може відбутися разом із однією з подій $B_1, B_2, \ldots, B_n$, що утворюють повну групу несумісних подій. Будемо ці події називати гіпотезами.

Доведемо, що в цьому випадку ймовірність події *A* обчислюється як сума добутків ймовірностей кожної гіпотези на умовну ймовірність події при цій гіпотезі. Ця формула має назву *формули повної ймовірності*.

Доведення. Оскільки гіпотези $B_1, B_2, \ldots, B_n$ утворюють повну групу, то подія *A* може з'явитись тільки в комбінації з якоюсь із цих гіпотез: $A = B_1A + B_2A + \ldots + B_nA$.

Так як гіпотези $B_1, B_2, \ldots, B_n$ несумісні, то і комбінації $B_1A + B_2A + \ldots + B_nA$ також несумісні; отже, за теоремою додавання отримаємо

$P(A) = P(B_1A) + P(B_2A) + \ldots + P(B_nA)$.

Використовуючи до подій $B_1A + B_2A + \ldots + B_nA$ теорему множення, отримаємо

$$\mathrm{P}(A) = p(B_1)p_{B_1}(A) + p(B_2)p_{B_2}(A) + \ldots + p(B_k)p_{B_k}(A) = \sum_{i=1}^{k} p(B_i)p_{B_i}(A),$$

що є *формулою повної ймовірності*.

Отже, ймовірність події *A*, яка може настати лише за умови появи однієї із несумісних подій $B_1, B_2, ..., B_n$, що утворюють повну групу, дорівнює сумі добутків ймовірностей кожної із усіх подій на відповідну умовну ймовірність події *A*.

$$\mathrm{P}(A) = p(B_1)p_{B_1}(A) + p(B_2)p_{B_2}(A) + ... + p(B_k)p_{B_k}(A) = \sum_{i=1}^{k} p(B_i)p_{B_i}(A) \qquad (2.4.1)$$

***Приклад 2.4.1*.** Партія деталей виготовлена трьома робітниками, причому перший виготовив 25 % продукції, другий робітник – 60% продукції, третій робітник – 15% продукції . Перший робітник допускає 5% браку, другий робітник – 2%, а третій робітник – 3%. Знайти ймовірність того, що випадково вибрана деталь буде бракованою

***Розв'язання*.** Нехай *A* – подія, при якій випадково відібрана деталь буде бракованою. Можливі наступні припущення (гіпотези):

B_1 – дану деталь виготовив перший робітник, $P(B_1)$ = 0,25;

B_2 – дану деталь виготовив другий робітник, $P(B_2)$ = 0,60;

B_3 – дану деталь виготовив третій робітник, $P(B_3)$ = 0,15.

Умовна ймовірність того, що випадково відібрана деталь буде бракованою за умовою, що вона виготовлена:

а) першим робітником, $P_{B1}(A)$ = 0,05 ; б) другим робітником, $P_{B2}(A)$ = 0,02; в) третім робітником, $P_{B3}(A)$ = 0,03.

Використаємо формулу повної ймовірності (2.3.1):

$P(A)=P(B_1)\cdot P_{B1}(A) + P(B_2)\cdot P_{B2}(A) + P(B_3)\cdot P_{B3}(A) = 0{,}25\cdot 0{,}05+$ $+0{,}60\cdot 0{,}02 + 0{,}15\cdot 0{,}03 = 0{,}029$.

***Приклад 2.4.2*. У** першій коробці знаходиться 20 радіоламп, із них 18 стандартних, в другій коробці - 10 ламп, із них 9 стандартних. Із другої коробки навмання взяли лампу та переклали в першу. Знайти ймовірність того, що лампа, яку дістали із першої коробки, буде стандартною.

Розв'язання. Позначимо через *A* подію – із першої коробки дістали стандартну лампу. З другої коробки можна перекласти стандартну лампу (подія B_1) або нестандартну (подія B_2). Ймовірність того, що із другої коробки дістали стандартну лампу, дорівнює $P(B_1) = \frac{9}{10}$. Ймовірність того, що із другої коробки дістали нестандартну лампу, дорівнює $P(B_2) = \frac{1}{10}$. Умовна ймовірність, що із першої коробки дістали стандартну лампу за умови, що із другої коробки в першу переклали стандартну лампу, дорівнює $P_{B_1}(A) = \frac{19}{21}$. Умовна ймовірність того, що з першої коробки дістали стандартну лампу за умови, що з другої коробки в першу переклали нестандартну лампу, дорівнює $P_{B_2}(A) = \frac{18}{21}$.

Шукана ймовірність того, що із першої коробки дістали стандартну лампу, за формулою повної ймовірності дорівнює

$$P(A) = P(B_1) \cdot P_{B1}(A) + P(B_2) \cdot P_{B2}(A) = \frac{9}{10} \cdot \frac{19}{21} + \frac{1}{10} \cdot \frac{18}{21} = 0{,}9.$$

Тепер ознайомимося з *формулами Байєса*.

За формулою повної ймовірності невідомо з якою із несумісних подій $B_1, B_2, \ldots, B_n$ з'явиться подія *A.* Тому кожну з подій $B_1, B_2, \ldots, B_n$ можна вважати гіпотезою. Тоді $P(B_i)$ - ймовірність *i*-ої гіпотези.

Якщо випробування проведене, і в результаті його подія *A* з'явилась, то умовна ймовірність $P_A(B_i)$ може не дорівнювати $P(B_i)$. Порівняння ймовірностей $P(B_i)$ та $P_A(B_i)$ дозволяє переоцінити ймовірність гіпотез за умови, що подія *A* з'явилася.

Якщо подія *A* може настати за умови появи однієї із несумісних подій B_i (i = *1,2,…, n*) гіпотез, то для обчислення умовної ймовірності використовуємо теорему множення ймовірностей залежних подій

$$P(A \cdot B_i) = P(A) \cdot P_A(B_i) = P(B_i) \cdot P_{B_i}(A).$$

Порівнюючи між собою ці формули, отримаємо *формули Байєса*:

$$\boldsymbol{P_A(B_i) = \frac{P(B_i)P_{B_i}(A)}{\sum_{i=1}^{k} P(B_i)P_B(A)} = \frac{P(B_i)P_{B_i}(A)}{P(A)}} \qquad (2.4.2)$$

Приклад 2.4.3. Два автомати виготовляють однакові деталі, які поступають на один конвеєр. Продуктивність праці першого автомата вдвічі більше продуктивності другого. Перший автомат виготовляє в середньому 60% деталей відмінної якості, а другий - 84%. Навмання взята з конвеєра деталь відмінної якості. Знайти ймовірність того, що ця деталь виготовлена першим автоматом.

Розв'язання. Подія A – виготовлена деталь відмінної якості. Можна зробити два припущення (гіпотези): B_1 – деталь виготовлена першим автоматом. Так як перший автомат виготовляє вдвічі більше деталей, то $P(B_1) = \frac{2}{3}$; B_2 – деталь виготовлена другим автоматом, $P(B_2) = \frac{1}{3}$.

Умовна ймовірність того, що деталь відмінної якості, якщо вона виготовлена першим автоматом, дорівнює $P_{B1}(A) = 0{,}6$. Умовна ймовірність того, що деталь відмінної якості, якщо вона виготовлена другим автоматом, дорівнює $\boldsymbol{P_{B_2}(A) = 0{,}84}$.

Ймовірність того, що навмання взята деталь буде відмінної якості, за формулою повної ймовірності, дорівнює

$$P(A) = P(B_1) \cdot P_{B1}(A) + P(B_2) \cdot P_{B2}(A) = \frac{2}{3} \cdot 0{,}6 + \frac{1}{3} \cdot 0{,}84 = 0{,}68.$$

Шукана ймовірність того, що взяли навмання деталь відмінної якості, яка виготовлена на першому автоматі, за формулою Байєса дорівнює

$$P_A(B_1) = \frac{P(B_1) \cdot P_{B1}(A)}{P(A)} = \frac{\frac{2}{3} \cdot 0{,}6}{0{,}68} \approx 0{,}59.$$

Правило-орієнтир для розв'язування задач за формулою повної ймовірності

Для обчислення ймовірності розглядуваної події за формулою повної ймовірності необхідно:

1. Усвідомити послідовність випробувань, які розглядаються в задачі.
2. Позначити подію, ймовірність появи якої треба знайти, якоюсь літерою, наприклад А.
3. Скласти множину попарно несумісних гіпотез H_1, H_2,..., H_n. Перевірити, що об'єднання їх співпадає з простором елементарних подій випробування, яке розглядається.
4. Обчислити ймовірності кожної із гіпотез та умовні ймовірності появи події А при умові, що відбулась подія H_i (і = 1, 2, ... ,n) (якщо вони не дані в умові задачі).
5. За формулою (2.4.1) обчислити ймовірність події А. Якщо з умови задачі відомо, що подія А вже відбулася, то треба скористатися формулою (2.4.2).

Питання для самоперевірки

1. Як означають та позначають суму, добуток випадкових подій?
2. Як формулюють і якими формулами записують теореми додавання ймовірностей сумісних та несумісних подій?

3. Які випадкові події називаються незалежними; залежними? Наведіть приклади залежних і незалежних подій.

4. Як визначають та позначають умовну ймовірність?

5. Сформулюйте теореми множення ймовірностей залежних і незалежних випадкових подій.

6. За якою формулою можна обчислити ймовірність появи хоча б однієї випадкової події з *n* сумісних випадкових подій ?

7. Які події складають повну групу подій? Як визначають та позначають повну групу подій?

8. Яким умовам повинна задовольняти подія, щоб її ймовірність можна було знаходити за формулою повної ймовірності? Який вигляд має ця формула?

9. В яких випадках можна використовувати формули Байєса та як їх записують?

Задачі та вправи

2.1. Продано 1000 білетів лотереї. На один білет випадає виграш у 500 грн., на 10 білетів - у 100 грн., на 50 білетів - у 20 грн., на 100 білетів - у 5 грн., інші білети – невиграшні. Гравець купив один лотерейний білет. Знайти ймовірність того, що він виграє не менше 20 грн.

2.2. Відомо, що в деякому регіоні N 40% компаній мають у штаті юриста (подія *A*) і 80% компаній мають у штаті бухгалтера (подія *B*). Вважається, що події *A* і *B* незалежні. Знайти ймовірність того, що фірма має одночасно в штаті і юриста і бухгалтера.

2.3. У цеху працюють 7 чоловіків і 3 жінки. За табельними номерами навмання відібрали 3-х робітників. Знайти ймовірність того, що всі відібрані робітники будуть чоловіками.

2.4. У двох ящиках знаходяться кульки, які відрізняються тільки кольором, причому в першому ящику 5 білих, 11 чорних, 8 червоних кульок, а в другому відповідно 10, 8 і 6. З обох ящиків навмання дістають по одній кульці. Яка ймовірність того, що обидві кульки одного кольору?

2.5. На ділянці три бригади. Ймовірність виконання плану першою бригадою дорівнює 0,8; другою – 0,85; третьою – 0,7. Знайти ймовірність виконання плану: а) ділянкою; б) тільки однією бригадою.

2.6. Фірма має можливість отримати два контракти. Ймовірність отримання першого контракту складає 0,9; а другого - 0,8. Вважаючи ці події незалежними, визначити ймовірність того, що фірма отримає: а) обидва контракти; б) жодного контракту; в) щонайменше один контракт; г) не більше одного контракту.

2.7. Коефіцієнти використання робочого часу у 3-х тракторів відповідно дорівнюють 0,8; 0,7; 0,6. Вважаючи, що зупинки в роботі кожного трактора виникають випадково та незалежно одна від однієї, визначити ймовірність: а) сумісної роботи усіх трьох тракторів; б) сумісної роботи двох тракторів; в) роботи тільки одного трактора; г) роботи хоча б одного трактора.

2.8. Ймовірність того, що необхідна книга знаходиться у фондах першої бібліотеки, дорівнює 0,8; другої – 0,9; третьої – 0,7. Знайти ймовірність того, що книга знаходиться у фондах хоча б двох бібліотек.

2.9. Ймовірність того, що виготовлена на першому верстаті деталь буде першосортна, дорівнює 0,7. При виготовленні такої деталі на другому верстаті ця ймовірність дорівнює 0,8. На першому верстаті виготовлені дві деталі, а на другому - три. Знайти ймовірність того, що всі деталі першосортні.

2.10. В автобусі їдуть 6 пасажирів. На наступній зупинці може вийти кожен з цих пасажирів з ймовірністю 0,2. Крім того, на зупинці в автобус може ввійти один новий пасажир з ймовірністю

0,3 або може не ввійти з ймовірністю 0,7. Знайти ймовірність того, що, коли автобус відійде від наступної зупинки, в ньому, як і раніше, буде 6 пасажирів.

2.11. Для сповіщення про аварію установлено два сигналізатори, що працюють незалежно. Ймовірність того, що при аварії запрацює перший сигналізатор дорівнює 0,95; другий – 0,9. Знайти ймовірність того, що при аварії запрацює: а) тільки один сигналізатор; б) хоча б один сигналізатор.

2.12. Користувач соціальних мереж три рази викликає кореспондента. Ймовірність прийняття першого виклику становить 0,2; другого виклику - 0,3; третього виклику - 0,4. Знайти ймовірність установлення зв'язку.

2.13. Три стрілка стріляють по мішені. Ймовірність влучення для першого стрілка рівна 0,7; для другого - 0,8; для третього - 0,85. Знайти ймовірність того, що влучать в ціль: а) два стрілка; б) тільки один; в) хоча б два стрілка; г) не влучить ні один.

2.14. У слюсаря є 16 деталей, що виготовлені заводом № 1, і 4 деталі – заводом № 2. Навмання взято дві деталі. Знайти ймовірність того, що хоча б одна із них буде виготовлена заводом № 1.

2.15. Схожість насіння кукурудзи складає 80%. У кожне гніздо висівається по 2 зерна. Чому дорівнює ймовірність того, що в кожному гнізді проросте хоча б одне зерно?

2.16. У трьох кошиках знаходиться картопля. У першому – 10% пошкоджених бульб, у другому – 15%, у третьому – 10%. Навмання беруть одну бульбу. Яка ймовірність того, що вона пошкоджена?

2.17. Компанія по страхуванню автомобілів розподіляє водіїв на три класи: клас I (малий ризик), клас II (середній ризик), клас III (великий ризик). Відомо, що з усіх застрахованих водіїв 30%

належить до класу I, 45% - до класу II та 25% - до класу III. Ймовірність того, що протягом 12 місяців водій I класу потрапить хоч би в одну дорожню транспортну пригоду (ДТП) складає 0,01; для водія II класу ця ймовірність дорівнює 0,03; а для водія класу III - 0,10. Визначити ймовірність подій: а) протягом 12 місяців водій *N* потрапить в ДТП; б) відомо що водій *N* потрапив в ДТП. Яка ймовірність того, що він водій класу I?

2.18. У групі спортсменів 20 лижників, 6 велосипедистів і 4 бігуна. Ймовірність виконати кваліфікаційну норму для лижника 0,95; для велосипедиста - 0,85; для бігуна – 0,75. Знайти ймовірність того, що спортсмен, якого визвали навмання, виконає норму.

2.19. Економічний університет провів обстеження працевлаштування своїх випускників. Ймовірність того, що людина, яка працює у сфері бізнесу, має прибуток вище К грн.; становить 0,9; а поза сферою бізнесу (з тим же прибутком) – 0,3. 80% випускників працюють у сфері бізнесу. Навмання вибраний випускник університету. Визначити ймовірність того, що: а) він отримує прибуток вище К грн.; б) з'ясувалось, що він отримує прибуток вище К грн. Яка ймовірність, що він працює у сфері бізнесу?

2.20. У двох ящиках знаходяться деталі. В першому ящику містяться 12 деталей, із них 1 нестандартна, в другому – 10 деталей, із них 1 нестандартна. З першого ящика навмання взяли деталь і переклали в другий. Знайти ймовірність того, що навмання взята деталь із другого ящика буде нестандартною.

2.21. Робітник обслуговує 3 верстати, на яких обробляються однотипні деталі. Ймовірність браку на першому верстаті дорівнює 0,02; на другому – 0,05; на третьому – 0,04. Оброблені деталі складають в один ящик. Продуктивність першого верстата в три рази більша ніж другого, а третього в три рази менша ніж другого. Знайти ймовірність того, що взята навмання деталь буде бракована.

2.22. У піраміді 10 гвинтівок, із яких 4 з оптичним прицілом. Ймовірність того, що стрілок влучить в мішень при пострілі із гвинтівки з оптичним прицілом, дорівнює – 0,95; для гвинтівки без оптичного прицілу ця ймовірність дорівнює 0,8. Стрілок влучив у мішень із навмання взятої гвинтівки. Що ймовірніше: стрілок вистрілив із гвинтівки з оптичним прицілом чи без нього?

2.23. Для участі в студентських відбірних спортивних змаганнях визначили з першої групи курсу – 4, з другої – 6, з третьої групи – 5 студентів. Ймовірність того, що студент першої, другої і третьої групи попаде в збірну університету, відповідно дорівнюють 0,9; 0,7 і 0,8. Навмання вибраний студент у результаті змагання попав у збірну. До якої із груп найімовірніше належав цей студент?

2.24. Лиття в болванках поступає з двох заготовчих цехів: 70% з першого і 30% з другого. При цьому матеріал першого цеху має 10% браку, а другого – 20%. Знайти ймовірність того, що взята навмання болванка без дефектів.

2.25. У господарстві є 6 гусеничних тракторів і 4 колісних. Ймовірність того, що за час виконання деякої роботи гусеничний трактор не вийде з ладу, рівна 0,95; а для колісного трактора – 0,8. Для виконання деякої роботи довільно вибирається трактор. Знайти ймовірність того, що до закінчення роботи трактор не вийде з ладу.

2.26. В урні три кулі білих, чорна і червона. З урни виймають кулі по одній п'ять раз, причому, після кожного виймання кулю повертають в урну. Знайти ймовірність того, що чорну і білу кулю візьмуть не менше, ніж по два рази кожну.

2.27. Робітник виготовляє з імовірністю 0,9 якісний виріб; з імовірністю 0,09 - виріб, дефект якого можна усунути, та з імовірністю 0,01 - виріб, дефект якого усунути неможливо. Виготовлено чотири вироби. Знайти ймовірність того, що серед цих виробів два якісних і хоча б один з дефектом, який можна усунути.

2.28. У першому ящику міститься 20 деталей, із них 15 стандартних. В другому – 30 деталей, із них 24 стандартних, в третьому – 10 деталей, із них 6 стандартних. Знайти ймовірність того, що дістали стандартну деталь із навмання взятого ящика.

2.29. З двадцяти стрільців четверо влучають у мішень з імовірністю 0,9; десять - з імовірністю 0,7 і шестеро - з імовірністю 0,5. Знайти ймовірність того, що навмання вибраний стрілець влучить у мішень. Навмання вибраний стрілець влучив у мішень. Яка ймовірність того, що він належить до третьої групи стрільців?

2.30. Страхова компанія поділяє застрахованих на три класи ризику: 1 клас - малий ризик, II клас - середній ризик і III клас - великий ризик. Серед клієнтів компанії 50% належать до першого класу ризику, 30% - до другого і 20% - до третього. Ймовірність того, що компанії доведеться виплачувати страхове винагородження клієнтам першого класу ризику, дорівнює 0,01; другого - 0,03 і третього — 0,08. Яка ймовірність того, що: а) застрахований отримає грошове винагородження за термін страхування; б) застрахований, що отримав грошове винагородження, належить до групи малого ризику?

2.31. Пасажир може купити квиток в одній із трьох кас. Ймовірності звернення в кожну касу залежать від їх місцезнаходження і дорівнюють відповідно 0,3; 0,5 і 0,2. Ймовірності того, що квитки до приходу пасажира будуть розпродані, дорівнюють відповідно 0,6; 0,8 і 0,7. Пасажир підійшов до однієї з кас і купив квиток. Знайти ймовірність того, що це була перша каса.

2.32. Для того, щоб здати випробування, студенти повинні підготувати 30 питань. З 25 студентів 10 підготували всі питання, 8 студентів підготували 25 питань, 5 студентів - 20 питань і 2 студенти - 15 питань. Знайти ймовірність того, що викликаний навмання студент відповість на поставлене питання. Викликаний студент відповів на питання. Яка ймовірність того, що цей студент: а) підготував всі питання; б) підготував тільки половину питань?

2.33. У першій урні знаходяться дев'ять чорних і одна біла кулі, у другій - одна чорна і п'ять білих куль. З кожної урни навмання забрали по одній кулі. Знайти ймовірність того, що ці кулі будуть різного кольору.

2.34. Продукцію, яку виготовляють на підприємстві, перевіряють 2 контролери, причому, перший контролер перевіряє 55% виробів, а другий - 45%. Імовірність того, що перший контролер недогляне нестандартний виріб, дорівнює 0,01; а другий - 0,02. Яка ймовірність того, що взятий навмання виріб буде нестандартний? Взятий навмання виріб виявився нестандартним. Яка ймовірність того, що цей виріб перевіряв другий контролер?

2.35. У першій з двох урн є 3 білих та 4 чорних кулі, у другій - 2 білих та 3 чорних. З першої урни до другої перекладають навмання 2 кулі, а потім з другої урни виймають одну кулю. Який склад перекладених куль найімовірніший, якщо куля, що взята з другої урни, виявилась білою?

2.36. Підкидають монету. Якщо випаде герб, то виймають одну кулю з першої урни. Якщо ж випаде цифра, то виймають кулю з другої урни. У першій урні знаходяться 1 біла і 3 червоні кулі. У другій урні є 3 білі та 1 червона кулі. Яка ймовірність того, що: а) вийнята куля червона? б) кулю вийнято з першої урни, якщо вона виявилась червоною?

2.37. У кожній із трьох урн знаходяться по 6 чорних і 4 білих кулі. З першої урни навмання вийняли одну кулю і переклали до другої, після чого з другої урни вийняли навмання одну кулю і переклали до третьої. Знайти ймовірність того, що вийнята з третьої урни куля - біла.

2.38. У групі з десяти студентів, що прийшли на випробування, троє підготували двадцять питань з 20, 4 студенти підготували 16 з 20 питань, 2 студенти - 10 з 20 питань і один студент - 5 питань з 20. Викликаний навмання студент відповів на три довільно задані питання. Знайти ймовірність того, що цей студент: а) підготував всі питання; б) підготував 5 питань.

2.39. У першій урні знаходиться 5 білих і 3 чорних кулі, у другій - 2 білих і 4 чорних. З першої до другої урни перекладають навмання 2 кулі. Після цього з другої урни беруть 2 кулі. Яка ймовірність того, що з другої урни взято дві кулі різного кольору? Які кулі найімовірніше були перекладені з першої до другої урни?

2.40. Група студентів складається з чотирьох відмінників, шести, що навчаються добре, і двадцяти, які навчаються як-небудь. Відмінники на випробуванні можуть отримати тільки відмінні оцінки; ті, що навчаються добре, можуть отримати відмінні та добрі оцінки, ті, що навчаються як-небудь можуть отримати з однаковою ймовірністю добрі, задовільні і незадовільні оцінки. Для відповіді навмання викликали студента. Знайти ймовірність того, що студент отримає відмінну або добру оцінку. Викликали студента та він отримав добру оцінку. З якої групи найімовірніше цей студент?

2.41. Три доньки: Катя, Марина і Оля - мають обов'язок у сім'ї - мити посуд. Через те, що Катя найстарша, вона виконує 40%

роботи. Решту - 60% роботи виконують Марина та Оля порівну. Якщо миє посуд Катя, то ймовірність того, що вона розіб'є тарілку, дорівнює 0,02. Для Марини та Олі ці ймовірності такі: 0,03 та 0,04. Батьки не знають, хто мив посуд, але вони почули дзвін розбитої тарілки. Яка ймовірність того, що мила посуд Катя? Марина? Оля?

2.42. Число вантажних автомашин, які проїжджають по шосе, де стоїть бензоколонка, відносяться до числа легкових автомашин, що теж проїжджають по шосе, як 3:2. Імовірність того, що вантажна автомашина буде заправлятися бензином, дорівнює 0,1; для легкової автомашини ця ймовірність дорівнює 0,2. До бензоколонки під'їхала машина. Знайти ймовірність того, що це вантажна машина.

2.43. З першого верстата поступає 40% деталей, другого - 30%, третього - 20% і з четвертого - 10%. Серед деталей, які випускає перший верстат, 2% бракованих; другий - 1%, третій - 0,5%, четвертий - 0,2%. Знайти ймовірність того, що деталь, яка виготовлена на одному з верстатів, бракована. Виявлено, що деталь бракована. З якого верстата вона найімовірніше поступила?

2.44. У першій бригаді 8 робітників мають третій розряд і 6 - четвертий. В іншій бригаді 5 робітників мають третій розряд і 5 - четвертий. З першої до другої бригади перевели двох робітників. Знайти ймовірність того, що два робітники, взяті навмання з другої бригади, мають третій розряд. Виявилось, що з другої бригади взяли двох робітників з третім розрядом. Яка ймовірність того, що з першої до другої бригади перевели двох робітників: а) з четвертим розрядом; б) з третім розрядом; в) один - четвертого і один - третього розрядів?

2.45. Перший цех випускає 40% всієї продукції заводу, а другий - 60%. В середньому 9 одиниць з 1000 одиниць продукції першого цеху виявляються браком, а другого - 2 одиниці з 500. Якась одиниця продукції, яку вибрали випадково для контролю, виявилась бракованою. Яка ймовірність того, що її було вироблено: а) другим цехом; б) першим цехом?

2.46. Чотири стрільці незалежно один від одного стріляють у одну і ту ж мішень. Ймовірність влучення для кожного з них відповідно дорівнює: 0,4; 0,6; 0,7; 0,8. Після того, як стрільці вистрілили у мішень по одному разу, виявилось, що у мішені є три пробоїни. Знайти ймовірність того, що не влучив у мішень четвертий стрілець.

2.47. До спеціалізованої лікарні приймають у середньому 50% хворих на хворобу К, 30% - на хворобу Н, 20% - на хворобу М. Імовірність повного вилікування хвороби К дорівнює 0,7; для хвороб Н і М ці ймовірності дорівнюють відповідно 0,8 і 0,9. Хворий, що пройшов лікування, виписався здоровим. Знайти ймовірність того, що він хворів хворобою К.

2.48. На базу поступили швейні вироби, з яких 20% виготовлені фабрикою № 1, 30% - фабрикою № 2 і 50% - фабрикою № 3. Ймовірність того, що виріб фабрики №1 має перший сорт, дорівнює 0,7; а для фабрик № 2 і № 3 ймовірності дорівнюють відповідно 0,8 і 0,9. Знайти ймовірність того, що навмання взятий виріб буде виробом першого сорту. Взятий виріб - першого сорту. З якої фабрики найімовірніше поступив цей виріб?

2.49. У відділі *А* працюють 5 інженерів і 3 старших інженери, а у відділі *В* - 8 інженерів і 2 старших інженери. З відділу *А* до відділу *В* перевели одного співробітника. Знайти ймовірність того, що: а) з трьох співробітників, яких вибрали з нового складу відділу *А,* всі три - старші інженери; б) з відділу *А* у відділ *В* перевели старшого інженера.

2.50. У трьох урнах є білі та чорні кулі. В першій урні - 2 білих та 3 чорних кулі, у другій - 2 білих та 2 чорних кулі, у третій - 3 білих та 1 чорна. З першої урни перекладають одну кулю у другу; потім з другої урни одну кулю перекладають у третю, і, насамкінець, з третьої урни перекладають одну кулю у першу. Який склад куль у першій урні після проведених операцій буде найімовірніший? Знайти ймовірність того, що у всіх урнах склад куль не зміниться.

2.51. Для переливання крові необхідно враховувати групу крові донора і хворого. Людині, яка має четверту групу крові, можна перелити кров будь-якої групи; людині, що має другу або третю групу крові, можна переливати кров або тієї ж групи, або першої; людині з першою групою крові можна переливати тільки кров першої групи. Серед населення 33,7% людей мають першу, 37,5% - другу; 20,9% - третю і 7,9% - четверту групу крові. Знайти ймовірність того, що : а) випадково вибраному хворому можна перелити кров випадкового взятого донора; б) переливання крові можна здійснити, якщо є 2 донори.

2.52. На фабриці виготовляють гвинти. Перша машина виготовляє 25%, друга - 35% і третя - 40% усіх виробів. Частка браку продукції цих машин складає відповідно 5%, 4% і 2%. Яка ймовірність того, що навмання взятий гвинт бракований? Навмання взятий гвинт виявився бракованим. Яка ймовірність того, що його виготовлено першою, другою, третьою машинами?

2.53. У рибалки є три улюблені місця для лову риби, які він відвідує з однаковою ймовірністю Якщо він закидає вудку на першому місці, то риба клює з ймовірністю 0,7; на другому місці - з ймовірністю 0,65; на третьому місці - з ймовірністю 0,75. Відомо, що рибалка, вийшовши на лов риби, три рази закинув вудку та риба клюнула тільки один раз. Знайти ймовірність того, що він ловив рибу на першому місці.

2.54. Завод виготовляє вироби, кожен з яких з ймовірністю 0,02 має дефект. У цеху виріб з однаковою ймовірністю оглядає один з двох контролерів. Перший контролер виявляє дефект з ймовірністю 0,85; а другий - з ймовірністю 0,9. Якщо в цеху виріб не забраковано, то він поступає до ВТК заводу, де дефект, якщо він є, виявляють з ймовірністю 0,95. Відомо, що виріб забраковано. Знайти ймовірність того, що його забраковано: а) першим контролером; б) другим контролером; в) ВТК заводу.

2.55. Прилад складається з двох дублюючих один одного вузлів і може випадково працювати у двох режимах: сприятливому та несприятливому. У сприятливому режимі надійність кожного вузла дорівнює 0,9; у несприятливому - 0,7. Імовірність того, що прилад буде працювати в сприятливому режимі, дорівнює 0,6; а у несприятливому - 0,4. Знайти повну надійність приладу.

РОЗДІЛ 3
ПОВТОРНІ НЕЗАЛЕЖНІ ВИПРОБУВАННЯ

При практичному використанні теорії ймовірностей часто зустрічаються задачі, в яких одне й те ж випробування зустрічається неодноразово. В результаті кожного випробування може з'явитися або не з'явитися деяка подія *A*. Причому, нас цікавить не результат кожного випробування, а загальне число появи події А в результаті серії випробувань. Наприклад, якщо стрілок проводить декілька пострілів по мішені, то, як правило, буде цікавити не результат кожного пострілу, а загальна кількість попадань. У подібних задачах потрібно вміти визначити ймовірність довільного заданого числа появи події в результаті серії дослідів. Такі задачі розв'язуються для незалежних повторних випробувань.

Декілька випробувань називаються *незалежними*, якщо ймовірність того чи іншого наслідку кожного із випробувань не залежить від того, які наслідки мали інші випробування. Наприклад, декілька послідовних підкидань монети або грального кубика є незалежними випробуваннями. Незалежні випробування можуть проводитися в різних або однакових умовах. У першому випадку ймовірність події *A* від випробування до випробування змінюється. Таку послідовність незалежних випробувань називають поліноміальною схемою. У другому випадку ймовірність події *A* у всіх випробуваннях однакова. Така послідовність незалежних випробувань називається схемою Бернуллі.

3.1. Схема Бернуллі

Розглянемо схему Бернуллі, яка складається з *n* незалежних випробувань, кожне з яких має тільки два наслідки: поява події *A* (успіх) або непоява $\overline{A}$ (невдача), причому ймовірність того, що подія відбулася $P(A) = p$ при кожному випробуванні стала і не залежить від кількості випробувань, а ймовірність протилежної події $\overline{A}$ (непоява події A) дорівнює $P(\overline{A}) = 1 - p = q$. Числа p і q називаються параметрами схеми Бернуллі.

Приклад 3.1.1. Нехай проводиться три незалежних постріли по мішені, ймовірність попадання в яку при кожному пострілі дорівнює p. Знайти ймовірність того, що при цих трьох пострілах ми отримаємо рівно два попадання.

Розв'язання. Позначимо через B_2 подію, яка означає два попадання по мішені. Ця подія може відбутися трьома способами:

1) попадання при першому пострілі, попадання при другому пострілі, промах при третьому;

2) попадання при першому пострілі, промах при другому пострілі, попадання при третьому;

3) промах при першому пострілі, попадання при другому пострілі, попадання при третьому.

Отже, подію B_2 можна представити як суму добутків подій:

$$B_2 = A_1 A_2 \bar{A}_3 + A_1 \bar{A}_2 A_3 + \bar{A}_1 A_2 A_3,$$

де A_1, A_2, A_3 – попадання при першому, другому, третьому пострілах відповідно; $\bar{A}_1$, $\bar{A}_2$, $\bar{A}_3$ - промах при першому, другому та третьому пострілах відповідно.

Враховуючи, що три перераховані варіанти події B_2 несумісні, а події, які входять у добуток, незалежні, то за теоремами додавання і множення отримаємо:

$$P(B_2) = p \cdot p \cdot (1 - p) + p(1 - p) \cdot p + (1 - p) \cdot p \cdot p$$

або $P(B_2) = 3p^2 q$, враховуючи те, що $1 - p = q$.

Аналогічно, перераховуючи всі можливі варіанти, в яких подія, що нас цікавить, може з'явитися задане число разів, можна розв'язати і наступну загальну задачу.

Виконуються n незалежних випробувань, в кожному з яких подія A може з'явитися або не з'явитися; ймовірність появи події A в кожному випробуванні рівна p, а ймовірність того, що подія A не з'явиться, рівна $q = 1 - p$. Треба знайти ймовірність події B_m, суть якої полягає в тому, що в n незалежних випробуваннях подія A настане рівно m раз. Цю подію можна подати різними способами. Розкладемо подію B_m як суму добутків подій, які складаються з появи або непояви події А в окремому випробуванні:

$B_m = A_1 A_2 \ldots A_m \bar{A}_{m+1} \ldots \bar{A}_n + + A_1 \bar{A}_2 A_3 \ldots \bar{A}_{n-1} A_n + \ldots + \bar{A}_1 \bar{A}_2 \ldots \bar{A}_{n-m} A_{n-m+1} \ldots A_n$,

причому в кожний добуток подія A повинна входити m раз, а подія $\bar{A}$ $(n\text{-}m)$ раз. Число всіх комбінацій такого роду дорівнює C_n^m. Ймовірність кожної такої комбінації, за теоремою множення для незалежних подій, дорівнює $p^m q^{n-m}$. Так як комбінації між собою несумісні, то, за теоремою додавання, ймовірність події B_m рівна

$$P_n(m) = C_n^m p^m q^{n-m} \qquad \text{(формула Бернуллі).} \qquad (3.1.1)$$

Приклад 3.1.2. Ймовірність того, що витрати електроенергії протягом доби не перевищують норми, дорівнює p = 0,75. Знайти ймовірність того, що в наступні шість діб витрати електроенергії протягом 4-х діб не перевищать норму.

Розв'язання*.* Імовірність перевитрат електроенергії дорівнює q = 1–p= 1 – 0,75 = 0,25. Шукана ймовірність за формулою Бернуллі дорівнює

$$P_6(4) = C_6^4 \cdot p^4 \cdot g^2 = C_6^2 \cdot p^4 \cdot g^2 = \frac{6 \cdot 5}{1 \cdot 2} \cdot (0{,}75)^4 \cdot (0{,}25)^2 = 0{,}30.$$

Приклад 3.1.3. Схожість насіння становить 95%. Відбирається навмання 6 зерен. Яка ймовірність того, що вони дадуть не менше п'яти сходів?

Розв'язання*.* Події “не менше п'яти сходів” відповідає наступна комбінація подій: “рівно п'ять сходів” або “рівно шість сходів”. На основі теореми додавання несумісних подій шукана ймовірність дорівнює

$P_6(m \geq 5) = P_6(5) + P_6(6).$

Ймовірності $P_6(5)$ і $P_6(6)$ визначаються за формулою Бернуллі (p = 0,95; q = 0,05) :

$P_6(5) = C_6^5 \cdot (0{,}95)^5 \cdot (0{,}05)^1 = 6 \cdot (0{,}95)^5 \cdot 0{,}01 \approx 0{,}23;$

$P_6(6) = C_6^6 \cdot (0{,}95)^6 \cdot (0{,}05)^0 = 1 \cdot (0{,}95)^6 \cdot 1 \approx 0{,}73.$

Отже, $P_6(m \geq 5) = 0{,}23+0{,}73 = 0{,}96$.

3.2. Біноміальний розподіл ймовірностей

Нехай виконуються n незалежних повторних випробувань, в кожному з яких подія A може з'явитися m раз. Якщо обчислити ймовірності для всіх значень m (m = 0, 1, 2, …n) за формулою Бернуллі, то отримаємо відповідну послідовність ймовірностей $P_n(0), P_n(1), P_n(2), \ldots, P_n(n-1), P_n(n)$, яка називається *біномним розподілом ймовірностей.* Порівнюючи формулу Бернуллі та розклад бінома Ньютона, бачимо, що права частина формули Бернуллі дорівнює загальному члену розкладу бінома Ньютона.

Біноміальний (біномний) розподіл ймовірностей дозволяє визначити не тільки ймовірність появи події A m раз при n незалежних випробуваннях, а і ймовірність $P_n(m_1 \leq m \leq m_2)$ того, що число m появи події A настане не менше ніж m_1 раз і не більше ніж m_2 разів. У силу несумісності подій при m_1, $m_1+1, \ldots, m_2-1, m_2$ має місце формула:

$$P_n(m_1 \leq m \leq m_2) = P(m_1) + P(m_1 + 1) + \cdots + P\left(m_2 - 1\right) + \\ +P(m_2). \qquad (3.2.1)$$

Біноміальний розподіл ймовірностей можна подати у вигляді таблиці, яка складається з двох рядків. У першому рядку розміщені числа появи події A, як значення випадкової величини

X, а в другому – відповідні їм імовірності, які обчислені за формулою Бернуллі.

Оскільки число появи події *A* при *n* незалежних повторних випробуваннях складає повну групу подій, то

$$P_n(0) + P_n(1) + P_n(2) + \ldots + P_n(n-1) + P_n(n) = 1.$$

Приклад 3.2.1. Проводяться чотири незалежні досліди, в кожному з яких подія *A* може з'явитись з імовірністю 0,4. Розглядається випадкова величина *X* – число появи події *A* у чотирьох дослідах. Записати ряд розподілу та побудувати многокутник розподілу випадкової величини *X*.

Розв'язання. У чотирьох дослідах подія *A* може не з'явитись або з'явитись 1, 2, 3 або 4 рази. Це означає, що випадкова величина *X* може приймати значення 0; 1; 2; 3; 4. Ймовірності цих подій обчислимо за формулою Бернуллі:

$p_0 = P(X = 0) = C_4^0 \cdot 0{,}4^0 \cdot 0{,}6^4 = 0{,}1296;$

$p_1 = P(X = 1) = C_4^1 \cdot 0{,}4 \cdot 0{,}6^3 = 0{,}3456;$

$p_2 = P(X = 2) = C_4^2 \cdot 0{,}4^2 \cdot 0{,}6^2 = 0{,}3456;$

$p_3 = P(X = 3) = C_4^3 \cdot 0{,}4^3 \cdot 0{,}6^4 = 0{,}1536;$

$p_4 = P(X = 4) = C_4^4 \cdot 0{,}4^4 \cdot 0{,}6^0 = 0{,}0256.$

Біноміальний закон розподілу величин *X* подано в таблиці (табл. 3.2.1).

Таблиця 3.2.1

X	0	1	2	3	4
p	0,1296	0,3456	0,3456	0,1536	0,0256

За даними таблиці побудуємо многокутник (полігон) розподілу випадкової величини *X* (рис. 3.2.1).

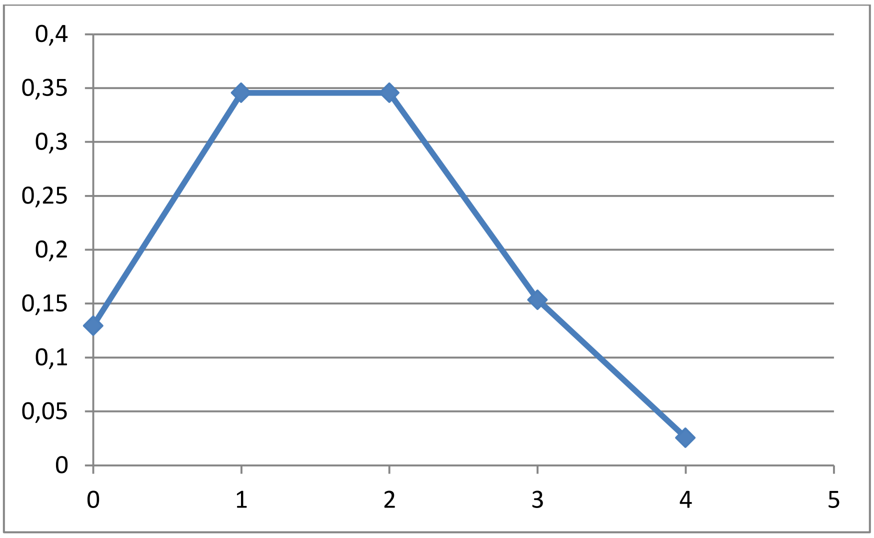


Рис. 3.2.1

3.3. Найімовірніше число настання події

Якщо виконуються n незалежних повторних випробувань, то в багатьох задачах виникає момент знаходження такого значення m_0, при якому ймовірність появи події найбільша серед імовірностей, які обчислені для значень m (m = 0, 1, 2, …,n).

Число m_0 називається *найімовірнішим числом настання події* в незалежних повторних випробуваннях, якщо ймовірність того, що подія з'явиться в цих випробуваннях m_0 разів, перевищує (або, принаймні, не менше) ймовірності решта можливих наслідків випробування, та знаходиться за формулою

$$np - q \le m_0 \le np + p, \qquad (3.3.1)$$

де m_0 – цілий розв'язок подвійної нерівності.

Зокрема: а) якщо число ($np - q$) – дробове, то існує одне найімовірніше число m_0; б) якщо число ($np - q$) – ціле, то існує два найімовірніших числа, а саме: m_0 і m_0+1; в) якщо np – ціле, то найімовірніше число $m_0 = np$.

Приклад 3.3.1. Товарознавець оглядає 24 зразки товару. Ймовірність того, що взятий зразок буде визнано придатним до продажу, дорівнює 0,7. Знайти найімовірніше число зразків, які товарознавець визнає придатними до продажу.

Розв'язання. Маємо p = 0,7; q = 0,3; n = 24.Найімовірніше число настання події знаходимо за формулою

$np - g \leq m_0 \leq np + p$;

$24 \cdot 0{,}7 - 0{,}3 \leq m_0 \leq 24 \cdot 0{,}7 + 0{,}7$;

$16{,}5 \leq m_0 \leq 17{,}5$.

Отже, $m_0 =$ 17, як цілий розв'язок подвійної нерівності, тобто 17 зразків товару найімовірніше будуть придатними до продажу.

3.4. Формула Пуассона

Якщо ймовірність p близька до нуля ($p < 0{,}1$) при великих n, тобто, коли в кожному випробуванні подія, яка нас цікавить, з'являється рідко, то ймовірність того, що подія з'явиться в n випробуваннях рівно m раз, обчислюється за *формулою Пуассона*:

$$P_n(m) \approx \frac{\lambda^m \cdot e^{-\lambda}}{m!}, \qquad \lambda = np, \qquad e \approx 2{,}71828\ldots \qquad (3.4.1)$$

Набір чисел $P_n(m)$, де m= 0, 1, …, n; а $P_n(m)$ обчислюються за формулою Пуассона, називається *розподілом Пуассона*.

Приклад 3.4.1. Пристрій складається із 1000 елементів, які працюють незалежно. Ймовірність відмови у роботі будь-якого елемента протягом часу Т дорівнює 0,002. Знайти ймовірність того, що за час Т відмовлять рівно три елементи.

Розв'язання. За умовою n =1000, p =0,002; m =3. Так як n - велике число, а ймовірність $p \leq 0{,}1$ (тобто подія малоймовірна), то скористаємося формулою Пуассона. Знайдемо:

$\lambda = n \cdot p = 1000 \cdot 0{,}002 = 2$; тоді:

$$P_{1000}(3) \approx \frac{2^3 \cdot e^{-2}}{3!} = \frac{4}{3} \cdot 0{,}13534 = 0{,}18.$$

Формула Пуассона поряд із задачею повторення випробувань використовується також для розрахунку ймовірності появи різного числа подій (наприклад, точок або інших елементів) у якій-небудь області (площі, об’ємі або у часі). При цьому потрібно дотримуватися наступної умови: події (точки) в області розподілені в загальному рівномірно; положення кожної події (точки) випадкове, незалежне одна від одної; події (точки) з'являються в області поодинці, а не парами, трійками і т. д. Типовими прикладами випадкової величини, що має розподіл Пуассона, є: число викликів на АТС за деякий час t; число відновлень складної апаратури за час t, якщо відомо, що відновлення незалежні один від одного і в середньому на одиницю часу випадає λ відновлень, і т.п.

Приклад 3.4.2. У середньому на 1 м^2 площі посівів зустрічається 0,5 стебел бур'янів. Визначити ймовірність того, що на площі 4 м^2 виявиться два стебла бур'янів.

Розв’язання. Кожне стебло бур'янів розглядається як точка, що з'являється в заданій площі. Застосовуємо формулу Пуассона. За умовою λ_1 = 0,5 стебел/м^2, S = 4 м^2. Параметр розподілу Пуассона $\lambda=\lambda_1 s=0{,}5\cdot 4 = 2$. Шукану ймовірність знаходимо за формулою Пуассона:

$$p_2 = \frac{\lambda^2}{2!}e^{-\lambda} = \frac{2^2}{1\cdot 2}e^{-2} = \frac{2}{e^2} \approx 0.37 .$$

Послідовність подій, які з'являються у випадкові моменти часу, називається *течією подій.* Течія подій є пуассонівською, якщо виконуються наступні умови:

1) середнє число подій, які з'являються за одиницю часу, є сталим і дорівнює *а* (інтенсивність течії) ;

2) події з'являються поодинці, а не групами ;

3) відсутня післядія, тобто ймовірність появи події не залежить від появи або непояви події раніше та не впливає на найближче майбутнє.

Якщо течія подій є пуассонівською, то ймовірність появи події *m* разів за час t знаходимо за формулою

$$P_t(m) \approx \frac{(at)^m \cdot e^{-at}}{m!}. \quad (3.4.2)$$

Приклад 3.4.3. Середнє число викликів на АТС за одну хвилину дорівнює двом. Знайти ймовірність того, що за 5 хвилин поступить: а)два виклики; б) менше двох викликів; в) не менше двох викликів.

Розв'язання. Течія подій є пуассонівською, тому що виконуються всі умови згідно з означенням.

а) $P_t(m) \approx \frac{(at)^m \cdot e^{-at}}{m!}$, де *a*=2, *t*=5, $P_5(2) \approx \frac{(10)^2 e^{-10}}{2!} = 0,00225;$

б) $P_5(m<2) = P_5(0) + P_5(1) = e^{-10} + 10e^{-10} = 0,000495$ (ця подія практично неможлива).

в) $P_5(m \geq 2) = 1 - P_5(m<2) = 1 - 0,000495 = 0,999505$.

3.5. Локальна та інтегральна теореми Лапласа. Ймовірність відхилення відносної частоти появи події від ймовірності події

При великих *n* обчислення за формулою Бернуллі громіздкі, тому виникає необхідність у побудові формули, яка давала б можливість у таких випадках з достатньо великою точністю обчислювати ймовірності подій у повторних незалежних випробуваннях. Таку формулу вперше довів Муавр у 1732 р. для окремого випадку схеми Бернуллі при $p=\frac{1}{2}$, а пізніше, в 1801 р., цю теорему узагальнив Лаплас на випадок довільного фіксованого *p*, відмінного від нуля та одиниці.

Локальна теорема Муавра-Лапласа. Ймовірність того, що в *n* незалежних випробуваннях, в кожному з яких ймовірність появи події дорівнює *p* (*0*< *p*<1), подія настане рівно *m* раз (послідовність ролі не відіграє) наближено дорівнює (тим точніше, чим більше *n):*

$$P_n(m) \approx \frac{1}{\sqrt{npg}}\varphi(x), \tag{3.5.1}$$

де $\varphi(x) = \frac{1}{\sqrt{2\Pi}}\text{e}^{-x^2/2}$, $x = \frac{m-np}{\sqrt{npg}}$.

Таблиця значень функції *φ(x)* для додатних значень *x* наведена у додатку 1; для від'ємних значень *x* користуються цією ж таблицею, враховуючи парність функції: *φ(-x) =φ(x)*; крім того функція спадна і *φ(x*≥4)≈ 0.

Графік функції $\varphi(x) = \frac{1}{\sqrt{2\Pi}}\text{e}^{-x^2/2}$ зображено на рис. 3.5.1. – це крива, яку в літературі називають кривою Гаусса, або нормальною кривою.

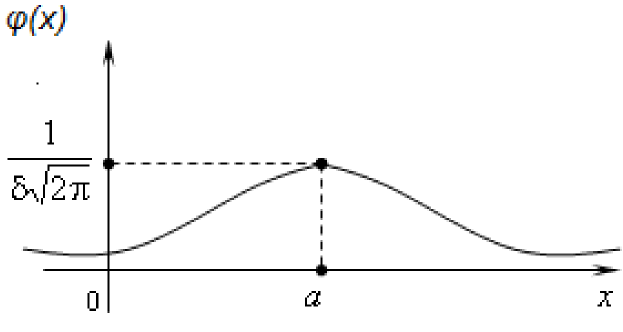


Рис. 3.5.1

Приклад 3.5.1. Ймовірність появи події в кожному з незалежних випробувань дорівнює 0,7. Знайти ймовірність того, що подія настане 1500 разів у 2100 випробуваннях.

Розв'язання. За умовою n=2100, m=1500, p=0,7; q=1-0,7=0,3. Так як n=2100 достатньо велике число, то скористаємося локальною теоремою Муавра-Лапласа. Знайдемо значення аргументу x:

$$x = \frac{m - np}{\sqrt{npg}} = \frac{1500 - 2100 \cdot 0,7}{\sqrt{2100 \cdot 0,7 \cdot 0,3}} = \frac{30}{21} \approx 1,43;$$

φ(1,43)=0,1435.

Шукана ймовірність$P_{2100}(1500) \approx \frac{1}{\sqrt{2100 \cdot 0,7 \cdot 0,3}} \varphi\,(1,43) \approx$

$\approx \frac{0,1434}{21} \approx 0,007$.

Багато задач теорії ймовірностей зводиться до обчислення сум виду $\sum_{m=m_1}^{m_2} b(m,n,p)$. Безпосереднє обчислення таких сум дуже громіздке, навіть коли можна застосувати локальну теорему Муавра-Лапласа. Крім того, при додаванні великої кількості наближених значень можуть нагромаджуватись великі похибки. Тому при розв'язуванні таких задач застосовують інтегральну теорему Лапласа.

Інтегральна теорема Муавра-Лапласа. Ймовірність того, що в n незалежних випробуваннях, в кожному із яких ймовірність появи події дорівнює p ($0<p<1$), подія поступить не менше m_1 раз і не більше m_2 раз, наближено дорівнює

$$P_{\mathrm{n}}(\mathrm{m}_1 \le \mathrm{m} \le \mathrm{m}_2) \approx \Phi(z_2) - \Phi(z_1). \qquad (3.5.2)$$

де $\Phi(x) = \frac{1}{\sqrt{2\Pi}} \int_0^x e^{\frac{-x^2}{2}} dx$; $z_1 = \frac{m_1 - np}{\sqrt{npg}}$; $z_2 = \frac{m_2 - np}{\sqrt{npg}}$.

Таблиця значень функції Лапласа для додатних значень x ($0 \le x \le 5$) наведена в додатку 2; $\Phi(x>5)$ = 0,5; функція непарна, тобто $\Phi(-x) = -\Phi(x)$. Функцію $\Phi(x)$ називають інтегральною функцією Лапласа. Графік інтегральної функції Лапласа Ф*(x)* зображений на рис. 3.5.2.

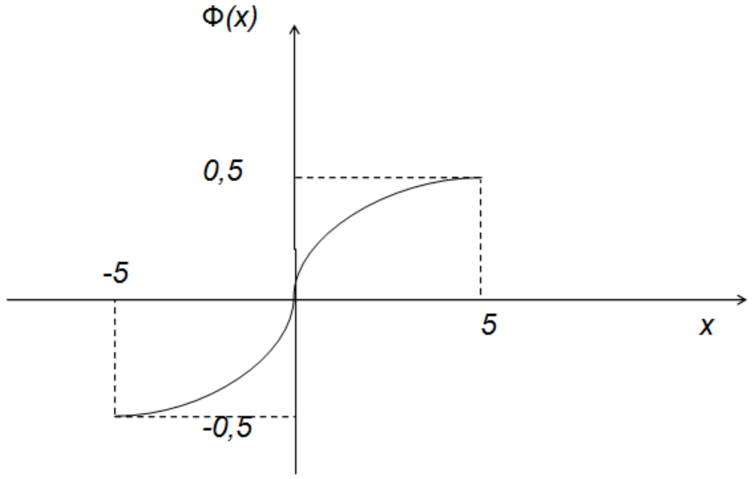


Рис. 3.5.2

Приклад 3.5.2. При встановленому технологічному процесі фабрика випускає в середньому 70% продукції першого сорту. Чому дорівнює ймовірність того, що в партії із 1000 виробів число виробів першого сорту не менше 652 і не більше 760?

Розв'язання*.* Число незалежних випробувань *n*=1000 і ймовірність появи події в окремому випробуванні *p*=0,7. Шукану ймовірність знайдемо за інтегральною теоремою Лапласа.

За умовою $m_1 = 652$ *і* $m_2 = 760$,отже ;

$$z_1 = \frac{652 - 1000 \cdot 0,7}{\sqrt{1000 \cdot 0,7 \cdot 0,3}} = -3,31$$

$$z_2 = \frac{760 - 1000 \cdot 0,7}{\sqrt{1000 \cdot 0,7 \cdot 0,3}} = 4,14;$$

$$z_1 = \frac{652 - 1000 \cdot 0,7}{\sqrt{1000 \cdot 0,7 \cdot 0,3}} = -3,31 P_{1000}\left(652 \le m \le 760\right) =$$

$$\Phi\left(4,14\right) - \Phi\left(-3,31\right) = \Phi\left(4,14\right) + \Phi(3,31) == 0,49997 + 0,49966 = 0,99963.$$

Оцінимо *ймовірність відхилення в*ідносної частоти появи події від сталої ймовірності в незалежних випробуваннях.

Нехай виконується *п* незалежних випробувань, в кожному з яких ймовірність події *А* стала та дорівнює *р.* Знайдемо ймовірність того, що відхилення $\left|\frac{m}{n}-p\right|$ відносної частоти від сталої ймовірності за абсолютною величиною не перевищує заданого числа $\varepsilon>0$. Тобто, знайдемо ймовірність виконання нерівності

$$\left|\frac{m}{n}-p\right|\le\varepsilon. \qquad (3.5.3)$$

Теорема. Ймовірність того, що в *n* незалежних випробуваннях, в кожному з яких ймовірність появи події дорівнює *p* ($0<p<1$), абсолютна величина відхилення відносної частоти появи події від ймовірності появи події в одному випробуванні не перевищить додатного числа ε, наближено дорівнює подвійній функції Лапласа:

$$P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)\approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right), \text{ при } x=\varepsilon\sqrt{\frac{n}{pq}}. \qquad (3.5.4)$$

Доведення. Замінимо нерівність (3.5.3) їй рівносильною $-\varepsilon\le\frac{m}{n}-p\le\varepsilon,\ -\varepsilon\le\frac{m-np}{n}\le\varepsilon.$

Після зведення до спільного знаменника помножимо нерівність на додатній множник $\sqrt{\frac{n}{pg}}$, отримаємо нерівність, яка рівносильна даній:

$$-\varepsilon\sqrt{\frac{n}{pq}}\le\frac{m-np}{\sqrt{npq}}\le\varepsilon\sqrt{\frac{n}{pq}}$$

Скористаємося інтегральною теоремою Лапласа, для цього позначимо через $x_1=-\varepsilon\sqrt{\frac{n}{pq}}$, $x_2=\varepsilon\sqrt{\frac{n}{pq}}$.

Маємо:

$$P(-\varepsilon\sqrt{\frac{n}{pq}} \le \frac{m-np}{\sqrt{npq}} \le \varepsilon\sqrt{\frac{n}{pq}}) = \Phi(\varepsilon\sqrt{\frac{n}{pq}}) - \Phi(-\varepsilon\sqrt{\frac{n}{pq}}) =$$

$$= \Phi(\varepsilon\sqrt{\frac{n}{pq}}) + \Phi(\varepsilon\sqrt{\frac{n}{pq}}) = 2\Phi(\varepsilon\sqrt{\frac{n}{pq}}).$$

Отже, ймовірність виконання нерівності $\left|\frac{m}{n} - p \le \varepsilon\right|$ приблизно дорівнює подвоєному значенню функції Лапласа:

$$P(\left|\frac{m}{n} - p\right| \le \varepsilon) \approx 2\Phi(\varepsilon\sqrt{\frac{n}{pq}}) \qquad (3.5.4)$$

Приклад 3.5.3. Посаджено 600 насінин кукурудзи. Ймовірність того, що зерно кукурудзи зійде, дорівнює 0,9. Знайти границю абсолютної величини відхилення відносної частоти зерен, які зійшли, від ймовірності *P*=0,9, якщо ця границя повинна бути гарантована з ймовірністю *P*=0,995.

Розв'язання. Якщо *n* - число незалежних випробувань і *p* - ймовірність появи події в окремому випробуванні, то при будь-якому *ε*>0 має місце наближена рівність $P(\left|\frac{m}{n} - p\right| \le \varepsilon) \approx 2\Phi(\varepsilon\sqrt{\frac{n}{pq}})$.

За умовою: *n* = 600, *p* = 0,9; *g* = 1-0,9 = 0,1; *P* = 0,995. Потрібно знайти ε, яке знайдемо із даної наближеної рівності $P(\left|\frac{m}{n} - 0{,}9\right| \le \varepsilon) \approx 2\Phi(\varepsilon\sqrt{\frac{600}{0{,}9\cdot 0{,}1}})$ або $2\Phi(81{,}65\varepsilon) = 0{,}995;\ \Phi(81.65\varepsilon) = 0.4975$.

Користуючись додатком 2, знаходимо $81{,}65\varepsilon = 2{,}81$, звідси *ε* = 0,034.

Правило-орієнтир для обчислення ймовірностей у повторних незалежних випробуваннях

1. Встановити, чи розглядуваний експеримент задовольняє схемі Бернуллі, тобто, необхідно перевірити, чи: а) випробування незалежні; б) кожне випробування має лише два результати; в) ймовірність появи події у кожному випробуванні стала і дорівнює *p* (а, отже, ймовірність непояви події буде *q=1-p*).

2. Ввести необхідні позначення і вибрати потрібну формулу для обчислення ймовірності. При виборі формули можна керуватись наступним:

3. а) Якщо число незалежних випробувань n мале, то для обчислення ймовірності появи події рівно m раз $P_n(m)$ використовується формула Бернуллі (3.1.1) – тут не виникає громіздких обчислень.

б) Якщо число незалежних випробувань n мале і потрібно знайти ймовірність появи події від m_1 до m_2 раз, тобто $P_n(m_1 \leq m \leq m_2)$, то використовується формула (3.2.1).

в) Якщо число незалежних випробувань n достатньо велике, а ймовірність появи події в кожному випробуванні мала, то для обчислення ймовірності появи події рівно m раз $P_n(m)$ використовується формула Пуассона (3.4.1).

г) Якщо число незалежних випробувань n достатньо велике, а ймовірність появи події в кожному випробуванні відмінна від 0 і 1, то для обчислення $P_n(m)$ використовується формула Лапласа (3.5.1).

д) Якщо число незалежних випробувань n достатньо велике, то для обчислення ймовірності появи події від m_1 до m_2 раз, тобто $P_n(m_1 \leq m \leq m_2)$, використовується формула з інтегральної теореми Лапласа (3.5.2).

Запитання для самоперевірки

1. Які випробування називаються повторними і незалежними?
2. Яка послідовність випробувань утворює схему Бернуллі?
3. Ймовірність якої події обчислюють за формулою Бернуллі?

4. Сформулюйте умови задач, для розв’язування яких при міняється формула Бернуллі.

5. За якими формулами знаходять ймовірність появи події *A* менше *m* або не менше *m* разів у *n* випробуваннях схеми Бернуллі?

6. Що називається біномним розподілом ймовірностей?

7. За якою формулою обчислюють найімовірніше число настання події у схемі Бернуллі ?

8. Яка асимптотична формула біноміального розподілу?

9. Що таке найімовірніше число появ події? Як воно визначається?

10. При яких умовах з формули Бернуллі виходить формула Пуассона?

11. Чому закон Пуассона називають законом рідкісних подій?

12. Коли доцільно використовувати формулу Пуассона?

13. Як визначаються та які мають властивості локальна та інтегральна функції Лапласа?

14. Вказати застосування локальної та інтегральної теорем Муавра-Лапласа.

15. Чому дорівнює ймовірність відхилення відносної частоти появи події від ймовірності події?

Задачі та вправи

3.1. У цеху є 6 моторів. Для кожного мотора ймовірність того, що він включений, дорівнює 0,7. Знайти ймовірність того, що в даний момент часу: а) включені 4 мотори; б) включені всі мотори; в) виключені всі мотори.

3.2. Ймовірність того, що телевізор потребує ремонту протягом гарантійного терміну, дорівнює 0,2. Знайти ймовірність того, що протягом гарантійного терміну із 6 телевізорів: а) не більше одного потребують ремонту; б) хоча б один потребує ремонту.

3.3. Два рівносильних противники грають у шахи. Що ймовірніше: а) виграти одну партію із двох чи дві партії із чотирьох? б) виграти не менше двох партій із чотирьох або не менше трьох партій із п'яти? (Нічия до уваги не береться).

3.4. Завод відправив на базу 500 виробів. Ймовірність пошкодження виробів у дорозі рівна 0,002. Знайти ймовірність того, що в дорозі буде пошкоджено менше трьох виробів.

3.5. Батарея зробила 6 пострілів по об’єкту. Ймовірність попадання в об’єкт при одному пострілі дорівнює 0,3. Знайти: а) найімовірніше число влучень; б) ймовірність найімовірнішого числа влучень; в) ймовірність того, що об’єкт буде зруйнований, якщо для цього достатньо хоча б двох влучень.

3.6. Ймовірність виграшу по лотереї дорівнює 1/3. Скласти біноміальний розподіл ймовірностей виграшу та побудувати полігон розподілу ймовірностей для п’яти лотерей.

3.7. Нехай ймовірність того, що покупцю необхідне взуття 41-го розміру, дорівнює 0,9. Знайти ймовірність того, що із 5 перших покупців взуття цього розміру буде необхідно: а) трьом покупцям; б) принаймні одному покупцю.

3.8. Скільки разів треба кинути гральний кубик, щоб найімовірніше число появи трійки дорівнювало 55?

3.9. Чому дорівнює ймовірність появи події в кожному випробуванні, якщо найімовірніше число появи події в 160 випробуваннях дорівнює 40?

3.10. Ймовірність того, що при 3-х випробуваннях подія настане не менше одного разу, дорівнює 0,992. Знайти ймовірність того, що в 5 випробуваннях ця подія настане 3 рази.

3.11. Здають 6 незалежних іспитів, у кожному з яких імовірність появи події А дорівнює 1/3. Знайти імовірність того, що подія А з'явиться: а) 2 рази; б) не менш одного разу.

3.12. Аудитор перевіряє 100 навмання вибраних рахунків. 5% рахунків торгових підприємств дають помилки. Якщо не буде жодної помилки, то рахунки підприємства далі не перевіряються. Яка ймовірність того, що в 100 рахунках: а) не буде жодної помилки; б) буде не більше трьох помилок?

3.13. Прилад складається з п'яти елементів, що працюють незалежно. Ймовірність відмови елемента в момент включення приладу дорівнює 0,2. Знайти: а) найімовірніше число елементів, які відмовили під час роботи; б) ймовірність найімовірнішого числа елементів, які відмовили під час роботи.

3.14. Ймовірність влучення в мішень одним пострілом дорівнює 0,6. Яке найімовірніше число влучень у мішень і ймовірність такої події, якщо проведено 10 пострілів?

3.15. Монету підкидають десять разів. Знайти ймовірність того, що герб з'явиться не менше чотирьох разів і не більше 7 разів. Яке найімовірніше число появи герба?

3.16. Робітник обслуговує 20 однотипних верстатів. Ймовірність того, що йому доведеться ремонтувати протягом години один з верстатів, дорівнює 0,2. Знайти: а) ймовірність того, що протягом однієї години робітнику потрібно буде ремонтувати чотири верстати; б) найімовірніше число верстатів, які потрібно буде ремонтувати протягом години.

3.17. Прилад вийде з ладу, якщо перегорять не менше п'яти елементів першого типу, або не менше двох елементів другого типу. Знайти ймовірність виходу приладу з ладу, якщо відомо, що перегоріло п'ять елементів, а ймовірність перегорання елементів першого та другого типів дорівнюють відповідно 0,7 і 0,3.

3.18. З партії виробів вибирають вироби вищого ґатунку. Ймовірність того, що навмання вибраний виріб - вищого ґатунку, дорівнює 0,8. Знайти ймовірність того, що з п'яти вибраних виробів будемо мати: а) два вироби вищого ґатунку; б) не менше двох виробів вищого ґатунку.

3.19. Ймовірність виготовлення стандартної деталі дорівнює 0,95. Скільки деталей необхідно виготовити, щоб найімовірніше число нестандартних деталей порівнювало 55?

3.20 Подія *B* може відбутися тільки в тому випадку, коли подія *A* відбудеться не менше трьох разів. Знайти ймовірність появи події *B*, якщо ймовірність події *A* в одному випробуванні дорівнює 0,3 і проведено: а) 5 незалежних випробувань; б) 7 незалежних випробувань.

3.21. Людина, що належить до певної групи населення, з ймовірністю 0,2 є брюнетом, з ймовірністю 0,3 - шатеном, з ймовірністю 0,4 - блондином і з ймовірністю 0,1 - рудим. Навмання вибирають групу з шести осіб. Знайти ймовірності таких подій: а) у складі групи не менше трьох блондинів; б) у складі групи хоча б один рудий; в) у складі групи однакове число блондинів і шатенів.

3.22. Завод виготовляє вироби, кожен з яких з імовірністю 0,01 має дефект. Скільки виробів з поверненням треба взяти на перевірку, щоб імовірність отримати хоча б один дефектний виріб була не менше ніж 0,95?

3.23. Ймовірність хоча б однієї появи події *A* у чотирьох незалежних випробуваннях дорівнює 0,59. Яка ймовірність появи події *A* в одному випробуванні, якщо в кожному випробуванні вона одна й та ж?

3.24. Ймовірність влучення у ціль одним пострілом дорівнює 0,4. Скільки треба зробити пострілів, щоб імовірність принаймні одного влучення була не меншою 0,9?

3.25. Скільки ізюму повинні містити в середньому здобні булочки для того, щоб ймовірність мати хоча б одну ізюминку в булці була не меншою за 0,99?

3.26. При обертанні антени радіолокатора за час опромінювання точкової цілі (наприклад, літака) встигають відобразитися 10 імпульсів. Знайти ймовірність знаходження цілі за одним оборотом радіолокатора, якщо для цього є необхідним проходження через приймальник не менше 8 імпульсів, а ймовірність придушення імпульсу перешкодою в приймальнику становить 0,1 і придушення різних імпульсів перешкодами є незалежними подіями?

3.27. Відомо, що 10% малих підприємств, які тільки починають свою діяльність, припинять роботу протягом першого року роботи. Яка ймовірність того, що з шести малих підприємств не більше двох підприємств припинять свою діяльність протягом року?

3.28. У сім'ї десятеро дітей. Вважаючи, що ймовірність народження хлопчика або дівчинки однакова, знайти ймовірність того, що у цій сім'ї: а) не менше трьох хлопчиків; б) не більше трьох хлопчиків.

3.29. На факультеті 800 студентів. Ймовірність народження кожного студента в даний день становить $\frac{1}{365}$. Знайти найімовірніше число студентів, які народилися 1 січня, та ймовірність того, що знайдуться 5 студентів з одним і тим же днем народження.

3.30. Мішень має центральне коло та два кільця. Ймовірність влучення у центральне коло під час одного пострілу дорівнює 0,4; у перше кільце - 0,3 і в друге кільце - 0,2. Ймовірність влучення у мішень дорівнює 0,1. У мішень зроблено п'ять пострілів. Знайти ймовірність того, що в результаті пострілів будуть два влучення в центральне коло та одне - у друге кільце.

3.31. Ймовірність народження хлопчика приймемо рівною 0,5. Знайти ймовірність того, що серед 200 новонароджених дітей будуть: а) 100 хлопчиків; б) 90 хлопчиків; в) від 90 до 110 хлопчиків.

3.32. Ймовірність того що покупцю погрібне взуття 43-го розміру, дорівнює 0,2. Знайти ймовірність того, що серед 100 покупців забажають купити взуття 43-го розміру: а) 25 покупців; б) від 10 до 30 покупців; в) не більше 30 покупців; г) не менше 35 покупців.

3.33. Сто верстатів працюють незалежно один від одного, причому ймовірність безперебійної роботи кожного з них протягом робочого дня дорівнює 0,8. Знайти ймовірність того, що протягом робочого дня безперебійно працюватимуть: а) 85 верстатів; б) від 75 до 85 верстатів.

3.34. Ймовірність помилки на сторінці книги дорівнює 0,002. Перевіряється книга, що має 500 сторінок. Знайти ймовірність того, що з помилками виявляться: а) 5 сторінок; б) від 3 до 5 сторінок.

3.35. Ймовірність появи події A в одному випробуванні дорівнює 0,4. Знайти ймовірність того, що для 100 проведених випробувань частота появи цієї події відхилиться від ймовірності p= 0,4 не більше, ніж на 0,5.

3.36. Проводиться 600 підкидань гральної кістки. Потрібно оцінити ймовірність того, що частота випадань шістки відхилиться від ймовірності випадання шістки при одному підкиданні менше, ніж: а) на 0,01; б) на 0,02.

3.37. Радіоприймач складається з 1000 мікроелементів. Ймовірність відмови кожного з елементів протягом доби дорівнює

0,001 і не залежить від стану інших елементів. Знайти ймовірність відмови: а) 2 елементів; б) не менше 2 елементів за добу.

3.38. Схожість насіння оцінюється ймовірністю 0,85. Знайти ймовірність того, що з 500 зерен посіяного насіння зійдуть: а) 425; б) 400; в) від 425 до 450.

3.39. Магазин отримав 1000 пляшок мінеральної води. Ймовірність того, що внаслідок перевезення пляшка розіб'ється, дорівнює 0,003. Знайти ймовірність того, що магазин отримає: а) 2 розбиті пляшки; б) менше двох розбитих пляшок; в) більше двох розбитих пляшок.

3.40. За деякий час конденсатор виходить з ладу з імовірністю 0,2. Знайти ймовірність того, за цей час зі 100 незалежно працюючих конденсаторів вийдуть з ладу: а) не менше 20; б) менше 28; в) від 14 до 26 конденсаторів.

3.41. Ймовірність появи події A у кожному незалежному випробуванні дорівнює 0,9. Скільки потрібно провести випробувань, щоб з ймовірністю 0,98 можна було очікувати, що не менше 150 випробувань дадуть позитивний результат?

3.42. Шульгами є у середньому 1% студентів. Яка ймовірність того, що серед 200-т студентів знайдуться: а) рівно 4 шульги; б) не менше ніж 4 шульги.

3.43. Завод відправив на базу 5000 доброякісних виробів. Імовірність пошкодження кожного виробу в дорозі дорівнює 0,0002. Знайти ймовірність того, що з 5000 виробів буде пошкоджено у дорозі: а) рівно 3 вироби; б) не більше 3 виробів; г) більше 3 виробів.

3.44. Ймовірність виготовлення бракованої деталі дорівнює 0,008. Знайти ймовірність найімовірнішого числа бракованих деталей серед вибраних навмання 100 деталей.

3.45. За статистичними даними 87% народжених доживають до 50 років. Знайти ймовірність того, що з 1000 новонароджених частка тих, хто дожив до 50 років, буде: а) знаходиться в межах від 0,9 до 0,95; б) відрізнятися від ймовірності цієї події не більше, ніж на 0,04 (за абсолютною величиною).

3.46. За результатами перевірок податковими інспекціями встановлено, що в середньому кожне друге мале підприємство даного району має порушення фінансової дисципліни. Знайти ймовірність того, що з 1000 зареєстрованих малих підприємств у районі порушують фінансову дисципліну: а) 400 підприємств; б) не менше 480; в) від 480 до 520.

3.47. У банк відправлено 40000 пакетів грошових знаків. Імовірність того, що пакет містить недостатнє або надлишкове число грошових знаків, дорівнює 0,0001. Знайти ймовірність того, що перевірка пакетів виявить: а) три помилково укомплектовані пакети; б) не більше трьох пакетів.

3.48. Будівельна компанія, яка виготовляє та встановлює літні будинки, розсилає рекламні листи у поштові скриньки. Досвід роботи компанії показав, що приблизно в одному випадку з двох тисяч поступає замовлення. Знайти ймовірність того, що розіславши 100 тисяч листівок, компанія отримає: а) рівно 48 замовлень; б) від 45 до 55 замовлень.

3.49. Відомо, що серед виготовлених заводом телефонних апаратів є 60% апаратів першого сорту. Чому дорівнює ймовірність того, що серед 200 виготовлених заводом апаратів виявиться 120 апаратів першого сорту?

3.50. Аудиторну роботу з теорії ймовірностей за першим разом успішно виконують 50% студентів. Знайти ймовірність того, що з 400 студентів роботу успішно виконають: а) 180 студентів; б) не менше 180 студентів.

3.51. Ймовірність того, що оператор набере текст з помилками, дорівнює 0,1. Знайти ймовірність того, що з 200 сторінок тексту буде набрано без помилок не менше 180.

3.52. Аналіз статутних фондів банків показав, що п'ята частина банків мають статутний фонд більше 100 млн. грн. Знайти ймовірність того, що серед 1800 банків мають статутний фонд більше 100 млн. грн.: а) не менше 300; б) від 300 до 400.

3.53. Ймовірність того, що пасажир запізниться на поїзд, дорівнює 0,01. Знайти найімовірніше число пасажирів, які запізняться, та ймовірність такої події, якщо білети взяли 800 пасажирів.

3.54. Ймовірність того, що деталь стандартна, дорівнює 0,9. Знайти: а) з імовірністю 0,9545 межі числа стандартних деталей серед перевірених 900 деталей; б) ймовірність того, що частка стандартних деталей серед них знаходиться в межах від 0,8 до 0,11.

3.55. У результаті перевірки якості підготовленого до сівби насіння пшениці встановлено, що в середньому 90% дають добру схожість. Скільки потрібно посіяти насіння, щоб з ймовірністю 0,991 можна було сподіватись, що частка насіння, яка зійшла, відрізнятиметься від ймовірності сходів кожної насінини не більше, ніж на 0,03?

3.56. Ймовірність того, що дилер, який торгує цінними паперами, продасть їх, дорівнює 0,7. Скільки повинно бути цінних паперів, щоб можна було стверджувати з ймовірністю 0,996, що частка проданих серед цих паперів відрізнятиметься від 0,7 не більше, ніж на 0,04?

3.57. Страхова компанія має 10 000 клієнтів. Кожний клієнт вносить 500 грн. на можливість нещасного випадку. Ймовірність нещасного випадку дорівнює 0,0055; а страхова сума, яку потрібно виплатити постраждалому клієнту, складає 50 000 грн. Яка ймовірність того, що: а) страхова компанія зазнає збитків; б) для виплати страхових сум потрібно більше половини всіх коштів, які поступили від клієнтів?

3.58. Ймовірність влучення у ціль при кожному пострілі дорівнює 0,001. Знайти ймовірність двох і більше влучень, якщо було зроблено 5000 пострілів.

3.59. Ймовірність виходу з ладу за час t одного приладу дорівнює 0,1. Знайти ймовірність того, що за час t зі 100 приладів вийдуть з ладу: а) не менше 20; б) менше 15; в) від 6 до 18 приладів.

3.60. В урні три кулі білі, чорна і червона. З урни виймають кулі по одній п'ять раз, причому, після кожного виймання кулю повертають в урну. Знайти ймовірність того, що чорну і білу кулю візьмуть не менше, ніж по два рази кожну.

3.61. В електропоїзд, що має шість вагонів, сідають дванадцять пасажирів, причому вибір вагону кожним пасажиром рівноможливий. Знайти ймовірність того, що: а) у кожний вагон ввійшли по два пасажири; б) в один вагон ніхто не ввійшов, у другий ввійшов один пасажир, у два вагони - по два пасажири і в останні два вагони ввійшли відповідно три та чотири.

3.62. Лінія зв’язку, що має 200 каналів, зв’язує пункт А з пунктом В, де налічується 2 000 абонентів, кожний із яких користується телефоном у середньому 6 хв. на годину. Знайти ймовірність безвідмовного обслуговування абонентів.

3.63. У водоймище випущено 100 мічених риб. Незабаром після цього з водоймища було виловлено 400 риб, серед яких виявилося 5 мічених. Визначити загальне число риб у цьому водоймищі з імовірністю а) 0,9; б) 0,6.

3.64. Ймовірність зупинки верстата протягом зміни дорівнює 0,1. Знайти ймовірність того, що серед 1000 верстатів працює: а) від 700 до 740 верстатів; б) не менше 700 верстатів.

3.65. Ймовірність появи події в кожному з 2100 незалежних випробувань дорівнює 0,7. Знайти ймовірність того, що подія з’явиться: а) не менше 1470 і не більше 1500 разів; б) не більше 1469 разів.

3.66. Ймовірність виготовлення бракованої деталі для деякого виробу дорівнює 0,015. Знайти ймовірність того, що доля (відносна частота) бракованих деталей серед 1000 виготовлених буде відрізнятись від імовірності виготовлення бракованої деталі не більше, ніж на 0,005 у той або інший бік. Те ж - для 500 штук. Порівняти і вказати, як впливає кількість на шукану ймовірність.

3.67. Ймовірність того, що механізм потребує регулювання, дорівнює 0,4. Навмання вибираються 500 механізмів. Знайти величину найбільшого відхилення відносної частоти механізмів, які потребують регулювання, від ймовірності 0,4, яку можна гарантувати з ймовірністю 0,9973.

3.68. Скільки зерен кукурудзи треба посіяти, що відносна частота зерен, що зійшли, з ймовірністю 0,99 відрізнялася від ймовірності проростання окремої зернини 0,02 за абсолютною величиною менше, ніж на 0,01?

3.69. Ймовірність попадання в мішень стрілком при одному пострілі дорівнює 0,75.Обчислити ймовірність того, що при 100 пострілах стрілок влучить в мішень: а) не менше 70 раз і не більше 80 раз; б) не менше 70 раз.

3.70. Гральну кістку кидають 600 разів. Яка ймовірність того, що число очок кратне трьом, випаде не менше 218 і не більше 256 разів?

3.71. Скільки разів треба кинути монету, щоб із ймовірністю 0,6 треба було очікувати, що відхилення відносної частоти від імовірності появи герба буде за абсолютною величиною не більше 0,01?

3.71. Ймовірність появи події в кожному із 10 000 незалежних випробувань дорівнює 0,75. Знайти ймовірність того, що відносна частота появи події відхилиться від його ймовірності за абсолютною величиною не більше, як на 0,01.

3.72. Ймовірність появи події в кожному із 900 незалежних випробувань дорівнює 0,7. Знайти таке додатне число ε, що з імовірністю 0,98 абсолютна величина відхилення відносної частоти появи події від його ймовірності не перевищує ε.

3.73. У деякій місцевості є 3% хворих на серцеву недостатність. Проводиться обстеження 500 осіб. З якою ймовірністю серед обстежених виявиться 3±0,5% хворих на серцеву недостатність?

3.74. Набір цукерок «Асорті» містить порівну цукерки чотирьох найменувань: «а», «б», «в», «г». Велика кількість наборів розфасовується в коробки по 20 цукерок у кожну. Яка ймовірність того, що із 25 коробок у 9 будуть по одній цукерці сорту «а»?

РОЗДІЛ 4
ВИПАДКОВІ ВЕЛИЧИНИ

При дослідженні багатьох проблем виникають такі випадкові події, наслідком яких є поява деякого числа, заздалегідь невідомого. Тому такі числові значення – випадкові. Наприклад: при киданні грального кубика могли з'явитися числа 1; 2; 3; 4; 5; 6. Наперед визначити число очок на верхній грані кубика неможливо, тому що воно залежить від багатьох випадкових причин, які повністю не можуть бути враховані. Отже, число очок є величина випадкова, а числа 1; 2; 3; 4; 5; 6 є можливі значення цієї величини.

Випадковою величиною називають таку величину, яка внаслідок випробування може прийняти лише одне числове значення, заздалегідь невідоме та обумовлене випадковими причинами. Випадкові величини бувають дискретними та неперервними.

4.1. Дискретна випадкова величина. Закони розподілу ймовірностей дискретної випадкової величини

Випадкова величина називається *дискретною,* якщо вона приймає значення деякої числової послідовності (скінченої або нескінченої). Випадкові величини позначають прописними великими літерами *X, Y, Z, …,* а їхні можливі значення *x, y, z, …* Якщо випадкова величина *X* має три можливих значення, то вони будуть позначені $x_1; x_2; x_3$.

Наведемо приклад дискретної випадкової величини: кількість зерен у колосі пшениці, кількість студентів, які прийдуть на лекцію; кількість яблук на дереві. У цих випадках випадкова величина приймає окремі, ізольовані одне від одного можливі значення.

На перший погляд можна вважати, що для задання дискретної випадкової величини достатньо перерахувати всі її можливі значення. Насправді ж це не так: випадкові величини можуть мати однаковий перелік можливих значень, а їхні ймовірності - різні. Тому для задання дискретної випадкової величини недостатньо перерахувати всі можливі її значення, треба ще й вказати їхні ймовірності.

Законом розподілу дискретної випадкової величини називають відповідність між можливими значеннями випадкової величини та відповідними їм ймовірностями. Його можна задати за допомогою таблиці, аналітично (у вигляді формули) і графічно.

При табличному заданні закону розподілу дискретної випадкової величини перший рядок таблиці містить можливі її значення, а другий - їх імовірності.

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

Беручи до уваги, що в одному випробуванні випадкова величина приймає одне і тільки одне можливе значення, заключаємо, що події $X = x_1; X = x_2; \dots; X = x_n$ утворюють повну групу подій; отже сума ймовірностей другого рядка дорівнює 1:

$p_1 + p_2 + \dots + p_n = 1$ або коротше: $\sum_{i=1}^{n} p_i = 1$.

Таблична форма задання закону дискретної випадкової величини називається ще *рядом розподілу*.

Для наочності ряд розподілу представляють графічно – у вигляді *полігона* (многокутника розподілу). Щоб побудувати полігон, потрібно у прямокутній системі координат побудувати точки з координатами $(x_i\,; p_i)$, де x_i - можливі значення випадкової величини X, p_i - відповідні їм імовірності, і послідовно сполучити їх відрізками прямих.

Приклад 4.1.1. Задано закон розподілу дискретної випадкової величини:

x_i	-1	6	13	20	27
p_i	0,2	0,1	0,4	0,2	0,1

Накреслити многокутник розподілу заданої випадкової величини.

Розв'язання. По осі абсцис відкладаємо в обраному масштабі значення випадкової величини, по осі ординат — відповідні ймовірності. Кінці ординат, що виражають імовірності можливих значень випадкової величини, з'єднуємо відрізками прямих ліній. Отримуємо многокутник розподілу заданої випадкової величини (рис. 4.1.1).

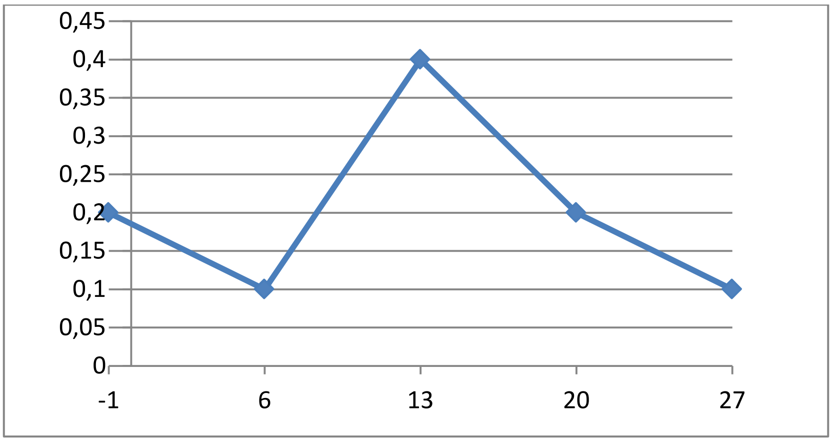


Рис. 4.1.1

Многокутник розподілу, як і ряд розподілу, є однією із форм задання закону розподілу дискретної випадкової величини. Многокутники розподілу можуть мати різноманітну форму, але всі вони мають одну спільну властивість: сума ординат вершин многокутника розподілу, що являє собою суму ймовірностей всіх можливих значень випадкової величини, завжди дорівнює одиниці.

Якщо ймовірність у законі розподілу дискретної випадкової величини *X* – числа появи події у *n* незалежних повторних випробуваннях, обчислюється за формулою Бернуллі (3.1.1), то відповідний закон розподілу називається *біноміальним* (п. 3.2).

Якщо число випробувань велике, а ймовірність *p* появи події в кожному випробуванні дуже мала, то використовують наближену формулу Пуассона (3.4.1), і говорять, що випадкова величина розподілена за законом Пуассона (п. 3.4).

У різноманітних задачах статистичного контролю якості виробів, у теорії надійності, страхових розрахунках та ін. застосовують *геометричний розподіл*, який має вигляд:

$$P(X = m) = pq^{m-1}, \tag{4.1.1}$$

де $p = P(A)$ – ймовірність появи події А в кожному випробуванні, $q = 1 - p$, *X* – кількість випробувань до появи події *A* в серії незалежних повторних випробувань. Ряд імовірностей цього розподілу буде нескінченно спадною геометричною прогресією зі знаменником q, сума якої дорівнює одиниці.

Існують також інші види розподілів, які застосовуються при розв'язуванні задач економіки: гіпергеометричний, поліноміальний та ін.

Характеристикою випадкової величини служать імовірності появи різних її значень. Для їх задання використовують *функцію розподілу ймовірностей* дискретної випадкової величини

$$F(x) = P(X < x), \tag{4.1.2}$$

де $P(X < x)$ - ймовірність виконання умови $X < x$, що розглядається як функція змінної *x* (аналітичний спосіб задання розподілу дискретної випадкової величини). Тут функція розподілу задається формулою

$$F(x) = \sum_{x_i<x} p_i, \qquad (4.1.3)$$

де символ $x_i < x$ під знаком суми означає, що сума поширюється на всі ті можливі значення випадкової величини, які за своєю величиною менші від аргументу х. З означення функції розподілу випливає, що вона існує для всіх випадкових величин: як дискретних, так і неперервних. Для дискретної випадкової величини *X,* яка набирає значення $x_1, x_2, \dots, x_n$, функція розподілу буде мати вигляд:

$$F(x) = \sum_{x_i<x} P(X = x_i). \qquad (4.1.4)$$

З виразу (4.1.4) випливає, що функція розподілу дискретної випадкової величини *X* розривна і зростає стрибками при переході через точки можливих її значень $x_1; x_2; \dots; x_n$. Причому величина стрибка рівна ймовірності відповідного значення (рис. 4.1.2).

Приклад 4.1.2. Задано закон розподілу дискретної випадкової величини:

X	0	1	2	3
P	0,216	0,432	0,288	0,064

Знайти функцію розподілу *F(x)* і накреслити її графік.

Розв'язання*.* Якщо $x \le 0$, то *F(x)*=0. Дійсно, значень, які менші за число 0, величина *X* не приймає.

Якщо $0 \le x \le 1$, то *F(x)*=0,216. Дійсно, *X* може прийняти значення 1 з імовірністю 0,216.

Якщо $1 < x \le 2$, то *F(x)*=0,648. Дійсно, *X* може прийняти значення 0 з імовірністю 0,216 і значення 1 з імовірністю 0,432. Отже, одне з цих значень, байдуже яке, *X* може прийняти (за теоремою додавання несумісних подій), з імовірністю 0,216+0,432=0,648.

Аналогічно, якщо $2 < x \leq 3$, то $F(x)$=0,216+0,432+0,288=0,936.

Якщо $x > 3$,$F(x)$=1. Дійсно, подія $x \leq 3$ - достовірна і ймовірність її дорівнює 1.

Отже, шукана функція розподілу матиме вигляд:

$$F(x) = \begin{cases} \sum\limits_{x_i<0} P(X = x_i) = 0 \ \text{при}\, x < 0, \\ \sum\limits_{x<1} P(X = x_i) = P(X = 0) = 0{,}216 \ \ \text{при}\, 0 < x \leq 1, \\ \sum\limits_{ix<2} P(X = x_i) = P(X = 0) + P(X = 1) = 0{,}648 \ \ \text{при}\, 1 < x \leq 2, \\ \sum\limits_{x<3} P(X = x_i) = P(X = 0) + P(X = 1) + P(X = 2) = 0{,}936 \ \ \text{при} \ 2 < x \leq 3, \\ P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 1 \ \ \text{при} \ x > 3. \end{cases}$$

Графік функції $F(x)$ зображено на рис. 4.1.2:

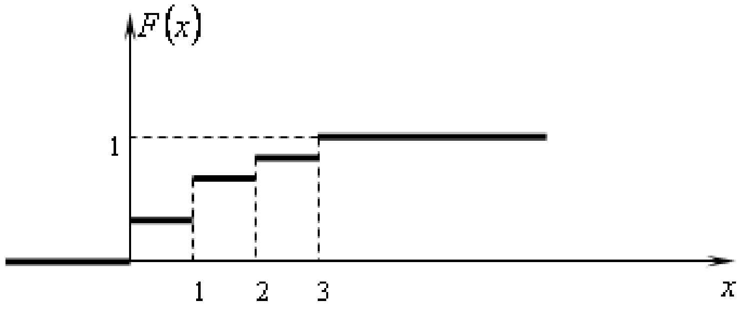


Рис. 4.1.2

Закони розподілу дискретної випадкової величини повністю характеризують випадкові величини та дозволяють розв'язувати всі пов'язані з ними задачі.

Правило-орієнтир для складання таблиці розподілу ймовірностей дискретної випадкової величини

1. Установити, що є випадковою величиною в розглядуваній задачі.
2. Перерахувати всі можливі значення випадкової величини.
3. З умови задачі встановити закон розподілу ймовірностей випадкової величини.

4. Використовуючи відповідну формулу, знайти ймовірності появи можливих значень випадкової величини.

5. Скласти таблицю розподілу ймовірностей випадкової величини і перевірити, що сума ймовірностей розподілу дорівнює одиниці: $\sum_{i=1}^{n} p_i = 1$.

4.2. Числові характеристики дискретної випадкової величини та їх властивості

У практичній діяльності не завжди вдається отримати закон розподілу, або закон надто складний для практичних розрахунків. Тому з'явилася потреба характеризувати дискретну випадкову величину за допомогою числових характеристик, які характеризують особливості випадкових величин .

Найчастіше використовують такі числові характеристики: математичне сподівання, дисперсію та середнє квадратичне відхилення.

Математичним сподіванням дискретної випадкової величини *X* називається сума добутків всіх можливих значень випадкової величини на ймовірності цих значень:

$$M(X) = x_1p_1 + x_2p_2 + \dots + x_np_n = \sum_{i=1}^{n} x_ip_i. \quad (4.2.1)$$

Математичне сподівання має такі *властивості:*

1) M(C)=C (де C - стала).

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

Якщо X і У - незалежні випадкові величини, то:

3) M(X+У)=M(X)+M(У), зокрема) M(X+C)=M(X)+C.

4) M(X·У)=M(X)·M(У) зокрема M(CX)=CM(X).

Математичне сподівання також називають своєрідним центром розподілу випадкової величини, навколо якого групуються її можливі значення, або середнім значення випадкової величини, або показником центральної тенденції. У фізиці математичному сподіванню відповідає поняття центру маси.

Але цієї характеристики недостатньо, оскільки часто має місце значне відхилення можливих значень від центру розподілу. Тому необхідно уміти виміряти *змінюваність* випадкової величини відносно її математичного сподівання. Для цього використовують дисперсію.

Дисперсією випадкової величини називається математичне сподівання квадрата різниці випадкової величини та її математичного сподівання.

$$D(X) = M(x - M(x))^2 \qquad (4.2.2)$$

Теорема. Дисперсія випадкової величини дорівнює різниці між математичним сподіванням квадрата цієї величини і квадратом її математичного сподівання

$$D(X) = M(x^2) - (M(x))^2. \qquad (4.2.3)$$

Властивості дисперсії*:*

1*. D(C) = 0.*

2. D(CX) = C^2 *D(X),* зокрема *D(-X) = D(-1X) = (-1)*^{2} *D(X)= D(X),*

3. D(X+У) = D(X)+D(У);

4. D(X-У) = D(X)+D(У).

На практиці зазвичай випадкова величина *X* має розмірність (метр, грам тощо). Розмірність дисперсії дорівнює квадрату розмірності випадкової величини, тому її не можна інтерпретувати геометрично. Цей недолік усуває середнє квадратичне відхилення випадкової величини.

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

$$\sigma(X) = \sqrt{D(X)}. \qquad (4.2.4)$$

Середнє квадратичне відхилення називають ще стандартним відхиленням, воно є мірою того, наскільки широко розкидані дані відносно їхнього середнього.

Дисперсія та середнє квадратичне відхилення є мірою *розсіювання* випадкової величини навколо математичного сподівання.

Математичне сподівання та дисперсія числа появи події в *n незалежних випробуваннях* дорівнюють

$$M(X) = n \cdot p, \quad D(X) = n \cdot p \cdot q. \tag{4.2.5}$$

Розподіл Пуассона залежить від одного параметра λ, який є математичним сподіванням випадкової величини X:

$$M(X) = \sum_{m=0}^{\infty} m \frac{\lambda^m \cdot e^{-\lambda}}{m\,!} = \lambda e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^{m-1}}{(m-1)\,!} = \lambda e^{-\lambda} e^{\lambda} = \lambda\,.$$

Дисперсія буде:

$$D(X) = M(X^2) - M^2(X) = \sum_{m=0}^{\infty} m^2 \frac{\lambda^m \cdot e^{-\lambda}}{m\,!} - \lambda^2 = \lambda e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^{m-1}}{(m-1)\,!} - \lambda^2 =$$

$$= \lambda e^{-\lambda} \sum_{m=0}^{\infty} \frac{(m-1)+1}{(m-1)\,!} \lambda^{m-1} - \lambda^2 = \lambda e^{-\lambda} (\lambda \sum_{m=0}^{\infty} \frac{\lambda^{m-2}}{(m-2)\,!} + \sum_{m=0}^{\infty} \frac{\lambda^{m-1}}{(m-1)\,!}) - \lambda^2 =$$

$$= \lambda e^{-\lambda} (\lambda e^{\lambda} + e^{\lambda}) - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda,$$

Отже, для пуассонівського розподілу:

$$M(X) = \lambda; \; D(X) = \lambda; \; \sigma(X) = \sqrt{\lambda}\,. \tag{4.2.6}$$

Приклад 4.2.1. Знайти математичне сподівання, дисперсію та середнє квадратичне відхилення випадкової величини, знаючи закон її розподілу.

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

Розв’язання.1) $M(x) = x_1p_1+x_2p_2+\ldots+x_np_n$.

M(X) = 3·0,1+5·0,6+2·0,3 = 3,9.

2) Дисперсію знайдемо двома способами:

а) $D(X) = M(X-M(X))^2$,

$D(X)=(3-3,9)^2 \cdot 0,1+(5-3,9)^2 \cdot 0,6+(2-3,9)^2 \; 0,3 =$
$=0,81 \cdot 0,1+1,21 \cdot 0,6+3,61 \cdot 0,3 = 1,89$;

б) $D(X) = M(X^2) - (M(X))^2$,

$D(X) = 3^2 \cdot 0,1+5^2 \cdot 0,6+2^2 \cdot 0,3-(3,9)^2 = 9 \cdot 0,1+25 \cdot 0,6+4 \cdot 0,3-15,21 =$
$=0,9+15+1,2-15,21 = 1,89$.

3) $\sigma (X) = \sqrt{Д(X)} = \sqrt{1,89} = 1,375$.

Приклад 4.2.2. Знайти математичне сподівання випадкової величини $Z = X+2У$,якщо відомі математичні сподівання $M(X) = 5$, $M(У) = 6$.

Розв'язання. Скористаємось властивостями числових характеристик: $M(Z) = M(X+2У) = M(X) + M(2У) = M(X) + 2M(У) =$ $=5+2 \cdot 6 = 17$.

Приклад 4.2.3. Знайти математичне сподівання та дисперсію дискретної випадкової величини *X* - числа відмови елемента деякого прикладу в 10 незалежних випробуваннях, якщо ймовірність відмови елемента в кожному випробуванні 0,9.

Розв'язання. $n = 10$, $p = 0,9$; $q = 1-0,9 = 0,1$.

$M(X) = np$; $M(X) =10 \cdot 0,9 = 9$; $D(X) = npq= 10 \cdot 0,9 \cdot 0,1= 0,9$.

4.3. Неперервна випадкова величина. Інтегральна та диференціальна функції розподілу випадкових величин

Нагадаємо, що *дискретна* випадкова величина задається переліком усіх її можливих значень та відповідних їм імовірностей. Такий спосіб задання не є загальним, і його не можна використовувати для інших видів випадкових величин.

Розглянемо випадкову величину *X,* можливі значення якої заповнюють повністю інтервал (*a; b)*. Таку випадкову величину *X* називають *неперервною*. Чи можливо скласти перелік усіх можливих значень *X*? Звичайно, цього зробити не можна. Цей факт вказує на доцільність дати загальний спосіб задання будь-яких типів випадкових величин.

Для задання любого типу випадкової величини вводять інтегральну функцію розподілу.

Нехай *x* – дійсне число. Ймовірність події, суть якої полягає в тому, що випадкова величина *X* приймає значення менше за *x*, тобто ймовірність події $X < x$, позначимо через $F(x)$. Якщо *x* буде змінюватися, то буде змінюватися і $F(x)$.

Інтегральною функцією розподілу (функцією розподілу) називають функцію $F(x)$, яка визначає для кожного значення *x* ймовірність того, що випадкова величина *X* прийме значення, яке менше *x,* тобто

$$F(x) = P(X < x). \tag{4.3.1}$$

Геометрично цю рівність можна пояснити так: $F(x)$ є ймовірність того, що випадкова величина прийме значення, яке зображується на числовій осі точкою, що лежить лівіше від точки *x*.

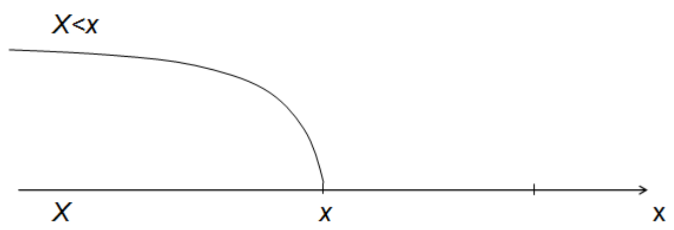


Рис. 4.3.1

Для дискретної випадкової величини, яка приймає значення x_1, x_2, …, x_n, виберемо деяке значення *x*, при якому виконується подвійна нерівність $x_k < x \le x_{k+1}$. Тоді лівіше числа *x* на числовій осі будуть тільки ті значення випадкової величини, які мають індекс 1, 2, 3, …, *k*. Причому нерівність $x<X$ виконується, якщо величина *X* прийме значення x_i, де i = 1, 2, …, *k*. Таким чином, подія $X<x$

настане, якщо настане будь-яка з подій $X=x_1, X=x_2, \ldots, X=x_k$. Так як ці події несумісні, то за теоремою додавання ймовірностей маємо $F(X) = P(X= x_1)+P(X= x_2)+\ldots+P(X= x_k)= \sum x_k < xP_i$.

Якщо будемо збільшувати *x*, тобто переміщувати точку *x* вправо по осі, то очевидно, що при цьому ймовірність того, що випадкова величина *X* попаде лівіше *x*, не може зменшуватися. Тому функція *F(x)* зі зростанням *x* спадати не може. Необмежено переміщуючи точку *x* вправо, переконуємося, що $F(\infty)=1$, так як подія $X < \infty$ стає достовірною.

Якщо *x* переміщувати необмежено вліво по числовій осі, то попадання випадкової величини *X* лівіше *x* стає неможливою подією, тому $F(-\infty)=0$.

Властивості інтегральної функції:

1. Значення інтегральної функції належить відрізку [0;1], отже $0 \le F(x) \le 1$.

Це твердження слідує з того, що ймовірність любої випадкової події є невід'ємне число, яке не перевищує 1.

2. $F(x)$ - неспадна функція, тобто $F(x_2) \ge F(x_1)$, якщо $x_2 > x_1$.

Нехай $x_2 > x_1$. Подію $(X< x_2)$ можна подати у вигляді суми двох несумісних подій $(X< x_1)$ та $(x_1 \le X< x_2)$, тобто $(X< x_2)=(X< x_1) \cup \cup (x_1 \le X<x_2)$. Тоді, за формулою додавання ймовірностей несумісних подій, маємо: $P(X< x_2)=P(X< x_1)+\mathrm{P}(x_1 \le X< x_2)$, звідси, за означенням інтегральної функції, отримаємо $F(x_2) = F(x_1) + P(x_1 \le X \le x_2)$.

3. Якщо можливі значення випадкової величини належать інтервалу *(a;b)* то:

а) $F(x) = 0$ при $x \le a$; б) $F(x) = 1$ при $x > b$.

Графік функції *F(x)* у загальному випадку є графіком неспадної функції, значення якої змінюються від 0 до 1, причому в окремих точках функція може мати розриви (рис. 4.3.2).

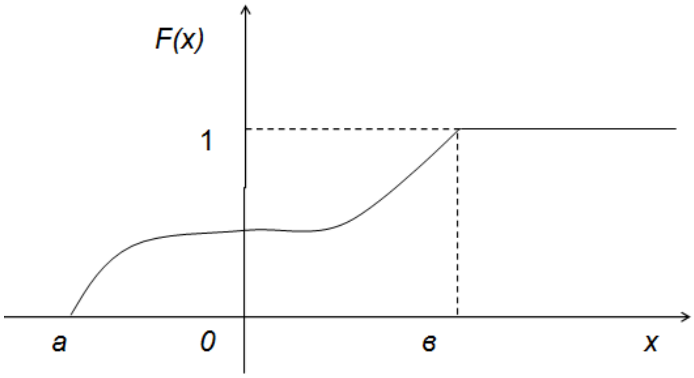


Рис. 4.3.2

Приклад 4.3.1. За заданим законом розподілу дискретної випадкової величини знайти інтегральну функцію розподілу випадкової величини *X* та побудувати її графік. Обчислити її математичне сподівання, дисперсію та середнє квадратичне відхилення.

X	0	1	2	3	4
p	0,12	0,36	0,34	0,15	0,03

Розв'язання.

$F(x) = P(x \le 0) = 0$;$F(1) = P(0 < x \le 1) = P(X = 0) = 0{,}12$;

$F(2) = P(< x \le 2) = P(X = 0) + P(X = 1) = 0{,}12 + 0{,}36 = 0{,}48$;

$F(3) = P(2 < x \le 3) = P(X = 0) + P(X = 1) + P(x = 2) = 0{,}48 + 0{,}34 = 0{,}82$;

$F(4) = P(3 < x \le 4) = P(X = 0) + P(X = 1) + P(x = 2) + P(X = 3) = 0{,}82 +$
$+ 0{,}15 = = 0{,}97$;

$F(\infty) = P(x > 4) = P(X = 0) + P(X = 1) + P(x = 2) + P(x = 3) + P(x = 4) =$
$= 0{,}97 + 0{,}03 = 1.$

Отже, $F(x) = \begin{cases} 0\ \ , \text{при}\ \ x \le 0, \\ 0{,}12,\ \text{при}\ 0 < x \le 1, \\ 0{,}48,\ \text{при}\ 1 < x \le 2, \\ 0{,}82,\ \text{при}\ 2 < x \le 3 \\ 0{,}97,\ \text{при}\ 3 < x \le 4, \\ 1, \text{при} x > 4. \end{cases}$

Графік функції розподілу наведено нижче (рис. 4.3.3).

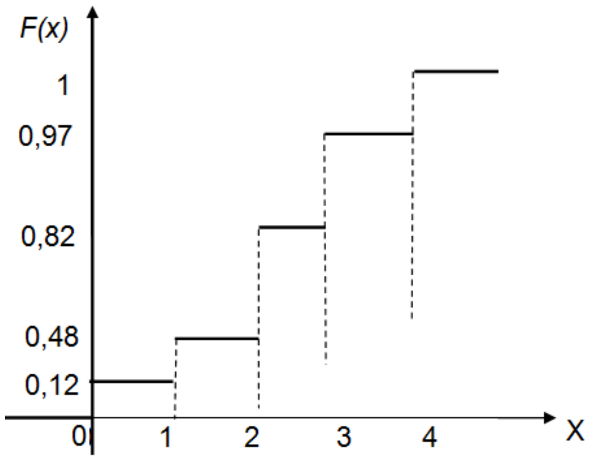


Рис. 4.3.3

Знайдемо математичне сподівання, дисперсію та середнє квадратичне відхилення.

$M(x) = x_1p_1+x_2p_2+\dots+x_np_n= 0\cdot 0{,}12 +1\cdot 0{,}36 + 2\cdot 0{,}34 +3\cdot 0{,}15 + +4\cdot 0{,}03 = 1{,}61.$

$D(X) = M(X-M(X))^2 = (0-1{,}61)^2 \cdot 0{,}12 + (1-1{,}61)^2 \cdot 0{,}36 + (2-1{,}61)^2 \cdot 0{,}34+ +(3-1{,}61)^2 \cdot 0{,}15 + (4-1{,}61)^2 \cdot 0{,}03 = 0{,}65.$

$\sigma(X) = \sqrt{D(X)} = \sqrt{0{,}65} = 0{,}81.$

Виявляється, що до побудови функції розподілу *F(x)* *неперервної* випадкової величини практично простіше підходити через іншу функцію - диференціальну.

Диференціальною функцією розподілу $f(x)$ називають першу похідну від інтегральної функції

$$f(x) = F'(x). \tag{4.3.2}$$

Функцію $f(x)$ також називають *щільністю* розподілу ймовірностей.

Властивості диференціальної функції

1. Диференціальна функція невід'ємна, тобто $f(x) \geq 0$, як похідна зростаючої функції.

2. Невласний інтеграл від диференціальної функції в межах $(-\infty;+\infty)$ дорівнює одиниці: $\int_{-\infty}^{\infty} f(x)dx = 1$ тому, що подія $-\infty < X < \infty$ є достовірною.

3. $f(x) = 0$ при *x<a* та *x≥ b,* тому що є похідною функції *F(x).*

Отже, інтегральна функція є первісною для диференціальної.

Зауважимо, що для опису розподілу ймовірностей дискретної випадкової величини диференціальна функція не використовується.

Графік щільності ймовірностей *f(x)* неперервної випадкової величини називають *кривою розподілу* (рис. 4.3.4).

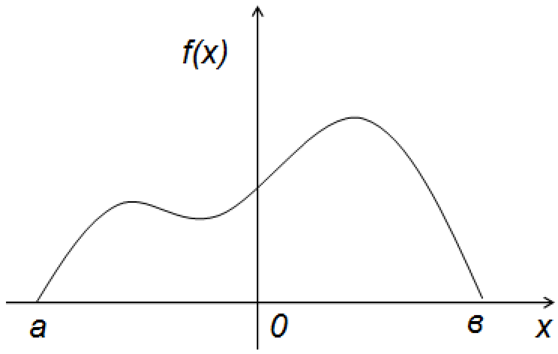


Рис. 4.3.4

Нагадаємо формули, які вказують на зв'язок між інтегральною та диференціальною функціями розподілу:

$$F(x) = \int_{-\infty}^{x} f(x)dx\,; \quad f(x) = F'(x)\,. \qquad (4.3.3)$$

Приклад 4.3.2. Щільність ймовірностей неперервної випадкової величини *X* має вигляд:

$$f(x) = \begin{cases} 0 \quad \text{при x} \leq \pi/6, \\ 3sin3x \quad \text{при } \pi/6 < \text{x} \leq \pi/3, \\ 0 \quad \text{при x} > \pi/3. \end{cases}$$

Знайти функцію розподілу *F(x).*

Розв'язання. Використаємо формулу знаходження інтегральної функції розподілу за відомою диференціальною $F(x)=\int_{-\infty}^{x} f(x)dx$.

1. $x \le \pi/6$; $F(x) = \int_{-\infty}^{x} 0dx = 0$.

2. $\pi/6 < x \le \pi/3$; $F(x) = \int_{-\infty}^{\Pi/6} 0dx + \int_{\Pi/6}^{x} 3\sin 3x dx = 3(-1/3)\cos 3x\Big|_{\pi/6}^{x} =$

$= -\cos 3x + \cos\frac{\pi}{2} = -\cos 3x.$

3. $x > \frac{\pi}{3}$; $F(x) = \int_{-\infty}^{\Pi/6} 0dx + \int_{\Pi/6}^{\Pi/3} 3\sin 3x dx + \int_{0/3}^{x} 0dx = -\cos 3x\Big|_{\Pi/6}^{\Pi/3} = -\cos\pi + +\cos\frac{\pi}{2} = 1.$

$$F(x) = \begin{cases} 0, & \text{при } x \le \pi/6, \\ -\cos 3x, & \text{при } \pi/6 < x \le \pi/3, \\ 1, & \text{при } x > \pi/3. \end{cases}$$

Знаючи обидві функції розподілу - інтегральну та диференціальну, можна обчислити ймовірність того, що неперервна випадкова величина прийме значення, яке належить заданому інтервалу.

Ймовірність того, що випадкова величина прийме значення, яке належить інтервалу (*a;b)* дорівнює приросту інтегральної функції на цьому інтервалі

$$P(a < X < b) = F(b) - F(a). \qquad (4.3.4)$$

Приклад 4.3.3. Неперервна випадкова величина *X* задана функцією розподілу:

$$F(x) = \begin{cases} 0, \text{при } x \le 0, \\ \dfrac{x}{2}, \text{при } 0 < x \le 2, \\ 1, \text{при } x \ge 2. \end{cases}$$

Знайти ймовірність того, що в результаті випробування випадкова величина *x* прийме значення, що належить інтервалу (1/4; 1).

Розв'язання. Шукана ймовірність дорівнює

$$P\ (a<X<b) = F(b) - F(a),$$

P (1/4<*X*<1) = *F*(1) – *F*(1/4). Так як за умовою *F(x)*=*x*/2 на інтервалі (1/4;1), то *F(*1) – *F*(1/4) = 1/2 -1/8 = 3/8

Отже, *P* (1/4<*X*<1) = 3/8.

4.4. Неперервна випадкова величина та її числові характеристики

Тепер дамо інше означення неперервної випадкової величини.

Випадкова величина називається *неперервною (абсолютно неперервною)*, якщо її інтегральну функцію розподілу можна подати у вигляді

$$F(x) = \int_{-\infty}^{x} f(x)dx,\ \ x \in R. \tag{4.4.1}$$

У випадку неперервних випадкових величин математичне сподівання, дисперсія та середнє квадратичне відхилення мають той же смисл і властивості, як і для дискретних випадкових величин, але обчислюються за іншими формулами.

Математичним сподіванням неперервної випадкової величини X, можливі значення якої належать інтервалу (*a;b),* називають визначений інтеграл

$$M(X) = \int_{a}^{b} xf(x)dx. \tag{4.4.2}$$

Якщо можливі значення належать всій осі, то

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx \qquad (4.4.2')$$

Якщо математичне сподівання M(X) існує і крива розподілу симетрична відносно прямої $x=C$, то $M(X)=C$.

Дисперсією неперервної випадкової величини X називають математичне сподівання квадрата її відхилення:

$$D(X) = \int_{a}^{b} (x - M(X))^2 f(x)dx. \qquad (4.4.3)$$

Якщо можливі значення *X* належать всій осі, то

$$D(X) = \int_{-\infty}^{\infty} (x - M(X))^2 \cdot f(x)dx. \qquad (4.4.3')$$

Дисперсію неперервної випадкової величини *X* можна також обчислити за такою формулою:

$$D(x) = \int_{a}^{в} x^2 f(x)dx - (M(x))^2. \qquad (4.4.4)$$

Середнє квадратичне відхилення неперервної випадкової величини означається рівністю

$$\sigma(X) = \sqrt{D(X)}. \qquad (4.4.5)$$

Приклад 4.4.1. Випадкова величина *X* задана диференціальною функцією розподілу

$$f(x) = \begin{cases} 0 \quad \text{при x} \le 0, \\ \frac{2}{9}x \quad \text{при } 0 < \text{x} \le 3, \\ 0 \quad \text{при x} > 3. \end{cases}$$

Знайти математичне сподівання та дисперсію.

Розв'язання. $M(x) = \int_{a}^{в} x \; f(x)dx = \int_{0}^{3} \frac{2}{9}x \; xdx = \frac{2}{9}\int_{0}^{3} x^2 dx = \frac{2}{9} \cdot \frac{x^3}{3} \Big|_0^3 =$

$= \frac{2 \cdot 3^3}{9 \cdot 3} - 0 = 2$;

D(x) = $\int\limits_a^в x^2 f(x)dx - (M(x))^2 = \int\limits_0^3 x^2 \frac{2}{9} xdx - 4 = \frac{2}{9} \cdot \frac{x^4}{4}\Big|_0^3 - 4 = 4{,}5 - 4 = 0{,}5$.

Приклад 4.4.2. Знайти математичне сподівання та дисперсію випадкової величини *X*, що задана інтегральною функцією

$$F(x) = \begin{cases} 0, & при \quad x \le 0, \\ x, & при \quad 0 \prec x \le 1, \\ 1, & при \quad x \succ 1. \end{cases}$$

Розв'язання. $$f(x) = F'(x) = \begin{cases} 0, & при \quad x \le 0, \\ 1, & при \quad 0 < x \le 1, \\ 0, & при \quad x > 1. \end{cases}$$

$$M(X) = \int\limits_0^1 x \cdot 1dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2}.$$

$$D(x) = \int\limits_a^в x^2 f(x)dx - (M(x))^2, \; D(x) = \int\limits_0^1 x^2 \cdot 1dx - \left(\frac{1}{2}\right)^2 = \frac{x^3}{3}\Big|_0^1 - \frac{1}{4} = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Теорема. Ймовірність того, що неперервна випадкова величина *X* прийме значення, яке належить інтервалу *(a;b),* дорівнює визначеному інтегралу від диференціальної функції, що взятий у межах від *a* до *b.*

$$P(a < X < b) = \int\limits_a^b f(x)dx. \qquad (4.4.6)$$

Геометрично ця ймовірність чисельно дорівнює площі заштрихованої криволінійної трапеції (рис. 4.4.1).

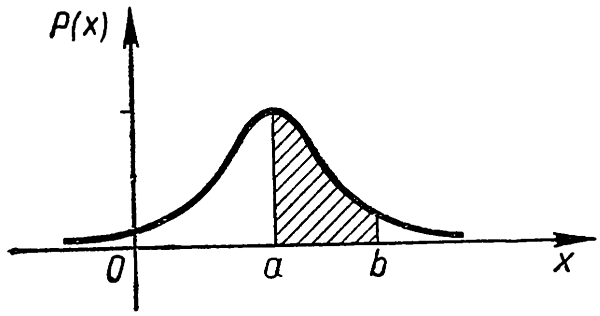


Рис. 4.4.1

Приклад 4.4.3. Випадкова величина *X* задана функцією розподілу

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

Знайти ймовірність того, що в результаті чотирьох незалежних випробувань величина *X* рівно три рази прийме значення, як належить інтервалу (0,25; 0,75).

Розв'язання. Знайдемо ймовірність попадання величини *X* у заданий інтервал (*a; b)* за формулою (4.4.6) $P(a<x<b)=\int_a^b f(x)dx$.

$$P(0{,}25<X<0{,}75) = \int_{0.25}^{0.75} 2x dx = x^2\Big|_{0{,}25}^{0{,}75} = (0{,}75)^2 - (0{,}25)^2 = 0{,}5.$$

За формулою Бернуллі знайдемо шукану ймовірність:

$$P_4(3) = C_4^3 \cdot p^3 \cdot q^1 = 4\cdot(0{,}5)^3\cdot 0{,}5 = 0{,}25.$$

4.5. Моменти випадкових величин. Коефіцієнт асиметрії та ексцес

Крім математичного сподівання, дисперсії та середнього квадратичного відхилення в теорії ймовірностей використовують також інші числові характеристики випадкової величини.

Узагальненням основних числових характеристик випадкової величини є поняття моментів випадкової величини.

Нехай *k* ≥ 0 - ціле число, *a* – дійсне число. *Моментом k–того порядку* випадкової величини *X* відносно точки *a* називається число (якщо воно існує), яке дорівнює математичному сподіванню величини $(X\text{-}a)^k$ і позначається

$$\nu_k(a) = M(X-a)^k \tag{4.5.1}$$

Якщо *a =0*, то момент називається *початковим.* Початковий момент *k* – того порядку називають математичним сподіванням величини X^k і позначають

$$\nu_k = M(X^k) \tag{4.5.2}$$

зокрема, $\nu_1 = M(X)$ – це математичне сподівання випадкової величини *X*, $\nu_2 = M(X^2)$, $\nu_3 = M(X^3)$, …

Центральним моментом називається момент відносно точки *a=M(X). Центральним моментом k–того порядку* випадкової величини *X* називають математичне сподівання величини $(X-M(X))^k$ і позначають:

$$\mu_k = M(X - M(X))^k. \tag{4.5.3}$$

Легко побачити, що μ_1 =0, а $\mu_2 = D(X)$.

Між центральними і початковими моментами існує певний зв'язок. Справді, за означенням

$$\mu_k = M(x - \boldsymbol{\nu}_k)^k. \tag{4.5.4}$$

Розкриваючи вираз у правій частині, дістанемо, зокрема, значення для перших чотирьох моментів: μ_1=0,

$$\mu_2 = V_2 - V_1^2,$$

$$\mu_3 = V_3 - 3V_2V_1 + 2V^3{}_1,$$

$$\mu_4 = V_4 - 4V_3V_1 + 6V_2V_1^2 - 3V_1^4. \tag{4.5.5}$$

Введені моменти називають *теоретичними*. На відміну від теоретичних моментів, моменти, які обчислюють за даними спостережень, називають *емпіричними*.

Вищенаведені формули теоретичних моментів мають місце для дискретної випадкової величини *X*. Для неперервного її розподілу теоретичні моменти означають такими формулами:

$$\boldsymbol{\nu_k} = \int_{-\infty}^{+\infty} x^k f(\mathrm{x})dx, \qquad (4.5.6)$$

$$\mu_{\boldsymbol{k}} = \int_{-\infty}^{+\infty} (x - M(X))^k f(\mathrm{x})dx \qquad (4.5.7)$$

де p_i - ймовірність, з якою випадкова величина *X* набуває значення x_i, а $f(x)$ – щільність розподілу випадкової величини *X.*

Значення μ_1 завжди дорівнює 0.

Наступним після нього центральний момент непарного порядку є μ_3, який характеризує відхилення розподілу випадкової величини *X* від симетрії. За міру *As* цього відхилення беруть відношення нормованого центрального моменту третього порядку до кубу середнього квадратичного відхилення, яке називають *коефіцієнтом асиметрії*, що є характеристикою скошеності (асиметрії) розподілу:

$$As = \frac{\mu_3}{\sigma^3}. \qquad (4.5.8)$$

Додатна асиметрія ($As > 0$) вказує на відхилення розподілу у бік додатних значень (правобічна асиметрія – права вітка кривої розподілу довша від лівої). Якщо $As < 0$, то є наявним відхилення у бік від'ємних значень (лівобічна асиметрія – ліва вітка кривої розподілу довша від правої).

Нормований центральний момент четвертого порядку служить характеристикою крутості (гостровершинності або плосковершинності) розподілу випадкової величини *X* у порівнянні з крутістю розподілу нормальної випадкової величини з математичним сподіванням і дисперсією такими ж, як і в *X*. За міру крутості беруть величину *Ex*, яка називається *ексцесом*:

$$Ex = \frac{\mu\,4}{V^4} - 3. \qquad (4.5.9)$$

Якщо $Ex > 0$, то крива розподілу більш гостровершинна, при $Ex < 0$ – більш плосковершинна, згладжена порівняно з кривою нормального розподілу.

Приклад 4.5.1. Дискретна випадкова величина *X* задана законом розподілу

X	-2	-1	0	1	2	3
p	0,1	0,2	0,3	0,25	0,1	0,05

Знайти початкові та центральні моменти цієї величини до четвертого порядку включно, коефіцієнт асиметрії та ексцес розподілу.

Розв'язання. За формулою $V_k = \sum_i x_i^k p_i$ знаходимо початкові моменти визначених порядків:

V_1 = - 2·0,1-1·0,2+0·0,3+1·0,25+2·0,1+3·0,05 =0,2;

$V_2 = (-2)^2 \cdot 0{,}1+(-1)^2 \cdot 0{,}2+0^2 \cdot 0{,}3+1^2 \cdot 0{,}25+2^2 \cdot 0{,}1+3^2 \cdot 0{,}05 = 1{,}7$;

$V_3 = (-2)^3 \cdot 0{,}1+(-1)^3 \cdot 0{,}2+0^3 \cdot 0{,}3+1^3 \cdot 0{,}25+2^3 \cdot 0{,}1+3^3 \cdot 0{,}05 = 1{,}4$;

$V_4 = (-2)^4 \cdot 0{,}1+(-1)^4 \cdot 0{,}2+0^4 \cdot 0{,}3+1^4 \cdot 0{,}25+2^4 \cdot 0{,}1+3^4 \cdot 0{,}05 = 1{,}4$.

Центральні моменти знаходимо за такими формулами: $\mu_1 = 0$,

$\mu_2 = V_2 - V_1^2 = 1{,}7 - (0{,}2)^2 = 1{,}66$;

$\mu 3 = V_3 - 3V_2 V_1 + 2V_1^3 = 1{,}4 - 3 \cdot 1{,}7 \cdot 0{,}2 + 2 \cdot (0{,}2)^3 = 0{,}40$;

$\mu_4 = V_4 - 4V_3 V_1 + 6V_2 V_1^2 - 3 V_1^4 = 7{,}7 - 4 \cdot 1{,}4 \cdot 0{,}2 + 6 \cdot 1{,}7 \cdot (0{,}2)^2 -$
$-3 \cdot (0{,}2)^4 = 6{,}17$.

Коефіцієнт асиметрії $As = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{(\sqrt{\mu 2})^3} = \frac{\mu_3}{\mu 2 \sqrt{\mu 2}} = \frac{0{,}4}{1{,}66 \cdot 1{,}29} = 0{,}19$.

Ексцес $Ex = \frac{\mu\, 4}{V^4} - 3 = \frac{\mu\, 4}{\mu 2^{\,2}} - 3 = \frac{6{,}17}{(1{,}66)^2} - 3 = -0{,}76$.

Приклад 4.5.2. Знайти початкові моменти випадкової величини з щільністю ймовірностей *f(x)*= $\begin{cases} 0 \text{ для } x < 1, \\ \frac{4}{x^5} \text{ для } x \geq 1. \end{cases}$

Розв'язання. За формулою(4.5.6)

$$\boldsymbol{\nu_k} = \int_{-\infty}^{+\infty} x^k f(\mathrm{x})dx = \int_1^{+\infty} x^k \frac{4}{\mathrm{x}^5} dx == 4\int_1^{+\infty} x^{k-5} dx = \left.\frac{4x^{k-4}}{k-4}\right|_1^{+\infty} = \frac{4}{4-k}.$$

Очевидно, що розрахунки правильні, коли $k<4$, тому що при $k\geq4$ даний інтеграл розбіжний. Отже, дана випадкова величина має початкові моменти першого, другого та третього порядків, а моментів четвертого порядку та вище не має.

$$\boldsymbol{\nu_1 = \frac{4}{3}, \nu_2 = 2,\ \nu_3 = 4.}$$

4.6. Мода та медіана

Дамо означення ще декількох числових характеристик випадкових величин, які часто використовуються в теорії та на практиці.

Модою $M_o(X)$ неперервної випадкової величини називають те її значення, якому відповідає найбільше значення диференціальної функції розподілу.

Очевидно, що для дискретної випадкової величини модою буде те її значення, якому відповідає найбільша ймовірність, тобто $M_o(X)=x_i$, де $P(X=x_i) = \max p_i$, $i = 1,\ldots,n$.

Якщо випадкова величина має одну моду, то такий розподіл ймовірностей називають одномодальним; якщо розподіл має дві моди – двомодальним (*бімодальним*.) і т. д. Існують і такі розподіли, які не мають моди. Їх називають антимодальними.

Приклад 4.6.1. Знайти моду випадкової величини, яка задана диференціальною функцією розподілу $f(x)=\frac{3}{4}\mathrm{x}^2 + \frac{9}{2}\mathrm{x} - 6$.

Розв'язання*.* Знайдемо максимум функції $f(x)$, для цього обчислимо похідну функції $f'(x) = \left(\frac{3}{4}\mathrm{x}^2 + \frac{9}{2}\mathrm{x} - -6\right)' = \frac{3}{2}\mathrm{x}+\frac{9}{2}$. Якщо $f''(x)=0$, то $x=3$. Отже, в точці $x=3$ функція щільності має максимальне значення. Тому $M_o(X)=3$.

Медіаною $M_e(X)$ неперервної випадкової величини X називають те її можливе значення, яке визначається рівністю $P(X<M_e(X))= P(X>M_e(X))= \frac{1}{2}$.

Можна сформулювати означення медіани й інакше.

Медіаною $M_e(X)$ випадкової величини X називають те її значення, для якого $F(M_e(X))\leq \frac{1}{2}$.

Геометрично, медіана – це абсциса точки, в якій площа, що обмежена кривою розподілу і віссю абсцис, ділиться навпіл.

При симетричному розподілі всі три значення центральної тенденції (математичне сподівання, медіана, мода) будуть співпадати. При зміщеному (асиметричному) розподілі всі три значення можуть бути різними.

Приклад 4.6.2. Випадкова величина X задана диференціальною функцією розподілу $f(x)=2\cos 2x$ в інтервалі $(0;\frac{\pi}{4})$. Знайти медіану величини X.

Розв'язання. За умовою можливі значення X додатні, тому

$$P(0 < X < Me(X)) = \frac{1}{2}.$$

$$2\int_0^{Me(X)} \cos 2x\, dx = \sin(2Me(X)) = \frac{1}{2}.$$ Звідси $2\big(Me(X)\big) = \arcsin\frac{1}{2} = \frac{\pi}{6}$.

Отже, $Me(X) = \frac{\pi}{12}$.

Приклад 4.6.3. Знайти медіану $M_e(X)$, якщо інтегральна функція розподілу має вигляд:

$$F(X) = \begin{cases} 0, \text{ при } x \leq -2, \\ \frac{28+24x+3x^2-x^3}{108}, \text{ при } -2 < x \leq 4, \\ 1, \text{ при } x > 4. \end{cases}$$

Розв'язання. Для знаходження медіани використаємо формулу $F(Me(X)) = \frac{1}{2}$.

$\frac{28+24\mathrm{Me}(\mathrm{X})+3\mathrm{Me}^2(\mathrm{x})-\mathrm{Me}^3(\mathrm{x})}{108} = \frac{1}{2}.$

$$24\mathrm{Me}(\mathrm{X}) + \mathrm{Me}^3(\mathrm{x}) - 3\mathrm{Me}^2(\mathrm{x}) - 24\mathrm{Me}(\mathrm{X}) + 26 = 0.$$

Отже, $M_e(X)=1$, як розв'язок даного рівняння.

Запитання для самоперевірки

1. Яка випадкова величина називається дискретною?
2. Дайте означення закону розподілу дискретної випадкової величини. Наведіть приклади.
3. Як означають числові характеристики дискретних випадкових величин? Що вони характеризують? За якими формулами їх обчислюють ?
4. Визначте, чи будуть дискретними такі випадкові величини:

а) кількість вибитих очок при стрільбі по мішені;

б) відстань від точки попадання кулі до центра мішені;

в) час роботи даної електронної лампи в даному електронному приладі;

г) кількість в даному приладі електронних ламп, які вийшли із ладу за даний проміжок часу;

д) момент виходу із ладу даної електронної лампи даного приладу.

5. Що називається функцією розподілу випадкової величини?
6. Наведіть означення неперервної випадкової величини.
7. Що називається щільністю ймовірностей випадкової величини? Як взаємопов’язані щільність ймовірностей і функція розподілу неперервної випадкової величини?
8. Як виражається ймовірність попадання випадкової величини у скінчений інтервал (*a;b*) через функцію розподілу *F(x)*?

9. Як виражається ймовірність попадання неперервної випадкової величини у скінчений інтервал (*a;b*) через щільність ймовірності *f(x)*?

10. Що називається математичним сподіванням дискретної випадкової величини? Що характеризує математичне сподівання?

11. Що називається математичним сподіванням неперервної випадкової величини?

12. Дайте означення дисперсії випадкової величини. Які властивості випадкової величини характеризує дисперсія?

13. Як означають інтегральну функцію розподілу та щільність імовірностей неперервної випадкової величини? Які властивості мають ці функції?

14. Який існує зв'язок між інтегральною функцією розподілу та щільністю ймовірності?

15. Як знаходять імовірність попадання випадкової величини в інтервал (*a;b*), використовуючи інтегральну функцію розподілу та щільність ймовірностей?

16. Яка випадкова величина називається неперервною?

17. Що називається математичним сподіванням, дисперсією та середнім квадратичним відхиленням неперервної випадкової величини?

18. Які формули використовуються при обчисленні ймовірності попадання випадкової величини у заданий інтервал?

19. Як означають початкові та центральні моменти випадкових величин, коефіцієнт асиметрії та ексцес?

20. Що називається модою та медіаною випадкової величини та які формули використовуються при їх обчисленні?

Вправи та задачі

4.1. Випадкова величина задана законом розподілу:

X	2	4	8	6
P	0,1	0,3	0,4	0,2

Знайти: а) дисперсію (двома способами); б) середнє квадратичне відхилення.

4.2. Знайти математичне сподівання випадкової величини та дисперсію випадкової величини *Z*, якщо відомі математичні сподівання *X* і *У: М(Х)* =5 і *М(У)* = 9; а) *Z* = 3*X*+4*У*; б) *Z* = 8у-7.

4.3. Незалежні випадкові величини *X* і *У* мають наступні закони розподілу :

X	3	2	4
P	0,2	0,3	0,5

Y	1	2	3	5
P	0,4	0,2	0,3	0,1

Скласти закон розподілу випадкової величини а) *X*+*Y*; б) *XУ.* Перевірити властивості математичного сподівання та дисперсії цих випадкових величин.

4.4. Скласти закон розподілу дискретної випадкової величини X, яка може приймати тільки два значення x_1 з імовірністю p_1 та x_2 з імовірністю p_2, причому $x_1 < x_2$, і $p_1 = 0,9$; *M(X)* = 4,1*; D(X)* = 0,09.

4.5. У партії 10% нестандартних деталей. Навмання відібрані 4 деталі. Скласти біноміальний закон розподілу дискретної випадкової величини *X*– числа нестандартних деталей серед 4 відібраних і побудувати многокутник отриманого розподілу.

4.6. Цінні папери на біржі можуть з ймовірністю p=0,5 подорожчати на 10%. Спостереження ведеться три дні. Початкова вартість цінних паперів 5000 грн. Треба: а) побудувати розподіл випадкової величини *X* (вартість цінних паперів), вважаючи, що вона має біноміальний закон розподілу; б) знайти середню сподівану вартість цінних паперів після трьох днів торгів; в) підрахувати стандартне відхилення.

4.7. Знайти математичне сподівання і дисперсію випадкової величини *X* – числа появи події в 100 незалежних випробуваннях, у кожному з яких ймовірність появи події рівна 0,7.

4.8. Знайти математичне сподівання числа виграшних лотерейних білетів, якщо куплено 60 білетів, причому ймовірність виграшу становить 0,05.

4.9. Цінні папери на біржі можуть з імовірністю *p*=0,5 подешевшати на 10%. Спостереження велося 2 дні. Початкова вартість цінних паперів 400 грн. Потрібно: а) побудувати розподіл випадкової величини *X* (вартість цінних паперів), вважаючи, що вона має біноміальний закон розподілу; б) знайти середню сподівану вартість цінних паперів після двох днів торгів; в) підрахувати стандартне відхилення.

4.10. Дискретна випадкова величина задана законом розподілу:

X	3	4	7	10
P	0,2	0,1	0,4	0,3

Знайти функцію розподілу *F(x)* і побудувати її графік.

4.11. Випадкова величина *X* задана інтегральною функцією розподілу: $F(x) = \begin{cases} 0, \text{при x} \leq 0, \\ \frac{x^2}{4}, \text{при } 0 < \text{x} \leq 2, \\ 1, \text{при x} > 2. \end{cases}$

Знайти щільність розподілу і побудувати графіки інтегральної та диференціальної функцій.

4.12. Задана диференціальна функція розподілу неперервної випадкової величини *X*: $f(x) = \begin{cases} 0, \text{при x} \le 0, \\ sinx, \text{при } 0 < \text{x} \le \pi/2, \\ 0, \text{при x} > \pi/2. \end{cases}$

Знайти функцію розподілу *F(x)* та побудувати графіки функцій *F(x)* і *f(x).*

4.13. Випадкова величина *X* задана інтегральною функцією:

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

Побудувати її графік і знайти ймовірність того, що *X* прийме значення, яке належить інтервалу: а) (1,2; 1,6); б) (1,1; 2,5).

4.14. Задана щільність розподілу перервної випадкової величини: $f(x) = \begin{cases} 0, \text{при x} \le 0, \\ \frac{2}{9}x, \text{при } 0 < \text{x} \le 3, \\ 0, \text{при x} > 3. \end{cases}$

Знайти ймовірність того, що *X* прийме значення, яке належить інтервалу: а) (1; 3); б) (2; 9).

4.15. Випадкова величина *X* задана функцією розподілу

$$F(x) = \begin{cases} 0, & \text{при x} \le 2, \\ 0{,}5x - 1, & \text{при } 2 < \text{x} \le 4, \\ 1, & \text{при x} > 4. \end{cases}$$

Знайти ймовірність того, що в результаті випробування *x* прийме значення: а) менше за 0,2; б) менше за 3; в) не менше, ніж 3; г) не менше п’яти.

4.16. Щільність розподілу неперервної випадкової величини *X* в інтервалі (0; π/3) приймає значення *f(x)* = (3/2)*sin3x*, зовні цього інтервалу *f(x)* = 0. Знайти ймовірність того, що в трьох незалежних випробуваннях *X* прийме рівно 2 рази значення, яке належить інтервалу (π/6; π/4).

4.17. Дискретна випадкова величина задана законом розподілу.

X	2	6	10
P	0,5	0,4	0,1

Знайти функцію розподілу *F(x)* і побудувати її графік.

4.18. Випадкова величина *X* задана інтегральною функцією *F*(*x*): $F(x)=\begin{cases}0, & \text{при x} \le 1,\\ 0{,}5(x-1), & \text{при } 1<\text{x}\le 3,\\ 1, & \text{при x} > 3.\end{cases}$

Знайти щільність розподілу. Побудувати графіки інтегральної та диференціальної функцій розподілу.

4.19. Випадкова величина *X* задана диференціальною функцією: $f(x)=\begin{cases}0, \text{при x} \le 0,\\ \frac{1}{2}sinx, \text{при } 0<\text{x}\le \pi,\\ 0, \text{при x} > \pi.\end{cases}$

Знайти функцію розподілу *F(x)* і побудувати графіки функцій *f(x)* та *F(x).*

4.20. Задана інтегральна функція розподілу:

$$F(x)=\begin{cases}0, & \text{при x} \le 2,\\ 0{,}5x-1, & \text{при } 2<\text{x}\le 4,\\ 1, & \text{при x} > 4.\end{cases}$$

Знайти ймовірність того, що *X* прийме значення, яке належить інтервалу: а) (3; 4); б) (4; 8).

4.21. Задана щільність розподілу перервної випадкової величини *X*

$$f(x)=\begin{cases}0, \text{при x} \le 1,\\ x-\dfrac{1}{2}, \text{при } 1<\text{x}\le 2,\\ 0, \text{при x} > 2.\end{cases}$$

Знайти ймовірність того, що *X* прийме значення, яке належить інтервалу: а) (1,4; 1,8); б) (1,2; 5).

4.22. Випадкова величина задана функцією розподілу

$$F(x)=\begin{cases}0, & \text{при } x \le 1,\\ \frac{1}{3}x-\frac{1}{3}, & \text{при } 1 < x \le 4,\\ 1, & \text{при } x > 4.\end{cases}$$

Знайти ймовірність того, що в результаті випробувань *X* прийме значення: а) менше –2; б) менше 1; в) не менше 1; г) не менше 5.

4.23. Випадкова величина задана диференціальною функцією

$$f(x)=\begin{cases}0, & \text{при } x \le 0,\\ \frac{1}{2}sinx, & \text{при } 0 < x \le \pi,\\ 0, & \text{при } x > \pi.\end{cases}$$

Знайти ймовірність того, що у п'яти незалежних випробуваннях *X* прийме рівно три рази значення, яке належить інтервалу (π/6; π/2).

4.24. Ймовірність влучення у мішень одним пострілом дорівнює 0,7. Проведено 4 постріли. Дискретна випадкова величина *X* – число влучень у мішень. Знайти: а) закон розподілу *X*; б) функцію розподілу *X*; в) ймовірність подій $X < 2$; $X \le 3$; $1 < X \le 3$; г) математичне сподівання, дисперсію та середнє квадратичне відхилення випадкової величини *X*.

4.25. Два стрільці виконують по одному пострілу в одну мішень. Імовірність влучення у мішень першим стрільцем дорівнює 0,5; другим – 0,4. Дискретна випадкова величина *X* – число влучень у мішень. а) Знайти закон розподілу і функцію розподілу величини *X*; б) побудувати многокутник розподілу; в) знайти ймовірність події $X \ge 1$; г) обчислити математичне сподівання, дисперсію та середнє квадратичне відхилення величини *X*.

4.26. У коробці є 7 олівців, з яких 4 червоні. З коробки навмання беруть 3 олівці. а) Знайти закон розподілу і функцію розподілу випадкової величини *X:* число червоних олівців, узятих із коробки; б) побудувати многокутник розподілу *X*; в) знайти ймовірність події $0<x\leq 2$; г) обчислити математичне сподівання, дисперсію і середнє квадратичне відхилення величини *X*.

4.27. З 25-ти контрольних робіт, 5 з яких написані на оцінку «відмінно», навмання беруть 3 роботи. Знайти закон розподілу і функцію розподілу випадкової величини *X*, що дорівнює числу оцінених на «відмінно» робіт серед вибраних. Чому дорівнює ймовірність події $X>0$? Обчислити математичне сподівання, дисперсію і середнє квадратичне відхилення величини *X*.

4.28. Робітник обслуговує 4 агрегати, що працюють незалежно. Ймовірність того, що протягом однієї години агрегат не потребує уваги робітника, дорівнює: для першого 0,7; для другого – 0,75; для третього – 0,8; для четвертого – 0,9. Знайти закон розподілу та функцію розподілу випадкової величини *X,* що дорівнює числу агрегатів, які не потребують уваги робітника.

4.29. Неперервну випадкову величину задано щільністю розподілу *f(x).* Знайти функцію розподілу *F(x)*, математичне сподівання, дисперсію та середнє квадратичне відхилення випадкової величини *X*, якщо

$$f(x)=\begin{cases} 0 & \text{для } x<-\frac{\pi}{2}, \\ 0{,}5\cos x & \text{для } -\frac{\pi}{2}\leq x<\frac{\pi}{2}, \\ 0 & \text{для } x>\frac{\pi}{2}. \end{cases}$$

4.30. Неперервну випадкову величину задано щільністю розподілу *f(x).* Знайти функцію розподілу *F(x),* математичне сподівання, дисперсію та середнє квадратичне відхилення випадкової величини *X,* якщо

$$f(x)=\begin{cases} 0 & \text{для } x\leq 0, \\ e^{x} & \text{для } x>0. \end{cases}$$

4.31. Задано функцію $f(x)=\begin{cases} 0 \text{ для } x\leq 0, \\ Ce^{-x} \text{ для } x>0. \end{cases}$

Для якого значення величина *C* ця функція буде щільністю розподілу деякої неперервної випадкової величини *X*? Знайти її функцію розподілу, математичне сподівання та середнє квадратичне відхилення.

4.32. Задано функцію розподілу випадкової величини *X*

$$F(x)=\begin{cases} 0, & \text{для} x < -1, \\ \frac{x+1}{2}, & \text{для} -1 \le x < 1, \\ 0, & \text{для} x \ge 1. \end{cases}$$

Знайти щільність розподілу *f (x)*, її математичне сподівання, дисперсію та середнє квадратичне відхилення.

4.33. Випадкову величину задано функцією розподілу

$$F(x)=\begin{cases} 0, & \text{для} x < 1, \\ \frac{1}{7}(x^3-1), & \text{для} \ 1 \le x < 2, \\ 1, & \text{для} x \ge 2. \end{cases}$$

Знайти щільність розподілу, математичне сподівання, дисперсію та середнє квадратичне відхилення цієї величини.

4.34. Задано щільність розподілу випадкової величини *X*

$$f(x)=\begin{cases} 0, & \text{для} x < 0, \\ \cos x, & \text{для} \ 0 \le x < \frac{\pi}{2}, \\ 0, & \text{для} x \ge \frac{\pi}{2}. \end{cases}$$

Знайти функцію розподілу випадкової величини *X*, її математичне сподівання, дисперсію та середнє квадратичне відхилення.

4.35. Задано функцію розподілу неперервної випадкової величини *X*

$$F(x)=\begin{cases} 0 & \text{для} x < 0, \\ \sin 2x & \text{для} \ 0 \le x < \frac{\pi}{2}, \\ 0 & \text{для} x \ge \frac{\pi}{4}. \end{cases}$$

Знайти щільність розподілу випадкової величини *X,* її математичне сподівання, дисперсію та середнє квадратичне відхилення.

4.36. Задано щільність розподілу випадкової величини *X*

$$f(x)=\begin{cases} 0, \text{для } x<0 \ \text{ і } x>1, \\ 4x^3, \ \text{ для } 0\le x \ <1. \end{cases}$$

Знайти функцію розподілу *X*, її математичне сподівання, дисперсію та середнє квадратичне відхилення.

4.37. Задана щільність розподілу випадкової величини. Знайти початкові та центральні моменти перших чотирьох порядків, асиметрію та ексцес, якщо

$$f(x)=\begin{cases} 0 \quad \text{для} x<0, \\ 1{,}5\ \text{x}-0{,}75 \quad \text{для } 0\le \text{x}<2, \\ 0 \quad \text{для} x\ge \ 2. \end{cases}$$

4.38. Знайти моду та медіану випадкової величини з щільністю розподілу ймовірностей $f(x)=3x^2$ при $x \in [0;1]$.

4.39. Дискретна випадкова величина *X* задана законом розподілу

X	1	3	5	7	9
P	0,1	0,4	0,2	0,2	0,1

Знайти початкові та центральні моменти цієї величини до четвертого порядку включно, коефіцієнт асиметрії та ексцес розподілу.

4.40. Знайти початкові та центральні моменти перших чотирьох порядків випадкової величини зі щільністю ймовірностей $f(x)= 0{,}5x$ на інтервалі (0;2), зовні цього інтервалу $f(x)=0$.

4.41. Дискретна випадкова величина *X* задана законом розподілу

X	2	4	6	8
p	0,4	0,3	0,2	0,1.

Знайти початкові та центральні моменти цієї величини до четвертого порядку включно, коефіцієнт асиметрії та ексцес розподілу.

4.42. Знайти початкові та центральні моменти випадкової величини, яка розподілена зі щільністю ймовірностей

$$f(x)=\begin{cases} 0 \text{для } x \le 1, \\ \frac{9}{x^{10}} \text{ для } x > 1. \end{cases}$$

4.43. Випадкова величина X має приймати лише наступні значення: -2; -1; 0; 1; 2 з імовірностями відповідно p_{-2} ; p_{-1}; p_0; p_1; p_2. Знайти ці ймовірності, якщо:

а) $M(X) = M(X^2) = 0$; $M(X^2) = 1$; $M(X^4) = 2$;

б) $M(X) = M(X^3) = 0$; $M(X^2) = 2$; $M(X^4) = 6$;

в) $M(X) = M(X^3) = 0$; $M(X^2) = a$; $M(X^2) = b$; чи будь-які значення *a* та *b* можна взяти у цьому випадку?

РОЗДІЛ 5
ЗАКОНИ РОЗПОДІЛУ НЕПЕРЕРВНИХ ВИПАДКОВИХ ВЕЛИЧИН

Випадкові величини, що характеризують природничі, фізичні, економічні та інші процеси і явища, можуть мати найрізноманітніші закони розподілу. *Розподіл випадкової величини* – це функція, яка однозначно визначає ймовірність того, що випадкова величина прийме задане значення або належить до певного заданого інтервалу. Він може бути розрахований як за функцією розподілу, так і, навпаки, функція розподілу – за розподілом. Основні закони розподілу дискретних випадкових величин розглянуті у розділах 3 та 4, вони часто використовуються при побудові теоретичних ймовірнісних моделей реальних соціально-економічних процесів. Залежно від того, за якою формулою обчислюються ймовірності P_i, ці закони мають свою назву: біноміальний, пуассонівський, геометричний і т.п.

При розв'язуванні практичних задач доводиться зустрічатись з різними розподілами неперервних випадкових величин. Щільність розподілу неперервних випадкових величин також називаються законом розподілу, а за виглядом диференціальних функцій розподілу (щільності ймовірностей) $f(x)$ розрізняються основні закони розподілу неперервних випадкових величин.

5.1. Рівномірний закон розподілу

Серед неперервних випадкових величин (НВВ) найбільш поширене вираження щільності ймовірностей має рівномірний розподіл випадкових величин.

Розподіл ймовірностей НВВ називають *рівномірним,* якщо на інтервалі, якому належать всі можливі її значення, щільність розподілу має стале значення:

$$f(x) = \begin{cases} C & \text{при } x \in [a;b], \\ 0 & \text{при } x \notin [a;b], \end{cases} \qquad (5.1.1)$$

де $C = \frac{1}{b-a} = \text{const}$. Отже, всі можливі значення рівномірно розподіленої випадкової величини лежать у межах деякого визначеного інтервалу. У межах цього інтервалу значення випадкової величини мають одну і ту ж імовірність, тобто одну і ту ж щільність ймовірностей. Тому запишемо

$$f(x) = \begin{cases} 0, & \text{при x} \le a, \\ \frac{1}{b-a}, & \text{при } a < \text{x} \le b, \\ 0, & \text{при x} > b. \end{cases} \qquad (5.1.2)$$

Числа *a* та *b* називаються параметрами розподілу.

Якщо ймовірність попадання випадкової величини на інтервал (*a; b*) пропорційна до довжини інтервалу і не залежить від розташування інтервалу на осі, то вона має рівномірний закон розподілу.

Рівномірний закон розподілу легко моделювати. За допомогою функціональних перетворень із величин, що розподілені рівномірно, можна діставати величини з довільним законом розподілу.

Знайдемо функцію розподілу *F(x)* за заданою щільністю розподілу ймовірностей. Маємо:

$$F(x) = \int_{-\infty}^{x} f(x)dx = \int_{a}^{x} \frac{1}{b-a}dx = \frac{x}{b-a}\Big|_a^x = \frac{x-a}{b-a}.$$

$F(x) = 0$ при $x \le a$ і $F(x) = 1$ при $x \ge b$. Отже,

$$F(x) = \begin{cases} 0, & \text{при x} \le a, \\ \frac{\text{x}-a}{b-a}, & \text{при } a < \text{x} \le b, \\ 1, & \text{при x} \ge b. \end{cases} \qquad (5.1.3)$$

Побудуємо графіки функцій розподілу *F*(*x*) та щільності ймовірностей *f*(*x*) рівномірного закону (рис. 5.1.1, 5.1.2).

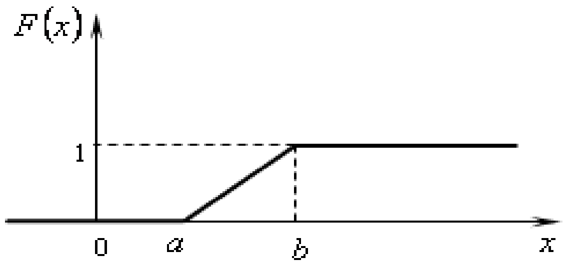


Рис. 5.1.1

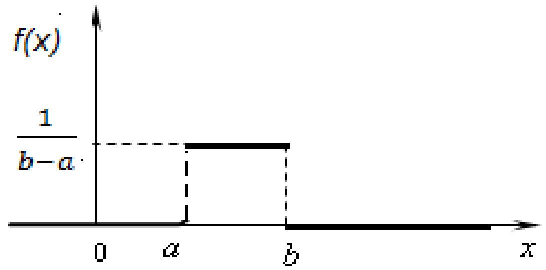


Рис. 5.1.2

Знайдемо основні числові характеристики даної випадкової величини. За відомою формулою знайдемо математичне сподівання

$$M(X) = \int_a^b Xf(x)dx = \int_a^b \frac{xdx}{b-a} = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{b+a}{2}. \tag{5.1.4}$$

Отже, *математичне сподівання* випадкової величини, що рівномірно розподілена на відрізку (a; b), є серединою цього відрізка.

Знайдемо *дисперсію* рівномірно розподіленої величини X:

$$D(X) = \int_a^b (x - M(X))^2 \cdot f(x)dx = \int_a^b \left(x - \frac{b+a}{2}\right)^2 \frac{1}{b-a}dx = \frac{(b-a)^2}{12}. \tag{5.1.5}$$

Середнє квадратичне відхилення випадкової величини

$$\sigma(X) = \sqrt{D(X)} = \frac{b-a}{2\sqrt{3}}. \tag{5.1.6}$$

Оскільки крива щільності розподілу симетрична, то *медіана*

$$Me(X) = M(X) \tag{5.1.7}$$

Моди закон рівномірної щільності не має.

В силу симетричності розподілу його центральні моменти непарного порядку дорівнюють нулю, а, отже, коефіцієнт *асиметрії* також дорівнює нулю : $As = 0$.

Для визначення *ексцесу* знаходимо центральний момент четвертого порядку:

$$\mu_4 = \int_{-\infty}^{+\infty} (x - \mathrm{M}(\mathrm{x}))^4 f(\mathrm{x}) dx = \int_a^в (x - \frac{a+в}{2})^4 \frac{1}{в-a} dx = \frac{(b-a)^4}{80},$$

звідки

$$Ex = \frac{\mu\ 4}{V^4} - 3 = \frac{(b-a)^4 144}{80(b-a)^4} - 3 = -1{,}2. \tag{5.1.8}$$

Для випадкової величини *X*, що рівномірно розподілена на відрізку (*a; b), ймовірність попадання в інтервал (α; β)*, який належить цьому відрізку, обчислюється за формулою:

$$P\ (\alpha < X < \beta) = \int_\alpha^\beta f(x) dx = \int_\alpha^\beta \frac{dx}{b-a} = \frac{\beta-\alpha}{b-a}, \tag{5.1.9}$$

тобто дорівнює відношенню довжини відрізка (*α; β*) до довжини всього відрізка (*a; b*).

Рівномірний закон розподілу задовольняють, наприклад, похибки заокруглень різноманітних розрахунків. Рівномірно розподілена випадкова величина зустрічається також у вимірювальній практиці при заокругленні показів вимірювальних приладів до цілих поділок шкал. Помилка при заокругленні до найближчої цілої поділки є випадковою величиною *X*, яка може приймати зі сталою ймовірністю будь-яке значення між двома сусідніми поділками.

5.2. Нормальний закон розподілу

На практиці найчастіше зустрічається так званий нормальний закон розподілу (або закон Гаусса).

Нормальним називають розподіл ймовірностей неперервної випадкової величини, який описується диференціальною функцією

$$f(x) = \frac{1}{\sigma\sqrt{2\Pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \qquad (5.2.1)$$

де *a* та *σ* – *параметри* розподілу.

Можна довести, що параметр *a* є математичним сподіванням випадкової величини *X*, тобто *a* = *M(x)*, а параметр *σ* - середнім квадратичним відхиленням цієї випадкової величини, тобто $\sigma = \sigma(x) = \sqrt{D(x)}$.

Дійсно, за означенням математичне сподівання неперервної випадкової величини

$$M(x) = \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} \frac{x}{\sigma\sqrt{2\Pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

Введемо нову змінну $z = \frac{x-a}{\sigma}$, звідси *x=σz+a, dx=σdz*. Враховуючи, що нові межі інтегрування дорівнюють старим, отримаємо

$$M(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma z e^{\frac{-z^2}{2}} dz + \frac{a}{2\pi} \int_{-\infty}^{\infty} e^{\frac{-z^2}{2}} dz.$$

Перший із доданків дорівнює нулю, оскільки підінтегральна функція – непарна та межі інтегрування симетричні відносно початку координат; другий доданок $\int_{-\infty}^{\infty} e^{\frac{-z^2}{2}} dz = \sqrt{2\pi}$ (інтеграл Пуассона). Отже, *M(x)=a.*

$$D(X) = \int_{-\infty}^{\infty} (x - M(X))^2 \cdot f(x) dx = \frac{1}{\sigma\sqrt{\pi}} \int_{-\infty}^{\infty} (\mathrm{x} - \mathrm{a})^2 \cdot \mathrm{e}^{\frac{-(\mathrm{x}-\mathrm{a})^2}{2\sigma^2}} dx.$$

Введемо нову змінну $z=\frac{x-a}{\sigma}$, звідси *x=σz+a, dx=σdz*. Враховуючи, що нові межі інтегрування дорівнюють старим, отримаємо

$$D(x) = \int_{-\infty}^{\infty} ze^{\frac{-z^2}{2}}dz = ze^{\frac{-z^2}{2}} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{\frac{-z^2}{2}}dz.$$

Перший із доданків дорівнює нулю,так як підінтегральна функція – непарна та межі інтегрування симетричні відносно початку координат; другий доданок $\int_{-\infty}^{\infty} e^{\frac{-z^2}{2}}dz = \sqrt{2\pi}$ (інтеграл Пуассона). Звідси $D(x)=\sigma^2$.

Отже, якщо відомо, що розподіл імовірностей випадкової величини *X* нормальний, то досить знати лише дві числові характеристики цієї випадкової величини: *a* та *σ*, щоб повністю описати розподіл імовірностей.

Зазначимо, що в розглядуваному випадку математичне сподівання випадкової величини *X* є водночас і модою, і медіаною розподілу ймовірностей.

Більшість випадкових неперервних величин, що зустрічаються в економічній практиці, підкоряться закону нормального розподілу ймовірностей. Наприклад: урожайність, удійність, кількість опадів. Нормальний розподіл імовірностей використовується в математичній статистиці, теорії похибок вимірювань, теорії стрільби тощо. У локальній та інтегральній теоремах Муавра-Лапласа теж маємо справу з нормальним розподілом випадкової величини.

Розглянемо загальний і нормований закони розподілу випадкової величини.

Загальним називається закон нормального розподілу випадкової величини з довільними параметрами *a* і σ . Диференціальною функцією загального закону нормального розподілу є функція (5.2.1).

Нормованим називається закон розподілу випадкової величини з параметрами *a=0* і σ =1. Диференціальна функція нормованого закону нормального розподілу випадкової величини *X* має вид

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \qquad (5.2.2)$$

Диференціальна функція (5.2.2) нормованого закону нормального розподілу збігається з функцією *φ(x)* з локальної теореми Лапласа, значення якої знаходять за спеціальною таблицею. Функція *φ(x)* парна, тобто *φ(-x)= φ(x)*.

Інтегральна функція нормованого закону нормального розподілу має вид

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx.. \qquad (5.2.3)$$

Однак знайти точне значення $\int e^{-\frac{x^2}{2}} dx$ не вдається, тому для функції $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{x^2}{2}} dx$, яка має назву функції Лапласа (з інтегральної теореми), складено спеціальні таблиці. Функція $\Phi(x)$ непарна, тобто $\Phi(-x) = -\Phi(x)$.

Якщо випадкова величина *X* підпорядковується загальному нормальному розподілу з параметрами *a* і *σ*, то випадкова величина підкоряється нормованому нормальному розподілу з параметрами *a = 0, a(U)* =1, де $u = \frac{x-a}{\sigma}$. Дійсно,

$$M(u) = M\left(\frac{x-a}{\sigma}\right) = \frac{1}{\sigma}(M(x) - M(a) = \frac{1}{\sigma}(\,a - a) = 0,$$

$$\sigma(u) = \sqrt{D(u)} = 1.$$

5.3. Ймовірність попадання нормально розподіленої випадкової величини у заданий інтервал

Відомо, що якщо неперервна випадкова величина *X* задана диференціальною функцією розподілу *f(x)*, то ймовірність того, що значення цієї випадкової величини попадає в інтервал $(\alpha;\beta)$, дорівнює

$$\mathbf{P}(\alpha < x < \beta) = \int_{\alpha}^{\beta} f(x)dx.$$

Для нормального розподілу випадкової величини

$$\mathbf{P}(\alpha < x < \beta) = \frac{1}{\sigma\sqrt{2\Pi}}\int_{\alpha}^{\beta} e^{-\frac{(x-a)^2}{2\sigma^2}}dx.$$

Перетворимо цей вираз. Введемо нову змінну $z = \frac{x-a}{\sigma}$, тоді *x=σz + a, dx = σdz.* Перейдемо до нових меж інтегрування для змінної z: якщо $x = \alpha$, то $z = \frac{\alpha - a}{\sigma}$; якщо $x = \beta$, то $z = \frac{\beta - a}{\sigma}$;

$$P(\alpha < x < \beta) = \frac{1}{\sigma\sqrt{2\pi}}\int_{\frac{\alpha-a}{\sigma}}^{\frac{\beta-a}{\sigma}} e^{-\frac{z^2}{2}}\sigma dz.$$

Відповідно до властивостей визначеного інтеграла маємо:

$$P(\alpha < x < \beta) = \frac{1}{\sqrt{2\pi}}\int_{\frac{\alpha-a}{\sigma}}^{0} e^{-\frac{z^2}{2}}dz + \frac{1}{\sqrt{2\pi}}\int_{\frac{\alpha-a}{\sigma}}^{0} e^{-\frac{z^2}{2}}dz = \frac{1}{\sqrt{2\pi}}\int_{0}^{\frac{\beta-a}{\sigma}} e^{-\frac{z^2}{2}}dz - \frac{1}{\sqrt{2\pi}}\int_{0}^{\frac{\alpha-a}{\sigma}} e^{-\frac{z^2}{2}}dz.$$

Перший доданок – функція Лапласа *Ф(x)* за аргументом $x = \frac{\alpha - a}{\sigma}$; другий доданок - функція Лапласа за аргументом $x = \frac{\beta - a}{\sigma}$.

Отже, ймовірність попадання нормально розподіленої випадкової величини в заданий інтервал дорівнює

$$P(\alpha < x < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right). \qquad (5.3.1)$$

Приклад 5.3.1. Відоме математичне сподівання *a*=2 і середнє квадратичне відхилення $\sigma = 5$ нормально розподіленої випадкової величини *X*. Знайти ймовірність попадання цієї величини в інтервал (4; 9).

Розв'язання. Відомо (5.3.1), що $P(\alpha < x < \beta) = \Phi(\frac{\beta - a}{\sigma}) - \Phi(\frac{\alpha - a}{\sigma})$.

$\alpha = 4,\ \beta = 9,\ \ a = 2,\ \sigma = 5$. Тому

$P(4 < x < 9) = \Phi(\frac{9-2}{5}) - \Phi(\frac{4-2}{5}) = \Phi(1{,}4) - \Phi(0{,}4) = 0{,}4192 - 0{,}1554 = 0{,}2638$.

Приклад 5.3.2. Строк служби приладу являє собою випадкову величину, що підпорядковується нормальному закону розподілу, з гарантією на 15 років і середнім квадратичним відхиленням, що становить 3 роки. Знайти ймовірність того, що прилад прослужить від 10 до 20 років.

Розв'язання. За умовою задачі *a*=15, середнє квадратичне відхилення $\sigma = 3$, $\alpha = 10,\ \ \beta = 20$. Застосувавши формулу (5.3.1) і врахувавши непарність функції Лапласа *Ф(х)*, отримаємо:

$P(10 < x < 20) = \Phi(\frac{20-15}{3}) - \Phi(\frac{10-15}{3}) = \Phi(1{,}6667) - \Phi(-1{,}6667) = 2\Phi(1{,}6667)$.

З таблиці функції Лапласа *Ф(х)* знаходимо, що *Ф(х)*=0,4525, тому шукана ймовірність $P(10 < x < 20) = 0{,}9050$.

5.4. Нормальна крива

Побудуємо графік диференціальної функції $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ нормального розподілу, як його називають *нормальною кривою* або *кривою Гаусса*. Для цього дослідимо її методами диференціального числення.

1. Очевидно, що функція визначена на всій числовій осі.

2. При всіх значеннях *х,* які належать області визначення, функція набуває додатного значення, тобто нормальна крива розміщена над віссю *ОХ.*

3. $\lim\limits_{|x|\to\infty} y = 0,$ тобто вісь *OX* є горизонтальною асимптотою графіка диференціальної функції.

4. Знайдемо екстремум функції :

$y' = \dfrac{x-a}{\sigma^2\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}.;\; y' = 0,$ при *x*=*a*;

$y' > 0$ при *x* < *a*, $y' < 0$ при *x* >*a*.

Отже, при *x*=*a* функція має один максимум $y_{max} = \dfrac{1}{\sigma\sqrt{2\pi}}$.

5. Вираз *x-a* входить до аналітичного виразу функції у квадраті, тому графік функції симетричний відносно прямої *x=a.*

6. Знайдемо точки перегину:

$$y'' = (1 - \frac{(x-a)^2}{\sigma^2}).\,(-\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2}}).$$

При *x=a+σ* та *x=a-σ* друга похідна дорівнює нулю, а при переході через ці точки вона змінює знак. Таким чином, точки з координатами $(a-\sigma, \dfrac{1}{\sigma\sqrt{2\pi e}})$; $(a+\sigma, \dfrac{1}{\sigma\sqrt{2\pi e}})$ є точками перегину графіка функції.

7. Зміна математичного сподівання *a* при σ=const призводить до зміщення кривої Гаусса вздовж осі *X*.

Побудуємо графік диференціальної функції нормального розподілу (рис. 5.4.1).

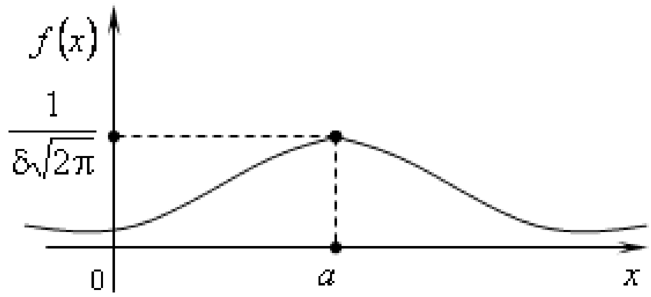


Рис. 5.4.1

При a=const зі збільшенням σ максимальна ордината нормальної кривої спадає, а сама крива стає більш пологою, тобто близько підходить до осі *OX*. При спаданні σ нормальна крива стає більш гостроверхою та розтягується в додатному напряму осі *OУ*. На рис. 5.4.2 для $a=0$ крива I відповідає σ_1 =2,5; крива II – σ_2 =1, а для кривої III – σ_3 = 0,4, тобто $\sigma_1 > \sigma_2 > \sigma_3$.

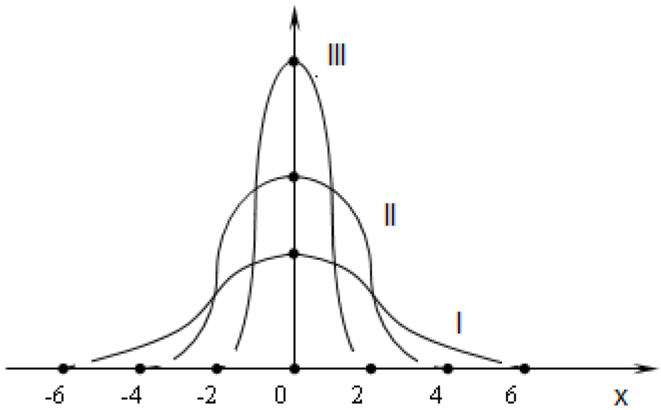


Рис. 5.4.2

Нормальний закон розподілу – це граничний закон, до якого наближаються інші закони розподілу при типових умовах, що часто зустрічаються. Сума великої кількості незалежних або слабо залежних випадкових величин, що розподілені за довільними законами, наближено підпорядковуються нормальному закону. Для цього повинна лише виконуватися умова, яка полягає в тому, щоб окремі доданки в сумі мали приблизно однакову значущість. У противному випадку в сумарному розподілі буде домінувати розподіл найбільш значущої складової.

5.5. Обчислення ймовірності даного відхилення

Часто потрібно обчислити ймовірність того, що відхилення нормально розподіленої випадкової величини *X* за абсолютною величиною менше від заданого додатного числа δ , тобто треба знайти ймовірність виконання нерівності $|X-a|<\delta$.

З даної нерівності маємо: $-\delta < X - a < \delta$ або $a - \delta < X < a + \delta$. Скористаємося формулою ймовірності попадання нормально розподіленої випадкової величини в заданий інтервал:

$$P(|x-a| \le \delta) = P(a-\delta < x < a+\delta) = \Phi(\frac{a+\delta-a}{\sigma}) - \Phi(\frac{a-\delta-a}{\sigma}) = \Phi(\frac{\delta}{\sigma}) - \Phi(-\frac{\delta}{\sigma}) = 2\Phi(\frac{\delta}{\sigma}).$$

Отже, ймовірність того, що відхилення випадкової нормально розподіленої величини *X* від її математичного сподівання за абсолютною величиною менше, ніж δ (δ>0), обчислюється за формулою:

$$P(|X-a| \le \delta) = 2\Phi(\frac{\delta}{\sigma}) \qquad (5.5.1)$$

Приклад 5.5.1. Задані математичне сподівання *a* = 7 і середнє квадратичне відхилення σ = 3 нормально розподіленої випадкової величини *X.* Знайти ймовірність того, що абсолютна величина відхилення │*x-a*│ буде менша за 3.

Розв'язання*.* Відомо (5.5.1), що $P(|X-a| \le \delta) = 2\Phi(\frac{\delta}{\sigma})$.

P(│*x*-7│)<3 = 2*Φ(*3/3) = 2*Φ*(1) = 2·0,3413 = 0,682.

Приклад 5.5.2. Розподіл маси консервних банок, що виробляються на заводі, розподіляється за нормальним законом із середньою масою 250 г і стандартним відхиленням 5 г. Знайти ймовірність того, що відхилення маси банок від середньої маси за абсолютною величиною не перевищить 8 г.

Розв'язання. За умовою задачі *a*=250, середнє квадратичне відхилення $\sigma = 5$, $\delta = 8$ *.* Потрібно знайти ймовірність події $|X-250| \le 8$. Застосувавши формулу(5.5.1), отримаємо:

$$P(|X-250| \le 8) = 2\Phi(\frac{8}{5}) = 2\Phi(1{,}6) = 2 \cdot 0{,}4452 = 0{,}8904.$$

Приклад 5.5.3. Довжина деталі, що виготовляється в цеху, являє собою випадкову величину, яка розподілена нормально, з математичним сподіванням (номінальним розміром), що дорівнює

12 мм, і середнім квадратичним відхиленням 0,06 мм. Знайти ймовірність того, що довжина готової деталі відхиляється від номінального розміру не більш ніж на 0,02 мм.

Розв'язання*.* Шукану ймовірність знайдемо за формулою (5.5.1).

$$P(|x-12|<0,02)=2\Phi\left(\frac{0,02}{0,06}\right)=2\Phi(0,333)=2\cdot 0,1293=0,2586.$$

5.6. Правило трьох сигм

Перетворимо формулу (5.5.1), поклавши $\delta=\sigma t$. Отримаємо $P(|x-a|<\sigma t)=2\Phi(t)$. Якщо $t=3$, то $\sigma t=3\sigma$ і $P(|X-a|<3\sigma)=2\Phi(3)=0,9973$; тобто якщо випадкова величина X розподілена нормально, то ймовірність того, що абсолютна величина її відхилення від математичного сподівання не перевищує потрійного середнього квадратичного відхилення, дорівнює 0,9973.

Інакше, ймовірність того, що абсолютна величина відхилення X від її математичного сподівання перевищує потроєне середнє квадратичне відхилення, дуже мала, а саме рівна 0,0027 (такі події практично неможливі).

Правило трьох сигм. Якщо випадкова величина розподілена нормально, то ймовірність того, що абсолютна величина її відхилення від математичного сподівання не більша від потроєного середнього квадратичного відхилення, - практично достовірна подія з імовірністю 0,9973.

На практиці правило трьох сигм використовується тоді, коли розподіл досліджуваної випадкової величини невідомий, але умова, що вказана у правилі, виконується. Тоді можна вважати, що досліджувана випадкова величина розподілена за нормальним законом.

Для випадкової величини *X*, що розподілена нормально, інтервалу *а* ±3σ відповідає 99,73% площі під нормальною кривою (рис. 5.6.1). Тобто: ймовірність того, що *X* знаходиться в інтервалі $a - 3\sigma < x < a + 3\sigma$, дорівнює 0,9973. Інакше: практично все розсіювання ймовірностей при нормальному розподілі випадкової величини *X* припадає на проміжок $(a - 3\sigma; a + 3\sigma)$.

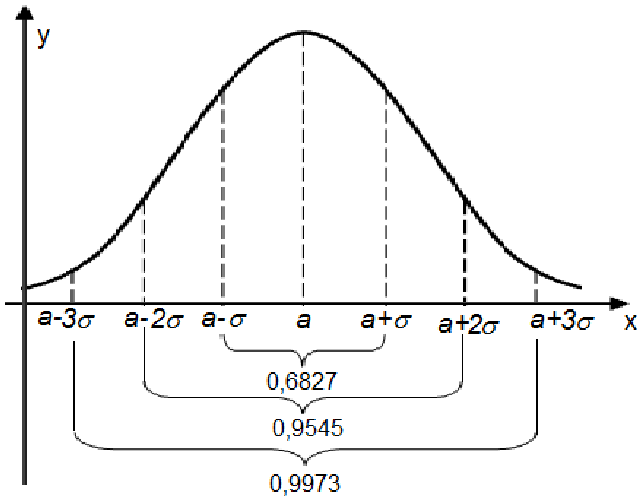


Рис. 5.6.1

5.7. Показниковий розподіл та його числові характеристики

Показниковим (експоненціальним) називають розподіл ймовірностей неперервної випадкової величини *X*, який описується щільністю розподілу

$$f(x)= \begin{cases} 0 & \text{для } x < 0; \\ \lambda e^{-\lambda x} & \text{для } x \geq 0, \end{cases} \tag{5.7.1}$$

де *λ* - стала додатна величина. Отже, показниковий розподіл визначається одним параметром *λ,* в цьому і є його перевага перед іншими розподілами, які залежать від багатьох параметрів.

Параметр λ має наступний статистичний зміст: *λ* є середнє число подій за одиницю часу, тобто 1/ *λ* є середній проміжок часу між двома послідовними подіями.

Прикладом неперервної випадкової величини, яка розподілена за показниковим законом і використовується в теорії масового обслуговування, може бути час між появою двох послідовних подій простої течії (час телефонних розмов, ремонту техніки, безвідмовної роботи технічних приладів, час затримки вильоту літака за вини технічних служб аеропорту тощо). Показниковий розподіл часто зустрічається також у теорії надійності (наприклад, строк служби електронної апаратури).

Знайдемо функцію розподілу $F(x)$ показникового закону. Якщо *x*≤0, то $F(x) = 0.$ Якщо *x*>0, то $f(x) = \lambda e^{-\lambda x}$, і тоді

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

Таким чином:

$$F(x) = \begin{cases} 0 \text{ для } x < 0; \\ 1 - e^{-\lambda x} \text{ для } x \geq 0. \end{cases} \quad (5.7.2)$$

Знайдемо ймовірність попадання в інтервал (*a,b)* неперервної випадкової величини X, яка розподілена за показниковим законом. Шукана ймовірність

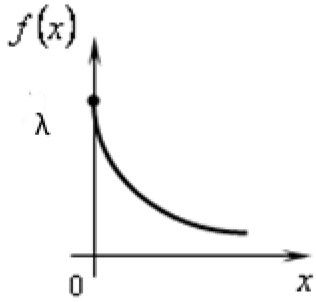
$$P(a < X < b) = F(b) - F(a) = 2e^{-0,6}\left(1 - e^{-\lambda b}\right) - \\ -\left(1 - e^{-\lambda a}\right) = e^{-\lambda b} - e^{-\lambda a}. \quad (5.7.3)$$

Значення функції $e^{-\lambda}$ можна взяти з таблиць значень цієї функції, *e*≈2,718281828… – основа натуральних логарифмів.

Покажемо графіки щільності $f(x)$ та функції розподілу $F(x)$ показникового закону (рис. 5.7.1; 5.7.2).

Приклад 5.7.1. Неперервна випадкова величина *X*, яка розподілена за показниковим законом, задана щільністю розподілу

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



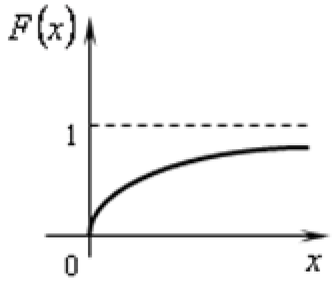


Рис. 5.7.1 Рис. 5.7.2

Знайти ймовірність того, що результаті випробувань *X* попадає в інтервал (0,3; 1).

Розв'язання. За умовою λ=-2.

$P(0{,}3 < X < 1) = F(1) - F(0{,}3) = 2e^{-0{,}6} - 2e^{-2} = 0{,}41$.

Знайдемо *числові характеристики показникового розподілу*.

Нехай неперервна випадкова величина описується щільністю розподілу

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

Математичне сподівання $M(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{\infty} xf(x)dx =$

$$= \lambda \int_0^{\infty} xe^{-\lambda x}dx = \lambda(-x\frac{e^{-\lambda x}}{\lambda}|_0^{\infty} + \int_0^{\infty} \frac{e^{-\lambda x}}{\lambda}dx) = -\frac{e^{-\lambda x}}{\lambda}|_0^{\infty} = \frac{1}{\lambda}.$$

Таким чином, математичне сподівання показникового розподілу дорівнює оберненій величині параметра λ.

$$M(X) = \frac{1}{\lambda}. \qquad (5.7.4)$$

Обчислимо дисперсію: $D(X) = \int_0^{\infty} x^2 f(x)dx - (M(X))^2 =$

$$= \lambda \int_0^{\infty} xe^{-\lambda x}dx = \lambda(-x\frac{e^{-\lambda x}}{\lambda}|_0^{\infty} + \int_0^{\infty} \frac{e^{-\lambda x}}{\lambda}dx) = -\frac{e^{-\lambda x}}{\lambda}|_0^{\infty} = -2x\frac{e^{-\lambda x}}{\lambda}|_0^{\infty} + \frac{2}{\lambda}\int_0^{\infty} e^{-\lambda x}dx - \frac{1}{\lambda^2} = -\frac{2}{\lambda^2}e^{-\lambda x}|_0^{\infty} - \frac{1}{\lambda^2} =$$

$$\frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Отже,

$$D(X) = \frac{1}{\lambda^2}. \quad (5.7.5)$$

Середнє квадратичне відхилення

$$\sigma(X) = \sqrt{\mathrm{D(X)}} = \sqrt{\frac{1}{\lambda^2}} = \frac{1}{\lambda}, \quad (5.7.6)$$

Звідси видно, що при показниковому розподілі неперервної випадкової величини $M(X) = \sigma(X) = \frac{1}{\lambda}$.

5.8. Інші види розподілів

Мають місце й інші види розподілів випадкових величин.

- *Розподілу Пуассона* або закону рідкісних явищ підпорядковуються події, ймовірність яких при кожному окремому випробуванні дуже мала, а кількість випробувань – велика. Ймовірність випадкової величини *X*, що розподілена за пуассонівський законом, як відомо, обчислюється за формулою

$$P(x) \approx \frac{a^x \cdot e^{-a}}{x\,!} \quad (5.8.1)$$

де *a* – математичне сподівання випадкової величини, e≈2,718281828… Важливою властивістю розподілу Пуассона є та, що його дисперсія дорівнює математичному сподіванню:

$$D(x) = M(x) = a \quad (5.8.2)$$

Ця властивість дозволяє на практиці перевіряти справедливість гіпотези про те, що розглядувана випадкова величина розподілена за пуассонівським законом.

- *Гамма-розподіл*. Випадкова величина *X* має гамма-розподіл з параметрами $a > 0$ та b> 0, якщо її щільність розподілу ймовірностей має вигляд:

$$f(x)=\begin{cases}0, & x\le 0,\\ \dfrac{b^a}{\Gamma(a)}\cdot x^{a-1}e^{-bx}, & x>0\end{cases} \qquad (5.8.3)$$

де $\Gamma(a)$ - гамма-функція Ейлера $\Gamma(\mathrm{a})=\int_0^\infty \mathrm{t}^{\mathrm{a}-1}\cdot \mathrm{e}^{-\mathrm{t}}\mathrm{dt}$.

Математичне сподівання і дисперсія випадкової величини, яка розподілена за гамма-розподілом, обчислюються за формулами:

$$M(X)=\frac{a}{b},\ \ D(X)=\frac{a}{b^2}. \qquad (5.8.4)$$

При $a>1$ гамма-розподіл має моду $Mo(X)=\frac{a-1}{b}$; це графічно означає, що крива розподілу має точку максимуму при $x=Mo$.

Частинним випадком гамма-розподілів є показниковий розподіл при $a=1$ і $\lambda=\frac{1}{b}$. При $a\in N$ гамма-розподіл називається розподілом Ерланга (К. Ерланг (1878-1929) – датський учений), з робіт якого почався розвиток теорії масового обслуговування, що займається ймовірнісно-статистичним моделюванням систем, в яких відбувається обслуговування течії заявок, з метою прийняття оптимальних рішень. Розподіл Ерланга використовують також у тих прикладних галузях, де мають місце показникові розподіли. Нормальний розподіл – це граничний випадок гамма-розподілу.

- *Розподіл* χ^2 *(хі-квадрат Пірсона)*: це розподіл випадкової величини $X=X_1^2+X_2^2+\cdots+X_n^2$, де X_i (*i*=1, 2, …, *n*) – незалежні нормальні, нормовані величини, які мають один і той же розподіл *N*(0,1), тобто їхнє математичне сподівання дорівнює 0, а середнє квадратичне відхилення дорівнює 1. Тоді сума квадратів $\chi^2=\sum_{i=1}^n X_i^2$ розподілена за законом χ^2 з параметром *k=n ступенями вільностей*. При зростанні *k* розподіл хі-квадрат прямує до нормального дуже повільно.

Диференціальна функція розподілу χ^2 має вигляд:

$$f(x) = \begin{cases} 0 \quad \text{при } x \le 0, \\ \frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} \cdot e^{-\frac{x}{2}} \cdot x^{\frac{k}{2}-1} \quad \text{при } x > 0. \end{cases} \quad (5.8.5)$$

де $\Gamma(x) = \int_0^\infty t^{x-1} \cdot e^{-t} dt$ – гамма-функція, $\Gamma(n+1) = n!$, тобто розподіл χ^2 є частинним випадком гамма-розподілу при $x = \frac{k}{2}$ $(n \in N)$ і $b = \frac{1}{2}$.

Розподіл хі-квадрат використовується для оцінки дисперсії (за допомогою довірчого інтервалу), при перевірці гіпотез згоди, однорідності, незалежності, насамперед для якісних змінних, що приймають скінчену кількість значень, та в інших задачах статистичного аналізу даних.

- Розподіл *t* Стьюдента. Нехай *X* – нормальна нормована випадкова величина *N*(0,1), а *Y* – незалежна від *X* величина, що розподілена за законом χ^2 з *k* ступенями волі. Тоді величина

$$T = \frac{X}{\sqrt{\frac{y}{k}}} \quad (5.8.6)$$

має розподіл, який називають t-розподілом або розподілом Стьюдента (псевдонім англійського статистика В Госсета, який розробляв ймовірнісно-статистичні методи для прийняття економічних і технічних рішень). При зростанні *k* розподіл Стьюдента швидко наближається до нормального розподілу.

Розподіл Стьюдента найчастіше використовують у задачах математичної обробки результатів вимірювань, для аналізу даних: при оцінюванні математичного сподівання, прогнозного значення та інших характеристик за допомогою довірчих інтервалів, для перевірки гіпотез про значення математичних сподівань, коефіцієнтів регресійної залежності, гіпотез однорідності вибірок тощо.

- Розподіл *F* Фішера – це розподіл випадкової величини

$$F = \frac{\frac{1}{k_1}X_1}{\frac{1}{k_2}\chi_2} \quad (5.8.7)$$

де випадкові величини X_1 та X_2 - незалежні та мають розподіл χ^2 з числом ступенів вільностей k_1 та k_2відповідно.

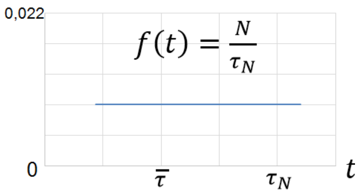
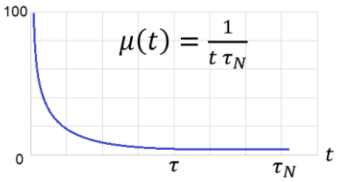
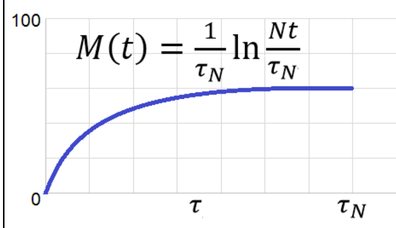
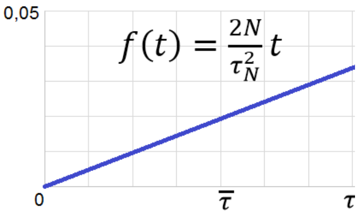
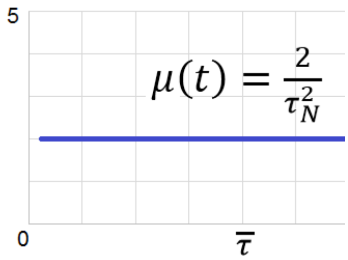
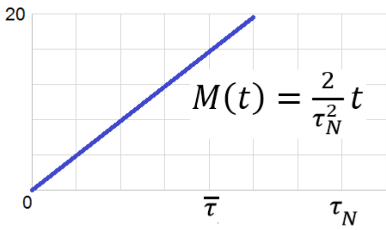
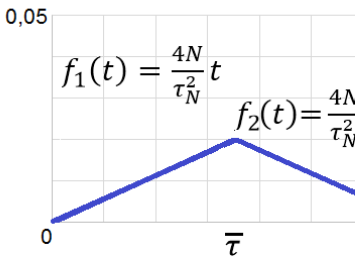
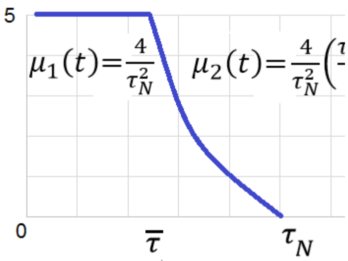
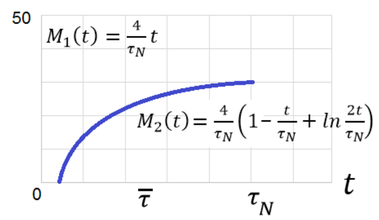
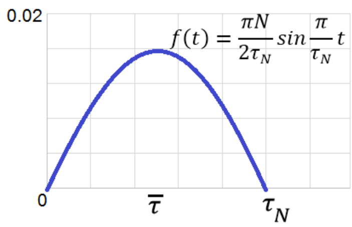
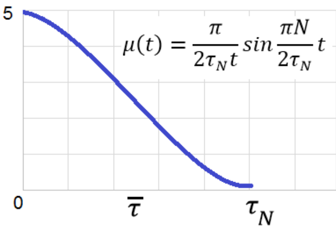
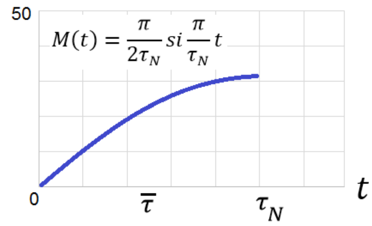
Розподіл Фішера використовують для перевірки гіпотез про адекватність моделі в регресійному аналізі, про рівність дисперсій тощо.

Для випадкових величин, що мають нормальний розподіл, розподіли Пірсона, Фішера, Стьюдента, мають місце такі співвідношення:

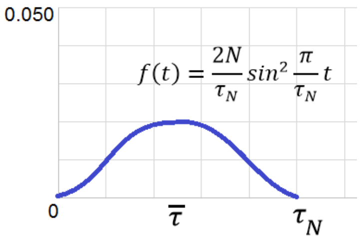
$$\chi^2(1) = X^2; \;\; F(n;\infty) = \frac{\chi^2(n)}{n}; \;\;\; T^2 = F(1;n).$$

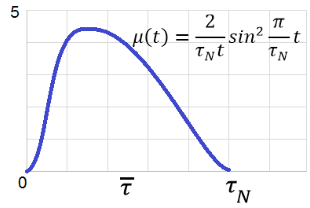
Ці та інші види розподілів широко використовуються у технічній, економічній практиці. Для них складені спеціальні таблиці, які необхідні для практичного використання і які можна знайти у спеціальній літературі. Дослідження розподілів випадкових величин – один із основних етапів у розв’язуванні задач імовірнісних і статистичних прогнозів та передбачень. Так, наприклад, пуассонівський розподіл відіграє важливу роль при вирішенні питань масового обслуговування. Велике практичне значення має аналіз розподілів при передбаченні відмов у теорії надійності (табл. 5.8.1) тощо.

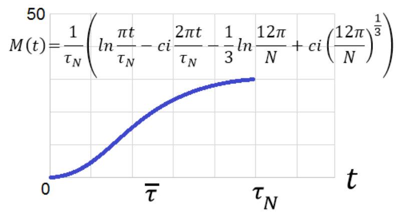
Таблиця 5.8.1

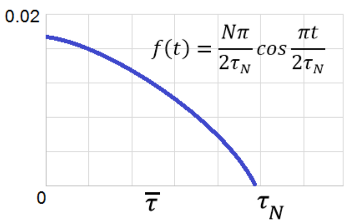
Характеристика / Закон	Функція щільності розподілу	Диференціальна частота відмов	Середня частота відмов
Рівномірний	 $f(t)=\frac{N}{\tau_N}$	 $\mu(t)=\frac{1}{t\tau_N}$	 $M(t)=\frac{1}{\tau_N}\ln\frac{Nt}{\tau_N}$
Лінійний	 $f(t)=\frac{2N}{\tau_N^2}t$	 $\mu(t)=\frac{2}{\tau_N^2}$	 $M(t)=\frac{2}{\tau_N^2}t$
Трикутний	 $f_1(t)=\frac{4N}{\tau_N^2}t$ $f_2(t)=\frac{4N}{\tau_N^2}(\tau_N-1)$	 $\mu_1(t)=\frac{4}{\tau_N^2}$ $\mu_2(t)=\frac{4}{\tau_N^2}\left(\frac{\tau_N}{t}-1\right)$	 $M_1(t)=\frac{4}{\tau_N}t$ $M_2(t)=\frac{4}{\tau_N}\left(1-\frac{t}{\tau_N}+ln\frac{2t}{\tau_N}\right)$
Синусоїдний	 $f(t)=\frac{\pi N}{2\tau_N}sin\frac{\pi}{\tau_N}t$	 $\mu(t)=\frac{\pi}{2\tau_N t}sin\frac{\pi N}{2\tau_N}t$	 $M(t)=\frac{\pi}{2\tau_N}si\frac{\pi}{\tau_N}t$

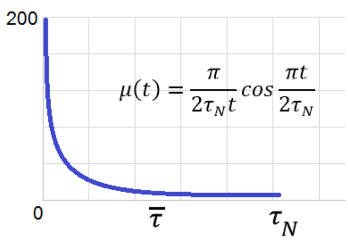
Таблиця 5.8.1 (продовження)

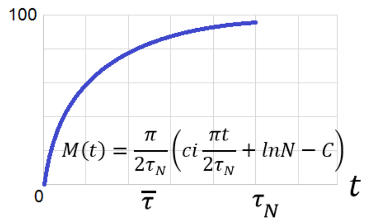
<table>
<tr>
<td>Sin^2</td>
<td>

$$f(t) = \frac{2N}{\tau_N} sin^2 \frac{\pi}{\tau_N} t$$</td>
<td>

$$\mu(t) = \frac{2}{\tau_N t} sin^2 \frac{\pi}{\tau_N} t$$</td>
<td>

$$t\ M(t) = \frac{1}{\tau_N}\left(ln\frac{\pi t}{\tau_N} - ci\frac{2\pi t}{\tau_N} - \frac{1}{3}ln\frac{12\pi}{N} + ci\left(\frac{12\pi}{N}\right)^{\frac{1}{3}}\right)$$</td>
</tr>
<tr>
<td>Косинусоїдний</td>
<td>

$$f(t) = \frac{N\pi}{2\tau_N} cos \frac{\pi t}{2\tau_N}$$</td>
<td>

$$\mu(t) = \frac{\pi}{2\tau_N t} cos \frac{\pi t}{2\tau_N}$$</td>
<td>

$$M(t) = \frac{\pi}{2\tau_N}\left(ci\frac{\pi t}{2\tau_N} + lnN - C\right)$$</td>
</tr>
</table>

Запитання для самоперевірки

1. Який розподіл ймовірностей випадкової величини називається рівномірним?

2. Диференціальна та інтегральна функції рівномірного розподілу.

3. Числові характеристики рівномірно розподіленої випадкової величини.

4. Нормальний розподіл та його параметри.

5. Ймовірнісний зміст параметрів *а* та σ нормального закону розподілу.

6. Чому дорівнює ймовірність попадання нормально

розподіленої випадкової величини в заданий інтервал?

7. Формула обчислення ймовірності даного відхилення.

8. Нормальна крива та її властивості.

9. Сформулювати правило трьох сигм. Показати графічно.

10. Показниковий розподіл та його числові характеристики.

11. Де застосовуються нормальний розподіл випадкової величини та інші розподіли (нормальний, експоненціальний, Пірсона, Фішера, Стьюдента та ін.)?

Задачі та вправи

5.1. Рівномірно розподілена величина *X* задана на інтервалі (-4; 4). Знайти:а) щільність даного розподілу; б) математичне сподівання та дисперсію.

5.2. Випадкова величина *X* задана щільністю розподілу *f(x)*=1/3 в інтервалі (0; 3), зовні інтервалу *f(x)*=0. Знайти математичне сподівання, дисперсію та середнє квадратичне відхилення даного розподілу.

5.3. Рівномірно розподілена випадкова величина задана функцією розподілу

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

Знайти: а) ймовірність того, що в результаті випробування *X* прийме значення, яке належить інтервалу (0,25; 0,4); б) математичне сподівання та дисперсію.

5.4. Випадкова величина задана щільністю розподілу *f(x)*=2cos2x в інтервалі $(0; \pi$ /4); зовні цього інтервалу *f(x)*=0. Знайти математичне сподівання та дисперсію.

5.5. Задані математичне сподівання *a*=12 і середнє квадратичне відхилення σ=8 нормально розподіленої випадкової величини *X*. Знайти ймовірність того що:

а) *X* прийме значення, яке належить інтервалу (8; 17);

б) абсолютне відхилення $|x\text{-}a|$ буде менше δ; $\delta = 6$.

5.6. Випадкова величина підлягає нормальному закону з математичним сподіванням *a*=0. Ймовірність попадання цієї величини в інтервал (-3; 3) рівна 0,5. Знайти середнє квадратичне відхилення σ і записати вираз функції щільності ймовірностей *f(x).*

5.7. Середній діаметр стовбурів дерев на деякій ділянці дорівнює 30 см, середнє квадратичне відхилення – 5 см. Враховуючи, що діаметр стовбура – величина випадкова, що розподілена нормально, визначити ймовірність того, що: а) діаметр стовбура навмання відібраного дерева попаде в інтервал від 25 до 37 см; б) діаметр стовбура дерева не менше 25 см; в) діаметр стовбура дерева не перевищить 36 см; г) визначити величину, яку не перевищить діаметр стовбура навмання відібраного дерева з імовірністю 0,95.

5.8. Керуючий банком установив, що відрізок часу, який затрачується клієнтом банку в черзі, розподілений за нормальним законом із середнім значенням 4 хвилини та стандартним відхиленням 1 хвилина. Визначити ймовірність того, що навмання вибраний клієнт стоїть у черзі більше 6 хвилин. Скільки осіб із 150 відвідувачів будуть чекати в черзі більше 6 хвилин?

5.9. Випадкова величина *X* задана інтегральною функцією*:*

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

Знайти математичне сподівання та дисперсію.

5.10. Рівномірно розподілена випадкова величина *X* задана щільністю розподілу

$$f(x) = \begin{cases} 0, \text{ при x } \leq 2, \\ \frac{1}{2}, \quad \text{при } 2 < \text{x} \leq 4, \\ 0, \text{ при x } > 4. \end{cases}$$

Знайти ймовірність того, що в результаті випробування *X* прийме значення, що належить інтервалу: а) (2; 3); б) (3; 5).

5.11. Випадкова величина *X* задана диференціальною функцією розподілу

$$f(x) = \begin{cases} 0, \text{ при x } \leq 0, \\ \frac{1}{4} sin \frac{1}{2} x, \text{ при } 0 < \text{x} \leq 2\pi, \\ 0, \text{ при x} > 2\pi. \end{cases}$$

Знайти математичне сподівання та дисперсію.

5.12. Задані математичне сподівання $a = 11$ і середнє квадратичне відхилення $\sigma = 5$ нормально розподіленої випадкової величини *X*. Знайти ймовірність того, що:

а) *X* прийме значення, яке належить інтервалу (8; 17);

б) абсолютне відхилення $|x\text{-}a|$ буде менше δ, $\delta = 9$.

5.13. Випадкова величина підлягає нормальному закону з математичним сподіванням *a* = 0. Ймовірність попадання цієї випадкової величини в інтервал (-1; 1) рівна 0,5. Знайти середнє квадратичне відхилення і записати функцію щільності ймовірностей *f(x).*

5.14. Середня сума, яку тратить відвідувач універсаму, розподілена за нормальним законом із середнім значенням 56 грн. та стандартним відхиленням 12 грн. Навмання вибирається відвідувач магазину. Яка ймовірність того, що він витратив менше 30 грн.? Скільки людей із 1000 відвідувачів витратять суму що менша за 30 грн.?

5.15. Час, який співробітники витрачають на шлях до роботи, розподілений за нормальним законом із середнім значенням 40 хвилин і стандартним відхиленням 6 хвилин. Яка ймовірність, що навмання вибраний співробітник витрачає на дорогу більше 50 хвилин? Яка очікувана кількість із 250 співробітників витрачає на дорогу більше 50 хвилин?

5.16. Зріст пацієнтів деякої групи здоров'я є величина, яка розподілена за нормальним законом, із середнім значення 170 см і середнім квадратичним відхиленням 5 см. Фактично зріст пацієнтів групи здоров'я знаходиться в інтервалі від 155 до 180 см. Знайти процент пацієнтів, які мають зріст: а) більше 160 см; б) менше 175 см.

5.17. Середня вага зерна 0,2 г, середнє квадратичне відхилення дорівнює 0,05 г. Знайти ймовірність того, що вага навмання взятого зерна буде в межах від 0,16 до 0,22 г, якщо вага зернини розподілена за нормальним законом.

5.18. Маса яблука, середня вага якого 150 г, є нормально розподіленою випадковою величиною із середнім квадратичним відхиленням 20 г. Знайти ймовірність того, що вага навмання взятого яблука виявиться від130 г до180 г.

5.19. Середній розмір деталі дорівнює 200 мм, а середнє квадратичне відхилення 0,25 мм. Стандартними вважаються деталі, розмір яких знаходиться в межах від 199,5 мм до 200,5 мм. Знайти процент стандартних деталей, якщо відхилення розміру деталі від номіналу розподілені нормально.

5.20. За результатами замірів середня відстань між двома населеними пунктами 16 км із середнім квадратичним відхиленням 100 м. Знайти ймовірність того, що відстань між цими пунктами не менше 15,8 км.

5.21. Вага упаковок розфасованого цукру розподілена за нормальним законом із середньою вагою 10 кг та середнім квадратичним відхиленням 0,05 кг. Яка ймовірність того, що взята навмання упаковка важить більше 10,1 кг?

5.22. Середній діаметр стовбурів дерев на деякій ділянці дорівнює 25 см. а середнє квадратичне відхилення дорівнює 5 см. Вважаючи діаметр стовбура величиною розподіленою за нормальним законом, знайти відсоток дерев, які мають діаметр стовбура більше 20 см.

5.23. Маса яблука, середня вага якого 120 г, є нормально розподіленою випадковою величиною із середнім квадратичним відхиленням 15 г. Знайти ймовірність того, що вага навмання взятого яблука виявиться від 110 г до130 г.

5.24. Знайти математичне сподівання та дисперсію випадкової величини *X,* якщо вона описується щільністю розподілу

а)$f(x)=\begin{cases} 0 & \text{для } x \le 0; \\ 2e^{-2x} & \text{для } x > 0. \end{cases}$ б)$f(x)=\begin{cases} 0 & \text{для } x \le 0; \\ 0{,}1e^{-0{,}1x} & \text{для } x > 0. \end{cases}$

в)$f(x)=\begin{cases} 0 & \text{для } x \le 0; \\ 16e^{-16\lambda x} & \text{для } x > 0. \end{cases}$ г)$f(x)=\begin{cases} 0 & \text{для } x \le 0; \\ 25e^{-25x} & \text{для } x > 0. \end{cases}$

5.25. Неперервна випадкова величина *X* розподілена за законом, що задано диференціальною функцією

$$f(x)=\begin{cases} 0 \text{для } x \le 0; \\ 3e^{-3x} \text{для } x > 0. \end{cases}$$

Знайти ймовірність того, що *X* потрапить в інтервал: а) (0,13; 0,7); б) (0,8; 1,2); в) ($\frac{1}{3}, \frac{2}{3}$); г) (0,2; 1).

5.26. Випадкова величина *X* розподілена нормально з параметрами *a* =1 та σ =2. Знайти щільність розподілу $У = X^3$.

5.27. Випадкова величина *X* розподілена рівномірно в інтервалі (0; 2). Знайти дисперсію випадкової величини *У* = 3- 2*X.*

5.28. Похибка спостереження *X* при обчисленні довжини розподілена нормально з параметрами *a* =5 мм і σ =4 мм. Знайти імовірність того, що значення, яке вимірюється, відхиляється від дійсного більше ніж на 10 мм.

5.29. Відхилення розміру деталі від стандартного є випадкова величина, яка розподілена нормально з *M(X)*=0 та *Д(X)*=36. Скільки потрібно виготовити деталей, щоб з імовірністю не меншою за 0,95 можна було стверджувати, що серед них буде хоча б одна стандартна, якщо дозволено відхилення розміру деталі від стандарту в інтервалі (-2; 4)?

5.30. Знайти початкові та центральні моменти перших чотирьох порядків випадкової величини, яка розподілена з щільністю ймовірностей

$$f(x)=\begin{cases} 0 \text{для } x \le 0, \\ 3e^{-3x} \text{для } x > 0. \end{cases}$$

5.31. Для випадкової величини *X* з щільністю розподілу знайти початкові та центральні моменти перших чотирьох порядків, коефіцієнт асиметрії та ексцес розподілу, якщо

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

РОЗДІЛ 6
СИСТЕМИ ВИПАДКОВИХ ВЕЛИЧИН

Вище були розглянуті випадкові величини, які при кожному випробуванні визначались одним можливим числовим значенням. Таку випадкову величину називають одновимірною. На практиці результат випробування часто описується двома, трьома і більшим числом випадковими величинами.

Якщо можливі значення випадкової величини визначаються у кожному випробуванні двома, трьома, ..., n числами, то такі величини називають відповідно *дво-*, *три-, ..., n-вимірними* відповідно.

Двовимірну випадкову величину будемо позначати $(X, У)$, де X та $У$ при цьому будуть компонентами. Величини X та $У$, що розглядаються одночасно, утворюють систему двох випадкових величин. Аналогічно можна розглядати систему n випадкових величин.

Сукупність n одночасно розглянутих випадкових величин $(X_1, X_2, ..., X_n)$ називають *системою (комплексом) випадкових величин*.

Наведемо приклади багатовимірних випадкових величин.

1. Двічі кидають гральний кубик. Позначимо через X - число появи очок при першому киданні, а через $У$ – число появи очок при другому. Пара $(X,У)$ буде системою випадкових величин.

2. Погода в даній місцевості у певний час доби може бути охарактеризована такими випадковими величинами: X_1 - температура, X_2 - тиск, X_3 - вологість, X_4 - швидкість вітру і т.д.

3. Врожайність сільськогосподарської культури залежить від кількості внесених добрив (X_1), якості насіння (X_2), якості ґрунту

(X_3), погодних умов (X_4), якості обробки ґрунту (X_5), якості догляду за рослинами (X_6) тощо.

4. Зняття показів електроприладів (напруга, частота, споживана потужність) залежить від метеорологічних умов, температури, навантаження та багатьох інших факторів.

Систему *n* випадкових величин $(X_1, X_2, \dots, X_n)$ можна розглядати як випадкову точку в *n*-вимірному просторі з координатами $(X_1, X_2, \dots, X_n)$, або як випадковий вектор, що напрямлений з початку системи координат у точку $M(X_1, X_2, \dots, X_n)$.

При *n* = 2 маємо систему двох випадкових величин *(X, У),* яку можна зобразити як випадкову точку *M(X,У)* на площині *XOУ*,або як випадковий вектор, складові якого за осями є випадкові величини *X* та *У*.

При *n* = 3 маємо систему трьох випадкових величин *(X, У, Z),* яку можна зобразити як випадкову точку *M(X,У,Z) у* просторі *OXУZ*,або як випадковий вектор, складові якого за осями є випадкові величини *X, У, Z.*

Властивості системи випадкових величин не вичерпуються властивостями окремих її складових. Потрібно ще враховувати залежності, зв'язок між цими випадковими величинами

Багатовимірні випадкові величини бувають дискретними та неперервними (компоненти цих величин відповідно будуть дискретними та неперервними).

6.1. Закон розподілу ймовірностей дискретної двовимірної випадкової величини

Законом розподілу дискретної двовимірної випадкової величини називають перелік можливих значень цієї величини - пар значень (x_i, y_k) та їх імовірностей $p(x_i, y_k)$, де i=1, 2,…,n; k= 1,2,…,m.

Закон розподілу може бути заданий у різних формах. Найчастіше закон розподілу двовимірної дискретної випадкової величини задають у вигляді таблиці з двома входами (табл. 6.1.1). У першому рядку таблиці записують усі можливі значення компоненти *X.* У першому стовпчику таблиці записують усі можливі значення компоненти *У.* На перетині *k*-го рядка та *i*-го стовпця записують імовірність $p(x_i, y_k)$ того, що двовимірна випадкова величина (*X, Y*) прийме значення (x_i, y_k), де $i = 1, 2, ..., n$; $k = 1, 2, ..., m$.

Події ($X = x_i$, $Y = y_k$),$i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$, утворюють повну групу, тому сума ймовірностей таблиці дорівнює одиниці, тобто: $\sum_{i=1}^{n}\sum_{j=1}^{m} p\,(x_i, y_k) = 1.$

Таблиця 6.1.1

У	*X*					
	x_1	x_2	…	x_i	…	x_n
y_1	$P(x_1, y_1)$	$P(x_2, y_1)$	…	$P(x_i, y_1)$	…	$P(x_n, y_1)$
y_2	$P(x_1, y_2)$	$P(x_2, y_2)$	…	$P(x_i, y_2)$	…	$P(x_n, y_2)$
…	…	…	…	…	…	…
y_j	$P(x_1, y_k)$	$P(x_2, y_k)$	…	$P(x_i, y_k)$	…	$P(x_n, y_k)$
…	…	…	…	…	…	…
$У_m$	$P(x_1, y_m)$	$P(x_2, y_m)$	…	$P(x_i, y_m)$	…	$P(x_n, y_m)$

Закон розподілу двовимірної випадкової величини дозволяє отримати закони розподілу кожної компоненти, тому що події (x_i, y_1), (x_i, y_2), …, (x_i, y_m) несумісні. Ймовірність $P(x_i)$ того, що *X* прийме значення x_i, за теоремою додавання ймовірностей, буде $P(x_i) = P(x_i, y_1) + P(x_i, y_2) + \cdots + P(x_i, y_m)$, тобто дорівнюватиме сумі ймовірностей *k*-го рядка таблиці. Ймовірність

$P(y_j) = P(x_1, y_j) + P(x_2, y_j) + \cdots + P(x_n, y_j)$, тобто буде дорівнювати сумі ймовірностей *j*-го стовпця таблиці.

Приклад 6.1.1. Знайти закони розподілу компонент двовимірної випадкової величини, закон розподілу якої заданий таблицею.

Y	*X*		
	x_1	x_2	x_3
y_1	0,1	0,4	0,2
y_2	0,08	0,12	0,10

Розв'язання*.* Знаходимо ймовірності значень *X* та *Y* відповідно. Для цього додамо спочатку ймовірності за стовпцями, отримаємо ймовірності можливих значень змінної *X*: $p(x_1) = 0{,}1 +$ $+0{,}08 += 0{,}18$; $p(x_2) = 0{,}4 + 0{,}12 = 0{,}52$; $p(x_3) = 0{,}2 + 0{,}10 = 0{,}30$. Тобто закон розподілу *X* має наступний вид.

X	x_1	x_2	x_3
P	0,18	0,52	0,30

Далі додамо ймовірності за рядками, отримаємо ймовірності можливих значень змінної *У*: $p(y_1) = 0{,}1 + 0{,}4 + 0{,}2 = 0{,}7$; $p(y_2) =$ $= 0{,}08 + 0{,}12 + 0{,}10 = 0{,}3$.

Тобто закон розподілу *Y* має наступний вид:

Y	y_1	y_2
P	0,7	0,3

Контроль: 0,18 + 0,52 + 0,30 = 0,7 + 0,3 =1.

6.2. Інтегральна і диференціальна функції розподілу двовимірної випадкової величини

Інтегральною функцією розподілу (функцією розподілу) двовимірної випадкової величини (*X, Y*) називають функцію двох змінних *F(x, y),* яка для кожної пари чисел (*X, Y*) визначає ймовірність виконання нерівностей *X < x, Y < y,* тобто:

$$F(x,y) = P\,(X < x;\ Y < y). \qquad (6.2.1)$$

Аналогічно означають функцію розподілу системи *n* випадкових величин:

$F\,(x_1.x_2,\ldots,x_n) = P(X_1 < x_1,\ X_2 < x_2,\ \ldots,\ X_n < x_n).$

Геометрично функція розподілу *F(x, y*) задає ймовірність попадання випадкової точки *(X, У)* у нескінчений прямий кут з вершиною в точці *M*(*x; y*), який розміщений нижче та лівіше цієї вершини; при цьому промені прямого кута в область не включаються (рис. 6.2.1).

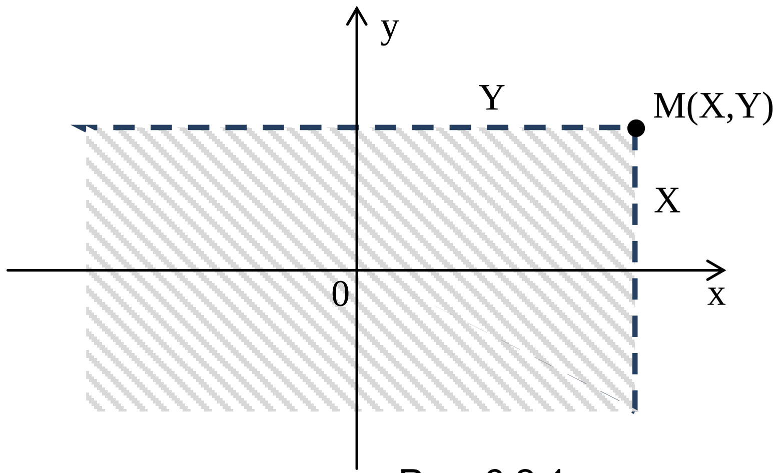


Рис. 6.2.1

Властивості ймовірностей *P* та означення функції розподілу дозволяють довести такі властивості функції розподілу, які у випадку двовимірної випадкової величини будуть наступними:

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

2) *F(x, y*) - неспадна функція за кожним аргументом, тобто $F(x_2,\ y) \ge F\,(x_1,\ y)$, якщо $x_2 > x_1$; $F(x,y_2) > F(x,\ y_1)$, якщо $y_2 > y_1$.

3) Мають місце граничні співвідношення

F(- ∞, y) = 0, *F(x, -∞)* = 0, *F(-∞, ∞) = 0, F (∞, ∞)* = 1.

4) Ймовірність влучення випадкової точки до прямокутника ($x_1 \le X \le x_2$, $y_1 \le Y \le y_2$), якщо відома інтегральна функція, можна знайти за формулою (6.2.2):

$$P(x_1 < X < x_2;\ y_1 < Y < y_2) = \big(F(x_2; y_2) - F(x_1; y_2)\big) -$$
$$-(F(x_2; y_1) - F(x_1; y_1)). \qquad (6.2.2)$$

Ймовірність влучення точки в область довільної форми означається через функцію щільності розподілу (6.2.3).

Приклад 6.2.1. Знайти ймовірність влучення випадкової точки (*X,Y)* у прямокутник, що обмежений лініями $x_1 = \frac{\pi}{6}$, $x_2 = \frac{\pi}{2}$, $y_2 = \frac{\pi}{3}$, $y_1 = \frac{\pi}{4}$, якщо задана функція розподілу має вигляд *F(x, y)=sinx·siny, 0 ≤ x ≤* $\frac{\pi}{2}$*; 0 ≤ y ≤* $\frac{\pi}{2}$.

Розв'язання*.* У даному випадку маємо $x_1 = \frac{\pi}{6}$, $x_2 = \frac{\pi}{2}$,$y_2 = \frac{\pi}{3}$, $y_1 = \frac{\pi}{4}$.

Згідно з формулою (6.2.2) одержуємо:

$$P\left(\frac{\pi}{6} < X < \frac{\pi}{2}; \frac{\pi}{4} < Y < \frac{\pi}{3}\right) =$$

$$= \left[sin\frac{\pi}{2}.\ sin\frac{\pi}{3} - \ sin\frac{\pi}{6}\ .sin\frac{\pi}{3}\right] - \left[sin\frac{\pi}{2}\ .sin\frac{\pi}{4}\ -\ sin\frac{\pi}{6}\ .sin\frac{\pi}{4}\right] =$$

$$= \left[\frac{\sqrt{3}}{2} - \frac{1}{2}.\frac{\sqrt{3}}{2}\right] - \left[\frac{\sqrt{2}}{2} - \frac{1}{2}.\frac{\sqrt{2}}{2}\right] = \frac{\sqrt{3} - \sqrt{2}}{4} \approx 0{,}08.$$

Диференціальною функцією розподілу (двовимірною щільністю ймовірностей) *f (x, y)* двовимірної випадкової величини (*X, Y)* називають мішану частинну похідну другого порядку від інтегральної функції розподілу

$$f(x,\ y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}. \qquad (6.2.3)$$

Аналогічно означають щільність ймовірностей *n*-вимірної випадкової величини, тобто

$$f(x_1, x_2, \ldots, x_n) = \frac{\partial^n F(x_1, x_2, \ldots x_n)}{\partial x_1 \cdot \partial x_2 \ldots \partial x_n}.$$

Таким чином, якщо функція розподілу *F(x, y)* двовимірної випадкової величини відома, то за формулою (6.2.3) можна знайти диференціальну функцію розподілу *f(x, y)* цієї випадкової величини.

І, навпаки, якщо відома щільність ймовірностей *f(x, y)* двовимірної випадкової величини, то її функцію розподілу знаходять за формулою

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x, y) dx dy, \tag{6.2.4}$$

тобто з використанням невласного двократного інтеграла.

Диференціальна функція розподілу $f(x, y)$ задовольняє наступним *властивостям*:

1) $f(x, y) \geq 0$, тобто вона невід'ємна;

2) $\iint_{-\infty}^{\infty} f(x, y) dx dy = 1$.

Цей інтеграл – це ймовірність влучення випадкової точки в площину *xOy*, тобто ймовірність достовірної події. Геометрично це означає, що повний об'єм тіла, що обмежене поверхнею розподілу і площиною *xOy*, дорівнює одиниці.

3) *Ймовірність влучення випадкової точки (X, У) у довільну область D,* якщо відома диференціальна функція розподілу, знаходять за формулою

$$P((X;Y) \in D) = \iint_D f(x; y) dx dy. \tag{6.2.5}$$

Приклад 6.2.2. Знайти функцію розподілу, якщо відома щільність сумісного розподілу $f(x, y) = \frac{1}{\pi^2} \cdot \frac{1}{(1+x^2)(1+y^2)}$.

Розв'язання.

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(x,y)dxdy=\frac{1}{\pi^2}\int_{-\infty}^{x}\frac{dx}{1+x^2}\int_{-\infty}^{y}\frac{1}{1+y^2}dy=$$

$$=\frac{1}{\pi^2}\lim_{u\to\infty}u\int_{u}^{x}\frac{dx}{1+x}\lim_{v\to\infty}v\int_{v}^{y}\frac{1}{1+y2}dy=$$

$$=\frac{1}{\pi^2}(\lim_{u\to\infty}u\,(\ arctgx|_{u}^{x})\lim_{v\to\infty}v\left(\ arctgy|_{u}^{y}\right)=$$

$$=\frac{1}{\pi^2}\left(\ arctgx+\frac{\pi}{2}\right)\left(\ arctgy+\frac{\pi}{2}\right)=\left(\frac{1}{\pi}arctgx+\frac{1}{2}\right)\left(\frac{1}{\pi}\ arctgy+\frac{1}{2}\right).$$

Приклад 6.2.3. Знайти сталу *а,* якщо диференціальна функція

$$f(x,y)=\begin{cases}a\left(\ 2-\sqrt{\mathrm{x}^2+\mathrm{y}^2}\right), & \text{якщо }(\mathrm{x},\mathrm{y})\in D=\{\mathrm{x}^2+\ \mathrm{y}^2\le 4\},\\ 0, & \text{поза цією областю }.\end{cases}$$

Розв'язання*.* Згідно з властивістю диференціальної функції $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dxdy=1,$

$$\int\int_{D}\ \mathrm{a}\left(\ 2-\sqrt{\mathrm{x}^2+\mathrm{y}^2}\right)dxdy=1,$$

$$a=\frac{1}{\int_0^{2\pi}d\varphi\int_0^{2}(2-\tau)\tau d\tau}=\ \frac{1}{2\pi(\ 2\frac{\tau^2}{2}-\frac{\tau^3}{3})}=\frac{1}{2\pi(\ 4-\frac{8}{3})}=\frac{3}{8\pi}.$$

При обчисленні даного інтеграла відбувся перехід до полярних координат:

$$x=\tau\cos\varphi\ ;\ \mathrm{y}=\tau\sin\varphi.$$

6.3. Знаходження щільності ймовірностей складових двовимірної випадкової величини

Нехай відома щільність сумісного розподілу ймовірностей системи двох випадкових величин *f(x, y).* Знайдемо щільність розподілу складової *X.* Позначимо через $\boldsymbol{F_1(x)}$ її функцію розподілу. За означенням $\boldsymbol{f(x)=\frac{\partial F_1(x)}{\partial x}}$.

Оскільки $\boldsymbol{F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(x,y)dxdy}$, то $\boldsymbol{F_1(x)=F(x,\infty\ )}$, і дістанемо

$$F_1(x) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(x,y)dxdy.$$

Знайдемо диференціали обох частин рівності по *x*,

$$\frac{\partial F_1}{\partial \mathrm{x}} = \int_{-\infty}^{\infty} f(x,y)dy,$$

тоді

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy. \qquad (6.3.1)$$

Аналогічно

$$f_2(x) = \int_{-\infty}^{\infty} f(x,y)d\mathrm{x}. \qquad (6.3.2)$$

Очевидно, що $\int_{-\infty}^{\infty} f_1(x)d\mathrm{x} = 1, \quad \int_{-\infty}^{\infty} f_2(x)dy = 1$.

Приклад 6.3.1. Двовимірна випадкова величина (*X, У*) задана щільністю розподілу

$$f(x,y) = \begin{cases} \frac{1}{6\pi}, & (x,y)\epsilon \{\frac{\mathrm{x}^2}{9} + \frac{\mathrm{y}^2}{4} \le 1\} \\ 0, & (x,y)\epsilon\{\frac{\mathrm{x}^2}{9} + \frac{\mathrm{y}^2}{4} > 1\}. \end{cases}$$

Знайти щільність розподілу складових *X* та *У.*

Розв'язання. $f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{6\pi}\int_{-2\sqrt{1-\mathrm{x}^2}}^{2\sqrt{1-\mathrm{x}^2}} dy = \frac{2}{9\pi}\sqrt{9-\mathrm{x}^2}$.

Отже, $f_1(x) = \begin{cases} \frac{2}{9\pi}\sqrt{9-\mathrm{x}^2}, & |\mathrm{x}| < 3, \\ 0, & |\mathrm{x}| \ge 3. \end{cases}$

$f_2(x) = \int_{-\infty}^{\infty} f(x,y)dy = \frac{2}{6\pi}\int_{0}^{3\sqrt{1-\mathrm{y}^2}} d\mathrm{x} = \frac{1}{2\pi}\sqrt{4-\mathrm{y}^2}$.

Отже, $f_2(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-\mathrm{y}^2}, & |\mathrm{y}| < 2, \\ 0, & |\mathrm{y}| \ge 2. \end{cases}$

Для контролю покажемо, що $\int_{-3}^{3} f_1(x)d\mathrm{x} = 1$. Маємо:

$$\int_{-3}^{3} \frac{2}{9\pi}\sqrt{9-\mathrm{x}^2}d\mathrm{x} = \frac{4}{9\pi}\int_{0}^{3}\sqrt{9-\mathrm{x}^2}d\mathrm{x};$$

Заміна $x = 3\cos\varphi$; $y = \sin\varphi$, тоді $dx = -3\sin\varphi d\varphi$, при x =0, $\varphi = \frac{\pi}{2}$; при x= 3, $\varphi = 0$.

$$\frac{4}{9\pi}\int_0^3 \sqrt{9-x^2}dx = \frac{4}{9\pi}\int_{\frac{\pi}{2}}^0 3\sin^2\varphi d\varphi =$$

$$= -\frac{2}{\pi}\left(\varphi - \frac{\sin 2\varphi}{2}\right)\Big|_{\frac{\pi}{2}}^0 = -\frac{2}{\pi}\left(-\frac{\pi}{2} + \frac{\sin 2\pi}{2}\right) = 1.$$

Аналогічно можна довести, що $\int_{-2}^2 f_2(x)dy = 1$.

6.4. Числові характеристики двовимірної випадкової величини

Математичне сподівання двовимірної випадкової величини (X, Y) характеризує координати *центру розподілу* випадкової величини. Ці координати у випадку неперервних величин знаходять за формулами

$$M(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dxdy, \tag{6.4.1}$$

$$\mathrm{M(У)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, f(x,y)dxdy. \tag{6.4.2}$$

Дисперсії Dx та *Dy* характеризують розсіювання випадкової точки *(X, Y)* вздовж координатних осей *Ox* та *Oy* відповідно. Їх знаходять за формулами

$$D(\mathrm{X}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 \cdot f(x,y)dxdy - m_x^2, \tag{6.4.3}$$

$$D(\mathrm{У}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2 \cdot f(x,y)dxdy - m_y^2. \tag{6.4.4}$$

$$\sigma_x = \sqrt{D_x}, \quad \sigma_y = \sqrt{D_y}. \tag{6.4.5}$$

Якщо випадкові величини *X* та *Y* дискретні, то в формулах (6.4.1) - (6.4.4) знаки інтегралів замінюють знаками суми за усіма можливими значеннями випадкових величин.

Приклад 6.4.1. Задано закон розподілу системи дискретних випадкових величин *Х* та *У*.

x/y	5,2	10,2	15,2
2,4	0,01	0,2	0,09
4,4	0,2	0,02	0,18
6,4	0,19	0,08	0,03

Обчислити: *М(Х), М(У), D (Х), D (У), σ(Х), σ(У).*

Розв'язання. $M(x)\sum_{\mathrm{i}=1}^{3}\sum_{j=1}^{3}x_i p_{ij} =$ 5,2 (0,2 + 0,01+ 0,19) + +10,2 (0,2+0,02+0,08) + 15,2(0,09+0,18+0,03) = = 5,2·0,4+10,2·0,3+ 15,2·0,3= 9,7;

$M(Y) = \sum_{\mathrm{i}=1}^{3}\sum_{j=1}^{3}\mathrm{y}_j p_{ij} =$ 2,4(0,01+0,2+0,09) +4,4 (0,2+0,18+0,02)+ + 6,4(0,19+0,08 + 0,03) = 2,4·0,3+4,4·0,4+6,4·0,3 =4,4;

$M(\mathrm{X}^2) = \sum_{\mathrm{i}=1}^{3}\sum_{j=1}^{3}\mathrm{x}_{\mathrm{i}}^2 p_{ij} =$ $(5,2)^2$ ·0,4+$(10,2)^2$·0,3+$(15,2)^2$·0,3= 111,34;

$M(Y^2) = \sum_{\mathrm{i}=1}^{3}\sum_{j=1}^{3}\mathrm{y}_j^2 p_{ij} =$ ($2,4)^2$·0,3+$(4,4)^2$·0,4+$(6,4)^2$ ·0,3 =21,76;

D (X)= *M*(X^2) − *(M(X)*^{2}=111,34 – $(9,7)^2$ =17,25;

D (y)= *M*(y^2) − *(M(y))*^{2}=21,76 – $(4,4)^2$ = 2,4;

σ(Х)=$\sqrt{17{,}25}$ =4,15; *σ(У)*=$\sqrt{2{,}4}$ =1,55.

6.5. Умовні закони розподілу складових системи дискретних і неперервних випадкових величин

Умовним розподілом складової *Х* при $\mathrm{У} = y_j$ називають сукупність умовних імовірностей *p(*$^{\mathrm{x}_1}/_{\mathrm{y}_j}$*), ...,p(*$^{\mathrm{x}_n}/_{\mathrm{y}_j}$*),* знайдених за умови, що подія *У=y$_j$* вже відбулась. Аналогічно визначається умовна ймовірність складової *У*. Якщо закон розподілу двовимірної дискретної випадкової величини відомий, то легко можна знайти закони розподілу складових. Згідно з теоремою множення ймовірностей маємо: $P_A(B)=\frac{P(AB)}{P(A)}$. Відповідно для випадкових величин

$$P({}^{x_i}/_{y_j}) = \frac{P(x_{i,}y_j)}{P(yi)} \qquad (6.5.1),$$

$$P({}^{y_j}/_{x_i}) = \frac{P(x_{i,}y_j)}{P(x_i)} \qquad (6.5.2),$$

де $i=1, \ldots, n$; $j=1,\ldots, m$.

Таким чином, умовний закон розподілу складової *X* за умови, що *У*=y_j , визначається за формулою (6.5.1), а умовний закон розподілу складової У за умови, що *X*= x_i , визначається за формулою (6.5.2).

Очевидні такі рівності: $\sum_{i=1}^{n} P\left({}^{x_i}/_{y_j}\right) = 1$, $\sum_{j=1}^{m} P\left({}^{y_j}/_{x_i}\right) = 1$, які використовуються для контролю правильності знайдених умовних законів розподілів складових.

Приклад 6.5.1. Дискретна двовимірна випадкова величина задана таблицею (табл. 6.5.1). Знайти: а) закони розподілу одновимірних випадкових величин *X* і *У*; б) умовні закони випадкової величини *X* за умови, що *у* =2; в) умовні закони випадкової величини *У* за умови, що *x* =1; г) *P (У*≤ X).

Таблиця 6.5.1.

x/y	**0**	**1**	**2**	**3**	**4**	**5**
0	0,02	0,08	0,1	0,06	0,04	0,02
1	0	0,03	0,07	0,09	0,1	0,12
2	0	0	0,02	0,04	0,06	0,08
3	0	0	0	0,01	0,02	0,04

Розв’язання. а) Додамо ймовірності за стовпцями і отримаємо ймовірності можливих значень *У*: p_1= 0,02; p_2 = 0,11; p_3 =0,19; p_4 = 0,2; p_5 = 0,22; p_6 = 0,26, тобто закон розподілу *У* має наступний вид (табл. 6.5.2):

Таблиця 6.5.2

У	0	1	2	3	4	5
P	0,02	0,11	0,19	0,2	0,22	0,26

$$\sum_{i=1}^{6} p_i = 1.$$

Додамо ймовірності за рядками і отримаємо ймовірності можливих значень *X*: p_1= 0,32; p_2 = 0,41; p_3 = 0,2; p_4 = 0,07, тобто закон розподілу *X* має наступний вид (табл. 6.5.3)

Таблиця 6.5.3

X	0	1	2	3
P	0,32	0,41	0,2	0,07

$$\sum_{i=1}^{4} p_i = 1.$$

б) Умовний закон розподілу *X* за умови, що *y* =2, отримаємо, якщо ймовірності p_{ij}, які стоять в стовпці з можливими значеннями *y* =2, розділити на їх суму, тобто $p_3 = P(\,y=2\,)$=0,19. Маємо:

$$P({}^{x_1}/_{y=2}) = \frac{p_{13}}{p_3} = \frac{0,1}{0,19} = \frac{10}{19};\ P({}^{x_2}/_{y=2}) = \frac{p_{23}}{p_3} = \frac{0,07}{0,19} = \frac{7}{19};$$

$$P({}^{x_3}/_{y=2}) = \frac{p_{33}}{p_3} = \frac{0,02}{0,19} = \frac{2}{19};\ P({}^{x_4}/_{y=2}) = \frac{p_{43}}{p_3} = \frac{0}{0,19} = 0.$$

Додамо для контролю знайдені умовні ймовірності, упевнимося, що їх сума дорівнює одиниці

$$\frac{10}{19} + \frac{7}{19} + \frac{2}{19} = 1.$$

Умовний закон розподілу має наступний вид (табл. 6.5.4)

Таблиця 6.5.4

X	0	1	2	3
$P(x_i/y=2)$	$\frac{10}{19}$	$\frac{7}{19}$	$\frac{2}{19}$	0

в) Аналогічно отримується умовний закон розподілу випадкової величини *У* за умови, що *х* =1, якщо ймовірності p_{ij}, які стоять у другому рядку з можливими значеннями *х* =1, розділити на їх суму, тобто $p_2 = P(x=1)=0{,}41$. Маємо:

$$P(y_1/x=1) = \frac{p_{12}}{p_2} = \frac{0}{0{,}41} = 0;\ P(y_2/x=1) = \frac{p_{22}}{p_2} = \frac{0{,}03}{0{,}41} = \frac{3}{41};$$

$$P\left(y_3/x=1\right) = \frac{p_{32}}{p_2} = \frac{0{,}07}{0{,}41} = \frac{7}{41};P(y_4/x=1) = \frac{p_{42}}{p_2} = \frac{0{,}09}{0{,}41} = \frac{9}{41};$$

$$P(y_5/x=1) = \frac{p_{52}}{p_2} = \frac{0{,}1}{0{,}41} = \frac{10}{41};P(y_6/x=1) = \frac{p_{33}}{p_3} = \frac{0{,}12}{0{,}41} = \frac{12}{41}.$$

Додамо для контролю знайдені умовні ймовірності, упевнимося, що їх сума дорівнює одиниці: $\frac{10}{19}+\frac{7}{19}+\frac{2}{19}=1$.

Умовний закон розподілу має наступний вид (табл. 6.5.5)

Таблиця 6.5.5

У	0	1	2	3	4	5
$P(y_i/x=1)$	0	$\frac{3}{41}$	$\frac{7}{41}$	$\frac{9}{41}$	$\frac{10}{41}$	$\frac{12}{41}$

г) Для знаходження ймовірності *Р (У*≤ X) необхідно додати ймовірності подій p_{ij} з табл. 6.5.1, для яких $y_j \le x_i$. Маємо:

Р (У≤ X) = 0,02 + 0,03 + 0,02 + 0,01 = 0,08.

Розглянемо умовні закони розподілу складових системи неперервних випадкових величин. Нехай (*X, У*) - неперервна випадкова величина.

Умовною щільністю $f(^{x}/_{y})$ *розподілу складової X* за умови, що *У*=y_j , називають відношення щільності сумісного розподілу $f(x, y)$ системи (*X,У)* до щільності розподілу $f_2(y)$ складової *У:*

$$f(^{x}/_{y}) = \frac{f(x,y)}{f_2(y)}. \quad (6.5.3)$$

Аналогічно,

$$f(^{y}/_{x}) = \frac{f(x,y)}{f_1(x)}. \quad (6.5.4)$$

Якщо відома щільність розподілу $f(x, y)$, то умовні щільності складових можна знайти наступним чином відповідно:

$$f(^{x}/_{y}) = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dx}, \qquad f(^{y}/_{x}) = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dy}. \quad (6.5.5)$$

Приклад 6.5.2. Двовимірна випадкова величина задана (*X,У)* задана щільністю розподілу

$$f(x, y) = \begin{cases} \frac{1}{4\pi}, (x, y) \in x^2 + y^2 \le 4, \\ 0, \ (x, y) \epsilon x^2 + y^2 > 4. \end{cases}$$

Знайти умовні закони розподілу ймовірностей складових.

Розв'язання*.* Знайдемо умовний закон складової *X* при $|y| \le \sqrt{4 - x^2}$:

$$f(^{x}/_{y}) = \frac{\frac{1}{4\pi}}{\int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} \frac{1}{4\pi} dy} = \frac{1}{2\sqrt{4-x^2}}.$$

Аналогічно знайдемо умовний закон складової *У* при $|x| \le \sqrt{4 - y^2}$:

$$f(^{y}/_{x}) = \frac{\frac{1}{4\pi}}{\int_{-\sqrt{4-y^2}}^{\sqrt{4-y^2}} \frac{1}{4\pi} dx} = \frac{1}{2\sqrt{4-y^2}}.$$

6.6. Залежні та незалежні випадкові величини

Дві випадкові величини є *незалежними*, якщо закон розподілу однієї з них не залежить від того, які можливі значення прийняла інша величина.

Випадкові величини є *залежними,* якщо закон розподілу однієї величини залежить від того, які значення прийняла інша величина.

Теорема. Щоб випадкові величини *X* та *Y* були незалежні, необхідно і достатньо, щоб інтегральна функція системи (*X, У*) дорівнювала добутку інтегральних функцій кожної з них:

$$F(x, y) = F_1(x) \cdot F_2(y). \tag{6.6.1}$$

Наслідок. Щоб неперервні випадкові величини *X* та *У* були незалежними, необхідно і достатньо, щоб диференціальна функція системи *(X, У)* дорівнювала добутку диференціальних функцій складових:

$$f(x, y) = f_1(x) \cdot f_2(y). \tag{6.6.2}$$

Для опису двовимірної випадкової величини, крім математичного сподівання, дисперсії та середнього квадратичного відхилення, використовують також інші характеристики. Одна з них – *кореляційний момент* (або *коваріація*):

$$cov\,(X, Y) = K_{XY} = M((X - m_x)\,(Y - m_y)). \tag{6.6.3}$$

Для неперервних величин *X* та *Y* кореляційний момент дорівнює

$$K_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - m_X)\,(y - m_Y)\,f\,(x,y)dxdy. \tag{6.6.4}$$

Кореляційний момент характеризує зв’язок між випадковими величинами *X* та *У*. Для незалежних випадкових величин кореляційний момент дорівнює нулеві.

Отже, описати систему двох випадкових величин *X* та *Y* можна за допомогою двох математичних сподівань, що характеризують їхні середні значення, або двох дисперсій, що характеризують розсіювання, або двох кореляційних моментів, що характеризують парний зв'язок між змінними системи. Дисперсія кожної випадкової величини є частинним випадком кореляційного моменту.

Всі кореляційні моменти для неперервних величин *X* та *Y* зазвичай подають у вигляді квадратної кореляційної матриці

$$\begin{pmatrix} K_{xx} & K_{xy} \\ K_{yx} & K_{yy} \end{pmatrix} \text{ або } \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}. \tag{6.6.5}$$

З означення кореляційного моменту очевидно, що $K_{ij} = K_{ji}$, тобто симетричні відносно головної діагоналі елементи кореляційної матриці-таблиці рівні. Тому матрицю іноді подають у формі

$$\begin{pmatrix} K_{11} & K_{12} \\ & K_{22} \end{pmatrix},$$

по головній діагоналі якої - дисперсії. Якщо випадкові величини не корельовані, то всі елементи матриці, крім діагональних, дорівнюють нулеві:

$$\left(K_{ij}\right) = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}.$$

З кореляційним моментом тісно пов'язаний коефіцієнт кореляції. *Коефіцієнт кореляції* є кількісна характеристика залежності випадкових величин *X* та Y, яка часто використовується у статистиці. *Коефіцієнт кореляції* обчислюється за формулою

$$r_{xy} = \frac{K_{XY}}{\sigma_x \sigma_y}. \tag{6.6.6}$$

Випадкові величини X та Y називаються *некорельованими,* якщо їх кореляційний момент або коефіцієнт кореляції дорівнює нулеві.

Властивості коефіцієнта кореляції

1) $|r_{xy}| \leq 1$.

2) якщо X та Y незалежні, то $r_{xy} = 0$;

3) якщо між X та Y існує функціональна лінійна залежність $У = aX + b$, де a та b - сталі, то $|r_{xy}| = 1$.

Якщо момент кореляції або коефіцієнт кореляції не дорівнює нулеві, тоді випадкові величини X та $У$ - корельовані. Дві корельовані величини обов'язково залежні. Але дві залежні випадкові величини можуть бути корельованими або некорельованими, тобто їх коефіцієнт кореляції може дорівнювати, а може і не дорівнювати нулеві.

З незалежності двох величин випливає їх некорельованість, але з некорельованості ще не випливає незалежність цих величин. У випадку нормального розподілу величин із некорельованості випадкових величин випливає їх незалежність.

Примітка. При розв'язуванні практичних задач мають справи з системами більше ніж двох випадкових величин. Повною характеристикою системи довільної кількості випадкових величин $(X_1, X_2, ..., X_n)$, як і двох випадкових величин, є закон розподілу, який може бути заданий функцією розподілу або щільністю розподілу.

Систему k випадкових величин, що взяті з системи n випадкових величин довільним чином ($k<n$), називають частковою системою. Описати систему n випадкових величин $X_1, X_2, ..., X_n$ можна за допомогою n математичних сподівань, що характеризують їхні середні значення, або n дисперсій, що характеризують розсіювання, або $n(n\text{-}1)$ кореляційних моментів, що характеризують попарний зв'язок всіх величин, що входять до

системи. Дисперсія кожної з випадкових величин є частинним випадком кореляційного моменту.

Всі кореляційні моменти для неперервних величин $X_1, X_2, ..., X_n$ зазвичай подають у вигляді прямокутної кореляційної матриці

$$\left(K_{ij}\right) = \begin{pmatrix} K_{11} & K_{12} & ... & K_{1n} \\ K_{21} & K_{22} & ... & K_{2n} \\ ... & ... & ... & ... \\ K_{m1} & K_{m2} & ... & K_{mn} \end{pmatrix}. \qquad (6.6.7)$$

З означення кореляційного моменту очевидно, що $K_{ij} = K_{ji}$, тобто симетричні відносно головної діагоналі елементи кореляційної матриці-таблиці рівні. Тому матрицю іноді подають у формі

$$\begin{pmatrix} K_{11} & K_{12} & ... & K_{1n} \\ & K_{22} & ... & K_{2n} \\ & & ... & ... \\ & & ... & K_{mn} \end{pmatrix},,$$

по головній діагоналі якої - дисперсії. Якщо випадкові величини не корельовані, то всі елементи матриці, крім діагональних, дорівнюють нулеві:

$$\begin{pmatrix} D_1 & 0 & ... & 0 \\ 0 & D_2 & ... & 0 \\ & & ... & ... \\ 0 & 0 & ... & D_n \end{pmatrix},.$$

Запитання для самоперевірки

1. Дайте поняття багатовимірної та двовимірної випадкових величин. Їхня геометрична інтерпретація.
2. Закон розподілу двовимірної дискретної випадкової величини.

3. Як за законом розподілу дискретної випадкової величини (*Х, У)* знайти одновимірні закони розподілу випадкових величин *Х* та *У*?

4. Що називається функцією розподілу двовимірної випадкової величини?

4. Дайте геометричну інтерпретацію функції розподілу.

5. Сформулюйте властивості функції розподілу.

6. Диференціальна функції розподілу двовимірної випадкової величини та її властивості.

7. Який існує зв'язок між інтегральною та диференціальною функціями розподілу двовимірної випадкової величини?

8. Як знайти щільність ймовірностей складових двовимірної випадкової величини?

9. Числові характеристики двовимірної випадкової величини.

10. Що називається умовним законом розподілу складових системи дискретних і неперервних випадкових величин і які формули для їх обчислення ?

11. Залежні та незалежні випадкові величини та їх властивості.

12. Кореляційний момент та коефіцієнт кореляції,його властивості.

13. Охарактеризуйте кореляційну матрицю системи випадкових величин.

Задачі та вправи

6.1. В урні є чотири кулі: 2 білі, 1 чорна і 1 синя. З урни навмання беруть дві кулі. Знайти закон розподілу двовимірної випадкової величини (*Х, У)*, де *у*- число чорних, а *х*– число синіх куль у вибірці.

6.2. Двовимірна дискретна випадкова величина задана законом розподілу

x/*y*	0,01	0,02	0,03	0,04
0,08	0,02	0,03	0,04	0,04
0,02	0,03	0,15	0,06	0,24
0,04	0,04	0,08	0,08	0,10
0,06	0,02	0,04	0,03	0,02

Знайти: *M(X), M(У), D(X), D(У), σ(X), σ(У).*

6.3. Закон розподілу випадкового вектора (*X,У)* дискретного типу задається таблицею

x/*y*	1	2	3
0	0,05	0,05	0,20
2	0,10	0,15	0,05
4	0,10	0,10	0,03
6	0,10	0,04	0,03

Знайти : *M(X), M(У), D(X), D(У), σ(X), σ(У).*

6.4. Дано щільність розподілу системи випадкових величин (*y, x)*

$$f(x,y) = \begin{cases} a\sin(x+y), 0 \le x \le \pi/2, \ 0 \le y \le \pi/2, \\ 0, \text{ в інших випадках}. \end{cases}$$

Знайти коефіцієнт *a* та функцію розподілу випадкової величини (*X, У).*

6.5. Система випадкових величин (*X, У)* підпорядкована закону розподілу з щільністю

$$f(x,y) = \begin{cases} a(x+y) & \text{в області Q}, \\ 0 & \text{поза цією областю.} \end{cases}$$

Треба знайти коефіцієнт *a* та функцію розподілу випадкової величини (*X, У),* якщо: а) Q – квадрат, що обмежений прямими *x*=0, *x* = 3, *y* = *0, y = 3;* б) $f(x,y) \in Q$, де Q – квадрат, що обмежений прямими *x* = 1, *x* = 2, *y* = 1, *y* = 2.

6.6. Система випадкових величин (*X, У)* розподілена за законом $f(x,y) = -\frac{a}{1+x^2+x^2y^2+y^2}$, де $(x;y)\epsilon R^2$.

Визначити коефіцієнт *a.* Перевірити,чи є величини *X* і *У* залежними. Знайти p_{y}, p_x та ймовірність попадання випадкової величини (*X, У*) у квадрат, центр якого збігається з початком координат, а сторони паралельні осям координат і мають довжину *b* = 2.

6.7. Два стрільці, незалежно один від одного, роблять по два поодиноких (незалежних) постріли кожен по своїй мішені. Випадкова величина *x* – число влучень першого стрільця, *y* – другого стрільця. Ймовірність влучення при одному пострілі для першого стрільця p_1 = 0,7, для другого стрільця p_2 = 0,4.

Побудувати закон розподілу двовимірної випадкової величини (*x, y)* і закони розподілу випадкових величин *X* та *У*. Знайти функцію розподілу.

6.8. Тричі незалежно підкидається гральний кубик. Нехай *y* – число появ четвірок, *x* – число появ парних чисел. Знайти ймовірність *P (2y >x).*

6.9. Закон розподілу системи двох випадкових величин (*X, У)* задано таблицею розподілу

x/y	0	1	2	3
-1	0,01	0,06	0,05	0,04
0	0,04	0,24	0,15	0,07
1	0,05	0,10	0,10	0,09.

Знайти:1) безумовні закони розподілу *X* та *У;* 2) умовний закон розподілу *y* за умови, що *x* = 0, і умовний закон розподілу *x* при *y* = 0.

6.10. Система випадкових величин (*X,У*) задана щільністю розподілу

$$f(x,y) = \begin{cases} \cos x \cos y, & 0 \le x \le \pi/2, \; 0 \le y \le \pi/2, \\ 0, & \text{в інших випадках}. \end{cases}$$

Знайти функцію розподілу системи (*X, У*).

6.11. Випадковий вектор (*X, У*) має закон розподілу

x/y	0	1	2
0	0,1	0,2	0,1
1	0,2	0,3	0,1

Знайти: 1) закони розподілу складових *X* та *У*; 2) *P(y < 1)*; *3) P (y < 1, x> 0).*

6.12. Якість продукції характеризується двома випадковими параметрами *X* та *У*. Закон розподілу двовимірного вектора задано таблицею. Знайти: 1) умовний розподіл випадкової величини *y* за умови, що *x* = 0,1; 2) умовний розподіл випадкової величини *x* за умови, що *y* = 6.

x/y	5	6	7
0	0,2	0,03	0
0,1	0,1	0,15	0
0,2	0,05	0,15	0,1
0,3	0,05	0,1	0,1.

6.13. Двовимірна випадкова величини (*X,У*) задана таблицею розподілу

x/y	-1	1
-1	1/8	5/24
0	1/6	1/6
1	5/24	1/8

Знайти: 1) закон розподілу випадкових величин *Х* та *У*; 2) умовний розподіл випадкової величини *х* за умови, що *у* = -1; 3) *P(y<x).*

6.14. Система випадкових величин (*Х,У)* має щільність розподілу

$f(x,y)=\frac{a}{\pi^2(16+x^2)(25+y^2)}$, де $(x,y)\in R^2$.

Треба знайти: 1) величину *а*; 2) функцію розподілу *F.*

6.15. Щільність розподілу двовимірної випадкової величини (*Х,У)* задана формулою

$f(x,y)=e^{-ax^2+bxy-cy^2}$, *a>0, c>0,* $(x,y)\in R^2$.

Знайти щільність розподілу p_y і p_x . Для яких значень параметрів *У* та *Х* є незалежними випадковими величинами?

6.16. Щільності розподілів випадкових величин *Х* та *У* задаються відповідно формулами:

f_y (*x*) = $\begin{cases} 0 \text{ при } x \le -1, \\ 1/3 \text{ при} -1 < x \le 2, \\ 0 \text{ при } x > 2, \end{cases}$ та f_x (*y*) = $\begin{cases} 0 \text{ при } x \le 2, \\ \frac{1}{5} \text{ при } 2 < x \le 7, \\ 0 \text{ при } x > 7. \end{cases}$

Випадкові величини *Х* та *У* незалежні. Знайти ймовірність того, що випадкова величина (*X, У*) попала в прямокутник
П ≡{(x, y), -2 ≤ x ≤ 1, 3 ≤ y ≤ 5}.

6.17. Пара випадкових величин (*X, У*) визначається таким чином: якщо при підкиданні грального кубика випадає парне число, то *у* = 1, в протилежному випадку *у = 0; х* = 1, коли число очок кратне трьом, у протилежному випадку *х* = 0.

Написати: а) закон розподілу випадкового вектора (*X, У*); б) безумовні та умовні закони розподілу *X* та *У.*

6.18. Нехай (*x, y)* – координати випадкової точки – розподілені рівномірно всередині прямокутника П ≡ *{(x, y)*| *a≤ x ≤b;*

c≤ *y* ≤*d}*. Знайти щільність розподілу та функцію розподілу двовимірної випадкової величини (*X, У*).

6.19. Нехай є двовимірна випадкова величина (X, У), де випадкова величина *У* розподілена за показниковим законом з параметром λ:

$$f_{\mathrm{y}}\ (x)=\begin{cases} 0 \text{ при } \mathrm{x}\ \leq 0 \\ \lambda\ \mathrm{e}^{-\lambda \mathrm{x}} \text{ при } \mathrm{x} > 0. \end{cases}$$

Випадкова величина *X* при заданому значенні *y* = *x, x* > 0, розподілена також за показниковим законом, але з параметром *x:*

$$f_x\ (y/x) = \begin{cases} 0 \text{ при } \mathrm{y} \leq 0 \\ x e^{-xy} \text{ при } \mathrm{y} > 0. \end{cases}$$

Знайти: 1) щільність розподілу двовимірної випадкової величини (*X, У*); 2) щільність розподілу f_x (*x/y).*

6.20. Система випадкових величин (*X, У)* підпорядкована закону розподілу зі щільністю

$$f(x,y) = \begin{cases} \mathrm{C}(x^2 + xy + y^2), & 0 \leq x \leq 1; 0 \leq y \leq 1, \\ 0, & \text{у решті випадків.} \end{cases}$$

Знайти C, *M(X), M(У), D(X), D(У),* $f(x,y),\ f(x),\ f(y).$

РОЗДІЛ 7
ГРАНИЧНІ ТЕОРЕМИ ТЕОРІЇ ЙМОВІРНОСТЕЙ

Теорія ймовірностей, як відомо, вивчає закономірності масових випадкових явищ. Наявність цих закономірностей пов'язана з великою кількістю виконаних однорідних випробувань, е результаті яких випадкова величина розподіляється за певним законом. При вивченні результатів випробувань над реальними масовими випадковими явищами має місце закономірність, яка називається властивістю стійкості. Суть цієї властивості полягає в тому, що конкретні властивості кожного конкретного явища майже не впливають на середній результат маси таких явищ. Ця стійкість середніх і є фізичним змістом *закону великих чисел,* суть якого в широкому понятті цього слова наступна: при дуже великій кількості випадкових явищ середній їхній результат практично перестає бути випадковим, і може бути вгаданий з великим ступенем визначеності.

Закон великих чисел грає важливу роль у різних процесах, що пов'язані з масовим виробництвом.

Під законом великих чисел розуміють ряд математичних (граничних) теорем, в кожній з яких для тих чи інших умов встановлюється факт наближення середніх характеристик великого числа випробувань до деяких певних сталих.

Граничні теореми теорії ймовірностей, які встановлюють відповідність між теоретичними та дослідними характеристиками випадкових величин або випадкових подій при великій кількості випробувань, об'єднують загальною назвою - *закон великих чисел.*

Граничні теореми описують також граничні закони розподілу. Граничні теореми, що встановлюють граничні закони розподілу випадкових величин, об'єднують загальною назвою - *центральна гранична теорема* (ЦГТ).

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

1) Коли сума багатьох випадкових величин мало відрізняється від сталої величини, тобто майже перестає бути випадковою величиною, і тому її поведінка може прогнозуватись із значною ймовірністю?

2) При яких умовах можна зі значною ймовірністю можна прогнозувати число появ деякої випадкової події при великій кількості незалежних випробувань?

3) При яких обмеженнях сума багатьох випадкових величин буде розподілена за нормальним законом?

7.1. Нерівність Чебишова

Для демонстрації ймовірнісно-статистичниого підходу до аналізу, перевірки вищеназваних задач є корисними нерівності, які вперше застосував у теорії ймовірностей російський математик П.Л.Чебишов (1821-1894). Ці нерівності широко використовуються в теорії математичної статистики, а також безпосередньо застосовуються в багатьох практичних задачах прийняття рішень, зокрема: у задачах статистичного аналізу процесів і якості продукції, коли невідомий вид функції розподілу результатів спостережень; у задачі виключення результатів спостережень, що різко відхиляються, тощо.

При розв'язанні подібних задач, а також при доведенні граничних теорем важливу роль відіграє нерівність Чебишова, яка має дві форми.

Перша форма нерівності Чебишова. Для довільної випадкової величини X, яка приймає невід'ємні значення ($X \geq 0$) та має скінчене математичне сподівання $M(X)=m$, для будь-якого ε>0 виконується нерівність

$$P(X \geq \varepsilon) \leq \frac{M(X)}{\varepsilon} \qquad (7.1.1)$$

Нехай X - дискретна випадкова величина, яка задана розподілом

X	x_1	x_2	…	x_n
P	p_1	p_2	…	p_n

Виберемо деяке число ε >0 і припустимо, що в даному ряду $x_1 \leq \varepsilon$, $x_2 \leq \varepsilon$, …, $x_k \leq \varepsilon$ і $x_{k+1} \geq \varepsilon$, $x_{k+2} \geq \varepsilon$, …, $x_n \geq \varepsilon$.

$$M(X) = \sum_{i=1}^{k} x_i p_i + \sum_{i=k+1}^{n} x_i p_i.$$

Так як за умовою всі $x_i \geq 0$ і $p_i \geq 0$, то, відкинувши першу суму, після чого $M(X)$ може тільки зменшитися, і замінивши в другій сумі всі x_i на меншу величину ε, отримаємо нерівність

$$M(X) \geq \varepsilon \sum_{i=k+1}^{n} p_i \qquad \text{звідки} \qquad \sum_{i=k+1}^{n} p_i \leq \frac{M(X)}{\varepsilon}.$$

Сума ймовірностей у лівій частині отриманої нерівності для подій $X=x_{k+1}$, …, $X=x_n$, тобто ймовірність події $X>\varepsilon$. Тому

$$P(X \geq \varepsilon) \leq \frac{M(X)}{\varepsilon} \qquad (7.1.1)$$

Якщо X – неперервна випадкова величина, а $f(x)$ – щільність її ймовірностей, то, виконуючи перетворення, аналогічні до попередніх, отримаємо послідовно

$$M(X) = \int_1^{\infty} xf(x)dx = \int_1^{\varepsilon} xf(x)dx + \int_{\varepsilon}^{\infty} xf(x)dx \geq \varepsilon \int_{\varepsilon}^{\infty} f(x)dx.$$

Звідси

$$\int_{\varepsilon}^{\infty} f(x)dx \leq \frac{M(X)}{\varepsilon} \qquad (7.1.2)$$

Наслідок. Якщо *X* приймає лише невід'ємні значення, *M(X)*<∞, *a* > *0*, то

$$P(X \geq a) \leq \frac{M(X)}{a}. \quad (7.1.3)$$

Дійсно, $P(X \geq a) = P\left(\frac{x}{a} \geq 1\right) \leq M\left(\frac{X}{a}\right) = \frac{M(X)}{a}$.

Нерівність (7.1.3) іноді називають *нерівністю Маркова*.

Приклад 7.1.1. Сума всіх вкладів у деякому банку складає $2 \cdot 10^6$ гривень, а ймовірність того, що випадково взятий вклад не перевищує 10^4 , дорівнює 0,8. Що можна сказати про число вкладників ощадного банку?

Розв'язання. Нехай *X* - величина випадково взятого вкладу, а *n* – число вкладників. З умови задачі випливає, що $M(X)=\frac{2 \cdot 10^6}{n}$.

Оскільки

$$P(X \leq 10^4\ ) \geq 1\text{-}\ \frac{2 \cdot 10^6}{10^4 n}, 0{,}8 \geq 1 - \frac{200}{n}; 0{,}2 \leq \frac{200}{n}.$$

Отже, $n\ \leq 100$.

Друга форма нерівності Чебишова. Якщо випадкова величина *X* має скінчені математичне сподівання та дисперсію, то для довільного ε > 0 має місце нерівність

$$P(|X - M(X)| > \varepsilon) \leq \frac{D(X)}{\varepsilon^2}. \quad (7.1.4)$$

або в іншому варіанті

$$P(|X - M(X)| < \varepsilon) \geq 1 - \frac{D(X)}{\varepsilon^2}. \quad (7.1.5)$$

Доведення. Спочатку розглянемо протилежну подію $P(|X - M(X)|) \geq \varepsilon$. Легко побачити, що ця подія еквівалентна події $(X - M(X)^2 \geq \varepsilon^2)$, тому до неї можна застосувати першу форму нерівності Чебишова:

$$P(|X - M(X)| \geq \varepsilon) = P\left(\frac{1}{\varepsilon^2}|X - M(X)|^2 \geq 1\right) \leq \frac{M\left(\left(X - M(X)\right)^2\right)}{\varepsilon^2} = \frac{D(X)}{\epsilon^2}.$$

Тепер імовірність протилежної події |*X-M(X)*| <*ε* задовольняє нерівність (7.1.4), що і треба було довести.

З (7.1.4), зокрема, випливає, що для випадкової величини *X* – числа настання події А в серії з *n* незалежних повторних випробувань, у кожному з яких подія А настає з імовірністю *p*, матимемо:

$$P(|k - np| \geq \varepsilon) \leq \frac{npq}{\varepsilon^2}, \tag{7.1.6}$$

звідки дістанемо *теорему Бернуллі* (7.2.3) або *закон великих чисел*

$$\lim_{n\to\infty} P\left(\left|\frac{k}{n} - p\right| \geq \varepsilon\right) = 0, \tag{7.1.7}$$

який стверджує, що частота $\frac{k}{n}$ настання події А в серії з *n* незалежних повторних випробувань наближається до ймовірності p події *A* зі зростанням *n*.

Закон великих чисел є основою для вивчення масових явищ. Сутність його полягає в тому, що кожне одиничне явище випадкове (може бути або не бути), але в об'єднанні великої кількості таких явищ у загальній характеристиці їх маси випадковість зникає в тім більшій мірі, чим більше об'єднано одиничних явищ. Математика, зокрема теорія ймовірностей і математична статистика, розглядає в чисто кількісному вираженні закон великих чисел, виражає його серією математичних теорем (Чебишова, Ляпунова, Маркова та ін.). Вони показують, при яких умовах можна розраховувати на відсутність випадковості в охоплюваних масових характеристиках, та як це пов'язане з чисельністю присутніх у ній індивідуальних явищ. Зокрема, статистика базується на цих теоремах у вивченні кожного масового явища.

Приклад 7.1.2. Дисперсія випадкової величини *X* дорівнює 0,003. Яка ймовірність того, що випадкова величина *X* відрізняється від її математичного сподівання *M(X)* більше ніж на 0,1?

Розв'язання. За нерівністю Чебишова (7.1.4) маємо:

$$P(|X - M(X)| > 0{,}1) \leq \frac{D(X)}{(0{,}1)^2} = \frac{0{,}003}{0{,}01} = 0{,}3.$$

Приклад 7.1.3. Ймовірність появи подій А в кожному з 2000 дослідів дорівнює 0,3. Користуючись нерівністю Чебишова, оцінити ймовірність того, що відхилення числа появ цієї події від математичного сподівання буде більшою за 50.

Розв'язання. Число появ події в *n*=2000 дослідах є випадкова величина X, яка розподілена за біноміальним законом. Тоді її математичне сподівання і дисперсія будуть знайдені за формулами (4.5.2):

$M(X) = 2000 \cdot 0{,}3 = 600; \quad D(X) = 2000 \cdot 0{,}3 \cdot 0{,}7 = 420.$

Користуючись нерівністю Чебишова (7.1.6) при ε=50, отримаємо:

$$P(|X - 600| \geq 50) \leq \frac{420}{50^2} = 0{,}168.$$

Приклад 7.1.4. Ймовірність народження хлопчиків становить 0,51. Оцінити з допомогою нерівності Чебишова ймовірність того, що серед 2000 новонароджених хлопчиків буде від 940 до 1100.

Розв'язання. Маємо справу з повторними незалежними випробуваннями. Випадкова величина X - число хлопчиків серед 2000 новонароджених – розподілена за біноміальним законом розподілу. Тому $M(X) = 2000 \cdot 0{,}51 = 1020;$

$D(X) = 2000 \cdot 0{,}51 \cdot 0{,}49 = 499{,}8.$ Оскільки числа 940 і 1100 – межі допустимих значень випадкової величини – симетричні відносно

$M(X) = 1020$, то нерівність $940 \leq \mathrm{X} \leq 1100 = 1020$ можна замінити рівносильною їй $|X - 1020| \leq 80$. Використовуючи нерівність (7.1.5) при ε=80, маємо:

$P(940 \leq \mathrm{X} \leq 1100) = P(|X - 1020| \leq 80) \geq 1 - \frac{499.8}{80^2} \approx 0{,}9219.$

Отже, ймовірність шуканої події не менше 0,92.

7.2. Важливі граничні теореми

Нерівності Чебишова дозволяють довести граничну теорему Бернуллі та інші важливі граничні теоремі про стійкість середніх.

Теорема Бернуллі. Нехай імовірність появи події *А* в кожному із *n* незалежних повторних випробувань дорівнює *p, m* - число появи події *А* (частота події) в *n* випробуваннях. Тоді

$$lim_{n\to\infty} P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) = 1, \quad \varepsilon > 0. \qquad (7.2.1)$$

Доведення. Частість $\frac{m}{n}$ можна розглядати як невід'ємну випадкову величину *X.* Знайдемо її математичне сподівання $M(X) = M\left(\frac{m}{n}\right) = \frac{1}{n}M(m) = \frac{1}{n}np = p.$

Отже, необхідно оцінити ймовірність відхилення випадкової величини *X* від її математичного сподівання. Для цього знайдемо дисперсію цієї випадкової величини

$$D(X) = D\left(\frac{m}{n}\right) = \frac{1}{n^2}D(m) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}.$$

За нерівністю Чебишова (7.1.4) одержимо

$$P|(\frac{m}{n} - p| \leq \varepsilon) = 1 - \frac{p(1-p)}{n\,\epsilon^2}$$

Звідси граничним переходом при *n*→∞ одержуємо (7.2.1), що й треба було довести. Це означає, що яким малим не було б ε, зі збільшенням числа випробувань стає дедалі вірогіднішим той факт, що частота появи події А відрізняється від ймовірності цієї події менше ніж на ε.

Теорема Чебишова. Нехай X_1, X_2, …, X_n – послідовність попарно незалежних випадкових величин, які задовольняють умовам: 1) *M* $(X_i) = a_i$, 2) $D(X_i) \le c$, де *i = 1, 2, …, n.*

Тоді

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \frac{\sum_{i=1}^{n} a_i}{n}\right| < \varepsilon\right) = 1 . \qquad (7.2.2)$$

Доведення. Знайдемо математичне сподівання та дисперсію середньої випадкової величин, тобто $\frac{X_1, X_2, \dots, X_n}{n}$:

$$M\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n} M\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} M\,(X_i) = \frac{1}{n}\sum_{i=1}^{n} a_i.$$

$$D\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2} D\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{c\,n}{n^2} = \frac{c}{n}.$$

Застосуємо для випадкової величини $\frac{1}{n}\sum_{i=1}^{n} X_i$ нерівність Чебишова (7.1.4), тоді:

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \frac{\sum_{i=1}^{n} a_i}{n}\right| < \varepsilon\right) \ge 1 - \frac{c}{n\varepsilon^2} . \qquad (7.2.3)$$

Границя цієї імовірності при *n*→∞ дорівнює одиниці, тобто нерівність (7.2.2) доведено.

Центральна гранична теорема (ЦГТ)*.* Нехай задана послідовність незалежних однаково розподілених випадкових величин X_1, X_2, …, X_n,таких що $M(X_i) = 0$, $D(X_i) = b$, де *i* = 1, 2,…, n.

Розглянемо випадкову величину $Y_n = \sum_{i=1}^{n} X_i$. Тоді

$M(Y_i) = \sum_{i=1}^{n} M(X_i) = 0$, $D(Y_n) = \sum_{i=1}^{n} M(x_i) = n\,b.$

При *n*→∞ функція розподілу

$$F_{Y_n}(x) = P(Y_n < x) \to \frac{1}{\sqrt{2\Pi nb}} \int_{-\infty}^{x} e^{-\frac{z^2}{2nb}}\, dz,$$

тобто сума Y_n буде розподілена за нормальним законом з математичним сподіванням *M*=0 та дисперсією $D = \sqrt{nb}$.

Для доведення цієї теореми треба знайти границю характеристичної функції, що побудована для нормованої випадкової величини $Z_n = \frac{Y_n}{\sqrt{nb}}$.

Наслідок. При n ≥ 30 розподіл суми однаково розподілених випадкових величин мало відрізняється від нормального розподілу.

Має місце і загальний випадок ЦГТ для різнорозподілених випадкових величин.

ЦГТ показує, що у випадку, коли результат вимірювань (спостережень) складається під дією багатьох причин, причому кожна з них вносить лише малий вклад, а сукупний результат визначається адитивно, тобто шляхом додавання (підсумовування), то розподіл результату вимірювань (спостережень) близький до нормального.

Отже, важливо, і яким чином ці причини діють. Якщо - адитивно, то X має наближено нормальний розподіл. Якщо мультиплікативно (дії окремих причин перемножуються), то розподіл *X* близький не до нормального, а до логарифмічного, тобто не *X*, а *lgX* має наближено нормальний розподіл. Якщо ж нема підстав вважати, що діє один із цих механізмів формування результату, то про розподіл *X* нічого певного сказати не можна. У таких випадках нормальність результатів вимірювань потрібно перевіряти за допомогою статистичних критеріїв або використовувати непараметричні статистичні методи, що не опираються на припущення належності функцій розподілу результатів вимірювань (спостережень) до певного виду.

Теорема Ляпунова. Нехай задана послідовність незалежних випадкових величин $\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_\mathrm{n}$ таких, що $M(X_i) = 0$, $(D(X_i) = b_i^2$, *i= 1, 2, …, n.*

Побудуємо суму випадкових величин $Y_n = \sum_{i=1}^{n} X_i$.

Позначимо $b_i^2 = \sum_{\mathrm{i=1}}^{\mathrm{n}} b_i^2$. Якщо виконується умова рівномірної малості величин, що утворюють суму, $\frac{1}{b_n^3}\sum_{i=1}^{n} M(X_i) \to 0$ при n→∞, то сума Y_n буде розподіленою нормально з математичним сподіванням $M(Y_n) = 0$ та дисперсією $D(Y_n) = b_n^2$.

Доведення цієї теореми досить складне, але відзначимо, що у випадку, коли $M(X_i) = a_i \neq 0$, можна розглядати випадкові величини $X_i^1 = X_i - a_i$, які будуть задовольняти умовам теореми Ляпунова.

Приклад 7.2.3. Скільки доданків треба взяти в теоремі Чебишова, щоб з надійністю 95% і точністю до 0,001 виконувалась наближена рівність $\frac{1}{n}\sum_{i=1}^{n} X_i \approx \frac{1}{n}\sum_{i=1}^{n} M(X_i)$?

Розв'язання. У цьому прикладі ε=0,001. Щоб одержати надійність 95%, згідно з формулою (7.2.2) достатньо підібрати таке *n*, яке задовольняє нерівності

$$\frac{c}{\varepsilon^2 n} \leq 0{,}05 \ \Rightarrow n \geq \frac{c}{0{,}05 \cdot 0{,}000001} = 20\,000\,000\text{c}.$$

Зауваження. Приклад показує, що навіть у випадку дуже великих точності та надійності, треба брати значну кількість доданків (*n* – досить велике число). Це означає, що оцінки, які одержані з використанням нерівності (7.2.2), - завищені. Більш точні оцінки можна одержувати за допомогою теореми Ляпунова.

Запитання для самоперевірки

1. Сформулюйте закон великих чисел у широкому та вузькому змісті.
2. Сформулюйте нерівність Чебишова.
3. Сформулюйте теорему Бернуллі.
4. Сформулюйте теорему Чебишова.
5. В чому суть центральної граничної теореми Ляпунова?

Вправи та задачі

7.1. Дискретна випадкова величина X задана законом розподілу

X	5	4	6	8
P	0,3	0,2	0,1	0,4

Використовуючи нерівність Чебишова, оцінити ймовірність того, що $P(X \geq \varepsilon) \leq \dfrac{M(X)}{\varepsilon}, \varepsilon = 1,2\sqrt{Д(X).}$

7.2. При виготовленні партії однакових деталей розміром l=10 мм існує допуск Δl =0,1мм. Оцінити ймовірність того,що випадково взята деталь – бракована, якщо σ_x^2 = 0,0025.

7.3. Ймовірність виготовлення нестандартної деталі рівна 0,04. Яке найменше число деталей треба відібрати, щоб з імовірністю 0,88 можна було стверджувати, що доля нестандартних деталей серед них буде відрізнятися від імовірності виготовлення нестандартної деталі по абсолютній величині не більше ніж на 0,02?

7.4. Середній термін служби мотора 4 роки. За нерівністю Маркова оцінити знизу ймовірність того, що даний мотор не прослужить більше 20 років.

7.5. Електростанція обслуговує мережу з 18 000 ламп, ймовірність увімкнення кожної з них у зимовий вечір дорівнює 0,9. Знайти ймовірність того, що число ламп, увімкнених у мережу в зимовий вечір, відрізняється від свого математичного сподівання за абсолютною величиною не більше, ніж на 200.

7.6. Середнє значення швидкості вітру біля землі в даному населеному пункті дорівнює 16 км/год. Оцінити за допомогою першої форми нерівності Чебишова ймовірність того, що в цьому пункті швидкість вітру (при одному спостереженні) не перевищить 80 км/год.

7.7. Середня витрата води в населеному пункті становить 50 000 л на день. Оцінити ймовірність того, що в цьому населеному пункті в даний день витрати води не перевищать 150 000 л.

7.8. Математичне сподівання кількості випадання протягом року в даній місцевості опадів становить 55 см. Оцінити ймовірність того, що в цій місцевості опадів випаде 175 см.

7.9. Число сонячних днів у році для даної місцевості є випадковою величиною з математичним сподіванням, що дорівнює 75 дням. Оцінити ймовірність того, що протягом року у цій місцевості буде не більше 200 сонячних днів.

7.10. Середнє споживання електроенергії за червень місяць населенням містаN становить 360 000 квт-год. а) Оцінити ймовірність того, що споживання електроенергії в червні поточного року перевищить 1 000 000 квт-год. б) Оцінити ту ж імовірність, якщо відомо, що середнє квадратичне відхилення споживання електроенергії в даному місті за червень становить 40 000 квт-год.

7.11. Розподіл випадкової величини X задано таблицею. Оцінити ймовірність P $(|X - M(X)|<2$, користуючись нерівністю Чебишова, якщо: а)

X	1	2	3	4	5	6
P	0,05	0,10	0,25	0,30	0,20	0,10

б)

X	0,3	0,6
P	0,2	0,8

7.12. Для випадкової величини X відомо $D(X)$ = 0,009 і $P(|X - M(X)|< \varepsilon)$ <0,1.Знайти ε.

7.13. Оцінити ймовірність того, що число осіб, які мають вищу освіту, у групі з 800 чоловік, відрізняється від свого математичного сподівання менше ніж на 30.

7.14. Оцінити, користуючись нерівністю Чебишова, ймовірність того, що серед 1200 новонароджених хлопчиків буде від 550 до 650 включно, якщо ймовірність народження хлопчика дорівнює 0,5.

7.15. Математичне сподівання швидкості вітру на даній висоті дорівнює 25км/год., а середнє квадратичне відхилення – 4,5 км/год. Яку швидкість вітру на цій висоті можна очікувати з імовірністю, не меншою ніж 0,9?

7.16. Ймовірність появи події в одному випробуванні дорівнює 0,5. Чи можна з імовірністю, яка більша за 0,97, стверджувати, що число появи події в 1000 незалежних випробуваннях буде в межах від 400 до 600?

7.17. Скільки разів треба виміряти дану величину, істинне значення якої дорівнює *а*, щоб з імовірністю, що не менша за 0,97, можна було стверджувати, що середнє арифметичне значення цих вимірювань відрізняється від *а* за абсолютною величиною менше, ніж на 2, якщо середнє квадратичне відхилення кожного з вимірювань менше 10?

7.18. Проведено 600 незалежних випробувань, у 200 із них імовірність появи події *А* дорівнювала 0,3; а у решті випадків – 0,5. Оцінити знизу ймовірність того, що відхилення частоти від середньої ймовірності не перевищить за абсолютною величиною 0,05.

7.19. Дисперсія кожної *з* $x_1, x_2, \ldots, x_{900}$ незалежних випадкових величин дорівнює 1. Оцінити ймовірність того, що середнє арифметичне цих величин відхиляється від свого математичного сподівання за абсолютною величиною не більше, ніж на 0,1: а) за допомогою нерівності Чебишова; б) за допомогою центральної граничної теореми.

7.20. Випадкова величина *X* має характеристики *M(X)* =1, *σ(X)*=0,2. Оцінити знизу ймовірності подій: *A* = (0,5<X<1,5); *B*=(0,75<X<1,3); *C* = (X<2).

7.21. Ймовірність випуску нестандартної деталі дорівнює 0,25. Оцінити знизу ймовірність того, що у партії з 1000 деталей число нестандартних відрізняється від 250 менше, ніж на 40.

7.22. З 4 000 проведених випробувань у 500 ймовірність появи очікуваного результату 0,4; а у 1 200 – 0,5; у 2 300 – 0,6. Знайти межі, в яких повинна знаходитися частота появи очікуваного результату, якщо це треба гарантувати з імовірністю 0,98.

7.23. Середнє значення витрат води в населеному пункті складає 500 000 літрів на день. Оцінити ймовірність того, що в даному населеному пункті витрати води не будуть перевищувати 120 000 літрів на день.

7.24. Для випадкової величини *X* відомо, що*D(X)* = 0,01 та *P* (|*X* – *M(X)*|< *ε*) >096. Знайти *ε*.

7.25. Ймовірність появи події при одному випробуванні дорівнює 0,5. Користуючись нерівністю Чебишова, оцінити ймовірність того, що число появи події в 100 незалежних випробуваннях буде в межах від 40 до 60.

7.26. Дискретна випадкова величина X задана законом розподілу

X	0,1	0,6	0,4
P	0,2	0,5	0,3

Оцінити ймовірність *P* (|*X* – *M(X)*|< $\sqrt{0{,}4}$, користуючись нерівністю Чебишова.

7.27. Дисперсія кожної з $x_1, x_2, \ldots, x_{4500}$ однаково розподілених незалежних випадкових величин дорівнює 5. Оцінити ймовірність того, що середнє арифметичне цих величин відхиляється від свого математичного сподівання за абсолютною величиною не більше, ніж на 0,04: а) за допомогою нерівності Чебишова; б) за допомогою центральної граничної теореми.

7.28. Ймовірність того, що покупцеві взуттєвого магазину потрібні черевики 41 розміру, становить 0,15. Оцінити межі відсотка покупців серед 2000 тих. що відвідали магазин, яким необхідні такі черевики, якщо ці межі треба гарантувати з імовірністю 0,98.

7.29. Нехай схожість насіння деякої рослини складає 70 відсотків. Оцінити знизу ймовірність того, що при посіві 10000 насінин відхилення частки тих, що зійшли, від ймовірності, що зійде кожна з них, не перевищить за абсолютною величиною 0,01.

7.30. Освітлювальна мережа складається з 20 паралельно ввімкнених ламп. Ймовірність того, що за час Т лампа буде ввімкненою, дорівнює 0,8. Користуючись нерівністю Чебишова, оцінити ймовірність того, що абсолютна величина різниці між числом увімкнених ламп і середнім числом (математичним сподіванням) увімкнених ламп за час Т виявиться: а) меншоюза три; б) не менше трьох.

ЧАСТИНА 2
МАТЕМАТИЧНА СТАТИСТИКА

Математична статистика - це розділ математики, присвячений вивченню закономірностей, що мають місце в масових явищах і статистичних сукупностях.

Математична статистика вивчає методи, які дають змогу за результатами випробувань робити певні ймовірнісні висновки.

Зміст математичної статистики складають математичні методи систематизації, обробки та аналізу масових статистичних даних незалежно від їх якісного змісту.

Методи математичної статистики можуть бути застосовані для обробки та аналізу будь-яких статистичних даних. Математичні методи безпосередньо пов'язані з імовірнісною оцінкою результатів спостереження та визначення величини математичної ймовірності. У зв'язку з цим висновки математичної статистики відносно масових явищ і процесів носять імовірнісний характер і опираються на апарат теорії ймовірностей.

Математична статистика вивчає однорідні сукупності незалежно від їхнього якісного змісту та розглядає їх як абстрактно-математичні сукупності.

Задача математичної статистики полягає у створенні методів збору і обробки статистичних даних для одержання наукових і практичних висновків.

У науці, техніці та на виробництві маємо справу із системами багатьох об’єктів,елементів, явищ. Такі великі системи досліджуються методами математичної статистики. Математична статистика також вивчає закономірності властивостей систем, до складу яких входять багато об’єктів. Вона являє собою науку про методи обробки великої кількості дослідних даних з метою отримання адекватних (достовірних) висновків і використовується в економіці, біології, агрономії, ветеринарії та інших галузях науки, техніки та виробництва.

В аграрних, природничих, технічних галузях, як і в інших, методами математичної статистики проводять систематизацію та аналіз числових даних, спостережень і досліджень над біологічними організмами, рослинами, тваринами, а також узагальнення виробничих показників сільського господарства, зокрема й у тваринництві.

Математична статистика як наука виникла (XVII ст.) з робіт К.Гаусса і почала розвиватися паралельно з теорією ймовірностей. На основі теорії ймовірностей він розробив, дослідив і обґрунтував метод найменших квадратів, який створив ще у 1795 р. і застосовував для обробки астрономічних даних (з метою уточнення орбіти малої планети Церера). Його іменем називають один із найпопулярніших розподілів ймовірностей – нормальний, а в теорії випадкових процесів основний об’єкт вивчення - гауссівські процеси.

В кінці XIX – на початку XX ст. значний вклад зробили англійські дослідники, зокрема К. Пірсон (1857-1936) і Р.А. Фішер (1890-1962). Пірсон розробив критерій “хі-квадрат” перевірки статистичних гіпотез, а Фішер – дисперсійний аналіз, теорію планування експерименту, метод максимальної правдоподібності оцінки параметрів.

У 1930-ті рр. поляк Є. Нейман (1894-1977) і англієць К. Пірсон розвинули загальну теорію перевірки статистичних гіпотез. Подальшим розвитком математична статистика зобов'язана П.Л. Чебишову, А.А. Маркову, О.М. Ляпунову, К.Ф. Гальтону та іншим.

У XX ст. найбільший вклад у математичну статистику зробили А. Вальд, В.С. Королюк, А.М. Колмогоров, Є. Нейман, К. Пірсон, В.І. Романовський, Є.Є. Слуцький, М.В. Смірнов, Стьюдент (псевдонім В. Госсета) та інші вчені.

Математична статистика стрімко розвивається і тепер. За останні 50 років можна виділити такі принципово нові напрямки досліджень:

- розробка і втілення математичних методів планування експериментів;
- розвиток статистики об’єктів нечислової природи як самостійного напрямку в прикладній математичній статистиці;
- розвиток статистичних методів, що стійкі до малих відхилень від використовуваної ймовірнісної моделі;
- широке розгортання робіт по створенню комп’ютерних пакетів програм для статистичного аналізу даних;
- розробка ймовірнісно-статистичних методів і оптимізації, зокрема оптимізації якості продукції та вимог стандартів у менеджменті і т. п.

РОЗДІЛ1
ВИБІРКОВИЙ МЕТОД

Нехай потрібно вивчити сукупність об'єктів відносно деякої якісної або кількісної ознаки, яка характеризує ці об'єкти. Кожен об'єкт, який спостерігають, має декілька ознак. Розглядаючи лише одну ознаку кожного об'єкта, припускається, що інші ознаки рівноправні.

Перевірку множини об'єктів можна провести двома способами:провести (контроль) перевірку всіх об’єктів або перевірити лише певну частину об’єктів.

Якщо об'єктів дуже багато, або перевірка пов'язана з певними пошкодженнями об'єкта, або більшість об’єктів недоступні, то перший спосіб перевірки не доцільний. Якщо дослідити всі елементи не можливо, то відбирають з усієї множини обмежене число елементів і перевіряють лише їх. Такий метод називається вибірковим; група об’єктів, відібраних випадково у відповідний спосіб, називається *вибірковою сукупністю* або *випадковою вибіркою*, а сукупність об’єктів, із яких проводиться вибірка, називають *генеральною сукупністю.*

1.1. Статистична сукупність та її розподіл.

Статистичною сукупністю називається сукупність однорідних об'єктів, схожих у будь-якому відношенні, і в той же час, які мають варіюючі (мінливі) ознаки, що є предметом статистичного дослідження. Прикладом статистичної сукупності може бути множина господарств, підприємств області, тварин однієї породи, рослин, продукції одного виду тощо.

Статистична сукупність складається з окремих елементів або одиниць, а їх загальне число називається *обсягом* (*об'ємом*) сукупності. Елементи сукупності можуть бути охарактеризовані однією або декількома ознаками. Якщо статистичні сукупності досліджуються за однією ознакою, то вони називаються *одновимірними*, за двома – *двовимірними*, за трьома і більше – *багатовимірними*. Ознаки, які набувають різних значень або видозмін у окремих одиниць сукупності, називаються *варіюючими*, а окремі їх значення - *варіантами*. Варіюючи, ознаки поділяються на атрибутивні (якісні) та кількісні. Атрибутивними називають ознаки, які не піддаються числовому вираженню. Наприклад, професія, національність, стать, сорт, порода тощо. Ознака називається кількісною, якщо окремі її варіанти виражаються числами (урожайність, продуктивність праці, заробітна плата, відстань тощо).

Нехай задано множину *N (генеральну сукупність)*, що складається із однотипних елементів (наприклад, партія деяких виробів). Треба вивчити *кількісну* чи *якісну* ознаку елементів множини *N.* Експеримент полягає в тому, що ми вибираємо навмання один із елементів множини *N* і реєструємо деяку його характеристику *X* (наприклад, довжину, масу). Випадково вибраний об’єкт після перевірки необхідної ознаки можна повернути (*повторна* вибірка) або не повертати (*безповторна* вибірка) у генеральну сукупність. За допомогою вибірки треба оцінити генеральну сукупність за ймовірнісними властивостями.

Результати вибірки розглядають як послідовність незалежних однаково розподілених випадкових величин

$$x_1,\ x_2,\ \cdots,\ x_n, \qquad (1.1.1)$$

які утворюють *статистичний ряд*. Він відіграє роль початкового (вихідного) числового матеріалу, який є основою для подальшої обробки та аналізу.

Нехай із генеральної сукупності зроблено вибірку об'єму *n*. Значення x_i (*i*=1, 2, …, *n*) досліджуваної ознаки X, що входять у вибірку, називають *варіантами.* Отримані в результаті дослідження статистичні дані необхідно систематизувати, привести їх до певного порядку. Розташувавши варіанти у зростаючому або спадному порядку, отримаємо *ранжирований ряд розподілу*. Але ранжирований ряд ще не дає загальної картини розподілу: не видно закономірності, яка закладена у розподілі, навколо якої величини концентруються варіанти, тощо. Тому виникає потреба подальшої обробки, узагальнення статистичних даних.

Перший етап обробки статистичного ряду (1.1.1) – це складання *варіаційного ряду*. Для цього серед усіх чисел (1.1.1) відбирають усі *різні* та розташовують їх у порядку зростання. При цьому x_i повторюється відповідно n_i раз ($i = 1, 2, ..., k$), тоді число n_i називається *частотою* варіанти x_i , а число W_i – *відносною частотою* (*частістю*) варіанти:

$$Wi = \frac{n_i}{n} \tag{1.1.2}$$

Відносні частоти можна виражати і у відсотках.

Відхилення значень ознаки одне від одного називається *варіаціями*.

Наступний етап обробки статистичного ряду (1.1.1) – це створення *емпіричного закону розподілу*, форма запису якого залежить від характеру ознаки (випадкової величини) X, що вивчається.

Якщо X – *дискретна* випадкова величина, то найуживанішою формою емпіричного закону розподілу є *розподіл частот*.

Статистичним розподілом (розподілом) вибірки називається перелік упорядкованих можливих варіант та відповідних їм частот або відносних частот (частостей). Його найчастіше подають у вигляді таблиці, у перший рядок якої заносять можливі варіанти x_i , які записані у зростаючому (для кількісних ознак) або спадному (для якісних ознак) порядку, а у другий - відповідні частоти (частості).

Приклад статистичного розподілу (розподілу) частот:

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

Статистичний розподіл (розподіл) частостей (відносних частот) має аналогічний вигляд:

x_i	x_1	x_2	...	x_k
$\frac{n_i}{n}$	$\frac{n_1}{n}$	$\frac{n_2}{n}$	...	$\frac{n_k}{n}$

Важливо пам’ятати, що сума всіх частот розподілу становить *n* – об’єм вибірки ($n_1 + n_2 + \cdots + n_k = \sum_1^k n_i = n$), а сума всіх відносних частот (частостей) розподілу дорівнює одиниці:

$(w_1 + w_2 + \cdots + +w_k = \sum_1^k w_i = 1)$ (у відсотках – 100%).

Приклад 1.1.1. Для заданої вибірки із генеральної сукупності скласти розподіли частот та частостей:

7, 4, 4, 8, 12, 12, 12, 7, 8, 12, 8, 12, 4, 12, 12, 4, 12, 4, 12, 12.

Розв’язання. Задано статистичний ряд. Випишемо можливі варіанти x_i , які запишемо у зростаючому порядку, тобто створимо *варіаційний ряд*: 4, 7, 8, 12. Він складається з чотирьох елементів (k=4). Знайдемо відповідні їм частоти n_i , тобто взнаємо, скільки разів зустрічається кожна варіанта. Створимо таблицю, у перший рядок якої занесемо варіанти x_i з варіаційного ряду, а у другий – відповідні їм частоти n_i :

x_i	4	7	8	12
n_i	5	2	3	10

Перевірка: $\sum_1^4 x_i = 5 + 2 + 3 + 10 = 20.$

За формулою (1.2) знайдемо відносні частоти (частості): $W_1 = \frac{n_1}{n} = \frac{5}{20} = \frac{1}{4} = 0{,}25$, $W_2 = \frac{n_2}{n} = \frac{2}{20} = \frac{1}{10} = 0{,}1$, $W_3 = \frac{3}{20} = 0{,}15$, $W_4 = \frac{10}{20} = \frac{1}{2} = 0{,}5$.

Тоді розподіл частостей буде наступним:

x_i	4	7	8	12
n_i	0,25	0,1	0,15	0,5

Перевірка: $\sum_1^4 w_i = 0{,}25 + 0{,}1 + 0{,}15 + 0{,}5 = 1.$

Коли *X – неперервна* величина та *п* - обсяг вибірки, то результати вибірки можна подати інтервальним (неперервним) рядом розподілу. Для цього область реалізацій розбивають на *п інтервалів (груп)* виду (x_1, x_2), (x_2, x_3), …, (x_{k-1}, x_k) однакової довжини (*крок*) $h=x_{i+1} - x_i$. Довжина інтервалу визначається за формулою

$$h = \frac{x_{\max} - x_{\min}}{k}, \qquad (1.1.3)$$

де $x_{\max}$, $x_{\min}$ - максимальне і мінімальне значення ознаки, *k* – число груп, яке може бути наближено обчислене за формулою Стерджесса:

$$k=1+3{,}322\lg N \qquad (1.1.4)$$

(N – чисельність сукупності). Формула Стерджесса може бути використана при умові, що розподіл одиниць сукупності за даною ознакою наближається до нормального закону розподілу. Для визначення *кількості* груп необхідно дотримуватися таких важливих умов побудови груп: 1) виділені групи мають відрізнятися якісною однорідністю; 2) кількість одиниць у кожній групі має бути досить великою, що відповідає закону великих чисел.

Для кожного інтервалу визначають суму частот тих варіант, які попадають у визначений інтервал (групу) – отримаємо згруповану частоту.

Інтервали бувають рівні, нерівні, відкриті або закриті. Вибір виду інтервалу залежить від характеру розподілу одиниць досліджуваної сукупності.

Рівні інтервали використовують у випадках, коли значення варіюючої ознаки *x* змінюється плавно, поступово, рівномірно. Найчастіше це виконують, коли досліджується кількісна ознака. Довжина (ширина) інтервалу обчислюється за формулою (1.1.3).

Приклад 1.1.2. Прибутковість активів комерційних банків коливається у межах від 5 до 35%. Побудувати неперервний розподіл прибутковості комерційних банків.

Розв'язання. При прийнятті кількості груп *k*=3 крок (ширина або довжина інтервалу) $h = \frac{35-5}{3} = 10$. Тоді межі інтервалів будуть відповідно 5-15, 15-25, 25-35. Оскільки межі інтервалів збігаються (кінець попереднього і початок наступного), то для виключення невизначеності віднесення межових значень ознаки до тієї чи іншої групи використовують правило: лівий інтервал не включає в себе кінцеве число значення ознаки, правий – включає. Тоді, наприклад, число 15 повинне бути віднесене до другої групи, а не першої; число 25 – до третьої. Це аналог математичного запису інтервалів [5;15), [15;25), [25;35]. Всі інтервали у даному прикладі є *закритими*.

Наведений розподіл прибутковості активів банків може бути представлений інакше: до 15, 15-25, 25-35, 35 і більше. Перший і останній інтервал мають лише одну межу, тому називаються *відкритими*. Для подальших досліджень відкриті інтервали перетворюють у закриті з тим же кроком.

Нерівні інтервали використовують у разі, коли діапазон значень ознаки надто широкий і розподіл сукупності за цією ознакою нерівномірний. Наприклад, розподіл населених пунктів за кількістю жителів (тис. осіб): до 3; 3-4,99; 5-9,99; 20-49,99.

Якщо інтервальний ряд будується за атрибутивною (якісною) ознакою, то виділяють стільки груп (класів), скільки є градацій ознаки.

Приклад 1.1.3.У результаті спостереження одержали дискретний розподіл ознаки *X* вибірки у вигляді

x_i	-2	0	1	2	3	5	7
n_i	4	5	7	8	6	2	1

Побудувати неперервний розподіл частот.

Розв'язання*.* Об'єм вибірки становить 33 одиниці. У даному випадку найменше значення варіанти $x_{min=}$ -2, а найбільше x_{max} =7, тому довжина проміжку $[x_{min}; x_{max}]$ дорівнює 9. Розіб'ємо цей відрізок на 4 рівних частини довжиною *h* = $\frac{9}{4} = 2{,}25$. Для кожного інтервалу визначимо суму частот тих варіант, які попадають у визначений інтервал (групу) – отримаємо згруповану частоту.

Отже, матимемо:

1) відрізку [-2;0,25] належать 9 значень варіант (-2 – чотири рази та 0 - п'ять разів);

2) відрізку [0,25;2,5] належать15 значень варіант (1 – сім разів та 2 - вісім разів);

3) відрізку [2,5;4,75] належать 6 значень варіант (3 – шість разів);

4) відрізку [4,75;7] належать 3 значення варіант (5 – двічі та 7 – один раз).

Отриманий неперервний (згрупований) розподіл буде наступним:

Відрізки довжини *h*=2,25: інтервали $(x_i - x_{i+1})$	[-2;0,25]	[0,25;2,5]	[2,5;4,75]	[4,75;7]
Кількість значень, n_i	9	15	6	3

Об'єм вибірки за неперервною ознакою становить теж 33 одиниці.

Накопичена частота даної варіанти отримується підсумовуванням всіх частот варіант, що передують даній, з частотою цієї варіанти. Те ж саме виконується і для згрупованого (інтервального) розподілу, а також для розподілу частостей дискретної або неперервної ознаки. Останньому (максимальному) значенню варіанти (інтервалу) відповідає накопичена частота, що дорівнює сумі всіх частот - об'єму вибірки (для частостей - одиниці).

Приклад 1.1.4. За умовою попереднього прикладу побудувати згрупований розподіл накопиченої частоти.

Розв'язання. Розбиття на інтервали залишимо таке ж.

Інтервали $(x_i - x_{i+1})$	[-2;0,25]	[0,25;2,5]	[2,5;4,75]	[4,75;7]
Накопичена частота, n_i	9	24	30	33
Накопичена частість, w_i	9/33	24/33	30/33	33/33=1

Правило-орієнтир для складання таблиці розподілу частот

1. Виписати всі можливі варіанти x_i, які потрібно записати у зростаючому порядку (або спадному – для якісної ознаки), тобто створити *варіаційний ряд*.

2. Знайти відповідні їм частоти n_i, тобто взнати, скільки разів зустрічається кожна варіанта.

3. Створити таблицю, у перший рядок якої занести варіанти x_i з варіаційного ряду, а у другий – відповідні їм частоти n_i.

4. Перевірити, щоб $n_1 + n_2 + \cdots + n_k = \sum_1^k n_i = n$.

Аналогічно складають розподіл відносних частот (частостей), врахувавши, що сума всіх відносних частот (частостей) розподілу дорівнює одиниці ($w_1 + w_2 + \cdots + +w_k = \sum_1^k w_i = 1$) , а також розподіл згрупованих частот або частостей.

1.2. Емпірична функція розподілу та її властивості

Щоб від окремих подій перейти до одночасного розгляду багатьох подій, використовують *накопичену частоту* – це відношення числа одиниць, для яких результати спостережень менші від заданого значення, до загального числа спостережень при умові, що результати спостережень – дійсні числа. Функцію, що виражає залежність між значеннями кількісної ознаки і накопиченою частотою, називають емпіричною функцією розподілу.

Емпіричною функцією розподілу називають функцію *F*(x),* яка визначає для кожного значення *x* відносну частоту події $X < x$.

$$\mathrm{F}^*(x) = \frac{n_x}{n}, \tag{1.2.1}$$

де n_x - число варіант, що менші за x; n – об'єм вибірки. Інакше: емпірична функція розподілу $F^*(x)$ - це доля елементів, що менші за x. Емпірична функція розподілу містить всю інформацію про результати спостережень.

Властивості емпіричної функції $F^*(x)$:

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

Це твердження витікає з того, що відносна частота довільної випадкової величини є невід'ємне число, яке не перевищує 1.

2. $F^*(x)$ – неспадна функція; тобто $F^*(x_2) \ge F^*(x_1)$, якщо $x_2 > x_1$.

3. Якщо x_1– найменша варіанта, то $F^*(x)=0$ при $x \le x_1$; якщо x_k - найбільша варіанта, то $F^*(x)=1$ *при* $x>x_k$.

Емпіричну функцію розподілу $F^*(x)$ можна використовувати як оцінку (наближене значення) функції розподілу $F(x)$ випадкової величини X (генеральної сукупності).

Приклад 1.2.1. Знайти емпіричну функцію за даним розподілом вибірки та побудувати її графік.

x_i	1	4	6
n_i	10	15	25

Розв'язання. n =10+15+25=50.

Найменша варіанта рівна 1, тому $F^*(x)=0$ при $x \le 1$.

При $1 < x \le 4$ значення $x<4$, (x_1=1) спостерігалось 10 разів, отже $F^*(x) = \frac{10}{50} = 0.2$.

При $4 < x \le 6$ значення $x<6$: x_1=1 і x_2=4, спостерігались 10+15=25 разів; отже $F^*(x) = \frac{25}{50} = 0{,}5$.

Так як x=6 – найбільша варіанта, то $F^*(x)=1$ при $x>6$.

Запишемо значення емпіричної функції:

$$F^*(x) = \begin{cases} 0; & x \le 1; \\ 0{,}2; & 1 < x \le 4; \\ 0{,}5; & 4 < x \le 6; \\ 1; & x > 6. \end{cases}$$

Графік емпіричної функції розподілу наведено нижче (рис. 1.2.1).

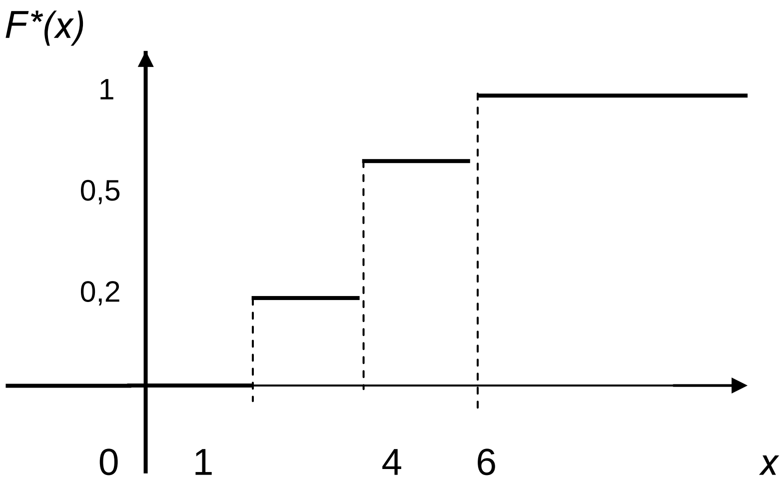


Рис. 1.2.1

1.3. Графічне зображення варіаційних рядів

Для наочності варіаційні ряди розподілу можуть бути зображені графічно у вигляді полігона, гістограми, огіви або кумуляти. Завдяки цьому можна побачити характерні зміни ряду розподілу, не користуючись аналізом цифрових даних. Зображення варіаційних рядів у табличній та графічній формах дозволяє одержати лише перше уявлення про найбільш загальні характерні властивості досліджуваного розподілу.

Графічне зображення варіаційних рядів треба розглядати окремо для незгрупованих і згрупованих даних.

Розглянемо спочатку зображення варіаційних рядів для незгрупованих даних у вигляді *гістограми*, *полігону* частот і частостей (відносних частот).

Нехай у результаті відбору ми отримали статистичний розподіл ознаки *X*, яку треба дослідити. Маємо перелік варіант ознаки x_1, x_2,…, x_k та відповідних їм частот n_1, n_2, …, n_k (або частостей W_1, W_2, ..., W_k).

Значення варіант і частот (або частостей) можна розглядати як координати точок (x_1, n_1); (x_2, n_2); …; (x_k, n_k) або $(x_1; W_1)$; $(x_2; W_2)$; …; $(x_k; W_k)$.

Полігоном частот називають ламану, відрізки якої з'єднують точки $(x_1; n_1)$; $(x_2; n_2)$; …; $(x_k; n_k)$.

Полігоном відносних частот називають ламану, відрізки якої з'єднують точки $(x_1; W_1)$; $(x_2; W_2)$; …; $(x_k; W_k)$.

Для побудови полігону частот і відносних частот відкладемо на осі абсцис варіанти x_i, а на осі ординат – відповідні частоти або відносні частоти. Знайдені точки $(x_i; n_i)$ або $(x_i; W_i)$ з'єднаємо відрізками та отримаємо полігон частот або відносних частот.

Полігони частот і частостей є аналогами щільності ймовірностей.

Приклад 1.3.1. Побудувати полігон відносних частот за даним розподілом вибірки

x_i	20	40	65	80
n_i	3	5	7	10

Розв'язання. Об'єм вибірки дорівнює n= 3+5+7 +10= 25. Знайдемо відносні частоти $W_i=\frac{n_i}{n}$, для чого розділимо кожну із частот на об'єм вибірки: $W_1=\frac{3}{25}$=0,12; $W_2=\frac{5}{25}$=0,2; $W_3=\frac{7}{25}$=0,28; $W_4=\frac{10}{25}$=0,4.

Напишемо розподіл відносних частот:

x_i	20	40	65	80
W_i	0,12	0,2	0,28	0,4

Побудуємо полігон. Відкладемо на осі абсцис декартової системи координат значення варіант x_i, а на осі ординат – відповідні відносні частоти. Знайдені точки (x_i; W_i) з'єднаємо відрізками та отримаємо полігон відносних частот (рис. 1.3.1).

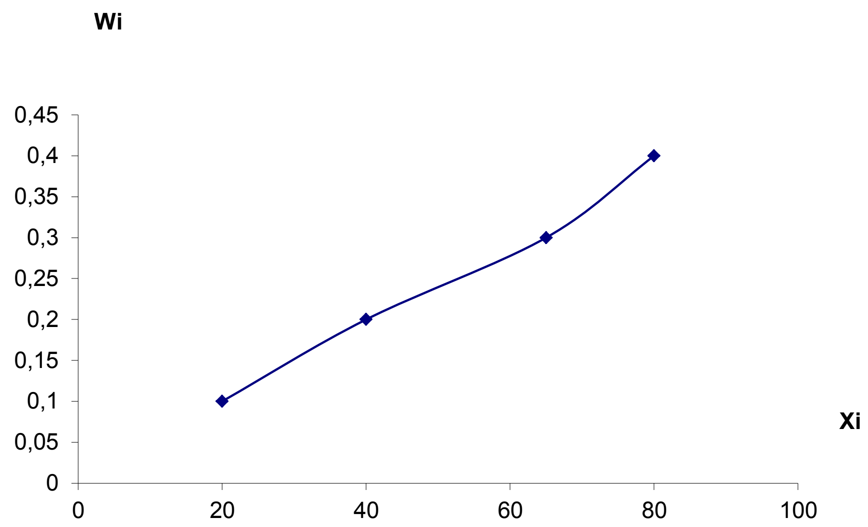


Рис. 1.3.1

Коли X – неперервна величина та n - обсяг вибірки, то результати вибірки можна подати інтервальним (неперервним) рядом розподілу. Для цього область реалізацій розбивають на n інтервалів виду (x_1, x_2), (x_2, x_3), …, (x_{k-1}, x_k) однакової довжини (крок) h=x_{i+1} - x_i, для кожного інтервалу визначають суму частот. Здобутий ряд геометрично подається гістограмою.

Гістограмою частот називають ступінчату фігуру, яка складається із прямокутників, основи яких є часткові інтервали довжиною $h = x_k - x_{k-1}$, а висоти H_i рівні відношенню $\frac{n_i}{h}$ (щільність частот)

$$H_i = \frac{n_i}{h} \tag{1.3.1}$$

Гістограмою відносних частот називають ступінчату фігуру, яка складається із прямокутників, основи яких є часткові інтервали довжиною $h = x_k - x_{k-1}$, а висоти H_iрівні відношенню $\frac{W_i}{h}$ (щільність частостей):

$$H_i = \frac{W_i}{h} \quad (1.3.2)$$

Площа гістограми частот дорівнює об'єму вибірки, а площа гістограми відносних частот – одиниці.

Для побудови гістограми частот (частостей) проміжок $(x_{max}\,; x_{min})$, тобто від найменшого значення $x_{min,}$ що спостерігали, до найбільшого x_{max} , розбивають на декілька відрізків рівної довжини *h.* Потім підраховують суму частот (частостей) тих значень варіант ознаки *X*, які належать кожному із одержаних відрізків. Далі на осі абсцис відкладають інтервали довжиною *h*, а на них, як на основах, будують прямокутники, висота яких дорівнює $H_i = \frac{n_i}{h}$ (для частостей $\frac{W_i}{h}$) – ця величина називається *щільністю* частот (частостей).

Приклад 1.3.2*.* У результаті спостереження одержали розподіл ознаки *X* вибірки у вигляді

x_i	-2	0	1	2	3	5	7
n_i	4	5	7	8	6	2	1

Побудувати гістограму частот цього розподілу.

Розв'язання*.* У даному випадку найменше значення варіанти $x_{min=}$ -2, а найбільше x_{max} $=7$, тому довжина проміжку $[x_{min}; x_{max}]$ дорівнює 9. Розіб'ємо цей відрізок на 4 рівних частини довжиною *h* = $\frac{9}{4} = 2{,}25$. (Див. приклад 1.1.3). Для побудови гістограми доцільно скласти таблицю, у перший рядок якої записують отримані відрізки, у другий – суму частот тих варіант, що належать відповідному відрізку, а у третій – висоти відповідних прямокутників.

Відзначимо, що:

1) відрізку [-2;0,25] належать 9 значень варіант (-2 – чотири рази та 0 - п'ять разів);

2) відрізку [0,25;2,5] належать15 значень варіант (1 – сім разів та 2 - вісім разів);

3) відрізку [2,5;4,75] належать 6 значень варіант (3 – шість разів);

4) відрізку [4,75;7] належать 3 значення варіант (5 – двічі та 7 – один раз).

Відрізки довжини h=2,25	[-2;0,25]	[0,25;2,5]	[2,5;4,75]	[4,75;7]
Кількість значень n_i	9	15	6	3
Щільність частот $\frac{n_i}{h}$	4	$\frac{20}{3}=6{,}67$	$\frac{8}{3}=2{,}67$	$\frac{4}{3}=1{,}33$

За отриманими даними побудуємо відповідну гістограму частот (рис. 1.3.2).

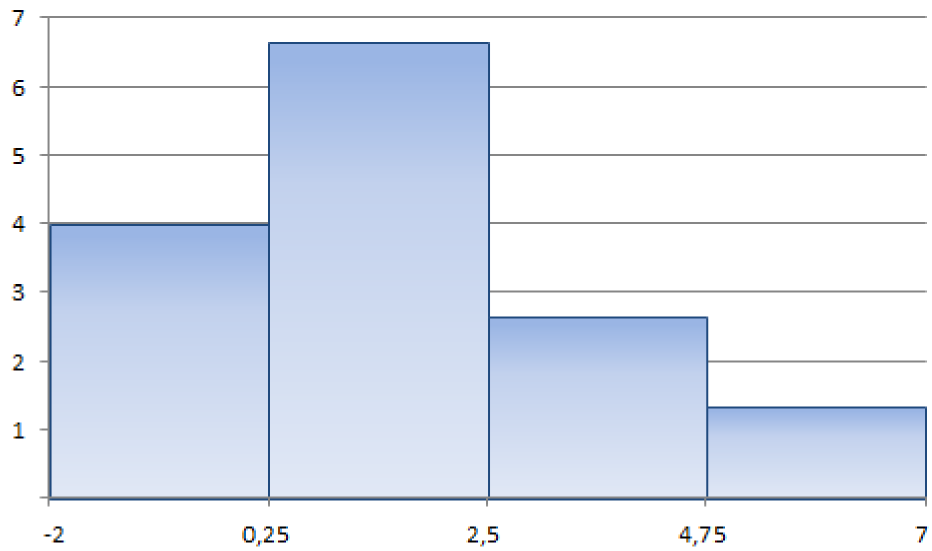


Рис.1.3.2

Приклад 1.3.3. На підставі сформованої статистичної сукупності побудувати інтервальний варіаційний ряд розподілу 100 господарств за урожайністю зернових культур (ц/га): 14,0; 14,7; 15,4; …; 41,5; 42,0.

Даний інтервальний варіаційний ряд зобразити графічно, побудувавши гістограму та полігон розподілу.

Розв'язання. Побудуємо ранжируваний ряд розподілу господарств за урожайністю зернових культур, розташовуючи їх в порядку зростання. Встановимо мінімальне та максимальне значення ознаки (урожайності) в ряду розподілу.

У запропонованій для розв'язання задачі у 100 відібраних у випадковому порядку господарствах мінімальне та максимальне значення урожайності становитимуть: x_{min} =14,0 ц/га; x_{max} = 42,0 ц/га.

Визначимо величину інтервалу *h*, виходячи з того, що досліджувану сукупність господарств внаслідок її великої чисельності можна розділити на *п* = 7 інтервалів (груп).

$$h = \frac{Xmax - Xmin}{n} = \frac{42{,}0 - 14{,}0}{7} = \frac{28{,}0}{7} = 4{,}0 \text{ ц/га}$$

Визначимо межі інтервалів. Мінімальне значення урожайності x_{min} = 14,0 ц/га приймаємо за нижню межу першого інтервалу, а верхню визначаємо як x_{min}+ h= 14,0 ц/га +4,0 ц/га = 18,0 ц/га. Тоді межі першого інтервалу складуть 14,0ц/га - 18,0 ц/га. Верхня межа першого інтервалу буде нижньою межею другого інтервалу. Знову, додаючи до неї значення величини інтервалу *(h),* знаходимо верхню межу другого інтервалу. Тоді другий інтервал має межі 18,0 ц/га — 22,0 ц/га. Аналогічно визначаємо межі всіх семи інтервалів. Підрахуємо кількість об'єктів у кожному інтервалі. Отриманий інтервальний ряд розподілу запишемо у вигляді таблиці розподілу 100 господарств за урожайністю зернових культур (табл. 1.3.1).

Таблиця 1.3.1

Інтервальний ряд розподілу 100 господарств за урожайністю зернових культур

№ групи	Межі інтервалів за урожайністю (ц/га)	Кількість господарств (частота)	Питома вага	
			у % до підсумку	у частках (частість)
I	14,0 - 18,0	9	9	0,09
II	18,0 - 22,0	15	15	0,15
III	22,0 - 26,0	16	16	0,16
IV	26,0 - 30,0	24	24	0,24
V	30,0 - 34,0	18	18	0,18
VI	34,0 - 38,0	12	12	0,12
VII	38,0 - 42,0	6	6	0,06
Разом		100	100	1

Далі розглянемо зображення варіаційних рядів для згрупованих даних у вигляді гістограми, полігону частот і частостей (відносних частот).

Для побудови полігону частот (частостей) треба знайти середини кожного інтервалу і над ними відкладається точка на висоті, відповідній частоті (відносній частоті) цього інтервалу. Після цього ці точки сполучаються відрізками прямих. Зобразимо варіаційний ряд розподілу господарств за урожайністю зернових культур у вигляді гістограми, відкладаючи значення інтервалів (урожайність x у ц/га) на осі абсцис, а кількість об'єктів n - на осі ординат. Побудуємо на цьому ж рисунку полігон розподілу, з'єднуючи прямими лініями середини верхніх сторін прямокутників (рис. 1.3.3).

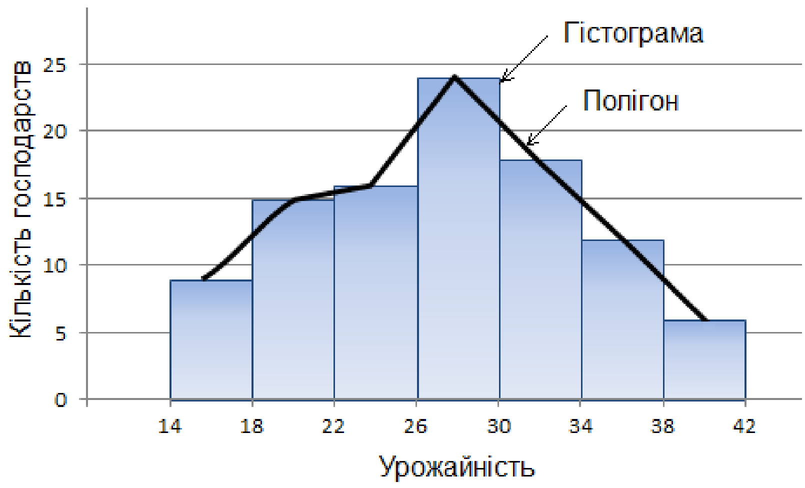


Рис. 1.3.3

Приклад 1.3.4. Побудувати полігон відносних частот за даним розподілом вибірки, поданої у вигляді таблиці частот.

інтервал	(-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

Розв'язання. Знайдемо середини інтервалів: -2,5; -1,5; -0,5; 0,5; 1,5; 2,5; 3,5; 4,5.

Об'єм вибірки n = 3+10+15+24+25+13+7+3 = 100. Обчислимо відносні частоти $W_i = \frac{n_i}{n}$, для чого розділимо кожну із частот на об'єм вибірки.

Середина інтервалу	-2,5	-1,5	-0,5	0,5	1,5	2,5	3,5	4,5
W_i	0,03	0,10	0,15	0,24	0,25	0,13	0,07	0,03

Зобразимо полігон графічно (рис. 1.3.4).

Для побудови гістограми згрупованого розподілу частот (відносних частот), щільності частоти (щільності відносної частоти) на горизонтальну вісь наносяться класи інтервалів. На кожному класі будується прямокутник, висота якого рівна значенню частоти (або відносної частоти, або щільності частоти, або щільності відносної частоти) на цьому інтервалі.

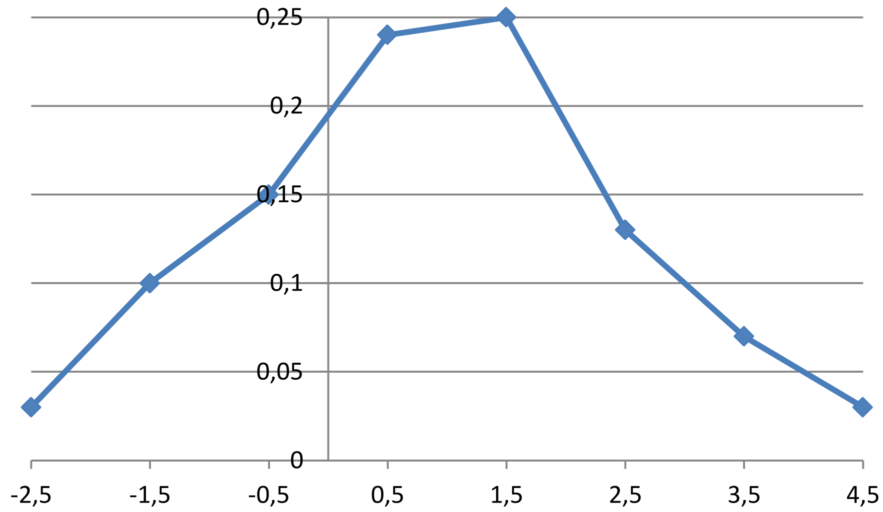


Рис. 1.3.4

Побудувавши полігон або гістограму, можна одержати перше уявлення про форму розподілу, під якою розуміють форму його графіка у межах розподілу, тобто форму кривої розподілу.

Кумулятивну криву (криву сум - *кумуляту*) отримують при зображенні варіаційного ряду з *накопиченими* частотами або відносними частотами у прямокутній системі координат. Накопичена частота даної варіанти отримується підсумовуванням всіх частот варіант, що передують даній, з частотою цієї варіанти. При побудові кумуляти дискретної ознаки у декартовій системі координат будуємо точки, абсциси яких мають значення ознаки (варіанти); ординати – відрізки, довжина яких пропорціональна накопиченій частоті або відносній частоті відповідної варіанти. З'єднавши точки відрізками прямої лінії, отримаємо ламану (криву) кумуляту.

При побудові кумуляти згрупованого розподілу вважається, що нижній межі першого інтервалу відповідає частота, що дорівнює нулю (0), а верхній – вся частота інтервалу. Верхній межі другого інтервалу відповідає накопичена частота перших двох інтервалів, тобто сума частот цих інтервалів, і т. д. Верхній межі останнього (максимального) інтервалу відповідає накопичена частота, що дорівнює сумі всіх частот (об'єму вибірки).

Отже, *полігон* служить для графічного зображення дискретного варіаційного ряду, *гістограма* – інтервального варіаційного ряду розподілу, *огіва* – ранжируваного ряду, *кумулята* – накопичених частот.

1.4. Середні значення сукупностей випадкових величин

У наукових і технічних дослідженнях та на виробництві виникає потреба порівняння показників різних сукупностей за однією й тією ж ознакою. Наприклад, порівняння показників виробництва, стану здоров'я тварин різних ферм, господарств, районів, областей, регіонів або результатів дослідів інших показників. Провести такі порівняння лише за даними варіаційних рядів – складна задача. Вивчення статистичних характеристик розподілів необхідно починати перш за все з простих і найбільш важливих та найчастіше використовуваних у статистичному аналізі показників, а саме - середніх величин і показників варіації, які необхідно розраховувати та аналізувати. Порівняння проводяться за відповідними статистичними показниками, до яких належать наступні типи середніх величин.

Середньою величиною у статистиці називається показник, який характеризує типовий розмір якісної або кількісної ознаки в однорідній сукупності.

У математичній статистиці розрізняють декілька видів середніх величин: арифметичну, квадратичну, кубічну, гармонічну, геометричну та ін.

Середнє арифметичне ознаки x_i позначається через $\overline{x}$ і обчислюється за формулами:

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i}{n}, \qquad \overline{x} = \frac{\sum_{i=1}^{k} n_i x_i}{n}; \qquad (1.4.1)$$

перша з яких називається простою, а друга – зваженою формою. Зауважимо, що у першому виразі сума береться за всіма значеннями x_i вибірки, а у другому – за згрупованими (класовими) значеннями.

Оскільки значення x_i у класі вибірки однакові, то їх вклад буде $n_i x_i$, де n_i - частота відповідного класу варіаційного ряду, а $n = \sum_{i=1}^{k} n_i$ - об'єм вибірки.

Із наведених формул випливає, що $\sum_{i=1}^{k} x_i = n\overline{x}$, $\sum_{i=1}^{k} n_i x_i = n\overline{x}$, тобто сума значень ознаки дорівнює добутку її середнього значення на об'єм вибірки. Вирази $\sum_{i=1}^{k} x_i = n\overline{x}$ та $\sum_{i=1}^{k} n_i x_i = n\overline{x}$ є умовами перевірки правильності розрахунків середніх величин.

Середні арифметичні найчастіше використовуються у порівняльних цілях.

Крім середнього арифметичного, виникає потреба обчислення *середнього арифметичного значення від степеня ознаки* $\overline{x^m}$, яке виражається за формулами:

$$\overline{x^m} = \frac{\sum_{i=1}^{k} x_i^m}{n}, \ \overline{x^m} = \frac{\sum_{i=1}^{k} n_i x_i^{\,m}}{n}; \qquad (1.4.2)$$

перша з яких називається простою, а друга – зваженою формою.

Середнє арифметичне від квадратаознаки $\overline{x^2}$ $(m = 2)$ має вигляд

$$\overline{x^2} = \frac{\sum_{i=1}^{k} x_i^2}{n} = \frac{\sum_{i=1}^{k} n_i x_i^2}{n}. \qquad (1.4.3)$$

Середнє квадратичне є корінь квадратний із середнього арифметичного від квадрата ознаки:

$$\bar{x} = \sqrt{\frac{\sum x^2}{n}}, \text{ або } \quad \bar{x} = \sqrt{\frac{\sum x^2 n_i}{n}} \tag{1.4.4}$$

Цей тип середніх використовується для визначення показників варіації (коливання) – дисперсії, середнього квадратичного відхилення. Середнє квадратичне застосовують також у випадках, якщо ознака виражається площею кола, для обчислення якої треба знайти його радіус або діаметр. Прикладом такої ознаки може бути радіус або діаметр колоній мікробів, висіяних у відповідних середовищах із діагностичною або лікувальною метою.

Якщо позначити радіуси колоній через x_i, загальна їх площа буде $S = \pi \sum_{i=1}^{n} x_i^2$.

Порівняємо загальну площу колоній S із площами:

$$S_{\bar{x}} = \pi n \overline{x^2} = \pi n \left(\frac{\sum_{i=1}^{k} x_i}{n} \right)^2 = \frac{\pi \left(\sum_{i=1}^{k} x_i \right)^2}{n} = \frac{\pi \left(\sum_{i=1}^{k} x_i^2 + 2 \sum_{i=1_{(i \neq i')}}^{k} x_i \sum_{i'=1}^{k} x_{i'} \right)}{n} = \frac{S}{n} + \frac{2}{n} \sum_{i=1_{(i \neq i')}}^{k} x_i \sum_{i'=1}^{k} x_{i'},$$

$S_q = \pi n x_q^2 = \dfrac{\pi n \left(\sum_{i=1}^{k} x_i^2 \right)}{n} = \pi \left(\sum_{i=1}^{k} x_i^2 \right) = S$, які розраховано за арифметичним і квадратичним середнім радіусами відповідно.

З наведених формул випливає, що дійсна загальна площа колонії *S* співпадає з S_q, а $S_{\bar{x}}$ - набагато менша. Таким чином, середню площу колоній треба обчислювати за середнім квадратичним радіусом.

Розглянемо використання середнього квадратичного на конкретному прикладі.

Приклад 1.4.1. Виміри радіусів (мм) колоній мікробів становили: 2,5; 5; 7,5; 10; 12,5. Знайти середню площу колоній.

Розв'язання. Обчислимо арифметичну і квадратичну середні радіусів колоній за формулами $\bar{x} = \frac{\sum_{i=1}^{k} x_i}{n}\pi$ та $x_q = \sqrt{\overline{x^2}}$.

Відповідно маємо: $\bar{x} = \frac{2,5+5+7,5+10+12,5}{5} = 7,5 \quad мм,$

$$x_q = \sqrt{\frac{2,5^2+5^2+7,5^2+10^2+12,5^2}{5}} = 8,3 \quad мм.$$

Використовуючи формули $S = \pi\sum_{i=1}^{k} x_i^2$,

$$S_{\bar{x}} = \pi n\overline{x^2} = \pi n\left(\frac{\sum_{i=1}^{k} x_i}{n}\right)^2 = \frac{\pi\left(\sum_{i=1}^{k} x_i\right)^2}{n} = \frac{\pi\left(\sum_{i=1}^{k} x_i^2 + 2\sum_{i=1_{(i\neq i')}}^{k} x_i\sum_{i'=1}^{k} x_{i'}\right)}{n} = \frac{S}{n} + \frac{2}{n}\sum_{i=1_{(i\neq i')}}^{k} x_i\sum_{i'=1}^{k} x_{i'}$$ і

$$S_q = \pi n x_q^2 = \frac{\pi n\left(\sum_{i=1}^{k} x_i^2\right)}{n} = \pi\left(\sum_{i=1}^{k} x_i^2\right) = S$$, розрахуємо площі S , $S_{\bar{x}}$ і S_q :

$$S = 3,4\cdot(2,5^2+5^2+7,5^2+10^2+12,5^2) = 1081,5 \quad мм^2$$

$$S_{\bar{x}} = 5\cdot 3,4\cdot 7,5^2 = 883 \quad мм^2,\ S_q = 5\cdot 3,4\cdot 8,3^2 = 1081,5 \quad мм^2.$$

Отже, $S = S_q$, що свідчить про те, що середню площу колоній треба обчислювати за середнім квадратичним радіусом.

Середнє кубічне. Цей тип середньої величини позначається через x_Q і обчислюється за формулами:

$$x_Q = \sqrt[3]{\frac{\sum_{i=1}^{k} x_i^3}{n}} = \sqrt[3]{\overline{x^3}},\ x_Q = \sqrt[3]{\frac{\sum_{i=1}^{k} n_i x_i^3}{n}} = \sqrt[3]{\overline{x^3}}, \tag{1.4.5}$$

$\overline{x^3} = \frac{\sum_{i=1}^{k} x_i^3}{n} = \frac{\sum_{i=1}^{k} n_i x_i^3}{n}$ – середнє арифметичне куба ознаки. Середнє кубічне використовується тоді, коли треба визначити середній розмір об'ємних ознак.

Середнє гармонічне – це обернена до середньої арифметичної із обернених значень ознак. Цей тип середньої величини використовується, якщо ознака характеризує середні витрати часу, праці, матеріалів на одиницю продукції; швидкість того чи іншого процесу, наприклад, швидкість доїння корів доярками або швидкість осідання еритроцитів у крові людини або тварини, приросту ваги риб, технологій виробництва тощо.

Середнє гармонічне виражається формулами:

$$x_h = \frac{n}{\sum_{i=1}^{k} \frac{1}{x_i}}, \quad x_h = \frac{n}{\sum_{i=1}^{k} \frac{n_i}{x_i}}. \tag{1.4.6}$$

Приклад 1.4.2. Дві доярки працюють зі швидкостями $x_1 = 30$ л/год. і $x_2 = 20$ л/год. відповідно. Обчислити середню швидкість та середній час доїння літрів молока за одну годину.

Розв'язання. Використовуючи формули $\bar{x} = \frac{\sum_{i=1}^{k} x_i}{n}$ і $\bar{x}_h = \frac{n}{\sum_{i=1}^{k} \frac{1}{x_i}} = \frac{n}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$, обчислимо $\bar{x}$ і $\overline{x_h}$,

$$\bar{x} = \frac{30+20}{2} = 25 \quad л/год, \; x_h = \frac{2}{\frac{1}{30} + \frac{1}{20}} = \frac{120}{5} = 24 \quad л/год.$$

Середнє геометричне. Обчислення середньої геометричної величини ознак проводиться за формулою

$$x_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[k]{\prod_{i=1}^{k} x_i} = \sqrt[k]{\prod_{i=1}^{k} x_i^{n_i}}. \tag{1.4.5}$$

Цей тип середньої величини застосовується, якщо ознака характеризує темп росту з часом (динамічний процес), наприклад, показників виробництва впродовж декількох років або швидкості осідання еритроцитів за відповідні проміжки часу.

Увівши початкове значення x_0 і відносні зміни ознаки $k_i = \frac{x_i}{x_{i-1}}$. за відповідні проміжки часу, отримаємо $x_i = k_i \cdot x_{i-1} = k_i \cdot k_{i-1} \cdot \ldots \cdot k_1 \cdot x_0$. Значення ознаки після n-го проміжку часу буде $x_n = k_n \cdot k_{n-1} \cdot \ldots \cdot k_1 \cdot x_0$

Якщо ввести середнє значення відносної зміни ознаки k_c, то вираз $x_n = k_n \cdot k_{n-1} \cdot \ldots \cdot k_1 \cdot x_0$ матиме вигляд $x_n = k_c^n \cdot x_0$. Із рівності правих сторін цих виразів випливає, що $k_n \cdot k_{n-1} \cdot \ldots \cdot k_1 \cdot x_0 = k_c^n \cdot x_0$, звідки

$$k_c = k_g = \sqrt[n]{k_n \cdot k_{n-1} \cdot \ldots \cdot k_1} = \sqrt[n]{\prod_{i=1}^{n} k_i}.. \qquad (1.4.6)$$

Таким чином, середня відносна зміна ознаки k_c визначається за формулою середнього геометричного відносних змін k_i.

Формула (1.4.6) використовується у випадках, коли відомі відносні зміни ознаки k_i, здебільшого для обчислення середніх коефіцієнтів (темпів) зростання у рядах динаміки.

Приклад 1.4.3. Унаслідок інфляції споживчі ціни за чотири роки зросли у 2,8 рази, зокрема: за перший рік – у 1,7 рази, за другий – у 1,3, за третій – у 1,1, за четвертий – у 1,15 рази. Який середньорічний темп зростання цін?

Розв'язання. Середнє арифметичне

$$\frac{1{,}7 + 1{,}3 + 1{,}1 + 1{,}15}{4} = 1{,}3125$$

не забезпечує визначеної властивості, оскільки за чотири роки за цією середньою ціни зросли б у 1,3125*1,3125*1,3125*1,3125=2,968 рази, а не у 2,8 рази. Визначену властивість забезпечує середнє геометричне:

$$\bar{x} = \sqrt[4]{1{,}7 \cdot 1{,}3 \cdot 1{,}1 \cdot 1{,}15} = \sqrt[4]{2{,}8} = 1{,}293.$$

Отже, споживчі ціни в середньому щороку зростали в 1,293 рази.

1.5. Числові характеристики генеральної та вибіркової сукупностей та їх властивості

У теорії ймовірностей велику роль відіграють різні числові характеристики випадкової величини: математичне сподівання, дисперсія, початкові та центральні моменти різних порядків. Аналогічні числові характеристики існують і для статистичних розподілів, причому кожній числовій характеристиці випадкової величини X відповідає її статистичний аналог. Оскільки ці аналоги обчисляються за даними вибірки, то їх називають *статистичними характеристиками* або *оцінками*.

Генеральною середньою називається середнє арифметичне значення ознаки генеральної сукупності, що вивчається. Позначення генеральної середньої *а* або $\overline{x_{Г.}}$

Якщо всі значення кількісної ознаки, що вивчається, різні, то генеральна середня дорівнює

$$\overline{x_{Г.}} = \frac{x_1 N_1 + x_2 N_2 + ... + x_k N_k}{N}, \qquad (1.5.1)$$

де x_1 спостерігається N_1 раз, x_2 спостерігається N_2 раз, ..., x_k спостерігається N_k раз; N - об'єм генеральної сукупності, $N = N_1 + N_2 + \ldots + N_k$. Цей результат справедливий, коли $N_1, N_2, \ldots, N_k$ однакові та коли вони різні.

У випадку неперервного розподілу деякої кількісної ознаки вважають, що $M(X) = \overline{x_{Г.}}$

Нехай для вивчення генеральної сукупності відносно деякої кількісної ознаки X зроблена вибірка об'ємом *n*. *Вибірковою середньою* називається середнє арифметичне значення ознаки вибірки даної генеральної сукупності. Якщо значення x_1, x_2, ..., x_k ознаки X вибірки об'ємом n різні між собою, то вибіркову середню знаходять за формулою

$$\overline{x}_в = \frac{\sum x_i}{n}. \tag{1.5.2}$$

Якщо значення x_1 зустрічається з частотою n_1, значення x_2 – з частотою n_2, ..., значення x_k – з частотою n_k, то середнє вибіркове обчислюють за формулою «зваженої середньої»

$$\overline{x}_в = \frac{\sum_{i=1}^{k} n_i \cdot x_i}{n} \tag{1.5.3}$$

де x_i - варіанта, n_i – частота, n - об'єм вибірки.

Генеральне (вибіркове) середнє є своєрідним центром, навколо якого «розкидані» всі випадки.

Для характеристики розсіювання значень кількісної ознаки X генеральної сукупності навколо свого середнього генерального значення, вводять поняття генеральної дисперсії.

Генеральною дисперсією ($D_Г$) називається середнє арифметичне квадратів відхилень значень ознаки X генеральної сукупності від його середнього генерального значення $\overline{x_{Г.}}$

$$D_{Г.} = \frac{\sum_{i=1}^{k}\left(x_i - \overline{x}_Г\right)^2}{N}. \tag{1.5.4}$$

Якщо варіанти x_i з'являються відповідно N_1, N_2,..., N_K разів, то дисперсія $D_{Г.}$ виражається формулою зваженої дисперсії:

$$D_{Г.} = \frac{\sum_{i=1}^{k} N_i\,(x_i - \overline{x}_Г)^2}{N}. \tag{1.5.5}$$

Для того, щоб охарактеризувати розсіювання дослідних значень кількісної ознаки вибірки навколо вибіркової середньої, вводять вибіркову дисперсію, яка зв'язана з відхиленням.

Вибірковою дисперсією *($D_в$)* називається середнє арифметичне квадратів відхилень спостережуваних значень ознаки розподілу *X* від вибіркової середньої. Якщо значення $x_1, x_2, x_3, \ldots, x_k$ кількісної ознаки *X* вибірки об'ємом *n* мають відповідні частоти $n_1, n_2, \ldots, n_k$, тоді обчислюють зважену дисперсію:

$$D_{в.}=\frac{\sum\limits_{i=1}^{k} n_i\left(x_i-\overline{x_в}\right)^2}{n}. \tag{1.5.6}$$

Перетворимо формулу (1.5.5) з врахуванням формули (1.5.1). Маємо:

$$D_{в.}=\frac{\sum\limits_{i=1}^{k} n_i\left(x_i-\overline{x_в}\right)^2}{n}=\frac{\sum\limits_{i=1}^{k} n_i(x_i^2-2x_i\overline{x_в}+\overline{x_в}^2)}{n}=\frac{\sum\limits_{i=1}^{k} n_i x_i^2-2\sum\limits_{i=1}^{k} x_i n_i\overline{x_в}+\sum\limits_{i=1}^{k} n_i\overline{x_в}^2}{n}=$$

$$\frac{\sum\limits_{i=1}^{r} n_i x_i^2}{n}-2\overline{x_в}^2+\overline{x_в}^2=\frac{\sum\limits_{i=1}^{k} x_i^2 n_i}{n}-\overline{x_в}^2.$$

Отже, вибіркову дисперсію також можна обчислити за такою формулою

$$D_{в.}=\frac{\sum\limits_{i=1}^{k} x_i^2 n_i}{n}-\overline{x_в}^2. \tag{1.5.7}$$

Дисперсія випадкової величини є зручною характеристикою розсіювання (варіації), але вона має той недолік, що її розмірність дорівнює квадрату випадкової величини (ознаки). Для більшої зручності вводиться характеристика, що має розмірність випадкової величини (ознаки), а саме – середнє квадратичне відхилення.

Генеральне середнє квадратичне відхилення σ, яке називається ще *стандартним відхиленням (стандартом)*, обчислюється за такою формулою:

$$\sigma=\sqrt{D_Г}=\sqrt{\frac{\sum\limits_{i=1}^{k}(x_i-\overline{x_Г})^2}{N}}=\sqrt{\frac{\sum\limits_{i=1}^{k}N_i(x_i-\overline{x_Г})^2}{N}}. \tag{1.5.8}$$

Розкривши у формулах (1.5.5) $D_{Г.}=\frac{\sum_{i=1}^{k}N_i(x_i-\overline{x}_Г)^2}{N}$ і (1.5.8) $\sigma=\sqrt{D_{Г.}}=\sqrt{\frac{\sum_{i=1}^{k}(x_i-\overline{x_Г})^2}{N}}=\sqrt{\frac{\sum_{i=1}^{k}N_i(x_i-\overline{x_Г})^2}{N}}$ квадрат різниці $(x_i-\overline{x_Г})^2$, маємо:

$$D_Г=\frac{\sum_{i=1}^{k}(x_i^2-2\overline{x}x_i+\overline{x}^2)}{n}=\frac{\sum_{i=1}^{k}x_i^2}{n}-2\overline{x}\frac{\sum_{i=1}^{k}x_i}{n}+\frac{n\overline{x}^2}{n}=\overline{x^2}-2\overline{xx_i}+\overline{x}^2=\overline{x^2}-2\overline{x}^2+\overline{x}^2=\overline{x^2}-\overline{x}^2.$$

Отже,

$$\sigma=\sqrt{\overline{x^2}-\overline{x}^2}. \tag{1.5.9}$$

Дисперсія на практиці ототожнюється з похибкою, а середнє квадратичне відхилення вибірки - з середньою похибкою.

Таким чином, вибіркові дисперсія $Dв.$ і середнє квадратичне відхилення σ є наближеними показниками характеристики генеральної сукупності, оскільки вони залежать від об'єму вибірки n, що зумовлює систематичну похибку. Зі збільшенням об'єму вибірки ці показники зменшуються, а тому вирази $Dв=\frac{\sum_{i=1}^{k}\left(x_i-\overline{x}\right)^2}{n}=\frac{\sum_{i=1}^{k}n_i(x_i-\overline{x})^2}{n}$ і $\sigma=\sqrt{D}=\sqrt{\frac{\sum_{i=1}^{k}(x_i-\overline{x})^2}{n}}=\sqrt{\frac{\sum_{i=1}^{k}n_i(x_i-\overline{x})^2}{n}}$ називаються *зсуненою дисперсією* та *зсуненим середнім квадратичним відхиленням*. З метою зменшення систематичних похибок у цих формулах *n* замінюється на *n*-1. У такій формі запису матимемо *незсунені* D і σ, які при великих об'ємах вибірки співпадають зі *зсуненими*.

Незсуненими оцінками генеральної дисперсії та генерального середнього квадратичного відхилення є відповідно *виправлена дисперсія* s^2 та *виправлене середнє квадратичне відхилення* s, які обчислюються за формулами відповідно:

$$s^2=\frac{n}{n-1}\cdot D_в \tag{1.5.10}$$

$$s=\sqrt{\frac{n}{n-1}\cdot D_в} \qquad (1.5.11)$$

Якщо у (1.5.9) підставити значення $D_в$ з (1.5.5), то отримаємо:

$$s^2=\frac{1}{n-1}\cdot\sum_{i=1}^{k} n_i(x_i-\overline{x_в})^2 \qquad (1.5.10')$$

Зауважимо, що при збільшенні числа спостережень вибірки всі статистичні характеристики будуть збігатись за ймовірністю до відповідних числових характеристик генеральної сукупності.

Приклад 1.5.1. Знайти вибіркову та виправлену дисперсії за даним розподілом вибірки об'ємом *n*=50:

x_i	18	19	20	21
n_i	5	10	20	15

Розв'язання. Знайдемо спочатку вибіркову середню за формулою (1.5.3): $\overline{x_в}=\frac{\Sigma n_i\cdot x_i}{n}$.

$$\overline{x_в}=\frac{18\cdot5+19\cdot20+20\cdot20+21\cdot15}{50}=\frac{90+380+400+315}{50}=23,7;$$

за формулою (1.5.7) обчислимо дисперсію:

$$D_в=\frac{\sum_{i=1}^{k}x_i^2\cdot n_i}{n}-(\overline{x_в})^2=\frac{18^2\cdot5+19^2\cdot10+20^2\cdot20+21^2\cdot15}{50}-(23,7)^2=0,89.$$

За формулою (1.5.10) обчислимо виправлену дисперсію:

$$s^2=\frac{50}{49}\cdot0,89=0,908.$$

Основні властивості вибіркової (генеральної) середньої та дисперсії

1. $\overline{C}=C,\ \ D(C)=0$, де С – стала.

2. Якщо кожну варіанту змінити на С одиниць, то середнє вибіркове зміниться на таку ж кількість одиниць, а дисперсія не зміниться:

$$\overline{x+C} = \bar{x} + C, \quad D(x+C) = D(x).$$

3. Якщо кожну варіанту змінити у k разів, то й середнє вибіркове зміниться у k разів, а дисперсія – у k^2 разів:

$$\overline{kx} = k\bar{x}, \quad D(kx) = k^2 D(x).$$

4. $\overline{x \pm y} = \bar{x} \pm \bar{y}, \quad D(x \pm y) = D(x) + D(y)$.

5. $\overline{x \cdot y} = \bar{x} \cdot \bar{y}$.

Часто при дослідженнях вибірку розбивають на групи, тоді характеристики варіації вибірки обчислюють за такими формулами.

Групову дисперсію обчислюють за формулою

$D_{г.р.} = \dfrac{\sum_{i=1}^{k} n_i \left(x_i - \overline{x_j}\right)^2}{N_j}$; $Nj = \sum_{i=1}^{k} n_i$ - об’єм групи, $\overline{x_j}$ - групова середня.

Внутрішньогрупову дисперсію знаходять за формулою

$D_{вгр.} = \dfrac{N_1 D_{1гр.} + N_2 D_{2гр.}}{n}$; де $n=N_1 + N_2$ - об'єм вибірки.

Міжгрупова дисперсія дорівнює

$D_{міжгр.} = \dfrac{\sum_{i=1}^{k} Ni\left(\bar{x} - x_i\right)^2}{n}$, де $\bar{x}$ - загальна середня.

Загальну дисперсію обчислюють за формулою

$D_{заг.} = D_{внгр.} + D_{міжгр.}$.

Приклад 1.5.2. Знайти внутрішньогрупову, міжгрупову та загальну дисперсії сукупності, яка складається з наступних двох груп:

перша група	x_i	2	4	5
	n_i	1	7	2

друга група	x_i	3	8
	n_i	2	3

Розв'язання. Спочатку знайдемо групові середні та дисперсії обох груп:

$$\overline{x_i}=\frac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i},\ \overline{x_1}=\frac{2\cdot 1+4\cdot 7+5\cdot 2}{10}=\frac{40}{10}=4,\ \overline{x_2}=\frac{3\cdot 2+8\cdot 3}{5}=6;$$

$$D_{1гр.}=\frac{\sum_{i=1}^{k} n_i\left(x_i-\overline{x_1}\right)^2}{N_j}=\frac{1(2-4)^2+7(4-4)^2+2(5-4)^2}{10}=0,6;$$

$$D_{2гр.}=\frac{2(3-6)^2+3(8-6)^2}{5}=6;$$

$$D_{вгр.}=\frac{\sum_{j=1}^{2} N_j D_{jгр.}}{n}=\frac{10\cdot 0.6+5\cdot 6}{15}=\frac{12}{5}=2,4.$$

$$D_{міжгр.}=\frac{\sum_{j=1}^{2} N_j\left(\overline{x_j}-\overline{x}\right)^2}{n}=\frac{10(4-\frac{14}{3})^2+5(6-\frac{14}{3})^2}{15}=\frac{8}{9},$$

де $\overline{x}=\frac{1\cdot 2+7\cdot 4+2\cdot 5+2\cdot 3+3\cdot 8}{15}=\frac{14}{3}.$

$$D_{заг.}=D_{вгр.}+D_{міжгр.}=\frac{12}{5}+\frac{8}{9}=\frac{148}{45}=3\frac{13}{45}.$$

$$D_{заг.}=\frac{1\cdot(2-\frac{14}{3})^2+7(4-\frac{14}{3})^2+2(5-\frac{14}{3})^2+2(3-\frac{14}{3})^2+3(8-\frac{14}{3})^2}{15}=\frac{148}{45}=3\frac{13}{45}.$$

1.6. Метод хибного (несправжнього) нуля. Умовні варіанти

Якщо значення варіант x_i- великі числа, то для зменшення громіздкості обчислень можна ввести умовні варіанти u_i виду

$$u_i = \frac{x_i - C}{h}, \qquad (1.6.1)$$

де *C* - константа, яка має спеціальну назву «хибний нуль». Значення "хибного нуля" вибирається самостійно в залежності від конкретної задачі.В якості хибного нуля *C* можна взяти або варіанту, яка має найбільшу частоту (моду), або середину варіаційного ряду (медіану), або, заокруглене до цілого, середнє арифметичне значень варіант. *h* – крок (різниця між двома сусідніми варіантами).

Знайдемо формулу переходу від умовних варіант u_i до варіант x_i. Для цього перетворимо формулу $\overline{x_в} = \frac{\Sigma n_i \cdot x_i}{n}$ наступним чином:

$$\overline{x_в} = \frac{1}{n} \frac{\sum_{i=1}^{k} (x_i - C + C)}{h} hn_i = C + h \cdot \overline{u_в}. \qquad (1.6.2)$$

Використовуючи властивості дисперсії, знайдемо формулу переходу від дисперсії, що обчислена за умовною варіантою u_i, до дисперсії змінної *X.*

$$D\left(x\right) = D(C + h \cdot u) = D(C) + D(h \cdot u) = 0 + h^2 D\left(u\right) = h^2 D(u). \qquad (1.6.3)$$

Отже, дисперсія змінної *X* не залежить від значення хибного нуля, а залежить тільки від значення кроку h.

Приклад 1.6.1. Знайти вибіркову середню та вибіркову дисперсію за даним розподілом вибірки об'єму n=10.

x_i	1250	1270	1280
n_i	2	5	3

Розв'язання. Варіанти – великі числа, тому перейдемо до умовних варіант $u_i = \frac{x_i - C}{h}$, де *C* =1270 – хибний (несправжній) нуль, *h* = 10, $u_i = \frac{x_i - 1270}{10}$.

Отримаємо розподіл умовних варіант

u_i	-2	0	1
n_i	2	5	3

$$\overline{u_в} = \frac{-2\cdot 2 + 0\cdot 5 + 1\cdot 3}{10} = \frac{-4+3}{10} = -\frac{1}{10},$$

звідки $\overline{x_в} = C + h\cdot\overline{u_в} = 1270 + 10\left(-\frac{1}{10}\right) = 1270 - 1 = 1269.$

$$D\ (u)\ =\ \frac{\sum_{i=1}^{k} n_i\left(u_i - \overline{u_в}\right)^2}{n} = \frac{2(-2+0,1)^2 + 5(0+0,1)^2 + 3(1+0,1)^2}{10} = 1,09.$$

Із формули (1.6.3) маємо, що *D (x)* = $h^2 D(u)$=10^2·1,09 =109.

Приклад 1.6.2. З плодового дерева випадково зірвано 10 плодів, їх маси подані в таблиці. Опрацювати статистичні дані вибірки:

i	1	2	3	4	5	6	7	8	9	10
x	225	274	305	253	220	245	211	234	230	231

Розв'язання. Варіанти – великі числа, тому перейдемо до умовних варіант $u_i = x_i - C$, C=250. Створимо розрахункову таблицю:

i	1	2	3	4	5	6	7	8	9	10	сума
x	225	274	305	253	220	245	211	234	230	231	2428
x-c	-25	24	55	3	-30	-5	-39	-16	-20	-19	-72
$(x-c)^2$	625	576	3025	9	900	25	1521	256	400	361	2356

$\overline{u} = \frac{-72}{10} = -7{,}2;\quad \mathrm{D(u)} = 235{,}6 - (-7{,}2)^2 = 183{,}76;\ \sigma(u) = \sqrt{183{,}76} \approx 13{,}56$.

Перейдемо до варіанти *x*: $\overline{x} = \overline{u} + C;\ D(x) = D(u);\ \sigma(x) = \sigma(u)$; тоді $\overline{x} = 242{,}8$; *D* = 183,76; $\sigma_в = 13{,}56$.

Отже, середня маса плода становить 242,8 г з похибкою 183,76 г^2 , середньою похибкою 13,56 г, тобто середня маса плода становить 242,8 г$\pm$3,56 г.

Якщо значення варіант x_i- малі числа, то для зменшення громіздкості обчислень можна ввести умовні варіанти u_i виду

$$\boldsymbol{u_i = kx_i} \tag{1.6.1}$$

де kϵR. Обчисливши числові характеристики для умовних варіант і скориставшись їхніми властивостями, переходять до числових характеристик варіанти x_i .

Приклад 1.6.3. Знайти вибіркову середню та вибіркову дисперсію за даним розподілом вибірки

x_i	0,04	0,06	0,09
n_i	3	5	2

Розв'язання. Варіанти – малі числа, тому перейдемо до умовних варіант $u_i = kx_i$, *k*=100. Створимо розрахункову таблицю:

i	1	2	3	Разом
x_i	0,04	0,06	0,09	-
n_i	3	5	2	10
$u_i = 100x_i$	4	6	9	19
u_i^2	16	36	81	133

$\overline{u} = \frac{19}{10} = 1{,}9;\ D(u) = 133 - (1{,}9)^2 = 129{,}39;\ \sigma(u) = \sqrt{129{,}39} \approx 11{,}375$.

Перейдемо до варіанти *x*, скориставшись властивостями числових характеристик: $\overline{x} = \frac{\overline{u}}{100}$; $D(x) = \frac{D(u)}{100^2}$; $\sigma(x) = \frac{\sigma(u)}{100}$; тоді $\overline{x} = 0{,}0019$; $D(x) = 0{,}012939 \approx 0{,}013$; $\sigma_в = 0{,}11375 \approx 0{,}114$.

1.7. Варіаційний ряд та його числові характеристики

Варіаційним рядом розподілу називається впорядкована статистична сукупність, у якій значення варіант розташовані в ранжируваний ряд і зазначенням для кожного інтервалу (групи) відповідних частот (частостей).

Всебічна характеристика рядів розподілу передбачає з'ясування умов, під впливом яких сформувався досліджуваний розподіл, вираження його основних особливостей числовими характеристиками.

Комплексний опис статистичних розподілів полягає в знаходженні перш за все таких найважливіших узагальнюючих характеристик: центра розподілу, ступеня варіації ознаки розподілу. Для їх визначення використовують відповідні кількісні характеристики. Так, для знаходження центра розподілу визначають різні типи середніх величин - моду та медіану; ступеня варіації - розмах варіації, середнє лінійне відхилення, дисперсію, середнє квадратичне відхилення, коефіцієнт варіації та інші показники.

Отже, *числовими характеристиками варіаційних рядів* є межі, розмах (інтервал), дисперсія, середнє квадратичне відхилення, коефіцієнт варіації, нормоване відхилення, мода, медіана, ексцес, асиметрія. Середні величини не дають чіткої характеристики генеральних сукупностей. Сукупності можуть мати однакові середні значення однієї й тієї ж ознаки, але відрізнятися за характером і величиною варіювання. Крім середніх величин, для повної характеристики генеральних сукупностей використовуються також інші показники варіаційних рядів.

Варіацією ознак називається наявність відмінностей у числових значеннях ознак одиниць сукупності. Основними показниками, що характеризують варіацію ознаки, є: розмах варіації (інтервал), середнє лінійне відхилення, дисперсія, середнє квадратичне відхилення, квартальне відхилення, коефіцієнт варіації та інші.

Межами варіаційних рядів називаються мінімальне $x_{\min}$ і максимальне $x_{\max}$ значення ознаки.

Розмах варіаційних рядів (інтервал) - це показник , який позначається через R і дорівнює різниці між максимальним $x_{\max}$ і мінімальним $x_{\min}$ значеннями ознаки варіаційних рядів

$$R = x_{\max} - x_{\min}. \tag{1.7.1}$$

Особливість цього показника полягає в тому, що він залежить від двох крайніх значень ознаки та не враховує проміжних значень і частот. Більш досконалою мірою виміру варіації є середнє лінійне та середнє квадратичне відхилення, які усувають указані вище недоліки розмаху варіації.

Дисперсія та середнє квадратичне (стандартне) відхилення характеризують розсіювання (відхилення) значень ознаки x_i відносно її середньої величини $\bar{x}$.

Розмах, дисперсія, середнє квадратичне відхилення є абсолютними показниками варіаційних рядів ознаки генеральних сукупностей. У зв’язку з цим вони не використовуються для порівняння *варіації* різноманітних ознак або однієї і тієї ж ознаки різних сукупностей.

Коефіцієнт варіації використовують для порівняння ступеня розсіювання значень різних ознак відносно своїх вибіркових середніх та варіювання однієї і тієї ж ознаки у різних вибірках. Коефіцієнт варіації обчислюється за формулою

$$V = \frac{\sigma}{\bar{x}}, \text{ (у долях одиниці)} \qquad (1.7.2)$$

або $V = \frac{\sigma}{\bar{x}} \cdot 100\%$ (у відсотках). (1.7.2')

Якщо підставити $\sigma = \sqrt{\overline{x^2} - \bar{x}^2}$ у $V = \frac{\sigma}{\bar{x}}$, отримаємо

$$V = \frac{\sqrt{\overline{x^2} - \bar{x}^2}}{\bar{x}} = \sqrt{\frac{\overline{x^2}}{\bar{x}^2} - 1}.$$

Приклад 1.7.1. Використовуючи інтервальний ряд розподілу, зробити розрахунок показників варіації:

Інтервали	14-18	18-22	22-26	26-30	30-34	34-38	38-42
n_i	9	15	16	24	18	12	6

Розв'язання. Дані для розрахунків наведемо у розрахунковій таблиці:

Інтервали	Середини інтервалів x_i	Частоти n_i	$x_i n_i$	$x^2_i n_i$
14-18	16	9	144	2304
18-22	20	15	300	6000
22-26	24	16	384	9216
26-30	28	24	672	18816
30-34	32	18	576	18432
34-38	36	12	432	15552
39-42	40	6	240	9600
Всього	-	100	2748	79920
Середні	-	-	27,48	799,2

Маємо: розмах варіації $R = x_{\max} - x_{\min}$. R=42-14=28;

середнє вибіркове $\overline{x} = \frac{2748}{100} = 27,48$;

дисперсія $D(x) = 799,2 - (27,48)^2 = 44,0496 \approx 44,05$;

середнє квадратичне (стандартне) відхилення $\sigma(x) = \sqrt{44,0496} \approx 6,63$,

конфіцієнт варіації $V = \frac{\sigma}{\overline{x}}, V = \frac{6,63}{27,48} = 0,2417$ або 24, 2%;

середнє лінійне відхилення $\bar{l} = \frac{\sum |x_i - \overline{x}| n_i}{100} = \frac{542,4}{100} = 5,424$.

Отже, істинне значення за даною сукупністю коливається у межах ±6,63, або на 24,2% від середньої.

Нормовані відхилення t_i визначаються за формулою

$$t_i = \frac{x_i - \overline{x}}{\sigma} \text{ (або } t_i = \left(\frac{x_i - \overline{x}}{\sigma}\right) \cdot 100\%) \quad (1.7.3)$$

і являють собою відхилення значень ознаки відносно середньої величини у відносних одиницях або у відсотках.

Використавши вираз $\sigma = \sqrt{\overline{x^2} - \overline{x}^2}$, нормовані відхилення $t_i = \frac{x_i - \overline{x}}{\sigma}$ запишуться у вигляді

$$t_i = \frac{x_i - \overline{x}}{\sqrt{\overline{x^2} - \overline{x}^2}}. \quad (1.7.3')$$

Нормовані відхилення використовуються для порівняння ознак однакової або різної природи у різних сукупностях.

З варіаційного ряду будується *кумулятивний* (нагромаджувальний, накопичений) ряд шляхом послідовного додавання частот: до частоти першого класу додається частота другого класу; до одержаної суми додають частоту третього класу і т.д.

Модою називається значення ознаки, яке має найбільшу частоту в статистичному ряду розподілу. У дискретних варіаційних рядах мода визначається без додаткових розрахунків за значенням варіанти, яка має найбільшу частоту. В інтервальних варіаційних рядах розподілу мода визначається за формулою

$$Mo = x_0 + h\frac{n_2 - n_1}{(n_2 - n_1) + (n_2 - n_3)}. \qquad (1.7.4)$$

де x_0 - нижня (мінімальна) межа модального інтервалу; h - величина інтервалу; n_1 - частота інтервалу, який передує модальному; n_2 - частота модального інтервалу; n_3 - частота інтервалу, що є наступним за модальним.

Медіаною називається таке значення ознаки, яке ділить ранжируваний ряд розподілу на дві рівні частини, тобто значення, яке знаходиться у середині ряду розподілу. Якщо в дискретному варіаційному ряду *2m* + 1 випадків, то значення ознаки для випадку *m* + 1 буде медіанним. Якщо в ряду парне число *2m* випадків, то медіану визначають як середню арифметичну величину з двох серединних значень.

Для обчислення медіани з непарною та парною кількістю варіант використовують такі формули:

$$Me = x_{m+1}; \qquad\qquad Me = \frac{x_{m+}x_{m-1}}{2}.$$

В інтервальних варіаційних рядах розподілу медіана обчислюється для середини медіанного інтервалу, за який приймається такий, де сума накопичених частот перевищує половину значень частот ряду розподілу, за формулою

$$Me = x_0 + h\frac{0{,}5\sum n_i - s_{x_0}}{n_m} \qquad (1.7.5)$$

де x_0 - нижня (мінімальна) межа медіанного інтервалу; h - величина медіанного інтервалу; $0{,}5\sum n_i$ – половина суми накопичених частот інтервального ряду; s_{x_0} – сума накопичених частот перед медіанним інтервалом; n_m - частота медіанного інтервалу.

Мода та медіана є структурними (позиційними) середніми. Їхні величини залежать лише від характеру частот, тобто від структури розподілу. На відміну від інших середніх, які залежать від усіх значень ознаки, мода і медіана не залежать від крайніх її значень. Це особливо важливо для незакритих крайніх інтервалів варіаційних рядів розподілу.

Для побудови інтервального варіаційного ряду розподілу необхідно визначити кількість груп і величину інтервалу. Кількість груп залежить від загальної чисельності одиниць сукупності та характеру групувальної ознаки.

Якщо ж інтервальний ряд будується за кількісною ознакою, яка змінюється неперервно і набуває в певних межах будь-які дробові значення, то при виділенні груп необхідно за кількісними змінами встановити якісні переходи для того, щоб виділити якісно відмінні одна від одної групи.

Величина інтервалу визначається за формулою

$$h = \frac{Xmax - Xmin}{n}, \tag{1.7.6}$$

де x_{max}, x_{min} - максимальне та мінімальне значення ознаки.

Ознаки елементів ряду можуть приймати як дискретні (окремі, ізольовані одне від одного числа), так і неперервні (будь-які числа з певного інтервалу) значення.

Залежно від того, дискретна чи неперервна ознака, у вузькому чи широкому інтервалі вона змінюються, матимемо безінтервальні та інтервальні ряди відповідно.

У безінтервальних рядах частота n_i відноситься до конкретних значень ознаки x_i , а в інтервальних – до окремих інтервалів, на які розбивається значення ознаки в межах від $x_{\min}$ до $x_{\max}$.

Приклад 1.7.2. Використовуючи умову прикладу 1.7.1, обчмслити медіану і моду.

Розв'язання. Складемо розрахункову таблицю, за якою знайдемо медіанний інтервал, для чого спрчатку знайдемо накопичувальні частоти:

Інтервали	Частоти n_i	Накопичувальні частоти
14-18	9	9
18-22	15	24
22-26	16	40
26-30	24	64
30-34	18	86
34-38	12	94
39-42	6	100

Медіанним є інтервал 26-30, оскільки на цей інтервал припадаєперша накопичувальна частота, що перевищує половину всього обєму сукупності (64 перевищує $\frac{\sum n_i}{2} = (100/2)$ =50). Розрахуємо медіану за формулою (1.7.5):

$$Me = 26 + 4\frac{50-40}{24} = 27{,}67,$$

де 26 – нижня межа медіанного інтервала; *h*=4 – крок (довжина інтервала); $0{,}5\sum n_i = 50$ – половина суми накопичених частот; 40 – сума накопичених частот інтервалу, що передує медіанному; 24 – частота медіанного інтервалу.

Розрахуємо модальне значення за формулою (1.7.4). Модальним є інтервал (26-30), оскільки він має найбільшу частоту.

$$Mo = 26 + 4\frac{24-16}{(24-16)+(14-18)} = 34,$$

де 26 - нижня межа модального інтервалу; *h*=4 – крок (довжина модального інтервала); $n_1 = 16$; $n_2 = 24$; $n_3 = 18$ - відповідно частот передмодального, модального і післямодального інтервалів.

Питання для самоперевірки

1. Що є предметом математичної статистики?
2. Вказати основні задачі математичної статистики.
3. Що називають статистичною, вибірковою та генеральною сукупністю, об'ємом (обсягом) цих сукупностей?
4. Що називається статистичним розподілом вибірки?
5. Емпірична функція та її властивості.
6. Графічне зображення варіаційних рядів для незгрупованих і згрупованих даних.
7. Який імовірнісний зміст мають гістограма та полігон частот для згрупованих даних вибірки?
8. Перелічити середні значення числових сукупностей.
9. Для чого потрібні середні величини?
10. Яка різниця між зваженими та простими середніми?
11. Коли використовують середнє арифметичне (формула)?
12. Коли використовують середнє квадратичне (формула)?
13. Коли використовують середнє гармонічне (формула)?
14. Коли використовують середнє геометричне (формула)?
15. Числові характеристики генеральної та вибіркової сукупностей.
16. За якими формулами обчислюють числові характеристики вибірки, яка розбивається на декілька груп?
17. У чому суть методу хибного нуля?
18. Варіаційний ряд та його числові характеристики.

19. Як розраховується розмах варіації? Яку він має суть?

20. Як розраховується дисперсія та середнє квадратичне відхилення, їхсуть?

21. Як розраховується коефіцієнт варіації і його суть?

22. Як розраховується нормоване відхилення і її суть?

23. Який показник варіації використовують для співставлень рівнів змінюваності однієї і тієї ж ознаки в різних вибірках?

24. Чи можна використовувати середнє квадратичне відхилення як порівняльну характеристику двох різних вибірок? Чому?

Задачі та вправи

1.1. Вибірка задана у вигляді розподілу частот:

а)

x_i	2	5	7
n_i	1	3	6

б)

x_i	4	7	8
n_i	5	2	3

Знайти розподіл відносних частот (частостей).

1.2. Знайти емпіричну функцію для даного розподілу вибірки та побудувати її графік:

а)

x_i	2	6	10
n_i	12	18	30

б)

x_i	2	5	7	8
n_i	4	3	1	2

1.2. Побудувати полігон частот за даним розподілом вибірки:

а)

x_i	2	3	5	6
n_i	10	15	5	20

б)

x_i	1	4	5	7
n_i	20	10	14	6

1.4. Побудувати полігон і гістограму відносних частот за даним розподілом вибірки:

а)

x_i	10	50	65	80
w_i	0,1	0,2	0,3	0,4

б)

x_i	2	4	5	7	10
w_i	0,15	0,2	0,1	0,1	0,45

1.5. Побудувати статистичний ряд і полігон для розподілу розмірів 45 пар чоловічого взуття, які були продані у магазині за день:

39 41 42 40 42 41 40 42 44 40 43 42
41 43 39 42 41 39 41 37 43 41 38 43
41 38 41 43 40 41 38 44 39 41 40 42
40 41 43 42 42 40 42 41 38.

1.6. Побудувати гістограму частот за даним розподілом вибірки:

а)

Номер інтервалу	Частковий інтервал X_i-X_{i+1}	Сума частот варіант інтервалу n_i.
1	2-7	5
2	7-12	10
3	12-17	25
4	17-22	6
5	22-27	4

б)

Номер інтервалу	Частковий інтервал Xi -Xi+1	Сума частот варіант інтервалу n_i
1	1-5	10
2	5-9	20
3	9-13	50
4	13 -17	12
5	17-21	8

1.7. У результаті випробувань випадкова величина *X* набуває таких значень: 2; 5; 7; 1; 10; 5; 9; 6; 8; 6; 2; 3; 7; 6; 8; 3; 8; 10; 6; 7; 3; 9; 4; 5; 6. Знайти статистичний розподіл вибірки та побудувати полігон частот.

1.8. У результаті випробувань випадкова величина *X* набуває таких значень: 16; 17; 9; 13; 21; 11; 7; 7; 19; 5; 17; 5; 20; 18; 11; 4; 6; 22; 21; 15; 15; 23; 19; 25; 11. Знайти таблицю статистичного розподілу вибірки, розбивши проміжок [0; 25] на п'ять розрядів, які мають одинакові довжини, та побудувати гістограму частот.

1.9. Знайти вибіркову середню за даним розподілом вибірки:

а)

x_i	250	1270	280	1320
n_i	2	15	3	10

б)

x_i	86	192	194	196	200
n_i	3	5	35	4	3

в)

x_i	18,4	18,9	19,3	19,6	20
n_i	5	2	3	6	4

1.10. Знайти вибіркову та виправлену дисперсії варіаційного ряду, складеного за даними вибірки:

x_i	12	15	17	20	23	26
n_i	6	2	8	4	3	2

1.11. Знайти внутрішньогрупову, міжгрупову та загальну дисперсії сукупності, яка складається з трьох груп:

перша група	x_i	5	2	8
	n_i	10	5	5

друга група	x_i	8	6
	n_i	10	15

третя група	x_i	3	8
	n_i	20	5

1.12. Знайти вибіркову середню за даним розподілом вибірки з об'ємом n=20.

x_i	2560	2600	2620	2650	2700	3000
n_i	2	3	10	4	1	5

13. Знайти внутрішньогрупову, міжгрупову та загальну дисперсії сукупності, яка складається з трьох груп:

перша група	x_i	15	10	8
	n_i	10	4	6

друга група	x_i	18	16
	n_i	10	5

третя група	x_i	15	5
	n_i	7	10

1.14. Знайти вибіркову дисперсію за даним розподілом вибірки об'єму n=100:

x_i	340	360	370	380	400	420
n_i	10	30	18	12	20	10

1.15. Знайти виправлену вибіркову дисперсію вибірки об'єму n=20:

x_i	102	104	108	112	116
n_i	2	3	5	6	4

1.16. Знайти внутрішньогрупову, міжгрупову та загальну дисперсії сукупності, яка складається з двох груп:

перша група	x_i	1	2	8
	n_i	30	15	5

друга група	x_i	1	6
	n_i	10	15

1.17. Обчислити моду та медіану статистичного розподілу проданого взуття за розміром:

x_i	36	37	38	39	40	41	42	43
n_i	1	1	5	8	17	21	18	8

1.18. Статистичний розподіл волокон бавовни за довжиною (у мм) має вигляд:

x_i	6	9	12	15	18	21	24	27	30	33	36	39
n_i	3	27	60	85	108	127	153	172	146	83	33	4

Знайти для цього розподілу моду та медіану.

1.19. Знайти моду та медіану для даного інтервального статистичного розподілу робітників за часом (у *хв.*), який витрачається на обробку однієї деталі:

l_i	(0; 2)	(2; 4)	(4; 6)	(6; 8)	(8; 10)	(10; 12)
n_i	133	237	373	362	270	125

1.20. Задано статистичний ряд:

а)

x_i	5	6	7	10
n_i	4	5	8	3

б)

x_i	2	5	7	8
n_i	4	3	8	5

Знайти його числові характеристики, побудувати графік емпіричної функції розподілу та полігон частот.

1.21. За даними чотирьох вступних іспитів складена таблиця:

Сума балів	12	13	14	15	16	17	18	19	20
Кількість абітурієнтів	1	3	7	15	21	30	12	8	3

Знайти вибіркову середню та виправлену вибіркову дисперсію для випадкової величини *X* – суми балів.

1.22. З метою вивчення середньої врожайності пшениці проведено вибіркове вимірювання на площі 2500 *га*. Результати вимірювань подано у таблиці.

Врожайність з 1 *га*	(15;17)	(17;19)	(19;21)	(21;23)	(23;25)	(25;27)
Кількість *га*	100	300	500	700	600	300

Знайти вибіркову середню врожайність з гектара та вибіркову дисперсію.

1.23. П'ятдесят спостережень за жирністю молока дали такі результати (у %):

3,86 4,06 3,67 3,97 3,76 4,04 3,61 3,96 3,84 3,87 4,07
3,57 3,98 3,76 3,87 3,69 3,99 3,71 3,82 4,16 3,94 3,88
3,76 4,00 4,08 4,01 3,46 3,933,72 3,814,02 4,17 3,72
4,09 3,78 4,02 3,73 4,18 3,52 3,89 3,92 4,26 4,03 4,14
3,72 3,82 4,03 3,62 3,91 4,33.

Побудувати за цими даними інтервальний варіаційний ряд з рівними інтервалами (перший інтервал [3,45; 3,55), другий - [3,55; 3,65) і т. д.) і побудувати гістограму.

1.24. Нижче наведено результати вимірювання зросту (у см) випадково відібраних 100 студентів:

Зріст	Число студентів	Зріст	Число студентів
[154;158)	20	[166;170)	18
[158;162)	14	[170;174)	10
[162;166)	26	[174;178)	12.

Знайти вибіркову середню, вибіркову дисперсію та виправлену вибіркову дисперсію зросту обстежуваних студентів.

1.25. У відділі технічного контролю було виміряно 200 втулок з партії, яка була виготовлена одним автоматичним верстатом. У таблиці дано відхилення діаметрів від номіналу (у мікронах) після групування.

Межі відхилень	Кількість втулок	Межі відхилень	Кількість втулок
[-20;-15)	7	[5; 10)	41
[-15;-10)	11	[10; 15)	26
[-10;-5)	15	[15; 20)	17
[-5; 0)	24	[20; 25)	7
[0; 5)	49	[25; 30)	3

Знайти вибіркову середню, вибіркову та виправлену дисперсії відхилення діаметрів від номіналу.

РОЗДІЛ 2
СТАТИСТИЧНІ ОЦІНКИ ПАРАМЕТРІВ РОЗПОДІЛУ

Однією із задач математичної статистики є визначення невідомих параметрів розподілу генеральної сукупності за даними вибірки, яка була відібрана з даної генеральної сукупності. Оскільки елементи вибірки є випадковими величинами, то випадковим буде і значення параметра, обчисленого за допомогою цієї вибірки. А якщо ми маємо декілька вибірок одного і того ж об'єму з однієї генеральної сукупності, то кожна з них дає своє значення параметра, який досліджується. Тому за вибіркою ми не можемо знайти значення цього параметра, а можемо лише оцінити цей параметр.

Задачу про оцінку можна розділити на дві частини: яку величину, підраховану за даними вибірки, можна прийняти за наближене значення характеристики генеральної сукупності (точкова оцінка), і в якому інтервалі знаходиться з даною ймовірністю оцінювана величина (інтервальна оцінка).

2.1. Точкові оцінки параметрів розподілу

Нехай задана генеральна сукупність *N,* кожен об'єкт якої має одну й ту саму кількісну ознаку *X*. При випадковому виборі об'єкта генеральної сукупності стає відомим конкретне значення кількісної ознаки *X*. Це значення кількісної ознаки *X* позначимо через *x.* При цьому *X* - це випадкова величина, *x* - це одне з можливих значень цієї випадкової величини.

Оскільки досліджується не вся генеральна сукупність, а тільки вибірка, то в розпорядженні дослідника завжди будуть тільки дані про вибірку. Тобто він матиме тільки значення варіант $x_1, x_2, x_3, \ldots,$

x_n, якщо об'єм вибірки *n*. Через ці дані і виражаються параметри, які необхідно оцінити.

Знайти оцінку (статистичну оцінку)невідомого параметра теоретичного розподілу - це означає знайти функцію від випадкових величин x_1, x_2, ..., x_k, що спостерігаються. Ця функція і дає наближене значення параметра, який оцінюється.

Статистична оцінка невідомого параметра генеральної сукупності одним числом називається *точковою.*Точкові оцінки бувають зсунені і незсунені, ефективні та слушні (обґрунтовані).

Нехай θ^* є статистична оцінка невідомого параметра θ теоретичного розподілу*. Зсуненою* називають статистичну оцінку θ^*, математичне сподівання якої не дорівнює параметру, який оцінюється. Використання такої оцінки призводить до систематичних помилок (похибок).

Незсуненою називають статистичну оцінку θ^*, атематичне сподівання якої дорівнює параметру, що оцінюється, тобто $M(\theta^*)=\theta$. Дотримання цієї вимоги усуває систематичні похибки.

Але незсунена оцінка не завжди дає добре наближення параметра, який оцінюється. Тому до статистичної оцінки ставиться вимога ефективності. *Ефективною* називають статистичну оцінку, яка (при заданому об'ємі вибірки *n*) має найменшу можливу дисперсію.

При аналізі вибірок великого об'єму (n>30) до статистичних оцінок ставиться вимога слушності (змістовності, спроможності, обґрунтованості). *Спроможною* називають статистичну оцінку, яка при $n \to \infty$ прямує за ймовірністю до параметра, що оцінюється. Для виконання цієї вимоги досить, щоб дисперсія незсуненої оцінки прямувала до нуля, коли $n \to \infty$. В цьому легко переконатись, якщо скористатись нерівністю Чебишова.

Прикладом спроможної оцінки можуть слугувати закони великих чисел, наприклад, теорема Бернуллі. Очевидно, такій умові повинна задовольняти всяка оцінка, що придатна для практичного використання.

Найчастіше параметрами, які необхідно оцінити,є: математичне сподівання; дисперсія та середнє квадратичне відхилення.

Незсунена оцінка генеральної середньої є вибіркова середня. За оцінку величини математичного сподівання звичайно беруть середнє арифметичне (вибіркове середнє) значень вибірки

$$\overline{x_в} = \frac{\sum_{i=1}^{k} n_i x_i}{n}. \qquad (2.1.1)$$

Оскільки вибіркове середнє являє собою суму випадкових величин, то воно також буде випадковою величиною. Тому мова йде про закон розподілу вибіркового середнього, його математичного сподівання та дисперсії. Математичне сподівання вибіркового середнього співпадає зі значенням M(x)=*m:*

$$M(\overline{x_в}) = \frac{1}{n}\sum_{i=1}^{к} M(x_i n_i) = \frac{1}{n} m \cdot n = m.$$

Це говорить про незсуненість оцінки генеральної середньої за вибірковою середньою.

Вибіркові дисперсія $Dв$ і середнє квадратичне відхилення $\sigma = \sqrt{\overline{x^2} - \overline{x}^2}$ є наближеними показниками характеристики генеральної сукупності, оскільки вони залежать від об’єму вибірки n, що зумовлює систематичну похибку. Зі збільшенням об’єму вибірки ці показники зменшуються, а тому вирази

$$Dв = \frac{\sum_{i=1}^{k}\left(x_i - \overline{x}\right)^2}{n} = \frac{\sum_{i=1}^{k} n_i (x_i - \overline{x})^2}{n} \quad \text{і} \quad \sigma = \sqrt{D} = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \overline{x})^2}{n}} = \sqrt{\frac{\sum_{i=1}^{k} n_i (x_i - \overline{x})^2}{n}}$$

є зсуненими оцінками. З метою зменшення систематичних похибок у цих формулах n замінюється на $n-1$. У такій формі запису матимемо *незсунені* $D_в$ і σ , які при великих об'ємах вибірки *співпадають зі зсуненими.*

Вибіркова дисперсія $D_в$ *є зсуненою оцінкою генеральної дисперсії* $D_г$, оскільки математичне сподівання $M(D)=\frac{n-1}{n}D_В \neq D_в$.

Незсуненою оцінкою генеральної дисперсії є *виправлена дисперсія*

$$S^2 = \frac{n}{n-1} D_в. \qquad (2.1.2)$$

Незсуненою оцінкою генерального середнього квадратичного відхилення є *виправлене середнє квадратичне відхилення S*, яке є коренем квадратним із виправленої вибіркової дисперсії

$$S = \sqrt{S^2} = \sqrt{\frac{n}{n-1} D_В} = \sqrt{\frac{n}{n-1}\left(\overline{x^2} - (\bar{x})^2\right)}. \qquad (2.1.3)$$

Приклад 2.1.1. З генеральної сукупності зроблена вибірка

x_i	2	5	7	10
n_i	16	12	8	14

Дати статистичну оцінку генеральній сукупності.

Розв'язання. Об'єм вибірки n =16 + 12 + 8 +14= 50.

$$\overline{x_в} = \frac{\sum_{i=1}^{k} n_i x_i}{n} = = \frac{16\cdot 2 + 12\cdot 5 + 7\cdot 8 + 10\cdot 14}{50} = 5,76.$$

$$D_в = \frac{\sum_{i=1}^{k} x_i^2 \cdot n_i}{n} - (\overline{x_в})^2 = \frac{2^2\cdot 16 + 5^2\cdot 12 + 7^2\cdot 8 + 10^2\cdot 14}{50} - (5,76)^2 = 2,895.$$

Виправлена вибіркова дисперсія дорівнює $S^2 = \frac{n}{n-1}\cdot D_в$,

$$\mathrm{S}^2 = \frac{50}{49}\cdot 2,895 = 2,95.$$

Виправлене середнє квадратичне відхилення дорівнює $S = \sqrt{S^2} = \sqrt{2,95} = 1,72.$

2.2. Інтервальні оцінки параметрів розподілу. Довірчі інтервали

Так як вибірка є випадковою, то в багатьох випадках, особливо при малих об'ємах вибірки, точкові оцінки невідомих параметрів генеральної сукупності одним числом можуть виявитися далекими від дійсних значень параметрів. Тому бажано мати дані про надійність цієї оцінки. Наприклад, вказати інтервал, в якому з імовірністю, що близька до одиниці, знаходиться точне значення оцінюваного параметра. Задачу отримання такого інтервалу називають *інтервальним оцінюванням*.

Для її розв'язання вибирають ймовірність γ, і за визначеними правилами знаходять таке число $\delta>0$, щоб виконувалось співвідношення

$$P(\left|\theta-\theta^{+}\right|<\delta)=\gamma. \qquad (2.2.1)$$

Число δ називається *точністю оцінки* (чим менше δ, тим вища точність оцінки).

Надійністю γ *(довірчою ймовірністю)* оцінки θ за статистичною характеристикою θ^{*} називають ймовірність γ, яка задовольняє відношенню $P(\left|\theta-\theta^{*}\right|<\delta)=\gamma$, і з якою оцінюваний параметр θ покривається інтервалом $(\theta^{*}-\delta, \theta^{*}+\delta)$.

Нехай $x_1, x_2, \dots, x_k$ - вибірка з генеральної сукупності, θ - невідомий параметр генеральної сукупності X. При інтервальному оцінюванні невідомого параметра θ шукають такі дві функції $h_1=h_1(x_1, x_2, \dots, x_k)$ і $h_2=h_2(x_1, x_2, \dots, x_k)$, для яких завжди $h_1<h_2$ і при заданому $\gamma є(0; 1)$ виконується умова $P(h_1<\theta<h_2)>\gamma$.

Тоді інтервал $(h_1; h_2)$ називається надійним (довірчим) інтервалом, γ - рівнем надійності, а h_1, h_2 – відповідно нижньою та верхньою межами надійності.

Рівень надійності γ задається наперед, причому значення γ беруть близькими до одиниці. Найчастіше за рівень надійності (надійність) беруть числа: 0,9; 0,95; 0,99 та інші.

Довірчим називають інтервал ($\theta^* - \delta$; $\theta^* + \delta$), який покриває невідомий параметр θ із заданою надійністю γ.

У прикладних статистичних задачах довжина надійного інтервалу грає важливу роль: чим менша його довжина, тим точніша його оцінка. Якщо довжина цього інтервалу велика, то значимість такої оцінки незначна.

Надійний інтервал для оцінки математичного сподівання нормального розподілу при відомій дисперсії

Нехай кількісна ознака *X* генеральної сукупності розподілена за нормальним законом і відоме середнє квадратичне відхилення σ. Треба знайти довірчий інтервал, що покриває математичне сподівання *a* генеральної сукупності з заданою надійністю γ.

Надійний інтервал для математичного сподівання будь-якого розподілу легко знайти за допомогою центральної граничної теореми. Якщо $x_1, x_2, \dots, x_k$ - вибірка з генеральної сукупності *X*, розподіленої за довільним законом з математичним сподіванням *a* і дисперсією σ^2, то $x_1, x_2, \dots, x_k$ *є* взаємно незалежними й однаково розподіленими випадковими величинами з математичним сподіванням *a* і дисперсією σ^2. З центральної граничної теореми випливає, що при довірчій ймовірності $P\left(|\overline{x} - a| < \frac{\sigma}{\sqrt{n}} t\right) \approx 2\Phi(t)$,де $\Phi(t) = \frac{1}{\sqrt{2\Pi}} \int_0^t e^{\frac{-t^2}{2}} dt$ - функція Лапласа, тобто при рівні надійності $\gamma = 2\Phi(t)$ надійним інтервалом для математичного сподівання *при відомому* σ є інтервал

$$P(\overline{x}_{в.} - \delta < a < \overline{x}_{в.} + \delta) = 2\Phi(t) = \gamma;$$

де δ - точність оцінки; $\delta = t\dfrac{\sigma}{\sqrt{n}}$.

$$P(\overline{x}_{в.} - t\frac{\sigma}{\sqrt{n}} < a < \overline{x}_{в.} + t\frac{\sigma}{\sqrt{n}}) = 2\Phi(t) = \gamma. \qquad (2.2.2)$$

Ця рівність визначає ймовірність того, що з надійною ймовірністю γ середня вибіркова відхиляться від математичного сподівання а генеральної сукупності на величину $t\dfrac{\sigma}{\sqrt{n}}$.

При даному рівні надійності γ значення *t* знаходиться з рівності $\Phi(t) = \frac{\gamma}{2}$.

Приклад 2.2.1. Знайти довірчий інтервал для оцінки математичного сподівання *а* нормального розподілу з надійністю γ=0,95, якщо $\overline{x}_в = 75{,}13$; *n* =100, σ =10.

Розв'язання. З рівності $\Phi(t) = \dfrac{0{,}95}{2} = 0{,}475$ за таблицею функції Лапласа знайдемо *t* = 1,96. Підставивши в $\overline{x}_{в.} - t\dfrac{\sigma}{\sqrt{n}} < a < \overline{x}_{в.} + t\dfrac{\sigma}{\sqrt{n}}$, відповідні значення, одержимо довірчий інтервал

$$75{,}13 - 1{,}96\frac{10}{\sqrt{100}} < a < 75{,}13 + 1{,}96\frac{10}{\sqrt{100}}$$ або $a \in (73{,}17;\ 77{,}09)$.

Приклад 2.2.2. Знайти мінімальний об'єм вибірки, при якій, з надійністю γ=0,975; точність оцінки математичного сподівання *а* генеральної сукупності за вибірковою середньою рівна δ=0,3 і відомо, що σ =1,2 нормально розподіленої генеральної сукупності.

Розв'язання. За умовою задачі γ= 0,975; $\Phi(t)$ = 0,975/2 = =0,4875. За таблицею (додаток 2) знайдемо *t* = 2,24.

Мінімальний об'єм вибірки можна знайти з формули $\delta = \dfrac{t \cdot \sigma}{\sqrt{n}}$.

$$n = \frac{t^2\sigma^2}{\delta^2} = \frac{(2,24)^2(1,2)^2}{(0,3)^2} = 81.$$

Приклад 2.2.3. Визначити, з якою надійністю γ можна гарантувати точність оцінки, яка не перевищує δ=10, якщо σ =16 та *n* =15.

Розв'язання. З формули $\delta = \frac{t \cdot \sigma}{\sqrt{n}}$ знаходимо $t = \frac{\delta\sqrt{n}}{\sigma} = \frac{10\sqrt{15}}{16} = 2,42.$ 2Φ(2,42) = γ, отже, шуканаймовірність буде γ =2·0,4922= 0,9844.

Надійний інтервал для оцінки математичного сподівання нормального розподілу при невідомій дисперсії

Нехай потрібно оцінити математичне сподівання генеральної сукупності, яка має нормальний розподіл, при невідомій дисперсії σ^2. Так як σ невідома, то не можна скористатися результатами, в якому σ припускалось відомим. Обчислимо за вибіркою виправлену дисперсію $S^2 = \frac{n}{n-1} \cdot D_в$. Визначимося з надійною ймовірністю γ та знайдемо таке число *ε*, щоб виконувалося співвідношення $P(\overline{x}_{в.} - \varepsilon < a < \overline{x}_{в.} + \varepsilon) = \gamma.$

Виявляється, що за даними вибірки можна побудувати випадкову величину $T = \frac{(\overline{x}_в - a)\sqrt{n}}{s},$ що має розподіл Стьюдента зі степенями вільностей *k*= *n*-1, яка не залежить від параметра σ.

Ймовірність виконання нерівності $T < \varepsilon$ визначається наступним чином:

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

За таблицею розподілу Стьюдента (додаток 3) отримаємо значення ε =t_λ, при якому виконується рівність $P\left(\left|\frac{(\overline{x} - a)\sqrt{n}}{S}\right| < \frac{t_\gamma S}{\sqrt{n}}\right) = \gamma$.

При рівні надійності γ= 2Ф(t) для *a* беруть надійний інтервал

$$P(\overline{x}_{в.} - t_\gamma \frac{S}{\sqrt{n}} < a < \overline{x}_{в.} + t_\gamma \frac{S}{\sqrt{n}}) = \gamma, \qquad (2.2.3)$$

де $t_\gamma = t\ (\gamma,\ n)$ – значення, яке береться з таблиці розподілу Стьюдента (додаток 3).

Приклад 2.2.4. З генеральної сукупності дістали вибірку

x_i	-2	1	2	3	4	5
n_i	2	1	2	2	2	1

Оцінити з надійністю 0,95 істинне значення нормально розподіленої ознаки генеральної сукупності за вибіркової середньою за допомогою довірчого інтервалу.

Розв'язання. Істинне значення ознаки ототожнюється з математичним сподіванням *a*. Для оцінки *a* при невідомому *σ* скористаємося формулою (2.2.3). Об'єм вибірки *n*=2+1+2+2+2+1=10. Далі обчислимо: $\overline{x_в} = \frac{\sum_{i=1}^{k} n_i x_i}{n} = 2,$

$$D_{в.} = \frac{\sum_{i=1}^{k} n_i \left(x_i - \overline{x_в}\right)^2}{n} = \frac{2(-2-2)^2+(1-2)^2+2(2-2)^2+2(3-2)^2+2(4-2)^2+(5-2)^2}{10} = \frac{52}{10} =$$

$= 5{,}2.$

$S^2 = \frac{n-1}{n} \cdot D_в = \frac{10}{9} \cdot 5{,}2 = 5{,}78.$ $S = \sqrt{S^2} = \sqrt{5{,}78} = 2{,}4.$

Знайдемо $t_\gamma =$ 2,26, користуючись додатком 3, за такими даними: γ= 0,95 і *n*=10*.* Шуканий довірчий інтервал знаходимо за формулою $\overline{x_в} - t_\gamma \frac{S}{\sqrt{n}} < a < \overline{x_в} + t_\gamma \frac{S}{\sqrt{n}}$, $2 - 2{,}26\frac{2{,}4}{\sqrt{10}} < a < 2 + 2{,}26\frac{2{,}4}{\sqrt{10}}$,

Отже, 0,3 <*a* < 3,7.

Надійний інтервал для оцінки середнього квадратичного відхилення σ

Нехай треба оцінити середнє квадратичне відхилення σ нормально розподіленої генеральної сукупності *N* за виправленим середнім квадратичним відхиленням *S*. Для цього проведемо *k* спостережень і за результатами $x_1, x_2, \ldots, x_k$ обчислимо $\overline{x_в}$, оцінку S^2 невідомої дисперсії σ^2 і оцінку *S* для *σ*. Задаючи надійність γ інтервальної оцінки, знайдемо таке число *ε*, щоб виконувалося співвідношення $P(S-\varepsilon<\sigma<S+\varepsilon)=\gamma.$

Перетворимо подвійну нерівність до виду $S-\varepsilon<\sigma<S+\varepsilon,$ $S(1-q)<\sigma<S(1+q)$ де $q=\dfrac{\varepsilon}{S}.$

Отже шуканий довірчий інтервал для оцінки середнього квадратичного відхилення нормального розподілу за виправленим середньоквадратичним відхиленням *S* має такий вигляд:

$$S(1-q)<\sigma<S(1+q) \qquad (\text{при } q<1) \qquad (2.2.4)$$

$$0<\sigma<S(1+q) \qquad (\text{при } q>1), \qquad (2.2.4`)$$

де $q=q(\gamma,\ n)$ визначається з таблиці (додаток 4) за надійним рівнем γ та обсягом вибірки *n*.

Піднесемо обидві частини інтервальної оцінки середнього квадратичного відхилення до квадрату, отримаємо, при тій же надійності γ, довірчий інтервал для дисперсії:

$$S^2(1-q)^2<\sigma^2<S^2(1+q)^2 \qquad (\text{при } q<1),$$

$$0<\sigma^2<S^2(1+q)^2 \qquad (\text{при } q>1).$$

Приклад 2.2.5. Знайти з надійністю γ = 0,98 інтервальну оцінку для середнього квадратичного відхилення та дисперсії нормального розподілу, якщо за вибіркою об'єму *n* = 17 обчислена оцінка S^2 = 25.

Розв’язання. Спочатку знайдемо довірчий інтервал для оцінки середнього квадратичного відхилення.

$S = \sqrt{S^2} = \sqrt{25} = 5, \quad q \ = \ 0,564; \ 5(1-0,564) < \sigma < 5(1+0,564),$

$2,18 < \sigma < 7,82.$

Піднесемо обидві частини нерівності в квадрат, отримаємо, при тій же надійності γ, надійний інтервал для дисперсії:

$$(2,18)^2 < \sigma^2 < (7,82)^2, \ 4,7524 < \sigma^2 < 61,1524.$$

Отже, для оцінки параметрів нормально розподіленої ознаки генеральної сукупності з надійністю γ використовують наступні довірчі інтервали:

- для оцінки математичного сподівання при відомому σ:

$$(\bar{x} - t\frac{\delta}{\sqrt{n}} < a < \bar{x} + t\frac{\delta}{\sqrt{n}}) \qquad (2.2.2)$$

- для оцінки математичного сподівання, якщо σ - невідома величина:

$$(\bar{x} - t_\gamma \frac{S}{\sqrt{n}} < a < \bar{x} + t_\gamma \frac{S}{\sqrt{n}}) \qquad (2.2.3)$$

- для оцінки середнього квадратичного відхилення σ:

$S(1-q) < \sigma < S(1+q)$ (при q < 1)

$0 < \sigma < S(1+q)$ (при q > 1). (2.2.4)

Запитання для самоперевірки

1. Як визначають статистичні оцінки числових характеристик та умови їх незсуненості, ефективності, слушності?
2. Яка оцінка генеральної сукупності називається точковою?
3. Які види точкових оцінок вивчає статистика?
4. Яка оцінка для дисперсії має властивість спроможності і незсуненості?
5. Яка оцінка генеральної сукупності називається інтервальною?

6. Яка різниця між інтервальними та точковими оцінками?

7. Що називається надійним (довірчим) інтервалом і надійною ймовірністю (надійністю)?

8. Як будується інтервал надійності для математичного сподівання випадкової величини, яка розподілена за нормальним законом, при відомій та невідомій дисперсії?

9. Як будується інтервал надійності для середнього квадратичного відхилення та дисперсії випадкової величини, що розподілена за нормальним законом?

10. Що таке рівень значимості, яким він може бути?

11. Що таке точність оцінки, як вона пов'язана з надійним (довірчим)інтервалом?

12. Як визначити точність оцінки, якщо відомі об'єм вибірки та надійність оцінки?

13. Як визначити надійність оцінки, якщо відомі точність оцінки та об'ємвибірки?

14. Як визначити об'єм вибірки, щоб задана точність була гарантованаіз заданою надійністю, якщо об'єм генеральної сукупності невідомий?

15. Як визначити об'єм вибірки, щоб задана точність була гарантованаіз заданою надійністю при відомому об'ємі генеральної сукупності?

16. У яких випадках для визначення *t* потрібно користуватись таблицями Стьюдента?

Задачі та вправи

2.1. З генеральної сукупності взято вибірку. За даним розподілом вибірки знайти незсунену оцінку генеральної середньої:

а)

x_i	2	5	7	10
n_i	16	12	8	14

б)

x_i	1	3	6	26
n_i	8	40	10	2

2.2. З генеральної сукупності зроблена вибірка:

x_i	22	25	27	30	35
n_i	6	10	8	14	10

Дати статистичну оцінку параметрам генеральної сукупності.

2.3. Знайти довірчий інтервал для оцінки з надійністю γ=0,95 невідомого математичного сподівання *a* нормального розподіленої ознаки генеральної сукупності, якщо генеральне середнє квадратичне відхилення σ= 5, вибіркова середня $\overline{x}_b = 14$ і об'єм вибірки *n* = 25.

2.4. Оцінити з надійністю γ= 0,95 математичне сподівання *a* нормально розподіленої ознаки генеральної сукупності за допомогою довірчого інтервалу, якщо відомо такий розподіл вибірки :

x_i	12	13	16	20
n_i	8	30	10	2

2.5. Знайти мінімальний об'єм вибірки нормально розподіленої генеральної сукупності, при якій з надійністю γ =0,925 точність оцінки математичного сподівання *a* генеральної сукупності за вибірковою середньою $\delta = 0,2;$ а середнє квадратичне відхилення σ= 1,5.

2.6. За даними вибірки об'єму *n* = 16, знайдено виправлене середнє квадратичне відхилення *S*=1 нормально розподіленої кількості ознаки. Знайти довірчий інтервал, який покриває генеральне середнє квадратичне відхилення σ та дисперсію з надійністю γ= 0,95.

2.7. Виконано 12 вимірів одним приладом деякої фізичної величини (без систематичної помилки, причому виправлене середнє квадратичне відхилення S випадкових помилок вимірювань дорівнює 0,6. Знайти точність приладу з надійністю 0,99. Результати вимірів розподілені нормально.

2.8. З 500 випадків обстежених користувачів Інтернетом 300 відповіли, що віддають перевагу користуванню інтернетом після 12-ої години ночі. Знайдіть: а) 95%-й довірчий інтервал ймовірності появи вказаних користувачів інтернету; б) інтервал користувачів інтернету після 12 години.

2.9. Дано середнє квадратичне відхилення, вибіркова середня та об'єм вибірки нормально розподіленої величини. Знайти довірчий інтервал для оцінки невідомого математичного сподівання із заданою надійністю, якщо σ=2, n=10, γ=0,95; $\overline{x}_{b=}$5,4.

2.10. Знайти мінімальний об'єм вибірки, при якому з надійністю γ = 0,99 точність оцінки δ математичного сподівання нормально розподіленої ознаки дорівнює 0,2, якщо середнє квадратичне відхилення σ дорівнює 2.

2.11. Виконано 10 вимірів одним приладом (без систематичної помилки) деякої фізичної величини, при чому виправлене середнє квадратичне відхилення S випадкових помилок вимірів виявилось рівним 0,8. Знайти точність приладу з надійністю 0,95. Припускається, що результати вимірів розподілені нормально.

2.12. Прийом на роботу у фірму проводиться на конкурсній основі. Претендент оцінюється за 100-бальною системою. Вибірка із 150 чоловік показала, що середній бал конкурсанта складає 70. Відомо, що стандартне відхилення генеральної сукупності становить 10. З надійністю γ=0,99 збудувати довірчий інтервал для невідомого середнього значення кількості набраних балів.

2.13. У вибірці з 25 зернин пшениці середня маса зернини склала $\bar{x}_{в.} = 0{,}152$ при S=0,03 г. Вважаючи масу зерна нормально розподіленою випадковою величиною, встановити: а) інтервал надійності для середньої ваги зернини з гарантією γ = 0,99; б) який буде інтервал надійності, коли об'єм усієї вибірки вважати n = 50?

2.14. На фірмі з 200 чоловік, які звільнилися за останні два роки, 70 з них мотивували своє звільнення низькою зарплатою. Знайти 99%-й довірчий інтервал звільнених з цієї причини.

2.15. Досліджується вибірка з 40 елементів з нормально розподіленої сукупності. Вибіркова середня та виправлена дисперсія відповідно дорівнюють 25 і 3,7. Знайти 90%-й надійний інтервал для математичного сподівання та 95%-й надійний інтервал для середнього квадратичного відхилення та дисперсії.

2.16. При перевірці 180 одиниць товару з великої партії виявилось 12 бракованих. Знайти 95%-й надійний інтервал для частки бракованих одиниць у всій партії та оцінити мінімальний об'єм вибірки, щоб з ймовірністю 0,95 можна було б стверджувати, що частка браку у всій партії відрізняється від частки появи браку у вибірці не більше, ніж на 2% .

2.17. Глибина моря вимірюється приладом, систематична похибка якого дорівнює нулю, а випадкові похибки розподілені нормально із середнім квадратичним відхиленням σ=15 м. Скільки треба зробити незалежних вимірювань, щоб знайти глибину з похибкою не більше, ніж 5 м, при надійному рівні γ= 0,9?

2.18. За даними вибірки об'єму n генеральної сукупності нормально розподіленої кількісної ознаки знайдено виправлене середнє квадратичне відхилення S. Знайти надійний інтервал, який покриває генеральне середнє квадратичне відхилення σ та дисперсію σ^2 з надійністю γ= 0,999, якщо: а) n= 10, S= 5,1; б) n= 50, S= 14.

2.19. Для визначення середньої суми вкладів у банку, який має 2200 вкладників, проведено вибіркове дослідження 111 вкладників, результати якого подані в таблиці:

Сума вкладу тис. грн.	[10;30)	[30;50)	[50;70)	[70;90)	[90;110)	[110;130]
Число вкладників	1	3	10	30	60	7

Користуючись цими даними, знайти надійні межі для генеральної середньої, які можна було б гарантувати з імовірністю 0,96.

2.20. Для визначення середньої врожайності пшениці на площі 500000 га проведено вибіркове вимірювання врожайності на 2500 га. Результати вимірювань подано в таблиці:

Врожайність з 1 га (в ц)	[15;17)	[17;19)	[19;21)	[21;23)	[23;25)	[25;27]
Кількість га	100	300	500	700	600	300

Знайти надійні межі для середньої врожайності з надійністю: а) γ= 0,95; б) γ= 0,999.

2.21. Вимірювання твердості 16 зразків легованої сталі (в умовних одиницях) дали такі результати: 13,1; 12,8; 11,9; 12,4; 13,5; 12,0; 13,7; 13,8; 10,6; 12,4; 13,5; 11,7; 13,9; 11,9; 11,5; 12,5. Вважаючи, що вибірка отримана з нормально розподіленої генеральної сукупності, знайти надійні інтервали для математичного сподівання та середнього квадратичного відхилення при надійному рівні γ = 0,95.

2.22. З метою підвищення продуктивності праці на фірмі К було обстежено 20 робітників, які виконують складну технологічну операцію. Середній час, витрачений ними на виконання операції, складає13 хвилин. Відомо, що витрачений час має нормальний розподіл із стандартним відхиленням в 6 хвилин. Треба побудувати 99%-й надійний інтервал для часу виконання технологічної операції.

2.23. Фірма К планує відкрити кафе в районі станції Т. З метою оптимального планування можливої кількості відвідувачів аналітики фірми провели статистичне дослідження кількості відвідувачів 30 кафе, магазинів та інших установ масового обслуговування. Середня кількість відвідувачів за годину пік складає 70 чоловік. Оцінка стандартного відхилення генеральної сукупності дає кількість 24,5 чоловік. Знайти 99%-й надійний інтервал можливої кількості відвідувачів.

2.24. На телефонній станції обстежувалася тривалість телефонної розмови. Була зроблена вибірка із 100 телефонних розмов. З’ясувалося, що середня тривалість розмови складає 3 хвилини. Враховуючи, що середня тривалість телефонної розмови розподілена нормально зі стандартним відхиленням 0,6 хв., знайдіть 99%-й довірчий інтервал для тривалості телефонної розмови.

РОЗДІЛ 3
ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ

Перевірка статистичних гіпотез тісно пов’язана з теорією оцінки параметрів розподілу. В практичній і науковій діяльності часто для доведення справедливості того чи іншого факту висувають гіпотезу, яка може бути перевірена на основі даних вибіркового спостереження. Під *гіпотезою* будемо розуміти в широкому смислі слова деяке наукове припущення про властивості явищ, що вивчаються, яке вимагає перевірки та доведення. Під *статистичною гіпотезою* (*H*) розуміють припущення відносно параметрів або форми розподілу генеральної сукупності, яке перевіряється на основі даних вибіркового спостереження. З цього означення випливає, що статистична гіпотеза може відноситись або до законів розподілу, або до окремих параметрів розподілу.

Перевірка статистичних гіпотез тісно зв'язана з теорією оцінки параметрів розподілу. Прикладом статистичних гіпотез можуть бути припущення про те, що жива маса телят у господарствах району підпорядкована закону нормального розподілу; середня урожайність картоплі одного сорту перевищує середню урожайність другого сорту; середні розміри деталей, що виготовляються на однотипних верстатах в ремонтній майстерні, однакові тощо.

У ході перевірки статистичної гіпотези необхідно встановити, чи узгоджуються дані спостереження з висунутою гіпотезою, чи можливо відмінності між гіпотезою і результатами спостереження віднести до випадкових, або ж ці відмінності викликані впливом яких-небудь систематично діючих причин. У результаті перевірки гіпотеза приймається або відхиляється.

3.1. Статистичної гіпотези та їх різновиди

Найчастіше перевіряється припущення про те, що отримана за вибіркою величина незначно відрізняється від гіпотетичної (теоретично припущеної) або встановленої величини в генеральній сукупності. Для перевірки цього положення висувається гіпотеза про те, що істинна різниця між фактичними і гіпотетичними показниками дорівнює нулю. У зв'язку з цим гіпотезу, що перевіряється, називають *нульовою* і позначають H_0. Нульову гіпотезу ще називають *основною* або *робочою.*

У кожному випадку нульовій гіпотезі протиставляється *альтернативна (конкуруюча) гіпотеза (H_a),* яка заперечує нульову гіпотезу. Наприклад, якщо нульова гіпотеза полягає в припущенні, що середня врожайність зернових культур у генеральній сукупності дорівнює 35 ц/га, то альтернативна гіпотеза, зокрема, може полягати в тому, що середня врожайність у генеральній сукупності не дорівнює 35 ц/га, або середня врожайність у генеральній сукупності більша за 35 ц/га і т. д Зміст гіпотези записується після двокрапки, а сам запис має вигляд

$$H_0\colon \overline{x} = 35, \qquad H_a\colon \neq 35.$$

По відношенню до нульової гіпотези може бути вказана нескінченна множина альтернативних гіпотез. Вибір тих чи інакших H_0 або H_a визначається характером вибіркової сукупності, конкретними завданнями, що стоять перед менеджерами, економістами, інженерами, дослідниками і розв'язуються за допомогою перевірки статистичних гіпотез. Формально будь-яка з двох протилежних гіпотез може виступати як нульова. Однак практичні наслідки будуть різними у зв'язку з різним характером можливих помилок, отриманих у результаті неправильного прийняття або заперечення нульової гіпотези. Тому вибір нульової гіпотези має принциповий характер і потребує спеціального обґрунтування.

За формою побудови гіпотези можуть бути простими і складними. *Простою* називають гіпотезу, яка містить тільки одне припущення, *складною* - гіпотезу, яка містить два і більше припущень. Наприклад, гіпотеза H_0*: σ* = 3 є простою гіпотезою, а гіпотеза H_0: *σ* >3 і H_0*: σ* <3, яка включає деяку область ймовірних значень досліджуваного параметра, є складною. Очевидно, що складні гіпотези включають множину простих гіпотез.

3.2. Помилки перевірки гіпотез

У результаті перевірки статистичної гіпотези, що основана на даних вибірки обмеженого обсягу, можна відхилити або прийняти нульову гіпотезу (відповідно вибіркові дані суперечать або узгоджуються з H_0*).* Звідси видно, що перевірка статистичних гіпотез зв'язана **з** ризиком прийняття помилкових рішень.

Неправильне рішення може бути прийняте у двох випадках. У зв'язку з цим розрізняють помилки двох родів.

Помилка першого роду полягає в тому, що нульова гіпотеза H_0 відхиляється, хоча в дійсності вона правильна.

Помилка другого роду полягає в тому, що приймається нульова гіпотеза, хоча в дійсності правильна альтернативна гіпотеза H_a.

Правильні та неправильні рішення можуть бути отримані у двох випадках. Це наочно ілюструє таблиця 3.2.1.

Ймовірність зробити помилку першого роду (невиправдане відхилення H_0*)* отримала назву *рівня значущості* і позначається *α.* Ймовірність зробити помилку другого роду (прийняття неправильної гіпотези) позначається *β*. Отже, можна сказати, що при великому числі вибірок частка хибних висновків дорівнює α, якщо правильна H_0, і дорівнює β, якщо правильна H_a.

Таблиця 3.2.1

Можливі результати перевірки нульової гіпотези

Результат перевірки H_0	Можливий стан гіпотези, що перевіряється	
	Правильна гіпотеза H_0	Правильна гіпотеза H_a
H_0 відхиляється	Помилка першого роду	Правильне рішення
H_0 приймається	Правильне рішення	Помилка другого роду.

Помилки першого α та другого β роду за своїми наслідками нерівнозначні і ведуть до різних матеріальних втрат. Тому вибір рівня значущості повинен ґрунтуватись на обліку можливих втрат: чим більші ці втрати, тим меншим повинен бути рівень значущості. Однак, якщо знижується рівень значущості, збільшується ймовірність появи помилок другого роду. В цьому розумінні помилки першого та другого роду є конкуруючими.

Оскільки помилки першого та другого роду практично виключити неможливо, то в кожному випадку необхідно прагнути до зменшення втрат від цих помилок. При практичній перевірці гіпотез прагнуть до того, щоб за помилку першого роду прийняти ту із можливих помилок, яка матиме більш серйозні наслідки на практиці. Так, при статистичному аналізі контролю приймання продукції α - ризик виробника, β - ризик споживача; при статистичному регулюванні технологічного процесу α – ризик надлишкового налаштування, β - ризик непоміченого розладу.

Рівень значущості α встановлюється самим дослідником залежно від характеру і важливості завдань, що розв'язуються (за так званим принципом практичної впевненості). Рівень значущості являє собою ту мінімальну ймовірність, починаючи з якої можна визнати подію практично неможливою. Можна користуватися стандартними значеннями α= 0,10; 0,05; 0,01; 0,001; 0,0001 та ін. Найчастіше α встановлюють на рівні 0,05 і 0,01. При більш відповідальних рішеннях α підвищують до 0,001. Рівень значущості, наприклад, α=0,05; означає, що в середньому в 5 випадках із 100 є ризик допустити помилку першого роду, тобто відхилити правильну гіпотезу *(H_0).*

Встановлюючи певний рівень значущості, дослідник контролює ймовірність помилки першого роду: чим він нижчий, тим частіше H_0 буде визнаватися правильною. Однак, як було відзначено вище, зниження рівня значущості веде до появи помилок другого роду. В більшості випадків єдиним шляхом одночасного зменшення ймовірності появи помилок двох родів є збільшення чисельності вибірки.

3.3. Критерії узгодження для перевірки гіпотез

Для перевірки нульової гіпотези і прийняття висновку про сумісність вибіркових даних з висунутою гіпотезою використовують спеціальні *статистичні критерії***,** які являють собою список правил, за якими гіпотеза, що перевіряється, або приймається, або відхиляється. Для кожного виду гіпотез, що перевіряються, розроблені спеціальні критерії, серед яких найбільш часто використовуються *t*-критерії нормального розподілу і розподілу Стьюдента, *F-к*ритерій Фішера-Снедекора, χ^2 (хі-квадрат) розподілу Пірсона та інші.

Статистичні критерії, які використовуються для перевірки статистичних гіпотез, бувають двох видів: *параметричні і непараметричні.*

Параметричними називають критерії, які ґрунтуються на припущенні, що розподіл випадкової величини в сукупності підпорядкований деякому відомому закону (наприклад, нормальному, біноміальному, Пуассона та іншим). До таких критеріїв належать критерії *t, F,* χ^2 та інші.

*Непараметричними (порядковими***)** називають критерії, використання яких не пов'язане зі знанням закону розподілу випадкової величини, їх можна використовувати і тоді, коли досліджуваний розподіл значно відрізняється від нормального. До таких критеріїв належать, зокрема, критерій знаків, Вілкоксона, Уайта, Манна-Уїтні та інші.

Параметричні критерії більш ефективні порівняно з непараметричними. Однак вони можуть бути використані для сукупностей, які мають нормальний або близький до нормального розподіл. Непараметричні критерії можуть бути використані при будь-якій формі розподілу. Єдиною умовою їх застосування є взаємна незалежність даних спостереження.

Критерій перевірки гіпотези повинен бути підібраний так, щоб ризик допущення помилок був мінімальним. При цьому дуже важливо визначати ймовірність β того, що не буде допущена помилка другого роду (зазвичай за імовірність помилки другого роду використовують не β, а її доповнення до 1: 1-β). Ця ймовірність характеризує чутливість критерію до помилок другого роду і отримала назву *потужності критерію*.

Потужністю критерію називається ймовірність відхилення гіпотези H_0, що випробовується, коли правильна альтернативна гіпотеза H_a, і позначається (1-β). Отже, потужністю критерію є ймовірність того, що не буде допущена помилка другого роду. Очевидно, що бажано мати найбільш потужний критерій, оскільки це забезпечить мінімальну ймовірність допущення помилки II роду. Тому з усіх можливих критеріїв слід вибирати найбільш потужний.

Потужність (чутливість) критерію може бути підвищена двома способами: а) шляхом збільшення рівня значущості (проте цей шлях не зовсім прийнятний, оскільки не обґрунтовано підвищується ймовірність помилок першого роду; б) шляхом збільшення чисельності вибірки.

У більшості можливих значень вибраного критерію можна виділити дві неперетинні підмножини, одна з яких містить значення критерію, а друга ні. Перша підмножина називається критичною областю, а друга - областю допустимих значень.

Критичною областю називають ті значення критерію, при яких нульова гіпотеза відхиляється. Тому *потужністю* критерію називають ще ймовірність попадання критерію у критичну область при умові, що справедлива конкуруюча гіпотеза. Областю допустимих значень (областю прийняття H_0) називають сукупність значень критерію, при яких нульова гіпотеза приймається.

Точки, які відділяють критичну область від області допустимих значень, називаються *критичними точками*.

Основний принцип перевірки статистичних гіпотез: якщо фактичне (спостережуване) значення критерію належить критичній області, то нульову гіпотезу H_0 відхиляють; якщо фактичне значення належить області прийняття, то H_0 приймають.

Для кожного критерію складені спеціальні таблиці, за якими знаходять його табличне значення (критичні точки), які відділяють критичну область від області допустимих значень. Знайдене табличне значення критерію порівнюють з його фактичним значенням. Якщо фактичне значення критерію, що розраховане за даними вибірки, буде більшим від табличного значення, то нульова гіпотеза повинна бути відхилена і прийнята альтернативна гіпотеза. Якщо ж фактичне значення критерію буде меншим або рівним табличному, то робиться висновок про узгодження даних, що спостерігаються, з нульовою гіпотезою, тобто підстави для відмови від H_0 немає і тому вона повинна бути прийнята.

При цьому слід мати на увазі, що згода з нульовою гіпотезою не доводить її абсолютної справедливості. Це лише свідчення про необхідність подальшої її перевірки, зокрема, шляхом збільшення обсягу вибірки або поки більш переконливі дослідження не дозволять зробити протилежний висновок. Тому при формулюванні остаточних висновків у цьому випадку правильніше говорити про те, що дані спостереження не суперечать нульовій гіпотезі і, отже, не дають підстав для її відхилення.

Аналогічно, можна стверджувати, що відхилення нульової гіпотези не доводить її абсолютної неправильності; необхідна подальша її перевірка, зокрема, застосування інших методик дослідження, збільшення обсягу вибірки тощо.

Розрізняють односторонню та двосторонню критичні області. *Односторонньою н*азивають правосторонню або лівосторонню критичну область. Ці області визначаються такими нерівностями: для правосторонньої критичної області: $k > a_{кр}$, де $a_{кр.}$ - додатне число, для лівосторонньої $-k < a_{кр}$, *де* $a_{кр}$- від'ємне число.

Двостороння критична область визначається нерівностями: *k*<a_1, *k*>a_2, де a_2>a_1 або коротко $|k| > a_{кр}$, де $a_{кр} > 0$.

Вибір односторонньої або двосторонньої критичної області залежить від конкретних умов і мети завдань, що розв'язуються. Наприклад, при альтернативній гіпотезі H_a : $\overline{x_1}$ ≠ $\overline{x_2}$ слід користуватися двосторонньою критичною областю, а при гіпотезах H_a : $\overline{x_1}$ > $\overline{x_2}$ або H_a : $\overline{x_1}$ < $\overline{x_2}$ - односторонньою (відповідно правосторонньою і лівосторонньою) критичною областю.

Критичну область доцільно будувати так, щоб вона найкращим чином відрізняла нульову гіпотезу від альтернативної.

3.4. Схема (алгоритм) перевірки статистичної гіпотези

Узагальнюючи, можна навести загальну схему (алгоритм) перевірки статистичної гіпотези. Ця перевірка, як зазначалось вище, може бути проведена з використанням параметричних або непараметричних критеріїв. Наведемо схему перевірки гіпотези, яка передбачає знання закону розподілу генеральної сукупності, тобто для випадків застосування параметричних критеріїв.

Перевірка вказаної статистичної гіпотези передбачає послідовне виконання таких етапів:

1. Оцінка вихідної інформації та опис статистичної моделі вибіркової сукупності.

2. Формулювання нульової та альтернативної гіпотез.

3. Встановлення рівня значущості, за допомогою якого контролюється помилка першого роду.

4. Вибір найпотужнішого критерію для перевірки нульової гіпотези, що дозволить контролювати ймовірність появи помилки другого роду.

5. Розрахунок за певним алгоритмом фактичного (емпіричного, експериментального, спостережуваного) значення критерію $K_ф$.

6. Визначення критичної області та області згоди з нульовою гіпотезою, тобто встановлення табличного (теоретичного) значення критерію $K_т$.

7. Співставлення фактичного і табличного значень критерію та формулювання висновків за результатами перевірки нульової гіпотези.

На практиці частіше доводиться розв'язувати задачі двох типів з перевірки гіпотез. Задачі першого типу пов'язані з перевіркою гіпотез про істотність відмінностей між параметрами статистичних сукупностей. Прикладом таких задач може бути оцінка вірогідності відмінностей між середніми, дисперсіями, коефіцієнтами кореляції, регресії тощо.

Задачі другого типу пов'язані з перевіркою гіпотез про істотність відмінностей законів розподілу. До них належать задачі з визначення відповідності вибіркового розподілу теоретичному, частіше всього нормальному, оцінці близькості двох емпіричних розподілів, однорідності складу декількох сукупностей та ін.

Кожний із вказаних типів задач підрозділяють на різні види. Всі вони знаходять широке прикладне використання в дослідженнях, зокрема і в економіці, аграрному господарстві.

Розглянемо деякі важливі передумови і властивості перевірки статистичних гіпотез задач першого типу, які пов'язані із застосуванням параметричних критеріїв і припущення нормального розподілу в генеральній сукупності.

Вибір конкретного методу перевірки гіпотези залежить від характеру досліджуваної гіпотези, властивостей вихідної інформації та інших умов.

Розглянемо основні з них.

1. *Обсяг вибіркової сукупності*. При перевірці гіпотез за даними великих вибірок $(n>30)$ доцільно використовувати t-критерій нормального розподілу, а за даними малих вибірок $(n<30)$ t -критерій розподілу Стьюдента.

2. *Рівність вибірок за чисельністю*. Вибіркові сукупності за чисельністю можуть бути рівними і нерівними. Ці властивості необхідно враховувати при практичній перевірці гіпотез про істотність відмінностей між середніми, зокрема, при розрахунку середньої помилки двох вибіркових середніх.

3. *Схема формування вибірок*. Прийоми перевірки статистичних гіпотез залежать від характеру формування

вибіркових сукупностей. Якщо спостереження однієї вибірки не протиставляються спостереженням іншої вибірки, то такі вибірки називаються *незалежними.* Якщо ж спостереження однієї вибірки деякою мірою пов'язані зі спостереженням іншої вибірки, то такі вибірки називаються *залежними.* Прикладом незалежних вибірок може служити дослід з двома групами тварин, одна з яких є контрольною, а на другій випробовується якийсь препарат. При цьому дослідна і контрольна група формується випадково. Однак, якщо при проведенні цього ж досліду тварини будуть попередньо розподілені на групи за будь-якими ознаками (наприклад, біологічними властивостями, продуктивністю тощо), а потім від кожної пари аналогів буде відібрано по одному представнику в контрольну і дослідну групи, то спостереження за такими вибірками, вже не можна розглядати як незалежні.

Формування вибіркових сукупностей зумовлює різницю прийомів оцінки вірогідності між середніми двох малих вибірок. Якщо вибірки незалежні, то статистичній оцінці підлягає різниця середніх, якщо залежні - середня різниця.

У практичних дослідженнях часто виникає необхідність встановити характер розподілу генеральних сукупностей. При цьому можуть виникати задачі трьох видів:

1) про узгодженість емпіричного (вибіркового) і теоретичного (генерального) розподілу;

2) про незалежність розподілу однієї ознаки від іншої;

3) про однорідність двох або більше емпіричних розподілів.

Такі гіпотези також, як і гіпотези відносно параметрів розподілу, перевіряють за допомогою спеціальних критеріїв згоди.

Критерієм згоди називають критерій перевірки гіпотези щодо припущеного закону невідомого розподілу в генеральній сукупності.

Існує ряд критеріїв згоди: К. Пірсона, О.М. Колмогорова, М.В. Смирнова, Б.С. Ястремського та ін.

Ці критерії дозволяють встановити, узгоджуються чи не узгоджуються досліджувані розподіли з теоретичними розподілами, а також те, наскільки істотні розбіжності, які є між цими розподілами.

Оскільки нормальний розподіл зустрічається дуже часто, то найчастіше перевіряють гіпотези про відповідність вибіркового розподілу нормальному. Однак, поряд з нормальним розподілом генеральні сукупності можуть бути розподілені й за іншими законами. Тому вибір теоретичного закону розподілу повинен базуватися на глибокому розумінні характеру формування досліджуваного явища або процесу.

Певну роль у розв'язанні цього питання відіграє розрахунок статистичних характеристик вибіркових розподілів і побудова графіків (гістограми, полігону тощо). Так, якщо середня арифметична та дисперсії рівні або дуже близькі одна до одної, то можна припустити, що вибірковий розподіл відповідає розподілу Пуассона.

Для перевірки гіпотези щодо відповідності вибраних законів розподілу (нормальний, біноміальний, Пуассона тощо) розподілу в генеральній сукупності в більшості випадків при розрахунку критеріїв згоди використовуються відхилення емпіричних частот від теоретичних. Чим менше це відхилення, тим точніше теоретичний розподіл відтворює вибірковий і навпаки.

3.5. Критерій Пірсона

З множини критеріїв згоди, які використовуються при перевірці гіпотез щодо розподілів, частіше від інших застосовується найбільш потужний параметричний критерій

Пірсона. За цим критерієм можна з'ясувати, чи відноситься розподіл тієї чи іншої ознаки варіаційних рядів до нормального або іншого закону розподілу. Він розраховується як сума часток від ділення квадрата різниці між емпіричними та теоретичними частотами на теоретичні частоти:

$$\chi^2 фак. = \sum_{i=1}^{l} \frac{(n_i - n_i')^2}{n_i'}, \qquad (3.5.1)$$

де l - число інтервалів (класів, груп), на які розбито вибірковий розподіл; n_i - частоти емпіричного розподілу; n_i' - частоти теоретичного розподілу (вирівнювальні частоти).

З формули випливає, що чим менше розходження між n_i і n_i', тим ближче за значенням одна до одної емпіричні та теоретичні частоти. При повному співпаданні теоретичних та вибіркових частот χ^2 = 0, у противному випадку χ^2 > 0. Область зміни χ^2 простягається від 0 до ∞. При великому числі ступенів вільностей ($k \to \infty$) розподіл χ^2 набуває форму, яка близька до нормального розподілу.

Для оцінки близькості емпіричного та теоретичного розподілів необхідно розрахувати фактичне значення χ^2 і порівняти його з табличним значенням при заданому рівні значущості α і відповідному числі ступенів вільностей k.

Число ступенів вільностей визначають по-різному залежно від характеру гіпотези, що перевіряється, і особливостей вихідної інформації. Так, якщо перевіряється гіпотеза про узгодженість вибіркового та теоретичного розподілів, то число ступенів вільностей визначають за формулою $k=l-1-r$, де l — число інтервалів (класів, груп) вибірки; r - число параметрів генерального розподілу, які оцінюються за даними вибірки.

При оцінці відповідності емпіричного розподілу нормальному, число ступенів вільностей $k = l-1 - 2 = l-3$, так як для побудови кривої нормального розподілу оцінюється два параметри: середня арифметична і середнє квадратичне відхилення. Якщо перевіряється відповідність вибіркового розподілу Пуассона, то оцінюється один параметр λ. Тоді число ступенів вільностей *k* = *l* -1 -1 = *l* - 2.

Якщо вихідні дані подані у вигляді таблиці розподілу частот і необхідно перевірити гіпотезу про незалежність розподілу двох ознак, то число ступенів вільностей визначають за формулою *k = =(a-1)(b-1),* де *a* - число рядків; *b* - число стовпчиків.

Так, число ступенів вільностей

$$k = (a - 1)(b - 1) = (2 - 1)x(2 - 1) = 1$$

для таблиці 2x2, *k*= 4 для таблиці 3x3, *k* = 2 для таблиці 3x2 і т.д.

Якщо перевіряється гіпотеза про *однорідність* двох сукупностей, то число ступенів вільностей визначають за формулою *k = п - 1,* де *п* - число інтервалів (класів, груп).

Як видно, для всіх випадків число ступенів вільностей, крім обов'язкових обмежень, завжди зменшується на одиницю, тобто має місце один лінійний обмежуючий зв'язок - рівність сум емпіричних та теоретичних частот.

Якщо отримане за вибіркою значення $\chi^2_{факт.} \leq \chi^2_{т.}$, то нульова гіпотеза приймається. Якщо ж $\chi^2_{факт.} > \chi^2_{т.}$, то нульова гіпотеза відхиляється.

Якщо вихідні дані подані у вигляді таблиці з чотирьох клітинок розподілу частот за двома ознаками (2x2) з числами n_i,

n_1	n_2
n_3	n_4

то фактичне $\chi^2_{факт.}$ значення може бути визначене за формулою

$$\chi^2_{факт.} = \frac{(n_1 n_4 - n_2 n_3)^2 (n_1 + n_2 + n_3 + n_4)}{(n_1 + n_2)(n_3 + n_4)(n_1 + n_3)(n_2 + n_4)} \qquad (3.5.2)$$

Критерій Пірсона використовується для розв'язання ряду задач, зокрема, при перевірці гіпотез щодо згоди (відповідності) вибіркового та теоретичного розподілів, про незалежність розподілів, про однорідність сукупностей. Відповідно до цих задач критерій χ^2 називають критерієм згоди, критерієм незалежності та однорідності.

Використання критерію χ^2 потребує дотримання ряду умов, найважливішими серед яких є:

1) обсяг вибірки повинен бути достатньо великим (при n <50 потужність критерію χ^2 помітно знижується);

2) чисельність окремих інтервалів (класів) повинна включати не менше п'яти одиниць. Якщо ця умова не виконується, то проводиться об'єднання малочисельних інтервалів з числом одиниць менше 5 (як виняток таких інтервалів не повинно бути більше 20% від їх загального числа); при цьому об’єднують і відповідні частоти. Якщо відбулось чергове об’єднання частот, то при знаходженні числа ступенів вільностей треба за s брати кількість груп вибірки, що залишились після об’єднання частот.

3) частоти не можна перетворювати в частість, оскільки це може призвести до збільшення величини відхилень n_i- n_i'.

3.6. Перевірка гіпотези про нормальний розподіл генеральної сукупності

Нехай задано емпіричний розподіл (у вигляді послідовності *рівновіддалених* варіант і відповідних їм частот):

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

Потрібно за критерієм Пірсона перевірити гіпотезу про те, що генеральна сукупність *N* розподілена нормально (тобто визначається двома параметрами: *M(X)=a* та середнім квадратичним відхиленням σ).

Для того, щоб при заданому рівні значущості α перевірити гіпотезу про нормальний розподіл генеральної сукупності треба:

1. Обчислити безпосередньо (при малій кількості спостережень) або спрощеним методом (при великій кількості спостережень) $\overline{x_в}$ та σ.

2. Обчислити теоретичні частоти n_i' за формулою $n_i' = \frac{nh}{\sigma} \cdot \varphi(u_i)$, де *n* – об'єм вибірки, *h* – крок (різниця між двома сусідніми варіантами), $u_i = \frac{x_i - \overline{x_в}}{\sigma_в}$, а значення функції $\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$ знаходять за таблицею 1 додатків.

3. Знайти спостережуване (фактичне, емпіричне, експериментальне) значення критерію

$$\chi^2 фак. = \sum_{i=1}^{l} \frac{(n_i - n_i')^2}{n_i'}. \qquad (3.6.1)$$

4. За таблицею критичних точок розподілу χ^2 за даним рівнем значущості α і числом ступенів вільностей $k = s - 3$ (де s - число груп вибірки) знаходять критичну (табличну, теоретичну) точку $\chi^2_{кр}(\alpha\ ;\ k)$ правосторонньої критичної області (таблиця 5 додатків). Число α - наперед задана достатньо мала величина - та мінімальна ймовірність, починаючи з якої можна признати подію практично неможливою. Найчастіше беруть α =0,05; 0,01; 0,001. Це означає, що у середньому в п'яти (1; 1/10; …) випадках із 100 є ризик відхилити правильну нульову гіпотезу H_0.

4. Порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона.

Якщо $\chi^2_{сп.} \le \chi^2_{кр.}$, то нульова гіпотеза H_0 приймається. Тобто емпіричні та теоретичні частоти відрізняються дуже мало (випадково).

Якщо $\chi^2_{факт.} > \chi^2_{кр.}$, то нульова гіпотеза H_0 відхиляється, тобто емпіричні і теоретичні частоти відрізняються значно.

Фактичне значення $\chi^2_{сп..}$ можна також розрахувати за іншою формулою:

$$\chi_{сп.}^{2} = \sum_{i=1}^{l} \frac{n_i^2}{(n_i^{'})^2} - n \qquad (3.6.2)$$

Приклад 3.6.1. Використовуючи Критерій Пірсона, при рівні значущості 0,05 перевірити, чи узгоджується гіпотеза про нормальний розподіл генеральної сукупності *X* з емпіричним розподілом вибірки об'єму *n* = 200:

x_i	15	17	19	21	23	25	27	29	31
n_i	15	26	25	30	26	21	24	20	13

Розв'язання. Знайдемо: $\bar{x}_в$=

$$\frac{\sum_{i=1}^{k} x_i n_i}{n} = \frac{15\cdot 15+17\cdot 26+19\cdot 25+21\cdot 30+23\cdot 26+25\cdot 21+27\cdot 24+29\cdot 20+31\cdot 13}{200} == 22,63.$$

$$D_{в.} = \frac{\sum_{i=1}^{k} n_i \left(x_i\right)^2}{n} - \overline{(x_{в.})}^2 =$$

$$= \frac{15^2\cdot 15+17^2\cdot 26+19^2\cdot 25+21^2\cdot 30+23^2\cdot 26+25^2\cdot 21+27^2\cdot 24+29^2\cdot 20+31^2\cdot 13}{200} - (22,63)^2 =$$

$$= 534,16 - 512,1169 = 22,0431.$$

$\sigma_в$=4,695.

Для обчислення теоретичних частот $n'_i = \frac{n \cdot h}{\sigma_{в.}}$ $\cdot$ $\varphi(u_i) = \frac{200 \cdot 2}{4,695} = 85,2\varphi(u_i)$ складемо розрахункову таблицю (табл. 3.6.1):

Таблиця 3.6.1

i	$\overline{x_i}$	$u_i = \frac{x_i - \overline{x_в}}{\sigma_в}$	$\varphi(u_i)$	n'$_i$=85,2φ(u$_i$)
1	15	-1,63	0,1057	9,01
2	17	-1,2	0,1942	16,55
3	19	-0,77	0,2966	25,27
4	21	-0,35	0,3752	31,97
5	23	0,08	0,3977	33,88
6	25	0,5	0,3521	30
7	27	0,93	0,2589	22,06
8	29	1,36	0,1582	13,48
9	31	1,78	0,0818	6,97

Отримаємо розподіл теоретичних (вирівнювальних) частот:

n_i	15	26	25	30	26	21	24	20	13
$n_i^{/}$	9,01	16,6	25,27	32	33,9	30	22,1	13,5	6,97

Знайдемо $\chi^2_{сп.} = \sum_{i=1}^{l} \frac{(n_i - n'_i)^2}{n'_i}$. Складемо нову розрахункову таблицю (табл. 3.6.2).

З таблиці (додаток 5) знайдемо табличне значення критерію $\chi^2_{кр}(\alpha\ ;\ k)$: $\chi^2_{кр}(0,01;6) = 0,872.$ ($k = s - 3 = 9 - 3 = 6$ - число ступенів вільностей, а $s = 9$ - кількість спостережень).

Так як $\chi^2_{сп.} > \chi^2_{кр.}$, то нульову гіпотезу відхиляємо, отже емпіричні та теоретичні частоти розходяться значно. Дані спостереження не узгоджуються з гіпотезою про нормальний розподіл генеральної сукупності.

Таблиця 3.6.2

i	n_i	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$
1	15	9,01	5,99	35,88	3,98
2	26	16,55	9,45	89,3	5,4
3	25	25,27	-0,27	0,073	0,003
4	30	31,97	-1,97	3,881	0,12
5	26	33,88	-7,88	62,09	1,83
6	21	30	-9	81	2,7
7	24	22,06	1,94	3,764	0,17
8	30	13,48	16,52	272,9	20,25
9	13	6,97	6,03	36,36	5,22
сума					$\chi^2_{сп.}$=39,673

Приклад 3.6.2. Дано: емпіричні частоти

n_i	6	13	38	74	106	85

і теоретичні частоти

n_i'	3	14	42	82	99	76

При заданому рівні значущості α = 0,01 перевірити, чи випадкова відмінність між емпіричними та теоретичними частотами. Можливо, ця відмінність випадкова і пояснюється невеликим числом спостережень, або способом відбору або іншими причинами.

Розв'язання. В якості критерію перевірки нульової гіпотези приймаємо випадкову величину $\chi^2 сп. = \sum_{i=1}^{l} \frac{(n_i - n_i')^2}{n_i'}$.

$$\chi^2_{сп.} = \frac{(6-3)^2}{3} + \frac{(13-14)^2}{14} + \frac{(38-42)^2}{42} + \frac{(74-82)^2}{82} + \frac{(106-99)^2}{99}$$
$$+ \frac{(85-76)^2}{76} + \frac{(30-37)^2}{37} + \frac{(14-13)^2}{13} = 7{,}19.$$

Обчислимо число ступенів вільностей $k = s - 3 = 8 - 5 = 3$, де s =8 - кількість спостережень. Табличне значення критерію дорівнює $\chi^2_{кр}(\alpha;\ k) = \chi^2_{кр}(0{,}01;\ 5) = 11{,}1.$

Так як $\chi^2_{сп.} \le \chi^2_{кр.}$, то нульову гіпотезу приймаємо. Отже, емпіричні та теоретичні частоти розходяться незначно. Дані спостереження узгоджуються з гіпотезою про нормальний розподіл генеральної сукупності.

3.7. Порівняння двох дисперсій нормальних генеральних сукупностей

При перевірці гіпотез щодо дисперсій двох нормальних генеральних сукупностей можливі два випадки по відношенню до вибіркових дисперсій:

1) дисперсії рівні ($\sigma_1^2 = \sigma_2^2$);

2) дисперсії нерівні ($\sigma_1^2 \neq \sigma_2^2$).

Виникає спеціальне завдання перевірки гіпотези про істотність відмінностей двох дисперсій. Для перевірки гіпотези про рівність двох дисперсій генеральних сукупностей використовується критерій *F* розподілу Фішера, який базується на співвідношенні двох вибіркових виправлених дисперсій: S_1^2, S_2^2, що заміняють значення дисперсій у генеральних сукупностях, які, як правило, невідомі:

$$F_{сп} = \frac{S_1^2}{S_2^2}, \qquad (3.7.1)$$

де $S_1^2 > S_{2.}^2$

Критичні значення критерію Фішера *($F_{кр.}$)* знаходять за спеціальними таблицями (додаток 7) при відповідному числі ступенів вільностей та заданому рівні значущості *α*.

Перевірка H_0 зводиться до такого:

1. Спочатку визначають фактичне значення критерію $F_{сп.}$.

2. Потім за таблицями (додаток 7) знаходять його критичне значення $F_{кр.}$.

3. Якщо виявиться, що $F_{сп.} > F_{кр.}$, то H_0 відхиляється. Якщо ж $F_{сп.} < F_{кр.}$, то H_0 приймається.

Нехай генеральні сукупності *X* і *Y* розподілені нормально, відомі вибірки об'ємів n_1 (з *X*) і n_2 (з *У*). Треба за даними вибірками перевірити при заданому рівні значущості нульову гіпотезу: H_0 : $D(X) = D(Y)$, тобто необхідно вияснити значно чи незначно відрізняються виправлені дисперсії.

Належно від того, рівні чи нерівні дисперсії в генеральних сукупностях, необхідно видозмінювати схему перевірки гіпотези.

В якості критерію перевірки нульової гіпотези, приймається відношення більшої виправленої дисперсії до меншої

Якщо нульова гіпотеза справедлива, тобто генеральні дисперсії одинакові, то відмінність виправлених генеральних дисперсій пояснюється випадковими причинами, одна із яких – випадковий відбір об'єктів вибірки.

$$F_{\text{сп.}} = \frac{S^2_{Більше}}{S^2_{Менше}}. \qquad (3.7.1`)$$

Критична область будується в залежності від виду конкуруючої гіпотези. Розглянемо два випадки перевірки гіпотези.

І випадок: H_0: $D(X) = D(Y)$.

H_1: $D(X) > D(Y)$.

Будують правосторонню критичну область і знаходять $F_{кр}(\alpha\ ,\ k_1\ ,\ k_2)$ (критичне значення критерію), де $k_1 = n_1 - 1$ - число ступенів вільностей більшої дисперсії, $k_2 = n_2 - 1$ - число ступенів вільностей меншої дисперсії. Критичну точку $F_{кр}.(\alpha\ ,\ k_1\ ,\ k_2)$ знаходять за таблицею критичних точок Фішера-Снедекора.

Якщо $F_{сп} < F_{кр}$, то приймаємо нульову гіпотезу. В противному випадку ($F_{сп} > F_{кр}$) – нульову гіпотезу відхиляємо.

II випадок: $H_0:\ D(X) = D(Y).$

$H_1:\ D(X) \neq D(Y).$

Будують двосторонню критичну область. Щоб вибрати границі критичної області, треба рівень значущості *α* поділити на 2. Знаходять $F_{кр}(\frac{\alpha}{2},\ k_1,\ k_2)$ (критичне значення критерію), де $k_1 = n_1 - 1$ - ступінь вільностей більшої дисперсії, $k_2 = n_2 - 1$ - ступінь вільностей меншої дисперсії. Критичну точку $F_{кр}(\frac{\alpha}{2}\ ,\ k_1\ ,\ k_2)$ знаходять за таблицею критичних точок Фішера-Снедекора (додаток 7).

Якщо $F_{сп} < F_{кр}$, то приймаємо нульову гіпотезу. В противному випадку ($F_{сп} > F_{кр}$) нульову гіпотезу відхиляємо.

Приклад 3.7.1. За двома незалежними вибірками об'ємів n_1 =12 та n_2 =15, які отримали з нормальних генеральних сукупностей *X* і *У,* знайдені S^2_x = 11,41 та S^2_y = 6,52. При рівні значущості 0,05 перевірити нульову гіпотезу H_o: $D(X) = D(Y).$ про рівність генеральних дисперсій, при конкуруючій гіпотезі
H_1: $D(X) > D(Y).$

Розв'язання*.* Обчислимо фактичне значення критерію, як відношення більшої виправленої дисперсії до меншої:

$$Fcn. = \frac{S^2x}{S^2y} = \frac{11,41}{6,52} = 1,75.$$

Критична область правостороння ($D(X) > D(Y).$). За таблицею (додаток 7), за такими даними: α=0,05; k_1=12-1=11; k_2 =15-1=14 знайдемо *F кр*. (0,05; 11; 14)=2,57.

Оскільки *Fсп.* $<$ *Fкр*., то немає підстав відхилити нульову гіпотезу про рівність генеральних дисперсій.

Приклад 3.7.2. За двома незалежними вибірками об'ємів $n_1 = 10$ і $n_2 = 18$, які отримали із нормальних генеральних сукупностей *X* та *Y*, задані виправлені вибіркові дисперсії $S_x^2 = 1,23$; $S_y^2 = 0,41$. При рівні значущості $\alpha = 0,1$ перевірити нульову гіпотезу про рівність генеральних дисперсій H_0: $D(X) = D(Y)$ при конкуруючій гіпотезі H_1: $D(X) \neq D(Y).$

Розв'язання. Знайдемо фактичне значення критерію

$$F_{спост}. = \frac{S_{б.}^2}{S_{м.}^2} = \frac{1,23}{0,41} = 3.$$

Так як $D(X) \neq D(Y)$, то критична область двостороння:

$$F_{кр.}(\frac{\alpha}{2} = \frac{0,1}{2} = 0,05;\ k_1 = 10-1 = 9,\ k_2 = 18-1 = 17).\ F_{кр}(0,05;\ 9;17) = 2,5.$$

Оскільки $F_{сп} > F_{кр}$, то нульову гіпотезу про рівність генеральних дисперсій відхиляємо, тобто вибіркові виправлені дисперсії відрізняються значно.

3. 8. Порівняння двох середніх нормальних генеральних сукупностей, дисперсії яких невідомі та однакові

Нехай генеральні сукупності *X* та *Y* розподілені нормально, відомі вибірки об'ємів n (*X*) і m (*У*). Треба за даними вибірками перевірити при заданому рівні значущості α нульову гіпотезу: H_0 : $M(X) = M(Y),$ про рівність математичних сподівань двох нормальних генеральних сукупностей, дисперсії яких невідомі та однакові.

Щоб при заданому рівні значущості α перевірити нульову гіпотезу H_0: $M(X) = M(Y)$(про рівність математичних сподівань двох нормальних генеральних сукупностей з рівними дисперсіями) у випадку малих вибірок (об'єм яких менше 30), треба знайти фактичне значення критерію:

$$T_{спост..} = \frac{\overline{x} - \overline{y}}{\sqrt{(n-1)S_x^2 + (m-1)S_y^2}} \cdot \sqrt{\frac{(m+n-2)mn}{m+n}}., \qquad (3.8.1)$$

Потім за таблицею критичних точок Стьюдента (додаток 5), врахувавши рівень значущості α і число ступенів вільностей $k = m + n - 2$, знаходимо $t_{одн..кр.}(\alpha\ ;\ k)$ при конкуруючій гіпотезі $H_к$: $M(X) > M(Y)$, або $t_{двохстор..кр.}(\alpha\ ;\ k)$ при конкуруючій гіпотезі $H_к$: $M(X) \neq M(Y)$.

Якщо $|T_{\text{сп}}| < t_{\text{одност. кр}}(t_{\text{двостор. кр.}})$, то приймаємо нульову гіпотезу.

Якщо $|T_{\text{сп}}| > t_{\text{одност. кр}}(t_{\text{двостор. кр.}})$, то відхиляємо нульову гіпотезу.

Приклад 3.8.1*.* За двома незалежними малими вибірками об'ємів *n*=5 і *m*=6, які отримали із нормальних генеральних сукупностей *X* та *У,* знайдені вибіркові середні: $\overline{x}$=3,3; $\overline{y}$=2,48. При рівні значущості α=0,05 перевірити нульову гіпотезу H_0*: M(X)=M(У)* при конкуруючій гіпотезі H_1*: M(X)≠M(У).* Відомо, що S^2_x =0,25 і S^2_y=0,108.

Розв'язання. Так як вибіркові дисперсії різні, то спочатку перевіримо нульову гіпотезу H_o*: D(x)=D(y)* про рівність генеральних дисперсій, користуючись критерієм Фішера–Снедекора. Так як S^2_x значно більша від S^2_y, то H_1*: D(x)>D(y).*

$$F_{сп}. = \frac{0,25}{0,108} = 2,31.$$

За таблицею (додаток 7) при α =0,05; k_1=5-1=4 і k_2 =6-1=5, знаходимо *Fкр.*(0,05; 4; 5)=5,19. Оскільки *Fсп.<Fкр.,* то приймаємо нульову гіпотезу про рівність генеральних дисперсій.

Тепер порівняємо генеральні середні. Обчислимо фактичне значення критерію Стьюдента

$$T_{сп.} = \frac{\overline{x}-\overline{y}}{\sqrt{(n-1)S^2x+(m-1)S^2y}} \cdot \sqrt{\frac{nm(n+m-2)}{n+m}} = \frac{3,3-2,48}{\sqrt{(4\cdot 0,25+5\cdot 0,108}} \sqrt{\frac{5\cdot 6(5+6-2)}{5+6}} = 5,97.$$

За умовою критична область – двостороння. За рівнем значущості α = 0,05 і числом ступенів вільностей *k*=5+6-2=9, знаходимо із таблиці (додаток 6) критичну точку $t_{двост.кр.}$ (0,05; 9)= =2,26.

Так як $|T_{\text{сп}}| > t_{\text{двостор. кр. .}}$, то нульову гіпотезу про рівність генеральних середніх відхиляємо. Отже, генеральні середні відрізняються значно.

Запитання для самоперевірки

1. Які гіпотези називають статистичною, основною та альтернативною, простою та складною?
2. Що таке нульова гіпотеза, її суть?
3. Що таке помилки першого та другого роду при перевірці статистичної гіпотези?
4. Який зміст рівня значущості α?
5. Що називають статистичним критерієм, критичною областю та критичною точкою перевірки гіпотези?
6. В чому полягає задача статистичної перевірки?
7. Сформулюйте основний принцип перевірки гіпотез.
8. Як розрахувати фактичне значення t-критерію Стьюдента?
9. Які величини потрібно знати та як їх знайти, щоб визначити стандартнезначення t-критерію Стьюдента?
10. Як розрахувати фактичне значення F-критерію Фішера?

11. Які величини потрібно знати та як їх знайти, щоб визначити стандартнезначення F-критерію Фішера?

12. Як використовується F-критерій Фішера при порівнянні двох дисперсій нормальних генеральних сукупностей?

13. У чому полягає ідея застосування критеріїв при узгодженні теоретичного та емпіричного розподілів?

14. Який критерій використовують при перевірці двох дисперсій нормальних генеральних сукупностей?

15. Який критерій використовують при перевірці двох середніх нормальних генеральних сукупностей?

Задачі та вправи

3.1. За двома незалежними вибірками з об'ємами $n_1 = 11$ і $n_2 = 14$, які дістали із нормальних генеральних сукупностей *Х* та *У*, знайдені виправлені вибіркові дисперсії $S^2_x = 0{,}76$ і $S^2_y = 0{,}38$. При рівні значущості $\alpha = 0{,}05$ перевірити нульову гіпотезу H_0: $D(X) = D(У)$ про рівність генеральних дисперсій при конкуруючій гіпотезі H_1: $D(X) > D(У)$.

3.2. Двома приладами виміряні розміри 5 деталей. Отримані наступні результати (у мм):

$x_1 = 4$; $x_2 = 5$; $x_3 = 6$; $x_4 = 7$; $x_5 = 8$;

$y_1 = 5$; $y_2 = 5$; $y_3 = 8$; $y_4 = 4$; $y_5 = 6$.

Потрібно при рівні значущості $\alpha = 0{,}05$ перевірити нульову гіпотезу H_0: $D(X) = D(У)$ про рівність генеральних дисперсій при конкуруючій гіпотезі H_1: $D(X) > D(У)$.

3.3. За двома незалежними вибірками з об'ємами $n_1 = 14$ і $n_2 = 10$, які дістали із нормальних генеральних сукупностей *Х* та *У*, знайдені виправлені вибіркові дисперсії $S^2_x = 0{,}84$ та $S^2_y = 2{,}52$. При рівні значущості $\alpha = 0{,}1$ перевірити нульову гіпотезу H_0: $D(X) =$

$=D(У)$ про рівність генеральних дисперсій при конкуруючій гіпотезі H_1: $D(X) \neq D(У)$.

3.4. За двома незалежними малими вибірками об'ємами n_1=12 та n_2=18, які дістали із нормальних генеральних сукупностей X та $У$, знайдені вибіркові середні $\bar{y}$=29,2 та $\bar{x}$=31,2 і виправлені дисперсії $S^2x = 0,7$ і $S^2y = 0,4$. Потрібно при рівні значущості $\alpha = 0,05$ перевірити нульову гіпотезу H_0: $D(X) = D(У)$ при конкуруючій гіпотезі H_1: $D(X) \neq D(У)$.

3.5. Двома методами проведені вибірки однієї й тієї ж величини. Отримані результати:

а) у першому випадку: x_1=9,6; x_2=10; x_3=9,8; x_4=10,2; x_5=10,6;

б) у другому випадку: y_1=10,4; y_2=9,7; y_3=10; y_4=10,3.

Чи можна вважати, що обидва методи забезпечують однакову точність вимірів, якщо прийняти рівень значущості α=0,1? Припускається, що результати вимірів розподілені нормально та вибірки незалежні.

3.6. При рівні значущості α = 0,05 необхідно перевірити нульову гіпотезу H_0: $M(X) = M(У)$ про рівність генеральних середніх нормальних сукупностей X та $У$ при конкуруючій гіпотезі

H_1: $M(X) > M(У)$ за малими незалежними вибірками з об'ємами n =10 і m = 16.Отримані наступні результати:

x_i	12,3	12,5	12,8	13	13,5
n_i	1	2	4	2	1

y_i	12,2	12,3	13
n_i	6	8	2

3.7. Використовуючи критерій Пірсона при рівні значущості α= 0,05; перевірити чи згоджується гіпотеза про нормальний розподіл генеральної сукупності з емпіричним розподілом вибірки об'єму n = 200:

x_i	83	85	87	89	91	93	95	97	99	101
n_i	6	7	12	15	30	10	8	6	4	2

3.8. Використовуючи критерій Пірсона, при рівні значущості α=0,01 перевірити, чи згоджується гіпотеза про нормальний розподіл генеральної сукупності з емпіричним розподілом вибірки об'єму n =100:

x_i	8	16	40	32	36	18	10
n_i	16	12	14	17	13	18	10

3.9. За двома незалежними вибірками об'ємами n_1=9 і n_2=16, які отримані із нормальних генеральних сукупностей, знайдено виправлені вибіркові дисперсії S^2_x= 34,02 і S^2_y = 12,15. При рівні значущості α = 0,01 перевірити нульову гіпотезу H_0: $D(X) = D(У)$ про рівність генеральних дисперсій при конкуруючій гіпотезі H_1: $D(X) > D(У)$.

3.10. Для порівняння точності двох верстатів–автоматів взяли дві вибірки з об'ємами n_1= 10 і n_2= 8. У результаті вимірів контролюючого розміру відібраних деталей отримані наступні результати:

x_i	1,08	1,10	1,12	1,14	1,15	1,25	1,36	1,38	1,40	1,42
y_i	1,11	1,12	1,18	1,22	1,33	1,35	1,36	1,38		

Чи можна вважати, що верстати мають однакову точність при α=0,1?

3.11. За двома незалежними вибірками об'ємами n=5 і m=6, які дістали із нормальних сукупностей X та $У$, знайдено вибіркові середні $\bar{y}$=15,9 і $\bar{x}$=14,1. Відомо, що S^2_x = 14,76 і S^2_y = 4,96. При рівні значущості α=0,05 перевірити нульову гіпотезу H_0: $M(x) = M(y)$ при конкуруючій гіпотезі H_1: $M(x) \neq M(y)$.

3.12. З двох партій виробів, що виготовлені на двох однаково налаштованих верстатах, добули малі вибірки, об'єм яких n=10 і m=12. Отримати наступні результати:

контролюючий розмір виробів першого верстата:

	x_i	3,4	3,5	3,7	3,9;
частота (число виробів):	n_i	2	3	4	1;

контролюючий розмір виробів другого верстата:

	y_i	3,2	3,4	3,6;
частота:	m_i	2	2	8.

Чи можна вважати, що вироби мають однаковий розмір при α=0,1?

3.13. Використовуючи критерій Пірсона, при рівні значущості α=0,01 перевірити, чи узгоджується гіпотеза про нормальний розподіл генеральної сукупності з емпіричним розподілом 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

3.14. Використовуючи критерій Пірсона, при рівні значущості α=0,05 перевірити, чи узгоджується гіпотеза про нормальний розподіл генеральної сукупності з емпіричним розподілом вибірки об'єму n= 100:

x_i	30	35	37	39	41	43	45	47	49
n_i	16	9	7	13	10	15	14	6	10

3.15. Використовуючи критерій Пірсона, при рівні значущості α=0,05 встановити випадкове чи значне розходження між емпіричними частотами n_i та теоретичними частотами n_i', які обчислені, виходячи з гіпотези про нормальний розподіл генеральної сукупності:

а)

n_i	5	10	20	8	7
$n_i^{/}$	6	14	18	7	5

б)

n_i	6	8	13	15	20	16	10	75
$n_i^{/}$	5	9	14	16	18	19	9	67

3.16. Використовуючи критерій Пірсона, при рівні значущості: *а)* α=0,05, *б)* α=0,01 встановити випадкове чи значне розходження між емпіричними частотами n_i та теоретичними частотами n_i', які обчислені, виходячи з гіпотези про нормальний розподіл генеральної сукупності:

n_i	8	16	40	72	36	18	10
n_i'	6	18	36	76	39	18	7

РОЗДІЛ 4
ДИСПЕРСІЙНИЙ АНАЛІЗ

Дисперсійним аналізом називається вивчення впливу факторів на генеральну сукупність на основі аналізу певних дисперсій. Суть цього аналізу полягає в розкладанні загальної варіації випадкової змінної на незалежні доданки, кожен з яких характеризує вплив деякого фактора або їх взаємодію.

Дисперсійний аналіз – це математично-статистичний метод вивчення результатів спостереження, що залежні від різноманітних одночасно діючих факторів.

Дисперсійний аналіз може бути представлений у вигляді наступних етапів, що виконуються послідовно:

1) Визначення і розклад варіації;
2) Визначення числа ступенів вільностей варіації;
3) Визначення дисперсій;
4) Аналіз дисперсій та прийняття рішень про перевірку нульової гіпотези.

Факторами називаємо, взагалі кажучи, зовнішні умови, які супроводжують експерименти при формуванні вибірки. Вони можуть змінюватись під час експериментів, що дає можливість досліджувати їх вплив на результати експериментів. Для здійснення дисперсійного аналізу необхідно, щоб результати експериментів були незалежними нормально розподіленими випадковими змінними з однаковою дисперсією. Результати експериментів можна класифікувати та аналізувати за однією ознакою – *однофакторний аналіз*, і *многофакторний*: за двома ознаками – *двофакторний аналіз*, за трьома – *трифакторний аналіз* тощо.

За аналогією з комбінованими групуваннями многофакторні моделі дисперсійного аналізу мають значні переваги у порівнянні з однофакторними: вони дозволяють виявити ступінь впливу не лише кожного фактора окремо, а й їх взаємодію. Ця задача розв'язується шляхом побудови комбінаційних групувань і таблиць.

Відмінність многофакторного аналізу від однофакторного полягає в тому, що загальний обсяг варіації розкладається на більше число компонент (складових), ускладнюються розрахунки (в залежності від розкладання загального об'єму варіації на складові) і аналіз дисперсій. Многофакторні моделі з великою кількістю факторів (трьох і більше) доцільно досліджувати кореляційним методом з використанням інформаційно-комунікаційних технологій (ІКТ).

Схема розкладання загальної варіації залежить від формування груп, і може бути випадковим, якщо спостереження однієї групи не пов'язані зі спостереженнями іншої групи, або невипадковим (спостереження за обома вибірками мають спільні умови експерименту). Відповідно отримують незалежні і залежні вибірки.

На практиці найчастіше розглядаються залежні вибірки, коли умови для груп (підгруп) вирівнюються. Так, при польовому досліді вся ділянка розбивається на блоки з максимально рівними умовами. При цьому кожний варіант досліду отримує рівні можливості бути представленим у всіх блоках, що вирівнює умови для всіх варіантів досліду, що перевіряються. Такий метод побудови досліду має назву *рендомізації (рендомізованих блоків).*

Як відомо, вибіркова дисперсія не є незміщеною оцінкою дисперсії генеральної сукупності. Тому в дисперсійному аналізі доводиться застосовувати до вибіркових дисперсій (загальних, групових, міжгрупових тощо) поправки (множники виду $\frac{n}{n-1}$, де n – кількість елементів множини), які визначають міру розсіювання елементів цієї множини, що припадає на один ступіть вільностей. З означення виправленої вибіркової дисперсії випливає, що вона є якраз такою мірою. У зв’язку з тим, що виправлена вибіркова дисперсія є незміщеною і спроможною оцінкою дисперсії генеральної сукупності, будемо розглядати власне дисперсійний аналіз. Це, зокрема, ще і спрощує розрахункові формули. Аналізуючи певні виправлені вибіркові дисперсії, робимо об’єктивні висновки про вплив факторів, які нас цікавлять, на генеральну сукупність, при умові, що групи утворені по проявах цих факторів.

Нехай генеральна сукупність є продуктивністю праці робітників, що приготовляють одну і ту ж продукцію. Фактором будемо вважати освіченість робітників. Тоді рівнями (*p)* цього фактору є:незакінчена середня освіта, незакінчена вища освіта, середня освіта, вища освіта. Тут *p*=4. Визначимо продуктивність праці певної групи робітників. Тоді, наприклад, x_{35} - продуктивність праці п’ятого по порядку контрольованого робітника, що має середню освіту; а, x_{42} - другого по порядку контрольованого робітника, що має вищу освіту.

Критерій істинності у дисперсійному аналізі ґрунтується на тому, що генеральна сукупність нормально розподілена. Якщо ж вона не розподілена нормально, то оцінки середніх арифметичних не є незалежними від виправленої вибіркової дисперсії. Однак, при великих об’ємах вибірок самі вибіркові значення, як випадкові змінні, нормально розподілені. Отже, застосування в цьому випадку дисперсійного аналізу дає наближені висновки про вплив виділених факторів на генеральну сукупність, причому тим точніші, чим більший об’єм вибірки.

4.1. Однофакторний дисперсійний аналіз

Будь-який дисперсійний аналіз можна застосувати лише тоді, коли елементи вибірки можна групувати за ознаками за певними факторами.

Однакова кількість випробувань на різних рівнях. Нехай дано вибірку незалежних спостережень за деякою нормально розподіленою мінливою величиною *X*, на яку впливає фактор *F*, який має *p* постійних рівнів. Число випробувань на кожному рівні однакове і рівне *g.* Нехай спостерігалось *n=pg* значень x_{ij} ознаки *X,* де *i* – номер випробування (*i*= 1,2,…,*g)*, *j* - номер рівня фактора *(j*=1,2, …,*p).* Результати спостережень можна навести в таблиці.

Номер випробування	Рівні фактору F_j			
	F_1	F_2	…	F_3
1	x_{11}	x_{11}	…	x_{1p}
2	x_{21}	x_{22}	…	x_{2p}
…	…	…	…	…
n	x_{g1}	x_{g2}	…	x_{gp}
Групова середня	$\overline{x_1}$	$\overline{x_2}$	…	$\overline{x_p}$

Елемент таблиці x_{ij} - це той елемент даної вибірки, який має *i*-й рівень виділеного контрольованого фактора та *j*-й його прояв.

Обчислимо групові середні величини x_j:

$$\bar{x}_j = \frac{1}{g}\sum_{i=1}^{g} x_{ij} \qquad (4.1.1)$$

та загальновибіркове середнє

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{g}\sum_{j=1}^{p} x_{ij} = \frac{1}{p}\sum_{j=1}^{p}\bar{x}_j, \qquad (4.1.2)$$

так як $\frac{1}{g}\sum_{i=1}^{g} x_{ij} = \bar{x}_j$.

Розглянемо суму квадратів відхилень значень x_{ij} від загальної середньої:

$$\sum_{i=1}^{g}\sum_{j=1}^{p}(x_{ij}-\bar{x})^2=\sum_{i=1}^{g}\sum_{j=1}^{p}((x_{ij}-\bar{x}_J)+(\bar{x}_J-\bar{x}))^2=\sum_{i=1}^{g}\sum_{j=1}^{p}((x_{ij}-\bar{x}_J)^2+$$

$$+2(x_{ij}-\bar{x}_J)(\bar{x}_J-\bar{x})+(\bar{x}_J-\bar{x}))^2=$$

$$=\sum_{i=1}^{g}\sum_{j=1}^{p}(x_{ij}-\bar{x}_J)^2+\sum_{i=1}^{g}\sum_{j=1}^{p}(\bar{x}_J-\bar{x})^2+$$

$$+2\sum_{i=1}^{g}\sum_{j=1}^{p}(x_{ij}-\bar{x}_J)(\bar{x}_J-\bar{x})=$$

$$=\sum_{i=1}^{g}\sum_{j=1}^{p}((x_{ij}-\bar{x}_J)^2+2\sum_{i=1}^{g}(\bar{x}_J-\bar{x})\sum_{j=1}^{p}(x_{ij}-\bar{x}_J)+$$

$$+\sum_{i=1}^{g}(\bar{x}_J-\bar{x})\sum_{j=1}^{p}1=\sum_{i=1}^{g}g(\bar{x}_J-\bar{x})^2+\sum_{i=1}^{g}\sum_{j=1}^{p}(x_{ij}-\bar{x}_J)^2,$$

або в скороченому вигляді цю суму можна записати як

$$S_{заг.}=S_{зал.}+S_{факт.} \qquad (4.1.3)$$

При обчисленні використана властивість середнього: сума відхилень елементів вибірки від відповідного групового середнього рівна нулю, тобто $\sum_{j=1}^{p}(x_{ij}-\bar{x})=0$. Тут $S_{заг.}$ – загальна сума квадратів відхилень спостережуваних значень від загальної середньої; $S_{факт.}$ - факторна сума квадратів відхилень групових середніх від загальної середньої, що характеризує розсіювання між групами; $S_{зал.}$ – залишкова сума квадратів відхилень групових середніх від групової середньої, яка характеризує розсіювання всередині групи.

Отримаємо більш зручні формули для розрахунку $S_{заг}$ і $S_{факт.}$

Очевидно, що

$$S_{\text{заг.}}=\sum_{i=1}^{g}\sum_{j=1}^{p}(x_{ij}-\bar{x})^2=$$

$$\sum_{i=1}^{g}\sum_{j=1}^{p}(x_{ij}{}^2-2\bar{x}x_{ij}+\bar{x}^2)=\sum_{i=1}^{g}\sum_{j=1}^{p}x_{ij}{}^2- \qquad (4.1.4)$$

$$-2\sum_{i=1}^{g}\sum_{j=1}^{p}\bar{x}x_{ij}+\sum_{i=1}^{g}\sum_{j=1}^{p}\bar{x}^2=\sum_{i=1}^{g}\sum_{j=1}^{p}x_{ij}{}^2-$$

$-2\frac{1}{pg}(\sum_{i=1}^{g}\sum_{j=1}^{p}x_{ij})^2+\frac{1}{pg}(\sum_{i=1}^{g}\sum_{j=1}^{p}x_{ij})^2=\sum_{j=1}^{p}P_j-$

$-\frac{1}{n}(\sum_{j=1}^{p}R_j)^2.$

тоді

$$S_{\text{факт}}=g\sum_{i=1}^{g}(\overline{x_J}-\bar{x})^2$$
$$+g\sum_{i=1}^{g}(\overline{x_J}-2\bar{x}\overline{x_J}+\bar{x}^2)=g\sum_{j=1}^{p}x_{ij}{}^2$$
$$-gx\sum_{j=1}^{p}\overline{x_J}+gp\,\bar{x}^2=$$
$$=\frac{1}{g}\sum_{i=1}^{p}(\sum_{j=1}^{g}x_{ij})^2-2g\frac{1}{p}\sum_{j=1}^{p}\overline{x_J}\sum_{j=1}^{p}\overline{x_J}+ \qquad (4.1.5)$$
$$+\frac{g}{p}(\sum_{j=1}^{p}x_j)^2=\frac{1}{g}\sum_{i=1}^{p}(\sum_{j=1}^{g}x_{ij})^2$$
$$-\frac{1}{pg}\sum_{i=1}^{p}(\sum_{j=1}^{g}x_{ij})^2=$$
$$=\frac{1}{g}\sum_{j=1}^{p}(R_j)^2-\frac{1}{pg}(\sum_{j=1}^{p}R_j)^2$$

де $P_j=\sum_{j=1}^{g}x_{ij}{}^2$ - сума квадратів значень ознаки на рівні F_j, $R_j=\sum_{j=1}^{g}x_{ij}$- сума значень ознаки на рівні F_j.

Зауваження. Обчислення за формулами (4.1.3) та (4.1.4) можна спростити, якщо виконати заміну змінних $y_{ij}=x_{ij}-C$, де C – «хибний» нуль, що приблизно дорівнює загальній середній. Підставляючи в ці формули $x_{ij}=y_{ij}+C$ і $R_j=\sum_{j=1}^{g}x_{ij}=\sum_{j=1}^{g}(y_{ij}+C)==\sum_{j=1}^{g}y_{ij}+gC=T_j+gC$, де $T_j=\sum_{j=1}^{g}y_{ij}$, з урахуванням позначення $G_j=\sum_{j=1}^{g}y_{ij}{}^2$, отримаємо:

$$S_{заг.}=\sum_{i=1}^{p}(\sum_{j=1}^{g}y_{ij}+C)^2-\frac{1}{pg}(\sum_{j=1}^{p}T_j+gC)^2= \qquad (4.1.6)$$

$$\sum_{i=1}^{g}\sum_{j=1}^{p}({y_{ij}}^2 + +2Cy_{ij} + C^2\ ) - \frac{1}{pg}(\sum_{j=1}^{p} T_j +$$
$$pgC)^{\ \ 2} = \sum_{j=1}^{p}(Gj + 2\mathrm{C}T_j + gC^2) — \frac{1}{pg}((\sum_{j=1}^{p} T_j)^2 +$$
$$2pgC\sum_{j=1}^{p} T_j + p^2g^2C^2\big) = \sum_{j=1}^{p} Gj + 2C\sum_{j=1}^{p} T_j +$$
$$+pgC^2 - \frac{1}{pg}((\sum_{j=1}^{p} T_j)^2 - C\sum_{j=1}^{p} T_j +\ pgC^2 =$$

$$=\sum_{j=1}^{p} Gj - \frac{1}{n}(\sum_{j=1}^{p} T_j)^2,$$

факторна сума

$$S_{\text{факт.}} = \frac{1}{g}\sum_{j=1}^{p} T_j^{\ \ 2} - \frac{1}{pg}(\sum_{j=1}^{p} T_j)^2. \qquad (4.1.7)$$

Зауваження*.* Якщо спостережувані значення x_{ij} - десяткові дроби з *k* знаками після коми, то краще перейти до чисел виду (умовних варіант) $y_{ij} = 10^k x_{ij} - C$, де *C* приблизно дорівнює загальній середній чисел $10^k x_{ij}$. Хоча при цьому факторна і залишкові дисперсії збільшуються в 10^{2k} раз, їх відношення не зміниться.

У формулі (4.1.3) $S_{заг.} = S_{зал.} + S_{факт.}$ полягає основна ідея дисперсійного аналізу. Але в ній аналізуються не самі суми квадратів відхилень, а так звані середні квадрати, що є незміщеними оцінками відповідних дисперсій, які отримуються діленням сум квадратів відхилень на відповідне число ступенів вільностей.

Отже, факторна сума

$$S_{\text{факт.}} = \frac{\sum_{j=1}^{p} \mathrm{T}j^2}{g} - \frac{(\sum_{j=1}^{p} \mathrm{T}j)^2}{pg} \qquad (4.1.7)$$

характеризує вплив фактора *F*, який має *p* рівнів.

Залишкова сума відображає вплив випадкових величин на результати спостережень (випробувань).

$$S_{зал.} = S_{заг.} - S_{факт.}; \qquad S_{заг.} = \sum_{j=1}^{p} G_l - \frac{\left(\sum_{j=1}^{p} \mathrm{T}j\right)^2}{pg}$$

Згадаємо, що число ступенів вільностей визначається як загальне число спостережень мінус число рівнянь, що їх зв'язує.

Тому число ступенів вільностей загальної дисперсії рівне $pg - 1 = n - 1$, оскільки один ступінь вільностей губиться при визначенні середньої. Аналогічно, число ступенів вільностей факторної дисперсії рівне $p - 1$, так як групові середні варіюють біля однієї загальної середньої. Нарешті, число ступенів вільностей залишкової дисперсії рівне різниці між числом ступенів вільностей загальної та факторної дисперсій: $pg - 1 - (p - 1) = pg - p = n - p$, так як використовуються p відношень при обчисленні p групових середніх $\overline{x_j}$.

Використовуючи отримані значення сум квадратів і відповідних значень ступенів вільностей, можна обчислити незміщені оцінки трьох дисперсій. Якщо розділити суми квадратів відхилень на відповідне число ступенів вільностей, то отримаємо загальну, факторну та залишкову дисперсії:

$$S^2_{заг.} = \frac{S_{заг}}{n-1};\ S^2_{факт.} = \frac{S_{фак}}{p-1};\ S^2_{зал.} = \frac{S_{зал}}{n-p}, \qquad (4.1.8)$$

де: p – число рівнів фактору, g – число спостережень на кожному рівні, $n = pg$.

Кожну групу вибіркових даних можна вважати окремою вибіркою. Якщо фактор впливає на генеральну сукупність, то ці вибірки однорідні, тобто немає суттєвої різниці між середніми арифметичними окремих груп. Розподіл середнього арифметичного нормальної генеральної сукупності не залежить від розподілу варіанти. За припущенням, вибірки є однорідними, тому середні арифметичні груп незалежні в групах. Середнє арифметичне вибірки з нормально розподіленої популяції також має нормальний розподіл. Отже, обидві дисперсії $S^2_{факт.}$ та $S^2_{зал..}$ є незміщеними оцінками дисперсії σ^2 тієї ж самої нормально розподіленої генеральної сукупності, а їх відношення має розподіл Фішера. Тому для перевірки гіпотези про вплив на генеральну сукупність виділеного фактора можна скористатися критерієм Фішера.

Перевірка нульової гіпотези H_0 про рівність групових математичних сподівань базується на порівнянні дисперсій $S^2_{факт.}$ та $S^2_{зал..}$. Якщо гіпотеза H_0 правильна, то правильна й гіпотеза про рівність факторної та залишкової дисперсій, яка перевіряється за критерієм Фішера–Снедекора:

$$F_{спост.} = \frac{S^2_{фак}}{S^2_{зал}}, \qquad (4.1.9)$$

який має ступені вільностей $k_1=p-1,\ k_2=n-p.$

Якщо нульова гіпотези H_0 про рівність групових математичних сподівань хибна, то хибна й гіпотеза про рівність $S^2_{факт.}$ та $S^2_{зал.}$

Тоді алгоритм перевірки цієї гіпотези у випадку правосторонньої критичної області буде наступним:

1. H_0: $m_1 = m_1 = ... = m_p$.

2. Формуємо з даної вибірки таблицю, де x_{ij} – це *j*-те спостереження на *i*-тому рівні фактора.

3. Вибираємо рівень значущості *α.*

4. Знаходимо емпіричне (спостережуване) значення статистики Фішера (4.1.9) $F_{спост} = \frac{S^2_{факт.}}{S^2_{зал.}}$.

5. Обчислюємо ступені вільностей $k_1=p-1,\ k_2=n-p.$

6. При вибраному рівні значущості α та кількості ступенів вільностей знаходимо критичне значення $F_{.кр}\ (\alpha,\ k_1,\ k_2)$ статистики Фішера.

7. Якщо $F_{спост..} < F_{.кр}$, то виділений фактор не впливає на генеральну сукупність. В противному випадку цей вплив суттєвий.

Приклад 4.1. Проведено по 4 випробування на кожному із трьох рівнів. Методом дисперсійного аналізу при рівні значущості α = 0,05 перевірити нульову гіпотезу про рівність групових середніх. Вибірки дістали із нормальних сукупностей з однаковими дисперсіями.

Номер випробування *i*	*Рівні фактору F_j*		
	F_1	F_2	F_3
1	51	52	42
2	52	54	44
3	56	56	50
4	57	58	52
$\overline{x}_{гр.j}$	54	55	47

Розв'язання. Для спрощення підрахунків віднімемо *C*=52 від кожного спостережуваного значення, тобто введемо умовні варіанти $y_{ij} = x_{ij} - 52$.

Номер випробу-вання *i*	Рівні фактору F_j						Підсумок
	F_1		F_2		F_3		
	y_{i1}	y^2_{i1}	y_{i2}	y^2_{i2}	y_{i3}	y^2_{i3}	
1	-1	1	0	0	-10	100	
2	0	0	2	4	-8	64	
3	4	16	4	16	-2	4	
4	5	25	6	36	0	0	
$G_j=\sum_{i=1}^{4} y_{ij}^2$		42		56		168	$\sum_{j=1}^{3} Gj$=266
$T_j=\sum_{i=1}^{4} y_{ij}$	8		12		-20		$\sum_{j=1}^{3} Tj$=0
$(T_j)^2$	64		144		400		$\sum_{j=1}^{3} T_j^2$=608

$$S_{заг.} = \sum_{j=1}^{p} G_j - \frac{\left(\sum_{j=1}^{p} \mathrm{T}j\right)^2}{pg} = 266 - \frac{0^2}{3 \cdot 4} = 266.$$

$$S_{факт.} = \frac{\sum_{j=1}^{p} \mathrm{T}j^2}{g} - \frac{\left(\sum_{j=1}^{p} \mathrm{T}j\right)^2}{pg} = \frac{608}{4} - \frac{0}{12} = 152.. S^2_{факт.} = \frac{S_{фак}}{p-1} = \frac{152}{3-1} = 76.$$

$$S_{зал..} = 266 - 152 = 114. .S^2_{зал.} = \frac{S_{зал}}{n-p} = \frac{144}{12-3} = \frac{144}{9} = 12,76.$$

$$F_{спост} = \frac{S^2_{фак}}{S^2_{зал}} = \frac{76}{12,67} = 5,998.$$

Враховуючи, що число ступенів вільностей: чисельника k_1 = 2, знаменника k_2 = 9; рівень значущості α = 0,05; з таблиці (додаток 7) знаходимо $F_{кр.}(0,05; 2; 9) = 4,26$.

Так як $F_{спост.} < F_{кр.}$, то нульову гіпотезу про рівність групових середніх відхиляємо. Якщо потрібно порівняти середні попарно, то треба використати критерій Стьюдента.

Неоднакова кількість випробувань на різних рівнях фактору. Нехай число випробувань на різних рівнях фактору різне, а саме: проведено g_1 випробувань на рівні F_1, g_2 випробувань - на рівні F_2, …, g_p випробувань - на рівні $F_{p.}$ У такому випадку загальну суму квадратів відхилень також знаходять за формулою

$$S_{заг.}=\sum_{j=1}^{p} P_j - \frac{1}{n}\left(\sum_{j=1}^{p} R_j\right)^2,$$

де $P_1 = \sum_{i=1}^{g_1} x_{i1}{}^2$, $P_2 = \sum_{i=1}^{g_2} x_{i2}{}^2, \dots, P_{\mathrm{p}} = \sum_{i=1}^{g_1} x_{i\mathrm{p}}{}^2$ - суми квадратів спостережуваних значень ознаки відповідно на рівнях F_1, F_2 , …, F_p; $R_1 = \sum_{j=1}^{g_1} x_{i1}$, $R_2 = \sum_{j=1}^{g_2} x_{i2}, \dots, R_{\mathrm{p}} = \sum_{j=1}^{g_{\mathrm{p}}} x_{i\mathrm{p}}$ – суми спостережуваних значень ознаки відповідно на рівнях F_1, F_2 , …, F_p.

Факторну суму квадратів відхилень знаходять за формулою

$$S_{факт.} = \frac{\sum_{j=1}^{p} R_j{}^2}{g_j} - \frac{\left(\sum_{j=1}^{p} R_j\right)^2}{n}, \tag{4.1.10}$$

де $n=g_1+g_2+\dots+g_p$ - об'єм вибірки.

Якщо для спрощення обчислень вводиться заміна змінних (умовних варіант) $y_{ij} = x_{ij} - C$, де C приблизно рівна загальній середній, то формула для загальної суми квадратів відхилень має вид

$$S_{заг.} = \sum_{j=1}^{p} G_j - \frac{\left(\sum_{j=1}^{p} \mathrm{T}j\right)^2}{pg}, \qquad (4.1.11)$$

де $G_1 = \sum_{i=1}^{g_1} y_{i1}{}^2$, $G_2 = \sum_{i=1}^{g_2} y_{i2}{}^2, \dots, G_{\mathrm{p}} = \sum_{i=1}^{g_1} y_{i\mathrm{p}}{}^2$,

$T_1 = \sum_{j=1}^{g_1} y_{i1}, T_2 = \sum_{j=1}^{g_2} y_{i2}, \dots, T_{\mathrm{p}} = \sum_{j=1}^{g_{\mathrm{p}}} y_{i\mathrm{p}}$,

а формула для факторної суми відхилень має вид

$$S_{\text{факт.}} = \frac{1}{g_j}\sum_{j=1}^{p} T_j{}^2 - \frac{1}{n}((\sum_{j=1}^{p} T_j)^2. \qquad (4.1.12)$$

Решта обчислень виконують, як і у випадку однакового числа випробувань, за формулами:

$S_{зал.} = S_{заг.} - S_{факт.}$,

$S^2{}_{заг.} = \frac{S_{заг}}{n-1}$; $S^2{}_{факт.} = \frac{S_{фак}}{p-1}$; $\mathrm{S}^2{}_{зал..} = \frac{S_{зал}}{n-p}$.

Приклад 4.2. За відповідною вибіркою склали таблицю результатів 22-х випробувань, із яких: 5 значень - для першого рівня фактора F_1 ;6 значень - для другого рівня фактора F_2; 7 значень - для третього рівня фактора F_3; 4 значення - для четвертого рівня фактора F_4. Перевірити нульову гіпотезу про рівність групових середніх. Передбачається, що вибірки взяті з нормальних сукупностей з однаковими дисперсіями.

Розв'язання. Складемо таблицю за вибіркою і знайдемо групове середнє за формулою $\overline{x_j} = \frac{\left(\sum_{i=j}^{qj} x_{ij}\right)}{q}$,

де $q_1 = 5, q_2 = 6, q_3 = 7, q_4 = 4$.

Для спрощення обчислень віднімемо від кожного спостережуваного значення одне й те ж число С=70, приблизно рівне загальній середній. Складемо розрахункову таблицю (табл. 4.1.).

Таблиця 4.1

Номер випробу-вання i	Рівні фактору F_j								Підсумок
	F_1		F_2		F_3		F_4		
	y_{i1}	y^2_{i1}	y_{i2}	y^2_{i2}	y_{i3}	y^2_{i3}	y_{i4}	y^2_{i4}	
1	-3	9	0	0	0	0	10	100	
2	3	9	4	16	-1	1	-1	1	
3	-2	4	9	81	6	36	-1	1	
4	2	4	-5	25	1	1	-10	100	
5	-3	9	2	4	-7	49	-	-	
6	-	-	1	1	-5	25	-	-	
7	-	-	-	-	3	9	-	-	
$G_j=\sum_{i=1}^{4} y_{ij}^2$	-	35	-	127	-	121	-	202	$\sum_{j=1}^{4} Gj$=485
$T_j=\sum_{i=1}^{4} y_{ij}$	-3	-	11	-	-3	-	-2	-	$\sum_{j=1}^{4} Tj$ =3
$(T_j)^2$	9	-	121	-	9	-	4	-	
$\frac{T_j^2}{g_j}$	1,8	-	20,17	-	1,29	-	1	-	$\sum_{j=1}^{4} \frac{T_j^2}{g_j} = 24{,}26$

Використовуючи таблицю, знайдемо загальну і факторну суму квадратів відхилень:

$$S_{\text{заг.}} = \sum_{j=1}^{p} G_j - \frac{\left(\sum_{j=1}^{p} \mathrm{T}j\right)^2}{pg} = 485 - \frac{3^2}{22} = 485 - 0,41 = 484,59.$$

$$S_{\text{факт.}} = \sum_{j=1}^{p} \frac{T_j^2}{g_j} - \frac{\left(\sum_{j=1}^{p} T_j\right)^2}{n} = \frac{9}{5} + \frac{121}{6} + \frac{9}{7} + \frac{4}{4} - \frac{3^2}{22} = 24,26 - 0,41 = 23,84.$$

$$S_{\text{зал.}} = S_{\text{заг.}} - S_{\text{факт}} = 484{,}59 - 23{,}84 = 460{,}75.$$

Знайдемо факторну і залишкову дисперсії.

$$S^2_{\text{факт.}} = \frac{S_{\text{фак}}}{p-1} = \frac{23,84}{3} = 7,95. \; \mathrm{S}^2_{\text{зал.}} = \frac{S_{\text{зал}}}{n-p} = \frac{460,75}{22-4} = 25,6.$$

*Т*ак як $S^2_{\text{факт}} < S^2_{\text{зал.}}$, то звідси слідує справедливість гіпотези про рівність групових середніх і, значить, немає необхідності вдаватися до критерію Фішера–Снедекора.

Приклад 4.3. Проведено 13 випробувань, із них 4 – на першому рівні фактору, 4 – на другому, 3 – на третьому і 2 – на четвертому. Методом дисперсійного аналізу при рівні значущості α=0,05 перевірити нульову гіпотезу про рівність групових середніх.

Номер випробування i	Рівні фактору F_j			
	F_1	F_2	F_3	F_4
1	1,38	1,41	1,32	1,31
2	1,38	1,42	1,33	1,33
3	1,42	1,44	1,34	-
4	1,42	1,45	-	-
$\overline{x}_{гр.j}$	1,4	1,43	1,33	1,32

Розв'язання. Перейдемо до цілих чисел, використовуючи формулу $y_{ij} = 10^2 \cdot x_{ij} - 138$. Отримаємо таблицю значень відносно нової змінної y_{ij}.

Номер випробу-вання i	Рівні фактору F_j								
	F_1		F_2		F_3		F_4		
	y_{i1}	y^2_{i1}	y_{i2}	y^2_{i2}	y_{i3}	y^2_{i3}	y_{i4}	y^2_{i4}	
1	0	0	3	9	-6	36	-7	49	
2	0	0	4	16	-5	25	-5	25	
3	4	16	6	36	-4	16	-	-	
4	4	16	7	49	-	-	-	-	
$Q_j=\sum_{i=1}^{g_j} y_{ij}^2$		32		110		77		74	$\sum_{j=1}^{4} Qj$ =293
$T_j=\sum_{i=1}^{g_j} y_{ij}$	8		20		-15		-12		$\sum_{j=1}^{4} Tj$ =1
T_j^2	64		400		225		144		

Так як число випробувань різне на кожному факторі, то:

$$S_{факт.} = \frac{T_1^2}{g_1} + \frac{T_2^2}{g_2} + \ldots\ldots + \frac{T_p^2}{g_p} - \frac{\sum_{j=1}^{p} T_j^2}{n} = \frac{64}{4} + \frac{400}{4} + \frac{225}{3} + \frac{144}{2} - \frac{1}{13} = 263 - 0,08 = 262,92.$$

$$S^2_{факт.} = \frac{S_{фак}}{p-1} = \frac{262,92}{4-1} = \frac{262,92}{3} = 87,64.$$

$$S\text{заг.} = \sum_{j=1}^{4} Qj - \frac{\left(\sum \text{T}j\right)^2}{n} = 293 - \frac{1^2}{13} = 293 - 0,08 = 292,92.$$

$S_{зал.} = S_{заг.} – S_{факт.} = 292,92 – 262,92 = 30.$

$$\text{S}^2_{зал.} = \frac{S_{зал}}{n-p} = \frac{30}{13-4} = \frac{30}{9} = 3,33.$$

$F_{спост} = \frac{S^2_{фак}}{S^2_{зал}} = \frac{87,64}{3,33} = 26,32$. $k_1 = p-1 = 3$, $k_2 = n-p = 13-4 = 9$.

$F_{кр.}$ (0,05; 3; 9) = 3,86.

Так як $F_{сп.} > F_{кр.}$, то нульову гіпотезу про рівність групових середніх відхиляємо.

4.2. Двофакторний дисперсійній аналіз

Нехай необхідно виявити вплив двох факторів *A* та *B* та їх взаємодію на деяку ознаку *X*. Спостереження проводяться при фіксованих рівнях факторів *A* та *B*. Оскільки для кожного поєднання факторів спостереження проводять *n* разів, то відповідно отримаємо *n* значень ознаки. Дані спостережень зручно представити у вигляді таблиці (табл. 4.2.1), в якій значення ознаки *X* позначено через x_{ijk}, де $i = \overline{1,q}$ (де q – число спостережень фактора *A*); $j = \overline{1,p}$ (*p* – число спостережень фактора *B*); $k = \overline{1,n}$ (*k* – порядковий номер спостережень для кожного поєднання рівнів).

За даними таблиці отримаємо наступні середні:

загальна середня:

$$\bar{x} = \frac{1}{qpn}\sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n} x_{1jk} \qquad (4.2.1)$$

Таблиця 4.2.1

Рівні фактора A	Рівні фактора B				Сума P_i
	B_1	B_2	…	B_p	
A_1	x_{111} x_{112} x_{113} … x_{11n}	x_{121} x_{122} x_{123} … x_{12n}	…	x_{1p1} x_{1p2} x_{1p3} … x_{1pn}	$\sum_{j=1}^{p}\sum_{k=1}^{n} x_{1jk}$
A_2	x_{211} x_{212} x_{213} … x_{21n}	x_{221} x_{222} x_{223} … x_{22n}	…	x_{2p1} x_{2p2} x_{2p3} … x_{2pn}	$\sum_{j=1}^{p}\sum_{k=1}^{n} x_{2jk}$
…	…	…	…	…	…
A_q	x_{q11} x_{q12} x_{q13} … x_{q1n}	x_{q21} x_{q22} x_{q23} … x_{q2n}	…	x_{qp1} x_{qp2} x_{qp3} … x_{qpn}	$\sum_{j=1}^{p}\sum_{k=1}^{n} x_{qjk}$
Сума R_j	$\sum_{i=1}^{q}\sum_{k=1}^{n} x_{i1k}$	$\sum_{i=1}^{q}\sum_{k=1}^{n} x_{i2k}$		$\sum_{i=1}^{q}\sum_{k=1}^{n} x_{ipk}$	$\sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n} x_{1jk}$

середні за рядками:

$$\bar{P_\iota} = \frac{1}{pn} P_i = \frac{1}{pn} \sum_{i=1}^{p} \sum_{k=1}^{n} x_{ijk},\ i = \overline{1,q} \tag{4.2.2}$$

середні за стовбцями:

$$\bar{R_J} = \frac{1}{pn} R_j = \frac{1}{pn} \sum_{i=1}^{q} \sum_{k=1}^{n} x_{ijk},\ j = \overline{1,p}\ ; \tag{4.2.3}$$

середні для кожного окремого блока таблиці

$$\overline{Q_{\iota J}} = \frac{1}{n} Q_{ij} = \frac{1}{n} \sum_{k=1}^{n} x_{ilk}. \tag{4.2.4}$$

Суми квадратів відхилень від загальної середньої $\sum_{i=1}^{q} \sum_{j=1}^{p} \sum_{k=1}^{n} (x_{ijk} - \bar{x}\ )^2$ розкладемо на складові. Для цього представимо $x_{ijk} - \bar{x}$ в еквівалентній формі

$$x_{ijk} - \bar{x} = \left(\bar{R}_J - \bar{x}\right) + \overline{(R_\iota} - \bar{x}) + \left(\overline{Q_{\iota J}} - \bar{R}_J - \bar{P}_i + \bar{x}\right) + (x_{ijk} - \overline{Q_{\iota J}}).$$

Оскільки все перехресні добутки, при піднесенні правої частини останнього виразу в квадрат, рівні нулю, то в результаті отримаємо:

$$\sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}(x_{ijk} - \bar{x})^2 = \sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}(\bar{R}_J - \bar{x})^2 + \sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}(\bar{P}_\iota - \bar{x})^2 +$$

$$+\sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}(\overline{Q_{\iota J}} - \bar{R}_J - \bar{P}_\iota + \bar{x})^2 + \sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}(x_{ijk} - \overline{Q_{\iota J}})^2$$

або в скороченому вигляді

$$S = S_1 + S_2 + S_3 + S_4. \qquad (4.2.5)$$

У (4.2.5) чотири складові: сума квадратів, що пов'язана з впливом фактора *B*; сума квадратів, що пов'язана з впливом фактора *A*; їх взаємодія, і складова, яка характеризує залишкову суму квадратів (суму квадратів всередині кожного блока таблиці).

Загальне число ступенів вільностей, очевидно, дорівнює *N*–1, де *N* – загальне число спостережень (*N* = *qpn*). Число ступенів вільностей дорівнює: між стовбцями *p*-1, між рядками *q*-1; для взаємодії (*p*-1)(*q*-1), всередині клітинок *pq(n*-1*)*=N–*pq*. Для перевірки знайдемо суму *p*-1+*q*-1+*(p*-1*)(q*-1*)*+ *N-pq=N-1.*

Схема дисперсійного аналізу, що основана на отриманих вище даних, представлена в таблиці (табл. 4.2.2).

Оскільки при проведенні аналізу інтерес являє вплив кожного фактора порізно та вплив їх взаємодії, то знаходимо відповідно три значення:

$$F_B = S_1^2/S_4^2; \qquad F_{\mathrm{A}} = S_2^2/S_4^2; \qquad F_{\mathrm{A}B} = S_3^2/S_4^2. \qquad (4.2.6)$$

Таблиця 4.2.2

Фактор	Сума квадратів	Число ступенів вільностей	Оцінка дисперсії
В	$S_1 = qn\sum_{j=1}^{p}(\overline{R}_j - \bar{x})^2$	p-1 =k_1	$S_1^2 = \frac{S_1}{p-1}$
А	$S_2 = pn\sum_{i=1}^{q}(\overline{P}_i - \bar{x})^2$	q-1 =k_2	$S_2^2 = \frac{S_2}{q-1}$
А×В	$S_3 = n\sum_{i=1}^{q}\sum_{j=1}^{p}(\overline{Q_{ij}} + \overline{R}_j - \overline{P}_i + \bar{x})^2$	(p-1)×(q-1)=k_3	$S_3^2 = \frac{S_3}{(p-1)(q-1)}$
Залиш-ковий	$S_4 = \sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}(x_{ijk} - \overline{Q_{ij}})^2$	N-pq= k_4	$S_4^2 = \frac{S_4}{N-pq}$
Сума	$S_4 = \sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}(x_{ijk} - \bar{x})^2$	N-1	$S^2 = \frac{S}{N-1}$

Необхідні для аналізу суми квадратів відхилень (4.2.5) можемо отримати також за наступними формулами:

$$S_1 = \frac{1}{np}\sum_{j=1}^{p}R_j^2 - \frac{G^2}{N}; \qquad (4.2.7)$$

$$S_2 = \frac{1}{np}\sum_{i=1}^{q}P_i^2 - \frac{G^2}{N}; \qquad (4.2.8)$$

$$S_3 = \frac{1}{n}\sum_{i=1}^{q}\sum_{j=1}^{p}Q_{ij}^2 - \frac{1}{nq}\sum_{j=1}^{p}R_j^2 - \frac{1}{np}\sum_{i=1}^{q}P_i^2 + \frac{G^2}{N}; \qquad (4.2.9)$$

$$S_4 = \sum_{i=1}^{q}\sum_{j=1}^{p}\left(\sum_{k=1}^{n} x_{ijk}^2 - \frac{Q_{ij}^2}{n}\right), \qquad (4.2.10)$$

де

$$G = \sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n} x_{ijk}\ ;\ Q_{ij} = \sum_{k=1}^{n} x_{ijk}\ ;\ F_{ij} = \sum_{k=1}^{n} x_{jk}^2 - \frac{Q_{ij}^2}{n}. \qquad (4.2.11)$$

При рівні значущості α визначаємо критичні точки:

$F_{\text{кр}}(\alpha; k_1; k_4) = x_{\text{кр}}^{B}$; $F_{\text{кр}}(\alpha; k_2; k_4) = x_{\text{кр}}^{\text{A}}$; $F_{\text{кр}}(\alpha; k_3; k_4) = x_{\text{кр}}^{\text{A}B}$.

Якщо:

1) $F_{\text{B}} > x_{\text{кр}}^{B}$, то нульова гіпотеза про відсутність впливу фактора В відхиляється;

2) $F_{\text{A}} > x_{\text{кр}}^{\text{A}}$, то нульова гіпотеза про відсутність впливу фактора А відхиляється;

3) $F_{\text{AB}} > x_{\text{кр}}^{\text{A}B}$, то нульова гіпотеза про відсутність спільного впливу факторів *А* і *В* відхиляється.

Приклад 4.2.1. При рівні значущості *α*=0,05 перевірити, чи існує вплив факторів *А* і *В*, а також їх спільного впливу на ознаку *Х,* для результатів випробувань, що наведені в таблиці.

Рівень фактору *А*	Рівень фактору *В*		
	В₁	*В₂*	*В₃*
А₁	10; 8; 7; 10	8; 12; 14; 12	15; 8; 10; 10
А₂	12; 8; 8; 7	12; 13; 11;14	13; 15; 12; 10

Розв'язання. Використовуючи формули (4.2.7) ÷ (4.2.11), складемо таблицю та обчислимо суми квадратів відхилень $S_i(i = \overline{1,4})$ (табл. 4.2.3).

Таблиця 4.2.3

B / A	B_1		B_2		B_3		$\sum_{j=1}^{3} Q_{ij}^2$	P_i	P_i^2
	x_{i1k}	x_{i1k}^2	x_{i2k}	x_{i2k}^2	x_{i3k}	x_{i3k}^2			
A_1	10	100	8	64	15	225			
	8	64	12	144	8	64			
	7	49	14	196	10	100			
	10	100	12	144	10	100			
Q_{1j}	35	-	46	-	43	-	-	124	15376
Q_{1j}^2	1225	-	2116	-	1849	-	5190	-	-
$\sum_{k=1}^{4} x_{1jk}^2$	-	313	-	548		489	-	-	-
F_{1j}	6,75		19		26,75		$\sum_{j=1}^{3} F_{2j} = 52{,}5$		
A_2	12	144	12	144	13	169			
	8	64	13	169	15	225			
	8	64	11	121	12	144			
	7	49	14	196	10	100			
Q_{2j}	35	-	50	-	50	-	-	135	18225
Q_{2j}^2	1225	-	2500	-	2500	-	6225	-	-
$\sum_{k=1}^{4} x_{2jk}^2$	-	321	-	630		638	-	-	-
F_{2j}	14,75		5		26,75		$\sum_{j=1}^{3} F_{2j} = 32{,}75$		
R_j	70	-	96	-	93	-	-	G=259	$\sum_{i=1}^{2} P_i^2 = 33601$
R_j^2	4900	-	9216	-	8649	-	$\sum_{j=1}^{3} R_j^2 = 22765$		

З таблиці маємо:

$$S_1 = \frac{1}{8}\sum_{j=1}^{3} R_j^2 - \frac{G^2}{24} = \frac{1}{8} \cdot 22765 - \frac{259^2}{24} \approx 50{,}58;$$

$$S_2=\frac{1}{12}\sum_{i=1}^{2}P_i^2-\frac{G^2}{24}=\frac{1}{12}\cdot 33601-\frac{259^2}{24}\approx 5{,}04;$$

$$S_3=\frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{3}Q_{ij}^2-\frac{1}{8}\sum_{j=1}^{3}R_j^3-\frac{1}{12}\sum_{i=1}^{2}P_i^2+\frac{G^2}{24}=$$

$$=\frac{1}{4}(5190+6225)-2845{,}625-2800{,}083+2795{,}0417\approx 3{,}083;$$

$$S_4=\sum_{i=1}^{2}\sum_{j=1}^{3}\left(\sum_{k=1}^{4}x_{ijk}^2-Q_{ij}^2/4\right)=52{,}5+32{,}75=85{,}25.$$

Дані дисперсійного аналізу наведемо в таблиці 4.2.4.

Таблиця 4.2.4

Джерело варіації	S_i	Число ступенів вільностей	Оцінка дисперсії
Фактор *B*	50,58	2	25,29
Фактор *A*	5,04	1	5,04
A×*B*	3,083	2	1,54
Залишкова варіація	85,25	18	4,74
∑	143,95	23	-

Визначимо спостережувані значення критерію:

$F^{B}_{\text{спост.}}=25{,}29/4{,}74\approx 5{,}34;$ $F^{\text{A}}_{\text{спост.}}=\dfrac{5{,}04}{4{,}74}\approx 1{,}06;$

$F^{\text{A}B}_{\text{спост.}}=1{,}54/4{,}74\approx 0{,}32.$

Порівняємо їх з відповідними критичними значеннями, знайденими з таблиці (додаток 7), $F_{кр}$(0,05;2;18) = $x^{B}_{\text{кр}}$ = 3,55; $F_{кр}$(0,05;1;18) = $x^{\text{A}}_{\text{кр}}=4{,}41$;$F_{кр}$(0,05;2;18) = $x^{\text{AB}}_{\text{кр}}$=3,55.

Так як $F^{\text{B}}_{\text{спост.}}>x^{B}_{\text{кр}}$, то нульова гіпотеза про відсутність впливу фактора В відхиляється і слід зробити висновок про значимість впливу цього фактора. Так як $F^{\text{A}}_{\text{спост.}}>x^{\text{A}}_{\text{кр}}$ і $F^{AB}_{\text{спост.}}>x^{AB}_{\text{кр}}$, то немає підстав для відхилення відповідних нульових гіпотез.

Питання для самоперевірки

1. Що таке кореляційний та функціональний зв'язки?

2. У чому суть дисперсійного аналізу?

3. Що характеризують коефіцієнти кореляції та лінійної регресії?

4. В яких межах знаходяться коефіцієнти кореляції та лінійної регресії?

5. Що таке лінійна залежність та її рівняння?

6. Як визначаються коефіцієнти a і b лінійної залежності y = ax+b?

7. Що означають коефіцієнти a і b?

8. Які знаки мають a і b для зростаючої та спадаючої залежностей?

9. Наведіть класифікацію моделей дисперсійного аналізу за числом факторів і за метою дослідження.

10. Що називається груповою та загальною середньою? Наведіть формули для їх обчислення.

11. Вивести основну тотожність для суми квадратів відхилень спостережуваних значень x_{ij} від загальної середньої $\overline{x}$.

12. Вивести формули для розрахунку $S_{\text{заг.}}$ і $S_{\text{факт.}}$ через P_j і R_j

13. Вивести формули для розрахунку $S_{\text{заг.}}$і $S_{\text{факт.}}$ через Q_jі T_j.

14. Наведіть формули для обчислення незміщених оцінок трьох дисперсій $S^2_{\text{заг.}}$; $S^2_{\text{факт.}}$, $S^2_{\text{зал.}}$.

15. Наведіть схему для перевірки нульової гіпотези про рівність групових середніх.

16. Наведіть основні та спрощені формули для розрахунку $S^2_{\text{заг.}}$ і $S^2_{\text{факт.}}$ при неоднаковій кількості випробувань на різних рівнях.

17. Запишіть теоретично-вірогідну модель для двофакторного дисперсійного аналізу.

18. Наведіть формули для обчислення середніх при двофакторному дисперсійному аналізі.

19. Дайте характеристику чотирьом складовим розкладання суми квадратів відхилень від загальної середньої.

20. Наведіть формули для розрахунку S_1, S_2, S_3 та S_4.

Задачі та вправи

4.1. Виконано по 4 випробування на кожному із трьох рівнів фактору *F.* Методом дисперсійного аналізу при рівні значущості α=0,05 перевірити нульову гіпотезу про рівність групових середніх. Припускається, що вибірки взяті із нормальних сукупностей з однаковими дисперсіями. Результати випробувань наведені в таблиці.

Номер випробування *i*	Рівні фактору F_j		
	F_1	F_2	F_3
1	38	20	21
2	36	24	22
3	35	26	31
4	31	30	34
$\bar{x}_{гр.j}$	35	25	27

4.2. Виконано по п'ять випробувань на кожному із чотирьох рівнів фактору *F*. Методом дисперсійного аналізу при рівні значущості α=0,05 перевірити нульову гіпотезу про рівність групових середніх. Припускається, що вибірки взяті із нормальних сукупностей з однаковими дисперсіями. Результати випробувань наведені в таблиці.

Номер випробування *i*	Рівні фактору F_j			
	F1	F2	F3	F4
1	36	56	52	39
2	47	61	57	57
3	50	64	59	63
4	58	66	58	61
5	67	66	79	65
$\bar{x}_{гр.j}$	51,6	62,6	61,0	57,0

4.3. Проведені досліди по відбору сортів озимої пшениці дали такі результати:

Варіанти (сорти)	Повтори (ц\га)			
	1	2	3	4
1	48	47	52	43
2	54	50	45	47
3	47	42	43	48
4	48	47	46	44
5	42	40	43	44

Методом дисперсійного аналізу вияснити чи існує різниця між середніми за рівнем значущості α =0,05.

4.4. На текстильну фабрику привезли чотири партії сировини. З кожної партії взяли по 5 зразків та провели випробування на величину розривного зусилля. Результати випробувань подані в таблиці.

Номер партії	Розрив зусилля (повтори)				
	1	2	3	4	5
1	200	140	170	145	165
2	190	150	210	150	150
3	230	190	200	190	200
4	150	170	150	170	180

Треба вияснити, чи суттєвим є вплив партії сировини на величину розривного зусилля для рівня значущості α =0,01.

4.5. Проведені досліди по відбору типу суміші добрив на врожай гречки дали такі результати:

Варіант суміші добрив	Повтори (ц\га)			
	1	2	3	4
1	7,1	6,9	7,3	6,3
2	6,8	7,0	6,9	6,9
3	7,0	7,1	7,2	7,2

Провести дисперсійний аналіз результатів досліду на рівні значущості α =0,05.

4.6. Проведено 14 спостережень, із них 5 – на першому рівні фактору, 3 – на другому, 2 – на третьому, 3 – на четвертому і 1 – на п'ятому. Методом дисперсійного аналізу при рівні значущості α=0,05 перевірити нульову гіпотезу про рівність групових середніх. Вибірки взяті із нормальних сукупностей з однаковими дисперсіями. Результати випробувань наведені в таблиці.

Номер випробування *i*	Рівні фактору F_j				
	F_1	F_2	F_3	F_4	F_5
1	7,3	5,4	6,4	7,9	7,1
2	7,6	7,1	8,1	9,5	-
3	8,3	7,4	-	9,6	-
4	8,3	-	-	-	-
5	8,4	-	-	-	-
$\overline{x}_{гр.j}$	7,98	6,63	7,25	9,0	7,1

4.6. Проведено 13 спостережень, із них 4 – на першому рівні фактору, 6 – на другому, 3 – на третьому. Методом дисперсійного аналізу при рівні значущості α =0,01 перевірити нульову гіпотезу про рівність групових середніх. Вибірки взяті із нормальних сукупностей із однаковими дисперсіями. Результати випробувань наведені в таблиці.

Номер випробування i	Рівні фактору F_j		
	F_1	F_2	F_3
1	37	60	69
2	47	86	100
3	40	67	98
4	60	92	-
5	-	95	-
6	-	98	-
$\bar{x}_{гр.j}$	56	83	89

4.8. Проведено по сім випробувань на кожному з чотирьох рівнів фактору. Методом дисперсійного аналізу при рівні значущості α=0,05 перевірити нульову гіпотезу про рівність групових середніх. Припускається, що вибірки взяті із нормальних сукупностей із однаковими дисперсіями. Результати випробувань наведені в таблиці.

Номер випробування i	Рівні фактору F_j			
	F_1	F_2	F_3	F_4
1	51	52	56	54
2	59	58	56	58
3	53	66	58	62
4	59	69	58	64
5	63	70	70	66
6	69	72	74	67
7	72	74	78	69
$\bar{x}_{гр.j}$	60,9	65,9	64,3	62,9

4.9. Проведено по чотири випробування на кожному з трьох рівнів фактору. Методом дисперсійного аналізу при рівні значущості α= 0,05 перевірити нульову гіпотезу про рівність групових середніх. Припускається, що вибірки взяті із нормальних сукупностей із однаковими дисперсіями. Результати випробувань наведені в таблиці.

Номер випробування i	Рівні фактору F_j		
	F_1	F_2	F_3
1	27	24	22
2	23	20	21
3	29	26	36
4	29	30	37
$\bar{x}_{гр.j}$	27	25	29

4.10**.** Проведено 14 випробувань, із них 7 – на першому рівні фактору, 3 – на другому, 4 – на третьому. Методом дисперсійного аналізу при рівні значущості α = 0,01 перевірити нульову гіпотезу про рівність групових середніх. Вибірки взяті з нормальних сукупностей із однаковими дисперсіями. Результати випробувань наведені в таблиці.

Номер випробування i	Рівні фактору F_j		
	F_1	F_2	F_3
1	30,56	43,44	31,36
2	32,66	47,51	36,20
3	34,78	53,80	36,38
4	35,50	-	42,20
5	36,63	-	-
6	40,20	-	-
7	42,28	-	-
$\bar{x}_{гр.j}$	36,09	48,25	36,54

4.11. Проведено 26 випробувань, із них 7 – на першому рівні фактору, 5 – на другому, 8 – на третьому, 6 – на четвертому. Методом дисперсійного аналізу при рівні значущості α = 0,05 перевірити нульову гіпотезу про рівність групових середніх. Припускається, що вибірки дістали із генеральних сукупностей із однаковими дисперсіями.

Номер випробування *i*	Рівні фактору F_j			
	F_1	F_2	F_3	F_4
1	1600	1580	1460	1510
2	1610	1640	1550	1520
3	1650	1640	1600	1530
4	1680	1700	1620	1570
5	1700	1750	1640	1600
6	1700	-	1660	1680
7	1800	-	1740	-
8	-	-	1820	-
$\bar{x}_{гр.j}$	1677	1662	1638	1568

4.12. Виконано по чотири випробування на кожному з чотирьох рівнів фактору *F*. Методом дисперсійного аналізу при рівні значущості α =0,05 перевірити нульову гіпотезу про рівність групових середніх. Припускається, що вибірки взяті із нормальних сукупностей із однаковими дисперсіями. Результати випробувань наведені в таблиці.

Номер випробування *i*	Рівні фактору F_j			
	F_1	F_2	F_3	F_4
1	140	150	52	150
2	144	149	57	155
3	142	152	59	154
4	145	150	58	152
$\bar{x}_{гр.j}$	142,75	150,25	147,5	152,75

4.13. Зроблено по вісім випробувань на кожному з шести рівнів фактора. Методом дисперсійного аналізу при рівні значущості α = 0,01 перевірити нульову гіпотезу про рівність групових середніх. Передбачається, що вибірки взяті з нормальних сукупностей із однаковими дисперсіями. Результати випробувань наведені в таблиці.

Номер випробування *i*	Рівні фактора					
	F_1	F_2	F_3	F_4	F_5	F_6
1	100	92	74	68	64	69
2	101	102	87	80	83	71
3	126	104	88	83	83	80
4	128	115	93	87	84	80
5	133	119	94	96	90	81
6	141	122	101	97	96	82
7	147	128	102	106	101	86
8	148	146	105	127	111	99
$\bar{x}_{\text{гр}.j}$	128	116	93	93	89	81

4.14. Зроблено по чотири випробування на кожному з трьох рівнів фактора F. Методом дисперсійного аналізу при рівні значущостіα = 0,05 перевірити нульову гіпотезу про рівність групових середніх. Передбачається, що вибірки взяті з нормальних сукупностей із однаковими дисперсіями. Результати випробувань наведені в таблиці.

Номер випробування *i*	Рівні фактора		
	F_1	F_2	F_3
1	35	30	21
2	32	24	22
3	31	26	34
4	30	20	31
$\bar{x}_{\text{гр}\ j}$	32	25	27

4.15. Зроблено по сім випробувань на кожному з чотирьох рівнів фактора. Методом дисперсійного аналізу при рівні значущості α=0,05 перевірити нульову гіпотезу про рівність групових середніх. Передбачається, що вибірки взяті з нормальних сукупностей із однаковими дисперсіями. Результати випробувань наведені в таблиці.

Номер випробування i	Рівні фактора			
	F_1	F_2	F_3	F_1
1 2 3 4 5 6 7	51 59 53 59 63 69 72 51	52 58 66 69 70 72 74 52	56 56 58 58 70 74 78 56	58 62 64 66 67 69 54
$\bar{x}_{гр\ j}$	60,9	65,9	64,3	62,9

4.16. Зроблено 14 випробувань, із них 5 - на першому рівні фактора, 3 - на другому, 2 - на третьому, 3 - на четвертому і 1 - на п'ятому. Методом дисперсійного аналізу при рівні значущості α=0,05 перевірити нульову гіпотезу про рівність групових середніх. Передбачається, що вибірки взяті з нормальних сукупностей із однаковими дисперсіями. Результати випробувань наведені в таблиці.

Номер випробування i	Рівні фактора				
	F_1	F_2	F_3	F_4	F_5
1	7,3	5,4	6,4	7,9	7,1
2	7,6	7,1	8,1	9,5	-
3	8,3	7,4	-	9,6	-
4	8,3	-	-	-	-
5	8,4	-	-	-	-
$\bar{x}_{гр\ j}$	7,98	6,63	7,25	9,0	7,1

4.17. Зроблено 13 випробувань, із них 4 - на першому рівні фактора, 6 - на другому і 3 - на третьому. Методом дисперсійного аналізу при рівні значущості α=0,01 перевірити нульову гіпотезу про рівність групових середніх. Передбачається, що вибірки взяті з нормальних сукупностей із однаковими дисперсіями. Результативипробувань наведено в таблиці.

Номер випробування i	Рівні фактора		
	F_1	F_2	F_3
1	37	60	69
2	47	86	100
3	40	67	98
4	60	92	-
5	-	95	-
6	-	98	-
$\bar{x}_{гр\ j}$	46	83	89

4.18. Зроблено 14 випробувань, із них 7 - на першому рівні фактора, 3 - на другому і 4 - на третьому. Методом дисперсійного аналізу при рівні значущості α = 0,01 перевірити нульову гіпотезу про рівність груповихсередніх. Передбачається, що вибірки взяті знормальних сукупностей із однаковими дисперсіями.Результати випробувань наведені в таблиці.

Номер випробування i	Рівні фактора		
	F_1	F_2	F_3
1	30,56	43,44	31,36
2	32,66	47,51	36,20
3	34,78	53,80	36,38
4	35,50	-	42,20
5	36,63	-	-
6	40,20	-	-
7	42,28	-	-
$\bar{x}_{\text{гр}\ j}$	36,09	48,25	36,54

4.19. Зроблено 26 випробувань, з них 7 - на першому рівні фактора, 5 - на другому, 8 - на третьому і 6 – на четвертому. Методом дисперсійного аналізу при рівні значущості α=0,05 перевірити гіпотезу про рівність групових середніх. Передбачається, що вибірки взяті з нормальних генеральних сукупностей із однаковими дисперсіями. Результати випробувань наведені в таблиці.

Номер випробування i	Рівні фактора			
	F_1	F_2	F_3	F_4
1	1600	1580	1460	1510
2	1610	1640	1550	1520
3	1650	1640	1600	1530
4	1680	1700	1620	1570
5	1700	1750	1640	1600
6	1700	-	1660	1680
7	1800	-	1740	-
8	-	-	1820	-
$\bar{x}_{\text{гр}\ j}$	1677	1662	1638	1568

4.20. За відповідною вибіркою скласти таблицю результатів 22-х випробувань, із яких 5 значень для першого рівня фактора F_1 вибираються з номерами 1÷5; 6 значень для другого рівня фактора F_2 вибираються з номерами 6÷ 11; 7 значень для третього рівня фактора F_3вибираються з номерами 12÷18; 4 значення для четвертого рівня фактора F_4 вибираються з номерами 12÷22. Перевірити нульову гіпотезу про рівність групових середніх при α=0,05, якщо вибірки взяті з нормальних сукупностей із однаковими дисперсіями.

а) 93, 88, 77, 92, 92, 103, 85, 90, 83, 86, 104, 102, 85, 91, 87, 102, 94, 98, 85, 82, 94, 86;

б) 72, 95, 96, 103, 89, 72, 105, 85, 85, 91, 101, 82, 80, 95, 91, 86, 94, 72, 85, 85, 80, 82.

РОЗДІЛ 5
РЕГРЕСІЙНИЙ І КОРЕЛЯЦІЙНИЙ АНАЛІЗ

Доволі часто приходиться мати справу з більш складною залежністю, ніж функціональна. Такого роду залежності відносяться до *кореляційних* залежностей. Учення про кореляцію (від латинського correlatio – відповідність, взаємозв'язок) і регресію (від латинського regression – рух назад) широко використовується при аналізі зв'язків різних явищ. Так, в економіці вивчається залежність об'ємів виробництва від ряду різних факторів: розмір основних фондів, рівень механізації, потреби населення та інші. Інтенсивно використовується регресійний аналіз при вивченні залежності урожайності визначеної сільськогосподарської культури від природних і економічних факторів, які впливають на неї; при вивченні зв’язку між товщиною снігового покриву взимку і об’ємом стоків наступної повені тощо. Широко використовуються методи регресійного і кореляційного аналізу в психології, соціології, педагогіці та ряду інших областей науки та практики. Через складність процесів, особливо біологічних, залежність між двома або більшою кількістю величин (ознак) задати аналітично, тобто у вигляді формули, у більшості випадків неможливо. У таких випадках кореляційні зв'язки між величинами задаються табличними даними на основі проведених дослідів.

Математична обробка кореляційних зв'язків між дослідними величинами проводиться методами математичної статистики, а саме методами теорії регресії, яка дає можливість прогнозувати характер зміни однієї випадкової величини від інших.

Поняття кореляції та регресії з’явилося всередині XIX ст. завдяки науковим працям Ф. Гальтона і К. Пірсона.

5.1. Функціональна, статистична і кореляційна залежність

Існують два типи зв'язків (залежностей) між величинами: функціональний та кореляційний.

Функціональною залежністю називають залежність між ознаками, при якій кожному значенню аргумента *Х* (незалежної змінної) відповідає строго визначене значення функції *У* (залежної змінної). Такі зв'язки зустрічаються в математиці, фізиці, астрономії та інших науках. Наприклад, площа круга ($S = \pi r^2$) і довжина кола ($\mathrm{C} = 2\,\pi r$) повністю визначаються тільки радіусом.

Але частіше має місце залежність, коли кожному фіксованому значенню однієї змінної відповідає не одне, а множина значень іншої змінної. Це пояснюється тим, що на залежну змінну впливає ряд не контрольованих або неврахованих факторів, а також множина випадкових факторів. У такому випадку залежна змінна *У* є випадковою величиною. Змінна *Х* може бути як детермінованою (тобто, мати визначені значення), так і випадковою величиною. Такий зв'язок між ознаками отримав назву *статистичний (*або *стохастичний, ймовірнісний*). Наприклад, відомо, урожайність залежить від кількості внесених добрив, але на неї впливають також інші фактори (якість ґрунту, якість насіння, погодні умови, обробка грунту тощо).

Допустимо, що існує стохастична залежність випадкової змінної *У* від *Х.* При *Х*=*х* змінна *У,* у силу стохастичної залежності від *Х,* може приймати довільне значення з деякої множини. Середнє цієї множини називають *груповим* генеральним середнім змінної *У* при *Х*=*х* або *умовним математичним сподіванням* випадкової величини *У*, яке обчислене при умові, що *Х*= *х.* Ця величина позначається *М(У|Х = х).* В окремому випадку, коли стохастична залежність випадкової змінної *У* від *X* проявляється в тому, що при зміні *х* змінюється умовне математичне сподівання *М(У|X= х)*, то кажуть, що має місце *кореляційна залежність* величини *У* від *X*. Якщо ж *М(У|X= х)* залишаються незмінними, то кореляційна залежність величини *У* від *X* відсутня.

Якщо існує стохастична залежність випадкової змінної *X* від *У,* то, аналогічно попередньому, можна ввести поняття умовного математичного сподівання *М(X|У= у)* і кореляційної залежності величини *X* від *У.*

Розрізняють *прямолінійний* і *криволінійний*, *прямий* і *зворотний*, *простий* (визначення взаємозв'язків між двома ознаками) і *множинний* (визначення взаємозв'язків між трьома і більшою кількістю ознак) кореляційні зв'язки. Такі зв'язки виражаються через математичні рівняння.

За допомогою методу кореляційного аналізу вирішуються дві основні задачі:

1) визначення форми та параметрів рівняння зв'язку;

2) визначення тісноти зв'язку.

Перша задача розв'язується знаходженням рівняння зв'язку та визначенням його параметрів. Друга – за допомогою розрахунку різних показників тісноти зв'язку (коефіцієнта кореляції, індексу кореляції та інших).

Схематично кореляційний аналіз можна розділити на п’ять етапів:

1) постановка задачі, встановлення наявності зв'язку між досліджуваними ознаками;

2) відбір найбільш суттєвих факторів для аналізу;

3) визначення характеру зв'язку, його напрямку і форми, підбір математичного рівняння для вираження існуючих зв'язків;

4) розрахунок числових характеристик кореляційного зв'язку (визначення параметрів рівняння і показників тісноти зв'язку);

5) статистична оцінка вибіркових показників зв'язку та аналіз отриманих результатів.

5.2. Лінійна кореляційна залежність і прямі лінії регресії

Нехай *Х* та *У* – випадкові величини, зв'язок між якими треба вивчити. У результаті *n* випробувань дістали *n* пар значень цих величин (x_1, y_1), (x_2, y_2), …, (x_n, y_n). Вибір того чи іншого рівняння для вивчення зв'язку між ознаками є найбільш складним і відповідальним моментом у кореляційного зв'язку. При визначенні взаємозв'язків між двома ознаками (парна кореляція) математичне рівняння зв'язку може бути встановлене за допомогою побудови графіків, складання кореляційних таблиць, перегляду різних функцій тощо.

В економічних дослідженнях часто розглядається лінійна форма зв'язку, яка виражається рівнянням прямої лінії $\overline{y}_x = ax + b$ ($\overline{x}_y = a'y + b'$), де $\overline{y}_x$ ($\overline{x}_y$) - вирівняні значення результативної ознаки (залежна змінна), *x (y)* – значення факторної ознаки (незалежна змінна), *a (a') – коефіцієнт регресії*, який показує середню зміну результативної ознаки*y* при зміні факторної ознаки *x* на одиницю. Якщо *a>0 (a'>0),* то зв'язок *прямий*, якщо *a<0 (a'<0),* то зв'язок *зворотний*, якщо *a=0 (a'=0),*то зв'язок відсутній. Рівняння такого типу називається *рівнянням регресії У на Х (Х на У) або кореляційним рівнянням.* Основним його завданням є *встановлення кількісного взаємозв'язку між ознаками*.

Параметри рівняння *a* і *b* визначають за допомогою *методу найменших квадратів* (МНК), розробка якого належить К. Гауссу та П. Лапласу. Цей метод дає можливість знайти таку теоретичну лінію регресії, яка порівняно з іншими теоретичними лініями регресії проходить найближче до точок кореляційного поля, що зображають фактичні дані, тобто дає найменшу суму квадратів відхилень фактичних значень результативної ознаки *у* від теоретичних значень $\overline{y_x}$.

Розглянемо функцію $S(a,b)=\sum_{i=1}^{n}(y-\overline{y_x})^2=\sum_{i=1}^{n}(y-ax-b)^2.$

Необхідною умовою існування мінімуму цієї функції є рівність нулю частинних похідних за невідомими параметрами *а* та *b*. Прирівнюючи частинні похідні S_a' і S_b' до нуля, отримаємо систему рівнянь для визначення *а* та *b*:

$$\begin{cases} S_a' = \sum_{i=1}^{n} 2(y-ax-b)(-1)=0; \\ S_b' = \sum_{i=1}^{n} 2(y-ax-b)(-x)=0; \end{cases}$$

$$\begin{cases} a\sum_{i=1}^{n}x_i^2+b\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}x_iy_i; \\ a\sum_{i=1}^{n}x_i+bn=\sum_{i=1}^{n}y_i. \end{cases} \qquad (5.2.1)$$

Знайдемо шукані параметри *а* та *b* як розв'язок цієї системи:

$$a=\frac{n\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n\sum_{i=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2};\ b=\frac{\sum_{i=1}^{n}x_i^2\sum_{i=1}^{n}y_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}x_iy_i}{n\sum_{i=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2}. \qquad (5.2.2)$$

Примітка. Знайти розв’язок системи (5.2.1) можна, застосувавши будь-який із методів розв’язування системи лінійних рівнянь (підстановки, порівняння, гауссів і т.п.).

Якщо потрібно за результатами спостережень отримати лінійне рівняння регресії *Х* на *У,* то у рівнянні регресії $\overline{y_x}$ = *ах +b* треба поміняти місцями змінні *х* та *у*. При цьому отримаємо рівняння $\overline{x_y}$ =*а*’*у*+*b*’, де *а*’ та *b*’ обчислюються за формулами:

$$a' = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2};$$

(5.2.3)

$$b' = \frac{\sum_{i=1}^{n} y_i^2 \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}.$$

Відзначимо, що регресійні прямі $\overline{y}_x = ax + b$ і $\overline{x}_y = a' y + b'$ – різні. Перша пряма отримується в результаті розв'язування задачі про мінімізацію суми квадратів відхилень по *вертикалі*, а друга – при розв'язуванні задачі на мінімізацію суми квадратів відхилень по *горизонталі*. Але обидві прямі лінії регресії *X* на *Y* і *X* на *Y* проходять через одну точку з координатами $(\bar{x};\ \bar{y})$.

Приклад 5.2.1. Знайти рівняння прямих ліній регресії Y на X та X на *Y* за даними п'яти спостережень:

x_i	1,00	1,50	3,00	4,50	5,00
y_i	1,25	1,40	1,50	1,75	2,25

Розв'язання. Складемо наступну розрахункову таблицю.

x_i	1,00	1,5	3,00	4,50	5,00	15,00
y_i	1,25	1,4	1,50	1,75	2,25	8,15
x_i^2	1,00	2,25	9,00	20,25	25,00	57,50
y_i^2	1,5625	1,96	2,25	3,0625	5,0625	13,8975
$x_i y_i$	1,25	2,10	4,50	7,875	11,250	36,9875

$\sum_{i=1}^{5} x_i = 15$, $\sum_{i=1}^{5} y_i = 8,15;$ $\sum_{i=1}^{5} x_i y_i = 26,9875$, $\sum_{i=1}^{5} x_i^2 = 57,5;$ $\sum_{i=1}^{5} y_i^2 = 13,8975.$

Рівняння прямої лінії регресії *Y* на *X* має вид $\overline{y_x} = ax + b,$

де $a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{5 \cdot 26,975 - 15 \cdot 8,15}{5 \cdot 57,5 - 15^2} = 0,202$;

$$b = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}. = \frac{57,5 \cdot 8,15 - 15 \cdot 26,975}{62,5} = 1,024.$$

$$\overline{y_x} = 0,202x + 1,024. \qquad (*)$$

Рівняння прямої лінії регресії *X* на *Y* має вид $\overline{x_y} = a' y + b'$, де *a*' та *b*' обчислюються за аналогічними формулами:

$$a' = \frac{5 \cdot 26,975 - 15 \cdot 8,15}{5 \cdot 13,8975 - (8,15)^2} = \frac{12,625}{13,065} = 0,97;;$$

$$b' = \frac{13,8975 \cdot 15 - 8,15 \cdot 26,9875}{5 \cdot 13,8975 - (8,15)^2} = \frac{11,94}{13,065} = 0,91.$$

$$\overline{x_y} = 0,97y + 0,91. \qquad (**)$$

Перевірка. За даними спостережень $\bar{x} = \frac{\sum x_i}{n_x} = \frac{15}{5} = 3$; $\bar{y} = \frac{\sum y_i}{n_y} = \frac{8,15}{5} = 1,63$. Знайдемо значення $\overline{y_x}$ з рівняння (*) при $x = \bar{x}$: $\overline{y_x}$= 0,202*3+1,024=1,63. Отже, розрахунки проведено правильно, оскільки $\overline{y_x} = \bar{y}$. За аналогією виконується перевірка і для рівняння (**).

У системі координат побудуємо точки (x_i; y_i) (рис. 5.2.1).

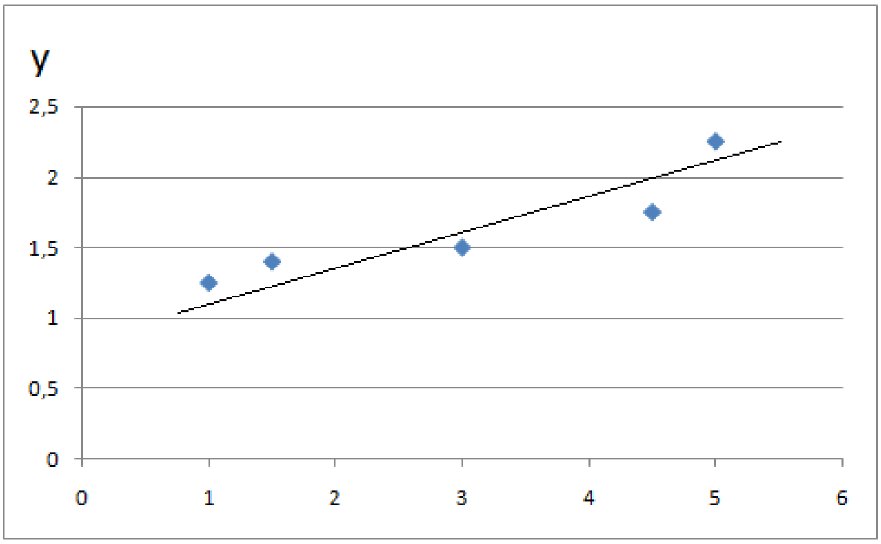


Рис. 5.2.1

Очевидно, що розташовані на площині точки групуються навколо деякої прямої, що вказує на лінійний характер зв'язку між ознаками, причому цей зв'язок прямий (більшому *x* відповідає більше *y*). На цій же координатній площині побудуємо лінію регресії (*) $\overline{y_x} = 0,202x + 1,024.$. Дійсно, точки (x_i; y_i) групуються навколо цієї прямої. Розрахуємо за цим рівнянням очікуване значення умовного середнього $\overline{y_x}$ та порівняємо їх з вибірковим середнім $\bar{y}$ при одних і тих же x_i. Для порівняння складемо таблицю:

x_i	y_i	$\overline{y_x}$	$\overline{y_x} - y_i$
1	1,25	1,226	-0,024
1,5	1,4	1,327	-0,073
3	1,5	1,63	0,13
4,5	1,75	1,933	0,183
5	2,25	2,034	-0,216

За значеннями відхилень $\overline{y_x} - y_i$ можна зробити висновок про те, що очікувані середні достатньо добре узгоджуються зі спостережуваними значеннями y_i (знак + чи – у цих відхиленнях свідчить про розташування точок відповідно нижче чи вище від прямої лінії регресії).

Примітка. Задачу складання рівняння прямої лінії регресії див. у Додатку 8.

При великій кількості спостережень ознак *X* та *Y* з метою спрощення розрахунків дослідні дані треба згрупувати в таблицю з двома входами, яка називається *кореляційною.* В ній одне і те ж значення *x* спостерігається n_x разів, одне і те ж значення *y* спостерігається n_y разів, а одна і та ж пара значень (x;y) спостерігається n_{xy} разів.

Y / *X*	y_1	y_2	…	y_k	n_{x_i}
x_1	n_{11}	n_{12}	…	n_{1k}	$n_{x_1} = \sum_{j=1}^{k} n_{1j}$

x$_2$	n_{21}	n_{22}	…	n_{2k}	$n_{x_2}=\sum_{j=1}^{k} n_{2j}$
…	…	…	…	…	…
x$_s$	n_{s1}	n_{s2}	…	n_{sk}	$n_{x_s}=\sum_{j=1}^{k} n_{sj}$
n_{y_i}	$n_{y_1}=\sum_{i=1}^{s} n_{i_1}$	$n_{y_2}=\sum_{i=1}^{s} n_{i_2}$	…	$n_{y_k}=\sum_{i=1}^{s} n_{i_k}$	$n=\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}$

У першому стовпцеві цієї таблиці (першому рядку) перераховуються у вибірці значення величини *X*: x_i, $i=1, \ldots, s$ (*У*: y_j, $j=1, \ldots, k$).

Якщо кількість різних значень величин *X* і *У* велика або ці величини розподілені неперервно, то відбувається групування їх значень за інтервалами. У цьому випадку x_i і y_j являють собою середини відповідних інтервалів.

На перетині стовпців і рядків указують частоти n_{ij}, які рівні числу появ у вибірці пари (x_i; y_j). Якщо пара значень ознак (x_i; y_j) не спостерігалась, то у відповідній клітинці ставиться риска.

В останньому рядку (в останньому стовпцеві) вказуються числа n_{x_i}, $i=1, \ldots, s$ (n_{yj}, $j=1, \ldots, k$), що дорівнюють кількості появ у вибірці значень x_i (y_j), незалежно від того, в парі з яким із значень *У* (*X*) воно з'явилось. Кореляційна таблиця містить всю інформацію, що отримана в результаті вибіркових спостережень величин *X* і *У*. Звідси, з врахуванням частот появи змінних x_i та y_j,отримаємо:

$$\sum_{i=1}^{n} x_i=\sum_{i=1}^{s} n_{x_i}\cdot x_i; \qquad \sum_{i=1}^{n} y_i=\sum_{j=1}^{k} n_{y_j}\cdot y_j; \qquad \sum_{i=1}^{n} x_i^2=\sum_{i=1}^{s} n_{x_i}\cdot x_i^2;$$
$$\sum_{i=1}^{n} y_i^2=\sum_{j=1}^{k} n_{y_j}\cdot y_j^2; \sum_{i=1}^{n} x_i y_i=\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} x_i y_j. \qquad (5.2.4)$$

Підставивши ці суми у формули (5.2.3) та (5.2.4), отримаємо:

$$a=\frac{n\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n\sum_{i=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2};\ b=\frac{\sum_{i=1}^{s}n_{x_i}x_i^2\sum_{j=1}^{k}n_{y_j}y_j-\sum_{i=1}^{s}n_{x_i}x_i\sum_{i=1}^{s}\sum_{j=1}^{k}n_{ij}x_iy_j}{n\sum_{i=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2}.$$

$$a'=\frac{n\sum_{i=1}^{s}\sum_{j=1}^{k}n_{ij}x_iy_j-\sum_{i=1}^{s}n_{x_i}x_i\sum_{j=1}^{k}n_{y_j}y_j}{n\sum_{j=1}^{k}n_{y_j}y_j^2-(\sum_{j=1}^{k}n_{y_j}y_j)^2}; \tag{5.2.5}$$

$$b'=\frac{\sum_{j=1}^{k}n_{y_j}y_j^2\sum_{i=1}^{s}n_{x_i}x_i-\sum_{j=1}^{k}n_{y_j}y_j\sum_{i=1}^{s}\sum_{j=1}^{k}n_{ij}x_iy_j}{n\sum_{j=1}^{k}n_{y_j}y_j^2-(\sum_{j=1}^{k}n_{y_j}y_j)^2}.$$

Як відомо, система рівнянь для визначення параметрів *a* та *b* рівняння прямої лінії регресії *У* на *Х* має вид (5.2.1). У даному випадку значення *Х* та відповідні їм значення *У* спостерігались по одному разу. Тепер, коли отримано велике число даних, серед яких є ті, що повторюються, і вони згруповані у вигляді кореляційної таблиці, запишемо систему так, щоб вона відображала дані кореляційної таблиці. Для цього використаємо такі тотожності:

$$\sum_{i=1}^{n}x_i=n\overline{x};\ \sum_{i=1}^{n}y_i=n\overline{y};\ \sum_{i=1}^{n}x_i^2=n\overline{x^2};\ \sum_{i=1}^{n}x_iy_i=n\overline{xy}; \tag{5.2.6}$$

де $\overline{x};\overline{y}$ – вибіркові середні; $\overline{x^2}$ – вибіркове середнє квадрата величин *Х*; $\overline{xy}$ – вибіркове середнє добутку.

Нагадаємо, що виправлене вибіркове відхилення визначається рівністю

$$S_{xy}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=\frac{1}{n}\sum_{i=1}^{n}x_iy_i-\overline{x}\overline{y}=\overline{xy}-\overline{x}\overline{y}. \tag{5.2.7}$$

Вибіркові дисперсії визначаються за такими формулами:

$$S_x^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\overline{x}^2=\overline{x^2}-\overline{x}^2, \tag{5.2.8}$$

$$S_y^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2=\frac{1}{n}\sum_{i=1}^{n}y_i^2-\overline{y}^2=\overline{y^2}-\overline{y}^2.$$

Вибірковий коефіцієнт кореляції дорівнює

$$\rho_{xy} = \frac{S_{xy}}{S_x S_y}. \qquad (5.2.9)$$

Підставимо праві частини рівностей (5.2.6) у систему (5.2.1), і отримаємо $\begin{cases} an\overline{x^2} + bn\overline{x} = n\overline{xy}, \\ an\overline{x} + bn = n\overline{y}, \end{cases}$

звідки

$$\begin{cases} a\overline{x^2} + b\overline{x} = \overline{xy}, \\ a\overline{x} + b = \overline{y}. \end{cases} \qquad (5.2.10)$$

Розв'язуючи систему (5.2.10) за формулами Крамера, отримаємо:

$$\Delta = \begin{vmatrix} \overline{x^2} & \overline{x} \\ \overline{x} & 1 \end{vmatrix} = \overline{x^2} - \overline{x}^2 = S_x^2; \qquad \Delta_a = \begin{vmatrix} \overline{xy} & \overline{x} \\ \overline{y} & 1 \end{vmatrix} = \overline{xy} - \overline{x}\,\overline{y} = S_{xy};$$

$$\Delta_b = \begin{vmatrix} \overline{x^2} & \overline{xy} \\ \overline{x} & \overline{y} \end{vmatrix} = \overline{x^2}\,\overline{y} - \overline{x}\cdot\overline{xy} = \overline{x^2}\,\overline{y} - \overline{x}^2\,\overline{y} + \overline{x}^2\,\overline{y} - \overline{x}\cdot\overline{xy} = \overline{y}(\overline{x^2} - \overline{x}^2) - \overline{x}(\overline{xy} - \overline{x}\,\overline{y}) = \overline{y}\cdot S_x^2 - \overline{x}\cdot S_{xy};$$

$$a = \frac{\Delta_a}{\Delta} = \frac{S_{xy}}{S_x^2}; \; b = \frac{\Delta_b}{\Delta} = \frac{\overline{y}S_x^2 - \overline{y}S_{xy}}{} = \overline{y} - \frac{S_{xy}}{S_x^2}\overline{x}.$$

Підставимо коефіцієнти *a* та *b* у рівняння регресії:

$$\overline{y_x} = \overline{y} - \frac{S_{xy}}{S_x^2}\overline{x} + \frac{S_{xy}}{S_x^2}x, \quad \overline{y_x} - \overline{y} = \frac{S_{xy}}{S_x^2}(x - \overline{x}).$$

Зробимо позначення: $\frac{S_{xy}}{S_x^2} = \frac{S_{xy}}{S_x S_y}\cdot\frac{S_y}{S_x} = \rho_{xy}\frac{S_y}{S_x}$). Отже, вибіркове рівняння прямої лінії регресії *У* на *X* має вид

$$\overline{y_x} - \overline{y} = \rho_{xy}\frac{S_y}{S_x}(x - \overline{x}). \qquad (5.2.11)$$

Аналогічно знайдемо рівняння прямої лінії регресії *X* на *У*

$$\overline{x_y} - \overline{x} = \rho_{xy}\frac{S_x}{S_y}(y - \overline{y}). \qquad (5.2.12)$$

З рівнянь (5.2.11) і (5.2.12) витікає, що прямі лінії регресії *У* на *X* та *X* на *У* проходять через точку з координатами $(\overline{x};\overline{y})$. Ці прямі співпадають, коли $|\rho_{xy}|^2 = 1$.

Величини $\rho_{xy}\dfrac{S_y}{S_x}$, $\rho_{xy}\dfrac{S_x}{S_y}$ називаються *вибірковими коефіцієнтами лінійної регресії* і позначаються

$$\rho_{y/x} = \rho_{xy}\frac{S_y}{S_x},\ \rho_{x/y} = \rho_{xy}\frac{S_x}{S_y}. \quad (5.2.13)$$

Перемноживши праві та ліві частини рівностей (5.2.13) і добувши корінь, отримаємо $\rho_{xy} = \pm\sqrt{\rho_{y/x}\rho_{x/y}}$, тобто *коефіцієнт кореляції є середнє геометричне коефіцієнтів лінійної регресії*.

Коефіцієнт регресії У на *X* (*X* на *У*) показує, на скільки одиниць в середньому змінюється змінна *У (X)* при зміні змінної *X (У)* на одну одиницю.

Вимірювання *тісноти зв'язку* кореляційної залежності – *друга основна задача теорії кореляції*. При великій тісноті зв'язку, знаючи аргумент, можна з великою точністю передбачити значення функції. Якщо ж тіснота зв'язку мала, то вплив аргументу *x* на функцію *y* виявляється лише в середньому за рівнянням регресії. Змінюваність функції під впливом змінюваності різних аргументів-факторів характеризується дисперсією.

Вибірковий лінійний коефіцієнт кореляції $\rho_{\text{в}}$ показує тісноту зв'язку між ознаками *X* та *Y*. Його можна обчислити за формулою

$$\rho_{\text{в}} = \frac{\sum n_{xy}xy - n\bar{x}\bar{y}}{n\sigma_x\sigma_y}, \quad (5.2.14)$$

де $|\rho_{\text{в}}| \leq 1$, а саме:

a) якщо $\rho_{\text{в}} = 0$, то *X* та *Y* не зв'язані кореляційною залежністю;

b) якщо $|\rho_{\text{в}}| = 1$, то *X* та *Y* зв'язані функціональною залежністю;

c) якщо $|\rho_{\text{в}}| < 1$, то між *X* та *Y* існує кореляційний зв'язок, причому він тим тісніший, чим $|\rho_{\text{в}}|$ ближчий до одиниці;

d) знак $\pm$.при коефіцієнті кореляції вказує на напрямок зв'язку: «+» - прямий, «-» - зворотний.

Під *тіснотою зв’язку* розуміють оцінку впливу факторної ознаки на результативну та встановлення адекватності теоретичної залежності між ознаками фактичним даним.

Коефіцієнт кореляції можна обчислити й за іншими формулами:

$$\rho = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}, \qquad (5.2.15)$$

де $\overline{xy} = \frac{\sum xy}{n}$; $\bar{x} = \frac{\sum x}{n}$; $\bar{y} = \frac{\sum y}{n}$; $\sigma_x = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$; $\sigma_y = \sqrt{\frac{\sum y^2}{n} - \bar{y}^2}$;

$$\rho = \frac{\sum (y - \bar{y})(x - \bar{x})}{n \sigma_x \sigma_y}, \qquad (5.2.16)$$

або

$$\rho = \frac{\sum (y - \bar{y})(x - \bar{x})}{\sqrt{\sum (y - \bar{y})^2 (x - \bar{x})^2}}, \qquad (5.2.17)$$

або

$$\rho = \frac{n \sum yx - \sum y \sum x}{\sqrt{(n \sum y^2 - (\sum y)^2)(n \sum x^2 - (\sum x)^2)}}, \qquad (5.2.18)$$

або

$$\rho = b \frac{\sigma_x}{\sigma_y}. \qquad (5.2.19)$$

Коефіцієнти регресії мають той же знак, що і коефіцієнт кореляції, і звязані співвідношенням

$$\rho^2 = \rho_{y/x} \cdot \rho_{x/e}. \qquad (5.2.20)$$

Тісноту зв’язку між ознаками оцінюють також за допомогою *коефіцієнта детермінації* – це квадрат коефіцієнта кореляції $R^2 = \rho^2$, який показує, яка доля загальної варіації ознаки визначається фактором, що досліджується. Він використовується як при лінійному, так і при нелінійному зв’язку між ознаками.

Коефіцієнт детермінації приймає значення від 0 до 1. Чим ближче R^2 до 1, тим тісніший зв'язок між ознаками. При $R^2 = 0$ відсутній лінійний зв'язок між ознаками; при $R^2 = 1$ не існує кореляційного зв’язку між ознаками.

Однак на підставі коефіцієнтів регресії не можна судити, яка з факторних ознак найбільше впливає на результативну, оскільки коефіцієнти регресії між собою непорівняльні, адже їх виражено різними одиницями, тому що найчастіше розглядаються залежності між ознаками різної природи. Наприклад, у рівнянні регресії маси людського тіла за зростом при вимірюванні зросту в сантиметрах коефіцієнт регресії *a* вийде у 100 разів менший, ніж при вимірюванні зросту у метрах. А при вимірюванні маси у грамах замість кілограмів коефіцієнт регресії *a* вийде у 1000 разів більшим. При довільному виборі одиниць вимірювання ознак коефіцієнт регресії *a* не може бути універсальним показником тісноти зв'язку. Якби ми мали у своєму розпорядженні таку систему вимірювання, використовуючи яку можна було б порівнювати дані за різними ознаками, то довільний вибір одиниць вимірювання не впливав би на коефіцієнт регресії. В якості такої системи використовується система вимірювання в «сигмах» (стандартизована). У ній за одиницю вимірювання ознаки береться її середнє квадратичне (стандартне) відхилення σ , початок відліку при цьому зазвичай переноситься в точку, що відповідає середньому арифметичному. Коефіцієнт регресії у стандартизованій системі одиниць (*стандартизований коефіцієнт регресії*) показує, на скільки сигм змінюється в середньому значення результативної ознаки *y* при зміні факторної ознаки *x* на одну сигму.

Після встановлення тісноти зв'язку дають оцінку *значимості зв'язку* між ознаками. Під терміном «значимість зв'язку» розуміють оцінку відхилення вибіркових змінних від своїх значень у генеральній сукупності за допомогою статистичних критеріїв. Оцінку значимості зв'язку здійснюють з використанням F-критерію Фішера і t-критерію Стьюдента.

Якщо дані спостережень за ознаками *Х* і *У* задані у вигляді кореляційної таблиці з рівновіддаленими варіантами, то можна перейти до умовних варіант:

$u_i = (x_i - C_1)/h_1, \quad \upsilon_i = (y_i - C_2)/h_2,$

де C_1, C_2 – несправжні („хибні”) нулі варіант Х та У відповідно, в якості яких вигідно взяти варіанти, які мають найбільшу частоту (моду) або медіану, або заокруглене до цілих середнє арифметичне значень ознак; h_1, h_2 – кроки, тобто різниця між двома сусідніми варіантами *Х* та *У*. Тоді відповідно

$$\overline{x} = h_1\overline{u} + C_1, \ \overline{y} = h_2 + \overline{\upsilon} + C_2, \qquad S_x^2 = \frac{h_1^2}{n}\sum_{i=1}^{s} u_i^2 n_{x_i} - (\overline{x} - C_1)^2 = h_1^2\overline{u^2} - (\overline{x} - C_1)^2. \qquad ,$$

$$S_y^2 = \frac{h_2^2}{n}\sum_{j=1}^{k} \upsilon_j^2 n_{y_j} - (\overline{y} - C_2)^2 = h_2^2\overline{\upsilon^2} - (\overline{y} - C_2)^2.$$

Обчислимо через умовні варіанти S_{xy},

$$S_{xy} = \frac{1}{n}\sum_{i=1}^{s}\sum_{j=1}^{k} x_i y_j n_{ij} - \frac{1}{n}\sum_{i=1}^{s} x_i n_{x_i} \frac{1}{n} = \frac{h_1 h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{k} \frac{(x_i - C_1 + C_1)}{h_1}\frac{y_j - C_2 + C_2}{h_2} n_{ij} -$$
$$\frac{h_1}{n}\sum_{i=1}^{s}\frac{(x_i - C_1 + C_1)}{h_1} n_{x_i} \frac{h_2}{n}\sum_{j=1}^{k}\frac{y_j - C_2 + C_2}{h_2} n_y = \frac{h_1 h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{r}(u_i + \frac{h_1}{C_1})\cdot(\upsilon_j + \frac{h_2}{C_2}) n_{ij} -$$
$$\frac{h_1 h_2}{n^2}\cdot\sum_{i=1}^{s}(u_i + \frac{C_1}{h_1}) n_{x_i}\cdot\sum_{j=1}^{k}(\upsilon_j + \frac{C_2}{h_2}) n_{y_j} = \frac{h_1 h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{k}(u_i\upsilon_j + \frac{C_1}{h_1}\upsilon_j + \frac{C_2}{h_2}u_i + \frac{C_1 C_2}{h_1 h_2}) n_{ij} -$$
$$\frac{h_1 h_2}{n^2}(\sum_{i=1}^{s} u_i n_{x_i} + \frac{C_1}{h_1}n)(\sum_{j=1}^{k}\upsilon_j n_{y_j} + \frac{C_2}{h_2}n) = \frac{h_1 h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{k} u_i\upsilon_j n_{ij} + \frac{C_1 h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{k}\upsilon_j n_{ij} + \frac{C_2 h_1}{n}\sum_{i=1}^{s}\sum_{j=1}^{k} u_i n_{ij} + C_1 C_2 -$$
$$\frac{h_1 h_2}{n^2}\sum_{i=1}^{s} u_i n_{x_i}\cdot\sum_{j=1}^{k}\upsilon_j n_{y_j} - \frac{C_1 h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{k}\upsilon_j n_{ij} - \frac{C_2 h_1}{n}\sum_{i=1}^{s}\sum_{j=1}^{k} u_i n_{ij} - C_1 C_2 = \frac{h_1 h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{k} u_i\upsilon_j n_{ij} - \frac{h_1 h_2}{n^2}\sum_{i=1}^{s} u_i n_{x_i}\sum_{j=1}^{k}\upsilon_j n_{y_j} =$$
$$h_1 h_2(\overline{u\upsilon} - \overline{u}\,\overline{\upsilon}),$$

де $n = \sum_{i=1}^{s} n_{x_i} = \sum_{j=1}^{k} n_y = \sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}, \overline{u} = \frac{1}{n}\sum_{i=1}^{s} u_i n_{x_i}; \overline{\upsilon} = \frac{1}{n}\sum_{j=1}^{k} n_j \upsilon_{y_j}; \overline{u\upsilon} = \frac{1}{n}\sum_{i=1}^{s}\ \sum_{j=1}^{k} u_i \upsilon_j n_{ij}.$

Отже,

$$S_{xy} = \frac{h_1 h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{k} u_i\upsilon_j n_{ij} - \frac{h_1 h_2}{n^2}\sum_{i=1}^{s} u_i n_{x_i}\sum_{j=1}^{k}\upsilon_j n_{y_j}. \qquad (5.2.21)$$

Приклад 5.2.2. Знайти вибіркові рівняння прямих лінії регресії У на *Х* та *Х* на *У* за даними кореляційної таблиці та побудувати їх графіки:

X / Y	10	20	30	40	50	60	n_{y_j}
15	5	7	-	-	-	-	12
25	-	20	23	-	-	-	43
35	-	-	30	47	2	-	79
45	-	-	10	11	20	6	47
55	-	-	-	9	7	3	19
n_{x_i}	5	27	63	67	29	9	n=200

Розв'язання. Перейдемо до умовних варіант $u_i = (x_i - C_1)/h_1$, $\upsilon_i = (y_i - C_2)/h_2$, $u_i = \frac{x_i - 40}{10}$ і $\upsilon_i = \frac{y_i - 35}{10}$, де в якості хибних нулів C_1 і C_2 візьмемо варіанти відповідно x= 40 і y=35, які розміщені приблизно посередині варіаційних рядів, тобто C_1 =40 і C_2 =35, а $h_1 = h_2$ = 10. Складемо кореляційну таблицю в умовних варіантах, де передостанні рядки і стопвчики містять наступні суми:

$$\sum_{i=1}^{6} u_i n_{x_i} = 18; \sum_{i=1}^{6} u_i^2 n_{x_i} = 214; \sum_{j=1}^{5} \upsilon_j n_{y_j} = -88; \sum_{j=1}^{5} \upsilon_j^2 n_{y_j} = 290.$$

Складемо розрахункову кореляційну таблицю 5.2.1. У правих верхніх кутах завантажених кліток запишемо добутки $u_i v_j$.

За даними таблиці розрахуємо потрібні статистики:

$$\overline{u} = \frac{\sum u_i n_u}{n}; \quad \overline{u} = \frac{-97}{200} = -0{,}485;$$

$$\overline{v} = \frac{\sum v_j n_v}{n}; \quad \overline{v} = \frac{18}{200} = 0{,}09;$$

$$\overline{u^2} = \frac{\sum u^2{}_i n_u}{n}; \quad \overline{u^2} = \frac{281}{200} = 1{,}405;$$

$$\overline{v^2} = \frac{214}{200} = 1{,}07;$$

$$\sigma_u = \sqrt{\overline{u^2} - (\overline{u})^2} = \sqrt{1{,}405 - (-0{,}485)^2} = 1{,}082;$$

$$\sigma_v = 1{,}030;$$

Таблиця 5.2.1

Розрахункова кореляційна таблиця

υ \ U	-3	-2	-1	0	1	2	n_{υ_j}	$\upsilon_j n_{\upsilon_j}$	$\upsilon_j^2 n_{\upsilon_j}$	$\sum_{j=1}^{5} u_i \upsilon_j n_{ij}$
-2	5 6	7 4	-	-	-	-	12	-24	48	58
-1	-	20 2	23 1	-	-	-	43	-43	43	63
0	-	-	30 0	47 0	2 0	-	79	0	0	0
1	-	-	10 -1	11 0	20 1	6 2	47	47	47	22
2	-	-	-	9 0	7 2	3 4	19	38	76	26
n_{u_i}	5	27	63	67	29	9	200	18	214	-
$u_i n_{u_i}$	-15	-54	-63	0	29	18	-97	-	-	-
$u_i^2 n_{u_i}$	45	108	63	0	29	36	281	-	-	-
$\sum_{i=1}^{6} u_i \upsilon_j n_{ij}$	30	68	13	0	34	24	-	-	-	169

$$\sum_{i=1}^{6} u_i \upsilon_j n_{ij} = 169.$$

$$S_x^2 = \overline{x^2} - \overline{x}^2,\ S_y^2 = \overline{y^2} - \overline{y}^2.$$

Знайдемо вибірковий коефіцієнт кореляції $\rho_{xy} = \frac{S_{xy}}{S_x S_y}$, а потім напишемо вибіркове рівняння прямої регресії X на У

$$\overline{x_y} - \overline{x} = \rho_{xy} \frac{S_x}{S_y}(y - \overline{y}).$$

Перейдемо до варіант *x* та *y*:

$\overline{x} = h_1 \overline{u} + c_1;$ $\qquad \overline{x} = 10 \cdot (-0{,}485) + 40 = 35{,}15;$

$\overline{y} = h_2 \overline{v} + c_2;$ $\qquad \overline{y} = 10 \cdot 0{,}09 + 35 = 35{,}9;$

$S_x^2 = h_1^2\overline{u^2} - (\overline{x} - c_1)^2;\quad S_x^2 = 100 \cdot 1{,}405 - (35{,}15 - 40)^2 = 116{,}9775;$

$S_x = 10{,}8156;$

$S_y^2 = h_2^2\overline{v^2} - (\overline{y} - c_2)^2;\quad S_x^2 = 100 \cdot 1{,}07 - (35{,}9 - 35)^2 = 106{,}19;$

$S_y = 10{,}3049;$

$$S_{xy} = \frac{h_1 h_2}{n}\sum_{i=1}^{s}\sum_{j=1}^{k} u_i v_j n_{ij} - \frac{h_1 h_2}{n^2}\sum_{i=1}^{s} u_i n_{x_i}\sum_{j=1}^{k} v_j n_{y_j} = \frac{100}{200}\cdot 169 -$$

$$- \frac{100}{40000}\cdot(-97)\cdot 18 = 88{,}865.$$

Знайдемо вибірковий коефіцієнт кореляції

$$\rho_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{88{,}865}{10{,}8156\cdot 10{,}3049} = 0{,}7973,$$

а потім напишемо вибіркове рівняння прямої регресії Х на У:

$$\overline{x_y} - \overline{x} = \rho_{xy}\frac{S_x}{S_y}(y - \overline{y});$$

$$\overline{x_y} - 35{,}15 = 0{,}7973 \cdot \frac{10{,}8156}{10{,}3049}\cdot(y - 35{,}9);$$

$$\overline{x_y} = 0{,}8368y + 5{,}1084.$$

Контроль: підставимо у праву частину знайденого рівняння значення $\overline{y} = 35{,}9$, тоді 0,8368*35,9+5,1084=35,14952≈35,15= $\overline{x}$, що свідчить про правильність складеного рівняння регресії $\overline{x_y}$.

Аналогічно знайдемо рівняння прямої регресії У на Х:

$$\overline{y_x} - \overline{y} = \rho_{xy}\frac{S_y}{S_x}(x - \overline{x}).$$

Маємо: $\overline{y_x} - 35{,}9 = 0{,}7973 \cdot \frac{10{,}3049}{10{,}8156}(x - 35{,}15)$, звідки

$$\overline{y_x} = 0{,}7596x + 9{,}1982.$$

Контрольна перевірка підтверджує правильність складеного рівняння прямої лінії регресії У на Х.

Побудуємо графіки отриманих прямих регресій (Рис. 5.2.2):

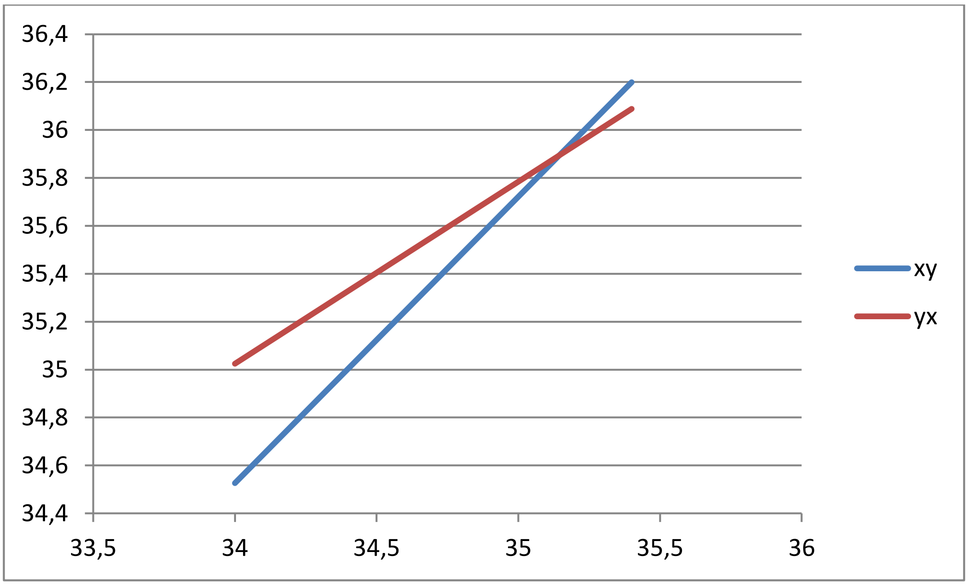


Рис. 5.2.2

Точка перетину графіків прямих ліній регресії має координати $\overline{(x;\bar{y})}$, а саме (35,15; 35,9).

Вибірковий лінійний коефіцієнт кореляції $\rho_в$ показує тісноту і напрямок зв'язку між ознаками *X* та *Y*. Його можна обчислити за формулою

$$\rho_в = \frac{\sum n_{xy}xy - n\bar{x}\bar{y}}{n\sigma_x\sigma_y}. \qquad (5.2.14)$$

Дійсно, $\rho_в = \frac{169-200\cdot(-0{,}425)\cdot 0{,}09}{200\cdot 1{,}082\cdot 1.030} = 0{,}7925$. Оскільки $\rho_в > 0$, то зв'язок між ознаками *X* та *Y* прямий (зі збільшенням *X* збільшується й Y), а також тісний, тому що $\rho_в$ близький до 1.

Оскільки $|\rho_в| < 1$, то кут нахилу прямої регресії *У* на *X* має менший нахил до осі *OX*, ніж пряма регресії *X* на *У*. Чим ближче $|\rho_в|$ до 1, тим менший кут між прямими регресії. Ці прямі співпадуть тоді і тільки тоді, коли $|\rho_в| = 1$. При $\rho_в = 0$ прямі регресії мають рівняння $y = b;\ \ x = a$.

Перевіримо умову (5.2.20) $\rho^2 = \rho_{y/x}{\cdot}\rho_{x/e}$:

0,8368*0,7596 =0,7925^2≈0,63.

5.3. Множинна лінійна кореляція

У тих випадках, коли досліджується кореляційний зв'язок між двома або більше величинами, використовують *множинну кореляцію.*

Розглянемо кореляційний зв'язок між трьома кількісними ознаками *X, Y* і *Z.* Можна ввести таке рівняння регресії

$$\overline{z_{xy}} = f\ (x,y),$$

де $\overline{z_{xy}}$ – середнє значення величини *Z*, яке відповідає визначеним значенням *x* та *y.* Геометричною інтерпретацією цього рівняння є деяка поверхня в прямокутній системі координат тривимірного простору.

У найпростішому випадку лінійної кореляційної залежності ознак *X, Y* і *Z* вибіркове рівняння регресії має вид

$$\overline{z_{xy}} = ax + by + c. \qquad (5.3.1)$$

Нехай у результаті незалежних спостережень над системою кількісних ознак (*X; Y; Z*), яка в даний момент вивчається; отримано *n* сукупностей чисел ($x_1; y_1; z_1$), ($x_2; y_2; z_2$), …, ($x_n; y_n; z_n$). Припустимо, що функція регресії лінійна, тобто

$$z = ax + by + c.$$

У формулі (5.3.1) $\overline{z_{xy}}$ замінено на z , так як різні значення x та y ознак X та У, і відповідне їм значення z ознаки Z спостерігались по одному разу. Якщо позначити за $z_i \ = ax_i + by_i + c$ наближене значення z_i , обчислене з рівняння (5.3.1), то величина $z_i - z_i$ є відхиленням наближеного z_i значення від точного z_i . Коефіцієнти a, b і c рівняння регресії (5.3.1) знайдемо з вимоги методу найменших квадратів:

$$S(a;\ b;\ c) = \sum_{i=1}^{n}(z_i - z_i) = \sum_{i=1}^{n}(z_i - ax_i - by_i + c)^2 \to \min .$$

Необхідні умови мінімуму функції *S* утворюють систему

$$\begin{cases} \dfrac{\partial S}{\partial a} = 2\sum\limits_{i=1}^{n}(z_i - ax_i - by_i - c)(-x_i) = 0, \\ \dfrac{\partial S}{\partial b} = 2\sum\limits_{i=1}^{n}(z_i - ax_i - by_i - c)(-y_i) = 0, \\ \dfrac{\partial S}{\partial c} = 2\sum\limits_{i=1}^{n}(z_i - ax_i - by_i - c)(-1) = 0, \end{cases} \quad (5.3.2)$$

яка в результаті тотожних перетворень набуде вигляду

$$\begin{cases} a\sum\limits_{i-1}^{n} x_i^2 + b\sum\limits_{i-1}^{n} x_i y_i + c\sum\limits_{i=1}^{n} x_i = \sum\limits_{i-1}^{n} x_i z_i, \\ a\sum\limits_{i=1}^{n} x_i y_i + b\sum\limits_{i=1}^{n} y_i^2 + c\sum\limits_{i=1}^{n} y_i = \sum\limits_{i=1}^{n} y_i z_i, \\ a\sum\limits_{i=1}^{n} x_i + b\sum\limits_{i=1}^{n} y_i + cn = \sum\limits_{i=1}^{n} z_i. \end{cases} \quad (5.3.3)$$

Розв'язавши систему довільним способом , підставимо значення *a, b* і *c* у рівняння (5.3.1) та отримаємо шукане рівняння регресії.

Зручніше знайти готові формули для розрахунку параметрів множинної регресії, а не кожного разу розв'язувати систему.

Коефіцієнти *a, b* і *c* рівняння регресії (5.3.1) обчислюються за наступними формулами:

$$a = \frac{\rho_{xz} - \rho_{yz}\rho_{xy}}{1-\rho_{xy}^2} \cdot \frac{S_z}{S_x}; \; b = \frac{\rho_{yz} - \rho_{xz}\rho_{xy}}{1-\rho_{xy}^2} \cdot \frac{S_z}{S_y}; \; c = \bar{z} - a\bar{x} - b, \quad (5.3.4)$$

де $\rho_{xy}, \rho_{xz}, \rho_{yz}$ - вибіркові коефіцієнти лінійної кореляції між відповідними ознаками; *Sx* ,*Sy*, *Sz* - середні квадратичні відхилення; $\overline{x}, \overline{y}, \overline{z}$ - середні значення відповідних кількісних ознак.

Для множинної лінійної регресії рівняння зв'язку матиме вигляд

$$\bar{y} = a_0 + a_1 x_1 + \cdots + a_n x_n \quad (5.3.5)$$

Методологія множинної кореляції базується на загальних принципах кореляційного аналізу. Однак, у ній ускладнюється змістовий аналіз, збільшується складність математичного апарату тощо. Тому такий кореляційний зв'язок доцільніше встановлювати з використанням комп’ютерних технологій.

Коефіцієнти множинної регресії показують ступінь середньої зміни результативної ознаки при зміні відповідної факторної ознаки на одиницю (одне своє значення) при умові, що решта факторів, що входять у рівняння регресії, залишаються сталими (фіксованими) на одному, найчастіше середньому, рівні.

З метою виявлення *порівняльноїсиливпливу* окремих факторів та їхніх резервів, статистика обчислює часткові коефіцієнти еластичності ε_i, а також бета-коефіцієнти β_i, за формулами:

$$\varepsilon_i = a_i \frac{\overline{X}_i}{\overline{Y}}; \qquad \beta_i = a_i \frac{\sigma_{x_i}}{\sigma_y}, \tag{5.3.6}$$

де a_i — коефіцієнт регресії при і-му факторі; $\overline{X}_i$ — середнє значення і-го фактора; $\overline{Y}$ — середнє значення результативної ознаки; σ_{x_i} — середнє квадратичне відхилення і-го фактора; σ_y — середнє квадратичне відхилення результативної ознаки.

Часткові коефіцієнти еластичності ε показують, на скільки процентів у середньому зміниться результативна ознака при зміні на 1% кожного фактора та фіксованому положенні інших факторів.

Для визначення факторів, які мають найбільші резерви поліпшення досліджуваної ознаки, з урахуванням ступеня варіації факторів, закладених у рівняння множинної регресії, обчислюють часткові β-коефіцієнти, які показують, на яку частину середнього квадратичного відхилення змінюється результативна ознака при зміні відповідної факторної ознаки на значення її середнього квадратичного відхилення.

Коефіцієнти множинної регресії, що характеризують зв'язок між результативною ознакою і фактором при фіксованому значенні решта факторів, називаються *коефіцієнтами чистої* *регресії*, а коефіцієнти парної регресії - *коефіцієнтами повної регресії*.

Для характеристики *щільності зв'язку* в множинній лінійній кореляції використовують *множинний коефіцієнт кореляції*

$$R_{yx_1x_2} = \sqrt{\frac{r^2{}_{yx_1} + r^2{}_{yx_2} - 2r_{yx_1}r_{yx_2}r_{x_1x_2}}{1 - r^2{}_{x_1x_2}}}, \tag{5.3.7}$$

де r_{yx_1}, r_{yx_2}, $r_{x_1x_2}$ — парні коефіцієнти лінійної кореляції двох ознак:

$$r_{yx_1} = \frac{\overline{YX}_1 - \bar{Y}\ \bar{X}_1}{\sigma_y \sigma_{x_1}};\ r_{yx_2} = \frac{\overline{YX}_2 - \bar{Y}\ \bar{X}_2}{\sigma_y \sigma_{x_2}};\ r_{x_1x_2} = \frac{\bar{X}_1\bar{X}_2 - \bar{X}_1\bar{X}_2}{\sigma_{x_1}\sigma_{x_2}}. \tag{5.3.8}$$

Множинний коефіцієнт кореляції показує, яку частину загальної кореляції складають коливання, під впливом факторів *X*$_1$*, X*$_2$*, ..., X*$_n$*,* щозакладені у багатофакторну модель для дослідження.

Множинний коефіцієнт кореляції коливається в межах від 0 до ± 1. При *R*= 0 зв'язку між досліджуваними ознаками немає, при *R*= 1 — зв'язок функціональний.

Перш ніж розраховувати множинний коефіцієнт кореляції, потрібно обчислити парні коефіцієнти кореляції, наприклад, для двох факторних ознак:

$$r_{yx_1} = \frac{\overline{YX}_1 - \bar{Y}\ \bar{X}_1}{\sigma_y \sigma_{x_1}};\ r_{yx_2} = \frac{\overline{YX}_2 - \bar{Y}\ \bar{X}_2}{\sigma_y \sigma_{x_2}};\ r_{x_1x_2} = \frac{\bar{X}_1\bar{X}_2 - \bar{X}_1\bar{X}_2}{\sigma_{x_1}\sigma_{x_2}}. \tag{5.3.9}$$

Високі значення парних коефіцієнтів кореляції свідчать про сильний вплив (окремо) факторів на результативну ознаку.

На основі парних коефіцієнтів кореляції можна обчислити часткові коефіцієнти кореляції першого порядку:

$$r_{yx_1(x_2)} = \frac{r_{yx_1} - r_{yx_2}r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_2})(1 - r^2_{x_1x_2})}};\ r_{yx_2(x_1)} = \frac{r_{yx_2} - r_{yx_1}r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_1})(1 - r^2_{x_1x_2})}}. \tag{5.3.10}$$

якіпоказують зв'язок кожного фактора з досліджуваним показником за умови комплексної взаємодії факторів.

Для виявлення щільності зв'язку між результативною ознакою і двома факторними ознаками водночас обчислюємо сукупний коефіцієнт множинної кореляції

$$R_{yx_1x_2} = \sqrt{\frac{r^2_{yx_1} + r^2_{yx_2} - 2r_{yx_1} r_{yx_2} r_{x_1x_2}}{1 - r^2_{x_1x_2}}}. \qquad (5.3.11)$$

Обчислений коефіцієнт множинної кореляції R, який близький до 1, показує, що між двома факторними і результативною ознаками існує достатньо щільний зв'язок.

Сукупний коефіцієнт множинної детермінації R^2, наприклад, 0,995 для двох ознак, свідчить про те, що варіація результативної ознаки на 99,5% зумовлюється двома факторами, що введені в кореляційну модель. Це означає, що вибрані фактори суттєво впливають на досліджуваний показник.

Аби поглибити економічний аналіз, збільшують кількість суттєвих факторів, які вводять у модель досліджуваного показника, і будують багатофакторні рівняння регресії, використовуючи сучасні методи і засоби обчислювальної техніки.

5.4. Нелінійна кореляційна залежність

Якщо лінії регресії відмінні від прямих, а графік регресії зображується кривою лінією, то таку кореляцію називають *криволінійною*. При криволінійній залежності система рівнянь будується так само, як і для прямолінійної.

Не існує загального правила для вибору емпіричної формули лінії регресії потрібного виду. Можна лише здогадуватися про підходящу форму рівняння за формою кривої, що зображують дані. Але існують способи, за допомогою яких можна перевірити правильність або неправильність здогадки.

Для залежностей з двома параметрами, що зустрічаються найчастіше, емпіричну формулу можна підібрати за допомогою таблиці 5.4.1.

Таблиця 5.4.1

№	$\bar{x}_s$	$\bar{y}_s$	Вигляд емпіричної формули
1	$\frac{x_1 + x_n}{2}$	$\frac{y_1 + y_n}{2}$	$y = ax + b$
2	$\sqrt{x_1 x_n}$	$\sqrt{y_1 y_n}$	$y = ax^b$
3	$\frac{x_1 + x_n}{2}$	$\sqrt{y_1 y_n}$	$y = ab^x$; $y = ae^{\beta x}$, де $\beta = lnb$
4	$\frac{2x_1 x_n}{x_1 + x_n}$	$\frac{y_1 + y_n}{2}$	$y = a + \frac{b}{x}$
5	$\frac{x_1 + x_n}{2}$	$\frac{2y_1 y_n}{y_1 + y_n}$	$y = \frac{1}{ax + b}$
6	$\frac{2x_1 x_n}{x_1 + x_n}$	$\frac{2y_1 y_n}{y_1 + y_n}$	$y = \frac{x}{ax + b}$
7	$\sqrt{x_1 x_n}$	$\frac{y_1 + y_n}{2}$	$y = algx + b$

Зокрема, у випадку *параболічної* кореляції другого порядку вибіркове рівняння регресії *У* на *Х* має вид

$$\overline{y_x} = ax^2 + bx + c. \qquad (5.4.1)$$

Невідомі параметри *a, b* і *c* знаходять методом Гаусса з системи рівнянь, що складена за МНК:

$$\begin{cases} a\sum_{i-1}^{s} n_{x_i} x_i^4 + b\sum_{i-1}^{s} n_{x_i} x_i^3 + c\sum_{i=1}^{s} n_{x_i} x_i^2 = \sum_{i-1}^{s} n_{x_i} \overline{y_x} x_i^2, \\ a\sum_{i=1}^{s} n_{x_i} x_i^3 + b\sum_{i=1}^{s} n_{x_i} x_i^2 + c\sum_{i=1}^{s} n_{x_i} x_i = \sum_{i=1}^{s} n_{x_i} \overline{y_x} x_i, \\ a\sum_{i=1}^{s} n_{x_i} x_i^2 + b\sum_{i=1}^{s} n_{x_i} x_i + cn = \sum_{i=1}^{s} n_{x_i} \overline{y_x}. \end{cases} \qquad (5.4.2)$$

Аналогічно знаходиться вибіркове рівняння регресії *Х* на *У*:

$$\overline{x_y} = ay^2 + by + c.$$

Приклад 5.4.1. За даними кореляційної таблиці знайти вибіркове рівняння регресії $\overline{y_x} = ax^2 + bx + c.$

X / *Y*	2	3	5	n_{y_j}
25	20	-	-	20
45	-	30	1	31
110	-	1	48	49
n_{x_i}	20	31	49	*n*=100

Розв'язання. Складемо наступну розрахункову таблицю:

x	n_{x_i}	$n_{x_i} x_i$	$n_{x_i} x_i^2$	$n_{x_i} x_i^3$	$n_{x_i} x_i^4$	$\overline{y_x}$	$n_{x_i} \overline{y_x}$	$n_{x_i} \overline{y_x} x_i$	$n_{x_i} \overline{y_x} x_i^2$
2	20	40	80	160	320	25	500	1000	2000
3	31	93	279	837	2511	47,1	1460	4380	13141
5	49	245	1225	6125	30625	108,67	5325	26624	133121
$\sum$	100	378	1584	7122	33456		7285	32004	148262

Далі підставимо числа останнього рядка таблиці в систему (5.4.2) та отримаємо систему рівнянь відносно невідомих коефіцієнтів *a, b, c.*

$$\begin{cases} 33456a + 7122b + 1584c = 148262, \\ 7122a + 1584b + 378c = 32004, \\ 1584a + 378b + 100c = 7285. \end{cases}$$

Розв'язавши цю систему (наприклад методом Гаусса), знайдемо

a= 2,94; *b*= 7,27; *c* = -1,25.

Підставимо знайдені коефіцієнти в рівняння регресії

$$\overline{y_x} = ax^2 + bx + c.$$

і остаточно отримаємо

$$\overline{y_x} = 2,94x^2 + 7,27x - 1,25.$$

У випадку *гіперболічної* регресії *У* на *Х* рівняння гіперболи має вигляд

$$\overline{y_x} = \frac{a}{x} + b.$$

Для знаходження коефіцієнтів *a* і *b,* одержуємо систему рівнянь

$$\begin{cases} a\sum_{i=1}^{s} \frac{1}{x_i} n_{x_i} + bn = \sum_{i=1}^{s} n_{x_i} \overline{y_x}, \\ a\sum_{i=1}^{s} \frac{1}{x_i^2} n_{x_i} + b\sum_{i=1}^{s} \frac{1}{x_i} n_{x_i} = \sum_{i=1}^{s} \frac{1}{x_i} n_{x_i} \overline{y_x}. \end{cases}$$

У випадку *гіперболічної* регресії *Х* на *У* вибіркове рівняння регресії має вид

$$\overline{x_y} = \frac{a}{y} + b.$$

.

Для *перевірки придатності вибраної емпіричної формули*, використовуючи початкові дані, знаходять значення $\overline{x_s}$ та $\overline{y_s}$. Потім порівнюють y_s, що відповідають $\overline{x_s}$ у початкових даних, зі значеннями $\overline{y_s}$. Якщо $\overline{x_s}$ не міститься серед початкових даних x_i, то відповідне значення можна визначити за допомогою лінійної інтерполяції:

$$\widehat{y_S} = y_i + \frac{y_{i+1} - y_i}{x_{i+1} - x_i}(\overline{x_s} - x_i),$$

де x_i та x_{i+1} - проміжні значення, між якими знаходиться $\overline{x_s}$ ($x_i < \overline{x_s} < x_{i+1}$).

Якщо величина $|\widehat{y_S} - \overline{y_s}|$ велика, то відповідна емпірична формула непридатна.

Залежності 1-7, що наведені в табл. 5.4.1, монотонні, і, отже, придатні лише в тому випадку, коли в початкових даних $x_{i+1} - x_i > 0$, а $y_{i+1} - y_i$ має сталий знак.

Приклад 5.4.2. Визначити вид емпіричної формули, що відповідає наступній таблиці:

X	2	3	4	5	6	7	8	9	10
y	12	35	75	125	210	315	445	600	800

Розв'язання. Підбір емпіричної формули за наведеними вище критеріями надано в табл. 5.4.2.

Таблиця 5.4.2

№	$\bar{x}_s$	$\bar{y}_s$	$\hat{y}_s$	$\lvert\hat{y}_s - \overline{y_s}\rvert$	Вигляд емпіричної формули
1	$\frac{x_1+x_n}{2} = \frac{2+10}{2} = 6$	$\frac{y_1+y_n}{2} = \frac{12+800}{2} = 406$	210	196	$y = ax + b$ не підходить
2	$\sqrt{x_1 x_n} = \sqrt{2\cdot 10} = 4{,}47$	$\sqrt{y_1 y_n} = \sqrt{12\cdot 800} = 98$	98,5	0,5	$y = ax^b$ підходить краще інших формул
3	$\frac{x_1+x_n}{2} = 6$	$\sqrt{y_1 y_n} = 98$	210	112	$y = ab^x$; $y = ae^{\beta x}$, де $\beta = lnb$ не підходить
4	$\frac{2x_1 x_n}{x_1+x_n} = 3{,}3$	$\frac{y_1+y_n}{2} = 406$	47	359	$y = a + \frac{b}{x}$ не підходить
5	$\frac{x_1+x_n}{2} = 6$	$\frac{2y_1 y_n}{y_1+y_n} = 23{,}6$	210	186,4	$y = \frac{1}{ax+b}$ не підходить
					Продовження табл.. 5.4.2
6	$\frac{2x_1 x_n}{x_1+x_n} = 3{,}3$	$\frac{2y_1 y_n}{y_1+y_n} = 23{,}6$	47	23,4	$y = \frac{x}{ax+b}$ не підходить
7	$\sqrt{x_1 x_n} = 4{,}47$	$\frac{y_1+y_n}{2} = 406$	98,5	307,5	$y = algx + b$ не підходить

Питання для самоперевірки

1. Яка залежність між ознаками Х та У називається функціональною, статистичною, кореляційною?

2. Що називається умовним середнім $\overline{y_x}\left(\overline{x_y}\right)$ спостережуваних значень У (Х) ?

3. Сформулюйте основні задачі регресійного і кореляційного аналізу.

4. Запишіть модель лінійної парної регресії У на Х та Х на У.

5. Що таке лінійна залежність і якою формулою вона визначається?

6. Що означають коефіцієнти a і b лінійної залежності?

7. Які знаки мають a і b для зростаючої та спадної залежностей?

5. У чому полягає суть метода найменших квадратів?

6. З яких умов підбираються a і b методом найменших квадратів?

7. Навести систему рівнянь відносно параметрів a і b лінії регресії.

8. Запишіть формули для знаходження параметрів *a* і *b* (a' *і b'*)прямої регресії.

9. Поясніть, як заповнюється кореляційна таблиця ?

10. Чому дорівнює вибірковий коефіцієнт кореляції ?

11. Чому дорівнюють вибіркові коефіцієнти лінійної регресії?

12. Установіть зв'язок між ρ_{xy} і $\rho_{y/x}$ $\rho_{x/y}$.

13. Яким чином визначають тип залежності (лінійний, криволінійний) між ознаками?

14. Як записується рівняння криволінійної параболічної залежності?

15. Яким чином знаходять координати перетину лінії з осями Х, У?

16. З якої умови знаходять екстремуми (максимум, мінімум) параболічноїзалежності?

17. При яких умовах параболічна залежність має максимум або мінімум?

18. Який вигляд має метод найменших квадратів для параболічної залежності?

19. Як розв'язувати методом детермінантів (визначників) систему лінійних рівнянь з трьома або більше невідомими?

Задачі та вправи

5.1. Знайти вибіркове рівняння прямої лінії регресії *У* на *Х* за даними п'яти спостережень.

x_i	1,40	1,90	3,40	4,90	5,40
y_i	1,25	1,50	1,75	2,25	2,50

5.2. Були проведені виміри довжини *Х* (мм) і ваги *У* (г) 10 яєць курки. Результати вимірів наступні:

Довжина	60	58	57	55	56	58	55	57	55	59
Вага	56	53	54	51	54	59	55	55	56	57

Обчислити коефіцієнт кореляції.

5.3. Знайти вибіркові рівняння прямих лінії регресії *У* на *Х* та *Х* на *У* за даними кореляційної таблиці та побудувати їх графіки.

X \ $У$	5	10	15	20	**25**	30	35	40	n_{y_j}
100	2	1	-	-	**-**	-	-	-	3
120	3	4	3	-	**-**	-	-	-	10
140	-	-	5	10	**8**	-	-	-	23
160	-	-	-	1	**-**	6	1	1	9
180	-	-	-	-	**-**	-	4	1	5
n_{x_i}	5	5	8	11	**8**	6	5	2	n=50

5.4. Дано кореляційну таблицю. Знайти вибіркові рівняння прямих регресії.

$У$ \ X	5	10	15	20	n_{y_j}
10	2	-	-	-	2
20	5	4	1	-	10
30	3	8	6	3	20
40	-	3	6	6	15
50	-	-	2	1	3
n_{x_i}	10	15	15	10	n=50

5.5. Знайти вибіркові рівняння прямих ліній регресії У на *X* та *X* на *У* за даними кореляційної таблиці та побудувати їх графіки.

У \ X	15	25	35	45	55	65	75	n_{y_j}
80	-	-	-	-	-	6	1	7
100	-	-	-	-	-	4	2	6
120	-	-	8	10	5	-	-	23
140	3	4	3	-	-	-	-	10
160	2	1	-	1	-	-	-	4
n_{x_i}	5	5	11	11	5	10	3	n=50

5.6. Задана кореляційна таблиця. Знайти вибіркові рівняння прямих регресії.

$Y \backslash X$	65	95	125	155	185	215	n_{y_j}
30	5	-	-	-	-	-	5
40	4	12	-	-	-	-	16
50	-	8	5	4	-	-	17
60	-	1	5	7	2	-	15
70	-	-	-	-	1	1	2
n_{x_i}	9	21	10	11	3	1	n=55

5.7. Дана кореляційна таблиця. Знайти вибіркові рівняння прямих регресії.

$Y \backslash X$	18	23	28	33	38	43	48	n_{y_j}
125	-	1	-	-	-	-	-	1
150	1	2	5	-	-	-	-	8
175	-	3	2	12	-	-	-	17
200	-	-	1	8	7	-	-	16
225	-	-	-	-	3	3	-	6
250	-	-	-	-	-	1	1	2
n_{x_i}	1	6	8	20	10	4	1	n=50

5.8. За даними кореляційної таблиці знайти вибіркове рівняння регресії $\overline{y_x} = ax^2 + bx + c.$

а)

$Y \backslash X$	0	4	6	7	10	n_{y_j}
7	19	1	1			21
13	2	14				16
40		3	22	2		27
80				15		15
200					21	21
n_{x_i}	21	18	23	17	21	n=100

б)

$Y \backslash X$	0	4	5	n_{y_j}
15	50	5	1	56
35		44		44
50		5	45	50
n_{x_i}	50	54	46	n=150

в)

$Y \backslash X$	0	1	2	3	4	n_{y_j}
10	20	5				25
11	7	15	3	1		26
20		3	17	4		24
35			8	13	7	28
50				5	42	47
n_{x_i}	27	23	28	23	49	n=150

г)

$Y \backslash X$	7	8	9	n_{y_j}
200	41	7		48
300	1	52	1	54
400		8	40	48
n_{x_i}	42	67	41	n=150

д)

$Y \backslash X$	0	1	2	3	4	n_{y_j}
0	18	1	1			20
3	1	20				21
5	3	5	10	2		20
10			7	12		19
17					20	20
n_{x_i}	22	26	18	14	20	n=100

5.9. За даними кореляційної таблиці знайти вибіркове рівняння регресії $\overline{x_y} = ay^2 + by + c.$

а)

$Y \backslash X$	6	30	50	n_{y_j}
1	15			15
3	1	14		15
4		2	18	20
n_{x_i}	16	16	18	*n*=50

б)

$Y \backslash X$	1	9	19	n_{y_j}
0	13			13
2	2	10		12
3	1	1	23	25
n_{x_i}	16	11	23	*n*=50

5.10. У таблиці наведено вибіркові дані про кількість випущених виробів *X* і повних витрат *У* на 20 однотипних підприємствах. Знайти рівняння регресії $\overline{y_x} = \frac{a}{x} + b.$

$У \backslash X$	3	4	5	7	8	10	12	13	14	19	20	24
2		1	1				1			1	1	1
4	1	1			2			1	1			
9	2			1		1	1				1	
18			1		1							

ТЕРМІНОЛОГІЧНИЙ СЛОВНИК

Абсолютний показник - це показник у формі абсолютної величини, яка відображає фізичні властивості, часові та вартісні характеристики соціально-економічних процесів та явищ.

Аксіоматине означення ймовірності: ймовірністю *P(A)* події *A* називається числова функція, що визначена на множині подій і задовольняє такі три умови (аксіоми ймовірності): 1) для довільної події $A \subset \Omega$ справедлива нерівніст $P(A) \geq 0$; 2) $P(\Omega) = 1$; 3) $P(\sum A_k) = \sum P(A_k)$.

Аналітичні групування – це групування, що дають змогу виявити наявність взаємозв'язку між явищами, що вивчаються, та їхніми ознаками.

Асиметричний розподіл — вершина розподілу зміщена. Вона виникає внаслідок обмеженої варіації в одному напрямі або під впливом домінуючої причини розвитку, яка призводить до зміщення центру розподілу.

Атрибутивний ряд розподілу — ряд розподілу, який будується за атрибутивною ознакою.

Багатовимірне групування — це складне групування, якщо воно проводиться за трьома і більше ознаками одночасно.

Багатоступенева вибірка — це вибірка, за якої з генеральної сукупності спочатку вибираються збільшені групи, потім — дрібні і так доти, доки не будуть відібрані ті одиниці, які підлягають спостереженню.

Багатофазна вибірка передбачає збереження постійної одиниці добору на всіх етапах його проведення. При цьому одиниці, які було відібрано на кожній стадії, підлягають обстеженню. На кожній наступній стадії добору програма обстеження розширюється.

Багатофакторні дисперсійні комплекси - методи вимірювання зв'язку результативної ознаки з двома і більше факторними ознаками і перевірки його істотності. Для цього використовують **аналітичні групування,** які дають змогу аналізувати залежність результативної ознаки від кожного з факторів за фіксованих значень інших.

Безповторний добір — це спосіб формування вибірки, за якого одиниця, що потрапила у вибірку, не повертається в сукупність, з якої здійснюється подальший добір.

Біном Ньютона – це формула

$$(a+b)^n = C_n^0 a^n b^0 + C_n^1 a^{n-1} b^1 + C_n^2 a^{n-2} b^2 + \cdots + C_n^{n-1} a^1 b^{n-1} + C_n^n a^0 b^n.$$

Варіанти — окремі значення ознаки, яких вона набуває у варіаційному ряду розподілу, тобто конкретні значення ознаки, що варіює.

Варіаційний ряд розподілу — ряд розподілу, який будується за кількісною ознакою.

Варіація — коливання значення ознаки окремих одиниць сукупності. Або: коливання, різноманітність, змінюваність значення ознаки окремих одиниць сукупності явищ.

Величина інтервалу (інтервал, інтервальна різниця, крок) – це різниця між верхньою та нижньою межами інтервалу.

Верхня межа інтервалу – це максимальне значення ознаки в інтервалі.

Вибірковий лінійний коефіцієнт кореляції$\rho_{\text{в}}$ показує тісноту зв'язку між ознаками *X* та *Y*, а саме:

a) якщо $\rho_{\text{в}} = 0$, то *X* та *Y* не зв'язані кореляційною залежністю;

b) якщо $|\rho_{\text{в}}| = 1$, то *X* та *Y* зв'язані функціональною залежністю;

c) якщо $|\rho_{\text{в}}| < 1$, то між *X* та *Y* існує кореляційний зв'язок, причому він тим тісніший, чим $|\rho_{\text{в}}|$ ближчий до одиниці;

d) знак $\pm$.при коефіцієнті кореляції вказує на напрямок зв'язку: «+» - прямий, «-» - зворотний.

Вибіркова сукупність — це сукупність одиниць, які вибрані навмання для обстеження.

Вибіркова частка (частість) — це питома вага одиниць, які мають певну ознаку у вибірковій сукупності.

Вибіркове спостереження — це несуцільне спостереження, за якого статистичному дослідженню підлягають одиниці сукупності, що вивчаються і вибираються випадковим чином.

Вибіркове спостереження — це обстеження, під час якого реєструється деяка частина одиниць сукупності, що відібрана у випадковому порядку.

Випадковою називається подія, яка при заданому комплексі умов вона може настати і може не настати.

Випадкова величина (ВВ) – це величина, значення якої змінюється від випробування до випробування випадковим чином.

Випадкові похибки — це наслідок випадковості вибору елементів для дослідження і пов'язаних з цим розбіжностей між структурами вибіркової та генеральної сукупностей щодо ознак, які вивчаються.

Випадкові явища — це явища з невизначеним наслідком, які відбуваються при повторному відтворенні одного й того ж випробування (досліду).

Випробування – це експеримент, який можна проводити в однакових умовах (принаймні теоретично) будь-яке число разів.

Відкриті інтервали – це інтервали, які мають тільки одну межу: верхня – у першого; нижня – в останнього.

Відносна похибка показує, на скільки відсотків вибіркова оцінка може відхилятися від параметра генеральної сукупності.

Відносна стандартна похибка середньої — це коефіцієнт варіації вибіркових середніх.

Відносний показник — показник у формі відносної величини — це результат порівняння одного абсолютного показника з іншим; характеризує співвідношення між кількісними характеристиками процесів і явищ, що вивчаються, або кількісне співвідношення різнойменних чи однойменних показників.

Відносні показники варіації — це коефіцієнт варіації лінійний та квадратичний, коефіцієнт осциляції.

Вірогідною називається подія, якщо при заданому комплексі умов вона обов'язково настане.

Генеральна сукупність — це сукупність одиниць, з яких вибирають елементи для обстеження.

Геометрична ймовірність. Нехай множина усіх елементарних наслідків нескінчена, і, як наслідок, займає деяку область *G*. Події А сприяє лише частина $g \in G$. Тоді ймовірність випадкової події А дорівнює відношенню міри *g* до міри *G*: $P(A) = \frac{m(g)}{m(G)}$.

Гістограма – графічне зображення варіаційних рядів для згрупованих даних, інтервального варіаційного ряду розподілу. Гістограмою частот (частостей) називають ступінчату фігуру, яка складається із прямокутників, основи яких є часткові інтервали довжиною $h = x_k - x_{k-1}$, а висоти (щільність частот або частостей) $H_i = \frac{n_i}{h}$ або $H_i = \frac{w_i}{h}$ відповідно. Площа гістограми частот дорівнює об'єму вібірки, а площа гістограми частостей дорівнює 1.

Гранична похибка вибірки — це максимально можлива похибка для взятої ймовірності, яка визначає розмір помилки залежно від того, з якою ймовірністю вона знаходиться.

Групова статистична таблиця – це таблиця, підметом якої є групування одиниць сукупності за однією (кількісною чи атрибутивною) ознакою.

Груповий добір передбачає формування вибіркової сукупності на основі добору груп одиниць з генеральної сукупності.

Дискретна випадкова величина (ДВВ)- це випадкова величина, яка приймає значення деякої числової послідовності (скінченої або нескінченої).

Дискретний варіаційний ряд — розподіл одиниць сукупності за дискретною ознакою.

Дискретний розподіл – це перелік можливих значень ДВВ і відповідних їм імовірностей (в теорії ймовірностей) або перелік можливих значень дискретної ознаки і відповідних їм частот (частостей) (у математичній статистиці).

Дисперсія — це середній квадрат відхилень індивідуальних значень ознаки від їхньої середньої величини. Для ДВВ формули $D(X) = M(x-(Mx)^2)$ або $D(X) = M(x^2) - (M(x))^2$; для НВВ $D(X) = \int_a^b (x - M(X))^2 f(x)dx$ або $D(X) = \int_a^в x^2 f(x)dx - (M(x))^2$.

Двовимірними називають статистичні сукупності, які групуються за двома ознаками.

Добуток двох випадкових подій *А та В* (позначається $A \cdot B$) - це така випадкова подія, яка полягає в тому, що відбуваються одночасно обидві події *А* та *В*. Для залежних випадкових подій *А* та *В* ймовірність $P(A \cdot B) = P(A) \cdot P_A(B) = P(B) \cdot P_B(A)$; для незалежних - $P(A \cdot B) = P(A) \cdot P(B)$.

Екстраполяція — визначення рівнів за межами ряду, що досліджується, тобто продовження ряду на основі виявленої закономірності рівнів за певний термін часу.

Ексцес розподілу — це ступінь зосередженості (згуртованості) елементів сукупності навколо центра розподілу. Характеризує ступінь гостро- або плосковершинності кривої нормального розподілу: якщо Ex>0, крива розподілу більш гостровершинна; при Ex<0 – більш плосковершинна.

Елементарні події (наслідки) - це найпростіші результати випробування, які не можна розкласти на більш прості події.

Емпіричне кореляційне відношення — це корінь квадратний з коефіцієнта детермінації:

Ефективною називають статистичну оцінку, яка (при заданому об'ємі вибірки *n*) має найменшу можливу дисперсію.

Єдиноможливими називають дві або кілька подій, якщо яка-небудь із них обов'язково настане у даному випробуванні.

Завдання статистичного дослідження — отримання узагальнюючих показників та виявлення закономірностей явищ і процесів у конкретних умовах місця та часу.

Закон розподілу (**ряд розподілу**) дискретної випадкової величини (ДВВ) – це відповідність між можливими значеннями ВВ та їхніми ймовірностями.

Закономірність — повторюваність, послідовність і порядок змін в явищах.

Закриті інтервали – це такі інтервали, які мають і нижню, і верхню межу.

Залежними називають випадкові події *A* та *B*, якщо ймовірність появи однієї з них залежить від появи або не появи іншої події. У противному випадку події називають **незалежними***.*

Змістовною (**спроможною, обгрунтованою, слушною**) називають статистичну оцінку, яка при $n\to\infty$ прямує за ймовірністю до параметра, що оцінюється.

Зображення лінії регресії можуть бути різні: табличне, аналітичне, графічне.

Індивідуальний добір передбачає формування вибіркової сукупності на основі добору окремих одиниць генеральної сукупності.

Індивідуальні абсолютні показники — показники, які отримують у процесі статистичного спостереження як результат замірювання, важення, підрахунку та оцінки зацікавленої кількісної ознаки.

Інтегральна теорема Муавра-Лапласа. - Ймовірність того, що в *n* незалежних випробуваннях, в кожному із яких ймовірність появи події дорівнює *p* ($0<p<1$), подія настане не менше m_1 раз і не більше m_2 раз, наближено дорівнює

$P_n(m_1 \le m \le m_2) \approx \Phi(z_2) - \Phi(z_1)$. Тут $\Phi(x) = \frac{1}{\sqrt{2\Pi}} \int_0^x e^{\frac{-x^2}{2}} dx$ - табульована функція; $z_1 = \frac{m_1 - np}{\sqrt{npg}}$; $z_2 = \frac{m_2 - np}{\sqrt{npg}}$.

Інтервал – це значення ознаки, що варіює і лежить у певних межах. Кожний інтервал має свою величину (крок), верхню і нижню межу чи хоч одну з них.

Інтервальна оцінка, тобто довірчий інтервал — це інтервал значень параметра, розрахований за даними вибірки для певної ймовірності. Чим він менший, тим точніша вибіркова оцінка.

Інтервальний варіаційний ряд — ряд, який зображає поточну (неперервну) варіацію ознаки. Якщо інтервали упродовж усього ряду зберігають одну і ту ж величину (крок), то вони називаються рівними, а варіаційний ряд — рядом з рівними інтервалами.

Інтерполяція — розрахунок приблизних рівнів, які знаходяться в середині ряду даних, але з яких-небудь причин невідомі.

Ймовірність події – це міра можливості її появи. *Математичною ймовірністю* випадкової події *А* називається відношення *m* - числа результатів випробування, сприятливих цій події *А*, до *n* - числа всіх єдиноможливих, рівноможливих і попарно несумісних результатів випробування: $P(A) = \frac{m}{n}$.

Ймовірність відхилення частоти появи події від ймовірності появи події в *п* незалежних випробуваннях, що не перевищить додатного числа *ε*, наближено дорівнює подвійній функції Лапласа: $P(\left|\frac{m}{n} - p\right| \le \varepsilon) \approx 2\Phi(\varepsilon\sqrt{\frac{n}{pq}})$, *при* $x = \varepsilon\sqrt{\frac{n}{pq}}$.

Квадратичний коефіцієнт варіації (коефіцієнт варіації) — це відсоткове відношення середнього квадратичного відхилення до середньої величини ознаки. Обчислюється за формулою $V = \frac{\sigma}{\bar{x}}$, (у долях одиниці) або $V = \frac{\sigma}{\bar{x}} \cdot 100\%$ (у відсотках).

Класифікація у математичній статистиці — це систематизований розподіл явищ та об'єктів за певними групами, класами, розрядами на підставі їхнього збігу або різниці. Основою класифікації, як правило, є якісна ознака. У сучасній статистичній практиці розрізняють економічні та соціальні класифікації, які об'єднуються загальним терміном **статистичні класифікації.**

Коефіцієнтом детермінації ρ^2 називається **к**вадрат коефіцієнта кореляції, який показує, яка доля загальної варіації ознаки визначається фактором, що досліджується.

Коефіцієнт регресії *У* на *Х* (*Х* на *У*) показує на скільки одиниць в середньому змінюється змінна *У (Х)* при зміні змінної *Х (У)* на одну одиницю.

Коефіцієнт осциляції — це відсоткове відношення розміру варіації до середньої величини ознаки.

Комбінаторика – один з розділів дискретної математики, присвячений розв’язанню задач вибору і розміщення елементів деякої скінченної множини згідно з певним правилом. Крім теорії ймовірностей, комбінаторика використовується в математичній логіці, теорії чисел, обчислювальній техніці, кібернетиці, теорії автоматів, у деяких задачах економіки, природознавства і т.д.

Комбінації – це різні невпорядковані *m*-елементні підмножини для множини, що містить *n* елементів ($m \leq n$). Кількість комбінацій із *n* елементів по *m* дорівнює $C_n^m = \frac{n!}{m!(n-m)!}$ або

$$C_n^m = \frac{n(n-1)(n-2)...(n-m+1)}{m!}.$$

Комбінаційне групування — це складне групування, в основі якого послідовно скомбіновано дві і більше ознак.

Комбінаційна статистична таблиця – це таблиця, підметом якої є групування одиниць сукупності одночасно за двома і більше ознаками.

Кореляційна залежність є підвидом стохастичної залежності: зі зміною факторної ознаки *х* змінюються групові середні результативної ознаки *у,* тобто замість умовних розподілів порівнюються середні значення цих розподілів.

Критерієм однорідності сукупності вважається квадратичний коефіцієнт варіації (V<0,33).

Критичне значення кореляційного відношення $\eta^2_к$ — це максимально можливе значення η^2, яке може виникнути випадково за відсутності кореляційного зв'язку.

Лінійний коефіцієнт варіації — це відсоткове відношення середнього лінійного відхилення до середньої величини ознаки.

Лінія регресії — це головна характеристика кореляційного зв'язку. Лінія регресії *у* на *х* — це функція, яка зв'язує середні значення ознаки ***у*** зі значенням ознаки *х.* Залежно від форми лінії регресії розрізняють **лінійний** і **нелінійний зв'язки.**

Локальна теорема Муавра-Лапласа. - Ймовірність того, що в *n* незалежних випробуваннях, в кожному з яких ймовірність появи події дорівнює *p* *(0*< *p*<1), подія настане рівно *m* раз (послідовність ролі не відіграє) наближено дорівнює (тим точніше, чим більше *n):* $P_n(m) \approx \frac{1}{\sqrt{npg}} \varphi(x)$, де $\varphi(x) = \frac{1}{\sqrt{2\Pi}}$ e $^{-x^2/2}$, $x = \frac{m-np}{\sqrt{npg}}$,

функція *φ(x)* – табульована.

Макет таблиці — це сукупність горизонтальних рядків і вертикальних граф таблиці без наведення числових даних.

Мала вибірка — це вибіркове спостереження, чисельність одиниць якого не перевищує 30.

Малоймовірна подія — це подія ймовірність якої близька до нуля ($p < 0{,}1$) при великій кількості випробувань *n*, тобто, коли в кожному випробуванні подія з'являється рідко.

Математична статистика – це наука, що використовує теорію ймовірностей для обробки численних одиниць інформації як наслідків експерименту. *Задача* математичної статистики полягає в тому, щоб за обмеженими даними (вибіркою) встановити з визначеним ступенем вірогідності характеристики, які притаманні генеральній сукупності, тобто всьому набору даних, які описують явище, що вивчається.

Математичне сподівання ВВ – це числова характеристика ВВ: для ДВВ $M(X) = x_1p_1 + x_2p_2 + \cdots + x_np_n = \sum_{i=1}^{n} x_ip_i$; для НВВ $M(X) = \int_a^b xf(x)dx.$ Характеризує середнє значення ВВ.

Медіана — у математичній статистиці це значення ознаки у тієї одиниці сукупності, яка знаходиться в середині упорядкованого ряду, тобто це варіанта, яка знаходиться в середині упорядкованого варіаційного ряду і поділяє його на дві рівні частини.

Мета спостереження — одержання достовірної інформації для виявлення закономірностей розвитку явищ і процесів. Кінцевою метою спостереження є підготовка рішень та вжиття відповідних заходів.

Механічна вибірка — це вибірка, за якої добір одиниць здійснюється механічно через певний інтервал. Суть її полягає в тому, що усю (генеральну) сукупність ділять на рівні частини відповідно до обраної ознаки і з кожної такої частини обстежують одну одиницю. За цього способу добору вивчають певне число одиниць сукупності через визначений інтервал (5%, 10% тощо).

Міра асиметрії (скіс**)** — це відносне відхилення, яке характеризує напрям і міру скісності в середині розподілу: за правосторонньої асиметрії *As* >0 (права вітка кривої розподілу довша від лівої), за лівосторонньої — *As* <0 (ліва вітка кривої розподілу довша від правої).

Мода — у математичній статистиці це значення ознаки (варіанти), що найчастіше зустрічається в сукупності (у якої найбільша частота).

Найімовірніше число настання події — це число випробувань, при яких ймовірність появи події найбільша. Число m_0 називається *найімовірнішим числом настання події* в незалежних повторних випробуваннях, якщо ймовірність того, що подія з'явиться в цих випробуваннях m_0 разів, перевищує (або, принаймні, не менше) ймовірності решта можливих наслідків випробування, та знаходиться за формулою $np - q \leq m_0 \leq np + p$, де m_0 – цілий розв'язок подвійної нерівності.

Наслідок або **Елементарна подія**– це найпростіший результат випробування.

Незалежними називають випадкові події *A* та *B*, якщо ймовірність появи однієї з них не залежить від появи або не появи іншої події. У противному випадку події називають **залежними**.

Незсуненою називають статистичну оцінку θ^*, математичне сподівання якої дорівнює параметру, що оцінюється, тобто $M(\theta^*)=\theta$.

Неможливою називають подію, якщо при заданому комплексі умов вона відбутись не може.

Нерівність Чебишова – для довільної ВВ X справедлива нерівність $P(|X - M(X)| > \varepsilon) \leq \frac{D(X)}{\varepsilon^2}$ або в іншому вигляді $P(|X - M(X)| \leq \varepsilon) \geq 1 - \frac{D(X)}{\varepsilon^2}$.

Несумісними називаються дві або кілька подій, якщо настання однієї з них виключає можливість настання кожної з інших у даному випробуванні. У противному випадку події називаються **сумісними**.

Несуцільне спостереження — спостереження, за якого дослідженню підлягає лише окрема частина досліджуваної сукупності.

Нижня межа інтервалу – це мінімальне значення ознаки в інтервалі (довірчому інтервалі).

Обґрунтованою називають статистичну оцінку, як і **змістовною, слушною, спроможною.**

Об'єкт спостереження — це статистична сукупність, в якій відбуваються явища і процеси, що досліджуються.

Одиниця спостереження — складовий (первинний) елемент об'єкта, що є носієм ознак, які підлягають реєстрації.

Одиниця статистичної сукупності — це кожний окремо взятий елемент даної чисельності, якому притаманні певні ознаки.

Одновимірними називають статистичні сукупності, які групуються за однією ознакою.

Однорідні сукупності — це сукупності, елементи яких мають спільні властивості та належать до одного типу, класу.

Основний простір (простір елементарних подій) – це множина всіх можливих наслідків випробування.

Перестановки – це впорядковані множини, які утворені з усіх елементів множини, що містить n елементів. Число перестановок із n елементів обчислюють за формулою $P_n=1\cdot 2\cdot 3\cdot \ldots \cdot n = n!$

Періодичне спостереження — спостереження, за якого реєстрація фактів відбувається за певний інтервал часу (декаду, місяць, квартал, півріччя, рік).

Підмет таблиці – це об’єкт дослідження: перелік елементів сукупності, їх групи, окремі одиниці або часові інтервали.

Повна група подій — це декілька подій в умовах даного випробування, якщо в результаті даного випробування обов'язково відбудеться хоча б одна з них. Або: П.Г.П. - це сукупність подій $A_1, A_2, A_3, \ldots, A_n$, якщо ці події єдиноможливі та несумісні.

Подія - будь-яке явище природи або суспільства, яке при заданому комплексі умов може відбутись або не відбутись. Подія настає в результаті реалізації комплексу умов, тобто в результаті випробування (досліду, експерименту, спостереження).

Показники варіації — це показники, які визначають міру варіації (коливання) окремих значень ознаки від середньої величини.

Полігон (многокутник) розподілу - це множина точок з координатами $(x_i\,; p_i)$, у прямокутній системі координат, які послідовно сполучені відрізками прямих (лінійна діаграма), де x_i - можливі значення випадкової величини *X*, p_i - відповідні їм імовірності.

Правило суми. Якщо деякий об'єкт N можна вибрати n різними способами, а об'єкт М – *m* різними способами, причому будь-який спосіб вибору об'єкта N відрізняється від будь-якого способу вибору об'єкта М, то вибір «*N або М*» можна зробити *n+m* способами.

Правило добутку. Якщо деякий об'єкт N можна вибрати n різними способами, а об'єкт М – *m* різними способами, причому будь-який спосіб вибору об'єкта N відрізняється від будь-якого способу вибору об'єкта М, то вибір пари об'єктів «*N і М*» можна зробити *n*m* способами.

Предмет математичної статистики — кількісна сторона якісно визначених масових явищ та процесів, які відображаються за допомогою статистичних показників.

Присудок таблиці – це показники, що характеризують підмет як об'єкт дослідження.

Програма спостереження — це перелік запитань (або ознак), які підлягають реєстрації в процесі спостереження.

Проста статистична таблиця – це таблиця, підметом якої є перелік об'єктів, елементів сукупності.

Просте групування – це групування одиниць сукупності за однією ознакою.

Простий випадковий добір — це вибірка, за якої добір одиниць (або груп одиниць) для обстеження здійснюється з генеральної сукупності не передбачено, а випадково.

Простір елементарних подій (основний простір) - множина усіх можливих елементарних подій. П.Е.П. може містити скінчену, злічену або незлічену множину елементів.

Протилежними називають дві події А і $\overline{\mathrm{A}}$ (не *А)*, якщо вони єдиноможливі та несумісні. $P(\overline{A}) = 1- P(A)$.

Прямі зв'язки — це зв'язки, за яких зі зростанням значень ознаки-фактора результативна ознака також збільшується, і навпаки: при зменшенні ознаки-фактора результативна ознака також зменшується, тобто направленість зміни результативної ознаки збігається з направленістю зміни ознаки-фактора.

Ранги — порядкові номери одиниць сукупності, упорядковані за зростанням значень ознаки.

Рівень значущості (істотності) – це така ймовірність, за якої ймовірність отримання значення η^2,більшого від критичного (за умови відсутності зв'язку між ознаками), була б достатньо малою.

Рівноможливими називають **д**ві або кілька подій, якщо немає підстав вважати, що яка-небудь з них з’явиться частіше, ніж яка-небудь інша.

Рівняння регресії показує типове в певних умовах співвідношення між розмірами ознаки-фактора і результативної ознаки, тоді як рівняння функціонального зв'язку справедливе лише для кожного окремого випадку.

Розмах варіації (інтервал) — це різниця між максимальним та мінімальним значеннями ознаки, яка варіює.

Розміщеннями з *n* елементів *m* елементів називають різні впорядковані *m*-елементні підмножини множини, що містить *n* елементів ($m \leq n$). Число розміщень обчислюють за формулою $A_n^m = \frac{n!}{(n-m)!} = n(n-1)(n-2)\dots(n-m+1)$.

Ряд розподілу — це упорядкований розподіл одиниць сукупності на групи за певною ознакою, яка варіює.

Середнє квадратичне відхилення — це корінь квадратний з дисперсії.

Середнє лінійне відхилення — це середня арифметична з абсолютних значень відхилень варіант ознаки від їхньої середньої.

Середня арифметична зважена використовується у випадках,коли окремі значення ознаки, за якою розраховується середня величина, можуть повторюватись по кілька разів, тобто в тих випадках, коли розрахунок здійснюється за згрупованими даними.

Середня арифметична проста використовується в тих випадках, коли розрахунок здійснюється за незгрупованими даними.

Середня величина — це узагальнююча характеристика сукупності однотипних явищ за ознакою, що варіює, тобто це узагальнюючий показник, який характеризує типовий рівень ознаки, що варіює, в розрахунку на одиницю однорідної сукупності.

Середня гармонічна — розраховується з відносних значень усередненої ознаки і за формою може бути простою і зваженою.

Серійний добір передбачає вивчення не окремих одиниць сукупності, а їх серій або гнізд.

Симетричний розподіл — рівновіддалені від центра значення ознаки мають однакові частоти.

Система показників — це сукупність взаємопов'язаних показників, які відображають стан та розвиток масових явищ з різних боків.

Система статистичних показників — це сукупність взаємопов'язаних показників, що має однорівневу чи багаторівневу структуру; має на меті розв'язання конкретного статистичного завдання.

Складне групування — це групування за двома і більше ознаками.

Складні події - які можна розкласти на **Елементарні** події.

Слушна статистична оцінка - те ж, що і О**бґрунтована, Змістовна**, С**проможна**.

Спеціальні обстеження — це несуцільні спостереження окремих масових явищ згідно з певною тематикою, що виходить за межі звітності. Можуть бути періодичними або одноразовими.

Сполуки – те ж, що і **комбінації**.

Стандартна (середня) похибка вибірки є середнім квадратичним відхиленням вибіркових оцінок від значення параметра в генеральній сукупності та характеризує середню величину можливих відхилень вибіркової і генеральної середньої.

Статистичною ймовірністю *(відносною частотою)* випадкової події називається відношення числа випробувань, у яких подія *A* настала, до числа всіх випробувань, якщо їх кількість достатньо велика: $W(A) = \frac{\mathbf{m}}{\mathbf{n}}$.

Статистична закономірність — форма виявлення причинного зв'язку який знаходить відображення в послідовності, регулярності, повторюваності подій з достатньо високим ступенем імовірності, якщо причини, що породжують подію, не змінюються або змінюються незначно.

Статистична сукупність — це безліч одиниць, яким властиві масовість, однорідність, певна цілісність, взаємозалежність станів окремих одиниць та наявність варіацій.

Статистична таблиця – це форма найбільш раціонального, наочного та систематизованого викладення результатів зведення і групування матеріалів статистичного спостереження.

Статистичне спостереження — це спланований, систематичний і науково організований збір масових даних про різноманітні явища та процеси.

Статистичний інструментарій — це набір статистичних формулярів, а також інструкцій і роз'яснень щодо проведення статистичного спостереження, реєстрації даних.

Статистичний показник — це кількісна характеристика явищ та процесів в умовах якісного визначення, тобто це міра якісного і кількісного відображення певної властивості явища чи процесу.

Статистичне групування — це основна ланка статистичного зведення, тобто поділ одиниць сукупності на групи, однорідні за певними ознаками.

Статистичні дані — це масові системні кількісні характеристики ознак явищ і процесів, що вивчаються.

Стохастичний зв'язок виявляється зміною умовних розподілів, тобто за цього зв'язку кожному значенню ознаки відповідає певна множина значень ознаки y, які варіюють і утворюють ряд розподілу.

Структурне групування — це групування, за якого відбувається розподіл однорідної сукупності на групи, що характеризують її структуру за певною ознакою, яка варіює (змінюється).

Сума двох випадкових подій A та B (позначається $A+B$) – це така подія, яка відбувається, коли відбувається принаймні одна з цих подій: A або B. Для несумісних подій ймовірність $P(A+B) = P(A) + P(B)$; для сумісних подій - $P(A+B) = P(A) + P(B) – P(A\ B)$.

Сумісними називаються дві або кілька подій, якщо настання однієї з них не виключає можливості настання кожної з інших у даному випробуванні. У противному випадку події називаються **несумісними**.

Суцільне спостереження — спостереження, за якого інформацію отримують про всі одиниці досліджуваної сукупності.

Схема Бернуллі - це сукупність з *n* незалежних випробувань, кожне з яких має тільки два наслідки: поява події *A* (успіх) або непоява $\overline{A}$ (невдача), причому ймовірність того, що подія відбулася *P(A) = p* при кожному випробуванні стала і не залежить від кількості випробувань, а ймовірність протилежної події $\overline{A}$ (непоява події *A*) дорівнює $P(\overline{A})=1-p=q$. Числа *p* і *q* називаються параметрами схеми Бернуллі.

Теорія ймовірностей - це математична наука, яка вивчає ймовірнісні закономірності масових, однорідних, випадкових подій стійкої частоти. Т.Й. вивчає математичні моделі експериментів з випадковими результатами (наслідками).

Типологічне групування — це розподіл якісно неоднорідної сукупності на класи, типи, однорідні групи згідно з правилами наукового групування.

Точкова оцінка — це значення параметра за даними вибірки, яке виражається одним числом, наприклад: вибіркова середня, вибіркова дисперсія тощо.

Умовна ймовірність – це ймовірність події *B*, обчислена при умові появи події *A;* позначають *P(B|A)* або $P_A(B)$.

Упорядкована множина – це множина, що складається з *n* елементів, якщо кожному її елементу можна поставити у відповідність число (номер елемента) від 1 до *n* так, що різним елементам відповідають різні номери.

Факторні ознаки – це групування одиниць сукупності за однією і більше ознаками.

Формула Байєса - це умовна ймовірність *i*-ої гіпотези $P_A(\boldsymbol{B_i})$, якщо подія *A* з'явилась: $p_A(B_i)=\dfrac{p(B_i)p_{B_i}(A)}{\sum_{i=1}^{k}p(B_i)p_B(A)}=\dfrac{p(B_i)p_{B_i}(A)}{p(A)}$.

Формула Бернуллі – обчислює ймовірність того, що в *n* повторних незалежних випробуваннях подія з’явиться рівно *m* раз (малі значення *n* та *m*): $P_n(m)=C_n^m p^m q^{n-m}$.

Формула повної ймовірності. Ймовірність події *A*, яка може відбутися разом із однією з подій $B_1, B_2, \ldots, B_n$, що складають **повну групу** несумісних подій (їх називають гіпотезами), обчислюється як сума добутків ймовірностей кожної гіпотези на умовну ймовірність події *A* при цій гіпотезі:

$$\mathrm{P}(A) = p(B_1)p_{B_1}(A) + p(B_2)p_{B_2}(A) + ... + p(B_k)p_{B_k}(A) = \sum_{i=1}^{k} p(B_i)p_{B_i}(A).$$

Формула Пуассона - ймовірність того, що подія з'явиться в *n* повторних незалежних випробуваннях рівно *m* раз при малих *p* (*p*$< \mathbf{0}$,1) і великих *n:* $P_n(m) \approx \dfrac{\lambda^m \cdot e^{-\lambda}}{m\,!}$, де $\lambda = np$, *e* ≈ 2,71828…

Функціональний зв'язок між явищами характеризується повною відповідністю між причиною і наслідком, факторною і результативною ознакою, тобто за цього зв'язку кожному можливому значенню факторної ознаки *x* відповідає чітко визначене значення результативної ознаки *y.*

Функція розподілу (інтегральна функція розподілу) ВВ – це функція $F(x) = P(X < x)$, що задає ймовірність події $X < x$, тобто ймовірність того, що ВВ *X* буде менша за деяке число *x*.

Частість (частка, відносна частота) — це частота, яка наведена відносною величиною у формі коефіцієнта чи відсотка суми. Сума усіх частостей варіаційного ряду дорівнює одиниці чи 100% відповідно.

Частота — це чисельність окремої варіанти кожної групи варіаційного ряду, тобто число, що показує, як часто зустрічається та або інша варіанта ряду розподілу. Сума усіх частот варіаційного ряду називається його обсягом (об’ємом).

ДОДАТКИ

Додаток 1

Таблиця значень функції $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3832	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	5332	3312	3292	3271	3252	3230	3209	3187	3166	3114
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	11582	1561	1539	1518
1,4	1497	1476	1456	1434	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	0846	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,5400	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034

3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Додаток 2

Таблиця значень функцій $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{x^2}{2}}dx$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	24566	21904	22240
0,6	22575	22907	23237	23865	23891	24215	14537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	0,34134	34375	34614	34850	35083	35314	35543	35769	35993	26214
1,1	36433	36650	36864	37076	37286	37493	37698	37800	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39976	39973	40147
1,3	40320	40490	40658	40824	40989	41149	41309	41466	41021	41774
1,4	41927	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46146	46327
1,8	46407	46485	46566	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	0,47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899

2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49688	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49831	49819	49825	49831	49837	49841	49846	49851	49856	49861
3,0	0,499865	3,3	49952	3,6	49984	3,9	49995	5,0	49999997	
3,1	49903	3,4	49966	3,7	40090	4,0	499968			
3,2	49931	3,5	49977	3,8	49993	4,5	499997	∞	500000	

Додаток 3

Таблиця значень $t_{\gamma}=t(\gamma, n)$.

$n \backslash \gamma$	0,95	0,99	0,999	$n \backslash \gamma$	0,95	0,99	0,099
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	3,461
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				

Додаток 4

Таблиця значень $q = q(\gamma, n)$

$n \setminus \gamma$	0,95	0,99	0,999	$n \setminus \gamma$	0,95	0,99	0,099
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	0,788	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

Додаток 5

Розподіл Пірсона. Таблиця значень χ^2.

α / k	0,99	0,98	0,95	0,90	0,80	0,20	0,10	0,05	0,02	0,01
1	0,00016	0,0063	0,393	0,0158	0,0642	1,642	2,706	3,841	5,412	6,635
2	0,0201	0,0404	0,103	0,211	0,440	3,219	4,605	5,991	7,824	9,210
3	0,115	0,185	0,352	0,584	1,005	4,642	6,251	7,815	9,837	11,341
4	0,297	0,429	0,711	1,064	1,649	5,989	7,779	9,488	11,668	13,277
5	0,554	0,752	1,145	1,610	2,349	7,289	9,236	11,070	13,388	15,086
6	0,872	1,134	1,635	2,204	3,070	8,558	10,645	12,592	15,033	16,812
7	1,239	1,564	2,167	2,833	3,822	9,803	12,017	14,067	16,622	18,475
8	1,646	2,032	2,733	3,490	4,594	11,030	13,362	15,507	18,168	20,090
9	2,088	2,532	3,325	4,168	5,380	12,242	14,684	16,919	19,679	21,666
10	2,558	3,059	3,940	4,865	6,179	13,442	15,987	18,307	21,161	23,209
11	3,053	3,609	4,575	5,578	6,989	14,631	17,275	19,675	22,618	24,725
12	3,571	4,178	5,226	6,304	7,807	15,812	18,549	21,026	24,054	26,217
13	4,107	4,765	5,892	7,042	8,634	16,985	19,812	22,362	25,472	27,688
14	4,660	5,368	6,571	7,790	9,467	18,151	21,064	23,685	26,873	29,141
15	5,229	5,985	7,261	8,547	10,307	19,311	22,307	24,996	28,259	30,578
16	5,812	6,614	7,962	9,312	11,152	20,465	23,542	26,296	29,633	32,000
17	6,408	7,255	8,672	10,085	12,002	21,615	24,769	27,587	30,995	33,409
18	7,015	7,906	9,390	10,865	12,857	22,760	25,989	28,869	32,346	34,805
19	7,633	8,567	10,117	11,651	13,716	23,900	27,204	30,144	33,687	36,191
20	8,260	9,237	10,851	12,443	14,578	25,038	28,412	31,410	35,020	37,566
21	8,897	9,915	11,591	13,240	15,445	26,171	29,615	32,671	36,343	38,932
22	9,542	10,600	12,338	14,041	16,314	27,301	30,813	33,924	37,659	40,289
23	10,196	11,293	13,091	14,848	17,187	28,429	32,007	35,172	38,968	41,638
24	10,856	11,992	13,848	15,659	18,062	29,553	33,196	36,415	40,270	42,980
25	11,524	12,697	14,611	16,473	18,940	30,675	34,382	37,652	41,566	44,314
26	12,198	13,409	15,379	17,292	19,820	31,795	35,563	38,885	42,856	45,642
27	12,879	14,125	16,151	18,114	20,703	32,912	36,741	40,112	44,140	46,963
28	13,565	14,847	16,928	18,939	21,588	34,027	37,916	41,337	45,419	48,278
29	14,256	15,574	17,708	19,768	22,475	35,139	39,087	42,557	46,693	49,588
30	14,953	16,306	18,493	20,599	23,364	36,250	40,256	43,773	47,962	50,892

Додаток 6

Критичні точки розподілу Стьюдента

k	Двостороння критична область (α)							
	0,2	0,1	0,05	0,02	0,01	0,005	0,002	0,001
1	3,08	6,31	12,71	31,82	63.7	127,32	318,30	636,61
2	1,89	2,92	4,30	6,97	9,92	14,09	22,33	31,60
3	1,64	2,35	3,18	4,54	5,84	4,75	10,22	12,92
4	1,53	2,13	2,78	3,75	4,00	5,60	7,17	8,61
5	1,48	2,02	2,57	3,36	4,03	4,77	5,89	6,87
6	1,44	1,94	2,45	3,14	3,71	4,32	5,21	5,96
7	1,41	1,89	2,36	3,00	3,50	4,03	4,79	5,41
8	1,40	1,86	2,31	2,90	3,36	3,83	4,50	5,04
9	1,38	1,83	2,24	2,82	3,25	3,69	4,30	4,78
10	1,37	1,81	2,23	2,76	3,17	3,58	4,14	4,59
11	1,36	1,80	2,29	2,72	3,11	3,50	4,02	4,44
12	1,36	1,78	2,18	2,68	3,05	3,43	3,93	4,32
13	1,35	1,77	2,16	2,65	3,01	3,37	3,85	4,22
14	1,34	1,76	2,14	2,62	2,98	3,33	3,79	4,14
15	1,34	1,75	2,13	2,60	2,95	3,29	3,73	4,07
16	1,34	1,75	2,12	2,58	2,92	3,25	3,69	4,02
17	1,33	1,74	2,12	2,57	2,90	3,22	3,65	3,97
18	1,33	1,73	2,10	2,55	2,88	3,20	3,61	3,92
19	1,33	1,73	2,09	2,54	2,86	3,17	3,58	3,88
20	1,33	1,72	2,09	2,53	2,85	3,15	3,55	3,85
21	1,32	1,72	2,08	2,52	2,83	3,14	3,53	3,82
22	1,32	1,72	2,07	2,54	2,82	3,12	3,51	3,79
23	1,32	1,71	2,07	2,50	2,81	3,10	3,48	3,77
24	1,32	1,71	2,06	2,49	2,80	3,09	3,47	3,75
25	1,32	1,71	2,06	2,49	2,79	3,08	3,45	3,73
26	1,32	1,71	2,06	2,48	2,78	3,07	3,44	3,71
27	1,31	1,70	2,05	2,47	2,77	3,06	3,42	3,69
28	1,31	1,70	2,05	2,47	2,76	3,05	3,41	3,67
29	1,31	1,70	2,05	2,46	2,76	3,04	3,40	3,66
30	1,31	1,70	2,04	2,46	2,75	3,03	3,39	3,65
40	1,30	1,68	2,02	2,42	2,70	2,97	3,31	3,55
60	1,30	1,67	2,00	2,39	2,66	2,91	3,23	3,46
120	1,29	1,66	1,98	2,36	2,62	2,85	3,16	3,37
∞	1,28	1,64	1,96	2,33	2,58	2,81	3,09	3,29
Одностороння критична область (α)								
	0,1	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005

Додаток 7

Критичні точки розподілуFФішера-Снедекора

(k_1 – число ступенів вільностей більшої дисперсії,

k_2 – число ступенів вільностей меншої дисперсії)

Критические точки распределения F Фишера — Снедекора

(k_1—число степеней свободы большей дисперсии,
k_2—число степеней свободы меньшей дисперсии)

Уровень значимости $\alpha = 0,01$

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	90,17	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	30,81	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	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	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	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	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	3,61	3,55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Уровень значимости $\alpha = 0,05$

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	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	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	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	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	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	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

Додаток 8

Методичні рекомендації

до розв'язання типових задач

Задача 1. У магазин надходить продукція з трьох підприємств у кількості 20, 50, 30 виробів відповідно. Ймовірності виготовлення неякісного виробу для кожного підприємства відповідно дорівнюють 0,01; 0,04; 0,03. Навмання вибраний виріб виявився неякісним. Якому підприємству, ймовірніше всього, належить цей виріб?

Розв'язання. Подія А - вибрано неякісний виріб. Гіпотези H_1, H_2, H_3 - це вибір виробу із продукції відповідного підприємства. Ймовірності цих подій дорівнюють:

$$P(H_1)=\frac{20}{100}=0.2;\qquad P(H_2)=\frac{50}{100}=0.5;\qquad P(H_3)=\frac{30}{100}=0.3.$$

Використовуючи формулу повної ймовірності знаходимо:

$$\mathrm{P}(A)=p(B_1)p_{B_1}(A)+p(B_2)p_{B_2}(A)+...+p(B_k)p_{B_k}(A)=\sum_{i=1}^{k}p(B_i)p_{B_i}(A)$$

P(A)= 0,2·0,01 + 0,5·0,04 + 0,3·0,03 = 0,031.

За формулами Байєса знаходимо умовні ймовірності гіпотез:

$$P_A(H_1)=\frac{P(H_1)\cdot P_{H_1}(A)}{P(A)}=\frac{0.002}{0.031}=\frac{2}{31};\qquad P_A(H_1)=\frac{P(H_1)\cdot P_{H_1}(A)}{P(A)}=\frac{0.002}{0.031}=\frac{2}{31};$$

$$P_A(H_3)=\frac{P(H_3)\cdot P_{H_3}(A)}{P(A)}=\frac{0.03}{0.031}=\frac{30}{31}.$$

Оскільки *max* $\left\{\frac{2}{31};\frac{20}{31};\frac{30}{31}\right\}=\frac{30}{31}$, то ймовірніше всього, що вибраний неякісний виріб належить третьому підприємству.

Задача 2. Ймовірність присутності студента на лекції дорівнює 0,8. Знайти ймовірність того, що із 100 студентів на лекції буде: а) 75 студентів; б) не менше 90 студентів.

Розв’язання.

а) за умовою n = 100, p = 0,8; q = 0,2; k = 75. Використовуємо локальну теорему Лапласа:

$$P_n(k) \approx \frac{\varphi(x)}{\sqrt{npg}}$$, де $x = \frac{k-np}{\sqrt{npg}}$ Знайдемо значення *x*:

$$\boldsymbol{x = \frac{75-80}{\sqrt{100 \cdot 0.8 \cdot 0.2}} = \frac{-5}{4} = -1.25.}$$

За таблицею (табл. 1) значень функції φ(x) знаходимо φ(1,25)= 0,1825. Оскільки φ (-x)= φ (x), то φ (-1,25) = 0,1825. Шукана ймовірність $\boldsymbol{P_{100}(75) \approx \frac{0{,}1825}{4} \approx 0{,}046}$.

б) використовуємо інтегральну теорему Лапласа:

$$\boldsymbol{P_n(k_1;k_2) \approx \Phi(x^{''}) - \Phi(x^{'}),}$$

де $x^{''} = \frac{k_2 - np}{\sqrt{npg}}$; $x^{'} = \frac{k_1 - np}{\sqrt{npg}}$.За умовою k_I = 90, k_2 = 100. Знаходимо x' і x":

$$x^{'} = \frac{90-80}{4} = \frac{10}{4} = 2.5; \quad x^{'} = \frac{100-800}{4} = \frac{20}{4} = 5.$$

За таблицею значень функції Лапласа Φ(x) знаходимо Φ(2,5)=0,4938; Φ(5)=0,5. Шукана ймовірність:

$$D_{100}(90;100) \approx 0.5 - 0.4938 = 0.0062.$$

Задача 3. Знайти: 1) математичне сподівання; 2) дисперсію; 3)середнє квадратичне відхилення дискретної випадкової величини X по даному закону її розподілу:

x_i	26	28	30	32
p_i	0.1	0.2	0.4	0.3

Розв’язання:

1) $M(X) = \sum_{i=1}^{n} x_i p_i = 26 \cdot 0.1 + 28 \cdot 0.2 + 30 \cdot 0.3 = 29.8$

2) Знайдемо дисперсію: D(*X*)=M(*X* – M(*X*))^2=$\sum_{i=1}^{n}$ $(x_i - a)^2 \cdot p_i$;

D(*X*)= $(26\text{-}29{,}8)^2 \cdot 0{,}1 + (28\text{-}29{,}8)^2 \cdot 0{,}2 + (30\text{-}29{,}8)^2 \cdot 0{,}4 + (32\text{-}29{,}8)^2 \cdot 0{,}3$ = $=(-3{,}8)^2 \cdot 0{,}1 + (-1{,}8)^2 \cdot 0{,}2 + (0{,}2)^2 \cdot 0{,}4 + (2{,}22)^2 \cdot 0{,}3 = 3{,}56$.

Знайдемо дисперсію за другою формулою: D(X)=M(X^2)–M^2(X)

$$M(X^2) = \sum_{i=1}^{n} x_i^2 p_i = 26^2 \cdot 0.1 + 28^2 \cdot 0.2 + 30^2 \cdot 0.4 + 32^2 \cdot 0.3 = 891.6$$

$$D(X) = 891{,}6 - (29{,}8)^2 = 891{,}6 - 888{,}04 = 3{,}56$$

Знайдемо середньоквадратичне (стандартне) відхилення:

$$\sigma(X) = \sqrt{D(X)} = \sqrt{3.56} = 1.89.$$

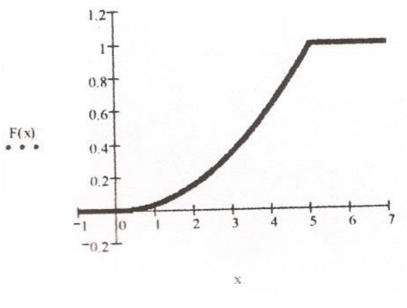
Задача 4. Випадкова величина *X* задана інтегральною функцією *F(X).* Знайти: I) диференціальну функцію розподілу; 2) математичне сподівання і дисперсію. Побудувати графіки інтегральної і диференціальної функцій розподілу ймовірностей випадкової величини *X.*

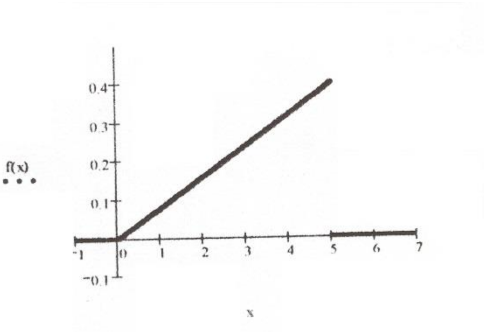
$$F(X) = \begin{cases} 0, & \text{при } x \le 0, \\ \dfrac{x^2}{25}, & \text{при } 0 < x < 5, \\ 1, & \text{при } x > 5. \end{cases}$$

Розв'язання. Знайдемо диференціальну функцію розподілу f(x):

$$f(x) = F'(x) = \begin{cases} 0, & x < 0; \\ \dfrac{2}{25}x, & 0 < x < 5, \\ 0, & x > 5. \end{cases}$$

Будуємо графіки інтегральної та диференціальної функцій:





За формулою $M(X) = \int_a^b xf(x)dx$ знаходимо математичне сподівання:

$$M(X) = \int_0^5 x \cdot \frac{2x}{25} dx = \frac{2}{25} \cdot \frac{x^3}{3}\Bigg|_0^5 = \frac{10}{3}.$$

Дисперсію знаходимо за формулою $D(X) = \int_a^b x^2 f(x)dx - M^2(X),$

тобто

$$D(X) = \int_0^5 x^2 \cdot \frac{2x}{25}dx - \left(\frac{10}{3}\right)^2 = \frac{2}{25}\cdot\frac{x^4}{4}\Big|_0^5 - \frac{100}{9} = \frac{25}{18}.$$

Середнє квадратичне відхилення дорівнює:

$$\sigma(X) = \sqrt{D(X)} = \sqrt{\frac{25}{18}} = \frac{5}{3\sqrt{2}} = \frac{5\sqrt{2}}{6} \approx 1.17.$$

Задача 5. Задані математичне сподівання а= 30 і середнє квадратичне відхилення σ= 2 нормально розподіленої випадкової величини X. Знайти: 1) ймовірність того, що X прийме значення, що належить інтервалу (28,39); 2) ймовірність того, що абсолютна величина відхилення $|X - a|$ виявиться меншою ε= 3.

Розв'язання. 1) Використовуємо формулу:

$$P(\alpha \le X \le \beta) = \Phi(\frac{\beta - a}{\sigma}) - \Phi(\frac{\alpha - a}{\sigma}).$$

$$\boldsymbol{P(28 \le X \le 39) = \Phi(\frac{39-30}{2}) - \Phi(\frac{28-30}{2}) = \Phi(4{,}5) - \Phi(-1) = \Phi(4{,}5) + \Phi(1) =}$$
$$\boldsymbol{= 0{,}5 + 0{,}3413 = 0{,}8413.}$$

2) Використовуємо формулу: $P(|X - a| \le \varepsilon) = 2\Phi(\frac{\varepsilon}{\sigma}).$

$$\boldsymbol{P(|X - 30| \le 3) = 2\Phi(\frac{3}{2}) = 2\Phi(1{,}5) = 2\cdot 0{,}4332 = 0{,}8664.}$$

Задача 6. Задані середнє квадратичне підхилення *σ* = 3 нррмально розподіленої випадкової величини *X*. вибіркова середня *X* = 3,24, об'єм вибірки *п* = 36. Знайти довірчі інтервали для оцінки невідомого математичного сподівання ***а*** з заданою надійністю γ = 0,95.

Розв'язання. 2*Ф*(*u*) = 0,95, звідки *Ф*(u) = $\frac{0.95}{2}$ = 0,475. За таблицями (табл. 2) функції Лапласа знаходимо значення u = 1,96. Далі знаходимо точність оцінки: $\varepsilon = \frac{u\cdot\sigma}{\sqrt{n}} = \frac{1.96\cdot 3}{\sqrt{36}} = 0.98$. Знаходимо границі інтервалу: $\boldsymbol{X - \varepsilon = 3{,}24 - 0{,}98 = 2{,}26;}$
$\boldsymbol{X + \varepsilon = 3{,}24 + 0{,}98 = 4{,}22.}$

Одержимо довірчий інтервал: (2,26; 4,22).

Задача 7. Знайти методом добутків вибіркову середню та вибіркову дисперсію заданої вибірки.

x_i	123	128	133	138
m_i	8	42	30	20

Розв'язання. Знайдемо об'єм вибірки:

$$n = \sum m_i = 8 + 42 + 30 + 20 = 100.$$

Крок між варіантами h = 5. Як «хибний нуль» виберемо варіанту з найбільшою частотою (мода): C = 128.

Побудуємо зведений варіаційний ряд, варіанти якого обчислимо за формулами:

$$y_i = \frac{x_i - C}{h} = \frac{x_i - 128}{5}$$

y_i	-1	0	1	2
m_i	8	42	30	20

Обчислимо вибіркову середню та вибіркову дисперсію зведеного ряду:

$$\overline{Y} = \frac{\sum_{i=1}^{k} y_i \cdot m_i}{n} = \frac{-1\cdot 8 + 0\cdot 42 + 1\cdot 30 + 2\cdot 20}{100} = \frac{-8+0+30+40}{100} = 0{,}62;$$

$$\overline{Y^2} = \frac{\sum_{i=1}^{k} y_i^2 \cdot m_i}{n} = \frac{(-1)^2\cdot 8 + 0^2\cdot 41 + 1^2\cdot 30 + 2^2\cdot 20}{100} = \frac{8+0+30+80}{100} = 1{,}18;$$

D(Y)=$\overline{Y^2} - \overline{Y}^2$= 1,18–(0,62)^2 = 0,7956; $\sigma(Y) = \sqrt{D(Y)} = \sqrt{0{,}7956} \approx 0{,}89196$.

Перейдемо до варіанти *X*: $\overline{x} = h\cdot\overline{y} + C = 5\overline{y} + 128$, тоді: $\overline{x} = 5\cdot 0{,}62 + 128 = 131{,}1$; $D(X) = h^2\cdot D(Y) = 25\cdot 0{,}7956 = 19{,}89$; $\sigma(X) = h\sigma(Y) \approx 5\cdot 0{,}89196 \approx 4{,}4598$.

Задача 8. Побудувати емпіричну функцію за даним розподілом вибірки:

x_i	3	5	12

m_i	4	6	10

Розв'язання. Об'єм вибірки *п* = 4 + 6 + 10 = 20. За означенням $F'(x) = \frac{m_x}{n}$, де m_x - сума частот тих варіант $X < x$.

Найменша варіанта *х* = 3, отже F'(x)= 0 при x $\leq$3.

Значення X<5 спостерігається 4 рази. Отже,

$$F'(x) = \frac{4}{20} = \frac{1}{5} \text{ при } 3<x\leq 5.$$

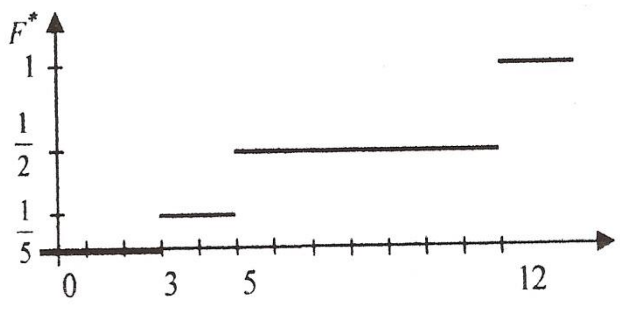
X<l2 спостерігається 4 + 6 = 10 раз, тому

$$F'(x) = \frac{10}{20} = \frac{1}{2} \quad \text{при } 5 < x \leq 12.$$

Оскільки *X* = 12 найбільша варіанта, то F' (x) = 1 при x> 12. Таким чином емпірична функція має вигляд:

$$F'(x) = \begin{cases} \mathbf{0}, & \textbf{при } x \leq 3; \\ \frac{1}{5}, & \textbf{при } 33 < x \leq 5; \\ \frac{1}{2}, & \textbf{при } 5 < x \leq 12 \\ \mathbf{1}, & \textbf{при } x > 12. \end{cases}$$

Зобразимо графік функції розподілу:



Задача 9. За даними двох незалежних вибірок об'єму n_1 = 10 та n_2 = 15 із нормальних сукупностей *X* та *У* знайдені виправлені вибіркові дисперсії S_1^2 = 15,42 та S_2^2 = 11,36. При рівні значущості α= 0,05 перевірити гіпотезу H_0: D(X) = D(Y) при альтернативній H_1: D(X) > D(Y).

Розв'язання. Знайдемо відношення більшої виправленої дисперсії до меншої: $F = \frac{S_1^2}{S_2^2} = \frac{15,42}{11,36} = 1,36.$ За таблицями розподілу

Фішера (табл. 7), за рівнем значущості α = 0,05 і числами ступенів вільності $k_{1,} = n_1 - 1 = 10 - 1 = 9$ і $k_2 = n_2 - 1 = 15 - 1 = 14$ знаходимо критичну точку $F_{кр.} = F_{табл.}(0,05;\ 9;\ 14) = 2,65$. Враховуючи, що $F<F_{кр.}$, робимо висновок, що немає підстав відкинути нульову гіпотезу про рівність генеральних дисперсій.

Задача 10. Знайти вибіркове рівняння прямої лінії регресії *У* на *Х* за даними кореляційної таблиці; перевірити значущість параметрів і тісноту кореляційного зв'язку.

y_i / x_i	2,1	2,2	2,3	2,4	m_x
-1,1		2		10	12
-1		4	2	9	15
-0,9	2	12	3		17
-0,8	21	14			35
-0,7	1				1
m_y	24	32	5	19	80

Розв'язання: Знайдемо середні значення величин *Х* і *Y:*

$$\overline{X} = \frac{\sum x\cdot m_x}{n} = \frac{-1,1\cdot 12 - 1\cdot 15 - 0,9\cdot 17 - 0,8\cdot 35 - 0,7\cdot 1}{80} = \frac{-72,2}{80} = -0,9025;$$

$$\overline{Y} = \frac{\sum y\cdot m_y}{n} = \frac{2,1\cdot 24 + 2,2\cdot 32 + 2,3\cdot 5 + 2,4\cdot 19}{80} = \frac{177,9}{80} = 2,22375.$$

Знайдемо середні квадратичні відхилення величин *Х* і *Y*, для цього:

$$\overline{X^2} = \frac{\sum x^2\cdot m_x}{n} = \frac{(-1,1)^2\cdot 12 + (-1)^2\cdot 15 + (-0,9)^2\cdot 17 + (-0,8)^2\cdot 35 + (-0,7)^2\cdot 1}{80} = \frac{66,18}{80} = 0,82725;$$

$$D(X) = \overline{X^2} - (\overline{X})^2 = 0,82725 - (-0,9025)^2 = 0.012744;$$

$$\sigma_x = \sqrt{D_x} = \sqrt{0,1274} = 0,112888$$

$$\overline{Y^2} = \frac{\sum y^2\cdot m_y}{n} = \frac{(2,1)^2\cdot 24 + (2,2)^2\cdot 32 + (2,3)^2\cdot 5 + (2,4)^2\cdot 19}{80} = \frac{396,61}{80} = 4,957625$$

$$D_y = \overline{Y^2} - (\overline{Y})^2 = 4{,}957625 - (2{,}22375)^2 = 0{,}012561.$$

$$\sigma_y = \sqrt{D_y} = \sqrt{0{,}012561} = 0{,}112076.$$

Знайдемо коваріацію між величинами *X* і *Y*:

$$\overline{X \cdot Y} = \frac{\sum x \cdot y \cdot m_{xy}}{n} = \frac{1}{80} \cdot (-1{,}1 \cdot (2{,}2 + 2{,}4 \cdot 10) - 1 \cdot (2{,}2 \cdot 4 + 2{,}3 \cdot 2 + 2{,}4 \cdot 9) -$$

$$- 0{,}9 \cdot (2{,}1 \cdot 2 + 2{,}2 \cdot 12 + 2{,}3 \cdot 3) - 0{,}8 \cdot (2{,}1 \cdot 21 + 2{,}2 \cdot 14) - 0{,}7 \cdot 2{,}1 \cdot 1) = \frac{-161{,}38}{80} = -2{,}01725;$$

$$K(X,Y) = \overline{X \cdot Y} - \overline{X} \cdot \overline{Y} = -2{,}01725 - (-0{,}9125) \cdot 2{,}22375 = -0{,}0103156.$$

Знайдемо коефіцієнт кореляції між функціями *X* і *Y*:

$$r = \frac{K(X,Y)}{\sigma_x \cdot \sigma_y} = \frac{-0{,}0103156}{0{,}112888 \cdot 0{,}112076} = -0{,}815335.$$

Знайдемо параметри лінійного рівняння $y = a \cdot x + b$ регресії *Y* на *X*.

Коефіцієнт рівняння регресії:

$$a = r \cdot \frac{\sigma_y}{\sigma_x} = -0{,}815335 \cdot \frac{0{,}112076}{0{,}112888} = -0{,}8095.$$

Вільний член рівняння регресії:

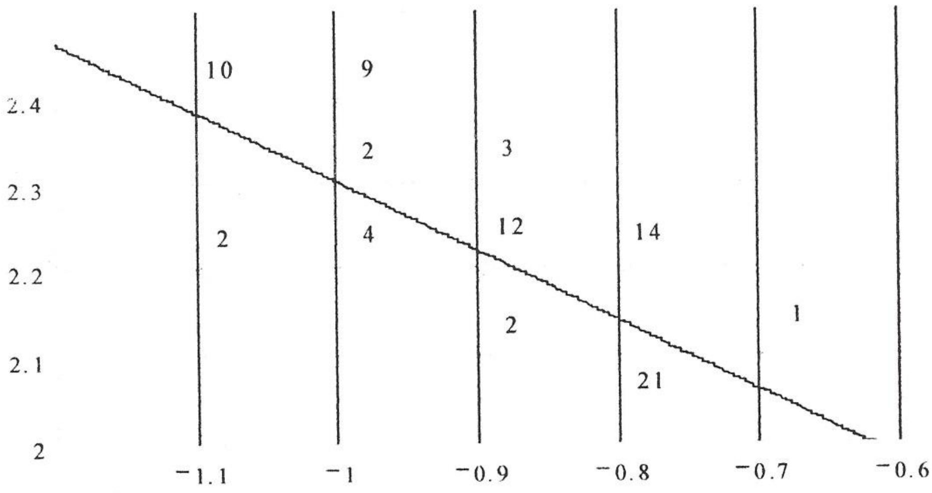
$$b = \overline{Y} - a \cdot \overline{X} = 2{,}22375 - (-0{,}8095) \cdot ((-0{,}9025) = 1{,}4932.$$

Одержимо рівняння прямої лінії регресії:

$$y = -0{,}8095 \cdot x + 1{,}4932.$$

Оскільки коефіцієнт кореляції *r* і коефіцієнт рівняння ресесії *a* від’ємний, то зв'язок між *X* і *Y* обернений, а близькість $|r|$ до 1 свідчить про тісний зв'язок між *X* і *Y.*

Побудуємо графік лінії регресії та експериментальних точок вибірки:


10
9
2.4
2
3
2.3
2
4
12
14
2.2
2
1
2.1
21
2
−1.1
−1
−0.9
−0.8
−0.7
−0.6

Додаток 9

Зразок завдань комплексної контрольної роботи з теорії ймовірностей та математичної статистики

Варіант №1

1. У ящику 10 деталей, з яких 6 пофарбовані. Навмання обирається 4 деталі. Знайти ймовірність того, що вийняті деталі пофарбовані.

2. Завод випускає кухонні набори білого та синього кольорів, що виготовляються двома цехами. Перший цех виробляє 35 % всієї продукції, серед яких 40 % наборів синього кольору. У продукції другого цеху 55 % синіх наборів. Яка ймовірність того, що навмання, вибраний набір синього кольору?

3. Неперервна випадкова величина X задана функцією розподілу $F(x)$. Визначити: а) математичне сподівання випадкової величини *X*; б) ймовірність того, що випадкова величина X матиме значення, що належить інтервалу (α,β).

$$F(x)=\begin{cases}0, & x\le -2;\\ \dfrac{x+2}{4}, & -2<x\le 2;\\ 1, & x>2.\end{cases}$$

$$\alpha=0{,}5;\quad \beta=4{,}5.$$

4. Оцінити з надійністю γ=0,95 математичне сподівання a і середнє квадратичне відхилення σ нормально розподіленої генеральної сукупності за допомогою надійного інтервалу, якщо заданий такий розподіл вибірки:

x_i	30	31	32	33	34	26
n_i	2	7	10	1	8	2

Варіант №2

1.У лотереї 100 білетів, серед яких 5 таких, що виграють. Обчислити ймовірність виграшу для того, хто має 2 білета.

2. На двох полицях стоять книжки: на першій – 15 книжок українською та 7 російською мовами, на другій – відповідно 10 і 8 книжок. З першої полиці навмання перекладено книжку на другу полицю. Яка ймовірність того, що навмання вибрана з другої полиці книжка виявиться українською?

3. Неперервна випадкова величина X задана функцією розподілу $F(x)$. Визначити: а) математичне сподівання випадкової величини *X*; б) ймовірність того, що випадкова величина X матиме значення, що належить інтервалу (α,β).

$$F(x)=\begin{cases}0, & x\le 0{,}5;\\ \dfrac{2x-1}{2}, & 0{,}5<x\le 1{,}5;\\ 1, & x>1{,}5.\end{cases}$$

$\alpha=1{,}0;\quad \beta=2{,}3.$

4. Оцінити з надійністю γ=0,95 математичне сподівання a і середнє квадратичне відхилення σ нормально розподіленої генеральної сукупності за допомогою надійного інтервалу, якщо заданий такий розподіл вибірки:

X_i	17	21	2	27
n_i	2	4	3	1

Варіант №3

1. Студент знає відповіді на 20 питань з 25. Знайти ймовірність того, що він зможе відповісти на 3 запитання, що задав екзаменатор.

2. 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 виготовлена продукція вищої якості?

3. Неперервна випадкова величина X задана функцією розподілу $F(x)$. Визначити: а) математичне сподівання випадкової величини X; б) ймовірність того, що випадкова величина X матиме значення, що належить інтервалу (α,β).

$$F(x)=\begin{cases}0, & x\le -0{,}5;\\ \dfrac{2x+1}{3}, & -0{,}5<x\le 1;\\ 1, & x>1.\end{cases}$$

$\alpha=0{,}2;\quad \beta=2{,}6.$

4. Оцінити з надійністю γ=0,95 математичне сподівання a і середнє квадратичне відхилення σ нормально розподіленої генеральної сукупності за допомогою надійного інтервалу, якщо заданий такий розподіл вибірки:

x_i	5	7	9	11
n_i	15	26	25	34

Варіант №4

1. У ящику 20 деталей, з яких 5 пофарбовані. Навмання обирається 4 деталі.Знайти ймовірність того, що 3 з них пофарбовані.

2. До каси підприємства надійшли банкноти в пачках з двох банків: 50 пачок з першого і 70 – з другого. Ймовірність допущення помилки касирами

першого банку становить 0,15 %, другого – 0,2 %. Яка ймовірність того, що навмання вибрану пачку сформовано без помилок?

3. Неперервна випадкова величина X задана функцією розподілу $F(x)$. Визначити: а) математичне сподівання випадкової величини *X*; б) ймовірність того, що випадкова величина X матиме значення, що належить інтервалу (α,β).

$$F(x)=\begin{cases}0, & x\le 0{,}25;\\ \dfrac{4x-1}{16}, & 0{,}25<x\le 4{,}25;\\ 1, & x>4{,}25.\end{cases}$$

$$\alpha=1{,}3;\quad \beta=4{,}8.$$

4. Оцінити з надійністю γ=0,95 математичне сподівання a і середнє квадратичне відхилення σ нормально розподіленої генеральної сукупності за допомогою надійного інтервалу, якщо заданий такий розподіл вибірки:

x_i	2	5	7	10
n_i	16	12	8	14

Варіант №5

1. Серед 10 виробів знаходяться 2 браковані. Знайти ймовірність того, що з 3 навмання взятих виробів, 1 бракований.

2. Податкові інспектори виконують перевірку діяльності підприємств: перший обслуговує 40 підприємств, серед яких 25 % не мають заборгованостей, другий – 60 підприємств, з них 40 % – без заборгованостей. Яка ймовірність того, що навмання обране підприємство не має заборгованості?

3. Неперервна випадкова величина X задана функцією розподілу $F(x)$. Визначити: а) математичне сподівання випадкової величини *X*; б) ймовірність того, що випадкова величина X матиме значення, що належить інтервалу (α,β).

$$F(x)=\begin{cases}0, & x\le 2;\\ \dfrac{x}{2}-1, & 2<x\le 4;\\ 1, & x>4.\end{cases}$$

$$\alpha=2{,}5;\quad \beta=4{,}0.$$

4. Оцінити з надійністю γ=0,95 математичне сподівання a і середнє квадратичне відхилення σ нормально розподіленої генеральної сукупності за допомогою надійного інтервалу, якщо заданий такий розподіл вибірки:

X_i	15	20	23	27

n_i	1	3	4	2

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Агапов Г.И. Задачник по теории вероятностей [Текст]: Уч. пособие для студ. Втузов. / Г.И. Агапов. – М.: ВШ, 1986. – 80 с.
2. Агекян Т.А. Теория вероятностей для астрономов и физиков [Текст]: Уч. пособие для студ. унив-в. / Т.А. Агекян. – М.: Наука, ГРФМЛ, 1974. – 264 с.
3. Барковський В.В. Теорія ймовірностей та математична статистика [Текст]: Навч. посібник для студ. екон. спец. / В.В.Барковський, Н.В. Барковська, О.К. Лопатін .– К.: ЦУЛ, 2002. – 448 с.
4. Валько Н.В. Методичні рекомендації та завдання для вивчення змістового модуля „Математична статистика” [Текст]: методичні рекомендації / Н.В.Валько, О.Г. Савченко, Л.В.Кузьмич, Г.М.Кавун. – Херсон: ХДАУ, 2015. – 58 с.
5. Венецкий И.Г. Теория вероятностей и математическая статистика [Текст]. / И.Г. Венецкий, Г.С. Кильдишев. – М.: Статистика, 1975. - 264 с.
6. Вентцель Е.С. Теория вероятностей [Текст]. / Е.С. Венцель. – М.: Наука, 1969. – 576 с.
7. Виленкин Н.Я. Комбинаторика [Текст]. / Н.Я. Виленкин. – М.: Наука, 1969. – 328 с.
8. Виленкин Н.Я. Задачник-практикум по теории верояностей с элементами комбинаторики и математической статистики [Текст]: Учеб. пос. для студ. пед. ин-в. / Н.Я. Виленкин, В.Г. Потапов. – М.: Просвещение, 1979. - 112 с.

9. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике [Текст]: Учеб. пос. для студ. вузов. / В.Е. Гмурман. – М.: ВШ, 2001. -. 400 с.
10. Гмурман В.Е. Теория вероятностей и математическая статистика [Текст]: Учеб. пос. для студ. вузов. / В.Е. Гмурман. – М.: ВШ, 1977. – 480 с.
11. Гнеденко Б.В. Курс теорії ймовірностей [Текст]. / Б.В. Гнеденко. – К.: Радянська школа, 1950. – 360 с.
12. Гнеденко Б.В. Элементарное введение в теорию вероятностей [Текст] / Б.В. Гнеденко, А.Я.Хинчин. – М.: Наука, ГРФМЛ, 1982. – 160 с.
13. Горбань С.Ф. Теория вероятностей и математическая статистика [Текст]: Уч. пособие / С.Ф. Горбань, Н.В. Снижко – К.: МАУП, 1999. – 168 с.
14. Єжов І.І. Елементи комбінаторики [Текст] / І. І. Єжов, А.В. Скороход, М.Й. Ядренко. - К. : ВШ, 1972. - 84 с.
15. Жалдак М. І. Елементи стохастики з комп'ютерною підтримкою [Текст] : посіб. для вчителів. / М. І. Жалдак, Г. О. Михалін. - К. : НПУ ім. М.П.Драгоманова, 2000. - 70 с.
16. Жалдак М. І. Початки теорії ймовірностей [Текст]: Пос. для самоосвіти вчителів. / М. І. Жалдак. – К.: РШ, 1978. – 143 с.
17. Жалдак М. І. Теорія ймовірностей і математична статистика з елементами інформаційної технології [Текст]: навч. посібник для студ. фіз.-мат. фак. пед. ін-тів / М. І. Жалдак [та ін.]. – К.: ВШ, 1995. – 351 с.
18. Жалдак М.И. Теория вероятностей с элемнтами информатики: Практикум: [Текст]: Учеб. пособие / М.И. Жалдак, А.Н.Квитко. Под общ. ред. М.И.Ядренко. – К.: Выща школа, 1989. – 263 с.

19. Жлуктенко В.І. Теорія ймовірностей і математична статистика [Текст]. / В.І. Жлуктенко, С.І. Наконечний. – К.: ІЗМН, 1975. – 408 с.
20. Кибзун А.И. Теория вероятностей и математическая статистика [Текст]: Базовый курс с примерами и задачами / А.И. Кибзун, Е.Р. Горяинова, А.В. Наумов, А.Н. Сиротин. - М.: ФИЗМАТЛИТ, 2002. – 224 с.
21. Колмогоров А.Н. Введение в теорию вероятностей [Текст]. / А.Н.Колмогоров, И.Г. Журбенко, А.В. Прохоров. – М.: Наука, ГРФМЛ, 1982. - 160 с. (Библиотека „Квант”. Вып. 23).
22. Кузьмич Л.В. Теорія ймовірностей. Методичні вказівки до вивчення курсу [Текст]: методичні вказівки / Л.В. Кузьмич. – Херсон: Айлант, 2000. – 36 с.
23. Лапа В.Г. Математические основы кибернетики [Текст] / В.Г. Лапа. – К.: ВШ, 1974. – 452 с.
24. Леоненко М.М. Теоретико-імовірнісні та статистичні методи в економетриці та фінансовій математиці [Текст]: Навч. посібник / М.М. Леоненко, Ю.С.Мішура, В.М. Пархоменко, М.Й. Ядренко – К.: Інформтехніка, 1995. - 380 с.
25. Майстров Л.Е. Теория вероятностей. Исторический почерк. / Л.Е. Майстров . – М.: Наука, 1967. – 320 с.
26. Мармоза А.Т. Практикум по математической статистике. Практикум: Учеб. пос. / А.Т.Мармоза. - К.: ВШ, 1990. – 191 с.
27. Мостеллер Ф. П’ятдесят занимательных вероятностных задач с решениями. Пер. с англ. / Под ред.Ю.В. Линника. 3-е изд. – М.: Наука, ГРФМЛ, 1985. – 88 с.
28. Орлов А.И. Вероятность и прикладная статистика: основные факты [Текст]: справочник. / А.И. Орлов. – М.: КНОРУС, 2010. – 192 с.

29. Орлов А.И. Прикладная статистика. – М.: Экзамен, 2004. – 625 с. [Интернет-ресурс]: Режим доступу www.antorlov.chat.ru, www.antorlov.euro.ru
30. Пустыльник Е.И. Статистические методы анализа и обработки наблюдений [Текст]. - М.: Наука, ГРФМЛ, 1968. – 288 с.
31. Сборник задач по теории вероятностей, математической статистике и теории случайных функций / Б.Г. Володин, М.П.Ганин, И.Я. Динер [и др.]. Под. ред. А.А. Свешникова. - М.: Наука, 1970. – 656 с.
32. Сборник задач по математике для втузов [Текст]: Ч. 3. Теория вероятностей и математическая статистика: Учеб. пос. для втузов / Под. ред. А.В. Ефимова. - М.: Наука, ГРФМЛ, 1990. – 428 с.
33. Сборник задач по теории вероятностей, математической статистике и теории случайных функций [Текст]: / Б.Г. Володин, М.П.Ганин, И.Я. Динер [и др.]. (Под. ред. А.А. Свешникова).- М.: Наука, 1970. – 656 с.
34. Севастьянов Б.А. Курс теории вероятностей и математической статистики [Текст]: Уч. для студ. вузов / Б.А.Севастьянов. – М.: Наука, ГРФМЛ, 1982. – 256 с.
35. Севастьянов Б.А. Сборник задач по теории вероятностей [Текст]. / Б.А.Севастьянов, В.П Чистяков., А.М.Зубков. – М.: Наука, ГРФМЛ, 1980. – 224 с.
36. Смирнов Н.В. Курс теории вероятностей и математической статистики для технических приложений [Текст]. / Н.В. Смирнов, И.В. Дунин-Барковский. – М.: Наука, ГРФМЛ, 1969. – 512 с.
37. Скороход А.В. Элементы теории вероятностей и случайных процессов. / А.В.Скороход. – К.: ВШ, 1980. – 344 с.

38. Соловьев В. И. Математика для специальностей «Государственное и муниципальное управление», «Менеджмент организации». Часть 3. Теория вероятностей и математическая статистика в экономике. Раздел 3.2. Рабочая тетрадь. - Москва — 2005. [Интернет-ресурс]: Режим доступу http://www.nauka.x-pdf.ru/17ekonomika/624251-1-v-solovev-matematika-dlya-specialnostey-gosudarstvennoe-municipalnoe-upravlenie-menedzhment-organizacii-chast-teor.php
39. Солодовников А.С. Теория вероятностей [Текст]: Уч. пособие для студ. пед. ин-в / А.С. Солодовников. - М.: Просвещение, 1963.- 207 с.
40. Феллер В. Введение в теорию вероятностей и ее приложения: В 2 т. Пер. с англ. / В. Феллер – М.: Мир, 1984. – Т. 1. - 528 с., Т.2. - 752 с.
41. Чистяков В.П. Курс теории вероятностей [Текст]: Уч. пособие для студ. втузов / В.П.Чистяков. – М.: Наука, ГРФМЛ, 1978 . – 224 с.

Навчальне видання

Савченко Олександр Григорович
Валько Наталія Валеріївна
Кавун Галина Михайлівна
Кузьмич Людмила Василівна

Теорія ймовірностей /та математична статистика

Базовий курс з прикладами та задачами

ISBN 978-966-630-165-2

Технічний редактор Савченко С.О.
Макет обкладинки Валько Н.В.

Підписано до друку 03.07.2017 р. Формат 60 х 84 1/16.
Папір офсетний. Друк різографія. Гарнітура Times New Roman.
Ум. друк. арк. 25. Наклад 300 прим.

Видруковано з готових оригінал-макетів у ТОВ "Айлант",
73000, Україна, м. Херсон, пров. Пугачова, 5/20.
Свідоцтво про реєстрацію ХС №1 від 20.08.2000 р.
Тел.: 49-33-48. 050-396-08-91