МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
“УЖГОРОДСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ”
КАФЕДРА ТЕОРІЇ ЙМОВІРНОСТЕЙ І МАТЕМАТИЧНОГО АНАЛІЗУ

М.С. Герич, О.О. Синявська

МАТЕМАТИЧНА СТАТИСТИКА

НАВЧАЛЬНИЙ ПОСІБНИК

Ужгород – 2021

УДК 519.22(075.8)
ББК 22.172я73
Г-37

Рецензенти:

А. О. Пашко, доктор фізико-математичних наук, професор, професор кафедри теоретичної кібернетики Київського національного університету ім. Тараса Шевченка;

Ю. Ю. Млавець, кандидат фізико-математичних наук, доцент, доцент кафедри кібернетики і прикладної математики ДВНЗ «Ужгородський національний університет».

Герич М.С., Синявська О.О.

Г-37 Математична статистика: навч. посіб. Ужгород: ДВНЗ "УжНУ", 2021. 146 с.

У навчальному посібнику викладено основні розділи математичної статистики (статистичні розподіли вибірки та їх графічне представлення, статистичні оцінки параметрів розподілу, задачі перевірки статистичних гіпотез, елементи кореляційного та регресійного аналізу) відповідно до робочої програми дисципліни. Навчальний матеріал поділений на 4 розділи, у кінці кожного розділу наведено 30 комплексних індивідуальних завдань із розв'язком типового варіанта завдання.

Рекомендовано для студентів математичних і технічних спеціальностей ДВНЗ «Ужгородський національний університет», студентів інших навчальних закладів усіх форм навчання, що мають схожу програму підготовки, а також для використання в якості додаткового матеріалу при організації викладачем практичних занять.

Рекомендовано до друку Вченою радою ДВНЗ «Ужгородський національний університет» від 21 квітня 2021 року, протокол № 5.

УДК 519.2(075.8)

ББК 22.17я73

ЗМІСТ

ВСТУП

Математична статистика – це розділ математики, що базується на поняттях та методах теорії ймовірностей і призначений для встановлення закономірностей, яким підкоряються масові однорідні випадкові явища, на основі збору, систематизації і обробки статистичних даних, одержаних у результаті спостережень, з метою формування обґрунтованих наукових і практичних висновків в умовах стохастичної невизначеності.

Предмет математичної статистики складають методи реєстрації, опису й аналізу статистичних експериментальних даних, які можна подати як сукупність значень одно- чи багатовимірної випадкової величини, що дозволяє застосовувати апарат теорії ймовірностей. У свою чергу, математична статистика служить базою для створення методів обробки й аналізу статистичного матеріалу в різних конкретних сферах людської діяльності

На сьогодні, у зв'язку з стрімким розвитком багатьох природничих та технічних наук, розвитком нових технологій й методів аналізу великих даних, актуальним є широке застосування імовірнісних та статистичних методів у багатьох галузях науки і техніки.

Серед основних задач математичної статистики виділяють:

1. *Оцінка невідомих параметрів розподілу ймовірностей значень випадкової змінної на основі отриманої вибірки.*

2. *Побудова довірчих інтервалів для параметрів розподілу ймовірностей генеральної сукупності.*

3. *Перевірка гіпотез про вигляд невідомого розподілу або про величину параметрів розподілу, вигляд якої відомий.*

4. *Задача виявлення тенденцій, закономірностей, залежностей між випадковими змінними на основі даних вибірок з цих змінних.*

Слово "статистика" походить від латинського "status" – стан. Звідси утворилися слова: "stato" – держава; "statista" – знавець держави; "statistica" – сума знань про державу. Спочатку статистика надавала лише словесний опис державних справ і лише з XIX ст. статистичні відомості стали описувати в кількісній формі.

Вивчення статистики розпочалось в стародавні часи з масових спостережень: з перепису населення у Давньому Китаї, Давній Греції, Римі, із опису земельних ділянок, робочої сили, ремесел, торгівлі, продуктів у Давньому Єгипті тощо. В античні часи у зв'язку із швидким ростом приватної власності статистичні поняття розглядались не тільки в державному обліку, але й при описі роботи банків, сфери торгівлі, власників майстерень.

Наступним кроком у розвитку теорії ймовірностей і статистики став період епохи Відродження, часу розвитку торгівельних та міжнародних товарно-грошових відносин, бухгалтерського обліку. У XIV ст. в Італії почали виникати перші страхові товариства, які займались розрахунком шанси, оцінювали ризик перевезень.

У другій половині XVII ст., в роботах Г. Галілея, Р. Декарта, І. Кеплера, Б. Спінози, Г. Лейбніца, І. Ньютона та інших, було закладено основи сучасної науки. Розвиток таких фундаментальних дисциплін як математика, фізика, астрономія, філософія допомогли усвідомити значення статистики як засобу пізнання, що призвело до виникнення статистики як науки в цілому.

Подальшим розвитком (кінець XIX – початок XX ст.) математична статистика зобов'язана П. Л. Чебишову, А. А. Маркову, О. М. Ляпунову, О. О. Чупрову, а також Ф. Гальтону, К. Пірсону та іншим.

У XX ст. найбільший вклад у математичну статистику зробили В. І. Романовський, Е. Е. Слуцький, А. Н. Колмогоров, М. В. Смірнов, Стьюдент (псевдонім У. Госсета), Е. Пірсон, Ю. Нейман, А. Вальд, Р. Фішер, А. В.Скороход, В. С.Королюк та інші вчені.

Навчальний посібник складено відповідно до робочої навчальної програми вивчення навчальної дисципліни «Математична статистика» для студентів математичних та нематематичних спеціальностей, яка може вивчатись як окрема дисципліна або є складовою курсу теорії ймовірностей і математичної статистики, вищої математики, а також для всіх зацікавлених осіб.

Посібник складається з чотирьох розділів та спрямований на викладення сучасних основ математичної статистики.

У першому розділі описано первісні поняття вибіркового методу в статистиці, а саме: поняття генеральної сукупності, вибірки, варіаційні ряди та їх графічне зображення, а також числові характеристики вибірки. У другому розділі викладено основні методи побудови точкових та інтервальних оцінок невідомих параметрів розподілу, зокрема для параметрів нормальних спостережень. Третій розділ присвячений перевірці статистичних гіпотез. У ньому наведено основні терміни та поняття з теорії перевірки гіпотез, головні положення щодо критерію хі-квадрат Пірсона, а також описано принципи перевірки статистичних гіпотез про значення параметрів нормальної вибірки. У четвертому розділі розглянуто методо кореляційного та регресійного аналізів, що є основним інструментом обробки результатів прикладних досліджень.

Також кожен розділ супроводжується 30 комплексними індивідуальними завданнями та прикладами їх виконання, які дозволяють формувати індивідуальну домашню роботу студентів. Кожний варіант – завдання з розділів: «Вибіркові ряди і їх числові характеристики», «Статистичні оцінки параметрів розподілу», «Перевірка статистичних гіпотез», «Елементи кореляційного і регресійного аналізу». Типове завдання складене так, що дотримується принцип однакової складності для усіх варіантів. Кожний варіант містить завдання з кожної теми різного рівня складності. Частину задач узято із навчальних посібників [3, 6, 7].

Наведені в посібнику вибірки можуть слугувати фактичним матеріалом для проведення лабораторних робіт з використанням сучасних інформаційно-комунікаційних технологій.

В списку літератури перераховані підручники та посібники, які автори використовували при написанні даного посібника, а також яку можна застосовувати для підготовки та самостійного опрацювання студентами до даної дисципліни. Підрозділи, формули, приклади, твердження і т. п. мають наскрізну нумерацію у кожному розділі.

РОЗДІЛ 1. СТАТИСТИЧНІ РОЗПОДІЛИ ВИБІРКИ ТА ЇХ ГРАФІЧНЕ ПРЕДСТАВЛЕННЯ

Література: [1, розділ 5, п. 5.2-5.8], [2, розділ 1, п. 1.1-1.4], [3, модуль 3, т. 1], [5, розділ 11], [7, розділ 8].

Базові поняття. Генеральна сукупність та вибірка. Варіаційний ряд. Статистичний розподіл вибірки. Полігон та гістограма. Емпірична функція розподілу. Числові характеристики вибірки.

Основні задачі. Побудова статистичних розподілів. Графічне представлення розподілів. Обчислення числових характеристик.

ЩО МАЄ ЗНАТИ ТА ВМІТИ СТУДЕНТ

1. Знання на рівні понять, означень, формулювань. Генеральна сукупність, вибірка. Статистичний розподіл. Полігон частот, відносних частот. Інтервальний статистичний розподіл. Гістограма частот та відносних частот. Емпірична функція розподілу та її властивості. Числові характеристики вибірки.

2. Уміння в розв'язуванні задач. Знаходити та будувати дискретний та інтервальний статистичний розподіл вибірки випадкової величини та системи випадкових величин. Будувати полігон та гістограму частот, знаходити та будувати функцію розподілу. За статистичними розподілами знаходити числові характеристики вибірки.

ОСНОВНІ ТЕОРЕТИЧНІ ВІДОМОСТІ

1.1. Генеральна сукупність та вибірка. Варіаційний ряд. Статистичний розподіл вибірки (дискретний та інтервальний)

Математична статистика вивчає методи збору і обробки результатів спостережень для одержання наукових і практичних висновків. Методи розв'язування задач в теорії ймовірностей вимагають знання різних ймовірнісних характеристик. При розв'язуванні практичних задач ці характеристики можуть бути невідомими. Математична статистика розглядає методи знаходження необхідних ймовірнісних характеристик за статистичними даними. Тобто основним предметом дослідження в математичній статистиці є випадкові величини, але висновки про властивості цих випадкових величин робляться на основі статистичних даних.

Одна із важливих задач математичної статистики пов'язана зі знаходженням розподілу випадкової величини за статистичними даними, до якої зводяться багато інших завдань. Основними завданнями математичної статистики є розробка методів знаходження оцінок (знайдених на основі статистичних даних характеристик досліджуваної величини), дослідження їх точності, розробка методів перевірки гіпотез, аналіз зв'язку між величинами.

Нехай задано сукупність об'єктів, що об'єднані спільною ознакою X (кількісною або якісною, якою характеризуються однорідні об'єкти). Її значення для кожного об'єкту передбачити неможливо, тому досліджувана ознака X є випадковою величиною. Вивчення властивостей величини X пов'язане із проведенням статистичних спостережень – знаходження значень величини X для деяких об'єктів. Таким чином можна обстежити всю сукупність або деяку її частину. Обстеження всієї сукупності не завжди доцільне, а інколи просто неможливе. Таким методом, зокрема, здійснюють контроль за якістю продукції, проводять соціологічні дослідження. У всіх таких випадках проводиться обстеження тільки частини об'єктів і по його результатах роблять висновки про всю сукупність. Такий метод називається *вибірковим*.

Означення 1.1. Множина значень $x_1, x_2, \dots, x_n$ величини X, що одержані в результаті n експериментів, називається *вибіркою обсягу* n. Множину всіх значень величини X ми будемо називати *генеральною сукупністю*.

Поняття вибірки вживається і більш широко – як n-вимірна випадкова величина. Вибірка є єдиним джерелом інформації про досліджувану величину. Ми будемо розглядати вибірку організовану методом випадкового відбору. При цьому кожен об'єкт сукупності має однакову ймовірність бути відібраним, така вибірка буде правильно відображати властивості всієї сукупності. На практиці частіше всього використовується вибір без повернення (безповторна вибірка), коли кожен відібраний об’єкт перед вибором наступних назад до сукупності не повертається. Вибір з поверненням розглядається частіше в теоретичних дослідженнях. Якщо обсяг вибірки значно менший обсягу досліджуваної сукупності, то обидві вибірки дають практично однакові результати.

При одночасному дослідженні двох (або більшої кількості) ознак вибірка буде складатись із упорядкованих пар (або упорядкованих наборів) чисел.

Для розв’язування багатьох задач важливо знати розподіл досліджуваної випадкової величини, бо на його основі приймають ті чи інші рішення.

Різні можливі значення x_i ознаки X, які потрапили до вибірки, називаються *варіантами*, а система (сукупність) варіант $x_1, x_2, \dots, x_k$, розміщена в зростаючому порядку, називається *варіаційним рядом.*

Якщо варіанти $x_1, x_2, \dots, x_k$ спостерігаються у вибірці відповідно $n_1, n_2, \dots, n_k$ раз, то значення $n_1, n_2, \dots, n_k$ називаються *частотами варіант,* і для них виконується умова

$$\sum_{i=1}^{k} n_i = n, \tag{1.1}$$

а відношення

$$\omega_1 = \frac{n_1}{n}, \omega_2 = \frac{n_2}{n}, \dots, \omega_k = \frac{n_k}{n}$$

називаються *відносними частотами варіант*, і для них виконується умова

$$\textstyle\sum_{i=1}^{k} \omega_i = 1. \tag{1.2}$$

Таблиця, в першому рядку якої наведено впорядковані варіанти x_i (варіаційний ряд), а в другому — відповідні їм частоти n_i (або відносні частоти ω_i), називається *дискретним статистичним розподілом* вибірки (табл. 1.1), де $n = n_1 + n_2 + \dots + n_k$ – об’єм вибірки.

Таблиця 1.1

Варіанти x_i	x_1	x_2	$\ldots$	x_{k-1}	x_k
Частоти n_i	n_1	n_2	$\ldots$	n_{k-1}	n_k

Якщо кількість варіант у вибірці надто велика або досліджувана ознака X є неперервною випадковою величиною, то будується *інтервальний статистичний розподіл* вибірки, аналогічний дискретному, у першому рядку якого замість дискретного варіаційного ряду записується інтервальний $(x_i;\ x_{i+1})$. Частоти n_i^* варіант, які потрапили в кожний частинний інтервал $(x_i;\ x_{i+1})$, записуються в другий рядок таблиці (табл. 1.2).

Таблиця 1.2

Інтервали $(x_i;\ x_{i+1})$	$(x_1;\ x_2)$	$(x_2;\ x_3)$	$\ldots$	$(x_{m-1};\ x_m)$	$(x_m;\ x_{m+1})$
Частоти n_i^*	n_1^*	n_2^*	$\ldots$	n_{m-1}^*	n_m^*

Зауваження 1.1. Кількість інтервалів m для вибірки обсягом n орієнтовно вважається такою, що дорівнює $\sqrt{n}$ (із округленням до цілого значення), а розмір (крок) h інтервалу обчислюється за формулою

$$h = \frac{x_k - x_1}{m}, \tag{1.3}$$

де x_1 і x_k — відповідно найменша і найбільша варіанти у вибірці. Якщо межа двох частинних інтервалів збігається зі значенням варіанти, то частота цієї варіанти або ділиться порівну між цими двома інтервалами, або в усіх таких випадках цілком заноситься у правий (або лівий) від варіанти інтервал. Якщо замість n_i^* поставити $\frac{n_i^*}{n}$, то будемо мати статистичний розподіл відносних частот.

Інколи для зручності подальшої обробки результатів спостережень від інтервального розподілу переходять до дискретного, замінюючи кожен інтервал його серединою.

1.2. Полігон та гістограма. Емпірична функція розподілу

Для створення наочного уявлення про статистичні розподіли застосовуються *полігон* та *гістограма* частот.

Означення 1.2. *Полігоном частот* (або відносних частот) називається ламана лінія, відрізки якої сполучають на площині точки з координатами $(x_i;\ n_i)$ (або $(x_i;\ \omega_{\ i})$). (див. рис. 1.1)

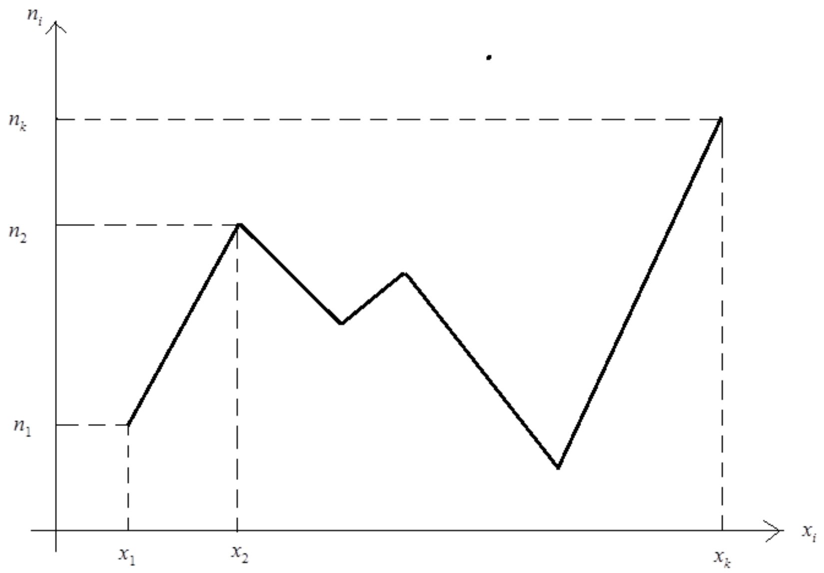


Рис. 1.1

Полігон найчастіше використовується для графічного зображення дискретного статистичного розподілу, але його можна застосувати також для графічного зображення інтервальних статистичних розподілів вибірки. У цьому випадку за абсциси x_i точок беруться центри частинних інтервалів.

Гістограма застосовується для графічного зображення інтервальних статистичних розподілів. Для побудови гістограми частот (або відносних частот) на осі абсцис відкладаються відрізки, що дорівнюють довжині (кроку) h частинних інтервалів, і на цих відрізках як на основах будуються прямокутники з висотами $\frac{n_i^*}{h}$ (або $\frac{\omega_i^*}{h} = \frac{n_i^*}{nh}$).

Оскільки величини $\frac{n_i^*}{h}$ ($\frac{\omega_i^*}{h}$) є щільностями частот (відносних частот) на відповідних інтервалах, то при достатньо великих обсягах вибірки n (і відповідно малих h) гістограма може бути достатньо близьким статистичним аналогом щільності ймовірності досліджуваної ознаки X у генеральній сукупності.

Нехай є статистичний розподіл частот деякої ознаки X. Позначимо як n загальну кількість спостережень. Нехай x – довільне дійсне число (точка числової прямої). Позначимо як n_x – кількість спостережень, при яких спостерігалися ознаки X менше x. Тоді відносна частота події $X < x$ дорівнює $\frac{n_x}{n}$. Якщо x змінюється, то, взагалі кажучи, змінюється й n_x, а отже й $\frac{n_x}{n}$, тобто $\frac{n_x}{n}$ є функцією від x. Ця функція знаходиться емпіричним (дослідним) шляхом, тому її називають емпіричною.

Означення 1.3. *Емпіричною функцією статистичного розподілу* називається функція $F^*(x)$, яка для кожного значення x дорівнює відносній частоті події $X < x$, тобто:

$$F^*(x) = \frac{n_x}{n},$$

де n_{x_i} – сумарна частота всіх варіант x_i, менших за x.

Означення 1.4. Інтегральну функцію розподілу $F(x) = P\{X < x\}$ генеральної сукупності називають *теоретичною функцією розподілу*.

Функція $F(x)$ відрізняється від функції $F^*(x)$ тим, що вона визначає ймовірність події $X < x$, а функція $F^*(x)$ – відносну частоту цієї події. При збільшенні числа випробувань n подія, яка полягає в тому, що емпірична функція розподілу буде мало відрізнятися від теоретичної функції розподілу, стає практично достовірною. А звідси випливає доцільність використання функції $F^*(x)$ для наближеного зображення теоретичної функції розподілу $F(x)$ генеральної сукупності. Можна сказати, що функція $F^*(x)$ є статистичним аналогом функції $F(x)$.

Такий висновок підтверджується також тим, що функція $F^*(x)$ має властивості, які аналогічні властивостям функції $F(x)$. А саме.

1. Всі значення функції $F^*(x)$ належать відрізку $[0,1]$: $0 \le F^*(x) \le 1$.
2. Функція $F^*(x)$ є неспадною, тобто, для довільних $x_1 < x_2$:

$$F^*(x_1) \le F^*(x_2).$$

3. $\lim\limits_{x \to -\infty} F^*(x) = 0, \ \lim\limits_{x \to +\infty} F^*(x) = 1.$

Ця властивість фактично означає наступне: якщо x_1 – найменша варіанта, то $F^*(x) = 0$ при $x \le x_1$; якщо x_k – найбільша варіанта, то $F^*(x) = 1$ при $x > x_k$.

Графік $F^*(x)$ можна розглядати як наближений графік теоретичної функції розподілу $F(x)$. Він демонструє перелічені вище властивості функції $F^*(x)$, а також наступні властивості: ця функція кусково-стала, у точках, що відповідають значенням ознаки, вона має розриви І роду і неперервна зліва в цих точках. Легко помітити, що ці властивості аналогічні властивостям функції розподілу дискретної випадкової величини.

1.3. Числові характеристики вибірки

Одним з важливим засобів обробки даних є обчислення їх числових характеристик. Найбільш важливі з них: середнє значення, вибіркова дисперсія, вибіркове середньоквадратичне відхилення. Ці характеристики можуть бути обчислені наближено за даними вибірки.

Вибіркове середнє, розмах вибірки. Нехай $x_1, x_2, \ldots, x_n$ – вибірка із генеральної сукупності об'єму n.

Означення 1.5. *Вибірковим середнім* називають величину

$$\bar{x}_{\text{в}} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i. \qquad (1.4)$$

Нехай маємо варіаційний ряд $x_1, x_2, \ldots, x_k$, де $x_1 = x_{min}$, а $x_k = x_{max}$.

Розмахом вибірки називається $R = x_{max} - x_{min}$.

Означення 1.6. *Вибірковим середнім* називають величину

$$\bar{x}_{\text{в}} = \frac{x_1 n_1 + x_2 n_2 + \cdots + x_k n_k}{n} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \sum_{i=1}^{k} x_i \omega_i. \qquad (1.5)$$

Вибіркове середнє є статистичним аналогом математичного сподівання.

Зауваження 1.2. Якщо варіанти вибірки розташовані в зростаючому порядку і утворюють арифметичну прогресію з різницею h (рівновіддалені варіанти), то ввівши умовну варіанту

$$u_i = \frac{x_i - C}{h}, \qquad (1.6)$$

де C – *хибний нуль* (для простоти обчислень C слід вибрати рівним варіанті, з найбільшою частотою); h – крок, тобто відстань між сусідніми варіантами. Тоді вибіркове середнє обчислюють за наступною формулою

$$\bar{x}_{\text{в}} = h\bar{u}_{\text{в}} + C. \tag{1.7}$$

Зауваження 1.3. Якщо маємо статистичний інтервальний розподіл (див. табл. 1.2), то в ролі x_i в формулах (1.4), (1.5) та (1.7) використовують середини інтервалів.

Зауваження 1.4. Обчислення по формулах (1.4), (1.5) можна спростити, якщо зробити заміну змінних $y_i = x_i - C$, де C вибирають як середину варіаційного ряду. Тоді формули (1.4), (1.5) перетворюються наступним чином

$$\bar{x}_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k} y_i + C = \bar{y}_{\text{в}} + C; \tag{1.8}$$

$$\bar{x}_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k} y_i n_i + C = \bar{y}_{\text{в}} + C. \tag{1.9}$$

Зауваження 1.5. Якщо варіанти $x_1, x_2, \dots, x_k$ є десятковими дробами з l знаками після коми, то доцільно перейти до умовних варіант $v_i = 10^l x_i$. Тоді:

$$\bar{x}_{\text{в}} = \frac{1}{10^l} \cdot \bar{v}_{\text{в}}. \tag{1.10}$$

Крім розглядуваних вибіркових середніх величин, які є аналітичними, в статистичному аналізу застосовують структурні і порядкові середні. Із них найбільш широко застосовується медіана і мода.

Медіана. *Медіаною вибірки (* позначається *Ме)* називається значення серединного елемента варіаційного ряду.

Для дискретного варіаційного ряду при непарному обсязі вибірки n медіана рівна серединному елементу, а при n парному – півсумі двох серединних елементів.

Якщо маємо статистичний інтервальний розподіл, то медіана обчислюється за наступною наближеною формулою:

$$Me = x_0 + h\frac{n/2 - T_{i-1}}{n_i}, \tag{1.11}$$

де x_0 – початок медіанного інтервала, тобто, інтервалу, в якому міститься серединний елемент; h – довжина медіанного інтервалу; n – об'єм вибірки; T_{i-1} – сума частот інтервалів, які передують медіанному; n_i – частота медіанного інтервалу.

Суттєвим плюсом медіани як міри центральної тенденції, є те, що на неї не впливає зміна крайніх членів ряду.

Мода. *Модою вибірки (* позначається *Mo)* для дискретного статистичного розподілу називається варіанта з найбільшою частотою.

У випадку статистичного інтервального розподілу, мода обчислюється згідно наступної формули:

$$Mo = x_0 + h\frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})}, \tag{1.12}$$

де x_0 – початок модального інтервалу, тобто інтервалу, який має найбільшу частоту; h – довжина модального інтервалу; n_{i-1} і n_{i+1} – частоти відповідно попереднього і наступного за модальним інтервалом; n_i – частота модального інтервалу.

Моду можна знайти графічним способом з допомогою гістограми. Для цього знаходимо прямокутник з найбільшою частотою (відносною частотою). З'єднуючи відрізками прямих вершини цього прямокутника з відповідними вершинами двох сусідніх прямокутників, отримаємо точку перетину цих відрізків, абсциса якої і буде модою ряду.

Особливість моди як міри центральної тенденції полягає в тому, що вона не змінюється при зміні крайніх членів ряду.

Вибіркова та виправлена дисперсія, вибіркове та виправлене середньоквадратичне відхилення. Нехай $x_1, x_2, \dots, x_n$ – вибірка із генеральної сукупності об'єму n. Якщо для даної вибірки складений дискретний розподіл (див. табл. 1.1), то

Означення 1.7. *Вибірковою дисперсією* називається величина:

$$D_B = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2 . \tag{1.13}$$

Іншими словами це середнє арифметичне квадратів відхилень варіант від вибіркового середнього з урахуванням відповідних частот. На практиці величину D_B, як правило, простіше обчислювати за формулою:

$$D_B = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - (\bar{x}_B)^2 , \qquad (1.14)$$

тобто середнє квадратів без квадрату середнього вибіркового.

При введені нових варіант, згідно (1.6), вибіркову дисперсію зручніше обчислювати за наступною формулою

$$D_B[\mathrm{x}] = h^2 \cdot \frac{1}{n}\sum_{i=1}^{k} n_i u_i^2 - h^2(\bar{u}_B)^2 = h^2\left(\overline{u^2}_B - (\bar{u}_B)^2\right). \qquad (1.15)$$

Для спрощення розрахунків, з урахуванням заміни $y_i = x_i - C,$ формули (1.13), (1.14) можна переписати наступним чином:

$$D_B[\mathrm{x}] = \frac{1}{n}\sum_{i=1}^{k} y_i^2 - (\bar{y}_B)^2 = \overline{y^2}_B - (\bar{y}_B)^2, \qquad (1.16)$$

$$D_B[\mathrm{x}] = \frac{1}{n}\sum_{i=1}^{k} n_i y_i^2 - (\bar{y}_B)^2 = \overline{y^2}_B - (\bar{y}_B)^2. \qquad (1.17)$$

Якщо переходити до нових варіант v_i наведених у зауваженні 1.5, то дисперсію доцільно обчислювати за наступною формулою:

$$D_B[\mathrm{x}] = \frac{1}{10^{2l}} \cdot D_B[v]. \qquad (1.18)$$

Якщо дані представлені у вигляді інтервального ряду, то необхідно його замінити дискретним рядом, а потім використати наведені вище формули.

Крім вибіркової дисперсії, користуються також вибірковим середньоквадратичним відхиленням.

Вибіркова дисперсія має один істотний недолік: якщо середнє вибіркове подається в тих же одиницях, що й значення випадкової величини, то, як випливає з формул, що задають дисперсію, остання подається вже у квадратних одиницях. Цього недоліку можна уникнути, узявши за міру розсіювання арифметичний квадратний корінь із дисперсії.

Корінь квадратний з вибіркової дисперсії називається *вибірковим середнім квадратичним відхиленням*, тобто

$$\sigma_{\text{в}} = \sqrt{D_{\text{в}}}\ . \tag{1.19}$$

Часто в математичній статистиці розглядається *виправлена дисперсія*

$$S^2 = \frac{n}{n-1} \cdot D_B. \tag{1.20}$$

Корінь квадратний з виправленої дисперсії називається *виправленим середньо квадратичним відхиленням*, тобто $S = \sqrt{S^2}$.

При достатньо великих значеннях n об'єму вибірки, вибіркова і виправлена дисперсія відрізняються мало. Тому на практиці використовують виправлену дисперсію, якщо приблизно $n < 30$.

Коефіцієнт варіації. Для порівняння величин розсіювання стосовно вибіркової середньої двох варіаційних рядів уводиться безрозмірна характеристика – *коефіцієнт варіації*, який дорівнює відсотковому відношенню середнього квадратичного відхилення до середнього вибіркового:

$$V = \frac{\sigma_{\text{в}}}{\bar{x}_{\text{в}}} \cdot 100\%, (\bar{x}_{\text{в}} \neq 0). \tag{1.21}$$

Із двох рядів той має більше розсіювання відносно вибіркової середньої, у якого V більша.

Початкові та центральні емпіричні моменти. Нехай $x_1, x_2, \dots, x_n$ – вибірка із генеральної сукупності об’єму n і для даної вибірки складений дискретний розподіл (див. табл. 1.1).

Означення 1.8. *Початковим емпіричним моментом* порядку m вибірки називають величину:

$$\alpha_m^* = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^m. \tag{1.22}$$

Означення 1.9. *Центральним емпіричним моментом* порядку m вибірки називають величину:

$$\mu_m^* = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^m. \tag{1.23}$$

Зокрема: $\alpha_1^* = \bar{x}_B$, $\mu_1^* = 0$, $\mu_2^* = D_B$.

Для центрального емпіричного моменту 2-го порядку маємо формулу:

$$\mu_2^* = \alpha_2^* - (\alpha_1^*)^2.$$

Для центральних емпіричних моментів 3-го та 4-го порядків справедливі наступні формули:

$$\mu_3^* = \alpha_3^* - 3\alpha_1^*\alpha_2^* + 2(\alpha_1^*)^3,$$
$$\mu_4^* = \alpha_4^* - 4\alpha_1^*\alpha_3^* + 6\alpha_2^*\,(\alpha_1^*)^2 - 3(\alpha_1^*)^4.$$

Асиметрія, ексцес вибірки. Для оцінки відхилень емпіричного (експериментального) розподілу від нормального використовуються різні характеристики, до числа яких відносяться вибіркові асиметрія і ексцес.

Означення 1.10. *Асиметрією вибірки* називається число

$$a_s = \frac{\tilde{\mu}_3}{\sigma_B^3}. \tag{1.24}$$

Означення 1.11. *Ексцесом вибірки* називається число

$$e_k = \frac{\tilde{\mu}_4}{\sigma_B^4} - 3. \tag{1.25}$$

Кажуть, що асиметрія емпіричного розподілу додатна (від'ємна), якщо його головна частина, тобто максимум, концентрується з лівої (правої) сторони від нормального розподілу.

Якщо ж максимум емпіричного розподілу розміщений вище (нижче), ніж у нормального розподілу, то кажуть, що емпіричний розподіл має додатний (від'ємний) ексцес.

Якщо асиметрія та ексцес близькі до нуля, то можна зробити припущення про нормальний закон розподілу в сукупності.

ВАРІАНТИ ІНДИВІДУАЛЬНИХ ЗАВДАНЬ (ІЗ № 1).

За вибірковими даними випадкової величини X (табл. 1.3) індивідуальних завдань до розділу 1 потрібно:

1) скласти варіаційний ряд, знайти дискретний та інтервальний статистичні розподіли вибірки;
2) знайти емпіричну функцію розподілу $F^*(x)$ за дискретним та інтервальним розподілами і накреслити їх графіки;
3) побудувати полігон частот та відносних частот для дискретного розподілу, а також для інтервального розподілу гістограму відносних частот;
4) знайти наступні числові характеристики вибірки: $R,\ \bar{x}_B,\ D_B,\ \sigma_B,\ S^2,\ V, Me, Mo.$

Таблиця 1.3

Номер варіанту	Дані вибірки									
1	112	117	105	121	118	115	119	120	118	120
	110	109	110	111	113	107	108	106	109	105
	101	107	112	103	101	98	93	100	95	91
2	-1,5	-1,0	1,0	0,5	1,5	0,5	-1,0	-1,0	-1,5	3,5
	0,5	2,0	-0,5	- 1,5	1,0	2,5	-2,5	-2,0	0,0	2,5
	-2,0	-1,0	-2,5	1,0	2,0	-0,5	- 1,5	0,0	0,5	1,5
3	180	100	400	170	520	680	250	1300	730	240
	350	460	630	570	450	520	310	240	540	450
	630	590	570	300	650	620	1200	730	350	430
4	779	763	769	774	758	750	742	793	784	763
	755	802	769	784	736	785	774	769	758	807
	750	742	763	774	755	816	763	758	784	785
5	0,56	0,59	0,57	0,58	0,57	0,62	0,61	0,60	0,61	0,64
	0,63	0,65	0,62	0,66	0,59	0,65	0,66	0,63	0,60	0,61
	0,60	0,58	0,64	0,59	0,62	0,60	0,61	0,64	0,66	0,56

6	10,0 9,8 11,4 11,0 13,0 14,0 14,7 12,3 9,0 12,3 10,0 13,8 11,2 13,6 13,0 12,5 12,3 11,4 11,2 12,0 10,0 13,0 12,3 14,2 14,1 15,0 14,7 14,0 9,8 11,0
7	200 300 250 450 350 600 650 550 450 350 400 550 500 450 500 350 500 650 300 350 450 250 400 550 400 450 300 400 500 450
8	37 8 88 103 25 71 37 121 132 103 121 71 132 148 88 56 71 103 8 25 132 121 103 37 71 121 37 25 121 103
9	194 195 200 191 190 191 197 193 200 192 194 195 196 198 192 198 191 190 197 193 194 195 191 190 198 193 199 196 190 199
10	6700 6800 6500 6530 7000 6950 6900 6930 6720 6720 6540 6790 6800 6750 6620 6650 7000 6830 6700 6690 6540 6970 6820 6670 6910 6930 6760 6710 6660 6760
11	6,5 7 6 8 7,5 6,5 6 8 5,5 8,5 9 7,5 7 6 7 6,5 7,5 7 6,5 10 9,5 8 8,5 6,5 7 7,5 5,5 6 6 8,5 7 11 8 8,5 8 7,5 7 6,5 7 7,5
12	79 63 69 74 75 75 74 79 84 76 55 80 69 78 76 85 77 76 75 80 75 74 76 77 75 81 76 75 78 85
13	670 680 655 653 700 695 691 693 654 670 654 679 680 670 662 655 700 682 655 679 654 670 682 670 691 691 670 691 700 676
14	6800 6500 6530 7300 6950 6900 6930 6720 6540 6790 6800 6750 6620 6650 7000 6830 6700 6690 6540 6970 6820 6670 6910 6930 6760 6710 6660 6760 6790 6820
15	340 316 325 329 351 348 330 345 352 331 318 332 341 318 341 353 356 320 347 349 352 342 337 341 350 348 327 339 340 339

16	235 230 231 234 230 232 239 238 240 230 232 237 234 240 239 231 235 230 238 239 240 235 233 240 239 233 240 236 234 239
17	51 52 51 55 54 56 53 54 56 58 55 57 60 53 55 58 54 50 59 53 54 56 57 53 58 55 59 56 55 54
18	26 30 38 35 31 34 29 28 37 36 30 31 26 36 29 27 30 35 29 34 35 38 31 34 28 37 36 28 38 29 27 29 35 37 30 26 28 36 34 31
19	0,5 -1,0 -0,5 2,0 3,0 -2,0 0,0 -1,5 -0,5 1,5 1,0 0,0 0,5 1,0 -3,0 -2,5 -0,5 2,0 2,5 3,0 0,5 1,5 -2,0 -1,0 3,0 -1,5 -2,5 2,5 -3,0 3,0 2,0 2,5 -0,5 1,5 1,0 0,0 -1,5 3,0 -2,5 -2,0
20	25 30 38 35 35 34 29 28 37 36 30 31 36 30 39 27 30 35 29 24 35 38 31 24 28 27 26 28 38 39 25 39 39 27 30 26 38 36 24 31
21	25,4 33,0 38,5 35,2 35,2 31,4 29,8 28,4 39,7 36,0 33,0 35,2 33,0 33,0 39,7 27,5 31,4 35,2 29,8 25,4 35,2 38,5 31,4 26,4 28,4 27,5 26,4 28,4 38,5 39,7 25,4 39,6 39,6 27,5 33,0 26,4 38,5 26,4 25,4 31,4
22	758 740 760 759 742 754 759 742 758 740 757 741 743 742 755 759 742 755 757 741 752 758 756 744 753 759 744 753 752 758
23	65 70 62 85 75 65 62 83 55 85 90 75 70 62 70 65 75 70 65 90 95 83 85 65 70 75 55 62 62 85 70 61 83 85 83 75 70 65 70 75

24	5,1	5,2	5,1	5,5	5,4	5,6	5,3	5,4	5,6	5,8
	5,5	5,7	6,0	6,0	5,5	5,8	5,4	5,0	5,9	5,3
	5,4	5,6	5,7	5,3	5,8	5,5	5,9	5,6	5,5	5,4
25	13	12	13	15	14	16	13	14	16	18
	15	17	16	13	15	18	14	10	19	13
	14	16	17	13	18	15	19	16	15	14
26	2,6	3,0	3,6	3,5	3,1	3,4	2,6	2,8	3,1	3,6
	3,0	3,5	2,7	3,6	2,9	3,5	2,7	2,7	3,5	3,0
	3,5	2,7	3,1	3,4	2,7	2,7	3,5	2,6	3,4	3,5
	2,7	2,9	3,5	3,7	2,8	3,7	2,9	2,6	2,7	2,7
27	20,2	20,4	21,0	20,8	19,5	21,0	21,8	20,4	21,8	20,2
	19,4	19,4	20,2	20,2	21,4	19,5	19,0	21,4	20,2	19,0
	20,4	21,2	21,2	18,8	20,4	20,2	20,0	20,4	20,4	20,0
	21,2	20,0	19,0	20,8	19,5	20,8	19,4	21,0	20,0	21,4
28	26	30	36	35	31	34	26	28	31	36
	30	35	27	36	29	35	37	27	35	30
	35	27	31	34	27	27	35	26	34	35
	27	29	35	37	28	37	29	26	27	27
29	2,0	0,0	1,2	1,0	0,5	0,5	2,2	0,8	0,5	2,8
	0,5	4,0	1,0	0,2	0,0	0,5	4,0	4,0	1,5	0,2
	0,7	0,0	0,2	2,1	1,7	1,5	2,0	0,2	2,0	0,8
	0,8	0,3	0,3	0,0	0,3	1,5	1,5	0,3	4,0	0,7
	1,5	0,7	0,4	0,5	0,7	0,8	1,0	2,1	2,1	2,2
30	6,5	7,0	62	85	75	6,5	62	83	55	85
	9,0	75	7,0	62	7,0	6,5	75	7,0	6,5	9,0
	9,5	83	85	6,5	7,0	75	55	62	62	85
	7,0	11	83	85	83	75	70	6,5	7,0	75

РОЗВ’ЯЗАННЯ ТИПОВОГО ВАРІАНТА ІЗ № 1

Умова завдання

За поданими далі вибірковими даними:

200	300	250	450	350	600	650	550	450	350
400	550	500	450	500	350	500	650	300	350
450	250	400	550	400	450	300	400	500	450

1) скласти варіаційний ряд, знайти дискретний та інтервальний статистичні розподіли вибірки;
2) знайти емпіричну функцію розподілу $F^*(x)$ за дискретним та інтервальним розподілами і накреслити їх графіки;
3) побудувати полігон частот та відносних частот для дискретного розподілу, а також для інтервального розподілу гістограму відносних частот;
4) знайти наступні числові характеристики вибірки: розмах R, вибіркове середнє $\bar{x}_B$, вибіркову дисперсію D_B, вибірковим середнім квадратичним відхиленням σ_B, виправлену вибіркову дисперсію S^2, варіацію V, медіану Me, моду Mo.

Хід розв’язання

1) Розмістивши значення варіант, які потрапили у вибірку, у порядку зростання, дістанемо дискретний варіаційний ряд і визначимо розмах вибірки $R = x_{max} - x_{min} = 450$:

$$200, 250, 300, 350, 400, 450, 500, 550, 600, 650$$

Знайдемо частоту кожної варіанти варіаційного ряду і складемо таблицю, тобто отримаємо дискретний статистичний розподіл вибірки (див. табл. 1.4). Знайдемо об’єм вибірки: $n = 1 + 2 + 3 + 4 + 4 + 6 + 4 + 3 + 1 + 2 = 30$.

Таблиця 1.4

x_i	200	250	300	350	400	450	500	550	600	650
n_i	1	2	3	4	4	6	4	3	1	2

Обчислимо відносні частоти варіант для даного дискретного розподілу:

$$\omega_1 = \frac{n_1}{n} = \frac{1}{30};\ \omega_2 = \frac{n_2}{n} = \frac{2}{30};\ \omega_3 = \frac{n_3}{n} = \frac{3}{30};\ \omega_4 = \frac{n_4}{n} = \frac{4}{30};\ \omega_5 = \frac{n_5}{n} = \frac{4}{30};$$

$$\omega_6 = \frac{n_6}{n} = \frac{6}{30};\ \omega_7 = \frac{n_7}{n} = \frac{4}{30};\ \omega_8 = \frac{n_8}{n} = \frac{3}{30};\ \omega_9 = \frac{n_9}{n} = \frac{1}{30};\ \omega_{10} = \frac{n_{10}}{n} = \frac{2}{30}.$$

Дістанемо дискретний розподіл відносних частот (табл.1.5.)

Таблиця 1.5

x_i	200	250	300	350	400	450	500	550	600	650
ω_i	$\frac{1}{30}$	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{4}{30}$	$\frac{4}{30}$	$\frac{6}{30}$	$\frac{4}{30}$	$\frac{3}{30}$	$\frac{1}{30}$	$\frac{2}{30}$

Складемо інтервальний статичний розподіл вибірки. Для цього знайдемо $x_{max} = 650$; $x_{min} = 200$. Розіб'ємо вибірку на $m = \sqrt{n} = \sqrt{30} \approx 5$ інтервалів. Довжина часткового інтервалу $h = \frac{x_{max} - x_{min}}{m} = \frac{450}{5} = 90$. Тоді інтервальний варіаційний ряд:

[200;290), [290;380), [380;470), [470;560), [560;650]

Інтервальний статистичний розподіл вибірки має вигляд (див. табл. 1.6):

Таблиця 1.6

(x_{i-1}, x_i)	[200;290)	[290;380)	[380;470)	[470;560)	[560;650]
n_i^*	3	7	10	7	3

Інтервальний статистичний розподіл відносних частот має вигляд (див. табл. 1.7):

Таблиця 1.7

(x_{i-1}, x_i)	[200;290)	[290;380)	[380;470)	[470;560)	[560;650]
$\omega_i^* = \dfrac{n_i^*}{n}$	$\dfrac{3}{30}$	$\dfrac{7}{30}$	$\dfrac{10}{30}$	$\dfrac{7}{30}$	$\dfrac{3}{30}$
ω_i^*/h	$\dfrac{1}{900}$	$\dfrac{7}{2700}$	$\dfrac{1}{270}$	$\dfrac{7}{2700}$	$\dfrac{1}{900}$

■

2) За дискретним статистичним розподілом (табл. 1.4) знайдемо емпіричну функцію розподілу та побудуємо її графік (див. рис. 1.2):

$$F^*(x) = \begin{cases} 0, & x \in (-\infty;\ 200]; \\ \dfrac{1}{30}, & x \in (200; 250]; \\ \dfrac{1}{30} + \dfrac{2}{30}, & x \in (250; 300]; \\ \dfrac{1}{30} + \dfrac{2}{30} + \dfrac{3}{30}, & x \in (300; 350]; \\ \dfrac{1}{30} + \dfrac{2}{30} + \dfrac{3}{30} + \dfrac{4}{30}, & x \in (350; 400]; \\ \dfrac{10}{30} + \dfrac{4}{30}, & x \in (400; 450]; \\ \dfrac{14}{30} + \dfrac{6}{30}, & x \in (450; 500]; \\ \dfrac{20}{30} + \dfrac{4}{30}, & x \in (500; 550]; \\ \dfrac{24}{30} + \dfrac{3}{30}, & x \in (550; 600]; \\ \dfrac{27}{30} + \dfrac{2}{30}, & x \in (600; 650]; \\ 1, & x \in (650; +\infty). \end{cases}$$

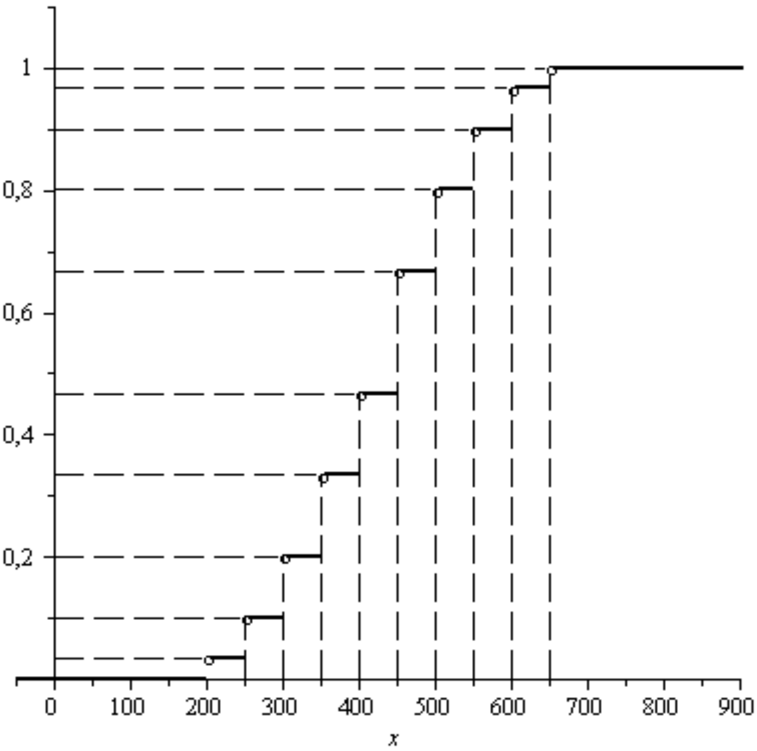


Рис. 1.2

За інтервальним статистичним розподілом (табл. 1.6) знайдемо емпіричну функцію розподілу та побудуємо її графік (див. рис. 1.3), прийнявши за значення x_i правий кінець інтервалу (крім останнього):

$$F^*(x) = \begin{cases} 0, & x \in (-\infty;\ 200]; \\ \dfrac{3}{30}, & x \in (200;\ 290] \\ \dfrac{3}{30} + \dfrac{7}{30}, & x \in (290;\ 380] \\ \dfrac{10}{30} + \dfrac{10}{30}, & x \in (380;\ 470] \\ \dfrac{20}{30} + \dfrac{7}{30}, & x \in (470;\ 560] \\ 1\,, & x \in (560; +\infty). \end{cases}$$

Графік емпіричної функції розподілу матиме наступний вигляд:

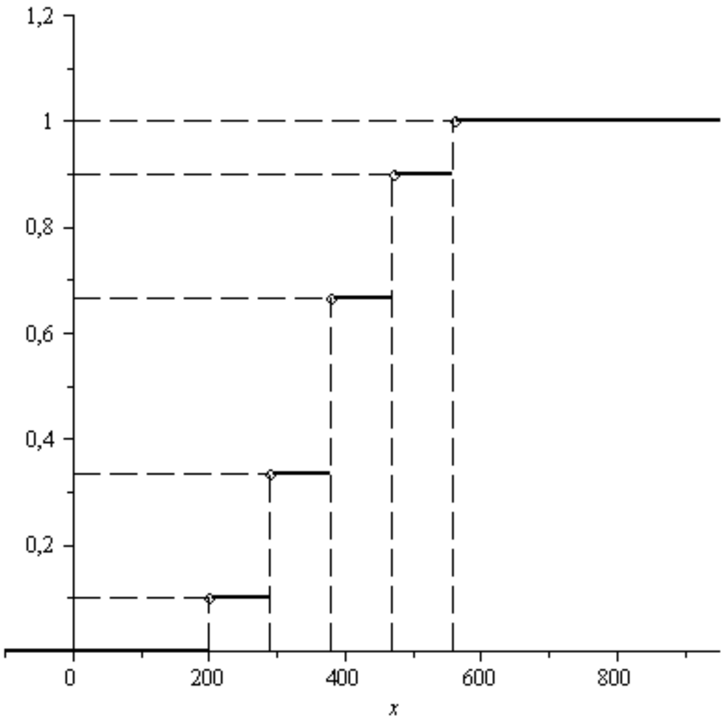


Рис. 1.3

■

3) Побудуємо полігон частот та відносних частот для дискретного статистичного розподілу. Для цього побудуємо ламану з вершинами $(x_i,\ n_i)$ (або $(x_i,\ \omega_i)$), де варіанти x_i та їх частоти n_i (відносні частоти ω_i) вибираємо з табл. 1.4 (табл. 1.5). Графіки відповідних полігонів зображені відповідно на рис.1.4 (рис. 1.5):

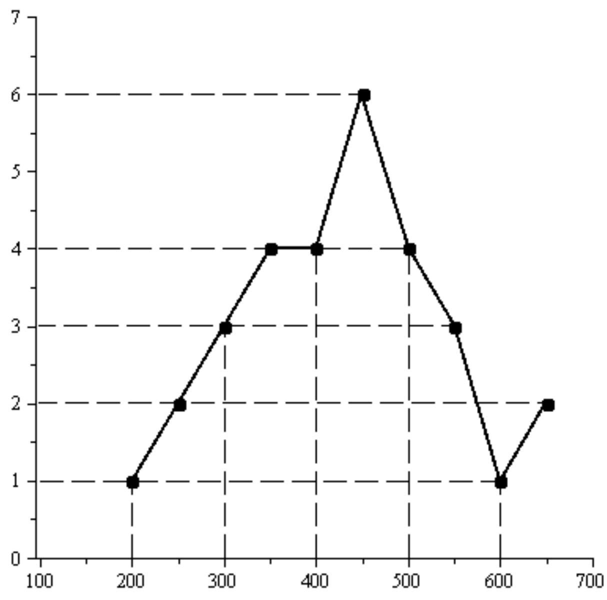


Рис. 1.4

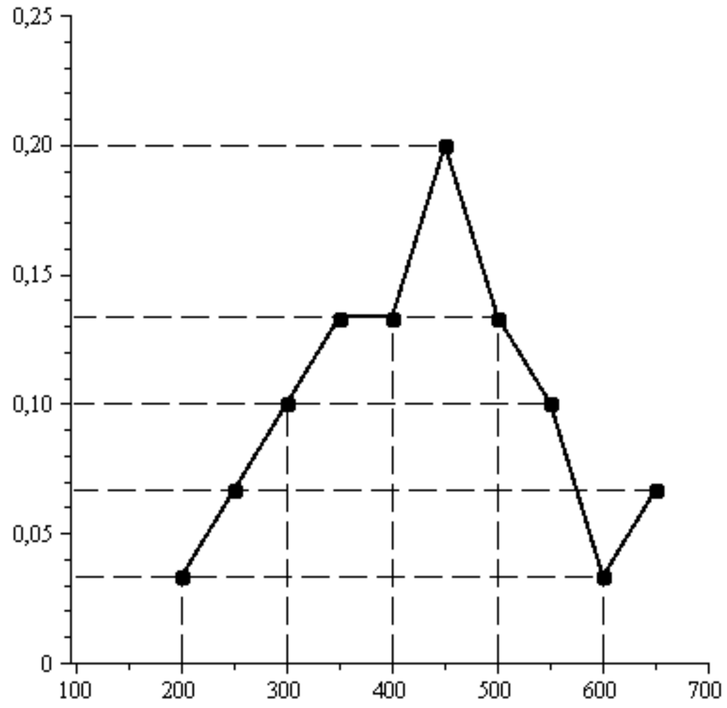


Рис. 1.5

Побудуємо полігон для інтервального статистичного розподілу, це ламана з вершинами $(x_i,\ n_i^*)$, де варіанти x_i вибираємо як середини інтервалів та відповідні їм частоти n_i^* з табл. 1.6. Графік полігону зобразимо на рис. 1.6:

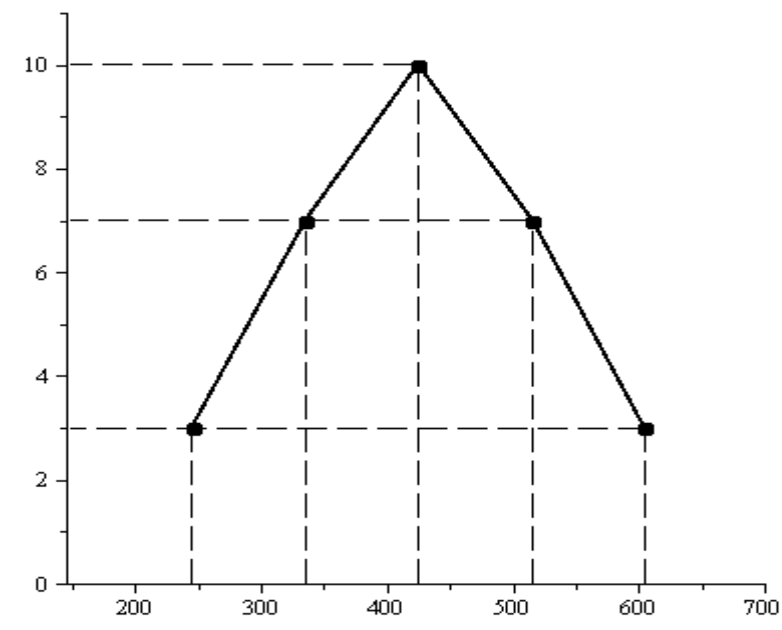


Рис. 1.6

Для інтервального розподілу гістограма відносних частот – прямокутники за даними з табл.1.7, а саме по 1 і 3 рядках (див. рис. 1.7)

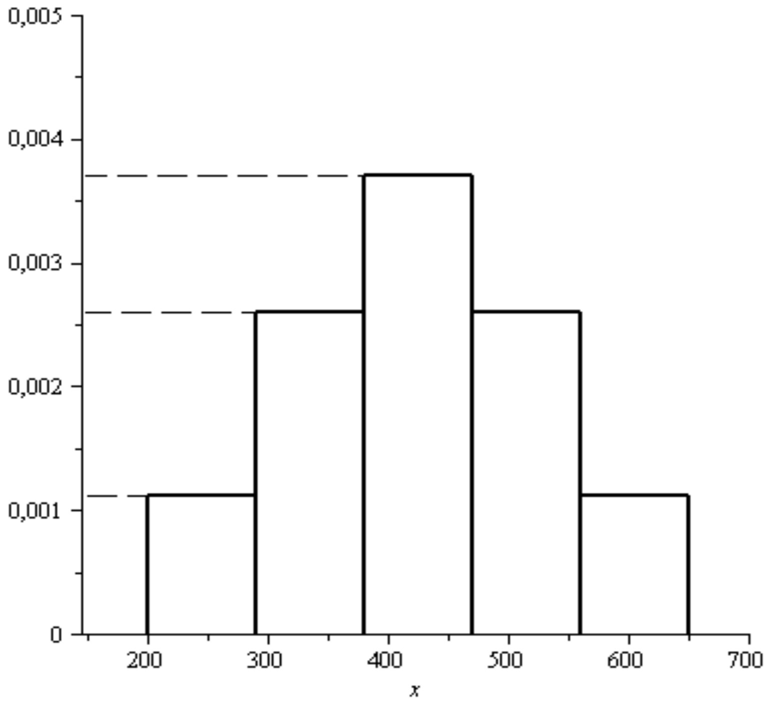


Рис. 1.7

■

4) Розмах вибірки $R = 450$ (обчислено в пункті 1).

1-ий спосіб) Знайдемо вибіркове середнє $\bar{x}_В$ для статистичного дискретного розподілу вибірки (див. табл. 1.4) за формулою (1.5):

$$\bar{x}_В = \frac{1}{30}(200 \cdot 1 + 250 \cdot 2 + 300 \cdot 3 + 350 \cdot 4 + 400 \cdot 4 + 450 \cdot 6 +$$

$$+500 \cdot 4 + 550 \cdot 3 + 600 \cdot 1 + 650 \cdot 2) = \frac{1}{30} \cdot 12850 = \frac{1285}{3} \approx 428{,}333;$$

Далі обчислимо вибіркову дисперсію $D_В$ для статистичного дискретного розподілу вибірки (див. табл. 1.4) за формулою (1.13):

$$D_B = \frac{1}{30}(\,(200 - 428{,}333)^2 \cdot 1 + (250 - 428{,}333)^2 \cdot 2 + (300 - 428{,}333)^2 \cdot 3 + (\,350 - 428{,}333)^2 \cdot 4 + (400 - 428{,}333)^2 \cdot 4 + (450 - 428{,}333)^2 \cdot 6 +$$

$$+(500 - 428{,}333)^2 \cdot 4 + (550 - 428{,}333)^2 \cdot 3 + (600 - 428{,}333)^2 \cdot 1 +$$

$$+(650 - 428{,}333)^2 \cdot 2) = \frac{1}{30} \cdot 388416{,}67 = 12947{,}22;$$

$$\sigma_{\text{В}} = \sqrt{D_B} = \sqrt{12947{,}22} \approx 113{,}786.$$

Згідно (1.20) виправлена дисперсія буде дорівнювати:

$$S^2 = \frac{30}{29} \cdot 12947,22 \approx 13393{,}676,\ S = \sqrt{S^2} = \sqrt{13393{,}676} \approx 115{,}73.$$

2-ий спосіб) Знайдемо вибіркове середнє $\bar{x}_{\text{В}}$, $D_{\text{В}}$, для статистичного дискретного розподілу вибірки (див. табл. 1.4) іншим способом. Оскільки варіанти в розподілі є досить великими числами і утворюють арифметичну прогресію, то для зручності обчислень зробимо заміну варіант згідно формули (1.6). Отже, перейдемо до нових варіант u_i за наступною формулою

$$u_i = \frac{x_i - 450}{50}.$$

Знайдемо умовні моменти розподілу і вибіркове середнє та вибіркову дисперсію на основі розрахунків і запишемо їх в таблиці 1.8:

Таблиця 1.8

x_i	n_i	u_i	$u_i n_i$	u_i^2	$u_i^2 n_i$
200	1	-5	-5	25	25
250	2	-4	-8	16	32
300	3	-3	-9	9	27
350	4	-2	-8	4	16
400	4	-1	-4	1	4
450	6	0	0	0	0
500	4	1	4	1	4
550	3	2	6	4	12
600	1	3	3	9	9
650	2	4	8	16	32
	$\sum_{i=1}^{10} n_i = 30$		$\sum_{i=1}^{10} u_i n_i = -13$		$\sum_{i=1}^{10} u_i^2 n_i = 161$

$$\bar{u}_{\text{в}} = \frac{1}{30} \cdot (-13) = \frac{-13}{30} \approx -0{,}4333;$$

Тоді вибіркове середнє $\bar{x}_{\text{в}}$ обчислюють за наступною формулою (1.7):

$$\bar{x}_{\text{в}} = h\bar{u}_{\text{в}} + C = 50 \cdot (-0{,}4333) + 450 = 428{,}333.$$

Далі обчислимо вибіркову дисперсію $D_{\text{В}}$ для статистичного дискретного розподілу вибірки (див. табл. 1.8) за формулою (1.15):

$$D_B[x] = \frac{1}{30} \cdot 161 - (-0{,}4333)^2 = 5{,}3667 - 0{,}1878 = 5{,}1789.$$

Коефіцієнт варіації обчислимо згідно формули (1.21):

$$V = \frac{\sigma_{\text{в}}}{\bar{x}_{\text{в}}} \cdot 100\% = \frac{113{,}786}{428{,}333} \cdot 100\ \% \approx 26{,}56\ \%.$$

Моду обчислимо за формулою (1.12), враховуючи, що: x_0 – початок модального інтервала (якому відповідає найбільша частота), x_0 = 380; n_i – частота у модальному інтервалі, n_i= 10; n_{i-1} , n_{i+1}– частоти в попередньому і наступному інтервалах відповідно, n_{i-1}= 7, n_{i+1} = 7; h – довжина інтервала, h = 90. Отже,

$$Mo = x_0 + h\frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} = 380 + 90\frac{10 - 7}{(10 - 7) + (10 - 7)} =$$

$$= 425.$$

Медіану обчислимо за формулою (1.11), враховуючи, що: x_0 – фактична нижня границя медіанного інтервала, x_0 = 380 (оскільки всього даних 30, то їх половина $30\,/\,2\ =\ 15$; у статистичному ряді до початку третього інтервала міститься 10 даних, тому медіана повинна знаходитися в третьому інтервалі); T_{i-1} – сума частот, що накопичена до початку медіанного інтервала, $T_{i-1} = 3 + 7 = 10$; n_i – частота в медіанному інтервалі, $n_i = 10$. Тоді:

$$Me = x_0 + h\frac{n/2 - T_{i-1}}{n_i} = 380 + 90\frac{15-10}{10} = 425.\blacksquare$$

РОЗДІЛ 2. СТАТИСТИЧНІ ОЦІНКИ ПАРАМЕТРІВ РОЗПОДІЛІВ

Література: [6, розділ 6, т. 13], [7, розділ 9], [14, розділ 13]. [16, розділ 9, п. 9.2-9.3], [17, розділи 5-9].

Базові поняття. Статистичні оцінки невідомих параметрів розподілу. Довірчі інтервали.

Основні задачі. Статистичне оцінювання (точкове, інтервальне) параметрів генеральної сукупності.

ЩО МАЄ ЗНАТИ ТА ВМІТИ СТУДЕНТ

1. Знання на рівні понять, означень, формулювань. Статистичні оцінки та їх властивості. Статистичні оцінки для математичного сподівання та дисперсії, моментів випадкових величин. Ефективні, оптимальні оцінки; нерівність Крамера-Рао. Методи одержання статистичних оцінок. Довірчі інтервали невідомих параметрів розподілу. Побудова довірчого інтервалу параметрів нормального розподілу із заданою надійністю.

2. Уміння в розв'язуванні задач. За статистичними розподілами знаходити точкові оцінки параметрів розподілу. Застосовувати методи побудови статистичних оцінок невідомих параметрів розподілу. Будувати довірчі інтервали для параметрів нормальних спостережень із заданою надійністю.

ОСНОВНІ ТЕОРЕТИЧНІ ВІДОМОСТІ

2.1. Точкові оцінки

Однією із задач математичної статистики є одержання на основі результатів спостережень випадкової величини певної інформації про значення параметра розподілу або про значення деякої функції від цього параметра. Задача полягає у побудові певного наближення до невідомого параметра, причому такі наближення називаються *оцінками*, а підхід – *точковим оцінюванням*.

Нехай у результаті n незалежних експериментів отримано вибірку $X = (x_1, \dots, x_n)$. Виникає питання, як на основі цієї вибірки оцінити невідомий параметр θ (наприклад, математичне сподівання, дисперсію тощо) розподілу випадкової величини *X*?

Означення 2.1. *Статистикою* або *оцінкою* $\hat{\theta}$ невідомого параметра θ називається функція $\hat{\theta} = f(x_1, \dots, x_n)$ від змінних $x_1, \dots, x_n$.

Зауваження 2.1. Якщо при кожному n спостерігається кратна вибірка *X* об'єму n, а спосіб, яким утворена оцінка, один і той самий (не залежить від об'єму вибірки n), то поняття "оцінка" використовують також у розумінні як "послідовність оцінок", що утворені за одним правилом при різних значеннях об'єму вибірки. Послідовності оцінок позначаються як $\hat{\theta}_n$.

Означення 2.2. Оцінка $\hat{\theta}$ називається *незміщеною*, якщо її математичне сподівання збігається з точним значенням θ:

$$M\hat{\theta} = \theta.$$

Якщо ж $M\hat{\theta} \neq \theta$, то оцінка називається *зміщеною* і величина $\hat{\theta} - \theta$ називається *зміщенням* оцінки.

Означення 2.3. Оцінка $\hat{\theta}_n$ називається *асимптотично незміщеною*, якщо має місце асимптотична збіжність середніх:

$$M\,\hat{\theta}_n \to \theta, n \to \infty.$$

Означення 2.4. Оцінка $\hat{\theta}_n$ називається *консистентною* (або *спроможною*), якщо вона збігається за ймовірністю до істинного значення θ:

$$\hat{\theta}_n \xrightarrow{p} \theta, n \to \infty,$$

тобто $P_\theta\left(\left|\hat{\theta}_n - \theta\right| \geq \varepsilon\right) \to 0, n \to \infty, \forall\, \varepsilon > 0.$

Означення 2.5. Оцінка $\hat{\theta}_n$ називається *сильно консистентною*, якщо вона збігається з імовірністю 1 до істинного значення θ :

$$\hat{\theta}_n \to \theta \text{ з ймовірністю 1 при } n \to \infty,$$

тобто $P_\theta(\exists \lim_{n\to\infty} \hat{\theta}_n = \theta) = 1.$

Зауваження 2.2. Наведені властивості стосуються оцінок $\widehat{\theta}$ для значення невідомого параметру *θ*. У випадку, коли цей параметр – векторний, доцільно розглядати також оцінки $\widehat{\tau}$ для значень деякої функції $\tau=\tau(\theta)$ від параметра *θ*. Сформульовані вище означення поширюються також і на дану схему, якщо замінити θ на $\tau(\theta)$, а $\widehat{\theta}$ на $\widehat{\tau}$.

Оцінюваний параметр може мати кілька точкових незміщених статистичних оцінок.

Для невідомого математичного сподівання $M(X)$ досліджуваної ознаки X *вибіркове середнє*

$$\bar{x}_{\text{В}} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

є незміщеною і консистентною оцінкою.

Для невідомої дисперсії $D(X)$ консистентною оцінкою є *вибіркова дисперсія*

$$D_{\text{В}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_{\text{В}})^2.$$

Проте ця оцінка є зміщеною. Тому для оцінки невідомої дисперсії застосовують *виправлену дисперсію*

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_{\text{В}})^2.$$

Оцінкою для середнього квадратичного відхилення $\sigma(X)$ є *виправлене вибіркове середнє квадратичне відхилення*

$$\sigma_{\text{В}} = \sqrt{S^2}.$$

2.2. Ефективні оцінки

При деяких припущеннях щодо функцій розподілу $\{F(x,\theta), \theta \in \Theta\}$ може бути отримана зручна оцінка для оцінювання, яка відома як *нерівність Рао-*

Крамера. Нехай $X = (x_1, \dots, x_n)$ – вибірка об'ємом n із розподілу з щільністю $p(x, \theta)$. Оскільки $x_1, \dots, x_n$ незалежні у сукупності, то сумісна щільність розподілу дорівнює

$$L(x, \theta) = \prod_{k=1}^{n} p(x_k, \theta).$$

У статистичних задачах $L(x, \theta)$ розглядається як функція двох аргументів $x \in R^n, \theta \in \Theta$, і називається *функцією вірогідності*.

Означення 2.6. *Вибірковою функцією вірогідності* називається випадкова величина, що отримується в результаті підстановки у функцію вірогідності замість аргумента x значення вибірки як випадкового вектора:

$$L(X, \theta) = L(x, \theta)|_{\,x=X}.$$

Якщо вибірку $x_1, \dots, x_n$ утворюють дискретні випадкові величини, то функція вірогідності визначається як добуток ймовірностей

$$L(x, \theta) = \prod_{k=1}^{n} P(x_k, \theta).$$

Припустимо, що для функції вірогідності виконуються *умови регулярності:*

1. Множина тих значень вибірки X, для яких функція вірогідності $L(X, \theta)$ строго додатна, не залежить від θ.
2. Функція вірогідності $L(X, \theta)$ двічі неперервно диференційована за параметром θ.
3. Функція впливу $U(X, \theta) = \frac{\partial}{\partial \theta} \ln L(X, \theta)$ інтегровна у квадраті: $MU^2(X, \theta) < \infty \; \forall \theta$.
4. Знак похідної за параметром θ можна внести під знак інтегралів $ML(X, \theta)$ за аргументом x від функції вірогідності $L(x, \theta)$, а також від її похідних за θ.

Теорема 2.1 (про нерівність та критерій Крамера-Рао). Нехай виконані умови регулярності.

(а) Для довільної незміщеної оцінки $T = T(X)$ функції $\tau = \tau(\theta)$ невідомого параметра $\theta \in \Theta$ справедлива нерівність Крамера-Рао

$$M(T-\tau)^2 \equiv DT \geq \frac{\tau_\theta^2(\theta)}{I(\theta)},$$

де $\tau_\theta(\theta) = \frac{d}{d\theta}\tau(\theta),\ I(\theta) = -M\frac{\partial^2}{\partial\theta^2}\ln L(X,\theta)\,,\ \forall\theta \in \Theta$ – інформація за Фішером у вибірці X.

(б) Рівність у нерівності (а) має місце тоді й тільки тоді, коли оцінка T є лінійною функцією від функції впливу вибірки:

$$T(X) - \tau(\theta) = c(\theta)U(X,\theta)\ \text{м. н.}, \forall\theta \in \Theta,$$

для деякої дійсної $c(\theta)$, причому у випадку рівності $c(\theta) = \tau_\theta(\theta)/I(\theta)$.

Означення 2.7. Оцінка $T(X)$ функції $\tau(\theta)$ невідомого параметра $\theta \in \Theta$ називається *ефективною* оцінкою параметричної функції τ_θ, якщо нерівність Крамера-Рао для неї є рівністю, тобто у випадку, коли ця оцінка має найменше можливе значення дисперсії у класі всіх незміщених оцінок.

2.3. Достатні статистики та оптимальність

Означення 2.8. Статистика $T = T(X)$ називається *достатньою статистикою* для вибірки X, якщо вибіркова функція вірогідності має вигляд

$$L(X,\theta) = g(T(X),\theta)\,h(X)\,j(\theta)$$

для деяких вимірних невід'ємних функцій g, h, j.

Теорема 2.2 (про властивість достатньої статистики). Якщо $T(X)$ – достатня статистика, то умовний розподіл

$$P(X = x|T(X) = t) = l(x,t)$$

не залежить від θ.

Означення 2.9. Достатня статистика $T = T(X)$ називається *повною статистикою*, якщо для кожної вимірної функції $\varphi(\cdot)$ із тотожностей $M\varphi(T) = 0,\ \forall\theta \in \Theta$, випливає, що $\varphi(T) = 0$ майже напевне.

Означення 2.10. Статистика T^* з класу всіх незміщенних оцінок дійсної параметричної функції $\tau(\theta)$ називається *оптимальною незміщеною оцінкою* значення $\tau(\theta)$, або ж *незміщеною оцінкою з рівномірно найменшою дисперсією*, якщо вона має найменшу дисперсію у класі всіх незміщених оцінок.

Теорема 2.3 (про оптимальність повної достатньої статистики). Якщо існує повна достатня статистика, то довільна вимірна функція від неї є оптимальною оцінкою свого математичного сподівання.

2.4. Методи знаходження статистичних оцінок

Розглянемо два методи одержання статистичних оцінок: метод моментів і метод максимальної вірогідності.

1. *Метод моментів* є спеціальним методом побудови оцінок невідомих числових параметрів, який спирається на асимптотичні властивості вибіркових моментів.

Нехай розподіл $F(x,\theta)$ випадкової величини X залежить від векторного параметра $\theta = (\theta_1, \dots, \theta_k)$. Тоді на основі розподілу $F(x,\theta)$ можна знайти, за умови їх існування, теоретичні моменти $MX, \dots, MX^k$. Ці теоретичні моменти будуть залежати від невідомих параметрів, тобто $MX^k = \alpha_k(\theta_1, \dots, \theta_k)$. За результатами спостережень $x_1, \dots, x_n$ можна знайти відповідні емпіричні моменти $\hat{\alpha}_k = \overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} x_i^k$. Крім того $\hat{\alpha}_k$ є незміщеними оцінками відповідних теоретичних моментів α_k і $\hat{\alpha}_k \xrightarrow{p} \alpha_k, n \to \infty$, тобто при великих n має місце наближена рівність $\alpha_k \approx \hat{\alpha}_k$. Тому *оцінками за методом моментів* для невідомих параметрів $(\theta_1, \dots, \theta_k)$. називають розв'язок системи рівнянь (2.1):

$$\begin{cases} \alpha_1(\theta_1, \dots, \theta_k) = \overline{X}, \\ \alpha_2(\theta_1, \dots, \theta_k) = \overline{X^2}, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \alpha_k(\theta_1, \dots, \theta_k) = \overline{X^k}. \end{cases} \qquad (2.1)$$

Якщо теоретичні моменти є неперервними функціями параметрів, то оцінки за методом моментів будуть консистентними.

2. *Метод максимальної вірогідності.* Визначення оцінки максимальної вірогідності грунтується на принципі максимальної вірогідності – “те, що спостерігається, є найбільш імовірним серед усіх можливих альтернатив”.

Означення 2.11. *Оцінкою максимальної вірогідності* (ОМВ, EML) параметру θ за вибіркою X називається статистика

$$\hat{\theta} = \hat{\theta}(X) \equiv \arg\max_{\theta\in\Theta} L(X,\theta),$$

де $L(X,\theta)$ – вибіркова функція вірогідності, тобто така статистика $\hat{\theta} = \hat{\theta}(X)$, що

$$L(X,\theta) \le L\big(X,\hat{\theta}\big), \forall\theta \in \Theta.$$

У випадку кратної вибірки ОМВ позначається як θ_n, де n – об’єм вибірки.

Необхідною умовою існування ОМВ є припущення існування точки максимуму. ОМВ визначена однозначно за умови єдиності цієї точки.

Теорема 2.4 (про рівняння максимальної вірогідності). Якщо $\Theta \subset R^d$, максимум в означенні оцінки максимальної вірогідності досягається всередині параметричної множини, а функція вірогідності диференційовна, то ОМВ задовольняє рівняння максимальної вірогідності:

$$\frac{\partial}{\partial\theta}\ln L(X,\theta)|_{\theta=\hat{\theta}} = 0,$$

тобто ОМВ є коренем функції впливу $U(X,\theta)$:

$$U\big(X,\hat{\theta}\big) = 0.$$

У випадку векторного параметра рівняння максимальної вірогідності є векторним, і перетворюється на систему рівнянь максимальної вірогідності для кожної координати вектора-градієнта.

Для знаходження ОМВ, знаходимо спочатку стаціонарні точки функції $U(X,\theta)$ і вибираємо ті із них, в яких ця функція має максимум. Якщо стаціонарна точка одна, то у більшості випадків вона і буде точкою максимуму. Таким чином, одержуємо систему:

$$\begin{cases} \frac{\partial \ln L(X,\theta)}{\partial \theta_1} = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots, \\ \frac{\partial \ln L(X,\theta)}{\partial \theta_k} = 0. \end{cases} \tag{2.2}$$

Ці рівняння і є рівняннями максимальної вірогідності. У більшості випадків, розв'язки системи (2.2) і є оцінками максимальної вірогідності невідомих параметрів $\theta_1, \dots, \theta_k$. При розв'язуванні рівнянь вірогідності необхідно відкидати розв'язки вигляду $\theta = const$ і розглядати лише ті розв'язки, які залежать від $X_1, \dots, X_n$ і попадають в область допустимих значень параметра Θ.

2.5. Деякі розподіли функцій від нормально розподілених випадкових величин

Нехай $X_1, \dots, X_n$ – незалежні в сукупності стандартні нормальні випадкові величини: $X_k \simeq N(0,1), k = 1, \dots, n$.

Означення 2.12. Випадкова величина

$$\chi^2(n) = X_1^2 + \cdots + X_n^2$$

має *розподіл Пірсона або χ^2-розподіл* ("хі" - квадрат) з n ступенями вільності.

Число ступенів вільності (свободи) визначають як різницю між числом підсумованих випадкових величин і лінійних зв'язків, які обмежують вільність зміни цих величин.

Означення 2.13. Випадкова величина $T(n)$ має $t-$ розподіл, або *розподіл Стьюдента*, із n ступенями вільності, якщо її можна подати у вигляді

$$T(n) = \frac{\zeta}{\sqrt{\frac{1}{n}\chi^2(n)}},$$

де $\zeta \simeq N(0,1)$ – стандартна нормальна величина, а $\chi^2(n)$ – незалежна від неї величина із χ^2-розподілом та n ступенями вільності.

Означення 2.14. Випадкова величина F має *розподіл Фішера*, або ж *Снедекора – Фішера*, із (n, m) ступенями вільності, якщо її можна подати у вигляді

$$F = \frac{\frac{1}{n}\chi^2(n)}{\frac{1}{m}\chi^2(m)},$$

де $\chi^2(n),\ \chi^2(m)$ – незалежні величини із χ^2 −розподілом та (n, m) ступенями вільності відповідно.

2.6. Інтервальні оцінки

Точкові оцінки параметрів розподілу не дають можливості зробити висновки про їхню точність та надійність, оскільки є випадковими величинами. Щоб мати уявлення про точність та надійність оцінки, у математичній статистиці використовують рівень довіри та довірчі інтервали.

Означення 2.15. *Інтервальною оцінкою* (або *довірчим інтервалом*) невідомого параметра θ називається випадковий інтервал $(\hat{\theta}_1, \hat{\theta}_2)$, який покриває невідоме точне значення θ із заданою ймовірністю (надійністю) γ, тобто

$$P\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = \gamma.$$

Випадкові величини $\hat{\theta}_1$ та $\hat{\theta}_2$, які є функціями вибіркового вектору $X = (x_1, \dots, x_n)$, називаються відповідно *нижньою* та *верхньою* межами довірчого інтервалу; ймовірність γ ще називають *довірчою ймовірністю*.

Точність оцінювання параметра θ характеризується довжиною довірчого інтервалу $\hat{\theta}_2 - \hat{\theta}_1$ і залежить від обсягу вибірки n та надійності γ.

На практиці надійність оцінки γ беруть такою, щоб подію з ймовірністю γ можна було вважати практично вірогідною, наприклад, γ = 0,95; 0,99; 0,999. Часто величину β записують у вигляді $\gamma = \alpha - 1 \in (0,1)$. Мале число α називається *рівнем довіри* оцінки.

Розглянемо оцінки, які найбільш часто використовують.

1) *Інтервальною оцінкою з надійністю γ невідомої ймовірності успіху p у схемі Бернуллі за відносною частотою $\hat{p} = \nu_n/n$* називають довірчий інтервал

$$\frac{\hat{p}+\frac{1}{2}t^2-t_\gamma\sqrt{\hat{p}(1-\hat{p})n+\frac{1}{4}t_\gamma^2}}{n+t_\gamma^2} < p < \frac{\hat{p}+\frac{1}{2}t^2+t_\gamma\sqrt{\hat{p}(1-\hat{p})n+\frac{1}{4}t_\gamma^2}}{n+t_\gamma^2}, \qquad (2.3)$$

де n – загальна кількість випробувань; t_γ – значення аргументу функції Лапласа (див. додаток 1), при якому $\Phi(t_\gamma) = \frac{\gamma}{2}$, а $\hat{p} = \nu_n/n$ – відносна частота.

Зауваження 2.3. При великих значеннях n для оцінки ймовірності p у схемі Бернулі, можна прийняти такий довірчий інтервал

$$\hat{p} - t_\gamma\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + t_\gamma\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}. \qquad (2.4)$$

Нехай $X_1, \dots, X_n$ – незалежні і нормально розподілені випадкові величини: $X_k \simeq N(a, \sigma^2), k = 1, \dots, n.$

2) *Інтервальною оцінкою з надійністю γ невідомого математичного сподівання a нормально розподіленої випадкової величини X з відомою дисперсією σ^2* називають довірчий інтервал

$$\bar{x}_\text{В} - t_\gamma\frac{\sigma}{\sqrt{n}} < a < \bar{x}_\text{В} + t_\gamma\frac{\sigma}{\sqrt{n}}, \qquad (2.5)$$

де $\bar{x}_\text{В} = \frac{1}{n}\sum_{i=1}^{n} x_i$ – вибіркове середнє, $t_\gamma\frac{\sigma}{\sqrt{n}} = \delta$ – точність оцінки, t_γ – значення аргументу функції Лапласа (див. додаток 2), при якому $\Phi\left(t_\gamma\right) = \frac{\gamma}{2}$, n – обсяг вибірки.

3) *Інтервальною оцінкою з надійністю γ невідомого математичного сподівання a нормально розподіленої випадкової величини X з невідомою дисперсією σ^2* називають довірчий інтервал

$$\bar{x}_\text{В} - t_{\gamma,k}\frac{s}{\sqrt{n}} < a < \bar{x}_\text{В} + t_{\gamma,k}\frac{s}{\sqrt{n}}, \qquad (2.6)$$

де $t_{\gamma,k}\frac{s}{\sqrt{n}} = \delta$ – точність оцінки, $t_{\gamma,k}$ знаходиться згідно додатку 4 за заданими надійністю γ і числом ступенів вільності $k = n - 1$, $s = \sigma_\text{В}$ – виправлене вибіркове середнє квадратичне відхилення.

4) *Інтервальною оцінкою з надійністю γ невідомої дисперсії σ^2 нормально розподіленої випадкової величини X з відомим математичним сподіванням* a називають довірчий інтервал

$$\frac{n\hat{\sigma}^2}{h''} < \sigma^2 < \frac{n\hat{\sigma}^2}{h'}, \tag{2.7}$$

де n – обсяг вибірки, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)$ – оцінка дисперсії, числа h' і h'' вибираються за допомогою додатку 3 так, щоб виконувались умови

$$P\{\chi^2(n) < h'\} = \frac{1-\gamma}{2} \text{ і } P\{\chi^2(n) < h''\} = \frac{1+\gamma}{2},$$

$\chi^2(n)$ – незалежна величина із χ^2-розподілом та n ступенями вільності.

5) *Інтервальною оцінкою з надійністю γ невідомої дисперсії σ^2 нормально розподіленої випадкової величини X з невідомим математичним сподіванням* a називають довірчий інтервал

$$\frac{(n-1)s^2}{h''} < \sigma^2 < \frac{(n-1)s^2}{h'}, \tag{2.8}$$

де n – обсяг вибірки, S^2 – виправлена вибіркова дисперсія, числа h' і h'' вибираються за допомогою додатку 3 так, щоб виконувались умови

$$P\{\chi^2(n-1) < h'\} = \frac{1-\gamma}{2} \text{ і } P\{\chi^2(n-1) < h''\} = \frac{1+\gamma}{2},$$

$\chi^2(n)$ – незалежна величина із χ^2 −розподілом та n ступенями вільності.

.

ВАРІАНТИ ІНДИВІДУАЛЬНИХ ЗАВДАНЬ (ІЗ № 2)

За вибірковими даними випадкової величини X (див. табл. 2.1) індивідуальних завдань потрібно:

1) знайти точкові незміщені оцінки математичного сподівання a, дисперсії σ^2 і оцінку середнього квадратичного відхилення σ генеральних сукупностей;

2) для заданого варіанта V за вибіркою X з надійністю

$$\gamma = \begin{cases} 0{,}9,\ V \le 10, \\ 0{,}95,\ 10 < V \le 20, \\ 0{,}99,\ V > 20, \end{cases}$$

побудувати:

а) довірчій інтервал для математичного сподівання, коли дисперсія невідома і за вибіркою знайдена виправлена вибіркова дисперсія S^2;

б) довірчий інтервал для невідомої дисперсії та середнього квадратичного відхилення.

Таблиця 2.1

Номер варіанту	Дані вибірки
1	2,0 4,8 5,2 3,8 3,5 3,2 3,2 3,9 4,9 2,8 3,7 1,8 3,4 2,3 3,2 4,5 0,5 3,3 2,8 2,5 1,4 3,2 3,5 2,2 2,3 3,5 3,5 4,1 4,4 2,3 1,9 2,2 3,8 3,4 2,2 3,1 2,1 2,1 3,2 2,5 2,1 2,9 2,8 3,1 4,3 2,8 4,0 2,3 2,7 2,4
2	34,0 36,1 34,3 34,4 34,1 35,6 35,9 34,4 35,2 34,2 35,8 35,2 34,3 34,8 35,1 34,5 34,6 34,2 34,4 34,2 34,8 35,0 34,8 37,7 34,3 36,0 36,0 35,1 34,2 34,2 34,4 34,3 34,0 34,0 36,4 34,1 35,1 34,7 34,1 34,1 34,2 34,2 34,6 35,0 36,5 34,1 34,4 34,2 34,6 35,0
3	12,8 12,3 14,7 12,2 13,2 12,0 15,2 13,2 12,3 13,7 14,3 12,5 12,2 13,9 16,2 14,4 13,0 12,3 12,3 15,1 14,2 12,5 15,9 12,0 14,8 12,1 19,9 12,8

	12,8 12,8 14,4 15,7 12,2 12,2 15,0 12,4 12,5 12,9 13,6 12,2 13,4 12,1 13,1 12,6 14,2 13,6 12,0 16,4 12,3 14,2
4	40,2 31,8 31,2 29,1 25,7 37,5 49,1 28,9 36,7 30,6 44,1 31,1 44,9 40,0 31,0 50,9 41,3 46,0 33,8 28,0 30,9 34,5 48,8 32,3 40,9 35,8 43,8 28,1 27,0 33,0 29,8 28,5 28,8 33,4 32,5 46,6 39,4 38,6 41,6 41,4 36,1 31,8 47,6 34,0 28,2 28,2 42,1 39,2 42,0 24,0
5	14,6 15,2 14,1 14,1 15,0 14,0 15,0 15,1 15,5 15,9 15,5 14,2 14,0 14,5 14,7 15,5 15,5 14,2 14,4 14,4 14,4 16,4 15,7 14,4 14,1 15,5 14,9 15,1 15,1 14,8 14,4 16,3 14,1 14,1 14,6 14,2 14,9 14,7 14,8 15,5 16,4 14,6 14,5 14,9 14,2 15,1 14,4 16,0 16,3 15,5
6	40,6 29,8 27,6 32,5 36,1 28,4 30,2 32,0 31,2 28,6 34,2 35,3 34,2 32,5 37,6 31,0 32,2 37,4 32,4 31,5 32,2 32,8 34,4 25,5 31,0 36,3 30,8 34,3 30,2 33,2 32,5 29,3 32,1 30,1 36,5 27,2 34,0 30,9 30,9 27,6 34,4 36,3 28,9 28,4 32,3 34,7 30,0 29,2 31,7 30,4
7	28,1 31,9 26,2 31,2 26,3 23,8 22,9 23,1 34,1 26,8 28,6 31,5 27,5 33,9 24,9 28,6 30,6 27,6 25,0 28,0 26,4 26,8 28,9 27,4 24,4 22,7 23,0 24,9 25,7 23,5 26,1 22,7 28,4 35,4 29,6 25,1 26,1 25,6 28,2 35,3 33,0 39,1 29,5 36,2 24,7 23,6 38,6 23,0 22,4 34,6
8	25,6 29,3 24,0 26,5 27,1 25,2 29,1 24,0 29,6 27,6 30,3 25,1 26,1 24,2 25,9 27,5 31,5 25,7 26,5 24,1 28,4 24,2 28,4 24,2 25,3 24,4 25,0 28,6 24,9 29,3 30,6 24,1 26,0 25,4 26,6 24,4 25,2 24,4 24,5 25,4 26,0 25,9 24,0 27,4 24,4 24,2 33,6 24,5 24,4 24,1
9	33,3 42,2 35,1 35,8 46,1 37,6 40,1 30,5 34,3 31,6 31,9 50,9 30,7 43,4 40,1 38,8 30,1 32,3 34,5 42,8 31,2 39,4 38,7 40,9 49,2 33,1 30,3 38,1 49,2 39,4 30,9 67,5 30,9 31,6 30,8 41,1 35,5 33,6 32,0 33,5 30,0 50,8 60,8 30,6 42,7 35,7 66,1 31,2 31,0 40,9
10	0,8 –0,1 –2,5 –1,0 –0,8 1,9 –2,1 0,3 3,5 0,5 –1,0 0,9 4,2 –3,2 –1,0 –5,4 –4,3 –6,1 –2,7 9,2 –3,4 –2,7 –1,9 –5,2 –12,8 –2,5 3,7

	–2,6 –1,5 0,3 0,7 –1,6 0,8 0,2 2,4 –3,4 1,9 0,7 –1,1 1,9 –5,0 0,0 10,2 3,7 –0,6 –6,1 –0,6 0,1 2,2 –3,2
11	38,8 39,2 44,5 43,6 36,5 39,2 42,2 55,2 43,7 42,3 38,9 39,5 36,7 37,1 37,0 44,8 39,1 41 ,9 38,0 42,4 44,2 39,1 40,1 37,6 36,1 44,9 36,5 38,3 36,4 37,1 40,3 40,3 45,6 58,8 38,1 38,1 56,0 41,6 44,3 49,9 41,1 40,6 39,9 38,4 38,0 39,3 36,4 39,1 50,9 43,8
12	30,3 34,3 33,2 33,9 33,1 28,8 33,0 32,6 34,4 30,8 33,7 28,8 36,6 28,2 31,6 34,0 32,9 35,3 33,5 31,5 31,9 33,5 33,2 35,7 32,0 31,0 30,3 31,5 33,5 29,4 32,8 34,3 29,6 32,9 31,1 33,7 32,6 33,2 34,1 29,5 30,6 31,5 32,8 33,2 29,3 36,0 32,8 34,8 34,8 32,5
13	33,1 57,2 26,5 20,3 25,5 20,2 20,0 26,1 25,4 18,9 29,5 27,5 34,5 18,1 21,7 22,0 21,7 21 ,1 22,9 18,8 42,2 18,1 18,0 27,9 34,5 20,2 27,8 18,8 20,4 19,2 28,8 23,6 35,8 30,0 23,1 19,3 19,9 19,6 37,1 21,4 32,9 20,8 22,2 38,6 24,0 41 ,3 29,1 21,7 24,7 21,0
14	22,4 17,5 17,6 18,7 18,3 24,3 18,5 43,2 20,2 20,0 21,7 15,1 18,3 24,3 15,0 15,1 29,3 14,6 18,2 14,4 19,3 15,0 18,6 14,1 18,9 27,0 14,5 16,0 20,0 30,3 17,0 27,7 15,9 28,4 22,0 17,2 21 ,5 14,9 24,8 14,1 15,1 21,9 14,8 18,1 14,8 14,2 15,0 14,6 14,5 15,8
15	32,4 33,5 29,1 32,7 30,6 31,1 30,6 30,2 30,9 32,1 32,5 33,8 33,5 32,2 35,1 31,5 33,5 28,7 32,6 30,6 30,5 27,9 33,0 30,8 31,5 28,8 30,9 33,4 31,4 27,4 25,5 29,5 30,4 32,5 32,4 36,5 32,3 30,8 31,7 31,1 28,1 27,7 28,5 31,3 31,6 35,1 29,8 33,0 28,7 28,6
16	46,4 57,6 35,6 36,1 50,3 58,1 56,6 52,9 35,2 57,4 44,8 56,7 53,4 38,9 58,4 57,4 44,2 55,9 51,4 56,1 54,3 57,5 50,3 47,3 35,2 53,6 35,9 40,6 54,0 35,9 44,5 49,8 50,6 36,9 47,4 50,2 36,8 42,6 35,6 55,2 52,1 57,8 50,2 37,4 37,6 51,6 46,1 49,6 36,6 44,9
17	10,4 7,2 10,7 –1,0 –3,3 2,9 0,7 –6,3 –5,1 –8,6 –5,4 –2,3 –0,1 4,4 9,7 2,3 –7,8 –3,8 2,5 4,3 10,9 –2,0 6,1 5,9 10,8 6,6 –2,1 7,2 –1,5 11,6 7,7

	0,4 –0,9 –4,0 –5,8 1,5 –1,1 7,2 –3,1 –1,2 –2,4 0,0 9,7 –0,8 –1,3 7,3 0,4 6,0 7,5 –3,1 –1,6
18	32,4 24,1 32,1 28,5 27,9 35,0 29,9 31,8 25,8 25,2 19,8 27,6 30,3 27,0 27,2 31,2 28,5 33,3 30,0 29,4 29,5 36,1 28,6 30,8 33,1 34,1 38,9 31,2 32,8 23,3 32,0 33,7 24,4 31,8 30,0 28,1 27,4 40,1 34,4 27,0 35,1 29,8 27,8 33,6 36,0 28,1 25,3 37,1 33,2 32,3
19	33,8 16,9 13,5 17,0 18,9 23,3 21,3 21,6 16,8 24,1 13,2 19,7 16,6 23,9 17,4 32,0 28,2 18,5 18,3 11,8 19,1 17,0 24,9 18,0 28,4 29,3 17,8 22,4 16,1 20,7 14,2 14,1 23,6 25,5 17,8 27,0 35,0 15,7 16,4 24,1 22,9 26,8 26,8 20,3 12,1 24,8 4,8 33,5 18,9 19,7
20	8,6 7,3 7,1 –4,9 6,1 –2,3 –10,7 –1,4 2,8 12,1 6,4 3,1 7,3 –18,1 –4,0 9,9 –1,7 –9,1 3,0 1,8 –15,7 1,9 12,3 2,9 4,7 13,2 –10,3 –1,8 5,3 0,2 1,8 –12,1 1,8 6,5 7,3 22,0 8,5 –4,8 13,9 –13,8 8,0 18,0 7,5 –4,6 11,4 6,9 –0,5 20,4 21,3 12,2
21	32,1 45,2 42,1 40,1 40,7 45,2 41,1 41,8 35,5 47,1 36,4 49,0 44,6 35,0 47,9 39,4 45,1 38,6 40,2 36,7 46,6 36,1 40,5 38,2 36,2 44,9 39,0 47,9 40,6 41,2 34,7 36,8 44,3 32,3 48,3 45,0 42,8 33,3 49,0 37,1 46,3 44,6 48,3 39,1 35,0 41 ,8 46,6 40,3 42,5 35,0
22	50,5 57,4 45,1 52,6 56,2 52,3 62,0 56,3 54,8 39,4 53,5 49,9 43,0 61,4 38,5 43,7 47,1 64,9 57,1 38,9 42,0 58,0 58,2 44,3 39,4 38,7 47,0 44,7 42,7 64,5 44,5 50,3 56,5 54,3 62,5 39,6 49,2 57,9 38,0 46,8 42,8 58,6 52,6 57,0 54,8 56,4 64,8 55,6 52,1 51,1
23	36,9 23,5 20,1 17,2 19,2 49,6 28,1 19,2 25,2 34,5 35,9 29,4 22,8 24,7 36,9 44,6 31,5 27,4 37,0 17,9 32,3 26,6 34,0 30,5 25,3 27,5 28,0 34,6 26,9 16,8 18,4 24,5 20,6 23,0 5,1 25,9 30,0 16,8 23,6 27,6 25,5 40,3 27,2 29,1 25,1 39,0 34,2 16,8 36,1 22,1
24	37,3 31,1 40,5 30,4 35,9 34,7 46,7 30,5 38,0 42,2 36,7 40,9 39,7 34,1 34,7 46,2 43,3 44,7 23,7 32,5 31 ,5 43,6 40,3 36,4 38,6 32,4 30,9 37,2

	36,0 42,9 38,3 40,1 43,7 37,7 41,0 25,6 51,5 43,9 35,7 34,5 42,0 20,8 41,5 42,0 48,1 34,5 44,4 33,1 38,4 34,5
25	59,0 40,9 40,2 56,8 39,0 42,1 49,3 44,1 58,8 52,2 52,3 43,6 50,0 54,9 50,4 49,1 56,9 43,6 42,9 56,5 46,8 45,5 58,2 39,2 44,2 47,9 51,0 42,5 42,8 52,3 45,4 57,6 54,6 50,7 40,7 39,4 57,1 58,8 49,6 58,3 49,6 50,0 57,4 39,4 51,2 57,7 58,3 50,8 57,7 56,0
26	21,4 20,0 23,6 27,5 20,3 20,4 23,4 26,3 21,3 22,6 26,3 19,1 19,8 23,2 15,9 19,6 25,1 11,8 16,3 17,9 22,3 17,7 13,7 20,8 19,3 22,4 14,6 29,9 18,7 16,2 17,3 21,5 16,1 26,3 20,6 21,3 22,9 23,1 16,9 19,5 21,1 24,6 24,6 24,3 13,2 16,0 26,4 14,6 21,3 14,0
27	38,1 34,6 31 ,3 36,4 34,4 43,2 40,3 42,3 35,3 31,7 31,6 43,5 30,5 30,1 37,1 28,1 38,2 34,6 28,8 29,0 28,2 37,6 42,9 32,5 42,8 37,6 32,9 29,3 42,7 39,7 30,4 40,6 37,1 42,8 41,8 44,0 33,5 34,9 44,7 30,7 36,6 34,0 33,0 43,2 44,9 33,9 40,1 42,1 32,0 38,1
28	31,2 30,7 31 ,3 30,6 30,2 28,9 30,3 30,8 32,7 27,6 29,5 31,0 32,2 29,7 31,3 30,7 33,1 29,2 28,2 26,5 28,3 28,9 30,7 27,3 27,5 34,8 35,3 32,3 28,6 31,4 29,0 31,1 28,5 30,4 27,2 27,1 29,5 30,8 28,2 36,6 31,7 28,6 30,0 25,9 25,6 29,4 32,5 32,6 25,6 28,9
29	36,8 37,0 43,0 45,5 34,8 35,6 35,2 38,1 49,5 46,0 37,2 42,7 46,9 49,3 50,6 35,9 35,7 46,3 38,3 48,4 41 ,2 50,1 35,5 49,3 42,9 42,8 46,5 49,1 48,0 34,8 47,2 50,4 36,5 47,2 43,9 43,9 47,2 45,2 41,5 46,6 45,5 42,8 35,2 44,8 37,5 41,2 36,7 37,3 46,1 46,6
30	11,9 46,0 17,1 11,9 11,8 13,8 13,2 10,5 15,2 21,3 16,1 16,0 10,7 11,4 31 ,3 18,3 11,1 27,0 13,0 10,2 12,3 15,6 13,6 11,1 14,6 12,9 15,5 17,8 32,7 15,7 10,1 10,4 10,2 17,7 15,4 33,5 31,6 20,5 12,4 12,4 11,4 17,3 14,4 16,6 10,2 24,4 21 ,1 15,9 10,2 15,4

3) За допомогою методів одержання статистичних оцінок знайти оцінки невідомих параметрів розподілів.

1. Нехай $x_1, \dots, x_n$ – кратна вибірка з біноміального розподілу з параметрами k, p. Оцінити параметр p за методом моментів.

2. Нехай $x_1, \dots, x_n$ – кратна вибірка з розподілу Пуассона з параметром $\lambda > 0$. Оцінити параметр λ за методом максимальної вірогідності.

3. Знайти оцінку максимальної вірогідності для невідомої кількості спостережень n за біноміальною одноелементною вибіркою $X \simeq B(n, p)$ при відомому p.

4. Методом моментів знайти оцінку параметра показникового розподілу λ при $p(x, \lambda) = \lambda e^{-\lambda x}$ $(\lambda > 0, x > 0)$, якщо в результаті n експериментів випадкова величина прийняла значення $x_1, \dots, x_n$.

5. Для групи страхових полісів кількість вимог для кожного полісу моделюється розподілом Пуассона з інтенсивністю λ на рік, незалежно для кожного полісу. Така група з 1600 полісів викликала 185 вимог протягом минулого року. Знайти оцінку максимальної вірогідності для параметра λ.

6. Нехай $x_1, \dots, x_n$ – кратна вибірка з біноміального розподілу з параметром $p \in (0, 1)$. Оцінити параметр p методом максимальної вірогідності.

7. Нехай $X = (x_1, \dots, x_n)$, $Y = (y_1, \dots, y_n)$ – незалежні кратні вибірки з нормальними розподілами $N(\mu_1, \sigma^2)$ та $N(\mu_2, \sigma^2)$) відповідно. Знайти оцінку максимальної вірогідності для параметра $\theta = (\mu_1, \mu_2, \sigma^2)$.

8. Обчислити оцінку максимальної вірогідності для параметра θ за кратною вибіркою з нормального розподілу $N(\theta, 2\theta^2)$.

9. За вибіркою $x_1, \dots, x_n$ обсягу n знайти оцінку параметру розподілу із щільністю $p(x) = \frac{(\theta-1)^x}{\theta^{x+1}}, \theta > 0$, користуючись методом максимальної правдоподібності.

10. Нехай задано логарифмічно нормальний закон розподілу, з щільністю $p(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln x - a)^2}, x > 0$. Оцінки параметрів a і σ^2 знайти методом моментів. Обчислюючи ймовірності, урахувати, що випадкова величина $Y = \ln x - a$ розподілена за нормальним законом з нульовим математичним сподіванням і одиничною дисперсією.

11. Нехай $x_1, \dots, x_n$ – вибірка із геометричного розподілу $p(k,p) = (1-p)^k p, k = 0,1, \dots$. Оцінити параметр p за методом моментів.

12. Обчислити оцінку максимальної вірогідності для параметра θ за вибіркою $x_1, \dots, x_n$, де $P\{x_k = m\} = e^{-\theta}\frac{\theta^m}{m!}, m = 0,1, \dots$.

13. Нехай $x_1, \dots, x_n$ – вибірка з генеральної сукупності зі щільністю $p(x,\theta) = k(\theta)x^2 e^{-\frac{x^2}{\theta^2}}$, $x \geq 0, \theta \geq 0$. Знайти функцію $k(\theta)$, оцінку параметра θ методом максимальної вірогідності.

14. Нехай $x_1, \dots, x_n$ – вибірка із гамма-розподілу $p(x,\alpha,\beta) = \frac{1}{\beta^{\alpha+1}\Gamma(\alpha+1)} x^\alpha e^{-x/\beta}$, $(x > 0, \alpha > -1, \beta > 0)$. Оцінити параметри α і β за методом моментів.

15. Нехай $x_1, \dots, x_n$ – вибірка із гамма-розподілу $p(x,\alpha,\beta) = \frac{1}{\beta^{\alpha+1}\Gamma(\alpha+1)} x^\alpha e^{-x/\beta}$, $(x > 0, \alpha > -1, \beta > 0)$. Оцінити параметр β методом максимальної вірогідності.

16. Методом максимальної правдоподібності знайти оцінку параметра показникового розподілу $p(x,\lambda) = \lambda e^{-\lambda x}$ $(\lambda > 0, x > 0)$, якщо в результаті n експериментів випадкова величина прийняла значення $x_1, \dots, x_n$.

17. За вибіркою $x_1, \dots, x_n$ знайти методом максимальної вірогідності оцінку ймовірності успіху в схемі Бернуллі $(\theta = p)$.

18. Використовуючи метод моментів, знайти за вибіркою $x_1, \dots, x_n$, де $P\{x_k = m\} = e^{-\theta}\frac{\theta^m}{m!}, m = 0,1, \dots$, оцінку параметра θ.

19. Нехай $x_1, \dots, x_n$ – вибірка з генеральної сукупності зі щільністю $p(x,\theta) = k(\theta)x^2 e^{-\frac{x^2}{\theta^2}}$, $x \geq 0, \theta \geq 0$. Знайти функцію $k(\theta)$, оцінку параметра θ методом моментів.

20. Методом максимальної вірогідності знайти оцінку параметра θ розподілу $P\{x_k = m\} = \frac{(\theta-1)^m}{\theta^{m+1}}, m = 0,1, \dots, \theta > 1$.

21. За вибіркою $x_1, \dots, x_n$ обсягу n знайти оцінку невідомого параметру нормального розподілу $N(0,\ \sigma)$, користуючись методом максимальної вірогідності.

22. За вибіркою $x_1, \dots, x_n$ обсягу n знайти оцінку невідомого параметру нормального розподілу $N(m,1)$, користуючись методом максимальної вірогідності.

23. За вибіркою $x_1, \dots, x_n$ обсягу n знайти оцінку параметру розподілу із щільністю $p(x) = \frac{\theta^x}{(\theta+1)^{x+1}}, \theta > 0$, користуючись методом максимальної вірогідності.

24. Випадкова величина має логарифмічний нормальний закон розподілу, $p(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln x - a)^2}, x > 0$. Оцінки параметрів a і σ^2 знайти методом максимальної вірогідності. Обчислюючи ймовірності, урахувати, що випадкова величина $Y = \ln x - a$ розподілена за нормальним законом з нульовим математичним сподіванням і одиничною дисперсією.

25. Випадкова величина має закон розподілу Релея, $p(x) = \frac{x}{\sigma^2} e^{-\frac{x}{\sigma^2}}, x > 0$. Оцінку параметра σ^2 знайти методом максимальної вірогідності.

26. Знайти оцінку максимальної вірогідності параметра p біномного розподілу $P_m(x_i) = C_m^{x_i} p^{x_i} (1-p)^{m-x_i}$, x_i – число настання події в i-му експерименті, m – кількість випробувань в одному досліді, n – число дослідів.

27. Нехай $x_1, \dots, x_n$ – вибірка з генеральної сукупності зі щільністю $p(x) = \begin{cases} 0, x < 1, \\ \theta \left(\frac{1}{x}\right)^{\theta+1}, x \geq 1, \end{cases}$ $\theta > 0$. Знайти оцінку максимальної вірогідності параметра θ.

28. Нехай $x_1, \dots, x_n$ – вибірка з генеральної сукупності із щільністю $p(x, \beta, m) = \frac{\beta^m}{\Gamma(m)} x^{m-1} e^{-\beta x}, x > 0, \beta > 0, m > 0$. Знайти оцінку невідомих параметрів β і m за допомогою методу моментів. За умови, що при $n = 10$ величини $x_1, \dots, x_n$ набули значень: 0,05; 0,02; 0,04; 0,1; 0,02; 0,09; 0,06; 0,5; 0,25; 0,08.

29. Нехай відібрано 10 приладів з метою контролю деяких параметрів. Результати вимірів напруги показали: 120; 100; 95; 99; 110. Методом максимальної вірогідності знайти оцінку параметра a, якщо напруга X є випадковою величиною, розподіленою за нормальним законом.

30. Методом максимальної вірогідності знайти оцінку параметра p біномного розподілу $P\{x = i\} = \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}$, якщо в n_1 незалежних дослідах подія A настала m_1 раз і в n_2 незалежних дослідах m_2 раз.

РОЗВ'ЯЗАННЯ ТИПОВОГО ВАРІАНТА ІЗ № 2

Умова завдання

Нехай задано вибірку X обсягом $n = 50$:

13,1 12,6 14,2 13,6 12,0 16,4 12,3 14,2 14,1 12,2 13,3 12,4 12,6

13,5 14,8 12,6 21,8 12,9 14,1 12,5 13,8 19,1 15,8 13,8 14,8 15,1

12,0 13,3 17,5 15,8 13,3 12,3 12,8 14,0 12,9 12,7 16,2 14,5 19,0

20,0 13,5 13,3 13,1 12,7 13,0 17,0 18,7 17,0 12,6 13,1.

1) Потрібно знайти точкові незміщені оцінки математичного сподівання a та дисперсії σ^2 і оцінку середнього квадратичного відхилення σ генеральних сукупностей.

2) Побудувати з надійністю $\gamma = 0{,}9$:

а) довірчій інтервал для математичного сподівання, коли дисперсія невідома і за вибіркою знайдена виправлена вибіркова дисперсія S^2;

б) довірчий інтервал для невідомої дисперсії та середнього квадратичного відхилення.

Хід розв'язання

1) Розмістивши значення варіант, які потрапили у вибірку, у порядку зростання, дістанемо дискретний варіаційний ряд:

12 12,2 12,3 12,4 12,5 12,6 12,7 12,8 12,9 13 13,1 13,3 13,5 13,6 13,8

14 14,1 14,2 14,5 14,8 15,1 15,8 16,2 16,4 17 17,5 18,7 19 19,1 20 21,8.

Обчислимо частоту кожної варіанти із варіаційного ряду і складемо таблицю, тобто знайдемо дискретний статистичний розподіл вибірки (див. табл. 2.2).

Таблиця 2.2

x_i	12	12,2	12,3	12,4	12,5	12,6	12,7	12,8	12,9	13	13,1	13,3
n_i	2	1	2	1	1	4	2	1	2	1	3	4

x_i	13,5	13,6	13,8	14	14,1	14,2	14,5	14,8	15,1	15,8	16,2	16,4
n_i	2	1	2	1	2	2	1	2	1	2	1	1

x_i	17	17,5	18,7	19	19,1	20	21,8
n_i	2	1	1	1	1	1	1

За дискретним статистичним розподілом (див. табл. 2.2) обчислюємо вибіркове середнє $\bar{x}_В$, виправлену вибіркову дисперсію S^2 і виправлене вибіркове середнє квадратичне відхилення $\sigma_В$, використовуючи відповідно формули (1.5), та (1.20):

$$\bar{x}_В = \frac{1}{50}(12 \cdot 2 + 12{,}2 \cdot 1 + 12{,}3 \cdot 2 + 12{,}4 \cdot 1 + 12{,}5 \cdot 1 + 12{,}6 \cdot 4 +$$

$$+12{,}7 \cdot 2 + 12{,}8 \cdot 1 + 12{,}9 \cdot 2 + 13 \cdot 1 + 13{,}1 \cdot 3 + 13{,}3 \cdot 4 + 13{,}5 \cdot 2 +$$

$$+13{,}6 \cdot 1 + 13{,}8 \cdot 2 + 14 \cdot 1 + 14{,}2 \cdot 2 + 14{,}4 \cdot 2 + 14{,}5 \cdot 1 + 14{,}8 \cdot 2 +$$

$$+15{,}1 \cdot 1 + 15{,}8 \cdot 2 + 16{,}2 \cdot 1 + 16{,}4 \cdot 1 + 17 \cdot 2 + 17{,}5 \cdot 1 + 18{,}7 \cdot 1 +$$

$$+19 \cdot 1 + 20 \cdot 1 + 21{,}8 \cdot 1) = 14{,}358;$$

$$S^2 = \frac{1}{50}((12 - 14{,}358)^2 \cdot 2 + (12{,}2 - 14{,}358)^2 \cdot 1 + (12{,}3 - 14{,}358)^2 \cdot 2 +$$

$$+(12{,}4 - 14{,}358)^2 \cdot 1 + (12{,}5 - 14{,}358)^2 \cdot 1 + (12{,}6 - 14{,}358)^2 \cdot 4 +$$

$$+(12{,}7-14{,}358)^2\cdot 2+(12{,}8-14{,}358)^2\cdot 1+(12{,}9-14{,}358)^2\cdot 2+$$

$$+(13-14{,}358)^2\cdot 1+(13{,}1-14{,}358)^2\cdot 3+(13{,}3-14{,}358)^2\cdot 4+$$

$$+(13{,}5-14{,}358)^2\cdot 2+(13{,}6-14{,}358)^2\cdot 1+(13{,}8-14{,}358)^2\cdot 2+$$

$$+(14-14{,}358)^2\cdot 1+(14{,}2-14{,}358)^2\cdot 2+(14{,}4-14{,}358)^2\cdot 2+$$

$$(14{,}5-14{,}358)^2\cdot 1+(14{,}8-14{,}358)^2\cdot 2+(15{,}1-14{,}358)^2\cdot 1+$$

$$+(15{,}8-14{,}358)^2\cdot 2+(16{,}2-14{,}358)^2\cdot 1+(16{,}4-14{,}358)^2\cdot 1+$$

$$+(17-14{,}358)^2\cdot 2+(17{,}5-14{,}358)^2\cdot 1+(18{,}7-14{,}358)^2\cdot 1+$$

$$+(19-14{,}358)^2\cdot 1+(20-14{,}358)^2\cdot 1+(21{,}8-14{,}358)^2\cdot 1)=5{,}24;$$

$$\sigma_{\text{В}}=\sqrt{5{,}24}=2{,}289.$$

Отже, для вибірки X отримано незміщену оцінку математичного сподівання $\bar{x}_{\text{В}}=14{,}358$, незміщену оцінку дисперсії $S^2=5{,}24$ і оцінку середнього квадратичного відхилення $\sigma_{\text{В}}=2{,}289$. ■

2) Знайдемо тепер довірчі інтервали за вибіркою X з надійністю $\gamma=0{,}9$.

а) Довірчий інтервал для математичного сподівання a знайдемо використовуючи нерівність (2.6) та наступну схему для знаходження точності δ та довірчих меж:

$$\left.\begin{matrix} n; s; \gamma \\ k=n-1 \end{matrix}\right\} \xrightarrow{\text{Додаток 4}} t_{\gamma,n} \to \delta=t_{\gamma,k}\frac{s}{\sqrt{n}} \to \bar{x}_{\text{В}}-\delta<a<\bar{x}_{\text{В}}+\delta.$$

За наведеною схемою та додатку 4 знаходимо:

$$\left.\begin{matrix} n=50; \\ s=2{,}289; \\ \gamma=0{,}9; \\ k=49 \end{matrix}\right\} \xrightarrow{\text{дод.4}} t_{\gamma,k}=t_{0,9;49}=1{,}303 \to \delta=\frac{1{,}303\cdot 2{,}289}{\sqrt{50}}=0{,}42\to$$

$$\to 14{,}358-0{,}42<a<14{,}358-0{,}42.$$

Звідси, $13{,}94<a<14{,}78$, тобто

$$P\{a\in(13{,}94;14{,}78)\}=0{,}9.$$

а) Довірчий інтервал для невідомої дисперсії σ, побудуємо використовуючи нерівність (2.8) та наступну схему для знаходження довірчих меж:

$$\left.\begin{matrix} n; s^2; \gamma \\ k = n-1 \end{matrix}\right\} \to \left.\begin{matrix} P\{\chi^2(k) < h'\} = \dfrac{1-\gamma}{2}; \\ P\{\chi^2(k) < h''\} = \dfrac{1+\gamma}{2}; \end{matrix}\right\} \xrightarrow{\text{дод.3}} h', h'' \to$$

$$\to \frac{(n-1)s^2}{h''} < \sigma^2 < \frac{(n-1)s^2}{h'}.$$

За наведеною схемою та додатком 3 отримаємо:

$$\left.\begin{matrix} n = 50; \\ s^2 = 5{,}24; \\ \gamma = 0{,}9; \\ k = 49 \end{matrix}\right\} \to \left.\begin{matrix} P\{\chi^2(49) < h'\} = 0{,}05; \\ P\{\chi^2(49) < h''\} = 0{,}95; \end{matrix}\right\} \xrightarrow{\text{дод.3}} h' = 34{,}8; h'' = 67{,}5 \to$$

$$\to \frac{49 \cdot 5{,}24}{67{,}5} < \sigma^2 < \frac{49 \cdot 5{,}24}{34{,}8}.$$

Звідси, $3{,}8 < \sigma^2 < 7{,}38$, тобто

$$P\{\sigma^2 \in (\,3{,}8; 7{,}38)\} = 0{,}9.$$

Добувши квадратний корінь із обох частин інтервальної оцінки для σ^2, одержимо, що при тій же надійності γ, довірчий інтервал для середнього квадратичного відхилення σ, тобто

$$1{,}95 < \sigma < 2{,}71.$$

Отже, справедлива рівність

$$P\{\sigma \in (\,1{,}95; 2{,}71)\} = 0{,}9. \blacksquare$$

Умова завдання

3) Знайти методом максимальної вірогідності за вибіркою $x_1, \dots, x_n$ оцінку параметра λ для показникового розподілу з щільністю $p(x, \lambda) = \lambda e^{-\lambda x}$ $(\lambda > 0, x > 0)$.

Хід розв'язання

Спочатку знаходимо вибіркову функцію вірогідності, визначену в означені 2.6, для невідомого параметра $\theta = \lambda$:

$$L(X,\lambda) = \lambda e^{-\lambda x_1} \cdots \lambda e^{-\lambda x_n} = \lambda^n e^{-\lambda(x_1+\cdots+x_n)}.$$

Знайдемо тепер логарифмічну функцію вірогідності

$$\ln L(X,\lambda) = \ln\left(\lambda^n e^{-\lambda(x_1+\cdots+x_n)}\right) = n\ln\lambda - \lambda\sum_{k=1}^{n} x_k.$$

Тоді рівняння максимальної вірогідності з теореми 2.4 набуває вигляду

$$U\left(X,\hat{\theta}\right) = \frac{d\ln L}{d\lambda} = \frac{n}{\lambda} - \sum_{k=1}^{n} x_k = 0.$$

Звідси знаходимо єдиний розв'язок

$$\hat{\lambda} = \frac{1}{\frac{1}{n}\sum_{k=1}^{n} x_k} = \frac{1}{\bar{x}_{\text{В}}}.$$

Знайдемо другу похідну по λ:

$$\frac{d^2\ln L}{d\lambda^2} = -\frac{n}{\lambda^2}.$$

Легко бачити, що при $\lambda = \frac{1}{\bar{x}_{\text{В}}}$ похідна $\frac{d^2\ln L}{d\lambda^2} \le 0$, тому ця точка і є точкою максимуму функції $\ln L$. Отже, оцінкою максимальної вірогідності для параметра λ є статистика $\hat{\lambda} = \frac{1}{\bar{x}_{\text{В}}}$. ■

РОЗДІЛ 3. ЗАДАЧІ ПЕРЕВІРКИ СТАТИСТИЧНИХ ГІПОТЕЗ

Література: [3, модуль 3, т. 2], [6, розділ 6, т. 14], [7, розділ 10], [14, розділ 15], [17, розділ 10].

Базові поняття. Статистичні гіпотези. Статистика критерію, критична область. Рівень та потужність критерію.

Основні задачі. Перевірка статистичних гіпотез про закон розподілу генеральної сукупності. Перевірка гіпотез про числові значення параметрів нормальних спостережень.

ЩО МАЄ ЗНАТИ ТА ВМІТИ СТУДЕНТ

1. Знання на рівні понять, означень, формулювань. Поняття статистичної гіпотези і статистичного критерію. Принципи перевірки гіпотез про ймовірності, про рівність середніх двох нормально розподілених випадкових величин, про рівність дисперсій двох нормально розподілених випадкових величин, про вигляд розподілу. Критерій Пірсона про нормальний, рівномірний, показниковий розподіли, про розподіли Бернуллі, Пуассона.

2. Уміння в розв'язуванні задач. Перевіряти статистичні гіпотези щодо властивостей вибірки. Перевіряти правильність статистичної гіпотези про припущення закону розподілу випадкової величини, про незалежність системи двох випадкових величин та про однорідність вибірок.

ОСНОВНІ ТЕОРЕТИЧНІ ВІДОМОСТІ

3.1. Основні принципи перевірки гіпотез

У багатьох випадках результати вибіркових спостережень або випробувань використовуються для перевірки припущень (гіпотез) щодо тих або інших властивостей розподілу генеральної сукупності. Зокрема, такого роду питання виникають у природознавсті, при розв'язанні технологічних задач, економіці та інших галузях науки і техніки.

Нехай X – спостережувана дискретна або неперервна випадкова величина.

Статистичною гіпотезою називається деяке припущення щодо вигляду невідомих розподілів випадкових величин або щодо параметрів відомих розподілів. В останньому випадку гіпотеза називається *параметричною*.

Наприклад, статистичними є гіпотези:

а) дисперсії двох нормальних сукупностей рівні між собою;

б) генеральна сукупність має нормальний закон розподілу.

Означення 3.1. Статистична гіпотеза називається *простою*, якщо вона однозначно визначає розподіл випадкової величини. У протилежному випадку гіпотеза називається *складною*.

Гіпотезу, яку перевіряють, називають *основною* або *нульовою* гіпотезою і позначають H_0. Наприклад, проста гіпотеза, яка стверджує рівність невідомого параметра заданій величині, записується так: $H_0: x = a$. У складних гіпотезах, наприклад, $H_0: x > a$ або $H_0: x < a$ зазначається деяка область можливих значень.

Поряд з нульовою гіпотезою H_0 розглядають *альтернативну* або *конкурентну* гіпотезу H_1. Зазвичай щодо нульової можна вказати нескінченну множину альтернативних гіпотез. Так, альтернативною до гіпотези $H_0: x = a$ може, наприклад, бути гіпотеза $H_1: x < a$, або гіпотеза $H_1: x \neq a$, або гіпотеза $H_1: x = b$ (b – число, яке не дорівнює a). Вибір альтернативної гіпотези визначається конкретним формулюванням задачі.

Статистичні гіпотези перевіряються на основі статистичних даних. Для перевірки статистичних гіпотез вибирається деяка випадкова величина, яка залежить від результатів спостережень, тобто є статистикою. Статистики, які вибирають для перевірки гіпотез, називають *статистиками критерію* або *критеріями* і позначають через K.

Зазвичай критерій перевірки гіпотези реалізують за допомогою деякої статистики. Далі обчислюють значення критерію за даними вибірок і називають його *спостережуваним значенням критерію* $K_{\text{сп}}$.

Множину значень критерію розбивають на дві області: *область прийняття гіпотези* – множина значень критерію, при яких гіпотеза приймається; *критичну область* S – множина значень критерію, при яких гіпотеза відкидається. Це рівносильно тому, що у вибірковому просторі R^n виділяється дві множини: множина тих точок із R^n, при яких критерій приймає значення із області прийняття гіпотези, і множина V тих точок із R^n, при яких критерій приймає значення із критичної області.

Ймовірність того, що критерій прийме значення із критичної області, називається *рівнем значущості* критерію. Частіше всього його задають близьким до 0 і покладають рівним, наприклад, 0,1; 0,05; 0,01; 0,005; 0,001.

При перевірці гіпотези вибіркові дані можуть суперечити гіпотезі H_0, тоді гіпотеза *відхиляється*. Якщо ж ці дані узгоджуються з нею, то вона *приймається*. Звідси випливає, що перевірка гіпотез, заснована на результатах вибірки, неминуче пов'язана з ризиком (імовірністю) прийняти помилкове рішення. При цьому можливі помилки (похибки) першого та другого роду.

Означення 3.2. *Помилкою (похибкою) першого роду* статистичного критерію називається відхилення нульової гіпотези H_0 за умови, що вона справджується. *Помилкою (похибкою) другого роду* називається прийняття нульової гіпотези H_0 за умови, коли справджується альтернатива H_1.

Вказані помилкові рішення пов'язані з відповідними випадковими подіями. Ймовірності цих подій називаються *ймовірностями помилок першого та другого роду.*

Ймовірність помилки першого роду – це *рівень значущості*. Позначимо її $\alpha = P_{H_0}(H_1)$, де $P_{H_0}(H_1)$ – ймовірність того, що буде прийнята гіпотеза H_1, тоді як насправді правильна гіпотеза H_0. Позначимо через β ймовірність помилки другого роду, тобто $\beta = P_{H_1}(H_0)$, де $P_{H_1}(H_0)$ – ймовірність того, що буде прийнята гіпотеза H_0, тоді як насправді правильною є гіпотеза H_1.

Оскільки виключити помилки першого роду й другого роду неможливо, то необхідно прагнути в кожній конкретній задачі мінімізувати втрати від цих помилок. Звичайно, бажано обидві помилки зменшити одночасно. Однак вони є конкуруючими, тобто зменшення ймовірності допустити одну з них спричиняє збільшення імовірності допустити другу. Тому в кожному конкретному випадку необхідно вибирати компромісний розв'язок. У більшості випадків єдиний шлях одночасного зменшення ризику помилок полягає в збільшенні об'єму вибірки.

Таким чином, сформулюємо *основний принцип перевірки статистичних гіпотез*: якщо спостережуване значення критерію $K_{\text{сп}}$ належить критичній області, то нульову гіпотезу H_0 відкидають, а якщо $K_{\text{сп}}$ належить області прийняття гіпотези H_0, то нульову гіпотезу H_0 відкидають.

У випадку одновимірності випадкової величини K, критична область, як правило, є множиною точок певних інтервалів на прямій, які відокремлені від області прийняття гіпотези так званими *критичними точками* $k_{\text{кр}}$.

Можливі три види розташування критичної області (залежно від вигляду нульової та альтернативної гіпотез, вигляду й розподілу критерію).

- *правобічна критична область* – це область, яка визначається нерівністю $K > k_{\text{кр}}$, де $k_{\text{кр}} > 0$ (рис. 3.1, а). Для її знаходження задають рівень значущості α, а потім шукають критичну точку $k_{\text{кр}}$, виходячи з вимоги, щоб за умови правильності нульової гіпотези ймовірність того, що $K > k_{\text{кр}}$, була рівна прийнятому рівню значущості:

$$P\{K > k_{\text{кр}}\} = \alpha.$$

У цьому випадку, відкидаючи правильну нульову гіпотезу, ми з ймовірністю α ризикуємо зробити помилку першого роду.

- *правобічна критична область* – це область, яка визначається нерівністю $K < -k_{\text{кр}}$, де $k_{\text{кр}} > 0$ (рис. 3.1, б). Задаючи значення рівня значущості α, $k_{\text{кр}}$ знаходять, виходячи з вимоги, щоб при правильності нульової гіпотези ймовірність того, що $K < -k_{\text{кр}}$, дорівнювала прийнятому рівню значущості:

$$P\{K < -k_{\text{кр}}\} = \alpha.$$

Правобічні й лівобічні критичні області називають *однобічними*.

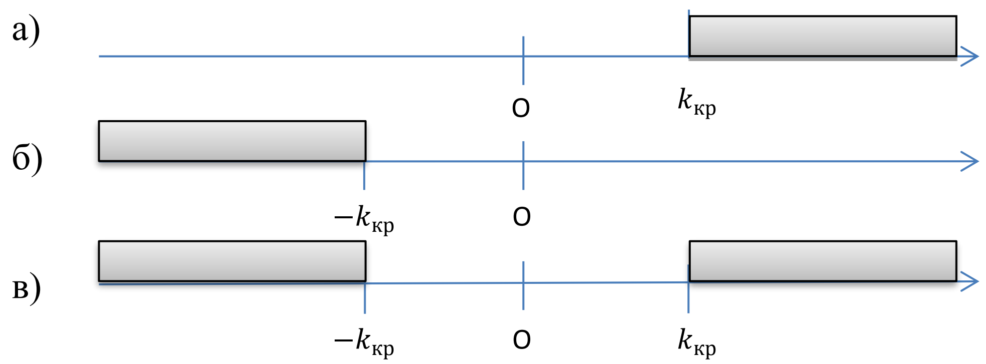


Рис. 3.1.

- *Двобічною* називають критичну область, що визначається нерівностями $K < k_1$, $K > k_2$, де $k_2 > k_1$.Зокрема, якщо критичні точки симетричні відносно нуля, то двобічна критична область визначається нерівностями $K < -k_{кр}$, $K > k_{кр}$ ($k_{кр} > 0$), або еквівалентною нерівністю $|K| > k_{кр}$ (рис. 3.1, в).

Означення 3.3. *Потужністю критерію* називають імовірність того, що нульова гіпотеза буде відкинута, якщо правильна альтернативна гіпотеза.

Звідси, якщо ймовірність помилки другого роду (приймати неправильну гіпотезу) дорівнює β, то потужність дорівнює $1-\beta$. Нехай потужність $1-\beta$ зростає; отже, зменшується ймовірність β зробити помилку другого роду.

Таким чином, якщо рівень значущості вже вибраний, то *критичну область потрібно будувати так, щоб потужність критерію була максимальною*. Виконання цієї вимоги повинне забезпечити мінімальну помилку другого роду.

Отже, перевірка статистичної гіпотези зводиться до таких етапів:

1) на основі вибіркових даних, формулюються основна H_0 та альтернативна H_1 гіпотези;

2) задається рівень значущості α;

3) вибирається статистика критерію для перевірки гіпотези H_0. Для її прийняття необхідно, щоб вона задовольняла наступні вимоги: а) повинен бути відомий закон розподілу статистики залежно від гіпотези; б) статистика повинна бути *незміщеною*, *обґрунтованою* й *ефективною*.

Умова незміщеності гарантує симетричність області прийняття гіпотези H_0; обґрунтованість забезпечує більш стислий розподіл статистики навколо центра, а отже, зменшення ймовірності помилки другого роду β при тому ж α; ефективність покращується за рахунок зменшення і α , і β за інших рівних умов.

4) визначається за таблицями, за рівнем значущості α і за альтернативною гіпотезою H_1 критична область;

5) отримується вибірка спостережень й обчислюється спостережуване значення критерію $K_{\text{сп}}$;

6) приймається рішення.

3.2. Перевірка гіпотез про параметри нормальних спостережень

Нехай $X = (x_1, \dots, x_n)$ – нормально розподілена вибірка спостережень $x_k \simeq N(a; \sigma^2), k = 1, \dots, n$ з параметрами $a = M(X)$ і $\sigma = \sqrt{D(X)}$, причому числові значення або одного, або обох параметрів невідомі.

За вибірковими незалежними спостереженнями $x_1, \dots, x_n$ обчислюють

$$\bar{x}_B = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ і } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_B)^2$$

які дають наближене уявлення відповідно про a і σ^2 і допомагають сформулювати гіпотези про те, які їхні числові значення.

1. Перевірка гіпотези про математичне сподівання при відомій дисперсії. Нехай дисперсія σ^2 – відома. Потрібно перевірити основну гіпотезу $H_0: a = a_0$. В якості альтернативи можна використати одну з наступних гіпотез $H_1: a > a_0$, $H_1: a < a_0$ або $H_1: a \neq a_0$.

Як критерій перевірки нульової гіпотези візьмемо статистику

$$Z = \frac{(\bar{x}_B - a_0)\sqrt{n}}{\sigma} \tag{3.1}$$

яка має стандартний нормальний розподіл – $Z \simeq N(0; 1)$.

Сформулюємо правила перевірки нульової гіпотези, позначивши значення критерію Z, обчислене за даними спостережень, через $Z_{\text{сп}}$.

Правило 1. Для того щоб при заданому рівні значущості α перевірити нульову гіпотезу $H_0: a = a_0$ при альтернативній гіпотезі $H_1: a > a_0$, необхідно обчислити спостережуване значення критерію $Z_{\text{сп}}$ за формулою (3.1). Потім з допомогою таблиці функції $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{1}{2}t^2}dt$ (див. додаток 2) знайти критичну точку правобічної критичної області за рівністю

$$\Phi\left(Z_{\text{кр}}\right) = (1 - 2\alpha)/2.$$

Якщо $Z_{\text{сп}} < Z_{\text{кр}}$ – немає підстав відкидати нульову гіпотезу. Якщо $Z_{\text{сп}} > Z_{\text{кр}}$ – нульову гіпотезу відкидають.

Правило 2. При альтернативній гіпотезі $H_1: a < a_0$ спочатку за формулою (3.1) необхідно обчислити спостережуване значення критерію $Z_{\text{сп}}$, а потім знайти критичну область $Z_{\text{кр}}$. Далі, якщо $Z_{\text{сп}} > -Z_{\text{кр}}$ – немає підстав відкидати нульову гіпотезу. Якщо $Z_{\text{сп}} < -Z_{\text{кр}}$ – нульову гіпотезу відкидають.

Правило 3. При альтернативній гіпотезі $H_1: a \neq a_0$ критичну точку двобічної критичної області знаходять за рівністю

$$\Phi\left(Z_{\text{кр}}\right) = (1 - \alpha)/2.$$

Якщо $|Z_{\text{сп}}| < Z_{\text{кр}}$ – немає підстав відкидати нульову гіпотезу. Якщо $|Z_{\text{сп}}| > Z_{\text{кр}}$ – нульову гіпотезу відкидають.

Розглянуті вище правила перевірки нульової гіпотези

$$H_0: a = a_0$$

дозволяють записати їх у вигляді наступної схеми:

$$\left.\begin{matrix} \bar{x}_B; n \\ a_0; \sigma \\ \alpha \end{matrix}\right\} \to H_0: a = a_0 \to Z_{\text{сп}} = \frac{(\bar{x}_B - a_0)\sqrt{n}}{\sigma} \to$$

$$\to \begin{cases} H_1: a > a_0 \to \Phi\left(Z_{\text{кр}}\right) = (1 - 2\alpha)/2 \xrightarrow{\text{дод.2}} Z_{\text{кр}} \to (1) \\ H_1: a < a_0 \to \Phi\left(Z_{\text{кр}}\right) = (1 - 2\alpha)/2 \xrightarrow{\text{дод.2}} Z_{\text{кр}} \to (2) \\ H_1: a \neq a_0 \to \Phi\left(Z_{\text{кр}}\right) = (1 - \alpha)/2 \xrightarrow{\text{дод.2}} Z_{\text{кр}} \to (3) \end{cases} \qquad (3.2)$$

$$(1) \to \begin{cases} Z_{\text{сп}} < Z_{\text{кр}} \to \boldsymbol{P}, \\ Z_{\text{сп}} > Z_{\text{кр}} \to \boldsymbol{V}, \end{cases}$$

$$(2) \to \begin{cases} Z_{\text{сп}} > -Z_{\text{кр}} \to \boldsymbol{P}, \\ Z_{\text{сп}} < -Z_{\text{кр}} \to \boldsymbol{V}, \end{cases} \quad (3) \to \begin{cases} |Z_{\text{сп}}| < Z_{\text{кр}} \to \boldsymbol{P}, \\ |Z_{\text{сп}}| > Z_{\text{кр}} \to \boldsymbol{V}. \end{cases}$$

Тут і далі $\boldsymbol{P}$ означає, що немає підстав відкидати нульову гіпотезу, а $\boldsymbol{V}$ – нульову гіпотезу відкидають.

2. Перевірка гіпотези про математичне сподівання при невідомій дисперсії. Якщо дисперсія генеральної сукупності невідома, то критерієм перевірки нульової гіпотези буде статистика

$$T = \frac{(\bar{x}_B - a_0)\sqrt{n}}{s}.$$

Величина T має розподіл Стьюдента з $k = n - 1$ ступенями вільності (додаток 4). Задаючи рівень значущості α, знайдемо область і перевіримо нульову гіпотезу $H_0: a = a_0$ для трьох, наведених вище, альтернативних гіпотез, аналогічно до схеми (3.2).

$$\left.\begin{matrix} \bar{x}_B; n \\ a_0; s \\ \alpha; k \end{matrix}\right\} \to H_0: a = a_0 \to T_{\text{сп}} = \frac{(\bar{x}_B - a_0)\sqrt{n}}{s} \to$$

$$\to \begin{cases} H_1: a > a_0 \to t_{\text{кр}}(\alpha; k) \xrightarrow{\text{дод.4}} t_{\text{кр}} \to (1) \\ H_1: a < a_0 \to t_{\text{кр}}(\alpha; k) \xrightarrow{\text{дод.4}} t_{\text{кр}} \to (2) \\ H_1: a \neq a_0 \to t_{\text{кр}}(\alpha; k) \xrightarrow{\text{дод.4}} t_{\text{кр}} \to (3) \end{cases}$$

$$\to (1) \to \begin{cases} T_{\text{сп}} < t_{\text{кр}} \to \boldsymbol{P}, \\ T_{\text{сп}} > t_{\text{кр}} \to \boldsymbol{V}, \end{cases}$$

$$(2) \to \begin{cases} T_{\text{сп}} > -t_{\text{кр}} \to \boldsymbol{P}, \\ T_{\text{сп}} < -t_{\text{кр}} \to \boldsymbol{V}, \end{cases}$$

$$(3) \to \begin{cases} |T_{\text{сп}}| < t_{\text{кр}} \to \boldsymbol{P}, \\ |T_{\text{сп}}| > t_{\text{кр}} \to \boldsymbol{V}. \end{cases}$$

3. Перевірка гіпотези про дисперсію при невідомому математичному сподіванні. Нехай випадкова величина X має нормальний закон розподілу, тобто $X \simeq N(a;\sigma^2)$, де параметр σ невідомий. Потрібно перевірити нульову гіпотезу: $H_0\colon \sigma^2 = \sigma_0^2$, де σ_0^2 – заздалегідь задане число.

Як критерій перевірки нульової гіпотези вибирають статистику

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2},$$

яка має χ^2–розподіл Пірсона із числом ступенів вільності $k = n-1$ (додаток 3).

Задаючи рівень значущості α, перевіримо нульову гіпотезу $H_0\colon \sigma^2 = \sigma_0^2$ для трьох видів альтернативної гіпотези H_1, відповідно до наступної схеми:

$$\left.\begin{matrix} s^2; n \\ \sigma_0^2; k \\ \alpha \end{matrix}\right\} \to H_0\colon \sigma^2 = \sigma_0^2 \to \chi^2_{\text{сп}} = \frac{(n-1)s^2}{\sigma_0^2} \to$$

$$\to \begin{cases} H_1\colon \sigma^2 > \sigma_0^2 \to \chi^2_{\text{кр}}(\alpha;k) \xrightarrow{\text{дод.3}} x_{\text{кр}}^{\text{пр}} \to (1) \\ H_1\colon \sigma^2 < \sigma_0^2 \to \chi^2_{\text{кр}}(1-\alpha;k) \xrightarrow{\text{дод.3}} x_{\text{кр}}^{\text{лів}} \to (2) \\ H_1\colon \sigma^2 \neq \sigma_0^2 \to \left[\begin{matrix} \chi^2_{\text{пр.кр}}(\alpha/2;k) \xrightarrow{\text{дод.3}} x_{\text{кр}}^{\text{пр}} \\ \chi^2_{\text{лів.кр}}(1-\alpha/2;k) \xrightarrow{\text{дод.3}} x_{\text{кр}}^{\text{лів}} \end{matrix}\right. \to (3) \end{cases}$$

$$\to (1) \to \begin{cases} \chi^2_{\text{сп}} < x_{\text{кр}}^{\text{пр}} \to \boldsymbol{P}, \\ \chi^2_{\text{сп}} > x_{\text{кр}}^{\text{пр}} \to \boldsymbol{V}, \end{cases}$$

$$(2) \to \begin{cases} \chi^2_{\text{сп}} > x_{\text{кр}}^{\text{лів}} \to \boldsymbol{P}, \\ \chi^2_{\text{сп}} < x_{\text{кр}}^{\text{лів}} \to \boldsymbol{V}, \end{cases} (3) \to \begin{cases} x_{\text{кр}}^{\text{лів}} < \chi^2_{\text{сп}} < x_{\text{кр}}^{\text{пр}} \to \boldsymbol{P}, \\ \chi^2_{\text{сп}} < x_{\text{кр}}^{\text{лів}} \cup \chi^2_{\text{сп}} > x_{\text{кр}}^{\text{пр}} \to \boldsymbol{V}. \end{cases}$$

Зауваження 3.1. Якщо число ступенів вільності $k > 30$, то критичну точку $\chi^2_{\text{кр}}(\alpha;k)$ можна знайти з рівності Уілсона-Гільферті

$$\chi^2_{\text{кр}}(\alpha;k) = k\left(1 - 2/(9k) + u_\alpha\sqrt{2(9k)}\right)^3,$$

де u_α знаходять, використовуючи функцію Лапласа (див. додаток 2) з рівності $\Phi(u_\alpha) = (1-2\alpha)/2$.

3.3. Перевірка гіпотез про рівність дисперсій і математичних сподівань двох нормальних розподілів.

Задачі перевірки гіпотез про рівність двох дисперсій виникають, наприклад, при аналізі стабільності виробничого процесу, при вивченні якості вимірювальних приладів, при вивченні ступеня однорідності двох сукупностей щодо будь-якої ознаки (кваліфікації робітників, стажу персоналу) й т.д. Перевірка гіпотез про рівність середніх значень двох вибірок виникає, наприклад, при вибірковому контролі якості виробів, виготовлених на різних установках, у фінансовому аналізі – при зіставленні рівня прибутковості різних активів і т.д.

1. **Перевірка гіпотези про рівність математичних сподівань двох нормально розподілених випадкових величин.** Нехай одночасно спостерігаються дві незалежні нормально розподілені вибірки $X = (\xi_1, \dots, \xi_n)$ та $Y = (\eta_1, \dots, \eta_m)$, що мають розподіли $\xi_1 \simeq N(a_X, \sigma_X^2), \eta_1 \simeq N(a_Y, \sigma_Y^2)$ з невідомими середніми $MX = a_X, MY = a_Y$ та відомими дисперсіями $\sigma_X^2, , \sigma_Y^2$. За незалежними вибірками з об’ємами n і m, отриманими відповідно з генеральних сукупностей X і Y, потрібно при заданому рівні значущості α перевірити нульову гіпотезу про рівність середніх $H_0: a_x = a_y$. Скористаємося тим, що вибіркові середні $\bar{x}_B = \frac{1}{n}\sum_{i=1}^{n} X_i$ і $\bar{y}_B = \frac{1}{n}\sum_{i=1}^{n} Y_i$ мають розподіли відповідно $N\left(a_X, \frac{\sigma_X^2}{n}\right)$ та $N\left(a_Y, \frac{\sigma_Y^2}{m}\right)$. Тоді з незалежності вибірок, отримаємо, що статистика

$$Z = \frac{\bar{x}_B - \bar{y}_B}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \simeq N(0, 1),$$

використовується як статистика критерію перевірки нульової гіпотези. Для заданого рівня значущості α перевіримо нульову гіпотезу $H_0: a_x = a_y$, залежно від виду альтернативної гіпотези H_1 відповідно до наступної схеми:

$$\left.\begin{matrix} \bar{x}_B;n;\sigma_X^2 \\ \bar{y}_B;m;\sigma_Y^2 \\ \alpha \end{matrix}\right\} \rightarrow H_0: a_x = a_y \rightarrow Z_{\text{сп}} = \frac{\bar{x}_B - \bar{y}_B}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \rightarrow$$

$$\rightarrow \begin{cases} H_1: a_x > a_y \rightarrow \Phi(Z_{\text{кр}}) = (1-2\alpha)/2 \xrightarrow{\text{дод.2}} Z_{\text{кр}} \rightarrow (1) \\ H_1: a_x < a_y \rightarrow \Phi(Z_{\text{кр}}) = (1-2\alpha)/2 \xrightarrow{\text{дод.2}} Z_{\text{кр}} \rightarrow (2) \\ H_1: a_x \neq a_y \rightarrow \Phi(Z_{\text{кр}}) = (1-\alpha)/2 \xrightarrow{\text{дод.2}} Z_{\text{кр}} \rightarrow (3) \end{cases} \quad (3.3)$$

$$(1) \rightarrow \begin{cases} Z_{\text{сп}} < Z_{\text{кр}} \rightarrow \boldsymbol{P}, \\ Z_{\text{сп}} > Z_{\text{кр}} \rightarrow \boldsymbol{V}, \end{cases}$$

$$(2) \rightarrow \begin{cases} Z_{\text{сп}} > -Z_{\text{кр}} \rightarrow \boldsymbol{P}, \\ Z_{\text{сп}} < -Z_{\text{кр}} \rightarrow \boldsymbol{V}, \end{cases} (3) \rightarrow \begin{cases} |Z_{\text{сп}}| < Z_{\text{кр}} \rightarrow \boldsymbol{P}, \\ |Z_{\text{сп}}| > Z_{\text{кр}} \rightarrow \boldsymbol{V}. \end{cases}$$

Нехай тепер дисперсії σ_X^2, ,σ_Y^2 невідомі, але рівні між собою, тобто $\sigma_X^2 = \sigma_Y^2 = \sigma^2$. За наближене значення дисперсії σ^2 величини X і Y беремо виправлені дисперсії $s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{x}_B)^2$ і $s_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{y}_B)^2$ відповідно. Тоді для перевірки гіпотези $H_0: a_x = a_y$ як критерій перевірки приймемо випадкову величину

$$T = \frac{\bar{x}_B - \bar{y}_B}{\sqrt{(n-1)s_X^2 + (m-1)s_Y^2}}\sqrt{\frac{nm(n+m-2)}{n+m}},$$

яка має розподіл Стьюдента з $k = n + m - 2$ ступенями вільності.

Для заданого рівня значущості α перевірку нульової гіпотези здійснюємо за наступною схемою:

$$\left.\begin{matrix} \bar{x}_B;n;s_X^2 \\ \bar{y}_B;m;s_Y^2 \\ D(X)=D(Y) \\ k=n+m-2;\alpha \end{matrix}\right\} \rightarrow H_0: a_x = a_y \rightarrow T_{\text{сп}} = \frac{\bar{x}_B - \bar{y}_B}{\sqrt{(n-1)s_X^2 + (m-1)s_Y^2}}\sqrt{\frac{nm(n+m-2)}{n+m}} \rightarrow$$

$$\rightarrow \begin{cases} H_1: a_x > a_y \rightarrow t_{\text{одн.кр}}(\alpha;k) \xrightarrow{\text{дод.4}} t_{\text{кр}} \rightarrow (1) \\ H_1: a_x < a_y \rightarrow t_{\text{одн.кр}}(\alpha;k) \xrightarrow{\text{дод.4}} t_{\text{кр}} \rightarrow (2) \\ H_1: a_x \neq a_y \rightarrow t_{\text{одн.кр}}(\alpha;k) \xrightarrow{\text{дод.4}} t_{\text{кр}} \rightarrow (3) \end{cases} \quad (3.4)$$

$$(1) \rightarrow \begin{cases} T_{\text{сп}} < t_{\text{кр}} \rightarrow \boldsymbol{P}, \\ T_{\text{сп}} > t_{\text{кр}} \rightarrow \boldsymbol{V}, \end{cases}$$

$$(2) \rightarrow \begin{cases} T_{\text{сп}} > -t_{\text{кр}} \rightarrow \boldsymbol{P}, \\ T_{\text{сп}} < -t_{\text{кр}} \rightarrow \boldsymbol{V}, \end{cases} (3) \rightarrow \begin{cases} |T_{\text{сп}}| < t_{\text{кр}} \rightarrow \boldsymbol{P}, \\ |T_{\text{сп}}| > t_{\text{кр}} \rightarrow \boldsymbol{V}. \end{cases}$$

2. **Перевірка гіпотези про рівність дисперсій двох нормально розподілених випадкових величин.** Нехай величини X та Y розподілені за законами відповідно $N(a_X, \sigma_X^2)$ та $N(a_Y, \sigma_Y^2)$. З генеральної сукупності X отримується вибірка обсягу n, а з генеральної сукупності Y – вибірка обсягу m. Вибірки незалежні. За цими вибірками необхідно перевірити гіпотезу про рівність дисперсій генеральних сукупностей X та Y, тобто гіпотезу H_0: $\sigma_X^2 = \sigma_Y^2 = \sigma^2$.

В основі перевірки цієї гіпотези лежить відношення s_X^2 / s_Y^2 виправлених вибіркових дисперсій. Якщо це відношення "значно" відрізняється від 1, то нульову гіпотезу природно відхилити. У протилежному разі гіпотезу H_0 природно не відхиляти. Отже, як критерій перевірки H_0: $D(X) = D(Y)$ приймемо відношення більшої виправленої дисперсії до меншої, тобто статистику

$$F = \frac{s_{\text{Б}}^2}{s_{\text{М}}^2}.$$

Величина F за умови правильності нульової гіпотези має розподіл Фішера-Снедекора зі ступенями вільності $k_1 = n - 1$ і $k_2 = m - 1$, де n – об'єм вибірки, при якій обчислена більша виправлена дисперсія, m – об'єм вибірки, при якій знайдено меншу виправлену дисперсію. Задаючи рівень значущості α, побудуємо критичну область і перевіримо нульову гіпотезу залежно від виду альтернативної гіпотези H_1, відповідно до наступної схеми:

$$\left. \begin{matrix} s_{\text{Б}}^2; n; k_1 \\ s_{\text{М}}^2; m; k_2 \\ \alpha \end{matrix} \right\} \rightarrow H_0: D(X) = D(Y) \rightarrow F_{\text{сп}} = \frac{s_{\text{Б}}^2}{s_{\text{М}}^2} \rightarrow \qquad (3.5)$$

$$\begin{cases} H_1: D(X) > D(Y) \to F_{\text{кр}}(\alpha; k_1; k_2) \xrightarrow{\text{дод.5}} \left[\begin{array}{l} F_{\text{сп}} < F_{\text{кр}} \to \boldsymbol{P}, \\ F_{\text{сп}} > F_{\text{кр}} \to \boldsymbol{V}. \end{array}\right. \\ H_1: D(X) \neq D(Y) \to F_{\text{кр}}(\alpha/2\,; k_1; k_2) \xrightarrow{\text{дод.5}} \left[\begin{array}{l} F_{\text{сп}} < F_{\text{кр}} \to \boldsymbol{P}, \\ F_{\text{сп}} > F_{\text{кр}} \to \boldsymbol{V}. \end{array}\right. \end{cases}$$

3.4. Перевірка гіпотези про рівність ймовірностей

Нехай A – випадкова подія, ймовірність p появи якої в кожному випробуванні невідома і нехай є підстави вважати, що p дорівнює заданому значенню p_0. Потрібно при рівні значущості α перевірити нульову гіпотезу $H_0: p = p_0$. В основі перевірки цієї гіпотези лежить порівняння числа p_0 з наближеним значенням імовірності p, наближенням якого є відносна частота $\mu_n = m/n$, де n – число випробувань у схемі Бернуллі, а m – число випробувань серед n випробувань, у яких відбулася подія A. Як критерій перевірки нульової гіпотези приймемо випадкову величину

$$Z = \frac{\mu_n - p_0}{\sqrt{np_0(1-p_0)}} \simeq N(0\,, 1).$$

Задаючи рівень значущості α, побудуємо критичну область і перевіримо нульову гіпотезу залежно від виду альтернативної гіпотези H_1, відповідно до наступної схеми:

$$\left.\begin{array}{r} n; \mu_n \\ p_0 \alpha \end{array}\right\} \to H_0: p = p_0 \to Z_{\text{сп}} = \frac{\mu_n - p_0}{\sqrt{np_0(1-p_0)}} \to$$

$$\begin{cases} H_1: p > p_0 \to \Phi(Z_{\text{кр}}) = \frac{1-2\alpha}{2} \xrightarrow{\text{дод.2}} Z_{\text{кр}} \to (1) \\ H_1: p < p_0 \to \Phi(Z_{\text{кр}}) = \frac{1-2\alpha}{2} \xrightarrow{\text{дод.2}} Z_{\text{кр}} \to (2) \\ H_1: p \neq p_0 \to \Phi(Z_{\text{кр}}) = \frac{1-\alpha}{2} \xrightarrow{\text{дод.2}} Z_{\text{кр}} \to (3) \end{cases} \tag{3.6}$$

$$(1) \to \begin{cases} Z_{\text{сп}} < Z_{\text{кр}} \to \boldsymbol{P}, \\ Z_{\text{сп}} > Z_{\text{кр}} \to \boldsymbol{V}, \end{cases}$$

$$(2) \to \begin{cases} Z_{\text{сп}} > -Z_{\text{кр}} \to \boldsymbol{P}, \\ Z_{\text{сп}} < -Z_{\text{кр}} \to \boldsymbol{V}, \end{cases} (3) \to \begin{cases} |Z_{\text{сп}}| < Z_{\text{кр}} \to \boldsymbol{P}, \\ |Z_{\text{сп}}| > Z_{\text{кр}} \to \boldsymbol{V}. \end{cases}$$

Нехай у двох генеральних сукупностях проводяться незалежні випробування Бернуллі великого об'єму n_1 і n_2, у кожному з яких подія A може з'явитися, або не з'явитися. Позначимо невідомі ймовірності появи події A у першій і у другій сукупностях відповідно p_1 і p_2. Припустимо, що в першій сукупності подія A спостерігалася m_1 разів, а в другій – m_2 разів.

При заданому рівні значущості α потрібно перевірити нульову гіпотезу про рівність ймовірностей p_1 і p_2, тобто $H_0: p_1 = p_2$. Як критерій перевірки нульової гіпотези приймемо статистику

$$Z = \frac{\frac{m_1}{n_1} - \frac{m_2}{n_{2}}_2}{\sqrt{\frac{m_1 + m_2}{n_1 + n_2}\left(1 - \frac{m_1 + m_2}{n_1 + n_2}\right)\left(\frac{1}{n_1} + \frac{1}{n_{2_2}}\right)}} \simeq N(0\,,1).$$

Критична область будується залежно від виду альтернативної гіпотези, тому наведемо лише схему перевірки нульової гіпотези:

$$\left.\begin{matrix} n_1; m_1 \\ n_2; m_2 \\ \alpha \end{matrix}\right\} \to H_0: p_1 = p_2 \to \hat{p} = \frac{m_1 + m_2}{n_1 + n_2} \to Z_{\text{сп}} \to$$

$$\to \begin{cases} H_1: p > p_0 \to \Phi\left(Z_{\text{кр}}\right) = \frac{1-2\alpha}{2} \xrightarrow{\text{дод.2}} Z_{\text{кр}} \to (1) \\ H_1: p < p_0 \to \Phi\left(Z_{\text{кр}}\right) = \frac{1-2\alpha}{2} \xrightarrow{\text{дод.2}} Z_{\text{кр}} \to (2) \\ H_1: p \neq p_0 \to \Phi\left(Z_{\text{кр}}\right) = \frac{1-\alpha}{2} \xrightarrow{\text{дод.2}} Z_{\text{кр}} \to (3) \end{cases} \qquad (3.7)$$

$$(1) \to \begin{cases} Z_{\text{сп}} < Z_{\text{кр}} \to \boldsymbol{P}, \\ Z_{\text{сп}} > Z_{\text{кр}} \to \boldsymbol{V}, \end{cases}$$

$$(2) \to \begin{cases} Z_{\text{сп}} > -Z_{\text{кр}} \to \boldsymbol{P}, \\ Z_{\text{сп}} < -Z_{\text{кр}} \to \boldsymbol{V}, \end{cases} (3) \to \begin{cases} |Z_{\text{сп}}| < Z_{\text{кр}} \to \boldsymbol{P}, \\ |Z_{\text{сп}}| > Z_{\text{кр}} \to \boldsymbol{V}. \end{cases}$$

3.5. Перевірка гіпотези про закон розподілу. Критерій Пірсона

Нехай X – випадкова величина, розподіл якої невідомий. Досить часто на основі вибірки і деяких додаткових міркувань можна висунути певну гіпотезу про вигляд цього розподілу. Найбільш простою і в той же час найбільш сильною гіпотезою є гіпотеза, що функція розподілу випадкової величини має повністю визначений вигляд $F(x)$.

Отже, якщо розподіл випадкової величини X невідомий, але є підстави припустити, що він має певний вигляд $F(x)$, то перевіряється нульова гіпотеза H_0: генеральна сукупність X має розподіл $F(x)$. Будемо писати $H_0: F_x(x) = F(x)$. Незалежно від того, справедлива гіпотеза H_0 чи ні, за вибіркою завжди можна знайти емпіричну функцію розподілу $F^*(x) = P^*\{X < x\}$, де $P^*\{X < x\}$ – відносна частота події $\{X < x\}$. Основну гіпотезу приймають, якщо "відхилення" емпіричного розподілу $F^*(x)$ від гіпотетичного $F(x)$ мале. У протилежному випадку гіпотезу H_0 відхиляють. Міру такого "відхилення" можна визначити багатьма способами, у відповідності до яких отримуються різні критерії для перевірки нульової гіпотези. Ці критерії носять назву *критеріїв згоди* або *узгодження*.

Підставою для висування гіпотези про закон розподілу генеральної сукупності можуть бути теоретичні передумови (наприклад, виконання умов центральної граничної теореми, яка обґрунтовує асимптотично нормальний розподіл), графічний аналіз розподілу (так, за формою полігона або гістограми можна стверджувати про закон розподілу). Одним з найбільш широко вживаних таких критеріїв є *критерій* χ^2 (*критерій Пірсона*).

Нехай при рівні значущості α необхідно перевірити нульову гіпотезу H_0 про те, що досліджувана генеральна сукупність підпорядковується деякому закону розподілу. Із цією метою будемо порівнювати емпіричні (спостережувані) та теоретичні (обчислені в припущенні нормального розподілу) частоти.

Нехай за вибіркою n об'єму отриманий емпіричний розподіл з рівновіддалених варіант і відповідних їм частот:

Таблиця 3.1

X	x_1	x_2	…	x_s
n	n_1	n_2	…	n_s

Загальна схема застосування критерію χ^2 (*критерій Пірсона*):

1. Усю множину значень X розбивають на r інтервалів, що не перетинаються: $\Delta_1, \Delta_2, \dots, \Delta_r$.

2. Обчислюють $n_1, n_2, \dots, n_r$ – кількість вибіркових значень, що потрапили $\Delta_1, \Delta_2, \dots, \Delta_r$ відповідно (тобто частоти варіант у частинних інтервалах $\sum_{i=1}^{r} n_i = n$).

3. Знаходять p_i – імовірності потрапляння ознаки X в i-й інтервал Δ_i ($\sum_{i=1}^{r} p_i = 1$) за теоретичним законом розподілу $F(x)$.

4. Обчислюють значення критерію R^* за формулою

$$R^* = \chi^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^{r} \frac{n_i^2}{np_i} - n. \qquad (3.8)$$

5. Використовуючи таблицю χ^2-розподілу (дод. 3) і заданий за умовою рівень значущості α знаходять $\chi^2_{\text{кр}}$ з $k = r - l - 1$ ступенями вільності (l — кількість параметрів розподілу $F(x)$, які оцінюються за вибіркою). Гіпотеза приймається на заданому рівні значущості, якщо $\chi^2_{\text{сп}} < \chi^2_{\text{кр}}$. У протилежному випадку гіпотеза відхиляється.

Зауваження 3.2. Застосування критерію χ^2 вимагає дотримання наступних умов: 1) експериментальні дані повинні бути незалежними; 2) об’єм вибірки повинен бути досить великим (не менше 50 одиниць), при цьому чисельність (частота) кожної групи повинна бути не менша 5. Якщо остання умова не виконується, то здійснюється попереднє об’єднання інтервалів із сусідніми.

Алгоритм перевірки гіпотези про біноміальний закон розподілу

Нехай проведено n дослідів. Кожен дослід складається з N незалежних випробувань, у кожному з яких ймовірність появи події A одна і та сама.

Фіксується число появ події A в кожному експерименті. В результаті отримано наступний розподіл дискретної випадкової величини X – числа появи події A (у першому рядку вказано число x_i появ події A в одному досліді, а в другому – частота n_i, тобто кількість дослідів, у яких зареєстровано x_i появи даної події):

x_i	0	1	2	...	N
n_i	n_0	n_1	n_2	...	n_N

Для того, щоб при рівні значущості α перевірити гіпотезу про розподіл випадкової величини X за біноміальним законом потрібно:

1. Обчислити за формулою Бернуллі $P_N(i) = C_N^i p^i (1-p)^{N-i}$ ймовірності $P_N(i)$ появи події A рівно i раз в N випробуваннях ($i = 0,1,2,\ldots,r$, де r – максимальна кількість появи події A в одному досліді, тобто $r \leq N$).
2. Знайти теоретичні частоти $n_i' = n \cdot P_N(i)$, де n – кількість дослідів.
3. Порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона, прийнявши число ступенів вільності $k = r - 1$ (при цьому припускається, що ймовірність p появи даної події задана, тобто не оцінювалась за вибіркою і не проводились об'єднання малочисельних частот). Якщо ж ймовірність p була оцінена за вибіркою, то $k = r - 2$. Якщо, крім того, було проведено об'єднання малочисельних частот, то r – кількість інтервалів, що залишилися після об'єднання.

Алгоритм перевірки гіпотези про розподіл Пуассона

Нехай задано точковий статистичний розподіл вибірки. Тоді перевірка при рівні значущості α гіпотези про розподіл генеральної сукупності за законом Пуассона проводиться за наступною схемою:

1. Знайти за заданим статистичнм розподілом вибіркове середнє $\bar{x}_в$.
2. Взяти в якості оцінки параметра λ розподілу Пуассона вибіркове середнє: $\hat{\lambda} = \bar{x}_в$.

3. Обчислити за формулою Пуассона $P_n(i) = e^{-\lambda}\frac{\lambda^i}{i!}$ ймовірності $P_n(i)$ появи події A рівно i раз в n випробуваннях ($i = 1,2,\dots,r$, де r – максимальна кількість подій, n – обсяг вибірки).

4. Визначити теоретичні частоти $n_i' = n \cdot P_n(i)$.

5. Порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона, прийнявши кількість ступенів вільності $k = r - 2$, де r – кількість варіант вибірки (якщо було проведено об'єднання малочисельних частот в одну групу, то r – кількість варіант, що залишилися після об'єднання частот).

Алгоритм перевірки гіпотези про рівномірний закон розподілу

Задано емпіричний розподіл неперервної випадкової величини X у вигляді послідовності r інтервалів $(x_i; x_{i+1})$ та відповідних їм частот n_i, причому $\sum_{i=1}^{n} n_i = n$. Тоді перевірка за допомогою критерію Пірсона при рівні значущості α гіпотези про те, що випадкова величина X має рівномірний розподіл зводиться до наступної схеми:

1. Оцінити параметри a і b – кінці інтервалу, у якому спостерігалися можливі значення X за формулами ($\hat{a}$ і $\hat{b}$ – оцінки параметрів):

$$\bar{x}_{\text{в}} = \frac{a+b}{2}, D_B = \frac{(b-a)^2}{12}.$$

2. Знайти щільність імовірності передбачуваного розподілу

$$f(x) = \begin{cases} \frac{1}{\hat{b} - \hat{a}}, x \in [\hat{a}, \hat{b}], \\ 0, x \notin [\hat{a}, \hat{b}]. \end{cases}$$

3. Визначити теоретичні частоти:

$$n_1' = np_1 = n\frac{1}{\hat{b}-\hat{a}}(x_1 - \hat{a});$$

$$n_i' = n\frac{1}{\hat{b}-\hat{a}}(x_i - x_{i-1}), i - 2,3,\dots,N-1;$$

$$n_N' = n\frac{1}{\hat{b}-\hat{a}}(\hat{b} - x_{N-1}).$$

4. Порівняти емпіричні те теоретичні частоти за допомогою критерію Пірсона, узявши кількість ступенів вільності $k = r - 3$, де r – кількість інтервалів у вибірці.

Алгоритм перевірки гіпотези про показниковий закон розподілу

Нехай задано емпіричний розподіл неперервної випадкової величини X у вигляді послідовності інтервалів $(x_i; x_{i+1})$ та відповідних їм частот n_i , причому $\sum_{i=1}^{n} n_i = n$. Тоді перевірка при рівні значущості α гіпотези про те, що випадкова величина X має показниковий розподіл зводиться до наступної схеми:

1. За даними емпіричного розподілу знаходять вибіркове середнє $\bar{x}_{\text{в}}$. Для цього, прийнявши в якості "представника" i-го інтервалу його середину $x_i^* = (x_i + x_{i+1})/2$, утворюють послідовність рівнорозподілених варіант і відповідних їм частот.

2. Приймають в якості оцінки параметра λ показникового розподілу величину, обернену до вибіркового середнього: $\hat{\lambda} = \frac{1}{\bar{x}_{\text{в}}}$.

3. Обчислюють ймовірності p_i потрапляння X до інтервалу $(x_i; x_{i+1})$ за формулою

$$p_i = P\{x_i < X < x_{i+1}\} = e^{-\lambda x_i} - e^{-\lambda x_{i+1}}$$

і знаходять шукані теоретичні частоти $n_i' = np_i$.

4. Порівнюють емпіричні та теоретичні частоти за допомогою критерію Пірсона, прийнявши число ступенів вільності $k = r - 2$, оскільки оцінюють один параметр λ, тому $s = 1$, для r – початкової кількості інтервалів вибірки; якщо ж було проведено об'єднання малочисельних частот та і самих інтервалів, то r – кількість інтервалів, що залишилися після об'єднання.

Алгоритм перевірки гіпотези про нормальний розподіл

Для того, щоб при заданому рівні значущості α перевірити гіпотезу про нормальний розподіл у випадку, коли емпіричний розподіл задано у вигляді послідовнності рівновіддалених варіант та відповідних їм частот, потрібно:

1. Обчислити вибіркове середнє $\bar{x}_в$ та вибіркове середнє квадратичне відхилення σ_B.

2. Знайти теоретичні частоти n_i', у припущенні нормального розподілу генеральної сукупності, за формулою

$$n_i' \coloneqq np_i = \frac{nh}{\sigma_B}\varphi(u_i), \sigma_B = \sqrt{D_B},$$

де n – об'єм вибірки, h – крок, рівний різниці між двома сусідніми варіантами, $u_i = (x_i - \bar{x}_в)/\sigma_B$, де $\varphi(u) = \frac{1}{\sqrt{2\pi}}e^{-u^2/2}$.

3. Порівняти емпіричні та теоретичні частоти за допомогою критерію Пірсона. Для цього, використовуючи таблицю χ^2-розподілу (дод. 3) і заданий за умовою рівень значущості α знаходять χ_α^2 з $k = r - 3$ ступенями вільності (оцінюють два параметри a і σ, тому $l = 2$). Якщо $\chi_{сп}^2 < \chi_{кр}^2$ – немає підстав відкидати нульову гіпотезу. Іншими словами, емпіричні і теоретичні частоти відрізняються несуттєво. Якщо $\chi_{сп}^2 > \chi_{кр}^2$ – нульову гіпотезу відкидають.

Якщо ж варіаційний ряд неперервний, то перевірка гіпотези про нормальний розподіл зводиться до такого алгоритму:

1. Весь інтервал спостережуваних значень X потрібно поділити на r часткових інтервалів $(x_i; x_{i+1})$ однакової довжини h. Знаходять $x_i^* = (x_i + x_{i+1})/2$ і як частоту n_i варіанти x_i^* приймають число варіант, які потрапили до i-го інтервалу. У кінцевому підсумку одержують послідовність рівнорозподілених варіант і відповідних їм частот:

$$\begin{matrix} x_1^* & x_2^* & \dots & x_r^* \\ n_1 & n_2 & \dots & n_r \end{matrix}$$

при цьому $\sum_{i=1}^{r} n_i = n$. Обчислюють вибіркову середню $\bar{x}_В^*$ і середнє квадратичне відхилення σ_B^*.

2. Нормують випадкову величину X, тобто переходять до величини $Z = (X - \bar{x}_В^*)/\sigma_B^*$ і обчислюють кінці інтервалів $(z_i; z_{i+1}]$: $z_i = (x_i - \bar{x}_В^*)/\sigma_B^*$; $z_{i+1} =$

$(x_{i+1} - \bar{x}_\text{В}^*)/\sigma_B^*$ при чому найменше значення Z, тобто z_1 вважають рівним $-\infty$, а найбільше, тобто z_{s+1} вважають рівним ∞.

3. Обчислюють теоретичні ймовірності p_i потрапляння X до інтервалу $(x_i; x_{i+1})$ за рівністю $p_i = \Phi(z_{i+1}) - \Phi(z_i)$ і знаходять шукані теоретичні $n_i' = np_i$, де $\Phi(z)$=$\frac{1}{\sqrt{2\pi}}\int_0^z e^{-\frac{t^2}{2}} dt$.

4. Використовуючи таблицю χ^2-розподілу (додаток 3) і заданий за умовою рівень значущості α знаходять $\chi^2_\text{кр}$ з $k = r - 3$ ступенями вільності та критичну точку $\chi^2_\text{кр}(\alpha; k) = x_\text{кр}$. Якщо $\chi^2_\text{сп} < \chi^2_\text{кр}$ – немає підстав відкидати нульову гіпотезу. Якщо $\chi^2_\text{сп} > \chi^2_\text{кр}$ – нульову гіпотезу відкидають.

.

ВАРІАНТИ ІНДИВІДУАЛЬНИХ ЗАВДАНЬ (ІЗ № 3)

За двома незалежними вибірками з обсягами n і m, які добуто з генеральних сукупностей X та Y з розподілами $N(a_X, \sigma_X^2)$ та $N(a_Y, \sigma_Y^2)$ відповідно, знайдено вибіркові середні $\bar{x}_B$ та $\bar{y}_B$ і виправлені вибіркові дисперсії s_X^2 та s_Y^2 (табл.3.2). Потрібно для заданого варіанта V при рівні значущості

$$\alpha = \begin{cases} 0{,}01,\ V \le 10, \\ 0{,}05,\ 10 < V \le 20, \\ 0{,}1,\ V > 20, \end{cases}$$

перевірити нульові гіпотези:

а) $H_0: a_X = a_Y$ за альтернативної гіпотези $H_1: a_X \neq a_Y$, якщо відомі дисперсії σ_X^2 та σ_Y^2;

б) $H_0: \sigma_X^2 = \sigma_Y^2$ при умові, що σ_X^2 та σ_Y^2 – невідомо за альтернативної гіпотези $H_1: \sigma_X^2 > \sigma_Y^2$, і якщо вона приймається, то потім перевірити гіпотезу $H_0: a_X = a_Y$ за альтернативної гіпотези $H_1: a_X > a_Y$.

Таблиця 3.2

V **(Номер варіанту)**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
n	5	11	10	10	7	5	10	4	8	35
m	6	16	16	10	9	6	9	9	7	45
$\bar{x}_B$	20,1	31,2	12,7	14,3	150	3,3	17,3	70,5	37	14,3
$\bar{y}_B$	19,8	29,2	12	12,2	142	2,5	14,4	70,2	32,3	12
σ_X^2	1,75	1,3	7,5	35	34,6	0,72	10	0,5	0,16	34
σ_Y^2	1,38	1,15	6,1	42,1	28,5	0,87	15	1,5	0,27	42
s_X^2	2,3	0,84	9,7	22,15	22,8	0,25	14	2	0,016	22
s_Y^2	2,8	0,4	10,3	18,3	22,2	0,108	18	2,7	0,09	18

V (Номер варіанту)	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
n	11	8	16	6	16	10	50	6	10	6
m	11	7	13	6	21	9	35	6	8	8
$\bar{x}_B$	3,53	37,5	8,5	17,8	10,57	17,4	35,4	35,5	145,3	201,7
$\bar{y}_B$	2,067	32,4	6,2	17,08	9,62	14,5	31,2	31,4	142,3	193,6
σ_X^2	17,3	0,16	3,7	7	2,2	10	37,3	37	3,5	19,2
σ_Y^2	20,5	0,27	3,9	9	2,8	15	42,7	42,8	3,1	16,1
s_X^2	19	0,016	4,2	10	2,7	14	46,97	46,7	2,7	19,36
s_Y^2	23	0,087	4,2	16	3,2	19	23,2	23	3,2	16,25
V (Номер варіанту)	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
n	65	35	50	50	6	45	35	10	10	16
m	50	45	70	50	6	45	45	9	16	25
$\bar{x}_B$	0,05	130	8,7	150	-30,5	8,25	145,3	17,4	13	37,5
$\bar{y}_B$	-0,03	125	6,1	141	-34,2	6,2	142,3	14,4	12	36,8
σ_X^2	0,32	60	100	34,6	63,3	3,7	3,5	10	7,4	1
σ_Y^2	0,38	80	74	28,6	58,5	3,9	3	15	6	1,1
s_X^2	0,0324	70	90	23	83	4,3	2,7	14	9,7	1,2
s_Y^2	0,0225	90	85	22	12,7	4,2	3,2	19	10	1,44

2) Використовуючи критерій Пірсона для заданого варіанта V при рівні значущості

$$\alpha = \begin{cases} 0{,}01,\ V \leq 10, \\ 0{,}05,\ 10 < V \leq 20, \\ 0{,}1,\ V > 20, \end{cases}$$

перевірити гіпотезу про нормальний розподіл генеральної сукупності, якщо з неї вилучена наступна вибірка (табл. 3.3):

Таблиця 3.3

Номер варіанту (**V**)	Дані вибірки
1	23 45 32 24 45 25 47 41 35 33 24 46 44 34 33 36 37 24 28 30 43 33 29 37 38 35 36 27 24 29 38 36 41 40 37 38 41 38 36 32 25 29 36 47 36 33 35 27 26 38
2	112 145 160 134 156 115 119 128 121 135 160 144 145 123 125 125 145 160 155 154 155 146 134 118 128 130 128 113 112 145 156 124 135 128 134 151 137 123 146 160 143 154 137 139 129 146 157 150 151 128
3	1 3 4 5 4 6 7 2 4 7 8 2 6 3 4 8 9 2 5 6 7 8 9 2 4 6 7 8 9 1 4 5 5 5 5 5 5 5 7 7 3 4 4 8 5 5 6 7 4 3 4 6 5 3 3 4 6 8 9 1
4	47 47 49 40 50 51 52 53 57 49 45 46 45 48 50 51 52 55 58 48 49 42 46 48 50 52 51 57 58 51 52 57 46 49 43 48 49 50 58 51 49 43 48 45 47 50 56 55 50 48
5	345 302 356 342 341 300 325 324 350 324 311 346 301 325 350 345 311 315 324 328 345 346 341 321 322 321 318 319 311 315 321 347 350 325 318 301 305 315 306 347 303 308 316 329 308 315 350 327 301 302
6	73 95 82 74 95 75 97 91 85 83 74 96 94 84 83 86 87 74 78 80 93 83 79 87 88 85 86 77 74 79 88 86 91 90 87 88 91 88 86 82 75 79 86 97 86 83 85 77 76 88

7	612 645 660 634 656 615 619 628 621 635 660 644 645 623 625 645 660 655 656 657 646 634 617 628 630 628 612 647 656 647 624 635 637 639 651 637 623 646 660 643 654 637 639 629 646 657 650 651 628 639
8	10 13 14 15 14 16 17 12 4 7 8 2 6 3 4 8 9 12 5 6 17 8 9 12 4 16 7 18 9 11 4 5 5 5 15 5 15 5 5 6 13 8 4 8 2 11 4 5 6 16 16 7 17 13 4 4 8 5 5 6
9	27 27 29 30 30 31 32 33 37 29 25 26 25 28 30 31 32 35 38 28 29 22 26 28 30 32 31 37 38 31 32 37 26 29 23 28 29 30 38 31 29 23 28 25 27 30 36 35 30 28
10	375 332 356 342 371 330 325 324 350 324 361 346 351 355 350 345 371 345 354 328 345 346 341 321 322 321 348 369 351 345 351 347 350 325 368 341 345 345 362 347 353 358 356 329 348 345 350 337 331 362
11	423 445 432 424 445 425 447 441 435 433 424 446 444 434 433 436 437 424 428 430 443 433 429 437 438 435 436 427 424 429 438 436 441 440 437 438 441 438 436 432 425 429 436 447 436 433 435 427 426 438
12	12 45 10 34 56 15 19 28 21 35 60 44 45 23 25 45 60 55 56 57 46 34 17 28 30 28 12 47 56 47 24 35 37 39 51 37 23 46 60 43 54 37 39 29 46 57 50 51 28 39
13	21 23 24 25 24 26 27 22 24 27 28 22 26 23 24 2829 22 25 26 27 28 29 22 24 26 27 28 29 21 24 25 25 25 25 25 25 25 25 26 23 28 24 28 22 21 24 25 26 26 26

14	4140 49 40 50 51 58 53 67 69 45 46 48 68 60 51 52 65 68 48 49 62 46 48 60 52 51 57 58 61 52 67 46 49 63 48 49 50 58 51 49 63 48 45 67 50 66 55 50 48 63
15	45 12 56 42 41 30 25 24 50 24 11 46 11 25 50 45 11 15 24 28 45 46 41 21 22 21 18 19 11 15 21 47 50 25 18 31 35 15 36 34 33 38 16 29 38 15 50 27 31 32.
16	61 71 72 74 75 75 67 71 75 83 74 76 74 74 83 76 67 64 68 70 83 83 69 67 78 75 76 77 74 69 68 66 81 80 67 78 81 68 76 72 75 69 66 67 76 73 75 77 76 68
17	212 245 260 234 256 215 219 228 221 235 260 244 245 223 225 245 260 255 256 257 246 234 217 228 230 228 212 247 256 247 224 235 237 239 251 237 223 246 260 243 244 237 239 229 246 257 250 251 228 239
18	5 3 4 5 4 6 7 2 4 7 8 7 6 3 4 8 9 8 8 6 7 8 9 2 4 6 7 8 9 7 4 5 5 3 9 8 5 9 7 6 3 8 4 8 6 7 4 5 6 6 6 7 7 3 4 9 8 5 5 6 7
19	547 547 549 540 560 551 552 553 557 549 545 546 545 548 550 551 552 555 558 548 569 542 560 548 550 552 551557 558551 552 557 546 549 553 548 549 550 558 551 549 543 548 535 537 550 556 555 550 548
20	45 72 564241 30 32 32 35 32 31 46 31 32 50 34 31 35 34 38 45 46 41 21 32 31 31 39 31 35 21 47 50 35 38 31 35 35 36 47 33 38 36 29 38 31 35 27 31 32 47
21	891 845 823 874 868 893 874 863 854 874 892 874 848 832 845 890 873 823 845 895 849 874 836 826 840 850 836 891 856 854 835 832 865 878 848 867 845 867 856 857 859 849 868 871 873 846 868 847 857 860
22	55 57 66 70 67 54 55 51 66 63 64 56 70 54 63 56 57 64 68 70 53 53 69 67 68 70 56 67 54 69 68 66 51 60 67 68 51 68 56 52 55 69 66 67 71 53 65 67 66 68

23	18 19 28 19 16 17 30 27 25 18 16 14 13 28 29 14 13 15 16 15 28 22 19 13 10 11 27 13 17 22 25 28 21 12 16 13 10 18 16 29 27 21 22 21 19 17 15 17 20 22
24	11 19 20 19 16 18 10 25 25 19 11 10 14 13 18 14 13 15 16 15 18 21 19 11 10 11 20 13 16 21 20 18 11 12 16 13 10 18 16 19 23 21 14 21 17 17 15 17 10 20
25	55 54 59 56 52 51 50 54 58 57 59 50 59 54 55 53 56 57 58 52 51 52 53 53 61 62 63 61 62 59 63 69 67 63 69 62 53 58 59 54 55 67 63 64 66 65 62 60 61 65
26	125 125 134 140 123 120 138 135 130 131 132 129 122 124 140 137 138 136 135 134 130 132 129 128 128 139 129 126 120 125 123 128 129 128 125 130 131 132 134 132 131 129 128 127 130 131 135 138 136 134
27	99 91 87 101 104 112 121 123 105 115 93 91 89 87 104 102 99 95 85 87 90 92 94 100 106 109 104 117 102 120 118 115 94 93 103 118 113 116 112 110 87 88 95 93 96 93 90 87 99 100 101 104
28	234 235 233 231 239 240 228 228 238 234 240 218 225 228 222 236 240 233 241 217 234 231 222 226 238 239 228 225 227 223 224 237 230 231 228 229 223 218 216 220 230 237 240 236 230 229 227 226 230 227
29	2 5 7 9 9 10 17 15 12 10 11 12 12 11 15 19 20 11 13 7 8 9 4 7 8 9 2 9 9 10 11 10 7 8 10 10 10 11 12 9 8 8 7 9 11 10 12 13 14 12 12 13 9 7 8 5 3 2 11 15 17
30	44 45 34 34 35 36 37 38 39 49 30 40 40 4141 42 45 46 48 44 43 42 45 46 47 48 49 50 50 49 48 48 45 33 32 38 35 45 38 39 33 30 33 34 32 36 37 45 44 40 43

3) Використовуючи критерій Пірсона, при рівні значущості α перевірити, чи узгоджується гіпотеза про закон розподілу генеральної сукупності X зі

статистичними даними, які подані у вигляді статистичного або інтервального варіаційних рядів (у першому рядку вказано варіанти x_i або часткові інтервали варіант $(x_i; x_{i+1}]$, у другому – відповідні їм частоти n_i). Номер варіанту співпадає з порядковим номером задачі.

1. $\alpha = 0{,}05$; рівномірний розподіл;

$(x_i; x_{i+1}]$	$(-4; -2]$	$(-2; 0]$	$(0; 2]$	$(2; 4]$	$(4; 6]$	$(6; 8]$	$(8; 10]$
n_i	28	27	34	23	25	31	32

2. $\alpha = 0{,}05$; показниковий розподіл;

$(x_i; x_{i+1}]$	$(0; 6]$	$(6; 12]$	$(12; 18]$	$(18; 24]$	$(24; 30]$	$(30; 36]$	$(36; 42]$
n_i	115	51	18	9	4	2	1

3. $\alpha = 0{,}05$; біноміальний розподіл;

x_i	0	1	2	3	4	5	6	7
n_i	4	23	31	23	11	5	2	1

4. $\alpha = 0{,}1$; розподіл Пуассона;

x_i	0	1	2	3	4	5	6	7
n_i	242	349	234	107	43	21	3	1

5. $\alpha = 0{,}05$; рівномірний розподіл;

$(x_i; x_{i+1}]$	$(4; 7]$	$(7; 10]$	$(10; 13]$	$(13; 16]$	$(16; 19]$	$(19; 22]$	$(22; 25]$
n_i	12	9	21	8	11	19	20

6. $\alpha = 0{,}05$; біноміальний розподіл;

x_i	0	1	2	3	4	5	6	7	8	9
n_i	2	12	31	42	39	35	27	5	4	2

7. $\alpha = 0{,}01$; розподіл Пуассона;

x_i	0	1	2	3	4	5	6	7
n_i	427	363	154	41	9	3	2	1

8. $\alpha = 0{,}05$; показниковий розподіл;

$(x_i; x_{i+1}]$	$(0; 4]$	$(4; 8]$	$(8; 12]$	$(12; 16]$	$(16; 20]$	$(20; 24]$	$(24; 28]$
n_i	121	95	76	56	45	36	21

9. $\alpha = 0{,}1$; біноміальний розподіл;

x_i	0	1	2	3	4	5	6	7
n_i	2	3	10	22	26	20	12	5

10. $\alpha = 0{,}05$; рівномірний розподіл;

$(x_i; x_{i+1}]$	$(22; 20{,}5]$	$(20{,}5; 21]$	$(21; 21{,}5]$	$(21{,}5; 22]$	$(22; 22{,}5]$	$(22{,}5; 23]$
n_i	91	76	75	74	65	69

11. $\alpha = 0{,}01$; показниковий розподіл;

$(x_i; x_{i+1}]$	$(0; 10]$	$(10; 20]$	$(20; 30]$	$(30; 40]$	$(40; 50]$	$(50; 60]$	$(60; 70]$
n_i	365	245	150	100	70	45	25

12. $\alpha = 0{,}05$; біноміальний розподіл;

x_i	0	1	2	3	4	5
n_i	72	77	34	14	2	1

13. $\alpha = 0{,}01$; показниковий розподіл;

$(x_i; x_{i+1}]$	$(0; 1]$	$(1; 2]$	$(2; 3]$	$(3; 4]$	$(4; 5]$	$(5; 6]$	$(6; 7]$
n_i	259	167	119	74	80	57	44

14. $\alpha = 0{,}05$; розподіл Пуассона;

x_i	0	1	2	3	4	5
n_i	505	336	125	24	8	2

15. $\alpha = 0{,}05$; біноміальний розподіл;

x_i	0	1	2	3	4	5	6
n_i	8	20	42	22	8	54	46

16. $\alpha = 0{,}1$; рівномірний розподіл;

$(x_i; x_{i+1}]$	$(-30; -20]$	$(-20; -10]$	$(-10; 0]$	$(0; 10]$	$(10; 20]$	$(20; 30]$
n_i	25	30	40	45	40	20

17. $\alpha = 0{,}05$; розподіл Пуассона;

x_i	0	1	2	3	4	5
n_i	132	43	40	20	7	8

18. $\alpha = 0{,}01$; рівномірний розподіл;

$(x_i; x_{i+1}]$	$(8; 9]$	$(9; 10]$	$(10; 11]$	$(11; 12]$	$(12; 13]$	$(13; 14]$	$(14; 15]$
n_i	12	45	28	16	28	6	11

19. $\alpha = 0{,}01$; розподіл Пуассона;

x_i	0	1	2	3	4	5	6
n_i	405	366	175	40	8	4	2

20. $\alpha = 0{,}025$; показниковий розподіл;

$(x_i; x_{i+1}]$	$(12; 15]$	$(15; 18]$	$(18; 21]$	$(21; 24]$	$(24; 27]$	$(27; 30]$	$(30; 33]$
n_i	10	23	30	65	55	24	33

21. $\alpha = 0{,}01$; розподіл Пуассона;

x_i	0	1	2	3	4	5	6	7
n_i	199	169	87	31	9	3	1	1

22. $\alpha = 0{,}05$; рівномірний розподіл;

$(x_i; x_{i+1}]$	$(0{,}5; 5{,}5]$	$(5{,}5; 10{,}5]$	$(10{,}5; 15{,}5]$	$(15{,}5; 20{,}5]$	$(20{,}5; 25{,}5]$
n_i	23	27	35	60	55

23. $\alpha = 0{,}05$; показниковий розподіл;

$(x_i; x_{i+1}]$	$(1{,}2; 5{,}2]$	$(5{,}2; 9{,}2]$	$(9{,}2; 13{,}2]$	$(13{,}2; 17{,}2]$	$(17{,}2; 21{,}2]$
n_i	18	28	38	66	42

24. $\alpha = 0{,}01$; розподіл Пуассона;

x_i	0	1	2	3	4
n_i	109	65	22	3	1

25. $\alpha = 0{,}05$; рівномірний розподіл;

$(x_i; x_{i+1}]$	$(2{,}6; 6{,}6]$	$(6{,}6; 10{,}6]$	$(10{,}6; 14{,}6]$	$(14{,}6; 18{,}6]$	$(18{,}6; 22{,}6]$
n_i	6	12	17	45	38

26. $\alpha = 0{,}1$; біноміальний розподіл;

x_i	1	2	3	4	5	6	7
n_i	300	110	55	13	10	9	5

27. $\alpha = 0{,}1$; розподіл Пуассона;

x_i	0	1	2	3	4	5	6	7
n_i	175	130	52	40	23	21	7	2

28. $\alpha = 0{,}025$; рівномірний розподіл;

$(x_i; x_{i+1}]$	$(2; 6]$	$(6; 10]$	$(10; 14]$	$(14; 18]$	$(18; 22]$	$(22; 26]$
n_i	26	16	15	25	23	15

29. $\alpha = 0{,}05$; показниковий розподіл;

$(x_i; x_{i+1}]$	$(10; 13]$	$(13; 16]$	$(16; 19]$	$(19; 22]$	$(22; 25]$	$(25; 28]$
n_i	12	14	30	29	39	16

30. $\alpha = 0{,}1$; біноміальний розподіл;

x_i	0	1	2	3	4	5
n_i	2	20	37	52	43	6

РОЗВ'ЯЗАННЯ ТИПОВОГО ВАРІАНТА ІЗ № 3

Умова завдання

1) Нехай за двома незалежними вибірками, об'єми яких становлять n =7, $m = 9$, які отримано з генеральних сукупностей X та Y, знайдено вибіркові середні $\bar{x}_{\text{в}} = 68{,}1$ та $\bar{y}_{\text{в}} = 67{,}6$ і виправлені дисперсії $\mathrm{s}_{\mathrm{X}}^2 = 30{,}2$ та $\mathrm{s}_{\mathrm{Y}}^2 = 29{,}2$. Потрібно при рівні значущості $\alpha = 0{,}05$ перевірити нульові гіпотези:

а) $H_0 : a_X = a_Y$ за альтернативної гіпотези $H_1 : a_X \neq a_Y$, якщо відомі дисперсії $\sigma_X^2 = 26{,}6$ та $\sigma_Y^2 = 24{,}3$;

б) $H_0 : \sigma_X^2 = \sigma_Y^2$ при умові, що σ_X^2 та σ_Y^2 – невідомо за альтернативної гіпотези $H_1 : \sigma_X^2 > \sigma_Y^2$, і якщо вона приймається, то потім перевірити гіпотезу $H_0 : a_X = a_Y$ за альтернативної гіпотези $H_1 : a_X > a_Y$.

Хід розв'язання.

а) За схемою (3.3) знаходимо:

$$\left.\begin{array}{c} \bar{x}_B = 68{,}1; n = 7; \sigma_X^2 = 26{,}6 \\ \bar{y}_B = 67{,}6; m = 9; \sigma_Y^2 = 24{,}3 \\ \alpha = 0{,}05 \end{array}\right\} \to H_0 : a_x = a_y \to$$

$$Z_{\text{сп}} = \frac{68{,}1 - 67{,}6}{\sqrt{\frac{26{,}6}{7} + \frac{24{,}3}{9}}} = 0{,}196 \to H_1 : a_x \neq a_y \to \Phi(Z_{\text{кр}}) =$$

$$= 0{,}475 \xrightarrow{\text{дод.2}} Z_{\text{кр}} = 1{,}96 \to |Z_{\text{сп}}| = 0{,}196 < Z_{\text{кр}} = 1{,}96 \to \boldsymbol{P}.$$

Отже, гіпотеза про рівність середніх значень заданих нормальних вибірок при рівні значущості $\alpha = 0{,}05$ є вірною. ■

б) Оскільки $\mathrm{s}_{\mathrm{X}}^2$ та $\mathrm{s}_{\mathrm{Y}}^2$ різні, то перевіримо нульову гіпотезу $H_0 : \sigma_X^2 = \sigma_Y^2$, використовуючи критерій Фішера $F = \frac{\mathrm{s}_{\text{Б}}^2}{\mathrm{s}_{\text{М}}^2}$ –відношення більш виправленої дисперсії до меншої, з $(n - 1; m - 1) = (6; 8)$ ступенями вільностей. Дисперсія $\mathrm{s}_{\mathrm{X}}^2$ більша за дисперсію $\mathrm{s}_{\mathrm{Y}}^2$, тому за схемою (3.5) маємо

$$\left.\begin{array}{c} s_{\text{Б}}^2 = 30{,}2; n = 7; k_1 = 6 \\ s_{\text{М}}^2 = 29{,}2; m = 9; k_2 = 8 \\ \alpha = 0{,}05 \end{array}\right\} \to H_0: \sigma_X^2 = \sigma_Y^2 \to F_{\text{сп}} = \frac{30{,}2}{29{,}2} \approx 1{,}03 \to$$

$$\to H_1: \sigma_X^2 > \sigma_Y^2 \to F_{\text{кр}}(0{,}05; 6; 8) \xrightarrow{\text{дод.5}} F_{\text{кр}} = 3{,}58 \to$$

$$\to F_{\text{сп}} = 1{,}03 < 3{,}58 \to \boldsymbol{P}.$$

Отже, гіпотезу H_0 про рівність дисперсій приймаємо.

б)Тоді за схемою (3.4) знаходимо

$$\left.\begin{array}{c} \bar{x}_B = 68{,}1; n = 7; s_X^2 = 30{,}2 \\ \bar{y}_B = 67{,}6; m = 9; s_Y^2 = 29{,}2 \\ D(X) = D(Y) \\ k = 14; \alpha = 0{,}05 \end{array}\right\} \to H_0: a_x = a_y \to$$

$$T_{\text{сп}} = \frac{68{,}1 - 67{,}6}{\sqrt{(7-1)\cdot 30{,}2 + (9-1)\cdot 29{,}2}}\sqrt{\frac{7\cdot 9(7+9-2)}{7+9}} =$$

$$= \frac{0{,}5}{\sqrt{414{,}8}} \cdot \sqrt{55{,}125} \approx 0{,}19 \to H_1: a_x > a_y \to t_{\text{одн.кр}}(0{,}95; 14) \xrightarrow{\text{дод.4}} t_{\text{кр}} = 1{,}761 \to$$

$$\to T_{\text{сп}} = 0{,}19 < t_{\text{кр}} = 1{,}761 \to \boldsymbol{P}.$$

Отже, гіпотезу H_0 про рівність математичних сподівань приймаємо. ■

Умова завдання

2) Використовуючи критерій Пірсона при рівні значущості α=0,05 потрібно перевірити гіпотезу про нормальний розподіл генеральної сукупності, якщо з неї вилучена наступна вибірка X обсягом $n = 50$:

346 345 329 348 356 369 359 328 337 339 345 360

320 329 339 348 333 344 345 356 355 346 349 350

327 328 347 333 333 334 354 345 365 345 356 358

348 342 341 326 329 330 350 351 331 348 349 327 359 330.

Хід розв'язання.

Обчислимо спочатку частоту кожної варіанти із варіаційного ряду і складемо таблицю, тобто отримаємо дискретний статистичний розподіл вибірки (див. табл. 3.4).

Таблиця 3.4

x_i	320	326	327	328	329	330	331	333	334	337	339	341
n_i	1	1	2	2	3	2	1	3	1	1	2	1

x_i	342	344	345	346	347	348	349	350	351	354	355	356
n_i	1	1	5	2	1	4	2	2	1	1	1	3

x_i	358	359	360	365	369
n_i	1	2	1	1	1

Оскільки найбільша варіанта рівна 369, а найменша 320, то вся вибірка попаде в інтервал (320; 369). Розіб'ємо варіаційний ряд на 7 інтервалів. Довжина кожного інтервалу дорівнюватиме $h = \frac{369-320}{7} = 7$. Отримаємо наступні сім інтервалів: [320; 327), [327; 333), [334; 341), …, [362; 369), а відповідний інтервальний варіаційний ряд представлено в таблиці 3.5:

Таблиця 3.5

x_i	[320; 327)	[327; 333)	[334; 341)	[341; 348)	[348; 355)	[355; 362)	[362; 369)
n_i	2	13	4	11	10	8	2

Обчислимо вибіркове середнє $\bar{x}_{\text{в}}$ та середньо-квадратичне відхилення вибірки σ_B, використовуючи таблицю 3.4:

$$\bar{x}_{\text{в}} = \frac{\sum_{i=1}^{m} x_i \cdot n_i}{n} = \frac{320\cdot 1 + 326\cdot 1 + 327\cdot 2 + \cdots + 365\cdot 1 + 369\cdot 1}{50} = 343{,}22,$$

$$D_B = \frac{1}{n}\sum_{i=1}^{m} x_i^2 \cdot n_i - \bar{x}_B^2 = \frac{1}{50}(320^2 \cdot 1 + 326^2 \cdot 1 + 327^2 \cdot 2 + \cdots + 369^2 \cdot 1) -$$

$$-(343{,}22)^2 = 130{,}97, \sigma_B = \sqrt{D_B} = 11{,}53.$$

Знайдемо нормовані інтервали $(z_i; z_{i+1}]$: $z_i = (x_i - \bar{x}_{\text{в}})/\sigma_B$; $z_{i+1} = (x_{i+1} - \bar{x}_{\text{в}})/\sigma_B$, враховуючи, що вибіркове середнє $\bar{x}_{\text{в}} = 343{,}22$ та середньо квадратичне відхилення $\sigma_B = 11{,}53$. Для цього складемо розрахункову таблицю 3.6, причому найменше значення, тобто лівий кінець першого інтервалу покладемо рівним $-\infty$, а найбільше, тобто правий кінець останнього інтервалу – рівним ∞.

Таблиця 3.6

i	x_i	x_{i+1}	$x_i - \bar{x}_{\text{в}}$	z_i	$x_{i+1} - \bar{x}_{\text{в}}$	z_{i+1}
1	320	327	-23,22	$-\infty$	-16,22	-1,40676
2	327	334	-16,22	-1,40676	-9,22	-0,79965
3	334	341	-9,22	-0,79965	-2,22	-0,19254
4	341	348	-2,22	-0,19254	4,78	0,414571
5	348	355	4,78	0,414571	11,78	1,021683
6	355	362	11,78	1,021683	18,78	1,628794
7	362	369	18,78	1,628794	25,78	∞

Визначимо теоретичні ймовірності $p_i = \Phi(z_{i+1}) - \Phi(z_i)$, де $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}dt$ і теоретичні частоти $n_i' = n \cdot p_i$. Для цього складемо розрахункову таблицю 3.7.

Таблиця 3.7

i	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	p_i
1	$-\infty$	-1,40676	-0,5	-0,4207	0,0793
2	-1,40676	-0,79965	-0,4207	-0,2881	0,1326
3	-0,79965	-0,19254	-0,2881	-0,0753	0,2128
4	-0,19254	0,414571	-0,0753	0,1591	0,2344
5	0,414571	1,021683	0,1591	0,3461	0,187
6	1,021683	1,628794	0,3461	0,4484	0,1023
7	1,628794	∞	0,4484	0,5	0,0516

Порівняємо емпіричні та теоретичні частоти, використовуючи критерій Пірсона. Для цього обчислимо спочатку спостережуване значення критерію Пірсона за допомогою розрахункової таблиці 3.8. Враховуючи, що в розглянутому емпіричному розподілі частоти 1-го, 3-го, 6-го та 7-го інтервалів ($n_1 = 2; n_3 = 4; n_6 = 8; n_7 = 2$) менше 5, то при використанні критерію Пірсона, відповідно до зауваження 3.2, доцільно об’єднати зазаначені інтервали з сусідніми.

Таблиця 3.8

<table>
<tr><th>i</th><th>n_i</th><th>n_i'</th><th>$n_i - n_i'$</th><th>$(n_i - n_i')^2$</th><th>$(n_i - n_i')^2/n_i'$</th><th>n_i^2/n_i'</th></tr>
<tr><td>1</td><td rowspan="2">$\left.\begin{matrix}2\\13\end{matrix}\right\}$</td><td rowspan="2">$\left.\begin{matrix}3{,}965\\6{,}63\end{matrix}\right\}$</td><td rowspan="2">4,405</td><td rowspan="2">19,40403</td><td rowspan="2">1,831432</td><td rowspan="2">21,23643</td></tr>
<tr><td>2</td></tr>
<tr><td>3</td><td rowspan="2">$\left.\begin{matrix}4\\11\end{matrix}\right\}$</td><td rowspan="2">$\left.\begin{matrix}10{,}64\\11{,}72\end{matrix}\right\}$</td><td rowspan="2">-7,36</td><td rowspan="2">54,1696</td><td rowspan="2">2,422612</td><td rowspan="2">10,06261</td></tr>
<tr><td>4</td></tr>
<tr><td>5</td><td>10</td><td>9,35</td><td>0,65</td><td>0,4225</td><td>0,045187</td><td>10,69519</td></tr>
<tr><td>6</td><td rowspan="2">$\left.\begin{matrix}8\\2\end{matrix}\right\}$</td><td rowspan="2">$\left.\begin{matrix}5{,}115\\2{,}58\end{matrix}\right\}$</td><td rowspan="2">2,305</td><td rowspan="2">5,313025</td><td rowspan="2">0,690452</td><td rowspan="2">12,99545</td></tr>
<tr><td>7</td></tr>
<tr><td>Σ</td><td>50</td><td>50</td><td></td><td></td><td>$\chi^2_{\text{сп}}$ =4,989683</td><td>54,98968</td></tr>
</table>

Контроль: оскільки

$$\sum_{i=1}^{4}\frac{n_i^2}{np_i} - n = 54{,}98968 - 50 = 4{,}98968 = \chi^2_{\text{сп}},$$

то обчислення проведено правильно.

Оскільки нове число інтервалів (з урахуванням об'єднання) $r = 4$, а нормальний закон розподілу визначається кількістю $s = 2$ параметрами, то число ступенів вільності

$$k = r - s - 1 = 1.$$

Тому *з*а таблицею критичних точок розподілу χ^2 (див. додаток 3) при рівні значущості α=0,05 і кількості ступенів вільності $k = 1$, знайдемо критичну точку правосторонньої критичної області:

$$\chi^2_{\text{кр}}(0{,}05; 1) = x_{\text{кр}} = 3{,}84.$$

Оскільки $\chi^2_{\text{сп}} > \chi^2_{\text{кр}}$, то гіпотезу про нормальний розподіл відхиляємо. ■

Умова завдання

3) Використовуючи критерій Пірсона, на рівні значущості $\alpha = 0.05$ перевірити гіпотезу про показниковий розподіл генеральної сукупності X зі статистичними даними, які подані у вигляді інтервального варіаційного ряду.

$(x_i; x_{i+1}]$	$[0; 5)$	$[5; 10)$	$[10; 15)$	$[15; 20)$	$[20; 25)$	$[25; 30)$
n_i	133	45	15	4	2	1

Хід розв'язання

Імовірність потрапляння випадкової величини, що має показниковий розподіл, в інтервал Δ_i обчислюється за формулою:

$$P\{x_i < X < x_{i+1}\} = e^{-\lambda x_i} - e^{-\lambda x_{i+1}},$$

де λ – параметр показникового розподілу – величина, обернена до математичного сподівання.

Тому для знаходження середнього $\bar{x}_B$, перетворимо інтервальний статистичний розподіл на точковий:

x_i	2.5	7.5	12.5	17.5	22.5	27.5
n_i	133	45	15	4	2	1

Отже,

$$\bar{x}_B = \frac{1}{200}(2.5 \cdot 133 + 7.5 \cdot 45 + 12.5 \cdot 15 + 17.5 \cdot 4 + 22.5 \cdot 2 + 27.5) = 5$$

і $\lambda = \frac{1}{\bar{x}_B} = \frac{1}{5} = 0.2$.

Оскільки в останніх трьох інтервалах частоти менші за 5, то об'єднаємо ці інтервали в один $[15;\ 30]$ із частотою 7. Подальші обчислення подаємо у вигляді розрахункової табл. 3.9:

Таблиця 3.9

i	$-\lambda x_i$	$e^{-\lambda x_i}$	p_i	np_i	$n_i - np_i$	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
1	0	1					
			0.6321	126.42	6.58	43.2964	0.3425
2	-1	0.3679					
			0.2326	46.52	-1.52	2.3104	0.0497
3	-2	0.1353					
			0.0855	17.1	-2.1	4.41	0.2579
4	-3	0.0498					
			0.0473	9.46	-2.46	6/0516	0.6397
	-6	0.0025					

Тоді $\chi^2 = 0.3425 + 0.0497 + 0.2579 + 0.6397 \approx 1.29$. Із таблиць розподілу χ^2 при $k = r - 2 = 4 - 2 = 2$ і $\alpha = 0.05$ знаходимо $\chi^2_{\text{сп}} = 6{,}0$ (додаток 3). Оскільки $\chi^2 < \chi^2_{\text{сп}}$, то гіпотеза про показниковий розподіл генеральної сукупності X стверджується при заданому рівні значущості. ■

РОЗДІЛ 4. ЕЛЕМЕНТИ КОРЕЛЯЦІЙНОГО ТА РЕГРЕСІЙНОГО АНАЛІЗУ

Література: [2, розділи 3-4], [3, модуль 3, т. 3], [6, розділ 4, т. 16], [7, розділ 12].

Базові поняття. Вибірка, двовимірний статистичний розподіл, парний статистичний розподіл. Статистичні оцінки параметрів системи. Перевірка гіпотези про незалежність системи двох випадкових величин. Перевірка гіпотези про однорідність вибірок. Рівняння лінійної регресії. Нелінійна регресія. Метод найменших квадратів.

Основні задачі. Виявлення форми функціональної залежності між змінними (лінійна чи нелінійна) за наявності кореляційного зв'язку між ними.

ЩО МАЄ ЗНАТИ ТА ВМІТИ СТУДЕНТ

1. Знання на рівні понять, означень, формулювань. Двовимірний статистичний розподіл вибірки та його числові характеристики. Умовні статистичні розподіли та їхні числові характеристики. Парний статистичний розподіл вибірки та його числові характеристики. Перевірка гіпотези про незалежність системи двох випадкових величин та гіпотези про однорідність вибірок. Поняття стохастичної залежності між змінними, коефіцієнт кореляції.
Поняття регресії. Лінійна і квадратична регресія та відшукання оцінок їхніх параметрів.

2. Уміння в розв'язуванні задач. Знаходити коефіцієнт кореляції. Знаходити рівняння лінійної та квадратичної регресії.

ОСНОВНІ ТЕОРЕТИЧНІ ВІДОМОСТІ

4.1. Елементи теорії кореляції. Функціональна, статистична та кореляційна залежності

На практиці ми зустрічаємось з різними типами зв'язку між випадковими величинами. В багатьох прикладних задачах необхідно виявити залежність між двома властивостями (ознаками) X і Y одного і того ж об'єкта або між певними ознаками різних об'єктів. Якщо вказані ознаки допускають кількісне вимірювання, і, виходячи з характеристики об'єкта, ознака Y залежить від ознаки X. Тоді X можна назвати *незалежною змінною* або *факторною ознакою*, а Y – *залежною змінною* або *результативною ознакою*.

Якщо кожному значенню факторної ознаки X відповідає одне і тільки одне значення результативної ознаки Y, то говорять, що між цими ознаками існує *функціональний зв'язок*: $Y = f(X)$. Така залежність виникає тоді, коли зв'язок між величинами настільки тісний, що, знаючи можливі значення однієї з величин, можна точно вказати можливі значення іншої. Нехай, наприклад, $Y = X^2$, можливі значення величини $X: 1, 2, 3$. Тоді можливі значення величини $Y: 1, 4, 9$.

До величин, що пов'язані функціональними залежностями, відносяться значна кількість тих, що зустрічаються у природознавстві. Наприклад, тиск P, температура T та об'єм газу V пов'язані функціональною залежністю $P = RT/V$, де R – універсальна газова стала.

Разом з цим точна функціональна залежність між величинами реалізується досить рідко, оскільки величини підлягають впливу багатьох випадкових факторів. Серед цих факторів можуть бути спільні для обох величин. В таких випадках виникають статистичні (стохастичні) залежності. Наприклад, якщо величина Y залежить від випадкових факторів Z_1, Z_2, Z_3, Z_4, а величина X – від випадкових факторів Z_3, Z_4, Z_5, то між Y та X є статистична залежність – серед випадкових факторів, що діють на величини Y та X, є спільні – Z_3, Z_4.

Означення 4.1. *Статистичною (стохастичною) залежністю* між величинами називають таку залежність, коли зміна однієї з величин викликає зміну розподілу іншої.

Вивчення статистичного зв'язку вважається дуже складним і трудомістким процесом, у якому потрібно аналізувати багатовимірні таблиці даних. Тому, зазвичай, вивчається не статистичний, а кореляційний зв'язок між X та Y.

Якщо кожному значенню факторної ознаки X відповідає певне середнє значення результативної ознаки Y, то говорять, що між цими ознаками існує *кореляційний зв'язок*. Тобто кореляційною є функціональна залежність між значеннями X і середніми значеннями Y: $\bar{Y} = f(\mathrm{X})$.

Кореляційні зв'язки є частинним випадком статистичних. Термін «кореляція» походить від англійського *correlation* – співвідношення, відповідність. При таких залежностях певному значенню однієї з величин може відповідати одразу декілька значень іншої.

Прикладами кореляційної зв'язку є:

1. Товарознавець цікавиться, чи залежить обсяг продажів в цьому місяці від обсягу реклами протягом цього періоду.

2. Викладач хоче з'ясувати, чи є залежність між кількістю годин, витрачених студентом на заняття, і оцінкою, отриманою під час іспиту.

3. Лікар досліджує, чи впливає кофеїн на серцево-судинні захворювання або чи існує зв'язок між віком людини і її артеріальним тиском.

4. Соціолог досліджує, наскільки сильний зв'язок між рівнем злочинності і рівнем безробіття в регіоні. Чи існує залежність між витратами на житло і сукупнім доходом сім'ї. Чи пов'язані дохід від професійної діяльності і ступеня освіти.

Показником, що вимірює стохастичний зв'язок між змінними, є *коефіцієнт кореляції,* який свідчить з певною ймовірності, наскільки зв'язок між змінними близький до строгої лінійної залежності.

Для двовимірного статистичного розподілу вибірки введемо поняття статистичної залежності між ознаками X та Y.

Статистичною залежністю ознаки X від *ознаки* Y називають таку залежність, за якої зі зміною значень ознаки $Y = y_i$ змінюється умовний статистичний розподіл ознаки X.

Статистичною залежністю ознаки Y від *ознаки* X називають таку залежність, за якої зі зміною значень ознаки $X = x_j$ змінюється умовний статистичний розподіл ознаки Y.

Умовним статистичним розподілом ознаки Y при фіксованому значенні $X = x_j$ називають перелік варіант ознаки Y та відповідних їм частот при фіксованому значенні X.

$$Y|X = x_j$$

$Y = y_i$	y_1	y_2	y_3	$\dots$	y_k
n_{ij}	n_{1j}	n_{2j}	n_{3j}	$\dots$	n_{kj}

Тут $\sum_{i=1}^{k} n_{ij} = n_{x_j}$.

Умовним статистичним розподілом ознаки X при фіксованому значенні $Y = y_i$ називають перелік варіант ознаки X та відповідних їм частот, узятих при фіксованому значенні Y.

$$X|Y = y_i$$

$X = x_j$	x_1	x_2	x_3	$\dots$	x_m
n_{ij}	n_{i1}	n_{i2}	n_{i3}	$\dots$	n_{im}

Тут $\sum_{j=1}^{m} n_{ij} = n_{y_i}$.

Якщо частота спільної появи ознак X і Y $n_{ij} = 1$ для всіх варіант, то в цьому разі двовимірний статистичний розподіл має наступний вигляд:

$X = x_i$	x_1	x_2	x_3	$\dots$	x_n
$Y = y_i$	y_1	y_2	y_3	$\dots$	y_n

У разі зміни умовних статистичних розподілів змінюватимуться й умовні числові характеристики.

Звідси випливає визначення кореляційної залежності між ознаками X і Y.

Означення 4.2. *Кореляційною залежністю ознаки* Y *від* X, *або регресією* Y *від* X, називається функціональна залежність умовного середнього $\bar{y}_{x_i}$ від аргументу x:

$$\bar{y}_x = \varphi(x). \tag{4.1}$$

Означення 4.3. *Кореляційною залежністю ознаки* X *від* Y, *або регресією* X *від* Y, називається функціональна залежність умовного середнього $\bar{x}_{y_i}$ від аргументу y:

$$\bar{x}_y = \psi(x). \tag{4.2}$$

Графіки функцій $\varphi(x)$ та $\psi(x)$ називають *кривими регресії*.

Означення 4.4. *Умовним середнім* $\bar{y}_x$ $(\bar{x}_y)$ називають середнє арифметичне спостережуваних значень Y (X), що відповідають $X = x$ $(Y = y)$.

Оскільки умовні математичні сподівання $M(Y|X = x)$ і $M(X|Y = y)$ є відповідно функціями від x і y, то їх оцінки, тобто умовні середні $\bar{y}_x$ і $\bar{x}_y$ також є функціями відповідно від x і y, тобто

$$\bar{y}_x = \tilde{\varphi}(x, b_0, b_1, \dots, b_k\ ), \tag{4.3}$$

$$\bar{x}_y = \tilde{\psi}(y, c_0, c_1, \dots, c_k\ ), \tag{4.4}$$

де $b_0, b_1, \dots, b_k$ і $c_0, c_1, \dots, c_k$ – параметри.

Рівняння (4.3) і (4.4) називаються *вибірковими рівняннями регресії* відповідно Y по X і X по Y, функції $\bar{y}_x = \tilde{\varphi}(x, b_0, b_1, \dots, b_k\ )$ і $\bar{x}_y = \tilde{\psi}(y, c_0, c_1, \dots, c_k\ )$ – *вибірковими функціями регресії,* а їхні графіки – *вибірковими лініями регресії.*

Зв'язок або кореляція двох змінних називається *парною.* Якщо в рівняннях (4.3) і (4.4) зі збільшенням x і y змінні $\bar{y}_x$ і $\bar{x}_y$ у середньому зростають (мають тенденцію до зменшення), то така парна кореляція буде *додатною (від'ємною)*. Нульова кореляція спостерігається за відсутності зв'язку між X і Y.

Діаграма, на якій зображується сукупність значень двох ознак, називається *кореляційним полем (діаграмою розсіювання)*. Кожна точка цієї діаграми має координати (x_i, y_i), що відповідають розмірам ознак в i-му спостереженні. Три варіанти розподілу точок на кореляційному полі наведено на рис. 4.1.

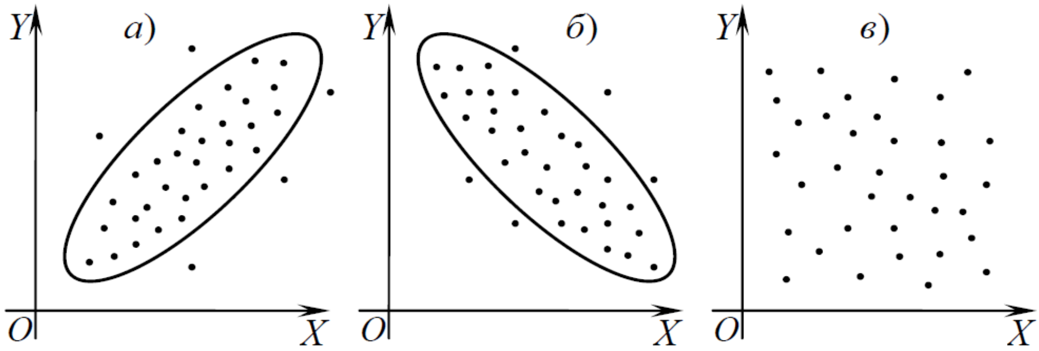


Рис. 4.1.

На першому з них основна маса точок укладається в еліпс, більша вісь якого утворює додатний кут з віссю OX (додатна кореляція). Другий варіант відповідає від'ємній кореляції. Рівномірний розподіл точок у просторі OXY свідчить про відсутність кореляційної залежності (рис. 4.1, в).

Статистичні зв'язки між змінними можна вивчати методами кореляційного і регресійного аналізу.

Основними задачами *кореляційного аналізу* є:

– вивчення сили зв'язку між двома і більше ознаками досліджуваного об'єкта;

– встановлення факторів, що найбільш суттєво впливають на результативну ознаку;

– виявлення невідомих причинно-наслідкових зв'язків між ознаками об'єкта.

Математико-статистичний апарат, що дозволяє встановити вид кореляційної залежності називається *регресійним аналізом*, а функція, яка описує цю залежність, називається *рівнянням регресії*.

Основною задачею *регресійного аналізу* є встановлення форми залежності за експериментальними даними, визначення функції регресії (процес *вирівнювання*) і вивчення залежності між змінними.

4.2. Лінійна парна регресія

Регресійний аналіз проводиться за такими етапами:

1) Встановлення виду кореляційної залежності результативної ознаки *Y* від факторної ознаки *X*.

2) Побудова регресійної моделі.

3) Перевірка статистичної значущості побудованої моделі.

Перший етап регресійного аналізу є найважливішим, оскільки помилки у виборі виду залежності призводять до побудови регресійної моделі, що не відповідає емпіричним даним і не може використовуватися для прогнозування.

Нехай у результаті незалежних спостережень над досліджуваною системою кількісних ознак $(X;Y)$, отримано n пар чисел $(x_1;y_1)$, $(x_2;y_2)$, … $(x_n;y_n)$. Згруповані дані зображуються графічно, що часто дозволяє визначити вид залежності Y від X

За даними спостережень знайдемо вибіркове рівняння прямої лінії регресії, для визначеності Y по X

$$y = \alpha + \beta x. \tag{4.5}$$

У рівняні (4.5) $\bar{y}_x$ замінене на y, тому що різні значення x ознаки X і відповідні їм значення y ознаки Y спостерігалися по одному разу і тому групувати дані немає необхідності.

Якщо позначити через $\tilde{y}_i = \alpha + \beta x_i$ наближене значення y_i, обчислене з рівняння регресії (4.5), то величина $y_i - \tilde{y}_i$ є відхиленням наближеного значення $\tilde{y}_i$ від точного значення y_i (див. рис. 4.2).

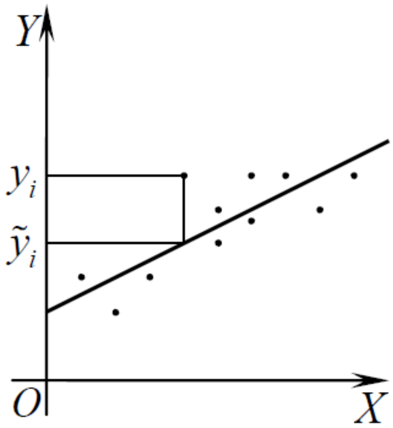


Рис. 4.2

Побудова лінійної регресійної моделі – це знаходження параметрів рівняння (4.5). Найбільш часте оцінювання параметрів α і β прямої регресії здійснюють на основі *методу найменших квадратів* (МНК). Цей метод набув широкого застосування в економіко-статистичних розрахунках після створення теорії регресії. Згідно з МНК параметри α і β прямої регресії вибирають так, щоб сума квадратів відхилень $y_i - \tilde{y}_i$ була мінімальною, тобто з умови мінімізації функції:

$$S(\alpha;\beta) = \sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 = \sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2.$$

Підберемо α і β з умови мінімуму функції $S(\alpha,\beta)$. Для цього прирівняємо до нуля частинні похідні цієї функції. Після цього одержимо систему рівнянь для визначення α і β:

$$\begin{cases} \dfrac{\partial S}{\partial \alpha} = 2\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)\,(-1) = 0; \\ \dfrac{\partial S}{\partial \beta} = 2\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)(-x_i) = 0; \end{cases} \Rightarrow \begin{cases} \beta\sum_{i=1}^{n} x_i^2 + \alpha\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i\,; \\ \beta\sum_{i=1}^{n} x_i + \alpha n = \sum_{i=1}^{n} y_i. \end{cases}$$

(4.6)

Тобто отримали систему двох лінійних алгебраїчних рівнянь відносно α і β. Систему (4.6) називають *системою методу найменших квадратів*. Розв'язавши цю систему, знайдемо шукані параметри:

$$\alpha = \frac{(\sum_{i=1}^{n} x_i^2)(\sum_{i=1}^{n} y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} x_i y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2};$$
$$\beta = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}. \qquad (4.7)$$

Якщо потрібно за результатами спостережень одержати лінійне рівняння регресії X по Y, то в рівнянні регресії $y = \alpha + \beta x$ треба поміняти місцями змінні x і y. При цьому одержимо рівняння $x = \alpha' + \beta' y$, де параметри α' і β' обчислюються за формулами:

$$\alpha' = \frac{(\sum_{i=1}^{n} y_i^2)(\sum_{i=1}^{n} x_i) - (\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i y_i)}{n(\sum_{i=1}^{n} y_i^2) - (\sum_{i=1}^{n} y_i)^2};$$
$$\beta' = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i)}{n(\sum_{i=1}^{n} y_i^2) - (\sum_{i=1}^{n} y_i)^2}. \qquad (4.8)$$

Зауважимо, що регресійні прямі $y = \alpha + \beta x$ і $x = \alpha' + \beta' y$ – різні. Перша пряма виходить у результаті розв’язку задачі про мінімізацію суми квадратів відхилень по вертикалі, а друга – при розв’язку задачі про мінімізацію суми квадратів відхилень по горизонталі.

На практиці для знаходження рівнянь регресії складають наступну розрахункову таблицю (табл. 4.1):

Таблиця 4.1

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	x_1	y_1	x_1^2	y_1^2	$x_1 y_1$
2	x_2	y_2	x_2^2	y_2^2	$x_2 y_2$
…	…	…	…	…	…
n	x_n	y_n	x_n^2	y_n^2	$x_n y_n$
$\sum$	$\sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n} y_i$	$\sum_{i=1}^{n} x_i^2$	$\sum_{i=1}^{n} y_i^2$	$\sum_{i=1}^{n} x_i y_i$

В останньому рядку цієї таблиці суми й визначають коефіцієнти α, β або α', β' у формулах (4.7) або (4.8) відповідно.

Якщо число вимірів велике, то з метою спрощення розрахунків експериментальні дані потрібно групувати, тобто поєднувати в таблицю 4.2, яка називається *кореляційною*.

Розглянемо, як заповнюється кореляційна таблиця.

У першому стовпці (першому рядку) перераховуються у вибірці значення величини X: x_i, $i = \overline{1, s}$ (Y: y_j, $j = \overline{1, k}$).

Таблиця 4.2

$X \backslash Y$	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}	$n_{x_2} = \sum_{j=1}^{k} n_{2j}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_s	n_{s1}	n_{s2}	…	n_{sk}	$n_{x_s} = \sum_{j=1}^{k} n_{sj}$
n_{y_j}	$n_{y_1} = \sum_{i=1}^{s} n_{i1}$	$n_{y_2} = \sum_{i=1}^{s} n_{i2}$	…	$n_{y_k} = \sum_{i=1}^{s} n_{ik}$	$n = \sum_{i=1}^{s} \sum_{j=1}^{k} n_{ij}$

Якщо кількість різних значень величин X і Y велике або ці величини розподілені неперервно, то здійснюється угруповання їхніх значень по інтервалах. У цьому випадку x_i і y_i являють собою середини відповідних інтервалів.

На перетині рядків і стовпців указуються частоти n_{ij}, які дорівнюють числу появ у вибірці пари $(x_i;\ y_j)$. Якщо пари значень ознак $(x_i;\ y_j)$ не спостерігалися, то у відповідних клітинках таблиці ставиться риска.

В останньому рядку (останньому стовпці) указуються числа n_{y_j}, $j = \overline{1,k}$ (n_{x_i}, $i = \overline{1,s}$), рівні кількості появ у вибірці значень y_j (x_i) незалежно від того, у парі з яким зі значень величин $X(Y)$ воно з'явилося.

Кореляційна таблиця містить всю інформацію, отриману в результаті вибіркових спостережень величин X і Y. Звідси з урахуванням частот появ змінних x_i і y_j, одержимо:

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{s} n_{x_i} x_i;\ \sum_{i=1}^{n} y_i = \sum_{j=1}^{k} n_{y_i} y_j;\ \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{s} n_{x_i} x_i^2;$$
$$\sum_{i=1}^{n} y_i^2 = \sum_{j=1}^{k} n_{y_j} y_j^2;\ \sum_{i=1}^{n} x_i\, y_i = \sum_{i=1}^{s} \sum_{j=1}^{k} n_{ij} x_i y_j.$$

Підставивши ці суми до формули (4.7) і (4.8), одержимо:

$$\alpha = \frac{\left(\sum_{i=1}^{s} n_{x_i} x_i^2\right)\left(\sum_{j=1}^{k} n_{y_i} y_j\right) - \left(\sum_{i=1}^{s} n_{x_i} x_i\right)\left(\sum_{i=1}^{s} \sum_{j=1}^{k} n_{ij} x_i y_j\right)}{n\left(\sum_{i=1}^{s} n_{x_i} x_i^2\right) - \left(\sum_{i=1}^{s} n_{x_i} x_i\right)^2};$$
$$\beta = \frac{n \sum_{i=1}^{s} \sum_{j=1}^{k} n_{ij} x_i y_j - \left(\sum_{i=1}^{s} n_{x_i} x_i\right)\left(\sum_{j=1}^{k} n_{y_i} y_j\right)}{n\left(\sum_{i=1}^{s} n_{x_i} x_i^2\right) - \left(\sum_{i=1}^{s} n_{x_i} x_i\right)^2}; \tag{4.9}$$

та

$$\alpha' = \frac{\left(\sum_{j=1}^{k} n_{y_j} y_j^2\right)\left(\sum_{i=1}^{n} x_i\right) - \left(\sum_{j=1}^{k} n_{y_i} y_j\right)\left(\sum_{i=1}^{s} \sum_{j=1}^{k} n_{ij} x_i y_j\right)}{n\left(\sum_{j=1}^{k} n_{y_j} y_j^2\right) - \left(\sum_{j=1}^{k} n_{y_i} y_j\right)^2};$$
$$\beta' = \frac{n \sum_{i=1}^{s} \sum_{j=1}^{k} n_{ij} x_i y_j - \left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n\left(\sum_{j=1}^{k} n_{y_j} y_j^2\right) - \left(\sum_{j=1}^{k} n_{y_i} y_j\right)^2}. \tag{4.10}$$

Як було зазначено вище, система рівнянь для визначення параметрів α і β рівняння прямої лінії регресії Y на X має вигляд (4.6). Передбачалося, що значення X і відповідні їм значення Y спостерігалися один раз. Тепер припустимо, що отримано велику кількість даних, серед яких є повторювані, і вони згруповані у вигляді кореляційної таблиці. Запишемо систему (4.6) так, щоб вона відображала дані кореляційної таблиці. Скористаємося тотожностями:

$$\sum_{i=1}^{n} x_i = n\bar{x};\ \sum_{i=1}^{n} y_i = n\bar{y};\ \sum_{i=1}^{n} x_i^2 = n\overline{x^2};\ \sum_{i=1}^{n} x_i\, y_i = n\overline{xy}, \tag{4.11}$$

де $\bar{x}$, $\bar{y}$ – вибіркові середні; $\overline{x^2}$ – вибіркове середнє квадрата; $\overline{xy}$ – вибіркове середнє добутку.

Нагадаємо, що *вибіркова коваріація* s_{xy} визначається рівністю

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y}) = \frac{1}{n}\sum_{i=1}^{n} x_i\, y_i - \bar{x}\cdot\bar{y} = \overline{xy} - \bar{x}\cdot\bar{y}. \tag{4.12}$$

Вибіркові дисперсії визначаються співвідношеннями:

$$\begin{aligned} s_x^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2 = \overline{x^2}-\bar{x}^2, \\ s_y^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n}y_i^2-\bar{y}^2 = \overline{y^2}-\bar{y}^2. \end{aligned} \qquad (4.13)$$

За визначенням *вибірковий коефіцієнт кореляції*

$$\rho_{xy} = \frac{s_{xy}}{s_x \cdot s_y}. \qquad (4.14)$$

Підставивши праві частини рівностей (4.11) у систему (4.6), одержимо

$$\begin{cases} \beta n\overline{x^2}+\alpha n\bar{x} = n\overline{xy}, \\ \beta n\bar{x}+\alpha n = n\bar{y}, \end{cases} \Leftrightarrow \begin{cases} \overline{x^2}\beta+\bar{x}\alpha = \overline{xy}, \\ \bar{x}\,\beta+\alpha = \bar{y}. \end{cases} \qquad (4.15)$$

Розв’язуючи систему (4.15) за формулами Крамера, одержимо:

$$\Delta = \begin{vmatrix} \overline{x^2} & \bar{x} \\ \bar{x} & 1 \end{vmatrix} = \overline{x^2}-\bar{x}^2 = s_x^2;$$

$$\Delta_\beta = \begin{vmatrix} \overline{xy} & \bar{x} \\ \bar{y} & 1 \end{vmatrix} = \overline{xy}-\bar{x}\cdot\bar{y} = s_{xy};\ \ \Delta_\alpha = \begin{vmatrix} \overline{x^2} & \overline{xy} \\ \bar{x} & \bar{y} \end{vmatrix} = \bar{y}\cdot s_x^2-\bar{x}\cdot s_{xy};$$

$$\alpha = \frac{\Delta_\alpha}{\Delta} = \frac{\bar{y}\cdot s_x^2-\bar{x}\cdot s_{xy}}{s_x^2} = \bar{y}-\frac{s_{xy}}{s_x^2}\cdot\bar{x};\ \ \beta = \frac{\Delta_\beta}{\Delta} = \frac{s_{xy}}{s_x^2}.$$

Підставивши коефіцієнти α і β до рівняння регресії $\bar{y} = \alpha + \beta x$, одержимо

$$\bar{y}_x = \bar{y}-\frac{s_{xy}}{s_x^2}\cdot\bar{x}+\frac{s_{xy}}{s_x^2}\cdot x \ \Leftrightarrow\ \bar{y}_x-\bar{y} = \frac{s_{xy}}{s_x^2}\cdot(x-\bar{x}),$$

$$\frac{s_{xy}}{s_x^2} = \frac{s_{xy}}{s_x s_y}\cdot\frac{s_y}{s_x} = |4.14| = \rho_{xy}\cdot\frac{s_y}{s_x}.$$

Таким чином, *вибіркове рівняння прямої регресії Y* по X має вигляд

$$\bar{y}_x-\bar{y} = \rho_{xy}\cdot\frac{s_y}{s_x}\cdot(x-\bar{x}). \qquad (4.16)$$

Аналогічно знайдемо, що рівняння прямої регресії X по Y запишеться у вигляді

$$\bar{x}_y-\bar{x} = \rho_{xy}\cdot\frac{s_x}{s_y}\cdot(y-\bar{y}). \qquad (4.17)$$

З рівнянь (4.16) і (4.17) випливає, що прямі регресії Y по X і X по Y проходять через точку $(\bar{x},\bar{y})$. Ці прямі збігаються, коли $|\rho_{xy}|^2 = 1$.

Величини $\rho_{xy}\cdot\frac{s_y}{s_x}$ і $\rho_{xy}\cdot\frac{s_x}{s_y}$ називаються *вибірковими коефіцієнтами лінійної регресії* та позначаються

$$\rho_{y|x} = \rho_{xy} \cdot \frac{s_y}{s_x}\ ;\ \rho_{x|y} = \rho_{xy} \cdot \frac{s_x}{s_y}. \qquad (4.18)$$

Перемножуючи праві й ліві частини рівностей (4.18), після добування кореня одержимо $\rho_{xy} = \pm\sqrt{\rho_{y|x} \cdot \rho_{x|y}}$, тобто коефіцієнт кореляції є середнє геометричне коефіцієнтів лінійної регресії. Знак у правій частині ρ_{xy} збігається зі знаками $\rho_{y|x}$ і· $\rho_{x|y}$.

Коефіцієнт регресії Y по X (X по Y) показує, на скільки одиниць у середньому змінюється змінна Y (X) при збільшенні змінної X (Y) на одну одиницю.

Якщо дані спостережень над ознаками X і Y задані у вигляді кореляційної таблиці з рівновіддаленими варіантами, то доцільно перейти до умовних варіант:

$$u_i = \frac{x_i - C_1}{h_1};\quad v_j = \frac{y_j - C_2}{h_2},$$

де C_1 і C_2 – хибні нулі варіант X і Y відповідно, h_1 і h_2 – кроки, тобто різниці між двома сусідніми варіантами X і Y.

Тоді відповідно до (1.7) і (1.15)

$$\bar{x} = h_1\bar{u} + C_1\ , \qquad (4.19)$$

$$\bar{y} = h_2\bar{v} + C_2\ , \qquad (4.20)$$

$$s_x^2 = h_1^2\overline{u^2} - (\bar{x} - C_1)^2, \qquad (4.21)$$

$$s_y^2 = h_2^2\overline{v^2} - (\bar{y} - C_2)^2, \qquad (4.22)$$

$$s_{xy} = h_1h_2(\overline{uv} - \bar{u} \cdot \bar{v}), \qquad (4.23)$$

де $\bar{u} = \frac{1}{n}\sum_{i=1}^{s} u_i n_{x_i}$; $\bar{v} = \frac{1}{n}\sum_{j=1}^{k} v_j n_{y_j}$; $\overline{uv} = \frac{1}{n}\sum_{i=1}^{s}\sum_{j=1}^{k} u_i\, v_j n_{ij}$.

4.3. Коефіцієнт кореляції

Для кількісної характеристики тісноти лінійного кореляційного зв'язку між двома величинами X і Y генеральної сукупності було введене поняття *коефіцієнта лінійної кореляції,* який визначався зі співвідношенням

$$r_{xy} = \frac{M(XY) - m_x m_y}{\sigma_x\sigma_y}, \qquad (4.24)$$

де $m_x = M(X), m_y = M(Y), \sigma_x = \sqrt{D(X)},\ \sigma_y = \sqrt{D(Y)}$.

Відомо, що якщо величини X і Y незалежні, то $r_{xy} = 0$; якщо $r_{xy} = \pm 1$, то X і Y зв'язані лінійною функціональною залежністю, причому, дорівнює $r_{xy} = 1$ у випадку зростаючої залежності та $r_{xy} = -1$ у випадку спадної; $|r_{xy}| \leq 1$. При цьому $r_{xy} > 0$, якщо при зростанні однієї величини (наприклад, X) $M(Y)$ збільшується, і $r_{xy} < 0$ у протилежному випадку.

На практиці для оцінювання тісноти (або сили) лінійного кореляційного зв'язку між величинами X і Y за результатами вибіркових спостережень використовується *вибірковий коефіцієнт лінійної кореляції* (*коефіцієнт кореляції Пірсона*), який визначається за формулою:

$$\rho_{xy} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{s_x \cdot s_y}, \qquad (4.25)$$

де $\bar{x} = \frac{1}{n}\sum_{i=1}^{s} n_{x_i} x_i$; $\bar{y} = \frac{1}{n}\sum_{j=1}^{k} n_{y_j} y_j$; $\overline{xy} = \frac{1}{n}\sum_{i=1}^{s}\sum_{j=1}^{k} x_i y_j n_{ij}$, а $s_x = \sqrt{\frac{1}{n}\sum_{i=1}^{s}(x_i - \bar{x})^2 n_{x_i}}$ і $s_y = \sqrt{\frac{1}{n}\sum_{j=1}^{k}\left(y_j - \bar{y}\right)^2 n_{y_j}}$.

Формулу (4.25) зазвичай застосовують у дещо перетвореному вигляді. Так, підставивши до неї розгорнуті вирази для s_x, s_y і s_{xy}, легко одержати наступну формулу:

$$\rho_{xy} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{n\sum_{i=1}^{s}\sum_{j=1}^{k} x_i y_j n_{ij} - \left(\sum_{i=1}^{s} n_{x_i} x_i\right)\cdot\left(\sum_{j=1}^{k} n_{y_j} y_j\right)}{\sqrt{n\sum_{i=1}^{s} n_{x_i} x_i^2 - \left(\sum_{i=1}^{s} n_{x_i} x_i\right)^2}\cdot\sqrt{n\sum_{j=1}^{k} n_{y_j} y_j^2 - \left(\sum_{j=1}^{k} n_{y_i} y_j\right)^2}}. \qquad (4.26)$$

У випадку незгрупованих даних розрахункова формула суттєво спрощується:

$$\rho_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\cdot\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\cdot\sqrt{n\sum_{i=1}^{n} y_j^2 - \left(\sum_{i=1}^{n} y_j\right)^2}}.$$

Коефіцієнт кореляції ρ_{xy} є безрозмірною величиною (оскільки розмірності чисельника і знаменника є розмірністю добутку XY); його величина не залежить від вибору одиниць виміру обох змінних; величина ρ_{xy} набуває значення на відрізку $[-1;\ 1]$. Наближена до нуля величина коефіцієнта кореляції говорить про відсутність лінійного зв'язку змінних, але не заперечує можливість існування іншої форми залежності між ними.

Властивості вибіркового коефіцієнта кореляції

1) Коефіцієнт кореляції Пірсона приймає значення на проміжку $[-1;1]$, тобто $-1 \le \rho_{xy} \le 1$ або $|\rho_{xy}| \le 1$.

2) Якщо $|\rho_{xy}| = 0$, то зв'язок відсутній; якщо $0{,}1 \le |\rho_{xy}| \le 0{,}3$, то зв'язок вважається слабким; якщо $0{,}3 \le |\rho_{xy}| \le 0{,}5$, то зв'язок вважається помірним; якщо $0{,}5 \le |\rho_{xy}| \le 0{,}7$, то зв'язок вважається середнім (помітним); якщо $0{,}7 \le |\rho_{xy}| \le 0{,}9$, то зв'язок вважається високим; якщо $0{,}9 \le |\rho_{xy}| \le 1$, то зв'язок вважається сильним (дуже високим); якщо $|\rho_{xy}| = 1$, то характер зв'язку – функціональний.

3) Якщо величини X і Y незалежні, то $\rho_{xy} = 0$.

4) Якщо $\rho_{xy} = \pm 1$, то X і Y зв'язані лінійною функціональною залежністю, причому, дорівнює $\rho_{xy} = 1$ у випадку зростаючої залежності та $\rho_{xy} = -1$ у випадку спадної.

5) Якщо $\rho_{xy} > 0$, то зв'язок називається додатнім, тобто зі збільшенням значень X значення Y також збільшуються. Якщо $\rho_{xy} < 0$, то зв'язок називається від'ємним, тобто зі збільшенням значень X значення Y зменшуються.

Зауваження 4.1. Слід пам'ятати, що коефіцієнт кореляції Пірсона показує силу лінійного зв'язку. Якщо між X і Y існує сильний нелінійний зв'язок, коефіцієнт кореляції може дорівнювати нулю.

Хоча вибірковий коефіцієнт кореляції ρ_{xy} являє собою обґрунтовану оцінку для r_{xy}, однак більш надійну оцінку наближеності ρ_{xy} до r_{xy} за даними вибірки можна дати лише в тому випадку, коли розподіл величин X і Y достатньо наближений до нормальної форми. Наприклад, для оцінки r_{xy} нормально розподіленої генеральної сукупності, у випадку великих вибірок ($n \ge 50$), можна скористатися формулою

$$\rho_{xy} - u_\gamma \frac{1-\rho_{xy}^2}{\sqrt{n}} \le r_{xy} \le \rho_{xy} + u_\gamma \frac{1-\rho_{xy}^2}{\sqrt{n}}, \qquad (4.27)$$

де u_γ корінь рівняння $2\Phi(\varepsilon) = \gamma$, що визначається з таблиці значень функції $\Phi(x)$ (додаток 2) за заданою довірчою ймовірністю γ, а величина

$$\delta = u_\gamma \frac{1-\rho_{xy}^2}{\sqrt{n}}, \qquad (4.28)$$

називається точністю оцінки.

Наведемо схему находження точності δ і довірчих границь, які відповідають надійності γ:

$$\left.\begin{matrix} n \\ \rho_{xy} \\ \gamma \end{matrix}\right\} \to 2\Phi(\varepsilon) = \gamma \overset{\text{Д.4}}{\to} \varepsilon = u_\gamma \to \delta = u_\gamma \frac{1-\rho_{xy}^2}{\sqrt{n}} \to \begin{matrix} \Rsh\ \rho_{xy} - \delta, \\ \Lsh\ \rho_{xy} + \delta. \end{matrix} \qquad (4.29)$$

Застосовуючи коефіцієнт кореляції як міру зв'язку, потрібно мати на увазі, що він отриманий на основі даних вибірки і, отже, схильний до впливу випадковості. Якщо об'єм вибірки невеликий, то знайти вибіркову помилку цієї величини досить складно, тому на практиці зазвичай замість визначення помилки коефіцієнта кореляції перевіряють *гіпотезу про його значущість*, тобто, чи істотно ρ_{xy} відрізняється від нуля або цю відмінність можна приписати впливу випадковості, пов'язаної з вибіркою. Інакше кажучи, виникає необхідність при заданому рівні значущості α перевірити нульову гіпотезу $H_0\colon r_{xy} = 0$ при конкуруючій гіпотезі $H_1\colon r_{xy} \neq 0$.

Якщо нульова гіпотеза відкидається, то це означає, що ρ_{xy} істотно відрізняється від нуля, а X і Y корельовані, тобто зв'язані лінійною залежністю. Якщо нульова гіпотеза буде прийнята, то ρ_{xy} неістотно відрізняється від нуля, а X і Y некорельовані, тобто не зв'язані лінійною залежністю.

Як критерій перевірки нульової гіпотези візьмемо випадкову величину

$$T = \frac{\rho_{xy}\sqrt{n-2}}{\sqrt{1-\rho_{xy}^2}}, \qquad (4.30)$$

яка має розподіл Стьюдента із $k = n - 2$ ступенями вільності. Число ступенів вільності менше за число спостережень на 2, оскільки до формули вибіркового коефіцієнта кореляції входять середні вибіркові значення X і Y, для розрахунку яких використовуються дві лінійні формули їхньої залежності від спостережень випадкових величин. Якщо альтернативна гіпотеза має вигляд $H_1: r_{xy} \neq 0$, то критична область є двобічною. Перевірку нульової гіпотези будемо здійснювати за наступною схемою:

$$\left.\begin{matrix} n; k = n-2 \\ \rho_{xy} \\ \alpha \end{matrix}\right\} \to H_0: r_{xy} = 0 \to T_{\text{сп}} = \frac{\rho_{xy}\sqrt{n-2}}{\sqrt{1-\rho_{xy}^2}} \to$$

$$\to\ H_1: r_{xy} \neq 0\ \to\ t_{\text{кр}}(\alpha; k) \xrightarrow{\text{Д 7}} t_{\text{кр}}^{\text{дв}} \to \begin{cases} |T_{\text{сп}}| < t_{\text{кр}}^{\text{дв}} \to \boldsymbol{P}, \\ |T_{\text{сп}}| > t_{\text{кр}}^{\text{дв}} \to \boldsymbol{V}. \end{cases} \tag{4.31}$$

4.4. Перевірка гіпотези про незалежність системи двох випадкових величин

Гіпотезу про незалежність двох випадкових величин X та Y можна перевірити за критерієм χ^2. У цьому випадку величина χ^2 визначається формулою

$$\chi^2 = \sum_{i=1}^{l}\sum_{j=1}^{m}\frac{\left(n_{ij}-m_{ij}\right)^2}{m_{ij}}, \tag{4.32}$$

де n_{ij} – кількість випадків, коли одночасно спостерігались значення

$$X = x_i, Y = y_j, m_{ij} = \frac{n_{i0}n_{0j}}{n},$$

де n_{i0}, n_{0j} – загальна кількість випадків, коли спостерігались відповідно значення $X = x_i, Y = y_j$, l та m – кількість значень, що їх набувають відповідно величини X та Y за умови, що обсяг вибірки дорівнює n.

Гіпотеза H_0 про незалежність випадкових величин X та Y приймається на рівні значущості α, якщо $\chi^2 \leq \chi_\alpha^2$, де значення χ_α^2 береться з таблиць розподілу χ^2 з $k = (n-1)(m-1)$ ступенями свободи (додаток 3). Інакше кажучи, гіпотеза H_0 відхиляється на рівні значущості α, якщо $\chi^2 \leq \chi_{1-\alpha}^2$.

Для обчислення вибіркового значення статистики (4.32) критерію зручно використовувати формулу

$$\chi^2 = n\left(\sum_{i=1}^{l}\frac{1}{n_{i0}}\left(\sum_{j=1}^{m}\frac{\left(n_{ij}\right)^2}{n_{0j}}\right) - 1\right). \tag{4.32}$$

Результати перевірки гіпотези можна подати у вигляді таблиці спряженості ознак $l \times m$ (табл. 4.3), яка представляє собою сукупний результат послідовності повторень випадкового експерименту (при цьому результати класифікуються за двома змінними ознаками).

Нехай є k різних експериментів, що складаються з $n_1, n_2, \dots, n_k$ одиничних спостережень, тобто k різних вибірок обсягу $n_1, n_2, \dots, n_k$.

Потрібно перевірити гіпотезу про те, що k вибірок вибрано з однієї і тієї самої сукупності або, інакше кажучи, гіпотезу про те, що ці вибірки однорідні.

У кожному екперименті спостерігається деяка змінна ознака і результати кожного зі спостережень розбиваються за значеннями цієї ознаки на l груп. Кількість результатів спостережень в i-й групі j-го ряду позначимо n_{ij}. Тоді всі дані розміщуються в таблиці такого самого вигляду, як і таблиця 4.3, причому суми за стовпцями в ній дорівнюють n_j.

Проте в цьому випадку таблиця є результатом спостережень не однієї послідовності, як у випадку табл. 4.3, а k незалежних спостережень, кожному з яких відповідає один стовпець. Для перевірки гіпотези про однорідність використовують той самий критерій, що й для перевірки незалежності двох ознак.

Таблиця 4.3

X \ Y	y_1	y_2	…	y_m	$\sum_{i=1}^{m} n_{ij} = n_{i0}$
x_1	n_{11}	n_{12}	…	n_{1m}	n_{10}
x_2	n_{21}	n_{22}	…	n_{2m}	n_{20}
⋮	⋮	⋮	⋮	⋮	⋮
x_l	n_{l1}	n_{l2}	…	n_{lm}	n_{l0}
$\sum_{i=1}^{l} n_{ij} = n_{0j}$	n_{01}	n_{02}	…	n_{0m}	$n = \sum_{i=1}^{l}\sum_{j=1}^{m} n_{ij}$

4.5. Поняття про множинну кореляцію

У тих випадках, коли досліджується кореляційний зв'язок між величинами, число яких більше за два, уводять поняття *множинної кореляції*. Так, при дослідженні кореляційного зв'язку між трьома кількісними ознаками X, Y і Z можна ввести рівняння регресії

$$\bar{z}_{xy} = f(x; y),$$

де $\bar{z}_{xy}$ – середнє значення величини Z, що відповідає відповідним значенням x і y. Геометричною інтерпретацією цього рівняння є деяка поверхня в прямокутній системі координат тривимірного простору. У найбільш простому випадку лінійної кореляційної залежності ознак X, Y і Z вибіркове рівняння регресії має вигляд

$$\bar{z}_{xy} = ax + by + c.$$

Нехай у результаті незалежних спостережень над досліджуваною системою кількісних ознак (X; $Y; Z$), отримано n сукупностей чисел $(x_1;\ y_1;\ z_1)$, $(x_2;\ y_2;\ z_2), \dots, (x_n;\ y_n;\ z_n)$.

Припустимо, що функція регресії лінійна, тобто

$$z = ax + by + c. \qquad (4.33)$$

У (4.33) $\bar{z}_{xy}$ замінене на z, тому що різні значення x і y ознак X і Y, і відповідні їм значення z ознаки Z спостерігалися один раз.

Якщо позначити $\tilde{z}_i = ax_i + by_i + c$ наближене значення z_i, обчислене з рівняння регресії (4.33), то величина $z_i - \tilde{z}_i$ є відхиленням наближеного значення $\tilde{z}_i$ від точного z_i. Коефіцієнти a, b і c рівняння регресії (4.33) знайдемо з умови МНК:

$$S(a; b; c) = \sum_{i=1}^{n} (z_i - \tilde{z}_i)^2 = \sum_{i=1}^{n} (z_i - ax_i - by_i - c)^2 \longrightarrow min.$$

Необхідні умови мінімуму функції S утворять систему

$$\begin{cases} \dfrac{\partial S}{\partial a} = 2\sum_{i=1}^{n}(z_i - ax_i - by_i - c)(-x_i) = 0, \\ \dfrac{\partial S}{\partial b} = 2\sum_{i=1}^{n}(z_i - ax_i - by_i - c)(-y_i) = 0, \\ \dfrac{\partial S}{\partial c} = 2\sum_{i=1}^{n}(z_i - ax_i - by_i - c)(-1) = 0, \end{cases}$$

яка в результаті тотожних перетворень набуде

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i y_i + c\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i z_i\,, \\ a\sum_{i=1}^{n} x_i y_i + b\sum_{i=1}^{n} y_i^2 + c\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} y_i z_i\,, \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i + c\,n = \sum_{i=1}^{n} z_i\,. \end{cases} \quad (4.34)$$

Розв’язуючи систему (4.34) відносно невідомих параметрів а, b і c, а потім підставляючи їхні значення до (4.33), одержимо шукане рівняння регресії.

Зручніше, однак, знайти готові формули для розрахунку параметрів, а не розв’язувати систему щоразу. Коефіцієнти регресії а, b і параметр с у цьому випадку обчислюються за наступними формулами:

$$a = \frac{\rho_{xz} - \rho_{yz}\rho_{xy}}{1 - \rho_{xy}^2} \cdot \frac{s_z}{s_x};\; b = \frac{\rho_{yz} - \rho_{xz}\rho_{xy}}{1 - \rho_{xy}^2} \cdot \frac{s_z}{s_y};\; c = \bar{z} - a\bar{x} - b\bar{y}. \quad (4.35)$$

Тут ρ_{xy}, ρ_{xz}, ρ_{yz} – вибіркові коефіцієнти лінійної кореляції між відповідними ознаками; s_x, s_y, s_z – середні квадратичні відхилення; $\bar{x}, \bar{y}, \bar{z}$ – середні значення.

Тіснота зв'язку ознаки Z з ознаками X і Y, оцінюється *вибірковим сукупним коефіцієнтом кореляції*

$$R = \sqrt{\frac{\rho_{xz}^2 - 2\rho_{xy}\rho_{xz}\rho_{yz} + \rho_{yz}^2}{1 - \rho_{xy}^2}}, \quad (4.36)$$

причому $0 \le R \le 1$.

Тіснота зв'язку між Z і X (при постійному Y), між Z і Y (при постійному X) оцінюється відповідно *частковими вибірковими коефіцієнтами кореляції*:

$$\rho_{xz(y)} = \frac{\rho_{xz}-\rho_{xy}\rho_{yz}}{\sqrt{\left(1-\rho_{xy}^2\right)\left(1-\rho_{yz}^2\right)}};\ \rho_{yz(x)} = \frac{\rho_{yz}-\rho_{xy}\rho_{xz}}{\left(1-\rho_{xy}^2\right)\left(1-\rho_{xz}^2\right)}. \quad (4.37)$$

Ці коефіцієнти показують ступінь лінійної залежності між спостережуваними значеннями Z і X, а також Z і Y, якщо вплив 0третьої ознаки усунуто.

ВАРІАНТИ ІНДИВІДУАЛЬНИХ ЗАВДАНЬ (ІЗ № 4)

За даними спостережень X і У двох випадкових величин потрібно:

1) побудувати кореляційне поле (діаграму розсіювання), знайти вибірковий коефіцієнт кореляції та при рівні значущості $\alpha = 0.05$ перевірити його значущість, зробити висновок про залежність між ознаками X і У;

2) знайти рівняння прямої регресій У по X *(методом найменших квадратів* та за допомогою *вибіркового рівняння лінійної регресії Y на X),* побудувати лінії регресії.

1.

X	24	26	28	30	32	34	36	38
Y	30	32	34	36	38	40	42	44

2.

X	- 20, 2	- 20, 5	- 21, 4	- 21, 8	- 22	- 22, 5	- 22, 8	- 22, 8
Y	- 10, 2	- 11, 5	- 12, 4	- 12, 8	- 13	- 13, 5	- 14, 2	- 14, 6

3.

X	0, 689	0, 692	0, 695	0, 698	0, 69	0, 71	0, 720	0, 725
Y	0, 71	0, 725	0, 78	0, 79	0, 795	0, 8	0, 81	0, 85

4.

X	0, 531	0, 524	0, 541	0, 55	0, 559	0, 62	0, 632	0, 672
Y	0, 62	0, 58	0, 640	0, 65	0, 67	0, 68	0, 695	0, 7

5.

X	44	43	42	41	40	39	38	37
Y	10	25	68	136	152	162	170	180

6.

X	30	35	31	38	41	48	50	55
Y	45	25	48	52	54	55	59	60

7.

X	30	25	31	32	38	41	40	46
Y	480	510	530	540	555	564	570	575

8.

X	2, 07	2, 12	2, 11	2, 58	2, 89	2, 92	3, 01	3, 12
Y	2, 88	2, 91	2, 92	2, 96	3, 01	3, 11	3, 21	3, 3

9.

X	1, 8	2, 1	2, 8	3	3, 2	3, 8	3, 9	4, 2
Y	5, 4	5, 6	6, 2	6, 8	7, 1	7, 8	8, 5	9

10.

X	70	75	82	89	95	100	105	115
Y	10, 5	15, 8	17, 8	19, 5	20, 4	21, 5	22, 2	24, 3

11.

X	4	5	5, 5	6	6, 8	7, 5	8, 5	10, 8
Y	9, 35	9, 21	9, 18	9, 50	9, 1	9, 08	9, 05	9, 01

12.

X	5, 2	5, 8	5, 9	6, 2	6, 9	7, 2	7, 5	8, 5
Y	11	11, 6	12, 1	12, 5	13, 2	13, 9	14, 1	14, 6

13.

X	170	180	200	230	240	250	280	300
Y	240	200	190	180	170	160	150	140

14.

X	15	16	17	18	19	20	21	22
Y	26, 8	26, 5	26, 3	26, 1	25, 7	25, 3	24, 3	24, 6

15.

X	62, 1	61, 1	61, 0	60, 5	60, 0	59, 0	58, 5	58
Y	115	116	117	118	119	120	121	122

16.

X	2, 2	2, 35	2, 42	2, 58	2, 65	2, 69	2, 74	2, 88
Y	35, 4	35	35, 8	36, 2	36, 7	36, 9	37, 3	39

17.

X	- 23, 2	- 24, 1	- 24, 5	- 25, 1	- 25, 8	- 26	- 26, 5	- 27
Y	- 14, 7	- 15, 7	- 16, 4	- 17, 2	- 17, 5	- 18, 2	- 18, 6	- 18, 9

18.

X	56, 5	56	55, 5	55	54, 5	54	53, 5	53
Y	124	125	126	127	128	129	130	133

19.

X	1, 23	1, 33	1, 43	1, 53	1, 63	1, 73	1, 83	1, 93
Y	10, 36	11, 56	13, 29	14, 5	15, 6	14, 25	17, 36	16, 23

20.

X	83	92	112	132	144	154	162	189
Y	369	380	370	395	420	412	436	420

21.

X	0, 41	0, 48	0, 56	0, 66	0, 72	0, 79	0, 85	0, 86
Y	6, 02	6, 12	6, 22	6, 28	6, 3	6, 35	6, 39	6, 45

22.

X	50	50, 2	52, 8	53, 5	54	56, 8	58, 5	60
Y	10, 1	10, 3	10, 45	10, 9	11, 2	11, 35	11, 9	12, 45

23.

X	14, 5	15, 9	25	28, 5	30, 5	36, 8	40	45, 8
Y	8, 98	8, 94	8, 9	8, 88	8, 82	8, 78	8, 75	8, 6

24.

X	5	5, 5	6	6, 5	7	7, 5	8	8, 5
Y	2, 6	2, 3	2, 11	2	1, 92	1, 82	1, 55	1, 34

25.

X	16	19, 75	23, 10	26, 44	29, 79	33, 13	36, 89	44, 54
Y	29	38	39	54	62	70	79	98

26.

X	2	7, 5	12, 5	14, 5	16	18, 5	20,	20, 05
Y	2	2, 5	3	3, 5	4	4, 5	5	5, 5

27.

X	2, 5	7, 5	12, 5	17, 5	22, 5	27, 5	32, 5	37, 5
Y	2	6	10	14	18	22	26	30

28.

X	1330	1340	1350	1360	1370	1380	1390	1400
Y	0, 25	0, 26	0, 27	0, 28	0, 29	0, 3	0, 31	0, 32

29.

X	330	350	380	400	410	420	430	440
Y	110	100	90	80	70	65	60	55

30.

X	2, 95	2, 99	3, 05	3, 11	3, 21	3, 29	3, 34	3, 44
Y	39, 1	40, 5	42, 4	43, 8	45, 6	46, 9	48, 5	49, 4

РОЗВ’ЯЗАННЯ ТИПОВОГО ВАРІАНТА ІЗ № 4

Умова завдання

Вивчаючи залежність між показниками X і Y, проведено обстеження 13 об'єктів і отримані вибіркові дані (табл. 4.4.). Потрібно:

1) побудувати кореляційне поле (діаграму розсіювання), знайти вибірковий коефіцієнт кореляції та при рівні значущості $\alpha = 0.05$ перевірити його значущість, зробити висновок про залежність між ознаками X і У;

2) знайти рівняння прямої регресій У по X *(методом найменших квадратів* та за допомогою *вибіркового рівняння лінійної регресії* Y на X*)*, побудувати лінії регресії.

Таблиця 4.4

$X = x_i$	-19,2	-14,8	-19,6	-11,1	-9,4	-16,9	-13,7
$Y = y_i$	-21,8	-15,4	-20,8	-11,3	-11,6	-19,2	-13,0

Продовження таблиці 4.4

$X = x_i$	-4,9	-13,9	-9,4	-8,3	-7,9	-5,3
$Y = y_i$	-7,4	-15,1	-14,4	-11,1	-10,5	-7,2

Хід розв’язання

1. Побудуємо спочатку кореляційне поле для встановлення існування лінійного зв’язку між досліджуваними ознаками. Для цього, за даними таблиці 4.4, позначимо в прямокутній декартовій системі координат точки з координатами (x_i, y_i) – емпіричні дані. З кореляційного поля дійсно видно, що між ознаками X і Y дійсно спостерігається лінійний зв'язок.

Рис. 4.3

Знайдемо тепер вибірковий коефіцієнт кореляції за формулою (4.14) і допоміжних формул (4.12)-(4.13):

$$\rho_{xy} = \frac{s_{xy}}{s_x \cdot s_y},$$

де

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

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \overline{x^2} - \bar{x}^2,$$

$$s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2 = \overline{y^2} - \bar{y}^2.$$

Наступні обчислення виконаємо за допомогою розрахункової таблиці 4.5:

Таблиця 4.5

№ п/п i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	-19,2	-21,8	418,56	368,64	475,24
2	-14,8	-15,4	227,92	219,04	237,16
3	-19,6	-20,8	407,68	384,16	432,64
4	-11,1	-11,3	125,43	123,21	127,69
5	-9,4	-11,6	109,04	88,36	134,56
6	-16,9	-19,2	324,48	285,61	368,64
7	-13,7	-13	178,1	187,69	169
8	-4,9	-7,4	36,26	24,01	54,76
9	-13,9	-15,1	209,89	193,21	228,01
10	-9,4	-14,4	135,36	88,36	207,36
11	-8,3	-11,1	92,13	68,89	123,21
12	-7,9	-10,5	82,95	62,41	110,25
13	-5,3	-7,2	38,16	28,09	51,84
Сума:	**-154,4**	**-178,8**	**2385,96**	**2121,68**	**2720,36**

Обчислимо вибіркові середні за формулами (4.11) і даними останнього рядка таблиці 4.5:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{13}(-154{,}4) = -11{,}88;$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{13}(-178{,}8) = -13{,}75;$$

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i = \frac{1}{13}(2385{,}96) = 183{,}54.$$

Тепер обчислимо значення вибіркових дисперсій та вибіркової коваріації, які обчислюються за формулою (4.13) і при цьому також використаємо дані з таблиці 4.5:

$$s_{xy} = \overline{xy} - \bar{x} \cdot \bar{y} = 183{,}54 - (-11{,}88) \cdot (-13{,}75) = 183{,}54 - 163{,}35 = 20{,}19;$$

$$s_x^2 = \overline{x^2} - \bar{x}^2 = \frac{1}{13}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{13}\cdot(2121{,}68) - (-11{,}88)^2 =$$
$$= 163{,}21 - 141{,}13 = 22{,}08;$$

$$s_y^2 = \overline{y^2} - \bar{y}^2 = \frac{1}{13}\sum_{i=1}^{n} y_i^2 - \bar{y}^2 = \frac{1}{13}\cdot(2720{,}36) - (-13{,}75)^2 =$$
$$= 209{,}26 - 189{,}06 = 20{,}2.$$

Підставимо у формулу (4.15) отримані дані:

$$\rho_{xy} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{20{,}19}{\sqrt{22{,}08}\cdot\sqrt{20{,}2}} = \frac{20{,}19}{4{,}7\cdot 4{,}5} = \frac{20{,}19}{21{,}15} = 0{,}96.$$

Отже, зв'язок між ознаками *X* і *Y* існує сильний додатній зв'язок, близький до лінійного.

Перевіримо при рівні значущості $\alpha = 0{,}05$ гіпотезу про значущість коефіцієнта кореляції. Для цього використаємо схему перевірки гіпотези про статистичну значущість наведену в (4.31).

Із умови одержимо:

$$\left.\begin{matrix} n = 13; k = 11 \\ \rho_{xy} = 0{,}96 \\ \alpha = 0{,}05 \end{matrix}\right\} \to H_0\colon r_{xy} = 0 \to T_{\text{сп}} = \frac{0{,}96\sqrt{13-2}}{\sqrt{1-0{,}96^2}} = \frac{3{,}18}{0{,}28} = 11{,}35 \to$$

$$\to\ H_1\colon r_{xy} \neq 0\ \to\ t_{\text{кр}}(\alpha; k) = t_{\text{кр}}(0{,}05; 11) \xrightarrow{\text{Додаток 4}}$$

$$\xrightarrow{\text{Додаток 4}} t_{\text{кр}}^{\text{дв}} = 2{,}20 \to |T_{\text{сп}}| > t_{\text{кр}}^{\text{дв}} \to \boldsymbol{V}.$$

Отже, нульову гіпотезу відкидаємо, оскільки $|T_{\text{сп}}| > t_{\text{кр}}^{\text{дв}}$. Іншими словами, коефіцієнт кореляції між генеральними сукупностями X та У r_{xy} значуще відрізняється від нуля, тобто величини X та У корельовані, а отже пов’язані лінійною залежністю.

■

2.1. Із результатів, отриманих в першому пункті, висуваємо гіпотезу про лінійну залежність *Y* від X, тобто рівняння регресії (4.5) будемо шукати у вигляді $y = \alpha + \beta x$. Знайдемо параметри α і β, рівняння регресії за *методом* *найменших квадратів*, для чого складемо відповідну систему рівнянь для незгрупованих даних. Для проведення необхідних розрахунків використаємо розрахункову таблицю 4.5.

Отже, складемо систему (4.6) для знаходження параметрів рівняння регресії та розв'яжемо її за правилом Крамера:

$$\begin{cases} \beta \sum_{i=1}^{n} x_i^2 + \alpha \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i\,; \\ \beta \sum_{i=1}^{n} x_i + \alpha n = \sum_{i=1}^{n} y_i. \end{cases} \Rightarrow \begin{cases} 2121{,}68\beta - 154{,}4\alpha = 2385{,}96; \\ -154{,}4\beta + 13\alpha = -178{,}8. \end{cases}$$

Знайдемо детермінант основної матриці системи, яка складена із коефіцієнтів перед невідомими:

$$\Delta = \begin{vmatrix} 2121{,}68 & -154{,}4 \\ -154{,}4 & 13 \end{vmatrix} = 2121{,}68 \cdot 13 - (-154{,}4) \cdot (-154{,}4) = 3742{,}48 \neq 0.$$

Знайдемо допоміжні детермінанти, що отримуються із попереднього заміною відповідного стовпця коефіцієнтів на стовпець вільних членів:

$$\Delta_\beta = \begin{vmatrix} 2385{,}96 & -154{,}4 \\ -178{,}8 & 13 \end{vmatrix} = 2385{,}96 \cdot 13 - (-154{,}4) \cdot (-178{,}8) = 3410,76;$$

$$\Delta_\alpha = \begin{vmatrix} 2121{,}68 & 2385{,}96 \\ -154{,}4 & -178{,}8 \end{vmatrix} = 2121{,}68 \cdot (-178{,}8) - (-154{,}4) \cdot 2385{,}96$$
$$= -10964{,}16.$$

Знайдемо невідомі за формулами Крамера:

$$\beta = \frac{\Delta_\beta}{\Delta} = \frac{3410,76}{3742{,}48} = 0{,}911, \alpha = \frac{\Delta_\alpha}{\Delta} = \frac{-10964{,}16}{3742{,}48} = -2{,}93.$$

Отже, шукане рівняння лінії регресії має вигляд: $y = 0{,}911x - 2{,}93$.

2.2. Знайдемо тепер параметри β *і* α, рівняння регресії за допомогою рівняння вибіркової лінійної регресії Y на X (див. (4.16)):

$$\bar{y}_x - \bar{y} = \rho_{xy} \cdot \frac{s_y}{s_x} \cdot (x - \bar{x}).$$

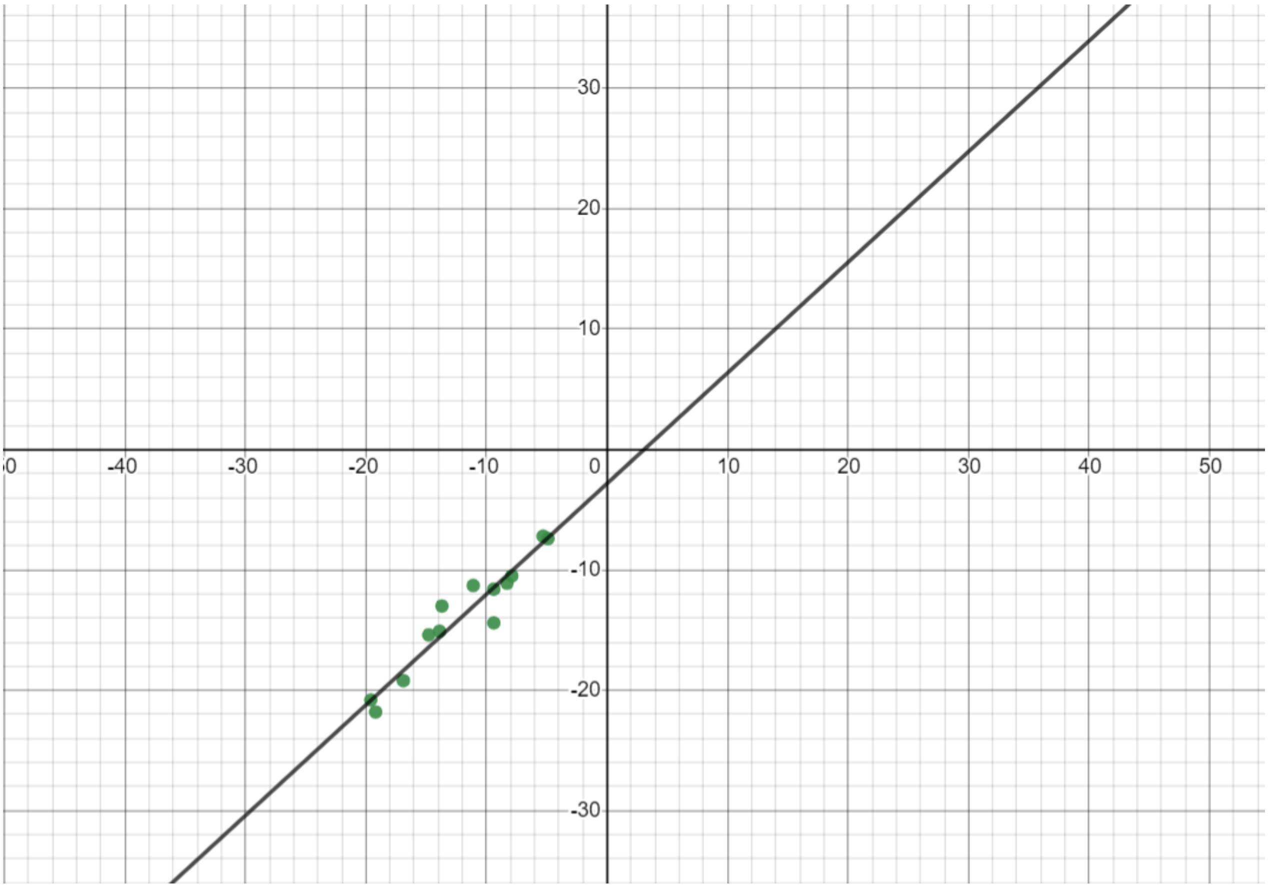
Підставимо значення вибіркових середніх та коефіцієнта кореляції, отримані в пункті 1:

$$\bar{y}_x - \bar{y} = \rho_{xy} \cdot \frac{s_y}{s_x} \cdot (x - \bar{x}) \Rightarrow$$

$$\Rightarrow \bar{y}_x - (-13{,}75) = 0{,}96 \cdot \frac{4{,}5}{4{,}7} \cdot \big(x - (-11{,}88)\big) \Rightarrow$$

$$\Rightarrow \bar{y}_x + 13{,}75 = 0{,}919\, x + 10{,}918 \Rightarrow \bar{y}_x = 0{,}92\, x - 2{,}83.$$

Отже, пряма регресії має вигляд:



.

■

ДОДАТКИ

Додаток 1

Таблиця значень функції Гаусса $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$.

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	3969	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1.0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229

2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2.7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
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

x	4,0	4,2	4,4	4,6	4,8	5,0
$\varphi(x)$	0,00013	0,0000589	0,0000249	0,0000101	0,0000040	0,0000015

Додаток 2

Таблиця значень функції $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}t^2} dt.$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	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	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	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	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41308	41466	41621	41774
1,4	41924	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	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	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	49638	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997

x	4,0	4,2	4,4	4,6	4,8	5,0
$\Phi(x)$	0,4999683	0,4999867	0,4999946	0,4999979	0,4999992	0,4999997

Додаток 3

Критичні точки розподілу $\chi^2(\alpha; k)$.

k	Рівень значущості α										
	0,005	0,01	0,025	0,05	0,1	0,5	0,9	0,95	0,975	0,99	0,995
1	$0{,}0^4 4$	$0{,}0^3 2$	0,001	0,004	0,016	0,46	2,71	3,84	5,02	6,63	7,88
2	0,010	0,020	0,051	0,103	0,211	1,39	4,61	5,99	7,38	9,21	10,6
3	0,072	0,115	0,216	0,352	0,584	2,37	6,25	7,81	9,35	11,3	12,8
4	0,207	0,297	0,484	0,711	1,06	3,36	7,78	9,49	11,1	13,3	14,9
5	0,412	0,554	0,831	1,15	1,61	4,35	9,24	11,1	12,8	15,1	16,7
6	0,676	0,872	1,24	1,64	2,20	5,35	10,6	12,6	14,4	16,8	18,5
7	0,989	1,24	1,69	2,17	2,83	6,35	12,0	14,1	16,0	18,5	20,3
8	1,34	1,65	2,18	2,73	3,49	7,34	13,4	15,5	17,5	20,1	22,0
9	1,73	2,09	2,70	3,33	4,17	8,34	14,7	16,9	19,0	21,7	23,6
10	2,16	2,56	3,25	3,94	4,87	9,34	16,0	18,3	20,5	23,2	25,2
11	2,60	3,05	3,82	4,57	5,58	10,3	17,3	19,7	21,9	24,7	26,8
12	3,07	3,57	4,40	5,23	6,30	11,3	18,5	21,0	23,3	26,2	28,3
13	3,57	4,11	5,01	5,89	7,04	12,3	19,8	22,4	24,7	27,7	29,8
14	4,07	4,66	5,63	6,57	7,79	13,3	21,1	23,7	26,1	29,1	31,3
15	4,60	5,23	6,26	7,26	8,55	14,3	22,3	25,0	27,5	30,6	32,8
16	5,14	5,81	6,91	7,96	9,31	15,3	23,5	26,3	28,8	32,0	34,3
17	5,70	6,41	7,56	8,67	10,1	16,3	24,8	27,6	30,2	33,4	35,7
18	6,26	7,01	8,23	9,39	10,9	17,3	26,0	28,9	31,5	34,8	37,2
19	6,84	7,63	8,91	10,4	11,7	18,3	27,2	30,1	32,9	36,2	38,6
20	7,43	8,26	9,59	10,9	12,4	19,3	28,4	31,4	34,2	37,6	40,0

21	8,03	8,90	10,3	11,6	13,2	20,3	29,6	32,7	35,5	38,9	41,4
22	8,64	9,54	11,0	12,3	14,0	21,3	30,8	33,9	36,8	40,3	42,8
23	9,26	10,2	11,7	13,1	14,8	22,3	32,0	35,2	38,1	41,6	44,2
24	9,89	10,9	12,4	13,8	15,7	23,3	33,2	36,4	39,4	43,0	45,6
25	10,5	11,5	13,1	14,6	16,5	24,3	34,4	37,7	40,6	44,3	46,9
26	11,2	12,2	13,8	15,4	17,3	25,3	35,6	38,9	41,9	45,6	48,3
27	11,8	12,9	14,6	16,2	18,1	26,3	36,7	40,1	43,2	47,0	49,6
28	12,5	13,6	15,3	16,9	18,9	27,3	37,9	41,3	44,5	48,3	51,0
29	13,1	14,3	16,0	17,7	19,8	28,3	39,1	42,6	45,7	49,6	52,3
30	13,8	15,0	16,8	18,5	20,6	29,3	40,3	43,8	47,0	50,9	53,7
35	17,2	18,5	20,6	22,5	24,8	34,3	46,1	49,8	53,2	57,3	60,3
40	20,7	22,2	24,4	26,5	29,1	39,3	51,8	55,8	59,3	63,7	66,8
45	24,3	25,9	28,4	30,6	33,4	44,3	57,5	61,7	65,4	70,0	73,2
50	28,0	29,7	32,4	34,8	37,7	49,3	63,2	67,5	71,4	76,2	79,5
75	47,2	49,5	52,9	56,1	59,8	74,3	91,1	96,2	100,8	106,4	110,3
100	67,3	70,1	74,2	77,9	82,4	99,3	118,5	124,3	129,6	135,6	140,2

Додаток 4

Критичні точки розподілу Стьюдента $t(\alpha; k)$.

k	Рівень значущості α						
	0,750	0,900	0,950	0,975	0,990	0,995	0,999
1	1,000	3,078	6,314	12,71	31,82	63,66	318
2	0,816	1,886	2,920	4,303	6,965	9,925	22,3
3	0,765	1,638	2,353	3,182	4,541	5,841	10,2
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297
10	0,700	1,372	1.812	2,228	2,764	3,169	4,144
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025
12	0,695	1,356	1,782	2,179	2,681	3,055	3,930
13	0,694	1,350	1,771	2,160	2,650	3,012	3,852
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552

21	0,686	1,323	1,721	2,080	2,518	2,831	3,527
22	0,686	1,321	1,717	2,074	2,508	2,819	3,505
23	0,685	1,319	1,714	2,069	2,500	2,807	3,485
24	0,685	1,318	1,711	2,064	2,492	2,797	3,467
25	0,684	1,316	1,708	2,060	2,485	2,787	3,450
26	0,684	1,315	1,706	2,056	2,479	2,779	3,435
27	0,684	1,314	1,703	2,052	2,473	2,771	3,421
28	0,683	1,313	1,701	2,048	2,467	2,763	3,408
29	0,683	1,311	1,699	2,045	2,462	2,756	3,396
30	0,683	1,310	1,697	2,042	2,457	2,750	3,385
40	0,681	1,303	1,684	2,021	2,423	2,704	3,307
60	0,679	1,296	1,671	2,000	2,390	2,660	3,232
120	0,677	1,289	1,658	1,980	2,358	2,617	3,160
∞	0,674	1,282	1,645	1,960	2,326	2,576	3,090

Додаток 5

Критичні точки розподілу F Фішера-Снедекора $F(\alpha; k_1; k_2)$ (k_1 – число ступенів вільності більшої дисперсії, k_2 – число ступенів вільності меншої дисперсії).

$$\alpha = 0{,}95$$

k_2	k_1								
	1	2	3	4	5	6	7	8	9
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38
3	16,13	9,55	9,78	9,12	9,01	8,94	8,89	8,85	8,81
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,99	5,14	4,75	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3.86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	4,85	3,98	3,59	3,36	3,20	3,09	3,01	2,93	2,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46
19	4,30	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39

21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,37	2,30
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88

$\alpha = 0{,}95$

k_2	k_1								
	10	12	15	20	24	30	40	60	120
1	241,9	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,3
2	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49
3	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55
4	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66
5	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40
6	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70
7	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27
8	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97
9	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75
10	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58
11	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45
12	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34
13	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25
14	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18
15	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11
16	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06
17	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01
18	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97
19	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93
20	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90

21	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87
22	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84
23	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81
24	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79
25	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77
26	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75
27	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73
28	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71
29	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70
30	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68
40	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58
60	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47
120	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35
∞	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22

$$\alpha = 0{,}99$$

k_2	k_1								
	1	2	3	4	5	6	7	8	9
1	4052	4999,5	5403	5625	5764	5859	5928	5982	6022
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46

21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07
40	7,31	5,18	4,31	3,53	3,51	3,29	3,12	2,99	2,89
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41

$$\alpha = 0{,}99$$

k_2	k_1								
	10	12	15	20	24	30	40	60	120
1	6056	6106	6157	6209	6235	6261	6287	6313	6339
2	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49
3	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22
4	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56
5	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11
6	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97
7	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74
8	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95
9	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40
10	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00
11	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69
12	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45
13	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25
14	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09
15	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96
16	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84
17	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75
18	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66
19	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58
20	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52

21	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46
22	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40
23	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35
24	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31
25	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27
26	3,09	2,96	2,81	2,66	2,58	2,50	2,42	2,33	2,23
27	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,20
28	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,17
29	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14
30	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11
40	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92
60	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73
120	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53
∞	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32

СПИСОК ВИКОРИСТАНИХ ТА РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

1. Барковський В. В., Барковська Н. В., Лопатін О. К. Теорія ймовірностей та математична статистика: навчальний посібник. К.: ЦНЛ, 2006. 424 с.

2. Василенко О. А., Сенча І. А. Матемачно-статистичні методи аналізу у прикладних дослідженнях: навч. посіб. Одеса: ОНАЗ ім. О. С. Попова, 2011. 166 с.

3. Вища математика: Модульна технологія навчання : У 4 ч. : навч. посіб. У Ч. 4. Теорія ймовірностей і математична статистика / В. П. Денисюк, В.М. Бобков, Т. А. Погребецька, В. К. Репета. 2-ге вид. доопрац. К. : Вид-во Нац. авіац. ун-ту «НАУ-друк», 2009. 256 с.

4. Глеч С.Г., Ледяєв С.Ф., Ольшанська І.В. Теорія ймовірностей та математична статистика: навч. посібник. Севастополь: СевНТУ, 2011. 176 с.

5. Гончаренко Я. В. Теорія ймовірностей і математична статистика. Практикум К.: НПУ імені М. П. Драгоманова, 2011. 145 с.

6. Жлуктенко В. І., Наконечний С. І. Теорія ймовірностей і математична статистика: навчально-методичний посібник у 2-х частинах. Ч. 2. Математична статистика. К.: КНЕУ, 2001. 336 с.

7. Зайцев Є. П. Теорія ймовірності і математична статистика. Базовий курс з індивідуальними завданнями і розв'язком типових варіантів: навч. посібн. К. : Алерта, 2013. 440 с.

8. Кармелюк Г.І. Теорія ймовірностей та математична статистика. Посібник з розвязання задач. К.: Центр учбової літератури, 2007. 576 с.

9. Кушлик-Дивульська О. І., Поліщук Н. В., Орел Б. П., Штабалюк П. І. Теорія ймовірностей та математична статистика: навч. посіб. К.: НТУУ «КПІ», 2014. 212 с.

10. Малютіна Т. І., Дахер К. А. Вища математика для економістів. Ч. 4. Теорія ймовірностей та математична статистика: навчальний посібник : у 4-х ч. Суми : ДВНЗ «УАБС НБУ», 2008. 181 с.

11. Медведєв М. Г., Пащенко І. О. Теорія ймовірностей та математична статистика: підручник. К.: Кондор, 2008. 536 с.

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

13. Оленко А.Я. Ймовірність і статистика. Задачі. К.: НаУКМА, 2002. 53 с.

14. Практикум з теорії ймовірностей та математичної статистики: навчальний посібник / за ред. Р. К. Чорнея. Київ: МАУП, 2003. 328 с.

15. Руденко В. М. Математична статистика. Навч. посіб. К.: Центр учбової літератури, 2012. 304 с.

16. Слюсарчук П. В. Теорія ймовірностей та математична статистика: підруч. для студ. вищ. навч. закл. Ужгород : вид-во «Карпати», 2005. 181 с.

17. Синявська О. О. Додаткові розділи статистики. Конспект лекцій. Ужгород: вид-во УжНУ «Говерла», 2017. 124 с.

18. Турчин В.М. Математична статистика в прикладах і задачах. У 2-х ч. Дніпропетровськ: ДДУ, 1998. Ч.1 68 с. Ч.2 228 с

Навчальне видання

МАТЕМАТИЧНА СТАТИСТИКА

НАВЧАЛЬНИЙ ПОСІБНИК

Відповідальний за випуск: завідувач кафедри теорії ймовірностей і математичного аналізу доктор фіз.-мат. наук, доц. Сливка-Тилищак Г.І.

Автори: канд. фіз.-мат. наук Герич М.С.
канд. фіз.-мат. наук, доц. Синявська О.О.

Рецензенти: д-р. фіз.-мат. наук, проф. Пашко А.О.,
канд. фіз.-мат. наук, доц. Млавець Ю.Ю.