

**Міністерство освіти і науки України
Центральноукраїнський національний технічний університет
Харківський національний автомобільно-дорожній університет
Кафедра будівельних, дорожніх машин і будівництва
Кафедра будівельних і дорожніх машин**

СТАТИСТИЧНІ МЕТОДИ В ІНЖЕНЕРНИХ ДОСЛІДЖЕННЯХ
Методичні рекомендації до самостійного вивчення та виконання
практичних завдань для здобувачів спеціальності 192 –
Будівництво та цивільна інженерія, 133 – Галузеве
машинобудування, 073 – Менеджмент

Затверджено на засіданні кафедри
будівельних, дорожніх машин і
будівництва 16 січня 2024 року,
протокол № 5

Кропивницький, Харків 2024

Статистичні методи обробки даних. Методичні рекомендації до самостійного вивчення та виконання практичних завдань для здобувачів спеціальності 192 – Будівництво та цивільна інженерія, 133 – Галузеве машинобудування 073 – Менеджмент. – Кропивницький, Харків: ЦНТУ, ХНАДУ 2024. – 79 с.

Методичні рекомендації призначені для виконання практичних завдань і самостійного вивчення дисципліни «Статистичні методи обробки даних» здобувачами спеціальності 192 – Будівництво та цивільна інженерія та 073 – Менеджмент дистанційної та денної форм навчання. Можуть бути корисні для здобувачів при виконанні бакалаврських і магістерських кваліфікаційних робіт у частині обробки експериментальних даних.

Укладачі – к.т.н., доцент Хачатурян С.Л., доцент кафедри будівельних, дорожніх машин і будівництва ЦНТУ

к.т.н., доцент Пімонов І.Г., доцент кафедри будівельних і дорожніх машин ім. А.М. Холодова ХНАДУ

к.т.н., доцент Щербак О.В., доцент кафедри будівельних і дорожніх машин ім. А.М. Холодова ХНАДУ

к.е.н. Хачатурян О.С., доцентка кафедри менеджменту авіаційної діяльності ЛА НАУ

Рецензент – д.т.н., професор Кулешков Ю.В., професор кафедри експлуатації та ремонту машин ЦНТУ

Відповідальний за випуск – к.т.н., професор Настоящий В.А., завідувач кафедри будівельних, дорожніх машин і будівництва ЦНТУ

© ЦНТУ, Кропивницький, пр. Університетський, 8
© Хачатурян С.Л., Пімонов І.Г. Щербак О.В.,
Хачатурян О.С.

З М І С Т

	с.
Вступ.....	5
1. Статистичний аналіз експериментальних даних.....	6
1.1. Елементи теорії ймовірностей.....	8
1.2. Числові характеристики випадкової величини.....	11
1.3. Числові характеристики положення (мода, медіана, квантилі)....	14
1.4. Типові закони розподілу.....	16
1.4.1. Геометричний розподіл.....	16
1.4.2. Біноміальний розподіл.....	16
1.4.3. Розподіл Пуассона.....	17
1.4.4. Рівномірний розподіл.....	18
1.4.5. Експоненціальний розподіл.....	19
1.4.6. Нормальний закон розподілу.....	20
1.4.7. Розподіл χ^2 (хі-квадрат).....	24
1.4.8. Розподіл Стьюдента.....	25
1.5. Числові характеристики системи випадкових величин (коваріація та кореляція).....	28
1.6. Нормальний розподіл системи випадкових величин.....	29
1.7. Елементи математичної статистики.....	30
1.7.1. Генеральна сукупність і випадкова вибірка.....	30
1.7.2. Точкові оцінки параметрів нормального розподілу.....	31
1.7.3. Класифікація помилок вимірювання.....	32
1.7.4. Закон додавання помилок.....	33
1.7.5. Помилки непрямих вимірювань.....	34
1.8. Довірчі інтервали і довірча ймовірність.....	35
1.9. Визначення необхідної кількості дослідів.....	38
1.10. Перевірка статистичних гіпотез.....	39
1.10.1. Відсів грубих похибок спостережень.....	41
1.10.2. Порівняння двох рядів спостережень.....	42
1.10.3. Перевірка однорідності дисперсій.....	43
1.10.4. Перевірка однорідності декількох дисперсій.....	44
1.10.5. Перевірка гіпотез про числових значеннях математичних очікувань.....	45
1.11. Критерії згоди. Перевірка гіпотез про вид функції розподілу...	47

1.11.1. Критерій Пірсона.....	48
1.11.2. Критерій Колмогорова.....	49
1.11.3. Критерій однорідності статистичного матеріалу.....	50
2. Методика виконання практичного завдання.....	54
2.1. Мета практичного завдання.....	54
2.2. Приклад виконання практичного завдання.....	54
2.2.1. Задача.....	54
2.2.2. Зміст практичного завдання.....	54
2.2.3. Методичні рекомендації до виконання практичного завдання...55	
2.2.4. Методика виконання практичного завдання.....	58
Література.....	69
Варіанти завдань.....	71
Додаток 1.....	75
Додаток 2.....	75
Додаток 3.....	76
Додаток 4.....	77
Додаток 5.....	78

ВСТУП

Науковому співробітнику, інженеру-будівельнику, менеджеру нерідко доводиться розв'язувати задачі з планування, проведення експериментальних досліджень, а також обробки результатів статистичних досліджень. Розширення можливостей сучасної комп'ютерної техніки дозволяють досліднику мати на озброєнні загальні підходи, методи та процедури планування та обробки їх результатів.

У теперішній час статистична обробка результатів досліджень не мислима без використання комп'ютерної техніки і пакетів прикладних програм. При величезному різноманітті статистичних пакетів, якими характеризується сучасний світовий і вітчизняний ринок, досліднику важливо вміти правильно зорієнтуватись на цьому ринку, правильно розуміти область використання статистичних методів розв'язання того чи іншого класу задач.

Одним з важливих етапів аналізу статистичних даних є їх попередня обробка, котра необхідна для того, щоб у подальшому, при побудові емпіричних залежностей (функцій відгуку), з найбільшою ефективністю використовувати статистичні методи та коректно аналізувати отримані результати.

Зміст попередньої обробки полягає у відсіюванні грубих похибок і оцінці достовірності результатів досліджень. Іншими важливими моментами попередньої обробки даних є перевірка відповідності результатів вимірювання нормальному закону та визначення параметрів цього розподілення. Якщо гіпотеза про те, що відгук не суперечить нормальному розподіленню, виявиться непридатною, то слід визначити, якому закону розподілення підкоряються дослідні дані чи, якщо це можливо, перетворити отримане розподілення до нормального виду.

1. СТАТИСТИЧНИЙ АНАЛІЗ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ

При виконанні вимірювань дослідник намагається визначити значення тієї чи іншої величини. Якщо використовувати досить точні прилади, то можна побачити, що повторне вимірювання однієї й тієї ж величини призводить іноді до результатів, які в тій або іншій мірі відрізняються від результатів початкового виміру. Це явище характерне як для простих, так і для складних вимірювань.

Умови проведення дослідження весь час змінюються, і в умовах реальної розвідки від розкиду та змін його результатів позбутися неможливо. Дослідник виконує вимірювання величин, які ніколи не залишаються постійними. Тому постановка питання про значення певної величини може бути некоректною, потрібна постановка такого питання, яке відбивало б цю властивість мінливості.

Рішення полягає в тому, щоб характеризувати досліджувану величину не одним значенням, а ймовірністю знайти в дослідженні те чи інше значення. Для цього вводиться функція, котра називається розподілом ймовірності виявлення шуканої величини, що показує, які значення частіше зустрічаються в дослідженні.

Функція розподілу в більшості досліджень є досить простою й має дві характеристики. Перша – середнє значення шуканої величини, друга – показує область навколо цієї середньої величини, в якій зосереджено більшість результатів дослідження. Вона характеризує ширину цього розподілу й називається похибкою. Ця ширина має сувору інтерпретацію в термінах теорії ймовірностей, тобто можна вказати, з якою ймовірністю ми повинні виявити істинне значення в заданій області навколо вимірюваного середнього значення. Цю похибку називають природною.

Для дослідника побудова функції розподілу вимагає проведення багаторазових (нескінченного числа) досліджень, що буває дорого і не потрібно. Тому доводиться обмежуватися кінцевим числом досліджень, що привносить додаткову похибку.

Тут з'являється й інша проблема: в кожному дослідженні присутній вимірювальний прилад або так званий «людський фактор», який вносить зміни в початкову функцію розподілу, приводячи до додаткової (приладної) похибки.

Поділ похибки на природну і приладову досить умовне, воно дозволяє краще зрозуміти природу похибки.

Дослідник повинен завжди ставити собі два питання: як виміряти досліджувану величину, тобто як визначити її характеристики – середню та ширину, й до якої міри вдасться розумно зменшити похибку дослідження? Тому важливо розуміти взаємозв'язок між трьома складовими похибки:

- природну похибка можна зменшити, змінюючи умови проведення дослідження,

- похибку, пов'язану з кінцевістю числа вимірювань – збільшуючи їх число,

- приладову – використовуючи більш точні методи й інструменти вимірювань, за можливості звести до мінімуму «людський фактор».

Разом з тим неможливо зменшити похибка до нуля. Для неї існує нижня межа, оцінка якої – принципове фізичне питання. Тому завданням є визначити ті методи дослідження, які адекватні бажаній і досяжній точності. Залежно від бажаної точності можуть виникнути різні ситуації:

- якщо ми бажаємо отримати порядок вимірюваної величини, то й похибка повинна оцінюватися грубо;

- якщо ми бажаємо отримати точність порядку декількох відсотків, тоді необхідно й більш акуратно визначати похибки;

- якщо необхідно отримати точність, порівнянну з точністю еталонних вимірювань, то проблема визначення похибки може стати більш важливою й складною, ніж проблема вимірювання самої величини.

Крім зазначених у дослідженні можуть мати місце й інші джерела помилок, які викликають так звані систематичні помилки. Виявлення їх і аналіз набагато складніший, ніж випадкових. Можна вказати три основних джерела систематичних помилок: методика, обрана для проведення дослідження, погана робота вимірювальних приладів, і, нарешті, помилки самого дослідника.

Оскільки відгук через вплив неконтрольованих факторів є випадковою величиною, то при обробці результатів дослідження

широко використовується апарат теорії ймовірності та математичної статистики, тому необхідно згадати необхідні поняття та визначення цього розділу математики.

1.1. Елементи теорії ймовірностей

Випадковою називається подія, результат якої при певному комплексі умов неможливо передбачити заздалегідь.

Коли мова йде про експеримент, мається на увазі, що він має певні результати. Список цих результатів часто буває досить невеликим.

Випадкова величина – величина, котра може приймати будь-яке значення з встановленого числа й з якою пов'язаний імовірнісний розподіл.

Випадкова величина може бути дискретною чи безперервною.

Дискретна випадкова величина – може приймати значення тільки з кінцевої чи рахункової множини дійсних чисел.

Безперервна випадкова величина – може приймати будь-які значення кінцевого або нескінченного інтервалу.

Експеримент і його результати часто мають певні числові характеристики. Саме наявність такого роду числових характеристик і дає підстави для використання математичних методів при вивченні випадкових подій.

Якщо зафіксувати рівні контрольованих факторів і провести n вимірювань відгуку X , то в результаті буде отримано низку близьких, але відмінних один від одного значень x_i , ($i=1, 2, \dots, n$), де x_i – i -тий вимір величини X ; x_1, x_2, \dots, x_n – реалізація випадкової величини X .

Однією з найважливіших числових характеристик випадкової події є її ймовірність, яка є деяким числом, відповідним даній випадковій події. Потрібно розуміти, що це – фундаментальна характеристика й тому простого визначення, що застосовувалося б до всіх випадкових подій, просто не може бути.

Для визначення ймовірності деякої випадкової події A виділимо деяку (кінцеву) множину випадків, які вважаються (чи передбачаються) рівно імовірними. Позначимо число цих результатів через n . Далі встановлюється, що заданій події A сприяє певне число, скажімо m , з цих n результатів.

Тоді вважають за визначенням, що частотою реалізації події A є $w=m/n$.

Імовірність $p(A)$ випадкової події A – число від нуля до одиниці, котре являє собою межу частоти реалізації події A при необмеженій кількості повторень одного й того ж комплексу умов.

Для дискретної випадкової величини можна вказати імовірність, з якою вона приймає кожне із своїх можливих значень кінцевої чи рахункової безлічі дійсних чисел. Для неперервної випадкової величини задають імовірність її попадання в один із заданих інтервалів області її визначення, оскільки імовірність того, що вона прийме якесь певне значення, прямує до нуля.

Випадкові величини можна задавати різними способами. Дискретні випадкові величини зазвичай задаються своїм законом розподілу в табличному чи графічному вигляді. Кожному можливому значенню x_1, x_2, \dots випадкової величини X зіставляється імовірність p_1, p_2, \dots цього значення. В результаті утворюється таблиця, що складається з двох рядків:

$$x_1 \ x_2 \ x_3 \ \dots$$

$$p_1 \ p_2 \ p_3 \ \dots$$

Це і є закон розподілу випадкової величини, під яким розуміють зв'язок між можливими значеннями випадкової величини та відповідними їм імовірностями.

Безперервні випадкові величини законом розподілу в вигляді таблиці задати неможливо, так як за визначенням їх значення неможливо перенумерувати. Однак для безперервних випадкових величин є інший спосіб задання (який можна застосовувати, до речі, і для дискретних величин) – це функція розподілу. Зазвичай використовується два способи опису розподілів ймовірностей випадкових величин: інтегральний (за допомогою функції розподілу) і диференційний (задається щільністю розподілу).

Функція розподілу $F(x)$ визначає для всіх дійсних x імовірність того, що випадкова величина X приймає значення не більше ніж x :

$$F(x)=P(X \leq x). \quad (1.1)$$

Функція розподілу $F(x)$ має такі властивості:

1. $F(x)$ приймає значення від 0 до 1 : $0 \leq F(x) \leq 1$.

2. Її ордината, що відповідає довільній точці x_1 , є ймовірність того, що випадкова величина X буде менше, ніж x_1 , тобто $F(x_1)=P(X<x_1)$.

3. Функція розподілу прагне до нуля при необмеженому зменшенні x і прагне до одиниці при необмеженому зростанні x , тобто

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

4. Функція розподілу являє собою монотонно зростаючу криву, тобто $F(x_1) > F(x_2)$, якщо $x_1 > x_2$.

5. Її приріст на довільному інтервалі (x_1, x_2) дорівнює ймовірності того, що випадкова величина X потрапить у даний інтервал:

$$F(x_2) - F(x_1) = P(X \leq x_2) - P(X \leq x_1) = P(x_1 \leq X \leq x_2).$$

Часто замість функції розподілу зручно використовувати іншу функцію – щільність розподілу випадкової величини X . Її ще іноді називають диференціальною функцією розподілу. Щільність розподілу $f(x)$ – це перша похідна (якщо вона існує) функції розподілу:

$$f(x) = \frac{dF(x)}{dx}.$$

Щільність функції розподілу $f(x)$ має такі властивості:

1. Щільність розподілу ймовірностей є невід'ємною функцією, тобто $f(x) \geq 0$. Ця властивість є наслідком того, що функція $F(x)$ є неубутна функція.

2. Функція розподілу ймовірностей випадкової величини дорівнює X дорівнює визначеному інтегралу від щільності розподілу ймовірностей у межах від $-\infty$ до x :

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

3. Імовірність події, що полягає в тому, що випадкова величина X прийме значення, розташоване в інтервалі (x_1, x_2) , дорівнює визначеному інтегралу від щільності розподілу ймовірностей на цьому інтервалі:

$$P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x) dx$$

4. Інтеграл щільності розподілу на нескінченно великому інтервалі $(-\infty, +\infty)$ дорівнює одиниці:

$$\int_{-\infty}^{\infty} f(x) dx = P(-\infty < X < +\infty) = 1,$$

так як попадання випадкової величини в інтервал $-\infty < X < +\infty$ є достовірною подією.

У більшості випадків при обробці експериментальних даних, ґрунтуючись на властивостях досліджуваної випадкової величини, вдається записати функцію її розподілу (щільність розподілу) з точністю до деяких невідомих параметрів. Так, для випадкової величини, котра задовольняє нормальному закону розподілу (закону Гауса), функція розподілу записується в вигляді:

$$F(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma_x^2}} \int_{-\infty}^x e^{-\frac{(x-M_x)^2}{2 \cdot \sigma_x^2}} dx.$$

В цьому випадку константи M_x , σ_x^2 є параметрами розподілу і визначають двохпараметричний закон розподілу.

Параметр розподілу – постійна величина, від якої залежить функція розподілу.

Отже, якщо встановлено, що випадкова величина не суперечить тому чи іншому закону розподілу, то для того, щоб однозначно охарактеризувати цю випадкову величину, досить знати параметри її розподілу. Так, для нормального закону параметрами розподілу є M_x – математичне очікування (що характеризує центр розсіювання) і σ_x^2 – дисперсія (характеризує ступінь розсіювання).

1.2. Числові характеристики випадкової величини

Функція розподілу ймовірностей повністю описує випадкову величину з імовірнісної точки зору. Однак у багатьох практичних задачах немає необхідності будувати закон розподілу, досить буває вказати лише окремі числові характеристики, котрі до деякої міри характеризують істотні риси розподілу випадкової величини.

Такі характеристики, призначення яких – виразити в стислій формі найбільш істотні особливості розподілу, називають числовими характеристиками випадкової величини.

Як основні числові характеристики випадкової величини виступають, так звані, моменти випадкової величини. Найчастіше застосовуються моменти двох видів: початкові та центральні. Для дискретної випадкової величини початковий момент k -ого порядку визначається формулою:

$$m_k = \sum_{i=1}^n x_i^k \cdot p_i, k=1, 2, \dots,$$

для неперервної випадкової величини – формулою

$$m_k = \int_{-\infty}^{\infty} x^k \cdot f(x) \cdot dx.$$

Початковий момент першого порядку ($k=1$) називається математичним очікуванням (середнім значенням) випадкової величини. Математичне сподівання прийнято позначати різним чином:

$$M_x, M[X], m_x, m.$$

Для дискретних випадкових величин

$$m_1 = \sum_{i=1}^n x_i \cdot p_i, k=1, 2, \dots,$$

для безперервних

$$m_1 = M[X] = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx.$$

Користуючись знаком математичного очікування, можна об'єднати формули для моментів k -ого порядку:

$$m_k = M[X^k],$$

тобто початковим моментом k -ого порядку випадкової величини X називається математичне сподівання k -го ступеня цієї випадкової величини.

Перед тим, як дати визначення центрального моменту, введемо поняття центрованої випадкової величини.

Нехай є випадкова величина X з математичним очікуванням m_x . Центрованою випадковою величиною, що відповідає величині X , називається відхилення випадкової величини X від її математичного очікування:

$$\dot{X} = X - m_x.$$

Неважко переконатися, що математичне очікування центрованої випадкової величини дорівнює нулю. Центрування випадкової величини рівносильно переносу початку координат в середню точку, абсциса якої дорівнює математичному очікуванню.

Моменти центрованої випадкової величини називаються центральними. Центральний момент k -ого порядку для дискретної

випадкової величини визначається формулою

$$\mu_k = \sum_{i=1}^n (x_i - m_x)^k p_i,$$

для неперервної випадкової величини

$$\mu_k = \int_{-\infty}^{\infty} (x - m_x)^k f(x) \cdot dx.$$

Таким чином, центральним моментом порядку k випадкової величини X називається математичне сподівання k -го ступеня відповідної центрованої випадкової величини:

$$\mu_k = M[\dot{X}^k] = M[(X - m_x)^k].$$

Перший центральний момент завжди дорівнює нулю. Другий центральний момент називається дисперсією. Дисперсією випадкової величини називається математичне сподівання квадрата відхилення випадкової величини від її математичного очікування:

$$D[X] = M[(X - m_x)^2],$$

для дискретної випадкової величини

$$D[X] = \sum_{i=1}^n (x_i - m_x)^2 p_i,$$

для безперервної

$$D[X] = \int_{-\infty}^{\infty} (x - m_x)^2 f(x) \cdot dx.$$

Інші позначення для дисперсії: D_x , σ_x^2 , σ^2 . Дисперсія випадкової величини має розмірність квадрата випадкової величини; для наочної характеристики розсіювання зручніше користуватися величиною, розмірність якої збігається з розмірністю випадкової величини. Тому часто використовується середньоквадратичне відхилення (СКВ або стандарт), рівне квадратному кореню з дисперсії і позначається σ_x , σ .

Математичне сподівання та дисперсія є числовими характеристиками випадкової величини, котрі найбільш часто використовуються. Вони характеризують найбільш важливі риси розподілу: його положення та ступінь розкиданості. Для більш детального опису розподілу використовуються моменти вищих

порядків.

Третій центральний момент служить для характеристики асиметрії (скошеності) розподілу. Якщо розподіл симетричний щодо математичного очікування, то всі моменти непарного порядку (якщо вони існують), дорівнюють нулю. Тому найбільш логічно прийняти 3-й центральний момент, а щоб отримати безрозмірну характеристику, його ділять на куб середнього квадратичного відхилення. Отримана величина називається коефіцієнтом асиметрії:

$$\gamma_1 = \frac{\mu_3}{\sigma^3}.$$

Четвертий центральний момент служить для характеристики «Крутості», тобто гостровершиності чи плосковершиності розподілів. Ці властивості розподілу описуються за допомогою ексцесу. Ексцесом випадкової величини називають відношення

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3.$$

Число 3 віднімається з відношення тому, що для нормального розподілу відношення μ_4/σ^4 дорівнює 3. Таким чином, для нормального розподілу ексцес дорівнює нулю.

1.3. Числові характеристики положення (мода, медіана, квантили)

З характеристик положення найважливішу роль відіграє математичне очікування, котре називають просто середнім значенням випадкової величини. Відомо, що при великій кількості дослідів середнє арифметичне значення випадкової величини наближається до її математичного очікування. Властивість стійкості при великому числі дослідів легко перевірити експериментально: в результаті зважування будь-якого зразка на точних вагах кілька разів виходить нове значення, усереднюючи ці значення, отримуємо середнє арифметичне. При подальшому збільшенні числа дослідів (зважувань) середнє арифметичне реагує на збільшення числа дослідів все менше й менше.

На практиці іноді застосовують і інші характеристики становища – моду та медіану.

Модою випадкової величини називається її найбільш ймовірне значення. Для безперервних величин – те значення випадкової величини, при котрому значення щільності ймовірності максимальне.

Часто застосовується характеристика положення – медіана. Використовується зазвичай для безперервних випадкових величин, хоча формально може бути використана й для дискретних.

Медіаною випадкової величини X називається таке її значення Me , для котрого

$$P(X < Me) = P(X > Me).$$

Тобто однаково вірно, чи виявиться випадкова величина менше або більше Me . Геометрично медіана – це абсциса точки, в якій площа, обмежена кривою розподілу, ділиться навпіл.

У разі симетричного розподілу медіана збігається з модою та математичним очікуванням.

Квантіль (термін був вперше використаний Кендаллом у 1940 р.) вибірки являє собою число x_p , нижче котрого знаходиться p -а частина (частка) вибірки.

Наприклад, квантіль $0,25$ для деякої змінної – це таке значення (x_p), нижче котрого знаходиться 25% значень змінної.

Аналогічно квантіль $0,75$ – це таке значення, нижче котрого потрапляють 75% значень вибірки.

Квартілі. Нижня та верхня квартілі, від слова кварта – чверть (термін вперше використав Гальтон у 1882) рівні відповідно 25 -й і 75 -й процентилям розподілу.

25 -а центиль змінної – це значення, нижче котрого розташовуються 25% значень змінної.

Аналогічно, 75 -а центиль дорівнює значенню, нижче котрого розташоване 75% значень змінної.

Отже, 3 точки – нижня квартіль, медіана та верхня квартіль – ділять вибірку на 4 рівні частини.

$1/4$ спостережень лежить між мінімальним значенням і нижньою квартіллю, $1/4$ – між нижньою квартіллю та медіаною, $1/4$ – між медіаною та верхньою квартіллю, $1/4$ – між верхньою квартіллю та максимальним значенням вибірки.

1.4. Типові закони розподілу

Для вивчення основних законів розподілу ймовірностей введемо поняття індикатора випадкової події A – це дискретна випадкова

величина X , яка дорівнює 1 при здійсненні події A і 0 при здійсненні \bar{A} :

$$X = \left\{ \begin{array}{l} 1, \quad A \\ 0, \quad \bar{A} \end{array} \right\}.$$

Ряд розподілу ймовірностей індикатора випадкової події:

x_i	0	1
p_i	q	p

де p – ймовірність здійснення A ;

$q=1-p$ – ймовірність здійснення \bar{A} .

Числові характеристики індикатора випадкової події:

$$m_x=p, D_x=q \cdot p.$$

1.4.1. Геометричний розподіл

Геометричний розподіл має дискретна випадкова величина X , якщо вона приймає значення $0, 1, \dots, \infty$ з ймовірностями:

$$p(X = i) = p_i = q \cdot p,$$

де p – параметр розподілу ($0 \leq p \leq 1$), $q=1-p$.

Числові характеристики геометричного розподілу:

$$m_x=q/p, D_x=q/p^2.$$

Умови виникнення. Проводиться низка однакових незалежних дослідів до першої появи деякої події A . Випадкова величина X – число проведених безуспішних дослідів до першої появи події A .

1.4.2. Біноміальний розподіл

Біноміальний розподіл має дискретна випадкова величина X , якщо вона приймає значення $0, 1, \dots, n$ з такими ймовірностями

$$p(X = i) = p_i = \frac{n!}{i!(n-i)!} p^i \cdot q^{n-i},$$

де n, p – параметри розподілу ($0 \leq p \leq 1$), $q=1-p$.

Числові характеристики біноміального розподілу:

$$m_x=n \cdot p; D_x=n \cdot q \cdot p.$$

Умови виникнення. Проводиться n однакових незалежних випробувань, у кожному з яких подія A з'являється з ймовірністю p . Випадкова величина X – число дослідів, у яких відбулася подія A .

1.4.3. Розподіл Пуассона

Розподіл Пуассона має дискретна випадкова величина X , якщо вона приймає значення $0, 1, \dots, \infty$ з такими ймовірностями:

$$p(X = i) = p_i = \frac{a^i}{i!} e^{-a},$$

де a – параметр розподілу ($a > 0$).

Числові характеристики пуассонової випадкової величини:

$$m_x = a, D_x = a.$$

Умови виникнення:

1. Розподіл Пуассона є граничним випадком біноміального, коли число дослідів n необмежено збільшується, а ймовірність p події A в одному досліді прагне до 0 , так що існує межа $\lim_{\substack{n \rightarrow \infty \\ p \rightarrow 0}} n \cdot p = a$.

2. Випадкова величина X – число подій пуассонового потоку, що з'явилися впродовж інтервалу τ , причому параметр $a = \tau \cdot \lambda$, де λ – інтенсивність потоку.

Розглянемо тимчасову вісь, на якій будемо відзначати моменти виникнення випадкових подій (наприклад, відмови компонентів у складному технічному пристрої, заявки на обслуговування і т. п.). Послідовність таких моментів називається потоком випадкових подій.

Потік випадкових подій називається стаціонарним, якщо число подій, що припадають на інтервал τ , у загальному випадку не залежить від розташування цієї ділянки на тимчасовій осі й визначається тільки його тривалістю, тобто середнє число подій в одиниці часу λ (інтенсивність потоку) постійне.

Потік випадкових подій називається ординарним, якщо ймовірність попадання в деяку ділянку Δt двох і більше випадкових подій значно менша, ніж ймовірність попадання 1-єї події. У потоці відсутня післядія, якщо ймовірність попадання подій на ділянку τ не залежить від того, скільки подій потрапило на інші ділянки, що не перетинаються з даною.

Потік випадкових подій називається *пуассоновим* або найпростішим, якщо він є стаціонарним, ординарним і без післядії.

1.4.4. Рівномірний розподіл

Рівномірний розподіл має безперервна випадкова величина X , якщо її щільність ймовірності в деякому інтервалі $[a; b]$ постійна, тобто якщо всі значення X у цьому інтервалі рівно ймовірні:

$$f(x) = \begin{cases} 0, & x < a; \\ \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & x > b. \end{cases} \quad F(x) = \begin{cases} 0, & x < a; \\ \frac{x-a}{b-a}, & a \leq x \leq b, \\ 1, & x > b. \end{cases}$$

Нижче наведено графік щільності рівномірного розподілу.

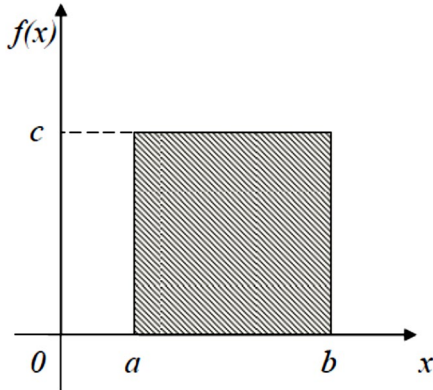


Рис.1.1. Графік щільності ймовірності рівномірного розподілу

Числові характеристики рівномірно розподіленої випадкової величини:

$$m_x = \frac{a+b}{2}, \quad D_x = \frac{(b-a)^2}{12}.$$

При необхідності визначення параметрів a і b за відомими m_x , D_x використовують такі формули:

$$\begin{aligned} a &= m_x + \sigma_x \sqrt{3}, \\ b &= m_x - \sigma_x \sqrt{3}. \end{aligned}$$

Умови виникнення:

1. Випадкова величина X – помилки округлення при обмеженій розрядній сітці:

- округлення до меншого цілого, $X \in [-1; 0]$, $m_x = -0,5$;
- округлення до більшого цілого, $X \in [0; 1]$, $m_x = 0,5$;
- округлення до найближчого цілого, $X \in [-0,5; 0,5]$, $m_x = 0$, де 1 – вага молодшого розряду.

2. Випадкова величина X – похибка зчитування значень з аналогової шкали вимірювального приладу, $X \in [-0,5; 0,5]$, $m_x=0$, де 1 – ціна поділки шкали.

3. Генератори псевдовипадкових величин, наприклад *RANDOM*, вбудовані в мови програмування високого рівня.

1.4.5. Експоненціальний розподіл

Експоненціальний розподіл або **показовий розподіл** має безперервна випадкова величина T , яка приймає тільки позитивні значення, якщо її щільність ймовірності та функція розподілу рівні:

$$f(t) = \begin{cases} \lambda \cdot e^{-\lambda \cdot t}, & t \geq 0, \\ 0, & t < 0. \end{cases} \quad F(t) = \begin{cases} 1 - \lambda \cdot e^{-\lambda \cdot t}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

де λ – параметр розподілу ($\lambda > 0$).

Нижче наведені графіки щільності і функції експоненціального розподілу.

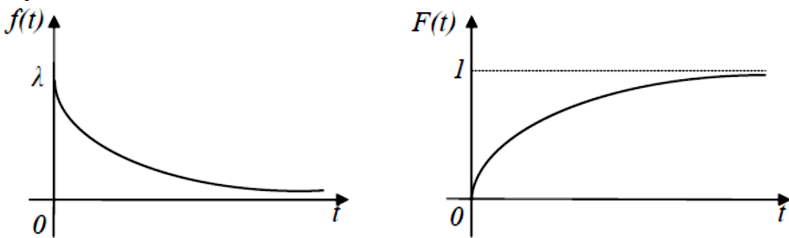


Рис.1.2. Щільність ймовірності і функція експоненціального розподілу

Числові характеристики експоненціальної випадкової величини:

$$m_T = 1/\lambda; D_T = 1/\lambda^2.$$

Умови виникнення. Випадкова величина T – інтервал часу між двома сусідніми подіями в найпростішому чи пуассоновому потоці випадкових подій, причому параметр розподілу λ – інтенсивність потоку.

1.4.6. Нормальний закон розподілу

З усіх вивчених до теперішнього часу випадкових величин, найбільш часто при обробці експериментальних даних дослідники використовують нормальний (Гауса) розподіл. Відзначимо, що згідно з центральною граничною теоремою, котра говорить, що за певних умов

розподіл нормованої суми n незалежних випадкових величин, розподілених за довільним законом, прагне до нормального, якщо n прагне до нескінченності. Умови, при котрих теорема виявляється справедливою, полягають у тому, що різні випадкові величини повинні мати кінцеві дисперсії і дисперсія будь-якої випадкової величини не повинна бути занадто великою в порівнянні з дисперсіями інших.

При обробці статистичних даних ця теорема має дуже велике значення, оскільки відгук є випадковою величиною в результаті впливу неконтрольованих факторів, число яких, у загальному випадку, прямує до нескінченності.

Отже, якщо при плануванні дослідження враховані всі найбільш істотні фактори і потім, при проведенні його, вони контролюються, то при обробці статистичних даних можна припускати, що відгук не повинен суперечити нормальному закону розподілу.

Більшість інших розподілів (Ст'юдента, Фішера, Кохрена, Пірсона та ін.), які використовуються в математичній статистиці, отримані на основі нормального розподілу.

Але з іншого боку, не можна абсолютизувати значення нормального закону. Не всі випадкові величини розподілені за нормальним законом. Але якщо явище піддається багатьом випадковим факторам, то їх сумарний вплив можна описати за допомогою нормального закону.

Для випадкової величини, що не суперечить нормальному закону, функція розподілу задається формулою

$$F(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma_x^2}} \int_{-\infty}^x e^{-\frac{(x-M_x)^2}{2 \cdot \sigma_x^2}} dx,$$

а відповідна їй щільність розподілу має вигляд:

$$f(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma_x^2}} e^{-\frac{(x-M_x)^2}{2 \cdot \sigma_x^2}},$$

і визначається двома параметрами M_x – математичним очікуванням і σ_x^2 – дисперсією. На рис. 1. 3 показаний графік щільності розподілу ймовірності нормального закону – крива розподілу, котра називається

нормальною кривою або кривою Гауса.

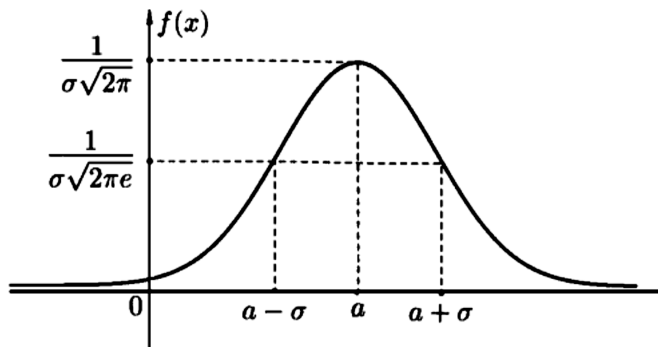


Рис. 1.3. Крива Гауса

Основні властивості нормального закону розподілу:

1. Крива щільності розподілу симетрична відносно значення $M_x = a$, званого центром розподілу.

2. При великих значеннях σ_x^2 крива $f(x)$ пологіша, тобто σ_x^2 є мірою величини розсіювання значення випадкової величини біля значення M_x .

3. Максимум ординати кривої щільності розподілу визначається виразом

$$f_{\max}(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma_x^2}}$$

4. Можна переконалися, що точки $(a - \sigma, \frac{1}{\sigma\sqrt{2\pi e}})$ і $(a + \sigma, \frac{1}{\sigma\sqrt{2\pi e}})$ є точками перегину графіка функції $f(x)$.

5. Для нормального розподілу математичне сподівання, мода та медіана збігаються.

У низці випадків зручно розглядати не саму випадкову величину, а її відхилення від математичного очікування:

$$\dot{X} = X - M_x.$$

Така випадкова величина називається **центрованою**. Відношення випадкової величини до її середнього квадратичного відхилення X/σ_x називається **нормованою** випадковою величиною.

Очевидно, що математичне очікування центрованої випадкової величини дорівнює нулю, а дисперсія нормованої випадкової величини – одиниці.

Приведена (стандартна) випадкова величина – це центрована й нормована випадкова величина

$$Z = \frac{X - M_x}{\sigma_x}$$

Математичне сподівання і дисперсія приведені випадкової величини дорівнюють відповідно, нулю та одиниці. Нормальний розподіл з $M_x=0$ і $\sigma_x^2 = 1$ називається стандартним і позначається $N(0,1)$.

Для наведеної випадкової величини нормальний стандартний розподіл набуває вигляду

$$F(z) = \frac{1}{\sqrt{2 \cdot \pi}} \int_{-\infty}^z e^{-\frac{z^2}{2}} dz$$

На рис. 1. 4 показані функція розподілу та щільності нормального закону розподілу ймовірності.

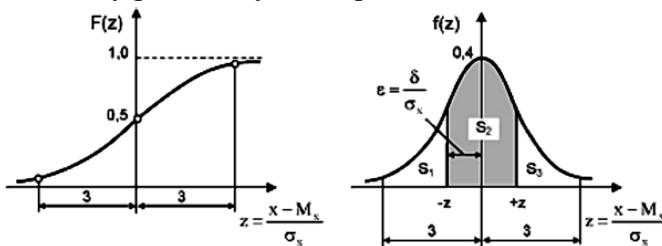


Рис. 1.4. Основні властивості нормального розподілу

Значення нормованої функції нормального розподілу табульовані й наведені в різних підручниках і довідниках. Відзначимо, що іноді замість функції $F(z)$ табулюється функція $\Phi(z)$

$$\Phi(z) = \frac{1}{\sqrt{2 \cdot \pi}} \int_0^z e^{-\frac{z^2}{2}} dz.$$

Так як $\frac{1}{\sqrt{2 \cdot \pi}} \int_{-\infty}^0 e^{-\frac{z^2}{2}} dz = \frac{1}{2}$ (відомий у математиці інтеграл Пуассона), то $F(z) = \frac{1}{2} + \Phi(z)$.

Функція $\Phi(z)$ непарна, тобто $\Phi(-z) = -\Phi(z)$. У низці випадків важливо знати, що випадкова нормально розподілена величина X не відрізнятиметься від свого математичного очікування M_x не більше, ніж на величину ε :

$$(|X - M_x| < \varepsilon) = P(M_x - \varepsilon < X < M_x + \varepsilon).$$

Виразивши ймовірність через щільність ймовірності, отримаємо

$$P(|X - M_x| < \varepsilon) = \int_{-\varepsilon}^{\varepsilon} f(x) dx.$$

Для приведеної випадкової величини

$$P(|z| < \varepsilon) = \Phi(\varepsilon) - \Phi(-\varepsilon) = 2 \cdot \Phi(\varepsilon).$$

$$z = \frac{X - M_x}{\sigma_x},$$

так як

$$P(|X - M_x| < \varepsilon) = P\left(|z| < \frac{\varepsilon}{\sigma_x}\right) = 2 \cdot \Phi\left(\frac{\varepsilon}{\sigma_x}\right).$$

Позначивши $\varepsilon = k \cdot \sigma_x$, отримаємо $P(|X - M_x| < k \cdot \sigma_x) = 2 \cdot \Phi(k)$.

Оскільки $\Phi(k)$ табульована функція, то її значення можна визначити для $k=1, 2, 3$:

$$P(|X - M_x| < \sigma_x) = 2 \cdot \Phi(1) = 0,6826;$$

$$P(|X - M_x| < 2 \cdot \sigma_x) = 2 \cdot \Phi(2) = 0,9544;$$

$$P(|X - M_x| < 3 \cdot \sigma_x) = 2 \cdot \Phi(3) = 0,997.$$

Таким чином, для нормально розподіленої випадкової величини ймовірність того, що вона прийме таке значення, яке не відрізнятиметься від її математичного очікування більш ніж на одне середнє квадратичне відхилення, дорівнює **0,68**. Тобто, при нормальному розподілі приблизно $2/3$ всіх значень випадкової

величини лежить в інтервалі $M_x \pm \sigma_x$, а 95% значень випадкової величини лежать в діапазоні $M_x \pm 2 \cdot \sigma_x$. Інтервалу $M_x \pm 2 \cdot \sigma_x$ відповідає ймовірність 0,95, а інтервалу $M_x \pm 3 \cdot \sigma_x$ – ймовірність 0,997.

Отже, *відмінність будь-якого із значень нормально розподіленої випадкової величини від її математичного очікування не перевищує потрєсного середнього квадратичного відхилення з ймовірністю 0,99. Це властивість нормально розподіленої випадкової величини носить назву «правило трьох сигм».*

1.4.7. Розподіл χ^2 (хі-квадрат)

Розглянемо розподіл деяких випадкових величин, що представляють функції нормальних величин, які використовуються в математичній статистиці.

Нехай випадкова величина Y розподілена за нормальним законом $Y \in N(a, \sigma^2)$. Тоді випадкова величина $U = \frac{Y-a}{\sigma} = \chi$ розподілена за нормальним законом з параметрами $M(U)=0$ і $\sigma(U)=1$, тобто $U \in N(0, 1)$. Квадрат такої стандартизованої випадкової величини

$$U^2 = \left(\frac{Y - a}{\sigma} \right)^2 = \chi^2$$

називається **випадковою величиною χ^2** (хі-квадрат) з одним ступенем свободи.

Розглянемо n незалежних випадкових величин Y_1, Y_2, \dots, Y_n , розподілених за нормальним законом з $M(Y_i)=a_i$ і середніми квадратичними відхиленнями $\sigma_i, i=1, \dots, n$.

Утворимо для кожної з цих випадкових величин стандартизовану випадкову величину

$$U_i = \frac{Y_i - a_i}{\sigma_i}, \quad i = \overline{1, n}.$$

Сума квадратів стандартизованих змінних

$$\chi^2 = U_1^2 + U_2^2 + \dots + U_n^2 = \left(\frac{Y_1 - a_1}{\sigma_1} \right)^2 + \left(\frac{Y_2 - a_2}{\sigma_2} \right)^2 + \dots + \left(\frac{Y_n - a_n}{\sigma_n} \right)^2$$

називається випадковою величиною χ^2 з n ступенями свободи. Щільність розподілу випадкової величини χ^2 має вигляд:

$$f(\chi^2) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} (\chi^2)^{\frac{n}{2}-1} e^{-\frac{\chi^2}{2}}, & \text{якщо } \chi^2 > 0, \\ 0, & \chi^2 \leq 0 \end{cases}$$

де $\Gamma(p) = \int_0^{\infty} t^{p-1} \cdot e^{-t} \cdot dt$ – гамма-функція Ейлера і є узагальненням поняття факторіала: $\Gamma(p) = (p-1)!$ для цілих позитивних p .

Отже, розподіл χ^2 залежить від одного параметра n – числа ступенів свободи. Із зростанням n розподіл χ^2 наближається до нормального закону розподілу (при $n \geq 30$ розподіл χ^2 практично не відрізняється від нормального).

На практиці, як правило, використовуються не $f(\chi^2)$ і $F(\chi^2)$, а квантілі χ^2 -розподілу $\chi_{\alpha,n}^2$. Квантілем $\chi_{\alpha,n}^2$, що відповідає заданому рівню імовірності α , називається таке значення $\chi^2 = \chi_{\alpha,n}^2$, при котрому

$$P(\chi^2 > \chi_{\alpha,n}^2) = \int_{\chi_{\alpha,n}^2}^{\infty} f(\chi^2) d(\chi^2) = \alpha.$$

Знаходження квантіля, з геометричної точки зору, полягає в тому, щоб вибрати таке значення $\chi^2 = \chi_{\alpha,n}^2$, при котрому площа заштрихованої криволінійної трапеції (див. рис. 1. 5) дорівнювала б α .

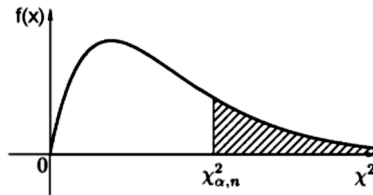


Рис. 1. 5. Знаходження квантіля розподілу Пірсона

1.4.8. Розподіл Стьюдента

Розподіл Стьюдента (t -розподіл) має важливе значення при статистичних обчисленнях, пов'язаних з нормальним законом, а саме тоді, коли середнє відхилення σ невідомо і підлягає визначенню за дослідними даними.

Нехай $Y: Y_1, Y_2, \dots, Y_n$ – незалежні випадкові величини, мають

нормальний розподіл з параметрами $M(Y)=M(Y_j)=0$ і $\sigma_Y=\sigma_{Y_i}=1, i = 1, n$.

Випадкова величина

$$t = \frac{Y}{\sqrt{\frac{1}{n \sum_{i=1}^n Y_i^2}}} = \frac{Y}{\sqrt{\frac{1}{n} \chi_n^2}}$$

є функцією нормально розподілених випадкових величин і називається безрозмірним дробом Стьюдента.

Щільність розподілу випадкової величини t має вигляд:

$$f(t) = S(t, n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi \cdot n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < t < +\infty,$$

де n – число доданків у підкореневому виразі дробу Стьюдента.

З формули видно, що розподіл випадкової величини t залежить тільки від одного параметра – числа ступенів свободи n , рівного числу доданків у підкореневому виразі дробу Стьюдента.

Відомо, що математичне очікування та дисперсія випадкової величини t відповідно дорівнюють

$$M(t) = 0; D(t) = \frac{n}{n-2}; (n > 2).$$

На рис. 1. 6 зображено графік щільності розподілу Стьюдента при різних ступенях свободи. Помічаємо, що при збільшенні числа ступенів свободи n він наближається до кривої Гауса.

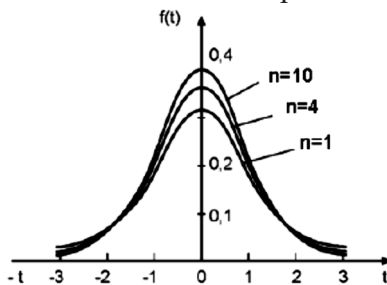


Рис. 1.6. Розподіл Стьюдента

У статистичних розрахунках використовуються квантілі t -

розподілу $t_{\frac{\alpha}{2};n}$. Значення квантилів знаходяться з розв'язання рівняння:

$$P\left(|t| > t_{\frac{\alpha}{2};n}\right) = 2 \int_{\frac{t_{\alpha}{2};n}{2}}^{\infty} f(t) dt = \alpha.$$

З геометричної точки зору, знаходження квантилів $t_{\frac{\alpha}{2};n}$ полягає в тому виборі значення $t=t_{\frac{\alpha}{2};n}$, при котрому сумарна площа заштрихованих на рис. 1. 7 криволінійних трапецій дорівнювала б α .

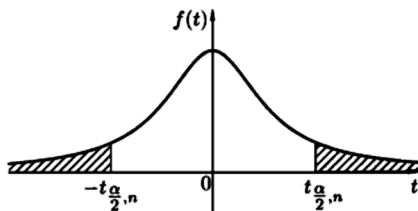


Рис. 1.7. Квантілі розподілу Стюдента

На рис. 1. 8 графічно представлено співвідношення між основними законами розподілу ймовірностей.

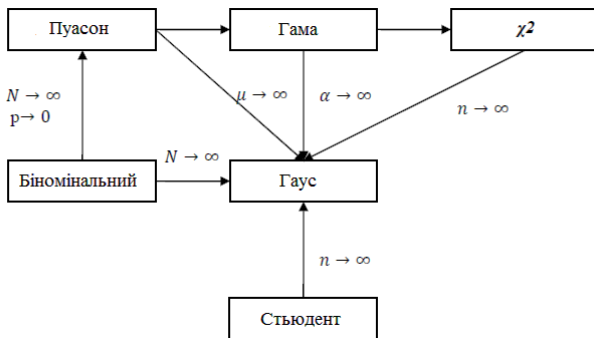


Рис. 1.8. Співвідношення між різними законами розподілу

1.5. Числові характеристики системи випадкових величин (коваріація та кореляція)

Особливу роль при дослідженні системи випадкових величин грає другий змішаний центральний момент, який називається

кореляційним моментом (коваріація). Він зазвичай позначається:

$$k_{x,y} = \mu_{1,1} = \iint_{-\infty}^{+\infty} (x - m_x)(y - m_y)f(x, y)dz.$$

Цей момент, який визначається як математичне очікування добутку відхилень двох випадкових величин від їх математичних очікувань, характеризує взаємний вплив цих випадкових величин. Для оцінки ступеня цього впливу використовують **коефіцієнт кореляції** випадкових величин Y і X :

$$r_{x,y} = \frac{k_{x,y}}{\sigma_x \cdot \sigma_y}.$$

Якщо випадкові величини Y і X незалежні, кореляційний момент і коефіцієнт кореляції дорівнюють нулю. У загальному випадку рівність нулю коефіцієнта кореляції є необхідною, але не достатньою умовою незалежності випадкових величин X і Y .

Якщо є система, що складається з l випадкових величин, можна ввести матрицю кореляційних моментів (коваріаційну матрицю):

$$\begin{matrix} k_{1,1} & k_{1,2} & \dots & k_{1,l} \\ k_{2,1} & k_{2,2} & \dots & k_{2,l} \\ k_{l,1} & k_{l,2} & \dots & k_{l,l} \end{matrix}$$

Оскільки $k_{1,1} = \sigma_1^2$, а з визначення центрального моменту слідує, що $k_{i,j} = k_{j,i}$, тому має місце трикутна матриця

$$\begin{matrix} \sigma_1^2 & k_{1,2} & \dots & k_{1,l} \\ & \sigma_2^2 & \dots & k_{2,l} \\ & & \dots & \sigma_l^2 \end{matrix}$$

Якщо випадкові величини некорельовані, то має місце діагональна матриця, елементами котрої є відповідні дисперсії випадкових величин.

Якщо перейти від кореляційних моментів до коефіцієнтів кореляції, то виходить кореляційна матриця:

$$\begin{matrix} 1 & r_{1,2} & \dots & r_{1,l} \\ & 1 & \dots & r_{2,l} \\ & & \dots & 1 \end{matrix}$$

Кореляційна матриця одна з **найважливіших**

характеристик, які описують систему випадкових величин. На основі кореляційної матриці можна отримати значення множинного коефіцієнта кореляції R , що характеризує статистичну залежність деякої змінної від інших змінних.

1.6. Нормальний розподіл системи випадкових величин

Так само, як і в одновимірному випадку найважливішим законом розподілу є нормальний багатовимірний закон розподілу, для котрого справедливо таке положення: якщо нормально розподілені випадкові величини некорельовані, то вони незалежні. Крім того, показано, що для нормально розподілених випадкових величин рівняння регресії мають вигляд:

$$y = m_y - r_{x,y} \frac{\sigma_y}{\sigma_x} (x - m_x),$$
$$x = m_x - r_{x,y} \frac{\sigma_x}{\sigma_y} (y - m_y).$$

Наведені вище теоретичні положення визначають умови застосовності коефіцієнта кореляції як показника, що дозволяє оцінювати тісноту зв'язку досліджуваних змінних.

Для коректного використання даного показника необхідно, щоб розглянуті змінні являли собою систему випадкових величин, які мають нормальний спільний закон розподілу. Тоді величина парного та множинного коефіцієнта кореляції може трактуватися як показник, який характеризує рівень статистичної лінійної залежності випадкових величин. Для парного коефіцієнта кореляції маємо:

$$-1 \leq r_{x,y} = \frac{k_{x,y}}{\sigma_x \cdot \sigma_y} \leq +1.$$

При $r_{x,y}=1$ змінні пов'язані прямою лінійною залежністю – при $r_{x,y}=-1$ – зворотною лінійною залежністю. Множинний коефіцієнт кореляції $0 \leq R \leq 1$ (0 – лінійна залежність відсутня; 1 – має місце функціональна лінійна залежність).

1.7. Елементи математичної статистики

Математична статистика – розділ математики, що вивчає методи збору, систематизації та обробки результатів спостережень.

Математична статистика вирішує такі завдання:

1. Впорядкування даних, подання їх в зручному для аналізу вигляді.

2. Оцінка цікавлячих нас характеристик випадкової величини, котра спостерігається.

3. Перевірка статистичних гіпотез, тобто вирішення питання узгодження оцінювання з даними дослідження (наприклад, перевірка гіпотези про те, що спостережувана випадкова величина підпорядковується нормальному закону).

Найважливішим завданням статистики є розробка методів, які дозволяють за результатами дослідження вибірки зробити висновки про параметри розподілу всієї сукупності.

1.7.1. Генеральна сукупність і випадкова вибірка

На практиці дослідник має лише обмежений об'єм значень випадкової величини, що представляє собою деяку **вибірку з генеральної сукупності**. Під генеральною сукупністю розуміємо всі допустимі значення випадкової величини. При аналізі неперервної випадкової величини (наприклад, температура, тиск) під значеннями випадкової величини, котра спостерігається, розуміють такі дискретні значення, розділені певним інтервалом часу, при котрому проведені виміри можна вважати незалежними.

Вибірка називається **репрезентативною**, якщо вона дає досить повне уявлення про генеральну сукупність.

У математичній статистиці доведено (теорема Гливінко), що при досить великій вибірці функцію розподілу ймовірностей генеральної сукупності можна замінювати функцією розподілу вибірки.

Числові характеристики, визначені при обмеженому обсязі інформації, називаються **оцінками**.

Іншими словами, на практиці ми завжди маємо справу з **оцінками** числових характеристик випадкових величин. Нехай \tilde{a} є оцінкою параметра a .

До оцінок числових характеристик пред'являються такі вимоги:

1. **Спроможність** – при збільшенні числа дослідів оцінка сходиться за ймовірністю до оцінюваного параметру, тобто виконується умова

$$P(|\tilde{a}_n - a| < \varepsilon) = 1$$

при збільшенні обсягу вибірки n .

2. **Незміщеність** – математичне очікування оцінки одного оцінюваного параметру, тобто при збільшенні обсягу вибірки її математичне очікування прагне до оцінюваного параметру: $M(\tilde{a}_n) = a$ при збільшенні n .

3. **Ефективність** – незміщена оцінка повинна мати мінімальну дисперсію в порівнянні з іншими оцінками, тобто

$$D(\tilde{a}_n) = D_{min}.$$

1.7.2. Точкові оцінки параметрів нормального розподілу

Як відомо, параметрами нормального розподілу є математичне очікування та дисперсія. Як оцінку для математичного очікування природно припустити середнє арифметичне спостережених значень (**вибіркове середнє**), тобто

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

яке виходить з раніше наведеної залежності для математичного очікування, якщо покласти

$$p_i = p = 1/n.$$

У математичній статистиці доведено, що вибіркове середнє є найкращою (змістовною, незміщеною та ефективною) оцінкою математичного очікування випадкової величини, котра підкоряється нормальному закону розподілу.

На перший погляд природною оцінкою для дисперсії $D[X]$ буде

$$s_1^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n},$$

Але ця оцінка виходить дещо зміщеною:

$$M[s_1^2] = \frac{n-1}{n} D[X].$$

Тому для оцінки дисперсії використовується незміщена оцінка:

$$s^2 = s_1^2 \frac{n}{n-1} = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{(n-1)}.$$

Зменшення знаменника на одиницю безпосередньо пов'язано з тим, що величина \bar{x} , щодо якої беруться відхилення, сама залежить від обсягу вибірки. Кожна величина, що залежить від елементів вибірки і входить у формулу, називається **зв'язком**. У статистиці доводиться, що знаменник вибіркової дисперсії завжди дорівнює різниці між обсягом вибірки і числом зв'язків l , накладених на цю вибірку. Ця різниця

$$f=n-l$$

називається **числом ступенів свободи** вибірки. У практичних обчисленнях для оцінки дисперсії часто використовується більш зручна формула:

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \right).$$

Перевага цієї формули в тому, що в ній немає операцій віднімання близьких чисел, що призводять до втрати точності.

1.7.3. Класифікація помилок вимірювання

Кожен результат вимірювання – випадкова величина. Відхилення реального результату від істинного називається **помилкою спостереження**. Помилка спостереження також є випадковою величиною. Вона є результатом впливу неврахованих чинників. Якщо позначити істинний результат через a , помилку – через ΔX , результат вимірювання X , то

$$X-a=\Delta X.$$

Розрізняють помилки трьох видів:

1. Грубі помилки, які виникають внаслідок порушення основних умов дослідження. Результат, який містить грубу помилку, різко відрізняється за величиною від інших вимірів. На цьому засновані деякі критерії щодо виключення грубих помилок.

2. Систематичні помилки постійні у всій серії вимірювань або змінюються за певним законом. Виявлення їх вимагає спеціальних досліджень, але як тільки систематичні помилки виявлені, вони можуть бути усунені шляхом введення відповідних поправок у результати дослідження.

3. Випадкові помилки – це ті помилки дослідження, котрі залишаються після усунення всіх виявлених грубих і систематичних

помилки. Вони викликаються великою кількістю факторів, ефекти впливу яких настільки незначні, що їх не можна виділити окремо (на даному рівні використаних методів і техніки дослідження). При цьому розподіл випадкових помилок симетричний відносно нуля: помилки, протилежні за знаком, але рівні за абсолютною величиною, зустрічаються однаково часто. З симетрії розподілу помилок слідує, що справжній результат спостереження є математичне очікування відповідної випадкової величини. Оскільки $X=a+\Delta X$, і при відсутності грубих і систематичних помилок $M[\Delta X]=0$, то $M[X]=a$.

Надалі будемо розглядати тільки випадкові помилки досліджень.

1.7.4. Закон додавання помилок

Для незалежних випадкових величин властивість адитивності мають дисперсії, а не середньоквадратичні помилки. Якщо X_1, X_2, \dots, X_n – незалежні випадкові величини; a_1, a_2, \dots, a_n – не випадкові величини і

$$Z=a_1 \cdot X_1+a_2 \cdot X_2+\dots+a_n \cdot X_n,$$

то вибіркова дисперсія величини Z визначається таким чином:

$$S_Z^2 = a_1^2 \cdot S_{X_1}^2 + a_2^2 \cdot S_{X_2}^2 + \dots + a_n^2 \cdot S_{X_n}^2.$$

Якщо покласти $a_1=a_2=\dots=a_n$, то

$$Z = \frac{X_1 + X_2 + \dots + X_n}{n} = \bar{X}.$$

У цьому випадку

$$S_{\bar{X}}^2 = \frac{S_{X_1}^2 + S_{X_2}^2 + \dots + S_{X_n}^2}{n^2} = \frac{\bar{S}_X^2}{n},$$

де $\bar{S}_X^2 = \frac{\sum_{i=1}^n S_{X_i}^2}{n}$.

Якщо X_1, X_2, \dots, X_n інтерпретувати як n незалежних спостережень однієї й тієї ж випадкової величини X , то $S_{X_1}^2 = S_{X_2}^2 = \dots = S_{X_n}^2 = S_X^2$, тоді отримаємо

$$S_{\bar{X}}^2 = \frac{S_X^2}{n}.$$

З отриманого виразу впливає один дуже важливий практичний висновок: при оцінці точності двох методів вимірювань слід

враховувати тривалість аналізу. Застосовуючи менш точні методи можна зробити більше число досліджень і отримати більш точні результати, ніж при використанні трудомісткого точного методу. Можна зробити висновок про можливість зменшити похибку остаточного результату при збільшенні кількості n окремих вимірювань. Однак також слід пам'ятати, що підвищення точності ніколи не дається безкоштовно. Так, щоб дізнатися додаткову цифру в \bar{X} , тобто підвищити точність у 10 разів, кількість вимірювань необхідно збільшити в 100 разів!

1.7.5. Помилки непрямих вимірювань

Вимірювання поділяються на прямі та непрямі. У першому випадку безпосередньо вимірюється величина, котра визначається, при непрямих вимірах вона визначається як функція від безпосередньо вимірюваних величин. Нехай між випадковими величинами z і x_1, x_2, \dots, x_n існує відома функціональна залежність:

$$z = f(x_1, x_2, \dots, x_n).$$

Істинне значення величини може не збігатися з математичним очікуванням M_z , а визначається тим же законом:

$$a_z = f(m_{x_1}, m_{x_2}, \dots, m_{x_n}).$$

Величина a_z називається середнім непрямого вимірювання.

Дисперсія непрямого вимірювання σ_z^2 визначається так само, як звичайна дисперсія, тільки відхилення беруться від середнього непрямого вимірювання a_z . Її можна знайти, знаючи дисперсії окремих спостережень і вид функції f . На практиці визначають вибірккові дисперсії $s_{x_i}^2$ і по ним вибірккову дисперсію непрямого вимірювання s_z^2 , яка служить оцінкою генеральної дисперсії σ_z^2 . Щоб знайти s_z^2 , розкладемо функцію $z = f(x_1, x_2, \dots, x_n)$ в ряд Тейлора в точці $(m_{x_1}, m_{x_2}, \dots, m_{x_n})$, обмежившись членами першого порядку:

$$z \approx f(m_{x_1}, m_{x_2}, \dots, m_{x_n}) + \frac{\partial f}{\partial x_1}(x_1 - m_{x_1}) + \frac{\partial f}{\partial x_2}(x_2 - m_{x_2}) + \dots + \frac{\partial f}{\partial x_n}(x_n - m_{x_n}),$$

і визначимо s_z^2 за законом додавання дисперсій:

$$s_z^2 = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 s_{x_i}^2.$$

Отриманий вираз називають **законом накопичення помилок**.

1.8. Довірчі інтервали і довірча ймовірність

Точкові оцінки мають той недолік, що за ними не можна судити про точність отриманих оцінок. Тому виникає задача визначення на підставі вибірових значень такого інтервалу, який покривав би невідомі значення параметру із заданою ймовірністю.

На відміну від точкової оцінки, інтервальна оцінка дозволяє отримати вірогідну характеристику точності оцінюваного параметру.

Вибіркові параметри є випадковими величинами, їх відхилення від генеральних (тобто похибки їх визначення) також будуть випадковими. Оцінка цих відхилень носить імовірнісний характер – можна лише вказати ймовірність тієї чи іншої похибки. Для цього в математичній статистиці користуються **довірчими інтервалами** та **довірчими можливостями**.

Довірчий інтервал – інтервал, який із заданою ймовірністю накріє невідоме значення оцінюваного параметра розподілу.

Довірча ймовірність – ймовірність того, що довірчий інтервал накріє дійсне значення параметра, що оцінюється за вибіровими даними.

Оцінювання за допомогою довірчого інтервалу – спосіб оцінки, при котрому із заданою довірчою ймовірністю встановлюють межі довірчого інтервалу.

Нехай для генерального параметру a отримана з досліду незміщена оцінка a^* . Потрібно оцінити можливу при цьому помилку. Призначимо досить велику ймовірність β – таку, що подію з цією ймовірністю можна вважати практично достовірною, і знайдемо таке значення ε , для якого

$$P(|a^* - a| \leq \varepsilon) = \beta. \quad (1.1)$$

При цьому діапазон практично можливих значень помилки, що виникає при заміні a на a^* , буде $\pm\varepsilon$, більші за абсолютною величиною помилки будуть з'являтися тільки з малою вірогідністю

$$\alpha = 1 - \beta,$$

званої рівнем значущості або ризиком. Рівень значущості часто виражають у відсотках. Інакше формулу (1. 1) можна інтерпретувати як імовірність того, що справжнє значення параметру a лежить у межах

$$a^* - \varepsilon \leq a \leq a^* + \varepsilon.$$

Імовірність β називається **довірчою ймовірністю**, **довірчим рівнем** або **надійністю**, тому що вона характеризує надійність отриманої оцінки.

Інтервал $I_\beta = a^* \pm \varepsilon$ називається **довірчим інтервалом**. Межі інтервалу $a' = a^* - \varepsilon$ і $a'' = a^* + \varepsilon$ – довірчими межами. **Довірчий інтервал при даній довірчій ймовірності визначає точність оцінки параметру.**

При цьому зазначимо таке. Раніше ми розглядали ймовірність попадання випадкової величини на заданий (невипадковий) інтервал. У даному випадку справа інша: величина a не випадкова, зате випадковий інтервал I_β . Випадкове його становище на числовій прямій, яке визначається його центром a^* , випадкова й довжина інтервалу $2 \cdot \varepsilon$, так як величина ε обчислюється, як правило, за дослідними даними, тобто за результатами дослідження. Тому в даному випадку зручно тлумачити інтервал I як ймовірність того, що випадковий інтервал I_β накріє величину a .

Величина довірчого інтервалу залежить від довірчої ймовірності, з якою гарантується перебування параметру всередині довірчого інтервалу: чим більше величина β , тим більше і ε (тобто з чим більшою ймовірністю ми хочемо гарантувати отриманий результат, тим у більшому інтервалі він повинен перебувати).

Збільшення числа дослідів проявляється в скороченні довірчого інтервалу при постійній довірчій ймовірності чи в підвищенні довірчої ймовірності при збереженні довірчого інтервалу.

Зазвичай на практиці фіксується на певному рівні значення довірчої ймовірності ($0,9$; $0,95$; $0,99$; $0,999$). Виходячи з цього значення, визначають довірчий інтервал результату I_β .

При побудові довірчого інтервалу вирішується задача про абсолютне відхилення:

$$p(|a^* - a| \leq \varepsilon) = p(\Delta a \leq \varepsilon) = F(\varepsilon) - F(-\varepsilon) = \int_{-\varepsilon}^{+\varepsilon} f(x) dx = \beta. \quad (1.2)$$

Таким чином, якщо відомий закон розподілу оцінки a^* , то задача визначення довірчого інтервалу вирішується досить просто.

Розглянемо побудову довірчого інтервалу для математичного очікування нормально розподіленої випадкової величини з відомим генеральним стандартом σ_x .

Поняття генерального стандарту тісно пов'язане з поняттям точності приладу. **Клас точності приладу** – це виражена в відсотках відносна гранична похибка вимірювання величини, що дорівнює межі вимірювання приладу. У вимірювальній техніці в більшості галузей промисловості під граничною похибкою розуміється величина, котра дорівнює двом середньоквадратичним відхиленням.

Нехай ϵ вибірка обсягу n значень випадкової величини. Оцінку m_x ϵ середнє вибірки:

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

Для побудови довірчого інтервалу необхідно знати розподіл цієї оцінки. Для вибірок з генеральної сукупності, розподіленої нормально можна показати, що \bar{x} також має нормальний розподіл з математичним очікуванням m_x і середнім квадратичним відхиленням $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$. Тоді

$$P(|\bar{x} - m_x| \leq \epsilon) = \beta = 2\Phi\left(\frac{\epsilon}{\sigma_{\bar{x}}}\right). \quad (1.3)$$

Задавши собі довірчу ймовірність, визначимо за таблицею значення функції Лапласа $k_\beta = \frac{\epsilon\beta}{\sigma_{\bar{x}}}$. Тоді довірчий інтервал для математичного очікування матиме вигляд

$$\bar{x} - k_\beta \cdot \sigma_{\bar{x}} \leq m_x \leq \bar{x} + k_\beta \cdot \sigma_{\bar{x}} \text{ або } \bar{x} - k_\beta \frac{\sigma_x}{\sqrt{n}} \leq m_x \leq \bar{x} + k_\beta \frac{\sigma_x}{\sqrt{n}}.$$

З оцінки видно, що зменшення довірчого інтервалу обернено пропорційне квадратному кореню з числа спостережень. Отже, якщо треба зменшити можливу помилку в два рази, треба збільшити число спостережень у 4 рази.

Якщо закон розподілу оцінки невідомий, то в математичній статистиці застосовують зазвичай два методи:

1) наближений – при n більше 50 замінюють невідомі параметри їх оцінками;

2) від випадкової величини a^* переходять до іншої випадкової величини, закон розподілу котрої не залежить від оцінюваного параметру a , а залежить тільки від обсягу вибірки n і від виду розподілу величини X . Такого роду величини найбільш детально вивчені для нормального закону. За довірчі межі приймають симетричні квантілі

$$a_{(1-\beta)/2}^* \leq a \leq a_{(1+\beta)/2}^*$$

Якщо виразити через p ,

$$a_{p/2}^* \leq a \leq a_{1-p/2}^*$$

На практиці, як правило, число вимірів кінцеве і не перевищує **10...30**. При малому числі вимірів фактична дисперсія σ_x^2 невідома, тому для побудови довірчого інтервалу математичного очікування використовують вибіркову дисперсію S_x^2 і приведену випадкову величину:

$$t = \frac{\bar{x} - m_x}{s_x / \sqrt{n}}$$

t – випадкова величина, що має розподіл, відмінний від нормального, залежна від числа ступенів свободи (t – розподіл або розподіл Стьюдента). При великих значеннях n розподіл Стьюдента наближається до стандартного нормального розподілу. І, за аналогією, отримуємо побудову довірчого інтервалу

$$\bar{x} - \frac{t_{\alpha,m} \cdot S_x}{\sqrt{n}} \leq m_x \leq \bar{x} + \frac{t_{\alpha,m} \cdot S_x}{\sqrt{n}}$$

1.9. Визначення необхідної кількості спостережень

Необхідна кількість спостережень (зразків, проб і т. д.) n можна визначити заздалегідь у тому випадку, коли відомо дійсне значення середньоквадратичного відхилення, а експериментальні дані підкоряються нормальному закону розподілу.

Дійсно, при цих припущеннях число вимірювань можна визначити з системи нерівностей:

$$\bar{x} - k_\beta \frac{\sigma_{\bar{x}}}{\sqrt{n}} \leq m_x \leq \bar{x} + k_\beta \frac{\sigma_{\bar{x}}}{\sqrt{n}}$$

Аналізуючи формулу довірчого інтервалу, можна помітити, що:

а) збільшення обсягу вибірки n призводить до зменшення довжини довірчого інтервалу;

б) збільшення довірчої ймовірності β призводить до збільшення довжини довірчого інтервалу, тобто до зменшення точності $\varepsilon = k_\beta \frac{\sigma_x}{\sqrt{n}}$;

в) якщо задати точність ε і довірчу ймовірність β , то з співвідношення $\varepsilon = k_\beta \frac{\sigma_x}{\sqrt{n}}$ можна знайти мінімальний обсяг вибірки, який забезпечує задану точність.

Однак, у дослідженні значення σ_x оцінюють, виходячи з кінцевого числа вимірів, кількість яких звичайно не перевищує $5 \div 10$. Тому точність оцінювання σ_x невелика. Це вносить додаткову невизначеність в остаточний результат. Щоб її врахувати, необхідно розширити межі довірчого інтервалу, заданого для точно відомої величини σ_x . Зрозуміло, що меншій кількості окремих вимірювань повинен відповідати більш широкий довірчий інтервал. Тому на практиці використовується формула

$$\bar{x} - t_{\alpha,m} \cdot S_{\bar{x}} \leq m_x \leq \bar{x} + t_{\alpha,m} \cdot S_{\bar{x}},$$

де $t_{\alpha,m}$ – квантіль розподілу Стьюдента, який визначається рівнем значущості α і кількістю ступенів свободи $m=n-1$.

1.10. Перевірка статистичних гіпотез

Перевірка статистичних гіпотез є однією з основних задач математичної статистики. Суть цієї задачі полягає в тому, що на підставі вибірових даних має бути прийнято (чи відкинуто) деяке припущення (**статистична гіпотеза**) щодо генеральної сукупності.

Процедура зіставлення гіпотези з вибіровими даними називається **перевіркою гіпотези**.

Задача статистичної перевірки ставиться в такому вигляді: щодо деякої генеральної сукупності висловлюється та чи інша гіпотеза H . З генеральної сукупності формується вибірка. Потрібно вказати правило, за допомогою котрого можна було б за вибіркою вирішити питання, чи слід прийняти гіпотезу H , або відхилити її.

Висунута гіпотеза може бути правильною чи неправильною, тому виникає задача її перевірки.

Під **статистичною гіпотезою** розуміють всяке висловлювання

про генеральну сукупність, яке перевіряється за вибіркою.

Статистичні гіпотези діляться на **параметричні** (гіпотези про параметри розподілу) і **непараметричні** (про вид невідомого розподілу).

Одну з гіпотез висувають як основну H_0 , а іншу, що є логічним запереченням H_0 , тобто протилежну H_0 , – як конкуруючу (альтернативну) й позначають H_1 .

Маючи дві гіпотези H_0 і H_1 треба на основі вибірки X_1, X_2, \dots, X_n прийняти основну гіпотезу H_0 , або конкуруючу H_1 .

Правило, за яким приймається рішення прийняти чи відхилити гіпотезу, називають **статистичним критерієм** перевірки гіпотези.

Для перевірки гіпотези на підставі вибірки формують функцію вибірки $T_n = T(X_1, X_2, \dots, X_n)$, яка називається **статистикою критерію**.

Основний принцип перевірки гіпотез полягає в такому. Безліч можливих значень статистики критерію T_n розбивається на дві непересічних підмножини: **критичну область S** , тобто область відхилення гіпотези H_0 і область \bar{S} прийняття цієї гіпотези. Якщо фактично отримане за вибіркою значення статистики критерію потрапляє в критичну область, то основна гіпотеза H_0 відхиляється, і приймається альтернативна гіпотеза H_1 . Якщо значення критерію потрапляє в \bar{S} , то приймається H_0 , H_1 відхиляється.

При перевірці гіпотези можуть бути допущені помилки двох типів.

Помилка першого роду полягає в тому, що відкидається нульова гіпотеза, коли насправді вона правильна.

Помилка другого роду полягає в тому, що відкидається альтернативна гіпотеза H_1 , коли насправді вона правильна.

Імовірність помилки першого роду (позначається α) називається **рівнем значущості критерію**:

$$\alpha = P(H_1 | H_0).$$

Чим менше α , тим менше ймовірність відхилити вірну гіпотезу. Припустиму похибку першого роду зазвичай задають заздалегідь. Зазвичай для α використовують стандартні значення $\alpha=0,05$; $0,01$; $0,005$; $0,001$.

Ймовірність помилки другого роду (позначається β):

$$\beta = P(H_1|H_0).$$

Величину $1-\beta$, тобто ймовірність недопущення помилки другого роду (відкинути хибну гіпотезу H_0 , прийняти вірну H_1), називають **потужністю критерію**.

Чим більше потужність критерію, тим менше ймовірність помилки другого роду.

Наслідки помилок 1 -го, 2 -го роду зовсім різні:

- стосовно радіолокації кажуть, що α – ймовірність пропуску сигналу, β – ймовірність помилкової тривоги;

- стосовно виробництва – α – ризик постачальника (тобто бракування за вибіркою всієї партії виробів, які задовольняють стандарту), β – ризик споживача (тобто прийом за вибіркою всієї партії виробів, які не задовольняють стандарту);

- стосовно судової практики, помилка 1 -го роду призводить до виправдання винного, помилка 2 -го роду – осуду невинного.

Відзначимо, що одночасне зменшення помилок 1 -го та 2 -го роду можливо лише при збільшенні обсягу вибірок. Тому зазвичай при заданому рівні значущості відшукується критерій з найбільшою потужністю.

1.10.1. Відсів грубих похибок спостережень

У разі відсіву грубих похибок (помілок) нульова гіпотеза формулюється так:

H_0 : «Серед результатів спостережень (вибіркових, дослідних даних) немає значень, які різко виділяються (аномальні)».

Альтернативною гіпотезою може бути:

Або H_1 : «серед результатів спостережень є тільки одна груба помилка»,

Або H_1 : «серед результатів спостережень є дві чи більше грубих помилки».

Критерій Н.В. Смирнова

Якщо відомо, що є тільки одне аномальне значення, то воно буде крайнім членом варіаційного ряду (тобто ряду спостережень, розташованих в зростаючій послідовності: $x_1 \leq x_2 \leq \dots \leq x_n$). Тому

перевіряти вибірку на наявність однієї грубої помилки природно за допомогою статистики

$$u_1 = \frac{\bar{x} - x_1}{s_x}, \quad (1.4)$$

якщо сумнівні викликає перший член варіаційного ряду $x_1 = \min x_i$, або

$$u_1 = \frac{x_n - \bar{x}}{s_x}. \quad (1.5)$$

якщо сумнівний максимальний член варіаційного ряду $x_n = \max x_i$.

Н.В. Смирновим досліджувався розподіл зазначених статистик (1.4) і (1.5) і були складені таблиці точок $u_{\alpha,n}$ (квантілі порядку $p=1-\alpha$ для $\alpha=0,1; 0,05; 0,01$ при обсязі вибірки від 3 до 20 дослідів).

При обраному рівні значущості α критична область для критерію Н.В. Смирнова будується таким чином:

$$u_1 > u_{\alpha,n} \text{ або } u_n > u_{\alpha,n}$$

$u_{\alpha,n}$ – це табличне значення.

У разі, коли виконується умова (статистика потрапляє в критичну область), то нульова гіпотеза відхиляється, тобто викид x_1 або x_n не характерний для даної вибірки, після чого значення x_1 або x_n виключають з розгляду, а знайдені раніше оцінки піддаються корегуванню з урахуванням відкинутого результату.

1.10.2. Порівняння двох рядів спостережень

При проведенні та аналізі результатів експериментальних досліджень часто доводиться порівнювати дві партії виробів, показання двох або декількох приладів, аналізувати результати роботи однотипних установ, порівнювати результати проб матеріалів і т. д. Ось деякі приклади подібних ситуацій:

1. Необхідно порівняти показання двох приладів, що вимірюють одну й ту ж величину, коли цими засобами отримано два ряди спостережень даної величини. Чи однакова точність вимірювання одного й того ж технологічного параметру різними приладами.

2. Потрібно повірити робочий засіб вимірювання (тобто перевірити, чи виходить похибка приладу за межі регламентованих значень) за допомогою зразкового засобу виміру. Чи дорівнює математичне очікування показань приладу дійсним значенням вимірюваного параметру?

3. Два підприємства випускають одну й ту ж продукцію. Необхідно зробити висновок про те, яке з них краще чи гірше в будь-якому сенсі. Рішення подібних задач здійснюється з використанням апарату перевірки статистичних гіпотез.

1.10.3. Перевірка однорідності дисперсій

Таку операцію доводиться виконувати, коли зіставляються результати декількох вибірок. Величина розсіювання характеризує такі виключно важливі показники, як точність машин, приладів, стабільність продаж, якість готової продукції і т. д. Тому, наприклад, про переваги тієї чи іншої технології чи про якість продукції, що випускається, можна зробити висновок у результаті порівняння дисперсій тих параметрів, які їх характеризують.

Для розв'язання задач такого типу потрібно встановити, чи значимо відрізняються вибіркові дисперсії $S_1^2 \neq S_2^2$ із ступенями свободи m_1 і m_2 чи вони характеризують вибірки, взяті з однієї й тієї ж генеральної сукупності чи з генеральних сукупностей з однаковими дисперсіями. В цьому випадку нульова гіпотеза формулюється так: між двома дисперсіями відмінності немає при заданому рівні значимості α ($\sigma_1^2 = \sigma_2^2 = \sigma^2$).

Для перевірки цієї гіпотези використовується **критерій Фішера**, котрий залежить від числа ступенів свободи m_1 і m_2 .

Оскільки для перевірки нуль-гіпотези $\sigma_1^2 = \sigma_2^2$, тобто потрібно перевірити, що дві вибірки належать одній і тій же генеральній сукупності, то вираз можна представити як відношення вибірових дисперсій

$$F = \frac{S_1^2}{S_2^2},$$

де $S_1^2 > S_2^2$.

Очевидно, що F завжди більше одиниці. вибирається рівень значущості α . Нульову гіпотезу приймають, якщо $F \leq F_{\alpha, m_1, m_2}$. F_{α, m_1, m_2} визначається за таблицями квантилів F -розподілу Фішера для числа ступенів свободи $m_1 = n_1 - 1$ і $m_2 = n_2 - 1$ і рівня значущості.

1.10.4. Перевірка однорідності декількох дисперсій

Критерій Фішера використовується для порівняння тільки двох дисперсій, однак на практиці доводиться порівнювати між собою три й більше дисперсій.

При зіставленні дисперсій низки вибірок нульова гіпотеза полягає в тому, що k сукупностей, з яких взяті вибірки, мають рівні дисперсії. Тобто перевірці підлягає припущення, що всі емпіричні дисперсії $S_1^2, S_2^2, \dots, S_k^2$ відносяться до вибірок з сукупності з однією й тією ж генеральною дисперсією σ^2 .

Нехай серед вибіркових дисперсій виявлена така, котра значно більше від усіх інших S_{max}^2 . Завдання полягає в тому, щоб з'ясувати, чи можна вважати відмінність виділеної дисперсії S_{max}^2 істотною. Альтернативна гіпотеза може бути обрана як $H_1: \sigma_{max}^2 > \sigma^2$.

При рівному обсязі вибірок для всіх k вибірок може бути використаний **критерій Кохрена**. Статистика критерію Кохрена G розраховується як

$$G = \frac{S_{max}^2}{\sum_{i=1}^k S_i^2}.$$

Далі для обраного рівня значущості α визначається табличне значення цього критерію, котрий залежить від числа ступенів свободи $m=n-1$ і числа порівнюваних дисперсій k : $G_{\alpha; m; k}$. Критична область будується як $G \geq G_{\alpha; m; k}$. При $G < G_{\alpha; m; k}$ – нульова гіпотеза приймається, тобто відмінність виділеної дисперсії вважається несуттєвою.

У разі підтвердження однорідності дисперсій можна зробити оцінку узагальненої дисперсії σ^2 :

$$S^2 = \frac{\sum_{i=1}^k S_i^2}{k}.$$

Критерій Кохрена використовується тільки в тих випадках, коли всі порівнювані дисперсії мають однакове число ступенів свободи (однакові обсяги вибірок). Якщо ж число вимірювань у різних серіях неоднаково, то для перевірки однорідності дисперсій зазвичай вибирається **критерій Бартлета**. Введемо позначення для загального числа ступенів свободи: $f=f_1+f_2+\dots+f_n$ і середньозваженої дисперсії:

$S_y^2 = \frac{S_1^2 \cdot f_1 + \dots + S_n^2 \cdot f_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum_{i=1}^n S_i^2 \cdot f_i}{f}$. Бартлет показав, що в умовах нульової гіпотези відношення $\frac{B}{C}$, де

$$B = 2.303 \left(f \cdot \lg(S_y^2) - \sum_{i=1}^n f_i \cdot \lg(S_i^2) \right),$$

$$C = 1 + \frac{1}{3(n-1)} \left(\sum_{i=1}^n \frac{1}{f_i} - \frac{1}{f} \right)$$

розподілено приблизно як χ^2 з $n-1$ ступенями свободи, якщо всі $f_i > 2$.

Гіпотеза рівності генеральної дисперсії приймається, якщо

$$\frac{B}{C} \leq \chi_\alpha^2$$

при обраному рівні значущості α .

У цьому випадку відмінності між вибірковими дисперсіями можна вважати незначними, а самі вибіркові дисперсії однорідними.

Так як $C > 1$, якщо $B \leq \chi_\alpha^2$, то нульову гіпотезу слід прийняти. Якщо $B > \chi_\alpha^2$, то критерій Бартлета обчислюють повністю.

1.10.5. Перевірка гіпотез про числові значення математичних очікувань

Для вирішення питання про відповідність виробленої продукції певним вимогам (наприклад, *ДЕСТ* або *ТУ*), або виявленні переваг нової розробки в порівнянні з існуючими аналогами, виникає необхідність за вибірковими середніми значеннями досліджуваних випадкових величин робити висновок про відповідні їм генеральні значення математичних очікувань.

При цьому може виникнути задача (1) порівняння невідомого математичного очікування M_1 , для котрого отримана оцінка через вибіркове середнє \bar{x}_1 з конкретним числовим значенням M (наприклад, з відомим математичним очікуванням) або задача (2) порівняння двох математичних очікувань M_1 і M_2 , які оцінені за двома вибірковими середніми \bar{x}_1 і \bar{x}_2 .

У першому випадку як нульова гіпотеза висувається

припущення про те, що оцінене математичне очікування M_1 дорівнює відомому математичному очікуванню M ($H: M_1=M$). Як альтернативну приймемо $H_1: M_1 \neq M$.

Якщо генеральна дисперсія σ^2 невідома й для неї зроблена оцінка S^2 , то використовується t -критерій (розподілу Стюдента). t -статистика має вигляд: $t = \frac{\bar{x}-M}{S} \sqrt{n}$. Як і при побудові довірчого інтервалу для математичного очікування, вибирається рівень значущості α . Для числа ступенів свободи $m=n-1$ (з яким зроблена оцінка дисперсії) встановлюються межі критичної області за табличними значеннями квантілів t -розподілу. Нульову гіпотезу приймають, тобто вважають, що $M_1=M$ при виконанні нерівності: $|t| \leq t_{\alpha, m}$.

У задачі (2), де порівнюються два невідомих математичних очікування M_1 і M_2 , перш за все необхідно переконатися, що досліджувані вибірки незалежні між собою. Після чого, для двох нормально розподілених генеральних сукупностей з невідомими параметрами M_1, σ_1^2 і M_2, σ_2^2 , які характеризуються незалежними вибірками обсягом, відповідно, n_1 і n_2 , для порівняння вибірових середніх \bar{x}_1 і \bar{x}_2 висувається нульова гіпотеза про рівність математичних очікувань: $H: M_1=M_2$. Альтернативну можемо сформулювати як $H_1: M_1 \neq M_2$.

Як і в попередній задачі, використовуємо t -критерій. Вид t -статистики залежить від того, рівні $\sigma_1^2 = \sigma_2^2 = \sigma^2$, або не рівні $\sigma_1^2 \neq \sigma_2^2$ між собою генеральні дисперсії (для відповіді на це питання можна скористатися критерієм Фішера).

У першому випадку (коли дисперсії не мають значущої відмінності) статистика набирає вигляду $t = \frac{\bar{x}_1 - \bar{x}_2}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ – двох вибіровий t -критерій з рівними дисперсіями, де S – узагальнене середнє квадратичне відхилення.

У другому випадку, коли дисперсії значимо відрізняються одна від другої, статистика має вигляд: $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1}{n_1} + \frac{S_2}{n_2}}}$ – двох вибіровий t -

критерій з нерівними дисперсіями.

Залежно від умови розв'язуваної задачі вибирається необхідний рівень значущості α . Межі критичної області встановлюються за табличними значеннями квантилів t -розподілу. При цьому число ступенів свободи розраховується як $m=n_1+n_2-2$.

Нульову гіпотезу приймають при виконанні нерівності $|t| \leq t_{\alpha, m}$.

1.11. Критерій злагоди. Перевірка гіпотез про вид функції розподілу

Розглянуті раніше методи перевірки статистичних гіпотез виконувалися в припущенні, що відома функція розподілу (нормальний закон). Однак у більшості випадків вид закону вимагає статистичного підтвердження.

Найбільш простим і досить наближеним методом перевірки злагоди результатів експерименту з тим чи іншим законом розподілу є графічний метод. Він полягає в оцінці емпіричної функції розподілу й зіставленні її з функцією передбачуваного теоретичного закону. Якщо побудовані експериментальні точки лежать поблизу теоретичного графіку, то можна вважати, що отримані в досліді данні не суперечать заданому теоретичному закону розподілу. Графічний метод є в значній мірі суб'єктивним і використовується на практиці як перше наближення при розв'язанні подібних задач.

Більш об'єктивні методи встановлення виду розподілу випадкової величини будуються на апараті перевірки статистичних гіпотез – критеріях злагоди.

Нульова гіпотеза в даному випадку полягає в тому, що H_0 : – досліджувана генеральна сукупність не суперечить пропонованому теоретичному закону розподілу. При цьому альтернативна гіпотеза зазвичай формулюється як H_1 : випадкова величина має будь-яке інше розподіл, відмінне від пропонованого.

Порівняння експериментального матеріалу з деяким видом теоретичного розподілу здійснюється за допомогою різних критеріїв злагоди: хі-квадрат (Пірсона), Колмогорова-Смірнова та ін.

1.11.1. Критерій Пірсона

Використовується для перевірки узгодженості розподілів,

отриманих за вибіркою з деякою теоретичною щільністю розподілу.

Для стандартного нормального розподілу теоретична ймовірність попадання випадкової величини в інтервал $\Delta z = z_{i+1} - z_i$ визначається за формулою

$$P_i^* = F(z_{i+1}) - F(z_i) = \frac{1}{2 \cdot \pi} \int_{z_i}^{z_{i+1}} e^{-u^2/2} du.$$

Відмінність оцінки закону розподілу від теоретичного закону розподілу P^* можна охарактеризувати величиною

$$\chi^2 = \sum_{i=1}^n C_i (P_i - P_i^*)^2,$$

де P_i, P_i^* – оцінка та теоретична ймовірність випадкової величини для i -ого інтервалу;

C_i – вагові коефіцієнти, які з великою вагою враховують відхилення для менших P_i .

Пірсон вибрав вагові коефіцієнти таким чином:

$$C_i = \frac{n}{P_i^*}.$$

Пірсон показав, що при такому виборі C_i закон розподілу слабо залежить від n і $P(x)$, а визначається в основному числом розрядів k (кількістю інтервалів).

Отже,

$$\chi^2 = n \sum_{i=1}^k \frac{(P_i - P_i^*)^2}{P_i^*} = n \sum_{i=1}^k \frac{\left(\frac{\bar{m}_i}{n} - P_i^*\right)^2}{P_i^*} = \sum_{i=1}^k \frac{(\bar{m}_i - n \cdot P_i^*)^2}{n \cdot P_i^*}.$$

Очевидно, що при ідеальному відповідно експериментальних даних нормальному закону, експериментальне значення критерію Пірсона дорівнюватиме нулю, тому що $P_i = P_i^*$.

Алгоритм використання критерію Пірсона полягає в такому:

1. Висувається нуль-гіпотеза: «Відмінність експериментальних даних від нормального закону розподілу не суттєва» і альтернативна

гіпотеза: «Відмінність експериментальних даних від нормального закону розподілу істотна, тобто експериментальні дані не підкоряються закону нормального розподілу».

2. За результатами експериментальних вимірювань і припущенням нормального закону їх розподілу визначається розрахункове значення критерію Пірсона.

3. Визначають число ступенів свободи m , задаються рівнем значущості α і визначають теоретичне значення критерію Пірсона $\chi^2_{\alpha; m}$.

4. Якщо $\chi^2 < \chi^2_{\alpha; m}$, то нуль-гіпотеза про нормальний закон розподілу експериментальних даних приймається з довірчою ймовірністю $p=1-\alpha$. В іншому випадку нуль-гіпотеза відкидається і приймається альтернативна.

1.11.2. Критерій Колмогорова

Поряд з критерієм злагоди Пірсона застосовуються й інші критерії, наприклад, Колмогорова, Романовського та ін.

Колмогоров довів, що незалежно від функції розподілу ймовірностей при необмеженому зростанні числа незалежних спостережень вірогідність нерівності

$$D\sqrt{n} \geq \lambda$$

прагне до межі

$$P(\lambda) = 1 - \sum_{k=-\infty}^{\infty} (-1)^k e^{-2 \cdot k^2 \cdot \lambda^2}.$$

Значення цієї ймовірності табульовані.

Суть критерію злагоди Колмогорова полягає в такому. Встановлюється максимальна величина модуля різниці між статистичною та теоретичною функціями розподілу ймовірностей

$$D = |F^0(x) - F(x)|_{max}$$

і визначається величина

$$\lambda = D\sqrt{n},$$

де n – число незалежних спостережень, і за таблицею знаходяться ймовірності $P(\lambda)$.

Величина цієї ймовірності $P(\lambda)$ свідчить про те, що за рахунок випадкових причин ймовірність максимальної розбіжності між функціями розподілу буде не менше $P(\lambda)$.

Якщо ймовірність мала, гіпотезу слід відкинути, при великих значеннях ймовірності цю гіпотезу слід вважати як таку, що не суперечить дослідним даним.

1.11.3. Критерій однорідності статистичного матеріалу

Критерій однорідності ще носить назву **критерій приналежності** вибірок до однієї генеральної сукупності.

Суть цього критерію зводиться до такого. У практиці обробки результатів спостережень не завжди ці результати отримані в одному експерименті. Однорідні результати, тобто результати однієї фізичної величини можуть бути отримані при проведенні різних експериментів i , може бути, навіть у різних умовах. І завдання зводиться до вирішення питання, чи є ці результати **однорідними** й чи можна їх обробляти спільно? Якщо це відобразити візуально, то отримаємо картину, показану на рис. 1. 9. Якщо говорити мовою теорії множини, задача зводиться до встановлення критерію, за яким можна встановити, чи належать підмножини $B_i, i=1, 2, \dots, n$ одній і тій же множині, котра називається генеральною сукупністю.

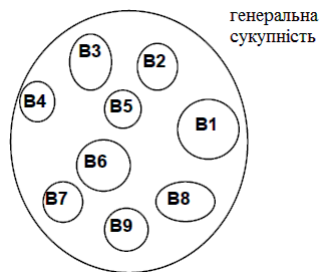


Рис. 1.9. До визначення критерію однорідності

Тепер перейдемо до математичної постановки задачі. Припустимо, що проведено s послідовних експериментів, які складаються відповідно з n_1, n_2, \dots, n_s одиничних спостережень. При цьому числа n_i не випадкові, а розглядаються як задані. У кожному експерименті спостерігається деяка змінна ознака, і результати кожної

низки спостережень розбиваються за значенням цієї ознаки на r груп. Кількість результатів спостережень в i -ій групі j -го ряду позначимо v_{ij} . Тоді отримані дані можуть бути розташовані в таблиці виду:

Таблиця 1.1

Ознака	Ряд 1 2 3 ... s	Сума
1	$v_{11} v_{12} v_{13} \dots v_{1s}$	$v_{1@}$
2	$v_{21} v_{22} v_{23} \dots v_{2s}$	$v_{2@}$
3	$v_{31} v_{32} v_{33} \dots v_{3s}$	$v_{3@}$
...
...
...
r	$v_{r1} v_{r2} v_{r3} \dots v_{rs}$	$v_{r@}$
Сума	$v_{@1} v_{@2} v_{@3} \dots v_{@s}$	N

Таким чином, таблиця представляє результат s незалежних рядів спостережень, кожному з яких відповідає один стовпчик.

Задача зводиться до перевірки гіпотези про те, що s вибірок, представлених стовпцями таблиці, взяті з однієї і тієї ж сукупності, чи, інакше кажучи, ці дані є однорідними.

Така гіпотеза еквівалентна (рівносильна) гіпотезі про те, що існує r постійних p_1, p_2, \dots, p_r і таких, що $\sum p_i = 1$, і ймовірність приналежності окремого результату до i -ої групи у всіх s послідовностях дорівнює p_i .

Для перевірки цієї гіпотези скористаємося розподілом Пірсона й напишемо його в вигляді

$$\chi^2 = n \sum_{i,j} \frac{(v_{ij} - v_{i@} \cdot v_{@j}/n)^2}{v_{i@} \cdot v_{@j}} = n \left[\sum_{i,j} \frac{v_{ij}^2}{v_{i@} \cdot v_{@j}} - 1 \right],$$

який має $(r-1)(s-1)$ число ступенів свободи.

Цю залежність можна поширити й на випадок, коли розглядається s незалежних вибірок по n_1, n_2, \dots, n_s елементів, розбитих на однакове число r груп, і за допомогою методу мінімуму χ^2 , застосованого до виразу

$$\chi^2 = \sum_{i,j} \frac{(v_{ij} - n_j \cdot p_i)^2}{n_j \cdot p_i}$$

p_i визначається як деяке число t невідомих параметрів.

Відомо, що закон Пірсона має граничний розподіл з $(r-1)(s-t)$ ступенями свободи, і в загальному випадку маємо справу з гіпотезою про те, що всі s вибірок взяті з однієї й тієї ж сукупності без подальшого уточнення виду розподілу цієї сукупності, так що параметрами є самі ймовірності. Завдяки співвідношенню $\sum p_i = 1$ маємо $t=r-1$ параметрів, так що отримуємо $(r-1)(s-1)$ ступенів свободи.

Розподіл Пірсона (розподіл χ^2) можна використовувати також для перевірки гіпотези про те, що задані чи наявні s вибірок взяті (належать) з однієї й тієї ж сукупності заданого типу, наприклад, мають розподіл Гауса, Пуассона або якийсь інший.

У цьому випадку застосування методу мінімуму χ^2 показує, що параметри розподілу ймовірностей відшуковуються так само, як і в разі однієї вибірки з груповими частотами, рівними сумах рядків $v_i, i=1, 2, \dots, r$ у наведеній вище таблиці. В окремому випадку при $r=2$ таблицю можна записати у вигляді, наведеному нижче.

$v_1 \quad v_2 \quad \dots \quad v_s$	$\sum v_j$
$n_1 - v_1 \quad n_2 - v_2 \quad \dots \quad n_s - v_s$	$n - \sum v_j$
$n_1 \quad n_2 \quad \dots \quad n_s$	N

У цьому випадку отримуємо s послідовностей спостережень, у кожному з яких деяка подія, скажімо, E здійснюється відповідно v_1, v_2, \dots, v_s раз, і треба встановити, чи є підстави вважати, що подія E в усіх цих спостереженнях має одну й ту ж постійну, хоча й невідому, ймовірність p ? Оцінку для цієї ймовірності може служити частота події E в усій сукупності даних

$$p^* = 1 - q^* = \frac{1}{n} \sum_j v_j$$

і тоді розподіл ймовірностей за Пірсоном запишеться в вигляді

$$\chi^2 = \sum_j \frac{(v_j - n_j \cdot p^*)^2}{n_j \cdot p^* \cdot q^*} = \frac{1}{p^* \cdot q^*} \sum_j \frac{v_j^2}{n_j} - n \frac{p^*}{q^*}$$

з $s-1$ ступенями свободи.

Величина

$$Q = \sqrt{\frac{n-1}{n(s-1)} \chi^2}$$

називається коефіцієнтом розбіжності.

Розглянемо випадок, коли $s=2$, тобто є дві незалежні вибірки і потрібно встановити, чи належать вони одній і тій ж сукупності? Для цього випадку таблицю можна представити в вигляді

μ_1	v_1	μ_1+v_1
μ_2	v_2	μ_2+v_2
...
...
μ_r	v_r	μ_r+v_r
m	n	$m+n$

Тут є $r-1$ ступенів свободи, і розподіл Пірсона запишеться в вигляді

$$\chi^2 = m \cdot n \sum_i \frac{1}{\mu_i + v_i} \left(\frac{\mu_i}{m} + \frac{v_i}{n} \right)^2.$$

Позначивши в цьому виразі

$$\frac{\mu_i}{\mu_i + v_i} = \omega_i, \quad \frac{m}{m+n} = \omega,$$

отримаємо зручну для розрахунку залежність

$$\chi^2 = \frac{1}{\omega(1-\omega)} \left(\sum_i \mu_i \cdot \omega_i - m \cdot \omega \right).$$

2. МЕТОДИКА ВИКОНАННЯ ПРАКТИЧНОГО ЗАВДАННЯ

2.1. Мета практичного завдання

Оволодіння способами побудови рядів розподілу, емпіричної та теоретичної кривої розподілу, методами розрахунку числових характеристик, набуття вмій і навичок використання критеріїв злагоди для перевірки висунутої статистичної гіпотези.

2.2. Приклад виконання практичного завдання

Виконання практичного завдання розглянемо на прикладі такої задачі.

2.2.1. Задача

Є дані про кількість залізобетонних виробів, виготовлених за зміну різними бригадами (у % до плану):

61,2	61,4	60,2	61,2	61,3	60,4	61,4	60,8	61,2	60,6
61,6	60,2	61,3	60,3	60,7	60,9	61,2	60,5	61,0	61,4
61,1	60,9	61,5	61,4	60,6	61,2	60,1	61,3	61,1	61,3
60,3	61,3	60,6	61,7	61,6	61,2	60,8	61,3	61,0	61,2
60,5	61,4	60,7	61,3	60,9	61,2	61,1	61,3	60,9	61,4
60,7	61,2	60,3	61,1	61,0	61,5	61,3	61,9	61,4	61,3
61,6	61,0	61,7	61,1	60,9	61,5	61,6	61,4	61,5	61,2
61,6	61,3	61,8	61,1	61,7	60,9	62,2	61,1	62,1	61,0
61,5	61,7	62,3	62,3	61,7	62,9	62,5	62,8	62,6	61,5
62,1	62,6	61,6	62,5	62,4	62,3	62,1	62,3	62,2	62,1

2.2.2. Зміст практичного завдання

На основі сукупності даних дослідження виконати таке:

1. Побудувати ряди розподілу (інтервальний і дискретний варіаційні ряди). Зобразити їх графіки.
2. Побудувати графік накопичених частот – кумуляту.
3. Скласти емпіричну функцію розподілу та зобразити її графічно.
4. Розрахувати моду, медіану, вибірккову середню, вибірккову дисперсію, вибірккове середнє квадратичне відхилення, коефіцієнт варіації, асиметрію, ексцес.
5. Побудувати довірчі інтервали для дійсного значення величини, що вимірюється, та середнього квадратичного відхилення генеральної сукупності.
6. Розкрити змістовний бік кожної характеристики.
7. Побудувати емпіричну (полігон) і теоретичну (нормальну) криву розподілу.
8. Перевірити узгодженість емпіричного розподілу з теоретичним нормальним, використовуючи три критерії:
 - а) критерій Пірсона;

- б) один з критеріїв: Колмогорова, Романовського, Ястремського;
- в) наближений критерій.

2.2.3. Методичні рекомендації до виконання практичного завдання

1.1. Побудувати інтервальний ряд. Для цього знайти:

- а) розмах варіювання ознаки за формулою

$$R = X_{max} - X_{min},$$

де X_{max} – найбільша варіанта в даній вибірковій сукупності;

X_{min} – найменша варіанта в даній вибірковій сукупності;

- б) число інтервалів варіаційного ряду, користуючись одним з наведених нижче співвідношень:

$$k \approx \sqrt{n}, \quad 6 < k < 12, \quad k \approx 1 + 3,21 \lg n,$$

де n – об'єм вибірки;

- в) довжину h часткових інтервалів за формулою

$$h = \frac{R}{K}$$

і, якщо необхідно, округлити це значення до деякого числа;

- г) записати отриманий інтервальний варіаційний ряд, заповнивши табл. 2. 1. Виконати контроль, переконавшись, що

$$\sum n_i = n.$$

Таблиця 2.1

Інтервальний варіаційний ряд

Варіанти – інтервали, $(X_{i-1};$ $X_i)$	$(X_0; X_1)$	$(X_1; X_2)$...	$(X_{k-1}; X_k)$
Частоти, n_i	n_1	n_2	...	n_k

- 1.2. Побудувати дискретний варіаційний ряд, взявши за варіанти середини варіант-інтервалів безперервного варіаційного ряду, а як частоти – частоти безперервного варіаційного ряду (табл. 2. 2).

Таблиця 2.2

Дискретний варіаційний ряд

Варіанти, X_i	X_1	X_2	...	X_k
Частоти, n_i	n_1	n_2	...	n_k

1.3. Зобразити графічно інтервальний і дискретний варіаційні ряди (побудувати гістограму та полігон частот).

2. Побудувати графік накопичених частот – кумуляту, тобто ламану, що переходить через точки з координатами X_i і відповідними накопиченими частотами. Попередньо скласти табл. 2. 3.

Таблиця 2.3

Розрахункова таблиця

Варіанти, X_i	X_1	X_2	...	X_k
Відносні частоти, $\omega_i = \frac{n_i}{n}$	$\omega_1 = \frac{n_1}{n}$	$\omega_2 = \frac{n_2}{n}$...	$\omega_k = \frac{n_k}{n}$
Накопичені відносні частоти, $W_i = W_{i-1} + \omega_i$	$W_1 = \omega_1$	$W_2 = \omega_2$...	$W_k = W_{k-1} + \omega_k$

3. Знайти емпіричну функцію розподілу та зобразити її графічно.

4.1. Знайти моду M_0X і медіану M_eX (моду M_0X називають варіанту, що має найбільшу частоту; медіаною M_eX називають варіанту, що ділить варіаційний ряд на рівні за числом варіант частини).

4.2. Для обчислення інших статистик скористатися методом добутоків. Ввести умовні варіанти

$$U_i = \frac{x_i - c}{h},$$

де $c = M_0X$;

h – крок (довжина інтервалу).

Скласти розрахункову табл. 2. 4.

Розрахункова таблиця

X_i	n_i	U_i	$n_i \cdot U_i$	$n_i \cdot U_i^2$	$n_i \cdot U_i^3$	$n_i \cdot U_i^4$	Контрольний стовпчик $n_i(U_i+1)^2$
Строчка суми	$\Sigma =$	$\Sigma =$	$\Sigma =$	$\Sigma =$	$\Sigma =$	$\Sigma =$	$\Sigma =$

Контроль розрахунків виконати за формулою:

$$\Sigma n_i + 2\Sigma n_i \cdot U_i + \Sigma n_i \cdot U_i^2 = \Sigma n_i (U_i + 1)^2.$$

4.3. Користуючись табл. 2. 4, розрахувати початкові моменти:

$$M_1^* = \frac{1}{n} \Sigma n_i \cdot U_i, M_2^* = \frac{1}{n} \Sigma n_i \cdot U_i^2, M_3^* = \frac{1}{n} \Sigma n_i \cdot U_i^3, M_4^* = \frac{1}{n} \Sigma n_i \cdot U_i^4.$$

4.4. Знайти вибіркву середню

$$\bar{X} = M_1^* \cdot h + c.$$

4.5. Знайти вибіркву дисперсію

$$D(X) = (M_2^* - M_1^{*2}) h^2.$$

4.6. Знайти вибіркве середнє квадратичне відхилення

$$\sigma_B = S = \sqrt{D(X)}.$$

4.7. Знайти коефіцієнт варіації:

$$V = \frac{S}{\bar{X}}.$$

4.8. Знайти центральні моменти:

$$m_3 = (M_3^* - 3 \cdot M_2^* \cdot M_1^* + 2 \cdot M_1^{*3}) h^3,$$

$$m_4 = (M_4^* - 4 \cdot M_3^* \cdot M_1^* + 6 \cdot M_2^* \cdot M_1^{*2} - 3 \cdot M_1^{*4}) h^4.$$

4.9. Розрахувати асиметрію:

$$A_S = \frac{m_3}{S^3}$$

і ексцес:

$$E_X = \frac{m_4}{S^4} - 3.$$

5. Довірчі інтервали для a і τ знайти за формулами:

$$\bar{X} - t_\gamma \frac{S_X}{\sqrt{n}} < a < \bar{X} + t_\gamma \frac{S_X}{\sqrt{n}} \text{ при } \gamma = 0,95.$$

Величину t , знайти за додатком 1, а за додатком 2 знайти величину q , яка задовольняє одній з умов:

$$S(1-q) < \tau < S(1+q) \text{ при } q < 1$$

$$0 < \tau < S(1+q) \text{ при } q > 1.$$

6. Розкрити змістовий бік кожної характеристики.
7. Будуємо емпіричну та теоретичну криві розподілу.
8. Перевіряємо узгодженість емпіричного розподілу з теоретичним нормальним за критерієм Пірсона.
9. Проводимо перевірку близькості емпіричного розподілу до нормального за критерієм Романовського.
10. Перевірити узгодженість емпіричного розподілу з теоретичним нормальним за критерієм Колмогорова.
11. Провести перевірку близькості розглянутої вибірки до нормального розподілу за наближеним критерієм.
12. Зробити висновки про узгодженість емпіричного розподілу з теоретичним нормальним.

2.2.4. Методика виконання практичного завдання

Позначимо через X кількість залізобетонних виробів, виготовлених за зміну різними бригадами.

1.1. За даними вибірки будуємо інтервальний варіаційний ряд.

а) Оскільки, як легко з'ясувати,

$$x_{max}=62,8, x_{min}=60,1,$$

то розмах варіювання ознаки X дорівнює

$$R=x_{max}-x_{min}=62,8-60,1=2,7.$$

б) Визначаючи число k інтервалів (число стовпчиків у таблиці) варіаційного ряду, покладемо $k=10$.

в) Довжина h кожного частинного інтервалу дорівнює

$$h=k/R=2,7/10=0,27.$$

Так як вихідні дані мало відрізняються один від одного та містять один десятковий знак, то величину h округляємо до одного десятинного знаку: $h=0,3$.

г) підраховуємо число варіант, які потрапляють у кожний інтервал, за даними вибірки. Значення x_i , що потрапляє на межу інтервалу, відносимо до правого інтервалу. За початок x_0 першого інтервалу беремо величину

$$x_0 = x_{min} - 0,5 \cdot h = 60,1 - 0,5 \cdot 0,3 = 59,95 \approx 60.$$

Кінець x_k останнього інтервалу знаходимо за формулою

$$x_k = x_{max} + 0,5 \cdot h = 62,8 + 0,15 = 62,95 \approx 63,0.$$

Сформований інтервальний варіаційний ряд записуємо у вигляді табл. 2. 5.

Таблиця 2.5

Варіанти-інтервали	60,0-60,3	60,3-60,6	60,6-60,9	60,9-61,2	61,2-61,5	61,5-61,8	61,8-62,1	62,1-62,4	62,4-62,7	62,7-63,0
Частоти, n_i	3	6	9	18	29	16	2	10	5	2

Контроль: $\sum n_i = 100$, і об'єм вибірки $n = 100$.

1.2. Записуємо дискретний варіаційний ряд (табл. 2. 6). Як варіанти x_i беремо середини інтервалів інтервального варіаційного ряду.

Таблиця 2.6

варіанти, x_i	60,15	60,45	60,75	61,05	61,35	61,65	61,95	62,25	62,55	62,85
частоти, n_i	3	6	9	18	29	16	2	10	5	2

Зображуємо інтервальний і дискретний варіаційні ряди графічно, побудувавши гістограму та полігон частот у одній системі координат (рис. 2. 1).

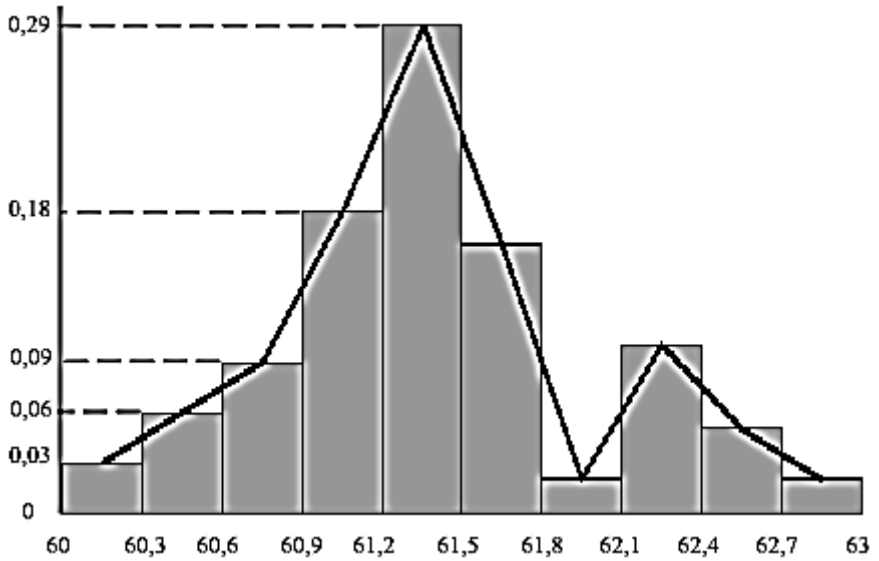


Рис. 2.1 – Гістограма та полігон

2. Будуємо графік накопичених частот – кумуляту (рис. 2. 2). Попередньо складаємо розрахункову таблицю 2. 7.

Таблиця 2.7

Варіанти, x_i	60,15	60,45	60,75	61,05	61,35	61,65	61,95	62,25	62,55	62,85
Відносні частоти, $\omega_i=n/n$	0,03	0,06	0,09	0,18	0,29	0,16	0,02	0,1	0,05	0,02
Накопичені відносні частоти, $W_i=W_{i-1}+\omega_i$	0,03	0,09	0,18	0,36	0,65	0,81	0,83	0,93	0,98	1

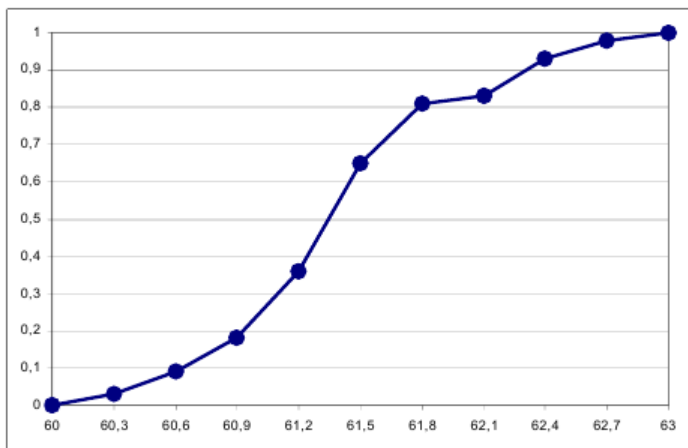


Рис. 2.2 – Кумулятивна крива

3. Знаходимо емпіричну функцію розподілу. Скористаємося формулою

$$F_B(x) = \frac{n_x}{n}.$$

Якщо $x \leq 60,15$, то $F_B(x) = 0$ – за властивістю емпіричної функції розподілу.

$$\text{Якщо } 60,15 < x \leq 60,45, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{3}{100} = 0,03.$$

$$\text{Якщо } 60,45 < x \leq 60,75, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{3+6}{100} = 0,09.$$

$$\text{Якщо } 60,75 < x \leq 61,05, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{3+6+9}{100} = 0,18.$$

$$\text{Якщо } 61,05 < x \leq 61,35, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{18+18}{100} = 0,36.$$

$$\text{Якщо } 61,35 < x \leq 61,65, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{36+29}{100} = 0,65.$$

$$\text{Якщо } 61,65 < x \leq 61,95, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{65+16}{100} = 0,81.$$

$$\text{Якщо } 61,95 < x \leq 62,25, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{81+2}{100} = 0,83.$$

$$\text{Якщо } 62,25 < x \leq 62,55, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{83+10}{100} = 0,93.$$

$$\text{Якщо } 62,55 < x \leq 62,85, \text{ то } F_B(x) = \frac{n_x}{n} = \frac{93+5}{100} = 0,98.$$

Якщо $x > 62,85$, то $F_B(x) = 1$ – за властивістю емпіричної функції розподілу.

Запишемо отриману емпіричну функцію в вигляді:

$$F_B = \left\{ \begin{array}{l} 0,00, x \in (-\infty, 60,15], \\ 0,03, x \in (60,15; 60,45], \\ 0,09, x \in (60,45; 60,75], \\ 0,18, x \in (60,75; 61,05], \\ 0,36, x \in (61,05; 61,35], \\ 0,65, x \in (61,35; 61,65], \\ 0,81, x \in (61,65; 61,95], \\ 0,83, x \in (61,95; 62,25], \\ 0,93, x \in (62,25; 62,55], \\ 0,98, x \in (62,55; 62,85], \\ 1,00, x \in (62,85; +\infty). \end{array} \right\}$$

Графік функції $F_6(x)$ представлений на рис. 2. 3.

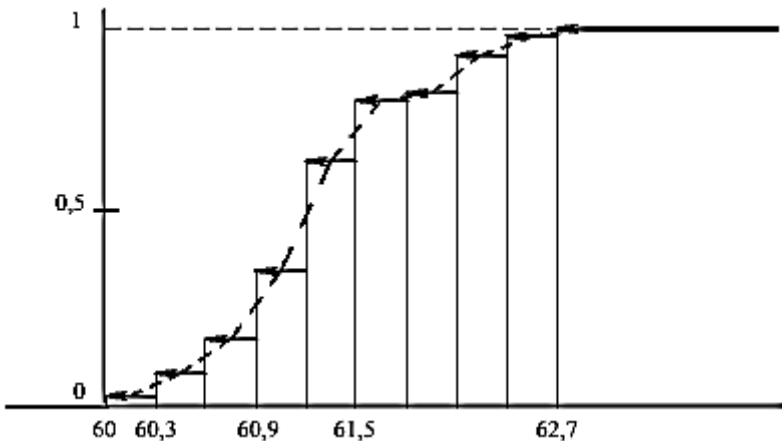


Рис. 2.3 – Кумулята та емпірична функція розподілу

З'єднавши середини вертикальних частин ступінчатої кусочно-постійної кривої, що є графіком функції $F_6(x)$, отримуємо плавну криву (на рис. 2. 3 це штрихова лінія). Абсцисами точок цієї кривої слугують значення кількості залізобетонних виробів, виготовлених за зміну різними бригадами, а ординатами – значення емпіричної функції розподілу, що характеризує оцінку ймовірності події $X \leq x_i$, тобто

ймовірності попадання можливих значень кількості залізобетонних виробів, виготовлених за зміну різними бригадами, на проміжок $(-\infty, x_i]$.

Для знаходження числових характеристик ознаки X – кількості залізобетонних виробів (*незміщених оцінок* для $M(X)=a$, $D(X)$, а також $M_e X$, $M_o X$, A_s , E_x) скористаємося табл. 2. 6.

4.1. Так як варіанта $x=61,35$ в табл. 2. 6 зустрічається з найбільшою частотою $n_5=29$, то $M_o X=61,35$, тобто це значення кількості залізобетонних виробів, яке зустрічається в дані вибірці з найбільшою частотою.

Знаходимо $M_e X$. Так як табл. 2. 6 містить парне число стовпчиків, то

$$M_e X = \frac{61,35 + 61,65}{2} = 61,5.$$

Це значення кількості залізобетонних виробів, яке ділить дані вибірки ознаки X на рівні частини.

Для знаходження інших статистик, які характеризують кількості залізобетонних виробів, скористаємося методом добутоків. Введемо у відповідності до формули

$$M_e X = \frac{x_k + x_{k+1}}{2}$$

(тут x_k – варіанта, що знаходиться зліва від середини варіаційного ряду, а x_{k+1} – справа від неї) *умовні варіанти*

$$u_i = \frac{x_i - C}{h}; C=M_o X=61,35, h=0,3.$$

Складемо розрахункову таблицю 2. 8.

Контроль розрахунків проводимо за формулою:

$$\sum n_i + 2 \sum n_i \cdot u_i + \sum n_i \cdot u_i^2 = \sum n_i (u_i + 1)^2,$$

тобто $100+2 \cdot 14+400=528=27+24+9+29+64+18+160+125+72$.

Таким чином, розрахунки виконані вірно.

4.3. Користуючись результатами останньої строчки табл. 2. 8, знаходимо умовні початкові моменти:

$$M_1^* = \frac{1}{n} \sum n_i \cdot u_i = \frac{14}{100} = 0,14,$$

$$M_2^* = \frac{1}{n} \sum n_i \cdot u_i^2 = \frac{400}{100} = 4,$$

Таблиця 2.8

x_i	n_i	u_i	$n_i \cdot u_i$	$n_i \cdot u_i^2$	$n_i \cdot u_i^3$	$n_i \cdot u_i^4$	контрольний стовпчик $n_i(u_i + 1)^2$
60,15	3	-4	-12	48	-192	768	27
60,45	6	-3	-18	54	-162	486	24
60,75	9	-2	-18	36	-72	144	9
61,05	18	-1	-18	18	-18	18	0
61,35	29	0	0	0	0	0	29
61,65	16	1	16	16	16	16	64
61,95	2	2	4	8	16	32	18
62,25	10	3	30	90	270	810	160
62,55	5	4	20	80	320	1280	125
62,85	2	5	10	50	250	1250	72
	100		14	400	428	4804	528

$$M_3^* = \frac{1}{n} \sum n_i \cdot u_i^3 = \frac{428}{100} = 4,28,$$

$$M_4^* = \frac{1}{n} \sum n_i \cdot u_i^4 = \frac{4804}{100} = 48,04.$$

4.4. Знаходимо вибірккову середню:

$$\bar{x} = M_1^* \cdot h + C = 0,14 \cdot 0,3 + 61,35 = 61,392 \approx 61,39,$$

котра характеризує середню кількість залізобетонних виробів, виготовлених за зміну різними бригадами, у даній вибірці, що складає **61,39%**.

4.5. Знаходимо вибірккову дисперсію:

$$S^2 = (M_2^* - M_1^{*2})h^2 = (4 - 0,14^2)0,09 = 0,358236.$$

4.6. Розраховуємо вибірккове середнє квадратичне відхилення:

$$S = \sqrt{S^2} = \sqrt{0,358236} = 0,59853.$$

4.7. Величина $S=0,599$ характеризує ступінь розсіювання значень кількості залізобетонних виробів відносно середньої кількості. Для визначення коливання значень кількості залізобетонних виробів у відсотковому відношенні розраховуємо коефіцієнт варіації:

$$V = \frac{S}{\bar{x}} = \frac{0,599}{61,39} = 0,00976.$$

Величина коефіцієнту варіації мала (складає $0,01$), що означає тісне згрупування значень кількості залізобетонних виробів біля центру розсіювання, тобто біля середньої кількості залізобетонних виробів.

4.8. Для попередньої оцінки відхилення значень кількості залізобетонних виробів від нормального розподілу розраховуємо асиметрію та ексцес. Спочатку знаходимо центральні моменти третього та четвертого порядків:

$$m_3 = (M_3^* - 3 \cdot M_2^* \cdot M_1^* + 2 \cdot M_1^{*3})h^3$$

$$= (4,28 - 3 \cdot 4 \cdot 0,14 + 2 \cdot 0,14^3)0,3^3 = 0,07035.$$

$$m_4 = (M_4^* - 4 \cdot M_3^* \cdot M_1^* + 6 \cdot M_2^{*2} \cdot M_1^{*2} - 3 \cdot M_1^{*4})h^4$$

$$= (48,04 - 4 \cdot 4,28 \cdot 0,14 + 6 \cdot 4 \cdot 0,14^2 - 3 \cdot 0,14^4)0,3^4 = 0,37023.$$

4.9. Тоді знаходимо:

$$A_s = \frac{m_3}{S^3} = \frac{0,07}{0,599^3} = 0,328,$$

$$E_x = \frac{m_4}{S^4} - 3 = \frac{0,37}{0,599^4} - 3 = -0,115.$$

Резюме. Значення A_s і E_x мало відрізняються від нуля. Тому можна припустити близькість даної вибірки, що характеризує кількість залізобетонних виробів, до нормального розподілу. Ця гіпотеза буде перевірятися нижче.

5. Виконаємо оцінку генеральної середньої $M(X)=a$ та генерального середньоквадратичного відхилення $\sigma=S$ за вибірковими статистиками x і S , використовуючи теорію довірчих інтервалів для нормального розподілу.

Довірчий інтервал для дійсного значення кількості залізобетонних виробів з надійністю $\gamma=0,95$ знаходимо:

$$\bar{x} - \frac{S}{\sqrt{n}} t_\gamma < a < \bar{x} + \frac{S}{\sqrt{n}} t_\gamma.$$

У відповідності до додатку 1, при $n=100$ і $\gamma=0,95$ знаходимо

$t_{\gamma}=1,984$. Записуємо довірчий інтервал:

$$61,39 - \frac{0,599}{10} 1,984 < a < 61,39 + \frac{0,599}{10} 1,984,$$

або $61,27 < a < 61,51$.

Таким чином, середня кількість залізобетонних виробів, виготовлених за зміну різними бригадами (у %) за даними вибірки повинна знаходитися в проміжку (61,27; 61,51).

Запишемо довірчий інтервал для генерального середньоквадратичного відхилення $\sigma=S$. При заданих $\gamma=0,95$ і $n=100$ за таблицею додатку 2 знаходимо $q=0,143$. Так як $q < 1$, то довірчий інтервал записуємо в вигляді:

$$S(1-q) < \sigma < S(1+q),$$

або

$$0,599(1-0,143) < \sigma < 0,599(1+0,143),$$

або

$$0,51 < \sigma < 0,68;$$

отже, відхилення дійсних значень кількості залізобетонних виробів, виготовлених за зміну різними бригадами не повинні виходити за межі проміжку (0,15; 0,86).

Продовжимо ймовірно-статистичну обробку результатів експерименту. За основу беремо дискретний варіаційний ряд в табл. 2. 6 і значення $\bar{x}=61,39$ і $S=0,599$.

6. Емпірична крива розподілу являє собою полігон частот. Для побудови теоретичної (нормальної) кривої знайдемо координати точок (x_i, n_i') , для чого розрахуємо теоретичні частоти n_i' (табл. 2. 9).

Будуємо емпіричну та теоретичну криві (рис. 2. 4).

7. Перевіримо узгодженість емпіричного розподілу з теоретичним нормальним за критерієм Пірсона. Розрахуємо величину χ^2 за формулою

$$\chi^2 = \sum_{i=1}^l \frac{(n_i - n_i')^2}{n_i'}$$

Для відшукування суми складаємо розрахункову табл. 2. 10. Знаходимо число ступенів свободи

$$k = s - r - s - 3 = 10 - 3 = 7.$$

Таблиця 2.9

x_i	n_i	$x_i - \bar{x}$	$u_i = \frac{x_i - \bar{x}}{S}$	$\varphi(u_i)$	$y_i = \frac{n \cdot h}{S} \varphi(u_i)$	n'_i
60,15	3	-1,24	-2,07	0,00468	2,3	2
60,45	6	-0,94	-1,57	0,1163	5,8	6
60,75	9	-0,64	-1,07	0,2251	11,3	11
61,05	18	-0,34	-0,57	0,3391	17,0	17
61,35	29	-0,04	-0,07	0,3980	19,9	20
61,65	16	0,26	0,43	0,3637	18,2	18
61,95	2	0,56	0,93	0,2589	13,0	13
62,25	10	0,86	1,44	0,1415	7,1	7
62,55	5	1,16	1,94	0,0608	3,0	3
62,85	2	1,46	2,44	0,0203	1,0	1

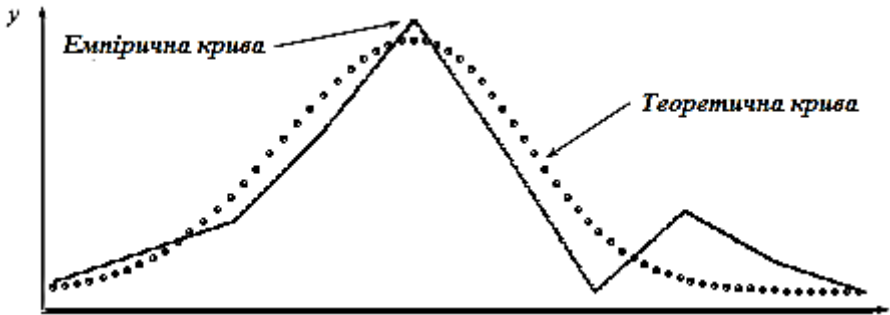


Рис. 2.4

8. Вибираємо рівень значущості $\alpha=0,95$. За таблицею критичних точок розподілу χ^2 (додаток 3) знаходимо $\chi_{кр}^2 = 2,17$. Так як $\chi_{кр}^2 < \chi_0^2$ ($2,7 < 18,12$), то робимо висновок, що дані вибірки, котрі характеризують кількість залізобетонних виробів, виготовлених за зміну, не підкоряються нормальному закону розподілу.

9. Проведемо перевірку близькості емпіричного розподілу до нормального за критерієм Романовського. Розрахуємо величину $\left| \frac{\chi^2 - k}{\sqrt{2 \cdot k}} \right|$. Так як $\chi^2 = \chi_0^2 = 18,12$, $k=7$, то

$$\left| \frac{\chi^2 - k}{\sqrt{2 \cdot k}} \right| = \left| \frac{18,12 - 7}{\sqrt{14}} \right| = 2,97 < 3,$$

тобто розходження між емпіричним і теоретичним розподілом несуттєве, що дозволяє стверджувати, що данні вибірки, котрі характеризують кількість залізобетонних виробів, виготовлених за зміну, за критерієм Романовського підкоряються нормальному закону розподілу. До такого ж висновку ми приходимо, використовуючи критерій Колмогорова (перевірити самостійно).

Таблиця 2.10

n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\frac{(n_i - n'_i)^2}{n'_i}$
3	2	1	1	0,5
6	6	0	0	0
9	11	-2	4	0,363636
18	17	1	1	0,058824
29	20	9	81	4,05
16	18	-2	4	0,222222
2	13	-11	121	9,307692
10	7	3	9	1,285714
5	3	2	4	1,333333
2	1	1	1	1
				$\chi^2 = 18,12$

10. Насамкінець, проведемо перевірку близькості розглянутої вибірки до нормального розподілу за наближеним критерієм, використовуючи вибіркові статистики: асиметрію, ексцес і їх середні квадратичні відхилення. Раніше були знайдені $A_s=0,328$, $E_x=-0,115$. Середні квадратичні відхилення для асиметрії та ексцесу знаходимо за формулами:

$$S_{A_s} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}} = \sqrt{\frac{6 \cdot 99}{101 \cdot 103}} = 0,24,$$

$$S_{E_X} = \sqrt{\frac{24 \cdot n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}} = \sqrt{\frac{24 \cdot 100 \cdot 98 \cdot 97}{99^2 \cdot 103 \cdot 105}} = 0,46.$$

Так як

$$|A_S| = 0,328 > S_{A_S}$$

і

$$|E_X| = 0,115 < S_{E_X},$$

то робимо висновок, що дані вибірки, котрі характеризують кількість залізобетонних виробів, виготовлених за зміну, не підкоряються нормальному закону розподілу.

10. Отже, для перевірки узгодженості емпіричного розподілу з теоретичним нормальним ми використали 4 критерії, два з них підтвердили близькість вибіркової сукупності до нормального розподілу. Однак, враховуючи, що критерій Колмогорова є більш потужним, ніж критерій χ^2 Пірсона, і підтверджує близькість розглянутої вибірки до нормального розподілу, остаточно укладаємо, що за закон розподілу ознаки X – кількості залізобетонних виробів, виготовлених за зміну – можна прийняти нормальний розподіл.

Література

1. Бахрушин В.Є. Методи аналізу даних : навчальний посібник для студентів / В.Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с.
2. Василенко О.А. Математично-статистичні методи аналізу у прикладних дослідженнях: навч. посіб. / О.А. Василенко, І.А. Сенча. – Одеса: ОНАЗ ім. О.С. Попова, 2011. – 166 с.
3. Герич М.С., Синявська О.О. Математична статистика: навч. посіб. / М.С. Герич, О.О. Синявська – ДВНЗ «УжНУ»: Ужгород, 2021. – 146 с.
4. Горват А.А. , Молнар О.О., Мінькович В.В. Методи обробки експериментальних даних з використанням MS Excel: Навчальний посібник / А.А Горват, О.О. Молнар, В.В. Мінькович. – Ужгород: Видавництво УжНУ «Говерла», 2019. – 182 с.
5. Данілов В.Я. Статистична обробка даних: навчальний посібник / В.Я. Данілов – Київський національний університет імені Тараса Шевченка: Київ, 2019. – 156 с.

6. Єлейко В.І. Економетричний аналіз діяльності підприємств : навч. посіб. / [уклад. : В.І. Єлейко, Р.Д. Боднар, М.Я. Демчишин]. – Львів : Львівська комерційна академія, 2011. – 368 с.
7. Єріна А.М. Статистичне моделювання та прогнозування / А.М. Єріна. – Київ: КНТЕУ, 2001. – 196 с.
8. Інженерний аналіз експериментальних даних. Методичні вказівки до самостійної роботи студентів спеціальностей: 192 – Будівництво та цивільна інженерія; 133 – Галузеве машинобудування. – Кропивницький: ЦНТУ, 2017. – 82 с.
9. Кветний Р.Н. Методи комп'ютерних обчислень. Навч. посібник. – Вінниця: ВДТУ, 2001. – 148с.
10. Мамчич Т. Статистичний аналіз даних з пакетом STATISTICA / Т. Мамчич, А. Оленко, М. Осипчук, В. Шпортюк. – Дрогобич : Відродження, 2006. – 208 с.
11. Мельниченко О.П. Статистична обробка експериментальних даних: Навчальний посібник / О.П. Мельниченко, І.Л. Якименко, Р.Л. Шевченко – Біла Церква, 2006. – 34 с.
12. Пашинський В.А., Пашинський М.В. Статистичні методи в інженерних дослідженнях: навч. посіб. для здобувачів вищої освіти з інженерних спеціальностей / В.А. Пашинський, М.В. Пашинський – Кропивницький: ЦНТУ, 2020. – 106 с.
13. Присенко Г.В. Прогнозування соціально-економічних процесів: навч. посіб. / Г.В. Присенко, Є.І. Равікович. – Київ : КНЕУ, 2005. – 378 с.
14. Руська Р.В., Іващук О.Т. Методи економіко-статистичних досліджень: Навчальний посібник / Р.В. Руська, О.Т. Іващук – Тернопіль: Тайп, 2014. – 190 с.
15. Статистичні методи обробки результатів фізичного експерименту: курс лекцій: [навчальний посібник] / [І.М. Гасюк, Л.С. Кайкан]. – Івано-Франківськ: Видавництво Прикарпатського національного університету імені Василя Стефаника, 2011. – 159 с.

Варіанти завдань

У варіантах завдань приведена кількість залізобетонних виробів, виготовлених за зміну різними бригадами.

Варіант №1

71	76	79	86	78	76	84	78	74	76	99	87	82
78	84	81	76	75	82	85	80	76	79	76	86	86
86	89	77	80	74	86	87	74	79	84	75	85	81
88	77	74	93	85	83	80	75	93	95	91	88	85
85	83	85	82	86	79	84	88	74	92	95	76	

Варіант №2

85	76	80	84	88	89	91	88	84	85	75	82	86
89	88	84	90	89	85	91	87	81	78	85	91	89
87	74	81	87	90	88	86	76	84	88	77	82	83
84	74	80	84	91	93	90	88	87	77	83	89	89
91	92	88	94	90	88	81	83	89	94	96	88	95
99	86	78	81	86	90	92	93	90	83	79	86	90
79	82	87	85	91	97	88	85	87	90	89	95	89
90	98	93	84	88	96	92	88	95				

Варіант №3

52	33	10	22	28	34	39	29	21	27	31	12	28
40	46	51	44	32	16	11	29	31	38	44	31	24
9	17	32	41	47	31	42	15	21	29	50	55	37
19	57	32	7	28	23	20	45	18	29	25		

Варіант №4

14	13	18	15	12	13	14	12	13	16	16	15	12
13	13	14	16	18	13	15	14	15	14	13	15	12
13	12	14	16	12	13	15	15	15	13	14	15	18
15	12	15	13	13	15	15	15	17	17			

Варіант №5

30	19	21	28	27	29	31	24	25	28	28	32	34
26	24	19	23	27	30	29	25	18	18	24	28	31
33	18	21	26	30	32	34	29	26	23	25	27	32
23	20	21	26	22	20	27						

Варіант №6

261	260	258	263	257	260	259	264	261	260	264	261	265
261	260	263	260	260	259	260	258	265	259	265	261	258
259	259	258	262	264	258	259	263	266	259	261	266	262
259	262	261	266	262	259	262	261	259	262	262	261	266
259	262											

Вариант №7

75	88	74	80	76	82	86	76	93	74	72	82	71
82	87	81	87	79	78	87	82	87	82	74	77	83
86	85	86	76	81	86	76	71	80	85	73	75	88
89	84	85	85	81	82	85	83	76	87	87	76	76
73	78	87	80	78	72	83	91	82	93	76	83	80

Вариант №8

550	550	551	550	551	562	550	562	561	530	542	535	542
539	537	543	540	556	546	556	556	534	548	533	558	560
558	548	540	541	551	549	551	550	552	568	538	551	547
552	559	557	546	552	550	557	547	552	554	547	554	567
558	563	562	569	552	554	549	545	560	539	549	539	

Вариант №9

73	77	78	88	76	78	86	77	75	90	88	84	79
87	83	79	73	84	86	85	74	77	74	88	81	87
85	76	79	71	88	83	76	76	82	73	89	79	90
76	75	91	83	82	84	85	78	85	85	79	92	86
84	77	92	93	91	85	84	87	81	83	80	82	76
81	90	78	81	95	77	91	84	96	84	79	79	83
88	84	83	93	73	79	92	89	75	83	87	89	71
75	83	87	92	80	88	91	95	82				

Вариант №10

39	19	21	28	26	27	29	28	28	27	23	26	32
34	26	24	22	19	23	27	30	29	25	18	18,5	20
22	24	28	31	33	25	18	21	26	30	32	34	29
26	21	20	23	25	27	30	32	29	27	23		

Вариант №11

82	83	73	76	79	89	95	92	93	84	88	76	88
81	78	86	84	84	86	85	87	84	74	83	87	73
76	73	78	76	76	74	88	82	73	85	79	77	79
97	84	80	75	81	73	78	83	75	90	83	77	84
85	90	92	91	85	71	85	87	82	94	92	76	93
90	73	92	84	93	88	84	81	93	81	91	78	85
84	95	79	79	83	96	89	82	79	77	83	88	81
88	82	77	92	76	84	83	87	89				

Вариант №12

21	29	27	29	27	29	31	29	31	29	29	23	39
31	29	31	29	31	29	31	33	31	31	31	27	23
27	33	29	25	29	19	29	31	23	31	29	27	33
29	31	29	31	23	35	27	29	29	27	29	29	21
29	27	29	29	29	33	29	25	25	27	31	29	29
27	33	29	31	29	29	29	35	27	29	35	29	33
29	27	31	31	27	29	35	27	33	29	27	29	25
27	31	37	25	31	27	27	29	25				

Вариант №13

28	30	28	27	28	29	29	29	31	28	26	25	33
35	27	31	31	30	28	33	23	30	31	33	31	27
30	28	30	29	30	26	25	31	33	26	27	33	29
30	30	36	26	25	28	30	29	27	32	29	31	30
31	26	25	29	31	33	27	32	30	31	34	28	26
38	29	31	29	27	31	30	28	34	30	26	30	32
30	29	30	28	32	30	29	34	32	35	29	27	28
30	30	29	32	29	34	30	32	24				

Вариант №14

48	29	6	18	24	24	30	35	25	17	24	36	42	47
40	28	12	7	25	23	33	28	19	14	8	40	27	
20	27	15	6	16	25	34	17	25	46	6	51	13	
28	37	43	27	38	53	24	41	21	11	26			

Вариант №15

95	57	15	26	35	46	52	55	59	47	42	48	58	
55	102	96	45	54	56	60	10	16	20	49	48	43	
12	19	51	103	62	61	38	29	10	39	40	18	14	
41	58	63	59	60	63	68	70	71	75	82	87	92	
99	65	68	78	91	94	77	65	79	67	74	80	89	
69	81	83	100	90	36	64	97	50	76	72	31	55	
28	57	85	69	13	53	11	61	90	76	17	37		

Вариант №16

19	29	21	39	25	26	32	25	28	26	36	30	31	
29	35	23	32	27	27	26	26	30	27	25	28	28	
36	29	35	26	32	29	38	28	25	29	34	28	29	
32	34	28	28	29	33	27	34	25	28	26	30	38	
39	32	29	29	34	35	32	27	26	25	26	35	36	
30	28	33	26	28	26	28	27	33	33	29	32	25	
38	26	36	23	24	27	26	30	34	25	24	33		

Вариант №17

133	133	142	135	145	144	145	147	146	134	130	134	138	
144	141	141	134	141	136	140	143	139	141	137	140	145	
145	141	144	138	139	143	141	141	146	143	140	139	143	
143	139	140	139	138	138	135	141	141	140	138	145	135	
148	136	139	142	143	143	137	138	138	139	138	144	143	
138	142	138	140	140	137	139	140	139	137	136	136	135	
135	141	142	136	140	136	137	138	138	137	139	139	140	
139	140	140	139	139	139	140	140	146					

Вариант №18

16	13	11	15	18	19	21	18	17	15	13	16	18
17	19	15	13	12	14	16	17	20	17	17	20	19
18	22	24	1	15	14	10	12	16	18	18	19	21
23	20	22	24	17	16	14	15	18	15	11	16	17
15	13	16	17	18	14	15	19	17	18	16	13	15
17	21	23	26	19	22	24	25	20	21	24	19	22
23	20	25	21	20	22	26	19	22	25	28	23	20
21	27	19	15	22	23	18	22	22				

Вариант №19

11	15	20	25	29	34	19	25	16	21	29	20	21
22	23	26	28	30	18	13	17	22	29	26	39	14
16	24	27	25	31	32	23	37	23	27	37	36	42
32	34	39	38	44	28	33	23	35	36	34		

Вариант №20

25	29	33	21	29	25	29	28	31	23	31	27	29
27	27	29	31	27	29	29	29	31	25	29	29	27
29	31	29	27	25	28	27	31	31	29	27	27	33
29	33	31	33	25	27	35	37	35	27	27	29	27
29	31	29	27	29	31	29	21	23	29	37	29	31
29	31	29	31	29	39	29	39	39	27	31	37	29
31	29	27	23	29	27	31	29	29	31	29	35	29
19	29	27	29	29	31	33	29	25				

Додаток 1

Таблица значений $t_\gamma = t(\gamma, n)$

$n \backslash \gamma$	0,95	0,99	0,999	$n \backslash \gamma$	0,95	0,99	0,999
5	2,78	4,60	8,61	20	2,093	2,361	3,883
6	2,57	4,03	6,86	25	2,064	2,797	3,745
7	2,45	3,71	5,96	30	2,045	2,756	3,659
8	2,37	3,50	5,41	35	2,032	2,720	3,600
9	2,31	3,36	5,04	40	2,023	2,708	3,558
10	2,26	3,25	4,78	45	2,016	2,692	3,527
11	2,23	3,17	4,59	50	2,009	2,679	3,502
12	2,20	3,11	4,44	60	2,001	2,662	3,464
13	2,18	3,06	4,32	70	1,996	2,649	3,439
14	2,16	3,01	4,22	80	1,001	2,640	3,418
15	2,15	2,98	4,14	90	1,987	2,633	3,403
16	2,13	2,95	4,07	100	1,984	2,627	3,392
17	2,12	2,92	4,02	120	1,980	2,617	3,374
18	2,11	2,90	3,97	∞	1,960	2,576	3,291
19	2,10	2,88	3,92				

Додаток 2

Таблица значений $q_\gamma = q(\gamma, n)$

$n \backslash \gamma$	0,95	0,99	0,999	$n \backslash \gamma$	0,95	0,99	0,999
5	1,37	2,67	5,64	20	0,37	0,58	0,88
6	1,09	2,01	3,88	25	0,32	0,49	0,73
7	0,92	1,62	2,98	30	0,28	0,43	0,63
8	0,80	1,38	2,42	35	0,26	0,38	0,56
9	0,71	1,20	2,06	40	0,24	0,35	0,50
10	0,65	1,08	1,80	45	0,22	0,32	0,46
11	0,59	0,98	1,60	50	0,21	0,30	0,43
12	0,55	0,90	1,45	60	0,188	0,269	0,38
13	0,52	0,83	1,33	70	0,174	0,245	0,34
14	0,48	0,78	1,23	80	0,161	0,226	0,31
15	0,46	0,73	1,15	90	0,151	0,211	0,29
16	0,44	0,70	1,07	100	0,143	0,198	0,27
17	0,42	0,66	1,01	150	0,115	0,160	0,211
18	0,40	0,63	0,96	200	0,099	0,136	0,185
19	0,39	0,60	0,92	250	0,089	0,120	0,162

Додаток 3
Критичні точки розподілу χ^2

Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,98
1	6,6	5,024	3,841	0,0039	0,00098	0,00016
2	9,2	7,378	5,991	0,103	0,051	0,020
3	11,3	9,348	7,815	0,352	0,216	0,115
4	13,3	11,143	9,488	0,711	0,484	0,297
5	15,1	12,832	11,070	1,15	0,831	0,554
6	16,8	14,449	12,592	1,64	1,24	0,872
7	18,5	16,013	14,067	2,17	1,69	1,24
8	20,1	17,535	15,507	2,73	2,18	1,65
9	21,7	19,023	16,919	3,33	2,70	2,09
10	23,2	20,483	18,307	3,94	3,25	2,56
11	24,7	21,920	19,676	4,57	3,82	3,05
12	26,2	23,336	21,026	5,23	4,40	3,57
13	27,7	24,736	22,362	5,89	5,01	4,11
14	29,1	26,129	23,685	6,57	5,63	4,66
15	30,6	27,488	24,996	7,26	6,26	5,23
16	32,0	28,845	26,296	7,96	6,91	5,81
17	33,4	30,191	27,587	8,67	7,56	6,41
18	34,8	31,536	28,869	9,39	8,23	7,01
19	36,2	32,852	30,144	10,1	8,91	7,63
20	37,6	34,170	31,410	10,9	9,59	8,26
21	38,9	35,479	32,671	11,6	10,3	8,90
22	40,3	36,781	33,924	12,3	11,0	9,54
23	41,6	38,076	35,172	13,1	11,7	10,2
24	43,0	39,364	36,415	13,8	12,4	10,9
25	44,3	40,646	37,652	14,6	13,1	11,5
26	45,6	41,923	38,885	15,4	13,8	12,2
27	47,0	43,194	40,113	16,2	14,6	12,9
28	48,3	44,461	41,337	16,9	15,3	13,6
29	49,6	45,722	42,557	17,7	16,0	14,3
30	50,9	46,979	43,773	18,5	16,8	15,0

Додаток 4
Критичні точки розподілу Стюдента

Число степеней свободы k	Уровень значимости α (двусторонняя критическая область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,3138	12,7062	31,82	63,7	318,3	637,0
2	2,9200	4,3037	6,97	9,92	22,33	31,6
3	2,3534	3,1824	4,54	5,84	10,22	12,9
4	2,1318	2,7764	3,75	4,60	7,17	8,61
5	2,0150	2,5706	3,37	4,03	5,89	6,86
6	1,9432	2,4469	3,14	3,71	5,21	5,96
7	1,8946	2,3646	3,00	3,50	4,79	5,40
8	1,8595	2,3060	2,90	3,36	4,50	5,04
9	1,8331	2,2622	2,82	3,25	4,30	4,78
10	1,8125	2,2281	2,76	3,17	4,14	4,59
11	1,7959	2,2010	2,72	3,11	4,03	4,44
12	1,7823	2,1788	2,68	3,05	3,93	4,32
13	1,7709	2,1604	2,65	3,01	3,85	4,22
14	1,7613	2,1448	2,62	2,98	3,79	4,14
15	1,7530	2,1314	2,60	2,95	3,73	4,07
16	1,7459	2,1190	2,58	2,92	3,69	4,01
17	1,7396	2,1098	2,57	2,90	3,65	3,96
18	1,7341	2,1009	2,55	2,88	3,61	3,92
19	1,7291	2,0930	2,54	2,86	3,58	3,88
20	1,7247	2,0860	2,53	2,85	3,55	3,85
21	1,7207	2,0796	2,52	2,83	3,53	3,82
22	1,7171	2,0739	2,51	2,82	3,51	3,79
23	1,7139	2,0687	2,50	2,81	3,49	3,77
24	1,7109	2,0639	2,49	2,80	3,47	3,74
25	1,7081	2,0595	2,49	2,79	3,45	3,72
26	1,7056	2,0555	2,48	2,78	3,44	3,71
27	1,7033	2,0518	2,47	2,77	3,42	3,69
28	1,7011	2,0484	2,46	2,76	3,40	3,66
29	1,6991	2,0452	2,46	2,76	3,40	3,66
30	1,6973	2,0423	2,46	2,75	3,39	3,65
40	1,6839	2,0211	2,42	2,70	3,31	3,55
60	1,6706	2,0003	2,39	2,66	3,23	3,46
120	1,6577	1,9840	2,36	2,62	3,17	3,37
∞	1,6479	1,9647	2,33	2,58	3,09	3,29
	0,05	0,025	0,01	0,005	0,001	0,0005
Уровень значимости α (односторонняя критическая область)						

Додаток 5

Таблиця значень функції $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	39986	3984	3982	3980	3977	3973
0,1	0,3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	0,3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	0,3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	0,3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	0,3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	0,3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	0,3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	0,2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	0,2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	0,2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	0,1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	0,1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	0,1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	0,1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	0,1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0,0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0,0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0,0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0,0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0,0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0,0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0,0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0,0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0,0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0,0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0,0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0,0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0,0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0,0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0,0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0,0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0,0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0,0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0,0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0,0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0,0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Навчально-методичне видання

СТАТИСТИЧНІ МЕТОДИ ОБРОБКИ ДАНИХ

Методичні рекомендації до самостійного вивчення та виконання
практичних завдань для здобувачів спеціальностей 192 – Будівництво
та цивільна інженерія, 073 – Менеджмент

Електронне видання

Укладачі – к.т.н., доцент Хачатурян С.Л.
к.т.н., доцент Пімонов І.Г.
к.т.н., доцент Щербак О.В.
к.е.н., доцент Хачатурян О.С.

© ЦНТУ, Кропивницький, пр. Університетський, 8

© ХНАДУ, Харків, вул. Ярослава Мудрого, 25

© Хачатурян С.Л., Пімонов І.Г., Щербак О.В.,
Хачатурян О.С.