МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ

ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

КАФЕДРА КОМП'ЮТЕРНИХ НАУК

КОНСПЕКТ ЛЕКЦІЙ

з дисципліни

«ЕМПІРИЧНІ МЕТОДИ ПРОГРАМНОЇ ІНЖЕНЕРІЇ»

для студентів напрямку підготовки:

Програмна інженерія

ТЕРНОПІЛЬ – 2016

Вступ

У даний час до емпіричних методів відносять: математичну статистику, теорію кореляції; теорію вірогідності; метод статистичних випробувань; статистичну теорію ухвалення рішень. Всі ці методи аналізу і досліджень ставлять завдання визначення ступеня зв'язку між чинниками виробництва і економічними показниками роботи галузей міського господарства. Тому вони покликані не тільки визначити зміну показників, але і встановити величину зв'язки між ними.

Метод кореляційного аналізу є найпоширенішим при дослідженнях, ним визначається ступінь зв'язку ряду чинників з економічними показниками досліджуваної сукупності. Цей метод виходить із звітно-статистичних відомостей, що інформують тільки про деякі обставини виробництва і лише в середньому по підприємствах за декілька років (місяців). Він застосовується, коли є кількісно однорідна статистична сукупність великого обсягу, що включає зміну економічних показників і основних обставин матеріально-технічних і соціально-економічних умовах виробництва.

Мета курсу "Емпіричні методи програмної інженерії" - викласти властивості математичних методів, їх практичне застосування за допомогою ЕОМ при плануванні й управлінні виробництвом (наданням послуг) в галузях житлово-комунального господарства, для здійснення і прискорення аналізу інформації і підвищення ефективності оперативного керівництва складною динамічною системою.

Безпосереднім способом дослідження і аналізу виробничо-економічних явищ за допомогою математики і обчислювальної техніки є математичне формулювання інформації, що складається з таких етапів:

- збір (отримання) виробничо-економічних відомостей по процесах і ланках виробництва:
- обробка і формулювання інформації по процесах, ланках і обслуговуючих цехах.
- перевірка на практиці впливу встановлених причин виробництва на показники роботи або експериментування за допомогою ЕОМ впливу матеріально-технічних умов на економічні показники виробництва (послуги).

РОЗДІЛ 1 Застосування принципів дискретної імовірності в комп'ютингу

1.1.Теорія множин.

В основі теорії множин лежать первинні поняття: *множина* та елемент множини. Тео́рія множи́н — розділ математики, в якому вивчаються загальні властивості множин (переважно нескінченних). Виділення теорії множин в самостійний розділ математики відбулося на рубежі XIX і XX століть. Теорія множин зробила дуже великий вплив на розвиток сучасної математики — вона являється фундаментом ряду нових розділів математики, дозволила по-новому поглянути на класичні розділи математики і глибше зрозуміти сам предмет математики.

Елемент множини перебуває щодо множини у відношенні *бути елементом* множини (позначається як $x \in A$ — «х є елемент множини А»). Серед похідних понять найважливішими є наступні:

- *порожня множина* — множина, яка не містить елементів, позначається зазвичай $\varnothing$;
- *підмножина* і *надмножина* — множина, яка складається тільки з елементів іншої множини, та множина, до якої належать усі елементи іншої множини, відповідно;
- сімейство множин;
- *простір* (*універсум*) — множина, що є надмножиною всіх множин;
- *конституента*.

Над множинами визначені наступні операції:

- *об'єднання* (або *сума*) (позначається як $A \cup B$);
- *перетин* (або *добуток*) (позначається як $A \cap B$);
- *різниця* (позначається як $A \setminus B$,рідше $A - B$);
- *симетрична різниця* (позначається як $A \bigtriangleup B$,рідше $A \dot{-} B$).
- *доповнення* (позначається як $\setminus A$,або $-A$);

Для множин визначені наступні бінарні відношення:

- *відношення рівності* (позначається як $A = B$);
- *відношення включення* (позначається як $A \subset B$,або $A \subseteq B$).

1.2 Теорія імовірності

Одним з найпростіших випадкових явищ, що розглядаються у теорії ймовірності, є так звані ***випадкові події***.

Теорія ймовірності- це математична наука, що вивчає закономірності у випадкових подіях.

Подія – це все те, що трапляється чи може трапитись.

Випадковою подією (або просто ***подією***) будемо називати усякий факт, який в досліді з випадковим перебігом може відбутися, або не відбутися. Так, наприклад, можна говорити про подію, яка полягає в тому, що при підкиданні грального кубика з‘явиться цифра “6”.

Події будемо позначати великими літерами латинського алфавіту: ***A,B,C,U,V*** тощо.

Домовимося розглядати **елементарні події,** які не можна розкласти на простіші, та **складні події**, які можна розкласти на елементарні.

Сукупність усіх можливих елементарних подій називають **простором елементарних подій *Ω***, а самі елементарні події ω_i- **точками** цього простору:

$$\Omega = \{\omega_1, \omega_2, ..., \omega_i, ... \}$$

При побудові теоретико-множинної моделі теорії ймовірностей будемо ототожнювати поняття події з поняттям множини.

В подальшому подія **А** – це деяка підмножина **Ω**, яка складається з усіх тих точок **ω –** елементарних подій, які сприяють події А.

Якщо результат випробування можна описати точкою **ω** і $\omega \in A$(**ω** входить до **А)**, то в даному випробуванні подія **А** здійснилася, якщо $\omega \notin A$, то подія **А** в цьому випробуванні не здійснилася.

Відомо, що для множин визначене відношення порядку, і над ними можна проводити певні алгебраїчні операції.

1. Сама множина **Ω** розглядається, як подія, що характеризується тим, що в результаті випробування вона обов'язково здійсниться. Дійсно, ніякі інші результати випробування крім тих, що можна описати точками , по означенню не можливі. Множину **Ω** називають **вірогідною подією.**

2. Підмножиною будь-якої множини **Ω** вважають пусту множину **Ø** , яка не містить жодної точки .

Якщо **Ø** ототожнити з подією, то ця подія при випробуванні не має місця. Цю множину **Ø** називають **неможливою подією**.

Операції над подіями.

Оскільки події – це деякі підмножини всіх елементарних подій, то над ними можна ввести такі самі операції, як і над множинами у теорії множин.

Розглянемо деякі з цих операцій. Говоритимемо, що подія $\boldsymbol{A}$ сприяє появі події $\boldsymbol{B}$ $(\boldsymbol{A} \subset \boldsymbol{B})$, якщо з відбуванням події $\boldsymbol{A}$ відбувається також подія $\boldsymbol{B}$.

З точки зору теорії множин множина $\boldsymbol{A}$ є підмножиною множини $\boldsymbol{B}$. У цьому випадку, якщо навмання вибраний елемент належить множині $\boldsymbol{A}$ (пишуть $\boldsymbol{X} \in \boldsymbol{A}$), то він належить і множині $\boldsymbol{B}$($\boldsymbol{X} \in \boldsymbol{B}$) (рис.1).

$\boldsymbol{A} \subset \boldsymbol{B}$

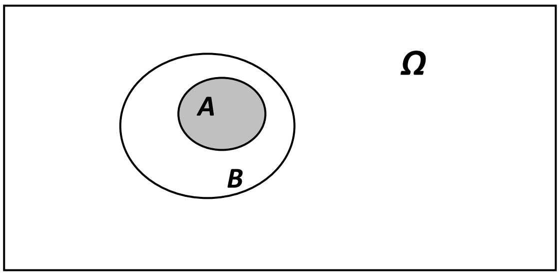

Рис. 1 Подія А сприяє події В

Якщо подія $\boldsymbol{A}$ сприяє появі події $\boldsymbol{B}$, і подія $\boldsymbol{B}$ сприяє появі події $\boldsymbol{A}$, то події $\boldsymbol{A}$ і $\boldsymbol{B}$ називають **еквівалентними** (рівносильними) і пишуть $\boldsymbol{A=B}$**.**

У цьому випадку відповідні множини збігаються (містять у собі одні й ті самі елементи).

Безпосередньо з означень випливає: $\boldsymbol{A} \subset \boldsymbol{A}$ для будь-якої події $\boldsymbol{A}$.

Якщо $\boldsymbol{A} \subset \boldsymbol{B}, \boldsymbol{B} \subset \boldsymbol{C}$,то $\boldsymbol{A} \subset \boldsymbol{C}$.

Отже, $\boldsymbol{A} = \boldsymbol{A}$ для будь-якої події.

Якщо $\boldsymbol{A} = \boldsymbol{B}$, $\boldsymbol{B} = \boldsymbol{C}$**,** то $\boldsymbol{A} = \boldsymbol{C}$.

Сумою двох подій $\boldsymbol{A}$ і $\boldsymbol{B}$ називають таку подію $\boldsymbol{C}$, яка відбувається, коли відбувається принаймні одна з подій $\boldsymbol{A}$, $\boldsymbol{B}$.

При цьому пишуть $\boldsymbol{C} = \boldsymbol{A}+\boldsymbol{B}$, або $\boldsymbol{C} = \boldsymbol{A} \cup \boldsymbol{B}$.

Сумі подій $\boldsymbol{A}$ і $\boldsymbol{B}$ відповідає множина $\boldsymbol{C}$, кожний елемент якої є елементом принаймні однієї з множин $\boldsymbol{A}$ і $\boldsymbol{B}$. В такому разі говорять, що множина $\boldsymbol{C}$ є **об'єднанням** множин $\boldsymbol{A}$ і $\boldsymbol{B}$ (рис.2).

$\boldsymbol{C} = \boldsymbol{A} \cup \boldsymbol{B}$

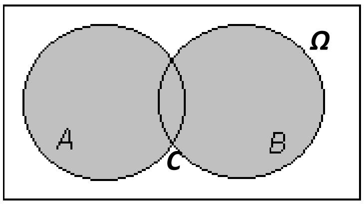

Рис. 2 Сума подій A і B

Добутком двох подій $\boldsymbol{A}$ і $\boldsymbol{B}$ називають таку подію $\boldsymbol{C}$, яка відбувається, коли одночасно відбуваються обидві події $\boldsymbol{A}$ і $\boldsymbol{B}$.

При цьому пишуть $C = A \cdot B$ (або $C = A \cap B$).

Добутку подій $\boldsymbol{A}$ і $\boldsymbol{B}$ відповідає множина $\boldsymbol{C}$ таких елементів, кожний з яких належить множині $\boldsymbol{A}$ і множині $\boldsymbol{B}$. У цьому випадку кажуть, що множина $\boldsymbol{C}$ є **перерізом** множин $\boldsymbol{A}$ і $\boldsymbol{B}$ (рис.3).

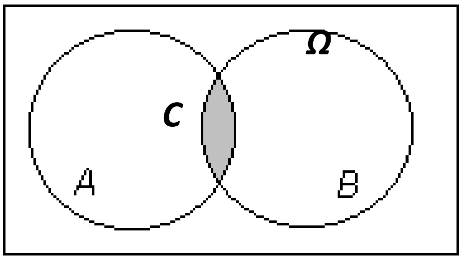

Рис. 3 Добуток подій A і B

Добутком декілька подій $A_1, A_2, ..., A_n$ називають подію $C = \prod_{i=1}^{n} A_i$ (або $C = \bigcap_{i=1}^{n} A$), яка полягає в тому, що одночасно відбуваються всі події: A_1, A_2, ..., A_n .

Різницею $\boldsymbol{A \backslash B}$ подій $\boldsymbol{A}$ і $\boldsymbol{B}$ називають таку подію $\boldsymbol{C}$, яка відбувається, коли відбувається подія $\boldsymbol{A}$ і не відбувається подія $\boldsymbol{B}$.

Різниці подій $\boldsymbol{A}$ і $\boldsymbol{B}$ відповідає множина, що є різницею множин $\boldsymbol{A}$ і $\boldsymbol{B}$ (із множини $\boldsymbol{A}$ вилучаються елементи множини $\boldsymbol{B}$)
(рис.4).

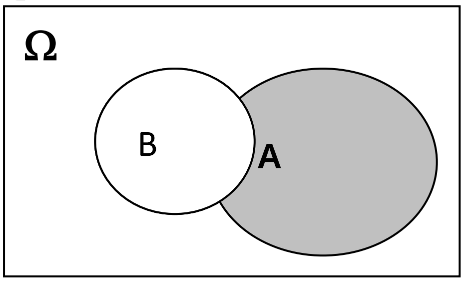

Рис. 4 Різниця подій A і B

Подією, **протилежною** до події $\boldsymbol{A}$, називають подію $\overline{A} = \Omega / A$, яка відбувається, коли не відбувається подія $\boldsymbol{A}$, і не відбувається, коли відбувається подія $\boldsymbol{A}$. Подію $\boldsymbol{A}$ визначає множина, яка є доповненням множини $\boldsymbol{A}$ до множини $\boldsymbol{\Omega}$.(рис.5).

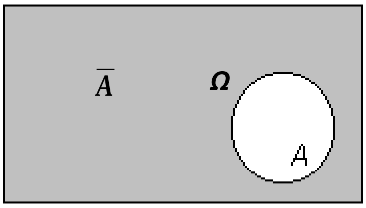

Рис. 5 Протилежна подія до події А

Основні **закони**, яким підлягають операції над подіями.

1. $\boldsymbol{A = A}$
2. $\boldsymbol{A + B = B + A}$– комутативний (переставний) додавання
3. $\boldsymbol{A\ B = B\ A}$– комутативний (переставний) множення
4. $\boldsymbol{(A+B) + C = A + (B+C)}$– асоціативний (сполучний) додавання
5. $\boldsymbol{(A\ B)C = A(B\ C)}$ –асоціативний (сполучний) множення
6. $\boldsymbol{(A+B)C = AC+BC}$ –1-й дистрибутивний
7. $\boldsymbol{AB+C = (A+C)(B+C)}$– 2-й дистрибутивний
8. $\boldsymbol{A + A = A}$
9. $\boldsymbol{A \cdot A = A}$
10. $\boldsymbol{A + \overline{A} = \Omega}$
11. $\boldsymbol{A \cdot \overline{A}} = \varnothing$
12. $\boldsymbol{A \cdot \Omega = A}$
13. $\boldsymbol{A + \varnothing = A}$
14. $\boldsymbol{A + \Omega = \Omega}$
15. $\boldsymbol{A \cdot \varnothing = \varnothing}$
16. $\boldsymbol{\overline{A + B} = \overline{A} \cdot \overline{B}}$
17. $\boldsymbol{\overline{AB} = \overline{A} + \overline{B}}$

Випробування (експеримент, іспит, дослід) – сукупність умов, при яких спостерігається те чи інше явище, фіксується той чи інший результат.
Якщо результат випробування при повторенні досліду змінюється, то говорять про випробування з ***випадковим перебігом.***

Приклад 1. Нехай випробування полягає в тому, що двічі підкидається монета, а подія А – в тому, що герб з'явиться при цьому хоча б один раз. Тоді

Ω={ГГ, ГЦ, ЦГ, ЦЦ}

A={ГГ, ГЦ, ЦГ}

де Г – поява "герба", Ц – поява "цифри".

Подія **А** містить в собі три елементарних події.

Події ***A*** і ***B*** називають **несумісними,** якщо вони не можуть відбуватися одночасно, тобто якщо $A \cap B = \varnothing$ **.** Говорять, що подія ***A*** розпадається на частинні випадки $\boldsymbol{\omega_1, \omega_2, \dots, \omega_n}$**,** якщо ***A*** можна подати у вигляді суми

$$A = \omega_1 + \omega_2 + \dots + \omega_n$$

причому події $\boldsymbol{\omega_i}$ $(\boldsymbol{i} = \overline{1, \boldsymbol{n}})$, попарно несумісні, тобто $\omega_i \cdot \omega_j = \varnothing$ при $\boldsymbol{i \neq j}$.

Групу подій $\boldsymbol{\omega_1, \omega_2, \dots, \omega_n}$ називають **повною групою подій**, якщо в результаті випробування принаймні одна з подій цієї групи обов'язково відбувається, тобто якщо $\boldsymbol{\omega_1 + \omega_2 + \dots + \omega_n = \Omega}$ **.**

Очевидно, множина всіх елементарних подій завжди є повною групою подій.

Будемо говорити, що дві події А і В **рівноймовірні**, якщо немає підстав вважати, що можливість появи однієї з цих подій більша або менша, ніж можливість появи іншої.

Розглянемо тепер повну групу рівноймовірних елементарних подій ω_1, $\omega_2,\ldots,\omega_n$, таких, що $\omega_1 + \omega_2 + \ldots + \omega_n = \Omega$.

Нехай подія А розпадається на m частинних випадків, що входять до повної групи з n рівноймовірних і попарно несумісних подій.

Ймовірністю події *А* називається число, що обчислюється за формулою, $P(A) = \frac{m}{n}$ де n – число всіх можливих взаємовиключаючих рівноймовірних елементарних подій, а m – число подій сприятливих до появи події ***А*.**

Це формулювання належить Лапласу.

__А Зауваження:__ Дане означення не відповідає на питання, що таке ймовірність випадкової події, а дає лише метод її обчислення у найпростіших випадках. Крім того, це означення непридатне для недискретного простору елементарних подій і навіть для дискретного простору нерівноймовірних подій.

Якщо припустити, що множина всіх можливих (рівноймовірних між собою) наслідків випробування **Ω** і множина ***А*** наслідків випробування, які сприяють події ***А***, мають скінченні міри $m(\Omega)$ і $m(A)$, то ймовірність події ***А*** (за означенням)

$$P(A) = \frac{m(A)}{m(\Omega)} \tag{1.1}$$

Під мірою множини розуміємо в одновимірному випадку загальну довжину проміжків, з яких складається ця одновимірна множина, у двовимірному випадку загальну площу областей, з яких складається відповідна двовимірна, у тривимірному випадку – об`єм і т.д.

В конкретних задачах, що зводяться до цієї ймовірної схеми, випробування інтерпретується як випадковий вибір точки в деякій області ,а подія ***А*** – як влучення обраної точки в деяку підобласть ***А*** області ***Ω***. При цьому вимагається, щоб всі точки мали однакові шанси бути обраними.

У двовимірному випадку будемо писати, $P(A) = \frac{S_A}{S_\Omega}$ де S –площа (рис.6).

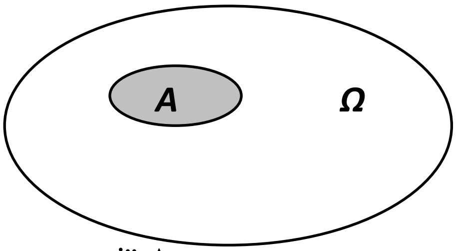

Рис. 6 Імовірність події А

Дамо загальне означення ймовірності, яке охоплює "класичне" і “геометричне" означення.

__Означення__. Непуста сукупність подій S_σ називається σ – алгеброю, якщо виконуються такі умови:

1). $A \in S_\sigma \Rightarrow \overline{A} \in S_\sigma$;

2). $A_i \in S_\sigma \ (i = 1,2,\ldots) \Rightarrow \bigcup\limits_{i=1}^{\infty} A_i \in S_\sigma$

З аксіом 1) і 2) випливає, що $\Omega \in S_\sigma, \varnothing \in S_\sigma$ і якщо $A_i \in S_\sigma$ $(i = 1,2,...)$, то $\bigcap\limits_{i=1}^{\infty} A_i \in S_\sigma$.

Будемо говорити, що на σ – алгебрі задано розподіл ймовірностей, якщо кожній події $A \in S_\sigma$ однозначно ставиться у відповідність деяке число ***P(A)***, яке називається ймовірністю події ***A***, так, що виконуються такі умови (аксіоми теорії імовірностей):

I). $P(A) \geq 0$;

II). $P(\Omega) = 1$;

III). якщо події $A_i \in S_\sigma$ $(i = 1,2,...)$ попарно несумісні $(A_i \cap A_j = \varnothing, \quad i \neq j, \quad i = 1,2,...)$,то $P\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = \sum\limits_{i=1}^{\infty} P(A_i)$

Сукупність трьох об'єктів $(\Omega,\ S_\sigma,\ P(A))$, в якій S_σ є σ – алгеброю подій, а функція ***P(A)*** задовольняє всім аксіомам теорії ймовірностей, називається ***імовірнісним простором.***

Величина називається випадковою, якщо вона формується під дією багатьох дрібних причин, не піддатливих до результату випробувань повному контролю і обліку, діючих відносно незалежно один від одного.

Дискретними називаються випадкові величини, які приймають окремі, строго визначені, ізольовані, кінцеві чисельні значення з певною вірогідністю, між якою не може бути проміжних (число робітників у бригаді, число перевезених за один рейс пасажирів і т.д.). При цьому число можливих значень дискретної випадкової величини може бути кінцевим і нескінченним.

Частіше зустрічаються безперервні випадкові величини, які можуть мати всі можливі значення в деякому кінцевому або нескінченному проміжку. Очевидно, число можливих значень безперервної випадкової величини нескінченне, наприклад, рівень собівартості перевезення одного пасажира, продуктивність праці і т.д.

Оскільки точність вимірювання або обліку завжди обмежена, то практично всі випадкові величини є дискретними.

Значна частина теорії вірогідності й математичної статистики пов'язана з необхідністю досліджувати і описувати велику сукупність об'єктів. Звичайно цю сукупність називають генеральною. Вона охоплює, наприклад, усіх мешканців великого міста, продукцію галузі народного господарства.

Якщо досліджувана сукупність об'єктів дуже численна або об'єкти вивчення труднодоступні, а також є інші причини, що не дозволяють вивчити всі об'єкти, то вдаються до вивчення якоїсь частини генеральної сукупності, що називається вибіркою.

Вибірка повинна бути представницькою або, як кажуть, репрезентативною. Якщо вибірка представляє не всю генеральну сукупність, а якусь її частину, то це називається зсувом вибірки. Зсув - одне з основних джерел помилок при використанні вибіркового методу.

Об'єктом дослідження теорії вірогідності є вибірка, що складається з n однорідних одиниць (елементів). Число n називається обсяг вибірки. Одиницями вибірки можуть бути різні економічні процеси і явища, результати виробничо-господарської діяльності підприємств: продуктивність праці, собівартість продукції, фондовіддача, рентабельність та ін.

Елементи комбінаторики

Сполученнями називають довільні групи, складені з будь-яких об’єктів (елементів). Існують три види сполучень: розміщення, перестановки або переставлення, комбінації.

Розміщеннями із n-елементів по m (m≤n)-взятих називають такі сполуки, які відрізняються між собою або самими об'єктами, або їх порядком розташування.

Число розміщень: $A_n^m = n(n-1)...(n-m+1)$ або $A_n^m = \frac{n!}{(n-m)!}$, де n!=1·2·3…(n-1)·n і читається “ен факторіал”.

Комбінаціями із n-елементів по m-взятих називають сполуки, які відрізняються між собою хоча б одним елементом.

Число комбінацій

$$\tilde{N}_n^m = \frac{A_n^m}{P_m} \equiv \frac{n!}{(n-m)!m!}. \qquad (1.2)$$

Важлива для обчислень властивість $\tilde{N}_n^m = C_n^{n-m}$.

Приклад 2. Серед 20 лампочок є 5 підвищеної якості. Навмання взяли 7 шт. Яка ймовірність, що серед них будуть 3 лампочки підвищеної якості?

Загальне число, рівноможливих наслідків

$$\tilde{N}_{20}^7 = \frac{20!}{7!13!} \equiv \frac{13!14\cdot15\cdot16\cdot17\cdot18\cdot19\cdot20}{13!1\cdot2\cdot3\cdot4\cdot5\cdot6\cdot7} \equiv 77520.$$

Сприятливі наслідки C_5^3, відбираючи довільні 3 із 5 лампочок підвищеної якості. До кожної трійки додаємо будь-які 4 шт., що беруться із 15 залишкових. Оскільки довільна трійка лампочок підвищеної якості може сполучатися з довільною четвіркою, взятою із групи залишкових лампочок, то всіх сприятливих наслідків всього буде $\tilde{N}_5^3 * \tilde{N}_{15}^4 = 13650$.

Шукана ймовірність є: $P_{(\grave{A})} = \frac{\tilde{N}_5^3 * \tilde{N}_{15}^4}{\tilde{N}_{20}^7} \equiv \frac{13650}{77520} \cong 0.176.$

Основні теореми класичної теорії ймовірностей

Теорема 1. Додавання ймовірностей несумісних подій

Ймовірність появи однієї з двох несумісних подій дорівнює сумі ймовірностей цих подій:

P(A+B)= P(A) +P(B).

Наслідок 1. Якщо події протилежні, то сума їх ймовірностей дорівнює одиниці, тобто

P(A)+P($\overline{A}$) =1.

Наслідок 2. Ймовірність появи однієї з декількох попарно несумісних подій дорівнює сумі ймовірностей цих подій.

$$P(\sum_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i).$$

Теорема 2. Додавання ймовірностей несумісних подій

Ймовірність появи хоча б однієї з двох сумісних подій дорівнює сумі ймовірностей цих подій без ймовірності їх сумісної появи.
P(A+B)=P(A)+P(B) P(A·B)
Зауваження. Теорема може бути узагальнена на довільне скінченне число сумісних подій. Якщо n=3, то
$P(A_1+A_2+A_3)=P(A_1)+P(A_2)+P(A_3)–P(A_1A_2)–P(A_1A_3)–P(A_2A_3)+P(A_1A_2A_3)$.

Приклад 2. Ймовірність того, що куплений товар випущено в Англії – 0,3, а ймовірність того, що товар випущено в Польщі дорівнює 0,6. Яка ймовірність того, що товар випущено в одній із цих країн?
Нехай подія А – товар випущено в Англії, подія В – товар випущено в Польщі, А і В несумісні. Застосувавши теорему додавання ймовірностей несумісних подій, маємо
P(A+B) = P(A) + P(B) =0,3 + 0,6 = 0,9

Приклад 3. Один лотерейний білет виграє з ймовірністю 0,0003. Яка ймовірність того, що володар одного білету нічого не виграє?
Нехай подія А – виграш. Тоді $\overline{A}$ означає, що білет не виграє. Події А і $\overline{A}$ протилежні. Використовуючи наслідок 1, маємо
$P(\overline{A})$ = 1 – P(A) = 1 0,0003 = 0,9997.
Теорема 3. Множення ймовірностей незалежних подій
Ймовірність сумісної появи двох незалежних подій дорівнює добутку ймовірностей цих подій:
P((A·B)=P(A)·P(B).
Узагальнення. Ймовірність появи кількох подій, незалежних в сукупності, дорвінює добутку ймовірностей цих подій:
$P((A_1·A_2·A_3·A_n)= P(A_1)P(A_2)P(A_3)·…·P(A_n)$.
Теорема 4. Множення ймовірностей залежних подій
Ймовірність сумісної появи двох залежних подій дорівнює добутку ймовірностей однієї з них на умовну ймовірність іншої:
$P(AB)=P_2\ (A)\ ·P_A\ (B)$.
Узагальнення. Ймовірність сумісної появи кількох залежних подій дорівнює добутку ймовірностей однієї з них на умовні ймовірності всіх інших, причому ймовірність кожної наступної події обчислюється в припущенні, що всі попередні події уже відбулися:
$P(A_1·A_2·A_3·…·A_n)= P(A_1)P_{A_1}(A_2)P_{A_1A_2}(A)_3·…·P_{A_1A_2A_3...A_{n-1}}(A_n)$.

1.3.Неперервні та дискретні випадкові величини.

Випадкова величина – це величина, яка приймає те чи інше значення в залежності від випадку. Позначаються випадкові величини великими літерами латинського алфавіту X,Y,Z..., або літерами грецького алфавіту ξ,η, ζ,....(додаток 3).

Означення. Нехай $(\Omega,\ S_\sigma,\ P(A))$ –імовірнісний простір. Будь яка дійсна функція $\xi=\xi(\omega)$ на Ω така, що для кожного дійсного *x:* $\{\omega:\xi(\omega)<x\}\in S_\sigma$ називається *випадковою величиною.*

Нехай ω_i (i=1,2,...,6) – випадкова подія, яка полягає в тому, що при киданні шестигранного кубика , на гранях якого нанесено цифри 1,2,3,4,5,6, на верхній грані випадає цифра *i*, $(i=\overline{1,6})$. Очевидно, множина всіх можливих наслідків випробування (елементарних подій) є: $\Omega=\{\ \omega_1,\omega_2,\omega_3,\omega_4,\omega_5,\omega_6\}$.

Випадкова величина ξ – число появи *i* очок – є функцією елементарної події, причому $\xi(\omega)=i$, якщо $\omega=\omega_i$.

Можливі значення випадкових величин будемо позначати : $x_1.x_2,...,x_n$

Випадкові величини поділяються на 3 типи:

- дискретні (перервні);
- неперервні.

<u>Означення.</u> Якщо множина можливих значень випадкової величини скінчена або зчисленна, то випадкова величина називається **дискретною.**

Приклади дискретних випадкових величин:

1) число пристроїв, які отримали пошкодження під час роботи;
2) число студентів, що не з’явились на лекцію;
3) довжина слова, кількість голосних або приголосних у реченні.

<u>Означення.</u> ***Неперервною*** випадковою величиною називається випадкова величина, можливі значення якої неперервно заповнюють деякій проміжок.

Приклади неперервних випадкових величин:

1) відстань від точки улучення у мішень до центра мішені;
2) інтенсивність звука, яка звичайно коливається в деяких межах;
3) працезатрати на відновлення роботи ПК.

1.4.Математичне сподівання та дисперсія випадкових величин.

Математичне сподівання ДВВ Х – це число, що дорівнює сумі добутків усіх можливих значень Х на відподвідні їм ймовірності. Позначають математичне сподівання М(Х) або m_X і обчислюють за формулою

$$M(X)=\sum_{k=1}^{n}x_k\,p_k \tag{1.3}$$

Зауваження. Математичне сподівання (МС) – це величина, навколо якої групуються можливі значення ВВ.

Основні властивості М (Х)

М(С)=С, де С=const.

М(СХ)=С М(Х), де С=const.

Якщо X_1, X_2,...X_n – взаємонезалежні ДВВ, тоді

$M(X_1\ X_2\ ...X_n)=M(X_1)\ M(X_2)\ ...M(X_n)$.

$M\ (X_1+X_2+...+X_n)=M\ (X_1)+M(X_2)+...M(X_n)$.

Дисперсія ДВВ Х – це міра розсіювання можливих значень Х відносно центру розподілу, і вона дорівнює математичному сподіванню квадрата відхилення ДВВ Х від її математичного сподівання. Позначають дисперсію D(X) або D_X. Математично означення дисперсії записується наступним чином

$$D(X) = M((X- M(X))^2) \tag{1.4}$$

Основні властивості D(X):

D(X)≥0, для будь-якої ДВВ Х.

D(C)=0, де С=const.

$D(CX)=C^2\,D(X)$, де С= const.

$$D(X)=M(X^{2})-(M(X))^{2} \quad (1.5)$$

або

$$D(X)=\sum_{K=1}^{n} x_k^2 p_k - (\sum_{ё=1}^{m} x_ё p_k)^2 \quad (1.6)$$

Зауважимо, що формулу (4) зручно використовувати для обчислення D(X): $D(X_1 \pm X_2 \pm ... \pm X_n)=D(X_1)+D(X_2)+...+D(X_n)$.

Середнє квадратичне відхилення ДВВ X визначається за формулою

$$\sigma=\sqrt{D(X)} \quad (1.7)$$

РОЗДІЛ 2 ОСНОВНІ ПОНЯТТЯ МАТЕМАТИЧНОЇ СТАТИСТИКИ

2.1. Поняття вибіркового методу в статистиці

Математична статистика – це розділ прикладної математики, предметом якого є розробка раціональних прийомів і методів отримання, опису та обробки експериментальних даних з метою вивчення закономірностей масових випадкових явищ.

Основними завданнями математичної статистики є:

– визначення за статистичними даними законів розподілу випадкових величин;

– визначення за статистичними даними параметрів розподілу випадкових величин;

– визначення за статистичними даними виду зв'язку між різними явищами (об'єктами) або властивостями одного і того ж явища (об'єкту);

– визначення сили (тісноти зв'язку) між різними явищами (об'єктами) або властивостями одного і того ж явища (об'єкту);

– перевірка вірогідності статистичних гіпотез;

– розробка рекомендацій щодо проведення експерименту та обробки його результатів.

У прикладних дослідженнях зазвичай необхідно вивчити сукупність однорідних об'єктів або спостережень за якою-небудь кількісною або якісною ознакою.

Сукупність об'єктів або спостережень, всі елементи якої підлягають вивченню при статистичному аналізі, називається **генеральною сукупністю.**

Генеральна сукупність може бути скінченою або нескінченною. Так, при вивченні розподілу населення за родом занять, розглядається велика, але скінчена генеральна сукупність об'єктів. При вивченні впливу яскравості освітлення робочого місця на продуктивність праці працівника генеральна сукупність спостережень теоретично нескінченна, оскільки яскравість освітлення може змінятся безперервно у межах певного інтервалу.

Число об'єктів (спостережень) генеральної сукупності називається її

об'ємом і позначається N.

На практиці рідко є можливість досліджувати кожен елемент генеральної сукупності, оскільки це зв'язано з великими витратами засобів і часу, а іноді з псуванням або знищенням досліджуваних об'єктів. У деяких випадках дослідити всі об’єкти генеральної сукупності взагалі неможливо. Тому при статистичному аналізі, як правило, вивчається не вся генеральна сукупність, а деяка її частина.

Частина об'єктів генеральної сукупності, використовувана в ході дослідження, називається **вибіркою**. Число об'єктів (спостережень) вибірки називається її **об'ємом** і позначається n.

Наприклад, продукція у кількості N одиниць, вироблена підприємством на протязі року, є генеральною сукупністю. Для дослідження якості продукції на практиці розглядається вибірка, що складається з n одиниць продукції. Ознакою якості в даному дослідженні служить відповідність вибраної одиниці товару сертифікатним вимогам.

Суть вибіркового методу в статистиці полягає в тому, що висновки, зроблені на основі вивчення вибірки, розповсюджуються на всю генеральну сукупність.

Слід зазначити, що незалежно від способу організації вибірки вона повинна правильно відображати кількісні співвідношення генеральної сукупності, тобто бути **репрезентативною**. Крім того, всі елементи генеральної сукупності повинні мати однакову ймовірність бути відібраними у вибірку, тобто вибірка повинна бути **випадковою**. Для результатів, що отримані при вибірковому дослідженні, необхідна перевірка на точність і статистичну значущість; спосіб формування вибірки та її об’єм повинні відповідати певному методу обробки даних.

2.2. Статистичні ряди та їх графічне зображення

Припустимо, що необхідно вивчити деяку ознаку генеральної сукупності X, для чого було проведено n вимірювань цієї ознаки і складено вибірку її значень $\{x_1, x_2, ..., x_n\}$ об'єму n.

Різні елементи вибірки називаються **варіантами**. Число n_i, що показує, скільки разів варіанта x_i зустрічається у вибірці, називається **частотою варіанти**. Число w_i, що дорівнює відношенню частоти варіанти n_i до об'єму вибірки n, називається **відносною частотою варіанти** x_i:

$$w_i = \frac{n_i}{n}. \qquad (2.1)$$

Ряд варіант, розташованих в порядку зростання їх значень, називається

варіаційним рядом. Ряд, що містить варіанти і відповідні ним частоти (відносні частоти) називається **статистичним рядом**. Групування кількісних результатів вимірювань у вигляді статистичних рядів є необхідним для застосування статистичних методів аналізу даних і побудови статистичних моделей.

Ознака X є випадковою величиною, а статистичний ряд – емпіричним (тобто отриманим у результаті експерименту або спостережень) законом її розподілу.

Статистичний ряд називається **дискретним**, якщо він є законом розподілу дискретної випадкової величини, та **інтервальним**, якщо він є законом розподілу неперервної випадкової величини.

Дискретний статистичний ряд у загальному вигляді можна представити таблицею (табл. 2.1):

Таблиця 2.1

Варіанти x_i	x_1	x_2	…	x_k
Частоти n_i (відносні частоти w_i)	n_1 (w_1)	n_2 (w_2)	…	n_k (w_k)

де k – кількість варіант.

Інтервальний статистичний ряд у загальному вигляді можна представити таблицею (табл. 2.2):

Таблиця 2.2

Інтервали $[a_i;a_{i+1})$	$[a_1;a_2)$	$[a_2;a_3)$	…	$[a_{\kappa-1};a_\kappa)$
Частоти n_i (відносні частоти w_i)	n_1 (w_1)	n_2 (w_2)	…	n_k (w_k)

де k – кількість інтервалів.

Для статистичних рядів повинні виконуватися рівності: $\sum_{i=1}^{k} n_i = n$, $\sum_{i=1}^{k} w_i = 1$.

Для побудови інтервального статистичного ряду множину значень варіант розбивають на інтервали $[a_i;a_{i+1})$, тобто проводять їх згрупування. Кількість інтервалів k рекомендується розраховувати за формулою Стерджерса:

$$k = 1 + 1,4 \ln n. \tag{2.2}$$

Довжина кожного із інтервалів Δ розраховується за формулою

$$\Delta = \frac{x_{\max} - x_{\min}}{k}, \tag{2.3}$$

де $x_{\max}$, $x_{\min}$ - максимальне і мінімальне значення у варіаційному ряді.

Підраховуючи кількість значень варіант, що потрапили в інтервал $[a_i;a_{i+1})$, отримують частоти n_i для $i = \overline{1,k}$.

Для наочності використовують графічне зображення статистичних рядів у вигляді полігону частот (відносних частот) та, виключно у випадку інтервального ряду, гістограми.

Полігоном частот (відносних частот) називається ламана лінія, що сполучає точки площини з координатами: (x_i; n_i) або (x_i; w_i) для $i=\overline{1,k}$ у разі дискретного статистичного ряду; (c_i; n_i) або (c_i; w_i) у разі інтервального ряду, де c_i – середина і-того інтервалу, $c_i = \dfrac{a_i + a_{i+1}}{2}$.

Гістограмою називається ступінчаста фігура, яка складається з прямокутників з основами, що дорівнюють довжині інтервалів Δ та висотами, що дорівнюють частотам n_i (відносним частотам w_i) на відповідних інтервалах.

За статистичним рядом можна встановити емпіричну функцію розподілу та емпіричну щільність розподілу випадкової величини X.

Емпіричною функцією розподілу називається функція

$$F_n(x) = \frac{1}{n}\sum_{x_i<x} n_i = \sum_{x_i<x} w_i . \qquad (2.4)$$

Відмітимо, що для інтервального ряду указуються не конкретні значення варіант, а тільки їх частоти на інтервалах. Тому емпірична функція розподілу визначена тільки на кінцях інтервалів. Її можна зобразити ламаною, такою, що проходить через точки (a_i; $F_n(a_i)$), де $i=\overline{1,k}$.

Емпіричною щільністю розподілу для інтервального ряду називається функція

$$f_n(x) = \begin{cases} \dfrac{n_i}{n\Delta} = \dfrac{w_i}{\Delta}, \text{ якщо } a_i \le x \le a_{i+1}, \ i=\overline{1,k} \\ 0, \quad \text{якщо } x < a_1 \text{ або } x > a_{k+1}, \ i=\overline{1,k} \end{cases}. \qquad (2.5)$$

ПРИКЛАД 2.1. За даними вибіркового дослідження було отримано розподіл родин за доходом на одного їх члена в умовних одиницях (табл. 2.3). Побудувати інтервальний статистичний ряд, полігон частот, гістограму, полігон відносних частот, емпіричні функцію і щільність розподілу та їх графіки.

Таблиця 2.3

28,92	27,54	22,36	29,09	32,19	26,04	17,06	26,83	24,55	33,22
17,53	30,07	36,27	24,24	26,03	31,05	13,94	14,56	21,40	23,04
13,09	38,84	25,57	22,87	6,11	27,79	25,68	16,30	17,93	24,37
28,92	27,54	22,36	29,06	32,19	26,04	17,06	26,83	24,55	33,22
17,53	30,07	36,27	24,24	26,03	31,05	13,94	14,56	21,40	23,04

Розв'язання. Таблиця 2.3 містить 50 даних, тобто $n = 50$. Для побудови інтервального статистичного ряду знаходимо: кількість інтервалів за формулою (2.2): $k = 1 + 1,4\ln 50 \approx 6,477 \approx 7$; $x_{\max} = 6,11$, $x_{\min} = 38,84$; довжина кожного інтервалу за формулою (2.3): $\Delta = \frac{x_{\max} - x_{\min}}{k} = \frac{38,84 - 6,11}{7} \approx 4,68$. Отже, за початок першого інтервалу обираємо $a_1 = x_{\min} = 6,11$. Тоді $a_2 = a_1 + \Delta = 6,11 + 4,68 = 10,79$. Аналогічно, $a_3 = 15,47; a_4 = 20,15; a_5 = 24,83; a_6 = 29,51; a_7 = 34,19; a_8 = 38,87$.

Підраховуючи кількість варіант, що попали в кожен інтервал, отримаємо інтервальний статистичний ряд (табл. 2.4).

Таблиця 2.4

$[a_i; a_{i+1})$	[6,11; 10,79)	[10,79; 15,47)	[15,47; 20,15)	[20,15; 24,83)	[24,83; 29,51)	[29,51; 34,19)	[34,19; 38,87)
n_i	1	5	6	12	15	8	3

За даними таблиці 2.4 будуємо гістограму (рис. 2.1).

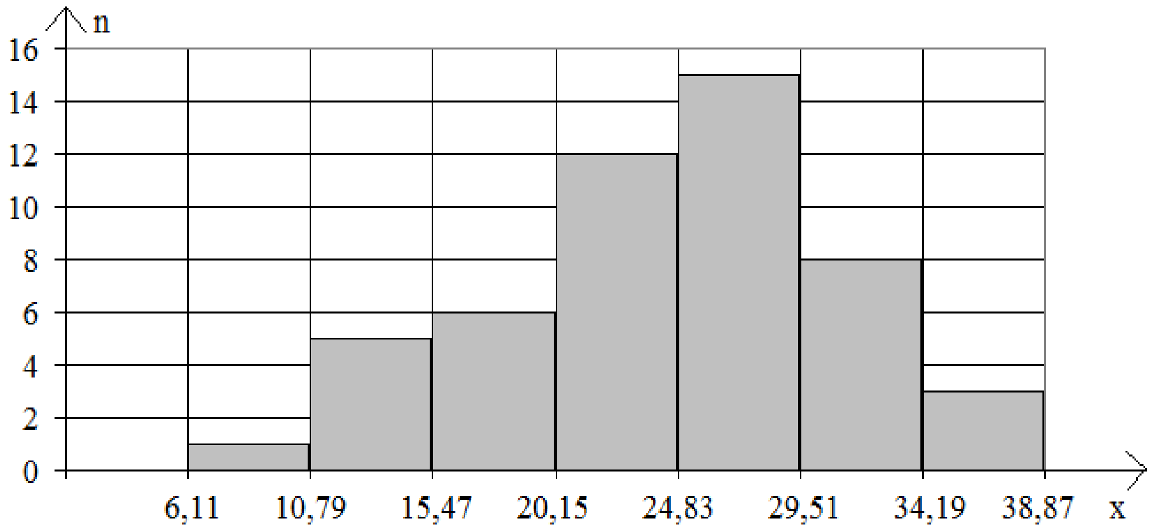

Рис. 2.1. Гістограма

Для побудови полігону частот і полігону відносних частот обчислимо середини кожного інтервалу за формулою $c_i = \frac{a_i + a_{i+1}}{2}$. Отримаємо: $c_1 = \frac{6,11 + 10,79}{2} = 8,45$; $c_2 = 13,13$; $c_3 = 17,81$; $c_4 = 22,49$; $c_5 = 27,17$; $c_5 = 27,17$; $c_6 = 31,85$; $c_7 = 36,51$.

Розрахуємо відносні частоти за формулою (1.1): $w_1 = \frac{n_1}{n} = \frac{1}{50} = 0,02$;

$w_2 = 0{,}1;\ w_3 = 0{,}12;\ w_4 = 0{,}24;\ w_5 = 0{,}3;\ w_6 = 0{,}16;\ w_7 = 0{,}06$.

Результати оформимо у вигляді таблиці (табл. 2.5), останній стовпець якої будемо використовувати для перевірки правильності розрахунків.

Таблиця 2.5

c_i	8,45	13,13	17,81	22,49	27,17	31,85	36,51	Перевірка
n_i	1	5	6	12	15	8	3	$\sum_{i=1}^{k} n_i = 50$
w_i	0,02	0,1	0,12	0,24	0,3	0,16	0,06	$\sum_{i=1}^{k} w_i = 1$

За даними таблиці 3.5 будуємо полігон частот (рис. 1.4) і полігон відносних частот (рис. 2.2).

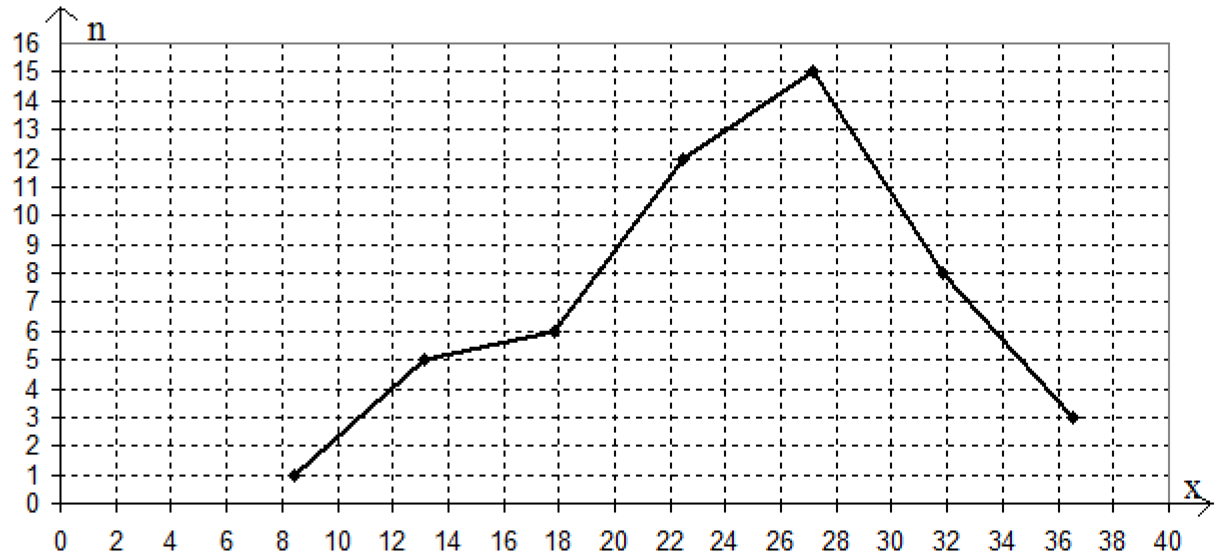

Рис. 2.2. Полігон частот

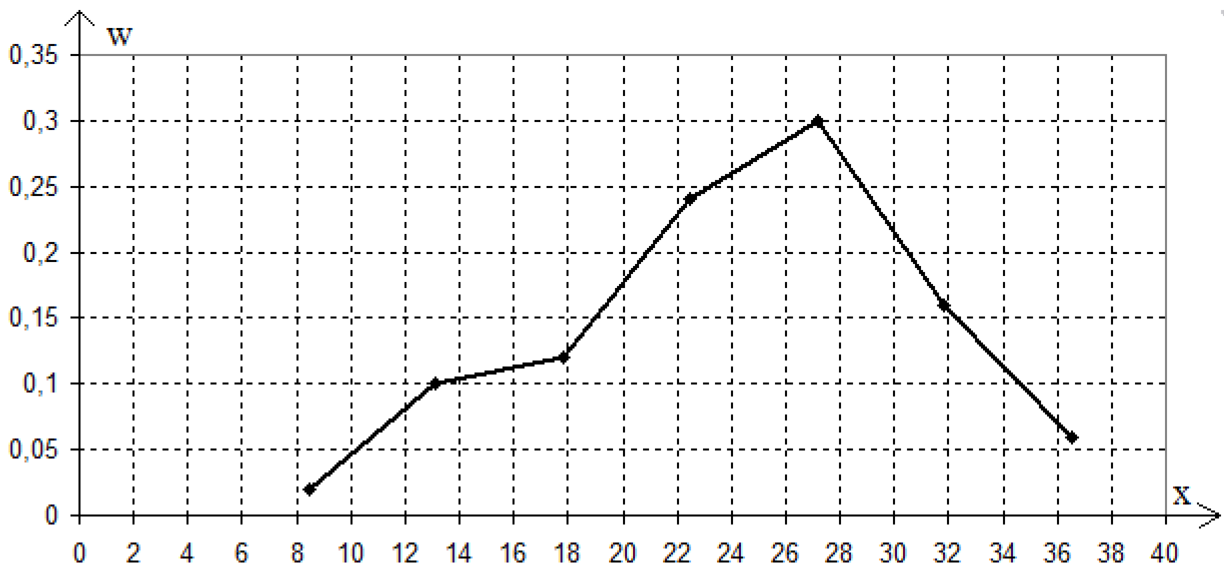

Рис. 2.3. Полігон відносних частот

Для побудови емпіричної функції розподілу обчислимо за формулою (2.4) суму відносних частот варіант, менших за x (тобто $F_n(x) = \sum\limits_{x_i < x} w_i$). Як x беремо ліву границю кожного інтервалу. Отримаємо:

якщо $x = a_1 = 6{,}11$, то $F_n(x) = \sum\limits_{x_i < 6{,}11} w_i = 0$, оскільки таких значень X немає;

якщо $x = a_2 = 10{,}79$, то $F_n(x) = \sum\limits_{x_i < 10{,}79} w_i = w_1 = 0{,}02$;

якщо $x = a_3 = 15{,}47$, то $F_n(x) = \sum\limits_{x_i < 15{,}47} w_i = w_1 + w_2 = 0{,}02 + 0{,}1 = 0{,}12$;

якщо $x = a_4 = 20{,}15$,то $F_n(x) = \sum\limits_{x_i < 20{,}15} w_i = w_1 + w_2 + w_3 = 0{,}02 + 0{,}1 + 0{,}12 = 0{,}24$;

якщо $x = a_5 = 24{,}83$, то $F_n(x) = \sum\limits_{x_i < 24{,}83} w_i = w_1 + w_2 + w_3 + w_4 =$

$$= 0{,}02 + 0{,}1 + 0{,}12 + 0{,}24 = 0{,}48;$$

якщо $x = a_6 = 29{,}51$, то $F_n(x) = \sum\limits_{x_i < 29{,}51} w_i = w_1 + w_2 + w_3 + w_4 + w_5 =$

$$= 0{,}02 + 0{,}1 + 0{,}12 + 0{,}24 + 0{,}3 = 0{,}78;$$

якщо $x = a_7 = 34{,}19$, то $F_n(x) = \sum\limits_{x_i < 34{,}19} w_i = w_1 + w_2 + w_3 + w_4 + w_5 + w_6 =$

$$= 0{,}02 + 0{,}1 + 0{,}12 + 0{,}24 + 0{,}3 + 0{,}16 = 0{,}94;$$

якщо $x = a_8 = 38{,}87$, то $F_n(x) = \sum\limits_{x_i < 38{,}87} w_i = w_1 + w_2 + w_3 + w_4 + w_5 + w_6 + w_7 =$

$$= 0{,}02 + 0{,}1 + 0{,}12 + 0{,}24 + 0{,}3 + 0{,}16 + 0{,}06 = 1.$$

Останнє значення функції розподілу позначає, що всі значення X менше за 38,87.

Для знаходження емпіричної щільності розподілу обчислимо $f_n(x) = \dfrac{w_i}{\Delta}$ за формулою (2.5). Результати обчислень надано у таблиці 1.8.

Таблиця 2.6

a_i	6,11	10,79	15,47	20,15	24,83	29,51	34,19	38,87
w_i	0,02	0,1	0,12	0,24	0,3	0,16	0,06	$\sum_{i=1}^{k} w_i = 1$
F_i	0	0,02	0,12	0,24	0,48	0,78	0,94	1
$[a_i; a_{i+1})$	[6,11; 10,79)	[10,79; 15,47)	[15,47; 20,15)	[20,15; 24,83)	[24,83; 29,51)	[29,51; 34,19)	[34,19; 38,87)	
f_i	0,0043	0,0214	0,0256	0,0513	0,0641	0,0342	0,0128	

За даними таблиці 2.6 будуємо графік емпіричної функції розподілу (рис. 2.4) та емпіричної щільності розподілу (рис. 2.5).

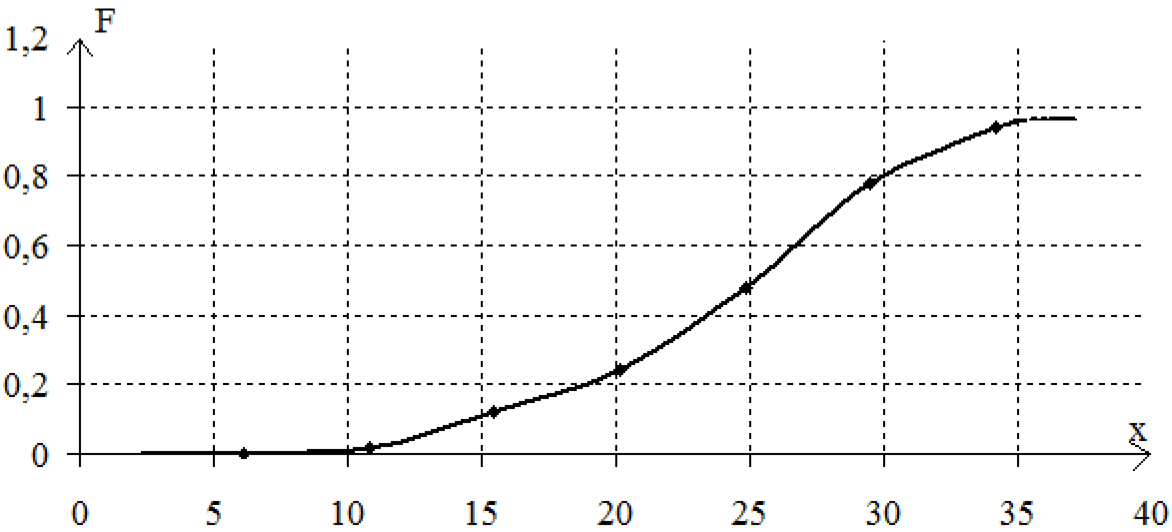

Рис. 2.4. Графік емпіричної функції розподілу

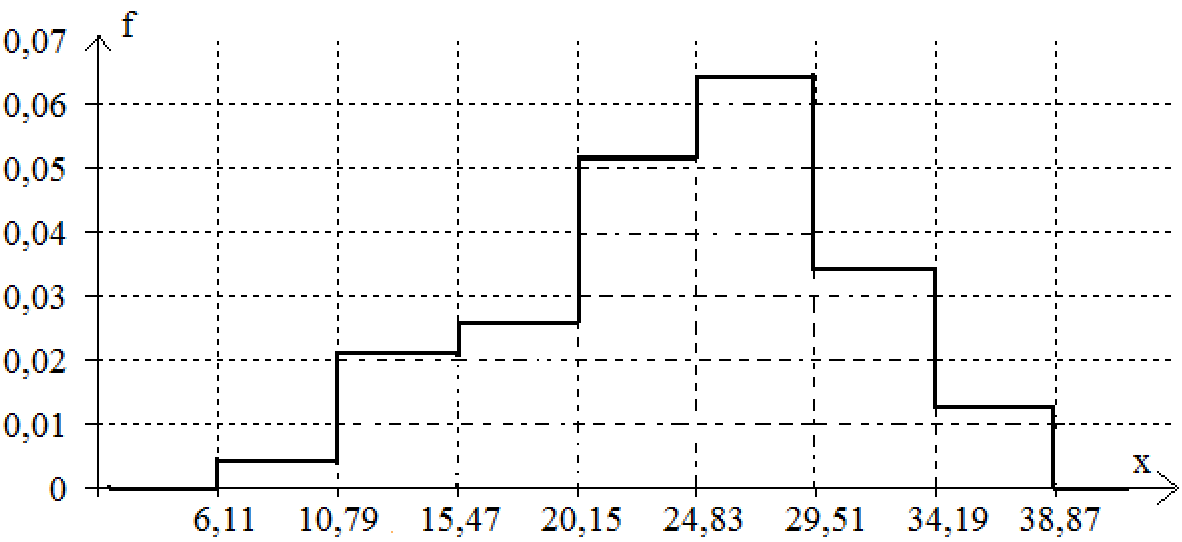

Рис. 2.5. Графік емпіричної щільності розподілу

2.3. Числові характеристики статистичних рядів

2.3.1. Поняття про оцінки параметрів

Генеральну сукупність X можна розглядати як випадкову величину. Тоді вибірка значень X – це емпіричний закон розподілу випадкової величини. Для дискретних і неперервних випадкових величин визначені числові характеристики, основними з яких є математичне сподівання, дисперсія і середнє квадратичне відхилення. Числові характеристики випадкових величин часто є параметрами їх розподілів. Аналогічно числові характеристики визначені і для статистичних рядів, це – вибіркове середнє, вибіркове середнє геометричне, вибіркова дисперсія, вибіркове середнє квадратичне відхилення і т. ін.

У прикладних задачах часто необхідно визначити за даними вибірки закон розподілу випадкової величини, що є однією із основних задач математичної статистики. При цьому вибіркове середнє вважається **оцінкою** (аналогом) математичного сподівання, вибіркова дисперсія – оцінкою дисперсії, вибіркове середнє квадратичне відхилення – оцінкою середнього квадратичного відхилення. При цьому виникає питання: наскільки правомірні такі оцінки?

Оцінки параметрів повинні відповідати таким вимогам.

Незсуненість. Це позначає, що при проведенні великої кількості спостережень (вимірювань) з вибірками одного об’єму оцінка параметру, отримана з кожної вибірки, прямує до істинного значення цього параметру генеральної сукупності.

Спроможність. Зі збільшенням об’єму вибірки оцінка прямує до значення відповідного параметру генеральної сукупності з ймовірністю, що дорівнює 1.

Достатність. Оцінка містить всю необхідну інформацію.

Ефективність. Оцінки, отримані за вибірками однакового об’єму, мають мінімальну дисперсію.

Зауваження. При використанні оцінок необхідно пам’ятати, що вони отримуються тільки при певних передмовах і, відповідно, дійсні тільки при виконанні цих передмов.

Для оцінювання параметрів розподілу за даними вибірки зазвичай використовується метод максимальної правдоподібності. Але він застосовується тільки тоді, коли відомий закон розподілу.

2.3.2. Числові характеристики положення

Основною числовою характеристикою статистичного ряду є середня арифметична, звана також вибірковим середнім.

Вибірковим середнім називається величина $\bar{x}$ яка обчислюється за формулою:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \sum_{i=1}^{k} x_i w_i . \qquad (2.6)$$

У разі інтервального статистичного ряду як x_i вибирається середина *і*-го інтервалу.

Якщо вибірка містить незгруповані дані, то вибіркове середнє розраховується за формулою:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (2.7)$$

<u>**Зауваження.**</u> Оскільки статистичний ряд є емпіричним законом розподілу величини *X*, то вибіркове середнє зазвичай вважається аналогом або оцінкою математичного сподівання випадкової величини *X*. Хоча це твердження безумовно вірне тільки для нормального закону розподілу.

Вибірковим середнім геометричним називається величина $\overline{x_G}$, яка обчислюється за формулою:

$$\overline{x_G} = \sqrt[n]{\prod_{i=1}^{n} x_i} . \qquad (2.8)$$

Середнє геометричне застосовується як центральна тенденція тоді, коли значення *X* змінюються с постійним співвідношенням між попереднім і наступним значеннями, тобто якщо $\frac{x_i}{x_{i-1}} = \frac{x_{i+1}}{x_i}$ (наприклад, збільшення капіталовкладень, експлуатаційні витрати і т. ін.).

Модою ***Mo*** називається таке значення величини *X*, яке спостерігається у вибірці з найбільшою частотою. У випадку інтервального статистичного ряду мода розраховується за формулою:

$$Mo = x_{Mo} + \frac{\Delta\left(n_{Mo} - n_{Mo-1}\right)}{2n_{Mo} - n_{Mo-1} - n_{Mo+1}}, \qquad (2.9)$$

де x_{Mo} – початок інтервалу, якому відповідає найбільша частота (такий інтервал називається модальним);

n_{Mo} – частота у модальному інтервалі;

n_{Mo-1} , n_{Mo+1} – частоти в попередньому і наступному інтервалах відповідно.

<u>**Зауваження.**</u> Мода не застосовується тоді, коли гістограма або полігон частот показують наявність двох або більше вершин („піків”).

Медіаною ***Me*** називається таке значення величини *X*, яке розділяє вибірку, елементи якої розташовані у порядку зростання, на дві рівні за об’ємом частини.

Якщо це вибірка значень дискретної випадкової величини, то медіаною є те її значення, яке розташовано всередині, якщо кількість членів ряду непарна: тобто це елемент з номером $\frac{n+1}{2}$. Якщо кількість елементів вибірки парна, то медіана дорівнює середньому арифметичному її членів з номерами $\frac{n}{2}$ та $\frac{n}{2}+1$.

Якщо розглядається вибірка неперервної випадкової величини, то медіана розраховується формулою:

$$Me = X_{Me} + \frac{\Delta\left(\frac{n}{2} - n_x^{\max}\right)}{n_m}, \tag{2.10}$$

де X_{Me} – фактична нижня границя медіанного інтервалу;

$n_x^{\max}$ – сума частот, що накопичена до початку медіанного інтервалу;

n_m – частота в медіанному інтервалі.

<u>**Зауваження.**</u> На значення медіани не впливають змінення значень крайніх елементів впорядкованої вибірки, тому її часто застосовують як центральну тенденцію тоді, коли крайні елементи вибірки значно відрізняються від інших її елементів.

2.3.3. Числові характеристики розсіювання

Варіаційним розмахом R називається різниця між максимальним і мінімальним елементом вибірки:

$$R = x_{\max} - x_{\min}. \tag{2.11}$$

Вибірковою дисперсією S^2 називається середня арифметична квадратів відхилень варіант від їх вибіркової середньої:

$$S^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i = \sum_{i=1}^{k}(x_i - \bar{x})^2 w_i \tag{2.12}$$

$$\text{або } S^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i. \tag{2.13}$$

Дисперсія є показником розсіювання елементів вибірки відносно їх середнього значення. Вибіркова дисперсія, отримана за формулою (2.13), називається незсуненою оцінкою дисперсії генеральної сукупності.

Різниця дисперсій, отриманих за формулами (2.12) та (2.13) зазвичай невелика, однак може вплинути на точність оцінок. Тому, якщо відомо точне значення математичного сподівання, використовують формулу (2.12), в іншому випадку – формулу (2.13).

Якщо дані не згруповані, то дисперсію можна розрахувати за формулою:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (2.14)$$

Вибірковим середнім квадратичним відхиленням S називається величина, що дорівнює кореню з вибіркової дисперсії:

$$S = \sqrt{S^2}\,. \qquad (2.15)$$

Вибіркове середнє квадратичне відхилення теж є показником розсіювання елементів вибірки відносно їх середнього значення, але, на відміну від дисперсії, воно має ті одиниці вимірювання, які мають елементи вибірки.

Коефіцієнтом варіації *v* називається величина, що дорівнює процентному відношенню вибіркового середнього квадратичного відхилення до модуля вибіркової середньої:

$$v = \frac{S}{|\bar{x}|}\cdot 100\%\ (\bar{x} \neq 0)\,. \qquad (2.16)$$

Якщо коефіцієнт варіації більший за 100%, то елементи вибірки неоднорідні і вона не може бути використана у подальших дослідженнях.

ПРИКЛАД 2.2. За даними вибіркового дослідження відомі ціни x_i певного товару у різних торгівельних організаціях (табл. 2.7). Знайти всі можливі числові характеристики за даними таблиці.

Таблиця 2.7

Організація	1	2	3	4	5	6	7	8
Ціна	100	110	115	125	140	145	145	150

Розв'язання. За незгрупованими даними таблиці 2.7 можна знайти: вибіркове середнє за формулою (2.6), медіану, розмах варіації за формулою (2.11), дисперсію за формулою (2.14), вибіркове середнє квадратичне відхилення за формулою (2.15), коефіцієнт варіації за формулою (2.16). Кількість елементів вибірки $n = 8$. Вибіркове середнє $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i = \frac{100+110+115+125+140+145+145+150}{8} = 128{,}75$.

Кількість елементів вибірки парна, тому медіана дорівнює середньому арифметичному її членів з номерами $\frac{n}{2}$ та $\frac{n}{2}+1$: $x_{n/2} = x_4 = 125$; $x_{n/2+1} = x_5 = 140$; $Me = \frac{x_4 + x_5}{2} = \frac{125+140}{2} = 132{,}5$.

Розмах варіації $R = x_{\max} - x_{\min} = 150 - 100 = 50$.

Для розрахунку вибіркової дисперсії складемо таблицю 2.10.

Таблиця 2.8

x_i	100	110	115	125	140	145	145	150
$x_i - \bar{x}$	-28,75	-18,75	-13,75	-3,75	11,25	16,25	16,25	21,25
$\left(x_i - \bar{x}\right)^2$	826,563	351,56	189,06	14,063	126,56	264,06	264,06	451,56

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{826{,}563 + 351{,}56 + 189{,}06 + 14{,}063 + 126{,}56 + 264{,}06 + 264{,}06 + 451{,}56}{8}$$
=310,938.

Вибіркове середнє квадратичне відхилення $S = \sqrt{S^2} = \sqrt{310{,}938} \approx 17{,}63$.

Коефіцієнт варіації $v = \frac{S}{\left|\bar{x}\right|} \cdot 100\% = \frac{17{,}63}{128{,}75} \cdot 100\% \approx 13{,}69\%$.

ПРИКЛАД 2.3. За даними вибіркового дослідження відомі ціни x_i певного товару у різних торгівельних організаціях (табл. 2.9). Визначити центральну тенденцію за даними таблиці.

Таблиця 2.9

Організація	1	2	3	4	5	6	7	8	9
Ціна	100	110	115	125	140	145	145	150	450

Розв’язання. Оскільки дані таблиці 2.9 незгруповані, то оцінкою центральної тенденції може слугувати вибіркове середнє або медіана.

Знайдемо вибіркове середнє за формулою (2.6):

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{100 + 110 + 115 + 125 + 140 + 145 + 145 + 150 + 450}{9} \approx 164{,}4.$$

Кількість елементів вибірки непарна, тому медіана дорівнює її члену з номером $\frac{n+1}{2}$: $Me = x_{n+1/2} = x_5 = 140$.

Зобразимо дані таблиці графічно, вкажемо на діаграмі положення середнього і медіани (рис. 2.6).

На рис. 2.6 видно, що як центральну тенденцію вибірки слід взяти медіану.

<u>**Зауваження.**</u> Цей приклад показує, що у випадках наявності у вибірці даних, які сильно відрізняються один від одного, або даних, які сильно відрізняються від всіх останніх (так званих викидів), медіана є більш усталеною оцінкою центральної тенденції, ніж вибіркове середнє.

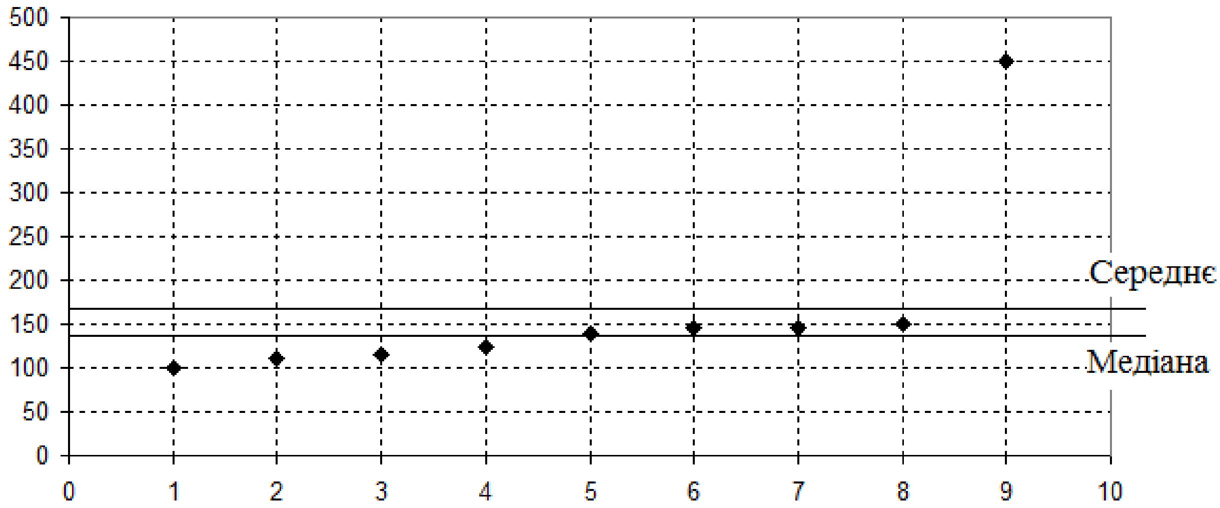

Рис. 2.6. Положення середнього і медіани відносно вибіркових даних

ПРИКЛАД 2.4. За даними вибіркового дослідження відома кількість людей, що відвідували лікарню протягом року. Дані згруповані залежно від віку відвідувачів (табл. 2.10). Знайти всі можливі числові характеристики за даними таблиці.

Таблиця 2.10

Вік	20-29	30-39	40-49	50-59	60-69
Кількість відвідувань	45	36	175	361	825

Розв'язання. Позначимо X – вік відвідувачів лікарні, n – загальна кількість відвідувань, $n = 1442$; n_i – кількість відвідувань залежно від віку; k – кількість досліджуваних вікових груп, $k = 5$. Тоді відповідно даним таблиці 2.11 отримаємо інтервальний статистичний ряд (табл. 2.11).

Таблиця 2.11

$[a_i;a_{i+1}]$	20-29	30-39	40-49	50-59	60-69
n_i	45	36	175	361	825

Розрахуємо вибіркове середнє за формулою (2.6). Для зручності розрахунки оформимо у вигляді таблиці (табл. 2.12).

Таблиця 2.12

$[a_i;a_{i+1}]$	20-29	30-39	40-49	50-59	60-69	Суми
x_i	24,5	34,5	44,5	54,5	64,5	
n_i	45	36	175	361	825	1442
$x_i n_i$	1102,5	1242	7787,5	19675	53213	83019

Тоді $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{83019}{1442} \approx 57,57$.

Моду обчислимо за формулою (2.10) враховуючи, що:

x_{Mo} – початок модального інтервалу (якому відповідає найбільша частота), x_{Mo} =60;

n_{Mo} – частота у модальному інтервалі, n_{Mo} = 825;

n_{Mo-1} , n_{Mo+1} – частоти в попередньому і наступному інтервалах відповідно, n_{Mo-1} = 361, n_{Mo+1} = 0 (оскільки модальний інтервал є останнім);

Δ - довжина інтервалу, Δ = 69-59 = 9.

Отже, $Mo = x_{Mo} + \dfrac{\Delta\left(n_{Mo} - n_{Mo-1}\right)}{2n_{Mo} - n_{Mo-1} - n_{Mo+1}} = 60 + \dfrac{10(825-361)}{2\cdot 825 - 365 - 0} = 65$.

Медіану обчислимо за формулою (2.11), враховуючи, що:

X_{Me} – фактична нижня границя медіанного інтервалу, X_{Me} = 60 (оскільки всього даних 1442, то їх половина 1442:2=721; у статистичному ряді до початку останнього інтервалу міститься 617 даних, тому медіана повинна знаходитися в останньому інтервалі);

$n_x^{\max}$ – сума частот, що накопичена до початку медіанного інтервалу, $n_x^{\max}$ =617;

n_m – частота в медіанному інтервалі, n_m =825.

Отже, $Me = X_{Me} + \dfrac{\Delta\left(\dfrac{n}{2} - n_x^{\max}\right)}{n_m} = 60 + \dfrac{10\left(\dfrac{1442}{2} - 617\right)}{825} = 61,26$.

Дисперсію обчислимо за формулою (2.13). Для зручності розрахунки оформимо у вигляді таблиці (табл. 2.13).

Таблиця 2.13

x_i	24,5	34,5	44,5	54,5	64,5	Суми
n_i	45	36	175	361	825	1442
$\left(x_i - \bar{x}\right)^2$	1093,6249	1190,25	1980,25	2970,25	4160,25	11394,62

Тоді $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i = \dfrac{11394,62}{1442-1} = 7,9$.

Вибіркове середнє квадратичне відхилення за формулою (2.15) дорівнює: $S = \sqrt{S^2}$ = 2,81.

Коефіцієнт варіації за формулою (2.16) дорівнює: $v = \dfrac{S}{\left|\bar{x}\right|}\cdot 100\% = \dfrac{2,81}{57,57}\cdot 100\% \approx 4,89\%$.

2.4. Довірчі інтервали і довірча ймовірність

Однією з основних задач математичної статистики є оцінка числових характеристик (параметрів) генеральної сукупності за вибірковими даними.

Для вибірки можна обчислити такі числові характеристики, як: вибіркове середнє, мода, медіана, вибіркова дисперсія та вибіркове середнє квадратичне відхилення. Для генеральної сукупності часто визначаються не самі ці параметри, а довірчі інтервали.

Довірчим інтервалом для певного параметру генеральної сукупності називається такий числовий інтервал, в межах якого знаходиться цей параметр. Ймовірність, з якою довірчий інтервал захватить істинне значення параметру, називається **довірчою ймовірністю** або **рівнем надійності** і позначається γ.

Значення довірчої ймовірності обирає дослідник залежно від того, яку ступінь точності розрахунків вимагає дослідження. Зазвичай це значення знаходиться в інтервалі від 0,9 до 0,999. Якщо вимоги точності дуже високі, то для довірчої ймовірності обирається значення 0,999; якщо підвищені – 0,99; звичайні – 0,95; знижені – 0,9.

Довірчі інтервали розраховуються з урахуванням певних вимог до генеральної сукупності. Зазвичай це вимога нормального розподілу її даних.

2.4.1. Довірчий інтервал для генерального середнього при відомій генеральної дисперсії

Нехай X – генеральна сукупність, що підкоряється нормальному закону розподілу; σ^2 – відома генеральна дисперсія; $\{x_1, x_2, ..., x_n\}$ – вибірка з генеральної сукупності об'єму *n*; $\overline{x}$ - вибіркове середнє. Потрібно знайти довірчий інтервал для генерального середнього *a* із заданим рівнем надійності γ.

Шуканий довірчий інтервал знаходиться за формулою:

$$\overline{x} - z_{\frac{1-\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}} < a < \overline{x} + z_{\frac{1-\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \qquad (2.17)$$

де значення $z_{\frac{1-\gamma}{2}}$ знаходиться з таблиці (табл. 2.16) або за допомогою вбудованої функції Excel НОРМСТОБР(γ). Величина $z_{\frac{1-\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ є шириною довірчого інтервалу.

Таблиця 2.14

γ	0,4	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005	0,001
$z_{\frac{1-\gamma}{2}}$	0,253	0,675	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,090

ПРИКЛАД 2.5. Автомат, що фасує чай в пачки, працює зі стандартним відхиленням σ =5 г. Проведено вибірку об'ємом *n*=30 пачок. Середня вага пачки чаю у вибірки $\overline{x}$ =101 г. Знайти довірчий інтервал для середньої ваги

пачки чаю в генеральної сукупності із рівнем надійності γ=0,95. Знайти об'єм вибірки, якщо потрібна ширина довірчого інтервалу ± 1 грам.

Розв'язання. Оскільки γ=0,95, то $\frac{1-\gamma}{2}=\frac{1-0,95}{2}=0,025$. З таблиці 1.16 знайдемо $z_{\frac{1-\gamma}{2}}=z_{0,025}=1,96$.

Тоді ширина довірчого інтервалу: $z_{\frac{1-\gamma}{2}}\cdot\frac{\sigma}{\sqrt{n}}=1,96\cdot\frac{5}{\sqrt{30}}\approx 1,79$ і за формулою (2.17) маємо довірчий інтервал:

$\bar{x}-1,79<a<\bar{x}+1,79$; $\quad 101-1,79<a<101+1,79$; $\quad 99,21<a<102,79$.

Отже, середня вага пачки чаю знаходиться в інтервалі від 99,21 до 102,79.

Знайдемо об'єм вибірки, необхідний для того, щоб ширина довірчого інтервалу дорівнювала 1 грам, тобто $z_{\frac{1-\gamma}{2}}\cdot\frac{\sigma}{\sqrt{n}}=1$. Знайдемо n з отриманого рівняння: $1,96\cdot\frac{5}{\sqrt{n}}=1\Rightarrow\sqrt{n}=1,96\cdot 5=9,8\Rightarrow n=9,8^2=96,04$. Отже, мінімальний об'єм вибірки для отримання довірчого інтервалу шириною 1 грам дорівнює 97 пачок.

Зауваження. У прикладі для розрахунку нового довірчого інтервалу потрібно знаходити вибіркове середнє для вибірки об'єму 97 пачок, а не середнє арифметичне середніх для вибірок об'єму 30 і 67 пачок.

1.4.2. Довірчий інтервал для генерального середнього при невідомій генеральної дисперсії

Нехай X – генеральна сукупність, що підкоряється нормальному закону розподілу; генеральна дисперсія σ^2 невідома; $\{x_1, x_2, ..., x_n\}$ – вибірка з генеральної сукупності об'єму n; $\bar{x}$ - вибіркове середнє; S – вибіркове середнє квадратичне відхилення. Потрібно знайти довірчий інтервал для генерального середнього a із заданим рівнем надійності γ.

Шуканий довірчий інтервал знаходиться за формулою:

$$\bar{x}-t_{\frac{1-\gamma}{2},n-1}\cdot\frac{S}{\sqrt{n-1}}<a<\bar{x}+t_{\frac{1-\gamma}{2},n-1}\cdot\frac{S}{\sqrt{n-1}}, \qquad (2.18)$$

де значення $t_{\frac{1-\gamma}{2},n-1}$ знаходиться з таблиці розподілу Стьюдента, яка є у будь-яких статистичних довідниках

2.4.3. Довірчий інтервал для генеральної частки

В прикладних дослідженнях часто потрібно визначити частку об'єктів, що мають певну властивість.

Частка об'єктів генеральної сукупності, що має певну властивість, називається **генеральною часткою**. Частка об'єктів вибірки, що має певну властивість, називається **вибірковою часткою**.

Нехай X – генеральна сукупність, що підкоряється нормальному закону розподілу; $\{x_1, x_2, ..., x_n\}$ – вибірка з генеральної сукупності об'єму n; m – кількість елементів вибірки, що мають задану властивість; $w = \frac{m}{n}$ – вибіркова частка. Потрібно знайти довірчий інтервал для генеральної частки W із заданим рівнем надійності γ.

Шуканий довірчий інтервал знаходиться за формулою:

$$w - z_{\frac{1-\gamma}{2}} \cdot \sqrt{\frac{w(1-w)}{n}} < W < w + z_{\frac{1-\gamma}{2}} \cdot \sqrt{\frac{w(1-w)}{n}}, \qquad (2.19)$$

де значення $z_{\frac{1-\gamma}{2}}$ знаходиться з таблиці 2.16 або за допомогою вбудованої функції Excel НОРМСТОБР(γ). Величина $z_{\frac{1-\gamma}{2}} \cdot \sqrt{\frac{w(1-w)}{n}}$ є шириною довірчого інтервалу.

Зауваження. Формула (2.19) використовується тоді, коли $nw \geq 5, \quad n(1-w) \geq 5$.

ПРИКЛАД 2.6. Проведено вибірку об'ємом n=2000 одиниць продукції. Серед обраних 150 одиниць виявилися бракованими. Знайти довірчий інтервал для генеральної долі бракованих виробів із рівнем надійності γ=0,95.

Розв'язання. Оскільки γ=0,95, то $\frac{1-\gamma}{2} = \frac{1-0,95}{2} = 0,025$. З таблиці 1.16 знайдемо $z_{\frac{1-\gamma}{2}} = z_{0,025} = 1,96$.

Знайдемо вибіркову частку бракованих виробів: $m = 150;\ w = \frac{m}{n} = \frac{150}{2000} = 0,075$.

Перевіримо можливість знаходження довірчого інтервалу:
$nw = 2000 \cdot 0,075 = 150 \geq 5, \quad n(1-w) = 2000(1-0,075) = 2000 \cdot 0,925 = 1850 \geq 5$.

Тоді ширина довірчого інтервалу: $z_{\frac{1-\gamma}{2}} \cdot \sqrt{\frac{w(1-w)}{n}} = 1,96\sqrt{\frac{0,075(1-0,075)}{2000}} \approx 0,012$ і за формулою (2.19) маємо довірчий інтервал:

$w - 0,012 < W < w + 0,012$; $0,075 - 0,012 < W < 0,075 + 0,012$; $0,063 < W < 0,087$.

Отже, доля бракованих виробів в генеральній сукупності знаходиться в межах від 0,063 до 0,087, тобто складає від 6,3% до 8,7% від об'єму продукції.

РОЗДІЛ 3 ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ

3.1. Поняття про статистичні гіпотези

При застосуванні певних статистичних методів обробки даних вибірки часто ставляться вимоги до розподілу даних або до числових характеристик.

Статистичною гіпотезою називається будь-яке припущення про властивості досліджуваної величини, висунуте на основі статистичних даних.

За змістом статистичні гіпотези можна віднести до таких типів:

1) Гіпотези про вид закону розподілу досліджуваної величини.
2) Гіпотези про числові характеристики досліджуваної величини.
3) Гіпотези про рівність числових характеристик досліджуваних величин.
4) Гіпотези про належність досліджуваних величин до одній генеральної сукупності.
5) Гіпотези про вид моделі, що описує взаємозв'язок між досліджуваними величинами.
6) Гіпотези про належність досліджуваних величин до одного класу.

Статистичні гіпотези позначаються латинськими буквами H_0, H_1, і т.д. Гіпотеза H_0 формулюється як основна в тому сенсі, що при перевірці бажано було б встановити її справедливість. Основної гіпотезі H_0 протиставляються інші гіпотези H_1, H_2, …, які називаються альтернативними.

Прийняття основної або одній з альтернативних гіпотез здійснюється на основі дослідження статистичних даних. Дослідження проводиться за певним **критерієм**, який обирається відповідно до змісту гіпотези і виду наявних статистичних даних.

Якщо сформульовані гіпотези H_0 – основна та H_1 альтернативна (конкуруюча) і обраний критерій перевірки справедливості основної гіпотези, то прийняття H_0 позначає відкидання H_1, а відкидання H_0 позначає справедливість H_1.

Оскільки прийняття гіпотези здійснюється на основі статистичних даних, то завжди існує ймовірність помилки.

Ймовірність відкидання гіпотези H_0, якщо вона справедлива, називається ймовірністю помилки першого роду або **рівнем значущості** і позначається α. Величина 1-α є ймовірністю прийняття справедливої гіпотези і називається **рівнем довіри**. Ймовірність прийняття гіпотези H_0, якщо вона не вірна, називається ймовірністю помилки другого роду і позначається β. Величина 1-β є ймовірністю відкидання невірної гіпотези і називається **потужністю критерію**.

Чим менше рівень значущості, тим менше ймовірність відкинути вірну гіпотезу. Зазвичай рівень значущості обирається дослідником рівним 0,1; 0,05; 0,01 або 0,001. Якщо, наприклад, обраний рівень значущості α =0,01, то риск відкинути вірну гіпотезу виникає в одному випадку із ста.

Зауваження. Перевірка статистичної гіпотези не надає точного висновку щодо її вірності або невірності. <u>Прийняття гіпотези позначає, що на прийнятому рівні значущості вона не протиричіть статистичним даним</u>.

Перевірка статистичних гіпотез зазвичай здійснюється за такими **етапами**:

1) Висунення припущень про вид розподілу досліджуваної величини (величин) або про її числові характеристики.
2) Формулювання статистичних гіпотез.
3) Вибір критерію перевірки відповідно до змісту гіпотез і статистичних даних.
4) Вибір рівня значущості залежно від вимог до точності результатів дослідження.
5) Розрахунок значення обраного критерію за статистичними даними.
6) Порівняння розрахованого значення критерію з його критичним значенням і прийняття або відкидання основної гіпотези.

3.2. Перевірка гіпотези про вид закону розподілу досліджуваної величини

Перевірка гіпотези про вид закону розподілу досліджуваної величини має велике значення для прикладних досліджень. Необхідність такої перевірки виникає при виборі критерію, оскільки для багатьох з них висувається вимога нормального розподілу статистичних даних. Означені гіпотези перевіряються при проектуванні систем масового обслуговування, перевірки якості продукції або праці і т. ін.

Припустимо, що з деякої генеральної сукупності X, яка розглядається як випадкова величина, обрана вибірка $\{x_1, x_2, \dots x_n\}$. За даними вибірки побудовано статистичний ряд (табл. 2.1), що містить варіанти x_i та відповідні частоти n_i, $i=\overline{1,k}$, k – кількість варіант у випадку дискретного ряду. У випадку інтервального ряду x_i – середини інтервалів, k – кількість інтервалів.

Таблиця 3.1

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

Отриманий на основі вибіркових даних статистичний ряд називається **емпіричним законом розподілу** величини *X*.

За даними статистичного ряду можна знайти числові характеристики, які є вибірковими параметрами закону розподілу *X*. Вид закону розподілу визначається відповідно до умов здобуття вибірки або залежно від виду графіка емпіричної щільності розподілу (гістограми) у випадку неперервної випадкової величини *X* і полігону частот, якщо величина *X* дискретна. Параметри обраного закону розподілу заміняються відповідними вибірковими параметрами.

Закон розподілу випадкової величини *X*, параметрами якого є відповідні вибіркові числові характеристики, називається **теоретичним законом розподілу**.

При здійсненні такої заміни немає впевненості, що закон розподілу обраний правильно. Тому розроблено процедуру, яка дозволяє оцінити ступінь відповідності обраного закону даним вибірки. Критерії здійснення такої перевірки називаються **критерії згоди**, найбільш відомим з яких є **критерій Пірсона** χ^2 (хи-квадрат).

Критерій Пірсона χ^2 обчислюється за формулою:

$$\chi^2=\sum_{i=1}^{k}\frac{\left(n_i-n_i'\right)^2}{n_i'}, \qquad (3.1)$$

де n_i' – частоти, отримані за теоретичним законом розподілу (теоретичні частоти).

З формули (3.1) видно, що у випадку, коли відповідні теоретичні та емпіричні частоти співпадають, χ^2=0. Тобто чим ближче χ^2 до нуля, тим краще узгоджуються вибіркові дані та обраний теоретичний закон розподілу.

Розраховане значення критерію χ^2 порівнюється з його критичним значенням $\chi^2_{\alpha,l}$, яке знаходиться за статистичними таблицями або за допомогою вбудованої статистичної функції Excel ХИ2ОБР(α, l). Параметрами функції ХИ2ОБР є: α – рівень значущості; l – ступені волі, $l=k-r-1$, де k – кількість груп емпіричного розподілу, r – кількість параметрів теоретичного розподілу (наприклад, для нормального розподілу r=2, оскільки параметрів два – a і σ). Якщо $\chi^2<\chi^2_{\alpha,l}$, то гіпотеза про закон розподілу приймається. У противному випадку гіпотеза відкидається.

Зауваження. У деяких статистичних таблицях критичне значення χ^2 надається залежно від рівня довіри γ, $\gamma = 1 - \alpha$.

Отже, перевірка гіпотези про закон розподілу величини *X* здійснюється за такими **етапами**:

1) З генеральної сукупності *X* здобувається вибірка і будується статистичний ряд.
2) Висувається гіпотеза про закон розподілу випадкової величини *X*.
3) Знаходяться вибіркові параметри обраного закону розподілу.
4) Розраховуються теоретичні частоти.
5) Розраховується критерій χ^2 за формулою (3.1).
6) Обирається рівень значущості α (або рівень довіри γ) і знаходиться критичне значення $\chi^2_{\alpha,l}$ (або $\chi^2_{\gamma,l}$).
7) Порівнюються розраховане і критичне значення критерію χ^2 і робиться висновок про справедливість висунутої гіпотези.

ПРИКЛАД 3.1. За наданим інтервальний статистичним рядом (табл. 2.2) знайти закон розподілу випадкової величини *X*.

Таблиця 3.2

$[a_i; a_{i+1})$	$[-2;-1,2)$	$[-1,2;-0,4)$	$[-0,4;0,4)$	$[0,4;1,2)$	$[1,2;2)$
n_i	6	11	21	7	5

Розв'язання. Для визначення виду закону розподілу побудуємо гістограму за даними таблиці 3.2 (рис. 3.1). За видом гістограми висуваємо гіпотезу про нормальний закон розподілу даної випадкової величини:

H_0 – випадкова величина X розподілена за нормальним законом;

H_1 – випадкова величина X не розподілена за нормальним законом.

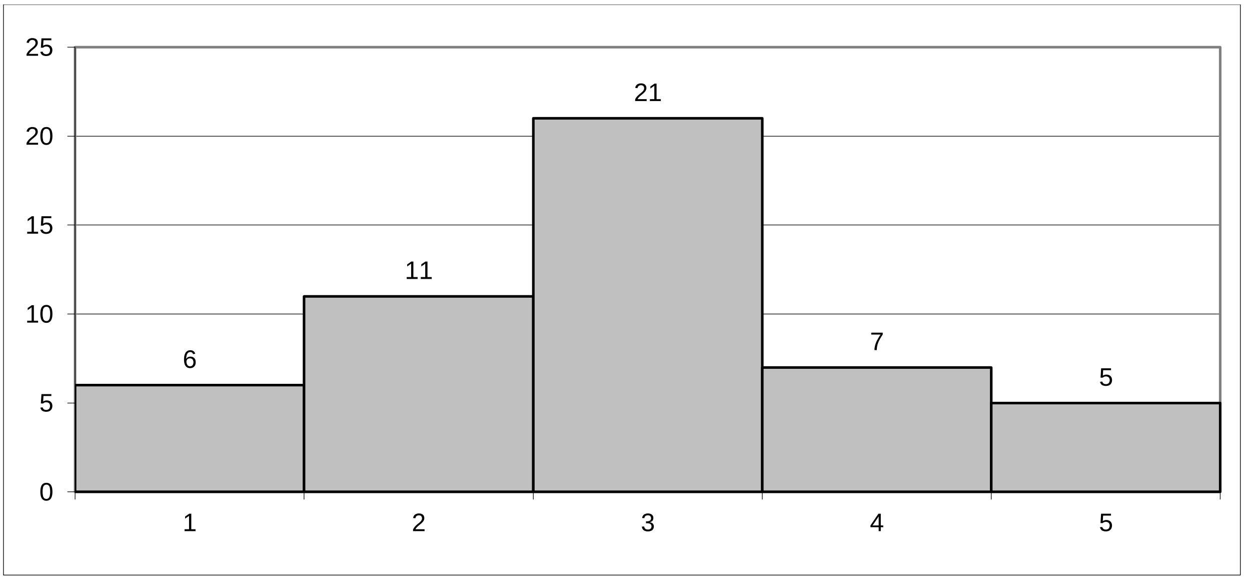

Рис. 3.1. Гістограма за даними таблиці 2.2

Щільність розподілу випадкової величини, розподіленої за нормальним законом має вид $f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$, де a і σ – параметри розподілу. Знайдемо означені параметри, враховуючи, що $\bar{x}=a;\quad S^2=\sigma^2$. Розрахунки оформимо у вигляді таблиці (табл. 3.3).

Таблиця 3.3

$[a_i;a_{i+1})$	[–2;–1,2)	[–1,2;–0,4)	[–0,4;0,4)	[0,4;1,2)	[1,2;2)
n_i	6	11	21	7	5
x_i	-1,6	0,8	0	0,8	1,6
$x_i\, n_i$	-9,6	8,8	0	5,6	8
$\left(x_i-\bar{x}\right)^2 n_i$	13,572	5,452	0,194	5,620	14,382

Знайдемо вибіркове середнє, вибіркову дисперсію і вибіркове середнє квадратичне відхилення за формулами (2.6), (2.12) та (2.15) відповідно:

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{k}x_i n_i=\frac{1}{50}(-1,6\cdot 6-0,8\cdot 11+0\cdot 21+0,8\cdot 7+1,6\cdot 5)=-0,096;$$

$$S^2=\frac{1}{n}\sum_{i=1}(x_i-\bar{x})^2 n_i=\frac{1}{50}(13,572+5,452+0,194+5,620+14,382)=0,7844;$$

$$S=\sqrt{S^2}\approx 0,886.$$

Отже, параметрами теоретичного закону розподілу є: $\bar{x}=a=-0,096;\quad S=\sigma=0,886$.

Для знаходження значення критерію χ^2 розрахуємо теоретичні частоти n_i'. Теоретичні частоти можна знайти за формулою $n_i'=np_i$, де p_i – ймовірності попадання випадкової величини в певний інтервал. Для нормального закону розподілу означені ймовірності знаходяться за формулою $P(a_i<X<a_{i+1})=p_i=\dfrac{1}{2}\left[\Phi\left(\dfrac{a_{i+1}-\bar{x}}{S}\right)-\Phi\left(\dfrac{a_i-\bar{x}}{S}\right)\right]$, де Ф – функція Лапласа, значення якої надані у статистичних таблицях. Для зручності обчислень побудуємо таблицю (табл. 2.4).

За формулою (3.1) маємо: $\chi^2=\sum_{i=1}^{k}\dfrac{\left(n_i-n_i'\right)^2}{n_i'}\approx 4,16$. Знайдемо критичне значення $\chi^2_{\alpha,l}$, враховуючи, що $l=k-r-1$ =5-2-1=2. Рівень значущості α оберемо рівним 0,1. За допомогою Excel знаходимо ХИ2ОБР(0,1; 2)=4,6.

Отже, оскільки $\chi^2<\chi^2_{\alpha,l}$, гіпотеза H_0 про нормальний розподіл приймається, гіпотеза H_1 відкидається.

Таблиця 3.4

$[a_i; a_{i+1})$	$[-2;-1,2)$	$[-1,2;-0,4)$	$[-0,4;0,4)$	$[0,4;1,2)$	$[1,2;2)$
n_i	6	11	21	7	5
x_i	-1,6	0,8	0	0,8	1,6
$x_i n_i$	-9,6	8,8	0	5,6	8
$\left(x_i - \overline{x}\right)^2 n_i$	13,572	5,452	0,194	5,620	14,382
$\frac{a_i - \overline{x}}{S}$	-2,1498	-1,2465	-0,3433	0,56	1,4633
$\Phi(\frac{a_i - \overline{x}}{S})$	-0,958	-0,785	-0,266	0,425	0,856
p_i	0,0856	0,2595	0,3455	0,2155	0,063
$n_i'=np_i$	4,325	12,975	17,275	10,775	3,15
$\frac{\left(n_i - n_i'\right)^2}{n_i'}$	0,649	0,301	0,803	1,323	1,087

3.3. Перевірка гіпотез про генеральні середні і дисперсії

В прикладних задачах часто виникає необхідність перевірки рівності середніх значень та дисперсій за даними двох або більше вибірок. Наприклад, коли визначається перевага одній з технологій виготовлення певної продукції, або наявність підвищення продуктивності праці після внесення змін в процес виробництва, або при перевірці якості продукції. Здійснення означеної перевірки виконується за критеріями, що обираються залежно від виду розподілу вибіркових даних і мети дослідження. Для деяких критеріїв перевірки рівності середніх значень висувається додаткова вимога про рівність генеральних дисперсій.

3.3.1. Перевірка гіпотези про рівність генеральних дисперсій. F-критерій (Фішера)

Перевірка гіпотези про рівність генеральних дисперсій здійснюється за F-критерієм (Фішера) тільки тоді, коли статистичні дані незалежні і розподілені за нормальним законом. Формулюються гіпотези:

H_0 – дисперсії двох нормально розподілених генеральних сукупностей рівні, тобто $S_1^2 = S_2^2$;

H_1 - дисперсії двох нормально розподілених генеральних сукупностей не рівні, тобто $S_1^2 \neq S_2^2$.

F-критерій (Фішера) розраховується за формулою:

$$F = \frac{S_1^2}{S_2^2}, \quad S_1^2 > S_2^2, \qquad (3.2)$$

Гіпотеза H_0 приймається, якщо розраховане значення F менше критичного значення розподілу Фішера $F_{крит}$, взятого із рівнем значущості α і ступенями волі l_1 та l_2 для чисельнику і знаменнику відповідно: $l_1 = n_1 - 1$, $l_2 = n_2 - 1$, де n_1, n_2 – об'єми вибірок. $F_{крит}$ можна знайти за допомогою вбудованої статистичної функції Excel FРАСПОБР (α; l_1; l_2).

Зауваження. Дисперсія у чисельнику дроби у формулі (2.2) повинна бути більше дисперсії у знаменнику, тобто значення F-критерію повинно бути більше одиниці.

ПРИКЛАД 3.1. Відомі дані про продуктивність праці (одиниць продукції за зміну) двох груп працівників: група 1 складається з працівників, що пройшли спеціальний навчальний курс; група 2 – із працівників, що не пройшли курсу (табл. 2.11). Враховуючи, що дані розподілені за нормальним законом, перевірити гіпотезу про рівність дисперсій.

Таблиця 3.5

	Група 1					Група 2				
Продуктивність праці	34	85	96	102	103	63	69	83	89	106
Кількість працівників	5	2	11	8	4	2	6	8	3	1

Розв'язання. Дані таблиці 3.5 є двома вибірками. Перша – вибірка значень величини X_1 – продуктивності праці робітників, що пройшли навчання, друга – вибірка величини X_2 – продуктивності праці робітників, що не пройшли навчання.

Сформулюємо гіпотези: H_0 – дисперсії генеральних сукупностей, з яких зроблено вибірки, рівні, $S_1^2 = S_2^2$; H_1 - дисперсії не рівні, $S_1^2 \neq S_2^2$. Перевіримо справедливість гіпотези H_0 за F-критерієм (Фішера).

Знайдемо за вибірковими даними оцінку дисперсії X_1 за формулою (2.13). Розрахунки оформимо у вигляді таблиці (табл. 2.12).

Таблиця 3.6

x_i	34	85	96	102	103	Суми
n_i	5	2	11	8	4	30
$x_i\, n_i$	170	170	1056	816	412	2624
$(x_i - \overline{x})^2 n_i$	14293,42	12,17	801,00	1689,74	965,14	17761,47

Отже, $\overline{x_1} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{1}{30} \cdot 2624 \approx 87,47$;

$$S_1^{\,2} = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{x})^2 n_i = \frac{1}{29}\cdot 17761,47 \approx 612,46 .$$

Аналогічно знайдемо оцінку дисперсії X_2. Розрахунки оформимо у вигляді таблиці (табл.. 2.6).

Таблиця 2.6

x_i	63	69	83	89	106	Суми
n_i	2	6	8	3	1	20
$x_i\ n_i$	126	414	664	267	106	1577
$(x_i - \overline{x})^2 n_i$	502,45	582,13	137,78	309,07	737,12	2268,55

Отже, $\overline{x_2} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{1}{20}\cdot 1577 \approx 78,85$;

$$S_2^{\,2} = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{x})^2 n_i = \frac{1}{19}\cdot 2268,55 \approx 119,40 .$$

Знайдемо значення F-критерію за формулою (3.2). Оскільки $S_1^2 > S_2^2$, то $F = \frac{S_1^2}{S_2^2} = \frac{612,46}{119,40} \approx 5,13$. Знайдемо $F_{крит}$, враховуючи, що $l_1 = n_1 - 1 = 30 - 1 = 29$; $l_2 = n_2 - 1 = 20 - 1 = 19$. Рівень значущості оберемо $\alpha = 0,05$. Тоді $F_{крит} =$ FРАСПОБР (0,05; 29; 19)=2,077.

Оскільки $F > F_{крит}$, то гіпотезу H_0 відкидаємо і приймаємо гіпотезу H_1 – дисперсії не рівні, тобто вибірки здобути з різних генеральних сукупностей.

Висновок: навчальний курс суттєво впливає на продуктивність праці робітників.

3.3.2. Перевірка гіпотези про рівність генеральних середніх. Критерій Стьюдента

Критерій Стьюдента використовується для перевірки гіпотез про рівність генеральних середніх, якщо статистичні дані розподілені за нормальним законом. Формулюються гіпотези:

H_0 – середні двох генеральних сукупностей рівні, тобто $\overline{x_1} = \overline{x_2}$;

H_1 - середні двох генеральних сукупностей не рівні, тобто $\overline{x_1} \neq \overline{x_2}$.

Перевірка виконується за даними двох вибірок об'ємом n_1 та n_2. При цьому можливі такі випадки.

Випадок 1. Генеральні дисперсії рівні ($S_1^2 = S_2^2$). Тоді t-критерій Стьюдента обчислюється за формулою:

$$t=\frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}. \qquad (3.3)$$

Розраховане значення t-критерію порівнюється з критичним значенням $t_{\text{крит}}$, де $t_{\text{крит}}$ – критичне значення розподілу Стьюдента з параметрами $\frac{\alpha}{2}$ і ступенем волі $l=n_1+n_2-2$, яке надається в статистичних таблицях або знаходиться за допомогою вбудованої функції Excel СТЬЮДРАСПОБР($\frac{\alpha}{2}$;l).

Випадок 2. Генеральні дисперсії не рівні ($S_1^2 \neq S_2^2$). Тоді t-критерій Стьюдента обчислюється за формулою:

$$t=\frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}. \qquad (3.4)$$

Розраховане значення t-критерію також порівнюється з критичним значенням $t_{\text{крит}}$, але ступені волі розраховуються за формулою:

$$l=\frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1+1}+\frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2+1}}+2. \qquad (3.5)$$

Випадок 3. Вибірки не є незалежними, оскільки на них впливає певний фактор і його вплив не відомий, або вибірки є даними, отриманими до і після проведення певного експерименту. Тоді формується парна вибірка, і для кожної пари елементів знаходиться d – різність їх значень. Подальша перевірка здійснюється над вибіркою різностей. t-критерій Стьюдента обчислюється за формулою:

$$t=\frac{\overline{x_d}}{S_d/\sqrt{n-1}}. \qquad (3.6)$$

де $\overline{x_d}$ – вибіркове середнє для вибірки різностей;

S_d – вибіркове середнє квадратичне відхилення для вибірки різностей;

n – об’єм вибірки різностей.

Розраховане значення t-критерію також порівнюється з критичним значенням розподілу Стьюдента з параметрами $\frac{\alpha}{2}$ і ступенями волі $l=n-1$.

У всіх випадках гіпотеза H_0 приймається, якщо розраховане значення t-критерію менше за абсолютною величиною критичного значення $t_{крит}$:
$|t| < t_{крит}$.

ПРИКЛАД 3.2. Для виробництва кожної з 10 деталей за першою технологією було витрачено, у середньому, 30с. Дисперсія часу складала 1с^2. Для виробництва кожної з 16 деталей за другою технологією було витрачено, у середньому, 28с із дисперсією часу 2с^2. Чи можна вважати, що у середньому для виробництва деталей за першою технологією потрібно більше часу?

Розв'язання. За умов задачі було зроблено дві вибірки: перша – вибірка об'єму n_1=10 значень величини X_1 – часу, потрібному для виготовлення деталей за першою технологією; друга – вибірка об'єму n_2=16 значень величини X_2 – часу, потрібному для виготовлення деталей за другою технологією. Відомі вибіркові середні $\overline{x_1}$ =30с та $\overline{x_2}$ =28с – середній час, необхідний для виготовлення деталей за першою і другою технологіями відповідно. Відомі дисперсії часу для вибірок: S_1^2 =1с^2 та S_1^2 =2с^2. За питанням задачі потрібно перевірити гіпотезу про рівність генеральних середніх.

Сформулюємо гіпотези:

H_0 – середні двох генеральних сукупностей рівні, тобто $\overline{x_1} = \overline{x_2}$;

H_1 - середні двох генеральних сукупностей не рівні, тобто $\overline{x_1} \neq \overline{x_2}$.

Перед вибором критерію для перевірки потрібно встановити, чи рівні генеральні дисперсії. Скористуємось критерієм Фішера. За формулою (3.2) обчислимо значення F-критерію:

оскільки $S_2^2 > S_1^2$, то $F = \dfrac{S_2^2}{S_1^2} = \dfrac{2}{1} = 2$.

Знайдемо критичне значення розподілу Фішера $F_{крит}$: оберемо рівень значущості α =0,05; врахуємо, що ступені волі l_1=n_1–1=9 та l_2=n_2–1=15. Тоді FРАСПОБР (α ; l_2; l_1)=FРАСПОБР (0,05; 15; 9)=3,006. Отже, F< $F_{крит}$, тому генеральні дисперсії можна вважати рівними.

Оскільки генеральні дисперсії рівні (випадок 1), то t-критерій Стьюдента розраховуємо за формулою (3.6):

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} = \frac{30 - 28}{\sqrt{\dfrac{10 \cdot 1 + 16 \cdot 2}{10 + 16 - 2}\left(\dfrac{1}{10} + \dfrac{1}{16}\right)}} \approx 7{,}032 .$$

Знайдемо критичне значення розподілу Стьюдента $t_{крит}$, враховуючи, що $l = n_1 + n_2 - 2 = 10 + 16 - 2 = 24$. Оберемо значення α =0,05. Тоді $t_{крит}$=СТЬЮДРАСПОБР($\frac{\alpha}{2}$;l)=СТЬЮДРАСПОБР(0,025;24)=2,39.

Отже, $|t| > t_{крит}$, тому гіпотеза H_0 про рівність генеральних середніх відкидається на рівні значущості 0,05 і приймається гіпотеза H_1.

Висновок: для вироблення деталей за першою технологією потрібно, у середньому, більше часу.

РОЗДІЛ 4 ОСНОВИ КОРЕЛЯЦІЙНОГО АНАЛІЗУ

Будь-який соціо-економічний об'єкт або явище зазвичай характеризується за декількома ознаками, тобто різними властивостями. Ці ознаки взаємозв'язані і впливають одна на одну. Крім того, може існувати зв'язок між ознаками різних об'єктів і явищ. Тому в математичній статистиці розроблений апарат для виявлення таких зв'язків і оцінки їх сили (тісноти). Цей математичний апарат називається кореляційним аналізом.

4.1. Поняття кореляційного зв’язку між досліджуваними величинами

В багатьох прикладних задачах необхідно виявити залежність між двома властивостями (ознаками) X і Y одного і того ж економічного об'єкту, або між певними ознаками різних об’єктів. Якщо вказані ознаки допускають кількісне вимірювання, і, з погляду економічної теорії, виходячи з економічної характеристики об'єкту, ознака Y залежить від ознаки X, тоді X можна назвати незалежною змінною, або **факторною ознакою**, або просто фактором, а Y – залежною змінною або **результативною ознакою**.

Якщо кожному значенню факторної ознаки X відповідає одне і тільки одне значення результативної ознаки Y, то говорять, що між цими ознаками існує **функціональний зв’язок**: $Y=f(X)$.

Якщо кожному значенню факторної ознаки X відповідає безліч значень результативної ознаки Y, то говорять, що між цими ознаками існує **статистичний зв’язок.**

Наприклад, якщо X приймає l значень $X = \{x_1, x_2, ..., x_l\}$ і кожному її значенню x_i відповідає множина значень Y, тобто:

значенню x_1 відповідає множина $\{y_{11}, y_{12}, ..., y_{1m_1}\}$; значенню x_2 відповідає множина $\{y_{21}, y_{22}, ..., y_{2m_2}\}$; значенню x_l відповідає множина $\{y_{l1}, y_{l2}, ..., y_{lm_l}\}$;

то між X та Y існує статистичний зв’язок.

Вивчення статистичного зв'язку дуже складний і трудомісткий процес, у якому потрібно аналізувати багатомірні таблиці даних. Тому зазвичай вивчається не статистичний, а кореляційний зв'язок між X та Y.

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

Наприклад, якщо X приймає l значень $X=\{x_1, x_2, ..., x_l\}$ і кожному її значенню x_i відповідає середнє множини значень Y, тобто:

значенню x_1 відповідає $\overline{y}_{x_1}=\dfrac{y_{11}+y_{12}+...+y_{1m_1}}{m_1}$;

значенню x_2 відповідає $\overline{y}_{x_2}=\dfrac{y_{21}+y_{22}+...+y_{1m_2}}{m_2}$;

…

значенню x_l відповідає $\overline{y}_{x_l}=\dfrac{y_{l1}+y_{l2}+...+y_{lm_l}}{m_l}$;

то між X та Y існує кореляційний зв’язок.

Наприклад відомо, що з однакових за площею ділянок землі при рівних кількостях внесеного добриву отримують різний урожай. Тому, якщо Y – урожайність зерна, а X – кількість внесеного добриву, то функціонального зв’язку між X та Y немає. Це пояснюється впливом таких випадкових факторів, як температура повітря, кількість опадів і т. ін. Однак досвід показує, що середній урожай є функцією від кількості добриву, тобто між X та Y існує кореляційний зв’язок.

Основними задачами кореляційного аналізу є:

- вивчення сили зв’язку між двома і більше ознаками досліджуваного об’єкту;
- встановлення факторів, що найбільш суттєво впливають на результативну ознаку;
- виявлення невідомих причинно-наслідкових зв’язків між ознаками об’єкту.

4.2. Групування даних для кореляційного аналізу

Вибіркові дані для вивчення кореляційного зв’язку між ознаками X та Y зазвичай мають вигляд пар їх значень: $(x_1;y_1)$, $(x_2;y_2)$, …, $(x_n;y_n)$, x_i – значення величини X, y_i – значення Y, n – кількість пар значень, $i=\overline{1,n}$.

Якщо кількість пар значень достатньо велика (принаймні n>20), то для зручності розрахунків дані групуються.

Для групування даних необхідно:

1) Розбити множини значень X та Y на інтервали, використовуючи формулу Стерджеса (форм. 1.2), кількість інтервалів для X та Y може бути різною (позначення: k – кількість інтервалів для X; m – кількість інтервалів для Y).

2) Зобразити дані графічно: побудувати на площині точки з координатами $\left(x_i; y_j\right)$. В результаті отримується площина, розбита на прямокутники, в кожному з яких може бути множина точок (рис. 3.1). Вказане графічне зображення вибіркових даних називається **полем кореляції**.

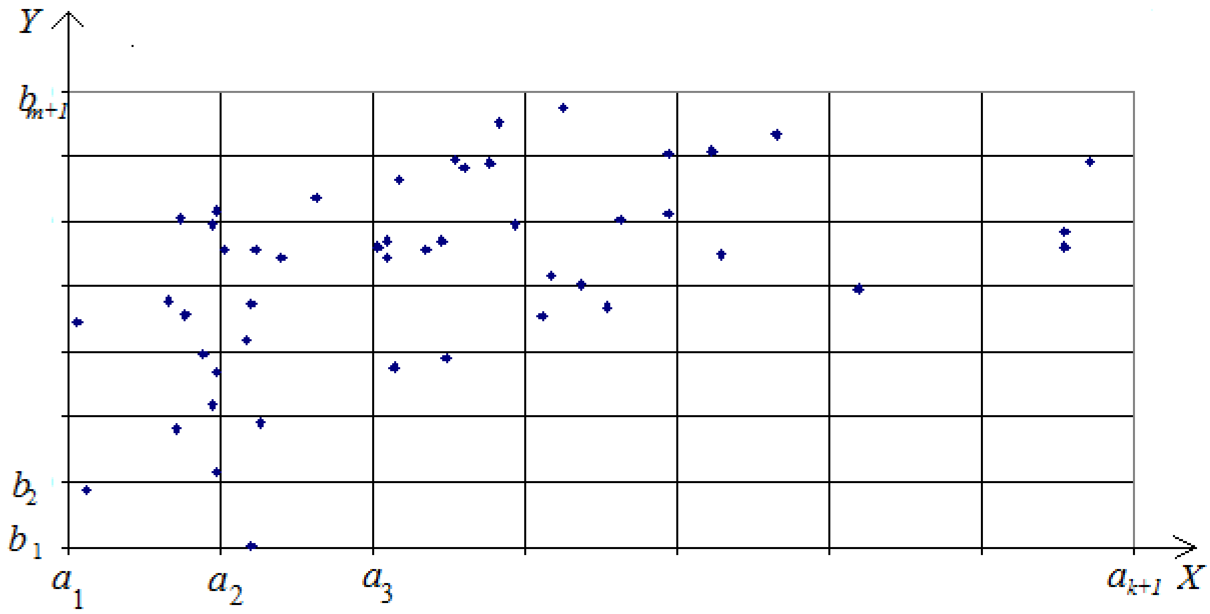

Рис. 3.1. Поле кореляції

3) Побудувати кореляційну таблицю (табл. 3.1). В першому рядку, розбитому на дві частини, записуються інтервали $\left[a_i; a_{i+1}\right)$ для X та їх середини x_i. У першому стовпці, розбитому на дві частини, записуються інтервали $\left[b_j; b_{j+1}\right)$ для Y та їх середини y_j. В центральній частині таблиці записуються частоти n_{ij} – кількість точок, що потратили в прямокутник, обмежений по X інтервалом $\left[a_i; a_{i+1}\right)$ і по Y інтервалом $\left[b_j; b_{j+1}\right)$. В останньому рядку таблиці записуються частоти n_i для X – кількість точок, що потратили в прямокутники, які відповідають інтервалу $\left[a_i; a_{i+1}\right)$, тобто $n_i = \sum_{j=1}^{m} n_{ij}$ – сума частот n_{ij} в стовпці з номером i. В останньому стовпці таблиці записуються частоти n_j для Y –

кількість точок, що потратили в прямокутники, які відповідають інтервалу $\left[b_j;b_{j+1}\right)$, тобто $n_j=\sum_{i=1}^{k} n_{ij}$ – сума частот n_{ij} в рядку з номером *j*.

Кореляційну таблицю можна розглядати як своєрідний подвійний статистичний ряд.

Таблиця 4.1

<table>
<tr><td colspan="2" rowspan="2">X (інтервали і їх середини)
Y (інтервали і їх середини)</td><td>$\left[a_1;a_2\right)$</td><td>$\left[a_2;a_3\right)$</td><td>…</td><td>$\left[a_k;a_{k+1}\right)$</td><td rowspan="2">$n_j=\sum_{i=1}^{k} n_{ij}$</td></tr>
<tr><td>x_1</td><td>x_2</td><td>…</td><td>x_k</td></tr>
<tr><td>$\left[b_1;b_2\right)$</td><td>y_1</td><td>n_{11}</td><td>n_{21}</td><td>…</td><td>n_{k1}</td><td>n_1</td></tr>
<tr><td>$\left[b_2;b_3\right)$</td><td>y_2</td><td>n_{12}</td><td>n_{22}</td><td>…</td><td>n_{k2}</td><td>n_2</td></tr>
<tr><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td></tr>
<tr><td>$\left[b_m;b_{m+1}\right)$</td><td>y_m</td><td>n_{1m}</td><td>n_{2m}</td><td>…</td><td>n_{km}</td><td>n_m</td></tr>
<tr><td colspan="2">$n_i=\sum_{j=1}^{m} n_{ij}$</td><td>$n_1$</td><td>$n_2$</td><td>…</td><td>$n_k$</td><td></td></tr>
</table>

4) За даними кореляційної таблиці будується ряд, що відображає залежність середнього значення *Y* від *X* (табл. 3.2). В першому рядку таблиці записуються середини інтервалів x_i. В другому – відповідні середні значення $\overline{y}_{x_i}$, що знаходяться за формулами:

$$\overline{y}_{x_1}=\frac{y_1n_{11}+y_2n_{12}+\ldots+y_mn_{1m}}{n_1}\ ;\ \overline{y}_{x_2}=\frac{y_1n_{21}+y_2n_{22}+\ldots+y_mn_{2m}}{n_2}\ ;\ \ldots\ ;\ \overline{y}_{x_k}=\frac{y_1n_{k1}+y_2n_{k2}+\ldots+y_mn_{km}}{n_k}.$$

Таблиця 4.2

x_i	x_1	x_2	…	x_k
$\overline{y}_{x_i}$	$\overline{y}_{x_1}$	$\overline{y}_{x_2}$	…	$\overline{y}_{x_k}$
n_i	n_1	n_2	…	n_k

В результаті отримується статистичний ряд, що містить значення *X*, відповідні середні значення *Y* та частоти. За даними такого ряду проводиться кореляційний аналіз.

4.3. Коефіцієнт кореляції Пірсона

Для оцінки тісноти (або сили) зв’язку між *X* та *Y* слугує коефіцієнт кореляції. У випадку, коли між *X* та *Y* існує <u>лінійний зв’язок та вибіркові дані розподілені за нормальним законом</u>, використовується **коефіцієнт кореляції Пірсона**, який зветься ще параметричним коефіцієнтом кореляції.

Коефіцієнт кореляції Пірсона розраховується за формулою:

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{S_x \cdot S_y}, \tag{4.1}$$

де $\overline{x}$ – вибіркове середнє величини X;

$\overline{y}$ – вибіркове середнє величини Y;

$\overline{xy}$ – вибіркове середнє величини XY;

S_x – вибіркове середнє квадратичне відхилення величини X;

S_y – вибіркове середнє квадратичне відхилення величини Y.

Враховуючи формули для знаходження вибіркових середніх і середніх квадратичних відхилень, а саме:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i; \quad \overline{y} = \frac{1}{n}\sum_{j=1}^{m} y_j n_j; \quad \overline{xy} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{m} x_i y_j n_{ij};$$

$$S_x = \sqrt{\frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - \left(\frac{1}{n}\sum_{i=1}^{k} x_i n_i\right)^2}; \quad S_y = \sqrt{\frac{1}{n}\sum_{j=1}^{m} y_j^2 n_j - \left(\frac{1}{n}\sum_{j=1}^{m} y_j n_j\right)^2};$$

отримують більш зручну для розрахунків формулу:

$$r = \frac{n\sum_{i=1}^{k}\sum_{j=1}^{m} x_i y_j n_{ij} - \left(\sum_{i=1}^{k} x_i n_i\right)\left(\sum_{j=1}^{m} y_j n_j\right)}{\sqrt{n\sum_{i=1}^{k} x_i^2 n_i - \left(\sum_{i=1}^{k} x_i n_i\right)^2}\sqrt{n\sum_{j=1}^{m} y_j^2 n_j - \left(\sum_{j=1}^{m} y_j n_j\right)^2}}. \tag{4.2}$$

У випадку незгрупованих даних розрахункова формула суттєво спрощується:

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}. \tag{4.3}$$

Властивості коефіцієнта кореляції Пірсона

1) Коефіцієнт кореляції Пірсона приймає значення на проміжку $[-1;1]$, тобто $-1 \le r \le 1$.

2) Якщо $|r| \le 0{,}5$, то зв’язок вважається слабким; якщо $|r| \le 0{,}7$, то зв’язок вважається середнім; $|r| > 0{,}7$, то зв’язок вважається сильним.

3) Якщо $r > 0$, то зв’язок називається додатнім, тобто зі збільшенням значень X значення Y також збільшуються. Якщо $r < 0$, то зв’язок називається від’ємним, тобто зі збільшенням значень X значення Y зменшуються.

Зауваження. Слід пам’ятати, що коефіцієнт кореляції Пірсона показує силу лінійного зв’язку. Якщо між X та Y існує сильний нелінійний зв’язок, коефіцієнт кореляції Пірсона може дорівнювати нулю.

Оскільки сила зв’язку між X та Y оцінюється за вибірковими даними, то необхідна перевірка її **статистичної значущості**, тобто оцінка можливості розповсюдити отримані результати на всю генеральну сукупність.

Перевірка статистичної значущості коефіцієнта кореляції Пірсона здійснюється за допомогою так званої t-статистики, яка розраховується за формулою:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}. \tag{4.4}$$

Розраховане значення t-статистики порівнюється з критичним значенням $t_{крит}$. $t_{крит}$ – табличне значення розподілу Стьюдента, яке також можна знайти за допомогою вбудованої статистичної функції Excel СТЬЮДРАСПОБР (α; l), де α – обраний дослідником рівень значущості, l – ступені волі, l=n-2.

Якщо розраховане значення t-статистики більше критичного $|t| > t_{крит}$, то коефіцієнт кореляції вважається значимим на обраному рівні α.

ПРИКЛАД 4.1. За наявними даними про рівень механізації праці X (%) і продуктивності праці Y (од. продукції/год.) для 14 однотипних підприємств (табл. 4.3) оцінити тісноту зв'язку між X і Y. Визначити можливість розповсюдження результатів розрахунків на всі підприємства такого типу.

Таблиця 4.3

X	32	30	36	40	41	47	56	54	60	55	61	67	69	76
Y	20	24	28	30	31	33	34	37	38	40	41	43	45	48

Розв’язання. Дані таблиці 4.3 є вибіркою значень X і відповідних значень Y. Оскільки кількість даних невелика (n=14), то їх можна не групувати. Для оцінки тісноти зв'язку між X і Y розрахуємо коефіцієнт кореляції Пірсона за формулою (4.3.) для незгрупованих даних. Розрахунки для зручності оформимо у вигляді таблиці (табл. 4.4).

Таблиця 4.4

x_i	y_i	x_i^2	y_i^2	$x_i\ y_i$
32	20	1024	400	640
30	24	900	576	720

36	28	1296	784	1008
40	30	1600	900	1200
41	31	1681	961	1271
47	33	2209	1089	1551
56	34	3136	1156	1904
54	37	2916	1369	1998
60	38	3600	1444	2280
55	40	3025	1600	2200
61	41	3721	1681	2501
67	43	4489	1849	2881
69	45	4791	2025	3105
76	48	5779	2304	3848
Суми				
724	492	40134	18138	26907

Отже,

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{n} y_j\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{j=1}^{n} y_j^2 - \left(\sum_{j=1}^{n} y_j\right)^2}} = \frac{14 \cdot 26907 - 724 \cdot 492}{\sqrt{14 \cdot 40134 - 724^2}\sqrt{14 \cdot 18138 - 492^2}}$$

$$= \frac{20490}{\sqrt{37700}\sqrt{11868}} \approx 0,969.$$

За значенням коефіцієнта кореляції можна зробити висновок, що між X і Y існує сильний додатній зв’язок.

Перевіримо статистичну значущість знайденого коефіцієнта кореляції Пірсона. Розрахуємо t-статистику за формулою (4.4): $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,969\sqrt{14-2}}{\sqrt{1-0,969^2}} \approx 13,59$. Знайдемо $t_{крит}$, враховуючи, що $l=n-2=14-2=12$. Оберемо рівень значущості α=0,01. Тоді $t_{крит}$=СТЬЮДРАСПОБР (0,01; 12)=3,055.

Оскільки розраховане значення t-статистики більше критичного $13,59 > 3,055$, то коефіцієнт кореляції можна вважати значимим на обраному рівні α=0,01.

Висновок. Між рівнем механізації праці та її продуктивністю на підприємствах, що досліджувалися, існує сильний додатній зв’язок: чим більше рівень механізації праці, тим вище її продуктивність. Висновок дійсний для всіх підприємств такого типу.

РОЗДІЛ 5. ПОБУДОВА РЕГРЕСІЙНИХ МОДЕЛЕЙ

При вивченні стохастичних зв'язків між різними ознаками економічного об'єкта головною задачею є встановлення виду кореляційної залежності результативної ознаки (Y) від факторної (X), тобто виду функціональної залежності $\overline{Y}=f(X)$. В першу чергу це пов'язано з необхідністю прогнозування досліджуваних процесів. Математико-статистичний апарат, що дозволяє встановити вид кореляційної залежності називається **регресійним аналізом**, а функція, що описує цю залежність, називається **рівнянням регресії**.

5.1. Встановлення виду кореляційної залежності

Регресійний аналіз проводиться за такими етапами:

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

2) Побудова регресійної моделі.

3) Перевірка статистичної значущості побудованої моделі.

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

Вибіркові дані для вивчення кореляційного зв'язку між ознаками X та Y зазвичай мають вигляд пар їх значень: $(x_1;y_1)$, $(x_2;y_2)$, …, $(x_n;y_n)$, x_i – значення величини X, y_i – значення Y, n – кількість пар значень, $i=\overline{1,n}$. Якщо їх кількість достатньо велика, то для зручності розрахунків дані групуються (див. п. 3.2) і будується статистичний ряд, що містить значення X, відповідні середні значення Y та частоти (табл. 4.1).

Таблиця 4.1

x_i	x_1	x_2	…	x_k
$\overline{y}_{x_i}$	$\overline{y}_{x_1}$	$\overline{y}_{x_2}$	…	$\overline{y}_{x_k}$
n_i	n_1	n_2	…	n_k

Згруповані дані (табл. 4.1) зображуються графічно, що часто дозволяє визначити вид залежності Y від X.

Ламана лінія, що сполучає крапки з координатами $(x_i;\ \overline{y_{x_i}})$, називається **емпіричною лінією регресії**.

Якщо емпірична лінія регресії значно наближається до прямої лінії, то висувається гіпотеза про наявність лінійного зв'язку між досліджуваними ознаками (рис. 4.1).

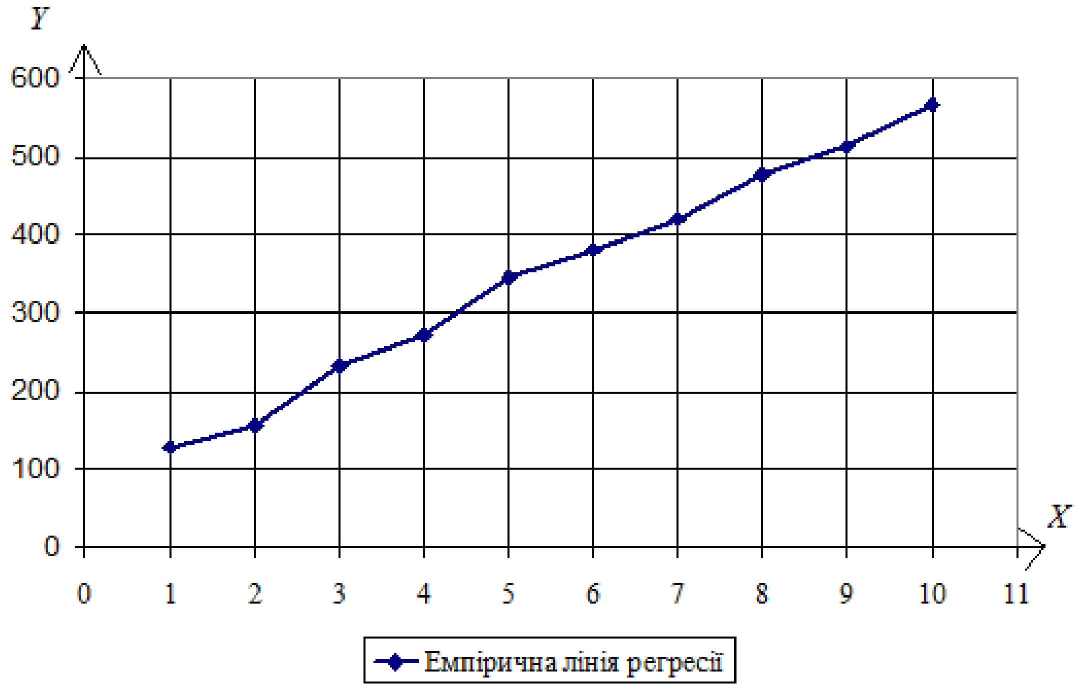

Рис. 4.1. Гіпотетична лінійна залежність

В іншому випадку висувається гіпотеза про наявність нелінійного зв'язку (рис.4.2).

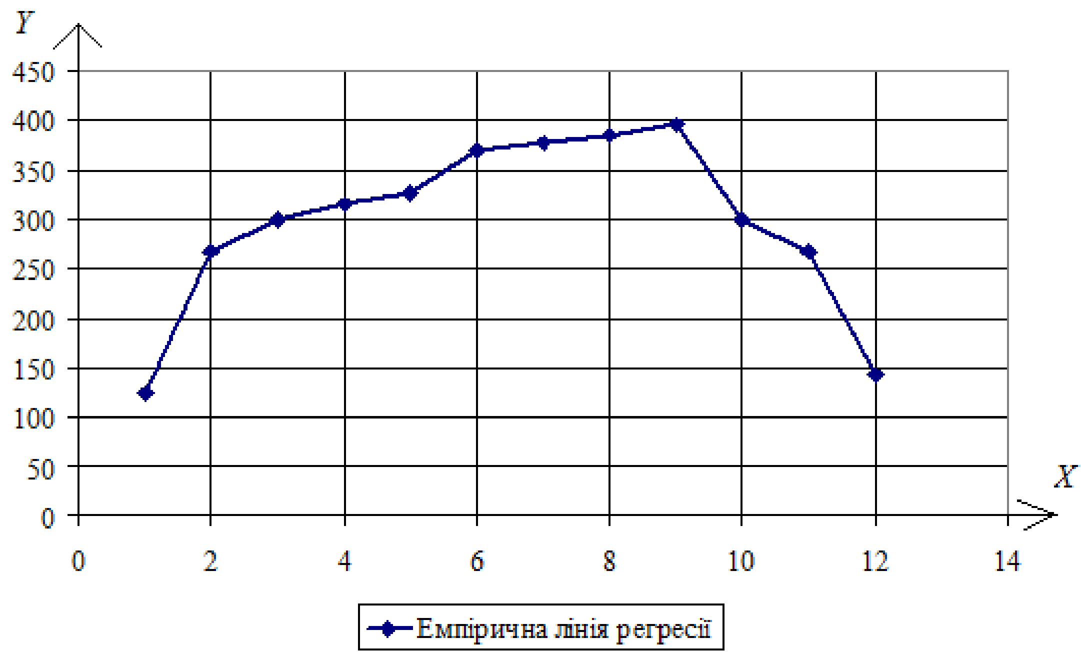

Рис. 4.2. Гіпотетична нелінійна залежність

5.2. Лінійна регресія

Якщо висунуто гіпотезу про наявність лінійної залежності результативної ознаки (*Y*) від факторної (*X*), то рівняння регресії має вид:

$$\overline{y_x} = ax + b, \tag{5.1}$$

де a, b - параметри моделі.

Побудова лінійної регресійної моделі – це знаходження параметрів рівняння (4.1). Параметри рівняння регресії зазвичай знаходяться за **методом найменших квадратів**.

Ідея методу найменших квадратів

Нехай при вивченні залежності *Y* від *X* було отримано вибіркові дані: $x_1, x_2, ..., x_n$ – значення величини *X*, $y_1, y_2, ..., y_n$ – відповідні значення *Y*. За вибірковими даними було побудовано рівняння регресії $y = ax + b$. Якщо в рівняння підставити замість *x* значення $x_1, x_2, ..., x_n$, то будуть отримані теоретичні значення *Y*: $y_{1,теор}, y_{2,теор}, ..., y_{n,теор}$, які відрізняються від $y_1, y_2, ..., y_n$. Різниця значень $y_{i,теор} - y_i$ називається помилкою регресійної моделі і позначається e_i. Якщо параметри рівняння підбираються так, щоб сума квадратів помилок була мінімальною, то говорять, що вони отримані за методом найменших квадратів.

У випадку лінійної регресії параметри рівняння регресії за методом найменших квадратів знаходяться з системи лінійних алгебраїчних рівнянь:

$$\begin{cases} a\sum_{i=1}^{k} x_i^2 n_i + b\sum_{i=1}^{k} x_i n_i = \sum_{i=1}^{k} x_i n_i \overline{y_{x_i}} \\ a\sum_{i=1}^{k} x_i n_i + b\sum_{i=1}^{k} n_i = \sum_{i=1}^{k} n_i \overline{y_{x_i}} \end{cases} \tag{5.2}$$

Якщо вибіркові дані не згруповані, то система (4.1) значно спрощується:

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i \end{cases} \tag{5.3}$$

Перевірка правильності побудови рівняння регресії здійснюється за основним варіаційним рівнянням:

$$Q = Q_p + Q_o, \tag{5.4}$$

де $Q = \sum_{i=1}^{k} \left(\overline{y_{x_i}} - \overline{y}\right)^2 n_i$ - загальна варіація, тобто сума квадратів відхилень емпіричних значень *Y* від середнього, $\overline{y} = \frac{\sum_{i=1}^{k} \overline{y_{x_i}} n_i}{n}$;

$Q_p = \sum_{i=1}^{k} \left(y_{i,теор} - \overline{y}\right)^2 n_i$ - варіація регресії, тобто сума квадратів відхилень теоретичних значень *Y* від середнього, що обумовлена регресією;

$Q_o = \sum_{i=1}^{k}\left(y_{i,теор} - \overline{y}_{x_i}\right)^2 n_i$ - варіація залишків, тобто сума квадратів відхилень теоретичних значень *Y* від емпіричних.

У випадку незгрупованих даних загальна варіація, варіації регресії і залишків знаходяться за формулами: $Q = \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2$; $Q_p = \sum_{i=1}^{n}\left(y_{i,\text{теор}} - \overline{y}\right)^2$; $Q_o = \sum_{i=1}^{n}\left(y_{i,\text{теор}} - y_i\right)^2$; а середнє значення за формулою $\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$.

Для перевірки статистичної значущості рівняння регресії розраховується F-статистика за формулою:

$$F = \frac{Q_p\left(n-l\right)}{Q_o\left(l-1\right)}, \tag{5.5}$$

де n – кількість наглядів, l – кількість груп у кореляційній таблиці або кількість параметрів моделі у випадку незгрупованих даних. Розраховане значення F-статистики порівнюється з критичним значенням $F_{кр}$ розподілу Фішера, яке можна знайти за статистичними таблицями або за допомогою вбудованої функції Excel $FРАСПОБР\left(\alpha, k_1, k_2\right)$, де $k_1 = l-1$; $k_2 = n-l$ - ступені волі, α - рівень значущості.

Адекватність моделі вибірковим даним можна оцінити за коефіцієнтом детермінації R^2, що показує частину варіації значень результативної ознаки *Y*, що пояснюється рівнянням регресії. Коефіцієнт детермінації розраховується за формулою:

$$R^2 = 1 - \frac{Q_o}{Q} = \frac{Q_p}{Q}. \tag{5.6}$$

Значення коефіцієнта детермінації знаходяться в інтервалі $\left[0;1\right]$, тобто $0 \le R^2 \le 1$. Чим ближче R^2 до 1, тим краще отримане рівняння регресії пояснює поведінку результативної ознаки. Наприклад, якщо R^2 =0,98, то 98% варіації результативної ознаки *Y* пояснюється рівнянням регресії.

ПРИКЛАД 5.1. Побудувати регресійну модель, що описує залежність сумарних виробничих затрат *Y* (тис. грн.) від об'ємів виробництва *X* (тис. од.). Відповідні статистичні дані надано у таблиці 4.2.

Таблиця 5.2

X	41	44	52	57	59	64	68	70	73	75
Y	670	657	713	736	778	812	833	876	911	932

Розв'язання. В таблиці 5.2 надано вибіркові дані: значення $x_i, i=\overline{1,n}$ величини X та відповідні значення $y_i, i=\overline{1,n}$; кількість пар – $n=10$ невелика, тому для проведення регресійного аналізу їх можна не групувати.

Перший етап аналізу: визначимо вид залежності *Y* від *X*. Побудуємо емпіричну лінію регресії (рис. 5.3).

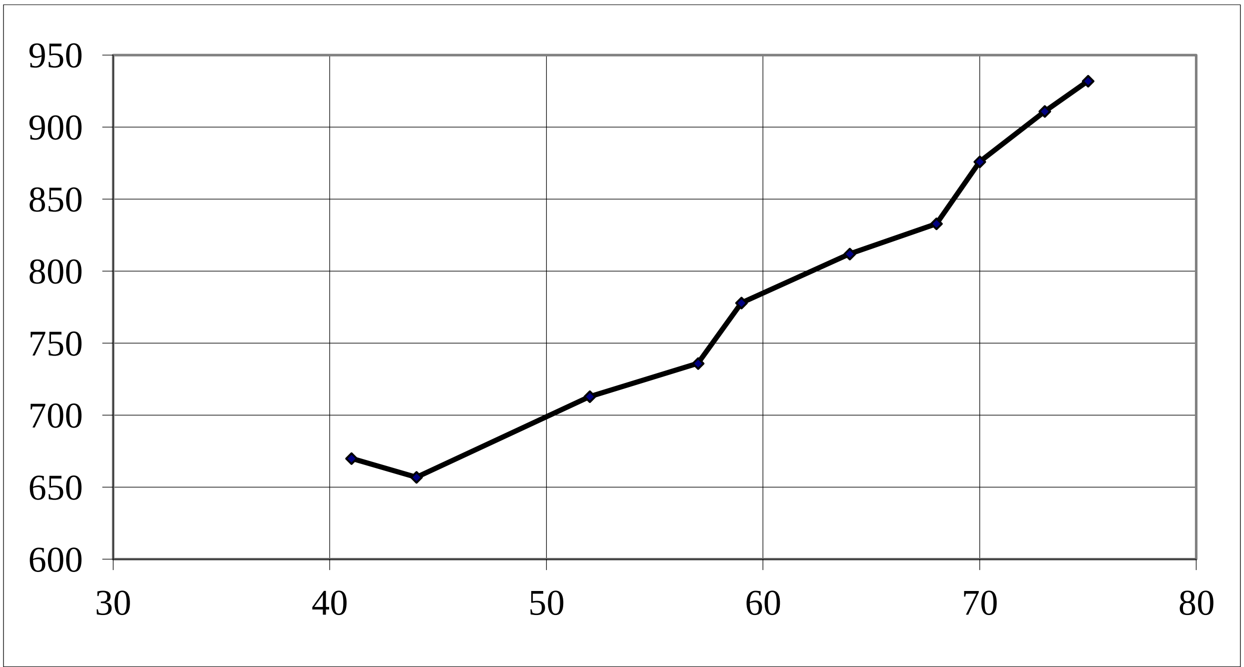

Рис. 5.3. Емпірична лінія регресії

Оскільки емпірична лінія регресії наближається до прямої лінії, то висуваємо гіпотезу про лінійну залежність *Y* від *X*, тобто рівняння регресії будемо шукати у вигляді $y = ax + b$.

Другий етап: знайдемо параметри a, b рівняння регресії, для чого складемо систему (5.3) для даних, що не згруповані. Необхідні розрахунки для зручності оформимо у вигляді таблиці (табл. 5.3).

Таблиця 5.3

Розрахункова таблиця											Суми
x_i	41	44	52	57	59	64	68	70	73	75	603
y_i	670	657	713	736	778	812	833	876	911	932	7918
x_i^2	1681	1936	2704	3249	3481	4096	4624	4900	5329	5625	37625
$x_i y_i$	27470	28908	37076	41952	45902	51968	56644	61320	66503	69900	487643

Отже, складемо систему для знаходження параметрів рівняння регресії та розв'яжемо її за правилом Крамера:

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i \end{cases} \Rightarrow \begin{cases} 37625a + 603b = 487643 \\ 603a + 10b = 7918 \end{cases}.$$

Знайдемо головний визначник системи, що складений із коефіцієнтів перед невідомими: $\Delta = \begin{vmatrix} 37625 & 603 \\ 603 & 10 \end{vmatrix} = 37625 \cdot 10 - 603^2 = 12641.$

Знайдемо допоміжні визначники, що отримуються із головного заміною відповідного стовпця коефіцієнтів на стовпець вільних членів:

$$\Delta a = \begin{vmatrix} 487643 & 603 \\ 7918 & 10 \end{vmatrix} = 487643 \cdot 10 - 603 \cdot 7918 = 101876;$$

$$\Delta b = \begin{vmatrix} 37626 & 487643 \\ 603 & 7918 \end{vmatrix} = 37626 \cdot 7918 - 48743 \cdot 603 = 3866021.$$

Знайдемо невідомі за формулами Крамера:

$$a = \frac{\Delta a}{\Delta} = \frac{101876}{12641} \approx 8,06; \quad b = \frac{\Delta b}{\Delta} = \frac{3866021}{12641} \approx 305,83.$$

Отже, шукане рівняння регресії має вигляд $y = 8,06x - 305,83$.

Третій етап: перевіримо правильність побудови моделі за рівнянням (5.4), її статистичну значущість за F-статистикою (5.5) і адекватність вибірковим даним за коефіцієнтом детермінації (5.6). Для чого знайдемо загальну варіацію, варіації регресії та залишків; необхідні розрахунки оформимо у вигляді таблиці (табл. 5.4).

Передусім знайдемо $\bar{y}$: $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \approx 791,8$.

Таблиця 5.4

x_i	y_i	$y_{i,\text{теор}}$	$\left(y_i - \bar{y}\right)^2$	$\left(y_{i,\text{теор}} - \bar{y}\right)^2$	$\left(y_{i,\text{теор}} - y_i\right)^2$
41	670	636,26	14835,24	24193,32	1138,52
44	657	660,44	18171,04	17256,64	11,80
52	713	724,91	6209,44	4474,42	141,82
57	736	765,20	3113,64	707,31	852,92
59	778	781,32	190,44	109,77	11,04
64	812	821,62	408,04	889,17	92,52
68	833	853,86	1697,44	3850,90	434,96
70	876	869,97	7089,64	6111,17	36,31
73	911	894,15	14208,64	10475,83	283,87

75	932	910,27	19656,04	14035,10	472,20
Суми			85579,6	82103,63	3475,974

Отже, $Q = 85579,6$; $Q_p = 82103,63$; $Q_o = 3475,974$; тоді основне варіаційне рівняння $Q = Q_p + Q_o$ для побудованої моделі має вигляд: $85579,6 = 82103,63 + 3475,974$ і є тотожністю, тому рівняння регресії побудовано правильно.

Для перевірки статистичної значущості рівняння регресії знайдемо F-статистику, враховуючи, що *n*=10, *l*=2 – оскільки шукали рівняння з двома параметрами:

$$F = \frac{Q_p(n-l)}{Q_o(l-1)} = \frac{82103(10-2)}{3475,974(2-1)} \approx 188,96 \cdot$$

Знайдемо $F_{кр}$: $F_{кр}$= $FРАСПОБР\left(0,001, 2-1, 10-2\right) \approx 25,41$. Розраховане значення F-статистики більше критичного, тому регресійна модель є статистично значущою на рівні 0,001.

Знайдемо коефіцієнт детермінації R^2: $R^2 = \frac{Q_p}{Q} = \frac{82103,63}{85579,6} \approx 0,96$. Значення коефіцієнта детермінації свідчить, що 96% варіації результативної ознаки *Y* пояснюються рівнянням регресії.

Висновок: Сумарні виробничі затрати *Y* (тис. грн.) лінійно залежать від об'єму виробництва *X* (тис. од.). Залежність описується рівнянням $y = 8,06x - 305,83$, яке є статистично значимим на рівні значущості 0,001 та описує 96% вибіркових даних.

5.3. Нелінійна регресія

Якщо висунуто гіпотезу про наявність нелінійної залежності результативної ознаки (*Y*) від факторної (*X*), то регресійний аналіз проводиться за тими ж етапами, як й у випадку лінійної залежності. Вид рівнянь регресії і системи для знаходження їх параметрів для нелінійних залежностей, що найчастіше зустрічаються, надано у таблиці 4.5.

Таблиця 5.5

Рівняння параболічної регресії: $\overline{y_x} = ax^2 + bx + c$	
Система для знаходження параметрів	
для згрупованих вибіркових даних:	для незгрупованих вибіркових даних:

$\begin{cases} a\sum_{i=1}^{k} x_i^4 n_i + b\sum_{i=1}^{k} x_i^3 n_i + c\sum_{i=1}^{k} x_i^2 n_i = \sum_{i=1}^{k} x_i^2 n_i \overline{y_{x_i}} \\ a\sum_{i=1}^{k} x_i^3 n_i + b\sum_{i=1}^{k} x_i^2 n_i + c\sum_{i=1}^{k} x_i n_i = \sum_{i=1}^{k} x_i n_i \overline{y_{x_i}} \\ a\sum_{i=1}^{k} x_i^2 n_i + b\sum_{i=1}^{k} x_i n_i + c\sum_{i=1}^{k} n_i = \sum_{i=1}^{k} n_i \overline{y_{x_i}} \end{cases}$	$\begin{cases} a\sum_{i=1}^{n} x_i^4 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2 y_i \\ a\sum_{i=1}^{n} x_i^3 + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i + cn = \sum_{i=1}^{n} y_i \end{cases}$
Рівняння гіперболічної регресії: $\overline{y_x} = \frac{a}{x} + b$	
Система для знаходження параметрів	
для згрупованих вибіркових даних: $\begin{cases} a\sum_{i=1}^{k} \frac{1}{x_i^2} n_i + b\sum_{i=1}^{k} \frac{1}{x_i} n_i = \sum_{i=1}^{k} \frac{1}{x_i} \overline{y_{x_i}} n_i \\ a\sum_{i=1}^{k} \frac{1}{x_i} n_i + b\sum_{i=1}^{k} n_i = \sum_{i=1}^{k} \overline{y_{x_i}} n_i \end{cases}$	для незгрупованих вибіркових даних: $\begin{cases} a\sum_{i=1}^{n} \frac{1}{x_i^2} + b\sum_{i=1}^{n} \frac{1}{x_i} = \sum_{i=1}^{n} \frac{1}{x_i} y_i \\ a\sum_{i=1}^{n} \frac{1}{x_i} + bn = \sum_{i=1}^{n} y_i \end{cases}$
Рівняння показникової регресії: $\overline{y_x} = ba^x$	
для згрупованих вибіркових даних: $\begin{cases} \lg a\sum_{i=1}^{k} x_i^2 n_i + \lg b\sum_{i=1}^{k} x_i n_i = \sum_{i=1}^{k} x_i n_i \lg \overline{y_{x_i}} \\ \lg a\sum_{i=1}^{k} x_i n_i + \lg b\sum_{i=1}^{k} n_i = \sum_{i=1}^{k} n_i \lg \overline{y_{x_i}} \end{cases}$	для незгрупованих вибіркових даних: $\begin{cases} \lg a\sum_{i=1}^{n} x_i^2 + \lg b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i \lg y_i \\ \lg a\sum_{i=1}^{n} x_i + n\lg b = \sum_{i=1}^{n} \lg y_i \end{cases}$

Перевірка статистичної значущості нелінійної регресійної моделі також здійснюється за F-статистикою. При цьому для параболічної регресії кількість параметрів l=3, для гіперболічної і показникової – l=2.

ПРИКЛАД 5.2. Дано розподіл однотипових підприємств за об'ємом виробництва X (тис. од.) і собівартістю одиниці продукції Y (грн.) (табл. 4.6). Знайти регресійну модель, що описує собівартості продукції від об'єму виробництва.

Таблиця 5.6

Y / X	10	15	20	25
25	-	-	1	2
50	-	2	2	-
75	-	5	3	1
100	1	3	-	-

125	3	1	1	-

Розв'язання. Для проведення регресійного аналізу за даними таблиці 4.6 побудуємо кореляційну таблицю (табл. 5.7).

Таблиця 5.7

x_i / y_j	25	50	75	100	125	n_j
10	0	0	0	1	3	4
15	0	2	5	3	1	11
20	1	2	3	0	1	7
25	2	0	1	0	0	3
n_i	3	4	9	4	5	$n=25$

За даними кореляційної таблиці побудуємо ряд, що відображає залежність середнього значення *Y* від *X* (табл. 4.2), для чого знайдемо середні значення $\overline{y}_{x_i}$ для кожного значення x_i, $i=\overline{1,5}$, і заповнимо таблицю 4.8:

$$\overline{y}_{x_1} = \frac{y_1 n_{11} + y_2 n_{12} + y_3 n_{13} + y_4 n_{14}}{n_1} = \frac{20\cdot 1 + 25\cdot 2}{3} \approx 23{,}33; \quad \overline{y}_{x_2} = \frac{15\cdot 2 + 20\cdot 2}{4} = 17{,}5;$$

$$\overline{y}_{x_3} = \frac{15\cdot 5 + 20\cdot 3 + 25\cdot 1}{9} = 17{,}78; \quad \overline{y}_{x_4} = \frac{10\cdot 1 + 15\cdot 3}{4} = 13{,}75;$$

$$\overline{y}_{x_5} = \frac{10\cdot 3 + 15\cdot 1 + 20\cdot 1}{5} = 13.$$

Таблиця 5.8

x_i	25	50	75	100	125
$\overline{y}_{x_i}$	23,33	17,5	17,78	13,75	13
n_i	3	4	9	4	5

Перший етап аналізу: визначимо вид залежності *Y* від *X*. Побудуємо емпіричну лінію регресії (рис. 5.4).

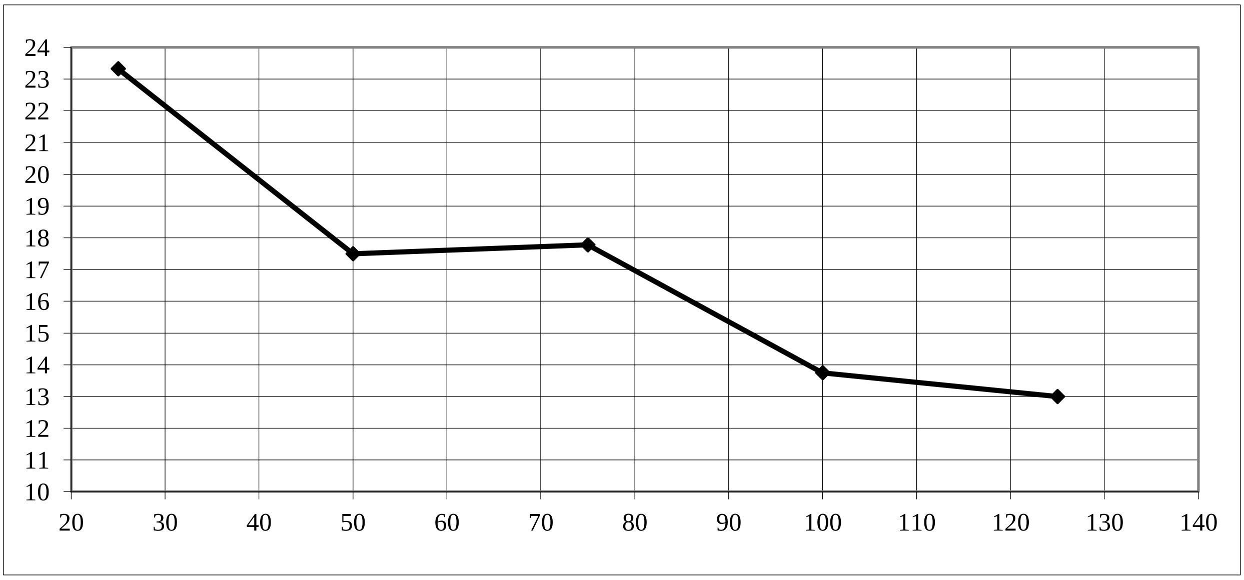

Рис. 5.4. Емпірична лінія регресії

Оскільки емпірична лінія регресії наближається до гіперболи, то висуваємо гіпотезу про гіперболічну залежність Y від X, тобто рівняння регресії будемо шукати у вигляді $\overline{y_x} = \dfrac{a}{x} + b$.

Другий етап: знайдемо параметри a, b рівняння регресії, для чого складемо систему для згрупованих даних. Необхідні розрахунки для зручності оформимо у вигляді таблиці (табл. 4.9), в останньому рядку якої знайдемо відповідні стовпцям суми.

Таблиця 5.9

x_i	$\overline{y}_{x_i}$	n_i	$\dfrac{1}{x_i}$	$\dfrac{1}{x_i} n_i$	$\dfrac{1}{x_i^2} n_i$	$\overline{y}_{x_i} n_i$	$\dfrac{1}{x_i} \overline{y}_{x_i} n_i$
25	23,33	3	0,04	0,12	0,0048	69,99	2,7996
50	17,5	4	0,02	0,08	0,0016	70	1,4
75	17,78	9	0,0133	0,12	0,0016	160,02	2,1336
100	13,75	4	0,01	0,04	0,0004	55	0,55
125	13	5	0,008	0,04	0,0003	65	0,52
Суми				0,4	0,0087	420,01	7,4032

Отже, складемо систему для знаходження параметрів рівняння регресії та розв'яжемо її за правилом Крамера:

$$\begin{cases} a\sum_{i=1}^{k}\dfrac{1}{x_i^2}n_i + b\sum_{i=1}^{k}\dfrac{1}{x_i}n_i = \sum_{i=1}^{k}\dfrac{1}{x_i}\overline{y_{x_i}}n_i \\ a\sum_{i=1}^{k}\dfrac{1}{x_i}n_i + b\sum_{i=1}^{k}n_i = \sum_{i=1}^{k}\overline{y_{x_i}}n_i \end{cases} \Rightarrow \begin{cases} 0,0087a + 0,4b = 7,4032 \\ 0,4a + 25b = 420,01 \end{cases}.$$

Головний визначник системи: $\Delta = \begin{vmatrix} 0,00872 & 0,4 \\ 0,4 & 25 \end{vmatrix} = 0,00872 \cdot 25 - 0,4^2 = 0,058$.

Допоміжні визначники: $\Delta a = \begin{vmatrix} 7,4032 & 0,4 \\ 420 & 25 \end{vmatrix} = 7,4032 \cdot 25 - 0,4 \cdot 420 = 17,076$;

$\Delta b = \begin{vmatrix} 0,00872 & 7,4032 \\ 0,4 & 420 \end{vmatrix} = 0,00872 \cdot 420 - 7,4032 \cdot 0,4 = 0,701207$.

Формули Крамера: $a = \dfrac{\Delta a}{\Delta} = \dfrac{17,076}{0,058} \approx 294,41$; $b = \dfrac{\Delta b}{\Delta} = \dfrac{0,7012207}{0,058} \approx 12,09$.

Отже, шукане рівняння регресії має вигляд $\overline{y_x} = \dfrac{294,41}{x} + 12,09$.

Третій етап: перевіримо правильність побудови моделі за рівнянням (5.4), її статистичну значущість за F-статистикою (5.5) і адекватність вибірковим даним за коефіцієнтом детермінації (5.6). Для чого знайдемо загальну варіацію, варіації регресії та залишків; необхідні розрахунки оформимо у вигляді таблиці (табл. 5.10).

Знайдемо $\bar{y}$: $\bar{y} = \dfrac{\sum_{i=1}^{k} \overline{y_{x_i}} n_i}{n} \approx 16,8$.

Таблиця 5.10

x_i	y_i	n_i	$y_{i,\text{теор}}$	$\left(y_i - \bar{y}\right)^2 n_i$	$\left(y_{i,\text{теор}} - \bar{y}\right)^2 n_i$	$\left(y_{i,\text{теор}} - y_i\right)^2 n_i$
25	23,33	3	23,866	127,907	149,782	0,863
50	17,5	4	17,978	1,958	5,547	0,914
75	17,78	9	16,015	8,637	5,547	28,028
100	13,75	4	15,034	37,220	12,482	6,594
125	13	5	14,445	72,215	27,737	10,441
Суми				247,936	201,096	46,840

Отже, $Q = 247,936$; $Q_p = 201,096$; $Q_o = 46,840$; тоді основне варіаційне рівняння $Q = Q_p + Q_o$ для побудованої моделі має вигляд: $247,936 = 201,096 + 46,840$ і є тотожністю, тому рівняння регресії побудовано правильно.

Для перевірки статистичної значущості рівняння регресії знайдемо F-статистику, враховуючи, що n=25, l=2 – оскільки шукали рівняння з двома параметрами:

$$F = \frac{Q_p(n-l)}{Q_o(l-1)} = \frac{201,096(25-2)}{46,840(2-1)} \approx 98,75 \cdot$$

Знайдемо $F_{кр}$: $F_{кр}$= $FРАСПОБР\left(0,001,2-1,25-2\right) \approx$ 14,20. Розраховане значення F-статистики більше критичного, тому регресійна модель є статистично значущою на рівні 0,001.

Знайдемо коефіцієнт детермінації R^2 : $R^2 = \dfrac{Q_p}{Q} = \dfrac{201,096}{247,936} \approx 0,81$. Значення коефіцієнта детермінації свідчить, що 81% варіації результативної ознаки Y пояснюється рівнянням регресії.

Висновок: Залежність собівартості одиниці продукції Y (грн.) від об'єму виробництва X (тис. од.) описується рівнянням $\overline{y_x} = \dfrac{294,41}{x} + 12,09$, яке є статистично значимим на рівні значущості 0,001 та описує 81% вибіркових даних.

РОЗДІЛ 6 НАДІЙНІСТЬ СИСТЕМ

Надійністю називається властивість ТЗ виконувати задані функції, зберігаючи в часі значення встановлених експлуатаційних показників в заданих межах, що відповідають заданим режимам та умовам використання, технічного обслуговування, збереження і транспортування.

Надійність включає в себе такі властивості: безвідмовність (faultlessness), довговічність (longevity), збережність (safety) та ремонтопридатність.

Розглянемо основні терміни і означення, що використовуються в теорії надійності згідно з міжнародним стандартом ІСО 8402-86 та ГОСТ 27.002–83 «Надёжность в технике. Термины и определения».

Роботоздатність (capacity) ТЗ – стан технічного засобу, при якому він здатний виконувати задані функції з параметрами, встановленими вимогами нормативно-технічної та конструкторсько-технологічної документації.

Відмова (refuse) – подія, що вказує на порушення роботоздатності ТЗ.

Критерій відмови (criterion of refuse) – ознака, за якою оцінюється надійність різних ТЗ.

Безвідмовність (faultlessness) – властивість ТЗ безупинно зберігати роботоздатний стан протягом деякого часу.

Напрацювання (work) – тривалість роботи ТЗ в годиннах, циклах, календарних днях та ін.

Напрацювання до відмови (work is to the refuse) – напрацювання ТЗ від початку його експлуатації до виникнення першої відмови.

Граничний стан (maximum state) – стан ТЗ, при якому його подальше застосування за призначенням стає неприпустимим чи недоцільним.

Довговічність – властивість ТЗ зберігати роботоздатний стан до настання граничного стану при встановленій системі технічного обслуговування і ремонтів.

Ремонтопридатність – властивість ТЗ, яка полягає в можливості попередження і виявлення причин виникнення відмов, підтримання і відновлення роботоздатного стану шляхом проведення технічного обслуговування і ремонтів.

Збережність – властивість ТЗ зберігати значення показників безвідмовності, довговічності і ремонтопридатності протягом експлуатації, зберігання та транспортування.

Ресурс (resource) – напрацювання ТЗ від початку його експлуатації чи відновлення після ремонту до переходу в граничний стан.

Термін експлуатації (term of exploitation) ТЗ – календарна тривалість від початку експлуатації ТЗ чи відновлення після ремонту до переходу в граничний стан.

Середній час відновлення (mean time of renewal) – це математичне сподівання часу відновлення роботоздатного стану.

Конструктивно всі ТЗ можна розділити на невідновлювані та відновлювані. **Невідновлюваними (unrefurbishable)** називають такі ТЗ, що у процесі виконання своїх функцій не можуть ремонтуватися, а **відновлювані** – ремонтуються. З огляду на цю властивість окремо розраховують і нормують показники надійності для відновлюваних та невідновлюваних ТЗ.

Показник надійності (reliability index) – це кількісна характеристика однієї або декількох властивостей, що визначають надійність ТЗ.

Метрологічна справність (metrology good condition) технічного засобу – це стан ТЗ, що визначає відповідність його нормованих метрологічних характеристик встановленим вимогам.

Метрологічна надійність (metrology reliability) ТЗ – це надійність ТЗ в частині збереження його метрологічної справності.

Метрологічна відмова (metrology refuse) ТЗ – це відмова ТЗ, що полягає у втраті його метрологічної справності.

Нестабільність (instability) метрологічної характеристики ТЗ – це зміна метрологічної характеристики ТЗ за встановлений інтервал часу.

Довірчі межі (confiding limits) нестабільності метрологічної характеристики ТЗ – це верхня і нижня межі інтервалу, що охоплює нестабільність метрологічної характеристики ТЗ з деякою довірчою вірогідністю.

Вірогідність (authenticity) метрологічної справності ТЗ – це вірогідність того, що в заданий момент часу ТЗ виявиться метрологічно справним.

Середній час (середнє напрацювання) до метрологічної відмови ТЗ – це математичне сподівання календарного часу експлуатації (напрацювання) ТЗ до першої метрологічної відмови.

Напрацювання на метрологічну відмову ТЗ – це відношення сумарного напрацювання ТЗ в стані метрологічної справності на заданий період експлуатації до математичного сподівання числа його метрологічних відмов за цей період.

Інтенсивність (intensity) метрологічних відмов ТЗ – це умовна щільність вірогідності метрологічної відмови ТЗ, яка визначається для даного моменту часу за умови, що до цього моменту відмови не відбулося.

Базою математичного апарату теорії надійності є

Теорія ймовірностей — розділ математики, що вивчає закономірності випадкових явищ: випадкові події, випадкові величини, їхні функції, властивості і операції над ними. Математичні моделі в теорії ймовірностей описують з деяким ступенем точності випробування (експерименти, спостереження, вимірювання), результати яких неоднозначно визначаються умовами випробування.

Математична статистика — розділ математики, в якому на основі дослідних даних вивчаються імовірнісні закономірності масових явищ. Основними задачами математичної статистики є статистична перевірка гіпотез, оцінка розподілу статистичних імовірностей та його параметрів, вивчення статистичної залежності, визначення основних числових характеристик випадкових вибірок, якими є: вибіркове середнє, вибіркові дисперсії, стандартне відхилення. Прикладом перевірки таких гіпотез є з'ясування питання про те, змінюється чи не змінюється виробничий процес з часом. Прикладом оцінки параметрів є оцінка середнього значення статистичної змінної за дослідними даними. Для вивчення статистичної залежності використовують методи теорії кореляції.

Математи́чна ло́гіка є наукою про закони математичного мислення. Предметом математичної логіки є математичні теорії в цілому, які вивчаються за допомогою логіко-математичних мов. При цьому в першу чергу цікавляться питаннями несуперечливості математичних теорій, їх розв'язності та повноти.

Теорія випадкових процесів — підрозділ математики (а саме теорії імовірностей), який займається вивченням випадкових процесів, їх властивостей та застосування. В рамках цієї теорії запроваджено концепцію інтеграла від випадкового процесу відносно випадкового процесу. Використовується для моделювання систем, які поводяться випадково

Теорія масового обслуговування (теорія черг) — розділ теорії ймовірностей, метою досліджень якого є раціональний вибір структури системи обслуговування та процесу обслуговування на основі вивчення потоків вимог на обслуговування, що надходять у систему і виходять з неї, тривалості очікування і довжини черг [1]. У теорії масового обслуговування використовуються методи теорії ймовірностей та математичної статистики.

Теорія графів — розділ математики, що вивчає властивості графів. Останні спрощено можна розглядати як сукупність точок (вершини) сполучених лініями (ребрами). Визначення графу є настільки загальним, що цим терміном можна описувати безліч подій та об'єктів повсякденного життя. Високий рівень абстракції та узагальнення дозволяє використовувати типові алгоритми теорії графів для вирішення зовнішньо несхожих задач у транспортних ікомп'ютерних мережах, будівельному проектуванні, молекулярному моделюванні тощо.

Залежність інтенсивності відмов від часу

Будь-який технічний засіб з початку і до кінця експлуатації має три найбільш характерних періоди роботи:

1) припрацювання ($0 < t < t_1$);

2) нормальна експлуатація ($t_1 < t < t_2$);

3) старіння чи знос ($t > t_2$) (рис. 1).

Рисунок 1 – Залежність інтенсивності відмов від часу роботи ТЗ

Період припрацювання характеризується високою інтенсивністю відмов, викликаних відхиленням від вимог конструкторсько-технологічної документації, що розподіляються за законом розподілу Вейбулла й усуваються за рахунок введення технологічного припрацювання («технологічного прогону»). Як видно з рис. 1 інтенсивність відмов на першому періоді монотонно зменшується.

Період нормальної експлуатації характеризується мінімальною і постійною інтенсивностями відмов. Ці відмови називаються **раптовими**, носять випадковий характер і розподіляються як правило за експоненціальним законом розподілу. Тут інтенсивність відмов залишається приблизно однаковою (див. рис. 1).

Період старіння або зносу характеризується різким збільшенням інтенсивності **зносових** відмов, що розподіляються за нормальним законом розподілу (законом Гаусса). На третьому періоді, як видно з рис. 1, інтенсивність відмов постійно зростає.

Виходячи з вище викладеного розглянемо детальніше найчастіше використовувані для розрахунку надійності ТЗ закони розподілу, що характеризують безперервні випадкові величини.

Основними нормованими показниками надійності невідновлюваних ТЗ можуть бути такі показники:

– ймовірність безвідмовної роботи (probability of faultless work) P(t);

– частота відмов (frequency of refuses) a(t);

– інтенсивність відмов (intensity of refuses) λ(t);

– середнє напрацювання (middle work) до першої відмови $T_{ср}$.

Розподіл Вейбулла

Розподіл Вейбулла (distributing of Veybulla) – двопараметричний закон розподілу випадкового напрацювання до відмови з параметрами: λ_0 , що визначає масштаб, і k, що визначає асиметрію.

Показники надійності при такому законі розподілу будуть визначатися так:

$$P(t) = e^{-\lambda t^k}, \tag{6.1}$$

$$a(t) = -P'(t) = \lambda_0 k t^{k-1} e^{-\lambda_0 t^{k-1}}, \tag{6.2}$$

$$\lambda(t) = \frac{a(t)}{P(t)} = \lambda_0 k t^{k-1}, \tag{6.3}$$

$$T_{cp} = \int_0^\infty P(t)dt = \frac{\Gamma(\frac{1}{k}+1)}{\lambda_0^{1/k}}, \tag{6.4}$$

де $\Gamma(\frac{1}{k} + 1)$ – гамма-функція.

Гамма-функція визначається за виразом

$$\Gamma(X) = \int_0^\infty t^{X-1} e^{-t} dt. \tag{6.5}$$

Характеристика зміни ймовірності безвідмовної роботи, що залежить від часу напрацювання при перерозподілі відмов за законом Вейбулла, подана на рис. 6.2.

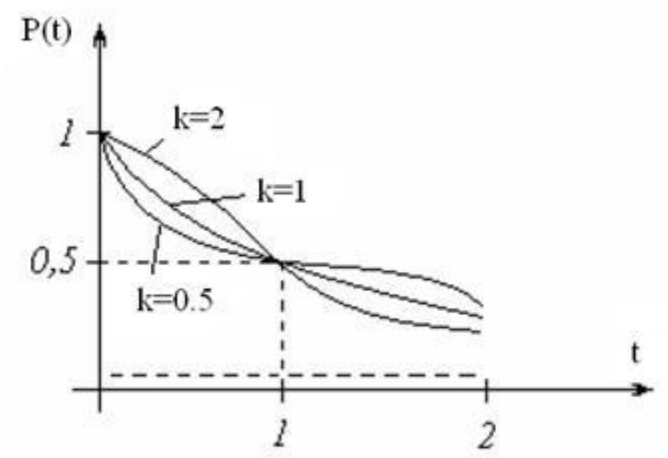

Рисунок 6.2 – Характеристика зміни ймовірності безвідмовної роботи для розподілу Вейбулла при $k = 2$, $k = 1$ та $k = 0.5$

Дисперсія часу безвідмовної роботи для розподілу Вейбулла описується виразом

$$\sigma_T^2 = \lambda_0^{-\frac{2}{k}}[\Gamma\left(\frac{2}{k} + 1\right) - \Gamma^2\left(\frac{2}{\mathrm{k}} + 1\right)]. \quad (6.6)$$

Отже, інтенсивність відмов $\lambda(t)$, що розподілені за законом Вейбулла, при $k < 1$ – монотонно зменшується, при $k > 1$ – монотонно збільшується і при $k = 1,\ \lambda = const$ (рис. 3).

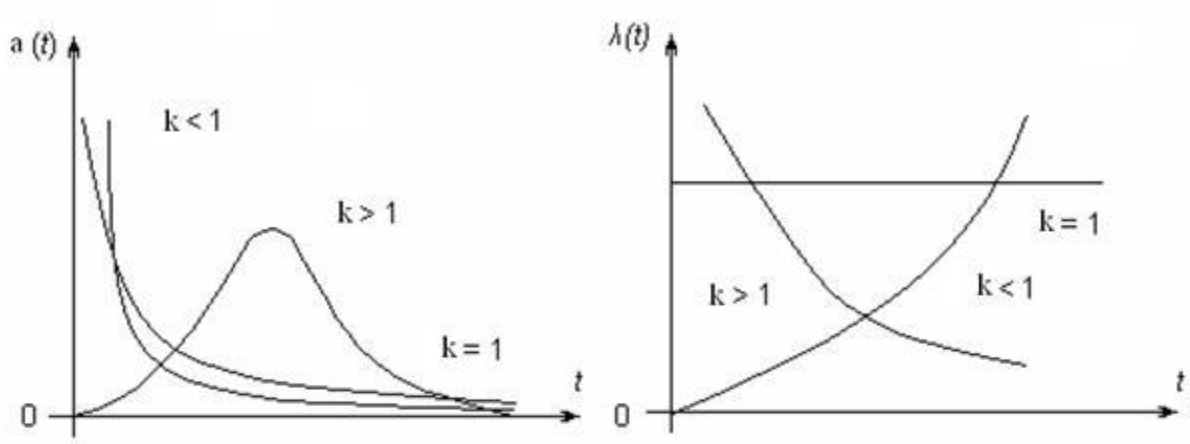

Рисунок 6.3 – Характеристики зміни частоти та інтенсивності відмов, що розподілені за законом Вейбулла.

Приклад 1. Визначіть середнє напрацювання T_{cp} та інтенсивність відмов $\lambda(t)$ для ТЗ, час безвідмовної роботи якого розподілений за законом Вейбулла з параметрами $k = 1.5$, $\lambda = 10^{-4}\text{год}^{-1}$ за час роботи $t = 100$ год.

Розв'язування:

Для розрахунку T_{cp} скористаємося формулою (4) для розподілу Вейбулла

$$T_{cp} = \frac{\Gamma(\frac{1}{k} + 1)}{\lambda_0^{1/k}} = (10 - 4) - 0{,}67\Gamma(1{,}67)$$

Знайшовши з довідникової таблиці значення гамма-функції $\Gamma(1.67) = 0{,}9033$ і провівши нескладні розрахунки, отримаємо T_ср=418 год.

Підставляючи у формулу (3) параметри розподілу Вейбулла k і λ, розрахуємо інтенсивність відмо в ТЗ за час t = 100 год.:

$\lambda(100) = \lambda k(100)^{k-1} = 1{,}5 \cdot 10^{-3}$ год$^{-1}$.

Експоненційний розподіл

Експоненційний закон розподілу – однопараметричний закон з постійною інтенсивністю відмов ($\lambda_0 = const$). Він є частковим випадком розподілу Вейбулла при $k = 1$.

Ймовірність безвідмовної роботи, частота відмов і середнє напрацювання до відмови при експоненційному розподілі визначаються за формулами:

$$P(t) = e^{-\lambda_0 t}, \tag{6.7}$$

$$a(t) = \lambda_0 e^{\lambda_0 t}, \tag{6.8}$$

$$\text{Т}_{\text{ср}} = {}^{1}\!/_{\lambda_0}. \tag{6.9}$$

З виразу (3) випливає, що при $k = 1$ інтенсивність відмов $\lambda(t) \equiv \lambda_0$. Саме тому експоненційний закон визначається тільки одним параметром λ_0, що являє собою постійну інтенсивність відмов.

Замінюючи λ_0 на $1/\text{Т}_{\text{ср}}$, отримаємо

$$P(t) = e^{\frac{-t}{T}}. \tag{6.10}$$

Ймовірність безвідмовної роботи на інтервалі часу $t = T_{cp}$ при експоненційному розподілі дорівнює

$$P(T_{cp}) = e^{-1} = 0.368. \tag{6.11}$$

Дисперсія часу безвідмовної роботи для експоненційного закону розподілу розраховується за формулою

$$\sigma_T^2 = 2\int_0^{\infty} te^{-\lambda_0 t}dt - \frac{1}{\lambda_0^2} = \frac{1}{\lambda_0^2} = T_{cp}^2, \qquad (6.12)$$

Графік зміни ймовірності безвідмовної роботи від часу при експоненційному розподілі відмов зображено на рис. 6.4.

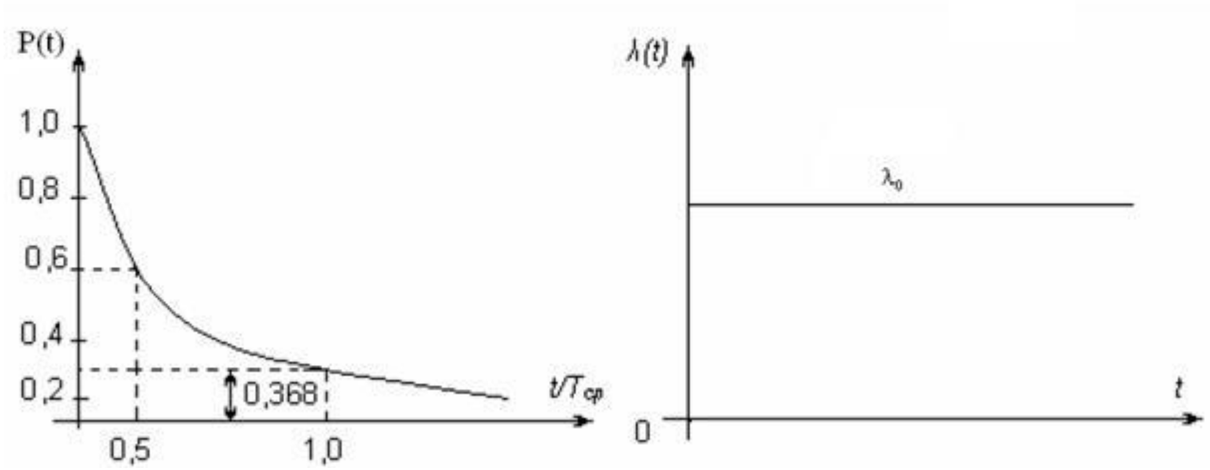

Рисунок 6.4 – Експоненційний розподіл ймовірності безвідмовної роботи та інтенсивності відмови

Знайдемо умовну ймовірність того, що для експоненційної моделі ТЗ пропрацює безвідмовно на інтервалі часу t, після того як він безвідмовно пропрацював на інтервалі T. В цьому випадку маємо

$$P\left({}^{t}\!/_{T}\right) = \frac{P(t+T)}{P(T)} = \frac{e^{-\lambda_0(t+T)}}{e^{-\lambda_0 T}} = e^{-\lambda_0 T}. \qquad (6.13)$$

Звідси випливає важливий висновок: *для експоненційного закону розподілу ймовірності безвідмовної роботи розподіл часу безвідмовної роботи не залежить від того, скільки часу ТЗ пропрацював до початку відліку від моменту першого ввімкнення.* Інші закони розподілу такої властивості не мають, оскільки в інших розподілах $\lambda_0 \neq const$, а залежить від часу.

Модель експоненційного розподілу широко використовується для апріорного аналізу надійності. При апріорному аналізі надійності необхідно проводити перевірку відповідності експоненційної моделі результатам випробувань.

Приклад 2. Напрацювання ТЗ до відмови описується експоненційним розподілом з параметром $\lambda_0 = 10^{-4}$ год$^{-1}$. Визначіть Р(t) і $a(t)$ ТЗ за час роботи $t = 2000$ год., а також визначіть середнє напрацювання Т$_{\text{ср}}$.

Розв'язування:

Відповідно до формули (7) отримаємо

$P(2000) = e^{\frac{-2000}{10^4}} = 0.819.$

Відповідно до (8) отримаємо

$a(2000) = 10^{-4}e^{\frac{-2000}{10^4}} = 8.19 * 10^{-6}$год$^{-1}$.

На основі рівняння (9) середнє напрацювання до відмови складає

Т$_{\text{ср}} = {}^{1}/_{10^{-4}} = 10^4$год.

Розподіл Релея

При розподілі Релея ймовірність безвідмовної роботи на інтервалі $(0;\ t)$ дорівнює

$$P(t) = e^{\frac{-t^2}{2r^2}}, \qquad (6.14)$$

де r – параметр розподілу Релея, який одночасно є модою цього розподілу (рис.6. 5, а).

Мода неперервного розподілу – це точка максимуму щільності розподілу ймовірності а(t). Мода дискретного розподілу – це таке спектральне значення η_m, при якому попереднє і наступне спектральні значення мають ймовірність, меншу ніж Р(η_m).

Щільність розподілу напрацювання до відмови дорівнює (рис. 6, б)

$$a(t) = -P'(T) = \frac{t}{r^2}e^{\frac{-t^2}{2r^2}}. \qquad (6.15)$$

Інтенсивність відмов дорівнює (рис.6. 5, в)

$$\lambda(t) = \frac{a(t)}{P(t)} = \frac{t}{r^2}. \qquad (6.16)$$

Середній час безвідмовної роботи Т$_{\text{ср}}$ для розподілу Релея описується виразом

$$Т_{ср} = \int_0^\infty ta(t)dt = r\sqrt{\frac{\pi}{2}} = 1.253r. \qquad (6.17)$$

Відповідно дисперсія часу безвідмовної роботи описується виразом

$$\sigma_T^2 = \left(2 - \frac{\pi}{2}\right) r^2 = 0.429 r^2. \qquad (6.18)$$

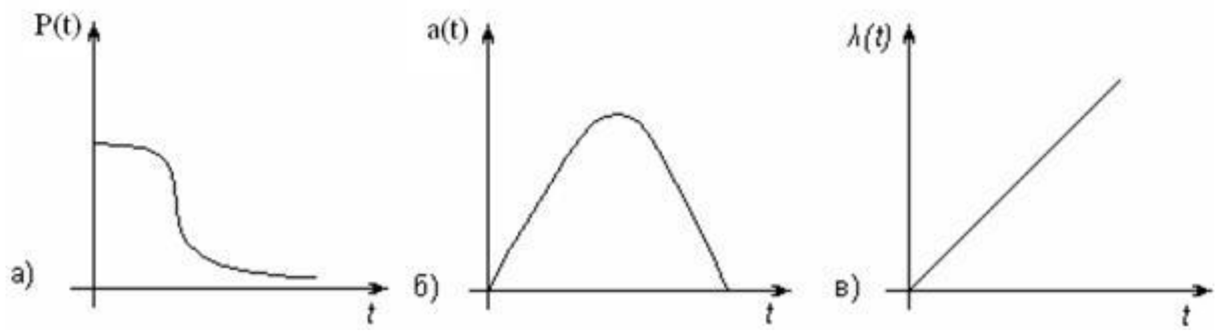

Рисунок 6.5 – Характеристики зміни ймовірності безвідмовної роботи, щільності розподілу напрацювання до відмови та інтенсивності відмови розподіленими за законом Релея

Гамма-розподіл

При гамма-розподілі щільність розподілу напрацювання до відмови описується виразом

$$a(t) = \frac{\lambda_0^2}{\Gamma(r)} t^{r-1} e^{-\lambda_0 t}, \qquad (6.19)$$

де $\Gamma(r)$ – повна гамма-функція.

В теорії надійності гамма-розподіл, як правило, використовується при цілому значенні параметра r. Якщо $r = 1$, то гамма-розподіл пере-творюється в експоненційний розподіл. Якщо r – ціле число більше 1, то гамма-розподіл є розподілом суми незалежних випадкових величин, кожна з яких має експоненційний розподіл.

Гамма-розподіл при цілому значенні r інколи називають *розподілом Ерланга*. Для такого розподілу ймовірність безвідмовної роботи на інтервалі $(0;\ t)$ описується виразом (рис. 6, а):

$$P(t) = e^{-\lambda_0 t} \sum_{i=0}^{r-1} \frac{(\lambda_0 t)^i}{i!}. \qquad (6.20)$$

Щільність розподілу напрацювання до відмови в цьому випадку описується виразом (рис. 6.6, б)

$$a(t) = \lambda_0 \frac{(\lambda_0 t)^{r-1}}{(r-1)!} e^{-\lambda_0 t}. \qquad (6.21)$$

Інтенсивність відмов визначається за формулою (6, в)

$$\lambda(t) = \lambda_0 \frac{(\lambda_0 t)^{r-1}}{(r-1)! \sum_{i=0}^{r-1} \frac{(\lambda_0 t)^i}{i!}}. \quad (6.22)$$

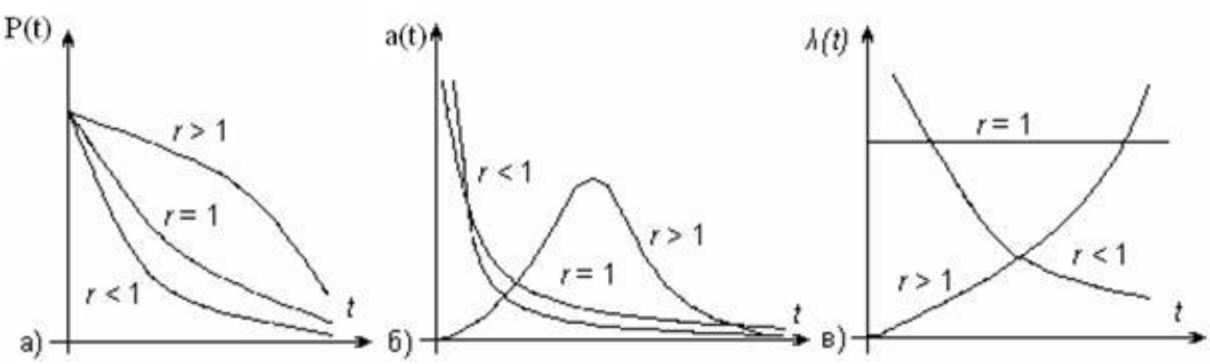

Рисунок 6.6 – Характеристики зміни гамма-розподілу: а) – функція надійності; б) – щільність розподілу напрацювання до відмови; в) – інтенсивність відмов

Середній час безвідмовної роботи і дисперсія часу безвідмовної роботи відповідно описуються виразами:

$$T_{cp} = \frac{r}{\lambda_0}, \quad (6.23)$$

$$\sigma_T^2 = \frac{r}{\lambda_0^2}. \quad (6.24)$$

При великих значеннях параметра r – гамма-розподіл наближається до нормального закону розподілу з параметрами: $\mu_{t0} = rT_{cp}$, $\sigma_{T0}^2 = r\sigma_T^2$.

Прикладом використання гамма-розподілу є резервна система, що складається з r однакових елементів. При цьому під навантаженням знаходиться один елемент. Інші елементи почергово автоматично вмикаються в роботу після відмови працюючого елемента. При експоненційному напрацюванні до відмови елементів їх сумарне напрацювання буде підпорядковуватися гамма-розподілу.

Нормальний розподіл

Нормальний закон розподілу (normal law of distributing) – це двопараметричний закон з параметрами розподілу: T – математичне сподівання і σ_T – СКВ (час безвідмовної роботи).

Ймовірність події в інтервалі часу від t_1 до t_2 визначається за формулою

$$P(t_1 < t < t_2) = \frac{1}{2}[\Phi(x_2) - \Phi(x_1)] = \frac{1}{2}\left[\Phi\left(\frac{t_2 - T_{cp}}{\sigma_T\sqrt{2}}\right) - \Phi\left(\frac{t_1 - T_{cp}}{\sigma_T\sqrt{2}}\right)\right], \quad (6.25)$$

Де $\Phi(x)$– інтеграл ймовірності (інтеграл Лапласа) виду

$$\Phi(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t}dt. \qquad (6.26)$$

При використанні центрованої і нормованої функції Лапласа $\Phi(z)$, де $z = {(t - T_{cp})}/{\sigma_T}$, ймовірність безвідмовної роботи визначається за формулою

$$P(t) = 0.5 - \Phi(\frac{t-T_{cp}}{\sigma_T}). \qquad (6.27)$$

Щільність розподілу напрацювання до відмови при нормальному розподілі має вигляд (рис.6.7)

$$a(t) = \frac{1}{\sigma_T\sqrt{2\pi}} e^{\frac{-(t-T_{cp})^2}{2\sigma_T^2}}. \qquad (6.28)$$

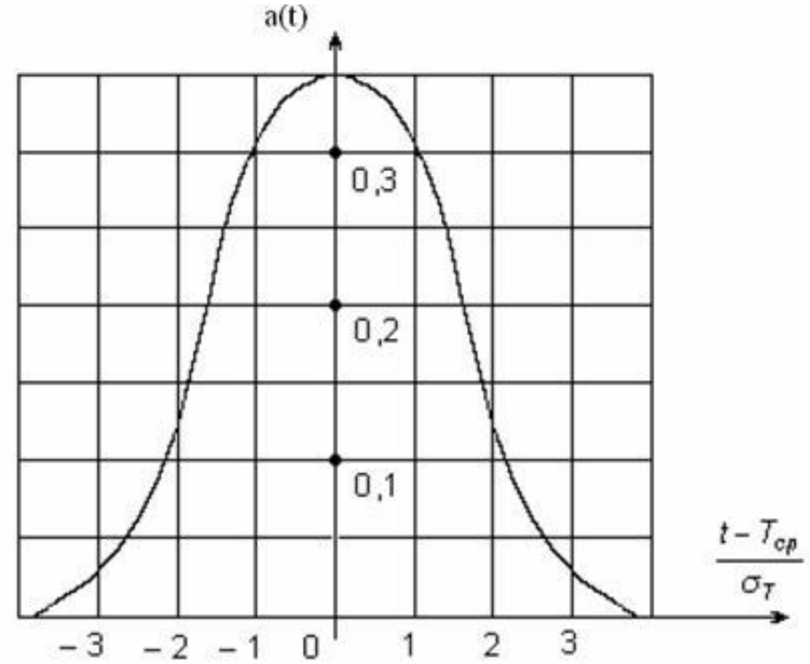

Рисунок 6.7 – Щільність розподілу напрацювання до відмови для нормального закону розподілу

Функцію надійності можна також записати у вигляді

$$P(t) = \int_t^\infty a(t)dt = \frac{1}{\sqrt{2\pi}}\int_{\frac{t-T_{cp}}{\sigma_T}}^\infty e^{\frac{-(t-T_{cp})^2}{2\sigma_T^2}} d\left(\frac{t-T_{cp}}{\sigma_T}\right) = \frac{1}{\sqrt{2\pi}}\int_{\frac{t-T_{cp}}{\sigma_T}}^\infty e^{\frac{-x^2}{2}}dx, \quad (29)$$

де $X = \frac{t-T_{cp}}{\sigma_T}$.

Ймовірність відмови на інтервалі $(0; t)$ (функція ненадійності) визначається за формулою

$$Q(t) = 1 - P(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{t-T_{cp}}{\sigma_T}} e^{\frac{-x^2}{2}}dx. \qquad (6.30)$$

РОЗДІЛ 7 ПЛАНУВАННЯ ЕКСПЕРИМЕНТУ

Історія планування експерименту.

Планування експерименту виникло в 20-х роках XX століття з потреби усунути або хоча б зменшити систематичні помилки в сільськогосподарських дослідженнях шляхом рандомізації умов проведення експерименту. Процедура планування виявилася спрямована не тільки на зменшення дисперсії оцінюваних параметрів, але також і на рандомізації щодо супутніх, спонтанно змінюються і неконтрольованих змінних. У результаті вдалося позбутися від зсуву в оцінках. З 1918 р. Р. Фішер почав свою відому серію робіт на Рочемстедсській агробіологічний станції в Англії. У 1935 році з'явилася його монографія «Design of Experiments», що дала назву всьому напрямку.

У 1942 році А. Кішен розглянув планування експерименту з латинським кубів, яке стало подальшим розвитком теорії латинських квадратів. Потім Р.Фішер незалежно опублікував відомості про ортогональних гіпер-греко-латинських кубах і гіпер-кубах. Незабаром після цього в 1946 р. Р. Рао розглянув їх комбінаторні властивості.

Подальшому розвитку теорії латинських квадратів присвячені роботи Х. Манна (1947 - 1950 рр..). Перше глибоке математичне дослідження блок-схеми виконано Р. Боуз в 1939 р. Спочатку була розроблена теорія збалансованих неповноблочних планів. Потім Р. Боуз, К. Нер і Р. Рао узагальнили ці плани і розробили теорію частково збалансованих неповноблочних планів. З тих пір вивчення блок-схем приділяється велика увага як з боку фахівців з планування експерименту (Ф. Йетс, Г. Кокс, В. Кохрен, В. Федерер, К. Гульден, О. Кемптгорн та інші), так і з боку фахівців за комбінаторним аналізу (Боуз, Ф. Шімамото, В. Клатсворсі, С. Шрікханде, А. Гофман та ін.)

Дослідження Р. Фішера знаменують початок першого етапу розвитку методів планування експерименту. Фішер розробив метод факторного планування. Також запропонував для цього методу просту обчислювальну схему. Факторний планування набуло широкого розповсюдження. Особливістю факторного експерименту є необхідність ставити відразу велику кількість дослідів. У 1945 р. Д. Фінні ввів дробові репліки від факторного експерименту. Це дозволило скоротити число дослідів і відкрило дорогу технічних програм планування. Інша можливість скорочення необхідного числа дослідів була показана в 1946 р. Р. Плакеттом і Д. Берманом, які ввели насичені факторні плани. Г. Хотеллінг в 1941 р. запропонував знаходити екстремум за експериментальними даними з використанням статечних розкладів і градієнта.

Наступним важливим етапом було введення принципу послідовного крокової експериментування. Цей принцип, висловлений в 1947 р. М. Фрідманом і Л. Севіджем, дозволив поширити на експериментальне визначення екстремуму - ітерацію.

Щоб побудувати сучасну теорію планування експерименту, не вистачало однієї ланки - формалізації об'єкта дослідження. Ця ланка з'явилося в 1947 р. після створення Н. Вінером теорії кібернетики. Кібернетичному поняття «чорний ящик», грає в плануванні важливу роль. У 1951 р. роботою американських вчених Дж. Боксу і К. Вілсона почався новий етап розвитку планування експерименту. У ній сформульована і доведена до практичних рекомендацій ідея послідовного експериментального визначення оптимальних умов проведення процесів з використанням оцінки коефіцієнтів статечних розкладів методом найменших квадратів, рух по градієнту і відшукання інтерполяційного полінома в області екстремуму функції відгуку (майже стаціонарної області).

У 1954 - 1955 рр.. Л.Ш. Бокс, а потім П. Юл показали, що планування експерименту можна використовувати при дослідженні фізико-хімічних процесів, якщо апріорі висловлені один або кілька можливих гіпотез. Напрямок отримав розвиток у роботах Н. П. Клепікова, С. Н. Соколова і В. В. Федорова у ядерній фізиці.

Третій етап розвитку теорії планування експерименту почався в 1957 р., коли Бокс застосував свій метод у промисловості. Цей метод став називатися «еволюційним плануванням». У 1958 р. Г. Шиффа запропонував новий метод планування експерименту для вивчення фізико-хімічних діаграм склад - властивість під назвою «симплексної решітки». Розвиток теорії планування експерименту в СРСР відображено в роботах В. В. Налімова, Ю. П. Адлера, Ю. В. Грановського, Е. В. Маркової.

Експеримент — спроба, дослід, які потребують підтвердження чи спростування, форма пізнання, один з основних методів наукового дослідження, в якому вивчення явищ відбувається в доцільно вибраних або штучно створених умовах, що забезпечують появу тих процесів, спостереження яких необхідне для встановлення закономірних зв'язків між явищами.

Експеримент в ході розвитку науки виступав потужним засобом дослідження явищ природи і технічних об'єктів. Але лише порівняно недавно він став предметом дослідження. Пильна увага вчених та інженерів до того, як краще і ефективніше проводити експеримент, виникло не випадково, а є наслідком досягнутого рівня і масштабу експериментальних робіт на сучасному етапі розвитку науки і техніки.

Загальними рисами експерименту є необхідність:

1) контролювати будь-який експеримент, тобто виключати вплив зовнішніх змінних, не прийнятих дослідником з тих чи інших причин до розгляду;

2) визначати точність вимірювальних приладів і одержуваних даних;

3) зменшувати до розумних меж число змінних в експерименті;

4) складати план проведення експерименту, найкращий з тієї чи іншої точки зору;

5) перевіряти правильність отриманих результатів і їх точність;

6) обирати спосіб обробки експериментальних даних і форму представлення результатів;

7) аналізувати отримані результати і давати їх інтерпретацію в термінах тієї області, де експеримент проводиться.

Постановка і організація експерименту визначають його призначення. Експерименти, які проводяться в різних галузях науки, є хімічними, біологічними, фізичними, психологічними, соціальними і т.д.

Вони відрізняються:

За способом формування умов:

- Природні (наприклад у школі — в процесі навчальних занять);
- Штучні (оптимізація відпочинку і праці на конвеєрних лініях).

За цілями дослідження:

- Перетворюють;
- Констатують;
- Контролюючі;
- Пошукові;
- Вирішальні.

За організацією проведення:

- Лабораторні;
- Натурні;
- Польові;
- Виробничі і т.д.

За структурою досліджуваних об'єктів і явищ:

- Прості;
- Складні.

За характером зовнішніх впливів на об'єкт дослідження:

- Речові;
- Енергетичні;
- Інформаційні.

За характером взаємодії кошти експериментального дослідження з об'єктом дослідження:

- Звичайний;
- Модельний.

За типом моделей, досліджуваних в експерименті:

- Матеріал;
- Уявний.

За контрольованим величин:

- Пасивний;
- Активний

За кількістю варійованих факторів:

- Однофакторний;

- Багатофакторний.

За характером досліджуваних об'єктів або явищ:

- Технологічні;

- Соціометричні.

Природний експеримент був запропонований у 1910 році російським ученим О.Ф.Лазурським. Він грунтується на керуванні поведінкою досліджуваних за умов повсякденного життя шляхом введення низки чинників, що впливають на їхню поведінку й контролюються дослідником. Самі по собі умови не містять у собі чогось штучного, незвичного для перебігу явищ повсякденного життя.

Логіка цього методу така: експериментальному впливу підлягають умови, у яких протікає досліджувана діяльність, сама ж діяльність досліджується в її природному перебігу. Замість того, щоб переводити досліджуване явище в лабораторні умови, дослідники намагаються врахувати вплив і підібрати природні умови, що відповідають їхнім цілям. Природний експеримент має особливе значення в ситуаціях, коли необхідна повна необізнаність його учасників із тим, що їх вивчають, що створена ситуація є експериментальною, бо результати можуть бути спотворені.

Лабораторний експеримент — це дослідження, проведене у штучно створених умовах (у лабораторії). На відміну від природного експерименту він передбачає організацію досить незвичної для досліджуваних ситуації. В умовах лабораторного експерименту досліджуваний знає, що його вивчають, але, як правило, не має інформації про характер завдань, які розв'язує дослідник в експерименті.

Штучним експериментом є такий, при проведенні якого незалежна змінна вводиться штучно.

Натурний експеримент (Natural experiment) - дослідження властивостей та поведінки об’єкта керування в певних умовах використовуючи сам об’єкт, є важливою складовою у сферах проектування та управління. Однак у багатьох випадках натурне моделювання є неможливим або недоцільним. Наприклад експерименти на об’єкті керування при управлінні технологічними процесами у режимі реального часу, проектуванні складних систем та пристроїв можуть бути економічно недоцільні або неможливі через неготовність самого об’єкту.

Польовий експеримент – це природний експеримент, в якому використовують обладнання, а учасників інформують про проведення дослідження.

Виробничі експерименти проводять у реальних умовах з урахуванням впливу різних факторів навколишнього середовища. Ці дослідження складніше лабораторних і вимагають ретельного планування. До виробничих досліджень відносяться різні польові випробування експлуатованих об'єктів.

Розрізняють два підходи планування експерименту:

- класичний, при якому по черзі змінюється кожен фактор до визначення часткового максимуму при постійних значеннях інших факторів,
- статистичний, де одночасно змінюють багато факторів.

При цьому суттєвим є: мінімізація числа дослідів; одночасне варіювання всіма параметрами; використання математичного апарата, який формалізує дії експериментатора; вибір чіткої стратегії, що дозволяє приймати обґрунтовані рішення після кожної серії експериментів. Загалом розрізняють такі експериментальні плани: дисперсного аналізу; відбору суттєвих факторів; багатофакторного аналізу; одержання поверхні відгуку; динамічних задач планування; вивчення механізмів явищ; побудови діаграм «склад — властивість», «склад — стан».

Для проведення експерименту будь-якого типу необхідно: розробити гіпотезу, що підлягає перевірці; створити програми експериментальних робіт; визначити способи і прийоми втручання в об'єкт дослідження; забезпечити умови для здійснення процедури експериментальних робіт; визначити способи і прийоми втручання в об'єкт дослідження; забезпечити умови для здійснення процедури експериментальних робіт; розробити шляхи та прийоми фіксації ходу і результатів експерименту (прилади, установки, моделі тощо); забезпечити експеримент необхідним обслуговуючим персоналом.

Особливе значення для проведення експерименту має правильно розроблена методика експерименту. Методика - це сукупність розумових і фізичних операцій, розміщених в певній послідовності, відповідно до якої досягається мета дослідження.

При розробці методик проведення експерименту необхідно передбачати: проведення попереднього цілеспрямованого спостереження над досліджуваним об'єктом або явищем з метою визначення вихідних даних (гіпотез, вибору варіюють факторів);

створення умов, в яких можливе експериментування (підбір об'єктів для експериментального впливу, усунення впливу випадкових факторів); визначення меж вимірювань; систематичне спостереження за ходом розвитку досліджуваного явища і точні описи фактів, проведення систематичної реєстрації вимірів і оцінок фактів різними засобами і способами; створення ситуацій, що повторюються, зміна характеру умов і перехресні впливу, створення ускладнених ситуацій з метою підтвердження або спростування раніше отриманих даних ; перехід від емпіричного вивчення до логічних узагальнень, до аналізу і теоретичній обробці отриманого фактичного матеріалу.

Перед кожним експериментом складається його план (програма), який включає: мета і завдання експерименту; вибір варіюють факторів, обґрунтування обсягу експерименту, числа дослідів; порядок реалізації дослідів, визначення послідовності зміни факторів; вибір кроку зміни факторів, завдання інтервалів між майбутніми експериментальними точками; обґрунтування способів обробки та аналізу результатів експерименту. Застосування математичної теорії експерименту дозволяє вже при плануванні певним чином оптимізувати обсяг експериментальних досліджень і підвищити їх точність.

Важливим етапом підготовки треба вибрати варійовані фактори, тобто встановити основні та другорядні характеристики, що впливають на досліджуваний процес, проаналізувати розрахункові (теоретичні) схеми процесу. На основі цього аналізу всі фактори класифікуються і складається з них регресний за важливістю для даного експерименту ряд. Правильний вибір основних і другорядних факторів відіграє важливу роль в ефективності експерименту, оскільки експеримент і зводиться до знаходження залежностей між цими чинниками. Іноді буває важко відразу виявити роль основних і другорядних факторів. У таких випадках необхідно виконувати невеликий за обсягом попередній пошуковий досвід.

Необхідно також обґрунтувати набір засобів вимірювань (приладів) іншого обладнання, машин та апаратів.

Планування експерименту включає в себе:

- визначення мети та завдань експерименту;
- місце, час проведення експерименту та його об`єм;
- характеристику вибірки та задіяних в експерименті груп;
- опис використовуваних для проведення експерименту матеріалів;
- опис методики проведення експерименту;
- опис додаткових змінних, що впливають на результати експерименту;
- опис методики фіксування, обробки та інтерпретації результатів експериментального дослідження.

Для одержання достовірної інформації важливим є виділення з об`єкту дослідження тої сукупності, яка буде підлягати вивченню, тобто формування вибіркової сукупності.

Вибіркова сукупність - частина генеральної сукупності, що виступає в якості основних об`єктів спостереження. Вибіркова сукупність повинна відображати властивості та ознаки генеральної сукупності.

Генеральна сукупність - це та сукупність об`єктів, на яку експериментатор поширює висновки дослідження, тобто та множина об`єктів, яка має спільну характеристику і вивчається в рамках дослідження на територіально-часових границях.

Процес формування вибіркової сукупності характеризується такими ознаками:

- числом ступенів відбору;
- типом виділення об`єктів репрезентації на проміжних етапах відбору;

- способом районування, виділених на проміжних етапах відбору, об`єктів репрезентації;
- способом відбору об`єктів репрезентації та одиниць спостереження на кожному етапі;
- об`ємом вибіркової сукупності (кількість одиниць спостереження).

Перші чотири ознаки описують тип вибірки, тобто особливості процесу відбору одиниць спостереження, п`ятий - об`єм вибіркової сукупності, що дозволяє розрізняти вибірку в рамках вибраного типу за кількістю одиниць спостереження.

Важливим розділом методики є вибір методів обробки і аналізу експериментальних даних. Обробка даних зводиться до систематизації всіх цифр, класифікації, аналізу. Результати експериментів повинні бути зведені в легким для читання форми запису - таблиці, графіки, формули, номограми, що дозволяють швидко і доброякісно зіставляти отримане і проаналізувати результати. Всі змінні повинні бути оцінені в єдиній системі одиниць фізичних величин.

Результати експериментів повинні відповідати трьом статистичним вимогам:

1) вимога ефективності оцінок, тобто мінімальність дисперсії відхилення щодо невідомого параметра;

2) вимога спроможності оцінок, тобто при збільшенні числа спостережень оцінка параметра повинна прагнути до його дійсного значення;

3) вимога незміщеності оцінок - відсутність систематичних помилок у процесі обчислення параметрів.

Для проведення експерименту необхідно створити певні умови. Сукупність умов, в яких відбувається експеримент, має назву експериментальної ситуації. Експериментальна ситуація має гарантувати, що саме досліджуваний у цьому експерименті чинник, а не будь-який інший, є причиною зафіксованих у ході експерименту змін в об'єкті.

Експеримент ґрунтується на розробці певної гіпотетичної моделі розглядуваного явища. На підставі цієї моделі явище описують як систему змінних, серед яких виокремлюють незалежні та залежні змінні. Незалежна змінна – це той новий чинник, що його вводять у діяльність експериментальної групи. Він повинен мати такі якості, як усталеність, самостійність, можливість справляти вплив на стан об'єкта дослідження. Залежною змінною називають такий чинник, який змінюється під впливом незалежної змінної.

Головним дослідницьким документом при проведенні експерименту є його протокол(журнал). У ньому зазначають найменування теми експериментального дослідження й час його проведення.

Спочатку аналізується обстановка до введення експериментального чинника.

У протоколі має бути відзначено, які показники виокремлені як незалежні та залежні змінні, а також за допомоги яких процедур фіксуються дані щодо всіх контрольованих змінних. Далі фіксуються початок дії експериментального чинника та те, як власне вплинула експериментальна ситуація на поведінку всієї системи

Основні етапи проведення експерименту:

- Висування експериментальної гіпотези. Експериментальна гіпотеза, на відміну від теоретичної, повинна бути сформульована у вигляді висловлення: «Якщо... те...». Крім того, вона повинна бути конкретизована. Це означає, що вхідні у висловлення «якщо А, то В» змінні А і В повинні контролюватися в експерименті: А – управлятися експериментатором, а В – реєструватися безпосередньо або за допомогою апаратури.
- Планування проведення експерименту. Планується час і місце проведення експерименту, вибирається експериментатор, складаються інструкції.
- Підготовка експерименту. Дослідник готує експериментальне помешкання й устаткування. Дослідник повинен вибрати експериментальний інструмент, що дозволяв би йому:
    - управляти незалежною перемінною;
    - реєструвати залежну перемінну.

Крім того, умови експерименту (помешкання, ситуація, час і ін.) повинні або повторювати вплив зовнішніх змінних, або зберігати константність розміру їхнього впливу на залежну змінну. Якщо це необхідно, проводиться декілька спробних дослідів для налагодження процедури експерименту.

Проведення експерименту. Експериментатор повинен чітко знати і дотримуватись порядку дій у ході дослідження (перед експериментатором можуть лежати інструкція, у якій зафіксований цей порядок) В експерименті може брати участь і асистент. Він бере на себе допоміжні задачі. Частіше усього саме асистент веде протокол. Експеримент у залежності від цілей дослідження може бути частково або цілком автоматизованим.

Статистична обробка експериментальних даних. Після проведення експерименту отримані в результаті дослідження дані опрацьовуються статистично. Звичайно методи опрацювання даних вибираються на стадії планування експерименту або ж ще раніше – при висуванні експериментальної гіпотези.

Висновки й інтерпретація результатів. Цей етап є завершальним у дослідницькому циклі. Результатом експериментального дослідження є підтвердження або спростування експериментальної гіпотези.

Електронно-обчислювальна машина (ЕОМ) є пристроєм, призначеним для виконання обчислювальних і логічних операцій у відповідності з програмою, що управляє її роботою.
ЕОМ поділяються на універсальні і спеціалізовані. Універсальні ЕОМ використовуються для вирішення будь-яких завдань, якщо вони мають алгоритм. Спеціалізовані ЕОМ призначені для задач певного призначення (керуючі, інформаційні та ін.)

Цифрові ЕОМ обробляють вводиться в них інформацію (дані) в дискретної формі у вигляді послідовних операцій (арифметичних і логічних) відповідно до заздалегідь підготовленою програмою. Після введення в пам'ять машини програма управляє роботою ЕОМ з урахуванням отриманої інформації (даних). Програма та дані вводяться в машину за допомогою пристрою введення. Результат вирішення завдань видається користувачеві в тій чи іншій формі за допомогою пристрою виведення. Послідовність операцій, певна програмою, витримується за допомогою пристрою управління. Вибравши чергову команду з пристрою пам'яті ЕОМ, керуючий пристрій готує арифметично-логічний пристрій для виконання відповідної операції, вказує адреси комірок пам'яті, з яких у арифметично-логічний пристрій мають надійти необхідні дані. Результат виконання операції вводиться в пам'ять.

Після виконання всієї програми на замовлення користувача результати видаються у вигляді роздруківки (таблиці) або виводяться на екран дисплея. Пристрої пам'яті ЕОМ поділяються на основну (оперативну) або основне запам'ятовуючий пристрій і зовнішній пристрій. У цих пристроях зберігаються програма, вихідні, проміжні та остаточні результати. Основною характеристикою основного запам'ятовуючого пристрою є ємність.

Пристрій управління, арифметично-логічний пристрій і пам'ять складають центральний процесор ЕОМ, що забезпечує керування послідовністю команд програми, виконання арифметичних і логічних операцій, виведення даних і введення результатів у пам'ять.

Обчислювальна машина також містить різноманітні за своїми функціями та принципам роботи периферійні пристрої. Сюди входять пристрої, призначені для зберігання обсягів інформації, пристрої введення в ЕОМ і виведення з неї інформації для реєстрації на носіях у вигляді печатки, перфорації і т.д. або шляхом індикації на екран (пристрої введення-виведення). Відомі в даний час пристрої введення інформації можна розділити на дві групи: ручні та автоматичні.

У групу пристроїв ручного введення входять пульти управління ЕОМ, електрифіковані друкарські машинки, дисплеї та ін
До групи автоматичних пристроїв входять пристрої для зчитування інформації з паперу й пристрої безпосереднього введення.

До пристроїв введення з проміжного носія інформації відносяться пристрої зчитування інформації з перфокарт, перфострічок і магнітних стрічок. До автоматичних засобів безпосереднього введення інформації відносяться пристрої, що зчитують інформацію зі спеціальних бланків, з друкованого тексту і з графіків. Ведуться інтенсивні розробки пристрою введення інформації з голосу. До автоматичних засобів безпосереднього введення відносяться також пристрої прийому інформації з ліній зв'язку. Пристрої виведення інформації підрозділяються на пристрої виведення: цифрової інформації на проміжний носій; на різного роду екрани (графобудівники, друкуючі пристрої); на зовнішнє середовище (пристрою видачі даних в лінії зв'язку та ін.)

Створення автоматизованих систем обробки даних, переробка інформації багатьох абонентів часто припускають використання багатомашинних обчислювальних систем. При цьому окремі ЕОМ повинні бути пристосовані до роботи з іншими машинами на відповідних рівнях організації обчислювальної системи.

Планування експерименту та обробка даних здійснюється за допомогою комп'ютерних програм: Mathcad, Statistica, Axum7, Statgraphics Plus, Simulink і ін.

Список використаної літератури

1. Адлер Ю.П. Введення в планування експерименту. - Москва: Металургія, 1968, - 155с.;
2. Адлер Ю.П., Маркова О.В., Грановський Ю.В. Планування експерименту при пошуку оптимальних умов. - М.: Наука, 1976. - 280 с.;

2. Бутник О.М. Економіко-математичне моделювання перехідних процесів у соціально-економічних системах: Монографія. – Х.: Видавничий Дім „ІНЖЕК”; СПД Лібуркіна Л.М., 2004. – 304 с.

3. Валентинов В.А. Эконометрика: Практикум. – М.: РДЛ, 2007. – 436 с.

4. Дослідження операцій: Навч. посіб. / М.Г.Медведєв, О.В.Колодінська. – 2-ге вид., перероб. і доп. – К.: Вид-во Європ. ун-ту, 2006. – 158 с.

5. 3. Красовський Г.І., Філаретов Г.Ф. Планування експерименту. - Мн.: Видавництво БГУ, 1982. - 302 с.;

6. Основи наукових досліджень: Підручник для технічних вузів / В. І. Крутов, І. М. Грушко, В. В. Попов та ін; Під ред. В. І. Крутова, В. В. Попова. - М.: Вища школа, - 1989. - 400с.: Іл.;

7. Планування експерименту в техніці / В. І. Барабащук, Б. П. Креденцера, В. І. Мірошниченко; Під ред. Б. П. Креденцера. - К.: Техніка, 1984. - 200с., Іл.

8. Дубина А.Г., Орлова С.С., Шубина И.Ю., Хромов А.В. Excel для экономистов и менеджеров. – СПб.: Питер, 2004. – 295 с.

9. Екимов С.В. Нетрадиционные подходы в экономико-математическом моделировании: Монография. – Днепропетровск: Наука и образование, 2004. – 240 с.

10. Карагодова О.О., Кігель В.Р., Рожок В.Д. Дослідження операцій: Навч. посібник. – К.: Центр учбової літератури, 2007. – 256 с.

11. Лапач С.Н., Чубенко А.В., Бабич П.Н. Статистика в науке и бизнесе. – К.: МОРИОН, 2002. – 640 с.

12. Макаренко Т.І. Моделювання та прогнозування у маркетингу: Навчальний посібник. – К.: Центр навчальної літератури, 2005. – 160 с.

13. Невежин В.П., Кружилов С.И. Сборник задач по курсу «Экономико-математическое моделирование». – М.: ОАО «Издательский дом „Городец”», 2005. – 320 с.

14. Просветов Г.И. Эконометрика: Задачи и решения: Учебно-методическое пособие. 4-е изд., доп. – М.: Издательство РДЛ, 2007. – 192 с.

15. Таха Х. Введение в исследование операций, 7-е издание.: Пер. с англ. – М.: Издательский дом "Вильямс", 2007. – 912 с.

16. Экономико-математические методы и прикладные модели: Учеб. пособие / Под ред. В.В. Федосеева. – М.: ЮНИТИ, 1999. – 321 с.

17. Сенча І.А. Економіко - математичні моделі та методи проектного менеджменту. Програма, тематичні плани і методичні рекомендації з навчально-методичного комплексу дисципліни для слухачів денної та заочної форми навчання спеціальності 8.000003 «Управління проектами» Видавництво Одеського регіонального інституту державного управління Національної академії державного управління при Президентові України