Частина 2

МАТЕМАТИЧНА СТАТИСТИКА

Математична статистика є розділом математики, безпосередньо пов'язаним з теорією ймовірностей, у якому вивчаються методи обробки й аналізу експериментальних даних, отриманих у результаті спостережень над масовими випадковими явищами. Сполучною ланкою між теорією ймовірностей і математичною статистикою є закон великих чисел і граничні теореми. Якщо ТЙ дозволяє знаходити ймовірності «складних» подій через імовірності «простих» подій (пов'язаних з ними будь-яким чином), то математична статистика за спостережуваним значенням (вибіркою) дозволяє оцінити ймовірності цих подій або здійснити перевірку припущень (гіпотез) щодо цих імовірностей.

Методи математичної статистики можна розділити на описові (дескриптивні), що дозволяють описати реальні спостереження за допомогою таблиць, графіків, характеристик положення, характеристик розсіювання й т. д., і аналітичні, що дозволяють на підставі вибіркових спостережень зробити статистично значущі висновки про наявність закономірностей для всієї сукупності. Широкому впровадженню цих методів дослідження сприяла поява в другій половині XX ст. електронних обчислювальних машин й, зокрема, персональних комп'ютерів, що дозволили зробити ці методи більш доступними й наочними.

А. Вальд[1] говорив, що «математична статистика – це теорія прийняття рішень в умовах невизначеності». Власне кажучи, математична статистика дає єдиний, математично обґрунтований, апарат для розв’язку задач керування й прогнозування за відсутності явних закономірностей у досліджуваних процесах.

[1]Вальд Абрагам (1902–1950) – американський математик.

Розділ 8

ВИБІРКОВІ РЯДИ І ЇХ ЧИСЛОВІ ХАРАКТЕРИСТИКИ

Оскільки вихідною базою для всіх побудов математичної статистики є розгляд результатів досвіду або спостережень як вибірки з деякої генеральної сукупності, то математична теорія вибірки є центральним розділом математичної статистики. Крім безпосереднього практичного застосування, теорія вибірки має фундаментальне значення для обґрунтування багатьох питань статистичного аналізу.

8.1 Генеральна сукупність і вибірка

У багатьох практичних задачах, пов'язаних з повторюваними випробуваннями, не можна провести всі можливі випробування, а можна провести лише доступну, вибіркову їхню частину, а потім зробити обґрунтований висновок. Наприклад, умовимося вважати деякий прилад стандартним, якщо тривалість його роботи становить 5000 год., у протилежному випадку − він уважається нестандартним. Досліджувати кожен прилад на тривалість роботи неможливо. Тому доводиться з усієї сукупності приладів відібрати випадково деяку частину й за тривалістю роботи відібраних приладів судити про якість усієї партії приладів. Аналогічна ситуація має місце при перевірці якості лампочок розжарювання, консервів, снарядів і т. д.

Сукупність усіх можливих значень, або реалізацій, досліджуваних ВВ, називається *генеральною сукупністю*. Вона може складатися зі скінченної або нескінченної множини значень, які називаються *елементами генеральної сукупності.*

Частина відібраних об'єктів з генеральної сукупності називається *вибірковою сукупністю*, або *вибіркою*.

Число N об'єктів генеральної сукупності й число n об'єктів вибірки будемо називати *об'єктами генеральної сукупності* й *вибірки* відповідно. При цьому будемо вважати, що $N \gg n$
(N значно більше n).

Задача математичної статистики полягає в дослідженні властивостей вибірки й узагальненні цих властивостей на всю генеральну сукупність. Отриманий при цьому висновок називається *статистичним*.

Для того щоб за вибіркою можна було досить упевнено судити про випадкову величину, вона повинна бути *представницькою* (*репрезентативною*), тобто її ймовірнісні властивості повинні збігатися або бути наближеними до властивостей генеральної сукупності.

Репрезентативну вибірку можна одержати, якщо вибирати об'єкти для дослідження випадково, тобто гарантувати всім об'єктам генеральної сукупності однакову ймовірність потрапити до вибірки. Існує кілька способів відбору, що забезпечують репрезентативність вибірки.

1. Якщо вибраний елемент, після того як над ним зроблено спостереження, повертається до генеральної сукупності перед вилученням наступного елемента, то така вибірка називається *вибіркою з поверненням*.

2. Якщо обраний елемент не повертається до генеральної сукупності, то вибірка називається *вибіркою без повернення*.

Якщо об’єм генеральної сукупності великий, а вибірка становить її невелику частину, то розходження між вибірками з поверненням і без повернення стирається. У таких випадках, як правило, використовують вибірку без повернення. Якщо генеральна сукупність має не занадто великий об’єм, то розходження між зазначеними вибірками буде істотним.

8.2 Варіаційні ряди і їхнє графічне зображення

Як правило, результати експерименту або спостереження дискретної ВВ зводять до таблиці, у першому рядку якої записується номер i експерименту, а у другому – відповідне спостереження x_i, яке називається *варіантою* ВВ X. Нижче наведено таблицю випуску приладів за одну зміну (табл. 8.1), де i – номер зміни, а варіанта x_i – число випущених приладів:

Таблиця 8.1

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x_i	5	3	5	5	6	4	7	4	5	6	6	4	5	3	6	5	4	6	5	5

Таблиці такого вигляду називаються *статистичними рядами незгрупованих даних*. Першим кроком до осмислення наявного статистичного матеріалу є розташування його варіант у порядку зростання (спадання), тобто *ранжування* варіант ряду. Ранжувальний статистичний ряд називається *варіаційним рядом*.

Нехай із генеральної сукупності витягнута вибірка, причому варіанта x_1 спостерігалася n_1 разів, $x_2 - n_2$ разів, …, $x_k - n_k$ разів. При цьому *об'єм* (*обсяг*) вибірки

$$n = \sum_{i=1}^{k} n_i , \tag{8.1}$$

де k – кількість варіант, що вирізняються числовими значеннями, числа $n_1,\ n_2, ..., n_k$ називаються *частотами* варіант відповідно $x_1,\ x_2, ..., x_k$, а їхні відношення до об'єму вибірки називаються *відносними частотами* (*частостями*), які позначаються $W_1,\ W_2, ..., W_k$, тобто

$$W_i = n_i / n, \quad i = \overline{1,k}. \tag{8.2}$$

Частоти й частості називаються *вагами*.

Для кожної вибірки виконується рівність

$$\sum_{i=1}^{k} W_i = 1. \tag{8.3}$$

При вивченні варіаційних рядів поряд з поняттям частоти використовується поняття *накопиченої частоти*, яка позначається $n_i^{\text{нак}}$. Вона показує, скільки спостерігалося варіант зі значенням ознаки, меншим за *x*. Відношення накопиченої частоти $n_i^{\text{нак}}$ до об'єму вибірки називається *накопиченою частістю*, тобто

$$W_i^{\text{нак}} = n_i^{\text{нак}} / n, \ i = \overline{1,k}. \tag{8.4}$$

Накопичені частоти (частості) для кожної варіанти знаходяться послідовним підсумовуванням частот (частостей) усіх попередніх варіант, включаючи дану.

Варіаційний ряд називається *дискретним*, якщо будь-які його варіанти вирізняються постійною величиною.

Виходячи з табл. 8.1, задамо дискретний варіаційний ряд, у якому вкажемо варіанти x_i й відповідні їм n_i, W_i, $n_i^{\text{нак}}$ і $W_i^{\text{нак}}$

(табл. 8.2).

Такі таблиці називаються *статистичними рядами згрупованих даних*.

Таблиця 8.2

x_i	n_i	W_i	$n_i^{\text{нак}}$	$W_i^{\text{нак}}$
3	2	0,1	2	0,1
4	4	0,2	6	0,3
5	8	0,4	14	0,7
6	5	0,25	19	0,95
7	1	0,05	20	1
Σ	20	1	–	–

Якщо кількість варіант занадто велика або наближена до об'єму вибірки, а також, якщо виробляється вибірка з неперервної генеральної сукупності, то доцільно складати варіаційний ряд за *інтервалами значень генеральної сукупності*.

Для побудови такого ряду весь інтервал варіювання спостережуваних значень ВВ розбивають на ряд часткових інтервалів і підраховують частоту потрапляння значень величини в кожен частковий інтервал.

Усі інтервали необхідно вибирати однакової довжини h, таким чином, щоб $x_{\min}$ увійшло до першого, $x_{\max}$ – до останнього інтервалу. Довжину часткового інтервалу h варто вибирати так, щоб побудований ряд не був громіздким й у той же час дозволяв виявити характерні риси зміни значень ВВ, тобто характерні риси досліджуваного явища.

Для більш точного визначення h можна скористатися формулою Стерджеса

$$h = (x_{\max} - x_{\min}) / (1 + 3{,}322 \lg n), \tag{8.5}$$

де $x_{\max} - x_{\min} = R$ називається *розмахом вибірки*. Якщо виявиться, що h –дробове число, то за h варто брати або найближче ціле число, або найближчий простий дріб.

За початок першого інтервалу рекомендується брати величину

$$x_{\text{поч}} = x_{\min} - 0{,}5h\,.$$

Кінець останнього інтервалу ($x_{\text{кін}}$) повинен задовольняти умову

$$x_{\text{кін}} - h \le x_{\max} < x_{\text{кін}}.$$

Тепер, переглядаючи результати спостережень, визначаємо, скільки значень ознаки потрапило до кожного конкретного інтервалу. При цьому до інтервалу включають значення ВВ, більші або рівні за нижню границю й менші за верхню границю. Такі величини називають *інтервальними частотами,* а їхнє відношення до загального числа спостережень – *інтервальними відносними частотами* (*частостями*).

При обчисленні інтервальних частостей округлення результатів варто робити таким чином, щоб загальна сума частостей дорівнювала б одиниці.

Іноді інтервальний варіаційний ряд для простоти досліджень умовно заміняють дискретним. У цьому випадку серединне значення i-го інтервалу беруть за варіанту x_i, а відповідну інтервальну частоту n_i – за частоту цієї варіанти.

Варто пам'ятати, що угруповання вибірки вносить похибку до подальших обчислень, яка росте зі зменшенням числа інтервалів.

Розглянемо приклад побудови інтервального варіаційного ряду.

Приклад 1. Подати вибірку 55 спостережень у вигляді таблиці частот. Вибірка: 17; 19; 23; 18; 21; 15; 16; 13; 20; 18; 15; 20; 14; 20; 16; 14; 20; 19; 15; 19; 16; 19; 15; 22; 21; 12; 10; 21; 18; 14; 14; 17; 16; 13; 19; 18; 20; 24; 16; 20; 19; 17; 18; 18; 21; 17; 19; 17; 13; 17; 11; 18; 19; 19; 17.

○ У наведеній вибірці

$$x_{\min} = 10,\; x_{\max} = 24,$$

тоді розмах вибірки $R = 14$.

Для вибору оптимальної величини інтервалу скористаємося формулою (8.5).

Маємо

$$h = 14/(1 + 3{,}322 \lg 55) \approx 2{,}064 \;\Rightarrow\; h = 2.$$

У цьому випадку за початок першого інтервалу рекомендується брати величину $x_{\text{поч}} = 9$.

Результати угруповання зведемо до таблиці 8.3.

Таблиця 8.3

Інтервали	Середина інтервалу	n_i	W_i	$W_i^{\text{нак}}$	n_i/h	W_i/h
9 – 11	10	1	1/55	1/55	1/2	1/110
11 – 13	12	2	2/55	3/55	1	1/55
13 – 15	14	7	7/55	2/11	7/2	7/110
15 – 17	16	9	9/55	19/55	9/2	9/110
17 – 19	18	14	14/55	33/55	7	7/55
19 – 21	20	15	3/11	48/55	15/2	3/22
21 – 23	22	5	1/11	53/55	5/2	1/22
23 – 25	24	2	2/55	1	1	1/55
Σ		55	1	–	–	–

●

1. Графіки варіаційних рядів. Для графічного зображення варіаційних рядів найбільш часто використовується полігон і гістограма.

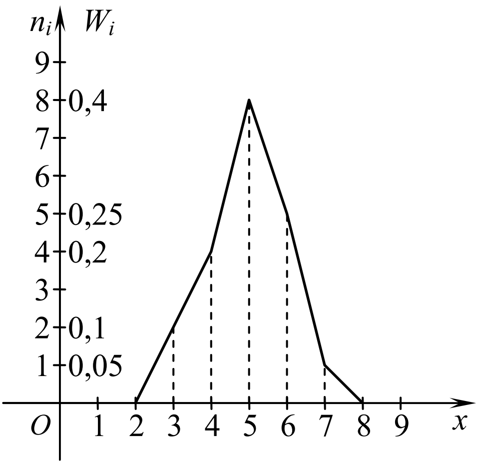

Рис. 8.1

Полігоном частот (*відносних частот* або *частостей*) називають ламану, відрізки якої з'єднують точки $(x_1; n_1)$, $(x_2; n_2)$, …, $(x_k; n_k)$, $((x_1; W_1), (x_2; W_2), ..., (x_k; W_k))$... На рис. 8.1 наведено полігони частот і частостей відповідно для згрупованих даних з табл. 8.2.

Полігон зазвичай використовують для зображення дискретного ряду.

Гістограмою частот (*відносних частот* або *частостей*) називають східчасту фігуру, що складається зі стикованих прямокутників, основами яких слугують часткові інтервали завдовжки h, а висоти дорівнюють відповідним відношенням n_i/h (W_i/h), які називаються *щільностями частот* (*відносних частот*).

Для побудови гістограми частот (частостей) на осі абсцис відкладають часткові інтервали, а над ними проводять відрізки, паралельні осі абсцис на відповідних відстанях n_i/h (W_i/h).

Площа *i*-го часткового прямокутника дорівнює $hn_i/h = n_i$ $(hW_i/h = W_i)$ – сумі частот (частостей) варіант *i*-го інтервалу, а площа гістограми частот (частостей) дорівнює об'єму вибірки (одиниці). Побудуємо гістограму частот (відносних частот) для прикладу 1 (рис. 8.2). Для графічного зображення інтервального варіаційного ряду можна використати полігон, якщо цей ряд перетворити на дискретний. У цьому випадку інтервали заміняють їхніми серединними значеннями й ставлять їм у відповідність інтервальні частоти (частості).

На рис. 8.2 полігон зображено пунктирною лінією.

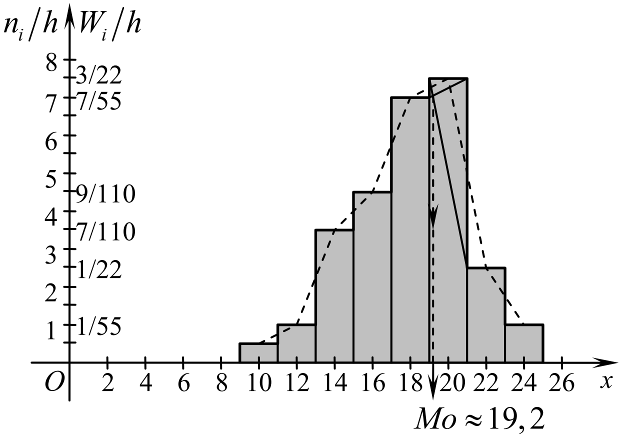

Рис. 8.2

2. Емпірична функція розподілу. Нехай є вибіркова сукупність значень випадкової величини *X*. Уведемо позначення: *x* – деяке дійсне число; n_x – число вибіркових значень ВВ *X*, менших *x*;

n – об'єм вибірки. Тоді число n_x/n є відносною частотою появи події $X < x$. Якщо x змінюється, то у загальному випадку буде змінюватися й величина n_x/n, тобто частість n_x/n є функцією від x. Оскільки ця функція знаходиться емпіричним шляхом, то її називають емпіричною.

Емпіричною функцією розподілу (*функцією розподілу вибірки*) називають функцію $F^*(x)$, що визначає для кожного значення x відносну частоту події $X < x$, тобто $F^*(x) = n_x/n$. На відміну від $F^*(x)$, функцію розподілу $F(x)$ генеральної сукупності називають *теоретичною функцією розподілу*. Відмінність між $F^*(x)$ і $F(x)$ полягає в тому, що $F(x)$ визначає ймовірність події $X < x$, а $F^*(x)$ – відносну частоту цієї ж події. Тому $F^*(x)$ можна використовувати для наближеного подання $F(x)$.

Функція $F^*(x)$ має всі властивості функції $F(x)$.

8.1° $0 \le F^*(x) \le 1$. **8.2°** $F^*(x)$ – неспадна функція.

8.3° $F^*(-\infty) = 0, \quad F^*(+\infty) = 1$.

Крива накопичених відносних частот (частостей) називається *кумулятивною кривою* (*кумулятою*) (від англ. accumulation – накопичення). Для дискретного ряду кумулята являє собою ламану, яка з'єднує точки $(x_i; W_i^{\text{нак}}), \quad i = \overline{1,k}$.

Приклад 2. Побудувати емпіричну функцію й кумуляту за даними табл. 8.2.

○ Оскільки значеннями $F^*(x)$ є накопичені частості, то аналітичне задання цієї функції має вигляд:

$$F^*(x) = \begin{cases} 0, & x \le 3; \\ 0,1, & 3 < x \le 4; \\ 0,3, & 4 < x \le 5; \\ 0,7, & 5 < x \le 6; \\ 0,95, & 6 < x \le 7; \\ 1, & x > 7. \end{cases}$$

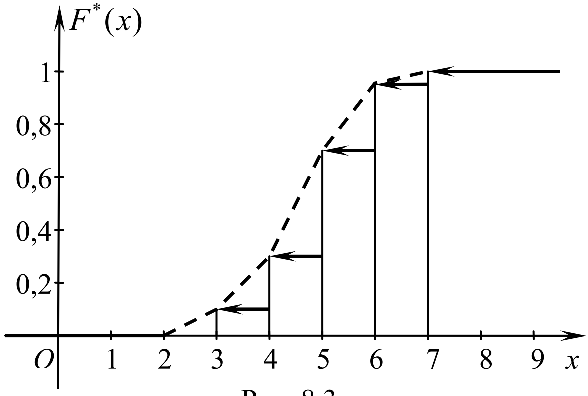

Рис. 8.3

Графік $F^*(x)$, що являє собою розривну східчасту лінію, будують так само, як і графік $F(x)$ дискретної випадкової величини (рис. 8.3), а кумуляту на цьому ж рисунку зображено пунктирною лінією. ●

Нехай варіаційний ряд складений за інтервалами значень, і як представник кожного інтервалу візьмемо його кінець. Приймаючи за координати точок кінці інтервалів і відповідні їм накопичені частості, з'єднуючи ці точки прямими, ми одержимо графік емпіричної функції розподілу (кумуляти) для цього випадку. Так, для інтервального варіаційного ряду (табл. 8.3) кумулятивну криву, яка дає наближене зображення графіка теоретичної функції $F(x)$, зображено на рис. 8.4.

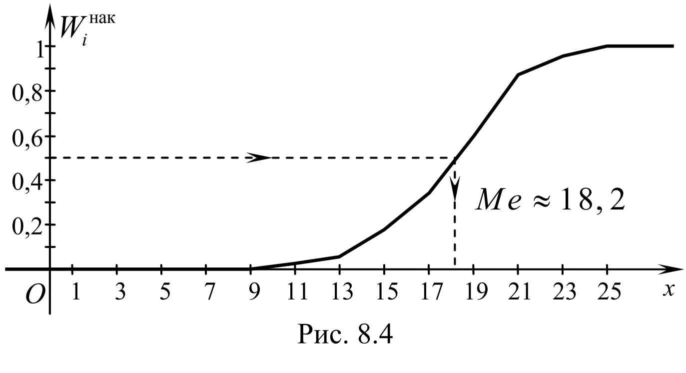

Рис. 8.4

8.3 Числові характеристики вибірки

У розділі 4 були введені різні числові характеристики: математичне сподівання; дисперсія; початкові й центральні моменти різних порядків. Вони відіграють важливу роль у теорії ймовірностей. Аналогічні числові характеристики існують і для статистичних розподілів, причому кожній числовій характеристиці ВВ X відповідає її статистична аналогія. Оскільки ці аналогії обчислюються за вибіркою, то їх зазвичай називають *статистичними характеристиками* або *оцінками*.

1. Середнє арифметичне. Нехай x_1, x_2, …, x_n – вибірка з генеральної сукупності.

Середнім арифметичним вибірки або *вибірковим середнім* називається число

$$\overline{x}_{\text{в}} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i\ . \qquad (8.6)$$

Якщо складено варіаційний ряд, у якому спостережувані варіанти x_1, x_2, …, x_k мають відповідно частоти n_1, n_2, …, n_k, причому $\sum_{i=1}^{k} n_i = n$, то

$$\overline{x}_{\text{в}} = \frac{1}{n}(x_1 n_1 + x_2 n_2 + \ldots + x_k n_k) = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \sum_{i=1}^{k} x_i W_i. \qquad (8.7)$$

Якщо варіанти вибірки, розташовані у зростаючому порядку, утворюють арифметичну проґресію з різницею h (*рівновіддалені варіанти*), то вводячи *умовну варіанту*

$$u_i = (x_i - C)/h, \qquad (8.8)$$

де C – *помилковий нуль* (для спрощення обчислень C варто вибрати рівним варіанті, яка має максимальну частоту); h – крок, тобто довжина інтервалу між сусідніми варіантами. Перетворимо (8.7) наступним чином:

$$\overline{x}_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k} \frac{x_i - C + C}{h} \cdot h n_i = \frac{h}{n}\sum_{i=1}^{k} u_i n_i + C. \qquad (8.9)$$

Якщо варіаційний ряд складений за інтервалами значень, то в ролі x_i у формулах (8.6), (8.7) і (8.9) використовують середини інтервалів.

Обчислення за формулами (8.6) і (8.7) можна спростити, якщо виконати заміну змінних $y_i = x_i - C$, де C вибирається поблизу середини варіаційного ряду. Формули (8.6) і (8.7) перетворяться наступним чином:

$$\overline{x}_{\text{в}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - C + C) = \frac{1}{n}\sum_{i=1}^{n}(x_i - C) + \frac{1}{n}\sum_{i=1}^{n} C = \frac{1}{n}\sum_{i=1}^{n} y_i + C; \qquad (8.10)$$

$$\overline{x}_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k} y_i n_i + C. \qquad (8.11)$$

Зауважимо, що середнє арифметичне − величина тієї ж розмірності, що й значення випадкової величини.

Зазначимо важливу властивість середнього арифметичного у вигляді теореми.

Теорема 8.1. Сума відхилень результатів спостережень x_i від їх середнього арифметичного $\overline{x}_{\text{в}}$ дорівнює нулю.

□ Нехай результати спостережень над ВВ X представлені у вигляді дискретного варіаційного ряду, причому кожна варіанта x_i має частоту n_i. Тоді,

$$\sum_{i=1}^{k}(x_i - \overline{x}_{\text{в}})n_i = \sum_{i=1}^{k} x_i n_i - \sum_{i=1}^{k} \overline{x}_{\text{в}} n_i = \sum_{i=1}^{k} x_i n_i - \overline{x}_{\text{в}} \cdot n = |\,8.7\,| = 0. \ \blacksquare$$

Крім розглянутих середніх величин, які називаються *аналітичними*, у статистичному аналізі застосовують *структурні* або *порядкові* середні. З них найбільш широко застосовують медіану й моду.

2. Медіана. *Медіаною вибірки* (Me) є значення серединного елемента варіаційного ряду.

Для дискретного варіаційного ряду при непарному об'єму вибірки n медіана дорівнює серединному елементу, а при n парному – півсумі двох серединних елементів.

Якщо варіаційний ряд складений за інтервалами значень, то медіана обчислюється за наступною наближеною формулою:

$$Me = x_0 + h\frac{n/2 - T_{i-1}}{n_i} \qquad (8.12)$$

де x_0 – початок медіанного інтервалу, тобто інтервалу, у якому знаходиться серединний елемент; h – довжина медіанного інтервалу; n – об’єм вибірки; T_{i-1} – сума частот інтервалів, що передують медіанному; n_i – частота медіанного інтервалу.

Медіану можна наближено визначити графічно за кумулятою. Для цього область зміни графіка кумуляти ділять навпіл. З отриманої точки проводять перпендикуляр до перетину з графіком кумуляти. Абсциса точки перетину є значенням медіани (див. рис. 8.4):

$$Me \approx 18{,}2.$$

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

3. Мода. *Модою вибірки* (*Mo*) для дискретного варіаційного ряду називається варіанта, якій відповідає найбільша частота. У випадку варіаційного ряду, складеного за інтервалами значень генеральної сукупності, мода обчислюється за наступною формулою:

$$Mo = x_0 + h\frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})}, \tag{8.13}$$

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

Моду можна також знайти графічним способом за допомогою гістограми. Для цього знаходимо прямокутник з найбільшою частотою (частістю). З'єднуючи відрізками прямі вершини цього прямокутника з відповідними вершинами двох сусідніх прямокутників (див. рис. 8.2), одержимо точку перетину цих відрізків, абсциса якої й буде модою варіаційного ряду: $Mo \approx 19{,}2$.

Особливість моди як міри центральної тенденції полягає в тому, що вона не змінюється при зміні крайніх членів ряду, тобто має певну стійкість до варіації ознаки.

Варіаційні ряди, у яких частоти варіант, рівновіддалених від середньої, рівні між собою, називаються *симетричними*. Особливість таких рядів складається в рівності трьох характеристик:

$$\overline{x}_{\text{в}} = Me = Mo \tag{8.14}$$

(це необхідна умова симетричності варіаційного ряду, але не достатня).

4. Вибіркова дисперсія. Для того щоб охарактеризувати розсіювання спостережуваних значень кількісної ознаки вибірки $x_1,\ x_2,...,x_n$ навколо свого середнього арифметичного $\overline{x}_{\text{в}}$, уводять зведену характеристику – вибіркову дисперсію, що пов'язана з *відхиленням* $x_i - \overline{x}_{\text{в}}$. Однак сума цих відхилень для будь-якого варіаційного ряду не може бути мірою розсіювання, тому що відповідно до **теореми 8.1** дорівнює нулю.

Щоб на оцінку міри розсіювання впливали всі зазначені відхилення $x_i - \overline{x}_{\text{в}},\ i = \overline{1,n}$ і їхні знаки при цьому не відігравали ролі, доцільно використати не самі відхилення, а їхні квадрати, тобто $(x_i - \overline{x}_{\text{в}})^2$.

Вибірковою дисперсією $D_{\text{в}}$ називається середнє арифметичне квадратів відхилень спостережуваних значень ознаки від їх середнього арифметичного $\overline{x}_{\text{в}}$, тобто

$$D_{\text{в}} = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x}_{\text{в}})^2 \ . \tag{8.15}$$

Перетворимо формулу (8.15) з урахуванням (8.6). Маємо

$$D_{\text{в}} = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i^{\ 2} - 2 \cdot x_i \cdot \overline{x}_{\text{в}} + \overline{x}_{\text{в}}^{\ 2}) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^{\ 2} - \frac{1}{n} \cdot 2 \cdot \overline{x}_{\text{в}} \cdot \sum_{i=1}^{n} x_i +$$

$$+ \frac{1}{n} \cdot \sum_{i=1}^{n} \overline{x}_{\text{в}}^{\ 2} = |\,8.6\,| = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^{\ 2} - 2 \cdot \overline{x}_{\text{в}}^{\ 2} + \overline{x}_{\text{в}}^{\ 2} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^{\ 2} - \overline{x}_{\text{в}}^{\ 2}.$$

Таким чином,

$$D_{\text{в}} = \frac{1}{n} \sum_{i=1}^{n} x_i^{\ 2} - \overline{x}_{\text{в}}^2. \tag{8.16}$$

Якщо значення x_i вибірки мають частоти n_i, $i = \overline{1,k}$, причому $\sum_{i=1}^{k} n_i = n$, то вибіркова дисперсія наводиться до вигляду

$$D_{\text{в}} = \frac{1}{n} \cdot \sum_{i=1}^{k} (x_i - \overline{x}_{\text{в}})^2 \cdot n_i = \frac{1}{n} \cdot \sum_{i=1}^{k} x_i^2 \cdot n_i - \overline{x}_{\text{в}}^2 \,. \tag{8.17}$$

Для рівновіддалених варіант із кроком *h* вибіркову дисперсію зручно обчислювати за формулою

$$D_{\text{в}} = \frac{h^2}{n} \cdot \sum_{i=1}^{k} u_i^2 \cdot n_i - (\overline{x}_{\text{в}} - C)^2, \tag{8.18}$$

де u_i $-$умовна варіанта (8.8), а C $-$помилковий нуль.

Для спрощення розрахунків, з урахуванням заміни змінної $y_i = x_i - C$, формули (8.15) і (8.17) можна перетворити наступним чином:

$$D_{\text{в}} = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - C - (\overline{x}_{\text{в}} - C))^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} ((x_i - C)^2 - 2 \cdot (x_i - C) \cdot (\overline{x}_{\text{в}} - C) +$$

$$+ (\overline{x}_{\text{в}} - C)^2) = \frac{1}{n} \sum_{i=1}^{n} (x_i - C)^2 - 2(\overline{x}_{\text{в}} - C) \frac{1}{n} \sum_{i=1}^{n} (x_i - C) + \frac{1}{n} \sum_{i=1}^{n} (\overline{x}_{\text{в}} - C)^2 =$$

$$= |8.10| = \frac{1}{n} \sum_{i=1}^{n} (x_i - C)^2 - 2(\overline{x}_{\text{в}} - C)^2 + (\overline{x}_{\text{в}} - C)^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - C)^2 - (\overline{x}_{\text{в}} - C)^2.$$

Таким чином,

$$D_{\text{в}} = \frac{1}{n} \sum_{i=1}^{n} y_i^2 - (\overline{x}_{\text{в}} - C)^2. \tag{8.19}$$

Аналогічно з формули (8.17), маємо

$$D_{\text{в}} = \frac{1}{n} \sum_{i=1}^{k} y_i^2 n_i - (\overline{x}_{\text{в}} - C)^2. \tag{8.20}$$

Якщо дані подані у вигляді інтервального ряду, то необхідно його замінити на дискретний, а потім скористатися наведеними вище формулами.

Часто в математичній статистиці розглядається *виправлена дисперсія*

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x}_{\text{в}})^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} x_i^2 - \overline{x}_{\text{в}} \sum_{i=1}^{n} x_i \right) = \frac{n}{n-1} D_{\text{в}}, \tag{8.21}$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{k} (x_i - \overline{x}_{\text{в}})^2 n_i = \frac{1}{n-1} \left(\sum_{i=1}^{k} x_i^2 n_i - \overline{x}_{\text{в}} \sum_{i=1}^{k} x_i n_i \right). \tag{8.22}$$

При досить великих значеннях *n* об'єму вибірки вибіркова й виправлена дисперсії розрізняються мало. На практиці користуються виправленою дисперсією, якщо наближено $n < 30$.

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

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

$$\sigma_{\text{в}} = \sqrt{D_{\text{в}}}. \qquad (8.23)$$

Виправлене середнє квадратичне відхилення знаходиться за формулою

$$S = \sqrt{S^2}. \qquad (8.24)$$

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

$$V = \frac{\sigma_{\text{в}}}{\overline{x}_{\text{в}}} \cdot 100\%, \; (x_{\text{в}} \neq 0). \qquad (8.25)$$

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

6. Вибіркові початкові й центральні моменти. Середнє арифметичне й вибіркова дисперсія є частковим випадком більш загального поняття − моменту варіаційного ряду.

Початковий момент $\tilde{v}_m$ m-го порядку варіаційного ряду визначається за формулою:

$$\tilde{v}_m = \frac{1}{n}\sum_{i=1}^{k} x_i^m n_i, \qquad (8.26)$$

де $\sum_{i=1}^{k} n_i = n$.

З визначення випливає, що $\tilde{v}_0 = 1$, а $\tilde{v}_1 = \overline{x}_{\text{в}}$, тобто середнє арифметичне є початковим моментом першого порядку варіаційного ряду.

Центральний момент $\tilde{\mu}_m$ m-го порядку варіаційного ряду визначається за формулою:

$$\tilde{\mu}_m = \frac{1}{n}\sum_{i=1}^{k} (x_i - \overline{x}_{\text{в}})^m n_i . \qquad (8.27)$$

З визначення випливає, що $\tilde{\mu}_0 = 1,$ $\tilde{\mu}_1 = 0$, а $\tilde{\mu}_2 = D_{\text{в}}$, тобто центральний момент другого порядку є дисперсією варіаційного ряду.

Центральні моменти зручно розраховувати за початковими, використовуючи наступні співвідношення між ними:

$$\tilde{\mu}_2 = \tilde{\nu}_2 - \tilde{\nu}_1^2; \tag{8.28}$$

$$\tilde{\mu}_3 = \tilde{\nu}_3 - 3\tilde{\nu}_2\tilde{\nu}_1 + 2\tilde{\nu}_1^3; \tag{8.29}$$

$$\tilde{\mu}_4 = \tilde{\nu}_4 - 4\tilde{\nu}_3\tilde{\nu}_1 + 6\tilde{\nu}_2\tilde{\nu}_1^2 - 3\tilde{\nu}_1^4. \tag{8.30}$$

На практиці зазвичай використовують моменти перших чотирьох порядків.

7. Вибіркові асиметрія й ексцес. Для оцінювання відхилення емпіричного розподілу від нормального використовують різні характеристики, до яких належать вибіркові асиметрія й ексцес.

Асиметрією вибірки називається число

$$a_s = \frac{\tilde{\mu}_3}{\sigma_{\text{в}}^3} = \frac{1}{n\sigma_{\text{в}}^3}\sum_{i=1}^{k}(x_i - \overline{x}_{\text{в}})^3 n_i. \tag{8.31}$$

Ексцесом вибірки називається число

$$e_k = \frac{\tilde{\mu}_4}{\sigma_{\text{в}}^4} - 3 = \frac{1}{n\sigma_{\text{в}}^4}\sum_{i=1}^{k}(x_i - \overline{x}_{\text{в}})^4 n_i - 3. \tag{8.32}$$

Говорять, що асиметрія емпіричного розподілу додатна (від’ємна), якщо її головна частина, тобто максимум, концентрується з лівого (правого) боку від нормального розподілу (рис. 8.5).

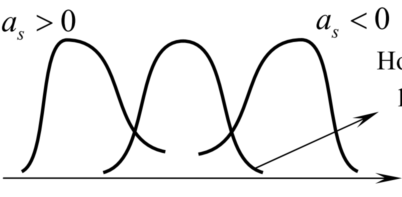

Рис. 8.5

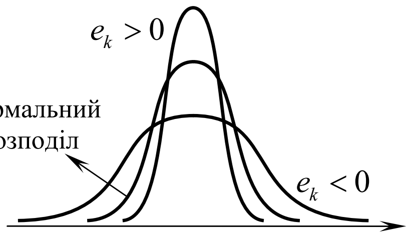

Рис. 8.6

Якщо ж максимум емпіричного розподілу розташований вище (нижче), ніж у нормального, то говорять, що емпіричний розподіл має додатний (від’ємний) ексцес (рис. 8.6).

8. Спрощений спосіб обчислення статистичних характеристик варіаційного ряду. Для спрощення обчислень зведених характеристик вибірки у випадку, якщо числові значення варіаційного ряду великі, необхідно замінити початкові варіанти на різниці $y_i = x_i - C$, де C–помилковий нуль.

Виразимо центральні моменти через

$$\tilde{v}'_m = \frac{1}{n}\sum_{i=1}^{k} y_i^m n_i. \tag{8.33}$$

З формули (8.20) випливає, що

$$D_{\text{в}} = \tilde{\mu}_2 = \frac{1}{n}\sum_{i=1}^{k} y_i^2 n_i - (\overline{x}_{\text{в}} - C)^2 = \tilde{v}'_2 - (\tilde{v}'_1)^2, \tag{8.34}$$

оскільки $\overline{x}_{\text{в}} - C = \frac{1}{n}\sum_{i=1}^{k} x_i n_i - C = \frac{1}{n}\sum_{i=1}^{k}(x_i - C)n_i = \tilde{v}'_1.$

Виразимо $\tilde{\mu}_3$ через (8.33). Маємо

$$\tilde{\mu}_3 = \frac{1}{n}\sum_{i=1}^{k}(y_i - (\overline{x}_{\text{в}} - C))^3 n_i = \frac{1}{n}\sum_{i=1}^{k} y_i^3 n_i - 3(\overline{x}_{\text{в}} - C)\frac{1}{n}\sum_{i=1}^{k} y_i^2 n_i +$$

$$+3(\overline{x}_{\text{в}} - C)^2 \frac{1}{n}\sum_{i=1}^{k} y_i n_i - (\overline{x}_{\text{в}} - C)^3 = |8.11| = \tilde{v}'_3 - 3\tilde{v}'_2\tilde{v}'_1 + 2(\tilde{v}'_1)^3.$$

Таким чином,

$$\tilde{\mu}_3 = \tilde{v}'_3 - 3\tilde{v}'_2\tilde{v}'_1 + 2(\tilde{v}'_1)^3. \tag{8.35}$$

Аналогічно одержимо, що

$$\tilde{\mu}_4 = \tilde{v}'_4 - 4\tilde{v}'_3\tilde{v}'_1 + 6\tilde{v}'_2(\tilde{v}'_1)^2 - 3(\tilde{v}'_1)^4. \tag{8.36}$$

Умовним емпіричним моментом m-го порядку називають початковий момент *m-го* порядку, обчислений для умовних варіант (8.8):

$$\tilde{v}^*_m = \frac{1}{n}\sum_{i=1}^{k} u_i^m n_i = \frac{1}{n}\sum_{i=1}^{k}\left(\frac{x_i - C}{h}\right)^m n_i. \tag{8.37}$$

Виразимо $\tilde{v}^*_m$ через $\tilde{v}'_m$, і навпаки. Маємо

$$\tilde{v}^*_m = \frac{1}{h^m}\frac{1}{n}\sum_{i=1}^{k} y_i^m n_i = \frac{\tilde{v}'_m}{h^m} \quad \Rightarrow \quad \tilde{v}'_m = h^m \tilde{v}^*_m. \tag{8.38}$$

Використовуючи (8.34)−(8.36) і (8.38), одержимо зручні для обчислень формули, що виражають центральні моменти через умовні варіанти:

$$\tilde{\mu}_2 = D_{\text{в}} = (\tilde{v}^*_2 - (\tilde{v}^*_1)^2)h^2; \tag{8.39}$$

$$\tilde{\mu}_3 = (\tilde{\nu}_3^* - 3\tilde{\nu}_2^*\tilde{\nu}_1^* + 2(\tilde{\nu}_1^*)^3)h^3; \quad (8.40)$$

$$\tilde{\mu}_4 = (\tilde{\nu}_4^* - 4\tilde{\nu}_3^*\tilde{\nu}_1^* + 6\tilde{\nu}_2^*(\tilde{\nu}_1^*)^2 - 3(\tilde{\nu}_1^*)^4)h^4. \quad (8.41)$$

Приклад 3. Обчислити $\overline{x}_{в}$, Me, Mo, $D_{в}$, $\sigma_{в}$, V, S^2, S, a_s і e_k для групованої вибірки (табл. 8.3).

○ Складемо розрахункову таблицю 8.4 відносно умовної варіанти (8.8). Як помилковий нуль виберемо варіанту $C = 18$ (ця варіанта розташована приблизно в середині варіаційного ряду), а $h = 2$.

Таблиця 8.4

x_i	n_i	u_i	n_iu_i	$n_iu_i^2$	$n_iu_i^3$	$n_iu_i^4$	$n_i(u_i+1)^4$
10	1	– 4	– 4	16	– 64	256	81
12	2	– 3	– 6	18	– 54	162	32
14	7	– 2	– 14	28	– 56	112	7
16	9	– 1	– 9	9	– 9	9	0
18	14	0	0	0	0	0	14
20	15	1	15	15	15	15	240
22	5	2	10	20	40	80	405
24	2	3	6	18	54	162	512
$\sum$	55	–	– 2	124	– 74	796	1291

Восьмий стовпець слугує для контролю обчислень за тотожністю:

$$\sum_{i=1}^{8} n_i(u_i+1)^4 = \sum_{i=1}^{8} n_iu_i^4 + 4\sum_{i=1}^{8} n_iu_i^3 + 6\sum_{i=1}^{8} n_iu_i^2 + 4\sum_{i=1}^{8} n_iu_i + n = G. \quad (8.42)$$

Контроль правильності обчислень: $\sum_{i=1}^{8} n_i(u_i+1)^4 = 1291.$

$$G = 796 - 4\cdot 74 + 6\cdot 124 - 4\cdot 2 + 55 = 1291.$$

Обчислимо умовні моменти до 4-го порядку включно:

$$\tilde{\nu}_1^* = \sum_{i=1}^{8} n_iu_i/n = -2/55 \approx -0,0364; \quad \tilde{\nu}_2^* = \sum_{i=1}^{8} n_iu_i^2/n = 124/55 \approx 2,2545;$$

$$\tilde{\nu}_3^* = \sum_{i=1}^{8} n_iu_i^3/n = -74/55 \approx -1,3455; \quad \tilde{\nu}_4^* = \sum_{i=1}^{8} n_iu_i^4/n = 796/55 \approx 14,4727.$$

Скориставшись формулами (8.9), (8.12), (8.13), (8.39), (8.23), (8.25), (8.22), (8.24), (8.40), (8.41), (8.31) і (8.32), одержимо:

$$\bar{x}_{\text{в}} = \tilde{v}_1^* h + C = -0,0364 \cdot 2 + 18 = 17,9272; \; Me = 17 + 2\frac{27,5-19}{14} \approx 18,2143;$$

$$Mo = 19 + 2\frac{15-14}{(15-14)+(15-5)} \approx 19,1818;$$

$$D_{\text{в}} = (2,2545 - (-0,0364)^2) \cdot 4 \approx 9,0127; \quad \sigma_{\text{в}} \approx 3,0021;$$

$$V = (3,0021/17,9272) \cdot 100\,\% \approx 16,75\,\%;$$

$$S^2 = (55/54)9,0127 \approx 9,1796; \quad S \approx 3,0298;$$

$$\tilde{\mu}_3 = (-1,3455 - 3 \cdot 2,2545 \cdot (-0,0364) + 2(-0,0364)^3) \cdot 8 \approx -8,7948;$$

$$\tilde{\mu}_4 = (14,4727 - 4(-1,3455)(-0,0364) + 6 \cdot 2,2545 \cdot (-0,0364)^2 -$$

$$-3(-0,0364)^4)16 \approx 228,7154; \; a_s = -8,7948/(3,0021)^3 \approx -0,325;$$

$$e_k = 228,7154/(3,0021)^4 - 3 \approx -0,1842. \; \bullet$$

Питання для самоперевірки

1. Чим займається математична статистика і яка її основна задача?

2. Що називається генеральною й вибірковою сукупностями?

3. Яка вибірка називається репрезентативною?

4. Дайте поняття статистичного ряду, а також характеристику його графічного зображення.

5. Назвіть числові характеристики статистичного розподілу.

6. Виразіть $\bar{x}_{\text{в}}$ й $D_{\text{в}}$ через умовні варіанти.

7. Виразіть $\tilde{\mu}_m$ через $\tilde{v}_m$, $m = \overline{1,4}$.

8. Що називається асиметрією й ексцесом вибірки.

9. Виразіть $\bar{x}_{\text{в}}$ й $D_{\text{в}}$ через y_i, а $\tilde{\mu}_m$ через $\tilde{v}'_m$, $m = \overline{1,4}$.

10. Виразіть $\tilde{\mu}_m$ через $\tilde{v}_m^*$.

Індивідуальні завдання 8. За вибірками A і B (табл. 8.5) необхідно: а) скласти варіаційні ряди і визначити розмахи вибірок; обчислити й занести до таблиці накопичені частоти, відносні частоти і накопичені відносні частоти; побудувати графіки варіаційних рядів (полігон і гістограму) та емпіричних функцій розподілу; б) обчислити числові характеристики варіаційних рядів: $\bar{x}_{\text{в}}$; Me; Mo; $D_{\text{в}}$; $\sigma_{\text{в}}$; V; S^2; S; $\tilde{\mu}_3$; $\tilde{\mu}_4$; a_s і e_k.

Таблиця 8.5

№	*Варіанти*											
	1		***2***		***3***		***4***		***5***		***6***	
	A	***B***	***A***	***B***	***A***	***B***	***A***	***B***	***A***	***B***	***A***	***B***
1	2	117	1	93	1	–29	2	117	1	93	1	–29
2	4	120	6	88	3	–22	4	120	6	88	3	–22
3	2	122	6	77	0	–16	2	122	6	77	0	–16
4	3	124	7	92	2	–20	3	124	7	92	2	–20
5	2	123	5	92	0	–16	2	123	5	92	0	–16
6	3	123	5	103	2	–18	3	123	5	103	2	–18
7	4	129	3	85	0	–28	4	129	3	85	0	–28
8	3	121	4	90	3	–20	3	121	4	90	3	–20
9	2	131	4	83	4	–32	2	131	4	83	4	–32
10	1	147	5	86	1	–22	1	147	5	86	1	–22
11	1	124	3	104	2	–23	1	124	3	104	2	–23
12	2	137	7	104	4	–26	2	137	7	104	4	–26
13	5	126	7	85	2	–10	5	126	7	85	2	–10
14	2	128	6	91	2	–25	2	128	6	91	2	–25
15	5	129	6	87	0	–25	5	129	6	87	0	–25
16	3	147	4	101	1	–19	3	147	4	101	1	–19
17	5	131	2	94	1	–12	5	131	2	94	1	–12
18	3	132	5	98	4	–26	3	132	5	98	4	–26
19	2	137	7	85	0	–18	2	137	7	85	0	–18
20	3	119	6	82	1	–20	3	119	6	82	1	–20
21	1	125	5	94	0	–9	1	125	5	94	0	–9
22	1	120	4	86	1	–24	1	120	4	86	1	–24
23	2	129	7	72	1	–11	2	129	7	72	1	–11
24	0	127	3	95	3	–30	0	127	3	95	3	–30
25	2	118	4	96	2	–23	2	118	4	96	2	–23
26	2	132	3	103	3	–30	2	132	3	103	3	–30
27	3	132	5	89	2	–18	3	132	5	89	2	–18
28	5	120	5	72	1	–20	5	120	5	72	1	–20
29	5	135	1	105	2	–13	5	135	1	105	2	–13
30	4	133	4	85	5	–17	4	133	4	85	5	–17
31	5	133	3	85	5	–24	5	133	3	85	5	–24
32	1	131	1	91	3	–28	1	131	1	91	3	–28
33	3	132	3	101	3	–21	3	132	3	101	3	–21
34	1	136	3	82	2	–21	1	136	3	82	2	–21
35	3	119	5	91	4	–36	3	119	5	91	4	–36
36	1	127	1	85	4	–15	1	127	1	85	4	–15
37	2	126	2	85	1	–30	2	126	2	85	1	–30
38	6	143	6	80	2	–33	6	143	6	80	2	–33
39	2	132	2	95	1	–13	2	132	2	95	1	–13
40	2	136	4	91	0	–14	2	136	4	91	0	–14

Продовження таблиці 8.5

№	***Варіанти***											
	7		***8***		***9***		***10***		***11***		***12***	
	A	***B***	***A***	***B***	***A***	***B***	***A***	***B***	***A***	***B***	***A***	***B***
1	2	60	4	78	2	48	2	60	4	78	2	48
2	2	50	4	85	1	62	2	50	4	85	1	62
3	1	58	8	52	2	50	1	58	8	52	2	50
4	3	71	7	53	0	47	3	71	7	53	0	47
5	4	61	5	62	1	80	4	61	5	62	1	80
6	2	55	5	56	1	67	2	55	5	56	1	67
7	1	75	5	58	2	87	1	75	5	58	2	87
8	1	68	3	68	3	78	1	68	3	68	3	78
9	3	65	10	98	3	88	3	65	10	98	3	88
10	3	63	6	58	0	46	3	63	6	58	0	46
11	4	60	7	94	2	57	4	60	7	94	2	57
12	3	52	10	84	1	65	3	52	10	84	1	65
13	2	69	7	57	3	60	2	69	7	57	3	60
14	4	58	7	68	4	72	4	58	7	68	4	72
15	2	56	6	64	4	28	2	56	6	64	4	28
16	1	54	10	61	3	51	1	54	10	61	3	51
17	4	61	8	64	0	34	4	61	8	64	0	34
18	3	71	10	62	0	42	3	71	10	62	0	42
19	1	51	1	53	2	43	1	51	1	53	2	43
20	4	57	3	89	3	81	4	57	3	89	3	81
21	5	55	4	66	1	87	5	55	4	66	1	87
22	1	49	9	54	0	67	1	49	9	54	0	67
23	3	67	7	62	1	65	3	67	7	62	1	65
24	3	72	4	57	2	38	3	72	4	57	2	38
25	4	54	4	35	2	58	4	54	4	35	2	58
26	7	53	5	53	4	42	7	53	5	53	4	42
27	3	61	9	73	3	61	3	61	9	73	3	61
28	4	65	5	46	2	57	4	65	5	46	2	57
29	0	66	6	44	0	72	0	66	6	44	0	72
30	3	52	4	63	0	77	3	52	4	63	0	77
31	0	62	2	71	1	73	0	62	2	71	1	73
32	3	70	5	77	0	61	3	70	5	77	0	61
33	6	72	7	60	3	91	6	72	7	60	3	91
34	2	71	3	43	0	65	2	71	3	43	0	65
35	4	71	8	76	0	80	4	71	8	76	0	80
36	3	62	8	53	3	66	3	62	8	53	3	66
37	2	63	7	58	1	69	2	63	7	58	1	69
38	0	68	4	85	3	71	0	68	4	85	3	71
39	3	60	7	87	4	73	3	60	7	87	4	73
40	1	53	5	71	2	51	1	53	5	71	2	51

Продовження таблиці 8.5

№	*Варіанти*											
	13		***14***		***15***		***16***		***17***		***18***	
	A	***B***	***A***	***B***	***A***	***B***	***A***	***B***	***A***	***B***	***A***	***B***
1	1	–51	5	53	0	149	1	–51	5	53	0	149
2	0	–37	6	46	1	166	0	–37	6	46	1	166
3	1	–72	6	59	2	174	1	–72	6	59	2	174
4	1	–64	11	64	0	161	1	–64	11	64	0	161
5	1	–77	8	50	0	164	1	–77	8	50	0	164
6	2	–58	7	57	2	158	2	–58	7	57	2	158
7	0	–50	3	68	2	163	0	–50	3	68	2	163
8	2	–39	2	48	1	160	2	–39	2	48	1	160
9	1	–48	9	43	1	154	1	–48	9	43	1	154
10	3	–44	7	62	0	168	3	–44	7	62	0	168
11	2	–65	9	40	0	172	2	–65	9	40	0	172
12	1	–69	5	44	1	166	1	–69	5	44	1	166
13	1	–57	10	45	2	177	1	–57	10	45	2	177
14	0	–84	8	60	4	169	0	–84	8	60	4	169
15	0	–65	9	38	0	153	0	–65	9	38	0	153
16	1	–49	9	72	0	179	1	–49	9	72	0	179
17	1	–21	10	43	4	164	1	–21	10	43	4	164
18	2	–40	3	33	1	151	2	–40	3	33	1	151
19	0	–27	10	60	1	170	0	–27	10	60	1	170
20	3	–67	5	50	0	175	3	–67	5	50	0	175
21	1	–41	7	61	2	170	1	–41	7	61	2	170
22	2	–46	6	42	1	151	2	–46	6	42	1	151
23	1	–79	8	41	0	155	1	–79	8	41	0	155
24	0	–37	9	47	2	173	0	–37	9	47	2	173
25	1	–52	3	54	1	162	1	–52	3	54	1	162
26	1	–50	11	64	1	158	1	–50	11	64	1	158
27	0	–70	4	48	1	149	0	–70	4	48	1	149
28	2	–90	6	66	0	171	2	–90	6	66	0	171
29	3	–72	8	42	3	160	3	–72	8	42	3	160
30	1	–61	7	70	1	147	1	–61	7	70	1	147
31	0	–39	7	59	2	149	0	–39	7	59	2	149
32	3	–44	3	46	1	148	3	–44	3	46	1	148
33	1	–36	6	71	0	173	1	–36	6	71	0	173
34	1	–43	12	60	1	158	1	–43	12	60	1	158
35	1	–58	10	63	2	161	1	–58	10	63	2	161
36	2	–35	6	72	3	152	2	–35	6	72	3	152
37	1	–38	8	54	0	157	1	–38	8	54	0	157
38	1	–27	2	77	2	148	1	–27	2	77	2	148
39	0	–72	3	49	4	169	0	–72	3	49	4	169
40	0	–34	8	38	0	162	0	–34	8	38	0	162

Продовження таблиці 8.5

№	*Варіанти*											
	19		***20***		***21***		***22***		***23***		***24***	
	A	***B***	***A***	***B***	***A***	***B***	***A***	***B***	***A***	***B***	***A***	***B***
1	0	89	8	93	5	115	0	89	8	93	5	115
2	1	126	7	81	3	110	1	126	7	81	3	110
3	2	161	4	78	4	116	2	161	4	78	4	116
4	2	140	0	92	5	120	2	140	0	92	5	120
5	0	133	4	126	3	136	0	133	4	126	3	136
6	0	151	6	104	2	129	0	151	6	104	2	129
7	0	132	5	110	3	132	0	132	5	110	3	132
8	1	120	2	120	3	118	1	120	2	120	3	118
9	0	105	4	91	4	119	0	105	4	91	4	119
10	1	127	8	101	4	131	1	127	8	101	4	131
11	0	137	6	88	5	126	0	137	6	88	5	126
12	1	120	2	69	1	119	1	120	2	69	1	119
13	4	145	5	65	4	103	4	145	5	65	4	103
14	1	154	3	71	2	109	1	154	3	71	2	109
15	0	107	3	67	5	132	0	107	3	67	5	132
16	0	106	6	112	5	108	0	106	6	112	5	108
17	2	126	6	112	3	11	2	126	6	112	3	11
18	0	95	8	86	4	135	0	95	8	86	4	135
19	1	91	9	99	4	131	1	91	9	99	4	131
20	1	99	5	46	0	115	1	99	5	46	0	115
21	0	160	5	72	2	128	0	160	5	72	2	128
22	1	131	10	73	6	126	1	131	10	73	6	126
23	2	143	3	106	7	104	2	143	3	106	7	104
24	1	157	7	92	1	126	1	157	7	92	1	126
25	1	138	5	138	3	132	1	138	5	138	3	132
26	2	117	3	130	2	117	2	117	3	130	2	117
27	1	122	7	98	3	119	1	122	7	98	3	119
28	0	115	4	94	3	122	0	115	4	94	3	122
29	0	161	9	120	6	111	0	161	9	120	6	111
30	1	107	3	114	0	101	1	107	3	114	0	101
31	1	99	2	75	6	113	1	99	2	75	6	113
32	0	102	1	100	2	139	0	102	1	100	2	139
33	2	115	4	84	4	122	2	115	4	84	4	122
34	0	117	4	121	6	124	0	117	4	121	6	124
35	0	122	2	123	1	141	0	122	2	123	1	141
36	2	96	4	132	5	139	2	96	4	132	5	139
37	1	112	3	137	4	135	1	112	3	137	4	135
38	1	144	6	75	4	118	1	144	6	75	4	118
39	3	181	1	102	5	120	3	181	1	102	5	120
40	0	155	5	105	2	97	0	155	5	105	2	97

Продовження таблиці 8.5

№	***Варіанти***											
	25		***26***		***27***		***28***		***29***		***30***	
	A	***B***	***A***	***B***	***A***	***B***	***A***	***B***	***A***	***B***	***A***	***B***
1	2	84	0	18	1	202	2	84	0	18	1	202
2	6	78	0	49	0	187	6	78	0	49	0	187
3	0	101	2	44	1	193	0	101	2	44	1	193
4	2	75	3	22	3	198	2	75	3	22	3	198
5	3	51	2	31	4	182	3	51	2	31	4	182
6	5	100	3	20	0	201	5	100	3	20	0	201
7	3	99	4	27	0	191	3	99	4	27	0	191
8	8	69	1	35	1	183	8	69	1	35	1	183
9	3	71	2	41	0	188	3	71	2	41	0	188
10	6	67	3	29	2	212	6	67	3	29	2	212
11	4	78	3	45	0	195	4	78	3	45	0	195
12	5	34	2	36	2	189	5	34	2	36	2	189
13	2	67	1	50	1	173	2	67	1	50	1	173
14	6	107	1	41	0	174	6	107	1	41	0	174
15	6	106	3	21	1	204	6	106	3	21	1	204
16	5	117	0	37	2	205	5	117	0	37	2	205
17	5	120	4	27	2	190	5	120	4	27	2	190
18	8	47	3	32	1	175	8	47	3	32	1	175
19	8	113	0	40	2	173	8	113	0	40	2	173
20	3	31	0	31	3	171	3	31	0	31	3	171
21	5	85	3	20	0	207	5	85	3	20	0	207
22	3	64	0	22	1	217	3	64	0	22	1	217
23	2	55	2	31	2	171	2	55	2	31	2	171
24	4	83	1	56	1	208	4	83	1	56	1	208
25	1	42	1	49	1	185	1	42	1	49	1	185
26	1	50	3	51	0	203	1	50	3	51	0	203
27	5	75	0	37	1	199	5	75	0	37	1	199
28	1	132	1	34	2	185	1	132	1	34	2	185
29	2	118	0	7	0	173	2	118	0	7	0	173
30	8	79	4	24	2	211	8	79	4	24	2	211
31	2	88	2	6	2	165	2	88	2	6	2	165
32	1	73	1	41	1	183	1	73	1	41	1	183
33	6	118	1	16	0	184	6	118	1	16	0	184
34	5	50	2	13	0	173	5	50	2	13	0	173
35	2	42	1	4	2	175	2	42	1	4	2	175
36	7	79	5	18	0	198	7	79	5	18	0	198
37	7	91	2	19	0	185	7	91	2	19	0	185
38	1	41	1	25	0	178	1	41	1	25	0	178
39	6	53	3	30	3	169	6	53	3	30	3	169
40	6	78	1	11	1	166	6	78	1	11	1	166

Розв'язок варіанта ІЗ

Розв'яжемо спочатку задачу за вибіркою *A*: 0; 2; 4; 3; 2; 2; 3; 3; 1; 3; 3; 3; 1; 1; 2; 3; 1; 4; 3; 1; 7; 4; 3; 4; 2; 3; 2; 3; 3; 1; 4; 3; 1; 4; 5; 3; 4; 2; 2; 4; 5; 3; 6; 4; 1; 3:

○ а) складемо варіаційний ряд і визначимо розмах вибірки. У цьому випадку $R = x_{\max} - x_{\min} = 7 - 0 = 7$. Оскільки розмах малий, то складемо дискретний варіаційний ряд (табл. 8.6).

Таблиця 8.6

x_i	n_i	$n_i^{\text{нак}}$	W_i	$W_i^{\text{нак}}$
0	1	1	0,0217	0,0217
1	8	9	0,1740	0,1957
2	8	17	0,1740	0,3697
3	16	33	0,3478	0,7175
4	9	42	0,1956	0,9131
5	2	44	0,0435	0,9566
6	1	45	0,0217	0,9783
7	1	46	0,0217	1
Σ	46	–	1	–

Для зображення дискретного варіаційного ряду зазвичай використовують полігон частот (відносних частот) (рис. 8.7).

Емпірична функція розподілу $F^*(x)$ визначається за значеннями накопичених відносних частот (частостей) співвідношенням:

$$F^*(x) = \sum_{x_i < x} (n_i / n).$$

У цьому випадку $F^*(x) = \begin{cases} 0, & x \le 0, \\ 0{,}0217, & 0 < x \le 1, \\ 0{,}1957, & 1 < x \le 2, \\ 0{,}3697, & 2 < x \le 3, \\ 0{,}7175, & 3 < x \le 4, \\ 0{,}9131, & 4 < x \le 5, \\ 0{,}9566, & 5 < x \le 6, \\ 0{,}9783, & 6 < x \le 7, \\ 1, & x > 7. \end{cases}$

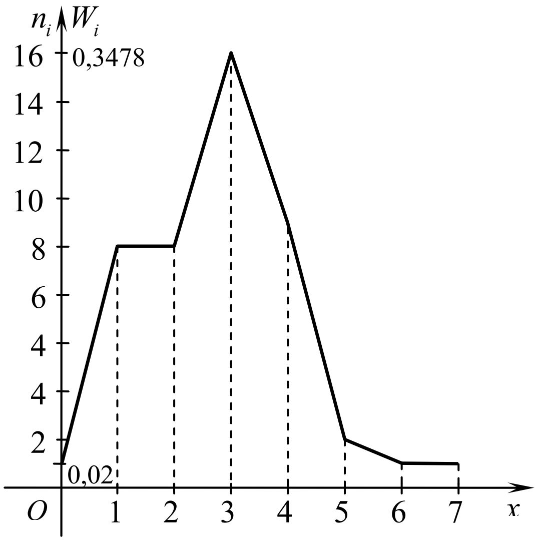

Рис 8.7

Графік $F^*(x)$ зображено на рис. 8.8, а кумуляту на цьому ж рисунку показано пунктирною лінією;

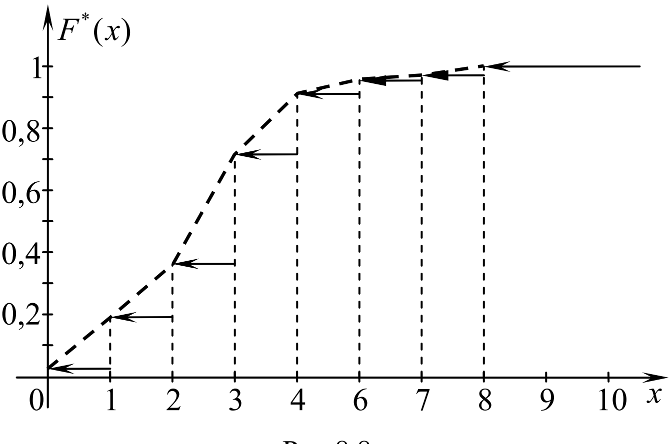

Рис 8.8

б) при обчисленні числових характеристик зручно спочатку скласти розрахункову таблицю 8.7 відносно умовних варіант (8.8). Як помилковий нуль виберемо варіанту $C = 3$, а крок таблиці $h = 1$.

Таблиця 8.7

x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$	$n_i u_i^3$	$n_i u_i^4$	$n_i(u_i+1)^4$
0	1	– 3	– 3	9	– 27	81	16
1	8	– 2	– 16	32	– 64	128	8
2	8	– 1	– 8	8	– 8	8	0
3	16	0	0	0	0	0	16
4	9	1	9	9	9	9	144
5	2	2	4	8	16	32	162
6	1	3	3	9	27	81	256
7	1	4	4	16	64	256	625
Σ	46	–	– 7	91	17	595	1227

Контроль правильності обчислень перевіримо за тотожністю (8.42): $\sum\limits_{i=1}^{8} n_i(u_i+1)^4 = 1227;$

$$G = 595 + 4\cdot 17 + 6\cdot 91 + 4(-7) + 46 = 1227.$$

Обчислимо умовні моменти до 4-го порядку включно

$$\tilde{v}_1^* = \sum_{i=1}^{8} n_i u_i \big/ n = -7/46 \approx -0{,}1522;\ \ \tilde{v}_2^* = \sum_{i=1}^{8} n_i u_i^2 \big/ n = 91/46 \approx 1{,}9783;$$

$$\tilde{v}_3^* = \sum_{i=1}^{8} n_i u_i^3 \big/ n = 17/46 \approx 0{,}3696;\ \ \tilde{v}_4^* = \sum_{i=1}^{8} n_i u_i^4 \big/ n = 595/46 \approx 12{,}9348.$$

Скориставшись формулами (8.9), (8.39), (8.23), (8.25), (8.22), (8.24), (8.40), (8.41), (8.31) і (8.32), одержимо

$$\overline{x}_{\text{в}} = \tilde{v}_1^* h + C = -0{,}1522 + 3 = 2{,}8478.$$

Оскільки у варіаційному ряді парне число варіант $n = 46$, то медіана дорівнює середньому арифметичному з двох серединних значень, тобто

$$Me = (x_{23} + x_{24})/2 = 3.$$

Мода для дискретного варіаційного ряду дорівнює варіанті, яка має найбільшу частоту, тобто $Mo = 3$.

$$D_{\text{в}} = 1{,}9783 - (0{,}1522)^2 \approx 1{,}9551;\ \ \sigma_{\text{в}} \approx 1{,}3982;$$

$$V = (1{,}3982/2{,}8478)\cdot 100\,\% \approx 49{,}1\,\%;$$

$$S^2 = (46/45)\cdot 1{,}9551 \approx 1{,}9985;\ \ S \approx 1{,}4137;$$

$$\tilde{\mu}_3 = 0{,}3696 - 3\cdot 1{,}9783(-0{,}1522) + 2(-0{,}1522)^3 \approx 1{,}2658;$$

$$\tilde{\mu}_4 = 12,9348 - 4 \cdot 0,3696(-0,1522) + 6 \cdot 1,9783 \cdot (-0,1522)^2 -$$
$$-3(-0,1522)^4 \approx 13,4332;$$
$$a_s = 1,2658 / (1,3982)^3 \approx 0,4631;$$
$$e_k = 13,4332 / (1,3982)^4 - 3 \approx 0,5148. \bullet$$

Тепер за вибіркою *B*: 67; 73; 68; 72; 67; 70; 74; 79; 65; 72; 71; 70; 69; 76; 71; 63; 65; 73; 80; 69; 69; 60; 70; 68; 74; 78; 75; 73; 72; 76; 64; 70; 68; 70; 70; 78; 73; 67; 62; 64; 70; 81; 72; 70; 70; 73, знайдемо $x_{\min} = 60$ й $x_{\max} = 81$.

Оскільки розмах вибірки досить великий $R = x_{\max} - x_{\min} = 21$, то складемо варіаційний ряд за інтервалами значень, узявши $h = 2$. За початок першого інтервалу рекомендується брати величину $x_{\text{нач}} = x_{\min} - 0,5h$. При складанні таблиці інтервального варіаційного ряду до інтервалу включають значення випадкової величини, більшої та рівної нижній границі й менші верхньої границі (табл. 8.8).

Таблиця 8.8

Інтервали	Середина інтервалу	n_i	n_i/h	W_i	$W_i^{\text{нак}}$	W_i/h
59–61	60	1	0,5	0,0217	0,0217	0,01085
61–63	62	1	0,5	0,0217	0,0434	0,01085
63–65	64	3	1,5	0,0652	0,1086	0,0326
65–67	66	2	1	0,0435	0,1521	0,02175
67–69	68	6	3	0,1304	0,2825	0,0652
69–71	70	12	6	0,2609	0,5434	0,13045
71–73	72	6	3	0,1304	0,6738	0,0652
73–75	74	7	3,5	0,1522	0,8260	0,0761
75–77	76	3	1,5	0,0652	0,8912	0,0326
77–79	78	2	1	0,0435	0,9347	0,02175
79–81	80	2	1	0,0435	0,9783	0,02175
81–83	82	1	0,5	0,0217	1	0,01085
Σ	–	46	–	1	–	–

Для зображення інтервальних варіаційних рядів використовується гістограма частот (відносних частот) (рис. 8.9).

Площа гістограми частот (відносних частот) дорівнює сумі всіх частот (відносних частот), тобто об'єму вибірки (одиниці).

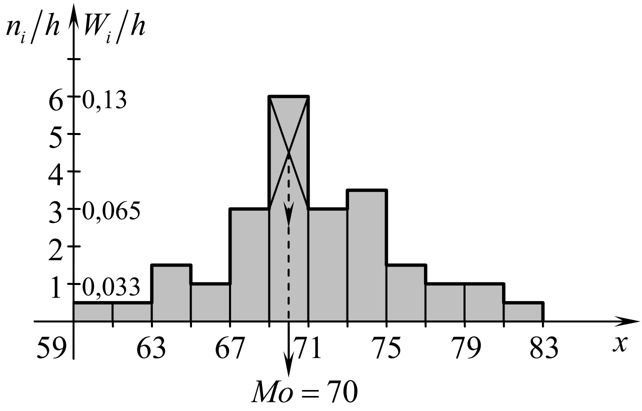

Рис. 8.9

Виходячи з табл. 8.8, побудуємо графік емпіричної функції розподілу $F^*(x)$. Як представник кожного інтервалу беремо його кінець. Уважаючи за координати точок кінці інтервалів і відповідні накопичені відносні частоти і з'єднуючи ці точки прямими, одержимо графік $F^*(x)$ (рис. 8.10), який називається кумулятою;

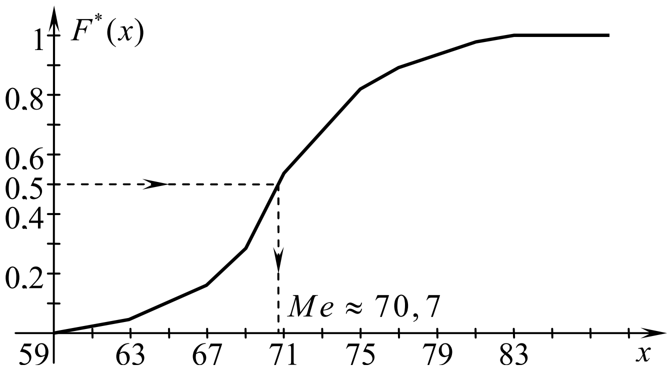

Рис. 8.10

б) складемо розрахункову таблицю 8.9 відносно умовних варіант (8.8). Як помилковий нуль виберемо варіанту $C = 70$, а крок таблиці $-$ $h = 2$.

Перевірка заповнення таблиці, за наведеною вище тотожністю (8.42), підтверджує правильність обчислень.

Таблиця 8.9

x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$	$n_i u_i^3$	$n_i u_i^4$	$n_i(u_i+1)^4$
60	1	–5	–5	25	–125	625	256
62	1	–4	–4	16	–64	256	81
64	3	–3	–9	27	–81	243	48
66	2	–2	–4	8	–16	32	2
68	6	–1	–6	6	–6	6	0
70	12	0	0	0	0	0	12
72	6	1	6	6	6	6	96
74	7	2	14	28	56	112	567
76	3	3	9	27	81	243	768
78	2	4	8	32	128	512	1250
80	2	5	10	50	250	1250	2592
82	1	6	6	36	216	1296	2401
Σ	46	–	25	261	445	4581	8073

Обчислимо умовні моменти до 4-го порядку включно за формулою (8.37):

$$\tilde{\nu}_1^* \approx 0,5435\,;\ \tilde{\nu}_2^* \approx 5,6739\,;\ \tilde{\nu}_3^* \approx 9,6739\,;\ \tilde{\nu}_4^* \approx 99,5870\,.$$

Скориставшись формулами (8.9), (8.12), (8.13), (8.39), (8.23), (8.25), (8.22), (8.24), (8.40), (8.41), (8.31) і (8.32), одержимо:

$$\overline{x}_{\text{в}} = \tilde{\nu}_1^* h + C = 2\cdot 0,5435 + 70 = 71,087\,;$$

$$Me = 69 + 2(23-13)/12 \approx 70,6667\,;$$

На рис. 8.10 показано графічне знаходження медіани.

$$Mo = 69 + 2\frac{12-6}{(12-6)+(12-6)} = 70\,;$$

На рис. 8.9 показано графічне знаходження моди.

$$D_{\text{в}} = (5,6739 - (0,5435)^2)4 \approx 21,514;\ \ \sigma_{\text{в}} \approx 4,6383\,;$$

$$V = (4,6383/71,087)\cdot 100\,\% \approx 6,52\,\%\,;$$

$$S^2 = (46/45)\cdot 21,514 \approx 21,992;$$

$$S \approx 4,6896\ \ \tilde{\mu}_3 = (9,6739 - 3\cdot 5,6739\cdot 0,5435 + 2(0,5435)^3)\cdot 8 \approx 4,6648;$$

$$\tilde{\mu}_4 = (99,587 - 4\cdot 9,6739\cdot 0,5435 + 6\cdot 5,6739\cdot (0,5435)^2 - 3\cdot (0,5435)^4)\cdot 16 \approx$$

$$\approx 1268,7952\,;\ a_s = 4,6648/(4,6383)^3 \approx 0,0467\,;$$

$$e_k = 1268,7952/(4,6383)^4 - 3 \approx -0,2587\,.\ \bullet$$