Частина II

МАТЕМАТИЧНА СТАТИСТИКА

Розділ 7

СТАТИСТИЧНІ РОЗПОДІЛИ ВИБІРКИ

Статистика (від італійського слова *stato* – держава) вивчає кількісний бік суспільних явищ і процесів у нерозривному зв'язку з їх якісним змістом.

Статистику поділяють на описову та пояснювальну. Прикладом описової статистики є книга рекордів Гіннеса. Пояснювальна статистика формулює певні висновки про ті чи інші явища, робить прогноз.

Основним змістом математичної статистики є систематизація, обробка і використання статистичної інформації для виявлення статистичних закономірностей ознаки або ознак певної сукупності елементів.

Математична статистика – розділ математики, в якому на основі дослідних даних вивчаються імовірнісні закономірності масових явищ.

Основними завданнями математичної статистики є статистична перевірка гіпотез, оцінка розподілу статистичних ймовірностей та його параметрів, вивчення статистичної залежності, визначення основних числових характеристик випадкових вибірок, якими є: вибіркове середнє, вибіркові дисперсії, стандартне відхилення.

Математична статистика виникла (XVII ст.) та почала розвиватись одночасно з теорією ймовірностей. Подальшим розвитком (кінець XIX – початок XX ст.) математичної статистики займалися П. Чебишов, А. Марков, О. Ляпунов, а також К. Гаусс, Ф. Гальтон, К. Пірсон та інші. У XX ст. найбільший внесок у математичну статистику зробили В. Романовський, Е. Слуцький, А. Колмогоров, Стьюдент (псевдонім У. Госсета), Е. Пірсон, Ю. Нейман, А. Вальд, А. Скороход, В. Королюк та інші вчені.

Прикладом перевірки статистичних 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. Загальні методи математичної статистики є основою теорії похибок.

Предмет математичної статистики полягає в розробці методів збору та обробки статистичних даних для одержання наукових та практичних висновків.

7.1. Генеральна та вибіркова сукупність

Основні поняття математичної статистики:

– **Генеральна сукупність** – це вся сукупність об'єктів, які досліджуються.

– **Вибірка або вибіркова сукупність** – це об'єкти, довільно або випадково відібрані з генеральної сукупності для дослідження.

– **Обсяг (об'єм) сукупності** – це кількість об'єктів цієї сукупності.

Приклад 7.1. Плоди дерева (200 шт.) обстежують на наявність специфічного смаку цього сорту. Для цього відбирають 10 штук. Тут 200 – обсяг генеральної сукупності, 10 – обсяг вибірки.

Вибірка повинна бути репрезентативною (представницькою), тобто правильно відображати ті властивості генеральної сукупності, які вивчаються. Досліджувані явища мають бути масовими. Лише тоді статистичні дані будуть достовірними.

Вибірки бувають повторні та безповторні. *Повторною* називають вибірку, за якої відібраний об'єкт повертається до генеральної сукупності перед відбором іншого об'єкта. Вибірку називають *безповторною*, якщо взятий об'єкт до генеральної сукупності не повертається. Найчастіше використовують безповторні вибірки.

План проведення статистичних досліджень:

1) формулюють завдання дослідження та визначають обсяг вибірки (мета, об'єкти вивчення, їх кількість, певні ознаки, характеристики);

2) збирають потрібні дані, вибирають доцільну форму їх подання для подальшого дослідження (водночас застосовують такі методи, як спостереження, порівняння, усне та письмове анкетування);

3) проводять остаточну обробку статистичного матеріалу та його вивчення;

4) за результатами формулюють певні висновки.

Компактною та наочною формою подання даних, отриманих у результаті статистичного дослідження, є: таблиці, графіки, стовпчасті та кругові діаграми.

7.2. Способи відбору статистичного матеріалу

У практичній діяльності використовують різноманітні способи відбору об'єктів із генеральної сукупності.

1. *Простий випадковий відбір.*

Навмання вибирається певна кількість об'єктів з усієї генеральної сукупності. Наприклад, вибираємо m об'єктів із n об'єктів генеральної сукупності. Для цього нумеруємо картки від 1 до n на об'єктах і вибираємо по одній картці і т. д. Простий відбір може бути повторним або безповторним. Для великого обсягу вибірки використовують готові таблиці «випадкових чисел».

2. *Типовий відбір.*

Об'єкти вибирають не з усієї сукупності, а з певної її типової частини. Наприклад, соціологи вивчають різні групи людей. Типовими ознаками можуть бути: місце проживання, вік, професія…

3. *Механічний відбір.*

Генеральну сукупність поділяють на стільки груп, частин, скільки планується об'єктів у вибірку. З кожної частини вибирають один об'єкт. Наприклад, якщо обстежується 20% телевізорів з партії, то беруть кожен п'ятий, якщо 10% – кожен десятий. Щоб механічний відбір був репрезентативним, треба враховувати специфіку технологічного процесу.

4. *Серійний відбір.*

Об'єкти з генеральної сукупності вибирають не по одному, а серіями, партіями. Серійний відбір використовують тоді, коли ознака, яку досліджують, мало змінюється в різних серіях.

На практиці розглянуті способи відбору поєднують і комбінують довільним способом.

7.3. Статистичний розподіл вибірки

Об'єкти обстежують за їх певними характеристиками або ознаками (вік, професія, …), які називають **варіантами**.

Сукупність значень ознаки (статистичної змінної), записаних у порядку їх зростання, називається **варіаційним рядом**.

Нехай X – статистична змінна, x_i $(i = 1, 2, \ldots, n)$ – її значення, то тоді $x'_1, x'_2, \ldots, x'_n$ – варіаційний ряд, якщо $x'_{i+1} \geq x'_i$.

Якщо кількість варіант велика, то сукупність їх значень для зручності розбивають на інтервали. Кожен інтервал характеризується одним числом – його серединою. Величину інтервалу ще називають інтервальною різницею. Частинні інтервали варіант, які розміщені у зростаючій послідовності, утворюють **інтервальний варіаційний ряд**.

Важлива характеристика – частота появи цієї варіанти.

Додатне число n_i, яке показує, скільки разів варіанта x_i $(i = 1, 2, \dots, m)$ трапляється в таблиці даних, називається **частотою** варіанти x_i.

Ряд $n_1, n_2, \dots, n_m$ називається **рядом частот**. Сума усіх частот повинна дорівнювати об'єму вибірки:

$$\sum_{i=1}^{m} n_i = n. \tag{7.1}$$

Статистичний розподіл вибірки встановлює зв'язок між рядом варіант, що зростає або спадає, і відповідними частотами. Він може бути зображений у вигляді таблиці

x_i	x_1	x_2	…	x_m
n_i	n_1	n_2	…	n_m

Відношення частоти n_i варіанти x_i до об'єму вибірки n називають **відносною частотою** і позначають

$$W_i = \frac{n_i}{n}. \tag{7.2}$$

Сума усіх відносних частот

$$\sum_{i=1}^{m} W_i = 1. \tag{7.3}$$

Залежність між упорядкованим рядом варіант і відповідними їм відносними частотами називається **статистичним розподілом відносних частот вибірки**:

x_i	x_1	x_2	…	x_m
W_i	$\frac{n_1}{n}$	$\frac{n_2}{n}$	…	$\frac{n_m}{n}$

Приклад 7.2. Під час дослідження кількісної ознаки X із генеральної сукупності було отримано вибірку

4; 3; 6; 4; 7; 2; 5; 1; 2; 5; 4; 4; 3; 5; 6; 3; 4; 1; 3; 4.

Знайти обсяг вибірки, побудувати її варіаційний ряд, статистичні розподіли частот та відносних частот.

Розв'язання. Оскільки вибірка містить 20 значень, то об'єм вибірки $n = 20$.

Побудуємо варіаційний ряд вибірки, тобто запишемо всі її значення у порядку зростання:

1; 1; 2; 2; 3; 3; 3; 3; 4; 4; 4; 4; 4; 4; 5; 5; 5; 6; 6;7.

У цій вибірці лише сім різних значень, тобто варіант:

1; 2; 3; 4; 5; 6; 7.

Знайдемо їх частоти:

$$n_1 = 2;\ \ n_2 = 2;\ \ n_3 = 4;\ \ n_4 = 6;\ \ n_5 = 3;\ \ n_6 = 2;\ \ n_7 = 1.$$

Запишемо шуканий статистичний розподіл частот вибірки:

x_i	1	2	3	4	5	6	7
n_i	2	2	4	6	3	2	1

Знайдемо відносні частоти варіант вибірки:

$$W_1 = \frac{2}{20} = 0{,}1;\ W_2 = \frac{2}{20} = 0{,}1;\ W_3 = \frac{4}{20} = 0{,}2;\ W_4 = \frac{6}{20} = 0{,}3;$$

$$W_5 = \frac{3}{20} = 0{,}15;\ \ W_6 = \frac{2}{20} = 0{,}1;\ \ W_7 = \frac{1}{20} = 0{,}05.$$

Отже, шуканий розподіл відносних частот має такий вигляд:

x_i	1	2	3	4	5	6	7
W_i	0,1	0,1	0,2	0,3	0,15	0,1	0,05

7.4. Емпірична функція розподілу та її властивості

За даними статистичного розподілу вибірки будується емпірична функція розподілу.

Емпіричною функцією розподілу (або функцією розподілу вибірки) називають функцію $F^*(x)$, яка визначає для кожного значення x відносну частоту події $X < x$.

Математично це означення записується:

$$F^*(x) = \frac{n_x}{n}, \qquad (7.4)$$

де n_x – кількість варіант, які менші від x, n – об'єм вибірки.

Емпірична функція розподілу $F^*(x)$ має такі властивості:

1. $0 \le F^*(x) \le 1$;
2. $F^*(x)$– зростаюча функція;
3. $F^*(x) = \begin{cases} 0, & x \le x_1; \\ 1, & x > x_m, \end{cases}$

де x_1 – найменше значення варіанти, x_m – найбільше значення варіанти.

Приклад 7.3. Нехай маємо таку частотну таблицю (розподіл частот):

x_i	3	5	7	10	15
n_i	2	4	7	4	3

Записати емпіричну функцію розподілу та побудувати її графік.

Розв'язання. Об'єм вибірки:

$$n = \sum_{i=1}^{m} n_i = 2 + 4 + 7 + 4 + 3 = 20.$$

Оскільки найменше значення варіанти $x_1 = 3$, то

$$F^*(x) = 0$$

для всіх $x \le 3$.

Значення $x < 5$, а саме: $x_1 = 3$, спостерігається двічі, тому

$$F^*(x) = \frac{2}{20} = 0{,}1,$$

при $3 < x \le 5$.

Значення $x < 7$, а саме: $x_1 = 3$, $x_2 = 5$, спостерігалися $2 + 4 = 6$ разів, тому

$$F^*(x) = \frac{6}{20} = 0{,}3,$$

при $5 < x \le 7$.

Значення $x < 10$, а саме, $x_1 = 3$, $x_2 = 5$, $x_3 = 7$, спостерігалися $2 + 4 + 7 = 13$ разів, тому

$$F^*(x) = \frac{13}{20} = 0{,}65,$$

при $7 < x \leq 10$.

Значення $x < 15$, а саме, $x_1 = 3$, $x_2 = 5$, $x_3 = 7$, $x_4 = 10$, спостерігалися $2 + 4 + 7 + 4 = 17$ разів, тому

$$F^*(x) = \frac{17}{20} = 0{,}85,$$

при $10 < x \leq 15$.

Оскільки $x_5 = 15$ – найбільше значення варіанти, то

$$F^*(x) = 1$$

при $x > 15$.

Отже, запишемо шукану емпіричну функцію розподілу:

$$F^*(x) = \begin{cases} 0, \text{при } x \leq 3; \\ 0{,}1, \text{при } 3 < x \leq 5; \\ 0{,}3, \text{при } 5 < x \leq 7; \\ 0{,}65, \text{при } 7 < x \leq 10; \\ 0{,}85, \text{при } 10 < x \leq 15; \\ 1, \text{при } x > 15. \end{cases}$$

Графік цієї функції зображено на рисунку 7.1.

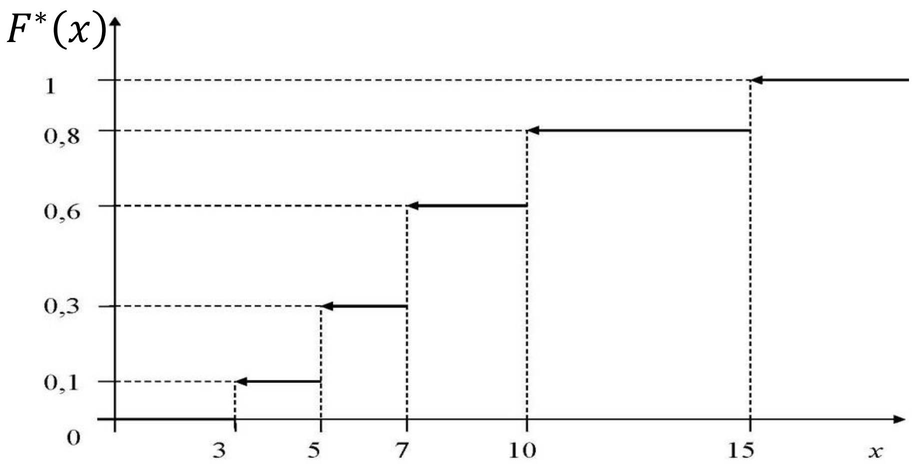

Рис. 7.1. Графік емпіричної функції розподілу

7.5. Згруповані розподіли вибірки

Часто для спрощення статистичного дослідження використовують згрупований розподіл вибірки.

Загальна схема побудови згрупованого розподілу частот

1. Визначити найбільше x_{max} та найменше x_{min} значення варіанти x_i і обчислити розмах варіант $R = x_{max} - x_{min}$.
2. Задати певне непарне число класів k. При загальному числі замірів $n \geq 100$ доцільно брати $9 \leq k \leq 15$, а при $n \leq 100$ можна вибрати $k = 7$.
3. Визначити ширину класу $h = \frac{R}{k}$.
4. Встановити границі класів і підрахувати кількість варіант у кожному класі. При підрахунку числа варіант значення x_i, що знаходиться на границі класів, слід відносити завжди до одного й того ж класу, а саме там, де це число трапилося вперше. Відтак воно стає нижньою границею класу.
5. Визначити частоту для кожного класу і записати ряд розподілу.

Згрупований розподіл накопиченої частоти та відносної частоти

Часто поряд з розподілом частот варіанти необхідно мати розподіл накопичених частот.

Розподіл накопиченої частоти одержують послідовним додаванням частот чергового інтервалу, починаючи з першого і закінчуючи останнім.

Розподіл накопиченої частоти (позначається F_i) дозволяє відповісти на питання: «Скільки існує варіант, які менші, ніж варіанта x_i ?».

Розподіл накопиченої відносної частоти одержують послідовним додаванням відносних частот чергового інтервалу, починаючи з першого і закінчуючи останнім.

Розподіл накопиченої відносної частоти (позначається $\frac{F_i}{n}$) дозволяє відповісти на питання: «Яка пропорція варіант, які менші, ніж x_i ?»

Згрупований розподіл щільності частот і щільності відносних частот

Якщо поділити всі частоти на ширину інтервалу h, то отримаємо **розподіл щільності частот вибірки** $\frac{n_i}{h}$.

Якщо поділити всі відносні частоти на ширину інтервалу h, то отримаємо **розподіл щільності відносних частот вибірки** $\frac{W_i}{h}$.

Приклад 7.4. У таблиці 1 наведена вибірка середньомісячної платні 100 співробітників фірми. Впорядкувати вибірку. Записати розподіл частот та відносних частот середньомісячної платні співробітників фірми; згрупований розподіл частот та відносних частот середньомісячної платні співробітників фірми; згрупований розподіл накопичених частот та накопичених відносних частот середньомісячної платні співробітників фірми; згрупований розподіл щільності частот та щільності відносних частот.

Таблиця 7.1

Вибірка середньомісячної платні 100 співробітників фірми									
338	348	304	314	326	314	324	304	342	308
336	304	302	338	314	304	320	321	322	321
312	323	336	324	312	312	364	356	362	302
322	310	334	292	362	381	304	366	298	304
381	368	304	298	368	290	340	328	316	322
302	314	292	342	321	322	290	332	298	296
296	298	324	338	352	326	318	304	332	322
360	312	331	331	304	316	332	282	342	338
342	322	324	325	302	328	354	330	316	324
334	350	334	324	332	340	324	314	326	323

Розв'язання. Запишемо розподіл частот та відносних частот цієї вибірки:

Таблиця 7.2

Розподіл частот та відносних частот середньомісячної платні співробітників фірми											
x_i	n_i	W_i	x_i	n_i	W_i	x_i	n_i	W_i	x_i	n_i	W_i
282	1	0,01	314	5	0,05	328	2	0,02	350	1	0,01
290	2	0,02	316	3	0,03	330	1	0,01	352	1	0,01
292	2	0,02	318	1	0,01	331	2	0,02	354	1	0,01
296	2	0,02	320	1	0,01	332	4	0,04	356	1	0,01
298	4	0,04	321	3	0,03	334	3	0,03	360	1	0,01
302	4	0,04	322	6	0,06	336	2	0,02	362	2	0,02
304	9	0,09	323	2	0,02	338	4	0,04	364	1	0,01
308	1	0,01	324	7	0,07	340	2	0,02	366	1	0,01
310	1	0,01	325	1	0,01	342	4	0,04	368	2	0,02
312	4	0,04	326	3	0,03	348	1	0,01	381	2	0,02

Наступним кроком в обробці даних, що веде до суттєвого спрощення досліджень, є групування.

Знаходимо максимальне та мінімальне значення варіанти вибірки:

$$x_{max} = 381, x_{min} = 282.$$

Розмах

$$R = x_{max} - x_{min} = 381 - 282 = 99.$$

Задамо число класів k =11.

Визначимо ширину класу $h = \frac{R}{k} = \frac{99}{11} = 9.$

Просумуємо частоти для кожного класу інтервалів, значення x_i, що знаходиться на границі класів, заносимо до того класу, де це число трапилось вперше. Результат запишемо у вигляді таблиці.

Таблиця 7.3

Згрупований розподіл частот середньомісячної платні співробітників фірми		
Інтервали платні	Частоти n_i	
282 – 291	3	=(1+2)
291 – 300	8	=(2+2+4)
300 – 309	14	=(4+9+1)
309 – 318	14	=(1+4+5+3+1)
318 – 327	23	=(1+3+6+2+7+1+3)
327 – 336	14	=(2+1+2+4+3+2)
336 – 345	10	=(4+2+4)
345 – 354	4	=(1+1+1+1)
354 – 363	4	=(1+1+2)
363 – 272	4	=(1+1+2)
372 – 381	2	=(2)
Разом:	100	

Запишемо розподіл накопичених частот, який одержимо послідовним додаванням частот чергового інтервалу.

Таблиця 7.4

Згрупований розподіл накопичених частот середньомісячної платні співробітників фірми		
Платня	Накопичені частоти F_i	
менше ніж 291	3	(=3)
менше ніж 300	11	(=3+8)
менше ніж 309	25	(=11+14)
менше ніж 318	39	(=25+14)
менше ніж 327	62	(=39+23)
менше ніж 336	76	(=62+14)
менше ніж 345	86	(=76+10)
менше ніж 354	90	(=86+4)
менше ніж 363	94	(=90+4)
менше ніж 372	98	(=94+4)
менше ніж 381	100	(=98+2)

Якщо поділити частоти та накопичені частоти на об'єм вибірки $n = 100$, то отримаємо, відповідно, розподіл відносних частот W_i та накопичених відносних частот $\frac{F_i}{n}$.

Запишемо розподіл відносних частот W_i та накопичених відносних частот $\frac{F_i}{n}$ (таблиця 7.5).

Таблиця 7.5

Згрупований розподіл відносних та накопичених відносних частот середньомісячної платні співробітників фірми			
Інтервали платні	Відносні частоти W_i	Платня	Накопичені відносні частоти $\frac{F_i}{n}$
282 – 291	0,03	менше ніж 291	0,03
291 – 300	0,08	менше ніж 300	0,11
300 – 309	0,14	менше ніж 309	0,25
309 – 318	0,14	менше ніж 318	0,39
318 – 327	0,23	менше ніж 327	0,62
327 – 336	0,14	менше ніж 336	0,76
336 – 345	0,1	менше ніж 345	0,86
345 – 354	0,04	менше ніж 354	0,9
354 – 363	0,04	менше ніж 363	0,94
363 – 272	0,04	менше ніж 372	0,98
372 – 381	0,02	менше ніж 381	1

Запишемо згрупований розподіл щільності частот та щільності відносних частот.

Для цього поділимо частоти та відносні частоти на ширину інтервалу $h = 10$. Щоб підсумувати результати, які одержали в *Прикладі 7.4*, зведемо разом в одну таблицю розглянуті розподіли *Таблиця 7.6*.

Ця таблиця містить усі важливі статистичні розподіли вибірки.

Таблиця 7.6

Статистичні розподіли вибірки							
x_i	n_i	W_i	$\frac{n_i}{h}$	$\frac{W_i}{h}$	x_i	F_i	$\frac{F_i}{n}$
282 – 291	3	0,03	$\frac{1}{3}$	$\frac{1}{300}$	менше, ніж 291	3	0,03
291 – 300	8	0,08	$\frac{8}{9}$	$\frac{2}{225}$	менше, ніж 300	11	0,11
300 – 309	14	0,14	$\frac{14}{9}$	$\frac{7}{450}$	менше, ніж 309	25	0,25
309 – 318	14	0,14	$\frac{14}{9}$	$\frac{7}{450}$	менше, ніж 318	39	0,39
318 – 327	23	0,23	$\frac{23}{9}$	$\frac{23}{900}$	менше, ніж 327	62	0,62
327 – 336	14	0,14	$\frac{14}{9}$	$\frac{7}{450}$	менше, ніж 336	76	0,76
336 – 345	10	0,1	$\frac{10}{9}$	$\frac{1}{90}$	менше, ніж 345	86	0,86
345 – 354	4	0,04	$\frac{4}{9}$	$\frac{1}{225}$	менше, ніж 354	90	0,9
354 – 363	4	0,04	$\frac{4}{9}$	$\frac{1}{225}$	менше, ніж 363	94	0,94
363 – 272	4	0,04	$\frac{4}{9}$	$\frac{1}{225}$	менше, ніж 372	98	0,98
372 – 381	2	0,02	$\frac{2}{9}$	$\frac{1}{450}$	менше, ніж 381	100	1

7.6. Полігон частот та відносних частот

Статистичні розподіли вибірки можуть бути представлені також графічно. Завдяки цьому ми можемо побачити характерні зміни ряду розподілу, не користуючись аналізом цифрових даних.

Якщо в результаті вибірки ми одержали статистичний розподіл ознаки X, яку треба дослідити, то будемо мати перелік варіант ознаки

$$x_1, x_2, x_3, \dots, x_m$$

та відповідних їм частот

$$n_1, n_2, n_3, \dots, n_m$$

або відносних частот

$$W_1, W_2, W_3, \dots, W_m.$$

Значення варіант та частот або відносних частот можна розглядати як координати точок

$$K_1(x_1;\ n_1),\ K_2(x_2;\ n_2),\ K_3(x_3;\ n_3),\ \dots,\ K_m(x_m;\ n_m)$$

або

$$L_1(x_1;\ W_1),\ L(x_2;\ W_2),\ L_3(x_3;\ W_3),\ \dots,\ L_m(x_m;\ W_m).$$

Полігоном частот називають ламану, відрізки якої з'єднують точки з координатами

$$K_1(x_1;\ n_1),\ K_2(x_2;\ n_2),\ K_3(x_3;\ n_3),\ \dots,\ K_m(x_m;\ n_m).$$

Полігоном відносних частот називають ламану, відрізки якої з'єднують точки з координатами

$$L_1(x_1;\ W_1),\ L(x_2;\ W_2),\ L_3(x_3;\ W_3),\ \dots,\ L_m(x_m;\ W_m).$$

Для побудови полігону частот на осі абсцис відкладають варіанти x_i ознаки X, а на осі ординат – відповідні їм частоти n_i. Точки $K_i(x_i;\ n_i)$ $(i = 1, 2, 3,\ \dots,\ m)$ з'єднують відрізками прямих і одержують полігон частот.

Для побудови полігону відносних частот на осі абсцис відкладають варіанти x_i ознаки X, а на осі ординат – відповідні їм відносні частоти W_i $(i = 1, 2, 3, \dots, m)$, а потім точки $L_i(x_i;\ W_i)$ з'єднують відрізками прямих.

При неперервному розподілі ознаки X у разі великої кількості спостережень увесь інтервал, в якому розміщені спостережні значення ознаки, як правило, розбивають на кілька частинних інтервалів однакової довжини h і знаходять

n_i – суму частот варіант, що потрапили в i-й інтервал. Для вибору оптимальної величини інтервалу рекомендовано використовувати формулу:

$$h \approx \frac{x_{max}-x_{min}}{1+3{,}2\cdot\lg n},$$

де x_{max}, x_{min} – відповідно найбільше і найменше значення вибірки, n – об'єм вибірки.

Якщо задано інтервальний статистичний розподіл вибірки, то для побудови полігону частот за даними вибірки з'єднують точки, абсциси яких є значення середин частинних інтервалів, а ординатами – відповідні їм значення частот.

Приклад 7.5. Вибірку задано інтервальним розподілом частот:

$(x_i\,;x_{i+1})$	(1; 3)	(3; 5)	(5; 7)	(7; 9)	(9; 11)
n_i	13	9	5	16	7

Побудувати полігон відносних частот.

Розв'язання. Знайдемо об'єм вибірки:

$$n = 13 + 9 + 5 + 16 + 7 + 7 = 50.$$

Для побудови інтервального статистичного розподілу відносних частот визначимо відносні частоти:

$$W_1 = \frac{n_1}{n} = \frac{13}{50} = 0{,}26; \qquad W_2 = \frac{n_2}{n} = \frac{9}{50} = 0{,}18;$$

$$W_3 = \frac{n_3}{n} = \frac{5}{50} = 0{,}1; \qquad W_4 = \frac{n_4}{n} = \frac{16}{50} = 0{,}32;$$

$$W_5 = \frac{n_5}{n} = \frac{7}{50} = 0{,}14.$$

Побудуємо інтервальний статистичний розподіл відносних частот:

$(x_i\,;x_{i+1})$	(1; 3)	(3; 5)	(5; 7)	(7; 9)	(9; 11)
W_i	0,26	0,18	0,1	0,32	0,14

Знайдемо середини частинних інтервалів:

$$x_1^{\text{сер}} = \frac{x_1+x_2}{2} = \frac{1+3}{2} = 2; \qquad x_2^{\text{сер}} = \frac{x_2+x_3}{2} = \frac{3+5}{2} = 4;$$

$$x_3^{\text{сер}} = \frac{x_3+x_4}{2} = \frac{5+7}{2} = 6; \qquad x_4^{\text{сер}} = \frac{x_4+x_5}{2} = \frac{7+9}{2} = 8;$$

$$x_5^{\text{сер}} = \frac{x_5+x_6}{2} = \frac{9+11}{2} = 10.$$

Відкладемо на осі абсцис значення середин частинних інтервалів $x_i^{\text{сер}}$, а на осі ординат – значення відповідних їм відносних частот W_i. Послідовно з'єднуючи між собою точки ($x_i^{\text{сер}}$; W_i) відрізками, отримаємо полігон відносних частот (рис. 7.2).

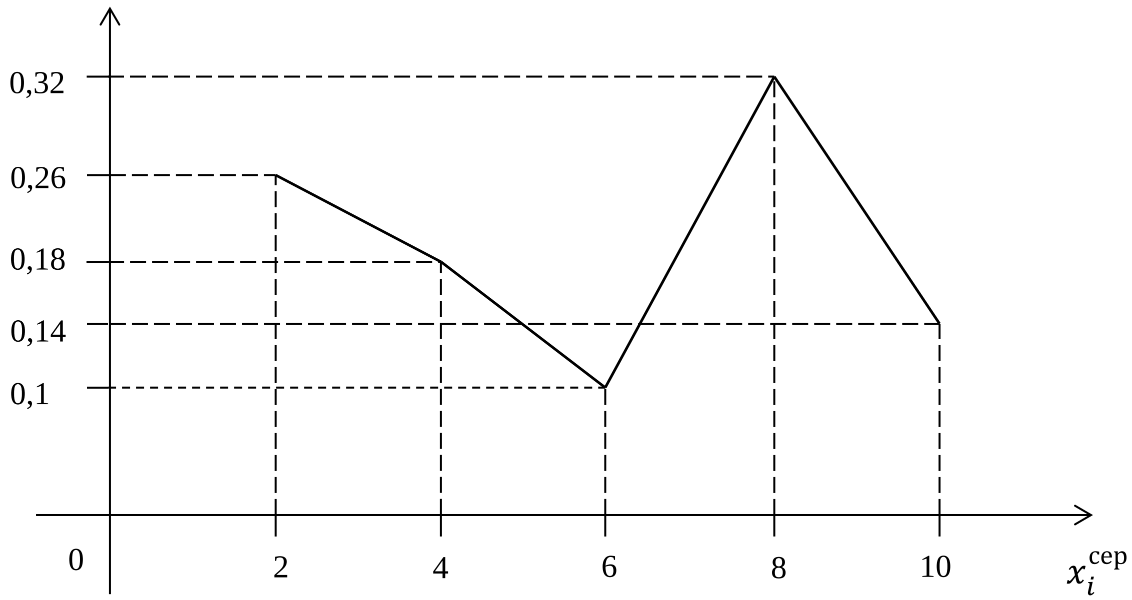

Рис. 7.2. Графік побудови полігону відносних частот

7.7. Гістограма частот та відносних частот

Гістограмою частот називають ступінчасту фігуру, яка складається з прямокутників, основами яких є частинні інтервали варіант довжиною $h = x_i - x_{i-1}$, а висоти дорівнюють $\frac{n_i}{h}$ (щільність частоти).

Гістограмою відносних частот називають ступінчасту фігуру, яка складається з прямокутників, основами яких є частинні інтервали варіант довжиною $h = x_i - x_{i-1}$, а висоти дорівнюють $\frac{W_i}{h}$ (щільність відносної частоти).

Площа гістограми частот дорівнює об'єму вибірки, а площа гістограми відносних частот – одиниці.

Для побудови гістограми частот (відносних частот) проміжок $[x_{min}; x_{max}]$, тобто від найменшого значення x_{min}, що спостерігали, до найбільшого значення x_{max}, розбивають на декілька відрізків рівної довжини h. Потім підраховують суму частот (відносних частот) тих значень варіант, які належать кожному із одержаних відрізків.

Якщо в i-му відрізку ($i = 1, 2, 3, \ldots, m$) кількість варіант, що спостерігали, з врахуванням їх частот дорівнює n_i, то будують прямокутник Π_i, основою якого буде i-ий відрізок довжиною h, а висотою буде відрізок довжиною $\frac{n_i}{h}$ (для відносних частот висота – $\frac{W_i}{h}$). Площа такого прямокутника дорівнює $h \cdot \frac{n_i}{h} = n_i$ (у випадку відносних частот площа дорівнює $h \cdot \frac{W_i}{h} = W_i$). Тому площа усіх прямокутників буде дорівнювати сумі n_i (W_i), тобто $S = \sum_{i=1}^{m} n_i = n$ об'єму вибірки.

У випадку гістограми відносних частот площа прямокутників буде дорівнювати сумі усіх відносних частот

$$S = \sum_{i=1}^{m} W_i = 1.$$

Приклад 7.6. Вибірку задано розподілом частот:

x_i	2	4	6	8	10	12	14
n_i	2	3	5	1	4	2	3

Побудувати гістограму відносних частот.

Розв'язання. Щоб побудувати гістограму відносних частот на основі цього статистичного розподілу частот, складемо інтервальний статистичний розподіл відносних частот і знайдемо їх щільність.

Обсяг вибірки:

$$n = 2 + 3 + 5 + 1 + 4 + 2 + 3 = 20.$$

Відносні частоти:

$$W_1 = \frac{n_1}{n} = \frac{2}{20} = 0{,}1; \qquad W_2 = \frac{n_2}{n} = \frac{3}{20} = 0{,}15;$$

$$W_3 = \frac{n_3}{n} = \frac{5}{20} = 0{,}25; \qquad W_4 = \frac{n_4}{n} = \frac{1}{20} = 0{,}05;$$

$$W_5 = \frac{n_5}{n} = \frac{34}{20} = 0{,}2; \qquad W_6 = \frac{n_6}{n} = \frac{2}{20} = 0{,}1;$$

$$W_7 = \frac{n_7}{n} = \frac{1}{20} = 0{,}05.$$

Знайдемо частинні інтервали, їх довжину і щільність відносних частот. Частинні інтервали визначимо з умови: задані в дискретному статистичному розподілі варіанти мають бути серединами частинних інтервалів.

Отже, довжина частинних інтервалів

$$h = x_i - x_{i-1} = 2.$$

Шуканий інтервальний статистичний розподіл відносних частот має вигляд:

$(x_i; x_{i+1})$	(1; 3)	(3; 5)	(5; 7)	(7; 9)	(9; 11)	(11; 13)	(13; 15)
W_i	0,1	0,15	0,25	0,05	0,2	0,1	0,15

Щільності відносних частот:

$$\frac{W_1}{h} = \frac{0{,}1}{2} = 0{,}05; \qquad \frac{W_2}{h} = \frac{0{,}15}{2} = 0{,}075;$$

$$\frac{W_3}{h} = \frac{0{,}25}{2} = 0{,}125; \qquad \frac{W_4}{h} = \frac{0{,}05}{2} = 0{,}025;$$

$$\frac{W_5}{h} = \frac{0{,}2}{2} = 0{,}1; \qquad \frac{W_6}{h} = \frac{0{,}1}{2} = 0{,}05;$$

$$\frac{W_7}{h} = \frac{0{,}15}{2} = 0{,}075.$$

Побудуємо гістограму відносних частот (рис. 7.3).

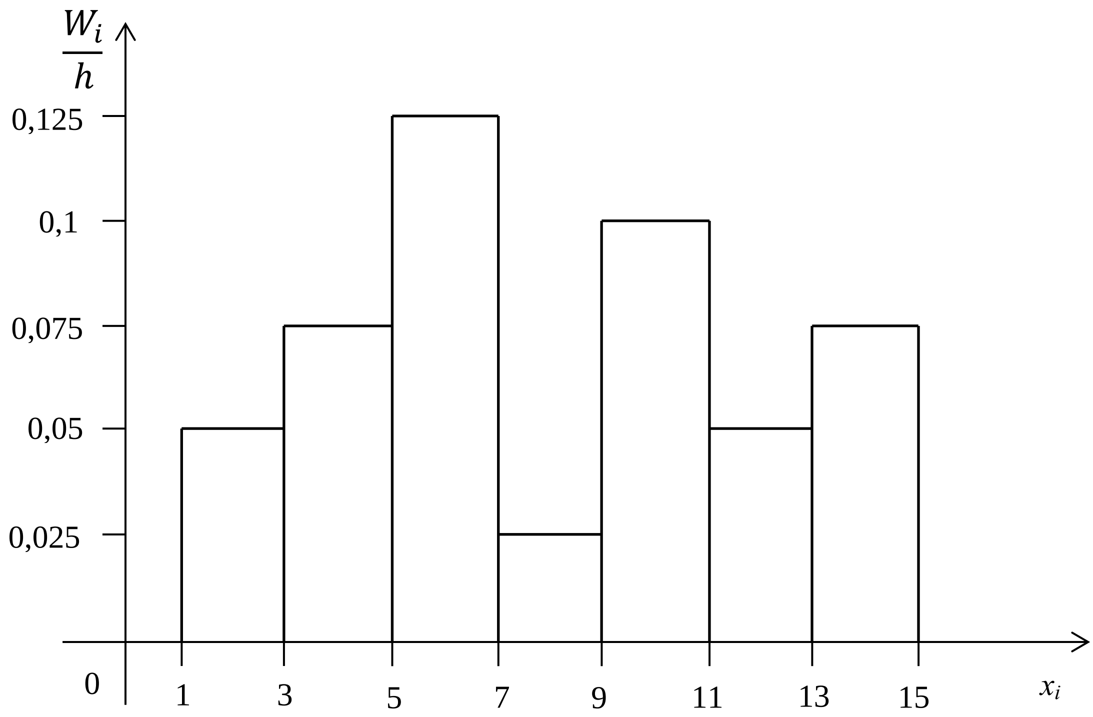

Рис. 7.3. Гістограма відносних частот

Запитання для самоконтролю

1. Що є предметом математичної статистики?
2. Вказати основні завдання математичної статистики.
3. Що називають генеральною та вибірковою сукупністю, об'ємом цих сукупностей?
4. Навести приклади генеральної сукупності та вибірки.
5. Способи відбору статистичного матеріалу (проілюструвати на прикладах).
6. Що називається варіантою, варіаційним рядом?
7. Що таке частота, відносна частота варіант?
8. Що називається емпіричною функцією розподілу?
9. Властивості емпіричної функція розподілу та її графік.
10. Статистичний розподіл частот вибірки.
11. Статистичний розподіл відносних частот вибірки.
12. Згрупований розподіл вибірки, загальна схема його побудови.
13. Що називають згрупованим розподілом накопичених частот вибірки; накопичених відносних частот вибірки?
14. Що називають згрупованим розподілом щільності частот; щільності відносних частот?