МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
ХЕРСОНСЬКИЙ ДЕРЖАВНИЙ АГРАРНИЙ УНІВЕРСИТЕТ

Кафедра прикладної математики та економічної кібернетики

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

до виконання лабораторно-практичних робіт
з дисципліни: «Біометрія»
для студентів 2 курсу факультету ФРГП
спеціальностей «лісове господарство», «садово –паркове господарство»

Змістовий модуль №1
Групування та статистична обробка науково-дослыдних даних

Херсон – 2017

УДК 681.3.04

Розглянуто і схвалено на засіданні кафедри прикладної математики та економічної кібернетики ХДАУ (протокол №___ від «____»_________________2017р.)

Рецензент: Тимофєєв К.В. – к.т.н., доцент кафедри технічної кібернетики Херсонського НТУ.

Лобода О.М. Методичні рекомендації до виконання лабораторно-практичних робіт з дисципліни: «Біометрія» для студентів 2 курсу факультету ФРГП спеціальностей «лісове господарство», «садово –паркове господарство». Змістовий модуль №1 -44с.

РОЗДІЛ 1. ОСНОВНІ ПОНЯТТЯ МАТЕМАТИЧНОЇ СТАТИСТИКИ

1.1. 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 зміняться безперервно у межах певного інтервалу.

Число об'єктів (спостережень) генеральної сукупності називається її **об'ємом** і позначається $\boldsymbol{N}$.

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 $\boldsymbol{n}$.

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.

1.2. 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:

$$w_i = \frac{n_i}{n}. \tag{1.1}$$

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. 1.1):

Таблиця 1.1

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

(відносні частоти w_i)				

де k – кількість варіант.

Інтервальний статистичний ряд у загальному вигляді можна представити таблицею (табл. 1.2):

Таблиця 1.2

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

$$k = 1 + 1,4\ln n. \qquad (1.2)$$

Довжина кожного із інтервалів Δ розраховується за формулою

$$\Delta = \frac{x_{\max} - x_{\min}}{k}, \qquad (1.3)$$

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, $c_i = \frac{a_i + a_{i+1}}{2}$.

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

$$F_n(x) = \frac{1}{n}\sum_{x_i<x} n_i = \sum_{x_i<x} w_i\ . \qquad (1.4)$$

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

$$f_n(x)=\begin{cases}\dfrac{n_i}{n\Delta}=\dfrac{w_i}{\Delta}, \text{ якщо } a_i \le x \le a_{i+1},\ i=\overline{1,k}\\ 0, \quad \text{якщо } x<a_1 \text{ або } x>a_{k+1},\ i=\overline{1,k}\end{cases}. \qquad (1.5)$$

ПРИКЛАД 1.1. У результаті тестування службовців деякої компанії були отримані такі результати (у баллах): 39, 41, 40, 42, 41, 40, 42, 44, 40, 43, 38, 42, 41, 43, 39, 37, 43, 41, 38, 42, 40, 41, 42, 40, 41. Побудувати дискретний статистичний ряд для випадкової величини X – оцінки службовців, полігон частот, емпіричну функцію розподілу та її графік.

Розв'язок. Для побудови дискретного статистичного ряду записуємо у порядку зростання різні значення випадкової величини X і відповідні частоти (табл. 1.3). Останній стовпець таблиці використовується для перевірки правильності побудови статистичного ряду (усього у тестуванні приймали участь 25 осіб, тому сума частот повинна дорівнювати 25).

Таблиця 1.3

x_i	37	38	39	40	41	42	43	44	Сума
n_i	1	2	2	4	6	5	4	1	$\sum_{i=1}^{k} n_i$ =25

Полігон частот даного розподілу зображено на рис. 1.1.

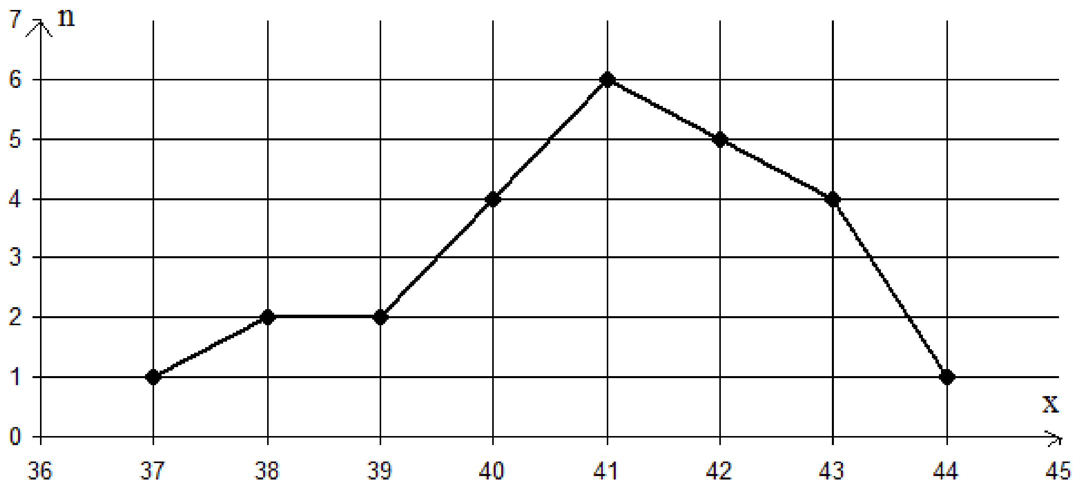

Рис. 1.1. Полігон частот

Для побудови емпіричної функції розподілу доповнимо таблицю двома рядками (табл. 1.4). В першому рядку обчислимо суму частот варіант, що менше x_i (тобто $\sum\limits_{x_i<x} n_i$).

Отримаємо:

якщо $x < x_1 = 37$, то $\sum\limits_{x_i<37} n_i = 0$, оскільки таких значень X немає;

якщо $x < x_2 = 38$, то $\sum\limits_{x_i<38} n_i = n_1 = 1$;

якщо $x < x_3 = 39$, то $\sum\limits_{x_i<39} n_i = n_1 + n_2 = 1 + 2 = 3$;

якщо $x < x_4 = 40$, то $\sum\limits_{x_i<40} n_i = n_1 + n_2 + n_3 = 1 + 2 + 2 = 5$;

якщо $x < x_5 = 41$, то $\sum\limits_{x_i<41} n_i = n_1 + n_2 + n_3 + n_4 = 1 + 2 + 2 + 4 = 9$;

якщо $x < x_6 = 42$, то $\sum\limits_{x_i<42} n_i = n_1 + n_2 + n_3 + n_4 + n_5 = 1 + 2 + 2 + 4 + 6 = 15$;

якщо $x < x_7 = 43$, то $\sum\limits_{x_i<43} n_i = n_1 + n_2 + n_3 + n_4 + n_5 + n_6 = 1 + 2 + 2 + 4 + 6 + 5 = 20$;

якщо $x < x_8 = 44$, то

$$\sum\limits_{x_i<44} n_i = n_1 + n_2 + n_3 + n_4 + n_5 + n_6 + n_7 = 1 + 2 + 2 + 4 + 6 + 5 + 4 = 24;$$

якщо $x > x_8 = 44$, то $\sum\limits_{x_i<x} n_i = 25$ – це позначає, що всі значення X менше числа, більшого за 44.

В другому рядку запишемо значення функції, обчислені за формулою (1.4). Графік отриманої емпіричної функції розподілу зображено на рис. 1.2.

Таблиця 1.4

x_i	37	38	39	40	41	42	43	44	
n_i	1	2	2	4	6	5	4	1	
$\sum\limits_{x_i<x} n_i$	0	1	3	5	9	15	20	24	25
F_i	0	0,04	0,12	0,2	0,36	0,6	0,8	0,96	1

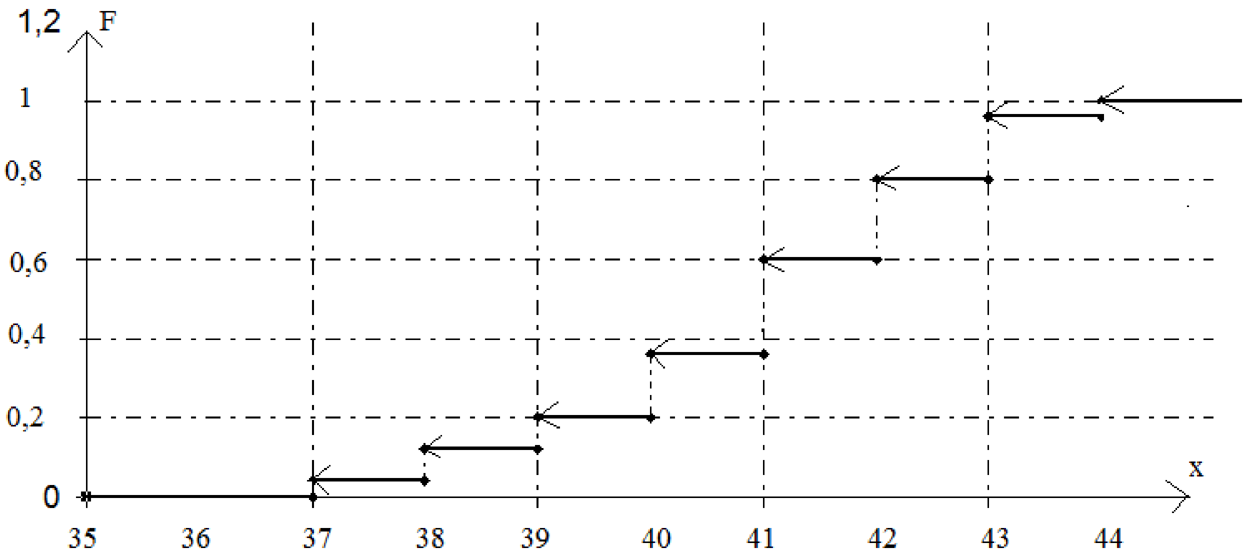

Рис. 1.2. Графік емпіричної функції розподілу

ПРИКЛАД 1.2. За даними вибіркового дослідження було отримано розподіл родин за доходом на одного їх члена в умовних одиницях (табл. 1.5). Побудувати інтервальний статистичний ряд, полігон частот, гістограму, полігон відносних частот, емпіричні функцію і щільність розподілу та їх графіки.

Таблиця 1.5

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'язок. Таблиця 1.5 містить 50 даних, тобто $n=50$. Для побудови інтервального статистичного ряду знаходимо: кількість інтервалів за формулою (1.2): $k=1+1,4\ln 50\approx 6,477\approx 7$; $x_{\max}=6,11$, $x_{\min}=38,84$; довжина кожного інтервалу за формулою (1.3): $\Delta=\dfrac{x_{\max}-x_{\min}}{k}=\dfrac{38,84-6,11}{7}\approx 4,68$. 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. 1.6).

Таблиця 1.6

$[a_i;a_{i+1})$	[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 1.6 будуємо гістограму (рис. 1.3).

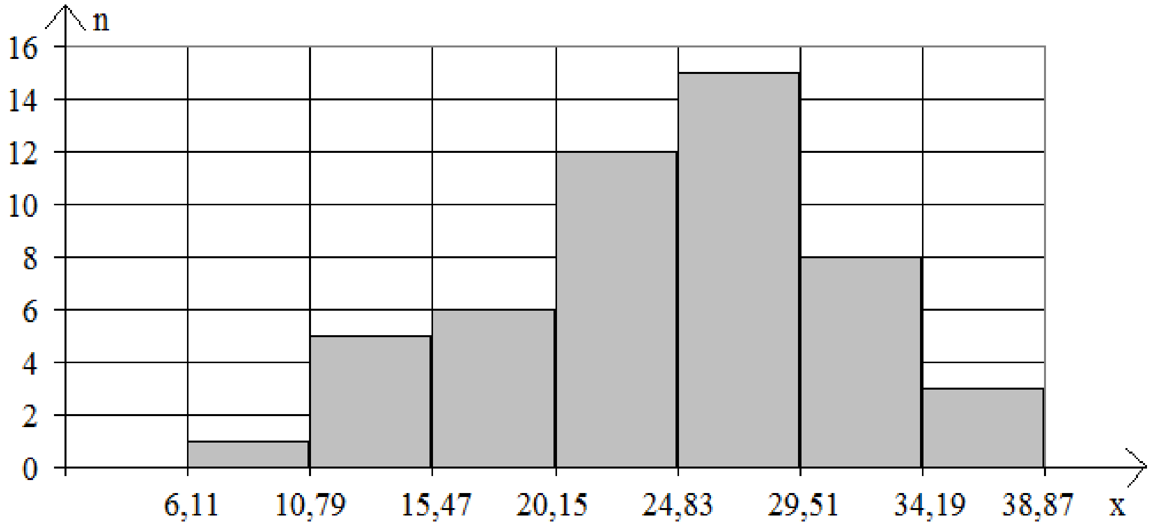

Рис. 1.3. Гістограма

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

Розрахуємо відносні частоти за формулою (1.1): $w_1 = \dfrac{n_1}{n} = \dfrac{1}{50} = 0{,}02$; $w_2 = 0{,}1$; $w_3 = 0{,}12$; $w_4 = 0{,}24$; $w_5 = 0{,}3$; $w_6 = 0{,}16$; $w_7 = 0{,}06$.

Результати оформимо у вигляді таблиці (табл. 1.7), останній стовпець якої будемо використовувати для перевірки правильності розрахунків.

Таблиця 1.7

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 1.7 будуємо полігон частот (рис. 1.4) і полігон відносних частот (рис. 1.5).

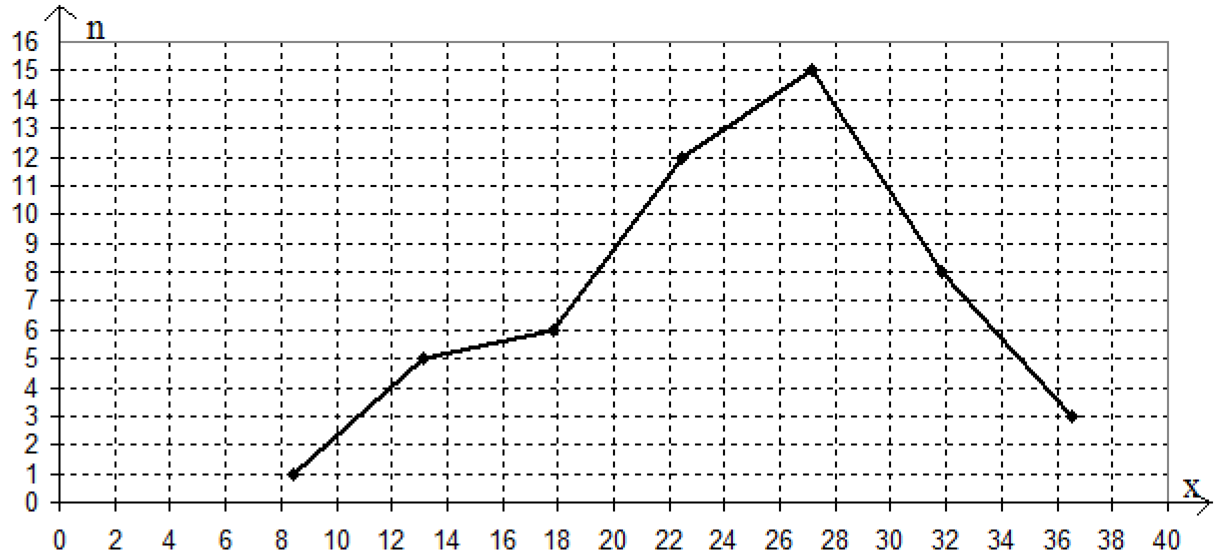

Рис. 1.4. Полігон частот

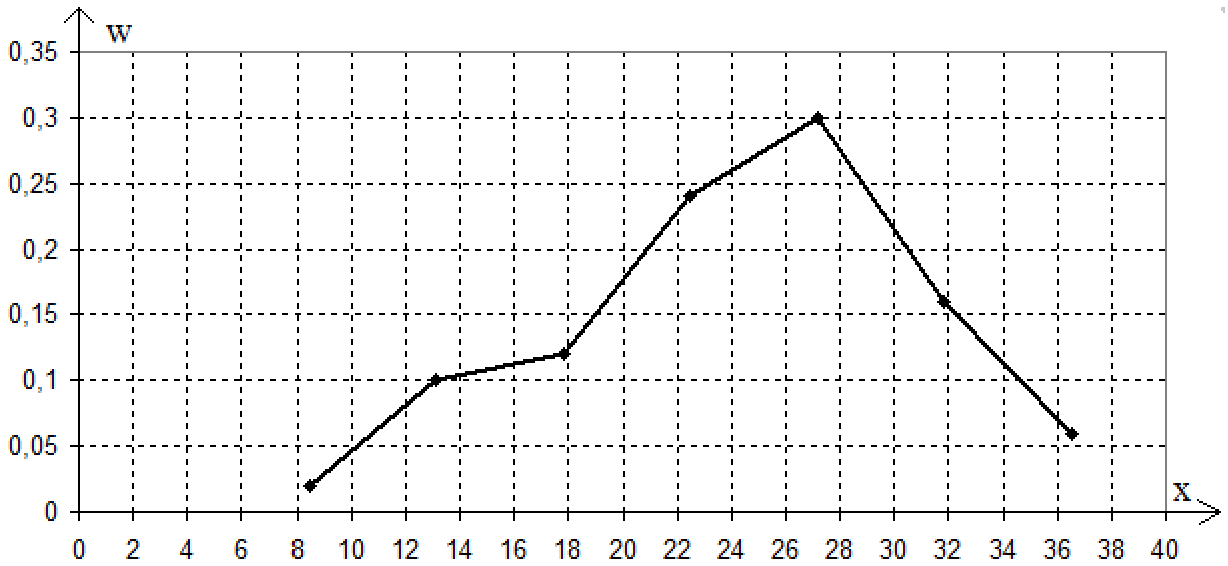

Рис. 1.5. Полігон відносних частот

Для побудови емпіричної функції розподілу обчислимо за формулою (1.4) 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 (1.5). Результати обчислень надано у таблиці 1.8.

Таблиця 1.8

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	$\sum\limits_{i=1}^{k} w_i = 1$
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 1.8 будуємо графік емпіричної функції розподілу (рис. 1.6) та емпіричної щільності розподілу (рис. 1.7).

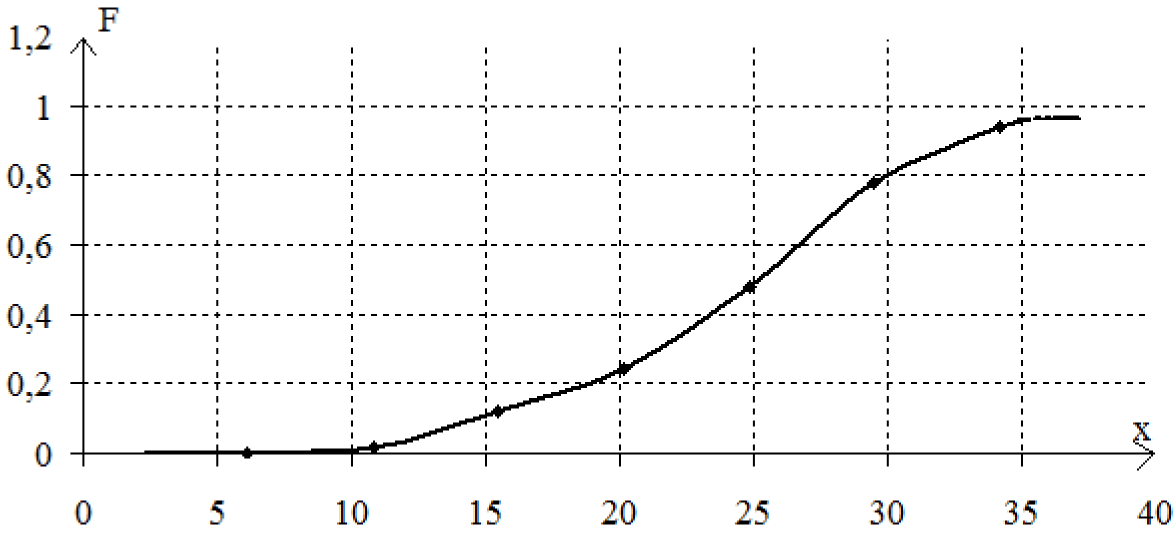

Рис. 1.6. Графік емпіричної функції розподілу

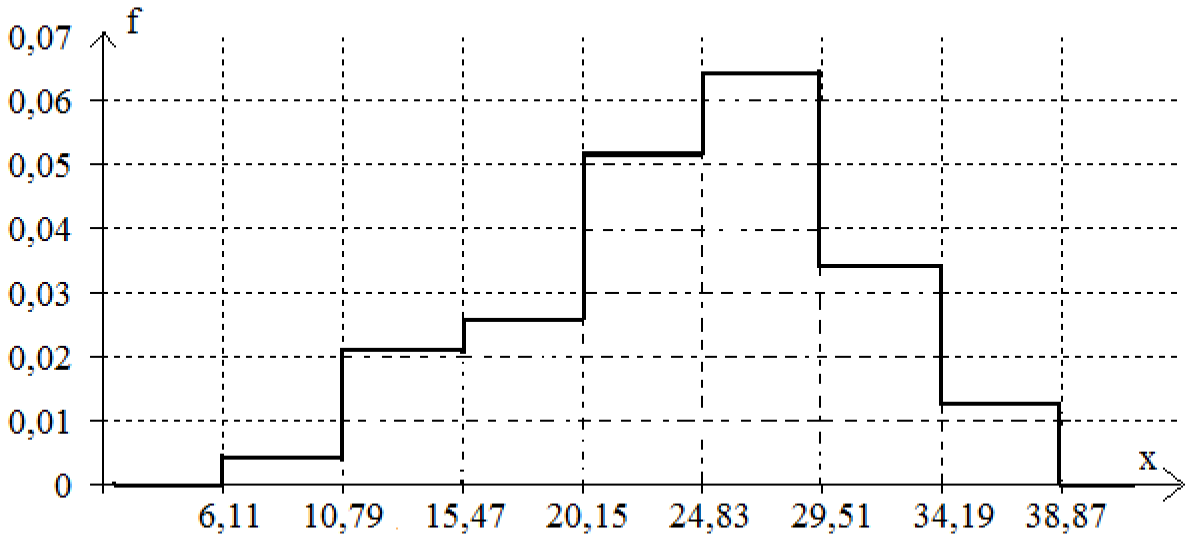

Рис. 1.7. Графік емпіричної щільності розподілу

1.3. Числові характеристики статистичних рядів

1.3.1. 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.

1.3.2. 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:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \sum_{i=1}^{k} x_i w_i . \qquad (1.6)$$

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

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

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1.7)$$

Зауваження. 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:

$$\overline{x_G} = \sqrt[n]{\prod_{i=1}^{n} x_i} . \qquad (1.9)$$

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:

$$Mo = x_{Mo} + \frac{\Delta\left(n_{Mo} - n_{Mo-1}\right)}{2n_{Mo} - n_{Mo-1} - n_{Mo+1}}, \qquad (1.10)$$

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.

Зауваження. 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:

$$Me = X_{Me} + \frac{\Delta\left(\frac{n}{2} - n_x^{\max}\right)}{n_m}, \qquad (1.11)$$

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.

Зауваження. 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.

1.3.3. Числові характеристики розсіювання

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

$$R = x_{\max} - x_{\min}. \qquad (1.12)$$

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

$$S^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i = \sum_{i=1}^{k}(x_i - \bar{x})^2 w_i \qquad (1.13)$$

$$\text{або } S^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i. \qquad (1.14)$$

Дисперсія є показником розсіювання елементів вибірки відносно їх середнього значення. Вибіркова дисперсія, отримана за формулою (1.14), називається незсуненою оцінкою дисперсії генеральної сукупності.

Різниця дисперсій, отриманих за формулами (1.13) та (1.14) зазвичай невелика, однак може вплинути на точність оцінок. Тому, якщо відомо точне значення математичного сподівання, використовують формулу (1.13), в іншому випадку – формулу (1.14).

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

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (1.15)$$

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

$$S = \sqrt{S^2}\,. \qquad (1.16)$$

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:

$$v = \frac{S}{|\bar{x}|}\cdot 100\%\ (\bar{x} \neq 0)\,. \qquad (1.17)$$

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

ПРИКЛАД 1.3. За даними вибіркового дослідження відомі ціни x_i певного товару у різних торгівельних організаціях (табл. 1.9). Знайти всі можливі числові характеристики за даними таблиці.

Таблиця 1.9

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

Розв'язок. За незгрупованими даними таблиці 1.9 можна знайти: вибіркове середнє за формулою (1.7), медіану, розмах варіації за формулою (1.12), дисперсію за формулою (1.15), вибіркове середнє квадратичне відхилення за формулою (1.16), коефіцієнт варіації за формулою (1.17). 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 1.10.

Таблиця 1.10

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

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

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

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

ПРИКЛАД 1.4. За даними вибіркового дослідження відомі ціни x_i певного товару у різних торгівельних організаціях (табл. 1.11). Визначити центральну тенденцію за даними таблиці.

Таблиця 1.11

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

Розв’язок. Оскільки дані таблиці 1.11 незгруповані, то оцінкою центральної тенденції може слугувати вибіркове середнє або медіана.

Знайдемо вибіркове середнє за формулою (1.7):

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

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. 1.8).

На рис. 1.8 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.

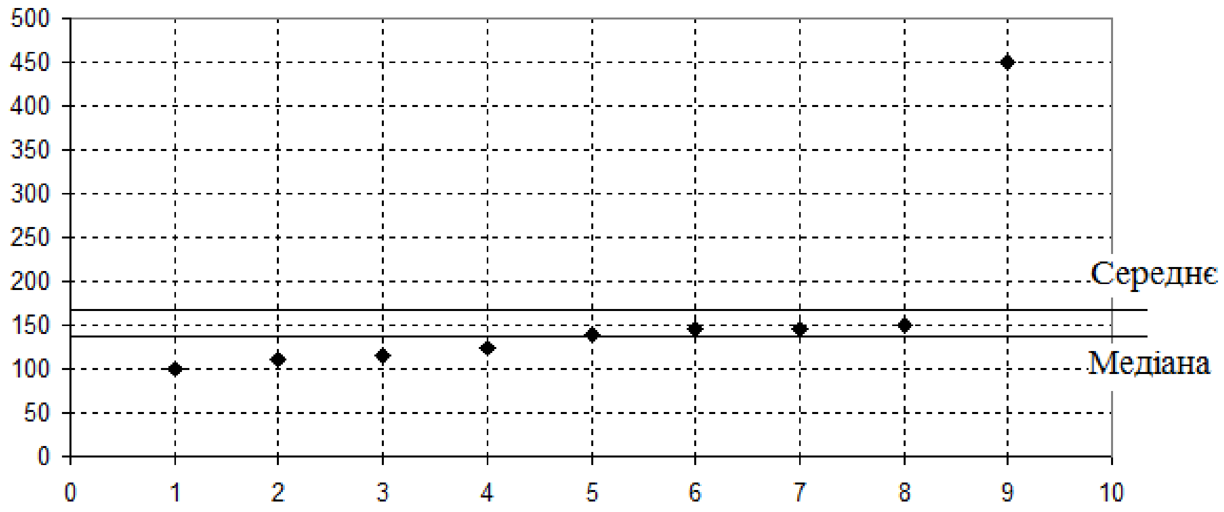

Рис. 1.8. Положення середнього і медіани відносно вибіркових даних

ПРИКЛАД 1.5. За даними вибіркового дослідження відома кількість людей, що відвідували лікарню протягом року. Дані згруповані залежно від віку відвідувачів (табл. 1.12). Знайти всі можливі числові характеристики за даними таблиці.

Таблиця 1.12

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

Розв'язок. 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 1.11 отримаємо інтервальний статистичний ряд (табл. 1.13).

Таблиця 1.13

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

Розрахуємо вибіркове середнє за формулою (1.6). Для зручності розрахунки оформимо у вигляді таблиці (табл. 1.14).

Таблиця 1.14

$[a_i;a_{i+1}]$	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 (1.10) 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 (1.11), 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 (1.14). Для зручності розрахунки оформимо у вигляді таблиці (табл. 1.15).

Таблиця 1.15

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 (1.16) дорівнює: $S = \sqrt{S^2}$ = 2,81.

Коефіцієнт варіації за формулою (1.17) дорівнює: $v = \dfrac{S}{\left|\bar{x}\right|}\cdot 100\% = \dfrac{2{,}81}{57{,}57}\cdot 100\% \approx 4{,}89\%\,.$

1.4. 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.

1.4.1. Довірчий інтервал для генерального середнього при відомій генеральній 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:

$$\overline{x} - z_{\frac{1-\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}} < a < \overline{x} + z_{\frac{1-\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \tag{1.18}$$

де значення $z_{\frac{1-\gamma}{2}}$ знаходиться з таблиці (табл. 1.16) або за допомогою вбудованої функції Excel НОРМСТОБР(γ). Величина $z_{\frac{1-\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ є шириною довірчого інтервалу.

Таблиця 1.16

Значення $z_{\frac{1-\gamma}{2}}$

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 1.6. 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 $\pm$ 1 грам.

Розв'язок. 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 (1.18) маємо довірчий інтервал:

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

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:

$$\bar{x}-t_{\frac{1-\gamma}{2},n-1}\cdot\frac{S}{\sqrt{n-1}}<a<\bar{x}+t_{\frac{1-\gamma}{2},n-1}\cdot\frac{S}{\sqrt{n-1}}, \qquad (1.19)$$

де значення $t_{\frac{1-\gamma}{2},n-1}$ знаходиться з таблиці розподілу Стьюдента, яка є у будь-яких статистичних довідниках, або за допомогою вбудованої функції Excel СТЬЮДРАСПОБР($\frac{1-\gamma}{2}$, n-1). Величина $t_{\frac{1-\gamma}{2},n-1}\cdot\frac{S}{\sqrt{n-1}}$ є шириною довірчого інтервалу.

ПРИКЛАД 1.7. Автомат фасує чай в пачки. Проведено вибірку об'ємом n=30 пачок. Середня вага пачки чаю у вибірки $\bar{x}$=101 г, вибіркове стандартне відхилення S=4 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'язок. Оскільки γ=0,95, то $\frac{1-\gamma}{2}=\frac{1-0,95}{2}=0,025$; n=30, тоді n-1=29. За допомогою Excel знайдемо $t_{\frac{1-\gamma}{2},n-1}=t_{0,025;29}$. Натиснемо $\boldsymbol{f_x}$ у командному рядку, виберемо категорію ***Статистические*** і функцію СТЬЮДРАСПОБР; задамо параметри 0,025 і 29 (див. рис. 1.9, зміст командного рядку). Отримаємо $t_{0,025;29}=2,3638$.

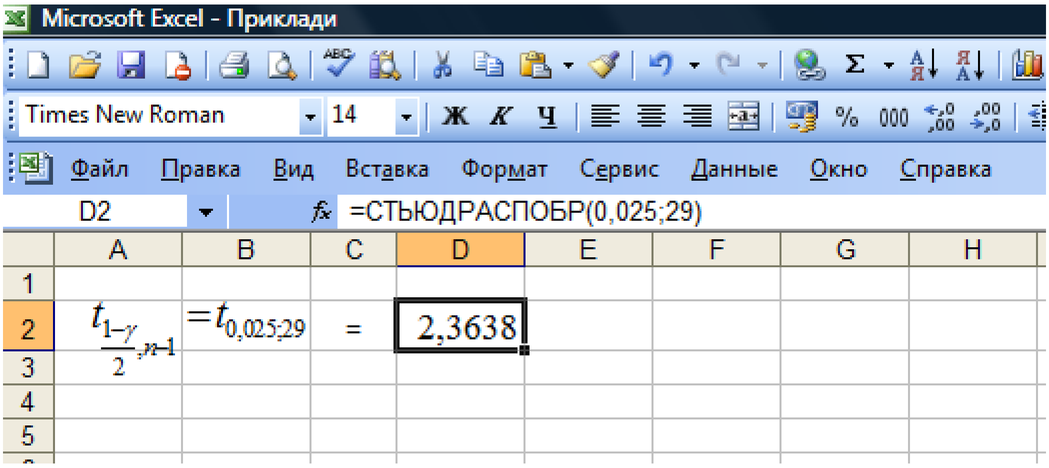

Рис. 1.9. Знаходження $t_{\frac{1-\gamma}{2},n-1}$ за допомогою Excel

Тоді ширина довірчого інтервалу $t_{\frac{1-\gamma}{2},n-1}\cdot\frac{S}{\sqrt{n-1}}=2,3638\cdot\frac{4}{\sqrt{29}}\approx 1,75$ і довірчий інтервал за формулою (1.19):

$$\bar{x}-1,75<a<\bar{x}+1,75;\quad 101-1,75<a<101+1,75;\quad 99,25<a<102,75.$$

Отже, середня вага пачки чаю знаходиться в інтервалі від 99,25 до 102,75.

Знайдемо об'єм вибірки, необхідний для того, щоб ширина довірчого інтервалу дорівнювала 1 грам, тобто $t_{\frac{1-\gamma}{2},n-1}\cdot\frac{S}{\sqrt{n-1}}$=1. Знайдемо n з отриманого рівняння:

$$2,3638\cdot\frac{4}{\sqrt{n-1}}=1\Rightarrow\sqrt{n-1}=2,3638\cdot 4\approx 9,46\Rightarrow.\quad\Rightarrow n-1=9,46^2=89,49\Rightarrow n=90,49\approx 91.$$

Отже, мінімальний об'єм вибірки для отримання довірчого інтервалу шириною 1 грам дорівнює 91 пачка.

1.4.3. 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:

$$w - z_{\frac{1-\gamma}{2}} \cdot \sqrt{\frac{w(1-w)}{n}} < W < w + z_{\frac{1-\gamma}{2}} \cdot \sqrt{\frac{w(1-w)}{n}}, \qquad (1.20)$$

де значення $z_{\frac{1-\gamma}{2}}$ знаходиться з таблиці 1.16 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 (1.20) використовується тоді, коли $nw \geq 5, \quad n(1-w) \geq 5$.

ПРИКЛАД 1.8. 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’язок. 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 (1.20) маємо довірчий інтервал:

$$w - 0{,}012 < W < w + 0{,}012; \qquad 0{,}075 - 0{,}012 < W < 0{,}075 + 0{,}012; \; 0{,}063 < W < 0{,}087.$$

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.

1.5. Шкали вимірювань

Статистичної обробці підлягають тільки ті ознаки об’єктів або фактори, які можна виміряти за деякою шкалою. Існують такі шкали вимірювань:

- шкала найменувань (класифікації);
- шкала порядку;
- шкала інтервалів;
- шкала відношень.

Шкала найменувань (класифікації)

Якщо дані вимірюються за шкалою найменувань, то над ними можливі тільки такі операції порівняння, як „рівні” або „не рівні”. Найменування слугує тільки для ідентифікації певного об’єкту – номер філіалу деякої фірми, номер методики навчання і т. ін.

Шкала порядку

Якщо дані вимірюються за шкалою порядку, їх можна порівняти за величиною – „більше”, „менше” або „рівні”. За такою шкалою вимірюються, наприклад, експертні оцінки, вік респондентів і т. ін.

Шкала інтервалів

Якщо дані вимірюються за шкалою інтервалів, їх можна порівняти за величиною – „більше”, „менше” або „рівні”, – та виконати операції додавання і віднімання, тобто визначити, наскільки більше або наскільки менше. Прикладом є шкали вимірювання температури.

Шкала відношень

Якщо дані вимірюються за шкалою відношень, їх можна порівняти за величиною та виконати всі арифметичні операції: додавання, віднімання, множення і ділення. За такою шкалою вимірюються вага, довжина, доход, об’єм виробництва і т. ін.

Від шкали вимірювання даних залежить можливість обчислити числові характеристики (табл. 1.17) і застосовувати певний статистичний метод.

Таблиця 1.17

Шкали вимірювань і числові характеристики

Назва шкали	**Числові характеристики**
Найменувань	Частоти
Порядку	Частоти, середнє, мода, медіана
Інтервалів	Частоти, середнє, мода, медіана, дисперсія
Відношень	Всі відомі

1.6. Визначення числових характеристик і довірчих інтервалів з використанням табличного процесору Microsoft Excel

Більшість числових характеристик у випадку незгрупованих даних можна обчислити з використанням табличного процесору Microsoft Excel. Основні вбудовані функції Excel, що застосовуються для таких розрахунків, надано у таблиці 1.18. Щоб викликати потрібну функцію, слід нажати кнопку $\boldsymbol{f_x}$ у командному рядку, обрати категорію ***Статистические*** та ім'я функції.

Крім того, часто корисні такі функції:

- НАИБОЛЬШИЙ (масив даних, *k*) – надає *k*-е найбільше значення в ряді даних;
- НАИМЕНЬШИЙ (масив даних, *k*) – надає *k*-е найменше значення в ряді даних.

Ширину довірчого інтервалу для генерального середнього можна знайти за допомогою вбудованої статистичної функції Excel ДОВЕРИТ (*альфа, станд_откл, размер*). Параметр *альфа* – це так званий рівень значущості, $\alpha = 1 - \gamma$; параметр *станд_откл* – це вибіркове середнє квадратичне відхилення *S*; параметр *размер* – це об'єм вибірки.

Таблиця 1.18

Статистичні функції Excel

Числові характеристики	**Назва функції**
Середнє	СРЗНАЧ (масив даних)
Середнє геометричне	СРГЕОМ (масив даних)
Мода	МОДА (масив даних)
Медіана	МЕДИАНА (масив даних)
Дисперсія	ДИСП (масив даних)
Середнє квадратичне відхилення	СТАНДОТКЛОН (масив даних)
Мінімальне значення	МИН (масив даних)
Максимальне значення	МАКС (масив даних)
Частота	ЧАСТОТА (масив даних; масив інтервалів)

ПРИКЛАД 1.9. За даними вибіркового дослідження відома заробітна платня (у грн.) 20-и службовців певної компанії (табл. 1.19). Знайти за допомогою вбудованих статистичних функцій Excel всі можливі числові характеристики за даними таблиці. Знайти довірчий інтервал для генерального середнього – середньої заробітної платні службовців компанії.

Таблиця 1.19

3560	2190	2390	3400
2180	2400	3350	2340
2900	2570	3300	3150
3680	3250	2250	3240
2180	2600	2870	3050

Розв'язок. Запишемо в лист Excel вхідні дані і числові характеристики, що можна знайти (рис. 1.10). Для знаходження характеристик введемо: в чарунку I3 формулу „=СРЗНАЧ(B2:B21)"; в чарунку I4 формулу „=МЕДИАНА(B2:B21)"; в чарунку I5 формулу „=ДИСП(B2:B21)”; в чарунку I6 формулу „=СТАНДОТКЛОН(B2:B21)”; в чарунку I7 формулу „=МАКС(B2:B21)”; в чарунку I8 формулу „=МИН(B2:B21)”.

Для знаходження довірчого інтервалу для генерального середнього знайдемо за допомогою функції ДОВЕРИТ його ширину (див. рис. 1.10, командний рядок). Параметрами візьмемо $\alpha = 1-\gamma = 1-0,95 = 0,05$; замість другого параметру надамо посилання на чарунку I6, що містить розраховане значення середнього квадратичного відхилення; *размер* – це об’єм вибірки, що дорівнює 20.

Для знаходження початку інтервалу запишемо в чарунку I10 формулу „=I3–I9”; для знаходження кінця – формулу „=I3+I9” в чарунку I11.

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

I9 =ДОВЕРИТ(0,05;I6;20)

	A	B	C	D	E	F	G	H	I
1		Вихідні дані		Числові характеристики					
2		3560							
3		2180		Середнє					2842,5
4		2900		Медіана					2885
5		3680		Дисперсія					257914
6		2180		Середнє квадратичне відхилення					507,853
7		2190		Максимальне значення					3680
8		2400		Мінімальне значення					2180
9		2570		Ширина довірчого інтервалу					222,572
10		3250		Початок довірчого інтервалу					2619,93
11		2600		Кінець довірчого інтервалу					3065,07
12		2390							
13		3350							
14		3300							
15		2250							
16		2870							
17		3400							
18		2340							
19		3150							
20		3240							
21		3050							

Рис. 1.10. Розрахунок числових характеристик

1.7. Побудова гістограми засобами Microsoft Excel

Excel надає два способи побудови гістограми.

Для побудови гістограми першим способом необхідно:

1) Внести в лист Excel вхідні дані і інтервали, за якими ці дані будуть групуватися.

2) Знайти частоти попадання даних в інтервали за допомогою функції ЧАСТОТА, для чого:

- виділити діапазон чарунок (на одну більше, ніж інтервалів), в яких будуть записані частоти;
- викликати f_x – ***Статистические*** – ЧАСТОТА;
- ввести посилання на чарунки, що містять вхідні дані і інтервали;
- натиснути **Ctrl+Shift+Enter**.

3) Викликати ***Вставка – Диаграмма – Гистограмма***, появиться діалогове вікно (див. рис. 1.11).

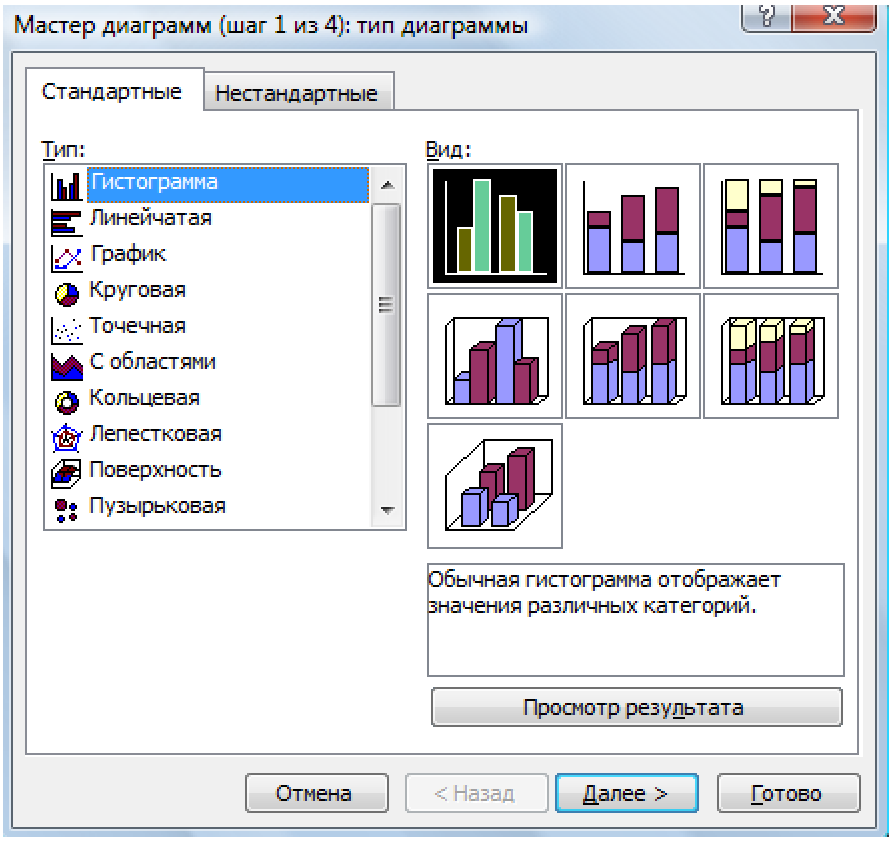

Рис. 1.11. Діалогове вікно майстра діаграм

4) Надати необхідні для побудови гістограми параметри:
- діапазон вхідних даних, спосіб їх групування (за рядками або стовпцями) та імена рядів даних, якщо це потрібно;
- якщо імена рядів надано, відмітити *Добавить легенду* і вказати її розміщення;
- якщо потрібно, добавити *Имена рядов*, або (та) *Имена категорій*, або (та) *Значения*;
- якщо потрібно, добавити *Заголовок*, *Линии сетки*, *Оси*, *Таблицу данных*.

Для побудови гістограми другим способом необхідно:

1) Внести в лист Excel вихідні дані.

2) Обрати в меню ***Сервис – Анализ данных – Гистограмма***, появиться діалогове вікно (див. рис. 1.12).

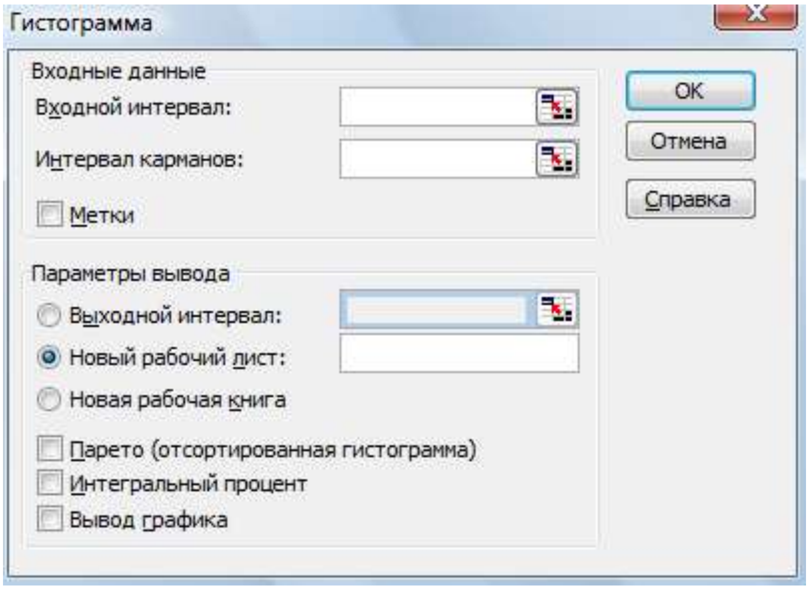

Рис. 1.12. Діалогове вікно для побудови гістограми

3) Задати необхідні для побудови гістограми параметри:

входной диапазон – задати посилання на чарунки, в яких знаходяться вхідні дані;

интервал карманов (параметр не є обов'язковим) – задати діапазон чарунок і набор граничних значень у порядку зростання; якщо параметр не введений, то буде автоматично створений набор відрізків, рівномірно розподілених між мінімальним і максимальним значеннями даних;

выходной диапазон – ввести посилання на верхню ліву чарунку діапазону, в який буде надано гістограму, або відмітити параметр *Новый рабочий лист* або *Новая рабочая книга*;

интегральный процент – якщо цей параметр відмічено, то будуть розраховані накопичені частоти і побудований їх графік;

вывод графика – якщо цей параметр відмічено, то буде створено автоматичну діаграму, при цьому обов'язково задається значення *Новая книга*.

ПРИКЛАД 1.10. За даними вибіркового дослідження відома кількість родин з дітьми дошкільного віку у селах деякої області (табл. 1.20). Побудувати за допомогою Excel гістограму за даними таблиці.

Розв'язок. Запишемо в лист Excel вхідні дані завдання (рис. 1.13), стовпець з границями інтервалів в чарунках M2:M11 (задається тільки початок інтервалів). Розрахуємо частоти попадання в інтервали (див. зміст командного рядку на рис. 1.13).

Таблиця 1.20

27	36	34	46	43	28	29	37	40	43
40	33	50	37	41	32	27	43	34	32
30	41	54	42	47	35	49	49	54	36
36	51	36	24	35	25	33	38	38	36
29	51	32	36	53	30	55	44	46	38

29	44	48	30	34	46	47	36	37	36
30	58	42	46	46	29	38	44	40	30
35	35	63	47	37	29	53	41	42	41

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1				Кількість родин з дітьми дошкільного віку									Інтервали	Частоти
2		27	36	34	46	43	28	29	37	40	43		24	1
3		40	33	50	37	41	32	27	43	34	32		28	4
4		30	41	54	42	47	35	49	49	54	36		32	13
5		36	51	36	24	35	25	33	38	38	36		36	17
6		29	51	32	36	53	30	55	44	46	38		40	11
7		29	44	48	30	34	46	47	36	37	36		44	13
8		30	58	42	46	46	29	38	44	40	30		48	9
9		35	35	63	47	37	29	53	41	42	41		52	5
10													56	5
11													60	1
12														1

Рис. 1.13. Вхідні дані для побудови гістограми

Викличемо ***Вставка – Диаграмма – Гистограмма***, задамо діапазон даних, тобто розраховані частоти і вкажемо групування за стовпцями (див. рис. 1.14, виділення діапазону чарунок, що містять частоти).

Для зручності читання діаграми добавимо *Заголовок* та *Значения*.

Уберемо відмітку *Легенда*, оскільки імена рядів не було надано – розглядається тільки один тип даних.

Після побудови діаграми можна у разі необхідності змінити шрифти, ширину стовпців гістограми, колір стовпців і фону. Для внесення змін потрібно двічі натиснути правою кнопкою миші на відповідне поле гістограми.

Зауважимо, що на горизонтальній вісі надаються не границі інтервалів, а їх порядковий номер.

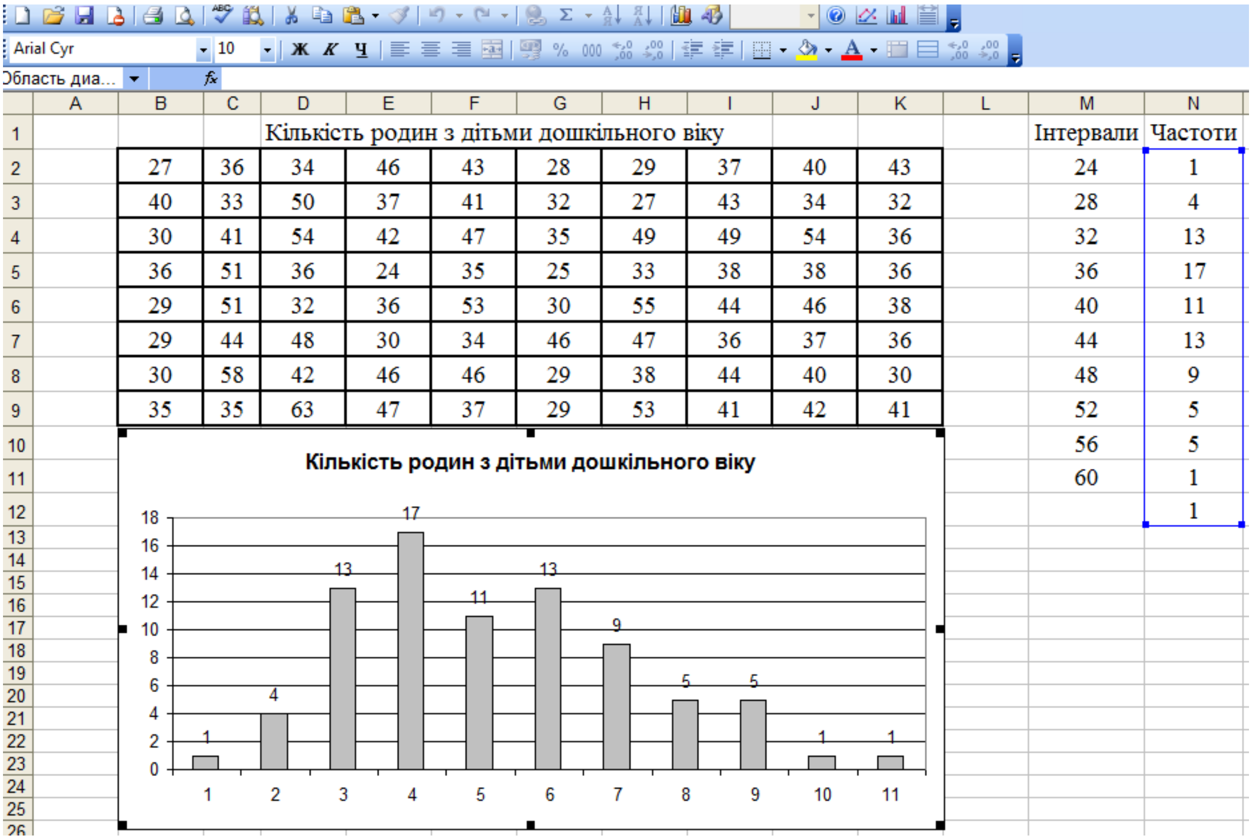

Рис. 1.14. Лист Excel з вхідними даними і гістограмою

1.8. Завдання для самостійного виконання

1.1. Відомі дані про безаварійну роботу автоматизованого комплексу (в місяцях) (табл. 1. 21). Побудувати статистичний ряд за даними вибірки, визначити середній час безаварійної роботи, дисперсію і середнє квадратичне відхилення часу. Побудувати полігон частот і гістограму.

Таблиця 1.21

0,000	0,000	0,002	0,006	0,023	0,084	0,382	0,810	0,003	0,864
1,033	0,912	0,093	0,324	0,194	0,522	2,336	0,057	0,654	0,250
0,877	0,276	0,037	0,537	0,183	1,306	0,752	0,198	1,623	0,875
0,185	0,274	0,613	0,356	0,645	0,676	1,079	0,500	0,902	0,191
0,250	0,348	0,320	0,182	0,458	0,936	1,204	0,576	0,303	0,522

1.2. Інтервал між потягами у метро складає 3 хв. В таблиці 1.21 надано час очікування пасажирами потягу. Скласти інтервальний статистичний ряд, знайти середній час очікування, медіану, дисперсію і середнє квадратичне відхилення часу. Побудувати функцію розподілу величини X – часу очікування.

Таблиця 1.22

0,787	1,004	0,941	0,612	1,200	1,692	1,354	0,908	1,245	1,292
0,617	0,828	1,413	1,030	1,459	2,483	2,769	1,563	2,661	1,635

1,654	0,838	1,143	0,618	2,317	1,853	1,555	0,653	1,922	1,653
1,747	2,677	0,341	2,952	0,545	1,297	1,981	0,214	2,452	2,087
0,001	0,007	0,025	0,312	1,068	2,604	0,014	0,045	2,340	2,001

1.3. Відомі дані про інтервал часу між появою покупців у касовому залі деякого магазину (табл. 1.23). Побудувати інтервальний статистичний ряд, знайти всі можливі числові характеристики. Побудувати графіки функції і щільності розподілу величини X – інтервалу часу.

Таблиця 1.23

0,002	1,004	0,007	0,612	0,091	1,692	1,527	0,908	2,590	1,292
4,134	3,647	0,374	2,150	0,778	5,223	3,344	2,001	3,492	4,011
3,507	0,838	0,148	0,618	0,704	1,853	3,007	0,653	3,600	1,653
0,738	1,069	2,453	1,447	2,614	3,742	4,314	1,211	1,949	5,001
1,000	0,007	1,272	0,312	1,832	2,604	2,267	0,045	4,450	2,001

1.4. В таблиці 1.24 наведено значення прибутку 50 фірм, що належать одній корпорації (в 1000 у. од.). Знайти середнє значення прибутку, дисперсію і середнє квадратичне відхилення. Побудувати полігон частот і гістограму.

Таблиця 1.24

4,744	9,127	7,201	8,650	11,534	9,013	10,390	9,268	7,354	10,255
6,232	6,739	6,088	8,671	15,103	9,124	11,902	10,216	10,954	11,470
7,351	9,832	7,126	9,715	10,744	10,687	10,582	12,271	11,047	13,190
5,536	8,917	9,823	8,383	14,212	15,031	13,001	11,089	12,091	10,321
9,766	5,854	2,917	6,379	6,748	7,024	11,587	11,101	10,954	10,387

1.5. Відомі дані про річний об’єм виробництва (тис. т) підприємств цементної промисловості (табл. 1.25). Побудувати інтервальний статистичний ряд, полігон частот і гістограму. Знайти всі можливі числові характеристики.

1.6. При формуванні портфелю поставок для фірми було обрано 100 поставщиків, які працювали із фірмою у минулому році. Знайти довірчий інтервал для долі поставщиків, що несвоєчасно здійснюють поставки на рівні надійності 0,999, якщо у вибірці таких 25.

Таблиця 1.25

11,240	18,545	17,750	22,560	18,355	20,424	20,650	10,780	15,590
13,720	28,505	23,170	20,360	22,450	21,590	14,565	24,295	25,140
27,655	17,786	27,045	28,650	18,670	31,445	18,540	15,598	19,720
15,230	21,240	19,535	12,934	18,195	19,074	17,037	19,610	20,970
22,075	15,090	20,754	10,195	13,580	21,490	13,987	22,645	21,218

1.7. Відомі дані про розмір вкладів в банку (табл. 1.26). Знайти з надійністю 0,96 довірчий інтервал для середнього розміру вкладу.

Таблиця 1.26

Розмір вкладу (тис. грн.)	10-30	31-50	51-70	71-90	91-110	111-130
Кількість вкладчиків	1	3	10	30	60	7

1.8. В результаті вимірювання продуктивності праці 150 робітників виявилося, що середня продуктивність дорівнює 256 од. продукції за зміну, а середнє квадратичне відхилення – 4,56. Знайти довірчий інтервал для вибіркового середнього з надійністю 0,99.

1.9. З 3000 одиниць продукції було перевірено 40%. Серед них виявлено 90% одиниць продукції першого сорту. Знайти довірчий інтервал, в якому з надійністю 0,99 знаходиться доля продукції першого сорту.

1.20. Серед 5000 автомобілів працівниками ДАІ було перевірено 600, серед яких виявлено 460 з різного роду несправностями. Знайти з надійністю 0,95 долю несправних автомобілів.

1.21 – 1.30. Автомат фасує цукор в пакети. Проведено вибірку n пакетів (табл. 1.27). Середня вага цукру виявилася рівною $\overline{x}$, вибіркове стандартне відхилення S. Знайти довірчий інтервал для середньої ваги цукру з надійністю γ, якщо:

1) стандартне відхилення автомата σ кг;

2) стандартне відхилення автомату невідоме.

Знайти об’єм вибірки, необхідний для того, щоб ширина довірчого інтервалу дорівнювала Δ.

Таблиця 1.27

№ варіанта	$\overline{x}$	n	σ	Δ	γ	S
1	0,99	30	0,01	0,10	0,95	0,05
2	0,98	34	0,07	0,15	0,99	0,10
3	0,97	33	0,03	0,18	0,95	0,04
4	0,96	35	0,06	0,12	0,99	0,08
5	0,95	36	0,09	0,19	0,95	0,02
6	1,01	32	0,02	0,11	0,99	0,09

7	1,02	37	0,08	0,13	0,95	0,06
8	1,03	38	0,04	0,16	0,99	0,03
9	1,04	39	0,10	0,14	0,95	0,07
10	1,05	31	0,05	0,17	0,99	0,01