

Лабораторна робота.

Ряди та їх статистичні характеристики

Мета – закріплення теоретичного й практичного матеріалу, придбання навичок роботи з варіаційними рядами в табличних редакторах *OpenOffice 3.2 Calc* та *Microsoft Office Excel*.

Завдання: необхідно провести аналіз статистичних характеристик дискретного та інтервального ряду у *Calc* або *Excel*.

1. Розрахувати статистичні характеристики ряду (середнє, дисперсію, середнє квадратичне відхилення, моду, медіану, коефіцієнти асиметрії та ексцесу).

2. Побудувати гістограму та полігон розподілу випадкової величини, зробити висновки щодо характеру закону розподілу.

3. За допомогою критеріїв Пірсона та Колмогорова – Смірнова перевірити гіпотезу про нормальний закон розподілу.

Лабораторна робота виконується в рамках вивчення теми 1 "Математико-статистична обробка вибірових даних".

Методичні рекомендації до виконання завдання

1. Запуск *OpenOffice* або *Microsoft Office* і підготовка даних.

Для початку роботи в пакеті після ініціювання ярлика *OpenOffice* необхідно обрати "*Електронная таблица*" (або після ініціювання ярлика *Microsoft Office* обрати *Microsoft Office Excel*).

Для введення даних (тексту чи числа) потрібно перейти до комірки, набрати дані і натиснути клавішу *Enter*. Кожен елемент даних займає одну комірку поля даних. Після заповнення всіх комірок одержимо таблицю, у якій зображено дискретний варіаційний ряд (рис. 1.1.).

Адреса кожної комірки складається з назви стовпця (буква алфавіту) та номера строки. Наприклад комірка A2, комірка F23.

Якщо треба вказати на блок комірок – слід надати адресу верхньої лівої та нижньої правої комірки блоку.

Наприклад, щоб вказати на необхідність роботи з даними, наведеними на рис. 1.1., слід вказати адресу, за якою вони розміщені – A2:B12.

	A	B	C
1	№ спостереження	X	
2	1	313,10	
3	2	1387,50	
4	3	438,40	
5	4	425,50	
6	5	290,50	
7	6	124,60	
8	7	262,90	
9	8	330,20	
10	9	470,20	
11	10	14772,90	
12	11	11854,00	
13	12	10735,20	

Рис. 1.1. Вихідні дані

2. **Розрахунок основних статистичних характеристик для дискретного ряду.** Для розрахунку основних статистичних характеристик дискретного варіаційного ряду, необхідно скористатися Майстром функцій, діалогове вікно якого зображено на рис. 1.2 для пакета *OpenOfficeCalc* чи на рис. 1.3 для *Microsoft Office Excel*.

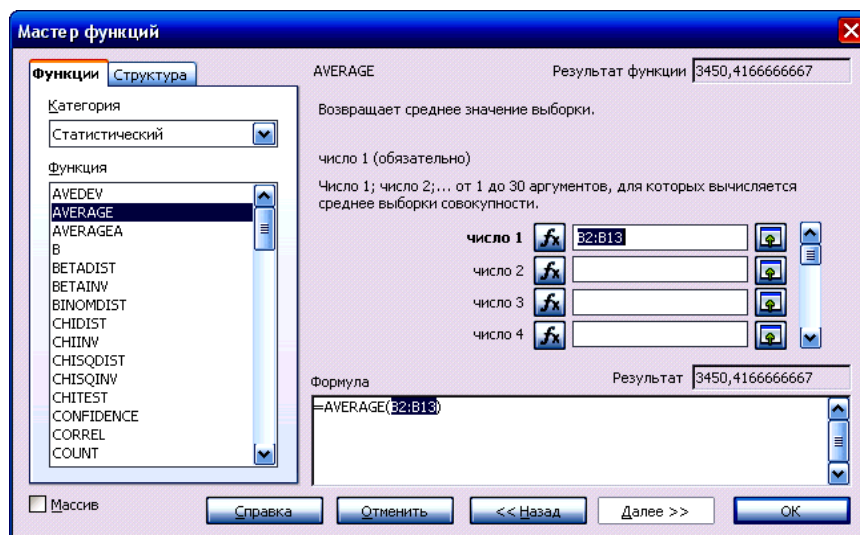


Рис. 1.2. Діалогове вікно Майстра функцій OpenOfficeCalc

Кожній із статистичних характеристик відповідає визначена функція, а саме: середнє значення ($=\text{AVERAGE}(\text{B2:B13})$); дисперсія ($=\text{VAR}(\text{B2:B13})$); стандартне відхилення ($=\text{STDEV}(\text{B2:B13})$); мода ($=\text{MODE}(\text{B2:B13})$); медіана ($=\text{MEDIAN}(\text{B2:B13})$); коефіцієнт асиметрії

(=SKEW(B2:B13)) та ексцесу (=KURT(B2:B13)).

У Microsoft Excel для розрахунку зазначених характеристик відповідають наступні функції: середнє значення – СРЗНАЧ; дисперсія – ДИСП; стандартне відхилення – СТАНДОТКЛОН; мода – МОДА; медіана – МЕДИАНА; коефіцієнт асиметрії – СКОС; коефіцієнт ексцесу – ЭКСЦЕСС.

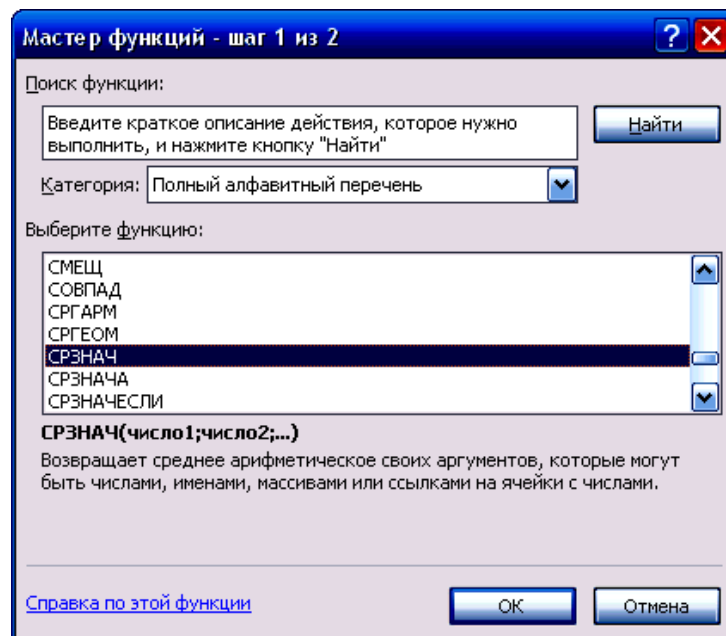


Рис.1.3. Діалогове вікно Майстра функцій Microsoft Office Excel

Результати розрахунку цих характеристик зображено на рис. 1.4.

	D	E
13		
14		Дискретний ряд
15	Середнє	3450,42
16	Дисперсія	30366499,27
17	Середнє квадратичне відхилення(СКВ)	5510,58
18	Мода	#ЗНАЧЕН!
19	Номер медіани	
20	Медіана	431,95
21	коефіцієнт асиметрії	1,42
22	коефіцієнт ексцесу	0,24
23		

Рис. 1.4. Результати розрахунку основних статистичних характеристик для дискретного варіаційного ряду

У даному випадку дискретний варіаційний ряд має множинне значення моди, тому що кожне зі значень цього ряду зустрічається однаково кількість разів.

3. Перетворення дискретного варіаційного ряду в інтервальний. Для побудови інтервального ряду необхідно розрахувати крок групування інтервалів за наступною формулою:

$$k = \frac{\max(X) - \min(X)}{n}, n = 1 + 3,22 \lg(N), \quad (1)$$

де k – крок групування інтервалів;

X – даний дискретний варіаційний ряд, $X = (x_1, x_2, \dots, x_N)$;

$\max(X) \min(X)$ – розмах варіаційного ряду X ;

n – кількість інтервалів;

N – кількість спостережень.

Для розрахунку елементів цієї формули необхідно скористатися функціями *Calc* MAX, MIN, ROUND (для округлення кількості інтервалів) (в *Microsoft Excel* відповідно – МАКС, МИН, ОКРУГЛ), що стають активними після введення у комірку знаку "=" у меню функцій, які зустрічаються більш часто (рис. 1.5.)

	B	C
1	313,10	
2	1387,50	
3	438,40	
4	425,50	
5	290,50	
6	124,60	
7	262,90	
8	330,20	
9	470,20	
10	14772,90	
11	11854,00	
12	10735,20	

Рис. 1.5. Меню функцій, що зустрічаються більш часто

Після отримання проміжних результатів потрібно розрахувати крок групування, як зображено на рис. 1.6.

	А	В	С
14			
15	Максимальне значення	14772,9	
16	Мінімальне значення	124,6	
17	Розмах	14648,3	
18	n	=ROUND(1+3,22*LOG10(12);0)	
19	Крок групування	3662,075	
20			

Рис. 1.6. Розрахунок кроку групування інтервалів ряду

Формування границь інтервалів відбувається наступним чином: нижня границя першого інтервалу дорівнює мінімальному значенню дискретного варіаційного ряду (для усіх наступних інтервалів дорівнює верхній границі попереднього інтервалу); верхня границя усіх інтервалів дорівнює сумі нижньої границі відповідного інтервалу та кроку групування. Кількість інтервалів збільшується до тих пір, поки верхня границя не охопить максимальне значення ряду (рис. 1.7 – 1.8)

	А	В
15	Максимальне значення	14772,9
16	Мінімальне значення	124,6
17	Розмах	14648,3
18	n	4
19	Крок групування	3662,075
20		
21		
22		
23		
24		
25		
26	1	2
27	Нижня межа інтервалу	Верхня межа інтервалу
28	124,6	=A28+B\$19

Рис. 1.7. Розрахунок верхньої границі інтервалу

	А	В
23		
24		
25		
26	1	2
27	Нижня межа інтервалу	Верхня межа інтервалу
28	124,6	3786,68
29	3786,68	7448,75
30	7448,75	11110,83
31	11110,83	14772,90
32	Сума	

Рис. 1.8. Сформовані інтервали варіаційного ряду

Після завершення формування границь інтервалів необхідно розрахувати частоти потрапляння до кожного інтервалу значень дискретного варіаційного ряду. Для цього можна скористатися функцією FREQUENCY (*OpenOffice*) або ЧАСТОТА (*Microsoft Excel*), обираючи у якості вихідного масиву заданий дискретний варіаційний ряд, а у якості масиву групування верхні межі побудованих інтервалів.

Остаточний вигляд інтервального ряду зображено на рис. 1.9.

	А	В	С
25			
26	1	2	3
27	Нижня межа інтервалу	Верхня межа інтервалу	Емпіричні частоти (f)
28	124,6	3786,68	9
29	3786,68	7448,75	0
30	7448,75	11110,83	1
31	11110,83	14772,90	2
32	Сума		12
33			

Рис. 1.9. Сформований інтервальний варіаційний ряд

4. Розрахунок основних статистичних характеристик для інтервального ряду. Для розрахунку основних статистичних характеристик для інтервального ряду необхідні наступні формули:

1. Середнє значення

$$\bar{X} = \frac{\sum x_i f_i}{\sum f_i}, \quad (2)$$

де x_i – середина відповідного інтервалу;

f_i – частота відповідного інтервалу, $i = \overline{1, n}$;

n – кількість інтервалів.

2. Дисперсія та стандартне відхилення:

$$D(x) = \frac{\sum (x - \bar{x})^2 f_i}{\sum f_i}, \quad (3)$$

$$\sigma = \sqrt{D}. \quad (4)$$

3. Мода

$$Mo = \frac{f_{mo} - f_{mo-1}}{(f_{mo} - f_{mo-1}) + (f_{mo} - f_{mo+1})} k_{mo} + x_{mo}, \quad (5)$$

де f_{mo} – частота модального інтервалу (модальний інтервал – це інтервал, якому відповідає найбільша частота ряду);

f_{mo-1} – частота інтервалу, що передує модальному;

f_{mo+1} – частота інтервалу, наступного за модальним;

k_{mo} – крок модального інтервалу;

x_{mo} – нижня границя модального інтервалу.

4. Медіана

$$Me = \frac{N_{me} - S_{me-1}}{f_{me}} k_{me} + x_{me}, \quad (6)$$

де f_{me} – частота медіанного інтервалу (медіанний інтервал – це інтервал, кумулятивна частота якого охоплює номер медіани, що розраховується за формулою $N_{me} = \frac{\sum f_i}{n}$);

S_{me-1} – кумулятивна частота інтервалу, що передує медіанному;

k_{me} – крок медіанного інтервалу;

x_{me} – нижня границя медіанного інтервалу

5. Коефіцієнт асиметрії

$$K_a = \frac{\bar{x} - Mo}{\sigma}. \quad (7)$$

6. Коефіцієнт ексцесу

$$K_3 = \frac{1}{n} \frac{\sum (x - \bar{x})^4 f_i}{\sigma^4}. \quad (8)$$

Таблиця проміжних результатів розрахунку зазначених статистичних характеристик зображена на рис. 1.10.

	A	B	C	D	E	F	G	H
25								
26	1	2	3	4	5	6	7	8
27	Нижня межа інтервалу	Верхня межа інтервалу	Емпіричні частоти (f)	Середина інтервалу (X)	X*f	(X-X _{ср}) ² *f	Накопичені частоти S	(X-X _{ср}) ⁴ *f
28	124,6	3786,68	9	1955,64	17600,74	53643173,22	9	319732225931016,00
29	3786,68	7448,75	0	5617,71	0,00	0,00	9	0,00
30	7448,75	11110,83	1	9279,79	9279,79	23841410,32	10	568412846099584,00
31	11110,83	14772,90	2	12941,86	25883,73	146028638,22	12	10662181589727300,00
32	Сума		12		52764,25	223513221,76		11550326661757900,00
33								

Рис. 1.10. Проміжні результати розрахунку статистичних характеристик інтервального варіаційного ряду

Результати розрахунків статистичних характеристик для дискретного та інтервального рядів зображені на рис. 1.11

	A	B	C	D	E	F	
14					Дискретний ряд	Інтервальний ряд	
15	Максимальне значення	14772,9		Середнє	3450,42	4397,02	
16	Мінімальне значення	124,6		Дисперсія	30366499,27	18626101,81	
17	Розмах	14648,3		Середнє квадратичне відхилення(СКВ)	5510,58	4315,80	
18	n	4		Мода	#ЗНАЧЕН!	1955,64	
19	Крок групування	3662,075		Номер медіани		6	
20				Медіана	431,95	2565,98	
21				коефіцієнт асиметрії	1,42	0,57	
22				коефіцієнт ексцесу	0,24	2,77	
23							
24							

Рис. 1.11. Результати розрахунку основних статистичних характеристик дискретного та інтервального рядів

4.1. Графічне зображення дискретного та інтервального варіаційних рядів. Розглянемо графічне зображення рядів даних окремо в пакеті *OpenOfficeCalc* та *Microsoft Excel*.

4.1.1. Побудова діаграм в OpenOfficeCalc. Для більш наочного представлення варіаційних рядів доцільно використовувати графіки їх розподілу. Для цього необхідно в меню Вставка обрати "Діаграма", після чого на екрані з'явиться меню Майстра діаграм, зображене на рис. 1.12.

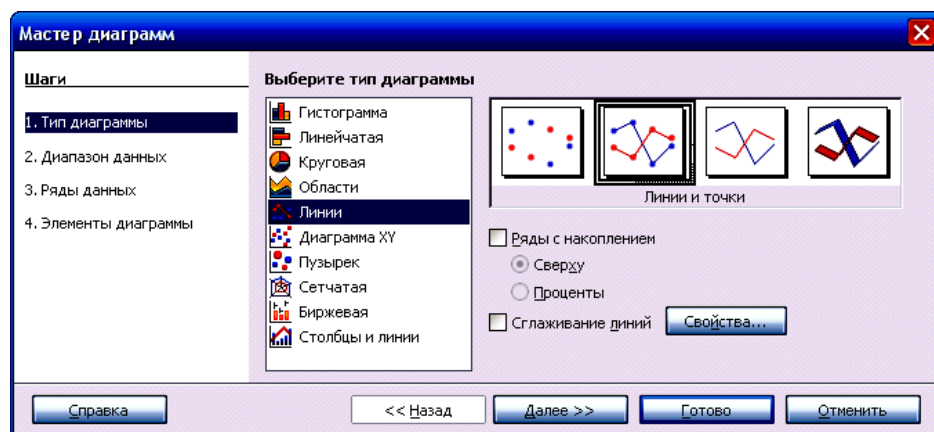


Рис. 1.12. Майстер діаграм OpenOfficeCalc

Для зображення дискретного ряду потрібно обрати тип діаграми Лінії (Лінії та точки) і натиснути Далее>>. Наступним кроком необхідно обрати діапазон даних, що необхідно зобразити, визначити розташування та підпис даних, як зображено на рис. 1.13. Після чого натиснути Готово. Графік зображення дискретного варіаційного ряду, у якого по осі X – номер об'єкта, а по осі Y – значення дослідженого показника представлено на рис. 1.14.

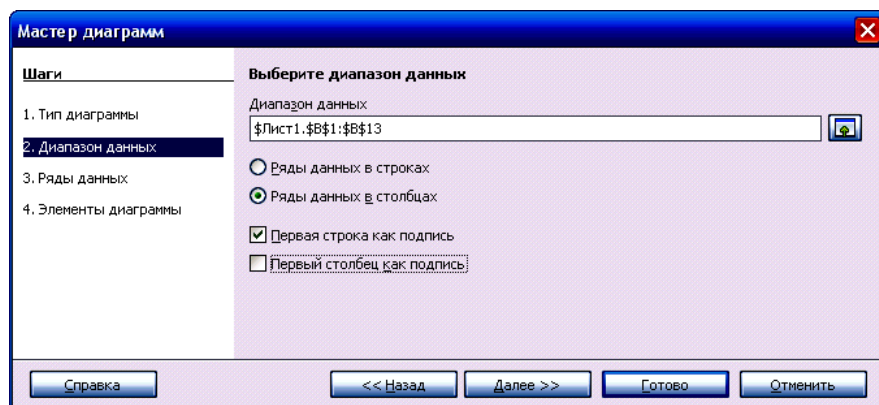


Рис. 1.13. Вікно визначення діапазону даних OpenOfficeCalc

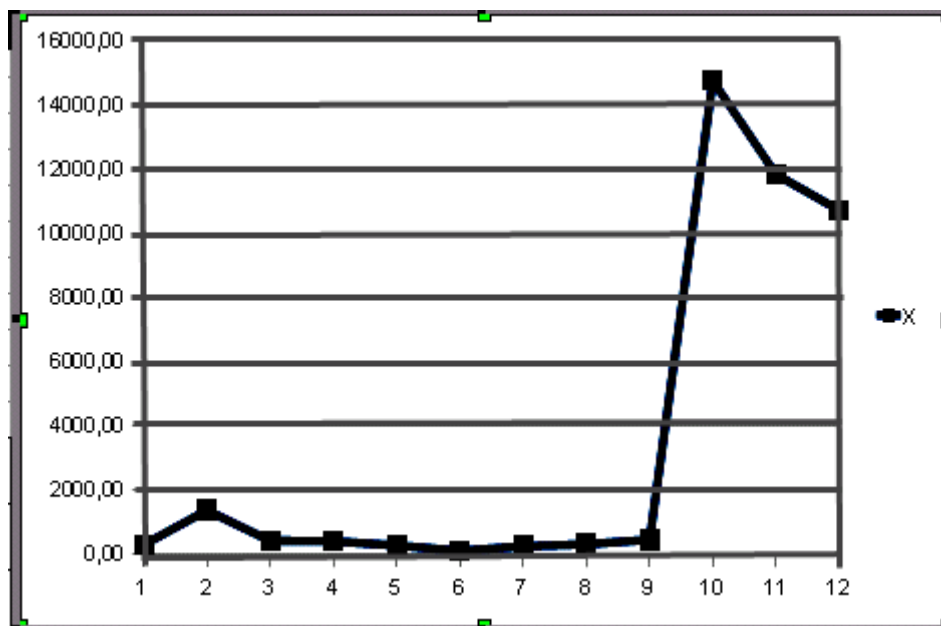


Рис. 1.14. **Графік зображення дискретного варіаційного ряду**
OpenOfficeCalc

Для зображення інтервального ряду у Майстрі діаграм потрібно обрати тип діаграми Гістограма (Звичайна) і натиснути Далее>>. Після обрання діапазону даних необхідно визначити комірку з назвою досліджуваного показника (Частоти (f)) та самі значення, як зображено на рис. 1.15.

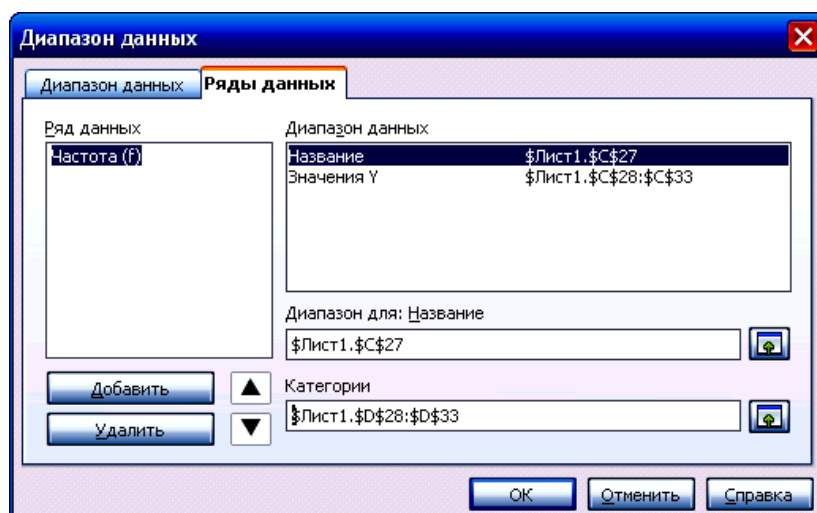


Рис. 1.15. **Визначення діапазону назви та значень досліджуваного показника за допомогою Майстра діаграм OpenOfficeCalc**

При ініціюванні кнопки "ОК" з'явиться гістограма розподілу частот (рис. 1.16)

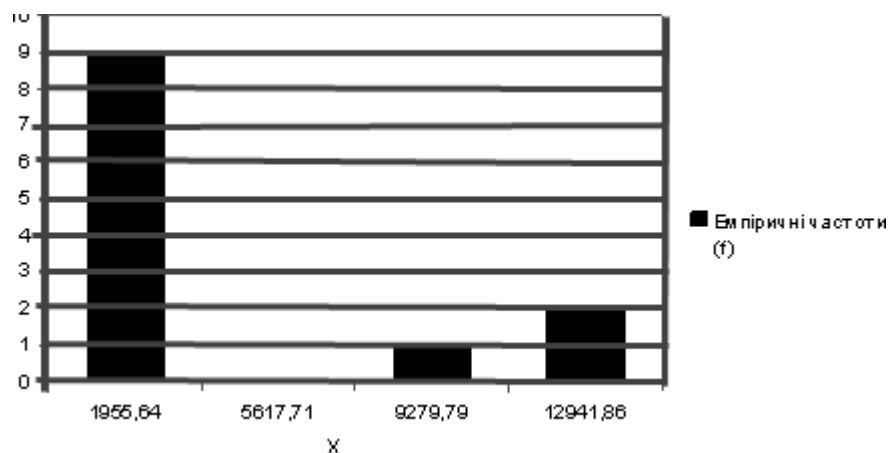


Рис. 1.16. Гістограма розподілу частот інтервального ряду OpenOfficeCalc

Наступним кроком доцільно графічно зобразити кумуляту частот інтервального ряду. Для цього у Майстрі діаграм необхідно обрати тип діаграми Лінії (Лінії та точки), а діапазоні даних визначити стовбцем, у якому знаходяться накопичені частоти інтервалів. У результаті з'явиться графік, зображений на рис. 1.17.

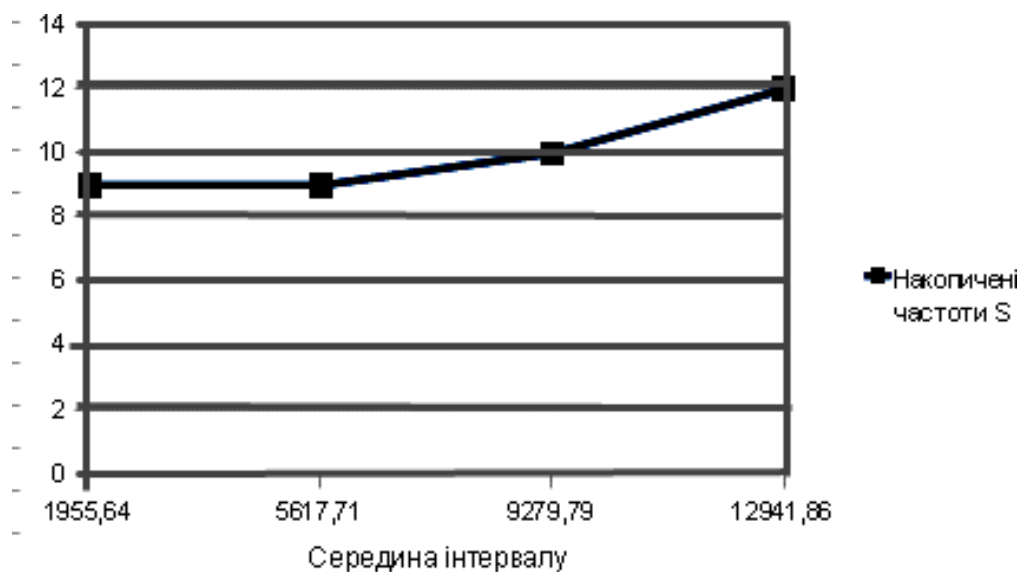


Рис. 1.17. Кумулята частот інтервального ряду OpenOfficeCalc

4.1.2. Побудова діаграм у Microsoft Office Excel. Для наочного представлення варіаційних рядів у середовищі *Microsoft Office Excel* необхідно в меню Вставка обрати "Диаграмма" після чого на екрані з'явиться меню Майстра діаграм, зображене на рис. 1.18.

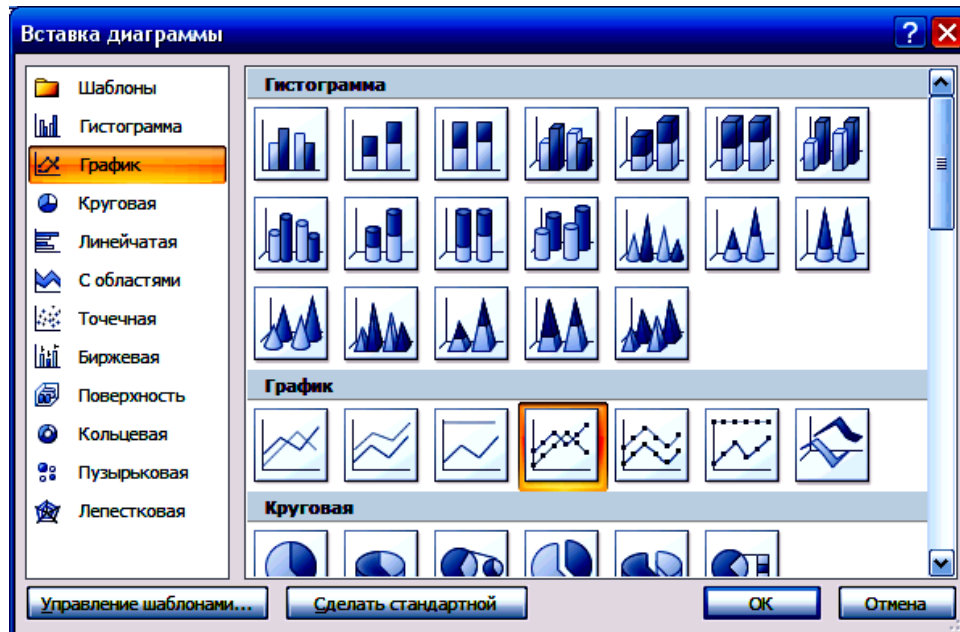


Рис. 1.18. Обрання типу діаграми у Майстра діаграм Excel

Для зображення дискретного ряду потрібно обрати тип діаграми "График (График с маркерами)" і натиснути ОК.

У контекстному меню порожнього рисунка, який щойно з'явився, потрібно обрати "Выбрать данные".

Наступним кроком необхідно обрати діапазон даних, що необхідно зобразити, визначити розташування та підпис, як зображено на рис. 1.19, 1.20. Після чого натиснути ОК.

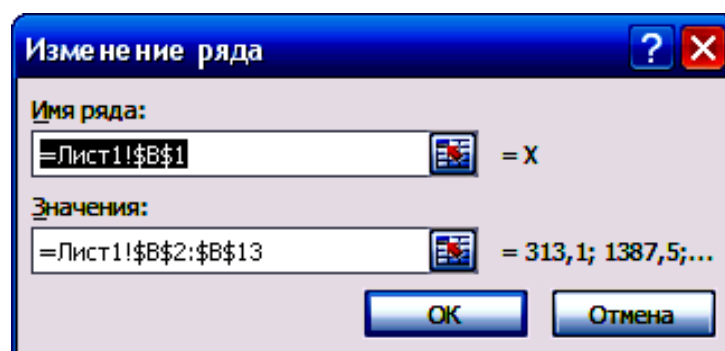


Рис. 1.19. Визначення підпису та діапазону показника у Excel

Графік зображення дискретного варіаційного ряду, у якого по осі Х – номер об'єкта, а по осі Y – значення дослідженого показника представлено на рис. 1.21

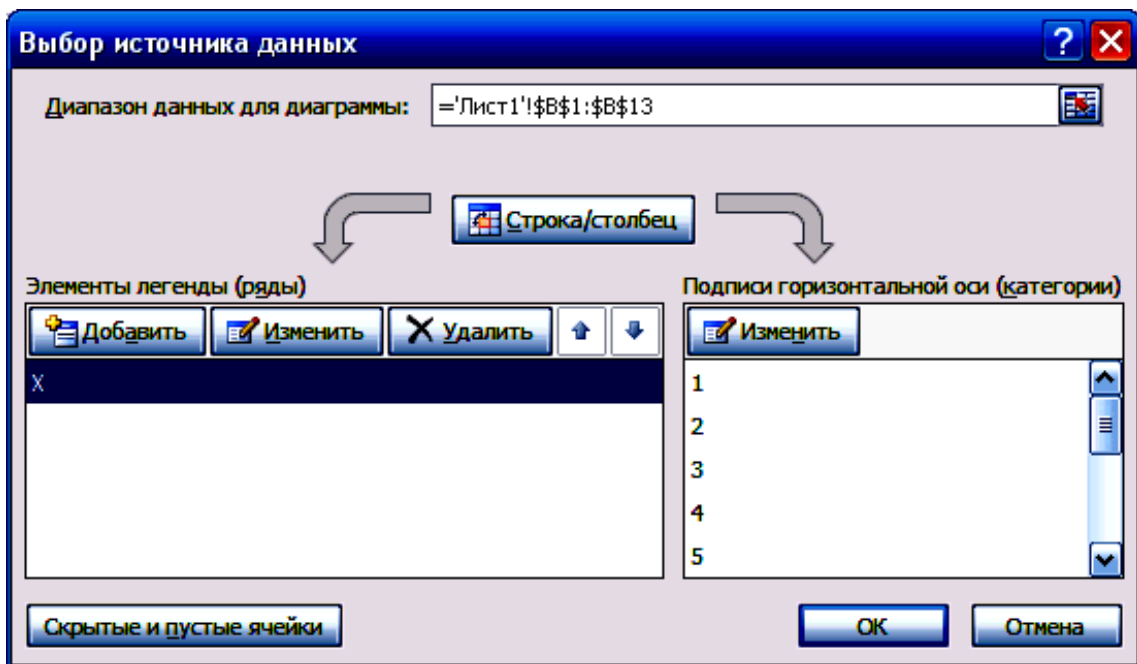


Рис. 1.20. Вікно заповнення вихідних даних

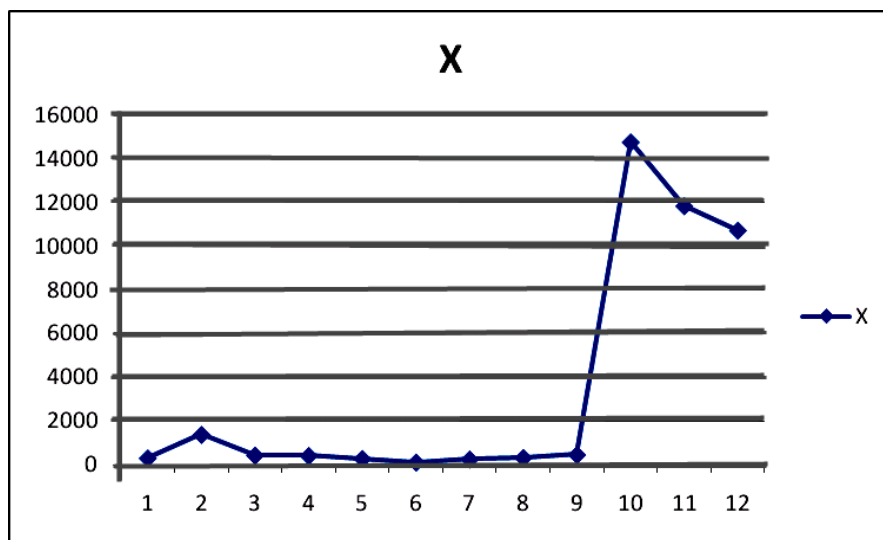


Рис. 1.21. Графік зображення дискретного варіаційного ряду у Microsoft Office Excel

Для зображення інтервального ряду у Майстра діаграм потрібно обрати тип діаграми "Гистограмма (Гистограмма с группировкой)" і

натиснути ОК. Наступним кроком здійснюється обрання діапазону даних і визначення комірки з назвою досліджуваного показника (Емпіричні частоти (f) та самі значення), а також визначення підписів горизонтальної осі, у якості якої виступають значення середини інтервалів, як зображено на рис. 1.22.

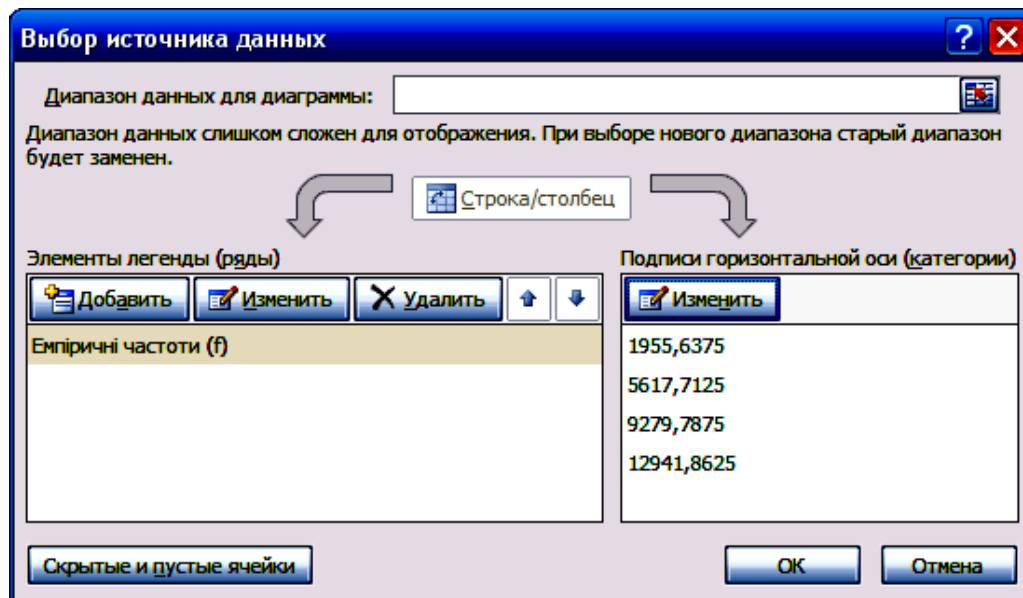


Рис. 1.22. Вікно заповнених вихідних даних інтервального варіаційного ряду у Майстрі діаграм у Microsoft Office Excel

При ініціюванні кнопки "ОК" з'явиться гістограма розподілу частот (рис. 1.23)

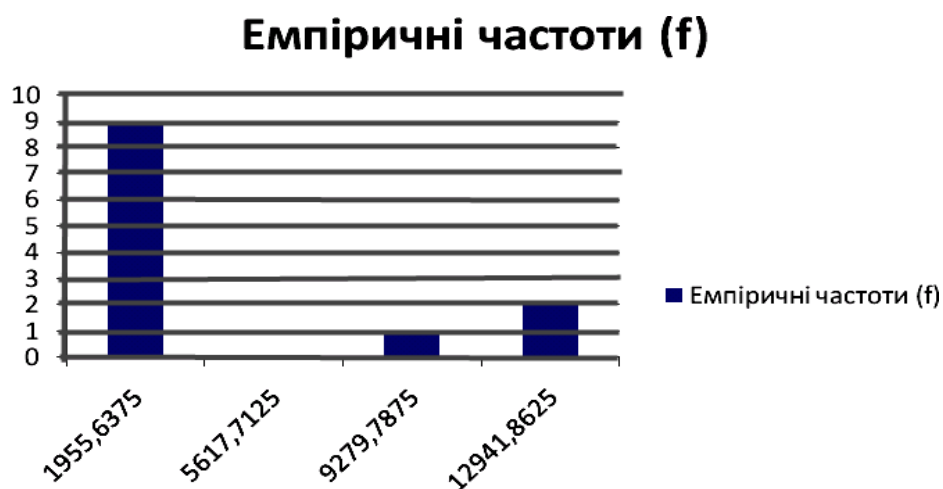


Рис. 1.23. Гістограма розподілу частот інтервального ряду у Microsoft Office Excel

Наступним кроком доцільно графічно зобразити кумуляту частот інтервального ряду. Для цього у Майстрі діаграм необхідно обрати тип діаграми "*График (График с маркерами)*", а діапазон даних визначити стовбцем, у якому знаходяться накопичені частоти інтервалів. У результаті з'явиться графік, зображений на рис. 1.24.



Рис. 1.24. Кумулята частот інтервального ряду у Excel

5. Перевірка гіпотези про нормальний закон розподілу ряду за допомогою критерію Пірсона та Колмагорова – Смірнова. Важливою умовою визначення характеру даного емпіричного ряду є побудова на базі емпіричних даних частот теоретичного нормального розподілу за наступною формулою:

$$f^* = \frac{nk}{\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, \quad (9)$$

де f^* – теоретичні частоти нормального розподілу;

n – кількість спостережень;

k – крок групування інтервалів,

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} - \text{табличне значення функції нормального розподілу } f(t), \text{ де}$$

$$t = \frac{x - \bar{x}}{\sigma} \quad (10)$$

На рис. 1.25 зображена таблиця, в якій розраховані теоретичні частоти досліджуваного ряду.

K28				=ROUND(J28*K\$24;0)
	H	I	J	K
23				
24		$n \cdot k / \text{CKB}$		9,30
25				
26	8	9	10	11
27	$(X - X_{cp})^4 \cdot f$	$t = (X - X_{cp}) / \text{CKB}$	$f(t)$	Теоретичні частоти f^*
28	319732225931016,00	-0,52	0,3485	3
29	0,00	0,26	0,3857	4
30	568412846099584,00	1,03	0,2347	2
31	10662181589727300,00	1,81	0,0775	1
32	11550326661757900,00			10
33				

Рис. 1.25. Розрахунок теоретичних частот нормального закону розподілу

Стовбець 9 даної таблиці розраховано за формулою (10), у стовбці 10 знаходяться відповідні t табличні значення функції нормального розподілу, стовбець 11 розраховується множенням 10 на постійне значення $\frac{nk}{\sigma}$ відповідно до формули (9) та наданням отриманим частотам форми цілого числа.

Для перевірки гіпотези про нормальність розподілу досліджуваного ряду найчастіше використовуються критерії Колмагорова – Смірнова (11) та критерій згоди Пірсона (12):

$$\lambda = \frac{\max |S - S^*|}{\sqrt{n}}. \quad (11)$$

$$\chi^2 = \sum \frac{(f - f^*)^2}{f^*}. \quad (12)$$

На рис. 1.26 зображено порядок розрахунку зазначених критеріїв, де стовбець 12 розраховується шляхом накопичення значень стовбця 11, $13 = 7 - 12$, $14 = (3 - 11)^2/11$, $\lambda = \text{M32/SQRT(12)}$; $\chi^2 = \text{SUM(N28:N33)}$, λ , $p(\chi^2)$ – табличні значення.

	I	J	K	L	M	N	O
23							
24	n*k/CKB		9,30				
25							
26	9	10	11	12	13	14	
27	t=(X-Xcp)/CKB	f(t)	Теоретичні частоти f*	Накопичені теоретичні частоти S*	S-S*	(f-f*)^2/f*	
28	-0,52	0,3485	3	3	6,0000	12,00	
29	0,26	0,3857	4	7	4,0000	4,00	
30	1,03	0,2347	2	9	1,0000	0,50	
31	1,81	0,0775	1	10	1,0000	1,00	
32			10	max	6,00		
33							
34							
35				λ	1,73		
36				p(λ)	0,0032		
37				χ^2	17,50		
38				p(χ^2) (k=4-1)	0,0002		
39							

Рис. 1.26. Розрахунок критеріїв Пірсона та Колмагорова – Смірнова

Згідно отриманих результатів гіпотеза про нормальність досліджуваного ряду відхиляється. Близькі до нуля ймовірності свідчать про не випадкові розходження між емпіричними та теоретичними частотами, що можна візуально побачити на рис 1.27.

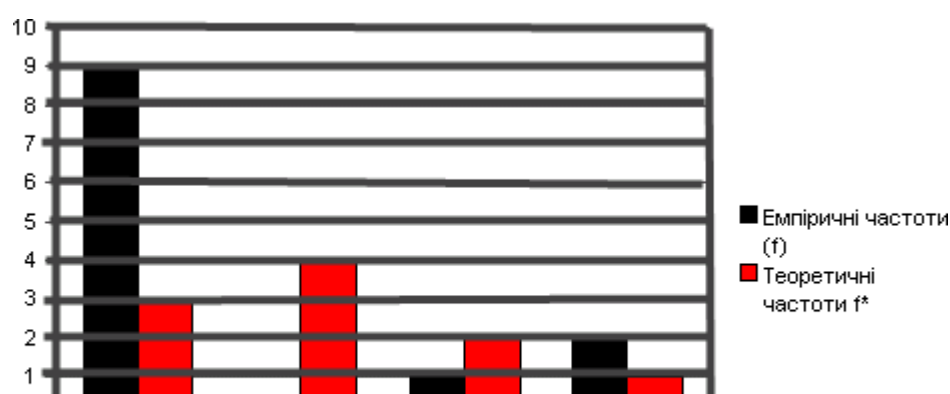


Рис. 1.27. Гістограми розподілу емпіричних та теоретичних частот досліджуваного ряду