

Сумський державний університет
Кафедра
Прикладної математики та моделювання складних систем

Звіт з практичної роботи №1

Дисципліна
Графові ймовірнісні моделі
Варіант 8

Студентка: Пороскун О. О.
Група: ПМ.м-21
Викладач: Хоменко О. В.

Суми, Сумська область
2023

Порядок виконання роботи

1. Згенеруємо вихідні дані.

Вихідні дані є вибіркою $\{x_1, x_2, \dots, x_n\}$, згенерованою за одним із законів розподілу в залежності від варіанту завдання (Додаток А, Табл. 1.2). Для генерації вихідних даних використовується функція [Генерація випадкових чисел] статистичної надбудови Microsoft Excel.

Згенеровані випадкові числа необхідно округлити до цілих і розмістити в другій стовпчик таблиці. Для цього можна скористатися математичною функцією ОКРУГЛ, що має 2 аргументи: число, що округлюється, і значення десяткового розряду, до якого його потрібно округлити. Число розрядів дорівнює 0 в разі округлення до цілого.

The screenshot shows a Microsoft Excel spreadsheet and a 'Генерация случайных чисел' (Generate Random Numbers) dialog box. The dialog box is open over the spreadsheet, which contains a table for task variants. The table has columns for '№ варіанта' (Variant number), 'Випадкова величина' (Random value), 'Параметри' (Parameters), and 'n'. Row 4 shows 'Нормальний розподіл' (Normal distribution) with parameters $\mu = 5000$ and $\sigma = 500$. Row 5 shows variant 8 with the description 'Витрати енергії в умовних одиницях при проведенні експерименту' (Energy consumption in conditional units during the experiment). The dialog box shows 'Число переменных' (Number of variables) set to 1, 'Число случайных чисел' (Number of random numbers) set to 170, and 'Распределение' (Distribution) set to 'Нормальное' (Normal). Other parameters like 'Среднее' (Mean) and 'Стандартное отклонение' (Standard deviation) are also visible. The main spreadsheet area shows rows 8 through 17 with various numerical values.

Рис. 1.1. Вибір параметрів генератора нормального розподілу

	Витрати енергії в умов. од.	Округлені значення витрат енергії
8		
9	3502.607923	3503
10	5617.618525	5618
11	5319.344053	5319
12	5054.094471	5054
13	5640.228563	5640
14	5126.270834	5126
15	4931.183083	4931
16	5432.78078	5433
17	4775.478727	4775
18	5074.188051	5074
19	4522.10942	4522
20	5540.80374	5541
21	4855.277224	4855
22	3957.055141	3957
23	6040.266397	6040
24	5055.402438	5055
25	5669.91106	5670
26	5726.022336	5726
27	4584.863417	4585
28	5597.218559	5597
29	4789.522462	4790
30	5696.365987	5696
31	5097.519433	5098
32	4611.749217	4612
33	4942.942281	4943
34	5284.546786	5285
		Аркуш1
		165 5049.556093 5050
		166 5177.155925 5177
		167 5704.219474 5704
		168 4600.217961 4600
		169 5422.743369 5423
		170 5017.27301 5017
		171 5465.004177 5465
		172 4853.361601 4853
		173 4619.050186 4619
		174 5211.857696 5212
		175 4892.754659 4893
		176 5141.256464 5141
		177 4505.91473 4506
		178 4452.419388 4452
		179

Рис. 1.2. Результати генератора розподілу та його округлені значення

2. Обчислимо основні статистичні показники трьома способами:

- За допомогою формул;
- За допомогою статистичних функцій;
- За допомогою статистичної надбудови.

Скористаємося спочатку статистичною надбудовою. Для обчислення показників варіації застосовується функція Описова статистика статистичної надбудови Microsoft Excel. У діалоговому вікні потрібно вибрати Вхідний інтервал, Мітки в першому рядку, Вихідний інтервал і Підсумкова статистика.

The screenshot shows a Microsoft Excel spreadsheet and a 'Описательная статистика' (Descriptive Statistics) dialog box. The spreadsheet has columns A through F. Column A contains row numbers 5 through 24. Column B contains the title 'Витрати енергii в умовних одиницях при проведеннi експерименту'. Column C contains values from 5000 down to 5055. Column D contains values from 500 down to 170. Row 8 contains the formula '=AVERAGE(B5:B24)' in cell B8 and the result '5000' in cell C8. Row 9 contains the formula '=STDEV(B5:B24)' in cell B9 and the result '500' in cell C9. Row 17 contains the formula '=COUNT(B5:B24)' in cell B17 and the result '170' in cell C17. The dialog box is centered over the data, with 'OK' selected.

	A	B	C	D	E	F
5	8	Витрати енергii в умовних одиницях при проведеннi експерименту	5000	500	170	
6						
7						
8	Витрати енергii в умов. од.	Округленi значення витрат енергii				
9	3502.607923	3503				
10	5617.618525	5618				
11	5319.344053	5319				
12	5054.094471	5054				
13	5640.228563	5640				
14	5126.270834	5126				
15	4931.183083	4931				
16	5432.78078	5433				
17	4775.478727	4775				
18	5074.188051	5074				
19	4522.10942	4522				
20	5540.80374	5541				
21	4855.277224	4855				
22	3957.055141	3957				
23	6040.266397	6040				
24	5055.402438	5055				

Рис. 1.3. Вибір параметрів описової статистики

	C	D	E
8		Описова статистика (з допомогою пакету Аналiз даних)	
9		<i>Округленi значення витрат енергii</i>	
10			
11	Среднее	4974.441176	
12	Стандартная ошибка	35.71324861	
13	Медiana	5027	
14	Мода	4775	
15	Стандартное отклонение	465.6437925	
16	Дисперсия выборки	216824.1415	
17	Эксцесс	0.011309245	
18	Асимметричность	-0.331411858	
19	Интервал	2537	
20	Минимум	3503	
21	Максимум	6040	
22	Сумма	845655	
23	Счет	170	
24			

Рис. 1.4. Результати описової статистики

Далі розрахуємо показники варіації, вказані в Табл 1.1 (Додаток А) за допомогою вбудованих в Microsoft Excel формул.

	F	G	H
9		Показники варіації згідно з табл. 1.1	
10			
11	Розмах варіації	2537	
12	Об'єм вибірки	170	
13	Медіана	5027	
14	Мода	4775	
15	Середнє	4974.441	
16	Середнє лінійне відхилення	370.7093	
17	Дисперсія	216824.1	
18	Середнє квадратичне відхилення	465.6438	
19	Коефіцієнт осциляції	51.0007	
20	Лінійний коефіцієнт варіації	7.452281	
21	Коефіцієнт варіації	9.360726	
22	Коефіцієнти асиметрії	-0.32559	
23	Коефіцієнт ексцесу	-0.05902	
24			

	G	H
9		
10	Показники варіації згідно з табл. 1.1	
11	Розмах варіації	=МАКС(\$B\$9:\$B\$178)-МИН(\$B\$9:\$B\$178)
12	Об'єм вибірки	=СЧЁТ(\$B\$9:\$B\$178)
13	Медіана	=МЕДИАНА(\$B\$9:\$B\$178)
14	Мода	=МОДА(\$B\$9:\$B\$178)
15	Середнє	=СРЗНАЧ(\$B\$9:\$B\$178)
16	Середнє лінійне відхилення	=СРОТКЛ(\$B\$9:\$B\$178)
17	Дисперсія	=ДИСП(\$B\$9:\$B\$178)
18	Середнє квадратичне відхилення	=СТАНДОТКЛОН(\$B\$9:\$B\$178)
19	Коефіцієнт осциляції	=\$H\$11/\$H\$15 * 100
20	Лінійний коефіцієнт варіації	=\$H\$16/\$H\$15 * 100
21	Коефіцієнт варіації	=\$H\$18/\$H\$15 * 100
22	Коефіцієнти асиметрії	=СУММ(((\$B\$9:\$B\$178 - \$H\$15)^3)/(\$H\$12*(\$H\$18^3)))
23	Коефіцієнт ексцесу	=СУММ(((\$B\$9:\$B\$178 - \$H\$15)^4)/(\$H\$12*(\$H\$18^4))) - 3
24		

Рис. 1.5. Результати показників варіації за допомогою вбудованих формул

Далі розрахуємо деякі показники варіації без застосування вбудованих формул, використаємо формули з Рис. 1.6.

Назва показників варіації	Розрахункові формули	
	Незгруповані дані	Згруповані дані
1. Розмах варіації	$R = X_{\max} - X_{\min}$	
2. Середнє лінійне відхилення	$\bar{d} = \frac{\sum x - \bar{x} }{n}$	$\bar{d} = \frac{\sum x - \bar{x} f}{\sum f}$
3. Середнє квадратичне відхилення	$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$	$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2 f}{\sum f}}$
4. Дисперсія	$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}$	$\sigma^2 = \frac{\sum (x - \bar{x})^2 f}{\sum f}$ $\sigma^2 = \bar{x}^2 - (\bar{x})^2$
5. Коефіцієнт варіації	$V_o = \frac{\sigma}{\bar{x}} \cdot 100$	
6. Лінійний коефіцієнт варіації	$V_d = \frac{\bar{d}}{\bar{x}} \cdot 100$	
7. Коефіцієнт осциляції	$V_R = \frac{R}{\bar{x}}$	

Рис. 1.6. Показники варіації та формули для їх обчислення

G	H	I	J	K	L	M
9	Показники варіації згідно з табл. 1.1					Деякі величини розраховані без використання будованих формул
10	Розмах варіації	2537				
11	Об'єм вибірки	170				
12	Медіана	5027				
13	Мода	4775				
14	Середнє	4974.441				
15	Середнє лінійне відхилення	370.7093	~	370.7093		
16	Дисперсія	216824.1	~	215548.7		
17	Середнє квадратичне відхилення	465.6438	~	464.2722		
18	Коефіцієнт осциляції	51.0007	~	51.0007		
19	Лінійний коефіцієнт варіації	7.452281	~	7.452281		
20	Коефіцієнт варіації	9.360726	~	9.333154		
21	Коефіцієнти асиметрії	-0.32559	~	-0.32848		
22	Коефіцієнт ексцесу	-0.05902	~	-0.02411		
23						
24						

G	H	I	J	K	L
9	Показники варіації згідно з табл. 1.1				Деякі величини розраховані без використання будованих формул
10	Розмах варіації	=МАКС(\$B\$9:\$B\$17)			
11	Об'єм вибірки	=СЧЁТ(\$B\$9:\$B\$17)			
12	Медіана	=МЕДИАНА(\$B\$9:\$B\$17)			
13	Мода	=МОДА(\$B\$9:\$B\$17)			
14	Середнє	=СРЗНАЧ(\$B\$9:\$B\$17)			
15	Середнє лінійне відхилення	=СРОТКЛ(\$B\$9:\$B\$17)	~	=СУММ(ABS(\$B\$9:\$B\$178 - \$H\$15)/\$H\$12)	
16	Дисперсія	=ДИСП(\$B\$9:\$B\$17)	~	=СУММ(((\$B\$9:\$B\$178-\$H\$15)^2)/\$H\$12)	
17	Середнє квадратичне відхилення	=СТАНДОТКЛОН(\$B\$9:\$B\$17)	~	=КОРЕНЬ(СУММ(((\$B\$9:\$B\$178-\$H\$15)^2)/\$H\$12))	
18	Коефіцієнт осциляції	=\$H\$11/\$H\$15 * 100	~	=\$H\$11/\$H\$15 * 100	
19	Лінійний коефіцієнт варіації	=\$H\$16/\$H\$15 * 100	~	=\$J\$16/\$H\$15 * 100	
20	Коефіцієнт варіації	=\$H\$18/\$H\$15 * 100	~	=\$J\$18/\$H\$15 * 100	
21	Коефіцієнти асиметрії	=СУММ(((\$B\$9:\$B\$178 - \$H\$15)^3)/(\$H\$12*((\$J\$18^3)))	~	=СУММ(((\$B\$9:\$B\$178 - \$H\$15)^4)/(\$H\$12*((\$J\$18^4))) - 3	
22	Коефіцієнт ексцесу	=СУММ(((\$B\$9:\$B\$178 - \$H\$15)^4)/(\$H\$12*((\$J\$18^4))) - 3	~		
23					

Рис. 1.7. Результати показників варіації без допомоги будованих формул (комірки J16 : J23)

3. Порівняймо результати розрахунків.

	D	E	F	G	H	I	J	K	L	M
9	<i>Округлені значення витрат енергії</i>			Показники варіації згідно з табл. 1.1			<i>Деякі величини розраховані без використання вбудованих формул</i>			
10										
11	Среднее	4974.441176		Розмах варіації	2537					
12	Стандартная ошибка	35.71324861		Об'єм вибірки	170					
13	Медiana	5027		Медіана	5027					
14	Мода	4775		Мода	4775					
15	Стандартное отклонение	465.6437925		Середнє	4974.441					
16	Дисперсия выборки	216824.1415		Середнє лінійне відхилення	370.7093	~	370.7093			
17	Экспесс	0.011309245		Дисперсія	216824.1	~	215548.7			
18	Асимметричность	-0.331411858		Середнє квадратичне відхилення	465.6438	~	464.2722			
19	Интервал	2537		Коефіцієнт осциляції	51.0007	~	51.0007			
20	Минимум	3503		Лінійний коефіцієнт варіації	7.452281	~	7.452281			
21	Максимум	6040		Коефіцієнт варіації	9.360726	~	9.333154			
22	Сумма	845655		Коефіцієнти асиметрії	-0.32559	~	-0.32848			
23	Счет	170		Коефіцієнт ексцесу	-0.05902	~	-0.02411			

Рис. 1.8. Результати показників варіації для 3 способів

Бачимо що всі значення або однакові або наближені, крім значення коефіцієнту ексцесу.

4. Зробіть висновок про однорідність вибірки.

5. Зробіть висновок про близькість до нормальногорозподілу.

При аналізі показників варіації можна використовувати такі правила:

- Вибірка вважається однорідною, якщо коефіцієнт варіації $V\sigma \leq 30\%$;
- Якщо коефіцієнти асиметрії та ексцесу близькі до нуля, то форму розподілу можна вважати близькою до нормальногорозподілу. Критичні значення A і E обчислюють за формулами:

$$D(A) = \frac{6(n - 1)}{(n + 1)(n + 3)} = 0,0343, \quad D(E) = \frac{24n(n - 2)(n - 3)}{(n + 1)^2(n + 3)(n + 5)} = 0,1293,$$

де $n = 170$.

Критерій згоди: $|A| \leq 3\sqrt{D(A)} \Rightarrow |A| \leq 0,5556; \quad |E| \leq 5\sqrt{D(E)} \Rightarrow |E| \leq 1,7979$

Отже, так як коефіцієнт варіації $V\sigma \approx 9,3\% (\leq 30\%)$, то вибірка вважається однорідною. А коефіцієнти асиметрії та ексцесу близькі до нуля, то форму розподілу можна вважати близькою до нормальногорозподілу.

6. Проведіть групування даних двома способами:

- За допомогою стандартних функцій Excel;
- За допомогою статистичної надбудови.

7. Побудуйте зведені таблиці і порівняйте результати групування, отримані двома способами.

Групування за допомогою статистичної надбудови

Спочатку створимо таблицю нижніх меж інтервалів групування.

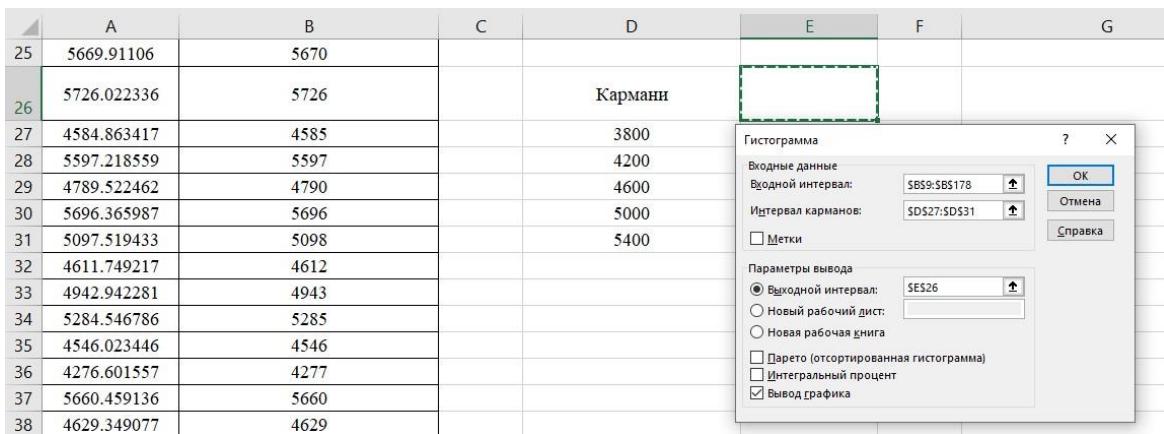


Рис.1.9 Встановлення функції Гістограма

Для групування даних за допомогою статистичної надбудови вибираємо меню [*Сервіс* → *Аналіз даних* → *Гістограма*]. Вказуємо наступні параметри:

Вхідні дані:

- Вхідний інтервал – вибірка вихідних даних;
- Інтервал кишень – нижні межі інтервалів групування.

Параметри виходу:

- Вхідний інтервал – розташування результатів групування на аркуші;
- Виведення графіка – побудова гістограмами.
- Інтегральний відсоток – обчислення накопичених частот.

Результат роботи функції Гістограма представлений на Рис. 1.10.

	D	E	F
25			
26	Кармани	Карман	Частота
27	3800	3800	2
28	4200	4200	8
29	4600	4600	24
30	5000	5000	48
31	5400	5400	54
32		Еще	34

Рис. 1.10 Результат виклику функції «Гістограма»

Згенеровану таблицю необхідно доповнити відсутніми стовпчиками. Графік необхідно настроїти для коректного відображення.

Таблицю необхідно доповнити наступними стовпцями:

- Інтервал - підписи для стовпців гістограми у вигляді текстових міток, що описують межі інтервалів групування;
- Частота n_i ;
- Відносна частота $n_i, (%)$;
- Накопичена частота $K_i, (%)$.

У графі **Всього** виконують підрахунок суми частот.

Для обчислення відсутніх характеристик інтервалів слід використовувати формули. Наприклад, для обчислення накопиченої частості для інтервалу 3800-4200 (комірка J28) використовується формула = $J27+\$I28$.

Обчислення відносної частоти для інтервалу <3800 (комірка I27) виконується за допомогою формули: = $100*\$H27/\$H\$33$.

Після обчислень слід переконатися у відсутності грубих помилок. Наприклад, накопичена частота повинна дорівнювати 100%.

Розташування стовпчиків гістограми має відповідати межам інтервалів групування даних. Для настройки графіка клацніть по стовпчику гістограми курсором і натисніть праву кнопку миші. Виберіть [*Формат рядів даних* → *Параметри*] і встановіть нульове значення параметрів Перекриття та Ширина зазору.

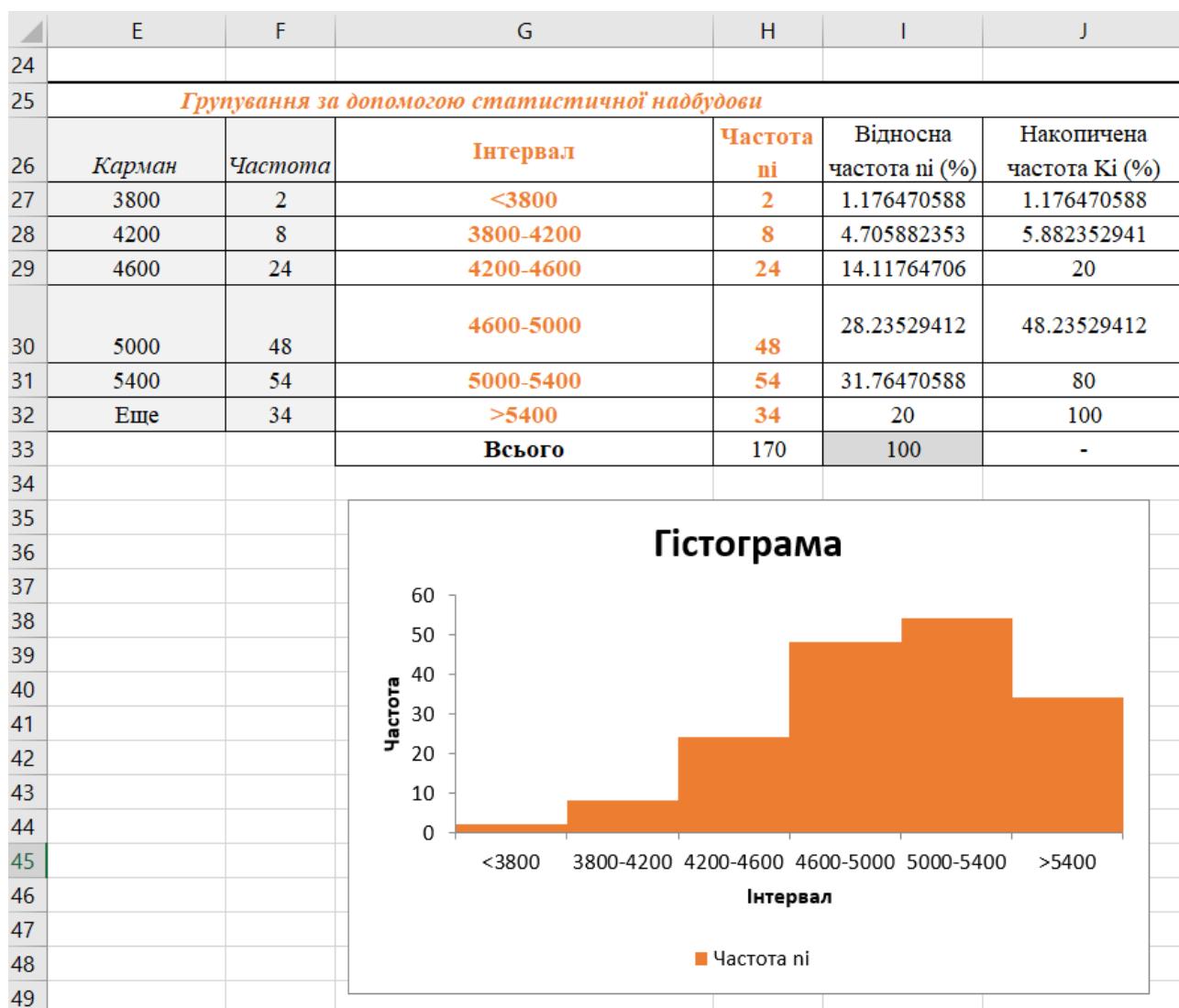


Рис. 1.11 Результат після доповнення таблиці та настройки графіка

Групування за допомогою формул

Приклад групування за допомогою формул наводиться на рис. 1.13. Підрахунок частоти потрапляння в інтервал значень визначається як різниця кількості значень менше верхньої межі і менше нижньої межі інтервалу. Наприклад, частота для першого інтервалу (комірка Q29) розрахована за допомогою функції *СЧЁТЕСЛИ*:

$$\text{СЧЁТЕСЛИ}(\text{B:B}; "<=" \& O29) - \text{СЧЁТЕСЛИ}(\text{B:B}; "<=" \& N29)$$

Відносна частина та накопичена частота розраховуються як у попередній частині про групування за допомогою статистичної надбудови.

За даною таблицею будується гістограма для цього вибираємо в меню *[Вставка → Діаграма → Гістограма → Звичайна гістограма]*. Переходимо на закладку *Ряд* і натискаємо кнопку *Додати*. Натискаємо кнопку *Значення* і вказуємо діапазон значень частот. Натискаємо кнопку *Підпису осі X* і вказуємо діапазон міток для осі X. натискаємо кнопки *Далі → Готово*. Як міток можна вказати середини інтервалів групування.



Рис. 1.12 Формули для зведені таблиці та гістограма

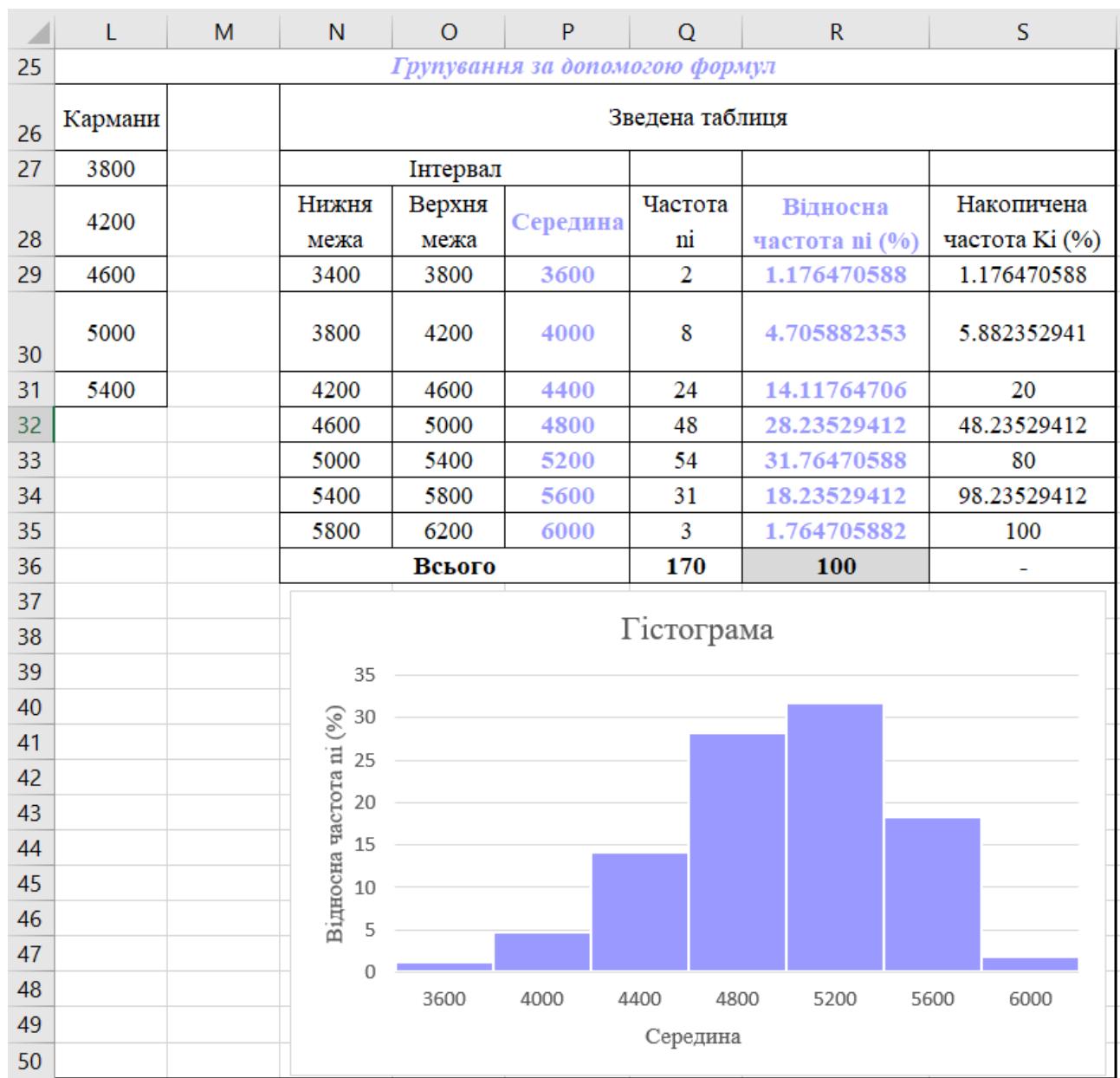


Рис. 1.13 Гістограма за результатом групування

Побудували зведені таблиці і далі порівняємо результати групування, отримані двома способами.

Поглянемо на таблиці отримані 2 способами (рис. 1.14).

	E	F	G	H	I	J
25	<i>Групування за допомогою статистичної надбудови</i>					
26	Карман	Частота	Інтервал	Частота ni	Відносна частота ni (%)	Накопичена частота Ki (%)
27	3800	2	<3800	2	1.176470588	1.176470588
28	4200	8	3800-4200	8	4.705882353	5.882352941
29	4600	24	4200-4600	24	14.11764706	20
30	5000	48	4600-5000	48	28.23529412	48.23529412
31	5400	54	5000-5400	54	31.76470588	80
32	Еще	34	>5400	34	20	100
33			Всього	170	100	-

	L	M	N	O	P	Q	R	S
25	Групування за допомогою формул							
26	Кармани		Зведена таблиця					
27	3800		Інтервал					
28	4200		Нижня межа	Верхня межа	Середина	Частота <i>n_i</i>	Відносна частота <i>n_i</i> (%)	Накопичена частота <i>K_i</i> (%)
29	4600		3400	3800	3600	2	1.176470588	1.176470588
30	5000		3800	4200	4000	8	4.705882353	5.882352941
31	5400		4200	4600	4400	24	14.11764706	20
32			4600	5000	4800	48	28.23529412	48.23529412
33			5000	5400	5200	54	31.76470588	80
34			5400	5800	5600	31	18.23529412	98.23529412
35			5800	6200	6000	3	1.764705882	100
36	Всього				170	100		-

Рис. 1.14 Гістограми за результатом групування 2 способами

Можемо побачити що відносні та накопичені частоти відповідних інтервалів співпадають.

8. Побудуйте гістограму, полігон і кумуляту двома способами:

- За допомогою статистичної надбудови
- «Вручну».

Використаємо відповідні дані отримані за допомогою статистичної надбудови та за допомогою формул відповідно до попереднього кроку.

В першому випадку кумулята – Інтегральний відсоток.

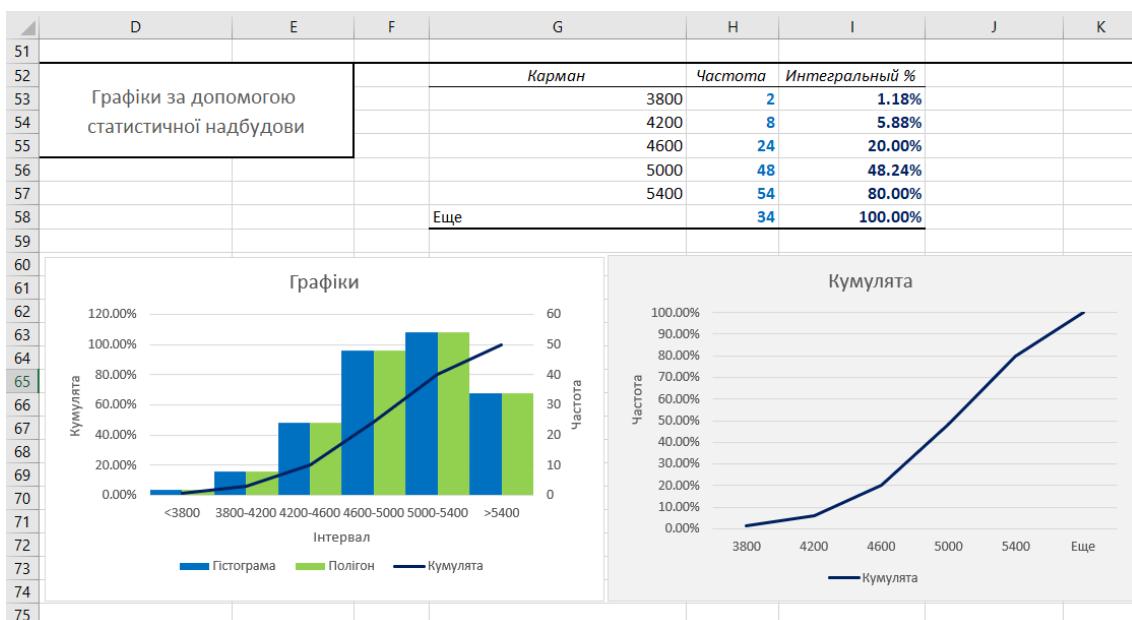


Рис. 1.15 Гістограма, полігон та кумулята за допомогою статистичної надбудови

У другому випадку кумулята - Накопичена частота, K_i (%).

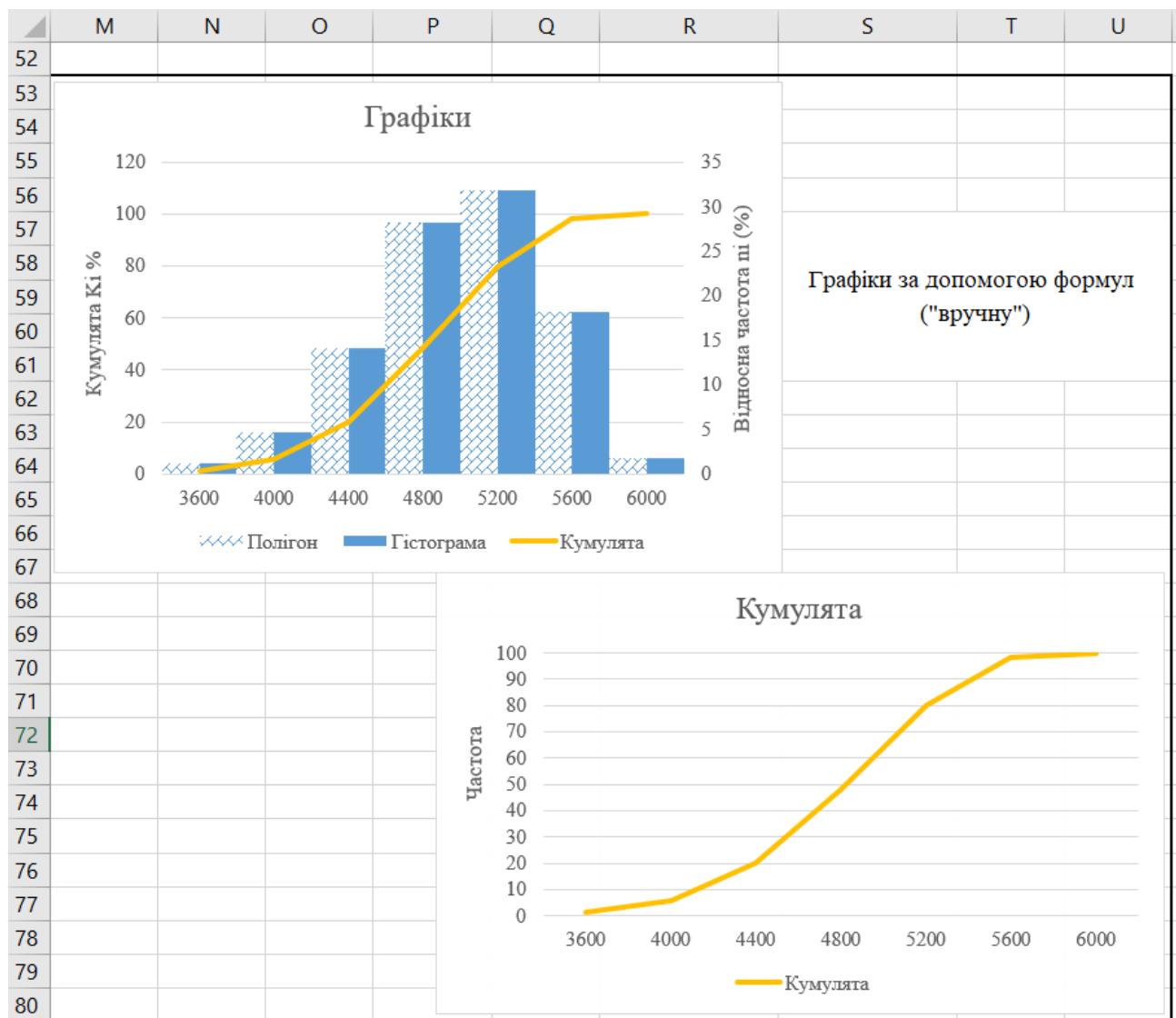


Рис. 1.16 Гістограма, полігон та кумулята за допомогою формул (вручну)

9. Побудуйте гістограми теоретичного та емпіричного розподілів на одному графіку.

Вихідна вибірка генерується за стандартним законом розподілу з параметрами відповідно до варіанта завдання. Щоб порівняти фактичний і теоретичний розподіл, необхідно побудувати їх графіки. Для роботи з теоретичними розподілами використовуються готові статистичні функції, наприклад, *NORMRASP* і *NORMOBR*.

Щоб обчислити значення теоретичної ймовірності попадання випадкової величини в інтервал $[x_1, x_2]$, необхідно знайти різницю ймовірності попадання в інтервали $[0, x_2]$ і $[0, x_1]$. Наприклад, ймовірність попадання випадкової величини в інтервал $[3400, 3800]$ для нормального закону розподілу із середнім 5000 і стандартним відхиленням 500, дорівнює:

= НОРМРАСП (3800; 5000; 500; ИСТИНА) - НОРМРАСП (3400; 5000; 500; ИСТИНА).

Побудуємо необхідні таблиці з даними для теоретичного та фактичного розподілу та графік розподілів.

T	U	V	W	X	Y	Z
25						
Теоретичний і фактичний розподіл						
27				Теоретичний розподіл	Фактичний розподіл	
28	Теоретичний розподіл нижньої межі	Теоретичний розподіл верхньої межі	Різниця теоретичних розподілів	<i>Різниця теоретичних розподілів, %</i>	<i>Відносна частота ni, %</i>	Середина інтервалу
29	0.000360832	0.005831368	0.005470536	0.547053631	1.176470588	3600
30	0.005831368	0.048139981	0.042308613	4.230861293	4.705882353	4000
31	0.048139981	0.210659069	0.162519088	16.25190879	14.11764706	4400
32	0.210659069	0.521886639	0.31122757	31.12275699	28.23529412	4800
33	0.521886639	0.819619231	0.297732592	29.77325923	31.76470588	5200
34	0.819619231	0.961880732	0.1422615	14.22615004	18.23529412	5600
35	0.961880732	0.995755387	0.033874655	3.38746555	1.764705882	6000

T	U	V	W	X	Y	Z
25						
26	Теоретичний і фактичний розподіл					
27				Теоретичний розподіл	Фактичний розподіл	
28	Теоретичний розподіл нижньої межі	Теоретичний розподіл верхньої межі	Різниця теоретичних розподілів	<i>Різниця теоретичних розподілів, %</i>	<i>Відносна частота ni, %</i>	Середина інтервалу
29	=HORM.PACP(\$N29; \$E\$11; \$E\$15; ИСТИНА)	=HORM.PACP(\$O29; \$E\$11; \$E\$15; ИСТИНА)	=\$V29 - \$U29	=W29 * 100	=SR29	=(SN29+\$O29)/2
30	=HORM.PACP(\$N30; \$E\$11; \$E\$15; ИСТИНА)	=HORM.PACP(\$O30; \$E\$11; \$E\$15; ИСТИНА)	=\$V30 - \$U30	=W30 * 100	=SR30	=(SN30+\$O30)/2
31	=HORM.PACP(\$N31; \$E\$11; \$E\$15; ИСТИНА)	=HORM.PACP(\$O31; \$E\$11; \$E\$15; ИСТИНА)	=\$V31 - \$U31	=W31 * 100	=SR31	=(SN31+\$O31)/2
32	=HORM.PACP(\$N32; \$E\$11; \$E\$15; ИСТИНА)	=HORM.PACP(\$O32; \$E\$11; \$E\$15; ИСТИНА)	=\$V32 - \$U32	=W32 * 100	=SR32	=(SN32+\$O32)/2
33	=HORM.PACP(\$N33; \$E\$11; \$E\$15; ИСТИНА)	=HORM.PACP(\$O33; \$E\$11; \$E\$15; ИСТИНА)	=\$V33 - \$U33	=W33 * 100	=SR33	=(SN33+\$O33)/2
34	=HORM.PACP(\$N34; \$E\$11; \$E\$15; ИСТИНА)	=HORM.PACP(\$O34; \$E\$11; \$E\$15; ИСТИНА)	=\$V34 - \$U34	=W34 * 100	=SR34	=(SN34+\$O34)/2
35	=HORM.PACP(\$N35; \$E\$11; \$E\$15; ИСТИНА)	=HORM.PACP(\$O35; \$E\$11; \$E\$15; ИСТИНА)	=\$V35 - \$U35	=W35 * 100	=SR35	=(SN35+\$O35)/2

Рис. 1.17 Таблиці значень для теоретичного та фактичного розподілів

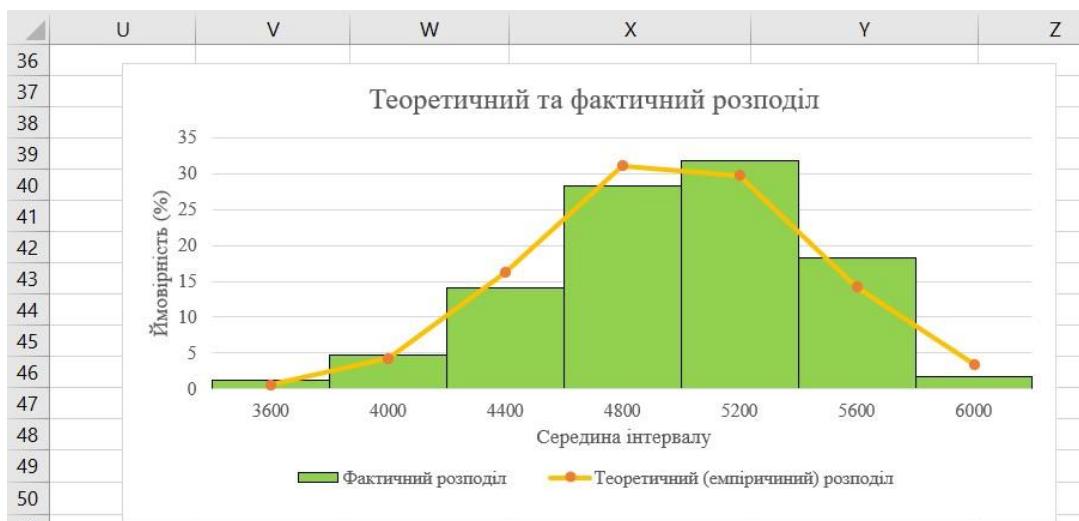


Рис. 1.18 Фактичний (емпіричний) та теоретичний розподіл

10. Зробіть висновок про близькість емпіричного розподілу до теоретичного. Емпіричний розподіл достатньо наближений до теоретичного.

Висновки

В ході виконання практичної роботи було проаналізовано статистичні дані нормального розподілу по витратам енергії в умовних одиницях при проведенні експерименту. Були розраховані основні показники варіації. Під час аналізу отриманих результатів було встановлено, що вибірка однорідна та форма розподілу близька до нормальнюї. Це підтверджується показниками коефіцієнта варіації, асиметрії та ексцесу. Також були побудовані гістограми теоретичного та емпіричного розподілів, які показують близькість даного за умовою завдання розподілу до нормального.

Контрольна задача до захисту практик 1 - 3

8 В потомстві від схрещування шести пар золотистих хом'ячків отримано 39 золотистих і 14 білих особин. Визначте частку тих і інших та розрахуйте середнє квадратичне відхилення із абсолютнох і відносних частот отриманих в потомстві особин.

Зробимо розрахунки на новому аркуші з назвою «Контрольна задача».

	A	B	C	D	E	F	G	H	I	J	K	L	
1	Контрольна задача												
2	8 В потомстві від схрещування шести пар золотистих хом'ячків отримано 39 золотистих і 14 білих особин. Визначте частку тих і інших та розрахуйте середнє квадратичне відхилення із абсолютнох і відносних частот отриманих в потомстві особин.												
3													
4													
5													
6	<i>Абсолютна частка</i>		<i>Відносна частка</i>			<i>Середнє квадратичне відхилення</i>							
7	Золотистих	Білих	Всього	золотистих	білих	Абсолютні частоти	Відносні частоти	золоті	блі	золоті	блі		
8	39	14	53	0.735849057	0.26415	0.86603	0.51887	0.11896	0.07127				
9	Кількість особин												
10													
11													

Рис. 1.19 Розраховані величини для контрольної задачі на аркуші «Контрольна задача».

	A	B	C	D	E	F	G	H	I	J	K		
1	Контрольна задача												
2	8 В потомстві від схрещування шести пар золотистих хом'ячків отримано 39 золотистих і 14 білих особин. Визначте частку тих і інших та розрахуйте середнє квадратичне відхилення із абсолютнох і відносних частот отриманих в потомстві особин.												
3													
4													
5													
6	<i>Абсолютна частка</i>		<i>Відносна частка</i>			<i>Середнє квадратичне відхилення</i>							
7	Золотистих	Білих	Всього	золотистих	білих	Абсолютні частоти	Відносні частоти	золоті	блі	золоті	блі		
8	39	14	=A8+B8	=A8/C8	=B8/C8	=КОРЕНЬ(A\$8/(\$C\$8-1))	=КОРЕНЬ(B\$8/(\$C\$8-1))	=КОРЕНЬ(D\$8/(\$C\$8-1))	=КОРЕНЬ(E\$8/(\$C\$8-1))				
9	Кількість особин												
10													

Рис. 1.20 Формули для розрахованих величин контрольної задачі

ДОДАТОК А

Таблиця 1.1

Показники варіації

Назва	Позначення	Назва у зводній таблиці	Метод обчислення	Формула Excel
1	2	3	4	5
Розмах варіації	R	Інтервал	Різниця максимального і мінімального значень	МАКС(інтервал)-МІН(інтервал)
Об'єм вибірки	n	Рахунок	Кількість статистичних одиниць	РАХУНОК(інтервал)
Медіана	Me	Медіана	Центральне значення відсортованої вибірки	МЕДІАНА(інтервал)
Мода	Mo	Мода	Найбільш часто зустрічається значення	МОДА(інтервал)
Середнє	\bar{x}	Середнє	Середнє арифметичне	СРЗНАЧ(інтервал)
Середнє лінійне відхилення	δ, d	-	Середній модуль відхилення від середнього значення	СРВІДХ(інтервал)
Дисперсія	σ^2	Дисперсія	Середній квадрат відхилення від середнього значення	ДИСП(інтервал)
Середнє квадратичне відхилення	σ	-	середнє квадратичне відхилення від середнього значення	СТАНДВІДХ(інтервал)
Середнє квадратичне відхилення(незміщена оцінка)	σ	Стандартне відхилення	середнє квадратичне відхилення від середнього значення з поправкою на обсяг вибірки	СТАНДВІДХ(інтервал)- незміщена оцінка

Закінчення таблиці 1.1

1	2	3	4	5
Коефіцієнт осциляції	V_R	-	$V_R = \frac{R}{\bar{x}} * 100$	-
Лінійний коефіцієнт варіації	V_d	-	$V_d = \frac{d}{\bar{x}} * 100$	-
Коефіцієнт варіації	V_σ	-	$V_\sigma = \frac{\sigma}{\bar{x}} * 100$	-
Коефіцієнти асиметрії	A	асиметричність	$A_s = \frac{\sum(x_i - \bar{x})^3}{n\sigma^3}$	-
Коефіцієнт ексцесу	E	ексцес	$E_x = \frac{\sum(x_i - \bar{x})^4}{n\sigma^4} - 3$	-

Варіанти завдання

Таблиця 1.2

№ варіанта	Випадкова величина	Параметри		n
	Рівномірний розподіл	min	max	
1	Зріст, см	155	190	150
2	Вага, кг	40	90	100
3	Витрати на споживання	2000	3500	170
4	Місячна зарплата, грн.	3000	7000	180
5	ВВП. трлн. дол.	2	5	180
	Нормальний розподіл	μ	σ	
6	Ціна автомобіля, тис.дол.	20	5	140
7	Число студентів у групі, чол.	31	10	190
8	Витрати енергії в умовних одиницях при проведенні експерименту	5000	500	170
9	Чисельність складових	20	4	150
10	Тривалість експерименту, хв.	240	60	140

Сумський державний університет
Кафедра
Прикладної математики та моделювання складних систем

Звіт з практичної роботи №2

Дисципліна
Графові ймовірнісні моделі
Варіант 8

Студентка: Пороскун О. О.
Група: ПМ.м-21
Викладач: Хоменко О. В.

Суми, Сумська область
2023

Порядок виконання роботи

1. Згенеруємо вихідні дані.

Для початку роботи потрібно згенерувати значення двох змінних x і y , відповідно Табл. 2.1. Обсяг вибірки – 100 елементів.

Таблиця 2.1 – Варіанти завдань

№	Фактор (x)	Результат (y)
1	Заробітна плата (грн) 3000 – 10000	Споживання (грн) $y = 1500 + 0,5 \cdot x + 500 \cdot e$
2	Дохід (грн) 3500 – 11000	Заощадження (грн) $y = -1000 + 0,5 \cdot x + 300 \cdot e$
3	Кількість студентів 2000...10000	Кількість викладачів ВНЗ $y = 220 + 0,09 \cdot x + 50 \cdot e$
4	Ціна товару (грн) 15 – 50	Попит, кг $y = 200000 \cdot x^{-0,85} + 500 \cdot e$
5	Валовий національний продукт (млрд. грн.) 1 – 8	Особисті доходи (млн. грн.) $y = -0,4 + 0,95 \cdot x + 0,4 \cdot e$
6	Витрати на рекламу (млн. грн.) 0 – 10	Прибуток(млн. грн.) $y = 10 + 6 \cdot x - 0,3 \cdot x^2 + 2 \cdot e$
7	Грошова маса (млн. грн.) 100 – 350	Індекс цін (%) $y = 38 + 0,3 \cdot x + 7 \cdot e$
8	Індекс трудовитрат (%) 100 – 160	Індекс обсягу продукції (%) $y = 7 \cdot x^{0,6} + 2 \cdot e$

Згадувана в таблиці випадкова складова e має нормальній розподіл з одиничною дисперсією і нульовим математичним очікуванням. Значення e слід згенерувати окремо, за допомогою функції «Генерація випадкових чисел» статистичної надбудови. Цей же спосіб можна використовувати для генерації значень x (тип розподілу – «рівномірний», ліва і права межа – відповідно до варіанту завдання). Отримані значення x і y доцільно округлити до того чи іншого знака після коми (або до цілого), в залежності від порядку отриманих величин (залежить від варіанту). Для округлення використовується функція ОКРУГЛ (число; число розрядів). Приклад результату генерації даних і округлення можна бачити на Рис. 2.1. У подальшій роботі використовуються тільки округлені значення x і y .

A	B	C
Варіант 8 з табл. 2.2		
8	Індекс трудовитрат (%) 100 – 160	Індекс обсягу продукції (%) $y = 7 \cdot x^{0.6} + 2 \cdot e$
Згенеровані значення		
e	<p>Генерація случайних чисел</p> <p>Число дірменних: 1</p> <p>Число случайних чисел: 100</p> <p>Распределение: Нормальне</p> <p>Параметри</p> <p>Среднее = 0</p> <p>Стандартное отклонение = 1</p> <p>Случайное рассеивание:</p> <p>Параметры вывода</p> <p><input checked="" type="radio"/> Выходной интервал: \$A\$8</p> <p><input type="radio"/> Новый рабочий лист</p> <p><input type="radio"/> Новая рабочая книга</p>	OK Отмена Справка

A	B	C	D	E
Варіант 8 з табл. 2.2				
8	Індекс трудовитрат (%) 100 – 160	Індекс обсягу продукції (%) $y = 7 \cdot x^{0.6} + 2 \cdot e$		
Згенеровані значення				
e	<p>Генерація случайних чисел</p> <p>Число дірменних: 1</p> <p>Число случайных чисел: 100</p> <p>Распределение: Равномерное</p> <p>Параметри</p> <p>Межу 100 и 160</p> <p>Случайное рассеивание:</p> <p>Параметры вывода</p> <p><input checked="" type="radio"/> Выходной интервал: \$B\$8</p> <p><input type="radio"/> Новый рабочий лист</p> <p><input type="radio"/> Новая рабочая книга</p>	OK Отмена Справка		
8	-0.14946	0.04013	0.17737	0.11281
9	-0.62412	0.17737	-1.52597	-1.19491
10	0.040134	-1.52597	0.94616	0.94616
11	0.17737	0.94616	-0.00172	-0.00172
12	0.11281	-0.00172	-0.47475	-0.47475
13	-1.52597	-0.47475	-1.24864	-1.24864
14	-1.19491	-1.24864	-0.54551	-0.54551
15	0.946163	-0.54551	0.62514	0.62514
16	-0.00172	0.62514		

A	B	C	D	E
Варіант 8 з табл. 2.2				
8	Індекс трудовитрат (%) 100 – 160	Індекс обсягу продукції (%) $y = 7 \cdot x^{0.6} + 2 \cdot e$		
Згенеровані значення		Округлені значення		
e	x	y	Індекс трудовитрат	Індекс обсягу продукції
8	-0.14946	125.5384991	126	127
9	-0.62412	116.4598529	116	120
10	0.040134	159.7289956	160	147
11	0.17737	156.5153966	157	146
12	0.112807	132.2916349	132	131
13	-1.52597	138.8323618	139	132
14	-1.19491	130.7681509	131	128
15	0.946163	110.7705924	111	120
16	-0.00172	121.8945891	122	125

A	B	C	D	E	F
97	0.84472	123.8795129	127.8424848	124	128
98	0.005087	132.526017	131.3748898	133	131
99	-0.96464	121.4697714	122.745608	121	123
100	-1.19428	141.833552	134.4364304	142	134
101	-0.91258	113.6161382	117.9486928	114	118
102	0.04312	146.2556841	139.4550827	146	139
103	0.326506	154.5469527	144.7099932	155	145
104	-0.51356	110.5563524	116.800823	111	117
105	0.37475	151.3699759	143.022303	151	143
106	0.946402	117.9082614	124.3613555	118	124
107	-0.83362	114.8155156	118.8636476	115	119
108					

A	B	C	D	E
1				
2	Варіант 8 з табл. 2.2			
3	Індекс трудовитрат	Індекс обсягу продукції (%)		
4	8	100 – 160	$y = 7 \cdot x^{0.6} + 2 \cdot e$	
5				
6	Згенеровані значення		Округлені значення	
7	e	x	y	Індекс трудовитрат
8	-0.1494595380791	125.538499099704	=7*\$B8^0.6 + 2*\$A8	=ОКРУГЛ(\$B\$8:\$B\$107; 0)
9	-0.624120275460882	116.459852900784	=7*\$B9^0.6 + 2*\$A9	=ОКРУГЛ(\$B\$8:\$B\$107; 0)
10	0.0401337274524849	159.728995635853	=7*\$B10^0.6 + 2*\$A10	=ОКРУГЛ(\$B\$8:\$B\$107; 0)
11	0.177369656739756	156.515396588031	=7*\$B11^0.6 + 2*\$A11	=ОКРУГЛ(\$B\$8:\$B\$107; 0)
12	0.112806901597651	132.291634876553	=7*\$B12^0.6 + 2*\$A12	=ОКРУГЛ(\$B\$8:\$B\$107; 0)
13	-1.52596840052865	138.832361827448	=7*\$B13^0.6 + 2*\$A13	=ОКРУГЛ(\$B\$8:\$B\$107; 0)
				=ОКРУГЛ(\$C\$8:\$C\$107; 0)

Рис. 2.1 Приклади результату генерації і округлення даних.

2. Розрахуйте коефіцієнт кореляції за допомогою надбудови і функції КОРРЕЛ.

Для обчислення коефіцієнтів кореляції можна використовувати як функцію «Кореляція» статистичної надбудови, так і функцію КОРРЕЛ (*діапазон_x*; *діапазон_y*). Отримане значення можна округлити з урахуванням числа значущих розрядів у вихідних даних. Результати розрахунку наведені на Рис.2.2

The screenshot shows a Microsoft Excel spreadsheet with two tables and a 'Correlation' dialog box.

Table 1:

	D	E	F	G	H	I	J
7	Індекс трудовитрат	Індекс обсягу продукці					
8	126	127					
9	116	120					
10	160	147					
11	157	146					
12	132	131					
13	139	132					
14	131	128					
15	111	120					
16	122	125					
17	149	140					
18	111	116					
19	102	111					

Table 2:

	D	E	F	G	H	I
7	Індекс трудовитрат	Індекс обсягу продукці				
8	126	127				
9	116	120			Столбец 1	Столбец 2
10	160	147			Столбец 1	1
11	157	146			Столбец 2	0.98127217
12	132	131				1
13	139	132			Показник	Значення
14	131	128			Регресія	0.981

Correlation Dialog Box:

Входній інтервал: \$D\$8:\$E\$107
Групування: по столбцам
 Метки в першій строкі
Параметри вивода: Вихідний інтервал: \$G\$9
 Новий робочий лист:
 Нова робоча книга

Formula in H14:

=ОКРУГЛ(КОРРЕЛ(\$D\$8:\$D\$107; \$E\$8:\$E\$107); 3)

Рис. 2.2 Приклад обчислення коефіцієнтів кореляції.

3. Зробіть висновок про тісноту зв'язку ознак.

Коефіцієнт кореляції приймає значення від -1 до +1, включно; його знак вказує на зворотній або прямий зв'язок показників. Величина коефіцієнта характеризує тісноту лінійного зв'язку (див. Табл. 2.2).

Таблиця 2.2

Оцінка тісноти лінійного зв'язку

Величина коефіцієнту кореляції	Характер зв'язку
$ r < 0,3$	Майже відсутній
$0,3 \leq r < 0,5$	Слабкий
$0,5 \leq r < 0,7$	Помірний
$0,7 \leq r < 1,0$	Сильний
$ r = 1,0$	Функціональний

Отже, коефіцієнт кореляції становить 0.981, маємо сильний характер зв'язку.

4. Розрахуйте коефіцієнти рівнянь регресії першого, другого і третього порядків за допомогою матричних функцій, функції ЛІНЕЙН і надбудови.

Модель зв'язку зазвичай будується в формі рівняння регресії. Парна регресія (зв'язок двох показників) може описуватися рівняннями:

Прямої: $\bar{y}_x = a_1 x + a_0$;

Параболи: $\bar{y}_x = a_2 x^2 + a_1 x + a_0$

Кубічного рівняння: $\bar{y}_x = a_3 x^3 + a_2 x^2 + a_1 x + a_0$

Невідомі коефіцієнти a_0, a_1, \dots, a_k можуть бути знайдені методом найменших квадратів (МНК), шляхом мінімізації суми квадратів:

$$\sum (\bar{y}_x - y_x)^2 \rightarrow \min$$

Системи рівнянь для обчислення коефіцієнтів регресії для поліномів різних ступенів виглядають наступним чином:

$$\begin{cases} \sum y = a_0 n + a_1 \sum x \\ \sum yx = a_0 \sum x + a_1 \sum x^2 \end{cases} \quad \text{для прямої};$$

$$\begin{cases} \sum y = a_0 n + a_1 \sum x + a_2 \sum x^2 \\ \sum yx = a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 \\ \sum yx^2 = a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 \end{cases} \quad \text{для параболи};$$

$$\begin{cases} \sum y = a_0 n + a_1 \sum x + a_2 \sum x^2 + a_3 \sum x^3 \\ \sum yx = a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 + a_3 \sum x^4 \\ \sum yx^2 = a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 + a_3 \sum x^5 \\ \sum yx^3 = a_0 \sum x^3 + a_1 \sum x^4 + a_2 \sum x^5 + a_3 \sum x^6 \end{cases} \quad \text{для кубічного рівняння}.$$

Загальний вигляд системи рівнянь у матричному записі: $Y = Z A$.

Для подальшої роботи доцільно обчислити проміжні значення, такі як $\sum x^2, \sum x^3$ і т.д. Використовуючи отримані суми, складаємо матриці для системи нормальних рівнянь. Наприклад, для побудови лінійного рівняння регресії будуть потрібні наступні матриці:

$$Y = \begin{pmatrix} \sum y \\ \sum yx \end{pmatrix}; \quad Z = \begin{pmatrix} n & \sum x \\ \sum x & \sum x^2 \end{pmatrix}$$

Таким чином, для знаходження значень матриці коефіцієнтів регресії A треба знайти матрицю, зворотну Z (тобто Z^{-1}), і помножити її зліва на матрицю Y .

На Рис. 2.3 показаний приклад обчислення проміжних значень, таких як $\sum x^2, \sum x^3$ і т.д. Використовується функція СУММПРОІЗВ, яка дозволяє обчислити суму попарних добутків декількох стовпців.

The screenshot shows a Microsoft Excel spreadsheet with data in columns D through L. Row 6 contains the header "Округлені значення". Row 7 contains "Індекс трудовитрат" and "Індекс обсягу продукції". Rows 8 through 15 contain data points. Row 16 is a summary row with formulas. The formulas used include the SUMMPROIZV function to calculate intermediate sums like $\sum x$, $\sum y$, etc. The final result in cell L16 is 232974002.

	D	E	F	G	H	I	J	K	L
6	Округлені значення								
7	Індекс трудовитрат	Індекс обсягу продукції							Проміжні значення
8	126	127						n	100
9	116	120		Столбец 1	Столбец 2			$\sum x$	13126
10	160	147	Столбец 1	1				$\sum y$	13023
11	157	146	Столбец 2	0.98127217	1			$\sum x^2$	1752306
12	132	131						$\sum x^3$	237794350
13	139	132	Показник	Значення				$\sum x^4$	32775249522
14	131	128	Регресія	0.981				$\sum x^*y$	1727416
15	111	120						$\sum y^*x^2$	232974002

Рис. 2.3 Приклад обчислення проміжних сум.

Для роботи з матрицями в пакеті *Excel* використовуються функції, що працюють з масивами. Матричні функції вводять в діапазон комірок, як описано нижче.

Дані функції повертають в якості результату не одне значення, а масиви чисел (діапазон комірок). Для того щоб отримати результат, виконайте наступні дії:

- оберіть діапазон комірок, в якому буде розташуватися матриця, що є результатом обчислень матричної функції;

- введіть формулу в клітинку, що є лівим верхнім кутом обраного діапазону, натиснути *Enter*;

- виділіть область осередків (обраний діапазон, де буде розрахована нова матриця, що є оберненою або є добутком матриць);

- натисніть *F2*;

- натисніть *Ctrl + Shift + Enter*.

Після введення матричних функцій (МОБР, МУМНОЖ), вони автоматично відображаються в фігурних дужках. На Рис. 2.4 наведено приклад матриць. Для знаходження оберненої матриці використовується функція МОБР (*матриця_Z*), для множення матриць – функція МУМНОЖ (*матриця_Z-1*; *матриця_Y*).

L24	:	X	✓	f _x	{=МУМНОЖ(J24:K25; G24:G25)}
	G	H	I	J	K L
17	Загальний вигляд системи рівнянь у матричному записі: $Y = Z * A$.				
18	Пряма				
19	$Y = \begin{pmatrix} \sum y \\ \sum yx \end{pmatrix}; \quad Z = \begin{pmatrix} n & \sum x \\ \sum x & \sum x^2 \end{pmatrix}$				
20	$A = Z^{-1} * Y$				
21	Y	Z	Z^{-1}		A
22	13023	100	13126	0.596281	-0.004466564
23	1727416	13126	1752306	-0.004467	3.40284E-05
24	49.75581987				
25	0.613089899				
	G	H	I	J	K L
17	Загальний вигляд системи рівнянь у матричному записі: $Y = Z * A$.				
18	Пряма				
19	$Y = \begin{pmatrix} \sum y \\ \sum yx \end{pmatrix}; \quad Z = \begin{pmatrix} n & \sum x \\ \sum x & \sum x^2 \end{pmatrix}$				
20	$A = Z^{-1} * Y$				
21	Y	Z	Z^{-1}		A
22	=\\$L\$10	=\\$L\$8	=\\$L\$9	=МОБР(H24:I25)	=МОБР(H24:I25)
23	=\\$L\$14	=\\$L\$9	=\\$L\$11	=МОБР(H24:I25)	=МОБР(H24:I25)
24	=МУМНОЖ(J24:K25; G24:G25)				
25	=МУМНОЖ(J24:K25; G24:G25)				

Рис. 2.4 Приклад роботи з матрицями для прямої

Далі, для побудови параболічного рівняння регресії будуть потрібні наступні матриці:

$$Y = \begin{pmatrix} \sum y \\ \sum yx \\ \sum yx^2 \end{pmatrix}; \quad Z = \begin{pmatrix} n & \sum x & \sum x^2 \\ \sum x & \sum x^2 & \sum x^3 \\ \sum x^2 & \sum x^3 & \sum x^4 \end{pmatrix}$$

Аналогічно до кроків для рівняння прямої, розраховуємо величини для параболи (Рис. 2.5).

K32	G	H	I	J	K	L	M	N
27	Парабола	$Y = \begin{pmatrix} \sum y \\ \sum yx \\ \sum yx^2 \end{pmatrix};$	$Z = \begin{pmatrix} n & \sum x & \sum x^2 \\ \sum x & \sum x^2 & \sum x^3 \\ \sum x^2 & \sum x^3 & \sum x^4 \end{pmatrix}$					
28								
29								
30								
31	Y	Z			Z ⁽⁻¹⁾		A	
32	13023	100	13126	1752306	48.06721961	-0.733362944	0.002750889	42.23573
33	1727416	13126	1752306	237794350	-0.733362944	0.011225927	-4.22388E-05	0.728558
34	232974002	1752306	237794350	3.2775E+10	0.002750889	-4.22388E-05	1.59411E-07	-0.00044

Рис. 2.5 Приклад роботи з матрицями для параболи

Далі, для побудови кубічного рівняння регресії будуть потрібні наступні матриці:

$$Y = \begin{pmatrix} \sum y \\ \sum yx \\ \sum yx^2 \\ \sum yx^3 \end{pmatrix}; \quad Z = \begin{pmatrix} n & \sum x & \sum x^2 & \sum x^3 \\ \sum x & \sum x^2 & \sum x^3 & \sum x^4 \\ \sum x^2 & \sum x^3 & \sum x^4 & \sum x^5 \\ \sum x^3 & \sum x^4 & \sum x^5 & \sum x^6 \end{pmatrix}$$

Далі подібним чином до попередніх рівнянь розраховуємо проміжні величини та величини для кубічного рівняння.

N11	K	L	M	N	O	P	Q
Проміжні значення							
7	n	100					
8	$\sum x$	13126	$\sum x^5$	4.583E+12			
9	$\sum y$	13023	$\sum x^6$	6.495E+14			
10	$\sum x^2$	1752306	$\sum y*x^3$	3.192E+10			
11	$\sum x^3$	237794350					
12	$\sum x^4$	32775249522					
13	$\sum x*y$	1727416					
14	$\sum y*x^2$	232974002					
15							

Рис. 2.6 Приклад обчислення проміжних сум.

P46	G	H	I	J	K	L	M	N	O	P
27	Парабола	$Y = \begin{pmatrix} \sum y \\ \sum yx \\ \sum yx^2 \end{pmatrix};$	$Z = \begin{pmatrix} n & \sum x & \sum x^2 \\ \sum x & \sum x^2 & \sum x^3 \\ \sum x^2 & \sum x^3 & \sum x^4 \end{pmatrix}$							
28										
29										
30										
31	Y	Z			Z ⁽⁻¹⁾		A			
32	13023	100	13126	1752306	48.0672196	-0.733362944	0.002750889	42.23573		
33	1727416	13126	1752306	237794350	-0.733362944	0.011225927	-4.22388E-05	0.7285577		
34	232974002	1752306	237794350	3.2775E+10	0.00275089	-4.22388E-05	1.59411E-07	-0.0004358		
35										
36	Кубічне	$Y = \begin{pmatrix} \sum y \\ \sum yx \\ \sum yx^2 \\ \sum yx^3 \end{pmatrix};$	$Z = \begin{pmatrix} n & \sum x & \sum x^2 & \sum x^3 \\ \sum x & \sum x^2 & \sum x^3 & \sum x^4 \\ \sum x^2 & \sum x^3 & \sum x^4 & \sum x^5 \\ \sum x^3 & \sum x^4 & \sum x^5 & \sum x^6 \end{pmatrix}$							
37	рівняння									
38										
39										
40										
41										
42	Y	Z			Z ⁽⁻¹⁾		A			
43	13023	100	13126	1752306	237794350	3113.993267	-72.30531146	0.5541532	-0.001401904	67.03159352
44	1727416	13126	1752306	237794350	3.2775E+10	-72.30531146	1.682024157	-0.0129143	3.27265E-05	0.149715237
45	232974002	1752306	237794350	3.2775E+10	4.5833E+12	0.554153206	-0.012914349	9.933E-05	-2.5213E-07	0.004023719
46	3.192E+10	237794350	3.2775E+10	4.5833E+12	6.495E+14	-0.001401904	3.27265E-05	-2.521E-07	6.41025E-10	-1.1338E-05

G	H	I	J	K	L	M	N	O	P
27	Парафола								
28		$Y = \begin{pmatrix} \sum y \\ \sum yx \\ \sum yx^2 \end{pmatrix};$		$Z = \begin{pmatrix} n & \sum x & \sum x^2 \\ \sum x & \sum x^2 & \sum x^3 \\ \sum x^2 & \sum x^3 & \sum x^4 \end{pmatrix}$					
29									
30									
31	Y	Z		Z'(-1)		A			
32	=L10	=L8	=L9	=L11	=МОБР(H32:J34)	=МОБР(H32:J34)	=МОБР(H32:J34)	=МУМНОЖ(K32:M34; G32:G34)	
33	=L14	=L9	=L11	=L12	=МОБР(H32:J34)	=МОБР(H32:J34)	=МОБР(H32:J34)	=МУМНОЖ(K32:M34; G32:G34)	
34	=L15	=L11	=L12	=L13	=МОБР(H32:J34)	=МОБР(H32:J34)	=МОБР(H32:J34)	=МУМНОЖ(K32:M34; G32:G34)	
35									
36	Кубічне								
37	рівняння								
38		$Y = \begin{pmatrix} \sum y \\ \sum yx \\ \sum yx^2 \\ \sum yx^3 \end{pmatrix};$		$Z = \begin{pmatrix} n & \sum x & \sum x^2 & \sum x^3 \\ \sum x & \sum x^2 & \sum x^3 & \sum x^4 \\ \sum x^2 & \sum x^3 & \sum x^4 & \sum x^5 \\ \sum x^3 & \sum x^4 & \sum x^5 & \sum x^6 \end{pmatrix}$					
39									
40									
41									
42	Y	Z		Z'(-1)		A			
43	=L10	=L8	=L9	=L11	=L12	=МОБР(H43:K46)	=МОБР(H43:K46)	=МОБР(H43:K46)	=МУМНОЖ(L43:O46; G43:G46)
44	=L14	=L9	=L11	=L12	=L13	=МОБР(H43:K46)	=МОБР(H43:K46)	=МОБР(H43:K46)	=МУМНОЖ(L43:O46; G43:G46)
45	=L15	=L11	=L12	=L13	=N9	=МОБР(H43:K46)	=МОБР(H43:K46)	=МОБР(H43:K46)	=МУМНОЖ(L43:O46; G43:G46)
46	=N11	=L12	=L13	=N9	=N10	=МОБР(H43:K46)	=МОБР(H43:K46)	=МОБР(H43:K46)	=МУМНОЖ(L43:O46; G43:G46)

Рис. 2.7 Приклад роботи з матрицями для параболи та кубічного рівняння (+ формули)

Коефіцієнти регресії можна також знайти за допомогою функції ЛІНЕЙН.

R9	...	X	✓	f _x	{=ЛІНЕЙН(\$E\$8:\$E\$107; \$D\$8:\$D\$107;;)}
R	S	T	U	V	W
7	ЛІНЕЙН				
8	m	b			
9	0.61309	49.75582			
10	y = mx + b				
11					

Рис. 2.8 Коефіцієнти регресії знайдені за допомогою функції ЛІНЕЙН

Далі використаємо надбудову «Аналіз даних» для регресії і виведемо показники.

R13	...	X	✓	f _x				
Q	R	S	T	U	V	W	X	Y
7	ЛІНЕЙН							
8	m	b						
9	0.61309	49.75582						
10	y = mx + b							
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								

Регресия

Входные данные

Входной интервал Y: \$E\$8:\$E\$107

Входной интервал X: \$D\$8:\$D\$107

Метки

Уровень надежности: 95 %

Выходной интервал: \$R\$13

Новый рабочий лист:

Новая рабочая книга

Остатки

Стандартизованные остатки

График остатков

График подбора

Нормальная вероятность

График нормальной вероятности

R	S	T	U	V	W	X	Y	Z
12	Надбудова Аналіз даних							
13	ВЫВОД ИТОГОВ							
14								
15	Регрессионная статистика							
16	Множественный R	0.981272						
17	R-квадрат	0.962895						
18	Нормированный R-квадрат	0.962516						
19	Стандартная ошибка	2.084092						
20	Наблюдения	100						
21								
22	Дисперсионный анализ							
23		df	SS	MS	F	Значимость F		
24	Регрессия	1	11046.05297	11046.1	2543.16	6.53752E-72		
25	Остаток	98	425.6570301	4.34344				
26	Итого	99	11471.71					
27								
28		Коэффициенты	Стандартная ошибка	t-статистика	P-значение	Нижние 95%	Верхние 95%	Нижние 95.0%
29	Y-пересечение	49.75582	1.609320122	30.9173	2.4E-52	46.56217647	52.9495	46.5622
30	Переменная X 1	0.61309	0.012157309	50.4297	6.5E-72	0.588964114	0.63722	0.58896
								0.63722

Рис. 2.9 Характеристики регресії знайдені за допомогою надбудови «Аналіз даних»

5. Побудуйте кореляційне поле.

6. Нанесіть на кореляційне поле лінії регресії.

При вивченні взаємозв'язків, необхідно побудувати діаграму розкиду (кореляційне поле): меню [Вставка → Діаграма]. На цій діаграмі вихідні дані (x , y) показані точками. Сюди ж наноситься лінія регресії. Для цього необхідно сформувати допоміжні стовпці x і y для кожного виду регресії.

Стовпець допоміжних значень факторної ознаки x повинен містити кілька значень з постійним кроком від мінімального до максимального. Для цього в першу комірку вводимо початкове значення, обираємо діапазон значень і викликаємо [Редагування → Заповнити → Прогресія]. При цьому потрібно обрати вид заповнення – За стовпцями, вид прогресії – Арифметична, крок та граничне значення. Кількість допоміжних проміжних значень фактора вибирають таким чином, щоб отримати на графіку гладку криву лінію.

Тип діаграми для ліній регресії – Точкова діаграма зі значеннями, з'єднаними гладкими лініями без маркерів, див. Рис. 2.11.

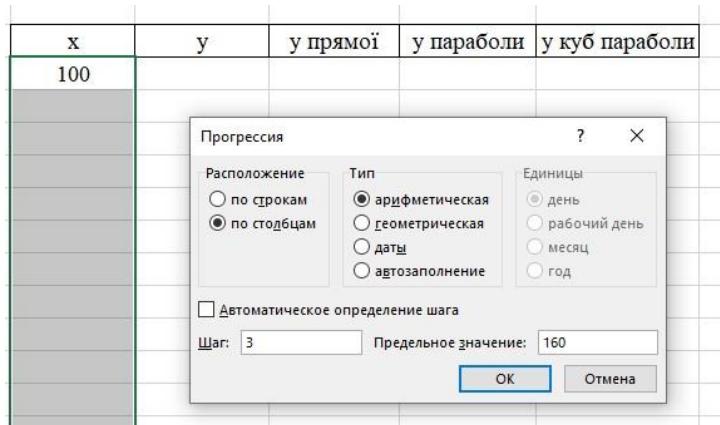
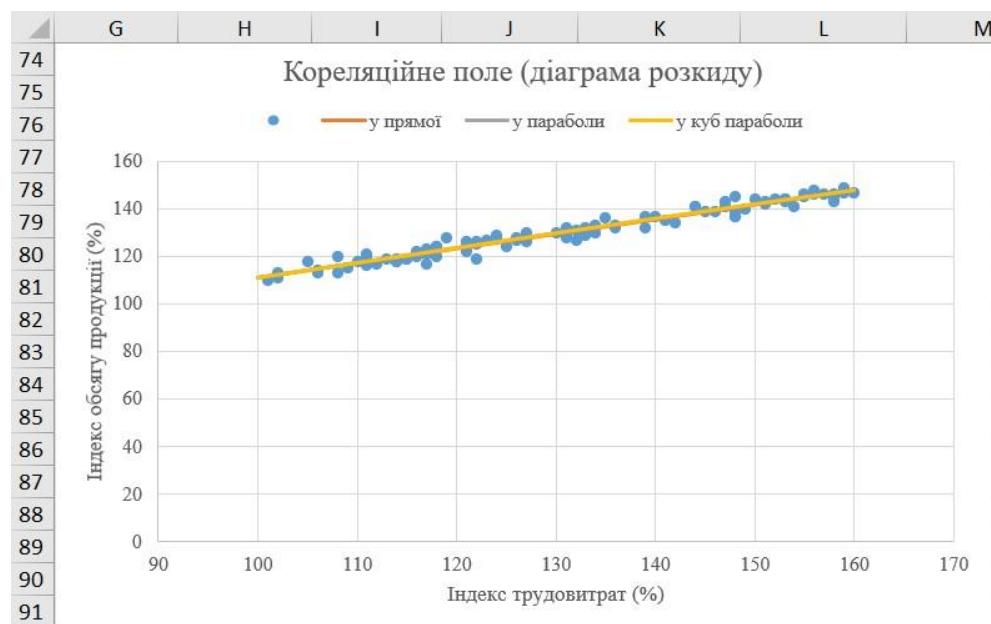


Рис. 2.10 Створення прогресії

	G	H	I	J	K
48					
49	Кореляційне поле (діаграма розкиду)				
50					
51	x		у прямої	у параболи	у куб параболи
52	100		111.06481	110.7336924	110.9023228
53	103		112.904079	112.6539749	112.7505762
54	106		114.743349	114.5664134	114.6081946
55	109		116.582619	116.4710078	116.4733414
56	112		118.421889	118.3677581	118.3441797
57	115		120.261158	120.2566644	120.2188729
58	118		122.100428	122.1377267	122.0955841
59	121		123.939698	124.0109448	123.9724767
60	124		125.778967	125.876319	125.8477138
61	127		127.618237	127.733849	127.7194587
62	130		129.457507	129.583535	129.5858747
63	133		131.296776	131.4253769	131.445125
64	136		133.136046	133.2593748	133.2953729
65	139		134.975316	135.0855286	135.1347815
66	142		136.814586	136.9038384	136.9615142
67	145		138.653855	138.7143041	138.7737341
68	148		140.493125	140.5169257	140.5696046
69	151		142.332395	142.3117033	142.3472889
70	154		144.171664	144.0986368	144.1049502
71	157		146.010934	145.8777263	145.8407518
72	160		147.850204	147.6489717	147.5528569



	G	H	I	J	K
49	Кореляційне поле (діаграма розкиду)				
50					
51	x	у прямої	у параболи	у куб параболи	
52	100	$=\$G52*\$L\$25 + \$L\$24$	$=\$N\$32 + \$N\$33 * \$G52 + \$N\$34 * (\$G52)^2$	$=\$P\$43 + \$P\$44 * \$G52 + \$P\$45 * (\$G52)^2 + \$P\$46 * (\$G52)^3$	
53	103	$=\$G53*\$L\$25 + \$L\$24$	$=\$N\$32 + \$N\$33 * \$G53 + \$N\$34 * (\$G53)^2$	$=\$P\$43 + \$P\$44 * \$G53 + \$P\$45 * (\$G53)^2 + \$P\$46 * (\$G53)^3$	
54	106	$=\$G54*\$L\$25 + \$L\$24$	$=\$N\$32 + \$N\$33 * \$G54 + \$N\$34 * (\$G54)^2$	$=\$P\$43 + \$P\$44 * \$G54 + \$P\$45 * (\$G54)^2 + \$P\$46 * (\$G54)^3$	
55	109	$=\$G55*\$L\$25 + \$L\$24$	$=\$N\$32 + \$N\$33 * \$G55 + \$N\$34 * (\$G55)^2$	$=\$P\$43 + \$P\$44 * \$G55 + \$P\$45 * (\$G55)^2 + \$P\$46 * (\$G55)^3$	
56	112	$=\$G56*\$L\$25 + \$L\$24$	$=\$N\$32 + \$N\$33 * \$G56 + \$N\$34 * (\$G56)^2$	$=\$P\$43 + \$P\$44 * \$G56 + \$P\$45 * (\$G56)^2 + \$P\$46 * (\$G56)^3$	
57	115	$=\$G57*\$L\$25 + \$L\$24$	$=\$N\$32 + \$N\$33 * \$G57 + \$N\$34 * (\$G57)^2$	$=\$P\$43 + \$P\$44 * \$G57 + \$P\$45 * (\$G57)^2 + \$P\$46 * (\$G57)^3$	

Рис. 2.11 Приклад кореляційного поля з лініями регресії та дані для нього

На графіку, що на Рис. 2.11, ми можемо бачити фактично одну криву, хоча їх насправді 3. Просто значення настільки подібні, що побачити різницю важко.

7. Нанесіть на кореляційне поле лінію емпіричної регресії.

Лінія умовного середнього (емпірична регресія)

Умовне середнє \bar{y}/x – це середнє арифметичне значень результативної ознаки у за умови, що відповідні значення факторної ознаки x потрапляють в заданий інтервал. Додайте інтервали за x , які обираються за загальними правилами групування даних (див. Лабораторну роботу №1).

Для знаходження умовного середнього можна використовувати функцію СУММЕСЛИ, яка дозволяє обчислити суму при виконанні заданої умови. Формат функції наступний:

СУММЕСЛИ (діапазон; критерій; діапазон_суммування).

Діапазон - комірки, значення яких перевіряються за допомогою умови;

Критерій - умови підсумовування, наприклад, " $<=$ " & H77;

Діапазон_суммування - комірки, значення яких складають при виконанні умови.

Отримана сума ділиться на кількість елементів, що потрапляють в діапазон. Для цього використовується функція СЧЕТЕСЛИ.

Формула для розрахунку умовного середнього може бути побудована в такий спосіб:

```
= (СУММЕСЛИ($D$8:$D$107; "<="&H77; $E$8:$E$107) - СУММЕСЛИ($D$8:$D$107; "<="&G77; $E$8:$E$107)) / (СЧЕТЕСЛИ($D$8:$D$107; "<="&H77) - СЧЕТЕСЛИ($D$8:$D$107; "<="&G77))
```

Лінія умовного середнього (емпірична регресія) наноситься на кореляційне поле, див. Рис. 2.12. Як значення x беруться середини інтервалів, точки з'єднуються прямими лініями.

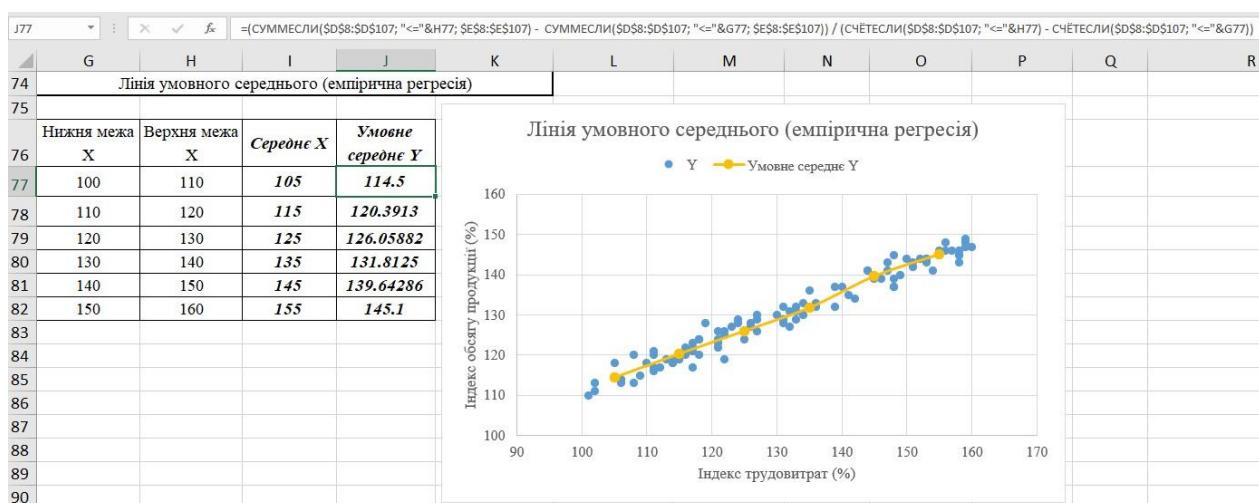


Рис. 2.12 Приклад кореляційного поля з лінією умовного середнього

9. Розрахуйте значення залишкової дисперсії для кожного рівняння регресії.

Для аналізу отриманої моделі зв'язку використовують показник залишкової дисперсії:

$$\sigma^2_{\text{зал}} = \frac{\sum(y_i - \hat{y}(x_i))^2}{n - k - 1}$$

де n – обсяг вибірки, k – число коефіцієнтів рівняння регресії.

Залишки – це різниця між фактичним значенням (Точками на графіку) і теоретичним прогнозом (лінією регресії). Облік числа коефіцієнтів k компенсує поступове наближення лінії регресії до початкових точок на кореляційному полі за рахунок підвищення порядку моделі. Рекомендується обирати рівняння регресії, що дає найменшу залишкову дисперсію.

AE17	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK
=СУММПРОІЗВ(\$E\$8:\$E\$107 - \$AB\$8:\$AB\$107; \$E\$8:\$E\$107 - \$AB\$8:\$AB\$107)) / (\$AE\$15 - \$AE\$16 - 1)										
Показник залишкової дисперсії										
6	у лін регресії	Теорет. прогноз (лінія регресії) використаємо з коефіцієнтами за формулою ЛІНЕЙН								
7	127.0051									
8	120.8742									
9	147.8502									
10	146.0109									
11	130.6837	$\sigma^2_{\text{зал}} = \frac{\sum(y_i - \hat{y}(x_i))^2}{n - k - 1}$								
12	134.9753									
13	130.0706									
14	117.8088									
15	124.5528	n	100							
16	141.1062	k	2							
17	117.8088	σ^2	4.388217							
18										

AB	AC	AD	AE
Показник залишкової дисперсії			
6	у лін регресії	Теорет. прогноз (лінія регресії) використаємо з коефіцієнтами за формулою ЛІНЕЙН	
7	= \$R\$9 * \$D8 + \$\$S9		
8	= \$R\$9 * \$D9 + \$\$S9		
9	= \$R\$9 * \$D10 + \$\$S9		
10	= \$R\$9 * \$D11 + \$\$S9		
11	= \$R\$9 * \$D12 + \$\$S9		
12	= \$R\$9 * \$D13 + \$\$S9		
13	= \$R\$9 * \$D14 + \$\$S9		
14	= \$R\$9 * \$D15 + \$\$S9		
15	= \$R\$9 * \$D16 + \$\$S9		
16	= СЧЁТ(R9:S9)		
17	= \$R\$9 * \$D17 + \$\$S9		
18	= \$R\$9 * \$D18 + \$\$S9		

Рис. 2.12 Розрахунок показника залишкової дисперсії

Отже, показник залишкової дисперсії $\sigma^2_{\text{зал}} = 4.388217$.

Контрольне питання

8. Що таке емпірична регресія?

Емпірична регресія – це лінія умовного середнього. Умовне середнє \bar{y}/x – це середнє арифметичне значень результативної ознаки у за умови, що відповідні значення факторної ознаки x потрапляють в заданий інтервал. Додаються інтервали за x , які обираються за загальними правилами групування даних.

Лінія умовного середнього наноситься на кореляційне поле. Як значення x беруться середини інтервалів, точки з'єднуються прямими лініями.

Висновки

В ході виконання практичної роботи було проаналізовано статистичні дані згенеровані за допомогою нормального та статистичного розподілів. Були розглянуті величини індексу трудовитрат та індексу обсягу продукції у відсотках. Були розраховані коефіцієнти кореляції та регресії для поліномів різних ступенів, значення умовного середнього, показник залишкової дисперсії. За допомогою коефіцієнта кореляції був встановлений сильний характер зв'язку досліджуваних величин. Також були побудовані діаграма розкиду та емпірична регресія.

Результати обчислені різними способами збігаються, це підтверджується графіками рівнянь для поліномів різних ступенів та числовими значеннями величин, зокрема показника регресії. За допомогою рівняння зв'язку можна побачити ймовірне значення величин, які відсутні у вибірці. Тобто спрогнозувати необхідну величину.

Сумський державний університет
Кафедра
Прикладної математики та моделювання складних систем

Звіт з практичної роботи №3

Дисципліна
Графові ймовірнісні моделі
Варіант 8

Студентка: Пороскун О. О.

Група: ПМ.м-21

Викладач: Хоменко О. В.

Суми, Сумська область

2023

Порядок виконання роботи

1. Згенеруйте ряди динаміки адитивної і мультиплікативної моделей згідно варіанту і побудуйте графіки.

Для знайомства з моделями динаміки згенеруйте ряди динаміки і побудуйте їх графіки відповідно до варіанту завдання, див. Табл. 3.1. Використовуйте різні листи Excel для роботи з адитивною і мультиплікативної моделями.

Адитивна модель будується за такими формулами:

$$\begin{aligned} T &= a + b * t; \\ S &= c * \sin(0.5236 * t); \\ E &= d * N(0.1); \\ y_t &= T + S + E; \end{aligned}$$

Мультиплікативна модель будується за наступними формулами:

$$\begin{aligned} T &= a + b * t; \\ S &= 1 + f * \sin(0.5236 * t); \\ E &= 1 + g * N(0; 1); \\ y_t &= T * S * E. \end{aligned}$$

Тут $N(0; 1)$ – нормальній розподіл з нульовим середнім значенням і одиничною дисперсією.

Створіть наступні стовпчики: t - час, міс., $t = 1 \dots n$;

T - тренд;

S - сезонні коливання;

E - випадкова складова;

$y(t)$ - ряд динаміки.

Варіанти параметрів моделей динаміки

Таблиця 3.1

№	n	a	b	c	d	f	g
1	60	2	0.3	3	0.5	0.3	0.05
2	70	3	0.2	2	0.5	0.5	0.07
3	80	4	0.1	2	0.5	0.4	0.06
4	90	5	0.2	4	0.6	0.3	0.06
5	100	6	0.3	3	0.4	0.3	0.06
6	110	7	0.5	4	0.8	0.2	0.05
7	80	3	0.2	4	0.7	0.1	0.03
8	90	5	0.1	5	0.5	0.2	0.05
9	80	6	0.4	6	0.4	0.5	0.04
10	70	8	0.3	3	0.6	0.7	0.07

Як видно з графіків, в адитивній моделі сезонності більша зміна амплітуди сезонних коливань, відповідно в мультиплікативній – менша.

2. Завантажте історичні дані відповідно до варіанту завдання і побудуйте графік за цінами закриття.

В якості вихідних даних використовуються денні та тижневі ряди котирувань акцій різних емітентів на фондовій біржі ММВБ.

Таблиця 3.2

Варіанти завдань

№	Ім'я файлу
1	BGDE_160101_161201.txt
2	DIXY_160101_161201.txt
3	GMKN_160101_161201.txt
4	GRNT_160101_161201.txt
5	HALS_160101_161201.txt
6	LNZL_160101_161201.txt
7	UTSY_160101_161201.txt
8	ZHIV_160101_161201.txt
9	ZILL_160101_161201.txt
10	ZVEZ_160101_161201.txt

Завантажте файл відповідно до варіанта завдання, див. Табл. 3.2. При спробі відкрити текстовий файл *.TXT через меню *Файл → Відкрити* автоматично запускається *Майстер імпорту текстових файлів MS Excel*. В процесі налаштування Майстра в нижній частині вікна можна бачити *Зразок розбору даних* і контролювати правильність імпорту.

Використовуйте наведене нижче

Крок 1 з 3: Формат вихідних даних: → з роздільниками.

Крок 2 з 3: Символом-роздільником є: → кома.

Крок 3 з 3: Детальніше → Роздільник цілої та дробової частини → Точка, див.

Рис. 3.2.

Переконайтесь, що при імпорті даних не була спотворена інформація. Для цього візуально порівняйте вихідний текстовий файл і таблицю Excel.

The screenshot shows two windows from the 'Text Import Wizard'.

Left Window: 'Мастер текстов (импорт) - шаг 3 из 3'

This window displays the 'Text Data Column Format' settings. It shows the 'General' radio button selected, with a note: 'Общий формат является наиболее универсальным. Числовые значения автоматически преобразуются в числа, даты — в даты, а все прочие значения — в текст.' Below this is a preview of the data with columns labeled: 'Общий', 'Общий', 'Общий', 'Общий', 'Общий', 'Общий', 'Общий', 'Общий', 'Общий'. The data rows show entries like '<TICKER> <PER> <DATE> <TIME> <OPEN> <HIGH> <LOW> <CLOSE> <VOL>', followed by specific values for ZHIV.

Right Window: 'Дополнительная настройка импорта текста'

This window is titled 'Настройка определения числовых данных'. It shows the 'Separator for integer and decimal parts:' dropdown set to a comma (','), and the 'Separator for thousands:' dropdown set to a space (' '). A note at the bottom states: 'Примечание. Числа отображаются в соответствии с настройкой элемента "Язык и стандарты" панели управления.' At the bottom are 'Сброс' and 'OK' buttons, with a checked checkbox for 'Знак минус в конце отрицательных чисел'.

	A	B	C	D	E	F
1	<DATE>	<OPEN>	<HIGH>	<LOW>	<CLOSE>	<VOL>
2	04.01.2016	26	26	26	26	40
3	05.01.2016	27	27	27	27	20
4	06.01.2016	27.1	27.9	27.1	27.9	20
5	13.01.2016	25.9	26.4	25.8	26	1.2
6	14.01.2016	26	26	25.8	25.8	90
7	15.01.2016	25.8	25.8	25.8	25.8	400
8	18.01.2016	25.8	26	25.8	26	350
9	20.01.2016	26.2	26.2	25.8	25.8	20
10	21.01.2016	25.6	25.9	25.6	25.9	100
11	22.01.2016	25.7	25.7	25.7	25.7	10
12	25.01.2016	25.7	26.2	25.5	25.5	400
13	26.01.2016	25.8	25.9	25.8	25.9	200
14	27.01.2016	25.6	25.6	24.8	25.4	690
15	28.01.2016	25.6	26.2	25.5	26.2	800
16	29.01.2016	26.2	26.3	26.2	26.3	320
17	01.02.2016	25.8	25.8	25.8	25.8	60
18	02.02.2016	26	26	26	26	30
19	02.02.2016	26	26	26	26	100

Рис. 3.2 – Налаштування майстра імпорту текстових файлів та результатом імпортування та редагування даних таблиці.

Після завершення імпорту видаліть малоінформативні колонки *TICKER*, *PER*, *TIME* і збережіть дані в форматі *.XLS.

3. За допомогою контекстного меню додайте лінію тренда на всі три графіка. Підберіть найбільш підходящий тип тренда.

Побудуйте графік за цінами закриття *CLOSE* в залежності від часу: Вставка → Діаграма → Тип → Точкова → Без маркерів.

Налаштуйте діаграму так, щоб оптимально використовувати все поле графіка, см. Рис. 3.3. При аналізі даних за кілька років можна використовувати скорочене значення цієї дати у форматі «місяць-рік».

Для побудови лінії тренду класніть правою кнопкою миші по лінії графіка. У контекстному меню виберіть пункт *Додати лінію тренда*. У діалоговому вікні виберіть вид апроксимуючої функції, див. Рис. 3.4.



Рис. 3.3 – Налаштування діаграми динаміки та вибір характеру тренду

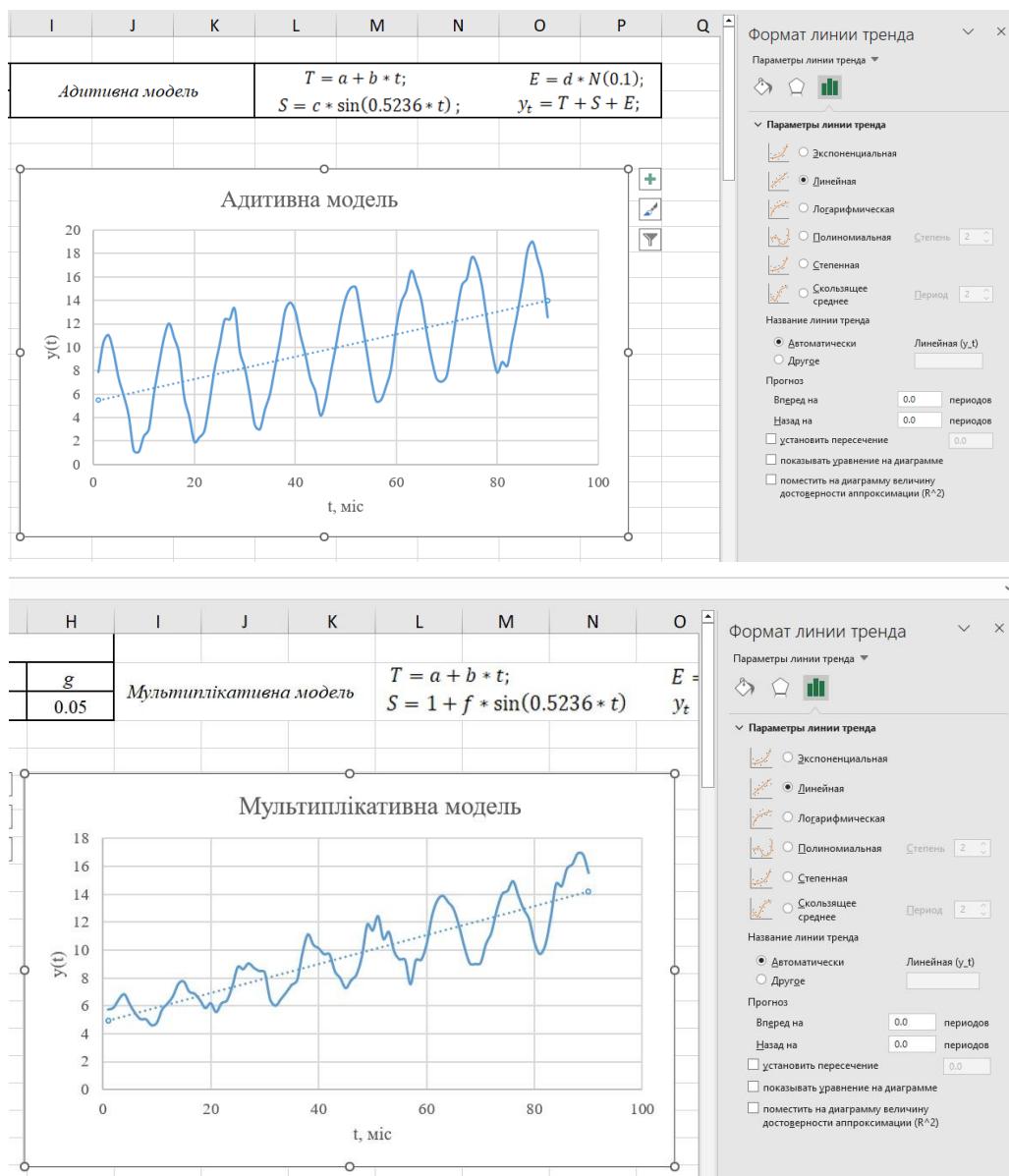


Рис. 3.4 – Налаштування діаграми динаміки та вибір характеру тренду для інших 2 графіків

4. Побудуйте рівняння тренда за допомогою функції **ЛИНЕЙН** і за допомогою статистичної надбудови і нанесіть лінію тренда на графік вихідних даних.

Рівняння тренду можна побудувати методами регресійного аналізу. При роботі з історичними даними можна створити допоміжний стовпець часу t . Побудуйте рівняння тренда за допомогою функції **ЛИНЕЙН** і за допомогою статистичної надбудови Аналіз даних, див. Опис лабораторної роботи №2. На основі рівняння тренда сформуйте стовпець лінійного прогнозу $T(t)$. Потім додайте цей ряд у вихідні дані діаграми з прив'язкою до фактичних дат (стовпець *DATE*). Виконайте наступні дії:

- оберіть діапазон комірок, в якому буде розташовуватися матриця, що є результатом обчислень матричної функції;
- введіть формулу в клітинку (J3), що є лівим верхнім кутом обраного діапазону, натиснути *Enter*;
- виділіть область осередків (обраний діапазон, де буде розрахована нова матриця, що є оберненою або є добутком матриць) (J3:K3);
- натисніть F2;
- натисніть *Ctrl + Shift + Enter*.

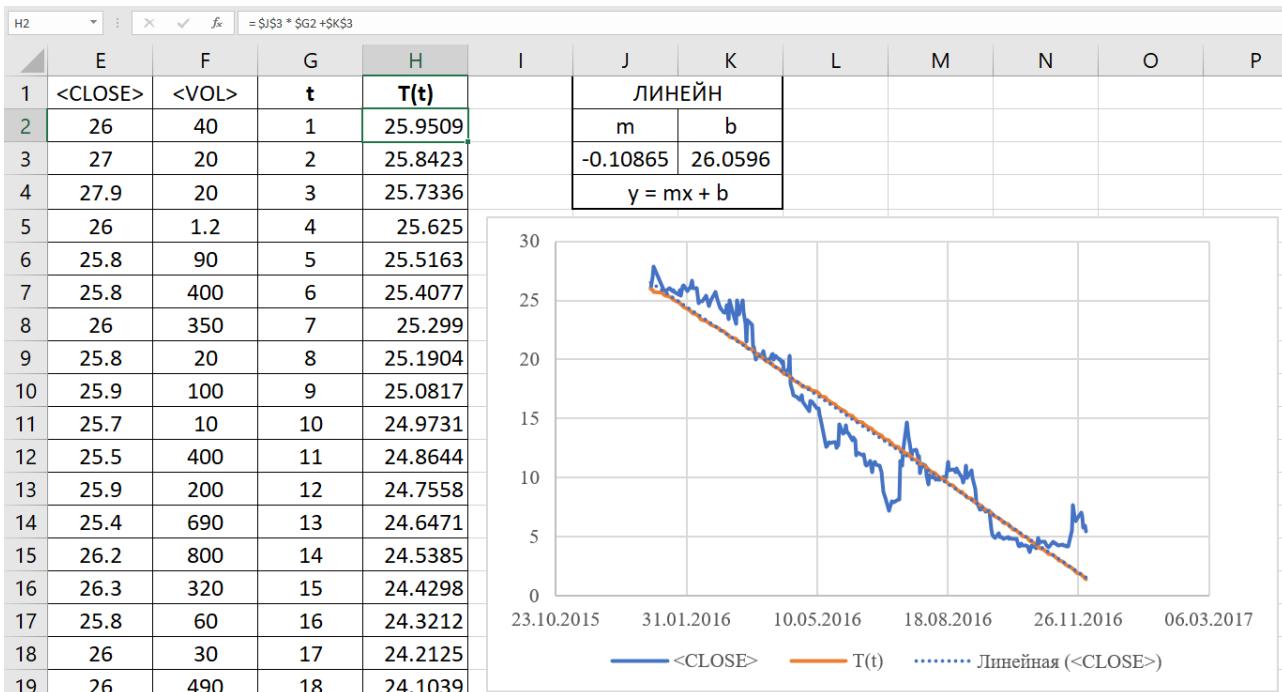
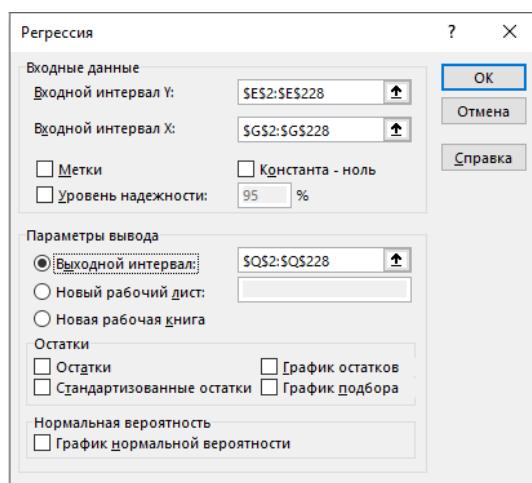


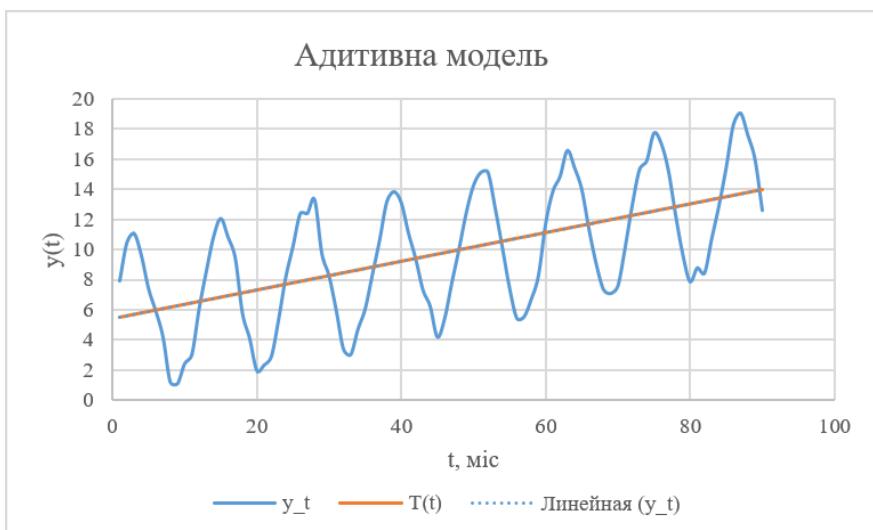
Рис. 3.5 – Побудова графіку тренду за допомогою формул ЛІНЕЙН (лінія $T(t)$)



	Q	R	S	T	U	V	W	X	Y
1									
2	ВЫВОД ИТОГОВ								
<i>Регрессионная статистика</i>									
5	Множественный R	0.958926092							
6	R-квадрат	0.919539249							
7	Нормированный R	0.919181646							
8	Стандартная ошибка	2.115383863							
9	Наблюдения	227							
10									
11	Дисперсионный анализ								
12		df	SS	MS	F	значимость F			
13	Регрессия	1	11506.60175	11506.60175	2571.3945	4E-125			
14	Остаток	225	1006.841	4.474848887					
15	Итого	226	12513.44275						
16									
17		Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95.0%	Верхние 95.0%
18	Y-пересечение	26.05957623	0.281736253	92.496354	5.07E-181	25.5044	26.6148	25.5044	26.6148
19	Переменная X 1	-0.108649965	0.00214262	-50.70891919	4.18E-125	-0.11287	-0.10443	-0.11287	-0.10443

Рис. 3.6 – Знаходження коефіцієнтів регресії за допомогою стат. надбудови Аналіз даних (Регресія)

I7											
1											
2											
3											
4											
5	t	T	S	N(0,1)	E	y_t	T(t)				
6	1	5.1	2.500005	0.586086	0.293043	7.893048	5.50064				
7	2	5.2	4.330133	1.817407	0.908703	10.43884	5.595846				
8	3	5.3	5	1.480739	0.74037	11.04037	5.691051				



G6											
1											
2											
3											
4											
5	t	T	S	N(0,1)	E	y_t	T(t)				
6	1	5.1	1.1	0.50824	1.025412	5.752562	4.943454				
7	2	5.2	1.173205	-0.69682	0.965159	5.888115	5.047459				
8	3	5.3	1.2	0.4449	1.022245	6.501478	5.151464				

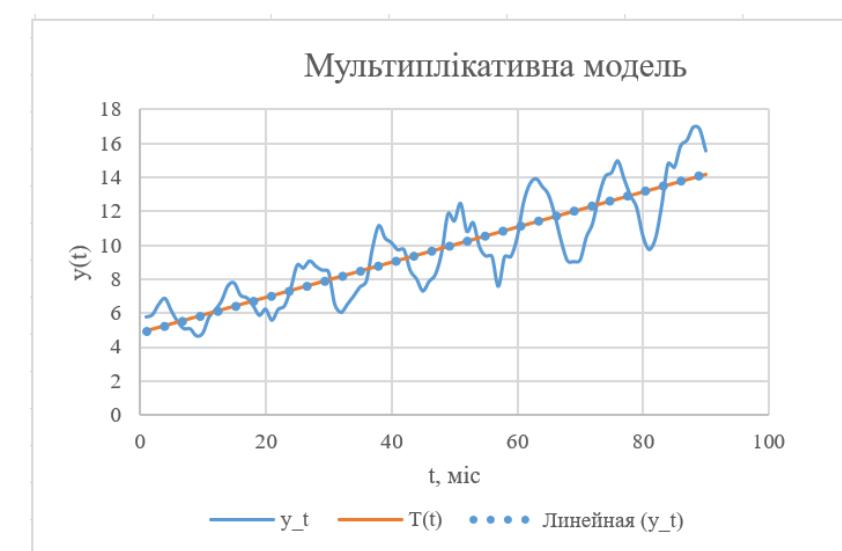


Рис. 3.7 – Побудова графіків тренду за допомогою формули ЛІНЕЙН ($T(t)$) для сезонних моделей

Як бачимо з рисунків, коефіцієнти з використанням формули ЛІНЕЙН і Статистичної надбудови Аналіз даних (Регресія) максимально наблизені тому не будемо будувати лінію регресії, бо вона просто буде співпадати з графіком $T(t)$.

5. Проведіть згладжування ряду динаміки з адитивною моделлю сезонності за допомогою простої ковзної середньої (трьома способами) і за допомогою ковзної медіани. Підберіть періоди усереднення для видалення випадкової складової і для видалення сезонності.

Для побудови простої ковзаючої середньої є кілька інструментів:

- Функція СРЗНАЧ (інтервал_згладжування_y);
- Надбудова Аналіз даних → Ковзаюче середнє;
- Контекстне меню Додати лінію тренда → Лінійна фільтрація.

При побудові ковзаючої середньої губляться кілька перших і останніх рівнів ряду. Побудуйте ковзаючу середню за допомогою функції СРЗНАЧ і віднесіть результат згладжування до середини інтервалу. Лінія тренда буде відповідати лінії регресії в традиційному розумінні.

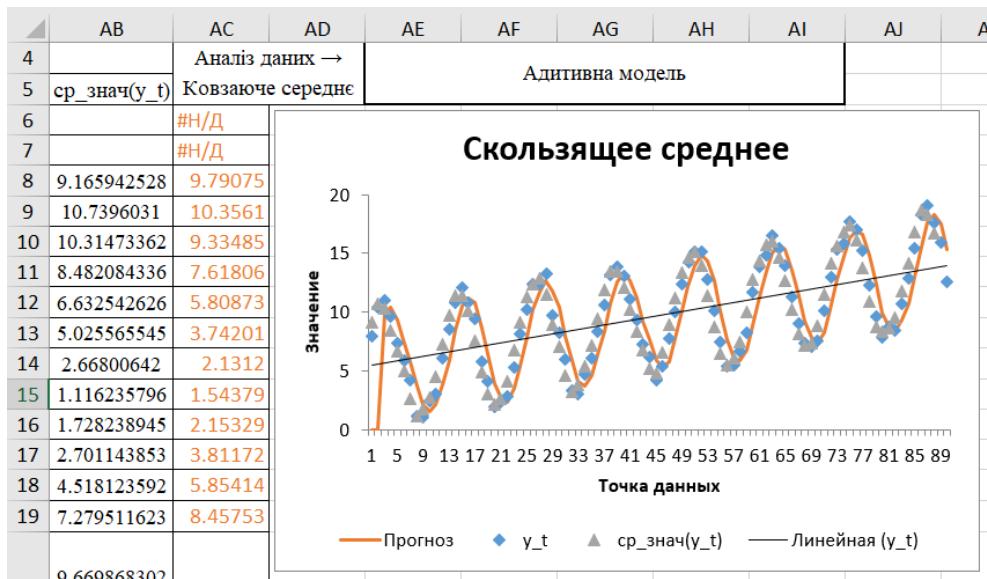


Рис. 3.9 – Лінія регресії і лінія підтримки

Побудуйте ковзаючу середню за допомогою надбудови Аналіз даних. В якості вихідного інтервалу вкажіть перший осередок вихідного стовпця. Зверніть увагу на автоматичну прив'язку до часу і діапазон невизначених згладжених значень. Нанесіть лінію на графік вихідного ряду. Підберіть інтервал згладжування так, щоб лінія «підвищуючого» тренда виступила в ролі лінії підтримки, див. Рис. 3.5. В цьому випадку локальні мінімуми ряду динаміки стосуються лінії підтримки.

При аналізі тренду замість середнього арифметичного може використовуватися медіана, див. Функцію МЕДІАНА (діапазон).

Згладжене значення \bar{y}_t обчислюється за поточним рівнем ряду y_t і попереднього згладженому рівню \bar{y}_{t-1} :

$$\bar{y}_t = \alpha * y_t + (1 - \alpha)\bar{y}_{t-1};$$

де α – параметр сгладжування, $\alpha < 1$;

$\alpha = \frac{2}{n+1}$ – еквівалентний період сгладжування ковзаючої середньої.

Якщо перетворити наведені вище співвідношення до стандартного вигляду ковзаючої середньої зваженої, то вагові коефіцієнти будуть спадати по експоненті.

Згладжені значення прив'язують до кінця інтервалу, див. Рис. 3.10.

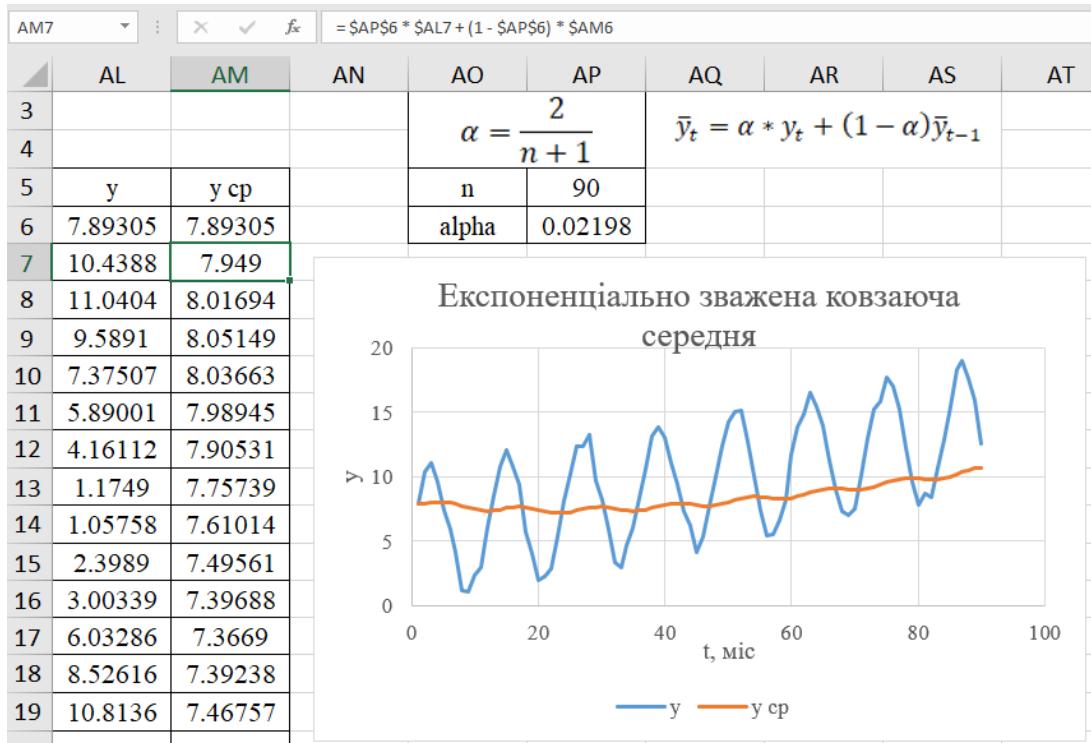


Рис. 3.10 – Експонентні вагові коефіцієнти

6. Побудуйте біржові графіки барів і свічок за різними діапазонами часу (місяць, рік).

За історичними даними про котируванням акцій в MS Excel можна побудувати два види біржових графіків: свічки і бари (штриховий графік). Для цього викликається меню *Вставка* → *Діаграма* → *Біржова*, див. Рис. 3.11.

На діаграмі японських свічок відображаються чотири значення ціни за інтервал часу:

- *OPEN* - ціна відкриття (на початок інтервалу часу);
- *HIGH* - максимальна ціна за інтервал часу;
- *LOW* - мінімальна ціна за інтервал часу;
- *CLOSE* - ціна закриття (на кінець інтервалу часу).

Якщо ціна закриття вище, ніж ціна відкриття (ринок росте), то тіло свічки біле. При падінні ціни тіло чорне, див. Рис. 3.12.

Мастер диаграмм (шаг 1 из 4): тип диаграммы



Стандартные

Нестандартные

Тип:

- Точечная
- С областями
- Кольцевая
- Лепестковая
- Поверхность
- Пузырьковая
- Биржевая

Вид:

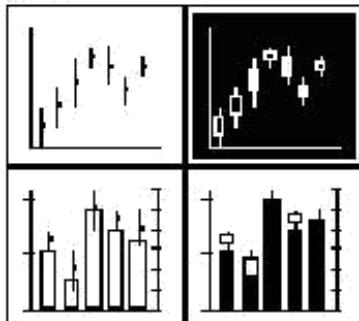


Рис. 3.11 – Вибір виду біржової діаграми

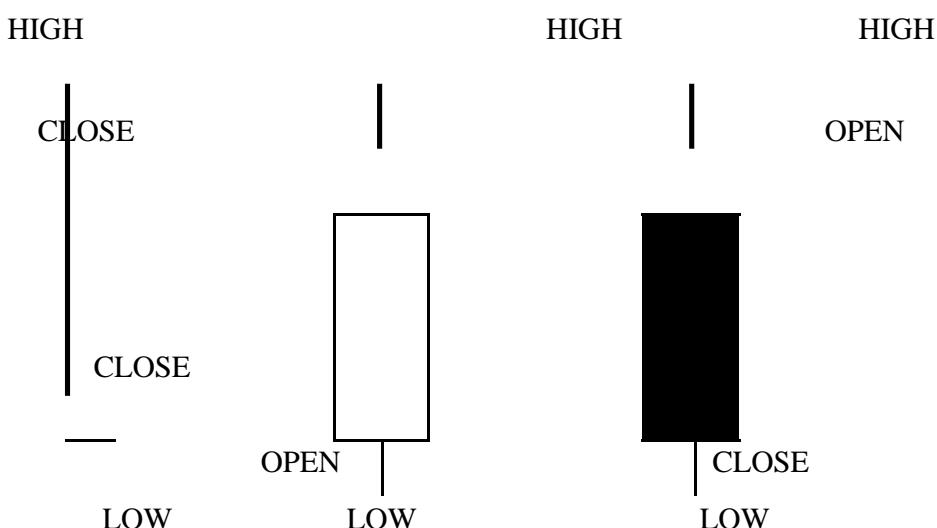


Рис. 3.12 – Біржові діаграми: бар, зростаюча свічка, падаюча свічка

На діаграмі барів MS Excel відображаються три значення ціни за інтервал часу: *HIGH*, *LOW*, *CLOSE*.

Таким чином, біржові діаграми відображають не тільки рух ціни, а й розмах значень за інтервал часу.

Після побудови діаграми можна налаштовувати товщину ліній. Виберіть різні діапазони по часу, щоб проаналізувати зміна напрямку динаміки цін.

Торгівля на фондовій біржі не проводиться по вихідних і свяtkових днях, тому на графіку з'являються «прогалини» в часі - від двох до десяти днів. Тому при аналізі можна видалити стовпець часу з вихідних даних діаграми і використовувати тільки порядковий номер свічки, див. Рис. 3.14.

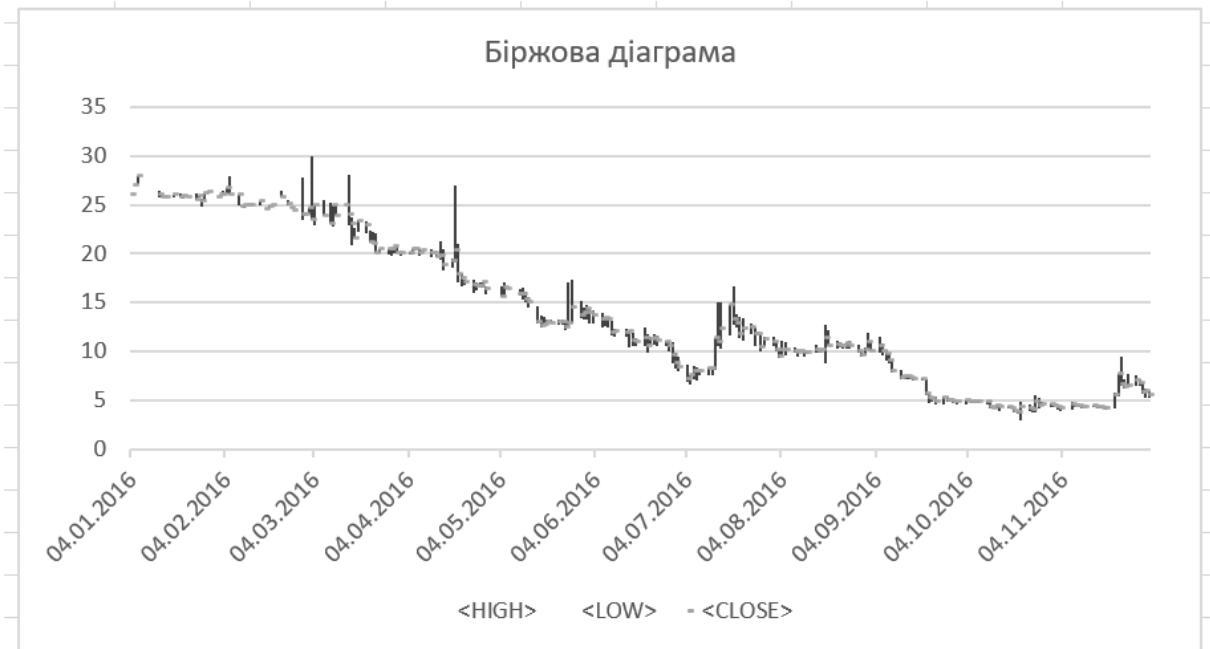


Рис. 3.13 – Діаграма барів с прив’язкою до календарних дат.

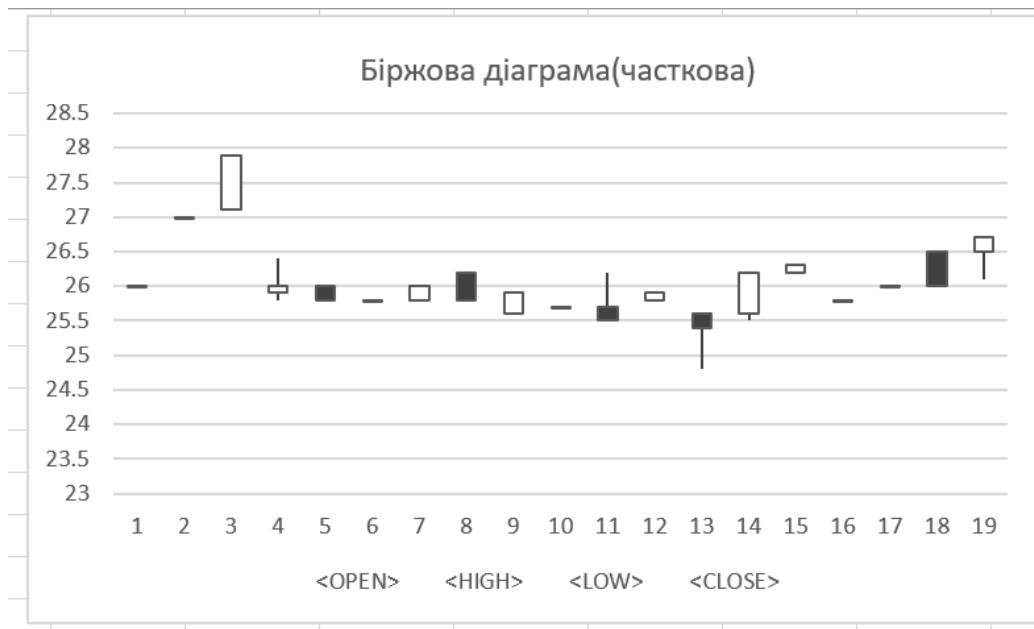


Рис. 3.14 – Діаграма японських свічок без вказівки часу

Контрольне питання

1. Що таке динаміка?
2. Які компоненти вивчають при аналізі рядів динаміки?
3. Чим відрізняються графіки адитивної і мультиплікативної моделей?
4. Що таке тренд?
5. Як побудувати рівняння тренда?

3. Чим відрізняються графіки адитивної і мультиплікативної моделей?

Дві моделі рядів динаміки (моделі сезонності), що найчастіше використовуються, подані наступними формулами:

- адитивна $y_t = T + S + E$
- мультиплікативна $y_t = T * S * E$

Адитивна модель будується за такими формулами:

$$\begin{aligned}T &= a + b * t \\S &= c * \sin(0.5236 * t) \\E &= d * N(0.1) \\y_t &= T + S + E.\end{aligned}$$

Мультиплікативна модель будується за наступними формулами:

$$\begin{aligned}T &= a + b * t \\S &= 1 + f * \sin(0.5236 * t) \\E &= 1 + g * N(0; 1) \\y_t &= T * S * E.\end{aligned}$$

Тому їхні графіки відрізняються зміною амплітуди сезонних коливань. В адитивній моделі сезонності більша зміна амплітуди сезонних коливань, відповідно в мультиплікативній – менша.

Висновки

В ході виконання практичної роботи було проаналізовано статистичні дані згенеровані за допомогою нормального розподілу та наведених формул для рядів динаміки адитивної і мультиплікативної моделей. Були побудовані графіки для цих моделей сезонності. Були розглянуті три складові компоненти ряду динаміки: T - тренд (основна тенденція); S - коливання; E - випадкова складова. Також були побудовані графіки тренду для рядів динаміки. Результати обчислені різними способами збігаються, це підтверджується графіками. . Були представлені графіки для змінної середньої за допомогою статистичних формул та статистичної надбудови «Аналіз даних» та експоненціально зважена ковзаюча середня, графіки згладженого ряду, тренда і рівняння моделей, біркові графіки.

Сумський державний університет
Кафедра
Прикладної математики та моделювання складних систем

Звіт з практичної роботи №4

Дисципліна
Графові ймовірнісні моделі
Варіант 8

Студентка: Пороскун О. О.

Група: ПМ.м-21

Викладач: Хоменко О. В.

Суми, Сумська область

2023

Порядок виконання роботи

Для даних зі свого варіанту необхідно проаналізувати чи впливає вказаний фактор на мінливість даних.

Для рослини *Abutilon* в різні години доби вивчали дані про вміст хлорофілу (у мг/дм³) (див. Табл. 1).

Таблиця 1

Години	Визначення			
	1	2	3	4
15	3,06	2,88	1,98	1,56
18	3,2	2,97	2,5	2,44
21	1,82	1,73	1,33	1,78
24	1,67	1,88	1,97	1,65
6	2,65	1,58	1,95	2,1
9	2,4	2,68	2,49	1,89
12	2,41	3,22	1,9	2,56

Для цього треба побудувати таблицю з вхідними даними та сумами рядів. Для підрахунку даних для фіксованого рівня фактору можна використати функцію *СЧЕТ* (див. Рис. 1).

Далі розрахувати суми квадратів відхилень, кількість відповідних степеней свободи та середні квадрати (див. Табл. 2).

Таблиця 2

Сума квадратів ss	Число степеней свободи df	Середній квадрат ms
$\sum_{i,j} x_{ij}^2 - \frac{T^2}{N}$	$N - 1$	$ms_0 = \frac{1}{N-1} \left(\sum_{i,j} x_{ij}^2 - \frac{T^2}{N} \right)$
$\sum_i \frac{T_i^2}{n_i} - \frac{T^2}{N}$	$a - 1$	$ms_1 = \frac{1}{a-1} \left(\sum_i \frac{T_i^2}{n_i} - \frac{T^2}{N} \right)$
$\sum_{i,j} x_{ij}^2 - \sum_i \frac{T_i^2}{n_i}$	$N - a$	$ms_2 = \frac{1}{N-a} \left(\sum_{i,j} x_{ij}^2 - \sum_i \frac{T_i^2}{n_i} \right)$

	A	B	C	D	E	F	G	H
1	Дані варіанту №8							
2	Рівень фактору	Спостереження						
3		1 (15)	2 (18)	3 (21)	4 (24)	5 (6)	6 (9)	7 (12)
4	1	3.06	3.2	1.82	1.67	2.65	2.4	2.41
5	2	2.88	2.97	1.73	1.88	1.58	2.68	3.22
6	3	1.98	2.5	1.33	1.97	1.95	2.49	1.9
7	4	1.56	2.44	1.78	1.65	2.1	1.89	2.56

	H	I	J	K	L	M	N
1							
2							
3	7 (12)	Суми за групами Ti	Кількість спостережень ni	Ti^2	Ti^2 / ni	$\sum x_{ij}^2$	T^2 / N
4	2.41	17.21	7	296.1841	42.312014	146.2319	138.3951
5	3.22	16.94	7	286.9636	40.9948		
6	1.9	14.12	7	199.3744	28.482057		
7	2.56	13.98	7	195.4404	27.920057		
8	4	62.25	28		139.70893		
9	a	T	N		$\sum (Ti^2 / ni)$		

	H	I	J	K	L	M	N
1							
2							
3	7 (12)	Суми за групами Ti	Кількість спостережень ni	Ti^2	Ti^2 / ni	$\sum x_{ij}^2$	T^2 / N
4	2.41	=СУММ(\$B4:\$H4)	=СЧЁТ(\$B4:\$H4)	=I4^2	=K4/J4	=СУММКВ(B4:H7)	=I8^2 / J8
5	3.22	=СУММ(\$B5:\$H5)	=СЧЁТ(\$B5:\$H5)	=I5^2	=K5/J5		
6	1.9	=СУММ(\$B6:\$H6)	=СЧЁТ(\$B6:\$H6)	=I6^2	=K6/J6		
7	2.56	=СУММ(\$B7:\$H7)	=СЧЁТ(\$B7:\$H7)	=I7^2	=K7/J7		
8	=СЧЁТ(\$A\$4:\$A\$7)	=СУММ(I\$4:I\$7)	=СУММ(J\$4:J\$7)		=СУММ(L4:L7)		
9	a	T	N		$\sum (Ti^2 / ni)$		

Рисунок 1 – Таблиця з вхідними даними, сумами рядів та кількістю спостережень.

Далі необхідно порівняти з табличним значенням розрахункове значення критерію Фішера:

$$F_{data} = \frac{ms_1}{ms_2}$$

Коли розрахункове значення критерію Фішера виявиться меншим табличного, то вплив досліджуваного фактору немає підстав вважати значимим. Коли ж розрахункове значення критерію Фішера виявиться більшим табличного, то цей фактор впливає на зміни середніх.

Додати до таблиці розрахункові та критичні значення критерію Фішера (див. Рис. 2) та зробити висновок щодо впливу фактора на мінливість даних.

	G	H	I	J	K	L
11	Сума квадратів <i>ss</i>	Число степеней свободи <i>df</i>	Середній квадрат <i>ms</i>	F_data	F_0,05	F_0,01
12	7.836811	27	0.290252	1.611338391	3.0087866	4.7180508
13	1.313839	3	0.437946	$F_{data} = \frac{ms_1}{ms_2}$		
14	6.522971	24	0.27179			

	G	H	I	J	K	L
11	Сума квадратів <i>ss</i>	Число степеней свободи <i>df</i>	Середній квадрат <i>ms</i>	F_data	F_0,05	F_0,01
12	= M4 - N4	= J8-1	= (1/H12) * G12	= I13 / I14	= F.ОБР.ПХ(0.05; H13; H14)	= F.ОБР.ПХ(0.01; H13; H14)
13	= L8 - N4	= H8-1	= (1/H13) * G13	$F_{data} = \frac{ms_1}{ms_2}$		
14	=M4 - L8	=J8-H8	= (1/H14) * G14			

Рисунок 2 – Вихідна таблиця дисперсійного аналізу.

Отже, розрахункове значення критерію Фішера F_{data} менше за табличні значення $F_{0,05}$ та $F_{0,01}$. Це означає, що вплив досліджуваного фактору немає підстав вважати значимим.

Контрольне питання

1. Що вивчає дисперсійний аналіз?
 2. Що таке критерій Фішера?
 3. З яких компонентів складається фактичне відхилення випадкової величини від середньої арифметичної?
 4. Що таке рівні (градації) факторів?
 5. Яка різниця між фіксованими і випадковими градаціями факторів?
 6. Що таке середній квадрат (дисперсія)?
 7. Напишіть робочі формули дисперсійного аналізу.
- 8. Які параметри оцінюють середні квадрати?*

Середні квадрати оцінюють такі параметри, як сума квадратів(ss) та число ступенів свободи(df). Тут використовуються значення суми за групами, кількість спостережень та спостереження як такі.

Висновки

В ході виконання практичної роботи були опановані методи дисперсійного аналізу для статистичного дослідження даних. Був скомпонований розподіл даних про вміст хлорофілу для рослини *Abutilon* за рівнями фактору. Були представлені такі величини як суми квадратів, число ступенів свободи та середні квадрати. Був проведений розрахунок критерію Фішера. Вказаний фактор, згідно з розрахунками, не впливає на мінливість даних.

