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

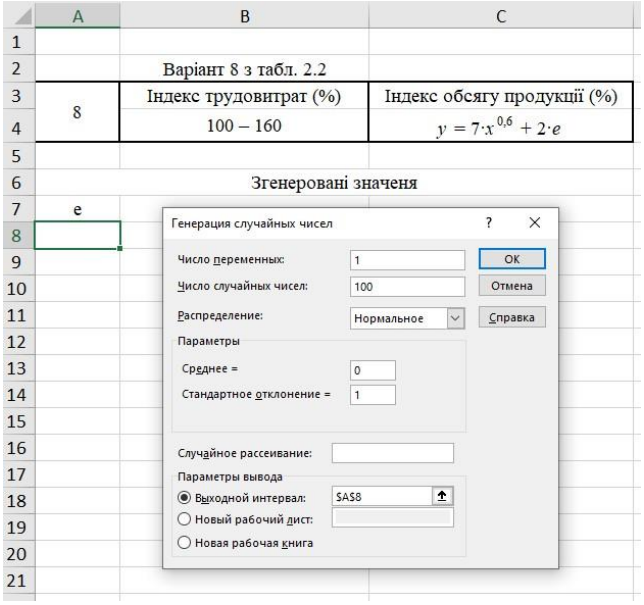
1. *Згенеруємо вихідні дані.*

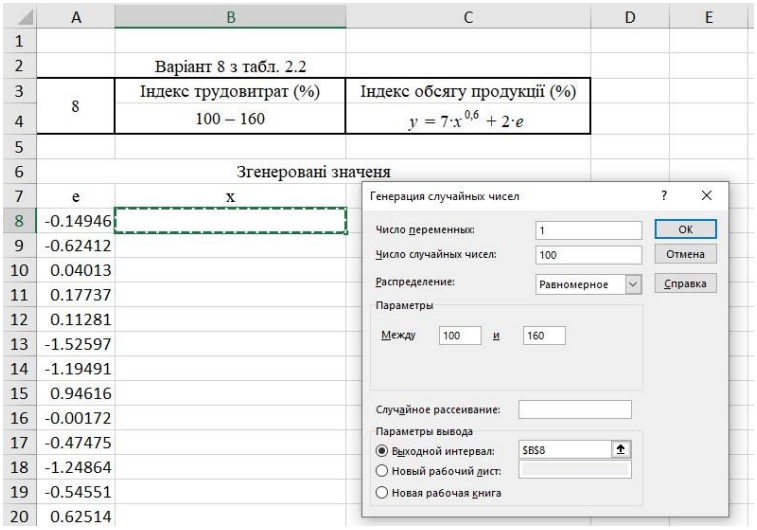
Для початку роботи потрібно згенерувати значення двох змінних x і y, відповідно Табл. 2.1. Обсяг вибірки – 100 елементів.

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 нормальний 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	D	E
1					
2	Варіант 8 з табл. 2.2				
3	8	Індекс трудовитрат (%)	Індекс обсягу продукції (%)		
4		100 – 160	$y = 7 \cdot x^{0,6} + 2 \cdot e$		
5					
6	Згенеровані значеня			Округлені значення	
7	e	x	y	Індекс трудовитрат	Індекс обсягу продукці
8	-0.14946	125.5384991	126.8650864	126	127
9	-0.62412	116.4598529	120.3154237	116	120
10	0.040134	159.7289956	147.0162957	160	147
11	0.17737	156.5153966	145.5098327	157	146
12	0.112807	132.2916349	131.4508823	132	131
13	-1.52597	138.8323618	132.0285034	139	132
14	-1.19491	130.7681509	127.9266311	131	128
15	0.946163	110.7705924	119.8572056	111	120
16	-0.00172	121.8945891	124.9328738	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 значеня			Округлені значення	
7	e	x	y	Індекс трудовитрат	Індекс обсягу продукці
8	-0.1494595380791	125.538499099704	=7*$B8^0.6 + 2*$A8	=ОКРУГЛ($B$8:$B$107; 0)	=ОКРУГЛ($C$8:$C$107; 0)
9	-0.624120275460882	116.459852900784	=7*$B9^0.6 + 2*$A9	=ОКРУГЛ($B$8:$B$107; 0)	=ОКРУГЛ($C$8:$C$107; 0)
10	0.0401337274524849	159.728995635853	=7*$B10^0.6 + 2*$A10	=ОКРУГЛ($B$8:$B$107; 0)	=ОКРУГЛ($C$8:$C$107; 0)
11	0.177369656739756	156.515396588031	=7*$B11^0.6 + 2*$A11	=ОКРУГЛ($B$8:$B$107; 0)	=ОКРУГЛ($C$8:$C$107; 0)
12	0.112806901597651	132.291634876553	=7*$B12^0.6 + 2*$A12	=ОКРУГЛ($B$8:$B$107; 0)	=ОКРУГЛ($C$8:$C$107; 0)
13	-1.52596840052865	138.832361827448	=7*$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

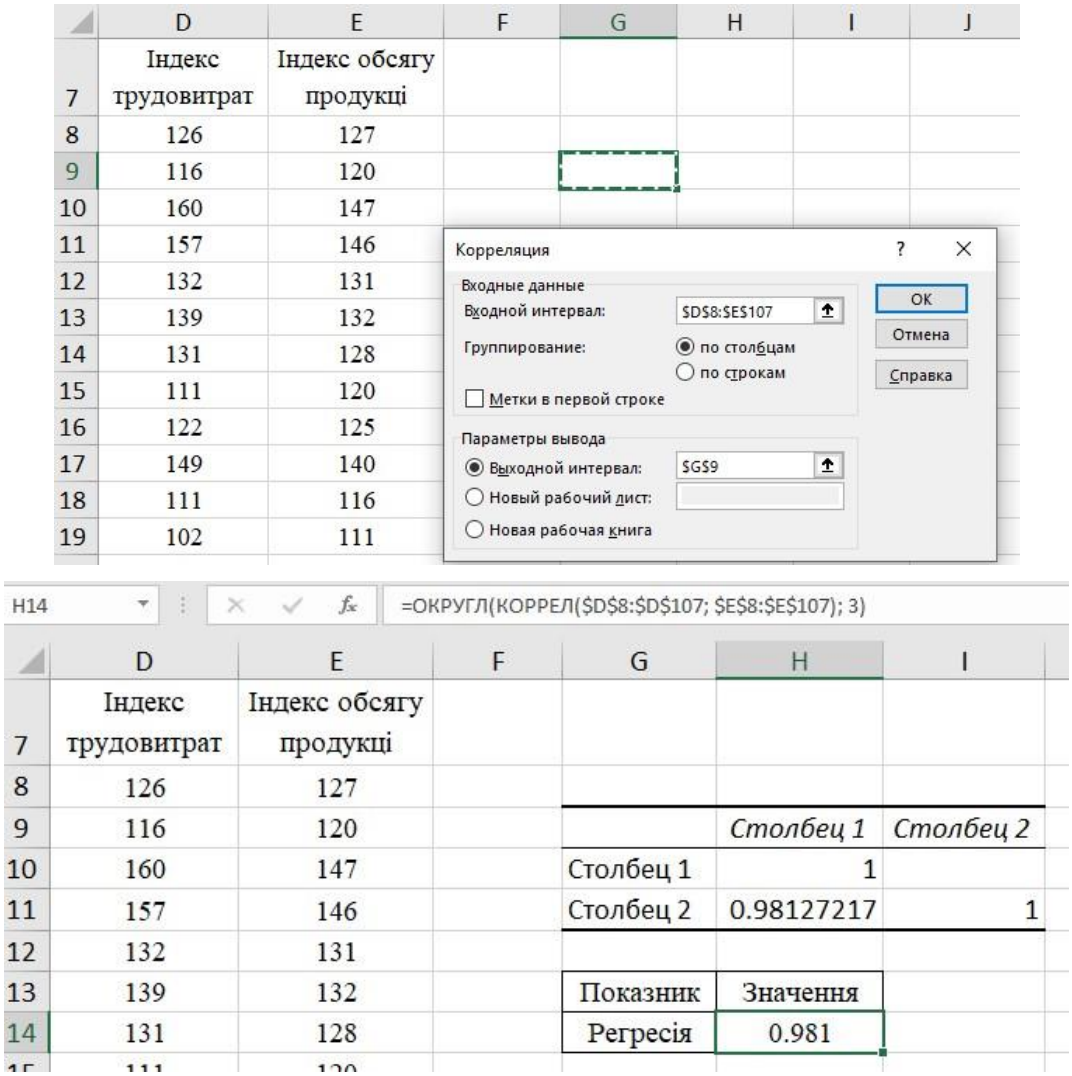


Рис. 2.2 Приклад обчислення коефіцієнтів кореляції.

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
$0{,}3 \le \|r\| < 0{,}5$	Слабкий
$0{,}5 \le \|r\| < 0{,}7$	Помірний
$0{,}7 \le \|r\| < 1{,}0$	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 ЛИНЕЙН і надбудови.*

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

Прямої: $\overline{y_x} = a_1 x + a_0$;

Параболи: $\overline{y_x} = a_2 x^2 + a_1 x + a_0$

Кубічного рівняння: $\overline{y_x} = a_3 x^3 + a_2 x^2 + a_1 x + a_0$

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

$$\sum (\overline{y_x} - y_x)^2 \to \min$$

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

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

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

$$\begin{cases} \sum y = a_0 n + a_1 \sum x + a_2 \sum x^2 + a_3 \sum x^3 \\ \sum yx = a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 + a_3 \sum x^4 \\ \sum yx^2 = a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 + a_3 \sum x^5 \\ \sum yx^3 = a_0 \sum x^3 + a_1 \sum x^4 + a_2 \sum x^5 + a_3 \sum x^6 \end{cases}$$ 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:

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

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 СУММПРОИЗВ, яка дозволяє обчислити суму попарних добутків декількох стовпців.

L15 =СУММПРОИЗВ($E$8:$E$107; $D$8:$D$107; $D$8:$D$107)

	D	E	F	G	H	I	J	K	L
6	Округлені значення								
7	Індекс трудовитрат	Індекс обсягу продукці						Проміжні значення	
8	126	127						n	100
9	116	120			*Столбец 1*	*Столбец 2*		∑ x	13126
10	160	147		Столбец 1	1			∑ y	13023
11	157	146		Столбец 2	0.98127217	1		∑ x^2	1752306
12	132	131						∑ x^3	237794350
13	139	132		Показник	Значення			∑ x^4	32775249522
14	131	128		Регресія	0.981			∑ x*y	1727416
15	111	120						∑ 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 розташовуватися 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 що 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 {=МУМНОЖ(J24:K25; G24:G25)}

	G	H	I	J	K	L
17–18	Загальний вигляд системи рівнянь у матричному записі: Y = Z * A.					
19	Пряма					
20–22	$Y = \begin{pmatrix} \sum y \\ \sum yx \end{pmatrix}$; $Z = \begin{pmatrix} n & \sum x \\ \sum x & \sum x^2 \end{pmatrix}$					***A = Z^(-1)*Y***
23	Y	Z		Z^(-1)		***A***
24	13023	100	13126	0.596281	-0.004466564	***49.75581987***
25	1727416	13126	1752306	-0.004467	3.40284E-05	***0.613089899***

	G	H	I	J	K	L
17–18	Загальний вигляд системи рівнянь у матричному записі: Y = Z * A.					
19	Пряма					
20–22	$Y = \begin{pmatrix} \sum y \\ \sum yx \end{pmatrix}$; $Z = \begin{pmatrix} n & \sum x \\ \sum x & \sum x^2 \end{pmatrix}$					***A = Z^(-1)*Y***
23	Y	Z		Z^(-1)		***A***
24	=$L$10	= $L$8	= $L$9	=МОБР(H24:I25)	=МОБР(H24:I25)	***=МУМНОЖ(J24:K25; G24:G25)***
25	=$L$14	= $L$9	= $L$11	=МОБР(H24:I25)	=МОБР(H24:I25)	***=МУМНОЖ(J24:K25; G24:G25)***

Рис. 2.4 Приклад роботи з матрицями для прямої

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

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

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

K32 {=МОБР(H32:J34)}

	G	H	I	J	K	L	M	N
27	Парабола							
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

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

Рис. 2.5 Приклад роботи з матрицями для параболи

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

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

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

N11 =СУММПРОИЗВ($E$8:$E$107; $D$8:$D$107; $D$8:$D$107; $D$8:$D$107)

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

Рис. 2.6 Приклад обчислення проміжних сум.

P46 {=МУМНОЖ(L43:O46; G43:G46)}

	G	H	I	J	K	L	M	N	O	P
27	Парабола									
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	-0.73336294	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									
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	3.1925E+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	Парабола									
28										
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										
39										
40										
41										
42	Y	Z				Z^(-1)				A
43	=L10	=L8	=L9	=L11	=L12	=МОБР(H43:K46)	=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)	=МОБР(H43:K46)	=МУМНОЖ(L43:O46; G43:G46)
45	=L15	=L11	=L12	=L13	=N9	=МОБР(H43:K46)	=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)	=МОБР(H43:K46)	=МУМНОЖ(L43:O46; G43:G46)

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

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

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

Коефіцієнти регресії можна також знайти за допомогою функції ЛИНЕЙН.

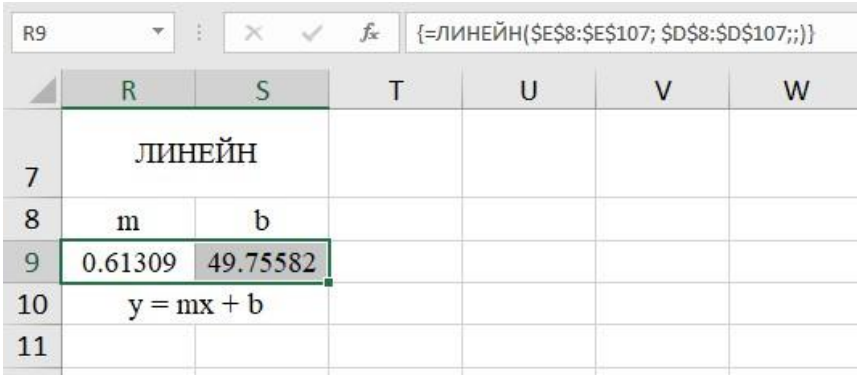
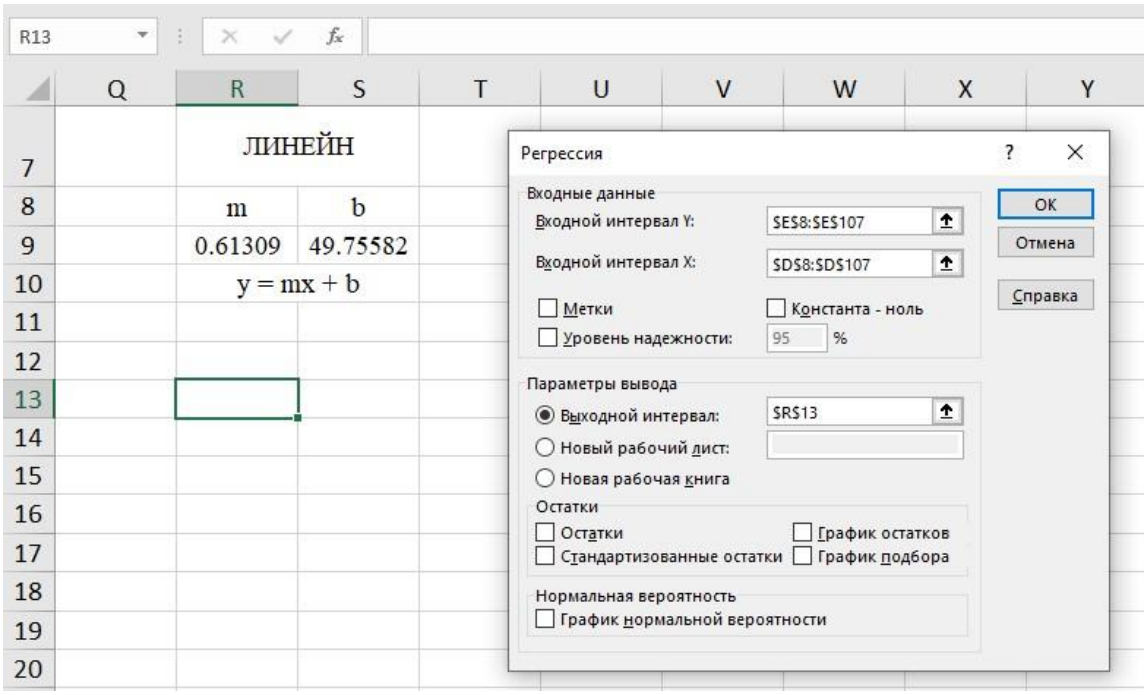


Рис. 2.8 Коефіцієнти регресії знайдені за допомогою функції ЛИНЕЙН

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



	R	S	T	U	V	W	X	Y	Z
12	Надбудова Анализ 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%*	*Верхние 95.0%*
29	Y-пересечение	49.75582	1.609320122	30.9173	2.4E-52	46.56217647	52.9495	46.5622	52.9495
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.

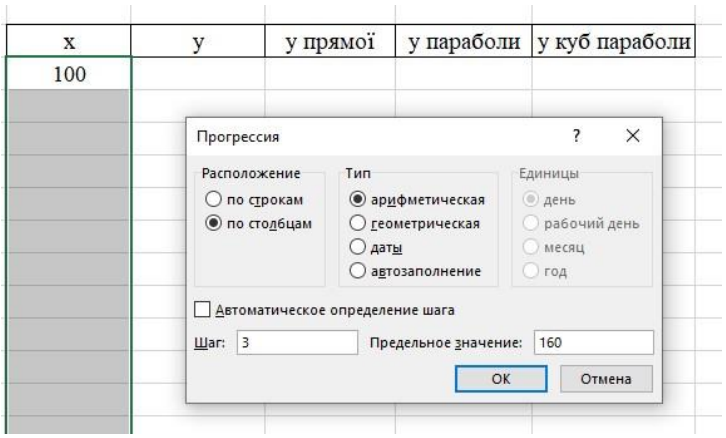
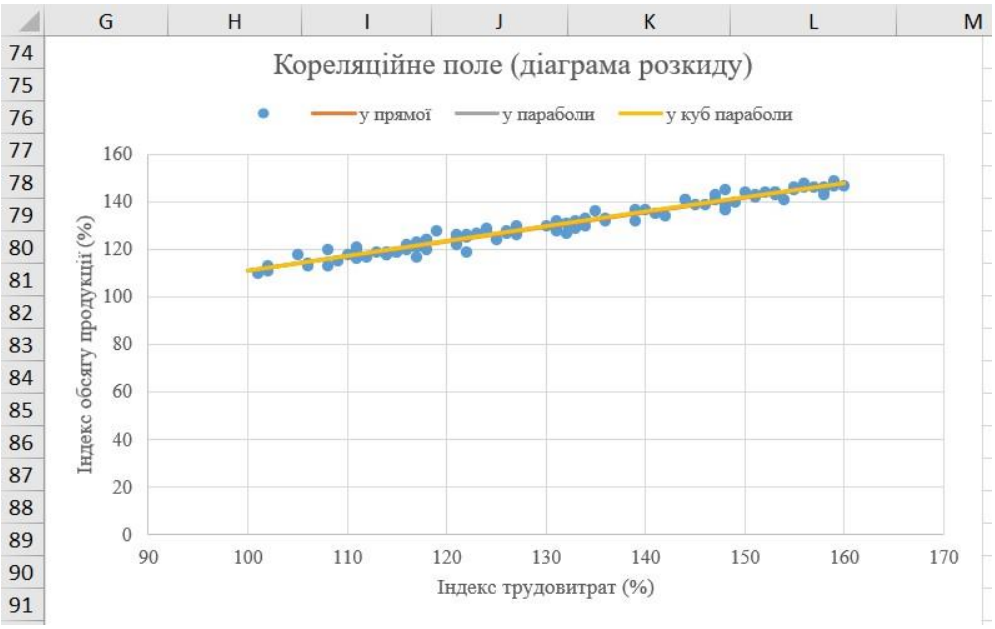


Рис. 2.10 Створення прогресії

	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 *y* 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;

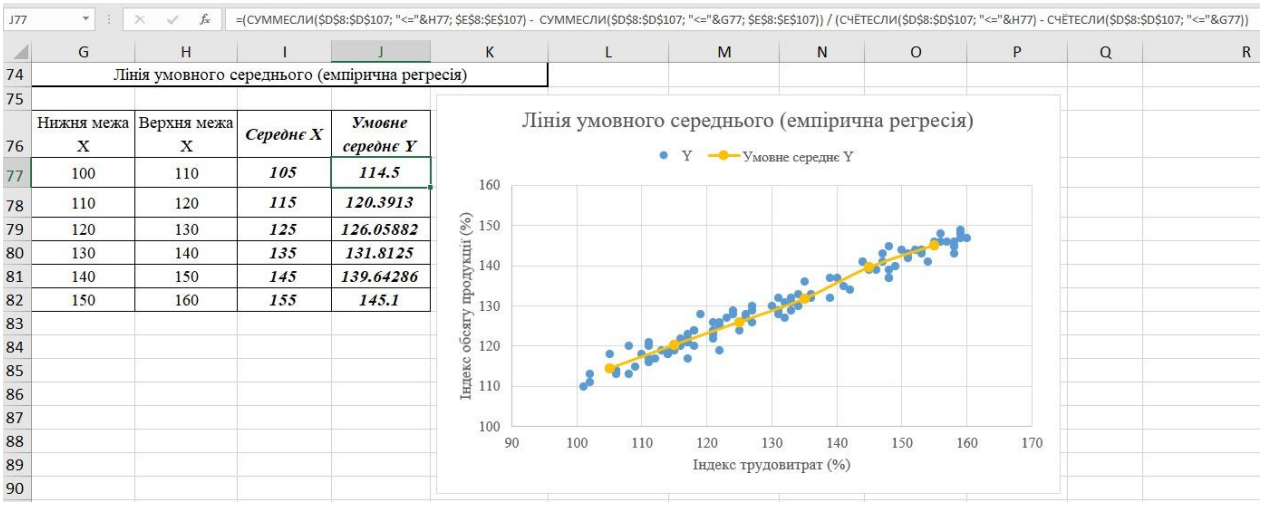
Діапазон_суммування - комірки, значення яких складають при виконанні умови.

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$107)) / (СЧЁТЕСЛИ($D$8:$D$107; "<="&H77) - СЧЁТЕСЛИ($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.



Нижня межа X	Верхня межа X	*Середнє X*	*Умовне середнє Y*
100	110	*105*	*114.5*
110	120	*115*	*120.3913*
120	130	*125*	*126.05882*
130	140	*135*	*131.8125*
140	150	*145*	*139.64286*
150	160	*155*	*145.1*

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:

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

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 | = (СУММПРОИЗВ($E$8:$E$107 - $AB$8:$AB$107; $E$8:$E$107 - $AB$8:$AB$107)) / ($AE$15 - $AE$16 - 1)

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

	AB	AC	AD	AE
6	Показник залишкової дисперсії			
7	у лін регресії	Теорет. прогноз (лінія регресії) використаємо з коефіцієнтами за формулою ЛИНЕЙН		
8	= $R$9 * $D8 + $S$9			
9	= $R$9 * $D9 + $S$9			
10	= $R$9 * $D10 + $S$9	Показник залишкової дисперсії		
11	= $R$9 * $D11 + $S$9	$\sigma^2{}_{зал} = \frac{\sum(y_i - \hat{y}(x_i))^2}{n-k-1}$		
12	= $R$9 * $D12 + $S$9			
13	= $R$9 * $D13 + $S$9			
14	= $R$9 * $D14 + $S$9			
15	= $R$9 * $D15 + $S$9		n	= $L$8
16	= $R$9 * $D16 + $S$9		k	=СЧЁТ(R9:S9)
17	= $R$9 * $D17 + $S$9		σ^2	= (СУММПРОИЗВ($E$8:$E$107 - $AB$8:$AB$107; $E$8:$E$107 - $AB$8:$AB$107)) / ($AE$15 - $AE$16 - 1)
18	= $R$9 * $D18 + $S$9			

Рис. 2.12 Розрахунок показника залишкової дисперсії

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

Контрольне питання

8. Що таке емпірична регресія?

Емпірична регресія – це лінія умовного середнього. Умовне середнє $\bar{y}/x$ – це середнє арифметичне значень результативної ознаки y 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.