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**

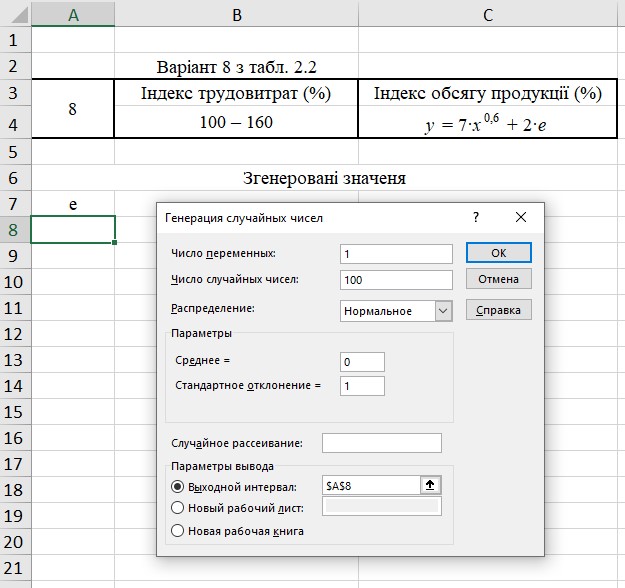
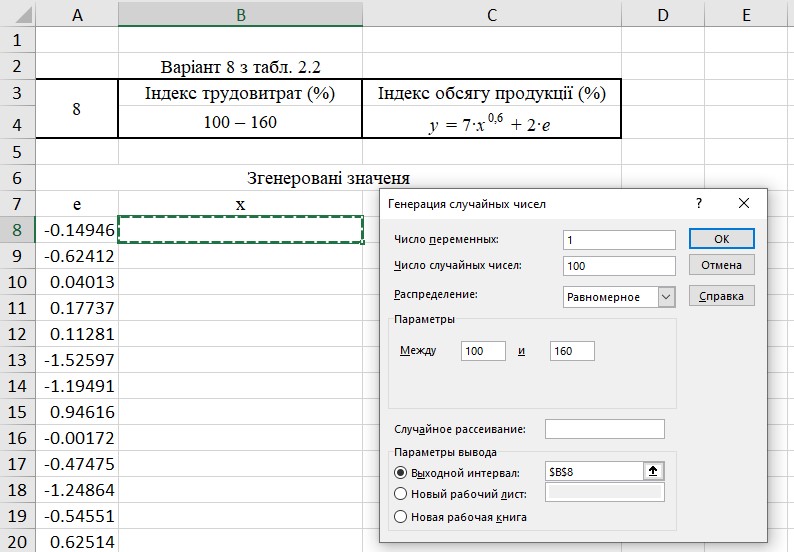
1. *Згенеруємо вихідні дані.*

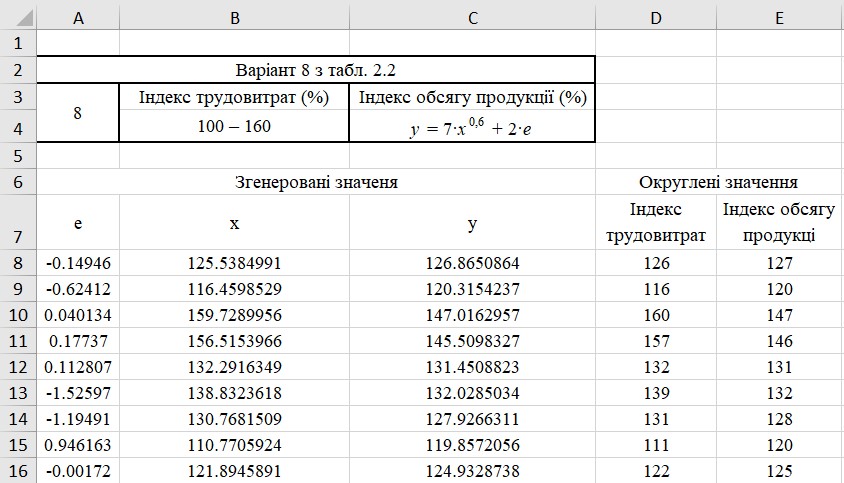
Для початку роботи потрібно згенерувати значення двох змінних *x* і *y,* відповідно Табл. 2.1. Обсяг вибірки – 100 елементів.

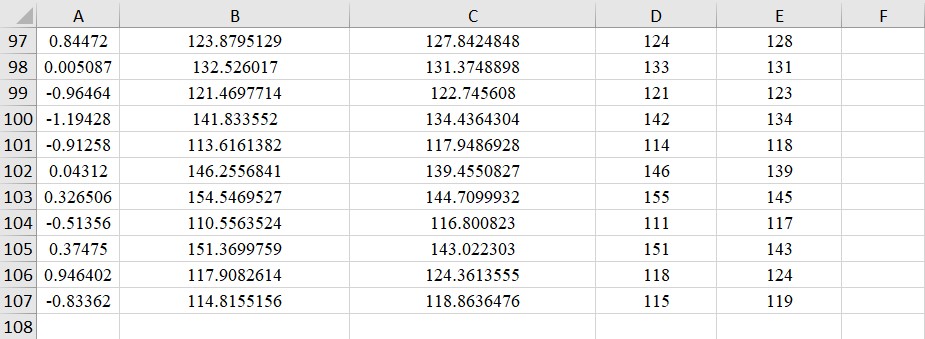
Таблиця 2.1 – Варіанти завдань

|  |  |  |  |
| --- | --- | --- | --- |
| **№** | **Фактор (*х*)** | **Результат (*у*)** | |
| 1 | Заробітна плата (грн)  3000 – 10000 | Споживання (грн)  *y* = 1500 + 0,5·*x* + 500·*e* | |
| 2 | Дохід (грн)  3500 – 11000 | Заощадження (грн)  *y* = –1000 + 0,5·*x* + 300·*e* | |
| 3 | Кількість студентів  2000…10000 | Кількість викладачів ВНЗ  *y* = 220 + 0,09·*x* + 50·*e* | |
| 4 | Ціна товару (грн)  15 – 50 | Попит, кг  *y* = 200000·*x*–0,85+ 500·*e* | |
| 5 | Валовий національний продукт  (млрд. грн.)  1 – 8 | Особисті доходи (млн. грн.)  *y* = –0.4 + 0,95·*x* + 0,4·*e* | |
| 6 | Витрати на рекламу (млн. грн.)  0 – 10 | Прибуток(млн. грн.)  *y* = 10 + 6·*x* – 0,3·*x*2+ 2·*e* | |
| 7 | Грошова маса (млн. грн.)  100 – 350 | Індекс цін (%)  *y* = 38 + 0,3·*x* + 7·*e* | |
| 8 | Індекс трудовитрат (%)  100 – 160 | Індекс обсягу продукції (%)  *y* = 7·*x*0,6+ 2·*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*.





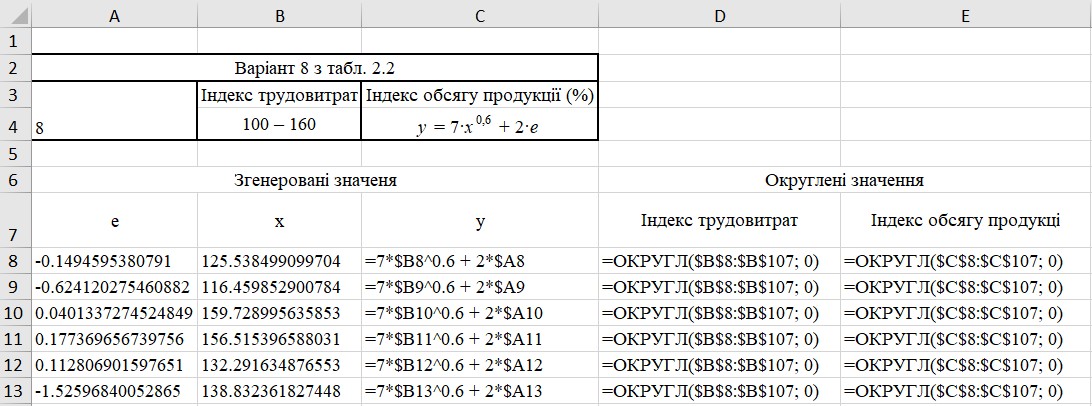
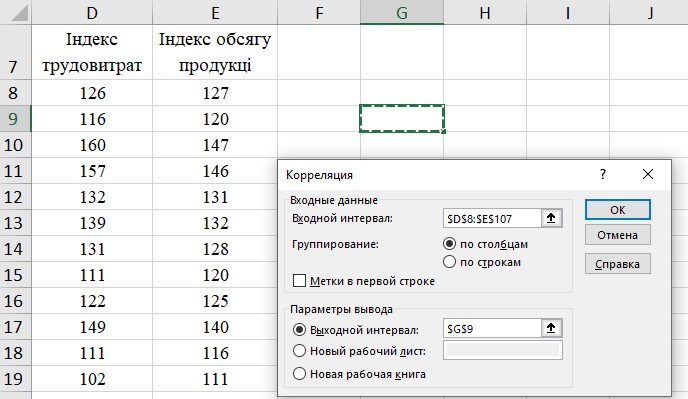


Рис. 2.1 Приклади результату генерації і округлення даних.

1. *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



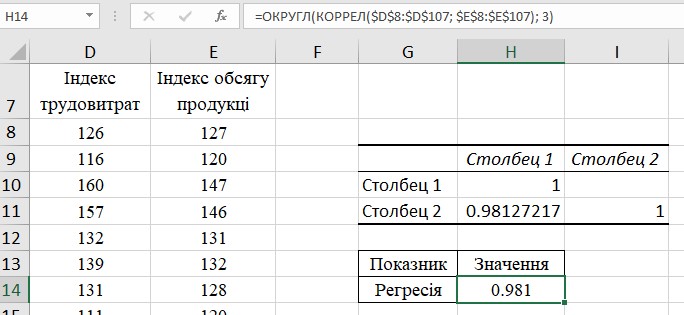


Рис. 2.2 Приклад обчислення коефіцієнтів кореляції.

1. *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 ≤ |*r*| < 0,5 | Слабкий |
| 0,5 ≤ |*r*| < 0,7 | Помірний |
| 0,7 ≤ |*r*| < 1,0 | Сильний |
| |*r*| = 1,0 | Функціональний |

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

1. *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:

Параболи:

Кубічного рівняння:

Невідомі коефіцієнти можуть бути знайдені методом найменших квадратів(МНК), шляхом мінімізації суми квадратів:

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

для прямої;

для параболи;

для кубічного рівняння.

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

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

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

На Рис. 2.3 показаний приклад обчислення проміжних значень, таких як і т.д. Використовується функція СУММПРОИЗВ, яка дозволяє обчислити суму попарних добутків декількох стовпців.

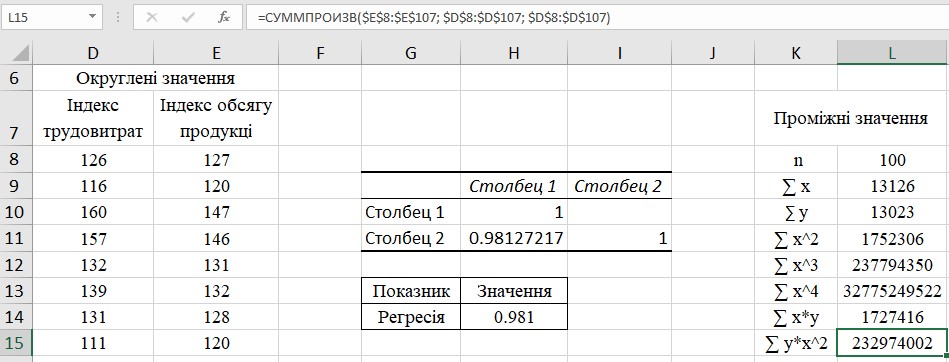


Рис. 2.3 Приклад обчислення проміжних сум.

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;

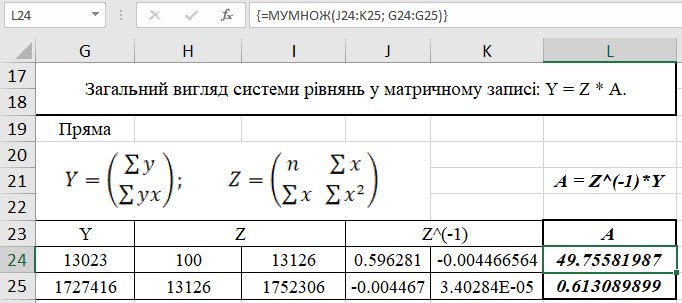
- введіть формулу в клітинку, що є лівим верхнім кутом обраного діапазону, натиснути *Enter*;

- 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*).*



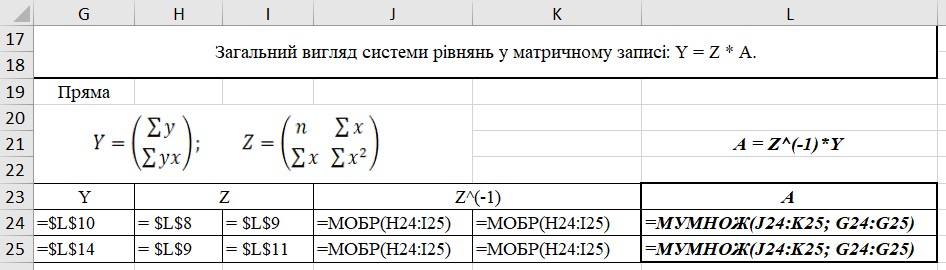


Рис. 2.4 Приклад роботи з матрицями для прямої

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

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

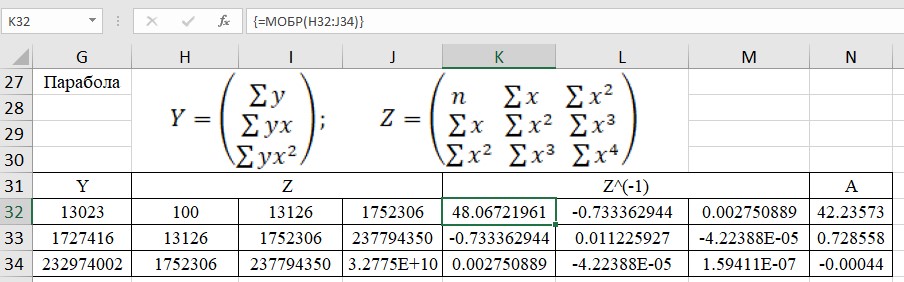


Рис. 2.5 Приклад роботи з матрицями для параболи

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

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

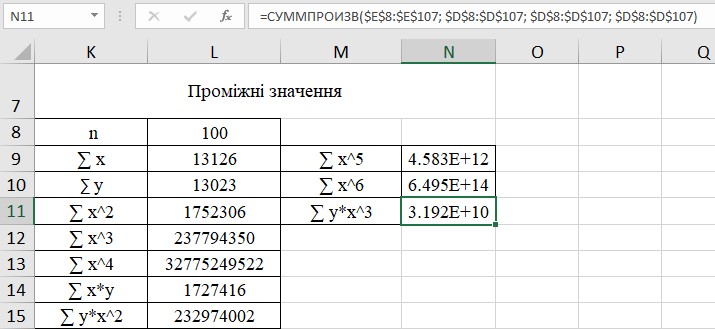
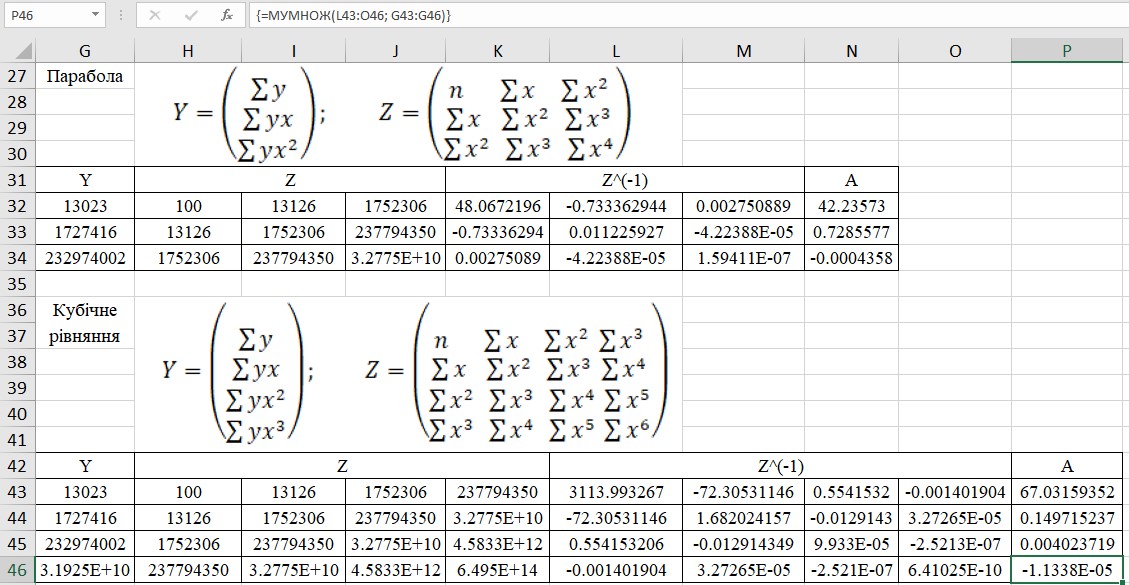


Рис. 2.6 Приклад обчислення проміжних сум.



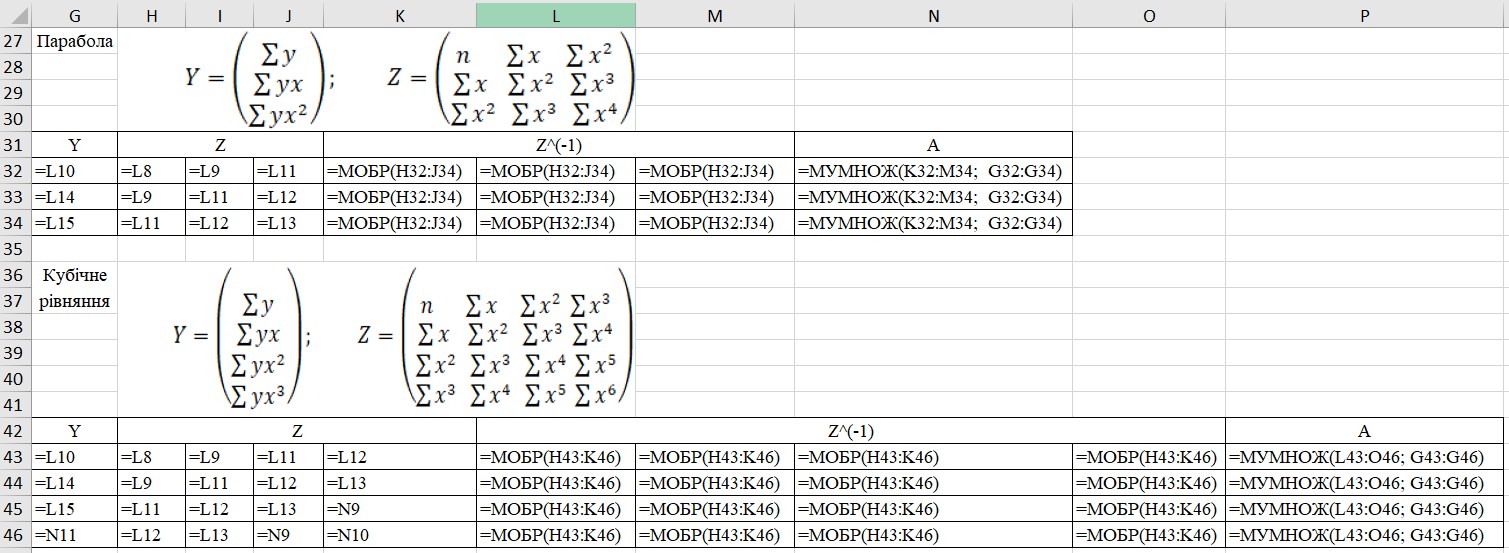


Рис. 2.7 Приклад роботи з матрицями для параболи та кубічного рівняння (+ формули)

Коефіцієнти регресії можна також знайти за допомогою функції ЛИНЕЙН.

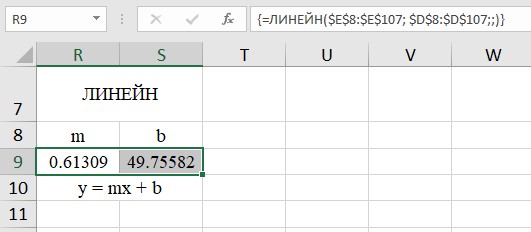
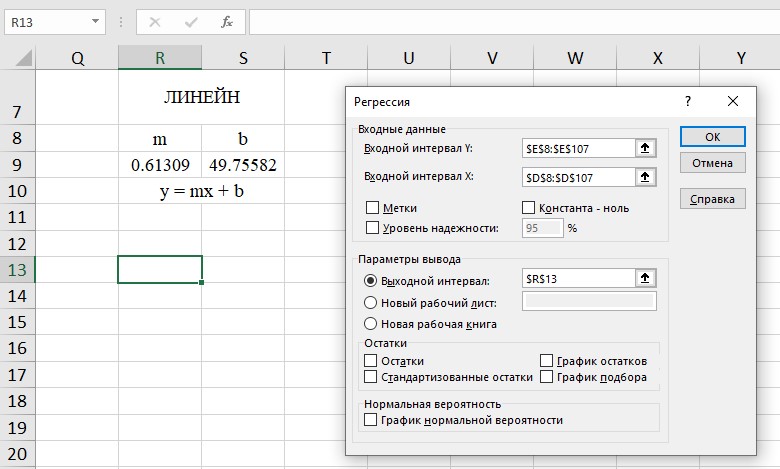


Рис. 2.8 Коефіцієнти регресії знайдені за допомогою функції ЛИНЕЙН

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



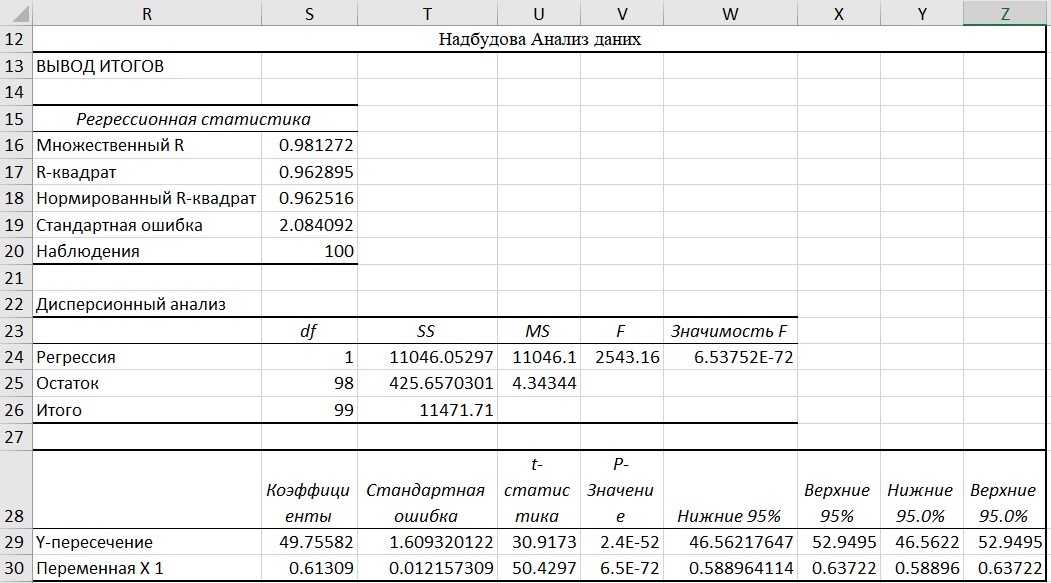


Рис. 2.9 Характеристики регресії знайдені за допомогою надбудови «Аналіз даних»

1. *Побудуйте кореляційне поле.*
2. *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.

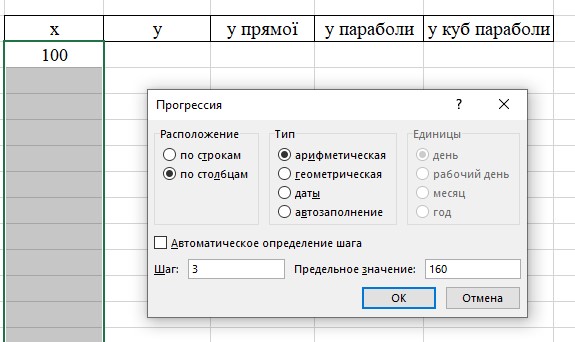
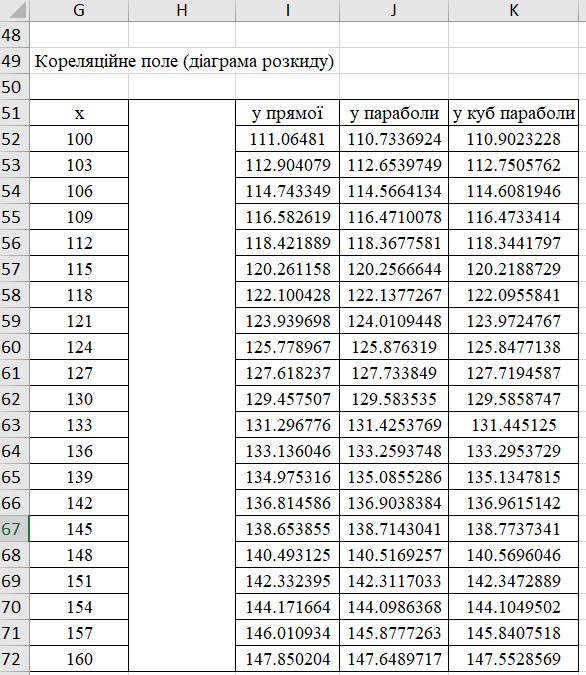
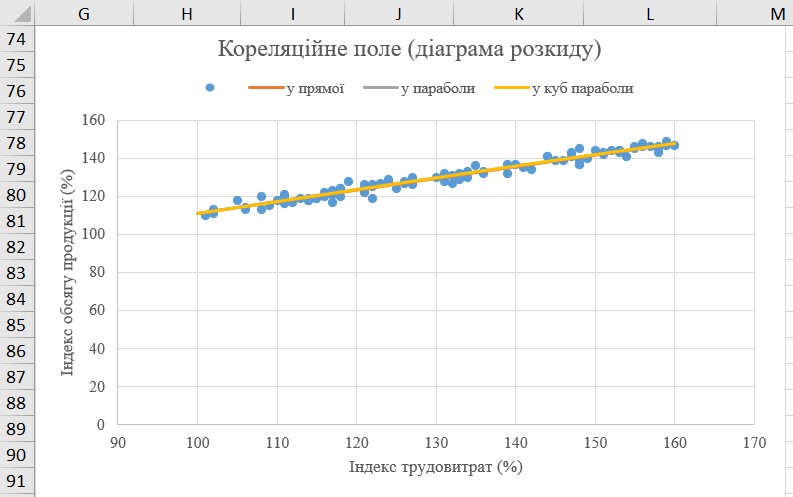


Рис. 2.10 Створення прогресії





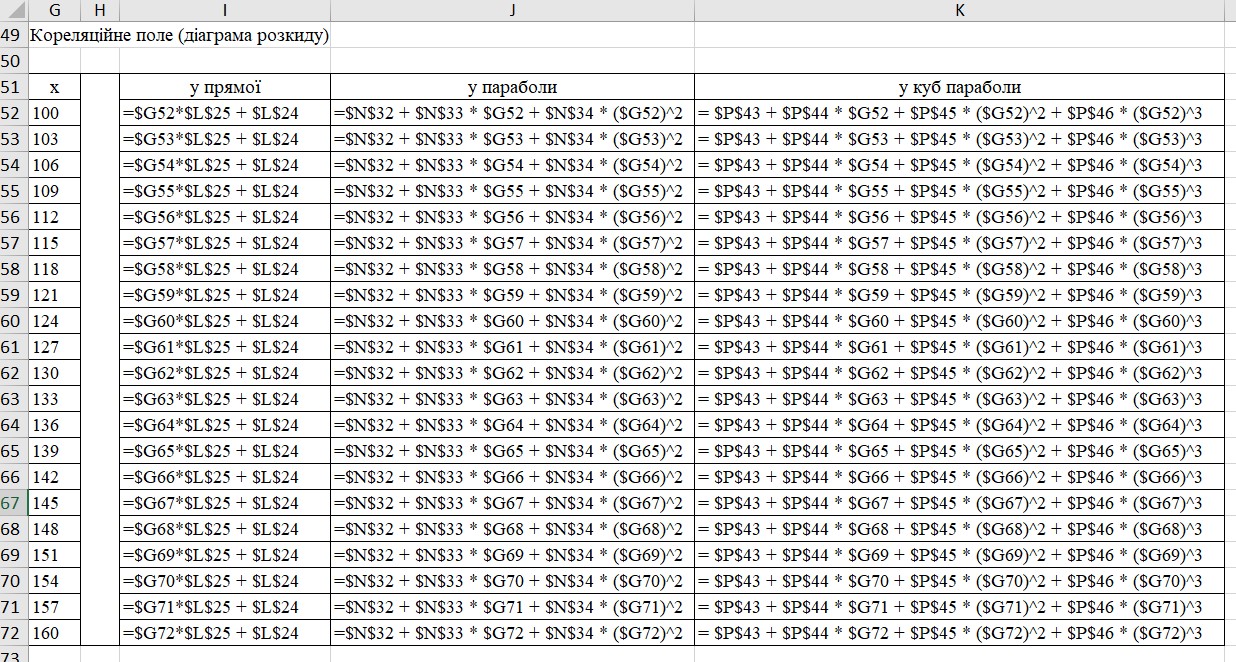


Рис. 2.11 Приклад кореляційного поля з лініями регресії та дані для нього

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.

1. *Нанесіть на кореляційне поле лінію емпіричної регресії.*

Лінія умовного середнього (емпірична регресія)

Умовне середнє – 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.

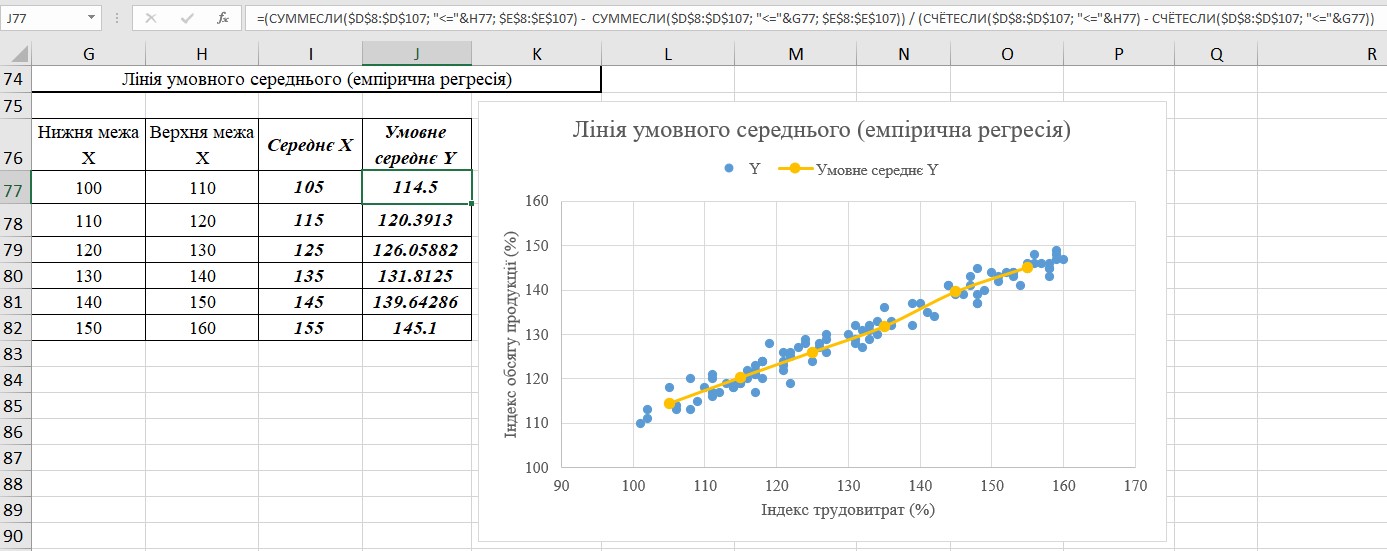


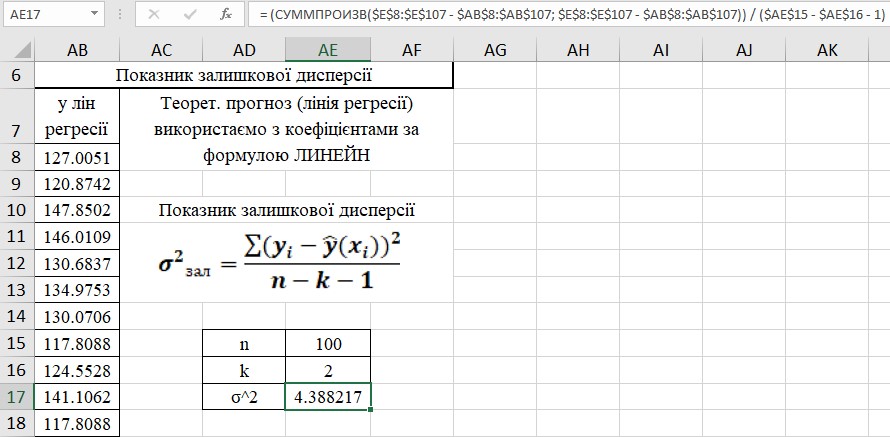
Рис. 2.12 Приклад кореляційного поля з лінією умовного середнього

*9. Розрахуйте значення залишкової дисперсії для кожного рівняння регресії.*

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

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.



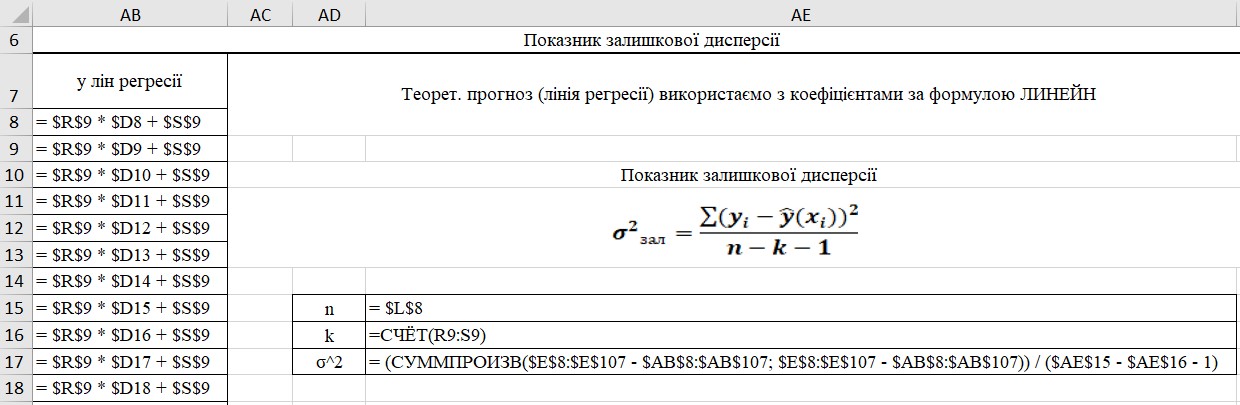


Рис. 2.12 Розрахунок показника залишкової дисперсії

Отже, показник залишкової дисперсії

**Контрольне питання**

*8. Що таке емпірична регресія?*

Емпірична регресія – це лінія умовного середнього. Умовне середнє – 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.