Тема 9. Регресійні моделі в прогнозуванні

Самостійна робота № 9.2

МОДЕЛЬ ПАРНОЇ ЛІНІЙНОЇ РЕГРЕСІЇ (ОДНОФАКТОРНИЙ ЛІНІЙНИЙ КОРЕЛЯЦІЙНО-РЕГРЕСІЙНИЙ АНАЛІЗ)

Мета роботи – ознайомлення з особливостями побудови математичної моделі парної регресії засобами Microsoft Excel, що передбачає вибір найкращого рівняння за допомогою обчислення помилки апроксимації.

1. ЗМІСТОВНА ПОСТАНОВКА ЗАВДАННЯ

На основі умови задачі провести дослідження. Для деякого регіону виконується дослідження залежності між вартістю основних засобів та прибутком підприємства. Дані статистичних спостережень наведені в таблиці 1.

Таблиця 1

Статистичні дані

№ підприємства	Вартість основних засобів, млн.грн.	Прибуток, млн.грн.
1	2,50	1,20
2	2,80	1,50
3	3,00	1,90
4	3,60	2,20
5	3,90	2,80
6	4,20	3,10
7	4,50	3,40
8	5,00	4,50
9	5,60	4,80
10	6,00	5,40

ЗАВДАННЯ:

1) виконати ідентифікацію змінних та специфікацію моделі: сформулювати гіпотезу та поставити економічну задачу використовуючи діаграму розсіювання (кореляційне поле) – детально це питання було розібрано в Задачі 9.1;

2) за допомогою надбудови «Аналіз даних / Data Analytics» інструменту аналізу «Кореляція /Correlation» побудувати кореляційну матрицю та проаналізувати отримані дані;

3) за допомогою надбудови «Аналіз даних / Data Analytics» інструменту аналізу «Описова статистика / Descriptive Statistics» отримати дані та проаналізувати їх;

4) за допомогою надбудови «Аналіз даних / Data Analytics» інструменту аналізу «Регресія / Regression» провести економіко-математичний аналіз моделі;

5) визначити точковий прогноз для заданого значення незалежної змінної – визначити прогнозне значення прибутку підприємства (y) якщо вартість ОЗ підприємства буде на 20% більшою за максимальне значення цієї ознаки;

6) визначити інтервальний прогноз для заданого значення незалежної змінної.

2. РОЗВ'ЯЗАННЯ

1) виконати ідентифікацію змінних та специфікацію моделі: сформулювати гіпотезу та поставити економічну задачу використовуючи діаграму розсіювання (кореляційне поле);

Висуваємо гіпотезу, що прибуток підприємства залежить від вартості основних засобів (ОЗ). Відповідно, незалежною (факторною) змінною x буде вартість ОЗ, а результативною y – прибуток підприємства.

Для визначення форми аналітичної залежності використаємо діаграму розсіювання (кореляційне поле). Результат кожного спостереження (x_i, y_i) деякого економічного процесу відображається точкою на площині. Сукупність цих точок утворює хмарку, яка відображає зв'язок між двома змінними. Діаграма розсіювання є геометричною формою систематизації інформаційної бази процесу дослідження (рис. 1).

За шириною розкиду точок можна зробити висновок про тісноту зв'язку сукупності. Якщо точки розміщені близько одна до одної (у вигляді вузької смужки), то можна стверджувати про наявність відносно тісного зв'язку. Якщо точки на діаграмі розкидані широко, то має місце слабкий зв'язок між змінними.

Розміщення точок на діаграмі розсіювання (рис. 1) дає можливість зробити припущення про існування лінійної форми зв'язку у вигляді функції

$\hat{y} = a_0 + a_1 x$,

де $\hat{y}$ - розрахунковий прибуток, млн.грн.; x – вартість ОЗ, млн.грн.

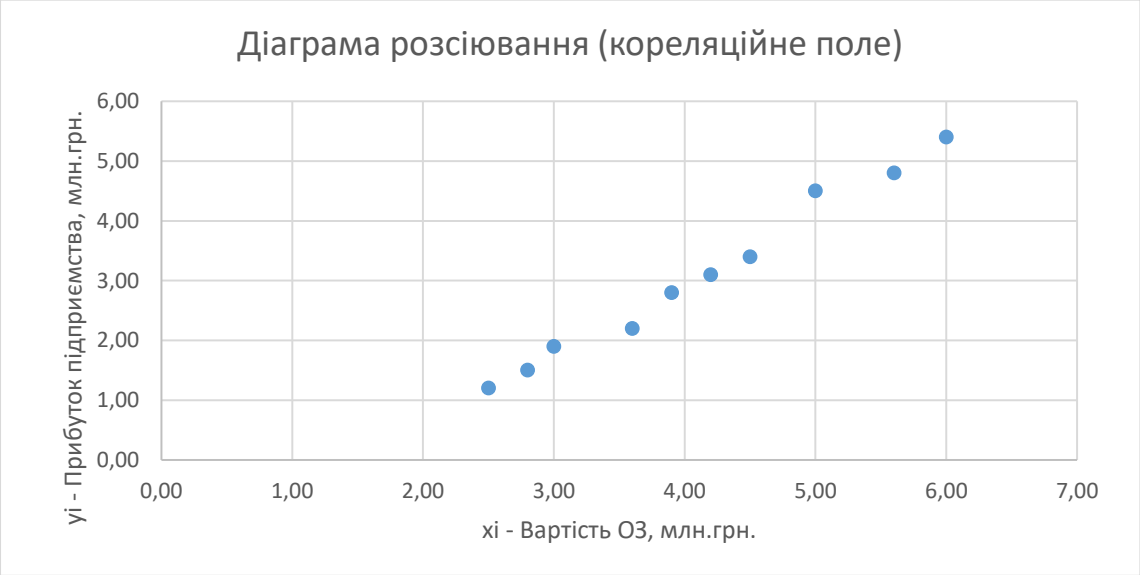

Рис. 1. Діаграма розсіювання (кореляційне поле)

Для отримання кривої емпіричних (фактичних) даних в табличному редакторі *MS Excel* можна поєднати крапки діаграми розсіювання наведені, на рис. 1. Для цього необхідно поставити курсор на область діаграми та у вкладці «Конструктор діаграм» за допомогою вікна «Змінити тип діаграми» обрати «Точкова діаграма з прямими лініями та маркерами». Отримаємо графік як на рис. 2.

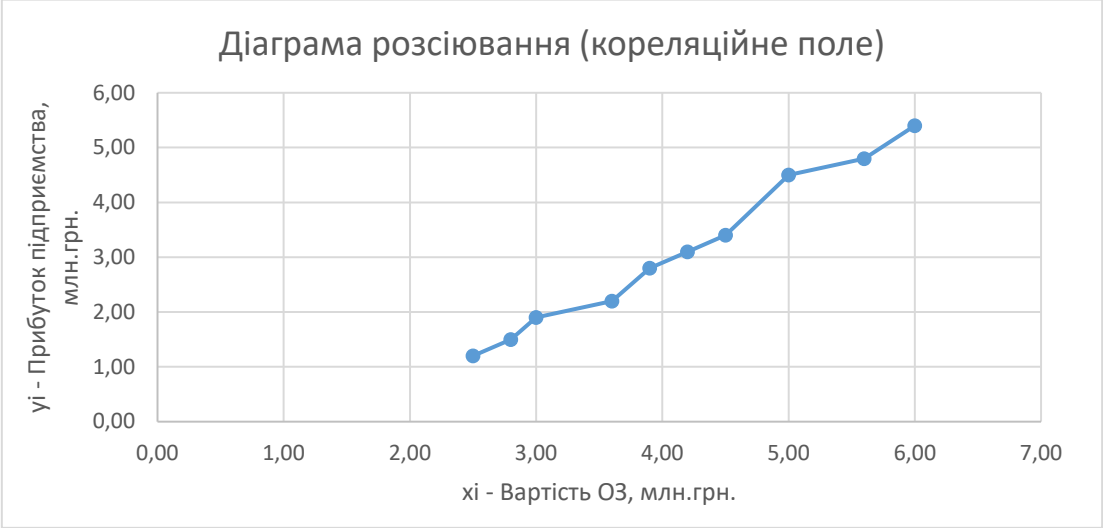

Рис. 2. Діаграма розсіювання (кореляційне поле) у вигляді графіку

Крім цього, в табличному редакторі *MS Excel* є можливість побудувати тренд автоматично. Для цього ставимо курсор на область діаграми – справа з'являються іконки (рис. 3), з них обираємо «+».

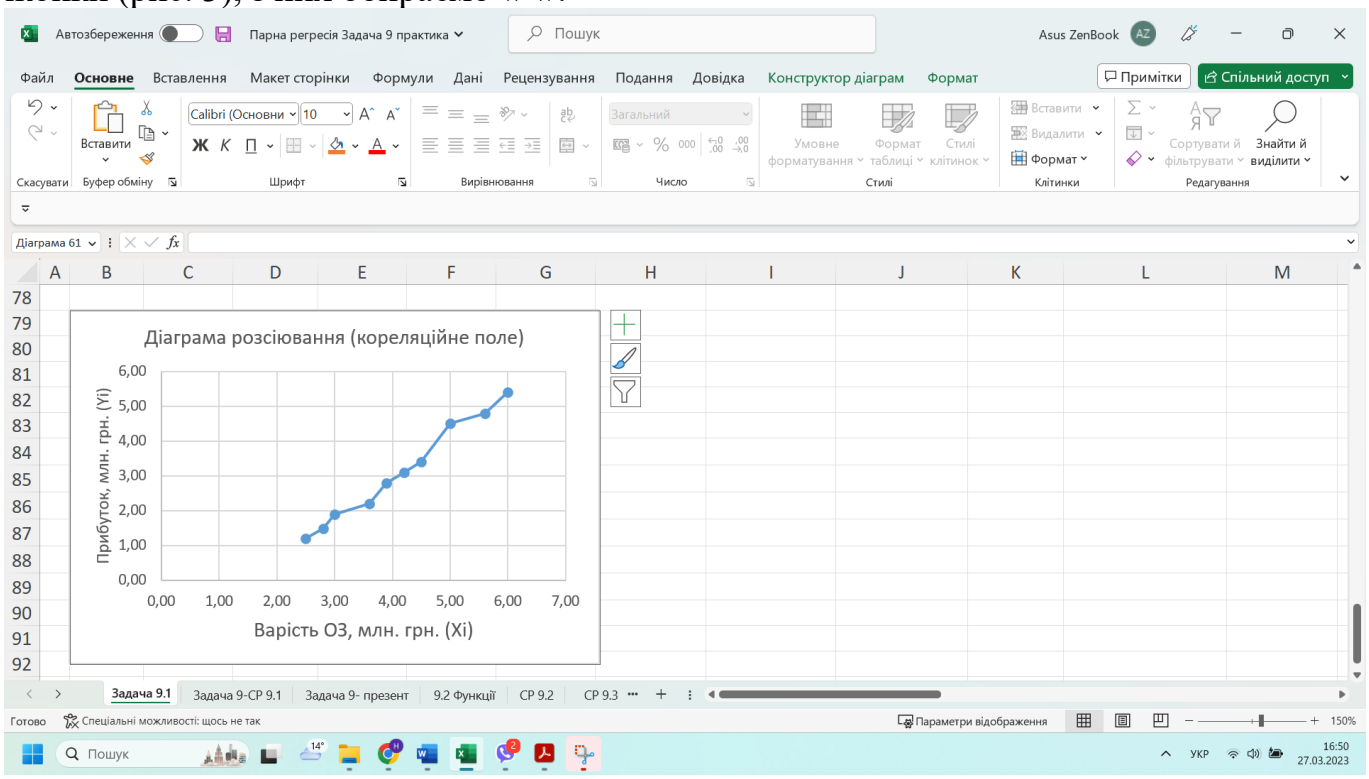

Рис. 3

З випадаючого меню «Елементи діаграми» поставити прапорець навпроти «Лінія тренду». За замовчуванням табличний редактор будує лінійний тренд (рис. 4).

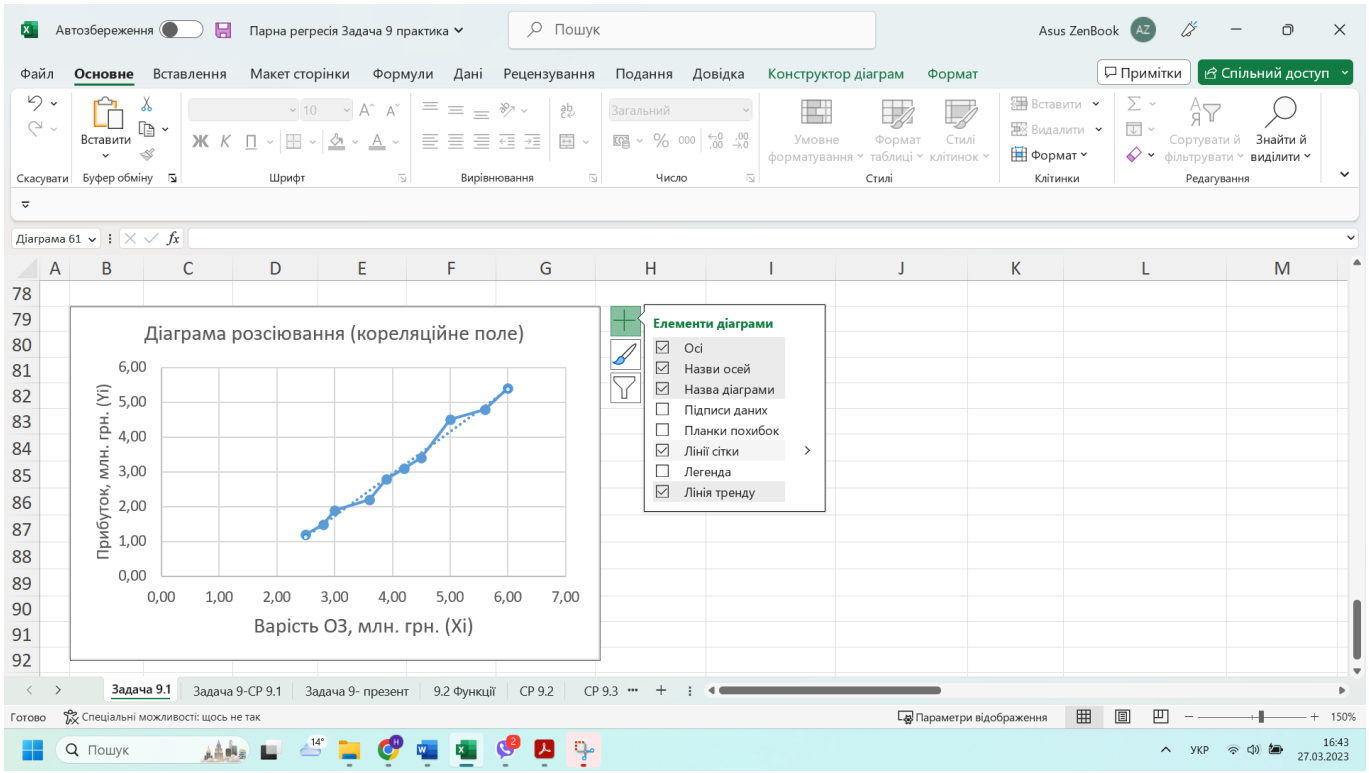

Рис. 4. Побудова лінійного тренду

Для того, аби на графіку з'явились написи з рівнянням тренду та значенням коефіцієнту детермінації R^2 необхідно у вікні «Формат лінії тренду» в розділі «Параметри лінії тренду» поставити прапорець навпроти значень як показано на рис. 5.

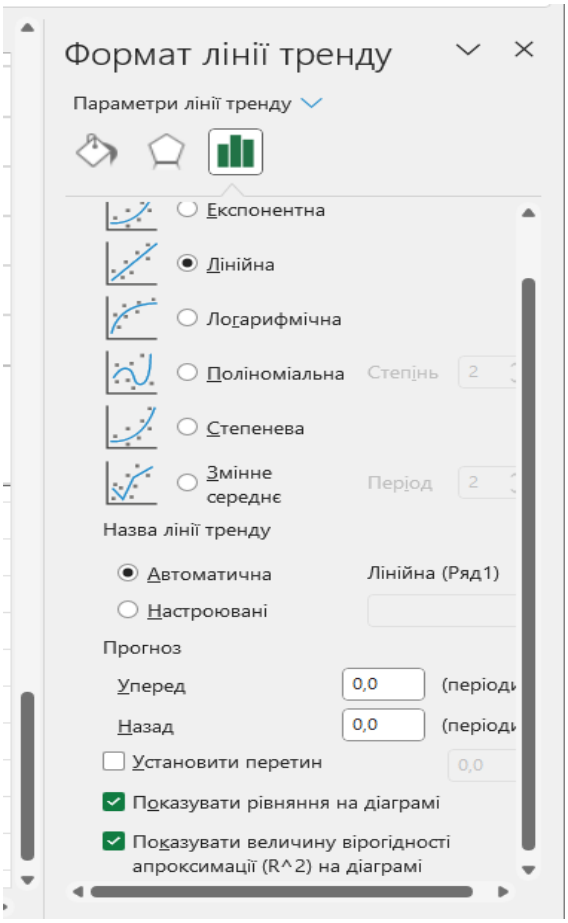

Рис. 5.

За допомогою меню «Елементи діаграми» ми можемо змінювати будь-які елементи діаграми. Наприклад, поміняти вид та колір трендової лінії на суцільну пряму червоного кольору.

В результаті отримаємо наступну діаграму (рис. 6).

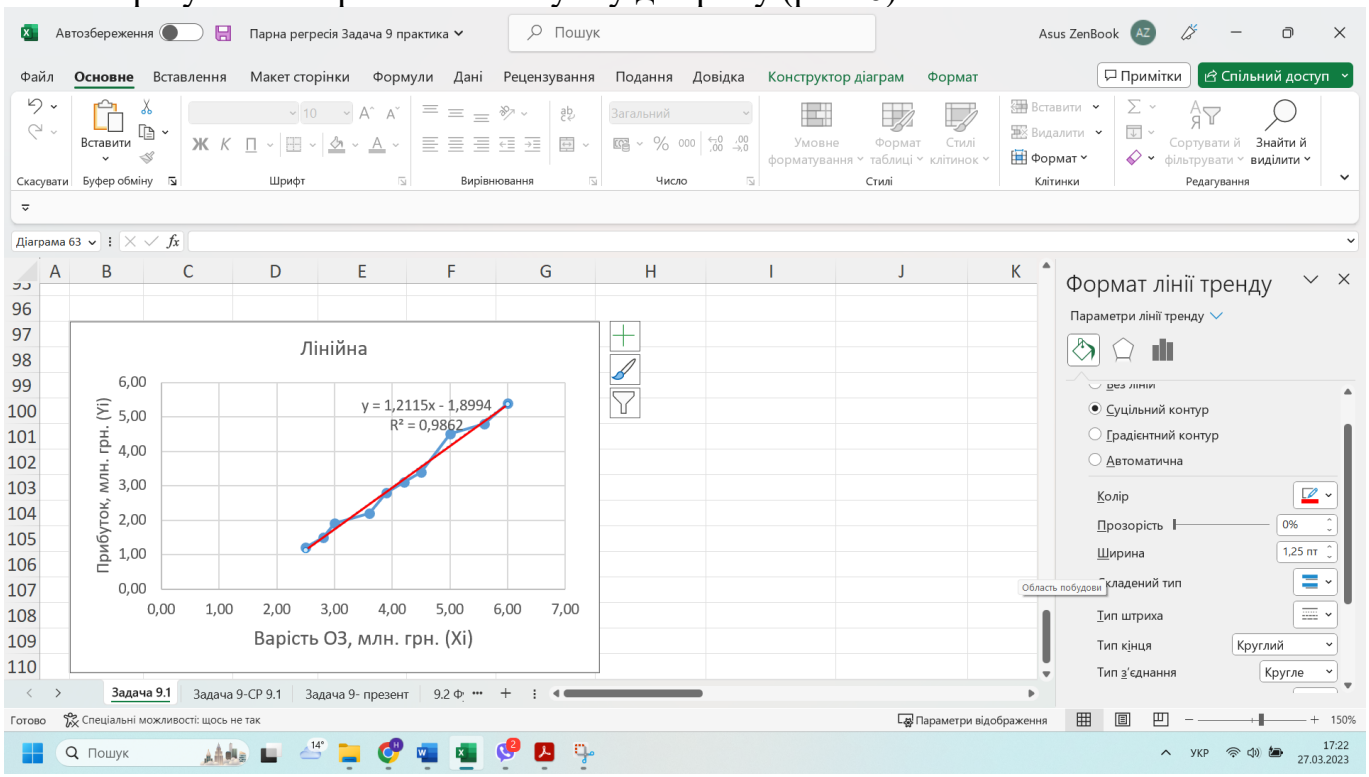

Рис. 6. Лінійний тренд

З даних рис. 6:

a_0= -1,8994

a_1= 1,2115

R^2= 0,9862

Висновки 1:

1) Розміщення точок на діаграмі розсіювання (кореляційному полі) дає можливість зробити припущення про існування лінійної форми зв'язку у вигляді функції $\hat{y} = a_0 + a_1 x$, де $\hat{y}$ – розрахунковий прибуток, млн.грн.; x – вартість ОЗ, млн.грн.

2) З лінійного тренду (рис. 6) перепишемо значення оцінок параметрів парної лінійної регресії

a_0= -1,8994

a_1= 1,2115

Отже, було отримано регресійне рівняння

$$\hat{y} = -1{,}8994 + 1{,}2115 \cdot x$$

3) Коефіцієнт детермінації R^2=0,9862. Він близький до одиниці, що вказує на добру загальну якість моделі. Також це означає, що теоретична пряма пояснює 98,62% варіації або можна сказати, що 98,62% варіації прибутку підприємства відбувається під впливом вартості основних засобів. А решта 1,38% - припадає на інші фактори та випадкові величини. Що свідчить про високий рівень адекватності моделі в цілому.

2) за допомогою надбудови «Аналіз даних / Data Analytics» інструменту аналізу «Кореляція /Correlation» побудувати кореляційну матрицю та проаналізувати отримані дані;

2.1. Активізація надбудови «Аналіз даних / Data Analytics».

Надбудова «Аналіз даних / Data Analytics» активується зі вкладки «Дані» в розділі «Аналіз / Analysis». Якщо ця команда відсутня в меню «Аналіз», її потрібно активізувати шляхом ряду команд. Відкриваємо вкладку «Файл». В лівій частині меню обираємо кнопку «Параметри». На екрані з'являється вікно «Параметри Excel». В лівій половини цього вікна обраємо кнопку «Надбудови» (на рис. 7 обведено червоним). В правій частині вікна «Параметри EXCEL» з'являється назва:

В цій половині вікна натискаємо кнопку «Перейти» (обведено синім на рис.7).

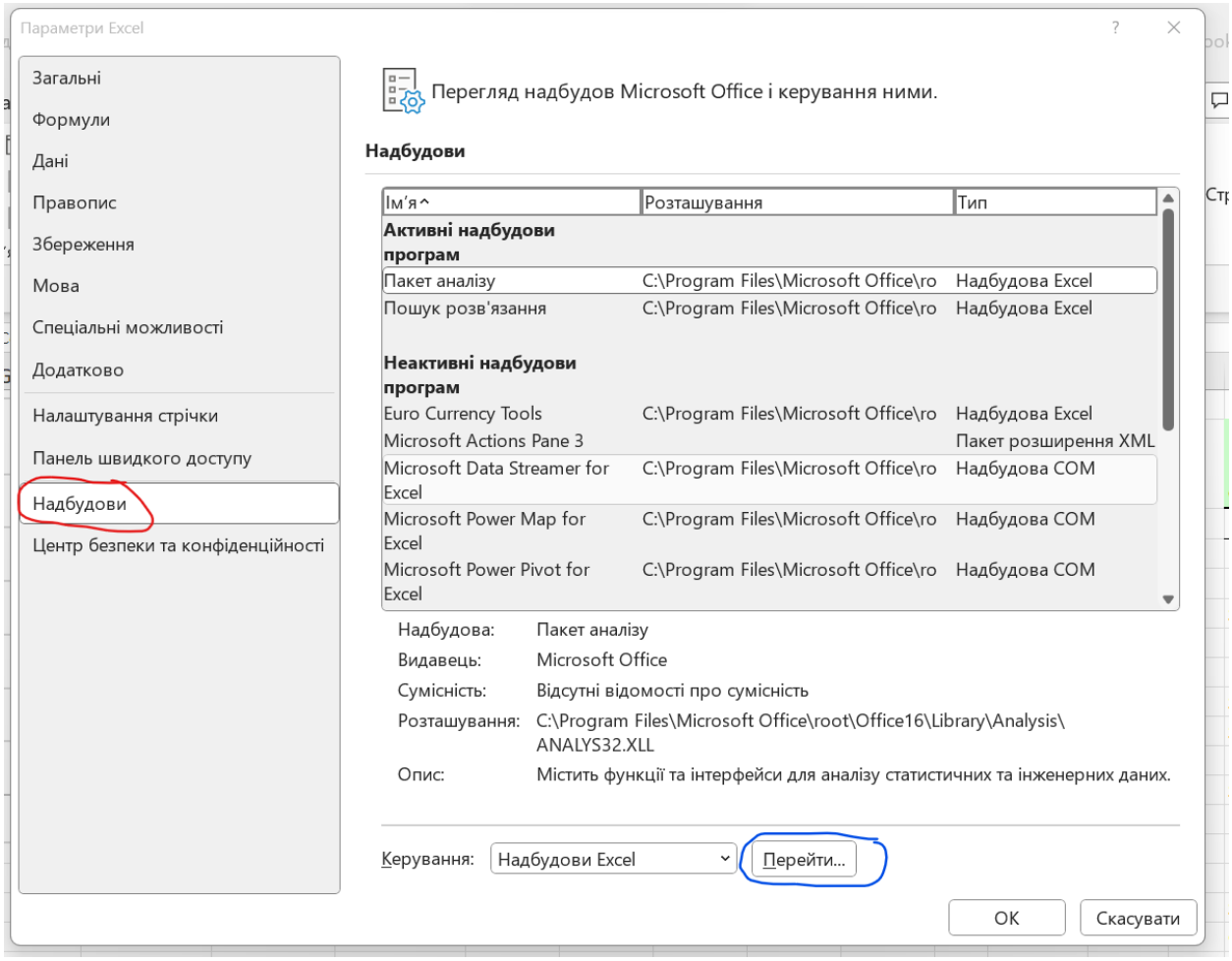

Рис.7

З'явиться вікно «Надбудови» (рис. 8).

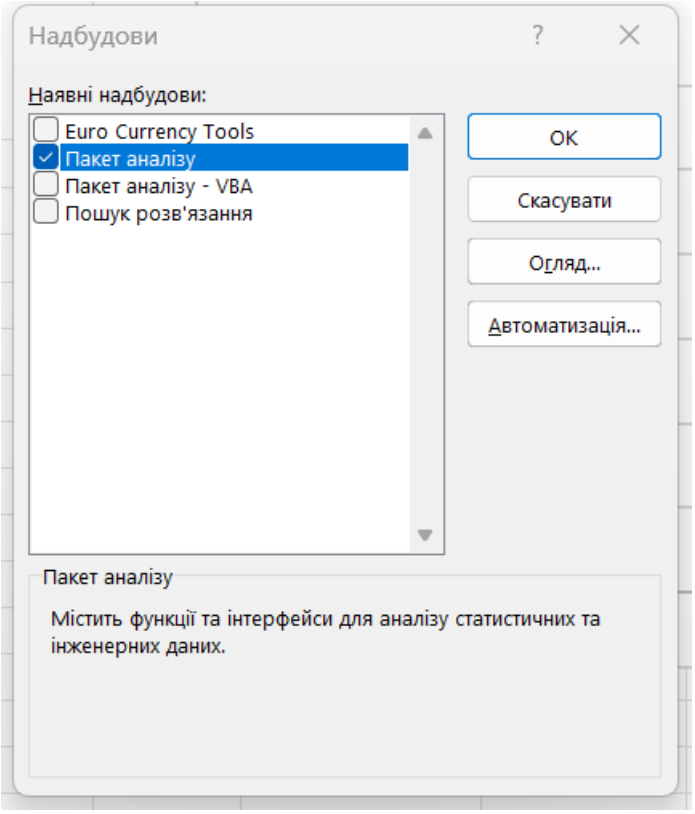

Рис. 8. Діалогове вікно «Надбудови»

Ставимо відмітку навпроти «Пакет аналізу». Натискаємо кнопку «ОК». Тепер пакет «Аналіз даних / Data Analytics» є активним. Про це свідчить поява напису «Аналіз даних / Data Analytics» у вкладці «Дані» (обведено червоним на рис. 9).

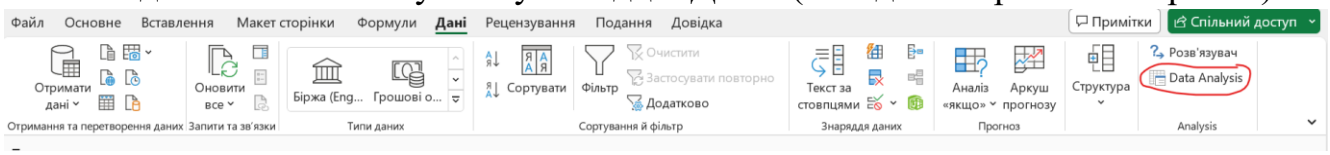

Рис. 9

2.2. Побудова та аналіз кореляційної матриці.

Для побудови кореляційної матриці використаємо інструмент «Кореляція /Correlation».

Обираємо вкладку «Дані», натискаємо на напис «Аналіз даних / Data Analytics». На екрані з'являється вікно «Аналіз даних / Data Analytics». Серед інструментів аналізу обраємо «Кореляція /Correlation» (рис. 10), «ОК».

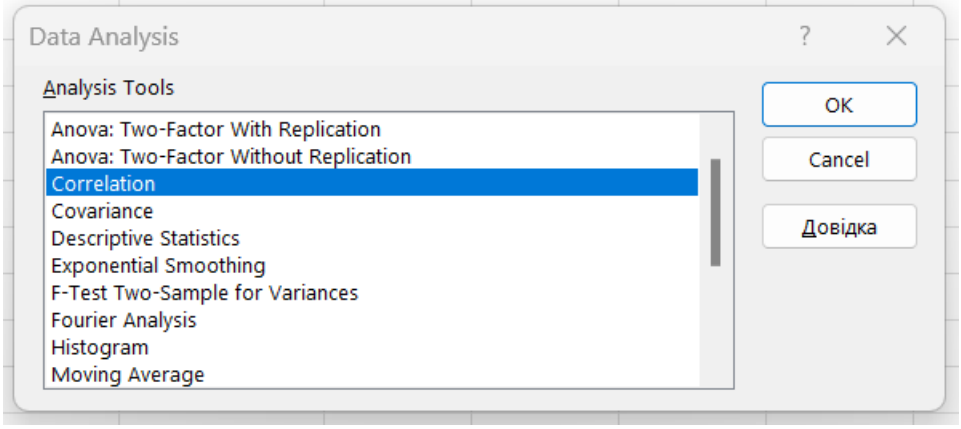

Рис. 10. Вибір інструменту аналізу «Кореляція /Correlation»

Після вибору інструменту аналізу «Кореляція / Correlation», відкривається вікно вводу вихідних даних задачі та параметрів виводу. Загальний вигляд і структура діалогового вікна інструменту «Кореляція / Correlation» наведено на рис. 11. У вікні даного інструменту вводимо вихідні дані та вказуємо спосіб виводу результатів. Натискаємо ОК.

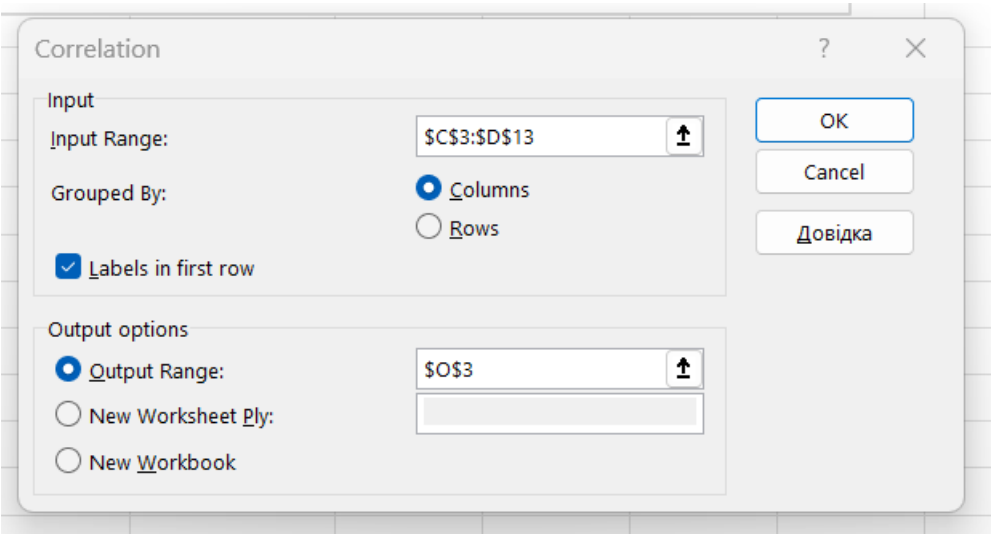

Рис. 11. Діалогове вікно інструменту «Кореляція / Correlation»

Опис елементів діалогового вікна інструменту «Кореляція / Correlation»:

- Зона «**Вхідні дані / Input»**
  - «Вхідний інтервал / Input Range» – вводяться посилання на діапазон, який містить всі вихідні дані (незалежної змінної X та залежної результативної змінної Y). Діапазон можна встановити або вводом з клавіатури або виділенням мишею цих клітинок на робочому аркуші.
  - «Групування / Grouped By:» – вихідні дані згруповані: «по стовпцям / Columns», «по рядках / Rows» – треба обрати один з варіантів в залежності від того, яким чином на робочому аркуші розміщені вихідні дані.
  - «Мітки в першому рядку / Labels in first row» – ставиться прапорець за умови, що крім безпосередньо самих даних в діапазоні «захоплені» назви стовпців або рядків
- Зона «**Параметри виводу / Output options»** місить варіанти способу виводу результатів:
  - «Вихідний інтервал / Output Range» – результати кореляційного аналізу будуть виводитися на цьому ж робочому аркуші. В активному вікні справа треба вказати ім’я комірки, яка буде лівим верхнім кутом діапазону, куди будуть виведені результати кореляційного аналізу;
  - «Новий робочий аркуш / New Worksheet Ply» – результати кореляційного аналізу будуть виводитись на новий робочий аркуш;
  - «Нова робоча книга / New Workbook» – результати кореляційного аналізу будуть виводитись в нову робочу книгу, тобто в новий файл.

В результаті отримуємо такі дані:

	Xi	*Yi*
Xi	1	
Yi	0,9931	1

Рис. 12. Кореляційна матриця

Кореляційна матриця (рис. 12) є симетричною відносно головної діагоналі, тому в верхній частині матриці дані відсутні. На головній діагоналі стоять 1. На перетині стовпчика *Xi* та рядка *Yi* знаходиться коефіцієнт кореляції між цими змінними. Отже, коефіцієнт кореляції r_{xy}=0,9931.

Висновки 2:

1) Оскільки коефіцієнт кореляції r_{xy}=0,9931 і він дуже близький до 1, він показує, що між вартістю ОЗ та прибутком підприємства є сильний лінійний зв'язок. А оскільки він ще й додатній – це вказує на те, між змінними прямий лінійний зв'язок. Прямий зв'язок означає, що зі зростанням фактору (змінної) *x* зростає і фактор (змінна) *y*).

2) Також варто зазначити, що при лінійній залежності коефіцієнт детермінації $R^2=(r_{xy})^2$. Отже перевіримо $0,9862=(0,9931)^2$. Рівність справджується. Отже розрахунки зроблені вірно.

3) за допомогою надбудови «Аналіз даних / Data Analytics» інструменту аналізу «Описова статистика / Descriptive Statistics» отримати дані та проаналізувати їх;

Обираємо вкладку «Дані», натискаємо на напис «Аналіз даних / Data Analytics». На екрані з'являється вікно «Аналіз даних / Data Analytics». Серед інструментів аналізу обраємо «Описова статистика / Descriptive Statistics» (рис. 13). Натискаємо ОК.

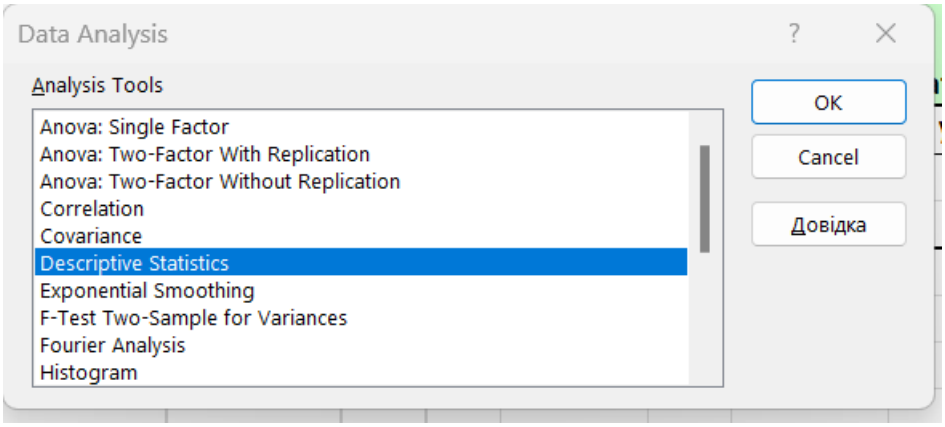

Рис. 13. Вибір інструменту аналізу «Описова статистика / Descriptive Statistics»

Після вибору інструменту аналізу «Описова статистика / Descriptive Statistics», відкривається вікно вводу вихідних даних задачі та параметрів виводу. Загальний вигляд і структура діалогового вікна інструменту «Описова статистика / Descriptive Statistics» наведено на рис. 14. У вікні даного інструменту вводимо вихідні дані та вказуємо спосіб виводу результатів. Натискаємо ОК.

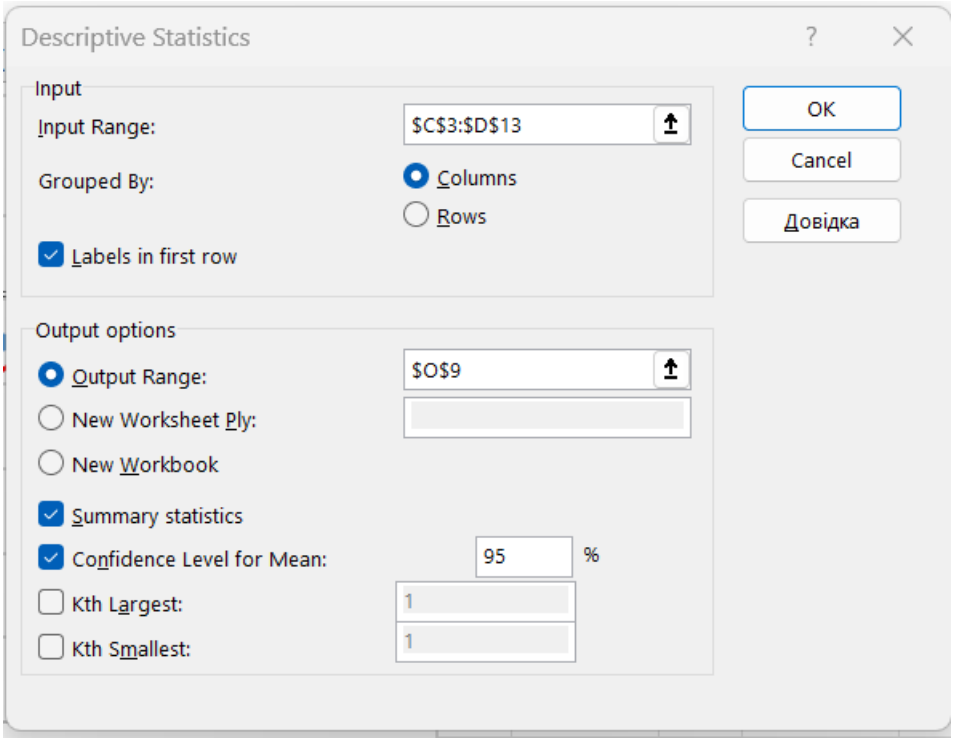

Рис. 14. Діалогове вікно інструменту «Описова статистика / Descriptive Statistics»

Опис елементів діалогового вікна інструменту «Описова статистика / Descriptive Statistics»:

- Зона «**Вхідні дані / Input»**
  - «Вхідний інтервал / Input Range» – вводяться посилання на діапазон, який містить всі вихідні дані (незалежної змінної X та залежної результативної змінної Y). Діапазон можна встановити або вводом з клавіатури або виділенням мишею цих клітинок на робочому аркуші.
  - «Групування / Grouped By:» – вихідні дані згруповані: «по стовпцям / Columns», «по рядках / Rows» – треба обрати один з варіантів в залежності від того, яким чином на робочому аркуші розміщені вихідні дані.
  - «Мітки в першому рядку / Labels in first row» – ставиться прапорець за умови, що крім безпосередньо самих даних в діапазоні «захоплені» назви стовпців або рядків
- Зона «**Параметри виводу / Output options»** місить варіанти способу виводу результатів:
  - «Вихідний інтервал / Output Range» – результати кореляційного аналізу будуть виводитися на цьому ж робочому аркуші. В активному вікні справа треба вказати ім’я комірки, яка буде лівим верхнім кутом діапазону, куди будуть виведені результати кореляційного аналізу;
  - «Новий робочий аркуш / New Worksheet Ply» – результати кореляційного аналізу будуть виводитись на новий робочий аркуш;
  - «Нова робоча книга / New Workbook» – результати кореляційного аналізу будуть виводитись в нову робочу книгу, тобто в новий файл;
  - «Загальна (зведена) статистика / Summary statistics»;

- «Рівень надійності / Confidence Level of Mean» – за замовчуванням рівень надійності встановлюється на рівні 95%, але якщо потрібно змінити, то у відповідне поле вводиться інший рівень надійності;
- «К-ий найбільший / Kth Largest»;
- «К-ий найменший / Kth Smallest».

В результаті отримуємо такі дані:

Xi		*Yi*	
Mean	4,11	Mean	3,08
Standard Error	0,374002377	Standard Error	0,45626503
Median	4,05	Median	2,95
Mode	#Н/Д	Mode	#Н/Д
Standard Deviation	1,182699361	Standard Deviation	1,442836712
Sample Variance	1,398777778	Sample Variance	2,081777778
Kurtosis	-1,038445182	Kurtosis	-1,163158671
Skewness	0,242907479	Skewness	0,348474234
Range	3,5	Range	4,2
Minimum	2,5	Minimum	1,2
Maximum	6	Maximum	5,4
Sum	41,1	Sum	30,8
Count	10	Count	10
Confidence Level(95,0%)	0,846052155	Confidence Level(95,0%)	1,032143206

Xi		*Yi*	
Середнє	4,11	Середнє	3,08
Стандартна похибка	0,374002377	Стандартна похибка	0,45627
Медіана	4,05	Медіана	2,95
Мода	#Н/Д	Мода	#Н/Д
Стандартне відхилення	1,182699361	Стандартне відхилення	1,44284
Дисперсія вибірки	1,398777778	Дисперсія вибірки	2,08178
Ексцес	-1,038445182	Ексцес	-1,16316
Асиметричність	0,242907479	Асиметричність	0,34847
Інтервал	3,5	Інтервал	4,2
Мінімум	2,5	Мінімум	1,2
Максимум	6	Максимум	5,4
Сума	41,1	Сума	30,8
Рахунок	10	Рахунок	10
Рівень надійності (95,0%)	0,846052155	Рівень надійності (95,0%)	1,03214

Рис. 15.

4) за допомогою надбудови «Аналіз даних / Data Analytics» інструменту аналізу «Регресія / Regression» провести економіко-математичний аналіз моделі.

Інструмент «Регресія / Regression» слугує для розрахунку параметрів рівняння лінійної регресії та перевірки його адекватності.

Зробити все аналогічно попереднім інструментам аналізу, але обрати інструмент «Регресія / Regression» (рис. 16).

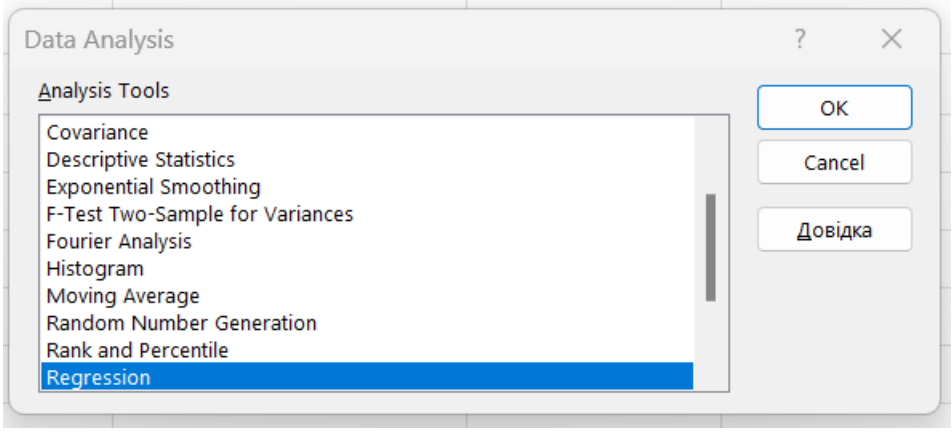

Рис. 16. Вибір інструменту аналізу «Регресія / Regression»

Загальний вигляд та структура діалогового вікна інструменту «Регресія / Regression» наведено на рис. 17. У вікні даного інструменту вводимо вихідні дані та вказуємо спосіб виводу результатів. Натискаємо ОК.

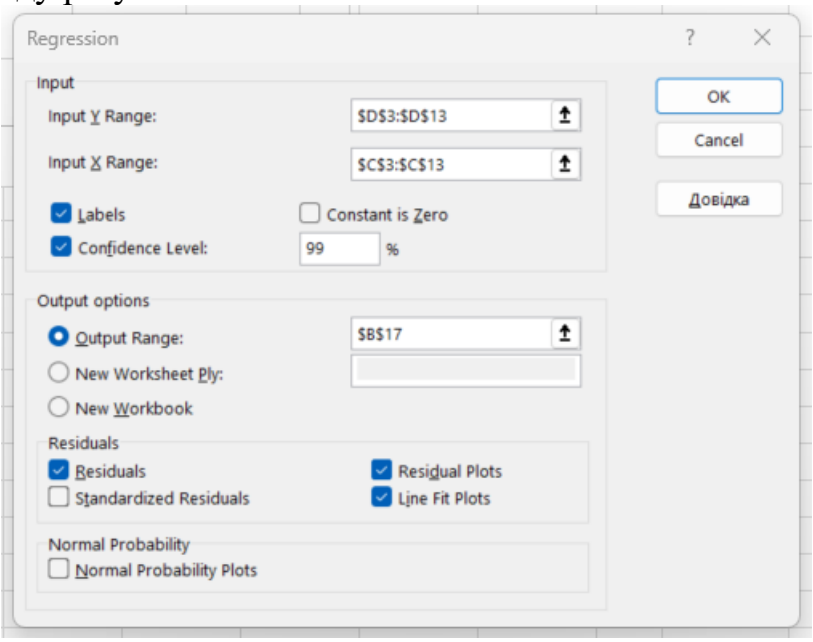

Рис. 17. Меню діалогового вікна інструменту «Регресія / Regression»

Опис елементів *діалогового вікна інструменту «Регресія / Regression»*:

❖ Зона «**Вхідні дані / Input»**

- «Вхідний інтервал Y / Input Y Range» – вводяться посилання на діапазон, який містить дані залежної (результативної) змінної, що аналізується. Діапазон можна встановити або вводом з клавіатури або виділенням мишею цих клітинок на робочому аркуші. Діапазон повинен містити лише один стовпець.
- «Вхідний інтервал X / Input Y Range» – вводяться посилання на діапазон незалежних (пояснюючих) даних, що аналізуються. Табличний редактор MS Excel розташовує незалежні змінні по стовпцях зліва направо у порядку зростання. Максимальна кількість вхідних діапазонів дорівнює 16.

- «Мітки / Labels» – якщо перший рядок або перший стовпець вхідного інтервалу містить заголовки, то робиться відповідна позначка. У випадку якщо заголовки відсутні, відповідні назви для даних вхідного діапазону створюються автоматично.
- «Рівень надійності / Confidence Level» – у відповідне поле вводиться рівень надійності, який буде використовуватися додатково до рівня 95%, яке встановлюється за замовчуванням;
- «Константа нуль / Constant is Zero» – у випадку, коли необхідно щоб лінія регресії пройшла через початок координат, робиться відповідна позначка.

❖ Зона «**Параметри виводу / Output options»** місить варіанти способу виводу результатів:

- «Вихідний інтервал / Output Range» – результати регресійного аналізу будуть виводитися на цьому ж робочому аркуші. В активному вікні справа треба вказати ім'я комірки, яка буде лівим верхнім кутом діапазону, куди будуть виведені результати аналізу;
- «Новий робочий аркуш / New Worksheet Ply» – результати регресійного аналізу будуть виводитись на новий робочий аркуш;
- «Нова робоча книга / New Workbook» – результати регресійного аналізу будуть виводитись в нову робочу книгу, тобто в новий файл;

❖ Зона «**Залишки / Residuals**»

- «Залишки / Residuals» – за потреби дозволяє включити залишки у вихідний діапазон; після підбору рівняння зазвичай здійснюється перевірка та аналіз залишків тому, що дуже великі відхилення суттєво спотворюють результати та призводять до помилкових висновків.
- «Стандартизовані залишки / Standardized Residuals» – дозволяють виводити значення стандартизованих залишків, які обчислюються як різниця між фактичними та прогнозними значеннями, що ділиться на квадратний корінь з середньоквадратичного значення залишків.
- «Графік залишків / Residual Plots» – виводить діаграму залишків для кожної незалежної змінної.
- «График підбору / Line Fit Plots» – будує діаграму фактичних та прогнозних значень для кожної незалежної змінної.
- «График нормальної ймовірності / Normal Probability Plots» – виводить діаграму нормальної ймовірності, яка будується наступним чином. Спочатку відбувається ранжування стандартизованих залишків. За цими рангами обчислюються стандартні значення нормального розподілу (z-значення) на основі припущення, що дані підпорядковуються нормальному розподілу. Ці z-значення і відкладаються на графіку.

В результаті отримаємо результати розрахунків «ВИВЕДЕННЯ ПІДСУМКІВ / SUMMARY OUTPUT» (рис. 18).

Регресія / Regression

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,993
R Square	0,986
Adjusted R Square	0,985
Standard Error	0,179
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	18,478	18,478	573,6707867	0,000000009838
Residual	8	0,258	0,032		
Total	9	18,736			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 99,0%*	*Upper 99,0%*
Intercept	-1,8994	0,2155	-8,814	0,0000216043	-2,396	-1,402	-2,623	-1,176
Xi	1,2115	0,0506	23,951	0,0000000098	1,095	1,328	1,042	1,381

ВИВЕДЕННЯ ПІДСУМКІВ

Регресійна статистика	
Множинний R	0,9931
R-квадрат	0,9862
Нормований R-квадрат	0,9845
Стандартна похибка	0,1795
Спостереження	10

Дисперсійний аналіз

	df	*SS*	*MS*	*F*	*Значимість F*
Регресія	1	18,478	18,478	573,67	0,0000000098
Залишки	8	0,258	0,032		
Загальне	9	18,736			

	Коефіцієнти	*Стандартна похибка*	*t-статистика*	*P-Значення*	*Нижні 95%*	*Верхні 95%*	*Нижні 99,0%*	*Верхні 99,0%*
Y-перетин	-1,8994	0,2155	-8,814	0,00002160433	-2,396	-1,402	-2,623	-1,176
Xi	1,2115	0,0506	23,951	0,00000000984	1,095	1,328	1,042	1,381

Рис. 18. Результати розрахунків інструменту «Регресія / Regression» *для задачі*

Результати розрахунків виводяться у вигляді трьох таблиць або блоків:

❖ **Перша таблиця *«Регресійна статистика / Regression Statistics»*** містить наступні поля:

➢ «Множинний R / Multiple R» – коефіцієнт множинної кореляції R, який визначає щільність зв'язку між залежними та незалежними змінними;

➢ «R-квадрат / R Square» – коефіцієнт детермінації R^2 показує частку впливу комбінації незалежних змінних на залежну змінну (у відносних величинах, які можна перевести у відсотки множенням на 100%);

➢ «Нормований R-квадрат / Adjusted R Square» – скоректований (адаптований (adjusted)) коефіцієнт детермінації

$$R^2_{adj} = 1 - (1 - R^2)\frac{n-1}{n-m-1}$$

де n – кількість спостережень; m – кількість пояснюючих змінних моделі. Враховує зв'язок кількості результатів спостережень і незалежних змінних та забезпечує інформацією про те, яке значення R^2 можна отримати в значно більшому наборі даних, ніж аналізований;

➢ «Стандартна похибка / Standard Error» – похибка моделі σ_ε, яка характеризує варіацію залишкових величин;

➢ «Спостереження / Observations» – кількість спостережень n.

❖ **Друга таблиця *«Дисперсійний аналіз / ANOVA»*** включає такі поля (див. назви стовпчиків):

- У стовпці *«df»* (degrees of freedom) – кількість ступенів свободи:
  в рядку «Регресія / Regression» для регресійної суми квадратів відхилень *df*=m-1;
  в рядку «Залишки / Residual» для залишкової суми квадратів відхилень *df*=n-m;
  в рядку «Загальне / Total» для загальної суми квадратів відхилень *df*=n-1.
- У стовпці «*SS*» (Sum of Squares) – сума квадратів (відхилень):
  в рядку «Регресія / Regression» – регресійна сума квадратів відхилень $S_Y = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$;
  в рядку «Залишки / Residual» – залишкова сума квадратів відхилень $S_e = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$;
  в рядку «Загальне / Total» - загальна сума квадратів відхилень $S_y = \sum_{i=1}^{n} (y_i - \bar{y})^2$.
- У стовпці «*MS*» (Mean of Squares) – середні суми квадратів відхилень з урахуванням числа ступенів вільності:

$$MS = \frac{SS}{df};$$

- У стовпці «*F*» – розрахункове значення критерію Фішера з рівнем довіри 95%;
- У п'ятому стовпці наведено «*Значимість F / Significance F*», яка показує, що при значенні цього показника менше 0,05 побудована регресійна модель відповідає реальній дійсності. Або це – теоретична ймовірність того, що при гіпотезі рівності нулю одночасно усіх коефіцієнтів моделі F-статистика є більшою емпіричного значення F. Якщо показник значущості F є меншим за 0,05, то отриманий результат є значимим; якщо значимість F менше 0,01, тоді отриманий результат є високо значимим.

❖ **Третя таблиця** містить наступні поля:

- У першому стопці «*Коефіцієнти / Coefficients*» наведені значення параметрів рівняння регресії (згори до низу):

в рядку «Y-перетин / Intercept» – оцінка параметру a_0;
в рядку «Змінна X1 / Xi» – оцінка параметру a_1;
в рядку «Змінна X2 / Xi» – оцінка параметру a_2 і т.д.

- В другому стовпці «*Стандартна похибка / Standard Error*» – стандартні похибки параметрів моделі σ_{a_0}, σ_{a_1}, σ_{a_2} і т.д., тобто середньоквадратичні відхилення параметрів моделі;
- В третьому стовпці «*t-статистика / t Stat*» наведені розрахункові значення t-критерію Стьюдента для кожного параметра: $t^*_{a_0}$, $t^*_{a_1}$, $t^*_{a_2}$ і т.д. Тобто це – стандартизовані (нормовані) параметри рівняння регресії, які знаходять діленням кожного фактично знайденого параметра (перший стовпець) на його стандартну похибку (другий стовпець);
- В четвертій колонці «*P-Значення / P-value*» наведено ймовірність, яка дозволяє визначити значимість параметра регресії.

Для рівня значимості α=0,05,

- якщо Р-значення ≥0,05, то параметр a_j ($j = \overline{0, k}$ де k-число параметрів моделі) є незначимим, відповідно гіпотеза $H_0 : a_j = 0$ (гіпотеза, про те, що всі коефіцієнти економетричної моделі $a_j = 0$) приймається;

- якщо Р-значення <0,05, то параметр ($j = \overline{0, k}$ де k-число параметрів моделі) є значимим, відповідно гіпотеза $H_0 : a_j = 0$ (гіпотеза, про те, що всі коефіцієнти економетричної моделі $a_j = 0$) відхиляється.

Таким чином, в цій колонці знаходять функції, які розраховуються за таким аргументами: стандартизованими t-критеріями Стьюдента, обчисленими шляхом ділення t-критеріїв на значення їх стандартних похибок; кількістю ступенів вільності *(n-m)*; числами 1 або 2 (якщо між залежною та незалежними змінними є позитивний або негативний зв'язок), то використовується число 1; якщо невідомо, якого зв'язку слід очікувати, то використовується число 2). Взагалі, якщо Р<0,05, то оцінки параметрів рівняння регресії є достовірними і модель відповідає реальній дійсності.

- Стовпці «*Нижні 95% / Lower 95%*» та «*Верхні 95% / Upper 95%* » – інтервали довіри для параметрів a_j, $j = \overline{0, k}$. Тобто ці стовпці вміщують нижні та верхні границі 95-відсоткового рівня довіри для кожного параметра регресії і виражають довірчі інтервали параметрів. Якщо довірчі інтервали не містять нуля, то з 95-відсотковою впевненістю можна стверджувати, що всі незалежні змінні x_j додають рівнянню регресії значущу інформацію і можна досить точно описувати розглянутий економічний процес чи явище.
- Стовпці «*Нижні 99% / Lower 99%*» та «*Верхні 99% / Upper 99%* » – інтервали довіри для параметрів a_j, $j = \overline{0, k}$. Рівень довіри був заданий вручну в елементі «Рівень надійності / Confidence Level» меню діалогового вікна інструменту «Регресія / Regression» (рис. 17) – 99%. Тобто ці стовпці вміщують нижні та верхні границі вже 99% рівня довіри для кожного параметра регресії і виражають довірчі інтервали параметрів. Якщо довірчі інтервали не містять нуля, то з 99-відсотковою впевненістю можна стверджувати, що всі незалежні змінні x_j додають рівнянню регресії значущу інформацію і можна досить точно описувати розглянутий економічний процес чи явище. Такий рівень довіри задають для серйозних наукових досліджень.

Проведемо аналіз даних, отриманих за допомогою інструменту *«Регресія / Regression»* (рис. 18).

4.1) Аналіз даних, наведених в блоці **«Регресійна статистика / Regression Statistics».**

Проведемо аналіз якості моделі за допомогою розділу «Регресійна статистика / Regression Statistics».

З рядку «Множинний R / Multiple R» (рис. 18) виписуємо значення коефіцієнта парної кореляції $r_{yx} = 0{,}9931$, тобто його значення наближується до одиниці. Це означає, що між вартістю ОЗ підприємства та прибутком існує прямий та сильний кореляційний зв'язок.

З рядку «R-квадрат / R Square» виписуємо значення коефіцієнта детермінації $R^2 = 0{,}9862$. Оскільки коефіцієнт детермінації дорівнює 0,9862, то це означає, що теоретична пряма пояснює 98,62% варіації або можна сказати, що 98,62% варіації

прибутку підприємства відбувається під впливом вартості основних засобів. А решта 1,38% - припадає на інші фактори та випадкові величини. Що свідчить про високий рівень адекватності моделі в цілому.

Оскільки значення коефіцієнтів кореляції та детермінації збігаються з розрахованими вище у завданні 2) за допомогою інструменту «Кореляція / Correlation» та лінії тренду (завдання 1), відповідно, і висновки збігаються.

З рядку «Стандартна похибка / Standard Error» виписуємо значення стандартної похибки моделі $\sigma_\varepsilon = 0{,}1795$.

Перевіримо статистичну значимість коефіцієнту кореляції r_{xy} за t-критерієм Стьюдента (більш детально було розібрано в Задачі 9.3):

1. Формулюємо і розглядаємо 2 гіпотези:

Перевіримо нульову гіпотезу H_0 : $r_{xy}=0$ (яка фактично означає, що між змінними y і x немає зв'язку) проти альтернативної гіпотези H_1 : $r_{xy}\neq 0$ (між змінними y і x є суттєвий зв'язок).

H_0: $r_{xy} = 0$;

H_1: $r_{xy} \neq 0$

2. Обчислюємо фактичне (розрахункове, емпіричне) значення t–критерію за формулою:

$$t_{розрах} = \frac{|r_{xy}|}{\sigma_r} = r_{xy} \cdot \sqrt{\frac{n-m}{1-r_{xy}^{\,2}}} = r_{xy}\sqrt{\frac{k}{1-r_{xy}^{\,2}}} \qquad (9.38)$$

Для розрахунку використовуємо значення з рядку «Множинний R / Multiple R»:

Отримуємо $t_{розрах}$=23,95.

3. Визначаємо критичне (теоретичне, табличне) значення t–критерію за допомогою вбудованої функції Ексель з урахуванням рівня значущості α і ступеня свободи $k=n-m$ (де n – обсяг вибіркової сукупності або кількість спостережень, m – кількість параметрів моделі).

Обираємо рівень значущості α=0,05 або 5%.

Ступінь свободи $k=n-m=10-2$.

Використаємо вбудовану в Ексель статистичну функцію

Функція	*СТЬЮДЕНТ.ОБР.2Х*	2010
	T.INV.2T	Англомовна версія
	СТЬЮДЕНТРАСПОБР	2007 та більш ранні версії

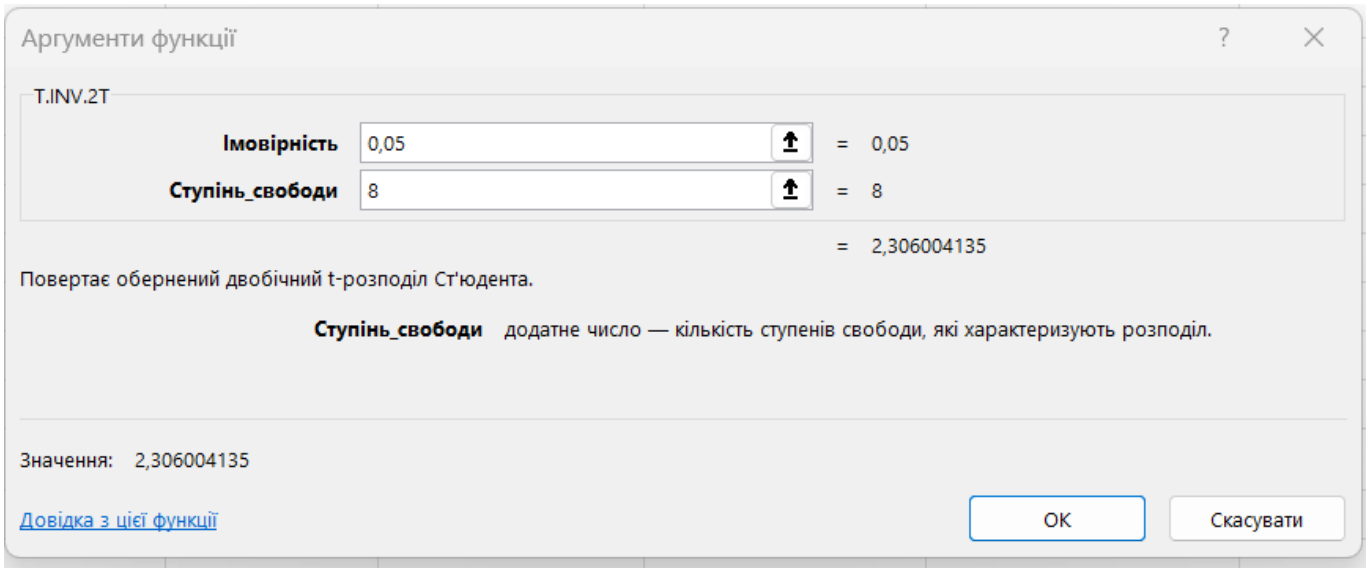

Рис. 19. Діалогове вікно функції T.INV.2T

В нашому випадку ми отримали t-критичне (табличне)=2,306 (див. рис.19).

4. Перевірка умови $|t_{розрах}| > t_{\alpha,k}$

В нашому випадку умова $|t_{розрах}| > t_{\alpha,k}$ (23,95>2,306) справджується, тому нульова гіпотеза не підтверджується з ймовірністю 95%. Тобто з ймовірністю 95% коефіцієнт кореляції r_{xy} не дорівнює 0, і відповідно в генеральній сукупності між змінними y та x існує суттєвий лінійний зв'язок, а отримане значення лінійного коефіцієнта кореляції є значимим.

4.2) Аналіз даних, наведених в блоці ***«Дисперсійний аналіз / ANOVA».***

За даними наведеними в цьому блоці перевіряємо модель на адекватність (більш детально було розібрано в Задачі 9.3).

Це можна зробити двома способами.

Перший спосіб. Для оцінки рівня адекватності побудованої економетричної моделі експериментальним даним користуємося F-критерієм Фішера. З комірки у стовпчику F виписуємо розрахункове значення F-критерію Фішера F^*=573,67.

За допомогою вбудованої статистичної функції

F.ОБР.ПХ	2010
FРАСПОБР	2007, 2003
F.INV.RT	Англомовна версія

для рівня значимості α=0,05 і ступенів вільності $k_1=m-1$ та $k_2=n-m$ (де n – обсяг вибіркової сукупності або кількість спостережень, m – кількість параметрів моделі) визначаємо критичне (табличне) значення F-критерію Фішера.

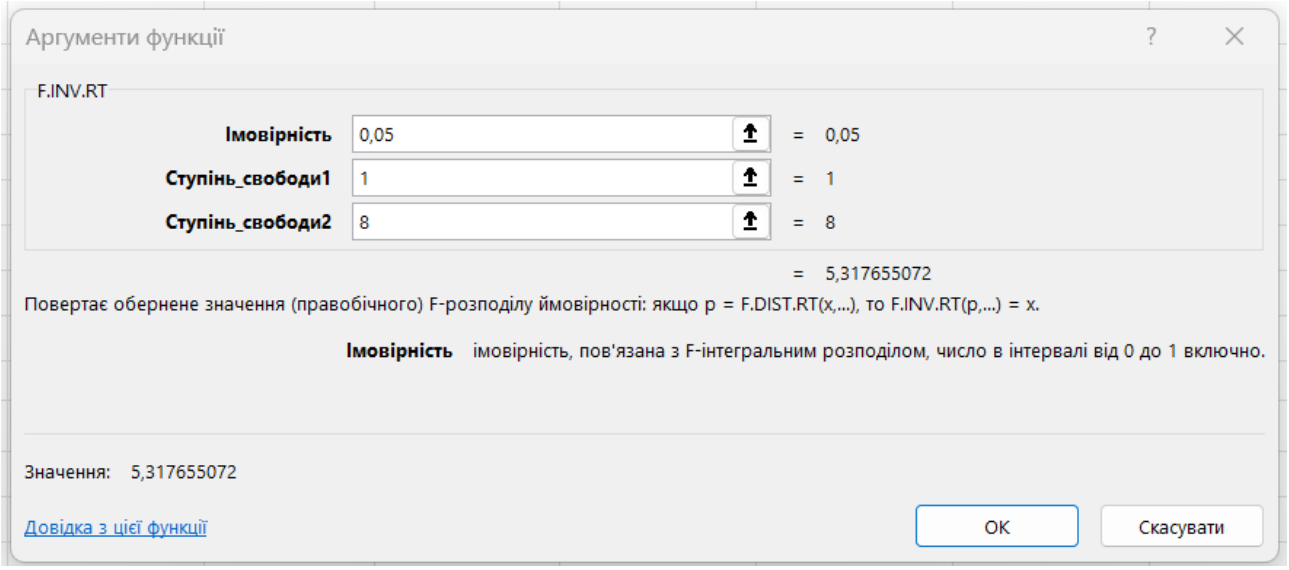

Рис. 20. Діалогове вікно функції «F.INV.RT» з даними моделі

Функція *«F.INV.RT»* має три поля для введення аргументів (рис. 20):

- Поле «Імовірність» призначене для введення рівня значимості α (в нашому прикладі $\alpha=0,05$);
- Поле «Ступінь свободи 1» призначене для введення значення ступеня вільності $k_1=m-1$ (де m – кількість параметрів моделі);
- Поле «Ступінь свободи 2» призначене для введення значення ступеня вільності $k_2=n-m$ (де n – обсяг вибіркової сукупності або кількість спостережень, m – кількість параметрів моделі).

В Ексель всі функції «зашиті» від зворотної ймовірності, тобто від ймовірності помилки. Тому для ймовірності 95% ми будемо брати зворотну від неї 1-0,95=0,05 або 5% помилки.

Отже, $F_{a,k1,k2}=5,32$

Оскільки виконується умова $F^* > F_{кр}$ (573,67>5,32), то робимо висновок про адекватність побудованої економетричної моделі експериментальним даним.

Другий спосіб. Поряд зі значенням F стоїть «Значимість F / Significance F», яка показує, що при значенні цього показника менше 0,05 побудована регресійна модель відповідає реальній дійсності. Або це – теоретична ймовірність того, що при гіпотезі рівності нулю одночасно усіх коефіцієнтів моделі F-статистика є більшою емпіричного значення F. Якщо показник значущості F є меншим за 0,05, то отриманий результат є значимим; якщо значимість F менше 0,01, тоді отриманий результат є високо значимим.

В нашому прикладі «Значимість F / Significance F»= 0,000000009838 значно менший за 0,01, отже отримана модель є високо значимою статистично.

4.3) Аналіз даних, наведених в третьому блоці ***(на рис. 18 він без назви).***

В третьому блоці в стовпчику «*Коефіцієнти / Coefficients*» наведені значення оцінок параметрів парної лінійної: a_0=-1,8994; a_1=1,2115.

В другому стовпці «*Стандартна похибка / Standard Error*» – стандартні похибки параметрів моделі σ_{a_0}, σ_{a_1} і т.д., тобто середньоквадратичні відхилення параметрів моделі:

Ба₀=0,2155 – це стандартна помилка оцінки параметру a_0;

Ба₁=0,0506 – це стандартна помилка оцінки параметру a_1

На основі даних третього блоку також можна перевірити статистичну якість параметрів моделі трьома способами.

Перший спосіб. **Перевірка на значущість за допомогою t-критерію Стьюдента**.

В третьому стовпці «*t-статистика / t Stat*» наведені розрахункові значення t-критерію Стьюдента для кожного параметра: t_{a_0}, t_{a_1} і т.д. Тобто це – стандартизовані (нормовані) параметри рівняння регресії, які знаходять діленням кожного фактично знайденого параметра (перший стовпець) на його стандартну похибку (другий стовпець).

Проведемо перевірку значущості для параметру ***a_0:***

1. Формулюємо та розглядаємо гіпотези:

H_0 : $a_0=0$ – оцінка параметра незначуща;

H_1 : $a_0\neq0$ – оцінка параметра значуща.

2. Розрахункове значення *t*–критерію Стьюдента для параметру a_0 беремо зі стовпчика «*t-статистика / t Stat*» $t_{a_0} = -8{,}814$.

3. Табличне значення беремо з попередніх розрахунків (коли оцінювали статистичну якість коефіцієнту кореляції r_{xy}). В нашому випадку ми отримали t-критичне (табличне)=2,306 (див. рис.19).

4. Перевірка умови $\left|t_{a_0}\right| > t_{кр}$

Умова справджується (8,8138>2,306), що означає, що нульова гіпотеза відкидається і параметр a_0 генеральної сукупності є статистично значущим.

Проведемо перевірку значущості для параметру ***a_1:***

1. Формулюємо та розглядаємо гіпотези:

H_0 : $a_1=0$ – оцінка параметра незначуща;

H_1 : $a_1\neq0$ – оцінка параметра значуща.

2. Розрахункове значення *t*–критерію Стьюдента для параметру a_0 беремо зі стовпчика «*t-статистика / t Stat*» $t_{a_1} = 23{,}951$.

3. Табличне значення беремо з попередніх розрахунків (коли оцінювали статистичну якість коефіцієнту кореляції r_{xy}). В нашому випадку ми отримали t-критичне (табличне)=2,306 (див. рис.19).

4. Перевірка умови $\left|t_{a_1}\right| > t_{кр}$

Умова справджується (23,951>2,306), що означає, що нульова гіпотеза відкидається і параметр a_1 генеральної сукупності є статистично значущим.

Другий спосіб. **Визначення інтервалів надійності для оцінок параметрів моделі за *t*–розподілом при рівні значущості 0,05.**

Використовуємо дані зі стовпців «*Нижні 95% / Lower 95%*» та «*Верхні 95% / Upper 95%* » – інтервали довіри для параметрів a_j, $j = \overline{0, k}$. Тобто ці стовпці вміщують

нижні та верхні границі 95-відсоткового рівня довіри для кожного параметра регресії і виражають довірчі інтервали параметрів. Якщо довірчі інтервали не містять нуля, то з 95-відсотковою впевненістю можна стверджувати, що всі незалежні змінні x_j додають рівнянню регресії значущу інформацію і можна досить точно описувати розглянутий економічний процес чи явище.

Довірчі межі для параметру a_0 : з 95% рівнем довіри =[-2,396; -1,402] з 99% = [-2,623; -1,176]. Вони не містять 0. Отже, параметр a_0 є статистично значущим на обох рівнях довіри – 95% та 99%.

Довірчі межі для параметру a_1 : з 95% рівнем довіри =[1,095; 1,328] з 99% = 1,042; 1,381]. Вони не містять 0. Отже, параметр a_1 є статистично значущим на обох рівнях довіри – 95% та 99%.

Третій спосіб. **За даними колонки «*P-Значення / P-value*».**

В четвертій колонці «*P-Значення / P-value*» наведено ймовірність, яка дозволяє визначити значимість параметра регресії.

Для рівня значимості α=0,05,

- якщо Р-значення ≥0,05, то параметр a_j ($j = \overline{0, k}$ де k-число параметрів моделі) є незначимим, відповідно гіпотеза $H_0 : a_j = 0$ (гіпотеза, про те, що всі коефіцієнти економетричної моделі $a_j = 0$) приймається;
- якщо Р-значення <0,05, то параметр ($j = \overline{0, k}$ де k-число параметрів моделі) є значимим, відповідно гіпотеза $H_0 : a_j = 0$ (гіпотеза, про те, що всі коефіцієнти економетричної моделі $a_j = 0$) відхиляється.

P-value
0,0000216043
0,0000000098

Отже, порівнюючи значення наведені в колонці «*P-Значення / P-value*» з 0,05, можемо сказати, що вони (0,0000216043<0,05 та 0,0000000098<0,05) є значно меншими за 0,05. З чого робимо висновок про статистичну значимість параметрів a_0 та a_1.

Оскільки в меню діалогового вікна інструменту *«Регресія / Regression»* (Рис. 17) ми поставили прапорець навпроти «Залишки / Residuals» була отримана така таблиця – Рис. 21.

В стовпчику «*Observation / Спостереження»* розміщено номери спостережень з вихідних даних.

В стовпчику *«Predicted Yi /Передбачене Yi»* знаходяться прогнозні значення результативної змінної y_i (прибутку підприємства), які розраховані за отриманою регресійною моделлю $\hat{y} = -1{,}8994 + 1{,}2115 \cdot x$ для кожного значення факторної змінної x_i в вихідних даних.

В стовпчику *«Residuals / Залишки»* розраховані залишки, тобто різниця між фактичним y_i та розрахованим за моделлю.

RESIDUAL OUTPUT /
ВИВЕДЕННЯ ЗАЛИШКА

Observation / Спостереження	*Predicted Yi /Передбачене Yi*	*Residuals / Залишки*
1	1,13	0,0706
2	1,49	0,0071
3	1,74	0,1648
4	2,46	-0,2621
5	2,83	-0,0256
6	3,19	-0,0890
7	3,55	-0,1525
8	4,16	0,3417
9	4,89	-0,0852
10	5,37	0,0302

Висновки 3:

1) Перевірка коефіцієнта кореляції r_{xy} показала його значущість *за t*-критерієм Стьюдента, оскільки справдилась умова $|t_{розрах}| > t_{\alpha,k}$ (23,95>2,306). Отже, між змінними *y* та *x* існує суттєвий лінійний зв'язок, а отримане значення лінійного коефіцієнта кореляції є значимим.

2) За даними дисперсійного аналізу (ANOVA-таблиця) було доведено адекватність побудованої економетричної моделі двома способами.

Оскільки в нашому випадку умова $F_p > F_{a,k1,k2}$ виконується (573>5,32) та «Значимість F / Significance F» менше 0,05 (0,000000009838<0,05), то це означає, що з ймовірністю 95% можемо сказати, що економетрична модель є адекватною і з 5% ймовірністю ми можемо помилятися.

3) За допомогою інструменту аналізу «Регресія / Regression» були отримані значення оцінок параметрів a_0=-1,8994 та a_1=1,2115. Їх значення збіглися зі значеннями, отримані при побудові тренду на кореляційному полі, отже, підтвердилось регресійне рівняння $\hat{y} = -1{,}8994 + 1{,}2115 \cdot x$.

4) Перевірка оцінок параметрів a_0 та a_1 показала їх значущість на трьох рівнях – на значущість *за t*-критерієм Стьюдента, шляхом визначення надійних інтервалів та за рівнем значущості.

В першому методі справдились умови для параметрів $a_0 : t_{a0} \geq t_{\alpha,k}$ (8,8138>2,306); $a_1 : t_{a1} \geq t_{\alpha,k}$ (23,95>2,306). В другому методі оцінка параметру a_0 попадає в інтервал [-2,3964; -1,4025], а параметру a_1 – в інтервал [1,0949; 1,3282] з рівнем надійності 95%. А з рівнем надійності 99% параметр a_0 попадає в інтервал [-2,623; -1,176], а параметр a_1 – в інтервал [1,042; 1,381]. В всіх випадках 0 в ці інтервали не входить, і це означає, що з ймовірністю 99% оцінки параметрів a_0 та a_1 не дорівнюватимуть 0.

В третьому методі «*P-Значення / P-value*» для обох параметрів значно менші за 0,05 (0,0000216043<0,05 та 0,0000000098<0,05). З чого робимо висновок про їх статистичну значимість.

Отже параметри a_0 та a_1 є статистично значущими та між змінними x та y існує суттєвий лінійний зв'язок.

Оскільки обидва коефіцієнти є статистично значущими, що було доведено вище, для обох параметрів можемо дати їх економічну інтерпретацію.

a_0 : Якщо коефіцієнт a_0 є значущим, тоді можна сказати, що результативна змінна ***y*** прийме значення a_0 за нульового значення факторної (незалежної) змінної ***x.***

Тобто, для нашого прикладу, якщо вартість ОЗ підприємства дорівнюватиме 0, то прибуток підприємства буде дорівнювати -1,8994 млн. грн. Постійна величина a_0=-1,8994 дає прогнозне значення y при $x=0$.

a_1 : Коефіцієнт при x показує, на скільки одиниць зміниться результативна змінна y при зміні змінної x на 1 одиницю. Отже, якщо x зросте на 1 одиницю, то y зросте на a_1 одиниць.

Отже для нашого прикладу коефіцієнт a_1 відображає, що якщо вартість ОЗ збільшити на 1 млн. грн., то прибуток підприємства зросте на 1,2115 млн. грн.

5) визначити точковий прогноз для заданого значення незалежної змінної – визначити прогнозне значення прибутку підприємства (*y*) якщо вартість ОЗ підприємства буде на 20% більшою за максимальне значення цієї ознаки;

Для визначення точкового прогнозу спочатку необхідно задати значення $x_{прогнозне}$ (його можна задавати будь-яким чином). В нашому випадку задамо таким чином – визначити прогнозне значення прибутку підприємства (y) якщо вартість ОЗ підприємства буде на 20% більшою за максимальне значення цієї ознаки.

$x_{прогнозне}$=6,0*1,2=7,2 млн. грн.

Прогнозне значення будемо визначати за рівнянням регресії:

$$\hat{y} = -1{,}8994 + 1{,}2115 \cdot x$$

Підставимо в рівняння $x_{прогнозне}$:

$y_{прогнозне}$=-1,8994+1,2115*7,2=6,82 млн.грн.

Висновки 4:

Значення $y_{прогнозне}$, яке дорівнює 6,82 млн. грн. можна інтерпретувати як точкову оцінку прогнозного значення математичного сподівання та індивідуального значення прибутку підприємства, за умови, що вартість ОЗ підприємства буде дорівнювати 7,2 млн. грн.

6) визначити інтервальний прогноз для заданого значення незалежної змінної.

Прогноз («істинне» значення прогнозу) буде знаходитись в межах:

$$\tilde{y}_{\text{прогнозне}} = \hat{y}_{прогнозне} \mp \Delta\hat{y}_{прогнозне}$$

або

$$\hat{y}_{\text{прогнозне}} - \Delta\hat{y}_{\text{прогнозне}} \ll \tilde{y}_{\text{прогнозне}} \ll \hat{y}_{прогнозне} + \Delta\hat{y}_{прогнозне}$$

В практиці прогнозування визначають для типи довірчих інтервалів:

1) для середнього значення;

2) для індивідуального значення.

Формули для їх розрахунків різняться тільки в тому, як розраховують значення під коренем. Для *середнього значення (математичного сподівання)* залежної змінної ознаки ***y*** похибку або граничну помилку ($\Delta\hat{y}_{прогнозне}$) визначають за формулою:

$$\Delta y_{\hat{p}} = \frac{t_{\alpha,k} \cdot \sigma_u}{\sqrt{n}} \cdot \sqrt{1 + \frac{(x_p - \bar{x})^2}{\sigma_x^2}}. \qquad (9.53)$$

Для *індивідуального значення* залежної змінної ознаки ***y*** похибку або граничну помилку ($\Delta\hat{y}_{прогнозне}$) визначають інакше – за формулою:

$$\Delta y_p = t_{\alpha,k} \cdot \frac{\sigma_u}{\sqrt{n}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sigma_x^2}} \qquad (9.54)$$

При щільній кореляції (більше 0,75) прогнозні значення можуть бути прийняті при бізнес-плануванні результативного показника на найближчу перспективу. При цьому слід пам'ятати, що чим далі від базисних показників взяте прогнозне значення фактора *x*, тим менш надійний прогноз, тим більша ймовірність значного відхилення від середнього розрахункового значення, тобто довірча зона прогнозування розширюється.

Для визначення *середнього значення (математичного сподівання)* похибка прогнозу дорівнює $\Delta\hat{y}_{прогнозне}$=0,1736 (формула 9.53), тому отримаємо відповідні нерівності *(прогнозний інтервал математичного сподівання)*:

Y т.прогноз-Δ Yпрогноз< =Y прогноз (середнє) <=Y т.прогноз+Δ Yпрогноз

6,8236-0,1736 < =*Y прогноз (середнє)* <= 6,8236+0,1736

6,6501 < =*Y прогноз (середнє)* <= 6,9972

Для визначення *прогнозного інтервалу індивідуального значення* похибка прогнозу дорівнюватиме $\Delta\hat{y}_{прогнозне}$=0,1784 (формула 9.54), тому отримаємо відповідні нерівності:

Y т.прогноз-Δ Yпрогноз < =Y прогноз (індив) <= Y т.прогноз+Δ Yпрогноз

6,8236-0,1784 < =*Y прогноз (індив)* <= 6,8236+0,1784

6,6452 < =*Y прогноз (індив)* <= 7,0021

Отже, з ймовірністю 95% (або на рівні значимості α=5%) прогноз математичного сподівання попадає в інтервал [6,6501; 6,9972], а прогноз індивідуального значення – в інтервал [6,6452; 7,0021].

Висновки 5:

Якщо в прогнозному періоді вартість ОЗ підприємства буде дорівнювати 7,2 млн.грн., то середній прибуток підприємства з ймовірністю 95% потрапляє в інтервал

[6,6501; 6,9972]. Водночас окреме (індивідуальне) значення прибутку підприємства буде міститися в ширшому інтервалі [6,6452; 7,0021].

ЗАГАЛЬНІ ВИСНОВКИ:

1) Розміщення точок на діаграмі розсіювання (кореляційному полі) дає можливість зробити припущення про існування лінійної форми зв'язку у вигляді функції $\hat{y} = a_0 + a_1 x$, де $\hat{y}$ – розрахунковий прибуток, млн.грн.; x – вартість ОЗ, млн.грн.

2) З лінійного тренду, побудованого на кореляційному полі отримали такі значення оцінок параметрів парної лінійної регресії

a_0= -1,8994

a_1= 1,2115

Отже, було отримано регресійне рівняння

$\hat{y} = -1{,}8994 + 1{,}2115 \cdot x$

3) Коефіцієнт парної кореляції або коефіцієнт кореляції Пірсона r_{xy}=0,9931. Його значення дуже близьке до 1 і це вказує на те, що між вартістю ОЗ та прибутком підприємства є сильний лінійний зв'язок. А оскільки він ще й додатній – це вказує на те, між змінними прямий лінійний зв'язок. Прямий зв'язок означає, що зі зростанням фактору (змінної) x зростає і фактор (змінна) y).

Перевірка коефіцієнта кореляції r_{xy} показала його значущість *за* t-критерієм Стьюдента, оскільки справдилась умова $|t_{розрах}| > t_{\alpha,k}$ (23,95>2,306). Отже, між змінними y та x існує суттєвий лінійний зв'язок, а отримане значення лінійного коефіцієнта кореляції є статистично значимим з рівнем довіри 95%.

4) Коефіцієнт детермінації R^2=0,9862. Він близький до одиниці, що вказує на добру загальну якість моделі. Також це означає, що теоретична пряма пояснює 98,62% варіації або можна сказати, що 98,62% варіації прибутку підприємства відбувається під впливом вартості основних засобів. А решта 1,38% - припадає на інші фактори та випадкові величини. Що свідчить про високий рівень адекватності моделі в цілому.

5) За даними дисперсійного аналізу (ANOVA-таблиця) було доведено адекватність побудованої економетричної моделі двома способами.

Оскільки в нашому випадку умова $F_p > F_{a,k1,k2}$ виконується (573>5,32) та «Значимість F / Significance F» менше 0,05 (0,000000009838<0,05), то це означає, що з ймовірністю 95% можемо сказати, що економетрична модель є адекватною і з 5% ймовірністю ми можемо помилятися.

6) За допомогою інструменту аналізу «Регресія / Regression» були отримані значення оцінок параметрів a_0=-1,8994 та a_1=1,2115. Їх значення збіглися зі значеннями, отримані при побудові тренду на кореляційному полі, отже, підтвердилось регресійне рівняння $\hat{y} = -1{,}8994 + 1{,}2115 \cdot x$.

7) Перевірка оцінок параметрів a_0 та a_1 показала їх значущість на трьох рівнях – *за* t-критерієм Стьюдента, шляхом визначення надійних інтервалів та за рівнем значущості.

В першому методі справдились умови для параметрів a_0 : $t_{a0} \geq t_{\alpha,k}$ (8,8138>2,306); a_1 : $t_{a1} \geq t_{\alpha,k}$ (23,95>2,306). В другому методі оцінка параметру a_0 попадає в інтервал [-2,3964; -1,4025], а параметру a_1 – в інтервал [1,0949; 1,3282] з рівнем надійності 95%. А з рівнем надійності 99% параметр a_0 попадає в інтервал [-2,623; -1,176], а параметр

a_1 – в інтервал [1,042; 1,381]. В всіх випадках 0 в ці інтервали не входить, і це означає, що з ймовірністю 99% оцінки параметрів a_0 та a_1 не дорівнюватимуть 0.

В третьому методі «*P-Значення / P-value*» для обох параметрів значно менші за 0,05 (0,0000216043<0,05 та 0,0000000098<0,05). З чого робимо висновок про їх статистичну значимість.

Отже параметри a_0 та a_1 є статистично значущими та між змінними *x* та *y* існує суттєвий лінійний зв'язок.

8) Оскільки обидва коефіцієнти є статистично значущими, що було доведено вище, для обох параметрів можемо дати їх економічну інтерпретацію.

a_0 : Якщо коефіцієнт a_0 є значущим, тоді можна сказати, що результативна змінна ***y*** прийме значення a_0 за нульового значення факторної (незалежної) змінної ***x.***

Тобто, для нашого прикладу, якщо вартість ОЗ підприємства дорівнюватиме 0, то прибуток підприємства буде дорівнювати -1,8994 млн. грн. Постійна величина a_0=-1,8994 дає прогнозне значення *y* при *x=0*.

a_1 : Коефіцієнт при *x* показує, на скільки одиниць зміниться результативна змінна *y* при зміні змінної *x* на 1 одиницю. Отже, якщо *x* зросте на 1 одиницю, то *y* зросте на a_1 одиниць.

Отже для нашого прикладу коефіцієнт a_1 відображає, що якщо вартість ОЗ збільшити на 1 млн. грн., то прибуток підприємства зросте на 1,2115 млн. грн.

9) Значення $y_{прогнозне}$, яке дорівнює 6,82 млн. грн. можна інтерпретувати як точкову оцінку прогнозного значення математичного сподівання та індивідуального значення прибутку підприємства, за умови, що вартість ОЗ підприємства буде дорівнювати 7,2 млн. грн.

10) Якщо в прогнозному періоді вартість ОЗ підприємства буде дорівнювати 7,2 млн.грн., то середній прибуток підприємства з ймовірністю 95% потрапляє в інтервал [6,6501; 6,9972]. Водночас окреме (індивідуальне) значення прибутку підприємства буде міститися в ширшому інтервалі [6,6452; 7,0021].