2 ПРОСТА ВИБІРКОВА ЛІНІЙНА РЕГРЕСІЯ

Економетрія вивчає кількісні взаємозв’язки і залежності між економічними показниками, явищами і процесами. При вивчені і описі таких взаємозв’язків і залежностей економічні показники повинні розглядатися як випадкові величини, що пов’язано з впливом на останні різноманітних випадкових факторів. Внаслідок цього залежності між економічними показниками не є однозначними, не є функціональними. Це означає, що кожному фіксованому значенню однієї економічної змінної (або фіксованому набору змінних) відповідає не одне єдине, а множина значень іншої змінної, тобто деякий імовірнісний розподіл. Тому в економіці спостерігаються і розглядаються так звані статистичні (або кореляційні) залежності.

Кореляційна залежність між економічними показниками (змінними) (або залежність у середньому) описується за допомогою функції (рівняння) регресії. Функція регресії описує залежність між незалежною змінною (або незалежними змінними) і умовним математичним сподіванням (середнім) залежної змінної. У якості змінних такої функції виступають економічні показники. *Регресією* прийнято називати функціональну залежність між незалежною змінною (або незалежними змінними) і умовним математичним сподіванням (середнім значенням) залежної змінної.

В залежності від числа змінних розрізняють модель *парної* і *багатофакторної* (множинної) регресії, а в залежності від виду функції регресії – лінійну та нелінійну моделі регресії.

Функція регресії має графічну форму представлення. Так для парної регресії графічне зображення функції регресії на площині $x0y$ представляє собою так звану криву (лінію) регресії. Для множинної регресії крива регресії перетворюється на поверхню (або гіперповерхню) регресії.

Фактична, статистична залежність між економічними показниками подається як модель регресії (або регресійна модель). *Модель регресії* – це математична модель, яка описує статистичну залежність деякого економічного показника (змінної) від інших. Іншими словами модель регресії – це математична модель, яка описує кореляційно-регресійні зв’язки між економічними показниками, один з яких розглядається як залежна змінна, а усі інші – як незалежні.

В залежності від статистичної бази розрізняють *теоретичну* і *вибіркову* (емпіричну) моделі регресії. Теоретична модель відповідає генеральній сукупності спостережень за змінними моделі, а вибіркова – побудована на основі окремої статистичної вибірки з генеральної сукупності спостережень.

Теоретична модель регресії є ідеалізованою конструкцією, оскільки у практиці економетричного моделювання, як правило, не доводиться мати справу з генеральною сукупністю спостережень, а тільки з деякою окремою статистичною вибіркою з неї. Тому реально, ніколи не можливо побудувати «дійсну» теоретичну регресію, а тільки вибіркову (емпіричну), яка є тільки наближенням (оцінкою) «дійсної» теоретичної регресії.

Питаннями побудови якісних і статистично надійних регресійних моделей займається *кореляційно-регресійний аналіз*, який представляє собою сукупність методів, за допомогою яких на основі вибіркових статистичних даних досліджуються та узагальнюються взаємозв'язки кореляційно пов'язаних змінних.

Кореляційно-регресійний аналіз відіграє дуже важливе значення в економетричному аналізі, оскільки саме його методи і підходи покладено в основу більшості економетричних методів. Основними задачами такого аналізу є:

– визначення наявності і сили кореляційного взаємозв'язку між економічними показниками;

– вибір виду (аналітичної форми) функції регресії;

– оцінювання (визначення) параметрів вибраного рівняння регресії на основі статистичної вибірки;

– аналіз якості регресійної моделі і перевірка адекватності моделі статистичним даним.

Просту лінійну регресію використовують для встановлення лінійного зв'язку між двома змінними: *незалежною* (*екзогенною* або *зовнішньою*) змінною *x*, значення яких є наперед визначеними перед використанням моделі, і *залежною* (*ендогенною* або *внутрішньою*) змінною *y*, значення яких визначаються тільки із самої моделі. Значення змінних є значеннями деяких ознак економічних процесів або явищ. При цьому ознаку *x* називають *факторною*, а *y* – результативною.

При регресійному аналізі визначаються два параметри: a_0 – *перетин* і a_1 – *нахил* регресії. Перетин показує значення *y* при умові, що *x* дорівнює нулю, а значення нахилу a_1 показує на скільки зміниться значення результативної ознаки *y* при умові, що значення факторної ознаки *x* збільшиться на 1.

Прикладом лінійної моделі може бути залежність попиту на ринку на певний товар від його ціни. В даному випадку, ціна – незалежна змінна, тобто, вхідна змінна, яка характеризує причину (або факторна), а попит на товар – це залежна змінна,тобто, вихідна змінна, що характеризує наслідок (або результативна).

На першому етапі побудови регресійної моделі для визначення форми зв'язку між *x* та *y* будують *діаграму розподілу* (будують точки з координатами $(x;\, y)$). Це можна зробити, наприклад, за допомогою ***Мастер диаграмм*** табличного редактора Microsoft Excel.

Розглянемо порядок побудови графіка:

1) відмічаємо необхідний для побудови графіка діапазон числових даних залежної змінної *y* (рис. 2.1). Якщо потрібно виділити декілька несуміжних діапазонів даних, то тримаючи натиснутою клавішу Ctrl , відмічаємо кожний із них. При переході до нового діапазону ліву клавішу миші відпускаємо.

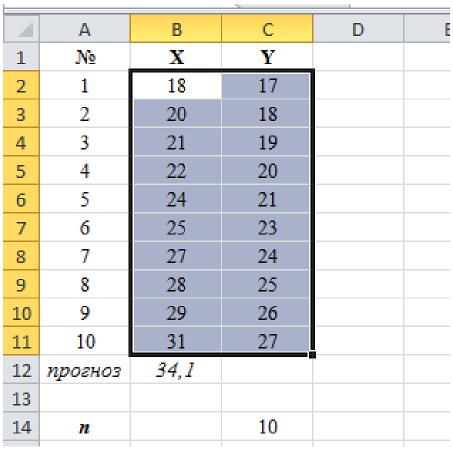

Рисунок 2.1 – Вікно виділених даних залежної змінної

2) на панелі інструментів наводимо курсор на вкладку ***Вставка*** ⇒ ***Точечная*** (або ***График***) (рис. 2.2) і натискуємо ліву клавішу мишки. З’являється графік (рис. 2.3).

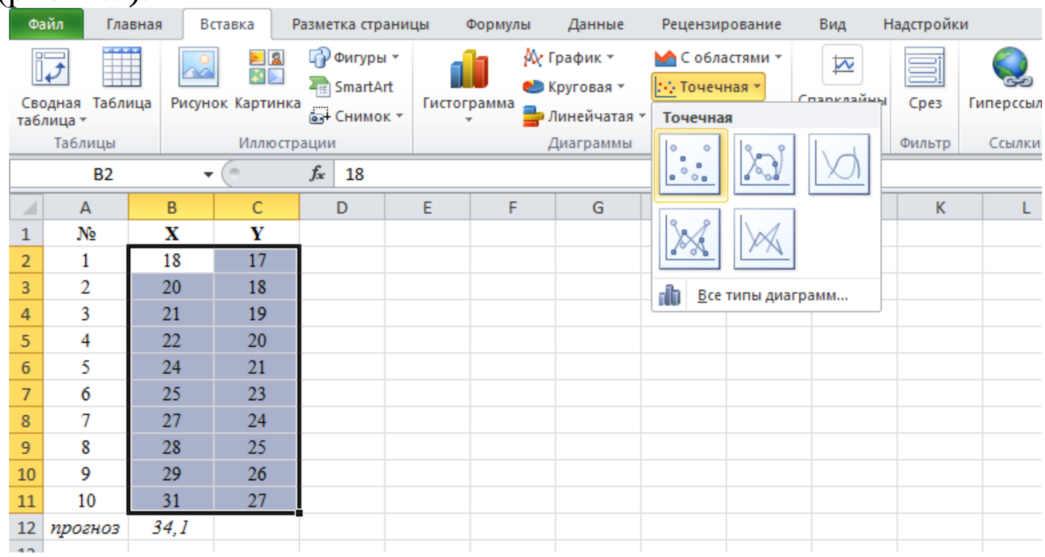

Рисунок 2.2 – Вікно типів діаграм та вихідних даних для діаграми ***Точечная***

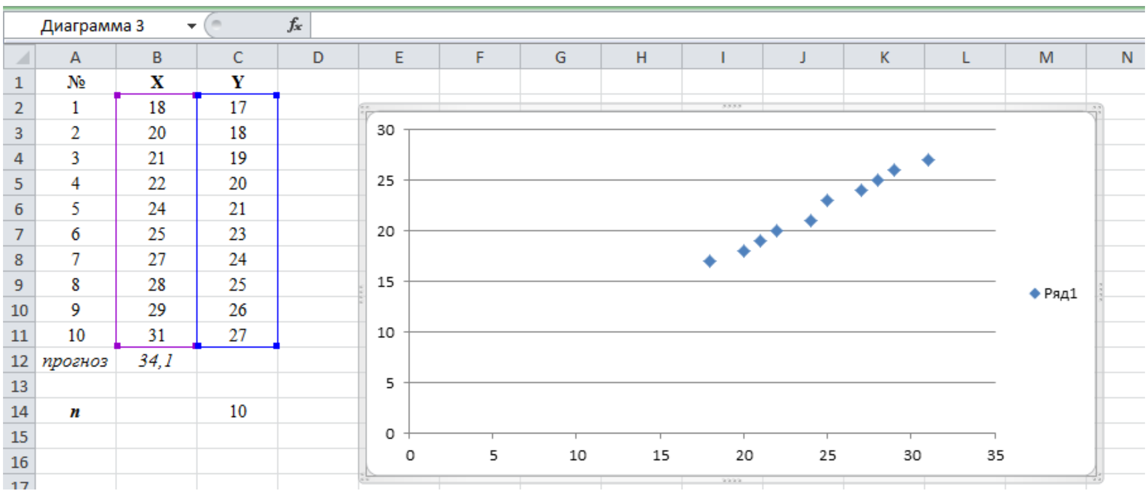

Рисунок 2.3 – Перший графік

3) виділяємо графік і переходимо на вкладку ***Работа с диаграммами*** ⇒ ***Макет*** ⇒ ***Названия осей***, для горизонтальної осі пишемо назву *Значення*

фактора X і для вертикальної – *Значення показника Y*, а назву діаграми *Діаграма розсіювання* за допомогою ***Работа с диаграммами*** ⇒ ***Макет*** ⇒ ***Название диаграммы*** (рис. 2.4). Для зміни назви ряду обрати ***Выбрать данные*** ⇒ ***Изменить***, потім уведемо ім’я ряду *Yf* в поле ***Имя*** (рис. 2.5). Діаграму розміщуємо на окремому листі через натискання правої кнопки миші і обрання ***Переместить диаграмму*** ⇒ ***Лист 2***, який можна назвати інакше, наприклад, діаграма.

Для додавання одного чи декількох рядів даних натиснемо праву кнопку миші на точки графіка, обрати ***Выбрать данные*** ⇒ ***Добавить***, потім уведемо ім’я ряду в поле **Имя** натиснемо у полі ***Значения Y*** і введемо діапазон даних.

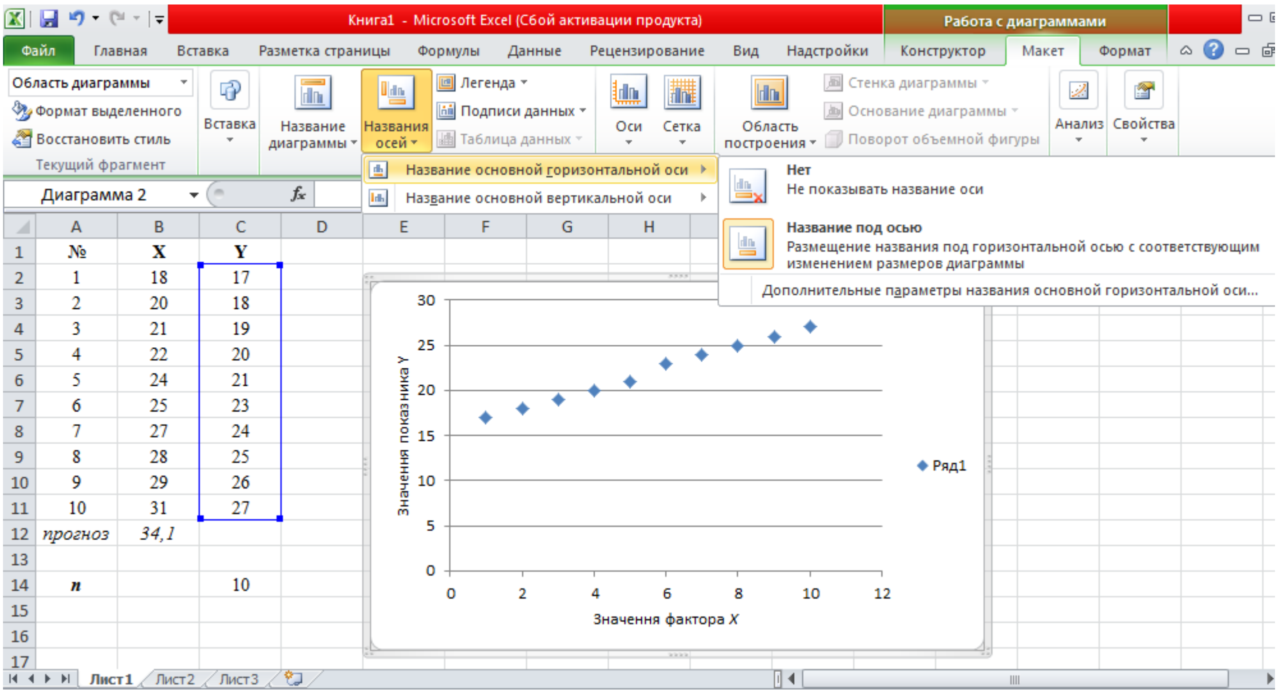

Рисунок 2.4 – Позначення назв осей

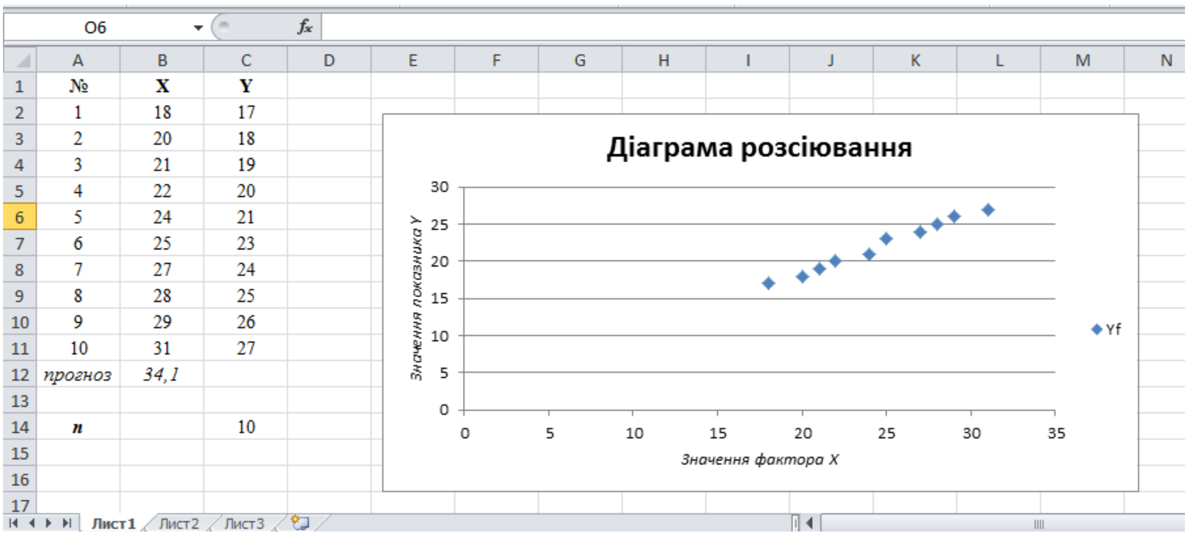

Рисунок 2.5 – Графік на робочому аркуші

Для визначення параметрів рівняння регресії найчастіше застосовується класичний *метод найменших квадратів* (МНК) основоположниками якого є К. Гаусс та П. Лаплас. МНК дозволяє підібрати певну неперервну аналітичну функцію для апроксимації (наближення) дискретного набору вихідних даних. Задача зводиться до знаходження такої лінії (прямої, кривої, ламаної тощо), яка мінімізує суму квадратів відхилень між фактичними y та отриманими розрахунковим шляхом за рівнянням регресії теоретичними значеннями $\hat{y}$:

$$F = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 \to \min .$$

Основні припущення моделі лінійної регресії:

1) значеннями незалежної змінної x є або фіксовані числа, або, якщо вони є випадковими змінними, то вони статистично незалежні від випадкової величини u (неспостережуваної випадкової величини);

2) випадкова величина u розподілена за нормальним законом з нульовим математичним сподіванням і постійною дисперсією;

3) випадкові величини u_i та x_i статистично незалежні.

Щоб оцінити щільність зв'язку між x та y, використовують *коефіцієнт кореляції* та *коефіцієнт детермінації*, які показують, наскільки варіація змінної x пояснює варіацію y. Щоб оцінити відносний ефект впливу фактора x на результативний показник y, використовують *коефіцієнт еластичності*, який характеризує % зростання показника y при збільшенні фактора x на 1%.

Щоб оцінити наскільки добре лінія регресії пояснює зв'язок між x та y, використовують *стандартні помилки оцінок параметрів моделі*, які показують відхилення емпіричних значень від лінії регресії. Стандартні похибки характеризують середні лінійні коливання оцінок параметрів моделі навколо свого математичного сподівання. Чим менші ці похибки, тим стійкіші оцінки параметрів моделі. Остаточні висновки стосовно стійкості оцінок можна зробити лише тоді, коли порівняти їх з абсолютними значеннями оцінок параметрів моделі. Вибіркова оцінка параметрів $\widehat{A}$ називається *незміщеною*, якщо вона задовольняє рівність $M\left(\widehat{A}\right) = A$. *Незміщеність* – це мінімальна вимога, яка ставиться до оцінок параметра $\widehat{A}$. Про наявність зміщеності оцінки можна стверджувати тоді, коли її стандартна похибка перевищує 10% від абсолютного значення оцінки.

При побудові регресійної моделі перевіряється гіпотеза про її *адекватність*. Для цього можна використовувати F –критерій Фішера.

Під час *оцінки параметрів регресії* перевіряються гіпотези, чи статистично значимо вони відрізняються від нуля. Для цього можна використовувати t -критерій Стьюдента. При цьому будують надійний інтервал для коефіцієнта кореляції.

Побудовану регресійну модель можна використовувати для *прогнозування* величини результативної ознаки y при заданому значенні факторної ознаки x, при цьому бажано будувати інтервал довіри для прогнозу.

Основні формули

1. *Узагальнена регресійна модель*:

$$Y = a_0 + a_1 X + u \,.$$

2. *Вибіркова лінійна регресійна модель*:

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 X \ ,$$

де a_0, a_1 – невідомі параметри моделі;

$\hat{a}_0, \hat{a}_1$ – їх оцінки;

$\hat{Y}$ – теоретичне (регресійне) значення результативної змінної;

$u = Y - \hat{Y}$ – вектор залишків (стохастична складова).

3. *Коефіцієнт кореляції*:

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x \sigma_y},$$

де $\sigma_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$, $\sigma_y^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n}$.

В табличному редакторі Microsoft Excel можна обчислити коефіцієнт кореляції за допомогою функції:

=КОРРЕЛ(*X*; *Y*) або =PEARSON(*Y*;*X*).

4. *Обчислення параметрів лінійної моделі методом найменших квадратів.*

Оцінимо параметри $\hat{a}_0, \hat{a}_1$ лінійної моделі $\hat{Y} = \hat{a}_0 + \hat{a}_1 X$ класичним методом найменших квадратів (МНК). Для мінімізації функції

$$F(a_0, a_1) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i - (a_0 + a_1 x_i) \right)^2$$

визначимо перші частинні похідні функції F по a_0 та a_1, прирівняємо їх до нуля та отримаємо систему нормальних рівнянь:

$$\begin{cases} n\hat{a}_0 + \hat{a}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \\ \hat{a}_0 \sum_{i=1}^{n} x_i + \hat{a}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i. \end{cases}$$

Використовуючи вихідні дані, зробимо необхідні обчислення і підставимо в систему знайдені значення сум. Розв’язати цю систему рівнянь відносно невідомих оцінок параметрів $\hat{a}_0, \hat{a}_1$ можна будь-яким з відомих математичних методів або за допомогою ***Поиск решения*** табличного редактора Microsoft Excel.

Для цього в Excel треба скласти розрахункову таблицю і записати необхідні формули у комірки, наприклад, як у табл. 2.1.

Таблиця 2.1 – Розрахункова таблиця для МНК

A	*B*	*C*	*D*	*E*
№	***X***	***Y***	***X*^2**	***XY***
1	вихідні дані		=B2^2	=B2*C2
2			аналогічно	аналогічно
…	…	…	…	…
сума	=СУММ(B2:BN)	=СУММ(C2:CN)	=СУММ(D2:DN)	=СУММ(E2:EN)

Далі застосуємо ***Поиск решения***. Для цього в деякі комірки, наприклад, G15:J15 запишемо відповідно:

G15 ^a_0
H15 ^a_1
I15 лев часть
J15 пр часть

Для подальшого застосування оператора ***Поиск решения*** запишемо необхідні функції згідно таблиці 2.2.

Таблиця 2.2 – Підготовка комірок до застосування оператора ***Поиск решения***

комірка	*формула*
I16	=C14*G16+H16*B13
J16	=C13
I17	=B13*G16+D13*H16
J17	=E13

Далі ставимо курсор в комірку I16 і обираємо ***Данные⇒ Поиск решения***. Заповнюємо вікно (рис. 2.6), натискаємо ***Найти решение*** і отримаємо розв’язок (рис. 2.7).

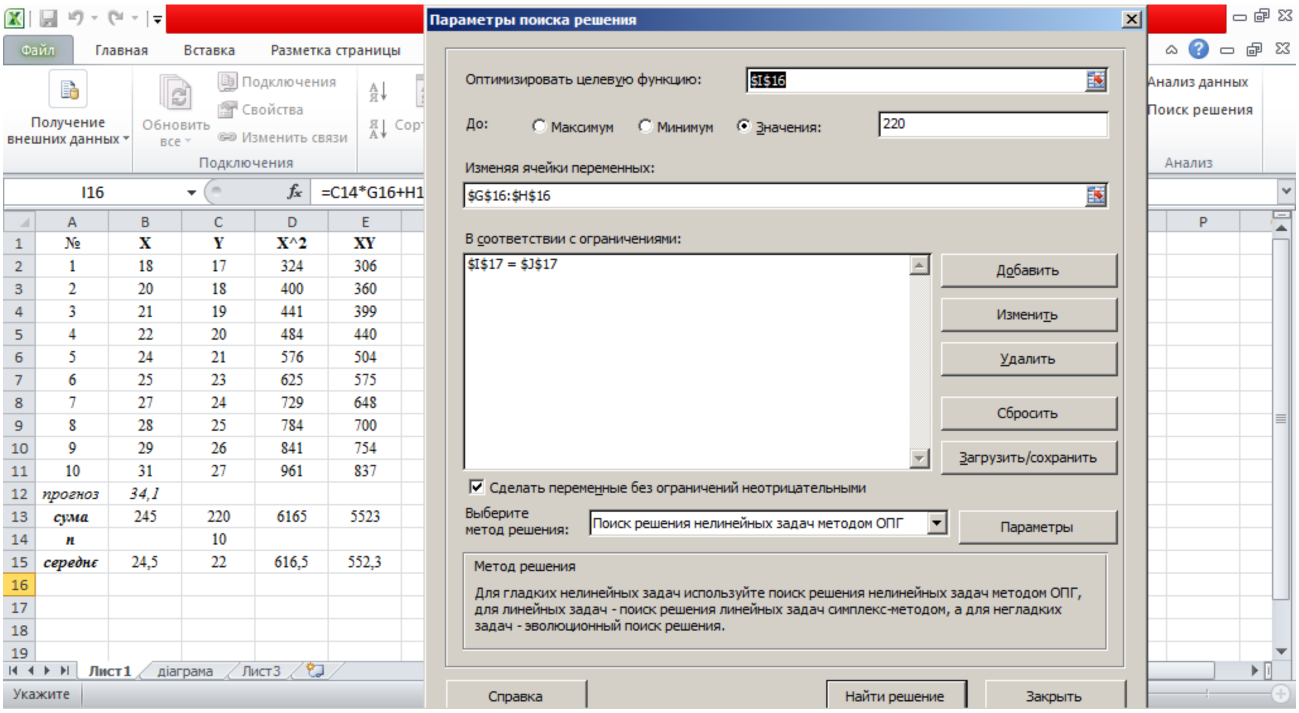

Рисунок 2.6 – Вікно оператора ***Поиск решения***

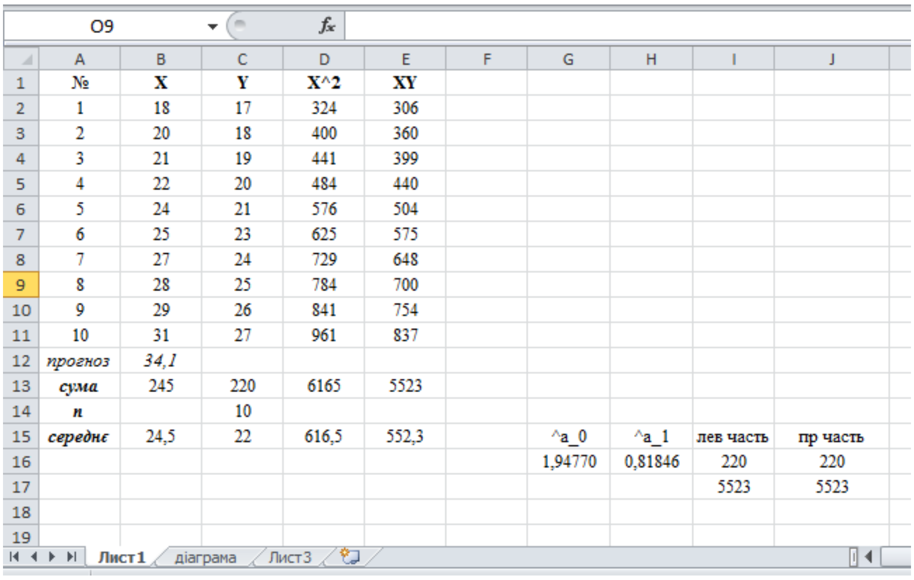

Рисунок 2.7 – Результат застосування оператора ***Поиск решения***

5. *Коефіцієнт детермінації:*

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}.$$

В табличному редакторі Microsoft Excel можна обчислити коефіцієнт детермінації за допомогою функції:

=КВПИРСОН(*Y*;*X*).

6. *Коефіцієнт еластичності:*

$$E_{\frac{Y}{X}} = \hat{a}_1 \cdot \frac{\bar{X}}{\bar{Y}}.$$

7. *Перевірка статистичної значущості коефіцієнта детермінації.*

Розглянемо нульову гіпотезу $H_0 : d = R^2 = 0$ проти альтернативної гіпотези $H_A : d = R^2 > 0$.

Порівняємо фактичне значення коефіцієнта детермінації $d = R^2$ із табличним для рівня значущості α та чисел ступенів свободи $k_1 = m - 1$ і $k_2 = n - m$: $d_{\text{табл}}(\alpha; k_1; k_2)$ (додаток Б). Якщо $d > d_{\text{табл}}$, то гіпотеза H_0 відхиляється і коефіцієнт детермінації є значущим.

8. *Перевірка статистичної значущості коефіцієнта кореляції.*

Для цього перевіримо нульову гіпотезу $H_0 : r_{xy} = 0$ проти альтернативної гіпотези $H_A : r_{xy} \neq 0$. Порівняємо отримане значення r_{xy} із табличним. Для n знаходимо $r_{xy\ \text{табл}}$ (додаток В). Якщо умова $r_{xy} > r_{xy\ \text{табл}}$ виконується, то гіпотеза H_0 відхиляється на користь альтернативної, що свідчить про значущість коефіцієнта кореляції.

9. *Перевірка значущості коефіцієнта кореляції за* t *- критерієм Стьюдента.*

Для цього перевіримо нульову гіпотезу $H_0 : r_{xy} = 0$ проти альтернативної гіпотези $H_A : r_{xy} \neq 0$.

Фактичне значення $t -$ критерію визначається за формулою:

$$t_{\text{факт}} = r_{xy} \sqrt{\frac{n-2}{1 - r_{xy}^2}}.$$

Для заданого рівня значущості α та числа ступенів свободи $k = k_2 = n - m$ визначимо табличне значення $t -$ статистики: $t_{\text{табл}} = t(\alpha; k)$ (додаток Д) або в табличному редакторі Microsoft Excel за допомогою функції =СТЬЮДЕНТ.ОБР.2Х(*alpha*; *k*).

Якщо $|t_{\text{факт}}| > t_{\text{табл}}$, то гіпотеза H_0 відхиляється на користь альтернативної та коефіцієнт кореляції вважається статично значущим.

10. *Визначення надійних інтервалів для коефіцієнта кореляції за* $t -$ *розподілом Стьюдента при рівні значущості* α.

Надійні інтервали коефіцієнта кореляції визначаються за формулою:

$$r_{xy} - \Delta r \le r \le r_{xy} + \Delta r,$$

$$r_{xy} - tS_r \le r \le r_{xy} + tS_r,$$

де $t = t_{\text{табл}} = \text{СТЬЮДЕНТ.ОБР.2Х}(\alpha; k)$, $S_r = \dfrac{1 - r_{xy}}{\sqrt{n}}$.

11. *Перевірка статистичної значущості оцінок параметрів моделі за* t *- критерієм Стьюдента.*

Розглянемо гіпотези:

$H_0 : \widehat{a}_j = 0 \; (j = 0, 1)$ – оцінка параметра незначуща;

$H_A : \widehat{a}_j \neq 0 \; (j = 0, 1)$ – оцінка параметра значуща.

Визначимо фактичні значення t –статистики:

$$t_0 = \frac{|\widehat{a}_0|}{S_{\widehat{a}_0}}, \qquad t_1 = \frac{|\widehat{a}_1|}{S_{\widehat{a}_1}},$$

де $S_{\widehat{a}_0} = \widehat{\sigma}_u \sqrt{\dfrac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \overline{x})^2}}$, $S_{\widehat{a}_1} = \widehat{\sigma}_u \dfrac{1}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}$, $\widehat{\sigma}_u^2 = \dfrac{1}{n-2}\sum_{i=1}^{n}(\widehat{y}_i - y_i)^2$.

Порівняємо знайдені значення з табличним значенням t– статистики:

$$t_{\text{табл}} = \text{СТЬЮДЕНТ.ОБР.2Х}(\alpha; k).$$

Якщо $t_0 > t_{\text{табл}}$ і $t_1 > t_{\text{табл}}$, то нульову гіпотезу для обох параметрів відхиляємо на користь альтернативної і параметри моделі статично значущі, тобто мають значний вплив на залежну змінну Y.

12. Порівняння стандартних помилок оцінок параметрів моделі зі значеннями цих оцінок:

$$\frac{S_{\widehat{a}_0}}{|\widehat{a}_0|} \cdot 100\%, \qquad \frac{S_{\widehat{a}_1}}{|\widehat{a}_1|} \cdot 100\%.$$

Якщо стандартна помилка оцінки параметра не перевищує 10%, то це свідчить про незміщеність даної оцінки параметра моделі. В противному випадку це означає, що даний параметр може мати зміщення (можу бути зумовлено невеликою сукупністю спостережень).

13. Визначення інтервалів надійності для оцінок параметрів моделі за t–розподілом при рівні значущості α:

– для параметра $\widehat{a}_0$: $\widehat{a}_0 - tS_{\widehat{a}_0} \leq \widehat{a}_0 \leq \widehat{a}_0 + tS_{\widehat{a}_0}$;

– для параметра $\widehat{a}_1$: $\widehat{a}_1 - tS_{\widehat{a}_1} \leq \widehat{a}_1 \leq \widehat{a}_1 + tS_{\widehat{a}_1}$,

де $t = t_{\text{табл}} = \text{СТЬЮДЕНТ.ОБР.2Х}(\alpha; k)$.

14. F -тест Фішера для перевірки моделі на адекватність.

Розглянемо гіпотезу $H_0 : R^2 = 0$ проти альтернативної $H_A : R^2 > 0$. Це рівнозначне тому, що перевіряється значущість водночас усіх параметрів моделі (гіпотеза $H_0 : \widehat{a}_0 = \widehat{a}_1 = 0$ проти альтернативної $H_A : \widehat{a}_0 \neq 0, \widehat{a}_1 \neq 0$).

Фактичне значення F–критерію:

$$F_{\text{факт}} = \frac{\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2}{\frac{1}{n-2}\sum_{i=1}^{n}(\widehat{y}_i - y_i)^2}$$

або табличному редакторі Microsoft Excel:

$$F_{\text{факт}} = \frac{\text{КВАДРОТКЛ}(\widehat{Y})}{\text{КВАДРОТКЛ}(Y)}.$$

Табличне значення для заданого рівня значущості α (рівня помилки) або $p = 1 - \alpha$ (довірчої ймовірності) та числа ступенів свободи $k_1 = m - 1$ і $k_2 = n - 2$ можна взяти зі статистичної таблиці (додаток Е) або, використавши функцію редактора Excel:

$$F_{\text{табл}} = \text{F.ОБР.ПХ}\left(\alpha;\, k_1;\, k_2\right).$$

Якщо $F_{\text{факт}} > F_{\text{табл}}$, то нульову гіпотезу відхиляємо і з заданою ймовірністю p економетричну модель вважаємо адекватною фактичним даним, тобто гіпотеза про значущість зв'язку між незалежною та залежною змінними підтверджується.

15. Визначення надійних інтервалів базисних значень.

Використовуючи раніше знайдені значення $\widehat{\sigma}_u$ і $t = t_{\text{табл}}$, визначаються надійні інтервали базисних значень $\widehat{Y}$ за формулою:

$$\widehat{y}_i \pm \Delta\widehat{y}_i,$$

де $\widehat{y}_i$ – теоретичні значення, обчислені за допомогою побудованої економетричної моделі;

$$\Delta\widehat{y}_i = t\widehat{\sigma}_{\widehat{y}_i} = t\widehat{\sigma}_u \sqrt{\frac{1}{n} + \frac{\left(x_i - \overline{x}\right)^2}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}}.$$

Виконавши відповідні розрахунки, визначимо надійні інтервали базисних значень за формулами:

$$\widehat{y}_{i,\,\min} = \widehat{y}_i - \Delta\widehat{y}_i,$$

$$\widehat{y}_{i,\,\max} = \widehat{y}_i + \Delta\widehat{y}_i.$$

16. Визначення точкового та інтервального прогнозів.

Для визначення точкової оцінки прогнозу підставимо задане прогнозне значення $X_{\text{прогн}}$ у рівняння регресії $\hat{Y} = \hat{a}_0 + \hat{a}_1 X$. Значення $\widehat{Y}_{\text{прогн}}$ можна інтерпретувати як точкову оцінку прогнозного значення математичного сподівання та індивідуального значення $\widehat{Y}_i$, коли відомо $X_{\text{прогн}}$.

Перевірку знайденого прогнозного значення можна здійснити за допомогою відповідної статистичної функції:

$$Y_{\text{прогн}}\text{=ПРЕДСКАЗ(}\, X_{\text{прогн}};\, Y;\, X\text{)}.$$

Для визначення прогнозного інтервалу математичного сподівання $M\left(\widehat{Y}_{\text{прогн}}\right)$ обчислимо:

– стандартну помилку прогнозного значення або стандартну похибку прогнозу математичного сподівання $M\left(\widehat{Y}_{\text{прогн}}\right)$, враховуючи, що стандартна помилка залишків $\widehat{\sigma}_u$: $\widehat{\sigma}_{\text{прогн}} = \widehat{\sigma}_u \sqrt{\dfrac{1}{n} + \dfrac{\left(x_{\text{прогн}} - \overline{x}\right)^2}{\sum\limits_{i=1}^{n}\left(x_i - \overline{x}\right)^2}}$;

– прогнозний інтервал математичного сподівання $M\left(\widehat{Y}_{\text{прогн}}\right)$ при $t = t_{\text{табл}} = \text{СТЬЮДЕНТ.ОБР.2Х}\left(\alpha;\, k\right)$:

$$\widehat{Y}_{\text{прогн}} - t\widehat{\sigma}_{\text{прогн}} \le M\left(\widehat{Y}_{\text{прогн}}\right) \le \widehat{Y}_{\text{прогн}} + t\widehat{\sigma}_{\text{прогн}}.$$

Для визначення прогнозного інтервалу індивідуального значення $\widehat{Y}_{\text{прогн}}$ обчислимо:

– дисперсію прогнозу індивідуального значення $\widehat{Y}_{\text{прогн}}$:

$$\widehat{\sigma}^2_{\text{прогн}(i)} = \widehat{\sigma}^2_u + \widehat{\sigma}^2_{\text{прогн}};$$

– стандартну похибку прогнозу індивідуального значення $\widehat{Y}_{\text{прогн}}$:

$$\sigma_{\text{прогн}(i)} = \widehat{\sigma}^2_{\text{прогн}(i)}.$$

Тоді прогнозний інтервал індивідуального значення $\widehat{Y}_{\text{прогн}}$ при $t = t_{\text{табл}} = \text{СТЬЮДЕНТ.ОБР.2Х}\left(\alpha;\, k\right)$ матиме вигляд:

$$\widehat{Y}_{\text{прогн}} - t\widehat{\sigma}_{\text{прогн}(i)} \le \widehat{Y}_{\text{прогн}} \le \widehat{Y}_{\text{прогн}} + t\widehat{\sigma}_{\text{прогн}(i)}.$$

17. Перевірка точності економетричної моделі за допомогою середньої відносної похибки апроксимації.

Визначимо абсолютну середню відносну похибку апроксимації $\overline{\varepsilon}$:

$$\overline{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y}_i}{y_i}\right| \cdot 100\%,$$

і порівняємо її зі значеннями таблиці та зробимо висновок (додаток Г).

При порівнянні різних моделей, побудованих на одній статистичній базі, перевагу надають моделям з меншою похибкою апроксимації $\overline{\varepsilon}$.

18. Економетричну модель можна побудувати *за допомогою функції* табличного редактора Microsoft Excel *ЛИНЕЙН*.

Виділивши область розміром 5 рядків і 2 стовпці ($m = 2$ – кількість параметрів моделі) і застосувавши функцію ЛИНЕЙН(Y;X;1;1), заповнимо таблицю 2.3.

Таблиця 2.3 – Застосування функції ЛИНЕЙН

$\widehat{a}_1$	$\widehat{a}_0$
$S_{\widehat{a}_1}$	$S_{\widehat{a}_0}$
R^2	σ_u

F	$k = n - m$
$SSR = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2$	$SSE = \sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2$

Перший рядок результатів розрахунку містить оцінки параметрів моделі: $\hat{a}_1$; $\hat{a}_0$.

Другий рядок містить стандартні похибки оцінок параметрів моделі: $S_{\hat{a}_1}$; $S_{\hat{a}_0}$.

Третій рядок містить коефіцієнт детермінації та стандартну помилку залишків: R^2; σ_u.

Четвертий рядок містить F–критерій та число ступенів свободи: F; $k = n - m$.

П’ятий рядок таблиці результатів містить суми квадратів: $SSR = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2$; $SSE = \sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2$.