Міністерство освіти і науки України
ДЕРЖАВНИЙ БІОТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ
Факультет економічних відносин та фінансів
Кафедра економіки та бізнесу

О.Д. Тімченко

ЕКОНОМЕТРИКА

Методичні вказівки та завдання
до виконання самостійної та індивідуальної роботи
для здобувачів першого (бакалаврського) рівня
вищої освіти спеціальності 051 Економіка
292 Міжнародні економічні відносини
071 Облік і оподаткування
072 Фінанси, банківська справа, страхування та фондовий ринок

Харків

2024

Міністерство освіти і науки України

ДЕРЖАВНИЙ БІОТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ

Факультет економічних відносин та фінансів

Кафедра економіки та бізнесу

О.Д. Тімченко

ЕКОНОМЕТРИКА

Методичні вказівки та завдання
до виконання самостійної та індивідуальної роботи
для здобувачів першого (бакалаврського) рівня
вищої освіти спеціальності 051 Економіка
292 Міжнародні економічні відносини
071 Облік і оподаткування
072 Фінанси, банківська справа, страхування та фондовий ринок

Затверджено рішенням науково –
методичної комісії факультету
економічних відносин та фінансів
Протокол № 1 від 25.09.2024 р.

Харків
2024

УДК 330.43(072)

Схвалено
на засіданні кафедри економіки та бізнесу,
протокол № 2 від 02.09.2024 р.

Рецензенти:
Т.О. Staверська, кандидат економічних наук, доцент, завідувач кафедри фінансів, банківської справи та страхування Державного біотехнологічного університету
Н.Б. Кащена, доктор економічних наук, професор, завідувач кафедри обліку, аудиту та оподаткування Державного біотехнологічного університету

Економетрика: методичні вказівки та завдання до виконання самостійної та індивідуальної роботи для здобувачів першого (бакалаврського) рівня вищої освіти спеціальності 051 Економіка, 292 Міжнародні економічні відносини, 071 Облік і оподаткування, 072 Фінанси, банківська справа, страхування та фондовий ринок / укладач: О.Д. Тімченко; ДБТУ.- Харків:, 2024. - 95 с.

Виконання індивідуальної роботи здобувачами є складовою частиною навчального процесу, активною формою самостійної роботи здобувачів.

Мета індивідуальної роботи полягає в закріпленні та поглибленні теоретичних знань, набутих здобувачем у процесі вивчення дисципліни, виробленні уміння самостійно працювати з навчальною, спеціальною літературою і статистичними матеріалами, робити узагальнення та висновки. Запропоновані методичні вказівки є доповненням до конспекту лекцій. Рекомендації охоплюють основні теми дисципліни.

УДК 330.43(072)

Відповідальний за випуск: О.Д. Тімченко, доц. кафедри економіки та бізнесу

ЗМІСТ

ПЕРЕДМОВА

Кінцевою метою вивчення дисципліни «Економетрика» є формування у майбутніх спеціалістів теоретичних знань та практичних навиків зі застосування статистичних методів та узагальнення емпіричних залежностей економічних змінних, а також побудови надійних прогнозів в економіці підприємства, фінансах, менеджменті з метою обґрунтування рішень, що приймаються.

Економетрика поєднує сукупність методів та моделей, що дозволяють на основі економічної теорії, економічної статистики та математико-статистичного інструментарію дослідити кількісні вирази якісних залежностей. Під час вивчення дисципліни «Економетрика» здобувач повинен володіти основами математичної статистики, матричної алгебри в обсязі курсу вищої математики для економічних спеціальностей.

Індивідуальні завдання є однією з форм організації самостійної роботи навчального процесу у вищих закладах освіти, яка передбачає створення умов для якнайповнішої реалізації творчих можливостей здобувачів і має на меті поглиблення, узагальнення та закріплення знань, які здобувачі одержують в процесі навчання, а також застосування цих знань на практиці.

Виконання індивідуальної роботи здобувачами є складовою частиною навчального процесу, активною формою самостійної роботи здобувачів.

Мета індивідуальної роботи полягає в закріпленні та поглибленні теоретичних знань, набутих здобувачем у процесі вивчення курсу, виробленні уміння самостійно працювати з навчальною, спеціальною літературою і статистичними матеріалами, робити узагальнення та висновки.

Запропоновані методичні вказівки є доповненням до конспекту лекцій. Рекомендації охоплюють основні теми дисципліни. Головна увага приділяється побудові економетричних моделей на основі просторових даних. Всі розділи даних вказівок мають ідентичну структуру:

- короткі основні теоретичні положення, які включають основні поняття, визначення, формули;
- питання для самоконтролю;
- методичні вказівки по виконанню індивідуального завдання з рішенням і висновками;
- індивідуальні завдання по реалізації типового завдання за допомогою пакетів прикладних програм Excel.

ПРОСТА ЛІНІЙНА РЕГРЕСІЯ ТА КОРЕЛЯЦІЯ В ЕКОНОМЕТРИЧНИХ ДОСЛІДЖЕННЯХ

Основні теоретичні положення

Специфікація моделі

Однією з головних перешкод застосування системного підходу у всіх сферах економічного аналізу є проблема невизначеності. Одним з можливих підходів до рішення цієї проблеми є регресійний аналіз.

Регресією називають стохастичну залежність однієї випадкової величини від іншої (або декількох інших) випадкових величин.

Стохастична залежність виражається за допомогою функції, яка називається ***функцією регресії***. У залежності від кількості факторів, які входять у рівняння регресії, прийнято розрізняти просту (парну) і множинну регресії.

Проста регресія являє собою регресію між двома змінними - y та x, тобто модель виду:

$$y = \hat{f}(x), \tag{1.1}$$

де y – залежна змінна (результативна ознака);

x – незалежна, або пояснююча, змінна (ознака-фактор).

Множинна регресія відповідно являє собою регресію результативної ознаки з двома і більшою кількістю факторів, тобто модель виду:

$$y = \hat{f}(x_1, x_2, ..., x_k). \tag{1.2}$$

Будь-яке економетричне дослідження починається зі ***специфікації моделі***, тобто з формулювання виду моделі, виходячи з відповідної теорії зв’язку між змінними. Іншими словами, дослідження починається з теорії, яка встановлює зв'язок між явищами.

Насамперед із кола факторів, що впливають на результативну ознаку, необхідно виділити найбільш вагомі. Парна регресія достатня, якщо мається домінуючий фактор, який і використовується в якості пояснюючої змінної.

Практично в кожному окремому випадку величина y складається з двох доданків:

$$y = \hat{y}_x + e, \tag{1.3}$$

де y – фактичне значення результативної ознаки;

$\hat{y}_x$ – теоретичне значення результативної ознаки, знайдене виходячи з відповідної математичної функції зв'язку *y* и *x*, тобто з рівняння регресії;

e – випадкова величина, що характеризує відхилення реального значення результативної ознаки від теоретичного, знайденого з рівняння регресії.

Випадкова величина *e* включає вплив не врахованих у моделі факторів, випадкових помилок і особливостей виміру. Її присутність у моделі покликана трьома джерелами: специфікацією моделі, вибірковим характером вихідних даних, особливостями виміру змінних.

Від правильно обраної специфікації моделі залежить величина випадкових помилок: вони тим менше, якщо теоретичні значення результативної ознаки $\hat{y}_x$ у більшій мері підходять до фактичних даних *y*.

У парній регресії вибір виду математичної функції $\hat{y}_x = f(x)$ може бути здійснений трьома методами:

- графічним;
- аналітичним, тобто виходячи з теорії досліджуваного взаємозв'язку;
- експериментальним.

При вивченні залежності між двома ознаками ***графічний метод*** підбору виду рівняння регресії досить наочний, який заснований на побудові поля кореляції.

Значний інтерес представляє ***аналітичний метод*** вибору типу рівняння регресії. Він заснований на вивченні матеріальної природи зв'язку досліджуваних ознак.

При обробці інформації на комп'ютері вибір виду рівняння регресії здійснюється ***експериментальним методом***, тобто шляхом порівняння величини залишкової дисперсії $D_{зал}$, розрахованої при різних моделях.

Якщо рівняння регресії проходить через усі крапки кореляційного поля, що можливо тільки при функціональному зв'язку, коли всі крапки лежать на лінії регресії $\hat{y}_x = f(x)$, то фактичні значення результативної ознаки збігаються з теоретичними $y = \hat{y}_x$, тобто вони цілком обумовлені впливом фактора *x*. У цьому випадку залишкова дисперсія $D_{зал} = 0$. У практичних дослідженнях, як правило, має місце деяке розсіювання крапок щодо лінії регресії. Воно зумовлено впливом інших факторів, які не враховуються в рівнянні регресії. Іншими словами, мають місце відхилення фактичних даних від теоретичних $(y - \hat{y}_x)$. Величина цих відхилень і лежить в основі розрахунку залишкової дисперсії:

$$D_{зал} = \frac{1}{n}\sum(y - \hat{y}_x)^2. \qquad (1.4)$$

Чим менше величина залишкової дисперсії, тим у меншій мірі спостерігається вплив інших факторів, що не враховуються в рівнянні регресії, тим краще рівняння регресії підходить до вихідних даних. При обробці статистичних даних на комп'ютері перебираються різні математичні функції в автоматичному режимі і з них обирається та, для якої залишкова дисперсія є найменшою.

Якщо залишкова дисперсія виявляється приблизно однаковою для декількох функцій, то на практиці перевага віддається більш простим видам функцій, тому що вони в більшому ступені піддаються інтерпретації і вимагають меншого обсягу спостережень.

Загальне поняття про вибіркову лінійну регресію. Оцінювання параметрів лінійної регресії

Лінійна регресія знаходить широке застосування в економетрії, тому що має чітку економічну інтерпретацію параметрів. Лінійна регресія зводиться до побудови рівняння виду

$$\hat{y}_x = a + bx \quad \textit{або} \quad y = a + bx + e. \tag{1.5}$$

Рівняння виду $\hat{y}_x = a + bx$ дозволяє за заданими значеннями фактора x отримати теоретичні значення результативної ознаки, підставляючи в рівняння фактичні значення фактора x.

Побудова лінійної регресії зводиться до оцінки її параметрів - a і b. Класичний підхід до оцінювання параметрів лінійної регресії заснований на ***методі найменших квадратів (МНК)***.

МНК дозволяє одержати такі оцінки параметрів a і b, при яких сума квадратів відхилень фактичних значень результативної ознаки y від розрахункових (теоретичних) $\hat{y}_x$ мінімальна:

$$\sum_i (y_i - \hat{y}_{x_i})^2 \to min. \tag{1.6}$$

Вирішуючи систему нормальних рівнянь, знайдемо оцінки параметрів a і b. Можна скористатися наступними готовими формулами:

$$b = \frac{\mathrm{cov}(x, y)}{\sigma_x^2}, \tag{1.7}$$

де *cov (x, y)* – коваріація ознак x та y;

σ_x – дисперсія ознаки x.

Через те, що $\mathrm{cov} = \overline{xy} - \bar{y} \cdot \bar{x}, \sigma_x^2 = \overline{x^2} - \bar{x}^2$, одержимо наступну формулу розрахунку оцінки параметра b:

$$b=\frac{\frac{1}{n}\sum_{i=1}^{n}(y-\bar{y})\cdot(x-\bar{x})}{\frac{1}{n}\sum_{i=1}^{n}(x-\bar{x})^2}\ \text{або}\ b=\frac{n\sum_{i=1}^{n}x\times y-\sum_{i=1}^{n}x\sum_{i=1}^{n}y}{n\sum_{i=1}^{n}x^2-\left(\sum_{i=1}^{n}x\right)^2};$$

$$a=\bar{y}-b\cdot\bar{x}\ . \tag{1.8}$$

Параметр b називається ***коефіцієнтом регресії***. Його величина показує середню зміну результату зі зміною фактора на одну одиницю. Так, якщо у функції витрат $\hat{y}_x = 725 + 1.5x$ (y - витрати (тис. грн.), x - кількість одиниць продукції), то зі збільшенням обсягу продукції (x) на 1 одиницю витрати виробництва зростають у середньому на 1,5 тис. грн., тобто додатковий приріст продукції на 1 од. викликає збільшення витрат у середньому на 1,5 тис. грн.

Можливість чіткої економічної інтерпретації коефіцієнта регресії зробила лінійне рівняння регресії досить розповсюдженим в економетричних дослідженнях.

Параметр a не має чіткого економічного змісту на відмінність від параметра b. Інтерпретувати можна лише знак при параметрі a. Якщо $a > 0$, то відносна зміна результату відбувається повільніше, ніж зміна фактора. Іншими словами, варіація результату менше варіації фактора - коефіцієнт варіації за фактором x вище коефіцієнта варіації для результату y: $Vx > Vy$.

Поняття тісноти зв'язку, оцінка коефіцієнтів кореляції та детермінації

Рівняння регресії завжди доповнюється показником тісноти зв'язку. При використанні лінійної регресії таким показником виступає ***лінійний коефіцієнт кореляції*** r_{xy}. Існують різні модифікації формули лінійного коефіцієнта кореляції. Деякі з них наведені нижче:

$$r_{xy}=b\frac{\sigma_x}{\sigma_y}=\frac{\operatorname{cov}(x,y)}{\sigma_x\sigma_y}=\frac{\overline{yx}-\bar{y}\cdot\bar{x}}{\sigma_x\sigma_y}\ . \tag{1.9}$$

Лінійний коефіцієнт кореляції знаходиться в інтервалі

$$-1 < r_{xy} < 1.$$

Якщо коефіцієнт регресії $b > 0$, то $0 < r_{xy} < 1$, і, навпаки, при $b < 0$, $-1 < r_{xy} < 0$.

Варто мати на увазі, що величина лінійного коефіцієнта кореляції оцінює тісноту зв'язку розглянутих ознак тільки в її лінійній формі. Тому близькість абсолютної величини лінійного коефіцієнта кореляції до нуля ще не означає відсутність зв'язку між ознаками. При іншій специфікації моделі зв'язок між ознаками може виявитися досить тісним.

Поряд з коефіцієнтом кореляції використовується ще один критерій, за допомогою якого також вимірюється щільність зв'язку між двома або більше показниками та перевіряється адекватність (відповідність) побудованої регресійної моделі реальній дійсності. Тобто дається відповідь на запитання, чи справді зміна значення *y* лінійно залежить саме від зміни значення *x*, а не відбувається під впливом різних випадкових факторів. Таким критерієм є ***коефіцієнт детермінації***.

Загальна сума квадратів відхилень розкладається на дві складові:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2, \qquad (1.10)$$

де $\sum_{i=1}^{n}(y_i-\bar{y}_i)^2$ – загальна сума квадратів, яка позначається через SST;

$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ – сума квадратів помилок, яка позначається через SSE;

$\sum_{i=1}^{n}(\hat{y}_i-\bar{y}_i)^2$ – сума квадратів, що пояснює регресію,тобто SSR.

Коефіцієнт детермінації можна записати у вигляді:

$$R^2=\frac{SSR}{SST}. \qquad (1.11)$$

Коефіцієнт детермінації завжди додатний і знаходиться в межах від нуля до одиниці ($0 \le R^2 \le 1$).

Розглянемо зв'язок між коефіцієнтом кореляції та нахилом регресійної лінії, тобто параметром b:

$$r_{xy}=b\frac{\sigma_x}{\sigma_y} \quad \text{або} \quad R^2=(r_{xy})^2. \qquad (1.12)$$

Наприклад, якщо величина R^2 = 0,96, то це означає, що рівнянням регресії пояснюється 96% дисперсії результативної ознаки, а на долю інших факторів приходиться лише 4 % її дисперсії (тобто залишкова дисперсія). Величина коефіцієнта детермінації є одним із критеріїв оцінки якості лінійної моделі. Чим більше частка поясненої варіації, тим відповідно менше роль інших факторів, і, отже, лінійна модель добре апроксимує вихідні дані і нею можна скористатися для прогнозу значень результативної ознаки.

Оцінка якості лінійного рівняння регресії

Оцінка якості лінійного рівняння регресії (адекватність, значимість) простої лінійної регресійної моделі можна перевірити за допомогою коефіцієнта детермінації. Якщо його значення близьке до одиниці, то можна вважати, що модель адекватна. Якщо його значення близьке до нуля, то модель неадекватна, тобто немає лінійного зв'язку між залежною та незалежною змінними. Але який висновок можна зробити, якщо значення коефіцієнта детермінації має не явно виражене граничне значення, тобто знаходиться в середині інтервалу від 0 до 1?

Зрозуміло, що в таких випадках важко зробити однозначний висновок про наявність зв'язку, тобто про адекватність моделі. Потрібен інший критерій, який би однозначно давав відповідь на запитання про адекватність побудованої моделі. Найбільш поширеним із таких критеріїв є ***критерій Фішера***. При цьому висувається нульова гіпотеза, за якою коефіцієнт регресії дорівнює нулю, тобто $b=0$, та фактор x не впливає на результат y.

Безпосередньому розрахунку F-критерію Фішера відбувається аналіз дисперсії, де загальна сума квадратів відхилень розкладується на факторну та залишкову:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2. \tag{1.13}$$

Кожна сума квадратів пов'язана з числом, яке називають її ***«ступенем вільності»***. Ступені вільності прийнято позначати через *DF* або *df*.

Це число показує, скільки незалежних елементів інформації, що утворились з елементів $y_1, y_2, \ldots, y_n$, потрібно для розрахунку даної суми квадратів.

Для утворення SST потрібно $(n - 1)$ незалежних чисел. Суму квадратів, що пояснює регресію (SSR), отримують, використовуючи тільки одну незалежну одиницю інформації. Сума квадратів помилок (SSE) має $(n - 2)$ ступенів вільності.

Поділивши кожну суму квадратів на відповідне їй число ступенів вільності, отримаємо середній квадрат відхилень, або дисперсію на одну ступень вільності:

$$D_{заг}=\frac{\sum(y-\bar{y})^2}{n-1};D_{регр}=\frac{\sum(\hat{y}-\bar{y})^2}{1};D_{пом}=\frac{\sum(y-\hat{y})^2}{n-2}. \tag{1.14}$$

Відношення факторної дисперсії до залишкової на одну ступень вільності дає розрахунок ***F-критерію Фішера***:

$$F=\frac{D_{регр}}{D_{пом}} \tag{1.15}$$

Якщо нульова гіпотеза справедлива, то факторна та залишкова дисперсії не відрізняються одна від одної. Визначене значення F-критерію буде достовірним, якщо воно більш табличного. У цьому випадку нульова гіпотеза про відсутність зв'язку між показниками відхиляється та робиться висновок щодо значимості цього зв'язку: $\boldsymbol{F_{факт}} > \boldsymbol{F_{табл}}$.

Якщо ж величина виявиться менш табличної $\boldsymbol{F_{факт}} < \boldsymbol{F_{табл}}$, то імовірність нульової гіпотези вище заданого рівня (наприклад, 0,05) і вона не може бути відхилена без ризику зробити неправильний висновок щодо наявності зв'язку. У цьому випадку рівняння регресії вважається статистично не значимим, нульова гіпотеза не відхиляється.

Оцінка значимості рівняння регресії зазвичай виконується у вигляді таблиці дисперсійного аналізу.

Таблиця 1.1 – **ANOVA - таблиця**

Джерело варіації	Кількість ступенів вільності	Сума квадратів відхилень	Дисперсія на одну ступень вільності	F- значення	
				фактичне	табличне
Загальна, **SST**	*n-1*	$\sum_{i=1}^{n}(y_i-\bar{y})^2$	-	-	-
Регресійна (факторна), **SSR**	*1*	$\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$	$\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$	$\dfrac{\sum(\hat{y}_i-\hat{y}_i)^2}{\frac{\sum(y_i-\hat{y}_i)^2}{n-2}}$	зі стат. таблиці
Помилкова (залишкова), **SSE**	*n-2*	$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$	$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2/n-2$		

Величина *F*- критерію пов'язана з коефіцієнтом детермінації R^2:

$$F=\frac{R^2}{1-R^2}\cdot(n-2)\,. \tag{1.16}$$

Оцінка значущості параметрів лінійної регресії та кореляції. Побудова інтервалів довіри

У випадку простої лінійної регресії оцінюється не тільки значимість рівняння регресії в цілому, а і окремих його параметрів. З цією метою по кожному з параметрів визначається його ***стандартна помилка: m_b та m_a.***

Стандартна помилка коефіцієнта регресії визначається за формулою:

$$m_b = \sqrt{\frac{\sum(y-\hat{y}_x)^2/(n-2)}{\sum(x-\bar{x})^2}} = \sqrt{\frac{S^2}{\sum(x-\bar{x})^2}}\,, \tag{1.17}$$

де S^2 – помилкова (залишкова) дисперсія на одну ступень вільності.

Величина стандартної помилки разом з ***t-розподілом Ст'юдента*** при $df = (n-2)$ застосовується для перевірки значимості коефіцієнта регресії та для побудови його інтервалів довіри.

Для оцінки значимості коефіцієнта регресії його величина порівнюється з його стандартною помилкою, тобто визначається фактичне значення t-критерію Ст'юдента:

$$t_b = \frac{b}{m_b}, \tag{1.18}$$

яке потім порівнюється з табличним значенням при певному рівні значимості α та кількості ступенів вільності $df = (n-2)$.

Між t-критерієм Ст'юдента та F-критерієм Фішера існує певна залежність, яка виражається формулою:

$$t_b = \sqrt{F}\,. \tag{1.19}$$

Інтервал довіри для коефіцієнта регресії визначається як

$$b \pm t_{табл} \cdot m_b \;\; або \quad b - t_{табл} \cdot m_b \langle\, b \,\langle\, b + t_{табл} \cdot m_b\,. \tag{1.20}$$

Оскільки коефіцієнт регресії в економетричних дослідженнях має чітку економічну інтерпретацію, то межі інтервалу довіри для коефіцієнта регресії не повинні мати суперечливих результатів, наприклад, нижня межа інтервалу має від'ємне значення, а верхня – додатне. Такий інтервал довіри містить в собі значення нуля, а це означає, що коефіцієнт регресії $b=0$, фактор x не впливає на результат y, коефіцієнт регресії є статистично не значимий.

Стандартна помилка для параметра a визначається за формулою:

$$m_a = \sqrt{\frac{\sum(y-\hat{y}_x)^2 \cdot \sum x^2}{(n-2)\cdot n\sum(x-\bar{x})^2}} = \sqrt{S^2 \cdot \frac{\sum x^2}{n \cdot \sum(x-\bar{x})^2}}\,. \tag{1.21}$$

Процедура оцінювання значимості даного параметра не відрізняється від оцінки значимості коефіцієнта регресії; розраховується *t*-критерій Ст'юдента для параметра *a*:

$$t_a = \frac{a}{m_a}, \tag{1.21}$$

його величина порівнюється з табличним значенням при *df = (n-2)* ступенів вільності.

Значимість лінійного коефіцієнта кореляції перевіряється на основі величині помилки коефіцієнта кореляції m_r:

$$m_r = \sqrt{\frac{1-r^2}{n-2}}. \tag{1.22}$$

Фактичне значення *t*-критерію Ст'юдента визначається як

$$t_r = \frac{r}{\sqrt{1-r^2}} \cdot \sqrt{n-2}. \tag{1.23}$$

Дана формула свідчить про те, що у випадку простої лінійної регресії $t_r^{\ 2} = F$. Таким чином,

$$t_r^{\ 2} = t_b^2 \tag{1.24}$$

Перевірка гіпотез про значимість коефіцієнта регресії, параметра *a* та коефіцієнта кореляції аналогічна перевірці гіпотези про значимість та адекватність лінійного рівняння регресії.

Для оцінки якості рівняння регресії застосовується також показник, який характеризує відхилення фактичних даних від теоретичних, які отримують з рівняння моделі. Фактичні значення результативної ознаки відрізняються від теоретичних, розрахованих по рівнянню регресії, тобто y і $\hat{y}_x$. Чим менше ця відмінність, тим ближче теоретичні значення підходять до емпіричних даних, тим краще якість моделі. Величина відхилень фактичних і розрахункових значень результативної ознаки (y- $\hat{y}_x$) за кожним спостереженням являє собою ***помилку апроксимації***. Їхнє число відповідає обсягу сукупності. Оскільки (y- $\hat{y}_x$) може бути як величиною позитивною, так і негативною, то помилки апроксимації для кожного спостереження прийнято визначати у відсотках по модулю.

Відхилення (y- $\hat{y}_x$) можна розглядати як абсолютну помилку апроксимації, а

$$\left| \frac{y - \hat{y}_x}{y} \right| \cdot 100$$

- як відносну помилку апроксимації. Щоб мати загальне поняття про якість моделі з відносних відхилень за кожним спостереженням, визначають ***середню помилку апроксимації*** як середню арифметичну просту:

$$\overline{A} = \frac{1}{n} \cdot \sum \left| \frac{y - \hat{y}_x}{y} \right| \cdot 100 \,. \tag{1.25}$$

Якщо значення середньої помилки апроксимації знаходиться в інтервалі від 5 % до 8 %, то це свідчить про високу якість побудованого рівняння моделі.

Побудова інтервалів прогнозу за лінійним рівнянням регресії

У прогнозних розрахунках за рівнянням регресії визначається прогнозне значення (y_{np}) шляхом підстановки в рівняння регресії $\hat{y}_x = a + bx$ відповідного значення x_{np}. Однак крапковий прогноз явно не реальний. Тому він доповнюється розрахунком стандартної помилки $\hat{y}_x$, тобто $m_{\hat{y}_X}$, і відповідно інтервальною оцінкою прогнозного значення (y^*)

$$\hat{y}_x - m_{\hat{y}_x} \le y^* \le \hat{y}_x + m_{\hat{y}_x} \,. \tag{1.26}$$

Помилка коефіцієнта регресії визначається за формулою:

$$m_b^{\ 2} = \frac{S^2}{\sum (x - \bar{x})^2} \,. \tag{1.27}$$

Вважаючи, що прогнозне значення фактора $x_{np} = x_к$, одержимо наступну формулу розрахунку стандартної помилки прогнозного значення по лінії регресії:

$$m^2_{\hat{y}_x} = \frac{S^2}{n} \cdot \frac{S^2}{\sum (x - \bar{x})^2} \cdot (x_k - \bar{x})^2 = S^2 \cdot \left(1 + \frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum (x - \bar{x})^2} \right) \tag{1.28}$$

Відповідно $m_{\hat{y}X}$ має вираз:

$$m_{\hat{y}_x} = S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum (x - \bar{x})^2}} \tag{1.29}$$

Розглянута формула стандартної помилки прогнозного значення y при заданому значенні $x_к$ характеризує помилку положення лінії регресії.

Для прогнозованого значення $\hat{y}_x$ інтервали довіри при заданому $x_к$ визначаються як

$$\hat{y}_{x_k} \pm t_\alpha \cdot m_{\hat{y}_x}. \tag{1.30}$$

Запитання для самоперевірки знань

1. Що є функцією регресії?
2. Чим регресійна модель відрізняється від функції регресії?
3. Назвіть причини наявності в регресійної моделі випадкового відхилення?
4. Назвіть основні етапи регресійного аналізу.
5. В чому полягають помилки специфікації моделі?
6. У чому різниця між теоретичним та емпіричним рівнянням регресії?
7. У чому суть методу найменших квадратів?
8. Які основні передумови МНК?
9. Поясните зміст коефіцієнта регресії, назвіть способи його оцінювання.
10. У чому сутність статистичної значущості коефіцієнтів регресії?
11. Які висновки можна зробити об оцінках коефіцієнтів регресії та випадкового відхилення, що знайдені по МНК?
12. Проінтерпретуйте коефіцієнти емпіричного парного лінійного рівняння регресії.
13. Поясніть суть коефіцієнту кореляції.
14. У яких межах змінюється коефіцієнт кореляції?
15. Поясніть суть коефіцієнту детермінації?
16. У яких межах змінюється коефіцієнт детермінації?
17. Як визначається дисперсія залишків, загальна дисперсія і дисперсія регресії? Який між ними зв’язок?
18. Що таке число ступенів волі і як воно визначається для факторної і залишкової сум квадратів?
19. Як визначається F-критерій? Для чого він застосовується?
20. Як оцінити достовірність коефіцієнта кореляції?
21. Як визначити довірчі інтервали для параметрів моделі?
22. У чому відмінність стандартної помилки положення лінії регресії від середньої помилки прогнозованого індивідуального значення результативної ознаки при заданому значенні фактора?
23. У чому зміст середньої помилки апроксимації і як вона визначається?

Методичні вказівки до виконання індивідуальної роботи

За даними таблиці 1.2 побудувати лінійне рівняння регресії, яке відображає залежність вартості однокімнатної квартири від її площі.

Таблиця 1.2 – **Вихідна інформація щодо вартості та площі квартир**

Номер з/п	Вартість квартири, дол. США / м. кв.	Площа квартири, м. кв.
1	30000	31,2
2	31200	32
3	32100	32
4	35280	37
5	32580	32
6	32580	32
7	32580	34
8	32100	29
9	34440	33
10	33420	31
11	33180	30
12	36120	34
13	42060	38
14	38520	31
15	42900	39
16	43140	39,5

Для побудованого рівняння необхідно обчислити:

1) коефіцієнти кореляції та детермінації;
2) F-критерій Фішера ;
3) стандартні помилки коефіцієнтів регресії;
4) t-статистики Ст’юдента ;
5) довірчі границі коефіцієнту регресії;
6) надати оцінку коефіцієнту регресії побудованої моделі.

Всі розрахунки провести в Excel з використанням вище наведених формул і «Пакета аналізу». Результати, отримані за формулами та за допомогою «Пакета аналізу», порівняти між собою.

Розв’язання:

Підготовимо дані для розрахунку оцінок коефіцієнтів регресії, які оформимо у вигляді табл. 1.3.

Таблиця 1.3 – **Дані для розрахунку оцінок коефіцієнтів регресії**

Номер з/п	**y**	**x**	**y^2**	**x^2**	**yx**
1	30000	31,2	900000000	973,44	936000
2	31200	32	973440000	1024	998400
3	32100	32	1030410000	1024	1027200
4	35280	37	1244678400	1369	1305360
5	32580	32	1061456400	1024	1042560
6	32580	32	1061456400	1024	1042560
7	32580	34	1061456400	1156	1107720
8	32100	29	1030410000	841	930900
9	34440	33	1186113600	1089	1136520
10	33420	31	1116896400	961	1036020
11	33180	30	1100912400	900	995400
12	36120	34	1304654400	1156	1228080
13	42060	38	1769043600	1444	1598280
14	38520	31	1483790400	961	1194120
15	42900	39	1840410000	1521	1673100
16	43140	39,5	1861059600	1560,25	1704030
Середнє значення	35138	33,42	1251636750	1127	1184766

Розрахунок коефіцієнтів регресії:

$$b = \frac{\overline{y \times x} - \overline{y} \times \overline{x}}{\sum_{i=1}^{n} x^2 - \overline{x}^2} = \frac{1184766 - 35138 \times 33{,}42}{1127 - 33{,}42^2} = 1060{,}147$$

$$a = \overline{y} - b \times \overline{x} = 35138 - 1060{,}147 \times 33{,}42 = -291{,}285$$

Побудована модель може бути записана в наступному вигляді:

$$y = -291{,}285 + 1060{,}147x$$

Коефіцієнт регресії *b* цієї моделі показує, що в середньому збільшення корисної площі на 1 кв. м. призводить до збільшення її вартості на 1060,147 дол. США

1. Розрахунок коефіцієнта кореляції й детермінації

$$r_{xy} = b \times \frac{\sigma_x}{\sigma_y} = b \times \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n} x^2 - \bar{x}^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} y^2 - \bar{y}^2}} = 1060{,}147 \times \frac{\sqrt{1127 - 33{,}42^2}}{\sqrt{1251636750 - 35138^2}} = 0{,}81$$

$$R^2 = r_{yx}^{\ 2} = 0{,}81^2 = 0{,}656$$

Коефіцієнт кореляції досить високий, що свідчить про істотну залежність вартості квартир від її площі. Коефіцієнт детермінації показує, що величина вартості квартири пояснюється величиною корисної площі тільки на 65,6 %.

2. Розрахунок F-критерію Фішера

$$F = \frac{r_{yx}^{\ 2}}{1 - r_{yx}^{\ 2}} \times (n-2) = \frac{0{,}81^2}{1 - 0{,}81^2} \times (16-2) = 26{,}69$$

Порівняння розрахункового значення F-критерію з табличним $F_{1;\,14} = 4{,}60$ для 95%-вого рівня значимості дозволяє зробити висновок про адекватність побудованої моделі.

3. Розрахунок стандартних помилок, у яких використовується середньоквадратична помилка $S_{зал}$, обчислена відповідно до даних табл. 1.4.

$$m_b = \sqrt{\frac{\sum (y - \hat{y}_x)^2 / (n-2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S^2}{\sum (x - \bar{x})^2}}$$

Таблиця 1.4 – **Дані для розрахунку стандартних помилок**

Номер з/п	Вартість квартири, дол. США , (y)	Площа квартири, м. кв., (x)	$(x-\bar{x})^2$	$\hat{y}$	$(y-\hat{y})^2$	x^2
1	30000	31,2	5	32785,3	7757904	973,44
2	31200	32	2	33633,42	5921528	1024
3	32100	32	2	33633,42	2351374	1024
4	35280	37	13	38934,15	13352841	1369
5	32580	32	2	33633,42	1109692	1024
6	32580	32	2	33633,42	1109692	1024

Продовження таблиці 1.4

7	32580	34	0	35753,71	10072454	1156
8	32100	29	20	30452,98	2712681	841
9	34440	33	0	34693,57	64295,72	1089
10	33420	31	6	32573,27	716948,3	961
11	33180	30	12	31513,13	2778472	900
12	36120	34	0	35753,71	134166,2	1156
13	42060	38	21	39994,3	4267112	1444
14	38520	31	6	32573,27	35363574	961
15	42900	39	31	41054,45	3406062	1521
16	43140	39,5	37	41584,52	2419513	1560,25
Сума	35138	33	159	-	93538310	18027,69

$$m_b = \sqrt{\frac{93538310/(16-2)}{159}} = 205{,}19$$

4. Розрахунок t-статистики Ст'юдента

$$t_b = \frac{b}{m_b} = \frac{1060{,}147}{205{,}19} = 5{,}1665$$

tтабл для числа ступенів вільності n-2=16-2=14 та α=0,05 дорівнює 2,145.

Розраховане значення t-статистики перевищує табличне значення що свідчить про статистичну значущість параметра b

5. Розрахунок довірчих границь для коефіцієнтів рівняння регресії

Інтервал довіри будується за схемою

$$b - t_{табл} \times m_b < b < b + t_{табл} \times m_b$$

1060,147-2,145×5,1665<b<1060,147+2,145×5,1665

6. Побудова лінійного рівняння регресії й розрахунок всіх його характеристик за допомогою «Пакета аналізу» табличного процесора Excel.

Для того, щоб мати можливість проводити регресійні розрахунку у табличному середовищі Excel, необхідно у панелі Меню «Сервіс» вибрати «Надстройки» (рис. 1.1).

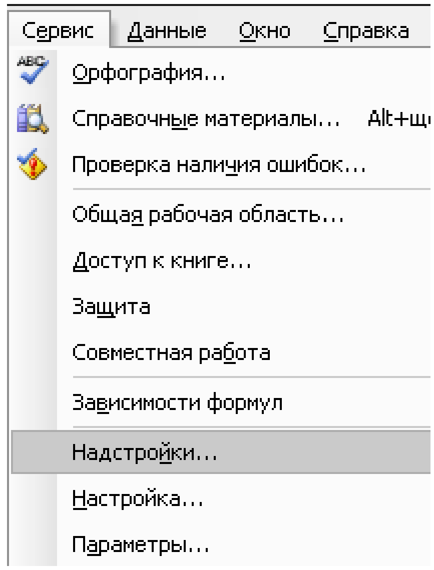


Рис. 1.1. - Вікно Меню Сервіс

З’явиться вікно «Надстройки», у якому навпроти надстройки «Пакет анализа» поставити позначку та натиснути ОК (рис. 1.2.)

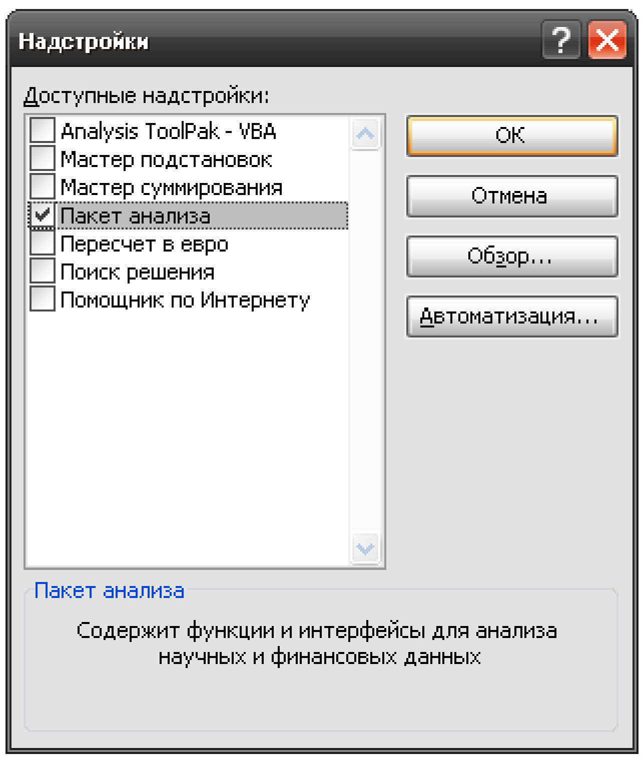


Рис. 1.2 – Вікно «Надстройки»

Після цього у меню «Сервіс» з’явиться пункт «Анализ данных» (рис. 1.3).

Рис 1.3. - Вікно меню «Сервіс»

Тепер внесемо усі вихідні дані у робочий лист Excel (рис. 1.4).

	A	B	C	D
1	Номер з/п	Вартість квартири, дол. США , (y)	Площа квартири, м. кв., (x)	
2	1	30000	30,2	
3	2	31200	32	
4	3	32100	32	
5	4	35280	37	
6	5	32580	30	
7	6	32580	30	
8	7	32580	30	
9	8	32100	29	
10	9	34440	33	
11	10	33420	31	
12	11	33180	30	
13	12	36120	34	
14	13	42060	38	
15	14	38520	31	
16	15	42900	39	
17	16	43140	39,5	
18				

Рис. 1.4 – Вікно вводу даних

Далі у меню «Сервіс» обираємо «Анализ даннях». З’явиться відповідне вікно у якому треба обрати пункт «Регрессия» (рис. 1.5).

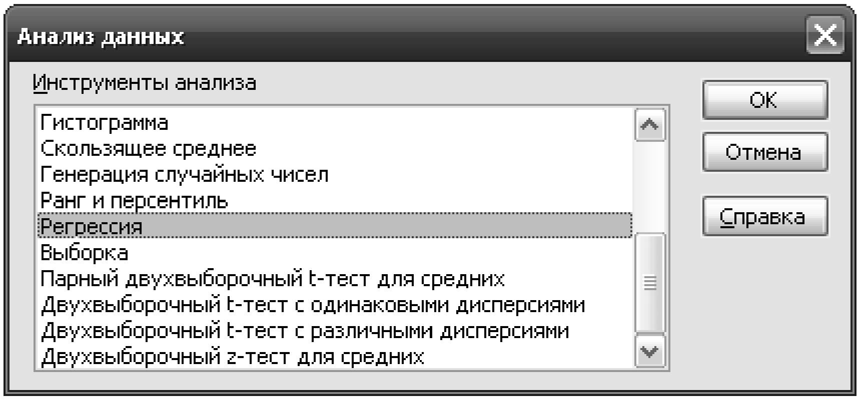


Рис. 1.5 – Вікно «Аналіз даних»

З'явиться вікно «Регрессия», у якому у вхідному інтервалі Y треба вказати діапазон даних, що відповідає y (у нашому випадку B2:B17) та відповідно вхідний інтервал X (C2:C17) і натиснути ОК (рис. 1.6).

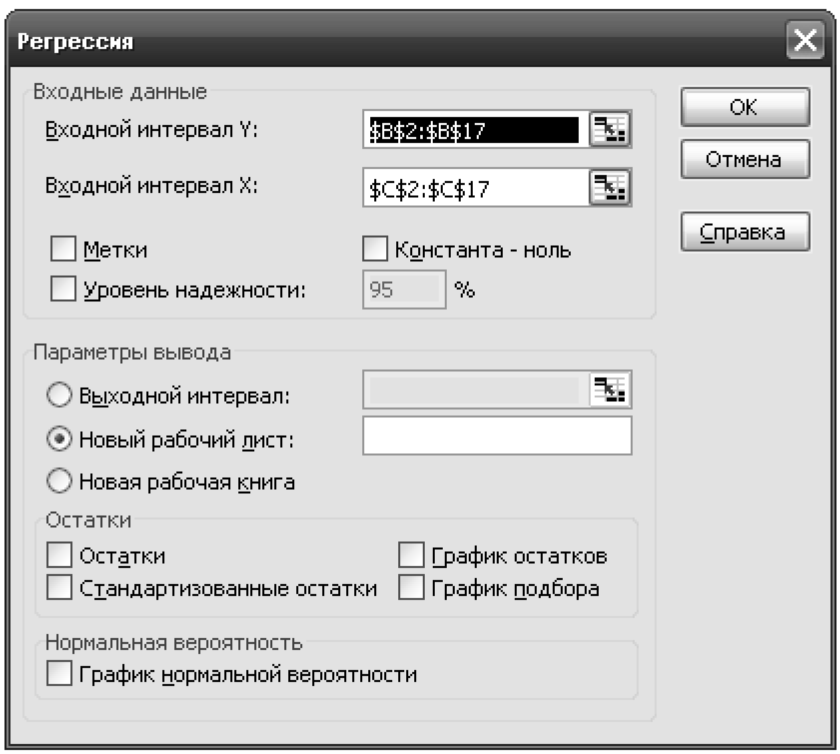


Рис. 1.6 – Вікно «Регресія»

Після цього на новому листі буде виведено результати розрахунків, які зображено на рис.1.7.

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,809916248
R-квадрат	0,655964329
Нормированный R-кв	0,631390353
Стандартная ошибка	2584,822601
Наблюдения	16

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	178347189,7	178347189,7	26,69345476	0,000143079
Остаток	14	93538310,33	6681307,88		
Итого	15	271885500			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	-291,2845861	6887,691852	-0,042290595	0,966864275	-15063,9143	14481,35
Переменная X 1	1060,146911	205,1935285	5,16657089	0,000143079	620,0505642	1500,243

Рис. 1.7 – Результати розрахунків регресії з використанням табличного процесора Microsoft Office Excel 2003

Порівняння результатів, отриманих за допомогою розрахункових формул, з результатами застосування інструментальних засобів Excel показує їхню повну ідентичність, що свідчить про правильне розуміння методу побудови лінійних регресійних рівнянь і методики оцінки його якості .

Індивідуальні завдання для самостійного виконання

Для відповідного варіанта (таблиця 1.5) на основі даних (таблиці 1.6, 1.7) за допомогою середовища Excel необхідно:

1) побудувати лінійне рівняння регресії;
2) надати оцінку коефіцієнту регресії побудованої моделі;
3) розрахувати коефіцієнти кореляції та детермінації;
4) розрахувати F-критерій Фішера та зробити висновки щодо якості побудованої моделі;
5) розрахувати стандартні помилки коефіцієнтів регресії;
6) розрахувати t-статистику Ст’юдента та оцінити значимість параметрів лінійного рівняння регресії;
7) побудувати інтервали довіри для коефіцієнта регресії.

Таблиця 1.5 – **Варіанти виконання індивідуального завдання**

Варіант	Номер стовбця для Y	Номера стовбців для пояснюючих змінних X						
		1	**2**	**3**	**4**	**5**	**6**	**7**
1	1	x_1						
2	1		x_1					
3	1			x_1				
4	1				x_1			
5	1					x_1		
6	1						x_1	
7	1							x_1
8	2	x_1						
9	2		x_1					
10	2			x_1				
11	2				x_1			
12	2					x_1		
13	2						x_1	
14	2							x_1
15	3	x_1						
16	3		x_1					
17	3			x_1				
18	3				x_1			
19	3					x_1		
20	3						x_1	
21	3							x_1
22	1	x_1						
23	2		x_1					
24	3			x_1				
25	1				x_1			

Таблиця 1.6 – **Роздрібний товарооборот за регіонами України за рік (Y)**

(млн. грн.)

Регіони	**Роздрібний товарооборот підприємств**	**Роздрібний товарооборот за продовольчими товарами**	**Роздрібний товарооборот за непродовольчими товарами**
	1	**2**	**3**
Вінницька	10897,8	4725,4	6172,4
Волинська	8515,1	3679,0	4836,1
Дніпропетровська	49591,7	20695,0	28896,7
Донецька	13852,8	5474,7	8378,1
Житомирська	10914,5	4843,4	6071,1
Закарпатська	10876,0	4286,5	6589,1
Запорізька	22794,9	8606,7	14188,2
Івано-Франківська	10185,4	3441,8	6744,2
Київська	37420,0	15960,5	21459,5
Кіровоградська	9283,3	4079,2	5204,1
Луганська	2813,8	1232,7	1581,0
Львівська	33154,6	13641,3	19512,7
Миколаївська	11668,2	4237,3	7430,2
Одеська	42585,8	18613,0	23972,8
Полтавська	14860,9	6277,7	8583,2
Рівненська	8359,0	3773,1	4586,4
Сумська	8649,3	4062,8	4588,0
Тернопільська	5849,4	2684,9	3164,5
Харківська	40943,1	18810,9	22132,2
Херсонська	11386,4	3742,1	7644,3
Хмельницька	10154,4	3300,1	6854,3
Черкаська	9968,2	3969,9	5998,3
Чернівецька	6396,9	2439,8	3956,7
Чернігівська	9087,7	4744,0	4343,7
м.Київ	121607,1	49692,6	71913,8

Таблиця 1.7 – **Макроекономічні показники за регіонами України за рік (Х)**

Регіони	Кількість об'єктів роздрібної торгівлі, тис.од (1)	Торгова площа магазинів, тис.м^2 (2)	Середньомісячна номінальна заробітна плата, грн. (3)	Доходи населення, млн.грн (4)	Чисельність населення, тис. осіб (5)	Обсяги імпорту, млн. дол. США (6)	Індекси споживчих цін, % (7)
Вінницька обл.	2,1	209,6	6121	69654	1583,7	371,5	113
Волинська обл.	1,4	160,1	5849	39359	1039,8	1238,7	114,6
Дніпропетровська обл.	8,2	701,2	6939	184138	3231,7	4165,5	112,8
Донецька обл.	3,1	232,2	7764	111547	4223,4	1794,4	115,9
Житомирська обл.	1,3	163,6	5836	51920	1236,3	419,3	113
Закарпатська обл.	1,2	184,5	6355	42235	1258,6	1232,4	113,9
Запорізька обл	3,7	323,0	6863	94160	1731,8	1179,2	114,1
Івано-Франківська обл.	1,8	167,8	6074	54492	1378,9	559,2	113,7
Київська обл.	5,0	476,5	7188	87937	1743,8	3109,3	113,8
Кіровоградська обл.	1,4	141,1	5792	40427	961,4	205,4	113,8
Луганська обл.	0,7	53,2	5862	38022	2182,2	241,0	114,9
Львівська обл.	4,2	466,8	6391	112697	2532,2	1984,5	113
Миколаївська обл.	2,0	207,5	6709	50728	1146,1	700,0	113,5
Одеська обл.	4,4	637,1	6542	115025	2384,6	1293,4	114,6
Полтавська обл.	2,1	235,6	6551	69789	1420,9	1040,6	113,1
Рівненська обл.	1,1	179,3	6013	45716	1161,9	305,2	115
Сумська обл.	1,3	160,6	5946	50951	1099,8	492,9	113,6
Тернопільська обл.	0,8	116,8	5554	38727	1056,0	324,0	113,3
Харківська обл.	5,9	582,8	6244	131681	2698,1	1472,3	113,8
Херсонська обл.	1,6	208,8	5842	42707	1051,6	183,3	114,4
Хмельницька обл.	1,4	216,8	5938	55542	1280,3	378,3	113,8
Черкаська обл.	1,8	158,1	6042	51710	1226,3	373,2	114,6
Чернівецька обл.	0,8	118,9	5621	32397	907,4	102,3	112,4
Чернігівська обл.	1,3	165,6	5636	44283	1027,3	407,4	113,9
м.Київ	24,3	1430,1	11135	333927	2929,6	17696,2	113,4

МНОЖИННА ЛІНІЙНА РЕГРЕСІЯ ТА КОРЕЛЯЦІЯ

Основні теоретичні положення

Поняття класичної багатофакторної регресії. Специфікація моделі

Побудова ***множинної регресії*** – один з найбільш розповсюджених методів в економетрії. Основна мета множинної регресії – побудувати модель з великим числом факторів, визначивши при цьому вплив кожного з них окремо, а також сукупний їхній вплив на показник, який досліджується.

Рівняння множинної регресії має вигляд:

$$y = a + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p + e \tag{2.1}$$

де y – залежна змінна;

a, b_1, b_2, b_p – невідомі параметри рівняння регресії;

x_1, x_2, x_p – незалежні змінні (фактори);

e – випадкова величина.

Побудова рівняння множинної регресії починається з рішення питання про специфікацію моделі. Суть проблеми специфікації містить у собі два кола питань: добір факторів і вибір виду рівняння регресії. Їхнє рішення при побудові моделі множинної регресії має деяку специфіку.

Включення в рівняння множинної регресії того чи іншого набору факторів пов'язано насамперед із представленням дослідника про природу взаємозв'язку показника з іншими економічними явищами.

Фактори, що включаються в множинну регресію, повинні відповідати наступним ***вимогам:***

1. Вони повинні бути кількісно виміряні. Якщо необхідно включити в модель якісний фактор, що не має кількісного виміру, то йому необхідно додати кількісну визначеність.

2. Фактори не повинні бути інтеркорельовані і тим більше знаходитися в точному функціональному зв'язку.

Включення в модель факторів з високою інтеркореляцією, коли $R_{yx1} < R_{x1x2}$ для залежності $y = a + b_1 x_1 + b_2 x_2 + e$ може привести до небажаних наслідків – система нормальних рівнянь може виявитися погано обумовленою і спричинити за собою нестійкість і ненадійність оцінок коефіцієнтів регресії.

Якщо між факторами існує висока кореляція, то не можна визначити їхній окремий вплив на результативний показник і параметри рівняння регресії виявляються неінтерпретованими. Так, у рівнянні $y = a + b_1 x_1 + b_2 x_2 + e$ передбачається, що фактори x_1 і x_2 незалежні друг від друга, тобто $r_{x1x2} = 0$. Тоді можна говорити, що параметр b_1 вимірює силу впливу x_1 на результат y при незмінному значенні фактора x_2. Якщо ж $r_{x1x2} = 1$, то зі зміною фактора x_1 фактор x_2 не може залишатися незмінним. Звідси b_1 і b_2 не можна інтерпретувати як показники роздільного впливу x_1 і x_2 на y.

3. Фактори, які входять до моделі, повинні пояснювати варіацію незалежної змінної. Якщо будується модель з набором p факторів, то для неї розраховується показник детермінації R^2, що фіксує частку поясненої варіації результативної ознаки за рахунок розглянутих у регресії p факторів. Вплив інших факторів, що не ввійшли до моделі, оцінюється як $(1–R^2)$ з відповідною залишковою дисперсією S^2.

При додатковому включенні в регресію $p+1$ фактора коефіцієнт детермінації повинний зростати, а залишкова дисперсія зменшуватися:

$$R_{p+1}^2 \geq R_p^2; S_{p+1}^2 \leq S_p^2 \tag{2.2}$$

Якщо ж цього не відбувається і дані показники практично мало відрізняються один від одного, то фактор x_{p+1}, який включається до аналізу, не поліпшує модель і практично є зайвим чинником.

Насичення моделі зайвими факторами не тільки не знижує величину залишкової дисперсії і не збільшує показник детермінації, але і приводить до статистичної незначимості параметрів регресії по t-критерію Ст'юдента.

Добір факторів здійснюється в двох стадіях: на першій обираються фактори, виходячи із сутності проблеми; на другій – на основі матриці показників кореляції визначають t-статистики для параметрів регресії.

Коефіцієнти кореляції між пояснюючими змінними дозволяють виключати з моделі дублюючі фактори. Вважається, що дві змінні явно колінеарні, тобто знаходяться між собою в лінійній залежності, якщо $r_{x_i\ x_j} \geq 0,7$.

4. Однією з умов, які висуваються до факторів множинної регресії, є перевірка однорідності вихідної інформації.

Оцінка параметрів багатофакторної регресії

Параметри рівняння множинної регресії оцінюються, як і у випадку простої регресії, за допомогою методу найменших квадратів (МНК). При його застосуванні будується система нормальних рівнянь, рішення якої і дозволяє одержати оцінки параметрів регресії:

$$\begin{cases} \sum y = n \times a + b_1 \times \sum x_1 + ... + b_p \times \sum x_p; \\ \sum yx_1 = a \times \sum x_1 + b_1 \times \sum x_1^2 + b_2 \times \sum x_1 x_2 + .. + b_p \times \sum x_p x_1; \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots; \\ \sum yx_p = a \times \sum x_p + b_1 \times \sum x_1 x_p + b_2 \times \sum x_2 x_p + ... + b_p \times \sum x_p^2. \end{cases}$$

Її рішення може бути здійснено методом визначників:

$$a = \frac{\Delta a}{\Delta}; \quad b_1 = \frac{\Delta b_1}{\Delta}; ......; b_p = \frac{\Delta b_p}{\Delta}, \tag{2.3}$$

де Δ – визначник системи;

Δa, Δb_1,..., Δb_p – приватні визначники.

Як і у випадку простої регресії, параметр a розраховується за допомогою середніх значень:

$$a = \bar{y} - b_1 \cdot \bar{x}_1 - ... - b_p \cdot \bar{x}_p. \tag{2.4}$$

У випадку лінійної множинної регресії $y=a+b_1 \cdot x_1+b_2 \cdot x_2+....+b_p \cdot x_p+\varepsilon$ параметри при x називаються ***коефіцієнтами "чистої" регресії***. Вони характеризують середню зміну результату зі зміною відповідного фактору на одну одиницю при незмінному значенні інших факторів, закріплених на середньому рівні.

Зважаючи на те, що коефіцієнти "чистої" регресії мають чітке економічне пояснення, лінійна множинна регресія отримала найбільше поширення в економетричних дослідженнях.

Можливий і інший підхід до визначення параметрів множинної регресії, коли на основі матриці парних коефіцієнтів кореляції будується рівняння регресії в стандартизованому масштабі:

$$t_y = \beta_1 . t x_1 + \beta_2 . t x_2 + ... + \beta_p . t_{xp} + \varepsilon, \tag{2.5}$$

де t_y, t_{x1},... , t_{xp} - стандартизовані змінні:

$$t_y = \frac{y - \bar{y}}{\sigma_y}, t_{x_i} = \frac{x_i - \bar{x}}{\sigma_{x_i}},$$

для яких середнє значення дорівнює нулю: $\bar{t}_y = \bar{t}_{x_i} = 0,$, а середнє квадратичне відхилення дорівнює одиниці $\sigma_{t_y} = \sigma_{t_x} = 1;$;

β - стандартизовані коефіцієнти регресії.

Коефіцієнти множинної кореляції та детермінації

Практична значимість рівняння множинної регресії оцінюється за допомогою показника множинної кореляції і його квадрата – ***коефіцієнта детермінації.***

Показник множинної кореляції характеризує тісноту зв'язку розглянутого набору факторів з досліджуваною ознакою, або, інакше, оцінює тісноту спільного впливу факторів на результат.

Незалежно від форми зв'язку показник множинної кореляції може бути знайдений як ***індекс множинної кореляції***:

$$R_{yx_1x_2\ldots x_p} = \sqrt{1 - \frac{\sigma^2_{зал}}{\sigma^2_y}},\ \sqrt{1 - \frac{\sum(y - \hat{y})^2}{\sum(y - \bar{y})^2}} \tag{2.6}$$

де σ^2_y – загальна дисперсія результативної ознаки;

$\sigma^2_{зал}$ – залишкова дисперсія для рівняння $y = f(x_1, x_2, \ldots, x_p)$.

Методика побудови індексу множинної кореляції аналогічна побудові індексу кореляції для парної залежності. Границі його зміни ті ж: від 0 до 1. Чим ближче його значення до 1, тим тісніше зв'язок результативної ознаки з усім набором досліджуваних факторів. Величина індексу множинної кореляції повинна бути більше або дорівнювати максимальному парному індексу кореляції:

$$R_{yx_1x_2\ldots x_p} \geq R_{yx_i(\max)} \left(i = \overline{1, p}\right).$$

При розрахунках показників множинної кореляції (індексу і коефіцієнту) використовується залишкова дисперсія, яка має систематичну помилку убік зменшення, тим більше значну, чим більше параметрів визначається в рівнянні регресії при заданому обсязі спостережень *n*. Якщо число параметрів при x_j дорівнює *m* і наближається до обсягу спостережень, то залишкова дисперсія буде близька до нуля і коефіцієнт (індекс) кореляції наблизиться до одиниці навіть при слабкому зв'язку факторів з результатом. Для того, щоб не допустити можливого перебільшення тісноти зв'язку, використовується скорегований індекс (коефіцієнт) множинної кореляції.

Скорегований індекс множинної кореляції містить виправлення на число ступенів вільності, а саме залишкова сума квадратів $\sum\left(y - \hat{y}_{x_1x_2\ldots x_p}\right)^2$ поділяється на число ступенів вільності залишкової варіації $(n - m - 1)$, а загальна сума квадратів відхилень $\sum\left(y - \overline{y}\right)^2$ – на число ступенів вільності в цілому по сукупності $(n - 1)$.

Формула скоректованого індексу множинної детермінації має вид:

$$\overline{R^2} = 1 - \frac{\sum(y - \hat{y})^2 : (n - m - 1)}{\sum(y - \bar{y})^2 : (n - 1)}, \tag{2.7}$$

де m – число параметрів при змінних x;

n – число спостережень.

Оскільки $\sum(y-\hat{y})^2 / \sum(y-\bar{y})^2 = 1 - R^2$, то величину скорегованого індексу детермінації можна представити у вигляді

$$\overline{R^2} = 1 - \left(1 - R^2\right) \cdot \frac{(n-1)}{(n-m-1)} . \qquad (2.8)$$

Чим більше величина m, тим сильніше розходження між $\overline{R^2}$ і R^2.

Частні рівняння множинної регресії

На основі лінійного рівняння множинної регресії

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_p \cdot x_p + \varepsilon$$

можуть бути знайдені ***частні рівняння регресії:***

$$\begin{cases} y_{x_1 \cdot x_2, x_3, ..., x_p} = f(x_1), \\ y_{x_2 \cdot x_1, x_3, ..., x_p} = f(x_2), \\ \dots\dots\dots\dots\dots\dots\dots\dots, \\ y_{x_p \cdot x_1, x_2, ..., x_{p-1}} = f(x_p), \end{cases} \qquad (2.9)$$

тобто рівняння регресії, які пов’язують результативну ознаку з відповідними факторами x при закріпленні інших факторів множинної регресії на середньому рівні. Частні рівняння регресії мають наступний вигляд:

$$y_{x_1 \cdot x_2, x_3, ..., x_p} = a + b_1 \cdot x_1 + b_2 \cdot \bar{x}_2 + b_3 \cdot \bar{x}_3 + ... + b_p \cdot \bar{x}_p + e;$$
$$y_{x_2 \cdot x_1, x_3, ..., x_p} = a + b_1 \cdot \bar{x}_1 + b_2 \cdot x_2 + b_3 \cdot \bar{x}_3 + ... + b_p \cdot \bar{x}_p + e;$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$y_{x_p \cdot x_1, x_2, ..., x_{p-1}} = a + b_1 \cdot \bar{x}_1 + b_2 \cdot \bar{x}_2 + ... + b_{p-1} \cdot \bar{x}_{p-1} + b_p \cdot x_p + e. \qquad (2.10)$$

Під час підстановки в ці рівняння середніх значень відповідних факторів вони приймають вигляд парних рівнянь лінійної регресії, тобто маємо:

$$\begin{cases} \hat{y}_{x_1 \cdot x_2, x_3, ..., x_p} = A_1 + b_1 \cdot x_1, \\ \hat{y}_{x_2 \cdot x_1, x_3, ..., x_p} = A_2 + b_2 \cdot x_2, \\ ...., \\ y_{x_p \cdot x_1, x_2, ..., x_{p-1}} = A_p + b_p \cdot x_p, \end{cases}$$

де

$$\begin{cases} A_1 = a + b_2 \cdot \bar{x}_2 + b_3 \cdot \bar{x}_3 + ... + b_p \cdot \bar{x}_p, \\ A_2 = a + b_1 \cdot \bar{x}_1 + b_3 \cdot \bar{x}_3 + ... + b_p \cdot \bar{x}_p, \\ ...., \\ A_p = a + b_1 \cdot \bar{x}_1 + b_2 \cdot \bar{x}_2 + ... + b_{p-1} \cdot \bar{x}_{p-1}. \end{cases} \quad (2.11)$$

На відміну від простої регресії *частні рівняння регресії* характеризують ізольований вплив фактора на результат, тому що інші фактори закріплені на незмінному рівні. Ефекти впливу інших факторів поєднані у них до вільного члену рівняння множинної регресії. Це дозволяє на основі часткових рівнянь регресії визначати ***частні коефіцієнти еластичності***:

$$Э_{y_{x_i}} = b_i \cdot \frac{x_i}{\hat{y}_{x_i \cdot x_1 x_2 ...... x_{i-1} x_{i+1} ... x_p}}, \quad (2.12)$$

де b_i – коефіцієнти регресії для фактора x_i в рівнянні множинної регресії;

$\hat{y}_{x_i \cdot x_1 x_2 ... x_{i-1} x_{i+1} ... x_p}$ – частне рівняння регресії.

Частні коефіцієнти еластичності показують, на скільки відсотків в середньому зміниться результат y, якщо відповідний фактор x_i зміниться на 1 %, при цьому інші фактори залишатимуться незмінними.

Оцінка надійності результатів множинної регресії та кореляції

Значимість рівняння множинної регресії в цілому, так само як і в парній регресії, оцінюється за допомогою *F-критерію Фішера*:

$$F = \frac{D_{факт}}{D_{зал}} = \frac{R^2}{1 - R^2} \times \frac{n - m - 1}{m}, \quad (2.13)$$

де $D_{факт}$ – факторна сума квадратів на одну ступінь вільності;

$D_{зал}$ – залишкова сума квадратів на одну ступінь вільності;

R^2 – коефіцієнт (індекс) множинної детермінації;

m – число параметрів при змінних x;

n – число спостережень.

Оцінюється значимість не тільки рівняння в цілому, але і фактора, додатково включеного в регресійну модель. Необхідність такої оцінки пов'язана з тим, що не кожен фактор, що ввійшов у модель, може істотно збільшувати частку поясненої варіації результативної ознаки. Крім того, при наявності в моделі декількох факторів вони можуть вводитися в модель у різній послідовності. Через кореляцію між факторами значимість того самого фактора може бути різною в залежності від послідовності його введення в модель. Мірою для оцінки включення фактора в модель служить ***частковий F-критерій, тобто*** $\boldsymbol{F_{x_i}}$.

Припустимо, що оцінюємо значимість впливу x_1 як додатково включеного в модель фактора. Використовуємо наступну формулу:

$$F_{x_1} = \frac{R^2_{yx_1x_2\ldots x_p} - R^2_{yx_2\ldots x_p}}{1 - R^2_{yx_1x_2\ldots x_p}} \times \frac{n-m-1}{1}, \tag{2.14}$$

де $R^2_{yx_1x_2\ldots x_p}$ – коефіцієнт множинної детермінації для моделі з повним набором факторів;

$R^2_{yx_2\ldots x_p}$ – той же показник, але без включення в модель фактора x_1;

n – число спостережень;

m – число параметрів у моделі (без вільного члена).

Якщо оцінюємо значимість впливу фактора x_p після включення в модель факторів $x_1, x_2, \ldots, x_{p-1}$, то формула частного F-критерію прийме вигляд:

$$F_{x_p} = \frac{R^2_{yx_1x_2\ldots x_p} - R^2_{yx_1x_2\ldots x_{p-1}}}{1 - R^2_{yx_1x_2\ldots x_p}} \times \frac{n-m-1}{1}. \tag{2.15}$$

У загальному виді для фактора x_i частний F- критерій визначиться як

$$F_{x_i} = \frac{R^2_{yx_1\ldots x_i\ldots x_p} - R^2_{yx_1\ldots x_{i-1}x_{i+1}\ldots x_p}}{1 - R^2_{yx_1\ldots x_i\ldots x_p}} \times \frac{n-m-1}{1}. \tag{2.16}$$

Фактичне значення частного F-критерію порівнюється з табличним при 5%-му або 1%-му рівні значимості і кількості ступенів вільності: 1 та $n-m-1$. Якщо фактичне значення F_{x_i} перевищує $F_{табл}$ (α, df_1, df_2), то додаткове включення фактора x_i у модель статистично виправдано і коефіцієнт чистої регресії b_i при факторі x_i статистично значимий. Якщо ж фактичне значення F_{x_i} менше табличного, то додаткове включення в модель фактора x_i не збільшує істотно частку поясненої варіації ознаки y, отже, недоцільно його включення у модель. Коефіцієнт регресії при даному факторі в цьому випадку статистично не значимий.

За допомогою частного F-критерію можна перевірити значимість усіх коефіцієнтів регресії в припущенні, що кожен відповідний фактор x_i вводився в рівняння множинної регресії останнім.

Частний F-критерій оцінює значимість коефіцієнтів чистої регресії. Знаючи величину F_{x_i}, можна визначити і t-критерій для коефіцієнта регресії при i-том факторі, t_{b_i}, а саме:

$$t_{b_i} = \sqrt{F_{x_i}}\ . \tag{2.17}$$

Оцінка значимості коефіцієнтів чистої регресії за *t-критерієм Ст'юдента* може бути проведена і без розрахунку частних F-критеріїв. У цьому випадку, як і в парній регресії, для кожного фактора використовується формула

$$t_{b_i} = \frac{b_i}{m_{b_i}}, \tag{2.18}$$

де b_i – коефіцієнт чистої регресії при факторі x_i;

m_{b_i} – стандартна помилка коефіцієнта регресії b_i.

Для рівняння множинної регресії

$$\hat{y} = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \ldots b_p \cdot x_p$$

стандартна помилка коефіцієнта регресії може бути визначена за наступною формулою:

$$m_{b_i} = \frac{\sigma_y \cdot \sqrt{1 - R^2{}_{yx_1 \ldots x_p}}}{\sigma_{x_i} \cdot \sqrt{1 - R^2{}_{x_i x_1 \ldots x_p}}} \cdot \frac{1}{\sqrt{n - m - 1}}, \tag{2.19}$$

де σ_y – середньоквадратичне відхилення для ознаки y;

σ_{x_i} – середньоквадратичне відхилення для ознаки x_i;

$R^2{}_{yx_1 \ldots x_p}$ – коефіцієнт детермінації для рівняння множинної регресії;

$R^2{}_{x_i x_1 \ldots x_p}$ – коефіцієнт детермінації для залежності фактора x_i з всіма іншими факторами рівняння множинної регресії;

$n-m-1$ – число ступенів вільності для залишкової суми квадратів відхилень.

Як бачимо, щоб скористатися даною формулою, необхідна матриця міжфакторної кореляції і розрахунок по ній відповідних коефіцієнтів детермінації $R^2{}_{x_i x_1 \ldots x_p}$. Так, для рівняння $y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + \varepsilon$ оцінка значимості коефіцієнтів регресії b_1, b_2, b_3 припускає розрахунок трьох міжфакторних коефіцієнтів детермінації, а саме: $R^2{}_{x_1 \cdot x_2 x_3}$, $R^2{}_{x_2 \cdot x_1 x_3}$, $R^2{}_{x_3 \cdot x_1 x_2}$.

Разом з тим, якщо врахувати, що

$$b_i = \frac{\sigma_y}{\sigma_{x_i}} \cdot \sqrt{\frac{R^2{}_{yx_1 \ldots x_p} - R^2{}_{yx_1 \ldots x_{i-1} x_{i+1} \ldots x_p}}{1 - R^2{}_{x_i x_1 \ldots x_p}}}, \quad (2.20)$$

то можна переконатися, що

$$t_{b_i} = \frac{b_i}{m_{b_i}} = \sqrt{F_{x_i}}. \quad (2.21)$$

Запитання для самоперевірки знань

1. Основні етапи множинного кореляційного аналізу.
2. Як визначається модель множинної лінійної регресії?
3. Назвіть, у чому полягає специфікація моделі множинної регресії?
4. Сформулюйте вимоги, які висуваються до факторів для включення їх в модель множинної регресії?
5. Перечисліть передумови МНК. Які наслідки, якщо вони не виконуються?
6. Що характеризують коефіцієнти регресії? Як вони інтерпретуються?
7. У чому суть МНК для побудови множинного лінійного рівняння регресії?
8. Опишіть алгоритм визначення коефіцієнтів множинної лінійної регресії.
9. Як визначається статистична значущість коефіцієнтів регресії?
10. Які коефіцієнти використовуються для оцінки порівняльної сили впливу факторів на результат?
11. У чому сутність коефіцієнта детермінації?
12. Чим скоригований коефіцієнт детермінації відрізняється від

звичайного?

13. Від чого залежить величина скорегованого коефіцієнта детермінації?

14. Як здійснити аналіз статистичної значущості коефіцієнта детермінації?

15. Як використовується *F*-статистика в регресійному аналізі.

16. Що таке частний *F*-критерій та чим він відрізняється від послідовного *F*-критерію?

17. Як пов'язані між собою *t*-критерій Ст'юдента для оцінки значимості b_i та частний *F*-критерій?

Методичні вказівки до виконання індивідуальної роботи

За даними табл. 2.1 побудувати лінійне рівняння множинної регресії, яке відображає залежність рентабельності активів підприємств від продуктивності праці та фондоозброєності.

Таблиця 2.1– **Вихідні дані**

№ з/п	**Рентабельність діяльності, % (y)**	**Продуктивність праці, тис. грн.(x_1)**	**Фондоозброєності, тис. грн. (x_2)**
1	2,93	0,31	8,42
2	5,27	0,98	10,24
3	6,85	1,21	10,81
4	7,01	1,29	9,89
5	7,02	1,12	13,72
6	8,35	1,49	13,92
7	4,33	0,78	8,54
8	5,77	0,94	12,36
9	7,68	1,29	12,27
10	3,16	0,48	11,01
11	1,52	0,24	8,25
12	3,15	0,55	9,31
13	5,81	0,98	7,65
14	4,98	0,86	8,32

Продовження таблиці 2.1

15	7,54	1,39	12,45
16	6,25	1,13	9,15
17	3,24	0,61	8,32
18	4,26	0,72	9,42
19	5,48	0,91	10,63
20	6,21	1,07	11,62

Для цього необхідно:

1) перевірити наявність мультиколінеарності та відібрати неколінеарні фактори;

2) побудувати рівняння лінійної множинної регресії;

3) оцінити коефіцієнти регресії для кожного фактора;

4) визначити коефіцієнт множинної кореляції;

5) перевірити значимість рівняння при рівнях значимості 0,05 та 0,01 за допомогою F-критерію Фішера;

6) побудувати частні рівняння регресії;

7) визначити середні частні коефіцієнти еластичності;

8) зробити висновки.

Всі розрахунки провести в Excel з використанням вище наведених формул і «Пакета аналізу». Результати, отримані за формулами та за допомогою «Пакета аналізу», порівняти між собою.

Розв'язання:

Побудуємо кореляційну матрицю, використовуючи функцію «Сервис.Анализ данных.Корреляция» табличного процесору MS Excel.

	Столбец 1 (y)	*Столбец 2 (x1)*	*Столбец 3 (x2)*
Столбец 1 (y)	1		
Столбец 2 (x1)	0,983761	1	
Столбец 3 (x2)	0,687872	0,633946	1

Із матриці виходить, що не спостерігається колінеарності між факторами x1 та x2, так як *r* x1 x2 = 0,633946.

2) Для побудови рівняння лінійної регресії використовуємо функцію «Сервис.Анализ данных.Регрессия». Результати розрахунків надано на рис. 2.1

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,987259658
R-квадрат	0,974681632
Нормированный R-квадрат	0,971703001
Стандартная ошибка	0,313419107
Наблюдения	20

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	2	64,28755888	32,14378	327,2246	2,6866E-14
Остаток	17	1,669936119	0,098232		
Итого	19	65,957495			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	-0,199915686	0,402374932	-0,49684	0,625666	-1,048852578	0,649021206
Переменная X 1	4,859285909	0,264803985	18,3505	1,22E-12	4,300598341	5,417973477
Переменная X 2	0,104897806	0,048750437	2,151731	0,046084	0,002043376	0,207752235

Рис. - 2.1. Результати розрахунків множинної регресії з використанням табличного процесора Microsoft Office Excel 2003

Таким чином рівняння регресії має вигляд: $y = -0{,}1999 + 4{,}8593x_1 + 0{,}1049x_2$

3. Коефіцієнти регресії цієї моделі показують, що збільшення продуктивності праці на 1 тис. грн. призводить до збільшення рентабельності активів в середньому на 4,8593 % при незмінної фондоозброєності, а збільшення фондоозброєності на 1 тис. грн. при незмінній продуктивності праці збільшує рентабельність активів в середньому на 0,1049 %.

4. Коефіцієнт множинної регресії, виходячи з даних розрахунків (рис. 2.1) дорівнює $R_{y,x_1,x_2,\dots x_p} = 0{,}987$

5. Перевірка значимості рівняння регресії заснована на використанні *F*-критерію Фішера. Фактичне значення критерію дорівнює *F*факт = 327,2246 (рис.2.1).

Для визначення табличних значень використаємо вбудовану функцію MS Excel «FРАСПОБР» (рис. 2.2), задаючи параметри k1 = 2, k2 = 20 – 2 – 1 = 17, α = 0,05 и α = 0,01.

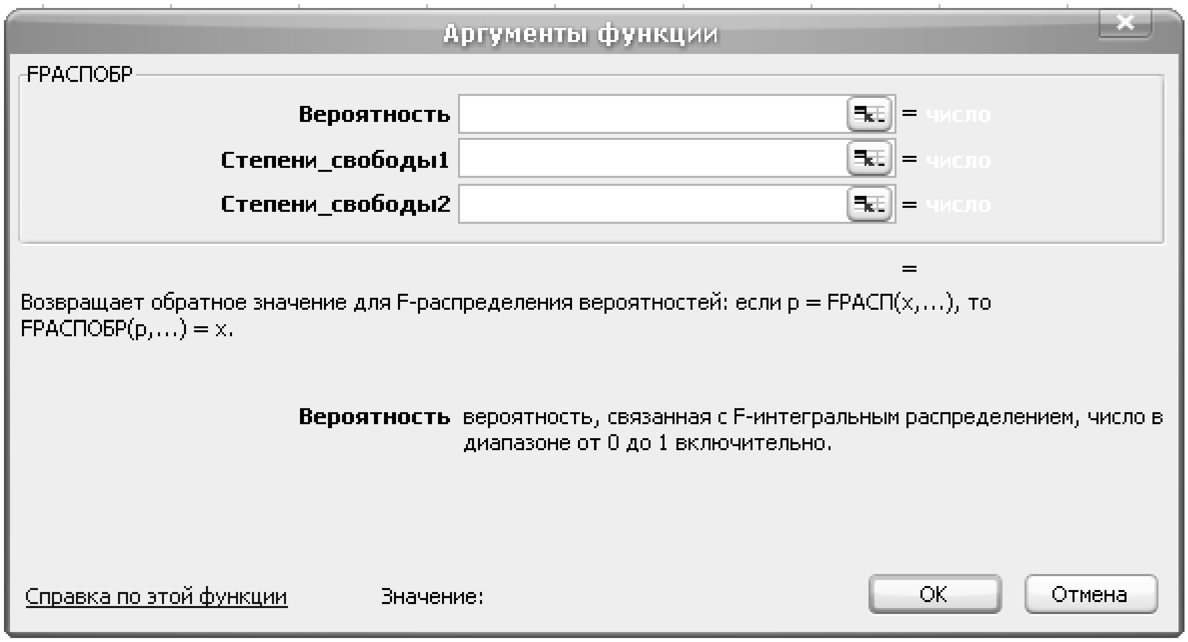


Рис.2.2 – Вікно функції «FРАСПОБР»

У результаті отримаємо $F_{\text{табл},0,05}$ = 3,5915, $F_{\text{табл},0,01}$ = 6,1121, які значно менші за $F_{факт}$, що свідчить про значимість рівняння регресії при α = 0,05, та α = 0,01.

6. Для розрахунку частних рівнянь регресії попередньо визначимо середні значення змінних $\overline{y} = 5{,}3405, \overline{x_1} = 0{,}9175, \overline{x_2} = 10{,}315$.

З урахуванням середніх значень побудуємо частні рівняння регресії

$$\hat{y}_{x_1,x_2} = -0{,}1999 + 4{,}8593x_1 + 0{,}1049 \times 10{,}315 = 0{,}882 + 4{,}8593x_1$$

$$\hat{y}_{x_2,x_1} = -0{,}1999 + 4{,}8593 \times 0{,}9175 + 0{,}1049x_2 = 4{,}2585 + 0{,}1049x_2$$

7. Середні частні коефіцієнти еластичності

$$\overline{Э_{yx_1}} = b_1 \times \frac{\overline{x_1}}{\overline{y}} = 4{,}8593 \times \frac{0{,}9175}{5{,}3405} = 0{,}83\,\%$$

$$\overline{Э_{yx_2}} = b_2 \times \frac{\overline{x_2}}{\overline{y}} = 0{,}1049 \times \frac{10{,}315}{5{,}3405} = 0{,}2026\,\%$$

Індивідуальні завдання для самостійного виконання

Для відповідного варіанта (таблиця 2.2) на основі даних (таблиці 2.3 та 2.4) за допомогою середовища Excel необхідно:

1) побудувати множинне рівняння лінійної регресії;

2) визначити коефіцієнт множинної кореляції;

3) перевірити значимість рівняння при рівнях значимості 0,05 і 0,01 за допомогою F-критерій Фішера;

4) побудувати частні рівняння регресії;

5) визначити середні частні коефіцієнти еластичності та зробити висновки.

Таблиця 2.2 – **Варіанти виконання індивідуального завдання**

Варіант	Номер стовбця для Y	Номера стовбців для пояснюючих змінних X						
		1	**2**	**3**	**4**	**5**	**6**	**7**
1	1	x_1	x_2	x_3				
2	1	x_1	x_2		x_3			
3	1	x_1	x_2			x_3		
4	1	x_1	x_2				x_3	
5	1	x_1	x_2					x_3
6	1	x_1		x_2	x_3			
7	1	x_1		x_2		x_3		
8	1	x_1		x_2			x_3	
9	1	x_1		x_2				x_3
10	1	x_1			x_2	x_3		
11	1	x_1			x_2		x_3	
12	1	x_1			x_2			x_3
13	1	x_1				x_2	x_3	
14	1	x_1				x_2		x_3
15	1	x_1					x_2	x_3
16	1		x_1	x_2	x_3			
17	1		x_1	x_2		x_3		
18	1		x_1	x_2			x_3	
19	1		x_1		x_2	x_3		
20	1		x_1		x_2		x_3	
21	1		x_1			x_2	x_3	
22	1			x_1	x_2	x_3		
23	1			x_1	x_2		x_3	
24	1			x_1		x_2	x_3	
25	1				x_1	x_2	x_3	

Таблиця 2.3 – **Роздрібний товарооборот за регіонами України за рік (Y)**

(млн. грн.)

Регіони	**Роздрібний товарооборот підприємств**	**Роздрібний товарооборот за продовольчими товарами**	**Роздрібний товарооборот за непродовольчими товарами**
	1	**2**	**3**
Вінницька	10897,8	4725,4	6172,4
Волинська	8515,1	3679,0	4836,1
Дніпропетровська	49591,7	20695,0	28896,7
Донецька	13852,8	5474,7	8378,1
Житомирська	10914,5	4843,4	6071,1
Закарпатська	10876,0	4286,5	6589,1
Запорізька	22794,9	8606,7	14188,2
Івано-Франківська	10185,4	3441,8	6744,2
Київська	37420,0	15960,5	21459,5
Кіровоградська	9283,3	4079,2	5204,1
Луганська	2813,8	1232,7	1581,0
Львівська	33154,6	13641,3	19512,7
Миколаївська	11668,2	4237,3	7430,2
Одеська	42585,8	18613,0	23972,8
Полтавська	14860,9	6277,7	8583,2
Рівненська	8359,0	3773,1	4586,4
Сумська	8649,3	4062,8	4588,0
Тернопільська	5849,4	2684,9	3164,5
Харківська	40943,1	18810,9	22132,2
Херсонська	11386,4	3742,1	7644,3
Хмельницька	10154,4	3300,1	6854,3
Черкаська	9968,2	3969,9	5998,3
Чернівецька	6396,9	2439,8	3956,7
Чернігівська	9087,7	4744,0	4343,7
м.Київ	121607,1	49692,6	71913,8

Таблиця 2.4 – **Макроекономічні показники за регіонами України за рік (Х)**

Регіони	**Кількість об'єктів роздрібної торгівлі, тис.од (1)**	**Торгова площа магазинів, тис.м^2 (2)**	**Середньомісячна номінальна заробітна плата, грн. (3)**	**Доходи населення, млн.грн (4)**	**Чисельність населення, тис. осіб (5)**	**Обсяги імпорту, млн. дол. США (6)**	**Індекси споживчих цін, % (7)**
Вінницька обл.	2,1	209,6	6121	69654	1583,7	371,5	113
Волинська обл.	1,4	160,1	5849	39359	1039,8	1238,7	114,6
Дніпропетровська обл.	8,2	701,2	6939	184138	3231,7	4165,5	112,8
Донецька обл.	3,1	232,2	7764	111547	4223,4	1794,4	115,9
Житомирська обл.	1,3	163,6	5836	51920	1236,3	419,3	113
Закарпатська обл.	1,2	184,5	6355	42235	1258,6	1232,4	113,9
Запорізька обл	3,7	323,0	6863	94160	1731,8	1179,2	114,1
Івано-Франківська обл.	1,8	167,8	6074	54492	1378,9	559,2	113,7
Київська обл.	5,0	476,5	7188	87937	1743,8	3109,3	113,8
Кіровоградська обл.	1,4	141,1	5792	40427	961,4	205,4	113,8
Луганська обл.	0,7	53,2	5862	38022	2182,2	241,0	114,9
Львівська обл.	4,2	466,8	6391	112697	2532,2	1984,5	113
Миколаївська обл.	2,0	207,5	6709	50728	1146,1	700,0	113,5
Одеська обл.	4,4	637,1	6542	115025	2384,6	1293,4	114,6
Полтавська обл.	2,1	235,6	6551	69789	1420,9	1040,6	113,1
Рівненська обл.	1,1	179,3	6013	45716	1161,9	305,2	115
Сумська обл.	1,3	160,6	5946	50951	1099,8	492,9	113,6
Тернопільська обл.	0,8	116,8	5554	38727	1056,0	324,0	113,3
Харківська обл.	5,9	582,8	6244	131681	2698,1	1472,3	113,8
Херсонська обл.	1,6	208,8	5842	42707	1051,6	183,3	114,4
Хмельницька обл.	1,4	216,8	5938	55542	1280,3	378,3	113,8
Черкаська обл.	1,8	158,1	6042	51710	1226,3	373,2	114,6
Чернівецька обл.	0,8	118,9	5621	32397	907,4	102,3	112,4
Чернігівська обл.	1,3	165,6	5636	44283	1027,3	407,4	113,9
м.Київ	24,3	1430,1	11135	333927	2929,6	17696,2	113,4

НЕЛІНІЙНА РЕГРЕСІЯ

Основні теоретичні положення

Специфікація моделі

Якщо між економічними явищами існують нелінійні співвідношення, то вони виражаються за допомогою відповідних нелінійних функцій.

Розрізняють ***два класи нелінійних регресій***:

1. регресії, нелінійні щодо включених в аналіз пояснюючих змінних, але лінійні за параметрами;

2. регресії, які нелінійні за оцінюваними параметрами.

Прикладом нелінійної регресії першого класу можуть служити наступні функції:

гіперболи - $y = a + \frac{b}{x}$;

параболи другого ступеня $y = a + bx + cx^2$;

поліноми різних ступенів та інші.

До нелінійних регресій за оцінюваними параметрами відносяться функції:

ступенева — $y = a \times x^b$;

показова — $y = a \times b^x$;

експонентна — $y = e^{a+bx}$ та інші.

Нелінійна регресія за включеними змінними не має будь-яких складностей в оцінці її параметрів. Вони визначаються, як і в лінійній регресії, методом найменших квадратів (МНК), тому що ці функції лінійні за параметрами. Так, в параболі другого ступеня $y = a + bx + cx^2$,

заміняючи змінні $x = x_1$, $x^2 = x_2$, одержимо двухфакторне рівняння лінійної регресії:

$y = a + bx_1 + cx_2$, для оцінки параметрів якого використовується МНК.

Відповідно для полінома третього порядку

$y = a + bx + cx^2 + dx^3$ при заміні $x = x_1$, $x^2 = x_2$, $x^3 = x_3$, одержимо трьохфакторну модель лінійної регресії.

Отже, поліном будь-якого порядку зводиться до лінійної регресії з її методами оцінювання параметрів і перевірки гіпотез. Як показує досвід більшості дослідників, серед нелінійної поліноміальної регресії найчастіше використовується парабола другого ступеня; в окремих випадках - поліном третього порядку. Обмеження у використанні поліномів більш високих ступенів пов’язані з вимогою однорідності досліджуваної сукупності: чим вище порядок полінома, тим більше вигинів має крива і відповідно менш однорідна сукупність за результативною ознакою.

Серед класу нелінійних функцій, параметри яких без особливих ускладнень оцінюються МНК, варто назвати добре відому в економіці функцію гіперболи $y = a + \frac{b}{x}$, класичним прикладом якої є крива Філіпса, яка характеризує нелінійне співвідношення між нормою безробіття x і відсотком приросту заробітної плати y.

Серед нелінійних функцій, які можуть бути приведені до лінійного виду, в економетричних дослідженнях широко використовується ступенева функція - $y = a \times x^b$. Пов'язано це з тим, що параметр b в ній має чітке економічне тлумачення та називається ***коефіцієнтом еластичності***. Величина коефіцієнта b показує, на скільки відсотків зміниться в середньому результат, якщо фактор зміниться на 1 %. Так, якщо залежність попиту від цін характеризується рівнянням виду $\hat{y}_x = 15{,}37\ x^{-0.14}$, то зі збільшенням цін на 1 % попит знижується в середньому на 0,14 %. Про правомірність подібного тлумачення параметра b для ступеневої функції $y = a \times x^b$ можна судити, якщо розглянути формулу розрахунку коефіцієнта еластичності:

$$E = f(x)\frac{x}{y} \ , \tag{3.1}$$

де $f'(x)$ - перша похідна, яка характеризує співвідношення приросту результату і фактора для відповідної форми зв'язку.

Коефіцієнт еластичності можна визначати і при наявності інших форм зв'язку, але тільки для ступеневої функції він являє собою постійну величину, рівну параметру b. В інших функціях коефіцієнт еластичності залежить від значень фактора x.

У зв'язку з тим, що коефіцієнт еластичності не є величиною постійною, а залежить від відповідного значення x, то розраховується середній показник еластичності за формулою:

$$\overline{E} = b \times \frac{\overline{x}}{\overline{y}} \tag{3.2}$$

Оскільки коефіцієнти еластичності становлять економічний інтерес, а види моделей не обмежуються тільки ступеневою функцією, наведемо формули розрахунку коефіцієнтів еластичності для найбільш розповсюджених типів рівнянь регресії .

Таблиця 3.1 - **Коефіцієнти еластичності для математичних функцій**

Вид функції	Коефіцієнт еластичності
Лінійна $y = a + bx$	$E = \dfrac{b \cdot x}{a + b \cdot x}$
Парабола другого ступеня $y = a + bx + cx^2$	$E = \dfrac{(b + 2 \cdot c \cdot x) \cdot x}{a + bx + cx^2}$
Гіпербола $y = a + \dfrac{b}{x}$	$E = \dfrac{-b}{a \cdot x + b}$
Показова $y = a \times b^x$	$E = x \cdot \ln b$
Ступенева $y = a \times x^b$	$E = b$
Логарифмічна $y = a + b \ln x$	$E = \dfrac{c \cdot x}{\frac{1}{b} \cdot e^{cx} + 1}$

Кореляція для нелінійної регресії

Рівняння нелінійної регресії, так само як і в лінійній залежності, доповнюється показником кореляції, а саме ***індексом кореляції (R)***:

$$R = \left(1 - \frac{\sigma^2_{зал}}{\sigma^2_y}\right)^{\frac{1}{2}}, \tag{3.3}$$

де σ^2_y - загальна дисперсія результативної ознаки y;

$\sigma^2_{зал}$ - залишкова дисперсія.

Тому що $\sigma^2_y = \frac{1}{n}\sum(y - \bar{y})^2$, а $\sigma^2_{зал} = \frac{1}{n}\sum(y - \hat{y})^2$, то індекс кореляції можна виразити як

$$R = \sqrt{1 - \frac{\sum(y - \hat{y}_x)^2}{\sum(y - \bar{y})^2}} \tag{3.4}$$

Величина даного показника знаходиться в межах $0 \le R \le 1$: чим ближче до одиниці, тим тісніше зв’язок розглянутих ознак, тим більш надійно знайдене рівняння регресії. Оцінка істотності індексу кореляції проводиться так само, як і оцінка надійності коефіцієнта кореляції.

Оскільки в розрахунку індексу кореляції використовується співвідношення факторної і загальної суми квадратів відхилень, то R^2 має той же зміст, що і коефіцієнт детермінації. У спеціальних дослідженнях величину R^2

для нелінійних зв'язків називають ***індексом детермінації***.

Індекс детермінації використовується для перевірки істотності в цілому рівняння нелінійної регресії за F-критерієм Фішера:

$$F_{факт} = \frac{R^2}{1-R^2} \times \frac{n-m-1}{m}, \qquad (3.5)$$

де R^2 - індекс детермінації;

n - число спостережень;

m - число параметрів при змінних x.

Величина m характеризує число ступенів вільності для факторної суми квадратів, а $(n-m-1)$ - число ступенів вільності для залишкової суми квадратів.

Для ступеневої функції $y = a \times x^b$ $m = 1$ і формула F-критерію прийме той же вигляд, що і при лінійній залежності:

$$F_{факт} = \frac{R^2}{1-R^2} \times (n-2)$$

Для параболи другого ступеня $y = a + bx + cx^2$ формула F-критерію

$$F_{факт} = \frac{R^2}{1-R^2} \times \frac{n-3}{2}$$

Індекс детермінації R^2 можна порівнювати з коефіцієнтом детермінації r^2 для обґрунтування можливості застосування лінійної функції. Чим більше кривизна лінії регресії, тим величина коефіцієнта детермінації r^2 менше індексу детермінації R^2. Близькість цих показників означає, що немає необхідності ускладнювати форму рівняння регресії і можна використовувати лінійну функцію. Практично якщо величина $(R^2 - r^2)$ не перевищує 0,1, то припущення про лінійну форму зв'язку вважається виправданим. В іншому випадку проводиться оцінка істотності розходження R^2, обчислених за тими самими вихідним даним, через *t-критерій Ст'юдента*:

$$t = \frac{R^2_{yx} - r^2_{yx}}{m_{|R-r|}}, \qquad (3.6)$$

де $m_{|R-r|}$ - помилка різниці між R^2 і r^2, яка визначається за формулою:

$$m_{|R-r|} = 2 \cdot \sqrt{\frac{(R^2 - r^2) - (R^2 - r^2)^2 \cdot (2 - (R^2 + r^2))}{n}}. \qquad (3.7)$$

Якщо $t_{факт} > t_{табл}$, то розходження між розглянутими показниками кореляції істотне і заміна нелінійної регресії рівнянням лінійної функції неможлива. Практично, якщо величина $t < 2$, то розходження між R^2 і r^2 несуттєві і можливе застосування лінійної регресії, навіть якщо є припущення про деяку нелінійність розглянутих співвідношень ознак фактора і результату.

Запитання для самоперевірки знань

1. Назвіть, які існують два класи нелінійних регресій? Наведіть приклади нелінійних функцій, які відносяться до першого та другого класу.

2. За допомогою яких методів нелінійна регресія перетворюється до лінійного вигляду?

3. Як називається процес перетворення нелінійних регресій в лінійну форму?

4. Якою нелінійною функцією може бути замінена парабола другого ступеня, якщо не спостерігається зміна спрямованості зв'язку ознак?

5. Запишіть усі види моделей, нелінійних відносно: включених змінних; оцінюваних параметрів.

6. У чому відмінність застосування МНК до моделей, нелінійним щодо включа змінних і оцінюваних параметрів?

7. Як визначаються коефіцієнти еластичності за різними видами регресійних моделей?

8. Назвіть показники кореляції, які використовуються при нелінійних співвідношеннях розглянутих ознак.

Методичні вказівки до виконання індивідуальної роботи

Використовуючи дані залежності двох показників (табл. 3.2) побудувати рівняння:

1) параболи другого ступеня;
2) гіперболічної функції;
3) експоненціальних функцій;
4) показової функції;
5) ступеневої функції;
6) логарифмічної функції.

Таблиця 3.2 – **Вихідні дані**

№ з/п	***x***	***y***
1	1	6
2	2	9
3	3	10
4	4	12
5	5	13

Розв'язання:

1. Побудова параболи другого ступеня

Необхідно побудувати систему нормальних рівнянь виду

$$\begin{cases} \sum y = n \times a + b \times \sum x + c \times x^2 \\ \sum yx = a \times \sum x + b \times \sum x^2 + c \times \sum x^3 \\ \sum yx^2 = a \times \sum x^2 + b \times \sum x^3 + c \sum x^4 \end{cases}$$

Для чого спочатку складемо розрахункову таблицю 3.3.

Таблиця 3.3 - **Дані для розрахунку оцінок коефіцієнтів регресії**

№ з/п	***x***	***y***	***x²***	***x³***	***x⁴***	***xy***	***yx²***
1	1	6	1	1	1	6	6
2	2	9	4	8	16	18	36
3	3	10	9	27	81	30	90
4	4	12	16	64	256	48	192
5	5	13	25	125	625	65	325
Сума	15	50	55	225	979	167	649

Таким чином, отримуємо таку систему рівнянь

$$\begin{cases} 50 = 5a + 15b + 55c \\ 167 = 15a + 55b + 225c \\ 649 = 55a + 225b + 979c \end{cases}$$

Яке вирішується методом визначників де $a = \frac{\Delta a}{\Delta}; b = \frac{\Delta b}{\Delta}; c = \frac{\Delta c}{\Delta}$.

Спочатку знаходиться визначник всієї системи за допомогою правила трикутників, для чого будується наступна матриця

$$\Delta = \begin{vmatrix} 5\_15\_55 \\ 15\_55\_225 \\ 55\_225\_979 \end{vmatrix} = 5 \times 55 \times 979 + 15 \times 225 \times 55 + 55 \times 225 \times 15 - 55 \times 55 \times 55 - 225 \times 225 \times 5$$

$$-15 \times 15 \times 979 = 700$$

Далі будуються матриці та відповідно знаходяться визначники коефіцієнтів a, b та c:

$$\Delta a = \begin{vmatrix} 50\_15\_55 \\ 167\_55\_225 \\ 649\_225\_979 \end{vmatrix} = 2380$$

$$\Delta b = \begin{vmatrix} 5\_50\_55 \\ 15\_167\_225 \\ 55\_649\_979 \end{vmatrix} = 2090$$

$$\Delta c = \begin{vmatrix} 5\_15\_50 \\ 15\_55\_167 \\ 55\_225\_649 \end{vmatrix} = -150$$

$$a = \frac{2380}{700} = 3{,}4; b = \frac{2090}{700} = 2{,}986; c = \frac{-150}{700} = -0{,}214$$

Таким чином отримуємо рівняння $y = 3{,}4 + 2{,}986x - 0{,}214x^2$

2. Побудова рівняння гіперболічної функції $y = a + \frac{b}{x}$

$$b = \frac{n\sum_{i=1}^{n}\frac{y}{x} - \sum_{i=1}^{n}\frac{1}{x}\sum_{i=1}^{n}y}{n\sum_{i=1}^{n}\frac{1}{x^2} - \left(\sum_{i=1}^{n}\frac{1}{x}\right)^2}$$

$$a = \frac{1}{n}\sum_{i=1}^{n}y - \frac{1}{n}b\sum_{i=1}^{n}\frac{1}{x}$$

Складемо розрахункову таблицю 3.4.

Таблиця 3.4 **- Дані для розрахунку оцінок коефіцієнтів регресії**

№ з/п	***x***	***y***	***1/x***	***y×1/x***	***1/x²***
1	1	6	1	6	1
2	2	9	0,5	4,5	0,25
3	3	10	0,33	3,33	0,11
4	4	12	0,25	3	0,0625
5	5	13	0,2	2,6	0,04
Сума	15	50	2,28	19,43	1,4625

$$b = \frac{5\times19{,}43 - 2{,}28\times50}{5\times1{,}4625 - 2{,}28^2} = -8{,}078; a = \frac{1}{5}\times50 - \frac{1}{5}\times(-8{,}078)\times2{,}28 = 13{,}69$$

Гіперболічне рівняння має вигляд: $y = 13{,}69 - \frac{8{,}078}{x}$

3. Побудова рівняння експоненціальної функції виду $y = e^{a+bx}$

$$b = \frac{n\sum_{i=1}^{n} x \times \ln y - \sum_{i=1}^{n} x \sum_{i=1}^{n} \ln y}{n\sum_{i=1}^{n} x^2 - \left(\sum_{i=1}^{n} x\right)^2}$$

$$a = \frac{1}{n}\sum_{i=1}^{n} \ln y - \frac{1}{n} b \sum_{i=1}^{n} x$$

Складемо розрахункову таблицю 3.5.

Таблиця 3.5 **- Дані для розрахунку оцінок коефіцієнтів регресії**

№ з/п	***x***	***y***	***lny***	***x×lny***	***x²***
1	1	6	1,791759	1,791759	1
2	2	9	2,197225	4,394449	4
3	3	10	2,302585	6,907755	9
4	4	12	2,484907	9,939627	16
5	5	13	2,564949	12,82475	25
Сума	15	50	11,34143	35,85834	55

$$b = \frac{5 \times 35{,}85834 - 15 \times 11{,}34143}{5 \times 55 - 15^2} = 0{,}183406; a = \frac{1}{5} \times 11{,}34143 - \frac{1}{5} \times 0{,}183406 \times 15 = 1{,}718066$$

Тоді $y = e^{1,718066 + 0,183406x}$

4. Побудова рівняння експоненціальної функції виду $y = a \times e^{bx}$

$$b = \frac{n\sum_{i=1}^{n} x \times \ln y - \sum_{i=1}^{n} x \sum_{i=1}^{n} \ln y}{n\sum_{i=1}^{n} x^2 - \left(\sum_{i=1}^{n} x\right)^2}$$

$$\ln a = \frac{1}{n}\sum_{i=1}^{n} \ln y - \frac{1}{n} b \sum_{i=1}^{n} x$$

$$a = e^{\ln a}, e = 2{,}718281828$$

Оскільки параметр b знаходиться аналогічно, необхідно розрахувати тільки параметр а:

$$a = 2{,}718281828^{1,718066} = 5{,}573741$$

Тоді рівняння матиме вигляд $y = 5{,}573741 \times e^{0,183406x}$

5. Побудова рівняння показової функції $y = a \times b^x$

$$ln\,b=\frac{n\sum_{i=1}^{n}x\times ln\,y-\sum_{i=1}^{n}x\sum_{i=1}^{n}ln\,y}{n\sum_{i=1}^{n}x^2-\left(\sum_{i=1}^{n}x\right)^2}$$

$$ln\,a=\frac{1}{n}\sum_{i=1}^{n}ln\,y-\frac{1}{n}ln\,b\sum_{i=1}^{n}x$$

$$a=e^{ln\,a},b=e^{ln\,b}$$

Складемо розрахункову таблицю 3.6.

Таблиця 3.6 - **Дані для розрахунку оцінок коефіцієнтів регресії**

№ з/п	***x***	***y***	***lny***	***x×lny***	***x²***
1	1	6	1,791759	1,791759	1
2	2	9	2,197225	4,394449	4
3	3	10	2,302585	6,907755	9
4	4	12	2,484907	9,939627	16
5	5	13	2,564949	12,82475	25
Сума	15	50	11,34143	35,85834	55

$$ln\,b=\frac{5\times 35{,}85834-15*11{,}34143}{5\times 55-15^2}=0{,}183406;\ ln\,a=\frac{1}{5}\times 11{,}34143-\frac{1}{5}\times 0{,}183406\times 15=1{,}718066$$

$$b=e^{0{,}183406}=1{,}201302;\ a=e^{1{,}718066}=5{,}573741$$

Тоді показове рівняння має вигляд $y=5{,}573741\times 1{,}201302^{x}$

6. Побудова рівняння ступеневої функції $y=a\times x^{b},(a>0)$

$$b=\frac{n\sum_{i=1}^{n}(ln\,x\times ln\,y)-\sum_{i=1}^{n}ln\,x\sum_{i-1}^{n}ln\,y}{n\sum_{i=1}^{n}(ln\,x)^2-\left(\sum_{i=1}^{n}ln\,x\right)^2}$$

$$ln\,a=\frac{1}{n}\sum_{i=1}^{n}ln\,y-\frac{1}{n}b\sum_{i=1}^{n}ln\,x$$

$$a=e^{ln\,a},e=2{,}718281828$$

Складемо розрахункову таблицю 3.7.

Таблиця 3.7 - **Дані для розрахунку оцінок коефіцієнтів регресії**

№ з/п	*x*	*y*	*lnx*	*lny*	*lnx×lny*	*lnx²*
1	1	6	0	1,791759	0	0
2	2	9	0,693147	2,197225	1,523	0,480453
3	3	10	1,098612	2,302585	2,529648	1,206949
4	4	12	1,386294	2,484907	3,444812	1,921812
5	5	13	1,609438	2,564949	4,128127	2,59029
Сума	15	50	4,787492	11,34143	11,62559	6,199504

$$b = \frac{5 \times 11{,}62559 - 4{,}787492 \times 11{,}34143}{5 \times 6{,}199504 - 4{,}787492^2} = 0{,}474278;$$

$$ln\, a = \frac{1}{5} \times 11{,}34143 - \frac{1}{5} \times 0{,}474278 \times 4{,}787492 = 1{,}814164; a = e^{1{,}814164} = 6{,}135947$$

Тоді ступеневе рівняння має вигляд $y = 6{,}135947 x^{0{,}474278}$

7. Побудова рівняння логарифмічної функції $y = a + b\, ln\, x$

$$b = \frac{n\sum_{i=1}^{n} ln\, x \times y - \sum_{i=1}^{n} ln\, x \sum_{i=1}^{n} y}{n\sum_{i=1}^{n} ln\, x^2 - \left(\sum_{i=1}^{n} ln\, x\right)^2}$$

$$a = \frac{1}{n}\sum_{i=1}^{n} y - \frac{1}{n} b \sum_{i=1}^{n} ln\, x$$

Складемо розрахункову таблицю 3.8.

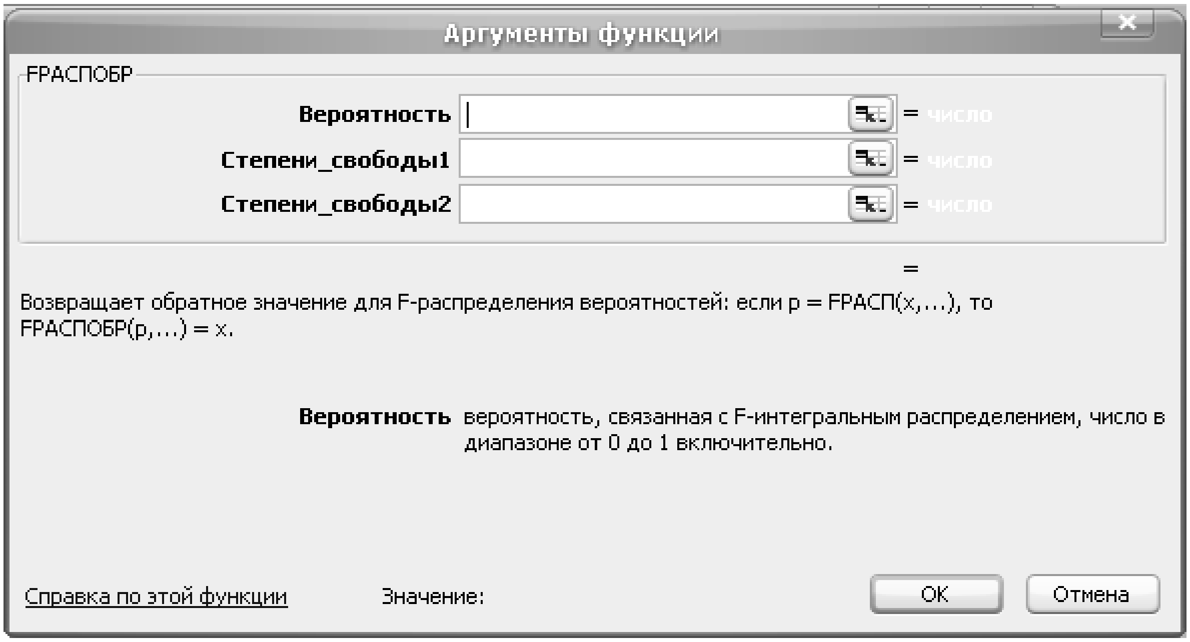
Таблиця 3.8 - **Дані для розрахунку оцінок коефіцієнтів регресії**

№ з/п	*x*	*y*	*lnx*	*lnx²*	*lnx×y*
1	1	6	0	0	0
2	2	9	0,693147	0,480453	6,238325
3	3	10	1,098612	1,206949	10,98612
4	4	12	1,386294	1,921812	16,63553
5	5	13	1,609438	2,59029	20,92269
Сума	15	50	4,787492	6,199504	54,78267

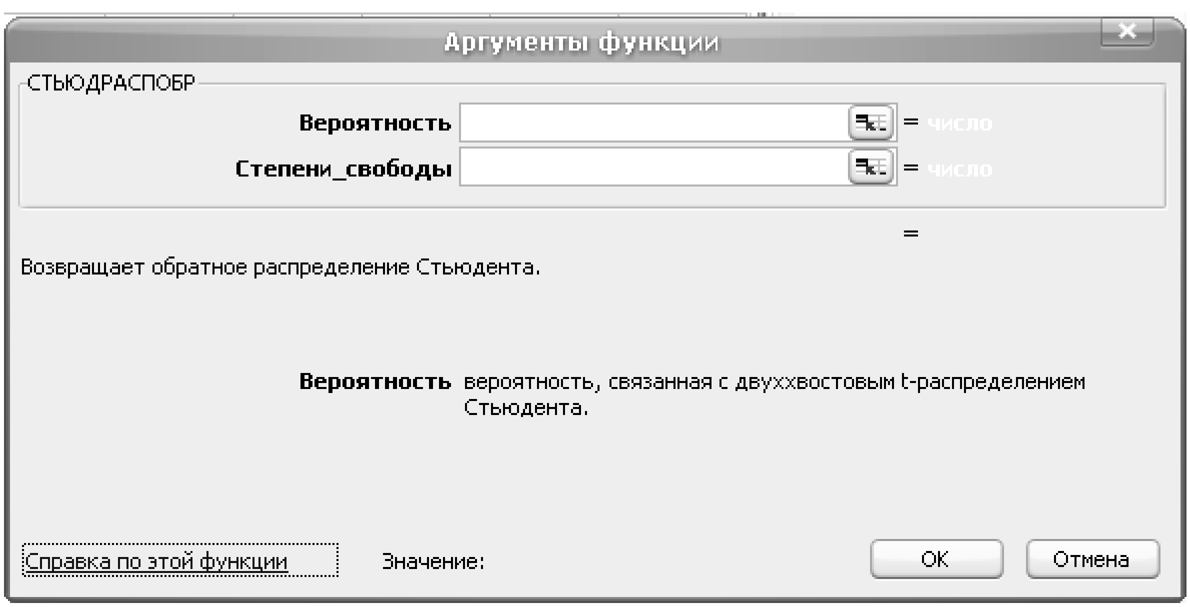
$$b = \frac{5 \times 54{,}78267 - 4{,}787492 \times 50}{5 \times 6{,}199504 - 4{,}787492^2} = 4{,}275953; a = \frac{1}{5} \times 50 - \frac{1}{5} \times 4{,}275953 \times 4{,}787492 = 5{,}905782$$

Тоді логарифмічне рівняння має вигляд $y = 5{,}905782 + 4{,}275953\, ln\, x$

Табличне значення може буде знайдене за допомогою статистичної функції табличного процесору Microsoft Excel – FРАСПОБР, алгоритм якої
FРАСПОБР(ймовірність;ступінь_вільності1;ступінь_вільності2).



Табличне значення може буде знайдене за допомогою статистичної функції табличного процесору Microsoft Excel – FРАСПОБР, алгоритм якої
СТЬЮДРАСПОБР(ймовірність; ступені_вільності)



Індивідуальні завдання для самостійного виконання

Для відповідного варіанта (таблиця 3.9) на основі даних (таблиці 3.10, 3.11) за допомогою середовища Excel необхідно:

1) побудувати нелінійні рівняння:
- параболи другого ступеня;
- гіперболічної функції;
- експоненціальних функцій;
- показової функції;
- ступеневої функції;
- логарифмічної функції;

2) розрахувати коефіцієнти кореляції та детермінації;

3) розрахувати F-критерій Фішера та зробити висновки щодо якості побудованої моделі;

4) розрахувати t-статистику Ст'юдента та оцінити значимість параметрів лінійного рівняння регресії.

Таблиця 3.9 – **Варіанти виконання індивідуального завдання**

Варіант	Номер стовбця для Y	Номера стовбців для X				
		1	2	3	4	5
1	1	x				
2	2		x			
3	3			x		
4	4				x	
5	5					x
6	6				x	
7	7			x		
8	8		x			
9	9	x				
10	10		x			
11	1			x		
12	2				x	
13	3					x
14	4	x				
15	5		x			
16	6			x		
17	7				x	

Продовження таблиці 3.9

18	8					х
19	9				х	
20	10			х		
21	1		х			
22	2	х				
23	3		х			
24	4			х		
25	5				х	

Таблиця 3.10 - **Макроекономічні показники України (Х)**

Роки	ВВП, млрд. грн.	Доходи населення, млрд. грн.	Сукупні витрати в середньому за місяць у розрахунку на одне домогоспо-дарство, грн.	Наявне населення, млн. осіб	Сукупні ресурси в середньому за місяць у розрахунку на одне домогоспо-дарство, грн.
	X1	**X2**	**X3**	**X4**	**X5**
1996	81,5	40,3	284,1	52,1	268,2
1997	93,4	50,1	356,7	51,7	297,2
1998	102,6	54,4	392,4	51,3	312,1
1999	130,4	61,9	426,5	49,8	332,0
2000	170,1	86,9	541,3	49,4	422,9
2001	204,2	157,9	607	48,9	520,8
2002	225,8	185,1	658,3	48,5	608,1
2003	267,3	215,7	736,8	48	708,6
2004	345,1	274,2	903,5	47,6	911,8
2005	441,5	381,4	1229,4	47,3	1321,4
2006	537,7	479,3	1442,8	46,9	1611,7

Таблиця 3.11 - **Наявність в домогосподарствах окремих товарів тривалого користування (Y)**

Роки	Забезпеченість населення окремими товарами тривалого користування (у середньому на 100 сімей, штук)									
	телевізори кольорові	магніто-фони	фото-апарати	холодиль-ники і морозиль-ники	пральні машини	електро-пилососи	швейні машини	мото-цикли	велосипеди і мопеди	годинники всіх видів
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2008	65	44	28	97	78	61	51	8	44	387
2009	58	43	26	96	76	58	50	8	42	338
2010	52	42	25	96	74	50	49	7	39	286
2011	51	41	23	95	73	53	48	7	41	225
2012	69	43	22	94	74	56	48	7	43	154
2013	71	43	23	93	74	55	46	6	43	84
2014	74	41	25	94	74	54	43	5	42	83
2015	79	39	27	95	75	56	40	5	40	94
2016	83	35	30	96	74	56	35	4	41	98
2017	91	34	36	99	77	62	35	4	44	101
2018	96	27	35	100	78	64	32	3	44	103

ГЕТЕРОСКЕДАСТИЧНІСТЬ

Основні теоретичні положення

Сутність гетероскедастичності та її наслідки

При практичному проведенні регресійного аналізу за допомогою МНК варто звернути серйозну увагу на проблеми, пов'язані з виконанням властивостей випадкових відхилень моделей. Властивості оцінок коефіцієнтів регресії прямо залежать від властивостей випадкової величини в рівнянні регресії. Для одержання якісних оцінок необхідно стежити за виконанням передумов МНК (умов Гауса - Маркова), тому що за їхніх порушень МНК може давати оцінки з поганими статистичними властивостями. При цьому існують інші методи визначення більше точних оцінок. Однієї із ключових передумов МНК є наступна умова: *дисперсія випадкових відхилень* ε_i *постійна.*

Виконання даної передумови називається ***гомоскедастичністю*** (сталістю дисперсії відхилень). Це можна записати як $\mathrm{var}(\varepsilon_i) = \sigma_\varepsilon^2 = const.$

Нездійсненність даної передумови називається *гетероскедастичністю* (мінливістю дисперсій відхилень). Це можна записати як $\mathrm{var}(\varepsilon_i) = \sigma_\varepsilon^2 \neq const.$

Сутність припущення гомоскедастичності полягає в тому, що σ_ε^2 не є функцією x_{ij}, тобто $\sigma_\varepsilon^2 \neq f(x_{1i}, x_{2i}, \ldots, x_{pi})$. Графічно випадок гомоскедастичності зображений на рис. 4.1.

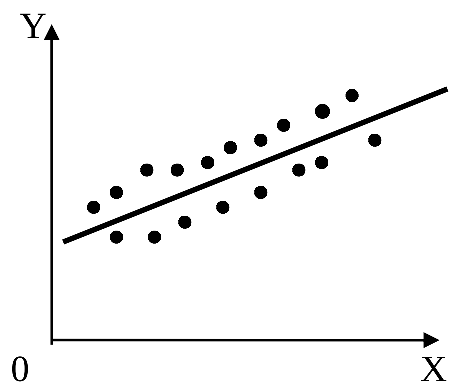


Рис 4.1. Гомоскедастичність

Якщо σ_ε^2 не є постійною, її значення залежать від значень x_{ij}, то можна записати $\sigma_\varepsilon^2 = f(x_{1i}, x_{2i}, \ldots, x_{pi})$. У цьому випадку має місце ***гетероскедастичність***. Один з випадків гетероскедастичності є випадок монотонно зростаючої дисперсії ε_i (зі зростанням x зростає й дисперсія ε_i).

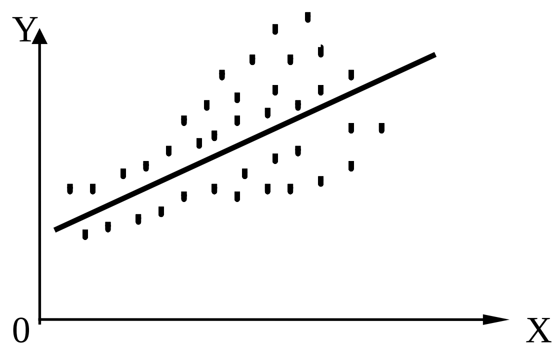


Рис 4.2. Гетероскедастичність (зростання дисперсії ε_i).

Проблема гетероскедастичності характерна для перехресних даних і досить рідко зустрічається при розгляді часових рядів. Це можна пояснити наступним чином. При перехресних даних ураховуються економічні суб'єкти (споживачі, домогосподарства, фірми, галузі, країни й ін.), що мають різні доходи, розміри, потреби й т. ін. У цьому випадку можливі проблеми, пов'язані з ефектом масштабу. У часових рядах звичайно розглядаються ті ж самі показники в різні моменти часу (наприклад, ВНП, чистий експорт, темпи інфляції й т. ін. у певному регіоні за певний період часу). Однак при збільшенні (зменшенні) розглянутих показників із часом може виникнути проблема гетероскедастичності.

При розгляді класичної лінійної регресійної моделі МНК дає найкращі лінійні незміщені оцінки (BLUE-оцінки) лише при виконанні ряду передумов, однієї з яких є сталість дисперсії відхилень (гомоскедастичність): $\operatorname{var}(\varepsilon_i) = \sigma_\varepsilon^2 = const$ для всіх спостережень *i, i = 1, 2, ..., n*.

При нездійсненності даної передумови (при гетероскедастичності) наслідки застосування МНК будуть наступними.

1. Оцінки коефіцієнтів, як і раніше, залишаться незміщеними та лінійними.
2. Оцінки не будуть ефективними, тобто вони не будуть мати найменшу дисперсію в порівнянні з іншими оцінками даного параметра. Збільшення дисперсії оцінок знижує ймовірність отримання максимально точних оцінок.
3. Дисперсії оцінок будуть розраховуватися зі зсувом. Зміщеність з'являється внаслідок того, що непояснена рівнянням регресії дисперсія $\sigma^2 = \dfrac{\sum e_i^2}{n-m-1}$ (*m* - число пояснюючих змінних), що використається при обчисленні оцінок дисперсій всіх коефіцієнтів, не є більше незміщеною.
4. Висновки, одержані на основі відповідних *t*- і *F*-статистик, а також інтервальні оцінки будуть ненадійними. Отже, статистичні висновки, одержані при стандартних перевірках якості оцінок, можуть бути помилковими та приведуть до невірних висновків за побудованою моделлю. Цілком імовірно, що стандартні помилки коефіцієнтів будуть занижені, а отже, *t*-статистики будуть

завищені. Це може привести до визнання статистично значимими коефіцієнтів, які такими насправді такими не є.

Виявлення гетероскедастичності. Методи усунення проблеми гетероскедастичності

У ряді випадків, знаючи характер даних, поява проблеми гетероскедастичністі можна передбачати та спробувати усунути цей недолік ще на етапі специфікації. Однак, значно частіше цю проблему доводиться вирішувати після побудови рівняння регресії.

Виявлення гетероскедастичності у кожному конкретному випадку є досить складним завданням, тому що для знання дисперсій відхилень $\sigma^2(e_i)$ необхідно знати розподіл випадкових величин *Y*, що відповідає обраному значенню x_i випадкових величин *X*. На практиці часто для кожного конкретного значення x_i визначається єдине значення y_i, що не дозволяє оцінити дисперсію випадкових величин *Y* для даного x_i.

Не існує якого-небудь однозначного методу визначення гетероскедастичності. Однак до теперішнього часу для такої перевірки розроблене досить велика кількість тестів і критеріїв для них.

Розглянемо найбільш популярні й наочні:

1. графічний аналіз відхилень;
2.тест рангової кореляції Спірмана;
3.тест Парка;
4.тест Глейзера,
5.тест Голдфелда-Квандта.

1. ***Графічний аналіз залишків***

Використання графічного подання відхилень дозволяє визначитися з наявністю гетероскедастичності. У цьому випадку по осі абсцис відкладаються значення (x_i) пояснюючої змінної *X* (або лінійної комбінації пояснюючих змінних $Y = a + b_iX_i + ... + b_mX_m$), а по осі ординат або відхилення e_i, або їхні квадрати e_i^2, $i= 1, 2, ..., n$. Приклади таких графіків наведені на рис. 4.3.

На рис. 4.3, *а* всі відхилення e_i^2 перебувають усередині напівсмуги постійної ширини, паралельної осі абсцис. Це говорить про незалежність дисперсій e_i^2 від значень змінної *X* та їхній сталості, тобто в цьому випадку виконуються умови гомоскедастичності.

На рис. 4.3, *б-д* спостерігаються деякі систематичні зміни в співвідношеннях між значеннями x_i змінної *X* і квадратами відхилень e_i^2. На рис. 4.3, *б* та *в* відбита лінійна, *г* – квадратична, *д* – гіперболічна залежності між квадратами відхилень і значеннями пояснюючої змінної *X*. Інакше кажучи, ситуації, представлені на рис. 4.3, *б - д,* відбивають більшу ймовірність наявності гетероскедастичності для розглянутих статистичних даних.

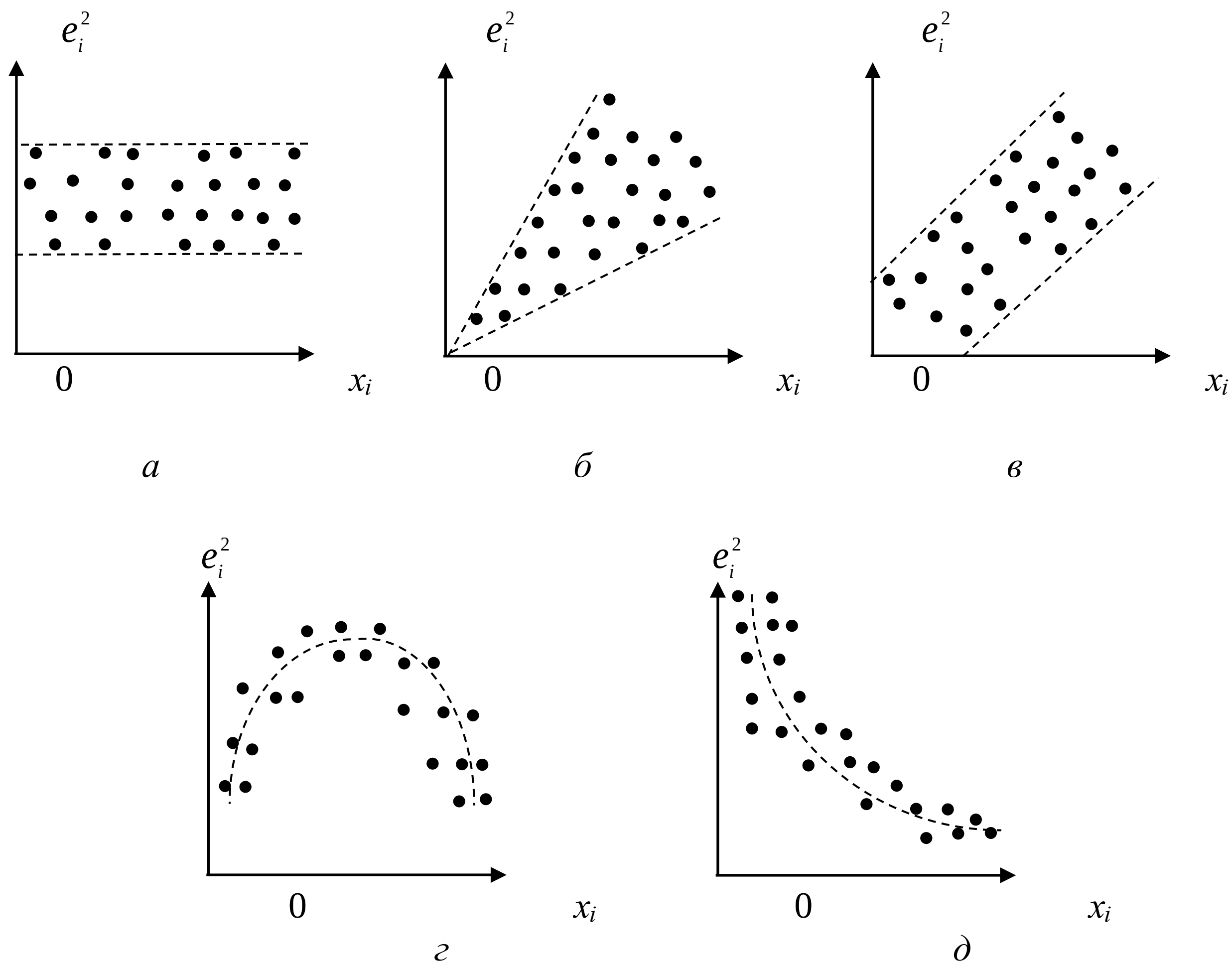


Рис. 4.3 - Графічний аналіз залишків

Графічний аналіз відхилень є зручним і досить надійним у випадку парної регресії. При множинній регресії графічний аналіз можливий для кожної з пояснюючих змінних X_j , $j = 1, 2, ..., m$, окремо. Частіше ж замість пояснюючих змінних X_j по осі абсцис відкладають значення $\hat{y}_i$, $i = 1, 2, ..., n$, одержані з емпіричного рівняння регресії. Оскільки за рівнянням множинної лінійної регресії $\hat{y}_i$ є лінійною комбінацією x_{ij}, $j = 1, 2, ... , m$, $i = 1, 2, ..., n$, то графік, що відбиває залежність e_i^2 від $\hat{y}_i$, може вказати на наявність гетероскедастичності аналогічно ситуаціям на рис. 4.3, *б-д.* Такий аналіз найбільш доцільний при великій кількості пояснюючих змінних.

2. *Тест рангової кореляції Спірмана*

При використанні даного тесту передбачається, що дисперсія відхилення буде або збільшуватися, або зменшуватися зі збільшенням значень X. Тому для регресії, побудованої за МНК, абсолютні величини відхилень e_i і значення x_i випадкових величин X будуть корельовані.

Тест проводиться в наступній послідовності:

1) Будують регресійну модель y від x і розраховують відхилення e_i.

2) Ранжуються абсолютні значення e_i і x_i у зростаючому або спадному порядку та підраховується коефіцієнт ***рангової кореляції Спірмана***:

$$r_s = 1 - 6\left[\frac{\sum_{i=1}^{n} d_i^2}{n(n^2-1)}\right], \tag{4.1}$$

де d_i – різниця між рангами e_i і x_i, $i = 1, 2, \ldots, n$, які приписуються двом характеристикам *і-го* об'єкта;

n – кількість об'єктів, що ранжуються.

3) Перевіряється значимість отриманого коефіцієнта за t-критерієм Ст'юдента із числом ступенів волі $\upsilon = (n-2)$:

$$t = \frac{r_{x,e}\sqrt{n-2}}{\sqrt{1-r_{x,e}^2}}, \tag{4.2}$$

де n – кількість спостережень.

Якщо спостережуване значення t-статистики, обчислене за формулою, перевищує $t_{кр} = t_{a/2,n-2}$ (обумовлене за таблицею критичних крапок розподілу Ст'юдента), це підтверджує гіпотезу про гетероскедастичність. У противному випадку гіпотеза про відсутність гетероскедастичності приймається.

Якщо в моделі регресії більше чим одна пояснююча змінна, то перевірка гіпотези може здійснюватися за допомогою t-статистики для кожної з них окремо.

3. *Тест Парка*

Р. Парк запропонував критерій визначення гетероскедастичності, що доповнює графічний метод деякими формальними залежностями. Передбачається, що дисперсія $\sigma_i^2 = \sigma^2(e_i)$ є функцією *і-го* значення x_i пояснюючої змінної. Парк запропонував наступну функціональну залежність:

$$\sigma_i^2 = \sigma^2 x_i^{\beta} e^{\upsilon_i}. \tag{4.3}$$

Прологарифмувавши (4.3), одержимо:

$$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln x_i + \upsilon_i. \tag{4.4}$$

Тому що дисперсії σ_i^2 звичайно невідомі, то їх заміняють оцінками квадратів відхилень e_i^2.

Критерій Парка включає наступні етапи:

1. Будується рівняння регресії $y_i = a + b_1 x_i + e_i$.
2. Для кожного спостереження визначаються $\ln e_i^2 = \ln(y_i - \hat{y}_i)^2$.
3. Будується регресія

$$\ln \sigma_i^2 = \alpha + \beta \ln x_i + \upsilon_i. \tag{4.5}$$

де $\alpha = ln\sigma^{2}$.

У випадку множинної регресії залежність (4.5) будується для кожної пояснюючої змінної.

4. Перевіряється статистична значимість коефіцієнта β рівняння (4.5) на основі t-статистики. Якщо коефіцієнт β статистично значимий, то це означає наявність зв’язку між $ln\, e_i^2$ та $ln\ x_i$, тобто гетероскедастичності в статистичних даних.

Використання в критерії Парка конкретної функціональної залежності (4.5) може привести до необґрунтованих висновків (наприклад, коефіцієнт β статистично незначимий, а гетероскедастичність має місце). Можлива ще одна проблема. Для випадкового відхилення υ_i, у свою чергу може мати місце гетероскедастичність. Тому критерій Парка доповнюється іншими тестами.

4. *Тест Глейзера*

Тест Глейзера за своєю суттю аналогічний тесту Парка й доповнює його аналізом інших (можливо, більше підходящих) залежностей між дисперсіями відхилень σ_i; і значеннями змінної x_i. За даним методом оцінюється регресійна залежність модулів відхилень $|\varepsilon_i|$ (тісно зв’язаних з σ_i^2) від x $|e_i|$. При цьому розглянута залежність моделюється наступним рівнянням регресії:

$$|e_i| = \alpha + \beta x_i^k + \upsilon_i. \tag{4.6}$$

Змінюючи значення k, можна побудувати різні регресії. Звичайно k = *..., -1, -0,5, 0,5, 1, ...* Статистична значимість коефіцієнта β у кожному конкретному випадку фактично означає наявність гетероскедастичності. Якщо для декількох регресій (7.6) коефіцієнт β виявляється статистично значимим, то при визначенні характеру залежності звичайно орієнтуються на кращу з них.

Так само, як й у тесті Парка, у тесті Глейзера для відхилень υ_i може порушуватися умова гомоскедастичності. Однак у багатьох випадках запропоновані моделі є досить гарними для визначення гетероскедастичності.

5. ***Тест Голдфелда-Квандта***

У цьому випадку передбачається, що стандартне відхилення $\sigma^2 = \sigma(\varepsilon_i)$ пропорційно значенню x_i змінної X у цьому спостереженні, тобто $\sigma_i^2 = \sigma^2 x_i^2$, $i = 1, 2, ..., n$. Передбачається, що ε_i має нормальний розподіл і відсутня автокореляція залишків.

Тест Голдфелда-Квандта полягає в наступному:

1) Ранжуються спостереження змінної X у порядку зростання або убування.

2) Задається величина C – кількість центральних спостережень за незалежними змінними X, які ми будемо виключати з подальшого аналізу. Оптимальною кількістю центральних спостережень є приблизно четверта частина всіх спостережень. Залишок $(n - C)$ спостережень ділиться на дві підвибірки однакові за розміром, одна з яких включає маленькі значення x, інша – більші.

3) Оцінюються окремі регресії для кожної підвибірки, розраховуються суми квадратів відхилень із $((n - C/2 - k)$ – ступенями вільності (k – загальна кількість оцінюваних параметрів у моделі).

Якщо припущення про пропорційність дисперсій відхилень значенням X вірно, то дисперсія регресії за першою підвибіркою (сума квадратів відхилень) буде істотно менше дисперсії регресії за другою підвибіркою.

4). Для порівняння відповідних дисперсій будується наступне F-відношення:

$$F = \frac{\sum e_2^2 \ / \left[\left\{(n-c)/2\right\} - k\right]}{\sum e_1^2 \ / \left[\left\{(n-c)/2\right\} - k\right]} = \frac{\sum e_2^2}{\sum e_1^2}, \tag{4.7}$$

де $\sum e_1^2$ – сума квадратів підвибірки з малими значеннями x;

$\sum e_2^2$ – відхилення від підвибірки з більшими значеннями x.

Якщо розраховані значення F-відношення більше табличного значення, то приймається гіпотеза про наявність гетероскедастичності, навпаки – про гомоскедастичності.

Природним є питання: якими повинні бути розміри підвибірок для прийняття обґрунтованих рішень? Для парної регресії Голдфелд і Квандт пропонують наступні пропорції: $n = 30$, $k = 11$; $n = 60$, $k = 22$.

Для множинної регресії даний тест звичайно проводиться для тієї пояснюючої змінної, котра найбільшою мірою пов'язана з σ_i. При цьому k повинне бути більше, ніж $(m + 1)$. Якщо немає впевненості щодо вибору змінної X_j, то даний тест може здійснюватися для кожної з пояснюючих змінних.

Цей же тест може бути використаний при припущенні про обернену пропорційність між σ_i і значеннями пояснюючої змінної. При цьому статистика Фішера прийме вид: $F = \dfrac{\sum e_1^2}{\sum e_2^2}$.

Гетероскедастичність приводить до неефективності оцінок, незважаючи на їх незміщеність. Це може обумовити необґрунтовані висновки про якість моделі. Тому при встановленні гетероскедастичності виникає необхідність перетворення моделі з метою усунення даного недоліку. Вид перетворення залежить від того, відомі чи ні дисперсії σ_i^2 відхилень ε_i, $i = 1, 2, ..., n$.

У ряді випадків для усунення гетероскедастичності необхідно змінити специфікацію моделі (наприклад, лінійну на лог-лінійну, мультиплікативну на адитивну і т. ін.).

На практиці має сенс застосувати кілька методів визначення гетероскедастичності та способів її коректування (перетворень, що стабілізують дисперсію).

Запитання для самоперевірки знань

1. У чому полягає сутність гетероскедастичності?
2. Яке з наступних тверджень вірно, помилково або не визначено:

а) внаслідок гетероскедастичності оцінки перестають бути ефективними й спроможними;

б) оцінки й дисперсії оцінок залишаються незміщеними;

в) виводи по t- і F-статистиках є ненадійними;

г) при наявності гетероскедастичності стандартні помилки оцінок будуть заниженими;

д) гетероскедастичність проявляється через низьке значення статистики Дарбина-Уотсона DW;

е) не існує загального тесту для аналізу гетероскедастичності;

ж) тест рангової кореляції Спірмана заснований на використанні t-статистики;

з) тест Парка є частковим випадком тесту Глейзера;

і) використання методу зважених найменших квадратів носить обмежений характер, тому що для його використання необхідно знати дисперсії відхилень;

к) якщо в парній регресії дисперсія випадкових відхилень пропорційна величині пояснюючої змінної (*x*), то для одержання ефективних оцінок необхідно всі спостережувані значення поділити на *x*?

3. Приведіть аргументи на користь графічного тесту, тесту Парка й тесту Глейзера.
4. Приведіть схему тесту Голдфелда-Квандта.
5. Є підстава вважати, що в регресії, побудованої за квартальним даними, випадкові відхилення в перші квартали більше, ніж відхилення в інших кварталах. Як це можна перевірити?

Методичні вказівки до виконання індивідуальної роботи

За даними табл. 4.1 необхідно перевірити гіпотезу про наявність або відсутність гетероскедастичності за допомогою критерію Голдфелда-Квандта.

Таблиця 4.1 – **Вихідні дані**

Район міста	Чисельність працівників, тис. чол	Надходження доходів до бюджету міста, млрд. грн.
1	115	99,1
2	125	114,2
3	95	88,3
4	149	156,1
5	112	122
6	49	48,7
7	39	37,5
8	20	15,5
9	132	150,6
10	79	90,5
11	60	68,6
12	18	20,8
13	3	4,4
14	6	8,1
15	23	28,8
16	8	12,9
17	106	132,4
18	282	342,9
19	74	104,6
20	157	209,5

Розв'язання:

Спочатку необхідно відсортувати вихідні дані за збільшенням змінної х (табл. 4.2), що можна зробити за допомогою функції «Сортировка» пункту меню «Данные».

Таблиця 4.2 – **Вихідні дані відсортовані за збільшенням змінної х**

Район міста	**Чисельність працівників, тис. чол**	**Надходження доходів до бюджету міста, млрд. грн.**
13	3	4,4
14	6	8,1
16	8	12,9
12	18	20,8
8	20	15,5
15	23	28,8
7	39	37,5
6	49	48,7
11	60	68,6
19	74	104,6
10	79	90,5
3	95	88,3
17	106	132,4
5	112	122,0
1	115	99,1
2	125	114,2
9	132	150,6
4	149	156,1
20	157	209,5
18	282	342,9
Разом	1652	1855,5

Відповідно до даної статистичної інформації була побудована лінійна регресійна модель залежності надходжень доходів до бюджету міста від чисельності працівників великих та середніх підприємств промисловості:

$$\hat{y}_x = -4{,}565 + 1{,}178x .$$

За даною моделлю були розраховані теоретичні значення $\hat{y}_x$ та відхилення від фактичних значень y, тобто e_t.

Таблиця 4.3 – **Результати розрахунків**

Район міста	x_i	y_i	$\hat{y}_x$	e_t
1	3	4,4	-1	5,4
2	6	8,1	2,5	5,6
3	8	12,9	4,9	8
4	18	20,8	16,6	4,2
5	20	15,5	19	-3,5
6	23	28,8	22,5	6,3
7	39	37,5	41,4	-3,9
8	49	48,7	53,2	-4,5
9	60	68,6	66,1	2,5
10	74	104,6	82,6	22
11	79	90,5	88,5	2
12	95	88,3	107,4	-19,1
13	106	132,4	120,4	12
14	112	122	127,4	-5,4
15	115	99,1	131	-31,9
16	125	114,2	142,7	-28,5
17	132	150,6	151	-0,4
18	149	156,1	171	-14,9
19	157	209,5	180,5	29
20	282	342,9	327,8	15,1
Разом	1652	1855,5	1855,5	0

Для застосування критерію Голдфелда-Квандта для перевірки наявності або відсутності гетероскедастичності необхідно:

1) виключити із вибірки деяку кількість центральних значень С, при цьому (n - С): 2 > *p*, де *p* – кількість параметрів, які оцінюються;

2) розподілити сукупність із (n - С) спостережень на дві групи – відповідно з меншим та більшим значенням чинника *x*;

3) визначити по кожній групі рівняння регресії;

4) визначити залишкову суму квадратів для першої (S_1) та другої (S_2) груп та розрахувати їх відношення $F = S_1 : S_2$.

При n = 20 кількість центральних значень приймаємо С = 4. Тоді в кожній групі буде по 8 спостережень ((20-4):2). Результати в табл. 4.4.

Таблиця 4.4 – **Розподіл вихідних даних на дві групи спостережень**

Район	x_i	y_i	$\hat{y}_x$	e_t	e_t^2
Перша група спостережень					
1	3	4,4	5,7	-1,3	1,69
2	6	8,1	8,5	-0,4	0,16
3	8	12,9	10,3	2,6	6,76
4	18	20,8	19,6	1,2	1,44
5	20	15,5	21,4	-5,9	34,81
6	23	28,8	24,2	4,6	21,16
7	39	37,5	38,9	-1,4	1,96
8	49	48,7	48,1	0,6	0,36
Разом					68,34
Друга група спостережень					
1	106	132,4	110,7	21,7	470,89
2	112	122	118,7	3,3	10,89
3	115	99,1	122,7	-23,6	556,96
4	125	114,2	136,1	-21,9	479,61
5	132	150,6	145,4	5,2	27,04
6	149	156,1	168,2	-12,1	146,41
7	157	209,5	178,9	30,6	936,36
8	282	342,9	346,1	-3,2	10,24
Разом					2638,4

Рівняння лінійної регресії склали: для першої групи - $\hat{y}_x = 2{,}978 + 0{,}921x$, для другої – $\hat{y}_x = 31{,}142 + 1{,}338x$.

Величина $F = S_1 : S_2 = 2638{,}4 : 68{,}34 = 19{,}3$ перевищує табличне значення F-критерію 4,41 при 5 %-му та 8,29 при 1 %-му рівні значимості для кількості ступенів вільності 6 для кожної залишкової суми квадратів ((20-4-2·2):2), тим самим підтверджується наявність гетероскедастичності.

Індивідуальні завдання для самостійного виконання

Для відповідного варіанта (таблиця 4.5) на основі даних (таблиці 4.6, 4.7) за допомогою середовища Excel необхідно перевірити гіпотезу про наявність або відсутність гетероскедастичності за допомогою критерію Голдфелда-Квандта.

Таблиця 4.5 – **Варіанти виконання індивідуального завдання**

Варіант	Номер стовбця для Y	Номера стовбців для пояснюючих змінних X						
		1	**2**	**3**	**4**	**5**	**6**	**7**
1	1	x_1						
2	1		x_1					
3	1			x_1				
4	1				x_1			
5	1					x_1		
6	1						x_1	
7	1							x_1
8	2	x_1						
9	2		x_1					
10	2			x_1				
11	2				x_1			
12	2					x_1		
13	2						x_1	
14	2							x_1
15	3	x_1						
16	3		x_1					
17	3			x_1				
18	3				x_1			
19	3					x_1		
20	3						x_1	
21	3							x_1
22	1	x_1						
23	2		x_1					
24	3			x_1				
25	1				x_1			

Таблиця 4.6 - **Роздрібний товарооборот за регіонами України за рік (Y)**

(млн. грн.)

Регіони	**Роздрібний товарооборот підприємств**	**Роздрібний товарооборот за продовольчими товарами**	**Роздрібний товарооборот за непродовольчими товарами**
	1	**2**	**3**
Вінницька	10897,8	4725,4	6172,4
Волинська	8515,1	3679,0	4836,1
Дніпропетровська	49591,7	20695,0	28896,7
Донецька	13852,8	5474,7	8378,1
Житомирська	10914,5	4843,4	6071,1
Закарпатська	10876,0	4286,5	6589,1
Запорізька	22794,9	8606,7	14188,2
Івано-Франківська	10185,4	3441,8	6744,2
Київська	37420,0	15960,5	21459,5
Кіровоградська	9283,3	4079,2	5204,1
Луганська	2813,8	1232,7	1581,0
Львівська	33154,6	13641,3	19512,7
Миколаївська	11668,2	4237,3	7430,2
Одеська	42585,8	18613,0	23972,8
Полтавська	14860,9	6277,7	8583,2
Рівненська	8359,0	3773,1	4586,4
Сумська	8649,3	4062,8	4588,0
Тернопільська	5849,4	2684,9	3164,5
Харківська	40943,1	18810,9	22132,2
Херсонська	11386,4	3742,1	7644,3
Хмельницька	10154,4	3300,1	6854,3
Черкаська	9968,2	3969,9	5998,3
Чернівецька	6396,9	2439,8	3956,7
Чернігівська	9087,7	4744,0	4343,7
м.Київ	121607,1	49692,6	71913,8

Таблиця 4.7 – **Макроекономічні показники за регіонами України за рік (X)**

Регіони	Виробництво продукції сільського господарства, млрд. грн. (1)	Торгова площа магазинів, тис.м^2 (2)	Товарні запаси, млн. грн. (3)	Доходи населення, млн.грн (4)	Чисельність населення, тис. осіб (5)	Виробництво товарів народного споживання, млрд. грн. (6)	Кількість ринків, тис.од (7)
Вінницька обл.	3,56	217	135	12084	1694,5	1,4	0,12
Волинська обл.	1,83	136	136	6852	1037,7	0,9	0,06
Дніпропетровська обл.	3,73	543	718	30136	3443,9	2,9	0,19
Донецька обл.	3,57	556	600	39566	4610,0	3,5	0,29
Житомирська обл.	2,02	185	101	9212	1330,8	0,9	0,08
Закарпатська обл.	1,38	131	106	7422	1242,6	0,5	0,11
Запорізька обл	2,34	238	304	15646	1860,2	1,9	0,1
Івано-Франківська обл.	1,65	124	118	8834	1386,2	0,5	0,1
Київська обл.	3,73	207	181	13806	1757,9	2,3	0,09
Кіровоградська обл.	2,68	131	91	7269	1060,8	0,4	0,08
Луганська обл.	1,6	238	197	17637	2404,5	0,8	0,14
Львівська обл.	2,86	356	336	19028	2558,7	1,7	0,14
Миколаївська обл.	2,06	151	127	9057	1218,9	0,9	0,09
Одеська обл.	3,28	474	468	18066	2391,2	1,7	0,21
Полтавська обл.	3,12	177	160	12453	1547,0	1,6	0,09
Рівненська обл.	1,85	122	104	7646	1155,4	0,3	0,07
Сумська обл.	2,11	139	98	8893	1224,1	1,0	0,06
Тернопільська обл.	1,89	91	69	6647	1108,8	0,6	0,07
Харківська обл.	3,73	383	446	22226	2813,4	3,9	0,13
Херсонська обл.	2,07	148	112	7321	1125,0	0,7	0,06
Хмельницька обл.	2,65	187	125	9570	1370,3	0,8	0,12
Черкаська обл.	2,87	151	121	9103	1337,8	1,2	0,06
Чернівецька обл.	1,24	78	76	5535	905,1	0,3	0,06
Чернігівська обл.	2,39	168	134	8523	1159,9	1,8	0,07
м.Київ	0,04	714	1473	41665	2651,9	4,2	0,1

АВТОКОРЕЛЯЦІЯ В ЕКОНОМЕТРИЧНИХ МОДЕЛЯХ

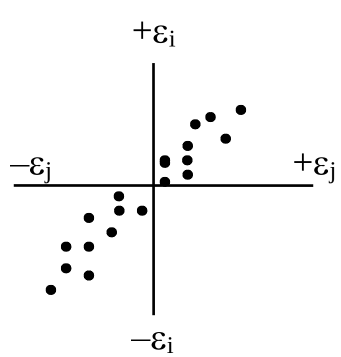
Основні теоретичні положення

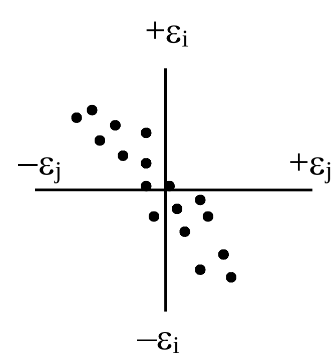
Сутність і причини виникнення автокореляції

Важливою передумовою побудови якісної регресійної моделі за МНК є незалежність значень випадкових відхилень ε_i від значень відхилень у всіх інших спостереженнях. Відсутність залежності гарантує відсутність корельованості між будь-якими відхиленнями ($\sigma(\varepsilon_i, \varepsilon_j) = \mathrm{cov}(e_i, \varepsilon_j) = 0$ при $i \neq j$) і, зокрема, між сусідніми відхиленнями ($\sigma(\varepsilon_{i-1}, \varepsilon_i) = 0$), $i = 2, 3, ..., n$.

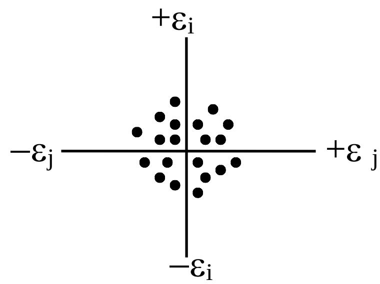
Автокореляція (послідовна кореляція) визначається як кореляція між спостережуваними показниками, упорядкованими в часі (часові ряди) або в просторі (перехресні дані). Автокореляція залишків (відхилень) звичайно зустрічається в регресійному аналізі при використанні даних часових рядів. При використанні перехресних даних наявність автокореляції (просторової кореляції) украй рідко. У силу цього в подальшому замість символу i порядкового номера спостереження будемо використати символ t, що відбиває момент спостереження. Обсяг вибірки при цьому будемо позначати символом T замість n. В економічних задачах значно частіше зустрічається так називана *позитивна автокореляція* ($\sigma(\varepsilon_{t-1}, \varepsilon_t) > 0$), ніж ***негативна автокореляція*** ($\sigma(\varepsilon_{t-1}, \varepsilon_t) < 0$).

На рис. 5.1 а) зображена наявність позитивної кореляції між випадковими величинами ε (позитивне значення ε_i супроводжується позитивним $-\varepsilon_j$ і навпаки). У більшості випадків позитивна автокореляція викликається спрямованим постійним впливом деяких неврахованих у моделі факторів. Негативна автокореляція фактично означає, що за позитивним відхиленням наступає негативне й навпаки. Можлива схема розсіювання крапок у цьому випадку представлена на рис. 8.1 б) негативне значення ε_i супроводжується позитивним $-\varepsilon_j$ і навпаки. На рис. 8.1 в) зображений класичний приклад відсутності кореляції між випадковими величинами, тобто немає систематичності в розміщенні випадкових значень ε, тому коваріація між ними дорівнює нулю.





а) *б)*



в)

Рис. 5.1.

Серед основних причин, що викликають появу автокореляції, можна виділити помилки специфікації, інерцію в зміні економічних показників, ефект павутини, згладжування даних.

Помилки специфікації. Неврахування в моделі якої-небудь важливої пояснюючої змінної або неправильний вибір форми залежності звичайно приводить до системних відхилень крапок спостережень від лінії регресії, що може обумовити автокореляцію.

Інерція. Багато економічних показників (наприклад, інфляція, безробіття, ВНП і т. ін.) мають певну циклічність, пов'язаної з хвилеподібністю ділової активності. Дійсно, економічний підйом приводить до росту зайнятості, скороченню інфляції, збільшенню ВНП і т. ін. Цей ріст триває доти, поки зміна кон'юнктури ринку й ряду економічних характеристик не приведе до зросту, потім зупинці й руху назад розглянутих показників. У кожному разі ця трансформація відбувається не миттєво, а має певну інертність.

Ефект павутини. У багатьох виробничих й інших сферах економічні показники реагують на зміну економічних умов із запізнюванням (часовим лагом). Наприклад, пропозиція сільськогосподарської продукції реагує на зміну ціни із запізнюванням (рівним періоду дозрівання врожаю). Більша ціна сільськогосподарської продукції в минулому році викличе (швидше за все) її надвиробництво цього року, а отже, ціна на неї знизиться й т. ін.

Згладжування даних. Найчастіше дані по деякому тривалому часовому періоді одержують усередненням даних за підінтервалами, які його складають. Це може привести до певного згладжування коливань, які були усередині розглянутого періоду, що у свою чергу може послужити причиною автокореляції.

Наслідки та виявлення автокореляції

Наслідки автокореляції деякою мірою подібні з наслідками гетероскедастичності. Серед них при застосуванні МНК звичайно виділяються наступні.

1. Оцінки параметрів, залишаючись лінійні і незміщеними, перестають бути ефективними. Отже, вони перестають мати властивості найкращих лінійних незміщених оцінок (BLUE-оцінок).

2. Дисперсії оцінок є зміщеними. Часто дисперсії, що обчислюють за стандартними формулами, є заниженими, що спричиняє збільшення *t*-статистик. Це може привести до визнання статистично значимими пояснюючі змінні, які в дійсності такими можуть і не бути.

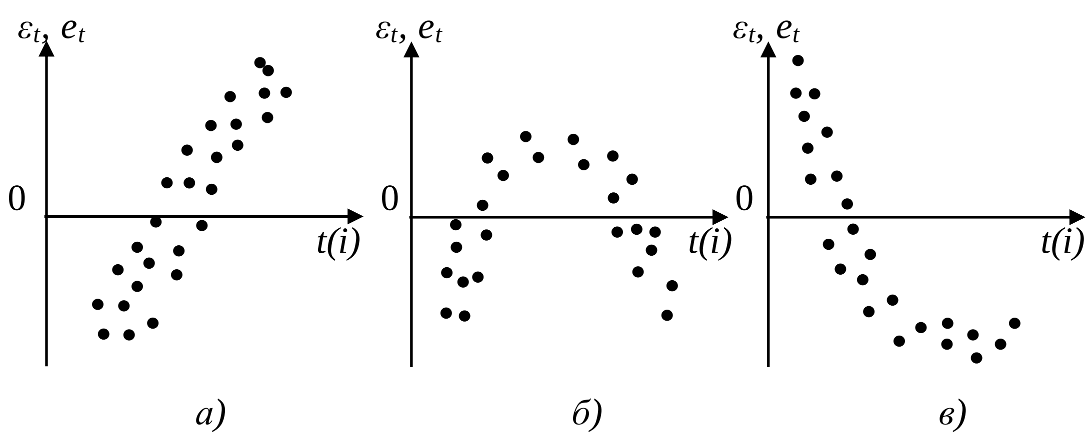
3. Оцінка дисперсії регресії $\sigma^2 = \sum_{t=1}^{T} \frac{e_t^2}{T - m - 1}$ є зміщеною оцінкою дійсного значення σ^2, у багатьох випадках занижуючи його.

4. У силу вищесказаного виводи за *t*- і *F*-статистиками, що визначають значимість коефіцієнтів регресії й коефіцієнта детермінації, можливо, будуть невірними. Внаслідок цього погіршуються прогнозні якості моделі.

В силу невідомості значень параметрів рівняння регресії невідомими будуть також і дійсні значення відхилень ε_t, $t = 1, 2, ..., T$. Тому висновки о їх незалежності здійснюються на основі оцінок e_t, $t = 1, 2, ..., T$, отриманих із емпіричного рівняння регресії. Існують декілька можливих методів визначення автокореляції.

1. Графічний метод

Існує кілька варіантів графічного визначення автокореляції. Один з них, що погоджує відхилення e_t з моментами t їхнього одержання (їхніми порядковими номерами *i)*, наведений на рис. 8.2. Це так названі послідовно-часові графіки. У цьому випадку по осі абсцис звичайно відкладаються або час (момент) одержання статистичних даних, або порядковий номер спостереження, а по осі ординат – відхилення ε_t (або оцінки відхилень e_t).



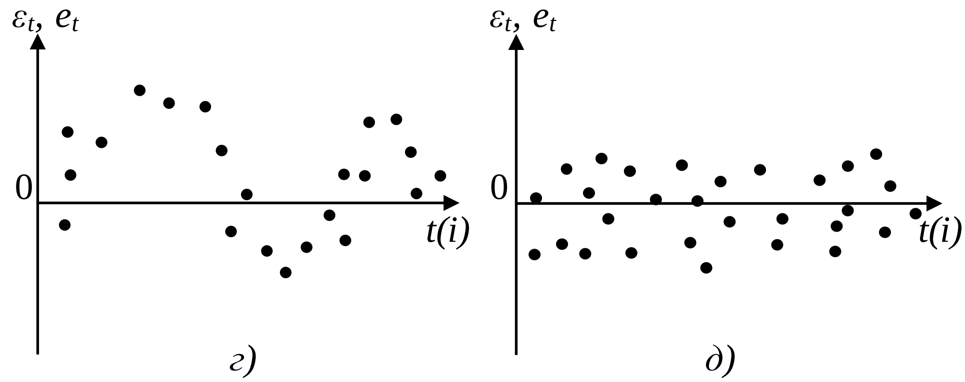


Рис. 5.2

Природно припустити, що на рис. 5.2, *а-г* є певні зв'язки між відхиленнями, тобто автокореляція має місце. Відсутність залежності на рис. 5.2, *д* швидше за все свідчить про відсутність автокореляції.

На рис. 5.2, *б* відхилення спочатку в основному негативні, потім позитивні, потім знову негативні. Це свідчить про наявність між відхиленнями певної залежності. Більше того, можна стверджувати, що в цьому випадку має місце позитивна автокореляція залишків. Вона стає досить наочною, якщо графік 5.2, *б* доповнити графіком залежності e_t від e_{t-1} (рис. 5.3).

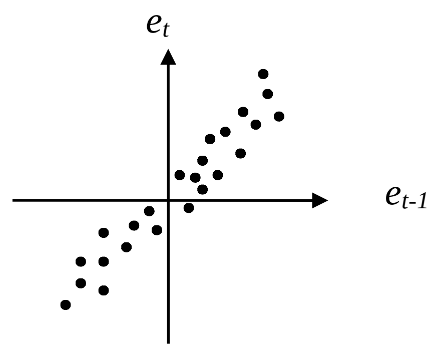


Рис. 5.3

Переважна більшість крапок на цьому графіку розташоване в *I* й *III* чвертях декартової системи координат, підтверджуючи позитивну залежність між сусідніми відхиленнями.

У сучасних комп'ютерних прикладних програмах для рішення завдань економетрії аналітичне вираження регресії доповнюється графічним поданням результатів. На графік реальних коливань залежної змінної накладається графік коливань змінної за рівнянням регресії. Зіставивши ці дві графіки, можна висунути гіпотезу про наявність автокореляції залишків. Якщо ці графіки перетинаються рідко, то можна припустити наявність позитивної автокореляції залишків.

2. Метод рядів

Цей метод досить простий: послідовно визначаються знаки відхилень e_t, $t = 1, 2, \ldots, T$. Наприклад,

(- - - - -)(+ + + + + + +)(- - -)(+ + + +)(-),

тобто 5 «-», 7 «+», 3 «-», 4 «+», 1 «-» при 20 спостереженнях.

Ряд визначається як безперервна послідовність однакових знаків. Кількість знаків у ряді називається *довжиною ряду*.

Візуальний розподіл знаків свідчить про невипадковий характер зв'язків між відхиленнями. Якщо рядів занадто мало в порівнянні з кількістю спостережень n, то цілком імовірна позитивна автокореляція. Якщо ж рядів занадто багато, то ймовірно негативну автокореляцію. Для більше детального аналізу пропонується наступна процедура. Нехай

n – обсяг вибірки;

n_1 – загальна кількість знаків «+» при n спостереженнях (кількість позитивних відхилень e_t);

n_2 – загальна кількість знаків «-» при n спостереженнях (кількість негативних відхилень e_t);

k – кількість рядів.

При досить великій кількості спостережень ($n_1 > 10$, $n_2 > 10$) і відсутності автокореляції ВВ k має асимптотично нормальний розподіл з

$$M(k) = \frac{2n_1 n_2}{n_1 + n_2} + 1;$$

$$D(k) = \frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}.$$

Тоді, якщо $M(k) - e_{a/2}D(k) < k < M(k) + e_{a/2}D(k)$, то гіпотеза про відсутність автокореляції не відхиляється.

Для невеликого числа спостережень ($n_1 < 20$, $n_2 < 20$) за таблицями критичних значень кількості рядів при n спостереженнях (додаток) на перетинанні рядка n_1 і стовпця n_2 визначаються нижнє k_1 і верхнє k_2 значення при рівні значимості $\alpha = 0.05$.

Якщо $k_1 < k < k_2$, то говорять про відсутність автокореляції.

Якщо $k \leq k_1$, то говорять про позитивну автокореляцію залишків.

Якщо $k \geq k_2$, то говорять про негативну автокореляцію залишків.

У нашому прикладі $n = 20$, $n_1 = 11$, $n_2 = 9$, $k = 5$. За таблицями (додаток) визначаємо $k_1 = 6$, $k_2 = 16$. Оскільки $k = 5 < 6 = k_1$, то приймається припущення про наявність позитивної автокореляції при рівні значимості $\alpha = 0,05$.

3. Критерій Дарбіна-Уотсона

Найбільш відомим критерієм виявлення автокореляції першого порядку є критерій Дарбіна–Уотсона. Статистика *DW* Дарбіна–Уотсона приводиться у всіх спеціальних прикладних комп'ютерних програмах як найважливіша характеристика якості регресійної моделі.

Суть методу полягає в тому, що на основі обчисленої статистики *DW* Дарбіна–Уотсона робиться висновок про автокореляцію.

$$DW = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \quad . \tag{5.1}$$

Статистика Дарбіна–Уотсона тісно пов'язана з вибірковим коефіцієнтом кореляції $r_{e_t e_{t-1}}$:

$$DW \approx \frac{2\left(\sum e_t^2 - \sum e_t e_{t-1}\right)}{\sum e_t^2} = 2(1 - r_{e_t e_{t-1}}) \, . \tag{5.2}$$

Якщо $e_t = e_{t-1}$, то $r_{e_t e_{t-1}}$ =1 й DW = 0. Якщо $e_t = -e_t$, то $r_{e_t e_{t-1}}$ = -1 й DW = 4. У всіх інших випадках 0 ≤ DW ≤ 4 і її значення можуть указати на наявність або відсутність автокореляції. Дійсно, якщо $r_{e_t e_{t-1}} \approx 0$ (автокореляція відсутня), то $DW \approx 2$. Якщо $r_{e_t e_{t-1}} \approx 1$ (позитивна автокореляція), то $DW \approx 0$. Якщо $r_{e_t e_{t-1}} \approx -1$ (негативна автокореляція), то $DW \approx 4$.

Для більш точного визначення, яке значення *DW* свідчить про відсутність автокореляції, а яке — про її наявність, була побудована таблиця критичних крапок розподілу Дарбіна—Уотсона. За неї для заданого рівня значимості α, числа спостережень *n* і кількості пояснюючих змінних *m* визначаються два значення: d_l – нижня межа й d_u – верхня межа.

Порядок тестування за критерієм Дарбіна–Уотсона:

1. За побудованим емпіричним рівнянням регресії

$$\hat{y}_t = b_0 + b_1 x_{t1} + ... + b_m x_{tm}$$

визначаються значення відхилень $e_t = y_t - \hat{y}_t$ для кожного спостереження t, $t = 1, 2, \ldots, T$.

2. За формулою (5.1) розраховується статистика DW.

3. За таблицею критичних крапок Дарбіна–Уотсона визначаються два числа d_l й d_u і здійснюють висновки за правилом:

$0 \le DW < d_l$ — існує позитивна автокореляція;

$d_u \le DW < d_u$ — висновок про наявність автокореляції невизначений;

$d_u \le DW < 4 - d_u$ — автокореляція відсутня;

$4 - d_u \le DW < 4 - d_l$ — висновок про наявність автокореляції невизначений;

$4 - d_l \le DW \le 4$ — існує негативна автокореляція.

Всі випадки зображені на рис. 5.4.

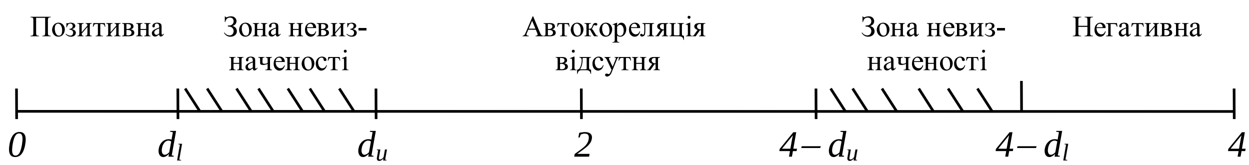


Рис. 5.4 – Шкала визначення автокореляції

При використанні критерію Дарбіна-Уотсона необхідно враховувати наступні обмеження.

1. Критерій DW застосовується лише для тих моделей, які містять вільний член.

2. Передбачається, що випадкові відхилення ε_t визначаються за ітераційною схемою: $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$, називаної авторегресійною схемою першого порядку AR(1). Тут v_t – випадковий член.

3. Статистичні дані повинні мати однакову періодичність (тобто не повинне бути пропусків у спостереженнях).

4. Критерій Дарбіна–Уотсона не застосуємо для регресійних моделей, що містять у складі пояснюючих змінних залежну змінну з часовим лагом в один період, тобто для так званих ***авторегресійних моделей*** виду:

$$y = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + \gamma y_{t-1} + \varepsilon_t . \qquad (5.3)$$

Причину четвертого обмеження пояснюється наступним. Нехай рівняння регресії має вигляд:

$$y_t = \beta_0 + \beta_1 x_t + \gamma y_{t-1} + \varepsilon_t . \qquad (5.4)$$

Нехай випадкове відхилення ε_t піддано впливу авторегресії першого порядку:

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t\,. \tag{5.5}$$

Тоді рівняння регресії (5.4) можна представити в наступному виді:

$$y_t = \beta_0 + \beta_1 x_t + \gamma\, y_{t-1} + \rho\varepsilon_{t-1} + v_t\,. \tag{5.6}$$

Але y_{t-1} залежить від ε_{t-1}, тому що якщо (5.4) вірно для t, то воно вірно й для t–1. Отже, є систематичний зв'язок між однією з пояснюючих змінних й одним з компонентів випадкового члена, тобто не виконується одна з основних передумов МНК (передумова 4^0) — пояснюючі змінні не повинні бути випадковими (не мати випадкової складової). Значення будь-якої пояснюючої змінної повинне бути екзогенним, повністю визначеним. У противному випадку оцінки будуть зміщеними навіть при більших обсягах вибірок.

Для авторегресійних моделей розроблені спеціальні тести виявлення автокореляції, зокрема h-статистика Дарбіна, що визначається за формулою

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n\,var(g)}}\,, \tag{5.7}$$

де $\hat{\rho}$ – оцінка ρ авторегресії першого порядку;

$var(g)$ – вибіркова дисперсія коефіцієнта при лаговій змінній y_{t-1},

n – число спостережень.

При великому обсязі вибірки n і справедливості нульової гіпотези H_0: $\rho = 0$ статистика h має стандартизований нормальний розподіл ($h \sim N(0, 1)$). Тому за заданим рівнем значимості α визначається критична крапка $u_{\alpha/2}$ з умови $\Phi(u_{\alpha/2}) = (1-\alpha)/2$ і рівняється h з $u_{\alpha/2}$. Якщо $h > u_{\alpha/2}$, то нульова гіпотеза про відсутність автокореляції повинна бути відхилена. У противному випадку вона не відхиляється.

Відзначимо, що звичайно значення $\hat{\rho}$ розраховується за формулою $\hat{\rho} = 1 - 0.5DW$, а var(g) дорівнює квадрату стандартної помилки S_g оцінки g коефіцієнта γ. Тому h легко обчислюється на основі даних оціненої регресії.

У такий спосіб можна записати

$$h = \left(1 - \frac{1}{2}DW\right)\sqrt{\frac{n}{1 - n\left[\mathrm{var}(g)\right]}}\,.$$

Якщо

а) h> 1.96, то є позитивна автокореляція;

б) h< –1.96, то є негативна автокореляція;

в) $-1.96 < h < 1.96$, то автокореляція відсутня.

Основна проблема при використанні цього тесту полягає в неможливості обчислення h при $nvar(\mathrm{g}) > 1$.

Методи усунення автокореляції

Основною причиною наявності випадкового члена в моделі є недосконалі знання про причини та взаємозв'язки, які визначаються, або інше значення залежної змінної. Тому властивості випадкових відхилень, у тому числі й автокореляція, у першу чергу залежать від вибору формули залежності й состава пояснюючих змінних. Тому що автокореляція найчастіше викликається неправильною специфікацією моделі, то необхідно насамперед скорегувати саму модель. Можливо, автокореляція викликана відсутністю в моделі деякої важливої пояснюючої змінної. Варто спробувати визначити даний фактор і врахувати його в рівнянні регресії. Також можна спробувати змінити формулу залежності (наприклад, лінійну на лог-лінійну, лінійну на гіперболічну й т. ін.).

Однак, якщо всі розумні процедури зміни специфікації моделі вичерпані, а автокореляція має місце, то можна припустити, що вона обумовлена якимись внутрішніми властивостями ряду $\{e_t\}$. У цьому випадку можна скористатися авторегресійним перетворенням. У лінійній регресійній моделі або в моделях, що зводяться до лінійної, найбільш доцільним і простим перетворенням є *авторегресійна схема першого порядку AR(1)*.

Для простоти викладу $AR(1)$ розглянемо модель парної лінійної регресії

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i. \tag{5.8}$$

Тоді спостереженням t й $(t-1)$ відповідають формули:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t\,. \tag{5.9}$$

$$y_{t-1} = \beta_0 + \beta_1 x_{t-1} + \varepsilon_{t-1}\,. \tag{5.10}$$

Нехай випадкові відхилення піддаються впливу авторегресії першого порядку (5.5):

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t\,.$$

де v_t, $t = 2, 3, \ldots, T$, — випадкові відхилення, що задовольняють всім передумовам МНК, а коефіцієнт ρ відомий.

Віднімемо з (5.9) співвідношення (5.10), помножене на ρ:

$$н_е - \rho н_{е-1} = \beta_0\,(1 - \rho) + \beta_1(ч_е - \rho\, ч_{е-1}) + (\varepsilon_е - \rho\varepsilon_{е-1})\,. \qquad (5.11)$$

Поклавши $y_t^* = y_t - \rho y_{t-1}$, $x_t^* = x_t - \rho x_{t-1}$, $\beta_0^* = \beta_0(1 - \rho)$, і з урахуванням (5.5) одержимо:

$$y_t^* = \beta_0^* + \beta_1 x_t^* + v_t\,. \qquad (5.12)$$

Тому що по припущенню коефіцієнт ρ відомий, то очевидно, y_t^*, x_t* , v_t обчислюються досить просто. У силу того що випадкові відхилення v_t задовольняють передумовам МНК, оцінки β_0^* й β_1 будуть мати властивості найкращих лінійних незміщених оцінок.

Однак спосіб обчислення y_t^* , x_t^* приводить до втрати першого спостереження (якщо ми не володієм попереднім йому спостереженням). Число ступенів волі зменшиться на одиницю, що при більших вибірках не так істотно, але при малих вибірках може привести до втрати ефективності. Ця проблема звичайно переборюється за допомогою *виправлення Прайса—Вінстена:*

$$x_1^* = \sqrt{1-\rho^2}\cdot x_1,\;\; y_1^* = \sqrt{1-\rho^2}\cdot y_1\,. \qquad (5.13)$$

Авторегресійне перетворення може бути узагальнене на довільне число пояснюючих змінних, тобто використано для рівняння множинної регресії.

Авторегресійне перетворення першого порядку $AR(1)$ може бути узагальнене на перетворення більш високих порядків $AR(2)$, $AR(3)$ і т. ін.:

$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + v_t,$$

$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \rho_3\varepsilon_{t-3} + v_t\,. \qquad (5.14)$$

Однак на практиці значення коефіцієнта ρ звичайно невідомо і його необхідно оцінювати. Існує кілька методів оцінювання. Розглянемо найбільш уживані.

? Запитання для самоперевірки знань

1. Що таке автокореляція?
2. Назвіть основні причини автокореляції.
3. Що може викликати негативну автокореляцію?
4. Яка передумова МНК порушується при автокореляції?
5. Які наслідки автокореляції?
6. Перелічите основні методи виявлення автокореляції.
7. Опишіть схему використання статистики *DW* Дарбіна–Уотсона.
8. Перелічите обмеження використання статистики *DW* Дарбіна– Уотсона.

9. Яка статистика використається для виявлення автокореляції в авторегресійних моделях?

10. Вірні або помилкові наступні твердження? Відповіді поясніть:

а) автокореляція характерна в основному для часових рядів;

б) при наявності автокореляції оцінки, отримані за мнк, є зміщеними;

в) статистика DW Дарбіна–Уотсона не використається в авторегресійних моделях;

г) статистика DW Дарбіна–Уотсона лежить у межах від 0 до 4;

д) для використання статистики DW статистичні дані повинні мати однакову періодичність;

е) авторегресійна схема першого порядку AR(1) усуває автокореляцію тільки у випадку, коли коефіцієнт автокореляції $\rho = 1$;

ж) при наявності автокореляції значення коефіцієнта детермінації R^2 буде завжди істотно нижче одиниці;

з) автокореляція завжди є наслідком неправильної специфікації моделі.

Методичні вказівки до виконання індивідуальної роботи

За вихідними даними попереднього завдання (табл. 4.1, 4.2) перевірити гіпотезу про наявність автокореляції в залишках.

Розв'язання:

За допомогою функції «Анализ данных. Регрессия» побудовано рівняння регресії $\hat{y}_x = -4{,}565 + 1{,}178x$.

Результати розрахунків теоретичного значення $\hat{y}$, залишки, відхилення залишків, інші проміжні розрахунки представлено в табл. 5.1.

Таблиця 5.1 – **Результати розрахунків**

t	x	y	$\hat{y}$	$e_t = y_t - \hat{y}_t$	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$	e_t^2
1	115	99,1	131	-31,9	-	-	1017,6
2	125	114,2	142,7	-28,5	3,4	11,56	812,25

Продовження таблиці 5.1

3	95	88,3	107,4	-19,1	9,4	88,36	364,81
4	149	156,1	171	-14,9	4,2	17,64	222,01
5	112	122	127,4	-5,4	9,5	90,25	29,16
6	49	48,7	53,2	-4,5	0,9	0,81	20,25
7	39	37,5	41,4	-3,9	0,6	0,36	15,21
8	20	15,5	19	-3,5	0,4	0,16	12,25
9	132	150,6	151	-0,4	3,1	9,61	0,16
10	79	90,5	88,5	2	2,4	5,76	4
11	60	68,6	66,1	2,5	0,5	0,25	6,25
12	18	20,8	16,6	4,2	1,7	2,89	17,64
13	3	4,4	-1	5,4	1,2	1,44	29,16
14	6	8,1	2,5	5,6	0,2	0,04	31,36
15	23	28,8	22,5	6,3	0,7	0,49	39,69
16	8	12,9	4,9	8	1,7	2,89	64
17	106	132,4	120,4	12	4	16	144
18	282	342,9	327,8	15,1	3,1	9,61	228,01
19	74	104,6	82,6	22	6,9	47,61	484
20	157	209,5	180,5	29	7	49	841
Разом						354,73	4382,8

Фактичне значення критерію Дарбіна-Уотсона складає:

$$d = 354{,}73 / 4382{,}82 = 0{,}081.$$

При заданому рівні значимості α = 0,05 за таблицями значень Дарбіна-Уотсона визначимо для числа спостережень n = 20 і числа незалежних перемінні моделі k = 1 критичні значення d_l = 1.2 і d_u = 1,41. Фактичне значення d = 0,081 попадає в проміжок від 0 до d_l. Отже, автокореляція в залишках присутня і вона позитивна.

Індивідуальні завдання для самостійного виконання

Для відповідного варіанта (таблиця 5.2) на основі даних (таблиці 5.3, 5.4) за допомогою середовища Excel необхідно перевірити гіпотезу про наявність автокореляції в залишках.

Таблиця 5.2 – **Варіанти виконання індивідуального завдання**

Варіант	Номер стовбця для Y	Номера стовбців для пояснюючих змінних X						
		1	**2**	**3**	**4**	**5**	**6**	**7**
1	1	x_1						
2	1		x_1					
3	1			x_1				
4	1				x_1			
5	1					x_1		
6	1						x_1	
7	1							x_1
8	2	x_1						
9	2		x_1					
10	2			x_1				
11	2				x_1			
12	2					x_1		
13	2						x_1	
14	2							x_1
15	3	x_1						
16	3		x_1					
17	3			x_1				
18	3				x_1			
19	3					x_1		
20	3						x_1	
21	3							x_1
22	1	x_1						
23	2		x_1					
24	3			x_1				
25	1				x_1			

Таблиця 5.3 - **Роздрібний товарооборот за регіонами України за рік (Y)**

(млн. грн.)

Регіони	**Роздрібний товарооборот підприємств**	**Роздрібний товарооборот за продовольчими товарами**	**Роздрібний товарооборот за непродовольчими товарами**
	1	**2**	**3**
Вінницька	10897,8	4725,4	6172,4
Волинська	8515,1	3679,0	4836,1
Дніпропетровська	49591,7	20695,0	28896,7
Донецька	13852,8	5474,7	8378,1
Житомирська	10914,5	4843,4	6071,1
Закарпатська	10876,0	4286,5	6589,1
Запорізька	22794,9	8606,7	14188,2
Івано-Франківська	10185,4	3441,8	6744,2
Київська	37420,0	15960,5	21459,5
Кіровоградська	9283,3	4079,2	5204,1
Луганська	2813,8	1232,7	1581,0
Львівська	33154,6	13641,3	19512,7
Миколаївська	11668,2	4237,3	7430,2
Одеська	42585,8	18613,0	23972,8
Полтавська	14860,9	6277,7	8583,2
Рівненська	8359,0	3773,1	4586,4
Сумська	8649,3	4062,8	4588,0
Тернопільська	5849,4	2684,9	3164,5
Харківська	40943,1	18810,9	22132,2
Херсонська	11386,4	3742,1	7644,3
Хмельницька	10154,4	3300,1	6854,3
Черкаська	9968,2	3969,9	5998,3
Чернівецька	6396,9	2439,8	3956,7
Чернігівська	9087,7	4744,0	4343,7
м.Київ	121607,1	49692,6	71913,8

Таблиця 5.4 – **Макроекономічні показники за регіонами України за рік (X)**

Регіони	**Виробництво продукції сільського господарства, млрд. грн. (1)**	**Торгова площа магазинів, тис.м² (2)**	**Товарні запаси, млн. грн. (3)**	**Доходи населення, млн.грн (4)**	**Чисельність населення, тис. осіб (5)**	**Виробництво товарів народного споживання, млрд. грн. (6)**	**Кількість ринків, тис.од (7)**
Вінницька обл.	3,56	217	135	12084	1694,5	1,4	0,12
Волинська обл.	1,83	136	136	6852	1037,7	0,9	0,06
Дніпропетровська обл.	3,73	543	718	30136	3443,9	2,9	0,19
Донецька обл.	3,57	556	600	39566	4610,0	3,5	0,29
Житомирська обл.	2,02	185	101	9212	1330,8	0,9	0,08
Закарпатська обл.	1,38	131	106	7422	1242,6	0,5	0,11
Запорізька обл	2,34	238	304	15646	1860,2	1,9	0,1
Івано-Франківська обл.	1,65	124	118	8834	1386,2	0,5	0,1
Київська обл.	3,73	207	181	13806	1757,9	2,3	0,09
Кіровоградська обл.	2,68	131	91	7269	1060,8	0,4	0,08
Луганська обл.	1,6	238	197	17637	2404,5	0,8	0,14
Львівська обл.	2,86	356	336	19028	2558,7	1,7	0,14
Миколаївська обл.	2,06	151	127	9057	1218,9	0,9	0,09
Одеська обл.	3,28	474	468	18066	2391,2	1,7	0,21
Полтавська обл.	3,12	177	160	12453	1547,0	1,6	0,09
Рівненська обл.	1,85	122	104	7646	1155,4	0,3	0,07
Сумська обл.	2,11	139	98	8893	1224,1	1,0	0,06
Тернопільська обл.	1,89	91	69	6647	1108,8	0,6	0,07
Харківська обл.	3,73	383	446	22226	2813,4	3,9	0,13
Херсонська обл.	2,07	148	112	7321	1125,0	0,7	0,06
Хмельницька обл.	2,65	187	125	9570	1370,3	0,8	0,12
Черкаська обл.	2,87	151	121	9103	1337,8	1,2	0,06
Чернівецька обл.	1,24	78	76	5535	905,1	0,3	0,06
Чернігівська обл.	2,39	168	134	8523	1159,9	1,8	0,07
м.Київ	0,04	714	1473	41665	2651,9	4,2	0,1

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. Березька К.М. Економетрика: основи теорії та комп'ютерний практикум. Тернопіль: ЗУНУ, 2022. 152 с.

2. Григорків, В.С. Моделювання економіки: підручник. Чернівці : ЧНУ ім. Ю. Федьковича, 2019. 360 с.

3. Диха М. В., Мороз В. С. Економетрія: Навчальний посібник. К.: Центр навчальної літератури (ЦУЛ), 2019. 206 с.

4. Економіко-математичне моделювання: навчальний посібник / за ред. О. Т. Іващука. – Тернопіль : ТНЕУ Економічна думка, 2008. – 704 с.

5. Єрьоменко В., Алілуйко А., Березька К., Мартинюк О. Економетрика : навчальний посібник. Тернопіль: Підручники і посібники, 2023. 168 с.

6. К.Ю. Величко, О.Д. Тімченко. Моделювання та прогнозування міжнародних економічних відносин: методичні вказівки для самостійного вивчення дисципліни для здобувачів першого (бакалаврського) рівня вищої освіти спеціальності 292 Міжнародні економічні відносини ОПП Міжнародна економіка [Електронний ресурс] – Х.: ДБТУ, 2024. – 82 с.

7. Ковальчук О. Я. Математичне моделювання та прогнозування в міжнародних відносинах: Підручник. Тернопіль: ТНЕУ, 2019. 412 с.

8. Козьменко О. В., Кузьменко О. В. Економіко-математичні методи та моделі (економетрика): Навч. посібник. Суми: Університетська книга, 2018. 406 с.

9. Лукяненко І. Г. Економетрика : підручник / І. Г. Лукяненко, Л. І. Краснікова – К. : Товариство «Знання», КОО, 1998. – 494 с.

10. Назаренко О. М. Основи економетрики : підручник вид. 2-ге, допов. та перероб. / О. М. Назаренко. – Київ : Центр навчальної літератури, 2005. – 395 с.

11. Руська Р. В. Економетрика: навчальний посібник. видання 2-е перероб. доп. Тернопіль: ЗУНУ, 2022. 224 с.

12. Тімченко О.Д. Економетрика: конспект лекцій для здобувачів першого (бакалаврського) рівня вищої освіти спеціальності 051 Економіка, 292 Міжнародні економічні відносини, 071 Облік і оподаткування, 072 Фінанси, банківська справа, страхування та фондовий ринок [Електронний ресурс] – Х.: ДБТУ, 2024. – 90 с.

ДОДАТКИ

Додаток А

Таблиця *t*-критерію *(t(p,k))*

	P							
k	***0.5***	***0.8***	***0.9***	***0.95***	***0.98***	***0.99***	***0.998***	***0.999***
1	1.00	3.078	6.314	12.706	31.821	63.657	318.3	636.619
2	0.816	1.886	2.920	4.303	6.965	9.925	22.33	31.598
3	0.765	1.638	2.353	3.182	4.541	5.841	10.210	12.941
4	0.741	1.533	2.132	2.776	3.747	4.604	1.173	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.689	1.333	1.74	2.110	2.567	2.898	3.646	3.965
18	0.688	1.33	1.734	2.101	2.552	2.878	3.610	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	0.679	1.299	1.676	2.009	2.403	2.678	3.261	3.496
100	0.677	1.290	1.660	1.984	2.364	2.626	3.174	3.390

k – число ступенів вільності;

p – довірча ймовірність.

Додаток Б

Таблиця *F*- розподілу *P*=0,95

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12	13
1	162	200	216	225	230	234	237	239	241	242	243	244	245
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.20	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26
9	5.12	4.26	3.68	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58
14	4.60	3.74	334	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45
16	4.49	3.63	3.24	30.1	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31
19	4.38	3.52	3.13	2.90	274	2.63	2.54	2.48	2.42	2.38	234	2.31	2.28
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	242	2.37	2.32	2.28	2.25	2.22
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.20
23	4.48	3.42	3.03	2.80	2.64	2.53	2.44	2.37	232	2.27	2.24	2.20	2.18
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	222	2.18	2.15
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.14
26	4.23	3.37	2.98	2.74	2.59	2.47	239	2.32	2.27	2.22	2.18	2.15	2.12
27	4.21	3.35	2.96	2.73	2.57	2.46	237	2.31	2.25	2.20	2.17	2.13	2.10
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12	1.09
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.08
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.97
50	4.03	318	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.99	1.95	1.92
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.89
70	3.98	3.13	2.74	2.50	2.35	223	2.14	2.07	2.02	1.97	1.93	1.89	1.86
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.9S	1.91	1.88	1.84
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.90	1.86	1.83
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.89	1.85	1.82
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.87	1.83	1.80
140	3.91	3.06	2.67	2.44	2.28	2.16	2.08	2.01	1.95	1.90	1.86	1.82	1.79
160	3.90	3.05	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.81	1.78
180	3.89	3.05	2.65	2.42	226	2.15	2.06	1.99	1.93	1.88	1.84	1.81	1.77
200	3.88	3.04	265	242	226	2.14	2.06	1.98	1.93	1.88	1.84	1.80	1.77

Продовження додатку Б

	k_1												
k_2	14	15	16	17	18	19	20	30	40	50	100	200	∞
1	245	246	246	247	247	248	248	250	251	252	253	254	254
2	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	8.71	8.70	8.69	8.68	8.67	8.67	8.66	8.62	8.60	8.58	855	8.54	8.53
4	5.87	5.86	5.84	5.83	5.82	5.81	5.80	5.75	5.72	5.70	5.66	5.65	5.63
5	4.64	4.62	4.60	4.59	4.58	4.67	4.56	4.50	4.46	4.44	4.41	4.39	4.37
6	3.96	3.94	3.92	3.91	3.9	3.88	3.87	3.81	3.77	3.75	3.71	3.69	3.67
7	3.53	3.51	3.49	3.48	3.47	3.46	3.44	3.38	3.34	3.32	3.27	3.25	3.23
8	3.24	3.22	3.20	3.19	3.17	3.16	3.15	3.08	3.04	3.02	2.97	2.95	2.93
9	3.03	3.01	2.99	2.97	2.96	2.95	2.94	2.86	2.83	2.80	2.76	2.73	2.71
10	2.86	285	283	2.81	2.8	2.79	2.77	2.70	2.66	2.64	2.59	2.56	2.54
11	2.74	2.72	2.70	2.68	2.67	2.66	2.65	2.57	2.53	2.51	2.46	2.43	2.4
12	2.64	2.62	2.60	2.58	2.57	2.56	2.54	2.47	2.43	2.40	2.36	2.32	2.3
13	2.55	2.53	2.51	2.50	2.48	2.47	2.46	2.38	2.34	2.31	2.26	223	2.21
14	2.48	2.46	2.44	2.43	2.41	2.40	2.36	2.31	2.27	2.24	2.19	2.16	2.13
15	2.42	2.40	2.38	2.37	235	2.34	2.33	2.25	2.20	2.18	2.12	2.10	2.07
16	2.37	2.35	2.33	2.32	2.30	2.29	2.28	2.19	2.15	2.12	2.07	2.04	2.01
17	2.33	2.31	2.29	2.27	2.26	2.24	2.23	2.15	2.10	2.08	2.02	1.99	1.96
18	2.29	2.27	2.25	223	2.22	2.20	2.19	2.11	2.10	2.08	1.98	1.95	1.92
19	2.26	2.23	2.21	2.00	218	2.17	2.16	2.07	2.03	2.00	1.94	1.91	1.88
20	2.22	2.20	2.18	2.17	2.15	2.14	2.12	2.04	1 99	1.97	1.91	1.88	1.84
21	2.20	2.18	2.16	2.14	2.12	2.11	2.10	2.01	1.96	1.94	188	1.84	1.81
22	2.17	2.15	2.13	2.11	2.10	2.08	2.07	1.98	1.94	1.91	1.85	1.82	1.78
23	2.15	2.13	2.11	2.09	1.08	2.06	2.05	1.96	1.91	1.88	1.82	1.79	1.76
24	2.13	2.11	1.09	2.07	2.05	2.04	2.03	1.94	189	1.86	1.80	1.77	1.73
25	2.11	2.09	2.07	2.05	2.04	2.02	2.01	1.92	1.87	1.84	1.78	1.50	1.71
26	2.09	2.07	2.05	2.03	2.02	2.00	1.99	1.90	1.85	1.82	1.76	1.73	1.69
27	2.08	2.06	2.04	2.02	2.00	1.99	1.97	1.88	1.84	1.81	1.74	1.71	1.67
28	2.06	2.04	2.02	2.00	1.99	1.97	1.96	1.87	1.82	1.79	1.73	1.69	1.65
29	2.05	2.03	2.01	1.99	1.97	1.96	1.94	1.85	1.81	1.77	1.71	1.57	1.64
30	2.04	2.01	1.99	1.98	1.96	1.95	1.93	1.84	1.79	1.76	1.70	1.66	1.62
40	1.95	1.92	1.90	1.89	1.87	1.85	1.84	1.74	1.69	1.66	1.63	1.60	1.51
50	1.89	1.87	1.85	183	1.81	1.80	1.78	1.69	1.63	1.60	1.59	1.55	1.44
60	1.86	1.84	1.82	1.80	1.78	1.76	1.75	1.65	1.59	1.56	1.48	1.44	1.39
70	1.84	181	1 79	1.77	1.75	1.74	1.72	1.62	1.57	1.53	1.45	1.40	1.35
80	1.82	1.79	1.77	1.75	1.73	1.72	1.70	1.60	1.54	1.51	1.43	1.38	1.32
90	1.80	1.78	1.76	1.74	1.72	1.70	1.69	1.59	1.53	1.49	1.41	1.38	1.30
100	1.79	1.77	1.75	1.73	1.71	1.69	1.68	1.57	1.52	1.48	1.39	1.34	1.28
120	1.78	1.75	1.73	1.71	1.69	1.67	1.66	1.55	1.50	1.46	1.37	1.32	1.25
140	1.76	1.74	1.72	1.70	168	166	1.65	1.54	1.48	1.44	1.35	1.30	1.23
160	1.75	1.73	1.71	1.69	1.67	1.65	1.64	1.53	1.47	1.43	1.34	128	1.21
180	1.75	1.72	1.70	1.68	1.66	1.64	1.63	1.52	1.46	1.42	1.33	1.27	1.20
200	1.74	1.72	1.69	1.67	1.66	1.64	1.62	1.52	1.46	1.41	1.32	1.26	1.19

Додаток В

Таблиця *F*- розподілу *P*=0,99

	k_1												
k_2	1	2	3	4	5	6	7	8	9	10	11	12	13
1	4052	4999	5403	5625	5764	5859	5928	5981	6023	6056	6083	6106	6126
2	98.5	99.0	99.2	99.3	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4
3	34.1	30.8	29.4	28.7	28.2	27.9	27.7	27.5	27.3	2.72	27.1	27.1	27.0
4	21.2	18.0	16.73	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.5	14.4	14.3
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.96	9.89	9.82
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.1	7.98	7.87	7.79	7.72	7.66
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.41
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67	5.61
9	10.6	8.02	6.99	6.42	6.06	5.8	5.61	5.47	5.35	526	5.18	5.11	5.05
10	10.0	7.56	6.55	5.99	5.64	5.39	5.2	5.06	4.94	4.85	4.77	471	4.65
11	9.64	7.2	6.21	5.67	5.31	5.07	4.88	4.74	4.63	4.54	4.46	4.39	4.34
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.5	4.39	4.30	4.22	4.16	4.10
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.3	4.19	4.10	4.02	3.96	3.90
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.75
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.8	3.73	3.67	3.61
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55	3.50
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46	3.40
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.50	3.51	3.43	3.37	3.32
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.24
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.18
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.24	3.17	3.12
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.07
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	3.02
24	7.82	5.64	4.72	4.22	3.90	3.7	3.50	3.36	3.26	3.17	3.09	3.03	2.98
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	3.06	2.99	2.94
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96	2.90
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.99	2.93	2.87
23	7.64	5.45	4.57	4.07	3.75	3.53	3.36	323	3.12	3.03	2.96	2.90	2.84
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.93	287	2.81
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.79
40	7.31	5.18	431	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.61
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.62	2.56	2.51
60	7.07	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.44
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59	2.51	245	2.40
80	6.96	4.88	4.04	3.56	3.25	3.04	2.87	2.74	2.64	2.55	2.48	2.42	2.36
90	6.92	4.85	4.01	3.53	3.23	3.01	284	2.72	2.61	2.52	2.45	2.39	2.33
100	6.89	4.82	3.98	3.51	321	2.99	2.82	2.69	2.59	2.50	2.43	2.37	2.31
120	6.84	4.79	3.95	3.48	317	2.96	2.79	2.66	2.56	2.47	2.40	2.34	2.28
140	6.81	4.76	3.92	3.46	3.15	2.93	2.77	2.64	2.54	2.45	2.38	2.31	2.26
160	6.79	4.74	3.90	3.44	3.13	2.92	2.75	2.62	2.52	2.43	236	2.30	224
180	6.77	4.72	3.89	3.43	3.12	2.9	2.74	2.61	2.51	2.42	2.35	2.28	2.23

Продовження додатку В

	k_1												
***k*₂**	***14***	***15***	***16***	***17***	***18***	***19***	***20***	***30***	***40***	***50***	***100***	***200***	∞
1	6143	6157	6169	6182	6192	6201	6209	6261	6287	6303	6335	6350	6366
2	89.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5	99.5
3	26.9	26.9	26.8	26.8	26.8	26.7	26.7	26.5	26.4	26.4	26.2	26.2	26.1
4	14.2	14.2	14.2	14.1	14.1	14.0	14.0	13.8	13.7	13.7	136	135	13.5
5	9.77	9.72	9.68	9.64	9.61	9.58	9.55	9.38	9.29	9.24	9.13	9.08	9.02
6	7.6	7.56	7.52	7.48	7.45	7.42	7.4	7.23	7.14	7.09	6.99	6.93	6.88
7	6.36	6.31	6.28	6.24	6.21	6.18	6 16	5.99	5.91	5.86	5.75	5.7	5.65
8	5.56	5.52	5.48	5.44	5.41	5.38	5.36	5.2	5.12	5.07	4.96	4.91	4.86
9	5.01	4.96	4.92	4.89	4.86	4.83	4.81	4.65	4.57	4.52	4.41	4.36	4.31
10	4.6	4.56	4.52	4.49	4.46	4.43	4.41	4.25	4.17	4.12	4.01	3.96	3.91
11	4.29	4.25	4.21	4.18	4.15	4.12	4.10	3.94	3.86	3.81	3.70	3.65	3.60
12	4.05	4.01	3.97	3.94	3.91	3.88	3.86	3.70	3.62	3.57	3.47	3.41	3.36
13	3.86	3.82	3.78	3.74	3.72	3.69	3.66	3.51	3.42	3.37	3.27	3.22	3.17
14	3.70	3.66	3.62	3.59	3.56	3.53	3.51	3.35	3.27	322	3.11	3.06	3.00
15	3.56	3.52	3.49	3.45	3.42	3.40	3.37	3.21	3.13	3.08	2.98	2.92	2.87
16	3.45	3.41	3.37	3.34	3.31	3.28	3.26	3.10	3.02	2.97	2.86	2.81	2.75
17	3.35	3.31	3.27	3.24	3.21	3.19	3.16	3.00	2.92	2.87	276	2.71	2.65
18	3.27	3.23	3.19	3.16	3.13	3,10	3.08	2.92	2.84	2.78	2.68	262	2.57
19	3.19	3.15	3.12	3.08	305	3.03	3.00	284	2.76	2.71	2.60	2.55	2.49
20	3.13	3.09	3.05	3.02	299	2.96	2.94	2.78	2.69	2.64	2.54	2.48	2.42
21	3.07	3.03	2.99	2.96	2.93	2.90	2.88	2.72	2.64	2.58	2.48	2.42	2.36
22	3.02	2.98	2.94	2.91	2.88	2.85	2.83	2.67	2.58	2.53	242	2.36	2.31
23	2.97	2.93	2.89	2.86	2.83	2.80	2.78	2.62	2.54	2.48	2.37	2.32	2.26
24	2.93	2.89	2.85	2.82	2.79	2.76	2.74	2.58	2.49	2.44	2.33	2.27	2.21
25	2.89	2.85	2.81	278	2.75	2.72	2.70	2.54	2.45	2.40	2.29	2.23	2.17
26	2.86	2.81	2.78	2.75	2.72	2.69	2.66	2.5	2.42	2.36	2.25	2.19	213
27	2.82	2.78	2.75	2.71	2.68	2.66	2.63	2.47	2.38	2.33	2.22	2.16	2.10
28	2.79	2.75	2.72	2.68	2.65	2.63	2.60	2.44	2.35	2.30	2.19	2.13	2.06
29	2.77	2.73	2.69	2.66	2.63	2.60	2.57	2.41	2.33	2.27	2.16	2.10	2.03
30	2.74	2.70	2.66	2.63	2.60	2.57	2.55	2.39	2.30	2.25	2.13	2.07	2.01
40	2.56	2.52	2.48	2.45	2.42	2.39	2.37	2.20	2.11	2.06	2.02	1.96	1.80
56	2.46	2.42	2.38	2.35	2.32	2.29	2.27	2.10	2.01	1.95	1.94	1.87	1.68
60	2.39	2.35	2.31	2.28	2.25	2.22	2.20	2.03	1.94	1.88	1.75	1.68	1.6
70	2.35	2.31	2.27	2.23	2.20	2.18	2.15	1.98	1.89	1.83	1.70	1.62	1.54
80	2.31	2.27	2.23	2.20	2.17	2.14	2.12	1.94	1.85	1.79	1.65	1.58	1.49
90	2.29	2.24	2.21	2.17	2.14	2.11	2.09	1.92	1.82	1.76	1.62	1.55	1.46
100	2.27	2.22	2.19	2.15	2.12	2.09	2.07	1.89	1.80	1.74	1.60	1.52	1.43
120	223	2.19	2.15	212	2.09	2.06	2.03	1.86	1.76	1.70	1.56	1.48	1.38
140	2.21	2.17	2.13	2.10	2.07	2.04	2.01	1.84	1.74	1.67	1.53	1.45	1.35
160	2.20	2.15	2.11	2.08	2.05	2.02	1.99	1.82	1 72	1.66	1 51	1.42	1.32
180	2.18	2.14	2.10	2.07	2.04	2.01	1.98	1.81	1.71	1.64	1.49	1.41	1.30
200	2.17	2.13	2.09	2.06	2.03	2.00	1.97	1.79	1.69	1.63	1.48	1.39	1.28
∞	2.08	2.04	2.00	1.97	1.93	1.90	1.88	1. 70	1. 59	1. 52	1.36	1.25	1.00

Додаток Г

Критичні значення d_N і d_L для коефіцієнта автокореляції за критерієм Дарбіна-Уотсона для P=0,95

Число спосте-режень	*Число факторів*									
	m=1		*m=2*		*m=3*		*m=4*		*m=5*	
	d_L	d_N	d_L	d_N	d_L	d_N	d_L	d_N	d_L	d_N
6	0.61	1.40								
7	0.70	1.36	0.47	1.90						
8	0.76	1.33	0.56	1.78	0.37	2.29				
9	0.82	1 32	0.63	1.70	0.46	2.13	0.30	2.59		
10	0.88	132	0.70	1.64	0.53	2.02	0.38	2,41	0.24	2.82
11	0.93	1.32	0.76	1.60	0.60	1.93	0.44	2.28	0.32	2.65
12	0.97	1,33	0.81	1.58	0.66	1.86	0.51	2.18	0.38	2.51
13	1.01	1.34	0.86	1.56	0.72	1.82	0.57	2.09	0.45	2.39
14	1.05	1.35	0.91	1.55	0.77	1.78	0.63	2.03	0.51	2.30
I5	1.08	1.36	0.95	1.54	0.81	1 75	069	1.98	0.56	2.22
16	1.11	1.37	0.98	1.54	0.86	1.73	0.73	1.94	0.62	2.16
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.66	2.10
18	1.16	1.39	1.05	1.54	0.93	1.70	0.82	1.87	0.71	2.06
19	1.18	1.40	1.07	1.54	0.97	1.69	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.89	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1 96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.06	1.86	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.65	1.04	1,77	0.95	1.89
26	1.30	1.46	1.23	1.55	1.14	1.65	1.06	1.76	0.98	1.87
37	1.32	1.47	1.24	1.56	1.16	1.65	108	1.75	1.00	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1,36	1,50	1,30	1.57	1.23	1.65	1.16	1,74	0.09	1.83
32	1,37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1,11	1.82
33	1,38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1,13	1,81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.14	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.53	1.35	1.59	1.30	1.65	1.24	1.72	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1 54	1.37	1.59	1.32	1.66	1.26	1.72	1.20	1.79
39	1.44	1.54	1.38	1.60	1 33	1.66	1.37	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1 48	1.57	1.43	1.62	1 38	1,87	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.37	1.77

Навчальне видання

ЕКОНОМЕТРИКА

Методичні вказівки та завдання
до виконання самостійної та індивідуальної роботи
для здобувачів першого (бакалаврського) рівня
вищої освіти спеціальності 051 Економіка
292 Міжнародні економічні відносини
071 Облік і оподаткування
072 Фінанси, банківська справа, страхування та фондовий ринок

ТІМЧЕНКО Ольга Дмитрівна

Формат 60х84/16. Гарнітура Times New Roman
Папір для цифрового друку. Друк ризографічний.
Ум. друк. арк._.
Наклад___пр.
ДБТУ
61002, м. Харків, вул. Алчевських, 44