

Міністерство освіти і науки, молоді та спорту України
Тернопільський національний економічний університет

Єрмоєнко В. О., Алілуйко А. М., Мартинюк О. М., Попіна С. Ю.

ЕКОНОМЕТРІЯ (ЕКОНОМЕТРИКА)

Навчальний посібник



Тернопіль
Видавництво «Підручники і посібники»
2012

УДК 330.43
Е 45

Рецензенти: *Ляшенко Оксана Миколаївна* — доктор економічних наук, доцент, завідувач кафедри економічної кібернетики та інформатики Тернопільського національного економічного університету;

Кривень Василь Андрійович — доктор фізико-математичних наук, професор, завідувач кафедри математичних методів в інженерії Тернопільського національного технічного університету ім. І. Пулюя.

*Розглянуто і рекомендовано до видання Вченою радою
Україно-нідерландського факультету економіки та менеджменту,
протокол № 3 від 18.11.2011 р.*

Єрмоменко В. О., Алілуйко А. М., Мартинюк О. М., Попіна С. Ю.

Е 45 Економетрія (економетрика). Навчальний посібник для студентів заочної форми навчання економічних спеціальностей. / Єрмоменко В. О., Алілуйко А. М., Мартинюк О. М., Попіна С. Ю. — Тернопіль: Підручники і посібники, 2012. — 116 с.

ISBN 978-966-07-2117-6

Посібник містить теоретичний матеріал з основних розділів економетрії (економетрики), а також приклади розв'язання типових задач.

Для студентів заочної форми навчання всіх економічних спеціальностей ВНЗ.

**УДК 330.43
Е 45**

ISBN 978-966-07-2117-6

© Єрмоменко В. О., Алілуйко А. М.,
Мартинюк О. М., Попіна С. Ю., 2012

Вступ

Попередні чотири десятиліття років характерні, зокрема, бурхливим розвитком наукової дисципліни, яка має назву економетрія або економетрика. Всесвітнє визнання робіт, виконаних з допомогою економетричних методів, відобразалося великим числом лауреатів Нобелівських премій по економіці.

Останні досягнення економічної науки висувають нові вимоги до вищої професійної освіти. Сучасна економічна освіта, — стверджує директор ЦЕМІ РАН академік В.Л. Макаров, — тримається на трьох китах: макроекономіці, мікроекономіці та економетриці [6].

Відсутність усталеної назви дисципліни економетрія (економетрика) характерна і для формулювання єдиного загальноприйнятого визначення її предмета. Термін «економетрія» означає вимірювання в економіці. Однак не всі вимірювання в економіці належать до економетрії. Ми будемо дотримуватися наступного означення.

Економетрія досліджує методи оцінювання кількісних взаємозв'язків між економічними показниками.

У зв'язку з тим, що основою математичного інструментарію економетрії є теорія імовірностей і математична статистика, в першому параграфі наведені основні закони розподілу випадкових величин, з якими доводиться постійно працювати (нормальний, Пірсона, Ст'юдента, Фішера-Снедекора). У викладі також здійснюється при необхідності посилання на потрібні означення та факти.

Основною вимогою до економетричних досліджень є коректність отриманих результатів, яка досягається за рахунок перевірки виконання всіх використаних гіпотез (передумов). У пропонованому посібнику така перевірка здійснюється (§3), починаючи із випадку класичної нормальної лінійної моделі парної регресії. Саме такий підхід на думку авторів дозволить студентам розібратися із важливими моментами досліджень найпростіших моделей з тим, щоб уникати помилкових висновків при курсовому, дипломному чи дисертаційному проектуванні.

§1. ПОНЯТТЯ ПРО ЕКОНОМЕТРИЧНІ МОДЕЛІ.

1. *Кореляційно-регресійний аналіз в економіці.*
2. *Економетрична модель та її елементи.*
3. *Інформаційна база економетричних досліджень.*
4. *Нормальний розподіл і основні вибіркові розподіли, пов'язані з ним.*
5. *Кореляція. Вибірковий коефіцієнт кореляції.*

1. У багатьох практично важливих задачах потрібно встановити й оцінити залежність деякого економічного показника від одного або кількох інших показників. Проте будь-які економічні показники, як правило, перебувають під впливом випадкових чинників, а тому з математичної точки зору їх інтерпретують як випадкові величини.

З теорії ймовірностей відомо, що випадкові величини можуть бути або незалежними, або пов'язані функціонально, або ж між ними існує стохастична залежність. Строга функціональна залежність в економіці реалізується рідко. Переважно спостерігається так звана стохастична залежність, коли зміна можливих значень однієї випадкової величини (внаслідок проведення випробувань) призводить до зміни **умовного закону розподілу ймовірностей** іншої [3, §7]. Частковим випадком стохастичної залежності є кореляційна залежність, коли зміна можливих значень однієї величини приводить до зміни **умовного математичного сподівання** [3, п.7.5] іншої випадкової величини.

Нагадаємо про два типи кореляційної залежності двох випадкових величин. У першому випадку вони рівноправні, тому зв'язок між ними двосторонній. У другому випадку змінні нерівноправні, тобто є економічний сенс розглядати лише односторонній зв'язок, коли зміна тільки однієї із них приводить до зміни умовного математичного сподівання іншої. Наприклад, середня урожайність залежить від кількості внесених добрив, а не навпаки.

Нехай з певних економічних міркувань встановлено, що деякий економічний показник x впливає на інший показник y . Статистичні дані по кожному із них інтерпретуються як деякі реалізації (можливі значення) випадкових величин X та Y . Тоді кореляційну залежність між ними або **залежність у середньому** можна зобразити у вигляді співвідношення

$$M(Y | x) = f(x), \quad (1.1)$$

де $M(Y | x) = M(Y | X = x)$ — умовне математичне сподівання випадкової величини Y при умові, що X набуває можливого значення x . Функція $f(x)$ називається **функцією регресії Y на X** , а її графік — **лінією регресії**. При цьому X називається **незалежною змінною**, а Y — **залежною**. Розглядаючи залежність двох випадкових величин, говорять про **парну регресію**.

Залежність Y від кількох змінних, що описується рівнянням

$$M(Y | x_1, x_2, \dots, x_m) = F(x_1, x_2, \dots, x_m), \quad (1.2)$$

називають **множинною регресією**.

Оскільки реальні (спостережені) значення залежної змінної не завжди співпадають із умовним математичним сподіванням, то ліві частини рівнянь (1.1) і (1.2) можна записати так:

$$M(Y | x) = Y - U, \quad M(Y | x_1, x_2, \dots, x_m) = Y - U,$$

де U — випадкові величини. Із урахуванням цих співвідношень рівняння (1.1) і (1.2) набудуть такого виду:

$$Y = f(x) + U, \quad (1.3)$$

$$Y = F(x_1, x_2, \dots, x_m) + U. \quad (1.4)$$

Означення. Зв'язки між залежною та незалежною (незалежними) змінними, що описуються співвідношеннями (1.3) або (1.4) називаються **економетричними (регресійними) моделями (рівняннями)**. При цьому модель (1.3) називається **однофакторною**, а (1.4) — **багатофакторною (m-факторною)**.

Виникає питання про причини обов'язкової присутності в регресійних моделях випадкової складової U , яку називають **відхиленням, залишком або збуренням**. Серед них виокремимо найістотніші.

- 1) **Уведення в модель не всіх пояснюючих змінних.** Будь-яка регресійна модель — це спрощення реальної ситуації. Остання завжди є складною композицією різних чинників, багато з яких у моделі не враховуються, що призводить до відхилення реальних значень залежної змінної від її модельних (розрахункових) значень. При цьому у деяких випадках заздалегідь невідомо, які чинники за умов, що склалися, насправді є визначальними, а якими можна знехтувати.
- 2) **Неправильний вибір функціональної форми моделі.** Визначальною особливістю моделей (1.3) та (1.4) є те, що **функції $f(x)$ та $F(x_1, x_2, \dots, x_m)$ є невідомими**, а тому вони шукаються у певному

класі функцій і немає ніяких гарантій, що обраний клас функцій є кращим за інші.

- 3) **Помилки вимірювання.** Якою б якісною не була модель, помилки вимірювання змінних впливатимуть на розбіжності між модельними та емпіричними даними.
- 4) **Обмеженість статистичних даних та їх випадковий характер.**
- 5) **Непередбачуваність людського фактора.**

2. Вище було показано, яким чином рівняння регресії приводять до економетричної моделі.

Означення. Економетрична модель — це функція (скалярна або векторна), яка описує кореляційно-регресійний зв'язок між економічними показниками, причому залежно від причинних зв'язків між ними один або кілька із цих показників розглядаються як залежні змінні, а інші — як незалежні.

У загальному випадку економетрична модель має такий вигляд:

$$Y = f(x_1, x_2, \dots, x_m, a_0, a_1, \dots, a_k, U), \quad (1.5)$$

де $Y = (Y_1, Y_2, \dots, Y_s)'$, $f = (f_1, f_2, \dots, f_s)'$, $(.)'$ — операція транспонування вектора, Y_1, Y_2, \dots, Y_s — залежні (пояснювані, регресанти, ендогенні) змінні, x_1, x_2, \dots, x_m — незалежні (пояснюючі, регресори, екзогенні) змінні, a_0, a_1, \dots, a_k — параметри (невідомі) моделі, U — l -вимірний випадковий вектор, яка називається відхиленням (залишком або збуренням).

Відмітимо, що у серйозних емпіричних економічних і соціальних дослідженнях вимірність s вектора Y часто буває дуже великою. Наприклад, економетричні моделі, які в теперішній час використовуються в Німеччині інститутами економічних досліджень і Федеральним банком для прогнозування кон'юнктури, містять, як правило, більше 100 рівнянь [1].

Означення. Дослідження опису явища чи процесу, тобто вибір конкретної функції f , називається **специфікацією моделі**.

Означення. Знаходження значень параметрів a_0, a_1, \dots, a_k обраної форми статистичного зв'язку змінних на підставі статистичних даних називається **параметризацією економетричної моделі або оцінюванням параметрів**.

Залежно від методу, з допомогою якого здійснюється оцінювання невідомих параметрів моделі, додатково накладаються певні обмеження як на параметри, так і на випадкову складову. Такого роду обмеження також

належать до специфікації моделі. Припускається, що параметри (числа) залишаються незмінними протягом усього періоду спостереження. Зміна незалежних змінних приводить модель у рух, зумовлює перехід системи до нового стану.

Зауваження. В багатьох економетричних моделях є такі незалежні змінні, які можуть бути змінені керівними органами (державним регулюванням чи керівництвом фірмою). Ці керовані змінні можуть впливати на подальший розвиток процесу.

Побудова якісної економетричної моделі, яка узгоджується з емпіричними даними і відповідає цілям дослідника, є досить складним процесом, для якого можна виділити наступні характерні етапи [9].

1. Початковий аналіз економічного процесу, що розглядається (його теоретичний опис з відображенням існуючих тенденцій і якісний економічний аналіз).
2. Визначення мети дослідження, досягнення якої вимагає залучення моделі; введення припущень та обмежень.
3. Вибір факторів, істотних з погляду мети дослідження (специфікація змінних).
4. Специфікація моделі.
5. Збір і попередній аналіз даних для змінних, що входять у модель.
6. Вибір методу оцінювання невідомих параметрів з урахуванням припущень про імовірнісні властивості випадкового відхилення.
7. Реалізація процедури оцінювання, яка забезпечує найкраще наближення модельних значень змінних до їхніх значень, що спостерігалися.
8. Перевірка виконання основних припущень.
9. Інтерпретація отриманих результатів, визначення їхньої адекватності поставленим цілям.
10. Аналіз неузгодженості і корегування моделі.
11. Прийняття рішень щодо наступного циклу дослідження з урахуванням альтернативних можливостей.

Основна мета дослідження економетричної моделі (1.5) — **теоретично обґрунтований і статистично надійний точковий чи інтервальний прогноз значень залежної змінної Y або її математичного сподівання (середньої).**

Економетричні моделі можуть бути статичними та динамічними. У статичних моделях зв'язки розглядаються у фіксований момент часу і ча-

сові зміни в них ролі не відіграють. У динамічній моделі час є необхідним фактором змін.

Моделі розрізняються також за рівнем агрегування змінних (мікро, мезо- чи макроекономічні показники), за способом відображення змінних (у поточних чи постійних цінах, у абсолютних значеннях чи приростах показників), за кількістю змінних (одно- чи багатофакторні моделі), за часом спостереження (річні, квартальні чи місячні дані).

Класифікують моделі також за призначенням та метою використання (аналітичні, імітаційні, прогностичні).

3. Будь-яке економетричне дослідження поєднує теорію (якісний економічний аналіз, на основі якого будуються математичні моделі) і практику (статистичні дані). За допомогою моделей принципово описують і пояснюють процеси, що вивчаються, а статистичні дані використовуються для параметризації та обґрунтування моделей.

Економічні дані звичайно поділяються на три види: **просторові, часові** (динамічні) **ряди** та **перехресні** дані.

Просторовими є дані за деякими економічними показниками, які отримані для різних однотипних об'єктів (фірм, регіонів) або в один і той самий період часу, або ж часова приналежність несуттєва.

Часові ряди характеризують один і той самий об'єкт у різні рівновіддалені моменти часу. Послідовні значення часових рядів можуть бути пов'язані між собою певними залежностями: спостерігаються деякі закономірності у відхиленнях від загальної тенденції розвитку (тренду) чи виявляються часові зсуви показників (часові лаги). У зв'язку з цим методи обробки таких даних дещо відрізняються від методів, що використовуються для опрацювання просторових даних.

Перехресні ряди є поєднанням просторових та часових рядів.

Коректність висновків, які можна зробити в результаті економетричного моделювання, зумовлюється якістю вхідних даних, тобто їх повнотою та достовірністю. Тому, формуючи сукупність спостережень, слід забезпечити порівнянність даних у просторі і часі. Це означає, що дані вхідної сукупності повинні мати:

- однаковий ступінь агрегування;
- однорідну структуру одиниць сукупності;
- одні і ті ж самі методи розрахунку показників у часі чи просторі;
- однакову періодичність обліку окремих змінних;
- порівнянні ціни та однакові інші зовнішні економічні умови.

Особливої уваги заслуговує обсяг вибіркової сукупності. Результати багатьох досліджень підтверджують [12], що число спостережень повинно у 6-7 разів перевищувати число параметрів моделі. Це означає, що досліджувати парну економетричну модель з двома параметрами при наявності менше семи спостережень взагалі не має сенсу. Враховуючи, що економетричні моделі часто будуються за даними рядів динаміки, обмеженими по тривалості (10, 20, 30 років), то при виборі специфікації моделі перевагу слід віддати моделі з меншим числом параметрів.

4. Економетрія використовує, зокрема, основні поняття і методи теорії імовірностей та математичної статистики. Із усього обсягу теоретичного матеріалу цієї дисципліни виокремимо основні закони розподілу випадкових величин, якими доведеться постійно користуватися.

Неперервна випадкова величина X розподілена за **нормальним законом**, якщо її густина розподілу імовірностей має такий вигляд:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}. \quad (1.6)$$

Імовірнісний зміст параметрів a і σ визначається рівностями: $a = M(X)$, $\sigma^2 = D(X)$. Надалі будемо використовувати компактний запис (1.6): $X \sim N(a, \sigma)$.

Нормальний закон розподілу з параметрами $a = 0$, $\sigma = 1$, тобто $N(0,1)$, називається **стандартним** або **нормованим**. Використовуючи властивості математичного сподівання і дисперсії, можна переконатись у тому, що коли $X \sim N(a, \sigma)$, то $U = \frac{X - a}{\sigma} \sim N(0,1)$, тобто випадкова величина U розподілена за стандартним нормальним законом.

Нормальний закон розподілу є найбільш вивченим, тому його намагаються використовувати і при дослідженні випадкових величин, розподіл яких відмінний від нормального. Один із шляхів — розподіл досліджуваної випадкової величини замінюють приблизно нормальним. Наприклад, для **достатньо великого обсягу вибірки** середня вибірка і вибірка частка для нефіксованої вибірки є нормально розподіленими із достатнім ступенем точності [4, с. 52, 58].

Інший підхід у припущенні нормального розподілу результатів спостережень полягає у побудові статистик, які мають відомі закони розподілу.

χ^2 -розподіл (розподіл Пірсона). Нехай Z_1, Z_2, \dots, Z_k — незалежні стандартні нормальні випадкові величини ($Z_i \sim N(0,1), i = \overline{1, k}$). Тоді випадкова величина

$$\chi^2 = \sum_{i=1}^k Z_i^2$$

розподілена за законом χ^2 (Пірсона) із k ступенями вільності. χ^2 -розподіл не містить невідомих параметрів і залежить тільки від k . При цьому $M(\chi^2) = k, D(\chi^2) = 2k$.

Виявляється, що деякі статистики вибірки мають χ^2 -розподіл. Наприклад, якщо σ^2 — дисперсія нормального розподілу і S^2 — її незміщена оцінка із $k = n - l$ ступенями вільності (тут n — обсяг вибірки, l — число зв'язків у виразі для S^2), то статистика

$$\chi^2 = \frac{kS^2}{\sigma^2} \quad (1.7)$$

має χ^2 -розподіл із k ступенями вільності [4, с.81].

Відмітимо, що поняття «число зв'язків» буде уточнене в §2 (п.4). Зокрема, якщо $X_i \sim N(a, \sigma)$ — спостережені значення ($i = \overline{1, n}$), то вибіркова

(виправлена) дисперсія $S^2 = \frac{n}{n-1} D_e$ має $k = n - 1$ ступенів вільності.

Відповідно статистика (1.7) має χ^2 -розподіл із числом ступенів вільності $k = n - 1$.

Для $k > 30$ випадкова величина $V = \sqrt{2\chi^2} - \sqrt{2k-1}$ приблизно розподілена за стандартним нормальним законом, тому при $k > 30$ значення χ_p^2 , які задовольняють рівнянню

$$P(\chi^2 \leq \chi_p^2) = p \quad (1.8)$$

і називаються **квантилем рівня p** (або **p -квантилем**), можна обчислювати за наближеною формулою

$$\chi_p^2 = \frac{1}{2} (\sqrt{2k-1} + u_p)^2, \quad (1.9)$$

де u_p — корінь рівняння $\Phi(u_p) + 0,5 = p$, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ — функція Лапласа.

Розподіл Ст'юдента (t -розподіл). Нехай Z — випадкова величина, розподілена за стандартним нормальним законом, а χ^2 — незалежна від Z випадкова величина, яка має χ^2 -розподіл із k ступенями вільності. Тоді розподіл випадкової величини

$$t = \frac{Z}{\sqrt{\frac{1}{k}\chi^2}} = \frac{Z}{\chi}\sqrt{k} \quad (1.10)$$

називається **розподілом Ст'юдента** або **t -розподілом** із k ступенями вільності.

При $k \rightarrow \infty$ t -розподіл наближається до стандартного нормального розподілу. При цьому практично вже при $k > 30$ t -розподіл можна вважати приблизно стандартним нормальним.

Побудуємо конкретні статистики Ст'юдента. Нехай деяка оцінка Θ^* параметра Θ лінійна за спостереженнями X_1, X_2, \dots, X_n , тобто

$$\Theta^* = c_0 + \sum_{i=1}^n c_i X_i.$$

Має місце таке твердження: якщо $X_i \sim N(a, \sigma) \forall i = \overline{1, n}$, тоді $\Theta^* \sim N(\Theta, \sigma_{\Theta^*})$, причому

$$M(\Theta^*) = c_0 + a \sum_{i=1}^n c_i, \quad \sigma_{\Theta^*}^2 = \sigma^2 \sum_{i=1}^n c_i^2, \quad (1.11)$$

а згідно із (1.6)

$$Z = \frac{\Theta^* - \Theta}{\sigma_{\Theta^*}} \sim N(0, 1). \quad (1.12)$$

В якості χ^2 -статистики візьмемо величину (1.7). Можна довести, що введені таким чином випадкові величини Z і χ є незалежними. А тому статистика

$$t = \frac{Z}{\chi} \sqrt{k} = \frac{\frac{\Theta^* - \Theta}{\sigma_{\Theta^*}}}{\sqrt{\frac{kS^2}{\sigma^2}}} \cdot \sqrt{k} = \frac{\Theta^* - \Theta}{\sigma_{\Theta^*}} \cdot \frac{\sigma}{S} \quad (1.13)$$

має розподіл Ст'юдента, причому згідно з другою рівністю (1.11) вона не залежить від генеральної дисперсії σ^2 .

Зокрема, якщо $\Theta^* = \bar{x}$, $\Theta = a$, тоді $\sigma_{\Theta^*} = \sigma / \sqrt{n}$, і статистика (1.13)

набуває такого вигляду:

$$t = \frac{\bar{x} - a}{S / \sqrt{n}}, k = n - 1. \quad (1.14)$$

Розподіл Фішера або F -розподіл (розподіл Фішера-Снедекора).

Нехай випадкові величини χ_1^2 і χ_2^2 мають χ^2 -розподіл із k_1 і k_2 ступенями вільності відповідно. Розподіл величини

$$F(k_1, k_2) = \frac{\chi_1^2}{k_1} : \frac{\chi_2^2}{k_2} \quad (1.15)$$

називається F -розподілом або розподілом Фішера із k_1 і k_2 ступенями вільності. Із (1.15) отримується:

$$F(k_2, k_1) = \frac{1}{F(k_1, k_2)}.$$

Універсальність F -розподілу підкреслюється зв'язками з іншими розподілами. При $k_1 = 1$, $k_2 = k$ корінь квадратний величини $F(1, k)$ має розподіл Ст'юдента із k ступенями вільності. Якщо ж $k_1 = k$, $k_2 = \infty$, то має місце тотожність

$$F(k, \infty) = \frac{\chi^2(k)}{k}.$$

Розподіл Фішера відіграє фундаментальну роль у математичній статистиці і досліджувався у першу чергу як розподіл частки двох вибірових дисперсій.

Нехай дві випадкові величини X та Y розподілені нормально за законом $N(a_1, \sigma_1)$ та $N(a_2, \sigma_2)$ відповідно, а S_1^2 та S_2^2 — незміщені оцінки із ступенями вільності k_1 і k_2 генеральних дисперсій σ_1^2 та σ_2^2 . Тоді згідно (1.7) випадкові величини $\chi_1^2 = \frac{k_1 S_1^2}{\sigma_1^2}$ і $\chi_2^2 = \frac{k_2 S_2^2}{\sigma_2^2}$ мають χ^2 -розподіл із

k_1 і k_2 ступенями вільності відповідно. На підставі (1.15) статистика

$$F(k_1, k_2) = \frac{\chi_1^2}{k_1} : \frac{\chi_2^2}{k_2} = \frac{S_1^2}{\sigma_1^2} : \frac{S_2^2}{\sigma_2^2} \quad (1.16)$$

розподілена за законом Фішера із k_1 і k_2 ступенями вільності.

Зокрема, якщо розглядається одна і та ж генеральна сукупність, тобто $a_1 = a_2$, $\sigma_1^2 = \sigma_2^2$, і S_1^2 та S_2^2 — вибіркові дисперсії при обсягах вибірок n_1 і n_2 відповідно, тоді з (1.16) одержуємо:

$$F(n_1 - 1, n_2 - 1) = \frac{S_1^2}{S_2^2}. \quad (1.17)$$

5. При умові наявності стохастичного зв'язку між двома випадковими величинами зміна можливих значень однієї з них приводить до зміни умовного закону розподілу іншої. Виявлення стохастичного зв'язку і оцінювання його сили — важлива і складна задача математичної статистики.

Якщо X та Y **незалежні** випадкові величини, тоді $D(X + Y) = D(X) + D(Y)$. Тому якщо $D(X + Y) \neq D(X) + D(Y)$, то це свідчить про наявність залежності між X та Y . У випадку виконання цієї нерівності будемо вважати, що випадкові величини X та Y **корельовані**, при цьому **кореляція** тим сильніша, чим більша різниця між лівою і правою частинами нерівності.

З'ясуємо аналітичний вираз для величини кореляції між X та Y . Використавши означення дисперсії, а також властивості математичного сподівання і дисперсії, отримаємо:

$$\begin{aligned} D(X + Y) &= M[X + Y - M(X + Y)]^2 = M\{[X - M(X)] + [Y - M(Y)]\}^2 = \\ &= M\{[X - M(X)]^2 + 2[X - M(X)][Y - M(Y)] + [Y - M(Y)]^2\} = \\ &= M[X - M(X)]^2 + 2M\{[X - M(X)][Y - M(Y)]\} + M[Y - M(Y)]^2 = \\ &= D(X) + 2\text{cov}(X, Y) + D(Y), \end{aligned}$$

де

$$\text{cov}(X, Y) = M\{[X - M(X)][Y - M(Y)]\}. \quad (1.18)$$

Таким чином, величина кореляції визначається величиною $\text{cov}(X, Y)$, яку називають **коваріацією** (спільною варіацією) або **кореляційним моментом**. Відмітимо також ще таку формулу для обчислення коваріації:

$$\text{cov}(X, Y) = M(X \cdot Y) - M(X) \cdot M(Y), \quad (1.18^*)$$

яка отримується із (1.18) з допомогою простих перетворень.

Коваріація залежить від одиниць виміру величин X та Y , тому доцільно використовувати безрозмірну величину

$$\rho = \frac{\text{cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)}, \quad (1.19)$$

яка називається **коефіцієнтом кореляції**.

Зауваження. Якщо випадкові величини X та Y незалежні, то $\rho = 0$. Проте обернене твердження неправильне: з рівності $\rho = 0$ не випливає незалежність випадкових величин (див. задачу 7.8 [3]).

Відмітимо такі властивості коефіцієнта кореляції [3, с.148]: 1) $|\rho| \leq 1$; 2) якщо $\rho = \pm 1$, то між X та Y існує **лінійна функціональна** залежність ($Y = \alpha_0 + \alpha_1 X$, де α_0 і α_1 — дійсні числа).

Якщо $\rho \neq 0$, то коефіцієнт кореляції своєю величиною характеризує не тільки наявність, але й силу **лінійного стохастичного** зв'язку між випадковими величинами: чим ближче $|\rho|$ до одиниці, тим сильніший лінійний зв'язок; чим ближче $|\rho|$ до нуля, тим слабший лінійний зв'язок.

Випадкові величини X та Y називаються **корельованими**, якщо $\rho \neq 0$, і **некорельованими**, якщо $\rho = 0$.

Недоліком коефіцієнта кореляції є те, що при $0 < |\rho| < 1$ між X та Y може бути як **стохастичний**, так і **функціональний нелінійний** зв'язок.

Оцінкою **невідомого** коефіцієнта кореляції ρ є **вибірковий коефіцієнт кореляції**, який обчислюється за формулою:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}. \quad (1.20)$$

Однак r є безпосередньою оцінкою ρ лише у **випадку двовимірного нормального закону розподілу** випадкових величин X та Y . Проте навіть і у цьому випадку при достатньо великому обсязі вибірки n оцінити похибку, що виникає, дуже важко. Але це і не обов'язково, тому що точне значення ρ у розрахунках практично не використовується, а потрібне лише як показник наявності кореляції між X та Y . Вибірковий коефіцієнт кореляції r застосовується в основному для перевірки статистичної гіпотези про наявність кореляції між величинами, що спостерігаються, не вдаючись у детальні оцінки сили цієї кореляції.

Випадковість вибірки може призвести до нерівності $r \neq 0$ навіть у випадку некорельованості випадкових величин. Тому для перевірки гіпотези про некорельованість величин потрібно перевірити, чи значуще r відрізняється від нуля. Відправною точкою такої перевірки є те, що при

некорельованості величин ($\rho = 0$) статистика $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ має t -розподіл

Ст'юдента із $k = n - 2$ ступенями вільності.

Вибірковий коефіцієнт кореляції r значущий на рівні α (гіпотеза $H_0 : r = 0$ відкидається), якщо

$$|t_{\text{крит.}}| = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} > t(1-\alpha; n-2), \quad (1.21)$$

де $t(1-\alpha; n-2)$ — табличне значення критерія Ст'юдента для рівня значущості α при числі ступенів вільності $n-2$.

І тільки у випадку, коли $H_0 : \rho = 0$ відкидається, будуються довірчі інтервали для оцінювання невідомого коефіцієнта кореляції r .

Нарешті, відмітимо інші модифікації формули для знаходження вибіркового коефіцієнта кореляції r :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x\sigma_y}, \quad (1.20^*)$$

$$r = \frac{n\sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{\sqrt{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \sqrt{n\sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i\right)^2}}. \quad (1.20^{**})$$

Для практичних розрахунків найбільш зручною є формула (1.20**), оскільки за нею r знаходиться безпосередньо із даних спостережень і на значення r не впливають заокруглення даних, зумовлені розрахунком середніх і відхилень від них.

§2. КЛАСИЧНА НОРМАЛЬНА ЛІНІЙНА МОДЕЛЬ ПАРНОЇ РЕГРЕСІЇ.

1. *Основні передумови нормальної класичної лінійної моделі парної регресії.*
2. *Знаходження оцінок параметрів.*
3. *Властивості МНК-оцінок.*
4. *Точкова оцінка дисперсії збурень. Число ступенів вільності.*
5. *Довірчий інтервал дисперсії збурень.*
6. *Перевірка значущості коефіцієнтів регресії.*
7. *Довірча зона функції регресії.*
8. *Довірчі інтервали коефіцієнтів регресії.*
9. *Критерії якості лінійної моделі. Коефіцієнт детермінації.*
10. *Прогнозування за класичною нормальною лінійною моделлю.*
11. *Задача.*

1. Припустимо, що попереднім результатом специфікації моделі (1.3) є висновок про лінійну залежність між X та Y . Тоді модель (1.3) набере такого вигляду:

$$Y = \alpha_0 + \alpha_1 X + U, \quad (2.1)$$

де α_0 , α_1 — невідомі (теоретичні) детерміновані параметри, U — невідома випадкова величина (збурення).

Приклади:

- 1) Y — річний обсяг заощаджень родини, X — річний дохід родини;
- 2) рівняння Кейнса: Y — індивідуальне споживання, X — наявний прибуток, α_0 — величина автономного споживання, α_1 — гранична схильність до споживання ($0 < \alpha_1 < 1$);
- 3) Y — річний товарообіг однієї філії торговельного підприємства, X — торговельна площа цієї філії;
- 4) Y — валовий випуск продукції, X — вартість основних виробничих фондів підприємства.

При моделюванні різних процесів розрізняють класичну і економетричну регресійні моделі.

У класичній лінійній моделі незалежна змінна вважається детермінованою величиною. В цьому параграфі будемо розглядати класичну лінійну модель

$$Y = \alpha_0 + \alpha_1 x + U. \quad (2.2)$$

Зауваження. Передумова стосовно детермінованості незалежної змінної не виконується для багатьох прикладних регресійних моделей в економіці і соціології, в які часто включаються випадкові неконтрольовані величини (наприклад, ціни і кількість пропонованих товарів або товарів, що користуються попитом).

Нехай x набирає значення x_1, x_2, \dots, x_n , де n — обсяг вибірки. Ці статистичні дані можуть бути або просторовими, або часовими рядами, або ж перехресними рядами. Тоді із (2.2) отримується система n рівнянь

$$Y_i = \alpha_0 + \alpha_1 x_i + U_i, \quad i = \overline{1, n}. \quad (2.3)$$

Повна специфікація моделі (2.2) передбачає виконання певних умов стосовно випадкової складової правої частини (2.3).

Передумова 1. Математичне сподівання збурень дорівнює нулю:

$$M(U_i) = 0, \quad i = \overline{1, n}. \quad (2.4)$$

Ця передумова означає, що збурення **в середньому** не здійснюють на Y ніякого впливу. Справді, за властивостями математичного сподівання для $i = \overline{1, n}$:

$$M(Y_i) = M(\alpha_0 + \alpha_1 x_i + U_i) = \alpha_0 + \alpha_1 x_i + M(U_i) = \alpha_0 + \alpha_1 x_i.$$

Передумова 2. Збурення мають однакову дисперсію:

$$D(U_i) = \sigma_u^2, \quad i = \overline{1, n}, \quad (2.5)$$

де σ_u^2 — **невідоме число**, яке підлягає оцінюванню.

Якщо виконуються рівності (2.5), то говорять, що збурення **гомоскедастичні**, якщо ж ні — то збурення **гетероскедастичні**. Наведені терміни зумовлені тим, що функція $g(x) = D(Y | x)$ називається **функцією скедастичності**.

Передумова 3. Збурення U_i і U_j при $i \neq j$ не корелюють між собою:

$$\text{cov}(U_i, U_j) = 0 \quad \forall i, j = \overline{1, n}, \quad i \neq j. \quad (2.6)$$

Враховуючи рівності (2.4) і формулу (1.18*), із (2.6) отримаємо рівності:

$$\text{cov}(U_i, U_j) = M(U_i \cdot U_j) - M(U_i)M(U_j) = M(U_i \cdot U_j) = 0,$$

тобто

$$M(U_i \cdot U_j) = 0, \quad i \neq j, \quad i, j = \overline{1, n}. \quad (2.7)$$

Передумова 4. Випадкові збурення U_i , $i = \overline{1, n}$, розподілені за нормальним законом.

Згідно із передумовами 1, 2: $U_i \sim N(0, \sigma_u)$, $i = \overline{1, n}$. Наслідком передумов 1-4 є нормальність розподілу випадкових величин Y_i :

$$Y_i \sim N(\alpha_0 + \alpha_1 x_i, \sigma_u), \quad i = \overline{1, n}. \quad (2.8)$$

Зауваження. Передумови 1-4 стосовно моделі (2.3) разом визначають класичну нормальну лінійну модель регресії. Якщо ж передумова 4 не виконується, то має місце класична лінійна модель.

2. Перший крок на шляху дослідження моделі (2.2) полягає у параметризації цієї моделі, тобто, щоб за **конкретною** вибіркою $\{(x_i, y_i), i = \overline{1, n}\}$ обсягом n знайти такі значення оцінок невідомих параметрів α_0, α_1 , для яких побудована пряма регресії була б найкращою в певному сенсі серед усіх інших прямих.

Нехай a_0 та a_1 — оцінки невідомих параметрів α_0 та α_1 відповідно. Тоді оцінкою моделі (2.3) по вибірці є n рівнянь

$$y_i = a_0 + a_1 x_i + u_i, \quad i = \overline{1, n}, \quad (2.9)$$

або

$$y_i = \hat{y}_i + u_i, \quad i = \overline{1, n}, \quad (2.10)$$

де \hat{y}_i — **групова середня**, або **умовна середня**, знайдена за рівнянням регресії

$$\hat{y} = a_0 + a_1 x, \quad (2.11)$$

тобто $\hat{y}_i = a_0 + a_1 x_i$,

u_i — **вбіркова оцінка збурення** U_i або **залишок регресії**, або ж **відхилення**.

Згідно із (2.8) \hat{y}_i можна назвати також **розрахунковим** або **оціночним значенням** Y_i при $x = x_i$, оскільки $M(Y_i) = \alpha_0 + \alpha_1 x_i$, а a_0 та a_1 — оцінки α_0 та α_1 відповідно.

Мірою якості оцінок a_0, a_1 можуть бути визначені композиції відхилень $u_i = y_i - \hat{y}_i$ [7, с.28]. Найпоширенішим і теоретично обґрунтованим є

метод, при якому мінімізується $\sum_{i=1}^n u_i^2$, і який має назву **метод найменших**

квадратів (МНК). Перевагами його є оптимальні властивості оцінок (незміщеність, ефективність, спроможність), а також зручність з обчислювальної точки зору.

Використаємо МНК для знаходження a_0, a_1 . Квадратична функція

$$Q(a_0, a_1) = \sum_{i=1}^n u_i^2 = \sum_{i=1}^n (a_0 + a_1 x_i - y_i)^2$$

є неперервною, опуклою і обмеженою знизу ($Q \geq 0$), тому має мінімум. Необхідною умовою існування екстремуму неперервно диференційованої функції двох змінних є рівність нулю її частинних похідних:

$$\begin{cases} \frac{\partial Q}{\partial a_0} = 2 \sum_{i=1}^n (a_0 + a_1 x_i - y_i) = 0, \\ \frac{\partial Q}{\partial a_1} = 2 \sum_{i=1}^n (a_0 + a_1 x_i - y_i) x_i = 0. \end{cases}$$

Цю систему рівнянь можна записати в такому вигляді:

$$\begin{cases} n a_0 + \left(\sum_{i=1}^n x_i \right) a_1 = \sum_{i=1}^n y_i, \\ \left(\sum_{i=1}^n x_i \right) a_0 + \left(\sum_{i=1}^n x_i^2 \right) a_1 = \sum_{i=1}^n x_i y_i. \end{cases}$$

Поділимо кожне із рівнянь на n :

$$\begin{cases} a_0 + \bar{x} a_1 = \bar{y}, \\ \bar{x} a_0 + \bar{x}^2 a_1 = \overline{xy}, \end{cases} \quad (2.12)$$

де $\bar{x} = \sum x_i / n$, $\bar{y} = \sum y_i / n$, $\overline{xy} = \sum x_i y_i / n$, $\bar{x}^2 = \sum x_i^2 / n$.

Використавши формулу Крамера, остаточно отримаємо єдиний розв'язок системи (2.12):

$$a_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - (\bar{x})^2}, \quad a_0 = \bar{y} - a_1 \bar{x}. \quad (2.13)$$

Отже, функція $Q(a_0, a_1)$ має єдину критичну точку. Виявляється [4, с.166], що в цій точці виконується і достатня умова існування мінімуму.

Оскільки оцінки (2.13) знайдені з допомогою МНК, то їх називають **МНК-оцінками**.

3. Для **нефіксованої** вибірки обсягом n згідно із (2.3) результуюча ознака Y набирає значення Y_1, Y_2, \dots, Y_n , які є випадковими величинами.

Тоді за формулами (2.13)

$$a_1 = \frac{\overline{xY} - \bar{x} \cdot \bar{Y}}{\bar{x}^2 - (\bar{x})^2}, \quad a_0 = \bar{Y} - a_1 \bar{x}, \quad (2.13^*)$$

тобто ці оцінки є також випадковими величинами.

Необхідно з'ясувати якість цих статистичних оцінок.

Властивість 1. МНК-оцінки a_0 , a_1 є лінійними комбінаціями спостережень Y_1, Y_2, \dots, Y_n :

$$a_1 = \sum_{i=1}^n \lambda_i Y_i, \quad a_0 = \sum_{i=1}^n \mu_i Y_i, \quad (2.14)$$

де

$$\lambda_i = \frac{x_i - \bar{x}}{n\sigma_x^2}, \quad \mu_i = \frac{1}{n} - \lambda_i \bar{x}, \quad i = \overline{1, n}, \quad \sigma_x^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / n. \quad (2.15)$$

□ Справді,

$$a_1 = \frac{\overline{xY} - \bar{x} \cdot \bar{Y}}{\overline{x^2} - (\bar{x})^2} = \frac{1}{n\sigma_x^2} \left(\sum_{i=1}^n x_i Y_i - \bar{x} \sum_{i=1}^n Y_i \right) = \frac{1}{n\sigma_x^2} \sum_{i=1}^n (x_i - \bar{x}) Y_i = \sum_{i=1}^n \frac{x_i - \bar{x}}{n\sigma_x^2} Y_i,$$

$$a_0 = \bar{Y} - a_1 \bar{x} = \frac{1}{n} \sum_{i=1}^n Y_i - \bar{x} \sum_{i=1}^n \frac{x_i - \bar{x}}{n\sigma_x^2} Y_i = \sum_{i=1}^n \left(\frac{1}{n} - \lambda_i \bar{x} \right) Y_i. \quad \square$$

Зауваження. Вагові коефіцієнти λ_i та μ_i залишаються незмінними при переході від вибірки до вибірки. Вони задовольняють такі співвідношення:

$$\sum_{i=1}^n \lambda_i = 0, \quad \sum_{i=1}^n \lambda_i x_i = 1, \quad \sum_{i=1}^n \lambda_i^2 = \frac{1}{n\sigma_x^2}, \quad \sum_{i=1}^n \mu_i = 1, \quad \sum_{i=1}^n \mu_i x_i = 0, \quad \sum_{i=1}^n \mu_i^2 = \frac{\overline{x^2}}{n\sigma_x^2}. \quad (2.16)$$

Пропонується самостійно перевірити правильність цих співвідношень.

Властивість 2. МНК-оцінки a_0 , a_1 є незміщеними оцінками відповідних параметрів α_0 , α_1 .

□ Використавши (2.8), (2.14), (2.16) і властивості математичного сподівання, отримаємо:

$$M(a_1) = M\left(\sum_{i=1}^n \lambda_i Y_i\right) = \sum_{i=1}^n \lambda_i M(Y_i) = \sum_{i=1}^n \lambda_i (\alpha_0 + \alpha_1 x_i) = \alpha_0 \sum_{i=1}^n \lambda_i + \alpha_1 \sum_{i=1}^n \lambda_i x_i = \alpha_1,$$

$$M(a_0) = M\left(\sum_{i=1}^n \mu_i Y_i\right) = \sum_{i=1}^n \mu_i M(Y_i) = \alpha_0 \sum_{i=1}^n \mu_i + \alpha_1 \sum_{i=1}^n \mu_i x_i = \alpha_0,$$

тобто

$$M(a_1) = \alpha_1, \quad M(a_0) = \alpha_0. \quad \square \quad (2.17)$$

Властивість 3 (теорема Гаусса-Маркова). *Із усіх лінійних незміщених оцінок параметрів α_0 , α_1 тільки МНК-оцінки a_0 , a_1 є ефективними і, отже, найкращими лінійними незміщеними оцінками.*

□ Обчислимо спочатку $D(a_1)$, $D(a_0)$, $\text{cov}(a_0, a_1)$. Для цього встановимо некорельованість випадкових величин Y_i та Y_j для $i \neq j$, $i, j = \overline{1, n}$, використавши (2.3), (2.4) та (2.7) і детермінованість α_0 , α_1 та x_i :

$$\begin{aligned} \text{cov}(Y_i, Y_j) &= M(Y_i \cdot Y_j) - M(Y_i)M(Y_j) = \\ &= M[(\alpha_0 + \alpha_1 x_i + U_i)(\alpha_0 + \alpha_1 x_j + U_j)] - M(\alpha_0 + \alpha_1 x_i + U_i)M(\alpha_0 + \alpha_1 x_j + U_j) = \\ &= M[(\alpha_0 + \alpha_1 x_i)(\alpha_0 + \alpha_1 x_j)] + M[(\alpha_0 + \alpha_1 x_i)U_j] + M[(\alpha_0 + \alpha_1 x_j)U_i] + \\ &\quad + M(U_i \cdot U_j) - (\alpha_0 + \alpha_1 x_i)(\alpha_0 + \alpha_1 x_j) = 0. \end{aligned}$$

А тому із (2.14)-(2.16) та (2.8) отримаємо:

$$D(a_1) = D\left(\sum_{i=1}^n \lambda_i Y_i\right) = \sum_{i=1}^n \lambda_i^2 D(Y_i) = \sum_{i=1}^n \lambda_i^2 \sigma_u^2 = \sigma_u^2 \sum_{i=1}^n \lambda_i^2 = \frac{\sigma_u^2}{n\sigma_x^2}; \quad (2.18)$$

$$D(a_0) = D\left(\sum_{i=1}^n \mu_i Y_i\right) = \sum_{i=1}^n \mu_i^2 D(Y_i) = \sum_{i=1}^n \mu_i^2 \sigma_u^2 = \sigma_u^2 \sum_{i=1}^n \mu_i^2 = \frac{\sigma_u^2 \bar{x}^2}{n\sigma_x^2}; \quad (2.19)$$

$$\begin{aligned} \text{cov}(a_0, a_1) &= M\{[a_0 - M(a_0)][a_1 - M(a_1)]\} = \\ &= M\left[\left(\sum_{i=1}^n \mu_i Y_i - \alpha_0\right)\left(\sum_{i=1}^n \lambda_i Y_i - \alpha_1\right)\right] = \\ &= M\left\{\left[\sum_{i=1}^n \mu_i (\alpha_0 + \alpha_1 x_i + U_i) - \alpha_0\right]\left[\sum_{i=1}^n \lambda_i (\alpha_0 + \alpha_1 x_i + U_i) - \alpha_1\right]\right\} = \\ &= M\left[\left(\alpha_0 \sum_{i=1}^n \mu_i + \alpha_1 \sum_{i=1}^n \mu_i x_i + \sum_{i=1}^n \mu_i U_i - \alpha_0\right)\left(\alpha_0 \sum_{i=1}^n \lambda_i + \alpha_1 \sum_{i=1}^n \lambda_i x_i + \sum_{i=1}^n \lambda_i U_i - \alpha_1\right)\right] = \\ &= M\left[\left(\sum_{i=1}^n \mu_i U_i\right)\left(\sum_{i=1}^n \lambda_i U_i\right)\right] = M\left[\sum_{i=1}^n \mu_i \lambda_i U_i^2 + \sum_{\substack{i,j=1 \\ i \neq j}}^n \mu_i \lambda_j U_i U_j\right] = \\ &= \sum_{i=1}^n \mu_i \lambda_i M(U_i^2) + \sum_{\substack{i,j=1 \\ i \neq j}}^n \mu_i \lambda_j M(U_i U_j) = \sigma_u^2 \sum_{i=1}^n \left(\frac{1}{n} - \lambda_i \bar{x}\right) \lambda_i = \end{aligned}$$

$$= \sigma_u^2 \left(\frac{1}{n} \sum_{i=1}^n \lambda_i - \bar{x} \sum_{i=1}^n \lambda_i^2 \right) = -\frac{\sigma_u^2 \bar{x}}{n\sigma_x^2} \neq 0.$$

Отже, оцінки a_0 і a_1 корелюють між собою, при цьому **коваріаційна матриця** цих оцінок має такий вигляд:

$$\Sigma_{a_0, a_1} = \begin{pmatrix} \sigma_{a_0}^2 & \sigma_{a_0, a_1} \\ \sigma_{a_1, a_0} & \sigma_{a_1}^2 \end{pmatrix} = \begin{pmatrix} \frac{\sigma_u^2 \bar{x}^2}{n\sigma_x^2} & -\frac{\sigma_u^2 \bar{x}}{n\sigma_x^2} \\ -\frac{\sigma_u^2 \bar{x}}{n\sigma_x^2} & \frac{\sigma_u^2}{n\sigma_x^2} \end{pmatrix}$$

або більш компактно:

$$\Sigma_{a_0, a_1} = \frac{\sigma_u^2}{n\sigma_x^2} \begin{pmatrix} \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}.$$

Зауваження. Символи $\sigma_{a_0}^2$ та σ_{a_0, a_1} позначають $D(a_0)$ і $\text{cov}(a_0, a_1)$ відповідно. Натомість символ Σ_{a_0, a_1} (велика літера «сигма» грецького алфавіту) позначає **дисперсійно-коваріаційну матрицю** (коротше: **коваріаційну матрицю**) двовимірної випадкової величини (a_0, a_1) , **втрачаючи значення символу суми**.

Розглянемо тепер довільні інші лінійні оцінки \hat{a}_0 , \hat{a}_1 (відмінні від a_0 , a_1) параметрів α_0 , α_1 такі, що

$$\hat{a}_0 = \sum_{i=1}^n B_i Y_i, \quad M(\hat{a}_0) = \alpha_0, \quad \hat{a}_1 = \sum_{i=1}^n C_i Y_i, \quad M(\hat{a}_1) = \alpha_1.$$

Згідно із формулюванням властивості 3, потрібно довести, що $D(\hat{a}_0) > D(a_0)$, $D(\hat{a}_1) > D(a_1)$, тобто ефективність оцінок a_0 , a_1 .

Доведемо, наприклад, другу нерівність. Нехай

$$C_i = \lambda_i + d_i, \quad \sum_{i=1}^n d_i^2 > 0,$$

де d_i — зсув ваги λ_i , визначеної (2.15), а друга нерівність зумовлена тим, що $a_1 \neq \hat{a}_1$.

Із лінійності оцінки \hat{a}_1 , рівнянь (2.3) та (2.4) отримуємо:

$$M(\hat{a}_1) = M\left(\sum_{i=1}^n C_i Y_i\right) = \sum_{i=1}^n C_i M(Y_i) = \sum_{i=1}^n C_i M(\alpha_0 + \alpha_1 x_i + U_i) =$$

$$= \alpha_0 \sum_{i=1}^n C_i + \alpha_1 \sum_{i=1}^n C_i x_i.$$

З другого боку, незміщеність \hat{a}_1 дає рівність $M(\hat{a}_1) = \alpha_1$, тому повинні виконуватись рівності

$$\sum_{i=1}^n C_i = 0, \quad \sum_{i=1}^n C_i x_i = 1 \quad \text{або} \quad \sum_{i=1}^n (\lambda_i + d_i) = 0, \quad \sum_{i=1}^n (\lambda_i + d_i) x_i = 1,$$

звідки із врахуванням перших двох формул (2.16) отримуємо необхідні умови для зсувів d_i :

$$\sum_{i=1}^n d_i = 0, \quad \sum_{i=1}^n d_i x_i = 0. \quad (2.20)$$

За аналогією з обчисленням $D(a_1)$:

$$D(\hat{a}_1) = \sigma_u^2 \sum_{i=1}^n C_i^2 = \sigma_u^2 \sum_{i=1}^n (\lambda_i + d_i)^2 = \sigma_u^2 \left(\sum_{i=1}^n \lambda_i^2 + 2 \sum_{i=1}^n \lambda_i d_i + \sum_{i=1}^n d_i^2 \right).$$

Але згідно із першою формулою (2.15) та (2.20):

$$\sum_{i=1}^n \lambda_i d_i = \sum_{i=1}^n \frac{(x_i - \bar{x})}{n\sigma_x^2} d_i = \frac{1}{n\sigma_x^2} \left(\sum_{i=1}^n d_i x_i - \bar{x} \sum_{i=1}^n d_i \right) = 0,$$

а відповідно до (2.18):

$$D(a_1) = \sigma_u^2 \sum_{i=1}^n \lambda_i^2 = \frac{\sigma_u^2}{n\sigma_x^2}.$$

Тому остаточно:

$$D(\hat{a}_1) = D(a_1) + \sum_{i=1}^n d_i^2 > D(a_1),$$

що свідчить про ефективність оцінки a_1 .

Аналогічно доводиться ефективність оцінки a_0 . ◻

Властивість 4. МНК-оцінки є спроможними оцінками.

◻ Нагадаємо [4, с.51], що оцінка Θ^* називається спроможною оцінкою параметра Θ , якщо для як завгодно малого $\varepsilon > 0$ має місце граничний перехід

$$\lim_{n \rightarrow \infty} P(|\Theta - \Theta^*| \leq \varepsilon) = 1.$$

Згідно із нерівністю Чебишева: $P(|Z - M(Z)| \leq \varepsilon) \geq 1 - \frac{D(Z)}{\varepsilon^2}$.

Для випадкової величини a_1 виконуються рівності $M(a_1) = \alpha_1$, $D(a_1) = \frac{\sigma_u^2}{n\sigma_x^2}$. Тому $D(a_1) \rightarrow 0$ при $n \rightarrow \infty$, тобто $P(|a_1 - \alpha_1| \leq \varepsilon) \rightarrow 1$ при $n \rightarrow \infty$ для як завгодно малого $\varepsilon > 0$.

Аналогічно доводиться спроможність оцінки a_0 . □

Висновки. При встановленні властивостей статистичних МНК-оцінок параметрів регресії в рамках класичної нормальної лінійної регресії були використані чотири умови стосовно випадкового збурення U , які часто називаються **умовами Гаусса-Маркова**, а саме — передумови 1-3 і умова $\text{cov}(U_i, x_i) = 0 \quad \forall i = \overline{1, n}$, яка виконується, оскільки за припущенням x є детермінованою змінною.

При невиконанні хоча б однієї з цих умов МНК-оцінки втрачають бажані властивості. Подолання складностей, які при цьому виникають, а також одержання більш надійних результатів — найважливіші задачі економетричного моделювання.

4. Вище були знайдені **теоретичні** числові характеристики МНК-оцінок. Для здійснення аналізу побудованої моделі (2.9), яка є оцінкою теоретичної моделі (2.3), потрібно знайти незміщені оцінки дисперсій (2.18), (2.19) для a_1 та a_0 відповідно. Кожна з цих дисперсій містить невідомий співмножник σ_u^2 . Тому необхідно знайти незміщену точкову оцінку σ_u^2 .

Вихідною інформацією для оцінювання σ_u^2 є значення $u_i = y_i - \hat{y}_i$, $i = \overline{1, n}$, тобто $\overline{u^2} - (\bar{u})^2$ — можлива оцінка σ_u^2 . Але

$$\bar{u} = \sum_{i=1}^n u_i / n = 0, \quad (2.21)$$

оскільки згідно із (2.13)

$$\sum_{i=1}^n u_i = \sum_{i=1}^n (y_i - a_0 - a_1 x_i) = \sum_{i=1}^n (y_i - \bar{y} + a_1 \bar{x} - a_1 x_i) = \sum_{i=1}^n (y_i - \bar{y}) + a_1 \sum_{i=1}^n (\bar{x} - x_i) = 0.$$

Відмітимо, що **рівність (2.21) відповідає передумові 1 і вона виконується**, взагалі кажучи, **тільки у випадку $\alpha_0 \neq 0$** (пропонується самостійно переконатися в цьому).

Отже, в оцінці σ_u^2 фігуруватиме тільки складова $\overline{u^2}$. Наступний крок — відкоригувати цю оцінку з тим, щоб вона стала незміщеною. Можна довести ([4, с.207]), що $M\left(\sum_{i=1}^n u_i^2\right) = \sigma_u^2(n-2)$.

Тому

$$S_u^2 = \frac{1}{n-2} \sum_{i=1}^n u_i^2 \left(= \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right) \quad (2.22)$$

є **незміщеною оцінкою** невідомої σ_u^2 .

Звернемо увагу на знаменник у (2.22). Виникає питання, як його **пояснити**, враховуючи, що незміщеною оцінкою (виправленою дисперсією) невідомої генеральної дисперсії кількісної ознаки X є

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (2.23)$$

тобто знаменник у цьому випадку вже дорівнює $n-1$. Справа у тому, що для знаходження (2.23) вхідною інформацією є n чисел x_1, x_2, \dots, x_n . Але серед чисел $x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x}$ незалежними є тільки $n-1$, оскільки $\sum_{i=1}^n (x_i - \bar{x}) = 0$, тобто одне із чисел x_1, x_2, \dots, x_n можна виразити з допомогою останньої рівності через решту $n-1$ чисел.

У випадку (2.22) вхідною інформацією є n чисел y_1, y_2, \dots, y_n , які задовольняють двом рівностям (2.14). Тепер уже два числа можна виразити через решту, використавши ці рівності, тобто лінійно незалежними є вже $n-2$ чисел.

Число l (для (2.22) $l=2$, а для (2.23) $l=1$) називається **числом зв'язків**, накладених на вибірку, при знаходженні дисперсії.

Число $k = n - l$ називається **числом ступенів вільності**.

5. Згідно із (1.7) та (2.22) випадкова величина

$$\chi^2 = \frac{(n-2)S_u^2}{\sigma_u^2} \quad (2.24)$$

розподілена за законом χ^2 із числом ступенів вільності $k = n - 2$. Оскільки розподіл χ^2 несиметричний [3, с.126], то це приводить до несиметричності і довірчого інтервалу для випадкової величини, розподіленої за цим законом.

Для побудови довірчого інтервалу (χ_1^2, χ_2^2) , в який з імовірністю γ потрапить можливе значення χ^2 , виберемо його межі таким чином, щоб

$$P(\chi^2 < \chi_1^2) = P(\chi^2 > \chi_2^2) = (1 - \gamma)/2. \quad (2.25)$$

Для цього значення χ_1^2 і χ_2^2 знайдемо за табл. 4 додатків, використовуючи рівняння

$$P(\chi^2(k) > \chi_1^2(p; k)) = p, \quad p = (1 + \gamma)/2, \quad (2.26)$$

$$P(\chi^2(k) > \chi_2^2(p; k)) = p, \quad p = (1 - \gamma)/2. \quad (2.27)$$

Із співвідношень (2.26), (2.27) з використанням протилежних подій і того, що для можливого значення χ_1^2 $P(\chi^2 = \chi_1^2) = 0$, впливає виконання рівностей (2.25).

Отже, за побудовою подвійна нерівність

$$\chi_1^2 < \frac{(n-2)S_u^2}{\sigma_u^2} < \chi_2^2 \quad (2.28)$$

виконується з імовірністю γ .

З першої нерівності (2.28) отримаємо рівносильну нерівність $\sigma_u^2 < \frac{(n-2)S_u^2}{\chi_1^2}$, а з другої: $\sigma_u^2 > \frac{(n-2)S_u^2}{\chi_2^2}$, об'єднання яких дає шуканий довірчий інтервал для оцінки σ_u^2 :

$$\frac{(n-2)S_u^2}{\chi_2^2} < \sigma_u^2 < \frac{(n-2)S_u^2}{\chi_1^2}. \quad (2.29)$$

6. Вище були знайдені точкові МНК-оцінки невідомих параметрів регресії α_0 , α_1 , а також їх числові характеристики. На перший погляд, наступний крок мав би полягати у побудові інтервальних оцінок для цих параметрів. Проте в полі зору необхідно тримати **правомірність самої специфікації вихідної моделі** (2.2), тобто теоретично можливі такі варіанти:

1) $\alpha_0 \neq 0$, $\alpha_1 \neq 0$; 2) $\alpha_0 = 0$, $\alpha_1 \neq 0$; 3) $\alpha_0 \neq 0$, $\alpha_1 = 0$; 4) $\alpha_0 = \alpha_1 = 0$.

При цьому **безпосередньо** на підставі вибірки неможливо діагностувати кожний із цих варіантів, оскільки вибірка організовується випадковим чином. У той же час немає сенсу будувати довірчі інтервали для α_1 у випадках 3 та 4, а для α_0 — у випадках 2 та 4.

Відмітимо важливість діагностування рівності $\alpha_1 = 0$. У цьому випадку взагалі слід переглянути висунуту модель. Разом з тим, якщо $\alpha_0 = 0$ (при $\alpha_1 \neq 0$), то це, як буде встановлено нижче, необхідно враховувати при розгляді питання про якість моделі.

Отже, актуальним є питання про **з'ясування значущості** коефіцієнтів регресії α_0 і α_1 на підставі наявних статистичних даних.

Висунемо статистичну нульову гіпотезу $H_0 : \alpha_m = 0$, де $m = 0, 1$. Альтернативна гіпотеза $H_1 : \alpha_m \neq 0$. Задамо також рівень значущості α .

Згідно із (2.17) і (2.14) МНК-оцінки a_m є незміщеними і лінійними відносно випадкових величин Y_1, Y_2, \dots, Y_n , які у відповідності із (2.8) розподілені нормально ($Y_i \sim N(\alpha_0 + \alpha_1 x_i, \sigma_u)$, $i = \overline{1, n}$). Тому і a_m мають нормальний розподіл:

$$a_m \sim N(\alpha_m, \sigma_{a_m}^2), \quad i = 0, 1,$$

де $\sigma_{a_m}^2$ визначаються формулами (2.18), (2.19). Незміщеною оцінкою σ_u^2 є S_u^2 (формула (2.22)). Тому із врахуванням (2.18), (2.19) незміщені оцінки $\sigma_{a_m}^2$ (вони невідомі, бо невідомим є σ_u^2 !) визначаються формулами

$$S_{a_0}^2 = \frac{S_u^2 \overline{x^2}}{n \sigma_x^2}, \quad S_{a_1}^2 = \frac{S_u^2}{n \sigma_x^2}. \quad (2.30)$$

Розглянемо випадкову величину (1.13), де $Z = Z_m = \frac{a_m - \alpha_m}{\sigma_{a_m}}$, $k = n - 2$,

$\chi = \sqrt{\chi^2}$, χ^2 визначається за формулою (2.24), тобто

$$t_m = \frac{a_m - \alpha_m}{\sigma_{a_m}} \cdot \frac{\sigma_u}{S_u}. \quad (2.31)$$

Вона має t -розподіл із $k = n - 2$ ступенями вільності і вже **не залежить від σ_u** .

Враховуючи (2.18), (2.19) та (2.30), із (2.31) отримаємо такі t -розподіли:

$$t_0 = \frac{a_0 - \alpha_0}{S_{a_0}}, \quad t_1 = \frac{a_1 - \alpha_1}{S_{a_1}}. \quad (2.32)$$

Зміст основної гіпотези $H_0(\alpha_m = 0)$ та альтернативної $H_1(\alpha_m \neq 0)$ дозволяє сформулювати **двосторонній критерій значущості оцінок a_0, a_1** :

якщо виконується нерівність

$$\left| \frac{a_m}{S_{a_m}} \right| > t_{кр.}, \quad (2.33)$$

де $t_{кр.} = t_{двосст.кр.}(\alpha, n-2)$, $\alpha = 1-\gamma$, — критична точка розподілу Ст'юдента (табл. 3 додатків), тоді на рівні значущості α приймається гіпотеза H_1 , тобто вважається, що $\alpha_m \neq 0$.

7. Нехай нерівність (2.33) виконується для $m = 1$, що означає існування стохастичного зв'язку між змінними Y та x .

Побудуємо **довірчий інтервал для функції регресії**, тобто для умовного математичного сподівання $M(Y | x)$, який із заданою надійністю γ покриває невідоме значення $\alpha_0 + \alpha_1 x$.

Знайдемо дисперсію групової середньої \hat{y} , яка є статистичною оцінкою $M(Y | x)$. Для цього рівняння регресії (2.11) запишемо у такому вигляді

$$\hat{y} = \bar{y} + a_1(x - \bar{x}), \quad (2.34)$$

підставивши другу рівність (2.13) в (2.11). Перевагою рівняння (2.34) є некорельованість його доданків справа. Доведення рівності $\text{cov}(\bar{y}, a_1) = 0$ (за припущенням $x - \bar{x}$ є детермінованою величиною) здійснюється з використанням (2.14), (2.4) і (2.16). Тому дисперсія лівої частини (2.34) дорівнює сумі дисперсій двох **незалежних** доданків правої:

$$D(\hat{y}) = D(\bar{y}) + (x - \bar{x})^2 D(a_1), \quad (2.35)$$

де $(x - \bar{x})^2$ отримується внаслідок винесення детермінованого множника за знак дисперсії, піднесеного до квадрату.

Згідно із (2.18) $D(a_1) = \frac{\sigma_u^2}{n\sigma_x^2}$, а використання (2.8) дає

$$\sigma_y^2 = D(\bar{y}) = D\left(\frac{\sum_{i=1}^n Y_i}{n}\right) = \frac{\sigma_u^2}{n}.$$

Отже, з (2.35)

$$D(\hat{y}) = \left[1 + \frac{(x - \bar{x})^2}{\sigma_x^2} \right] \frac{\sigma_u^2}{n},$$

а незміщена оцінка $D(\hat{y})$ знаходиться за формулою:

$$S_{\hat{y}}^2 = \left[1 + \frac{(x - \bar{x})^2}{\sigma_x^2} \right] \frac{S_u^2}{n}, \quad (2.36)$$

де S_u^2 визначена формулою (2.22).

Розглянемо статистику (1.10)

$$t = \frac{Z}{\sqrt{\frac{1}{k} \chi^2}} = \frac{Z}{\chi} \sqrt{k},$$

поклавши $Z = \frac{\hat{y} - M(Y|x)}{\sigma_{\hat{y}}}$, $\chi^2 = \frac{kS_{\hat{y}}^2}{\sigma_{\hat{y}}^2}$, $k = n - 2$.

Використання передумов 1-4 та (1.7) дозволяє показати, що $Z \sim N(0,1)$, а χ^2 має розподіл χ^2 із k ступенями вільності, причому Z та χ^2 незалежні випадкові величини. Тому випадкова величина

$$t = \frac{\hat{y} - M(Y|x)}{S_{\hat{y}}}$$

має t -розподіл Ст'юдента із $k = n - 2$ ступенями вільності. Оскільки густина $g_k(t)$ розподілу Ст'юдента парна, то

$$P\left(\left| \frac{\hat{y} - M(Y|x)}{S_{\hat{y}}} \right| < t\right) = P(|T| < t) = 2 \int_0^t g_k(t) dt = \gamma. \quad (2.37)$$

У табл. 2 додатків наведені значення $t = t(\gamma; k)$ як кореня рівняння (2.37) в залежності від заданої довірчої імовірності γ і від числа ступенів вільності k .

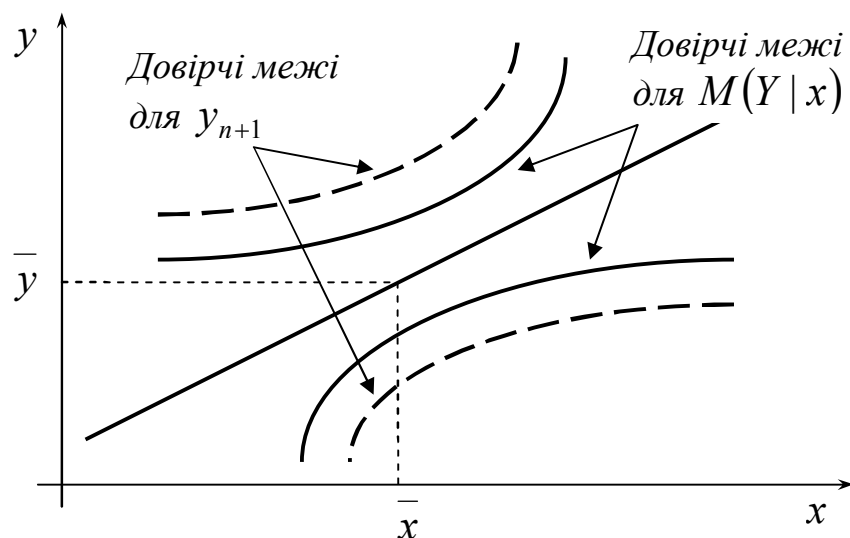
Таким чином, довірчий інтервал для невідомого умовного сподівання $M(Y|x)$ має такий вигляд:

$$\hat{y} - t(\gamma; n-2)S_{\hat{y}} < M(Y|x) < \hat{y} + t(\gamma; n-2)S_{\hat{y}}, \quad (2.38)$$

де $S_{\hat{y}} = \sqrt{S_{\hat{y}}^2}$, $S_{\hat{y}}^2$ визначається формулою (2.36).

Якщо врахувати, що $M(Y|x) = \alpha_0 + \alpha_1 x$, а x змінюється, то подвійна нерівність (2.38) дає **довірчу зону** для прямої регресії $y = \alpha_0 + \alpha_1 x$. Тому

потрібно з'ясувати, яка поведінка «ширини» цієї зони при зміні x . Із формул (2.36) і (2.38) видно, що «ширина» зони мінімальна при $x = \bar{x}$, а при віддаленні x від \bar{x} вона збільшується (мал. 2.1).



Малюнок 2.1.

Таким чином, **прогноз** значень залежної змінної Y (в середньому) може призвести до **значних похибок**.

Побудова довірчої зони для функції регресії (або $M(Y|x)$) у випадку **конкретної** вибірки (іншими словами, довірчої зони для значень регресії в базисних точках) передбачає побудову точок з координатами $\{x_i; \hat{y}_i - t(\gamma, n-2)S_{\hat{y}_i}\}$, $i = \overline{1, n}$, з наступним з'єднанням сусідніх (по індексу i) точок прямолінійними відрізками, а потім здійснення аналогічної процедури для послідовності точок $\{x_i; \hat{y}_i + t(\gamma, n-2)S_{\hat{y}_i}\}$.

8. Нехай МНК-оцінки a_0 та a_1 є значущими, тобто виконуються нерівності (2.33). Побудуємо довірчі інтервали для невідомих параметрів α_0 та α_1 , використавши випадкові величини (2.32), які розподілені за законом Ст'юдента із $k = n - 2$ ступенями вільності.

Позначимо $t_0(\gamma, k)$, $t_1(\gamma, k)$ корені рівнянь

$$P(|t_0| < t) = \gamma, \quad P(|t_1| < t) = \gamma$$

відповідно, які знаходяться за табл. 2 додатків. Тоді із надійністю γ виконуються нерівності

$$\left| \frac{a_0 - \alpha_0}{S_{a_0}} \right| < t_0(\gamma, k), \quad \left| \frac{a_1 - \alpha_1}{S_{a_1}} \right| < t_1(\gamma, k),$$

звідки отримуються довірчі інтервали

$$a_0 - t_0(\gamma, k)S_{a_0} < \alpha_0 < a_0 + t_0(\gamma, k)S_{a_0}, \quad (2.39)$$

$$a_1 - t_1(\gamma, k)S_{a_1} < \alpha_1 < a_1 + t_1(\gamma, k)S_{a_1}, \quad (2.40)$$

де S_{a_0} , S_{a_1} визначаються формулами (2.30).

9. Основна мета дослідження моделі (2.2) — це статистично надійний точковий чи інтервальний прогноз значень залежної змінної Y або її математичного сподівання. Для досягнення цієї мети необхідно з'ясувати **якість моделі** або **значущість моделі**, тобто встановити, чи відповідає модель експериментальним даним і чи достатньо включених в рівняння незалежних (пояснюючих) змінних (однієї або кількох) для описання залежної змінної. Перевірка значущості рівняння регресії проводиться на основі дисперсійного аналізу, який у даному випадку використовується як допоміжний засіб для вивчення якості регресійної моделі.

Розглянемо питання про **декомпозицію дисперсій**. Дисперсія спостережених значень залежної змінної має такий вигляд:

$$\sigma_y^2 = \sum_{i=1}^n (y_i - \bar{y})^2 / n.$$

Відхилення $y_i - \bar{y}$ запишемо таким чином:

$$y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i). \quad (2.41)$$

У статистиці різницю $y_i - \bar{y}$ називають **загальним відхиленням**, $\hat{y}_i - \bar{y}$ — **відхиленням, яке можна пояснити моделлю**, $y_i - \hat{y}_i$ — **непояснюваним відхиленням** (яке не можна пояснити моделлю).

Піднесемо обидві частини (2.41) до квадрату та підсумуємо за індексом i :

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

Якщо в моделі (2.2) $\alpha_0 \neq 0$, тоді другий доданок в правій частині дорівнює нулю із врахуванням системи нормальних рівнянь (2.12):

$$\begin{aligned} \sum (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) &= \sum \hat{y}_i (y_i - \hat{y}_i) - \bar{y} \sum (y_i - \hat{y}_i) = \\ &= \sum (a_0 + a_1 x_i)(y_i - a_0 - a_1 x_i) - \bar{y} \sum (y_i - a_0 - a_1 x_i) = \end{aligned}$$

$$= a_0 \sum (y_i - a_0 - a_1 x_i) + a_1 \sum (y_i - a_0 - a_1 x_i) x_i = 0.$$

Тому остаточно отримаємо рівність

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.42)$$

або

$$СКЗ = СКП + СКН,$$

де $СКЗ$ — загальна сума квадратів, $СКП$ — пояснена сума квадратів, $СКН$ — непояснена сума квадратів.

Поділивши (2.42) на n , отримаємо:

$$\sigma_{заг.}^2 = \sigma_{регр.}^2 + \sigma_{ном.}^2, \quad (2.43)$$

де $\sigma_{заг.}^2$ — загальна дисперсія, $\sigma_{регр.}^2$ — дисперсія, що пояснює регресію, $\sigma_{ном.}^2$ — дисперсія помилок.

Для одержання **незміщених** оцінок дисперсій, які фігурують в рівності (2.43), потрібно знайти відповідні ступені вільності.

Для обчислення $СКЗ$ використовуються n чисел $\{(y_1 - \bar{y}), (y_2 - \bar{y}), \dots, (y_n - \bar{y})\}$, серед яких лінійно незалежними є $n-1$, оскільки $\sum_{i=1}^n (y_i - \bar{y}) = 0$. Тому ступінь вільності $СКЗ$ дорівнює $n-1$.

Із врахуванням (2.13) рівняння регресії можна записати у такому вигляді:

$$\hat{y}_i - \bar{y} = a_1 (x_i - \bar{x}),$$

звідки $СКП = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = a_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$.

Отже, $СКП$ утворюється з використанням однієї одиниці незалежної інформації — a_1 , тому ступінь вільності її дорівнює 1 або $m-1$, де m — число параметрів моделі ($m = 2$).

Нарешті, $СКН = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ має (див. п.4) $n-2$ ступенів вільності.

Означення. Число, яке отримується діленням суми квадратів на відповідний ступінь вільності, називається **середнім квадратом**.

Середні квадрати

$$S_R^2 = \overline{СКП} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / 1, \quad S_u^2 = \overline{СКН} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-2) \quad (2.44)$$

є **незміщеними оцінками** дисперсій залежної змінної, обумовлених відповідно регресією (пояснюючою змінною) і дією неврахованих факторів та помилок.

Повернемося до рівності (2.43). Якщо обидві її частини розділити на $\sigma_{заг.}^2$, то отримаємо

$$1 = \frac{\sigma_{регр.}^2}{\sigma_{заг.}^2} + \frac{\sigma_{ном.}^2}{\sigma_{заг.}^2}. \quad (2.45)$$

Означення. Перший доданок в правій частині (2.45) називається **коєфіцієнтом детермінації** і позначається R^2 (або d):

$$R^2 = \frac{\sigma_{регр.}^2}{\sigma_{заг.}^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{n\sigma_y^2}. \quad (2.46)$$

Із рівності (2.45) отримаємо рівносильну формулу для означення коєфіцієнта детермінації:

$$R^2 = 1 - \frac{\sigma_{ном.}^2}{\sigma_{заг.}^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n\sigma_y^2} = 1 - \frac{\sum_{i=1}^n u_i^2}{n\sigma_y^2}. \quad (2.46^*)$$

R^2 є однією із **найбільш ефективних оцінок адекватності регресійної моделі, мірою її якості або характеристикою прогностичної сили моделі.**

Величина R^2 показує, яка частина (частка) варіації залежної змінної обумовлена варіацією пояснюючої змінної.

Оскільки доданки в рівності (2.45) невід'ємні, то з врахуванням (2.46) отримується подвійна нерівність $0 \leq R^2 \leq 1$.

Чим ближче значення R^2 до одиниці, тим краще модель апроксимує емпіричні дані, тим ближче спостереження знаходяться по відношенню до прямої регресії. Якщо $R^2 = 1$, то всі емпіричні точки (x_i, y_i) лежать на прямій регресії і між змінними Y та x існує лінійна функціональна залежність. Якщо ж $R^2 = 0$, то варіація залежної змінної повністю обумовлена дією неврахованих у моделі змінних і пряма регресії паралельна осі абсцис.

Зауваження. Коєфіцієнт детермінації R^2 є сенс розглядати тільки у випадку наявності вільного члена в моделі (2.2), тобто $\alpha_0 \neq 0$, оскільки

тільки у цьому випадку, як це вище відзначалося, виконується рівність (2.42), а, отже, і співвідношення (2.46*).

Зміст R^2 нагадує зміст вибіркового коефіцієнта кореляції r . Це обумовлено рівністю

$$R^2 = r^2, \quad (2.47)$$

яка отримується із використанням (2.46), (1.20), (2.13):

$$\begin{aligned} R^2 &= \frac{\sigma_{\text{регр.}}^2}{\sigma_{\text{заг.}}^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{n\sigma_y^2} = \frac{\sum_{i=1}^n a_1^2 (x_i - \bar{x})^2}{n\sigma_y^2} = \frac{a_1^2 \sigma_x^2}{n\sigma_y^2} = \\ &= \left(\frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2} \right)^2 \frac{\sigma_x^2}{\sigma_y^2} = \left(\frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y} \right)^2 = r^2. \end{aligned}$$

Числові характеристики R^2 та r є точковими статистичними оцінками відповідних невідомих чисел. У зв'язку із цим навіть у випадку $\alpha_1 = 0$ R^2 та $|r|$ можуть бути відмінними від нуля. А тому виникає необхідність перевірити значущість R^2 та r , отриманих для конкретної вибірки.

Якщо $\alpha_1 = 0$, тобто відсутня лінійна залежність між залежною і пояснюючою змінними, тоді випадкові величини

$$S_R^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / 1 \quad \text{та} \quad S_u^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-2)$$

мають χ^2 -розподіли відповідно із 1 та $n-2$ ступенями вільності, а їх відношення (згідно із (1.15)) — розподіл Фішера з тими ж ступенями вільності. Тому рівняння регресії значуще на рівні значущості α , якщо виконується нерівність

$$F_{\text{спост.}} = \frac{СКП \cdot (n-2)}{СКН \cdot 1} = \frac{S_R^2}{S_u^2} > F_{\text{кр.}}(\alpha; k_1; k_2), \quad (2.48)$$

де $F_{\text{кр.}}(\alpha; k_1; k_2)$ — табличне значення F -критерія Фішера-Снедекора, визначене на рівні значущості α при $k_1 = 1$ і $k_2 = n-2$ ступенях вільності.

Проте виявляється [8, с.97], що F -тест (2.48) рівносильний t -тесту Ст'юдента при перевірці значущості параметра α_1 (у випадку парної лінійної моделі).

В ряді задач потрібно оцінити значущість коефіцієнта кореляції r . На рівні значущості α він вважається значущим (тобто відкидається гіпотеза $H_0 : \rho = 0$), якщо виконується нерівність (1.21). Однак неважко пока-

зати, що отримувані значення t -критерію при перевірці гіпотез $\alpha_1 = 0$ по (2.33) і $\rho = 0$ по (1.21) однакові.

Отже, якщо на рівні $\alpha = 1 - \gamma$ зроблено висновок про значущість α_1 , то на тому ж рівні вважається значущим і генеральний (теоретичний) коефіцієнт кореляції ρ і навпаки.

Наведемо інші прості показники якості лінійної регресії, які використовуються як додаткова інформація при виборі найкращої моделі з можливих.

Абсолютна середня відсоткова помилка MAPE (mean absolute percentage error):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \cdot 100\%. \quad (2.49)$$

Цей показник використовується при порівнянні точності прогнозів різнорідних об'єктів, бо характеризує відносну точність прогнозу. При цьому вважається, що значення MAPE, менше 10%, дає високу точність прогнозу, а, отже, і якість моделі; від 10% до 20% — добру точність; від 20% до 50% — задовільну точність; понад 50% — незадовільну точність.

Середня відсоткова помилка MPE (mean percentage error):

$$MPE = \frac{1}{n} \sum_{i=1}^n \frac{\hat{y}_i - y_i}{y_i} \cdot 100\%. \quad (2.50)$$

Це показник незміщеності прогнозу. З точки зору практики для **якісних моделей** цей показник повинен бути «малим», тобто не перевищувати 5%.

Зауваження. Показники (2.49) та (2.50) — невизначені, якщо серед y_1, y_2, \dots, y_n є нульове значення.

10. Якщо встановлено, що побудована модель є адекватною, тоді можна знаходити **прогнознi значення залежної змінної**. При цьому можна отримати два типи прогнозів: точковий та інтервальний. Нехай задається значення x_{n+1} незалежної змінної. Тоді точковий прогноз для значення залежної змінної за моделлю (2.11) має такий вигляд:

$$\hat{y}_{n+1} = a_0 + a_1 x_{n+1}. \quad (2.51)$$

Разом з тим **дійсне** значення залежної змінної для прогнозного періоду згідно із (2.3) дорівнює:

$$Y_{n+1} = \alpha_0 + \alpha_1 x_{n+1} + U_{n+1}, \quad (2.52)$$

де стосовно випадкової величини U_{n+1} природно вимагати виконання передумов 1-4, тобто

$$\text{cov}(U_i, U_{n+1}) = 0, \quad i = \overline{1, n} \quad U_{n+1} \sim N(0, \sigma_u^2). \quad (2.53)$$

Отже, \hat{y}_{n+1} є точковою оцінкою **невідомого числа** y_{n+1} , яке є реалізацією (можливим значенням) випадкової величини (2.52).

Згідно із (2.51), (2.52) помилка прогнозу:

$$u_{n+1} = y_{n+1} - \hat{y}_{n+1} = (\alpha_0 - a_0) + (\alpha_1 - a_1)x_{n+1} + U_{n+1}. \quad (2.54)$$

Потрібно знайти числові характеристики u_{n+1} та закон розподілу. Незміщеність оцінок a_0, a_1 і (2.53) призводять до рівності $M(u_{n+1}) = 0$. Оскільки U_{n+1} не корелює із U_1, U_2, \dots, U_n , то згідно із (2.3) U_{n+1} не корелює із Y_1, Y_2, \dots, Y_n . Тому із врахуванням детермінованості $\alpha_0, \alpha_1, x_{n+1}$, (2.14), (2.18), (2.19) отримаємо:

$$\begin{aligned} D(u_{n+1}) &= D(-a_0) + D(-x_{n+1}a_1) + 2 \text{cov}(-a_0, -x_{n+1}a_1) + D(U_{n+1}) = \\ &= D(a_0) + x_{n+1}^2 D(a_1) + 2x_{n+1} \text{cov}(a_0, a_1) + \sigma_u^2 = \\ &= \sigma_u^2 \frac{\overline{x^2}}{n\sigma_x^2} + x_{n+1}^2 \frac{\sigma_u^2}{n\sigma_x^2} - 2x_{n+1} \frac{\bar{x}}{n\sigma_x^2} \sigma_u^2 + \sigma_u^2 = \\ &= \left\{ \overline{x^2} - (\bar{x})^2 + \left[(\bar{x})^2 - 2\bar{x}x_{n+1} + x_{n+1}^2 \right] \frac{1}{n\sigma_x^2} + 1 \right\} \sigma_u^2 = \left[1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{n\sigma_x^2} \right] \sigma_u^2. \end{aligned}$$

Неважко переконатися також у лінійній залежності u_{n+1} від збурень U_1, U_2, \dots, U_{n+1} . А тому остаточно отримуємо:

$$u_{n+1} \sim N(0, \sigma_{u_{n+1}}),$$

де

$$\sigma_{u_{n+1}} = \sigma_u \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{n\sigma_x^2}}.$$

Незміщена оцінка $D(u_{n+1}) = \sigma_{u_{n+1}}^2$ знаходиться за формулою

$$S_{u_{n+1}}^2 = S_u^2 \left[1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{n\sigma_x^2} \right], \quad (2.55)$$

де S_u^2 визначена формулою (2.22).

За аналогією із побудовою інтервальної зони функції регресії (п.7) остаточно можна отримати довірчий інтервал для прогнозного значення залежної змінної:

$$\hat{y}_{n+1} - t(\gamma; n - 2)S_{u_{n+1}} < y_{n+1} < \hat{y}_{n+1} + t(\gamma; n - 2)S_{u_{n+1}}. \quad (2.56)$$

11. Задача 2.1. Торгівельне підприємство має велику кількість філій і його керівництво вивчає питання про залежність Y (річний товарообіг однієї філії, млн. грн.) від x (торгівельної площі, тис. м²). Для десяти філій за певний рік зафіксовані такі значення показників Y і x :

i	1	2	3	4	5	6	7	8	9	10
y_i	1,5	2,9	3,1	3,2	4,3	5,7	5,8	7	7,2	7,5
x_i	0,2	0,3	0,5	0,6	0,8	1	1,1	1,2	1,3	1,4

На обсяг товарообігу впливають такі чинники: середньоденна інтенсивність потоку покупців, об'єм основних фондів, їх структура, середньоспискова чисельність працівників, площа підсобних приміщень тощо. Припускається, що в досліджуваній групі філій значення цих чинників приблизно однакові, тому вплив відмінностей їх значень на зміну обсягу товарообігу є незначним.

Вважаючи, що виконуються передумови 1-4, потрібно:

- 1) знайти статистичні оцінки параметрів лінійного рівняння регресії;
- 2) точкову оцінку та довірчий інтервал дисперсії збурень із надійністю $\gamma = 0,9$;
- 3) для рівня значущості $\alpha = 0,05$ перевірити значущість коефіцієнтів регресії α_0 та α_1 ;
- 4) знайти довірчі інтервали коефіцієнтів регресії з надійністю $\gamma = 0,95$;
- 5) знайти вибіркові коефіцієнт детермінації, коефіцієнт кореляції, а також інші показники якості лінійної регресії (MAPE, MPE);
- 6) знайти та побудувати довірчу зону функції регресії з надійністю $\gamma = 0,95$;
- 7) знайти прогнозне значення річного товарообігу для нової філії, торговельна площа якої складає 1,8 тис. м², а також із надійністю $\gamma = 0,95$ побудувати довірчий інтервал для цього прогнозного значення.

○ 1) Статистичні оцінки a_0 , a_1 параметрів α_0 та α_1 лінійного рівняння регресії задовольняють системі нормальних рівнянь (2.12):

$$\begin{cases} a_0 + \bar{x}a_1 = \bar{y}, \\ \bar{x}a_0 + \bar{x}^2a_1 = \overline{xy}. \end{cases}$$

Для знаходження коефіцієнтів цієї системи складемо розрахункову табл. 2.1, останній стовпець якої потрібний для обчислення σ_y .

Таблиця 2.1

i	x_i	y_i	x_i^2	$x_i y_i$	y_i^2
1	0,2	1,5	0,04	0,3	2,25
2	0,3	2,9	0,09	0,87	8,41
3	0,5	3,1	0,25	1,55	9,61
4	0,6	3,2	0,36	1,92	10,24
5	0,8	4,3	0,64	3,44	18,49
6	1	5,7	1	5,7	32,49
7	1,1	5,8	1,21	6,38	33,64
8	1,2	7	1,44	8,4	49
9	1,3	7,2	1,69	9,36	51,84
10	1,4	7,5	1,96	10,5	56,25
Σ	8,4	48,2	8,68	48,42	272,22

Використовуючи нижній рядок табл. 2.1, отримаємо (обсяг вибірки $n = 10$):

$$\begin{aligned} \bar{x} &= \sum_{i=1}^{10} x_i / n = 8,4 / 10 = 0,84; & \bar{y} &= \sum_{i=1}^{10} y_i / n = 48,2 / 10 = 4,82; \\ \overline{x^2} &= \sum_{i=1}^{10} x_i^2 / n = 8,68 / 10 = 0,868; & \overline{xy} &= \sum_{i=1}^{10} x_i y_i / n = 48,42 / 10 = 4,842; \\ \overline{y^2} &= \sum_{i=1}^{10} y_i^2 / n = 272,22 / 10 = 27,222; \end{aligned}$$

$$\begin{cases} a_0 + 0,84a_1 = 4,82, \\ 0,84a_0 + 0,868a_1 = 4,842. \end{cases}$$

Єдиний розв'язок цієї системи рівнянь згідно із формулами (2.13):

$$\begin{aligned} a_1 &= \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{4,842 - 0,84 \cdot 4,82}{0,868 - (0,84)^2} = \frac{0,7932}{0,1624} = 4,884, \\ a_0 &= \bar{y} - a_1 \bar{x} = 4,82 - 0,84 \cdot 4,884 = 0,717. \end{aligned}$$

Отже, емпіричне рівняння регресії має такий вигляд:

$$\hat{y} = 0,717 + 4,884x. \quad (2.57)$$

2) Незміщену точкову оцінку S_u^2 невідомої дисперсії збурень σ_u^2 знайдемо за формулою (2.22):

$$S_u^2 = \frac{1}{n-2} \sum_{i=1}^n u_i^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

попередньо обчисливши $\hat{y}_i = 0,717 + 4,884x_i$ та $u_i^2 = (y_i - \hat{y}_i)^2$, $i = \overline{1,10}$, (табл. 2.2).

Таблиця 2.2

i	1	2	3	4	5	6	7	8	9	10	Σ
y_i	1,5	2,9	3,1	3,2	4,3	5,7	5,8	7	7,2	7,5	–
\hat{y}_i	1,6938	2,1822	3,159	3,6474	4,6242	5,601	6,0894	6,5778	7,0662	7,5546	–
u_i	-0,1938	0,7178	-0,059	-0,4474	-0,3242	0,099	-0,2894	0,4222	0,1338	-0,0546	0,0044
u_i^2	0,0376	0,5152	0,0035	0,2002	0,1051	0,0098	0,0838	0,1783	0,0179	0,003	1,1544

Зауваження. Згідно із (2.21): $\bar{u} = 0$, в той час як у нашому випадку $\sum u_i / 10 = 0,00044$. Цим значенням можна ігнорувати (вважати практично рівним нулю). Разом з тим з'ясуємо причину такого відхилення від нуля. Значення a_1 та a_0 з точністю до шести знаків після коми відповідно складають 4,884236 та 0,717242, тобто обидва ці значення (хай і несуттєво) більші тих, які взяті у моделі (2.57). Накопичення додатних похибок у різницях $y_i - \hat{y}_i$ і привело до того, що $\sum (y_i - \hat{y}_i)$ незначно перевищує нуль. Відмітимо також, що значення $a_1 = 4,88$, $a_0 = 0,72$ приводять до рівності $\sum u_i = 0,008$.

Використавши підсумок останнього рядка, отримаємо:

$$S_u^2 = \frac{1}{10-2} \cdot 1,1544 = 0,1443.$$

Ліва і права межі довірчого інтервалу для σ_u^2 визначаються згідно (2.29) за формулами відповідно $\frac{(n-2)S_u^2}{\chi_2^2}$ і $\frac{(n-2)S_u^2}{\chi_1^2}$, де у відповідності

із (2.26) та (2.27) χ_1^2 та χ_2^2 є коренями рівнянь

$$P(\chi^2(k) > \chi_1^2(p; k)) = p, \quad p = (1 + \gamma)/2 = 0,95,$$

$$P(\chi^2(k) > \chi_2^2(p; k)) = p, \quad p = (1 - \gamma)/2 = 0,05.$$

За табл. 4 додатків для $k = n - 2 = 8$ знайдемо: $\chi_1^2(0,95; 8) = 2,73$ і $\chi_2^2(0,05; 8) = 15,51$. Тоді ліва межа довірчого інтервалу дорівнює

$$\frac{8 \cdot 0,1443}{15,51} = 0,0744, \quad \text{а права} \quad \frac{8 \cdot 0,1443}{2,73} = 0,4229. \quad \text{Тобто остаточно з на-}$$

дійністю 0,9

$$0,0744 < \sigma_u^2 < 0,4229.$$

3) Згідно з п.8, якщо виконується нерівність (2.33): $\left| \frac{a_m}{S_{a_m}} \right| > t_{кр.}$

($m = 0, m = 1$), тоді на рівні значущості α приймається гіпотеза $H_1 : a_m \neq 0$. Значення S_{a_0} та S_{a_1} знайдемо із виразів (2.30):

$$S_{a_0} = \sqrt{\frac{S_u^2 x^2}{n \sigma_x^2}} = \frac{S_u}{\sigma_x} \sqrt{\frac{x^2}{n}} = \frac{\sqrt{0,1443}}{\sqrt{x^2 - (\bar{x})^2}} \sqrt{\frac{x^2}{10}} = \frac{\sqrt{0,1443}}{\sqrt{0,868 - (0,84)^2}} \sqrt{\frac{0,868}{10}} = 0,2777$$

;

$$S_{a_1} = \sqrt{\frac{S_u^2}{n \sigma_x^2}} = \frac{S_u}{\sigma_x \sqrt{n}} = \frac{\sqrt{0,1443}}{\sqrt{10[0,868 - (0,84)^2]}} = 0,2981.$$

Тоді спостережені значення критерію:

$$\left| \frac{a_0}{S_{a_0}} \right| = \frac{0,717}{0,2777} = 2,582, \quad \left| \frac{a_1}{S_{a_1}} \right| = \frac{4,884}{0,2981} = 16,384.$$

Критична точка для двосторонньої критичної області $t_{кр.} = t_{двосм}(\alpha, k)$ при значеннях $\alpha = 0,05$, $k = n - 2 = 8$ знаходиться за верхньою частиною табл. 3 додатків: $t_{кр.} = 2,306$.

Оскільки $2,582 > t_{кр.} = 2,306$ і $16,384 > t_{кр.} = 2,306$, то на рівні значущості $\alpha = 0,05$ робимо висновки, що $\alpha_0 \neq 0$ і $\alpha_1 \neq 0$.

4) Згідно з (2.39) та (2.40) довірчі інтервали з надійністю γ для невідомих параметрів регресії a_0 та a_1 мають такий вигляд:

$$a_m - t_m(\gamma, k) S_{a_m} < \alpha_m < a_m + t_m(\gamma, k) S_{a_m},$$

де $m = 0, 1$, $t_m = t_m(\gamma, k)$ — корінь рівняння $P(|t_m| < t) = \gamma$, t_0 та t_1 — випадкові величини, розподілені за законом Ст'юдента.

У нашому випадку $\gamma = 0,95$, число ступенів вільності $k = n - 2 = 8$. За табл. 2 додатків знаходимо $t_0(0,95; 8) = t_1(0,95; 8) = 2,306$. Тоді з врахуванням знайдених значень $S_{a_0} = 0,1055$, $S_{a_1} = 0,1132$ отримаємо:

$$0,717 - 2,306 \cdot 0,2777 < \alpha_0 < 0,717 + 2,306 \cdot 0,2777,$$

$$4,884 - 2,306 \cdot 0,2981 < \alpha_1 < 4,884 + 2,306 \cdot 0,2981$$

або остаточно

$$0,0766 < \alpha_0 < 1,3574,$$

$$4,1966 < \alpha_1 < 5,5714.$$

5) Коефіцієнт детермінації R^2 знайдемо за формулою (2.46*):

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

Із табл. 2.2 (останнє число нижнього рядка) $\sum_{i=1}^{10} u_i^2 = \sum_{i=1}^{10} (y_i - \hat{y}_i)^2 = 1,1544$.

Для знаходження σ_y^2 використаємо табл. 2.1:

$$\sigma_y^2 = \overline{y^2} - (\overline{y})^2 = 272,22/10 - (48,2/10)^2 = 1,9896.$$

Отже,

$$R^2 = 1 - \frac{1,1544}{10 \cdot 1,9896} = 0,9711.$$

Таким чином, варіація залежної змінної Y на 97,11% пояснюється варіацією пояснюючої змінної.

Вибірковий коефіцієнт кореляції згідно із (2.47):

$$r = \sqrt{R^2} = \sqrt{0,9711} = 0,9854.$$

При цьому додатний знак цього числа обрано в зв'язку з тим, що $a_1 > 0$.

Обчислимо абсолютну середню відсоткову помилку MAPE за формулою (2.49):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \cdot 100\%.$$

Для цього використаємо другий і четвертий рядки табл. 2.2:

$$\begin{aligned} \sum_{i=1}^{10} \left| \frac{\hat{y}_i - y_i}{y_i} \right| &= \frac{0,1938}{1,5} + \frac{0,7178}{2,9} + \frac{0,059}{3,1} + \frac{0,4474}{3,2} + \frac{0,3242}{4,3} + \frac{0,099}{5,7} + \\ &+ \frac{0,2894}{5,8} + \frac{0,4222}{7} + \frac{0,1338}{7,2} + \frac{0,0546}{7,5} = 0,1292 + 0,2475 + 0,019 + 0,1398 + \\ &+ 0,0754 + 0,0174 + 0,0499 + 0,0603 + 0,0186 + 0,0073 = 0,7644. \end{aligned}$$

Отже, $MAPE = \frac{1}{10} \cdot 0,7644 \cdot 100\% = 7,644\% < 10\%$, тобто відповідає

високій точності прогнозу за моделлю.

Середню відсоткову помилку MPE знайдемо за формулою (2.50):

$$MPE = \frac{1}{n} \sum_{i=1}^n \frac{\hat{y}_i - y_i}{y_i} \cdot 100\%,$$

використавши розрахунки при обчисленні MAPE:

$$\begin{aligned} \sum_{i=1}^{10} \frac{\hat{y}_i - y_i}{y_i} &= 0,1292 - 0,2475 + 0,019 + 0,1398 + +0,0754 - 0,0174 + \\ &+ 0,0499 - 0,0603 - 0,0186 + 0,0073 = 0,0768. \end{aligned}$$

Остаточню

$$MPE = \frac{1}{10} \cdot 0,0768 \cdot 100\% = 0,768\% < 5\%.$$

Висновок: всі знайдені показники вказують на високу якість моделі.

б) Побудова довірчої зони для функції регресії передбачає побудову точок з координатами $\{x_i; \hat{y}_i - t(\gamma, n-2)S_{\hat{y}_i}\}$, $i = \overline{1, n}$, з наступним з'єднанням сусідніх (по індексу i) точок прямолінійними відрізками, а потім здійснення аналогічної процедури для послідовності точок $\{x_i; \hat{y}_i + t(\gamma, n-2)S_{\hat{y}_i}\}$.

Величину $S_{\hat{y}_i}$ знайдемо із формули (2.36) $S_{\hat{y}_i} = S_u \sqrt{\left[1 + \frac{(x_i - \bar{x})^2}{\sigma_x^2}\right] \frac{1}{n}}$.

Використовуючи табл. 2.1 і знайдене значення $S_u = \sqrt{0,1443} = 0,3799$, отримаємо:

$$\sigma_x^2 = \overline{x^2} - (\bar{x})^2 = 0,1624;$$

$$S_{\hat{y}_1} = 0,3799 \sqrt{\left[1 + \frac{(0,2 - 0,84)^2}{0,1624}\right] \frac{1}{10}} = 0,2255;$$

$$S_{\hat{y}_2} = 0,3799 \sqrt{\left[1 + \frac{(0,3 - 0,84)^2}{0,1624}\right] \frac{1}{10}} = 0,201;$$

$$S_{\hat{y}_3} = 0,1572; \quad S_{\hat{y}_4} = 0,1398; \quad S_{\hat{y}_5} = 0,1207; \quad S_{\hat{y}_6} = 0,1293;$$

$$S_{\hat{y}_7} = 0,143; \quad S_{\hat{y}_8} = 0,1611; \quad S_{\hat{y}_9} = 0,1823; \quad S_{\hat{y}_{10}} = 0,2057.$$

За табл. 2 додатків знайдемо $t(0,95;8) = 2,306$. Використовуючи табл. 2.2 і знайдені $S_{\hat{y}_i}$, отримаємо ординати точок нижньої межі довірчої зони:

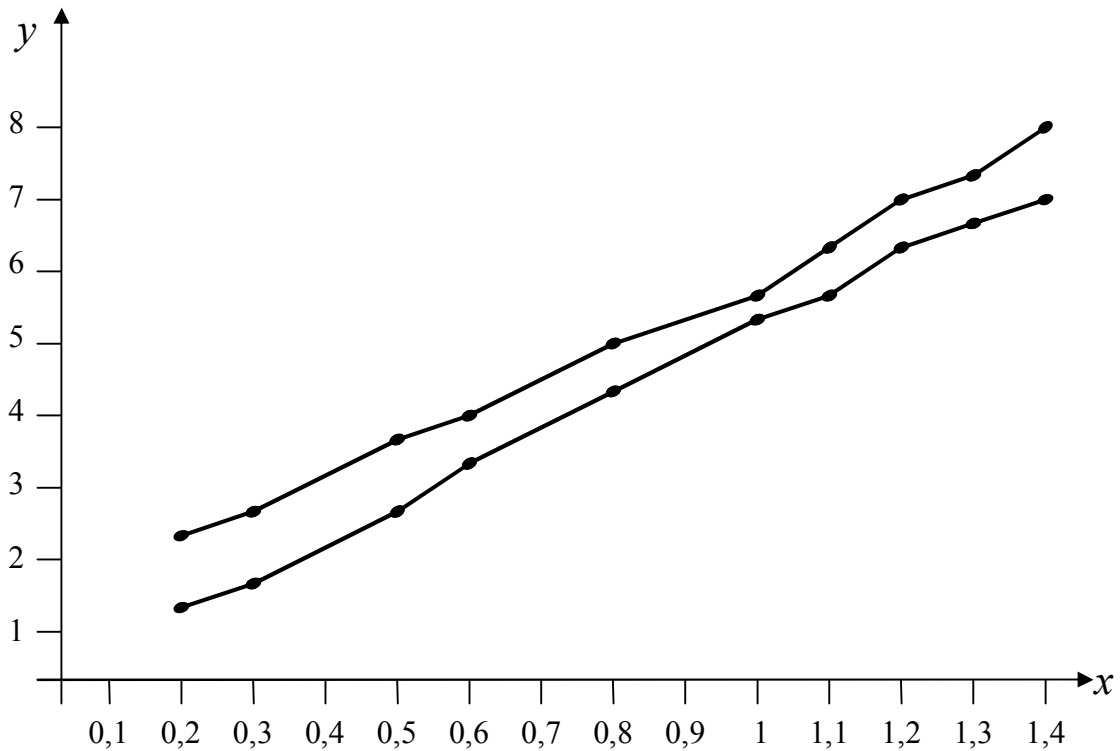
$$\hat{y}_1 - tS_{\hat{y}_1} = 1,6938 - 2,306 \cdot 0,2255 = 1,1738;$$

$$\begin{aligned}
\hat{y}_2 - tS_{\hat{y}_2} &= 2,1822 - 2,306 \cdot 0,201 = 1,7187; \\
\hat{y}_3 - tS_{\hat{y}_3} &= 3,159 - 2,306 \cdot 0,1572 = 2,7965; \\
\hat{y}_4 - tS_{\hat{y}_4} &= 3,6474 - 2,306 \cdot 0,1398 = 3,325; \\
\hat{y}_5 - tS_{\hat{y}_5} &= 4,6242 - 2,306 \cdot 0,1207 = 4,3459; \\
\hat{y}_6 - tS_{\hat{y}_6} &= 5,601 - 2,306 \cdot 0,1293 = 5,3028; \\
\hat{y}_7 - tS_{\hat{y}_7} &= 6,0894 - 2,306 \cdot 0,143 = 5,7596; \\
\hat{y}_8 - tS_{\hat{y}_8} &= 6,5778 - 2,306 \cdot 0,1611 = 6,2063; \\
\hat{y}_9 - tS_{\hat{y}_9} &= 7,0662 - 2,306 \cdot 0,1823 = 6,6458; \\
\hat{y}_{10} - tS_{\hat{y}_{10}} &= 7,5546 - 2,306 \cdot 0,2057 = 7,0803.
\end{aligned}$$

Тоді ординати точок верхньої межі довірчої зони набирають таких значень:

$$\begin{aligned}
\hat{y}_1 + tS_{\hat{y}_1} &= 1,6938 + 2,306 \cdot 0,2255 = 2,2138; \\
\hat{y}_2 + tS_{\hat{y}_2} &= 2,1822 + 2,306 \cdot 0,201 = 2,6457; \\
\hat{y}_3 + tS_{\hat{y}_3} &= 3,159 + 2,306 \cdot 0,1572 = 3,5215; \\
\hat{y}_4 + tS_{\hat{y}_4} &= 3,6474 + 2,306 \cdot 0,1398 = 3,9698; \\
\hat{y}_5 + tS_{\hat{y}_5} &= 4,6242 + 2,306 \cdot 0,1207 = 4,9025; \\
\hat{y}_6 + tS_{\hat{y}_6} &= 5,601 + 2,306 \cdot 0,1293 = 5,8992; \\
\hat{y}_7 + tS_{\hat{y}_7} &= 6,0894 + 2,306 \cdot 0,143 = 6,4192; \\
\hat{y}_8 + tS_{\hat{y}_8} &= 6,5778 + 2,306 \cdot 0,1611 = 6,9493; \\
\hat{y}_9 + tS_{\hat{y}_9} &= 7,0662 + 2,306 \cdot 0,1823 = 7,4866; \\
\hat{y}_{10} + tS_{\hat{y}_{10}} &= 7,5546 + 2,306 \cdot 0,2057 = 8,0289.
\end{aligned}$$

Довірча зона (з надійністю 0,95) для функції регресії зображена на мал. 2.2.



Малюнок 2.2.

7) Прогнозне значення річного товарообігу для нової філії із торгівельною площею 1,8 тис. м² знайдемо із рівняння (2.57):

$$\hat{y}_{n+1} = 0,717 + 4,884 \cdot 1,8 = 9,508.$$

Довірчий інтервал для прогнозного значення y_{n+1} із надійністю $\gamma = 0,95$ визначається (2.56). З допомогою виразу (2.55) знайдемо

$$S_{u_{n+1}} = S_u \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{n\sigma_x^2}} = 0,3799 \sqrt{1 + \frac{1}{10} + \frac{(1,8 - 0,84)^2}{10 \cdot 0,1624}} = 0,4906.$$

Тоді шуканий довірчий інтервал має вид

$$9,508 - 2,306 \cdot 0,4906 < y_{n+1} < 9,508 + 2,306 \cdot 0,4906$$

або остаточно

$$8,3767 < y_{n+1} < 10,6393.$$



ІНДИВІДУАЛЬНЕ ЗАВДАННЯ №1

Торгівельне підприємство має велику кількість філій і керівництво цього підприємства вивчає питання про залежність Y (річний товарообіг однієї філії, млн. грн.) від x (торгівельної площі, тис. м²). Для десяти філій за певний рік зафіксовані такі значення показників Y і x :

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
y_i	2,8	5,2	6,8	7,1	7,3	8,3	4,3	5,8	7,7	3,2	1,5	3,7	5,4	2,6
x_i	0,3	0,9	1,2	1,3	1,2	0,8	0,8	0,9	1,3	0,5	0,3	0,6	1,1	0,2
i	15	16	17	18	19	20	21	22	23	24	25	26	27	28
y_i	5,8	8,1	7,5	8,4	4,2	5,6	7,6	3,5	1,4	3,9	6,4	7,3	7,6	8,3
x_i	0,8	1,6	1,5	0,9	0,9	0,8	1,4	0,7	0,4	0,8	1,1	1,4	1,5	0,9
i	29	30	31	32	33	34	35	36	37	38	39	40	41	42
y_i	4,4	5,6	7,5	3,2	1,5	3,5	2,7	4,3	6,9	7,1	7,3	8,4	4,2	5,8
x_i	0,8	0,7	1,4	0,6	0,3	0,7	0,2	0,6	1,4	1,3	1,2	0,9	0,7	0,7
i	43	44	45	46	47	48	49	50						
y_i	7,5	3,5	1,4	4,2	2,6	5,8	4,2	3,1						
x_i	1,5	0,8	0,3	0,9	0,4	0,7	0,8	0,6						

На обсяг товарообігу впливають такі чинники: середньоденна інтенсивність потоку покупців, об'єм основних фондів, їх структура, середньоспискова чисельність працівників, площа підсобних приміщень тощо. Припускається, що в досліджуваній групі філій значення цих чинників приблизно однакові, тому вплив відмінностей їх значень на зміну обсягу товарообігу є незначним.

Потрібно:

1. знайти статистичні оцінки параметрів лінійного рівняння регресії;
2. точкову оцінку та довірчий інтервал дисперсії збурень із надійністю $\gamma = 0,9$;
3. для рівня значущості $\alpha = 0,05$ перевірити значущість коефіцієнтів регресії α_0 та α_1 ;
4. знайти довірчі інтервали коефіцієнтів регресії з надійністю $\gamma = 0,95$;
5. знайти вибіркові коефіцієнт детермінації, коефіцієнт кореляції, а також інші показники якості лінійної регресії (МАРЕ, МРЕ);
6. знайти та побудувати довірчу зону функції регресії з надійністю $\gamma = 0,95$;
7. знайти прогнозне значення річного товарообігу для нової філії, торговельна площа якої складає 1,8 тис. м², а також із надійністю $\gamma = 0,95$ побудувати довірчий інтервал для цього прогнозного значення.

§3. ПЕРЕВІРКА ВИКОНАННЯ ПЕРЕДУМОВ КЛАСИЧНОЇ НОРМАЛЬНОЇ ЛІНІЙНОЇ МОДЕЛІ ПАРНОЇ РЕГРЕСІЇ.

1. *Відповідність вибірки нормальному розподілу.*
2. *Гетероскедастичність та її наслідки.*
3. *Діагностування гетероскедастичності та її вилучення.*
 - 3.1. *Тест рангової кореляції Спірмена.*
 - 3.2. *Тест Голдфелда-Квандта.*
 - 3.3. *Тест Уайта.*
 - 3.4. *Тест Глейзера.*
 - 3.5. *Усунення гетероскедастичності.*
4. *Автокореляція залишків часового ряду.*
5. *Авторегресія першого порядку. Статистика Дарбіна-Уотсона.*
6. *Тести на наявність автокореляції.*

При розв'язуванні задачі §2 було зроблено висновок про високу якість побудованої лінійної моделі парної регресії. Проте суттєвою обставиною є те, що всі оцінки і висновки отримані при умові, що виконуються передумови 1-4. Тому для достовірності прийнятих рішень необхідно перевірити виконання цих передумов. При цьому слід мати на увазі, що перевірки при МНК-оцінюванні підлягають лише передумови 2, 3, 4 (див. зауваження до рівності (2.21)). У випадку констатації невиконання хоча б однієї передумови модель класичної парної регресії перетворюється на **економетричну модель** і актуальними стають шляхи її дослідження.

1. Розпочнемо із перевірки виконання передумови 4 про нормальність розподілу випадкового збурення U .

Якщо обсяг вибірки є дуже великим, тоді згідно з центральною граничною теоремою Ляпунова є підстави стверджувати виконання цієї передумови. У випадку малих вибірок ($n \leq 30$) вже неправомірно робити такий висновок. Проблема також полягає у тому, що випадкова величина U є **неспостережуваною**. Для отримання інформації про можливі значення цієї величини будемо виходити з двох обставин:

- 1) лінійності моделі (2.9)

$$y_i = a_0 + a_1 x_i + u_i, \quad i = \overline{1, n};$$

- 2) отримання оцінок коефіцієнтів регресії a_0, a_1 з допомогою МНК.

Тоді

$$u_i = y_i - \hat{y}_i = y_i - a_0 - a_1 x_i, \quad i = \overline{1, n} \quad (3.1)$$

можна тлумачити як «спостережені» значення збурення U . Наявність лапок в слові «спостережені» зумовлене врахуванням двох попередніх обставин (лінійна модель може виявитись не найкращою, а оцінки a_0 та a_1 можуть отримуватися з допомогою інших методів).

Отже, будемо вважати, що n чисел u_1, u_2, \dots, u_n , визначені (3.1), відомі. Вони є можливими (спостереженими) значеннями випадкової величини U . Ставиться задача про перевірку статистичної гіпотези $H_0: U$ розподілена за нормальним законом ($U \sim N(0, \sigma_u)$). Якщо обсяг вибірки великий (див. [4]), то можна використати критерії узгодженості Пірсона або Колмогорова. Проте в багатьох випадках обсяги вибірки є малими, тому використаємо критерій Фішера.

В якості основних характеристик розподілу зручніше всього брати коефіцієнт асиметрії і ексцес:

$$A = \frac{\mu_3}{\sigma^3}, \quad E = \frac{\mu_4}{\sigma^4} - 3,$$

де μ_3 та μ_4 — центральні моменти третього та четвертого порядків відповідно. Для випадку нормального розподілу $\mu_3 = 0$, $\mu_4 = 3\sigma^4$. Тому для цього розподілу виконуються рівності

$$A = 0, \quad E = 0. \quad (3.2)$$

Відповідні вибіркові (емпіричні) коефіцієнти асиметрії і ексцесу визначаються формулами:

$$A^* = \frac{\mu_3^*}{\sigma_u^3}, \quad E^* = \frac{\mu_4^*}{\sigma_u^4} - 3, \quad (3.3)$$

де центральний емпіричний момент m -го порядку μ_m^* визначається за формулою (для даного випадку позначень):

$$\mu_m^* = \frac{1}{n} \sum_{i=1}^n (u_i - \bar{u})^m, \quad m = 3, 4. \quad (3.4)$$

Емпіричні асиметрія і ексцес, як і всі числові характеристики нефіксованої вибірки, є випадковими величинами і тому навіть для нормального розподілу можуть відрізнятися від нуля. Закони розподілу A^* і E^* дуже складні і мало вивчені. Фішер запропонував модифікацію оцінок коефіцієнта асиметрії A^* та ексцесу E^* :

$$\hat{A}^* = \frac{\mu_3^*}{\sigma_u^3} \sqrt{\frac{n(n-1)}{(n-2)^2}}, \quad \hat{E}^* = \frac{n^2 - 1}{(n-2)(n-3)} \left(\frac{\mu_4^*}{\sigma_u^4} - 3 \frac{n-1}{n+1} \right). \quad (3.5)$$

При невеликих обсягах вибірок \hat{A}^* та \hat{E}^* помітно відрізняються від A^* та E^* . Виявляється, що у випадку нормального розподілу оцінки \hat{A}^* та \hat{E}^* мають з великим ступенем точності нормальні випадкові розподіли, причому $M(\hat{A}^*)=0$, $M(\hat{E}^*)=0$, а дисперсії визначаються виразами:

$$\sigma_{\hat{A}^*}^2 = \frac{6n(n-1)}{(n+3)(n+1)(n-2)}, \quad \sigma_{\hat{E}^*}^2 = \frac{24n(n-1)^2}{(n+5)(n+3)(n-2)(n-3)}. \quad (3.6)$$

Отже, задача полягає у відповіді на питання: чи значуще оцінки \hat{A}^* і \hat{E}^* відрізняються від своїх математичних сподівань, тобто від нуля?

На практиці можна користуватися таким наближеним критерієм згоди:

$$|\hat{A}^*| \leq 2\sigma_{\hat{A}^*}, \quad |\hat{E}^*| \leq 2\sigma_{\hat{E}^*}. \quad (3.7)$$

Задача 3.1. На основі статистичних даних задачі 2.1 здійснити перевірку виконання передумови 4 про нормальність розподілу випадкової величини U .

○ Для знаходження \hat{A}^* і \hat{E}^* згідно з формулами (3.5) потрібно обчислити μ_3 , μ_4 , σ_u^3 , σ_u^4 . Згідно з останнім рядком табл. 2.2 $\sum_{i=1}^{10} u_i^2 = 1,1544$. Тому

$$\sigma_u^2 = \overline{u^2} - (\bar{u})^2 = \frac{1}{10} \cdot 1,1544 - 0^2 = 0,11544, \quad \sigma_u = 0,33976, \quad \sigma_u^3 = 0,0392208,$$

$\sigma_u^4 = 0,0133263$. Значення μ_3^* та μ_4^* знайдемо за формулами (3.4). Для цього складемо розрахункову табл. 3.1, дані першого стовпця якої взяті з четвертого рядка табл. 2.2.

Таблиця 3.1

i	u_i	u_i^3	u_i^4
1	-0,1938	-0,0072788	0,0014106
2	0,7178	0,369837	0,265469
3	-0,059	-0,0002054	0,0000121
4	-0,4474	-0,0895546	0,0400667
5	-0,3242	-0,0340752	0,0110472
6	0,099	0,0009703	0,0000961
7	-0,2894	-0,0242379	0,0070145
8	0,4222	0,0752583	0,0317741
9	0,1338	0,0023953	0,0003205
10	-0,0546	-0,0001628	0,0000089
Σ	0,0044	0,2929462	0,3572197

Враховавши, що $\bar{u} = 0$, отримаємо:

$$\mu_3^* = \frac{1}{n} \sum_{i=1}^{10} u_i^3 = \frac{1}{10} \cdot 0,2929463 = 0,0292946,$$

$$\mu_4^* = \frac{1}{n} \sum_{i=1}^{10} u_i^4 = \frac{1}{10} \cdot 0,357219 = 0,0357220.$$

За формулами (3.5), (3.6) знаходимо:

$$\hat{A}^* = \frac{0,0292946}{0,0392208} \sqrt{\frac{10 \cdot 9}{8^2}} = 0,8857321,$$

$$\hat{E}^* = \frac{10^2 - 1}{8 \cdot 7} \left(\frac{0,0357219}{0,0133263} - 3 \cdot \frac{9}{11} \right) = 0,3997819,$$

$$\sigma_{\hat{A}^*}^2 = \frac{6 \cdot 10 \cdot 9}{13 \cdot 11 \cdot 8} = 0,4720280, \quad \sigma_{\hat{A}^*} = 0,6870429,$$

$$\sigma_{\hat{E}^*}^2 = \frac{24 \cdot 10 \cdot 9^2}{15 \cdot 13 \cdot 8 \cdot 7} = 1,7802198, \quad \sigma_{\hat{E}^*} = 1,3342488.$$

Обидві нерівності (3.7) виконуються:

$$0,8857321 < 2 \cdot 0,6870429, \quad 0,3997819 < 2 \cdot 1,3342488,$$

а тому згідно з цим критерієм згоди гіпотеза H_0 про нормальність закону розподілу U приймається. ◎

2. Друга передумова (гомоскедастичність або «однаковий розкид») передбачає виконання n рівностей

$$D(U_i) = \sigma_u^2 = const, \quad i = \overline{1, n}. \quad (3.8)$$

Якщо хоча б одна з цих рівностей не виконується, тобто

$$D(U_i) = f(x_i), \quad i = \overline{1, n}, \quad (3.9)$$

тоді має місце гетероскедастичність («неоднаковий розкид»).

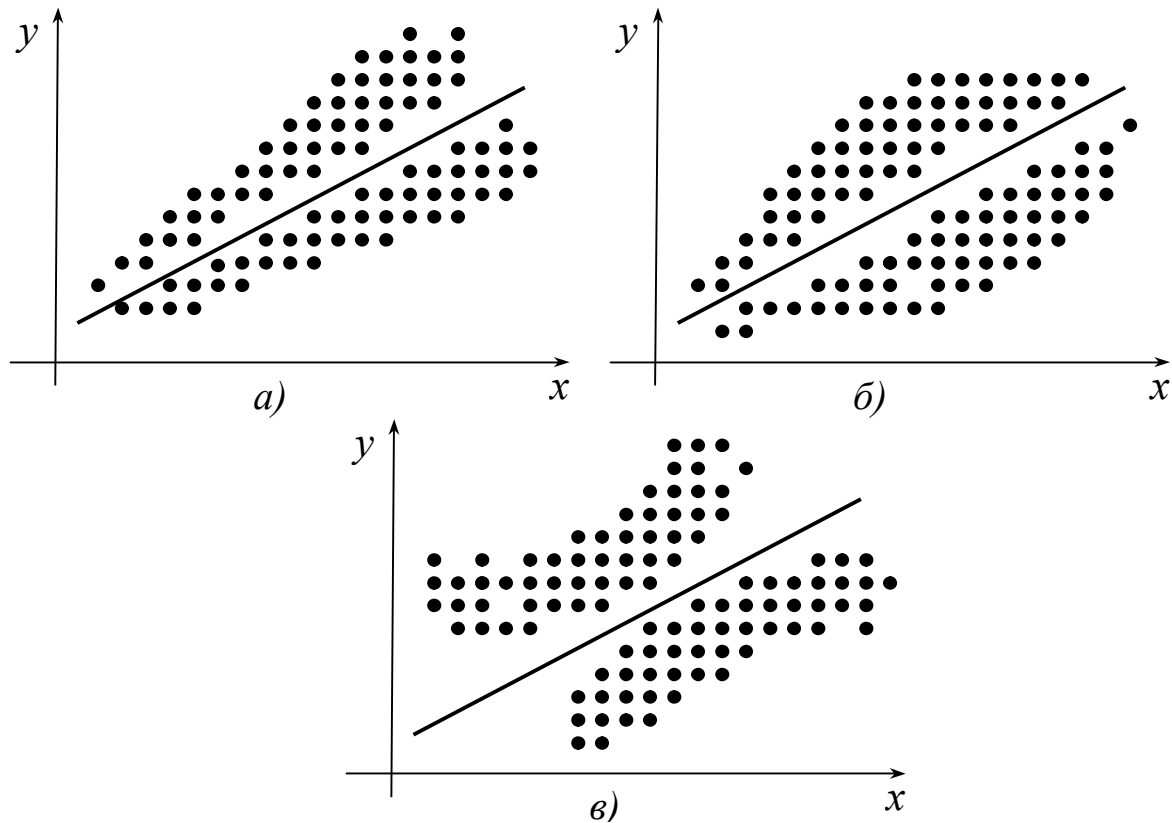
Появу проблеми гетероскедастичності часто можна передбачити заздалегідь, ґрунтуючись на значенні характеру даних. Припущення про гомоскедастичність виправдане в тих випадках, коли досліджувані об'єкти є достатньо **однорідними**. Якщо ж досліджуються неоднорідні об'єкти, то, як правило, виникає проблема гетероскедастичності.

Приклади:

1) Якщо вивчається залежність прибутку фірми від розміру основних фондів, то природно очікувати, що для великих фірм коливання прибутку буде вищим, ніж для малих.

2) Якщо досліджується залежність витрат на харчування в родині від загального доходу, то розкид у даних буде більшим для родин із більш високим доходом.

Графічна форма розкиду спостережень залежить від форми зв'язку між $\sigma_{u_i}^2$ та x_i . Приклади зображені на мал. 3.1.



Малюнок 3.1.

У прикладних дослідженнях, як правило, використовується зручне припущення, що $f(x_i) = k^2 x_i^2$, де k — стала, яку потрібно оцінити.

Розглянемо наслідки порушення передумови про гомоскедастичність. Можна довести, що оцінки коефіцієнтів регресії залишаються лінійними і незміщеними, але вже не володіють властивістю ефективності, тобто їх дисперсії вже не будуть мінімальними в класі лінійних оцінок. У зв'язку з цим розширюються довірчі інтервали. Як наслідок, тести Ст'юдента і Фішера-Снедекора дають неточні результати. Крім того, формулу для оцінки σ_u , строго кажучи, застосовувати вже не можна.

Отже, гетероскедастичність є серйозною проблемою. Досліднику потрібно знати: є вона чи немає. У випадку тестування гетероскедастичності вихідну модель необхідно модифікувати.

3. Виявляється, єдиних правил діагностування гетероскедастичності немає, а є різні тести з своїми недоліками та перевагами. Розглянемо найпростіші з них за змістом та розрахунками. В кожному тесті в якості нульової гіпотези розглядається H_0 — гіпотеза про відсутність гетероскедастичності.

3.1. Тест рангової кореляції Спірмена передбачає найбільш загальні припущення про залежність дисперсій помилок регресії від значень незалежної змінної:

$$\sigma_{u_i}^2 = f_i(x_i), \quad i = \overline{1, n}. \quad (3.10)$$

При цьому ніяких додаткових припущень відносно виду функцій f_i не робиться. Відмітимо також, що **відсутнє обмеження стосовно закону розподілу помилок**.

Ідея тесту полягає в тому, що абсолютні величини залишків регресії $|u_i|$ розглядаються як оцінки σ_{u_i} , тому при наявності гетероскедастичності $|u_i|$ і значення x_i будуть корелювати. Проте кореляція в цьому випадку передбачається **ранговою**.

Рангова кореляція досліджується тоді, коли необхідно встановити силу зв'язку між **ординальними (порядковими) змінними**. Прикладами ординальних змінних є житлові умови, тестові бали, екзаменаційні оцінки. Джерелом ординальних змінних можуть бути і **кількісні** змінні, для яких здійснюється процес ранжування. Наприклад, кожному з двох множин чисел $\{|u_1|, |u_2|, \dots, |u_n|\}$, $\{x_1, x_2, \dots, x_n\}$ можна ранжувати в порядку зростання. В результаті i -тий об'єкт характеризується двома рангами s_i та l_i по змінних U та x . Тоді **коефіцієнт рангової кореляції Спірмена** знаходиться за формулою

$$\rho = 1 - \frac{6 \sum_{i=1}^n (s_i - l_i)^2}{n^3 - n}. \quad (3.11)$$

Якщо ранги всіх об'єктів рівні між собою, тобто $s_i = l_i \quad \forall \quad i = \overline{1, n}$, то $\rho = 1$. Цей випадок називається **повним прямим зв'язком**. При **повному оберненому зв'язку**, коли ранги об'єктів по обох змінних розташовані в оберненому порядку, можна довести, що $\rho = -1$. У решті випадків $|\rho| < 1$.

При перевірці значущості ρ виходять із того, що у випадку правильності нульової гіпотези (про відсутність кореляційного зв'язку між змінними) при $n > 10$ статистика

$$t = \frac{\rho\sqrt{n-2}}{\sqrt{1-\rho^2}} \quad (3.12)$$

має t -розподіл Ст'юдента із $n-2$ ступенями вільності. Тому ρ значущий на рівні α , якщо

$$|t_{\text{спост.}}| > t_{\text{двост. кр.}}(\alpha; n-2). \quad (3.13)$$

Задача 3.2. На основі статистичних даних задачі 2.1 на рівні $\alpha = 0,05$ перевірити виконання передумови 2 з допомогою тесту рангової кореляції Спірмена.

○ Здійснимо ранжування змінних $|u_i|$ та x_i . Для цього складемо табл. 3.2, використавши дані табл. 2.1 (другий стовпець) та табл. 2.2 (передостанній рядок). При заповненні останнього стовпця вибирається найменше число в третьому стовпці і поруч з ним записується 1. Найменшому з тих чисел, що залишилися, відповідає 2, і т.д.

Таблиця 3.2

x_i	l_i (ранг x_i)	$ u_i $	s_i (ранг $ u_i $)
0,2	1	0,1938	5
0,3	2	0,7178	10
0,5	3	0,059	2
0,6	4	0,4474	9
0,8	5	0,3242	7
1	6	0,099	3
1,1	7	0,2894	6
1,2	8	0,4222	8
1,3	9	0,1338	4
1,4	10	0,0546	1

За формулою (3.11) знайдемо коефіцієнт рангової кореляції Спірмена:

$$\begin{aligned} \rho &= 1 - \frac{6}{10^3 - 10} \left[(5-1)^2 + (10-2)^2 + (2-3)^2 + (9-4)^2 + (7-5)^2 + \right. \\ &\quad \left. + (3-6)^2 + (6-7)^2 + (8-8)^2 + (4-9)^2 + (1-10)^2 \right] = \\ &= 1 - 0,00606(16 + 64 + 1 + 25 + 4 + 9 + 1 + 25 + 81) = -0,3696. \end{aligned}$$

Згідно із (3.12)

$$|t_{\text{спост.}}| = \frac{|-0,3696|\sqrt{10-2}}{\sqrt{1-(-0,3696)^2}} = 1,1251.$$

За табл. 3 додатків $t_{\text{двост. кр.}}(0,05;8) = 2,306$.

Оскільки на рівні $\alpha = 0,05$ нерівність (3.13) не виконується, то нульова гіпотеза про відсутність кореляційного зв'язку є правильною. Отже, на цьому ж рівні приймається і гіпотеза H_0 про відсутність гетероскедастичності (виконання передумови 2). \odot

3.2. Тест Голдфелда-Квандта використовується у тому випадку, коли помилки регресії U_i можна вважати **нормально розподіленими** випадковими величинами. При цьому спостережень має бути хоча б удвічі більше, ніж число оцінюваних параметрів. Як правило, тест застосовується до великих вибірок.

Припустимо, що середні квадратичні відхилення збурень σ_{u_i} пропорційні значенням пояснюючої змінної x . Це означає постійність **відносного**, а не **абсолютного**, як у класичній моделі, розкиду збурень регресійної моделі.

Впорядкуємо n спостережень в порядку зростання значень x_i і виберемо m перших і m останніх спостережень (число m визначимо пізніше). Тоді гіпотеза про гомоскедастичність буде рівносильна тому, що значення u_1, u_2, \dots, u_m та $u_{n-m+1}, u_{n-m+2}, \dots, u_n$ є вибірковими спостереженнями **нормально розподілених** випадкових величин, які мають **однакові дисперсії**.

Зауваження. Для знаходження $u_i = y_i - \hat{y}_i$ для двох груп ($i = \overline{1, m}$ та $i = \overline{n-m+1, n}$) необхідно попередньо знайти два емпіричні рівняння для кожної з груп.

Гіпотеза про рівність дисперсій двох нормально розподілених сукупностей, як відомо [4], перевіряється з допомогою критерія Фішера-Снедекора. Нульова гіпотеза про рівність дисперсій двох сукупностей по m спостережень (тобто гіпотеза про відсутність гетероскедастичності) відкидається на рівні α , якщо

$$F_{\text{спост.}} = \frac{\sum_{i=1}^m u_i^2}{\sum_{i=n-m+1}^n u_i^2} > F_{\text{кр.}}(\alpha; m-1; m-1). \quad (3.14)$$

Відмітимо, що чисельник і знаменник в (3.14) слід було розділити на відповідне число ступенів вільності, проте в даному випадку ці числа однакові і рівні $m - 1$.

Виявляється, що коли вибрати m порядку $n/3$, тоді **потужність тесту**, тобто імовірність відкинути гіпотезу про наявність гомоскедастичності, коли насправді гетероскедастичності немає, буде **максимальною**.

Задача 3.3. На рівні значущості $\alpha = 0,05$ для задачі 2.1 перевірити виконання передумови 2 (ігноруючи малість вибірки) з допомогою тесту Голдфелда-Квандта.

○ За аналогією із розв'язуванням задачі 2.1. (п.1) знайдемо статистичні оцінки параметрів $a_1^{(1)}$, $a_0^{(1)}$ та $a_1^{(2)}$, $a_0^{(2)}$, виходячи з двох груп даних ($m = 4$):

i	1	2	3	4
y_i	1,5	2,9	3,1	3,2
x_i	0,2	0,3	0,5	0,6

i	7	8	9	10
y_i	5,8	7	7,2	7,5
x_i	1,1	1,2	1,3	1,4

$$\bar{x}^{(1)} = (0,2 + 0,3 + 0,5 + 0,6)/4 = 0,4;$$

$$\bar{x}^{(2)} = (1,1 + 1,2 + 1,3 + 1,4)/4 = 1,25;$$

$$\bar{x}^{2(1)} = (0,2^2 + 0,3^2 + 0,5^2 + 0,6^2)/4 = 0,185;$$

$$\bar{x}^{2(2)} = (1,1^2 + 1,2^2 + 1,3^2 + 1,4^2)/4 = 1,575;$$

$$\bar{y}^{(1)} = (1,5 + 2,9 + 3,1 + 3,2)/4 = 2,675;$$

$$\bar{y}^{(2)} = (5,8 + 7 + 7,2 + 7,5)/4 = 6,875;$$

$$\overline{xy}^{(1)} = (1,5 \cdot 0,2 + 2,9 \cdot 0,3 + 3,1 \cdot 0,5 + 3,2 \cdot 0,6)/4 = 1,16;$$

$$\overline{xy}^{(2)} = (5,8 \cdot 1,1 + 7 \cdot 1,2 + 7,2 \cdot 1,3 + 7,5 \cdot 1,4)/4 = 8,66;$$

$$a_1^{(1)} = \frac{\overline{xy}^{(1)} - \bar{x}^{(1)} \cdot \bar{y}^{(1)}}{\bar{x}^{2(1)} - (\bar{x}^{(1)})^2} = \frac{1,16 - 0,4 \cdot 2,675}{0,185 - (0,4)^2} = 3,6;$$

$$a_0^{(1)} = \bar{y}^{(1)} - a_1^{(1)} \bar{x}^{(1)} = 2,675 - 3,6 \cdot 0,4 = 1,235;$$

$$a_1^{(2)} = \frac{\overline{xy}^{(2)} - \bar{x}^{(2)} \cdot \bar{y}^{(2)}}{\bar{x}^{2(2)} - (\bar{x}^{(2)})^2} = \frac{8,66 - 1,25 \cdot 6,875}{1,575 - (1,25)^2} = 5,3;$$

$$a_0^{(2)} = \bar{y}^{(2)} - a_1^{(2)} \bar{x}^{(2)} = 6,875 - 5,3 \cdot 1,25 = 0,25.$$

Отже, емпіричне рівняння для першої групи має такий вигляд:

$$\hat{y}^{(1)} = 1,235 + 3,6x,$$

а другої

$$\hat{y}^{(2)} = 0,25 + 5,3x.$$

Тоді

$$\begin{aligned} \sum_{i=1}^4 u_i^2 &= \sum_{i=1}^4 (y_i - \hat{y}_i^{(1)})^2 = (1,5 - 1,235 - 3,6 \cdot 0,2)^2 + (2,9 - 1,235 - 3,6 \cdot 0,3)^2 + \\ &+ (3,1 - 1,235 - 3,6 \cdot 0,5)^2 + (3,2 - 1,235 - 3,6 \cdot 0,6)^2 = \\ &= 0,207 + 0,3422 + 0,0042 + 0,038 = 0,5914; \end{aligned}$$

$$\begin{aligned} \sum_{i=7}^{10} u_i^2 &= \sum_{i=7}^{10} (y_i - \hat{y}_i^{(2)})^2 = (5,8 - 0,25 - 5,3 \cdot 1,1)^2 + (7 - 0,25 - 5,3 \cdot 1,2)^2 + \\ &+ (7,2 - 0,25 - 5,3 \cdot 1,3)^2 + (7,5 - 0,25 - 5,3 \cdot 1,4)^2 = \\ &= 0,0784 + 0,1521 + 0,0036 + 0,0289 = 0,263; \end{aligned}$$

$$F_{\text{спост.}} = \frac{\sum_{i=1}^4 u_i^2}{\sum_{i=7}^{10} u_i^2} = \frac{0,5914}{0,263} = 2,2487.$$

Врахувавши, що $m = 4$, $\alpha = 0,05$, за табл. 5 додатків знайдемо $F_{\text{кр.}}(0,05; 3; 3) = 9,28$.

Оскільки $F_{\text{спост.}} = 2,2487 < F_{\text{кр.}}(0,05; 3; 3) = 9,28$, тобто нерівність (3.14) не виконується, то робимо висновок, що гіпотеза H_0 про відсутність гетероскедастичності на рівні $\alpha = 0,05$ приймається.

Отже, за тестом Голдфелда-Квандта на рівні $\alpha = 0,05$ для статистичних даних задачі 2.1. передумова 2 виконується. ◎

3.3. Тест рангової кореляції Спірмена і тест Голдфелда-Квандта дозволяють лише **виявити наявність гетероскедастичності**, але вони не дають можливості з'ясувати кількісний характер залежності дисперсій помилок регресії від значень незалежної змінної, і, отже, не дають методів усунення гетероскедастичності.

Для досягнення цієї мети необхідні деякі додаткові припущення стосовно характеру гетероскедастичності. Справді, без цих припущень, очевидно, неможливо було б оцінити n дисперсій помилок регресії $\sigma_{u_i}^2$ ($i = \overline{1, n}$) з допомогою n спостережень.

Найбільш простий і часто використовуваний тест на гетероскедастичність — **тест Уайта**. При його використанні припускається, що дисперсії помилок регресії є **однією і тією ж функцією** від спостережених значень незалежної змінної, тобто рівняння (3.10) набирають такого виду:

$$\sigma_{u_i}^2 = f(x_i), \quad i = \overline{1, n}. \quad (3.15)$$

Найчастіше функція f обирається квадратичною:

$$f(x_i) = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2, \quad (3.16)$$

що відповідає тому, що σ_{u_i} залежить від x_i приблизно лінійно. У випадку гомоскедастичності $f = const$, тобто вибіркові коефіцієнти регресії a_1, a_2 , які є оцінками невідомих чисел α_1, α_2 відповідно, незначуще відрізняються від нуля.

Ідея тесту Уайта полягає в оцінці функції f з (3.15) за допомогою відповідного рівняння регресії для квадратів залишків:

$$\overline{u_i^2} = f(x_i) + \varepsilon_i, \quad i = \overline{1, n}, \quad (3.17)$$

де ε_i випадкова величина (за аналогією з U_i із рівняння (2.3)).

Відмітимо, що ліві частини рівнянь (3.15) та (3.17) співпадають, оскільки $\overline{u} = 0$ (див. 2.21)).

Гіпотеза про відсутність гетероскедастичності (умова $f = const$) приймається у випадку незначущості регресії (3.17) в цілому (тобто одночасної незначущості теоретичних коефіцієнтів регресії α_1 та α_2).

Якщо обрати функцію f у вигляді (3.16), тоді знаходити «вручну» оцінки a_0, a_1, a_2 , а також їх середні квадратичні відхилення — достатньо працемісткий процес. Оптимальний шлях — використання персонального комп'ютера із відповідним програмним забезпеченням.

3.4. Тест Глейзера аналогічний тесту Уайта, тільки в якості залежної змінної для вивчення гетероскедастичності вибирається не квадрат залишків, а їх абсолютна величина, тобто розглядається регресія

$$|u_i| = f(x_i) + \varepsilon_i, \quad i = \overline{1, n}. \quad (3.18)$$

В якості функції f зазвичай обирається функція виду

$$f(x) = \beta_0 + \beta_1 x^\delta, \quad (3.19)$$

Регресія (3.18) вивчається при різних значеннях δ , а потім вибирається те конкретне значення, при якому коефіцієнт β_1 виявляється найбільш

значущим, тобто має найбільше значення t -статистики. При цьому в якості значень δ беруться числа: 1, 2, 3, 1/2, 1/3 тощо. Якщо ж β_1 незначущий для всіх розглянутих значень δ (випадок $f \equiv const$), тоді робиться висновок про відсутність гетероскедастичності.

Задача 3.4. На рівні значущості $\alpha = 0,05$ для задачі 2.1 перевірити виконання передумови 2 за тестом Глейзера.

○ Покладемо у функції (3.19) $\delta = 1$. Нехай b_0 і b_1 — оцінки невідомих параметрів β_0 і β_1 відповідно. Тоді оцінкою моделі (3.18) за статистичними даними $\{|u_i|, x_i\}_{i=1, \overline{n}}$ ($|u_i|$ знайдені, а, отже, відомі) є n рівнянь

$$|u_i| = b_0 + b_1 x_i + \hat{\varepsilon}_i, \quad i = \overline{1, n},$$

де $b_0 + b_1 x_i = |\hat{u}_i|$, $\hat{\varepsilon}_i$ — вибіркова оцінка збурення ε_i .

За формулами (2.12), (2.13) знайдемо МНК-оцінки b_0 та b_1 , використавши дані табл. 2.1 і 2.2:

$$b_1 = \frac{\overline{x|u|} - \bar{x}\bar{|u|}}{x^2 - (\bar{x})^2} = \frac{0,198576 - 0,84 \cdot 0,27412}{0,868 - (0,84)^2} = -0,1951,$$

$$b_0 = \bar{|u|} - b_1 \bar{x} = 0,27412 - (-0,1951) \cdot 0,84 = 0,438.$$

Незміщену точкову оцінку S_ε^2 невідомої дисперсії збурень σ_ε^2 знайдемо за аналогом формули (2.22):

$$S_\varepsilon^2 = \frac{1}{n-2} \sum_{i=1}^n (|u_i| - |\hat{u}_i|)^2,$$

попередньо обчисливши

$$|\hat{u}_1| = |0,438 - 0,1951 \cdot 0,2| = 0,39899,$$

$$|\hat{u}_2| = |0,438 - 0,1951 \cdot 0,3| = 0,37948,$$

.....

$$|\hat{u}_{10}| = |0,438 - 0,1951 \cdot 1,3| = 0,16486.$$

Отже, $S_\varepsilon^2 = 0,0426$, $S_\varepsilon = 0,2064$.

Для визначення значущості β_1 знайдемо S_{b_1} за формулою (2.30):

$$S_{b_1} = \sqrt{\frac{S_\varepsilon^2}{n\sigma_x^2}} = \frac{S_\varepsilon}{\sqrt{n\sigma_x}} = \frac{0,2064}{\sqrt{10}\sqrt{0,868 - (0,84)^2}} = 0,162.$$

Тоді
$$\left| \frac{b_1}{S_{b_1}} \right| = \frac{0,1951}{0,162} = 1,2043.$$

Але оскільки $\left| \frac{b_1}{S_{b_1}} \right| = 1,2043 < t_{\text{довостор.}}(0,05;8) = 2,306$, то на рівні $\alpha = 0,05$

робимо висновок, що $\beta_1 = 0$. Таким чином, у випадку $\delta = 1$

$$f(x) = \beta_0 \equiv \text{const.}$$

При розгляді випадків $\delta = 2, 3, 1/2, 1/3$ доцільно досліджувати рівняння

$$|u_i| = \beta_0 + \beta_1 z_i + \varepsilon_i, \quad i = \overline{1, n},$$

де $z_i = x_i^\delta$, за вище розглянутою схемою. Відповідні чисельні розрахунки для цих випадків наведені в §5.

Виявляється, що і в цих випадках коефіцієнт регресії β_1 незначущий, тобто для статистичних даних задачі 2.1 передумова 2 виконується. ©

Зауваження. Припустимо, що значення t -статистики $|b_1/S_{b_1}|$ при $\delta = 1$ більше $t_{\text{довостор.}}(\alpha; n-2)$, що на рівні α означає наявність гетероскедастичності. Тоді необхідно з'ясувати, який вид функції (3.19), тобто, якому значенню δ слід віддати перевагу. Для цього потрібно при різних значеннях δ обчислити значення t -статистики або значення коефіцієнта детермінації. Знайдене максимальне значення i відповідатиме шуканому значенню δ .

3.5. Повернемося до моделі (2.3)

$$Y_i = \alpha_0 + \alpha_1 x_i + U_i, \quad i = \overline{1, n}, \quad (3.20)$$

для якої встановлено існування гетероскедастичності, а за припущенням збурення U_i не корелюють між собою ($\text{cov}(U_i, U_j) = 0 \quad \forall i, j = \overline{1, n}, i \neq j$).

Поділимо ліву і праву частину рівнянь (3.20) на σ_i , де $\sigma_1, \sigma_2, \dots, \sigma_n$ є фіксованими додатними числами, не рівними між собою. Тоді для моделі

$$\frac{Y_i}{\sigma_i} = \frac{\alpha_0}{\sigma_i} + \alpha_1 \frac{x_i}{\sigma_i} + \frac{U_i}{\sigma_i}, \quad i = \overline{1, n}, \quad (3.21)$$

вже виконується умова гомоскедастичності, оскільки при $\sigma_i = \sigma_{u_i}$

$$D\left(\frac{U_i}{\sigma_{u_i}}\right) = \frac{1}{\sigma_{u_i}^2} D(U_i) = \frac{1}{\sigma_{u_i}^2} \{M(U_i^2) - [M(U_i)]^2\} = \frac{1}{\sigma_{u_i}^2} \sigma_{u_i}^2 = 1.$$

Нехай a_0 і a_1 — оцінки невідомих параметрів α_0 і α_1 відповідно. Тоді оцінкою моделі (3.21) по вибірці $\{(x_i, y_i), i = \overline{1, n}\}$ є n рівнянь

$$\frac{y_i}{\sigma_{u_i}} = \frac{a_0}{\sigma_{u_i}} + a_1 \frac{x_i}{\sigma_{u_i}} + \frac{u_i}{\sigma_{u_i}}, \quad i = \overline{1, n}.$$

Оцінки a_0 , a_1 знайдемо із умови мінімізації функції

$$\tilde{Q}(a_0, a_1) = \sum_{i=1}^n \left(\frac{a_0 + a_1 x_i - y_i}{\sigma_{u_i}} \right)^2.$$

За аналогією із п.2 §2 отримаємо систему нормальних рівнянь

$$\begin{cases} \left(\sum_{i=1}^n \frac{1}{\sigma_{u_i}^2} \right) a_0 + \left(\sum_{i=1}^n \frac{x_i}{\sigma_{u_i}^2} \right) a_1 = \sum_{i=1}^n \frac{y_i}{\sigma_{u_i}^2}, \\ \left(\sum_{i=1}^n \frac{x_i}{\sigma_{u_i}^2} \right) a_0 + \left(\sum_{i=1}^n \frac{x_i^2}{\sigma_{u_i}^2} \right) a_1 = \sum_{i=1}^n \frac{x_i y_i}{\sigma_{u_i}^2}, \end{cases} \quad (3.22)$$

яка однозначно визначає невідомі оцінки a_0 , a_1 .

Описаний метод знаходження оцінок параметрів регресії називається **методом зважених найменших квадратів (МЗНК)**. Можна безпосередньо перевірити, що МЗНК покращує точність моделі: оцінки коефіцієнтів для моделі (3.21) (її називають **зваженою регресією**) більш ефективні в порівнянні з оцінками звичайної регресії (3.20).

Для практичної реалізації МЗНК потрібно в системі рівнянь (3.22) замість σ_{u_i} підставити знайдені за допомогою теста Глейзера значення $f(x_i)$, які є статистичними оцінками невідомих σ_{u_i} .

Зауваження. На практиці процедура усунення гетероскедастичності може викликати технічні труднощі. Справа в тому, що внаслідок використання оцінок σ_{u_i} модель (3.21) не обов'язково виявиться гомоскедастичною з огляду на такі причини. По-перше, далеко не завжди виявляється правильним саме припущення (3.16) або (3.18). По-друге, функція f у формулі (3.15), взагалі кажучи, не обов'язково степенева (і, тим більше, не обов'язково квадратична), і у цьому випадку її підбір може виявитися далеко не таким простим.

Другим недоліком тестів Уайта і Глейзера є те, що факт не виявлення ними гетероскедастичності, взагалі кажучи, не означає її відсутності. Справді, приймаючи гіпотезу H_0 , ми приймаємо лише той факт,

що відсутня певного виду залежність дисперсій помилок регресії від незалежної змінної.

4. Передумова 3 передбачає відсутність кореляції між збуреннями U_i та U_j при $i \neq j$, тобто $\text{cov}(U_i, U_j) = 0 \quad \forall i, j = \overline{1, n}, i \neq j$. Моделі, для яких не виконується ця передумова, називаються **моделями із наявністю автокореляції**.

Відмітимо, що у випадку просторової вибірки відсутність автокореляції **постулюється**. Разом з тим при використанні комп'ютерних регресійних пакетів (незалежно від виду вибірки) наводиться значення статистики Дарбіна-Уотсона стосовно наявності автокореляції між **сусідніми** членами вибірки.

Для часових рядів модель (2.3) запишеться у такому вигляді:

$$Y_t = \alpha_0 + \alpha_1 x_t + U_t, \quad t = \overline{1, n}. \quad (3.23)$$

Якщо збурення U_t в різні моменти часу t корелюють між собою, тоді між значеннями збурення існує залежність, яку у загальному випадку можна записати у вигляді

$$U_t = \rho_1 U_{t-1} + \rho_2 U_{t-2} + \dots + \rho_s U_{t-s} + \varepsilon_t, \quad t = \overline{2, n}, \quad (3.24)$$

де ρ_i — i -тий коефіцієнт кореляції, ε_t — випадкова величина, s — величина запізнення.

При виконанні співвідношення (3.24) із $\rho_i \neq 0, i = \overline{1, s}$, говорять, що послідовність $\{U_t\} \quad t = \overline{1, n}$ утворить **авторегресійний процес s -го порядку**. Назва «авторегресійний» зумовлена тим, що U_t визначається значеннями цієї ж самої величини в попередніх моментах часу (запізненнями).

МНК при наявності корельованості збурень регресії дає незміщені і спроможні оцінки коефіцієнтів регресії, але вони вже є неефективними. Більше того, оцінки їх дисперсій неспроможні і зміщені (як правило, у бік заниження), тобто результати тестування гіпотез виявляються недостовірними.

Таким чином, актуальними є такі питання:

- 1) як діагностувати наявність автокореляції?
- 2) яким чином модель із автокореляцією можна було б привести до класичної регресійної моделі?

Відповідь на друге питання пов'язана з необхідністю розгляду **узгальненого методу найменших квадратів (УМНК)**, який буде викладено пізніше.

5. Якщо в моделі (3.23) присутня автокореляція, тоді, як правило, найбільший вплив на наступне спостереження виявляє результат попереднього спостереження. Наприклад, коли розглядається ряд значень курсу деякої акції, то, очевидно, саме результат останніх торгів слугує відправною точкою для формування курсу на наступних торгах.

Ситуація, коли на значення спостереження Y_t виявляє основний вплив не результат Y_{t-1} , а більш ранні значення, є достатньо рідкісною.

Таким чином, відсутність кореляції між сусідніми членами служить достатньою підставою вважати, що кореляція відсутня в цілому, і звичайний МНК дає адекватні і ефективні результати.

Тест Дарбіна-Уотсона визначає наявність автокореляції між сусідніми членами (часового) ряду. Він ґрунтується на простій ідеї: якщо кореляція збурень (залишків) регресії U_t існує, то вона присутня у залишках регресії u_t , які отримуються в результаті використання звичайного МНК. У тесті Дарбіна-Уотсона для оцінки кореляції використовується статистика виду

$$d = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\sum_{t=1}^n u_t^2}. \quad (3.25)$$

Можна довести, що статистика Дарбіна-Уотсона пов'язана з вибірко-вим коефіцієнтом кореляції між сусідніми спостереженнями таким чином:

$$d \approx 2(1 - r).$$

Природно, що у випадку відсутності автокореляції вибірко-вий коефіцієнт кореляції r є близьким до нуля, а значення статистики d буде близьким до двох.

Якщо спостережене значення $d \approx 0$, то $r \approx 1$, тобто спостерігається додатна автокореляція, а якщо $d \approx 4$, то $r \approx -1$, тобто наявна від'ємна кореляція. У зв'язку з цим бажано мати відповідні порогові (критичні) значення, присутні у статистичних критеріях, і які або дозволяють прийняти гіпотезу, або примушують її відкинути. Проте, на жаль, такі критичні або порогові значення однозначно вказати неможливо.

Тест Дарбіна-Уотсона має один суттєвий недолік — розподіл статистики d залежить не тільки від числа спостережень n , але і від значень незалежної змінної t . Це означає, що тест Дарбіна-Уотсона, взагалі кажучи, **не можна віднести до статистичних критеріїв**, оскільки не можна вказати критичну область, яка дозволяла б відкинути гіпотезу про відсутність кореляції, якби виявилось, що в цю область потрапило б спостережене значення статистики d .

Проте існують два порогові значення d_g і d_n , які залежать **тільки** від n і рівня значущості, і такі, що виконуються наступні твердження.

Якщо спостережене значення d :

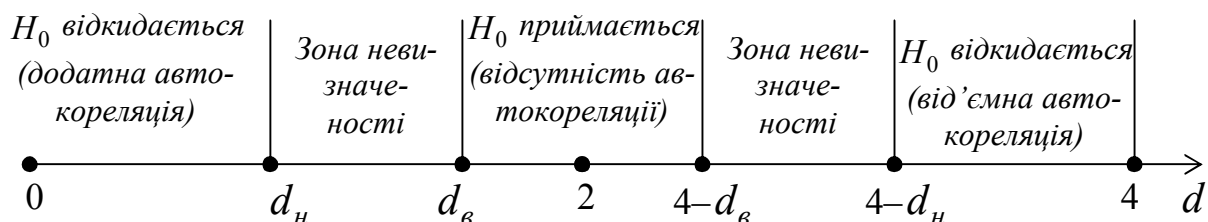
а) $d_g < d < 4 - d_g$, то гіпотеза про відсутність автокореляції приймається;

б) $d_n < d < d_g$ або $4 - d_g < d < 4 - d_n$, то питання про відхилення або прийняття гіпотези залишається відкритим (значення d потрапляє у область невизначеності критерія);

в) $0 < d < d_n$, то приймається альтернативна гіпотеза про додатну автокореляцію;

г) $4 - d_n < d < 4$, то приймається альтернативна гіпотеза про від'ємну автокореляцію.

Зобразимо наведені результати графічно:



Для d -статистики знайдені верхня d_g і нижня d_n межі на рівнях значущості $\alpha = 0,01$; $0,025$ і $0,05$.

У табл. 5 додатків наведені значення статистик d_g і d_n критерія Дарбіна-Уотсона для рівня значущості $\alpha = 0,05$.

Недоліком критерія Дарбіна-Уотсона є наявність зон невизначеності цього критерія, а також те, що критичні значення d -статистики визначені для обсягів вибірки $n \geq 15$. Тим не менше, тест Дарбіна-Уотсона є найбільш використовуваним.

При використанні комп'ютерних регресійних пакетів значення статистики d наводиться автоматично при оцінюванні моделі методом найменших квадратів.

Задача 3.5. На рівні значущості $\alpha = 0,05$ для задачі 2.1 перевірити виконання передумови 3 за тестом Дарбіна-Уотсона.

○ Підсумок останнього рядка табл. 2.2 визначає знаменник дроби (3.25):

$\sum_{t=1}^{10} u_t^2 = 1,1544$. Для обчислення чисельника цього ж дроби використаємо

дані передостаннього рядка табл. 2.2:

$$\begin{aligned} \sum_{t=2}^{10} (u_t - u_{t-1})^2 &= [0,7178 - (-0,1938)]^2 + (-0,059 - 0,7178)^2 + \\ &+ [-0,4474 - (-0,059)]^2 + [-0,3242 - (-0,4474)]^2 + [0,099 - (-0,3242)]^2 + \\ &+ (-0,2894 - 0,099)^2 + [0,4222 - (-0,2894)]^2 + (0,1338 - 0,4222)^2 + \\ &+ (-0,0546 - 0,1338)^2 = 0,831 + 0,6034 + 0,1509 + 0,0152 + 0,1791 + \\ &+ 0,1509 + 0,5064 + 0,0832 + 0,0355 = 2,5556. \end{aligned}$$

Спостережене значення статистики (3.25):

$$d = \frac{2,5556}{1,1544} = 2,2138.$$

За табл. 5 додатків при $n = 15$ критичні значення $d_n = 1,08$, $d_g = 1,36$, тобто $d = 2,2138$ знаходиться в межах від d_g до $4 - d_g$ ($1,36 < d < 2,64$). Як було відмічено вище, при $n < 15$ критичних значень d -статистики в таблиці немає, проте згідно з тенденцією їх змін із зменшенням n , можна припустити, що знайдене значення залишиться в інтервалі $(d_g; 4 - d_g)$, тобто для статистичних даних задачі 2.1 на рівні значущості $\alpha = 0,05$ гіпотеза про відсутність автокореляції збурень не відхиляється (приймається). ◎

6. Статистика Дарбіна-Уотсона є найбільш важливим індикатором наявності автокореляції. Проте недоліками цієї статистики, окрім наявності зон невизначеності, є також обмеженість результату, яка полягає у тому, що кореляція виявляється тільки між **сусідніми** членами. Ця обставина приводить до необхідності використання інших тестів на наявність автокореляції. У всіх цих тестах в якості основної гіпотези H_0 розглядається гіпотеза про відсутність автокореляції.

Тест серій (Бреуша-Годфрі) ґрунтується на такій ідеї: якщо існує кореляція між сусідніми спостереженнями, то природно очікувати, що у рівнянні

$$U_t = \rho U_{t-1} + \varepsilon_t, \quad t = \overline{2, n} \quad (3.26)$$

коефіцієнт ρ виявиться значуще відмінним від нуля. Відмітимо, що рівняння (3.26) є частковим видом рівняння (3.24) при $s = 1$.

Практичне використання тесту серій полягає в оцінюванні методом найменших квадратів моделі (3.26).

Припустимо, що випадкова величина ε_t задовольняє передумовам 1-4. Порівняння моделей (3.26) і (2.3) дозволяє зробити висновок, що $\alpha_0 = 0$, $\alpha_1 = \rho$. Крім цього, вибіркою обсягом $n - 1$ є дані

$$\{y_i = u_i; x_i = u_{i-1}\}, \quad i = \overline{2, n}. \quad (3.27)$$

Тоді згідно з МНК із другого рівняння (2.12), де $a_0 = 0$, отримаємо оцінку параметра ρ :

$$\hat{\rho} = \frac{\overline{xy}}{\overline{x^2}} = \frac{\sum_{i=2}^n x_i y_i}{\sum_{i=2}^n x_i^2} \quad (3.28)$$

або із врахуванням (3.27)

$$\hat{\rho} = \frac{\sum_{i=2}^n u_{i-1} u_i}{\sum_{i=2}^n u_{i-1}^2}. \quad (3.28^*)$$

За аналогією із формулою (2.22) незміщеною оцінкою дисперсії збурень ε_t є

$$S_\varepsilon^2 = \frac{1}{n-2} \sum_{i=2}^n (u_i - \hat{u}_i)^2, \quad (3.29)$$

де $\hat{u}_i = \hat{\rho} u_{i-1}$, а величина знаменника зумовлена тим, що обсяг вибірки $n - 1$ і число оцінюваних параметрів дорівнює 1.

Із (3.28) та (3.27) отримаємо з урахуванням передумов 2, 3 та детермінованості x_i і ρ

$$\sigma_{\hat{\rho}}^2 = D(\hat{\rho}) = \frac{\sum_{i=2}^n x_i^2}{\left(\sum_{i=2}^n x_i^2\right)^2} \sigma_y^2 = \frac{\sigma_\varepsilon^2}{\sum_{i=2}^n u_{i-1}^2}.$$

Незміщена оцінка $\sigma_{\hat{\rho}}^2$ із врахуванням (3.29) має такий вигляд:

$$S_{\hat{\rho}}^2 = \frac{\sum_{i=2}^n (u_i - \hat{u}_i)^2}{(n-2) \sum_{i=2}^n u_{i-1}^2}, \quad (3.30)$$

Зміст основної гіпотези $H_0(\rho = 0)$ та альтернативної $H_1(\rho \neq 0)$ дозволяє сформулювати двосторонній критерій значущості оцінки $\hat{\rho}$:

якщо виконується нерівність

$$\left| \frac{\hat{\rho}}{S_{\hat{\rho}}} \right| > t_{кр.}, \quad (3.31)$$

де $t_{кр.} = t_{двост.кр.}(\alpha; n-2)$ — критична точка розподілу Ст'юдента, тоді на рівні значущості α приймається гіпотеза H_1 , тобто вважається, що $\rho \neq 0$.

Перевагою тесту Бреуша-Годфрі в порівнянні з тестом Дарбіна-Уотсона є те, що він перевіряється з допомогою **статистичного критерія**. Друга перевага полягає у можливості досліджувати авторегресійний процес (3.24) s -го порядку.

Задача 3.6. На рівні значущості $\alpha = 0,05$ для задачі 2.1 перевірити наявність авторегресії першого порядку з допомогою тесту Бреуша-Годфрі.

○ Знайдемо оцінку $\hat{\rho}$ за формулою (3.28*), використавши дані двох останніх рядків табл. 2.2:

$$\hat{\rho} = -0,1474041/1,1168 = -0,132.$$

Для обчислення S_ε^2 за формулою (3.29) складемо допоміжну табл. 3.3, використавши дані передостаннього рядка табл. 2.2 і співвідношення $\hat{u}_i = -0,132u_{i-1}$, де $i = \overline{2,10}$.

Таблиця 3.3

i	u_i	\hat{u}_i	$u_i - \hat{u}_i$	$(u_i - \hat{u}_i)^2$
1	-0,1938	—	—	—
2	0,7178	0,02558	0,69222	0,47917
3	-0,059	-0,09475	0,03575	0,00128
4	-0,4474	0,00779	-0,45519	0,2072
5	-0,3242	0,05906	-0,38326	0,14689
6	0,099	0,04279	0,05621	0,00316
7	-0,2894	-0,01307	-0,2512	0,0631
8	0,4222	0,03820	0,384	0,14746
9	0,1338	-0,05573	0,18953	0,03592
10	-0,0546	-0,01766	-0,03694	0,00137
Σ	—	—	—	1,08555

Тоді

$$S_{\varepsilon}^2 = \frac{1}{10-2} \sum_{i=2}^{10} (u_i - \hat{u}_i)^2 = \frac{1}{8} \cdot 1,08555 = 0,1357.$$

За формулою (3.30) знайдемо незміщену оцінку дисперсії вибіркового коефіцієнта регресії $\hat{\rho}$:

$$S_{\hat{\rho}}^2 = \frac{\sum_{i=2}^{10} (u_i - \hat{u}_i)^2}{(10-2) \sum_{i=2}^{10} u_{i-1}^2} = \frac{1,08555}{8 \cdot 1,1168} = 0,1215,$$

звідки $S_{\hat{\rho}} = 0,34857$.

Тоді

$$\left| \frac{\hat{\rho}}{S_{\hat{\rho}}} \right| = \left| \frac{-0,132}{0,34857} \right| = 0,3787.$$

Але оскільки

$$\left| \frac{\hat{\rho}}{S_{\hat{\rho}}} \right| = 0,3787 < t_{\text{двост.кр.}}(0,05;8) = 2,306,$$

то на рівні $\alpha = 0,05$ робимо висновок, що $\rho = 0$, тобто для даних задачі 2.1 відсутня авторегресія першого порядку. \odot

Зауваження. В більшості сучасних комп'ютерних пакетів застосування тесту серії здійснюється спеціальною командою, а тому не обов'язково оцінювати регресію типу (3.26) «вручну».

Відмітимо також ефективність Q -тесту Л'юінга-Бокса. Розгляд цього тесту можна здійснити після наведення більш детальної інформації про часові ряди.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ №2

За умовами індивідуального завдання №1 для рівня значущості $\alpha = 0,05$ перевірити виконання передумов 1-4, використавши всі наведені вище тести. При наявності гетероскедастичності — усунути її за допомогою методу зважених найменших квадратів.

§4. МНОЖИННИЙ РЕГРЕСІЙНИЙ АНАЛІЗ.

1. Класична нормальна лінійна модель множинної регресії.
2. Оцінка параметрів класичної регресійної моделі методом найменших квадратів.
3. Коваріаційна матриця вектора оцінок параметрів регресії та її вибіркова оцінка.
4. Оцінка дисперсії збурень.
5. Знаходження довірчих інтервалів для коефіцієнтів і функції регресії.
6. Мультиколінеарність у множинних регресійних моделях.
7. Оцінка значущості множинної регресії. Коефіцієнти детермінації R^2 і \hat{R}^2 .

1. Економічні явища та процеси визначаються великим числом діючих чинників. Це зумовлює необхідність дослідження залежності однієї результуючої змінної y від пояснюючих (незалежних) змінних x_1, x_2, \dots, x_m . Така задача розв'язується з допомогою **множинного регресійного аналізу**.

Об'єктом вивчення цієї теми є така модель множинної лінійної регресії:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m + u, \quad (4.1)$$

де y та u — випадкові величини (u — збурення або залишок), $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_m$ — невідомі детерміновані параметри.

Нехай $x_{i1}, x_{i2}, \dots, x_{im}$ ($i = \overline{1, n}$) — спостережні значення пояснюючих змінних. Тоді модель (4.1) набере такого виду:

$$Y_i = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \dots + \alpha_m x_{im} + U_i, \quad i = \overline{1, n} \quad (4.2)$$

Систему n рівнянь (4.2) запишемо у векторно-матричному вигляді:

$$Y_M = X\alpha + U, \quad (4.2^*)$$

де

$$Y_M = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}, \quad \alpha = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix}, \quad U = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix}.$$

Оцінкою цієї моделі по вибірці $\{y_i, x_{i1}, x_{i2}, \dots, x_{im}\}$ обсягом n є векторно-матричне рівняння

$$Y = Xa + E, \quad (4.3)$$

де $Y = (y_1, y_2, \dots, y_n)'$, $a = (a_0, a_1, \dots, a_m)'$, $E = (e_1, e_2, \dots, e_n)'$, штрих означає операцію транспонування матриці.

Покладемо, що стосовно моделі (4.2*) виконуються такі припущення.

Передумова 1. U — випадковий вектор, змінні x_1, x_2, \dots, x_m — детерміновані величини, а тому X — детермінована матриця.

Передумова 2. $M(U) = \mathbf{0}_n = (0, 0, \dots, 0)'$.

Передумова 3. $\sum_U = M(UU') = \sigma^2 I_n$, де I_n — одинична матриця порядку n , σ — стала.

Передумова 4. U — нормально розподілений випадковий вектор, тобто $U \sim N_n(\mathbf{0}, \sigma^2 I_n)$.

Передумова 5. Ранг матриці X дорівнює $t + 1 < n$.

Означення. Модель (4.2*), яка задовольняє передумовам 1-5, називається *класичною нормальною лінійною моделлю множинної регресії*; якщо ж не виконується тільки передумова 4, то модель називається *класичною лінійною моделлю множинної регресії*.

Відмітимо, що згідно з передумовами 2 і 3 для компонент вектора U виконуються рівності

$$D(U_i) = M(U_i^2) - [M(U_i)]^2 = \sigma^2 - 0 = \sigma^2, \\ \text{cov}(U_i, U_j) = 0 \quad \forall i \neq j \quad i, j = \overline{1, n}.$$

Нагадаємо також, що \sum_U позначає, як і у §2, коваріаційну матрицю n -вимірної випадкової величини U , а

$$U \cdot U' = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix} (U_1, U_2, \dots, U_n) = \begin{pmatrix} U_1^2 & U_1 U_2 & \dots & U_1 U_n \\ U_1 U_2 & U_2^2 & \dots & U_2 U_n \\ \dots & \dots & \dots & \dots \\ U_n U_1 & U_n U_2 & \dots & U_n^2 \end{pmatrix}.$$

Нарешті, передумова 5 означає, що обсяг вибірки повинен бути більшим від $t + 1$, а всі стовпці матриці X є лінійно незалежними.

Таким чином, при $t = 1$ всі передумови лінійної моделі парної регресії виконуються.

2. Детерміновану складову моделі (4.3) позначимо $\hat{Y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)'$, тобто.

$$\hat{Y} = Xa, \quad (4.4)$$

Тоді критерієм вибору вектора оцінок a згідно з методом найменших квадратів є мінімізація суми квадратів залишків:

$$Q(a) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2 = E'E = (Y - Xa)'(Y - Xa) \rightarrow \min. \quad (4.5)$$

Оскільки для будь-яких матриць A та B узгодженої вимірності $(A \cdot B)' = B' \cdot A'$, то після розкриття дужок в (4.5) отримаємо

$$Q(a) = Y'Y - a'X'Y - Y'Xa + a'X'Xa.$$

Добуток $Y'Xa$ є скалярною величиною, тому він не змінюється від транспонування:

$$Y'Xa = (Y'Xa)' = a'X'Y.$$

З врахуванням цього умова (4.5) набуде такого вигляду:

$$Q(a) = Y'Y - 2a'X'Y + a'X'Xa \rightarrow \min. \quad (4.6)$$

Згідно з необхідною умовою екстремуму функції $Q(a)$ $m+1$ змінних a_0, a_1, \dots, a_m потрібно прирівняти до нуля всі частинні похідні першого порядку від $Q(a)$ по цих змінних або у матричній формі — вектор частинних похідних:

$$\frac{\partial Q}{\partial a} = \left(\frac{\partial Q}{\partial a_0}, \frac{\partial Q}{\partial a_1}, \dots, \frac{\partial Q}{\partial a_m} \right).$$

Нехай b та c — k -вимірні вектор-стовпці, A — симетрична матриця порядку k . Можна довести такі рівності:

$$\frac{\partial}{\partial b}(b'c) = c', \quad \frac{\partial}{\partial b}(b'Ab) = 2b'A.$$

Тому, покладаючи $b = a$, $c = X'Y$, $A = X'X$, $k = m+1$, із (4.6) отримаємо

$$\frac{\partial Q}{\partial a} = -2(X'Y)' + 2a'X'X = -2Y'X + 2a'X'X.$$

Врахувавши необхідну умову екстремуму

$$-2Y'X + 2a'X'X = (0, 0, \dots, 0),$$

або

$$-2X'Y + 2X'Xa = (0, 0, \dots, 0)'$$

прийдемо до системи нормальних рівнянь у матричній формі для визначення вектора a :

$$X'Xa = X'Y. \quad (4.7)$$

Згідно з передумовою 5 $(m+1) \times (m+1)$ — матриця $X'X$ є невиродженою, тому розв'язком рівняння (4.7) є вектор

$$a = (X'X)^{-1} X'Y, \quad (4.8)$$

де $(X'X)^{-1}$ — обернена матриця до матриці $X'X$.

Основні властивості отриманих оцінок (4.8) визначаються наступними твердженнями.

Теорема Гаусса-Маркова. Нехай стосовно моделі (4.2*) виконуються передумови 1-3, 5. Тоді оцінки (4.8) вектора параметрів α мають найменшу дисперсію в класі лінійних незміщених оцінок.

Доведення цієї теореми можна знайти, зокрема, в [6, стор. 94].

Якщо вектор a знайдено, тоді **вибіркове рівняння множинної регресії** можна зобразити у такому вигляді:

$$\hat{y} = X'_0 a = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_m x_m, \quad (4.9)$$

де \hat{y} — групова (умовна) середня змінної y при заданому векторі значень пояснюючих змінних

$$X'_0 = (1, x_{10}, x_{20}, \dots, x_{m0}) = (1, x_1, x_2, \dots, x_m).$$

Задача 3.1. Підприємство, що складається із багатьох філій, досліджує залежність свого річного товарообігу y (млн. грн.) від торгової площі своїх філій x_1 (тис. кв. м.) і середньоденної інтенсивності потоку покупців (тис. покупців за день). Дані 10 філій наведені у табл. 3.1. Оцінивши параметри лінійної регресійної моделі $y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + u$, знайти вибіркове рівняння множинної регресії.

Таблиця 3.1

Номер філії	Значення		
	y	x_1	x_2
1	6,9	1,2	10,8
2	7,1	1,3	9,9
3	7	1,1	13,7
4	8,4	1,5	13,9
5	4,3	0,8	8,5
6	5,8	0,9	12,4
7	7,7	1,3	12,3
8	3,2	0,5	11
9	1,5	0,2	8,3
10	3,1	0,6	9,3

○ Введемо позначення:

$$Y = \begin{pmatrix} 6,9 \\ 7,1 \\ 7 \\ 8,4 \\ 4,3 \\ 5,8 \\ 7,7 \\ 3,2 \\ 1,5 \\ 3,1 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 1,2 & 10,8 \\ 1 & 1,3 & 9,9 \\ 1 & 1,1 & 13,7 \\ 1 & 1,5 & 13,9 \\ 1 & 0,8 & 8,5 \\ 1 & 0,9 & 12,4 \\ 1 & 1,3 & 12,3 \\ 1 & 0,5 & 11 \\ 1 & 0,2 & 8,3 \\ 1 & 0,6 & 9,3 \end{pmatrix},$$

$$X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1,2 & 1,3 & 1,1 & 1,5 & 0,8 & 0,9 & 1,3 & 0,5 & 0,2 & 0,6 \\ 10,8 & 9,9 & 13,7 & 13,9 & 8,5 & 12,4 & 12,3 & 11 & 8,3 & 9,3 \end{pmatrix}.$$

Знаходимо добутки $X \cdot X'$ та $X'Y$. В результаті отримаємо:

$$X'X = \begin{pmatrix} 10 & 9,4 & 110,1 \\ 9,4 & 10,38 & 108,44 \\ 110,1 & 108,44 & 1249,23 \end{pmatrix}, \quad X'Y = \begin{pmatrix} 55 \\ 60,24 \\ 637,13 \end{pmatrix}.$$

Обернена матриця до матриці $X'X$ матиме вигляд:

$$(X'X)^{-1} = \begin{pmatrix} 3,6924 & 0,6007 & -0,3776 \\ 0,6007 & 1,1320 & -0,1512 \\ -0,3776 & -0,1512 & 0,0472 \end{pmatrix}.$$

Значення елементів вектора a знайдемо, перемноживши матрицю $(X'X)^{-1}$ на вектор $X'Y$:

$$a = (X'X)^{-1} X'Y = \begin{pmatrix} -1,3121 \\ 4,8961 \\ 0,1962 \end{pmatrix}.$$

Отже, ми отримали вибіркове рівняння множинної регресії:

$$\hat{y} = -1,3121 + 4,8961x_1 + 0,1962x_2.$$

Нагадаємо, що добуток $m \times n$ -матриці $A = (a_{ij})$, $i = \overline{1, m}$, $j = \overline{1, n}$, на $n \times k$ -матрицю $B = (b_{ij})$, $i = \overline{1, n}$, $j = \overline{1, k}$, є $m \times k$ -матриця $C = (c_{ij})$, $i = \overline{1, m}$, $j = \overline{1, k}$, для елементів якої виконуються рівності

$$c_{ij} = \sum_{s=1}^n a_{is} b_{sj}, \quad i = \overline{1, m}, \quad j = \overline{1, k}.$$

При цьому використовується позначення $C = AB$.

У випадку здійснення розрахунків «вручну» рекомендується використання табл. 4.1 та 4.2.

Таблиця 4.1

i	x_{i1}	x_{i2}	x_{i1}^2	$x_{i1}x_{i2}$	x_{i2}^2
1	1,2	10,8	1,44	12,96	116,64
2	1,3	9,9	1,69	12,87	98,01
3	1,1	13,7	1,21	15,07	187,69
4	1,5	13,9	2,25	20,85	193,21
5	0,8	8,5	0,64	6,8	72,25
6	0,9	12,4	0,81	11,16	153,76
7	1,3	12,3	1,69	15,99	151,29
8	0,5	11	0,25	5,5	121
9	0,2	8,3	0,04	1,66	68,89
10	0,6	9,3	0,36	5,58	86,49
Σ	9,4	110,1	10,38	108,44	1249,23

Таблиця 4.2

i	y	$y_i x_{i1}$	$y_i x_{i2}$
1	6,9	8,28	74,52
2	7,1	9,23	70,29
3	7	7,7	95,9
4	8,4	12,6	116,76
5	4,3	3,44	36,55
6	5,8	5,22	71,92
7	7,7	10,01	94,71
8	3,2	1,6	35,2
9	1,5	0,3	12,45
10	3,1	1,86	28,83
Σ	55	60,24	637,13

Елементи першого рядка матриці $X'X$ отримуються шляхом множення чисел першого рядка матриці X' на відповідні елементи стовпців матриці X з наступним додаванням добутоків, тобто 10; 9,4; 110,1 (див. останній рядок табл. 4.1).

Елементи другого рядка матриці $X'X$ — це сума добутоків чисел другого рядка на відповідні елементи стовпців матриці X : 9,4; 10,38; 108,44. Нарешті, елементи третього рядка матриці $X'X$ (110,1; 108,44; 1249,23) також визначаються із останнього рядка табл. 4.1.

Елементи вектор-стовпця $X'Y$ виписуються із останнього рядка табл. 4.2:

$$XY = \begin{pmatrix} 55 \\ 60,24 \\ 637,13 \end{pmatrix}.$$

Нагадаємо також послідовність дій при знаходженні оберненої матриці A :

1) обчислюється визначник $|A|$ (при цьому A^{-1} не існує, якщо $|A| = 0$);

2) для транспонованої матриці A' знаходиться приєднана матриця $\tilde{A} = (A_{ij})$, де A_{ij} — алгебраїчні доповнення елементів a_{ij} матриці A' , тобто $A_{ij} = (-1)^{i+j} |\cdot|$, $|\cdot|$ — визначник, який отримується із $|A'|$ шляхом виокреслення i -го рядка та j -го стовпця;

3) знаходиться $A^{-1} = \frac{1}{|A|} \tilde{A}$. ⊙

3. Дисперсії оцінок параметрів визначають точність рівняння множинної регресії. Для їх вимірювання розглядають так звану **коваріаційну матрицю вектора оцінок параметрів** Σ_a , яка є матричним аналогом дисперсії однієї випадкової величини:

$$\Sigma_a = \begin{pmatrix} \sigma_{00} & \sigma_{01} & \dots & \sigma_{0m} \\ \sigma_{10} & \sigma_{11} & \dots & \sigma_{1m} \\ \dots & \dots & \dots & \dots \\ \sigma_{m0} & \sigma_{m1} & \dots & \sigma_{mm} \end{pmatrix},$$

де елементи σ_{ij} — коваріації оцінок параметрів a_i та a_j :
 $\sigma_{ij} = \text{cov}(a_i, a_j)$, $i, j = \overline{0, m}$, $\sigma_{it} = D(a_i)$, $t = \overline{0, n}$.

Знайдемо матрицю Σ_a на підставі конкретної вибірки. Із урахуванням (4.2*) для нефіксованої вибірки отримаємо зображення вектора оцінок (4.8):

$$a = (X'X)^{-1} X'Y_M = (X'X)^{-1} X'(X\alpha + U) = (X'X)^{-1} X'X\alpha + \\ + (X'X)^{-1} X'U = I_{m+1}\alpha + (X'X)^{-1} X'U$$

або

$$a = \alpha + (X'X)^{-1} X'U. \quad (4.10)$$

Із цієї рівності випливають такі висновки:

1) оцінки параметрів (4.8), знайдені за нефіксованою вибіркою, будуть містити випадкові помилки;

2) згідно із передумовами 1 та 2

$$M(a) = \alpha,$$

тобто вектор a є незміщеною оцінкою вектора параметрів α .

У скороченому вигляді коваріаційна матриця вектора оцінок параметрів має такий вигляд:

$$\Sigma_a = M[(a - \alpha)(a - \alpha)'], \quad (4.11)$$

оскільки

$$\sigma_{ij} = \text{cov}(a_i, a_j) = M[(a_i - M(a_i))(a_j - M(a_j))] = M[(a_i - \alpha_i)(a_j - \alpha_j)],$$

$$(a - \alpha)(a - \alpha)' = \begin{pmatrix} a_0 - \alpha_0 \\ a_1 - \alpha_1 \\ \vdots \\ a_m - \alpha_m \end{pmatrix} (a_0 - \alpha_0, a_1 - \alpha_1, \dots, a_m - \alpha_m) =$$

$$= \begin{pmatrix} (a_0 - \alpha_0)(a_0 - \alpha_0) & (a_0 - \alpha_0)(a_1 - \alpha_1) & \dots & (a_0 - \alpha_0)(a_m - \alpha_m) \\ (a_1 - \alpha_1)(a_0 - \alpha_0) & (a_1 - \alpha_1)(a_1 - \alpha_1) & \dots & (a_1 - \alpha_1)(a_m - \alpha_m) \\ \dots & \dots & \dots & \dots \\ (a_m - \alpha_m)(a_0 - \alpha_0) & (a_m - \alpha_m)(a_1 - \alpha_1) & \dots & (a_m - \alpha_m)(a_m - \alpha_m) \end{pmatrix}.$$

Враховуючи (4.10), детермінованість матриці X , передумову 3 і симетричність матриці $X'X$, отримаємо із (4.11)

$$\Sigma_a = M\{[(X'X)^{-1}X'U][(X'X)^{-1}X'U]'\} = M\{(X'X)^{-1}X'UU'X[(X'X)^{-1}]'\} =$$

$$= (X'X)^{-1}X'M(UU')X(X'X)^{-1} = \sigma^2(X'X)^{-1}X'I_nX(X'X)^{-1} = \sigma^2(X'X)^{-1}.$$

Тобто, остаточно

$$\Sigma_a = \sigma^2(X'X)^{-1}. \quad (4.12)$$

Висновок: з допомогою матриці $(X'X)^{-1}$ визначається не тільки сам вектор a оцінок параметрів (4.8), але і дисперсії, і коваріації його компонент.

Із (4.12), зокрема отримуємо:

$D(a_i) = \sigma^2[(X'X)^{-1}]_{ii}$, $i = \overline{0, m}$, — відповідні діагональні елементи матриці Σ_a .

4. Параметр σ^2 , який фігурує у передумовах 3, 4, а також у рівності (4.12) — **невідомий**. Як і у випадку парної регресії його необхідно **оцінити**.

Розглянемо вектор залишків E , який визначається із (4.3):

$$E = Y - X\alpha.$$

Для **нефіксованої** вибірки $Y = Y_M$, а тому із врахуванням (4.2*) і (4.8)

$$\begin{aligned} E &= X\alpha + U - X[(X'X)^{-1}X'(X\alpha + U)] = \\ &= X\alpha + U - X(X'X)^{-1}X'X\alpha - X(X'X)^{-1}X'U = \\ &= X\alpha + U - X\alpha - X(X'X)^{-1}X'U = U - X(X'X)^{-1}X'U, \end{aligned}$$

тобто

$$E = U - X(X'X)^{-1}X'U,$$

звідки

$$E' = U' - U'X(X'X)^{-1}X'.$$

Тоді

$$\begin{aligned} M(E'E) &= M\{[U' - U'X(X'X)^{-1}X'] [U - X(X'X)^{-1}X'U]\} = \\ &= M(U'U) - M[U'X(X'X)^{-1}X'U] - M[U'X(X'X)^{-1}X'U] + \\ &\quad + M[U'X(X'X)^{-1}X'X(X'X)^{-1}X'U]. \end{aligned}$$

Оскільки два останні доданки знищуються, то

$$M(E'E) = M(U'U) - M[U'X(X'X)^{-1}X'U]. \quad (4.13)$$

Перший доданок правої частини (4.13) дорівнює $n\sigma^2$, оскільки згідно із передумовою 3

$$M(U'U) = M\left(\sum_{i=1}^n U_i^2\right) = \sum_{i=1}^n M(U_i^2) = \sum_{i=1}^n \sigma^2 = n\sigma^2. \quad (4.14)$$

Матриця $B = X(X'X)^{-1}X'$ симетрична, тому $U'BU$ — квадратична форма $\sum_{i,j=1}^n b_{ij}U_iU_j$, де b_{ij} — детерміновані величини. Тоді

$$\begin{aligned} M(U'BU) &= M\left(\sum_{i,j=1}^n b_{ij}U_iU_j\right) = \sum_{i,j=1}^n b_{ij}M(U_iU_j) = \\ &= \sum_{i=1}^n b_{ii}M(U_i^2) + \sum_{\substack{i,j=1 \\ (i \neq j)}}^n b_{ij}M(U_iU_j). \end{aligned}$$

Згідно із передумовою 3 останній доданок дорівнює нулю, а $M(U_i^2) = \sigma^2$ ($i = \overline{1, n}$). Позначимо $Sp B$ або слід матриці — сума діагональних елементів матриці B . Для будь-яких матриць K та M , для яких визначені добутки KM та MK , виконується рівність $Sp(KM) = Sp(MK)$, використавши яку, отримаємо

$$Sp B = Sp(X(X'X)^{-1}X') = Sp(X'X(X'X)^{-1}) = Sp I_{m+1} = m + 1.$$

Таким чином, $M(U'BU) = (m + 1)\sigma^2$, і остаточно із (4.13) та (4.14) отримаємо

$$M(E'E) = (n - m - 1)\sigma^2, \quad (4.15)$$

звідки визначається незміщена оцінка S_e^2 параметра σ^2 або **вибіркова залишкова дисперсія** S_e^2 :

$$S_e^2 = \frac{E'E}{n - m - 1} = \frac{\sum_{i=1}^n e_i^2}{n - m - 1} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - m - 1} \quad (4.16)$$

Число $n - m - 1$ називається **числом ступенів вільності**.

Можна довести, що оцінки a і S_e^2 параметрів α і σ^2 відповідно при виконанні передумови 4 некорелюють між собою як випадкові величини, тобто $cov(a, S_e^2) = 0$.

Задача 3.2. Для задачі 3.1 знайти точкову оцінку дисперсії збурень та точкову оцінку для коваріаційної матриці вектора оцінок параметрів.

○ Незміщену точкову оцінку S_e^2 невідомого параметра σ^2 знайдемо за формулою (4.16), попередньо обчисливши

$$\hat{y}_i = -1,3121 + 4,8961x_{i1} + 0,1962x_{i2} \text{ та } e_i^2 = (y_i - \hat{y}_i)^2, \quad i = \overline{1, 10}, \text{ (табл. 4.3).}$$

Таблиця 4.3

i	1	2	3	4	5	6	7	8	9	10	Σ
y_i	6,9	7,1	7	8,4	4,3	5,8	7,7	3,2	1,5	3,1	–
\hat{y}_i	6,6822	6,9952	6,7616	8,7592	4,2725	5,5273	7,4661	3,2942	1,2956	3,4502	–
e_i	0,2178	0,1048	0,2384	-0,3592	0,0275	0,2727	0,2339	-0,0942	0,2044	-0,3502	0,4959
e_i^2	0,0474	0,011	0,0568	0,129	0,0008	0,0744	0,0547	0,0089	0,0418	0,1226	0,5474

Зауваження. Раніше вже говорилося (див. задачу 2.1.) про накопичення додатних похибок у різницях $y_i - \hat{y}_i$, що приводить величину $\Sigma(y_i - \hat{y}_i)$ до незначного перевищення нуля. Значенням

$\bar{e} = \sum e_i / 10 = 0,04959$ можна ігнорувати (вважати практично рівним нулю).

Використавши підсумок останнього рядка, отримаємо:

$$S_e^2 = \frac{0,5474}{10 - 2 - 1} = 0,0782.$$

Оскільки в (4.12) параметр σ^2 невідомий, то використаємо його незміщену точкову оцінку S_e^2 і отримаємо оцінку для коваріаційної матриці Σ_a :

$$\begin{aligned} S_e^2 (X'X)^{-1} &= 0,0782 \cdot \begin{pmatrix} 3,6924 & 0,6007 & -0,3776 \\ 0,6007 & 1,1320 & -0,1512 \\ -0,3776 & -0,1512 & 0,0472 \end{pmatrix} = \\ &= \begin{pmatrix} 0,2887 & 0,045 & -0,0295 \\ 0,045 & 0,0885 & -0,0118 \\ -0,0295 & -0,0118 & 0,0037 \end{pmatrix}. \end{aligned} \quad \odot$$

5. З'ясуємо значущість коефіцієнтів регресії a_i ($i = \overline{0, m}$) і побудуємо довірчі інтервали для параметрів моделі α_i при умові їх значущості.

Згідно (4.12) і (4.16) незміщена оцінка $S_{a_i}^2$ дисперсії $\sigma_{a_i}^2$ коефіцієнта регресії a_i визначається за формулою

$$S_{a_i}^2 = S_e^2 [(X'X)^{-1}]_{ii}, \quad i = \overline{0, m},$$

де S_e^2 — незміщена оцінка параметра σ^2 , $[(X'X)^{-1}]_{ii}$ — i -й діагональний елемент матриці $(X'X)^{-1}$.

Середнє квадратичне відхилення або **стандартна помилка** коефіцієнта регресії a_i знаходиться за формулою

$$S_{a_i} = S_e \sqrt{[(X'X)^{-1}]_{ii}}, \quad i = \overline{0, m}. \quad (4.17)$$

Оскільки випадкова величина $(a_i - \alpha_i) / S_{a_i}$ для нефіксованої вибірки розподілена за законом Ст'юдента із $k = n - m - 1$ ступенями вільності, то a_i значуще відрізняється від нуля (гіпотеза $H_0: a_i = 0$ відхиляється) на рівні значущості α , якщо

$$|t_{\text{спост.}}| = \frac{|a_i|}{S_{a_i}} > t_{\text{кр.}}, \quad (4.18)$$

де $t_{\text{кр.}} = t_{\text{двост.кр.}}(1 - \alpha; n - m - 1)$ — табличне значення t — критерія Ст'юдента, визначене на рівні значущості α при числі ступенів вільності $k = n - m - 1$.

Якщо використовується нерівність (4.18), тоді є сенс будувати довірчий інтервал для параметра α_i , який визначається подвійною нерівністю

$$a_i - t_{\text{кр.}} S_{a_i} < \alpha_i < a_i + t_{\text{кр.}} S_{a_i}. \quad (4.19)$$

Для оцінки точності знаходження значень залежної змінної важливою є побудова **довірчого інтервала (зони) для функції регресії або умовного математичного сподівання залежної змінної** $M(Y | x_1, x_2, \dots, x_m)$, знайденого у припущенні, що пояснюючі змінні x_1, x_2, \dots, x_m набрали значення, які задаються вектором $X'_0 = (1, x_{10}, x_{20}, \dots, x_{m0})$. Для випадку парної регресії така довірча зона була побудована у §2 (див. (2.38)).

Узагальнюючи відповідні вирази на випадок множинної регресії, можна отримати довірчий інтервал

$$\hat{y} - t(\gamma; n - m - 1) S_{\hat{y}} < M(Y | x_1, x_2, \dots, x_m) < \hat{y} + t(\gamma; n - m - 1) S_{\hat{y}}, \quad (4.20)$$

де \hat{y} — групова середня, знайдена за рівнянням регресії:

$$\hat{y} = a_0 + a_1 x_{10} + a_2 x_{20} + \dots + a_m x_{m0}; \quad (4.21)$$

$$S_{\hat{y}} = S_e \sqrt{X'_0 (X'X)^{-1} X_0} \quad (4.22)$$

— стандартна помилка для величини \hat{y} .

Довірчий інтервал для **індивідуальних (базисних) значень** залежної змінної y_0 має такий вид:

$$\hat{y}_0 - t(\gamma; n - m - 1) S_{\hat{y}_0} < y_0 < \hat{y}_0 + t(\gamma; n - m - 1) S_{\hat{y}_0}, \quad (4.23)$$

де

$$S_{\hat{y}_0} = S_e \sqrt{X'_0 (X'X)^{-1} X_0}. \quad (4.24)$$

Задача 3.3. На рівні значущості $\alpha = 0,05$ для задачі 3.1 перевірити значущість коефіцієнтів регресії $\alpha_0, \alpha_1, \alpha_2$ та побудувати для них довірчі інтервали з надійністю $\gamma = 0,95$.

○ Знайдемо середні квадратичні помилки коефіцієнтів регресії a_0, a_1, a_2 за формулою (4.17):

$$S_{a_0} = S_e \sqrt{[(X'X)^{-1}]_{00}} = \sqrt{0,0782} \sqrt{3,6924} = 0,5374;$$

$$S_{a_1} = S_e \sqrt{[(X'X)^{-1}]_{11}} = \sqrt{0,0782} \sqrt{1,132} = 0,2975;$$

$$S_{a_2} = S_e \sqrt{[(X'X)^{-1}]_{22}} = \sqrt{0,0782} \sqrt{0,0472} = 0,0608.$$

Тоді спостережні значення критерію:

$$\frac{|a_0|}{S_{a_0}} = \frac{1,3121}{0,5374} = 2,4416, \quad \frac{|a_1|}{S_{a_1}} = \frac{4,8961}{0,2975} = 16,4575, \quad \frac{|a_2|}{S_{a_2}} = \frac{0,1962}{0,0608} = 3,227.$$

Критична точка для двосторонньої критичної області $t_{кр.} = t_{двост.кр.}(\alpha; n - m - 1)$ при значеннях $\alpha = 0,05$, $k = n - m - 1 = 7$ знаходиться за верхньою частиною табл. 3 додатків: $t_{кр.} = 2,365$.

Оскільки, $2,4416 > t_{кр.} = 2,365$, $16,4575 > t_{кр.} = 2,365$ і $3,227 > t_{кр.} = 2,365$, то на рівні значущості $\alpha = 0,05$ робимо висновок, що $a_0 \neq 0$, $a_1 \neq 0$ і $a_2 \neq 0$.

Згідно з (4.19) для параметрів регресії a_0 , a_1 та a_2 матимемо довірчі інтервали:

$$\begin{aligned} -1,3121 - 2,365 \cdot 0,5374 < \alpha_0 < -1,3121 + 2,365 \cdot 0,5374, \\ 4,8961 - 2,365 \cdot 0,2975 < \alpha_1 < 4,8961 + 2,365 \cdot 0,2975, \\ 0,1962 - 2,365 \cdot 0,0608 < \alpha_2 < 0,1962 + 2,365 \cdot 0,0608 \end{aligned}$$

або остаточно

$$-2,5831 < \alpha_0 < -0,0411,$$

$$4,1925 < \alpha_1 < 5,6,$$

$$0,0524 < \alpha_2 < 0,34. \quad \odot$$

6. Під мультиколінеарністю будемо розуміти високу корельованість незалежних змінних. Мультиколінеарність може проявлятися у **функціональній** (явній) і **стохастичній** (прихованій) формах.

При функціональній формі мультиколінеарності хоча б один із парних зв'язків між незалежними змінними є лінійним функціональним зв'язком. У цьому випадку матриця $X'X$ є виродженою, оскільки містить лінійно залежні вектор-стовпці. А тому передумова б регресійного аналізу не виконується. Це приводить до неможливості розв'язування системи нормальних рівнянь і отримання оцінок параметрів регресійної моделі.

Проте в економічних дослідженнях мультиколінеарність частіше проявляється у **стохастичній формі**, коли між хоча б двома пояснюючими змінними існує тісний кореляційний зв'язок. Матриця $X'X$ у цьому випадку є не виродженою, але її визначник — дуже мале число. Разом з тим вектор оцінок a і його коваріаційна матриця Σ_a у відповідності із формулами (4.8) і (4.12) пропорційні матриці $(X'X)^{-1}$, а тому їх елементи обернено пропорційні величині визначника $|X'X|$. В результаті отримують значні середні квадратичні відхилення (стандартні помилки) коефіцієнтів регресії a_0, a_1, \dots, a_m і оцінка їх значущості за критерієм Ст'юдента немає сенсу, хоча у цілому регресійна модель може виявитися значущою за критерієм Фішера-Снедекора.

Більше того, оцінки стають дуже чутливими до незначних змін результатів спостережень і обсягу вибірки. Рівняння регресії у цьому випадку, як правило, не має реального змісту, оскільки деякі із його коефіцієнтів можуть мати неправильні з точки зору економічної теорії знаки і невідповідно великі значення.

Точних кількісних критеріїв для визначення наявності або відсутності мультиколінеарності не існує. Разом з тим є деякі евристичні підходи по виявленню мультиколінеарності та видалення її або зменшення.

Один із таких підходів полягає в аналізі кореляційної матриці між незалежними змінними x_0, x_1, \dots, x_m і виявленні пар змінних, що мають високі коефіцієнти кореляції (по модулю) — як правило, більше 0,8. Якщо такі змінні існують, то говорять про мультиколінеарність між ними. Якщо виявлено таку пару змінних, тоді одну із них виключають із розгляду. При цьому, яку змінну залишити, а яку видалити, вирішують в першу чергу на підставі економічних міркувань. Якщо з економічної точки зору жодній із змінних неможливо віддати перевагу, то залишають ту із змінних, яка має більший коефіцієнт кореляції із залежною змінною.

Корисно також знаходити множинні коефіцієнти детермінації між однією із незалежних змінних та деякою групою решти незалежних змінних. Наявність відносно великого множника коефіцієнта детермінації ($\geq 0,6$) свідчить про мультиколінеарність.

Інший підхід полягає у дослідженні матриці $X'X$. Якщо визначник цієї матриці або її мінімальне власне значення λ_{\min} близькі до нуля (наприклад, одного порядку із накопиченими помилками обчислень), то це свідчить про наявність мультиколінеарності. Про це ж може свідчити і значне

відхилення максимального власного значення λ_{\max} матриці $X'X$ від її мінімального власного значення λ_{\min} .

7. Якість регресійної моделі, як і випадку одно факторних функцій регресії, будемо характеризувати коефіцієнтом детермінації:

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2}.$$

Коефіцієнт детермінації як критерій вибору функції регресії має істотний недолік, який полягає у тому, що з додаванням до регресії нових пояснюючих змінних він ніколи не зменшується, а у більшості випадків, навпаки, збільшується. Це означає, що рівняння з відносно великим числом незалежних змінних будуть давати кращі результати, ніж з відносно малою їх кількістю. Однак з додаванням нового фактора губиться один ступінь вільності, що не завжди є бажаним, оскільки довірчі і прогнознi інтервали будуть тим менші, чим більшим є число ступенів вільності $l = n - m - 1$. Крім того, при застосуванні критеріїв Ст'юдента і Фішера-Снедекора бажано мати можливо більше число ступенів вільності.

Скорегований коефіцієнт детермінації має такий вигляд:

$$\hat{R}^2 = 1 - \frac{\frac{1}{n - m - 1} \sum e_i^2}{\frac{1}{n - 1} \sum (y_i - \bar{y})^2}.$$

Зв'язок коефіцієнтів R^2 та \hat{R}^2 виражається співвідношенням:

$$\hat{R}^2 = 1 - \frac{n - 1}{n - m - 1} (1 - R^2).$$

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ №3

Для торговельного підприємства, яке має велику кількість філій, результуюча змінна y (річний товарообіг однієї філії, млн. грн.) лінійно залежить від x_1 (торгівельної площі, тис. м²) та від x_2 (середньоденної інтенсивності потоку покупців, тисяч людей за день). Для дванадцяти філій за певний рік зафіксовані такі значення показників y , x_1 та x_2 :

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
y_i	2,8	5,2	6,8	7,1	7,3	8,3	4,3	5,8	7,7	3,2	1,5	3,7	5,4	2,6
x_{i1}	0,3	0,9	1,2	1,3	1,2	0,8	0,8	0,9	1,3	0,5	0,3	0,6	1,1	0,2
x_{i2}	1,4	2,7	3,4	3,6	3,7	4,2	2,1	2,9	3,8	1,7	0,8	1,9	2,7	1,3
i	15	16	17	18	19	20	21	22	23	24	25	26	27	28
y_i	5,8	8,1	7,5	8,4	4,2	5,6	7,6	3,5	1,4	3,9	6,4	7,3	7,6	8,3
x_{i1}	0,8	1,6	1,5	0,9	0,9	0,8	1,4	0,7	0,4	0,8	1,1	1,4	1,5	0,9
x_{i2}	2,8	4,1	3,6	4,2	2,1	2,8	3,8	1,7	0,7	1,8	3,2	3,6	3,9	4,1
i	29	30	31	32	33	34	35	36	37	38	39	40	41	42
y_i	4,4	5,6	7,5	3,2	1,5	3,5	2,7	4,3	6,9	7,1	7,3	8,4	4,2	5,8
x_{i1}	0,8	0,7	1,4	0,6	0,3	0,7	0,2	0,6	1,4	1,3	1,2	0,9	0,7	0,7
x_{i2}	2,2	2,9	3,2	1,7	0,8	1,7	1,3	2,2	3,4	3,1	3,2	4,2	4,3	2,9
i	43	44	45	46	47	48	49	50						
y_i	7,5	3,5	1,4	4,2	2,6	5,8	4,2	3,1						
x_{i1}	1,5	0,8	0,3	0,9	0,4	0,7	0,8	0,6						
x_{i2}	3,7	1,8	0,7	2,0	1,4	2,9	2,0	1,6						

Потрібно:

1. Знайти статистичні оцінки параметрів теоретичної множинної лінійної регресії;
2. Обчислити вибірковий коефіцієнт детермінації;
3. Для рівняння значущості $\alpha = 0,05$ перевірити правильність статистичних гіпотез $a_1 = 0$, $a_2 = 0$;
4. З надійністю $\gamma = 0,95$ побудувати довірчі інтервали для параметрів теоретичного рівняння множинної регресії;
5. Знайти прогнозоване значення річного товарообігу для нової філії, введення в дію якої планується у відносно заселеному районі із середньоденною інтенсивністю потоку покупців 15 000 людей в день і торговельною площею 1200 м², а також із надійністю $\gamma = 0,95$ побудувати довірчий інтервал для прогнозованого значення.

§5. НЕЛІНІЙНІ ЕКОНОМЕТРИЧНІ МОДЕЛІ.

У реальних економічних умовах залежність між змінними може адекватно представлятися, як правило, у нелінійній формі. Ця залежність описується формулою

$$Y = f(x) + U,$$

де $f(x)$ — нелінійна функція аргумента x , U — випадковий чинник.

Відповідна економетрична модель має вид:

$$\hat{y} = f(x).$$

Вид економетричної моделі вибирається на основі графічного зображення у системі координат (x, y) статистичної інформації (побудови діаграми розсіювання).

Розглянемо найважливіші нелінійні економетричні моделі.

Гіперболічна (зворотна) залежність має вид

$$\hat{y} = a_0 + \frac{a_1}{x}.$$

Вона зводиться до лінійної заміною $z = \frac{1}{x}$. Одержимо

$$\hat{y} = a_0 + a_1 z.$$

Перевірка моделі на адекватність та побудова прогнозу здійснюється, як і для лінійної моделі, з урахуванням розглянутої заміни змінної x .

Задача 5.1. Використати гіперболічну модель для дослідження залежності собівартості Y (гр.од./шт.) від кількості виготовленої продукції x (шт.). Наведена статистична інформація для показників Y і x :

i	1	2	3	4	5	6	7	8	9	10
y_i	40	37	34	21	29	27	25	24	23	22
x_i	1	2	3	4	5	6	7	8	9	10

Потрібно знайти статистичні оцінки параметрів лінійного рівняння регресії.

○ Статистичні оцінки a_0 , a_1 параметрів α_0 та α_1 гіперболічного рівняння регресії, із врахуванням заміни $z = 1/x$, задовольняють системі нормальних рівнянь (2.12):

$$\begin{cases} a_0 + \bar{z}a_1 = \bar{y}, \\ \bar{z}a_0 + \bar{z}^2 a_1 = \bar{zy}. \end{cases}$$

Для знаходження коефіцієнтів цієї системи складемо розрахункову табл. 5.1.

Таблиця 5.1

i	x_i	y_i	z_i	z_i^2	$z_i y_i$
1	1	40	1	1	40
2	2	37	0,5	0,25	18,5
3	3	34	0,33	0,11	11,33
4	4	21	0,25	0,06	5,25
5	5	29	0,2	0,04	5,8
6	6	27	0,17	0,03	4,5
7	7	25	0,14	0,02	3,57
8	8	24	0,13	0,02	3
9	9	23	0,11	0,01	2,56
10	10	22	0,1	0,01	2,2
Σ		282	2,93	1,55	96,71

Використовуючи нижній рядок табл. 5.1, отримаємо (обсяг вибірки $n = 10$):

$$\bar{z} = \sum_{i=1}^{10} z_i / n = 2,93 / 10 = 0,293; \quad \bar{y} = \sum_{i=1}^{10} y_i / n = 282 / 10 = 28,2;$$

$$\bar{z}^2 = \sum_{i=1}^{10} z_i^2 / n = 1,55 / 10 = 0,155; \quad \bar{zy} = \sum_{i=1}^{10} z_i y_i / n = 96,71 / 10 = 9,671;$$

$$\begin{cases} a_0 + 0,293a_1 = 28,2, \\ 0,293a_0 + 0,155a_1 = 9,671. \end{cases}$$

Розв'язок цієї системи рівнянь згідно із формулами (2.13):

$$a_1 = \frac{\bar{zy} - \bar{z} \cdot \bar{y}}{\bar{z}^2 - (\bar{z})^2} = \frac{9,671 - 0,293 \cdot 28,2}{0,155 - (0,293)^2} = \frac{1,4084}{0,0691} = 20,382,$$

$$a_0 = \bar{y} - a_1 \bar{z} = 28,2 - 20,382 \cdot 0,293 = 22,228.$$

Отже, емпіричне рівняння регресії має такий вигляд:

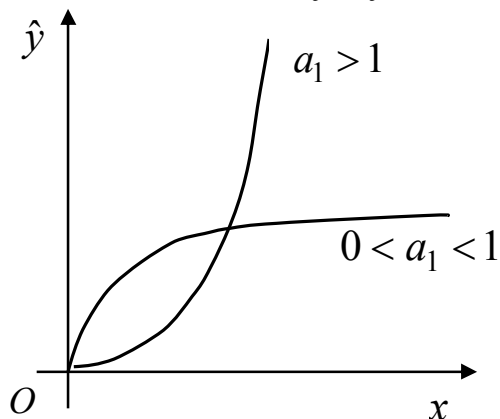
$$\hat{y} = 22,228 + \frac{20,382}{x}.$$

Для знаходження та оцінки значущості коефіцієнтів регресії α_0 та α_1 , точкової оцінки дисперсії збурень, вибіркового коефіцієнта детермінації, коефіцієнта кореляції та побудови для них довірчих інтервалів можна використати розглянуті в §2 методи дослідження лінійної моделі парної регресії. При цьому необхідно здійснити потрібну заміну переходу від нелінійної моделі до лінійної. ◎

Степенева (мультиплікативна) залежність має наступний вид

$$\hat{y} = a_0 \cdot x^{a_1}, \quad a_0 > 0, \quad x > 0.$$

Її графік зображено на малюнку 5.1. Степенева залежність використовується для моделювання ситуацій, в яких ріст витрат x деякого ресурсу обумовлює необмежене збільшення випуску Y .



Малюнок 5.1.

Вона зводиться до лінійної моделі логарифмуванням з довільною основою, наприклад, e . Тоді отримаємо співвідношення

$$\ln \hat{y} = \ln a_0 + a_1 \ln x.$$

Застосуємо такі заміни:

$$\ln \hat{y} = \hat{y}^*, \quad \ln a_0 = a_0^*, \quad \ln x = x^*.$$

Отримаємо рівняння

$$\hat{y}^* = a_0^* + a_1 x^*.$$

Експоненціальна (показникова) модель записується так:

$$\hat{y} = a_0 \cdot a_1^x, \quad a_0 > 0, \quad a_1 > 0, \quad a_1 \neq 1.$$

Для одержання лінійної залежності застосуємо логарифмування. Тоді

$$\ln \hat{y} = \ln a_0 + x \ln a_1.$$

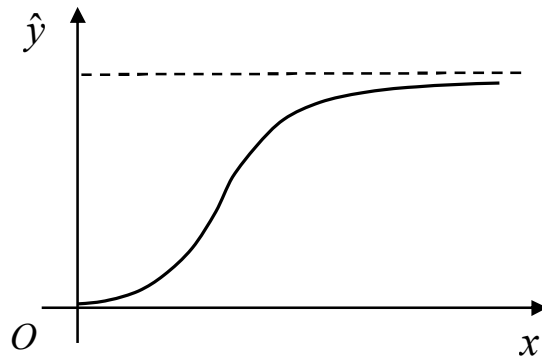
Здійснивши заміну змінних $\ln \hat{y} = \hat{y}^*$, $\ln a_0 = a_0^*$, $\ln a_1 = a_1^*$, отримаємо

$$\hat{y}^* = a_0^* + a_1^* x.$$

Криві з границею росту і точкою перегину часто використовуються для статистичного аналізу попиту на деякі нові товари. Такою кривою є, наприклад, **крива Джонсона**:

$$\hat{y} = e^{a_0 - \frac{a_1}{x}}, \quad a_0 > 0, \quad a_1 > 0.$$

Її графік зображено на малюнку 5.2.



Малюнок 5.2.

Знайдемо логарифми обох частин кривої Джонсона:

$$\ln \hat{y} = a_0 - \frac{a_1}{x}.$$

Замінивши $\ln \hat{y} = \hat{y}^*$, $\frac{1}{x} = z$, одержимо лінійну залежність

$$\hat{y}^* = a_0 - a_1 z.$$

Для моделювання **немонотонних (коливних)** процесів набули широкого використання многочлени (поліноми)

$$\hat{y} = a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m.$$

Якщо всі статистичні значення x_i ($i = 1, 2, \dots, n$) різні, то, як відомо з теорії інтерполяції, через n точок можна єдиним способом привести многочлен степені $n - 1$.

Для одержання лінійної моделі використаємо заміну $x^m = x_m^*$. Одержимо

$$\hat{y} = a_0 + a_1 x_1^* + a_2 x_2^* + \dots + a_m x_m^*.$$

Ця множинна лінійна залежність з числом змінних m , $m < n - 1$.

При дослідженні залежності обсягу податкових надходжень Y від величини податкової ставки x застосовують **криву Лаффера**

$$\hat{y} = a_0 \cdot e^{a_1(x-a_2)^2}.$$

Тут a_0 , a_1 , a_2 — невідомі коефіцієнти, які визначаються на основі статистичної інформації. Логарифмуємо обидві частини цієї залежності. Маємо

$$\ln \hat{y} = \ln a_0 + a_1 x^2 - 2a_1 a_2 x + a_1 a_2^2.$$

Використаємо заміни змінних $\ln \hat{y} = \hat{y}^*$, $\ln a_0 + a_1 a_2^2 = a_0^*$, $-2a_1 a_2 = a_3$. Матимемо многочлен степені 2

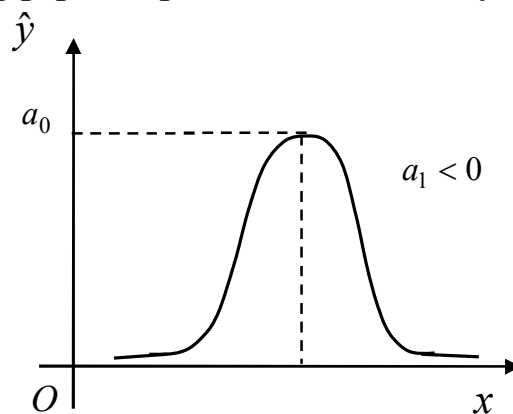
$$\hat{y} = a_0^* + a_3 x + a_1 x^2.$$

Коефіцієнти a_0^* , a_3 , a_1 знаходимо як розв'язок такої системи лінійних рівнянь

$$\begin{cases} a_0^* + a_3 \bar{x} + a_1 \bar{x}^2 = \bar{y}, \\ \bar{x} a_0^* + \bar{x}^2 a_3 + \bar{x}^3 a_1 = \overline{xy}, \\ \bar{x}^2 a_0^* + \bar{x}^3 a_3 + \bar{x}^4 a_1 = \overline{x^2 y}, \end{cases}$$

де $y_i^* = \ln y_i$, n — число статистичних значень кожної із змінних x , y .

Графік кривої Лаффера зображено на малюнку 5.3.



Малюнок 5.3.

Для опису процесів в демографії, маркетингу застосовують **криву Гомперця**

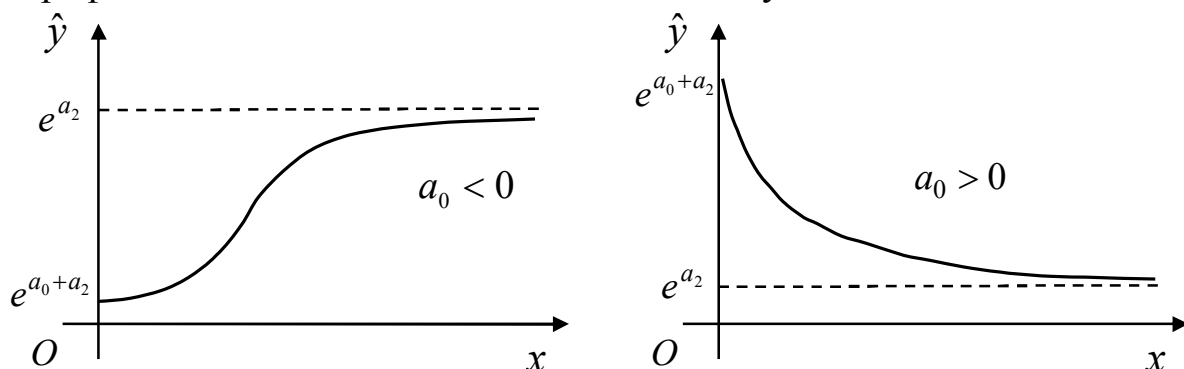
$$\hat{y} = e^{a_0 \cdot a_1^x + a_2}, \quad 0 < a_1 < 1.$$

Логарифмуванням ця крива зводиться до модифікованої експоненціальної моделі

$$\hat{y}^* = a_0 \cdot a_1^x + a_2,$$

де $\hat{y}^* = \ln \hat{y}$.

Графік цієї залежності наведено на малюнку 5.4.

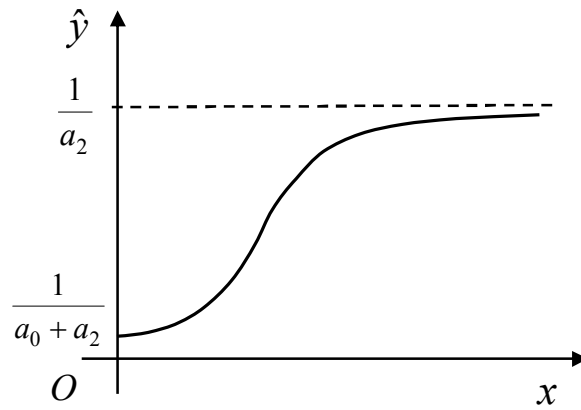


Малюнок 5.4.

Зворотною до модифікованої експоненти є **логістична крива**

$$\hat{y} = \frac{1}{a_0 \cdot a_1^x + a_2}, \quad 0 < a_1 < 1, \quad a_0 > 0, \quad a_2 > 0.$$

Її графік зображено на малюнку 5.5.



Малюнок 5.5.

§6. КОМПЮТЕРНА РЕАЛІЗАЦІЯ МЕТОДІВ ЕКОНОМЕТРІЇ.

Система Excel, що є складовою пакета програм Microsoft Office, є найпоширенішою серед табличних процесорів. Вона має потужні можливості в обчисленні даних у вигляді таблиць, у виконанні бухгалтерських розрахунків, аналізі даних, статистичних обчисленнях. Вона також дає змогу ілюструвати розрахунки графіками та діаграмами.

Розглянемо відповідні можливості цієї системи при розв'язуванні основних типів задач економетрії.

Парна лінійна регресія

На основі статистичних даних фактора x і показника Y із задачі 2.1. потрібно:

- 1) знайти статистичні оцінки параметрів лінійного рівняння регресії;
- 2) точкову оцінку та довірчий інтервал дисперсії збурень із надійністю $\gamma = 0,9$;
- 3) для рівня значущості $\alpha = 0,05$ перевірити значущість коефіцієнтів регресії α_0 та α_1 ;
- 4) знайти довірчі інтервали коефіцієнтів регресії з надійністю $\gamma = 0,95$;
- 5) знайти вибіркові коефіцієнт детермінації, коефіцієнт кореляції, а також інші показники якості лінійної регресії (MAPE, MPE);
- 6) знайти та побудувати довірчу зону функції регресії з надійністю $\gamma = 0,95$;
- 7) на рівні значущості $\alpha = 0,05$ перевірити виконання передумови 2 за тестом Глейзера.

1) Для роботи використовується пакет **Excel**. Блок вихідних даних формується, наприклад, в перших двох стовпцях (**B3:C12**). За блоком вихідних даних іде блок проміжних розрахунків (**D3:G12**).

Для знаходження добутку $x_1 \cdot y_1$ у комірку **F3** вводиться формула **=C3*D3**. Далі копіюємо одержану формулу в інші комірки стовпця **F**. Для **копіювання формули** необхідно: відмітити мишкою комірку **F3**, натиснути праву клавішу мишки та вибрати з меню команду **Копіювати**. Потім відмічаємо блок копіювання (**F4:F12**): ставимо покажчик миші на комірку **F4**, натискаємо ліву клавішу миші і, тримаючи її, рухаємо покажчик до комірки **F12** включно. Можна скопіювати іншим способом: ставимо курсор на комірку **F3**, а покажчик миші наводимо на маленький квадратик в правому нижньому куті комірки. Після перетворення покажчика миші в хрестик на-

тискуємо ліву клавішу миші і відмічаємо діапазон комірок копіювання до **F12** включно. Формули копіюються в помічений блок. Аналогічним чином обчислюється значення x_i^2 та y_i^2 ($i = \overline{1,10}$).

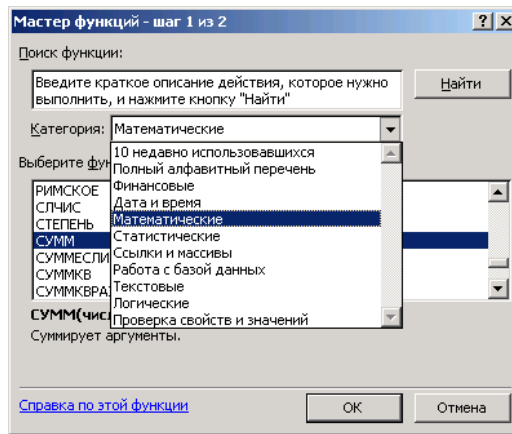
Умова і необхідні обчислення показані на наступних малюнках.

	A	B	C	D	E	F	G	H	I	J
2		i	x_i	y_i	x_i^2	$x_i y_i$	y_i^2			
3		1	0,2	1,5	0,04	0,3	2,25			
4		2	0,3	2,9	0,09	0,87	8,41			
5		3	0,5	3,1	0,25	1,55	9,61			
6		4	0,6	3,2	0,36	1,92	10,24			
7		5	0,8	4,3	0,64	3,44	18,49			
8		6	1	5,7	1	5,7	32,49			
9		7	1,1	5,8	1,21	6,38	33,64			
10		8	1,2	7	1,44	8,4	49			
11		9	1,3	7,2	1,69	9,36	51,84			
12		10	1,4	7,5	1,96	10,5	56,25			
13		Сума	8,4	48,2	8,68	48,42	272,22			
14										
15		$\bar{x} = \sum_{i=1}^{10} x_i / n = 0,84$			$\bar{y} = \sum_{i=1}^{10} y_i / n = 4,82$		$\overline{xy} = \sum_{i=1}^{10} x_i y_i / n = 4,842$			
16										
17										
18										
19		$\overline{x^2} = \sum_{i=1}^{10} x_i^2 / n = 0,868$			$\overline{y^2} = \sum_{i=1}^{10} y_i^2 / n = 27,222$					
20										
21										
22										
23		$a_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2} = 4,884$			$a_0 = \bar{y} - a_1 \bar{x} = 0,717$					
24										
25										

В режимі формул даних малюнок має вигляд:

	A	B	C	D	E	F	G	H	I	J
2		i	x_i	y_i	x_i^2	$x_i y_i$	y_i^2			
3		1	0,2	1,5	=C3^2	=C3*D3	=D3^2			
4		2	0,3	2,9	=C4^2	=C4*D4	=D4^2			
5		3	0,5	3,1	=C5^2	=C5*D5	=D5^2			
6		4	0,6	3,2	=C6^2	=C6*D6	=D6^2			
7		5	0,8	4,3	=C7^2	=C7*D7	=D7^2			
8		6	1	5,7	=C8^2	=C8*D8	=D8^2			
9		7	1,1	5,8	=C9^2	=C9*D9	=D9^2			
10		8	1,2	7	=C10^2	=C10*D10	=D10^2			
11		9	1,3	7,2	=C11^2	=C11*D11	=D11^2			
12		10	1,4	7,5	=C12^2	=C12*D12	=D12^2			
13		Сума	=СУММ(C3:C12)	=СУММ(D3:D12)	=СУММ(E3:E12)	=СУММ(F3:F12)	=СУММ(G3:G12)			
14										
15		$\bar{x} = \sum_{i=1}^{10} x_i / n =$	=C13/B12		$\bar{y} = \sum_{i=1}^{10} y_i / n =$	=D13/B12	$\overline{xy} = \sum_{i=1}^{10} x_i y_i / n =$	=F13/B12		
16										
17										
18										
19		$\overline{x^2} = \sum_{i=1}^{10} x_i^2 / n =$	=E13/B12		$\overline{y^2} = \sum_{i=1}^{10} y_i^2 / n =$	=G13/B12				
20										
21										
22										
23		$a_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2} =$	= (J16-C16*G16)/(C20-C16^2)		$a_0 = \bar{y} - a_1 \bar{x} =$	=G16-D24*C16				
24										
25										

Для **визначення сум** стовпців використовуємо кнопку автосумування на панелі інструментів Σ або вбудовану функцію **СУММ(ДІАПАЗОН КОМІРОК)**. Після встановлення курсору на комірку **C13** натискаємо клавіші **Shift+F3** або кнопку f_x на панелі інструментів. Відкривається вікно **Мастер функций** процесора **Excel**. У категорії активізуємо позицію **математические**, в функції — **СУММ**, і натискаємо клавішу **Enter** або **Ок**.



Діалогове вікно вбудованих функцій

У відкритому вікні відмічаємо діапазон комірок **C3:C12** і натискаємо клавішу **Enter** або **Ок**. Введена формула копіюється в необхідні комірки **I3**-го рядка. Середні значення x , y обчислюються в комірках **C16**, **G16**. Ці ж значення можна обчислити з використанням вбудованої статистичної функції **СРЗНАЧ** (діапазон комірок).

До комірок **D24**, **H24** вводяться формули для визначення оцінок параметрів відповідно a_1 і a_0 .

2) Для знаходження точкової оцінки та довірчого інтервалу дисперсії збурень необхідно виконати наступні обчислення:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
34		i	x_i	y_i	x_i^2	y_i^2	\hat{y}_i	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$	$\frac{\hat{y}_i - y_i}{y_i}$	$\frac{y_i - \hat{y}_i}{x_i}$	$S_{\hat{y}_i}$	$\hat{y}_i - t_{(y, n-2)} S_{\hat{y}_i}$	$\hat{y}_i + t_{(y, n-2)} S_{\hat{y}_i}$	
35		1	0,2	1,5	0,04	2,25	1,6938	-0,1938	0,0376	0,1292	0,1292	0,225444	1,1739	2,2137	
36		2	0,3	2,9	0,09	8,41	2,1822	0,7178	0,5152	-0,2475	0,2475	0,200848	1,7190	2,6454	
37		3	0,5	3,1	0,25	9,61	3,1590	-0,0590	0,0035	0,0190	0,0190	0,157167	2,7966	3,5214	
38		4	0,6	3,2	0,36	10,24	3,6474	-0,4474	0,2002	0,1398	0,1398	0,139814	3,3250	3,9698	
39		5	0,8	4,3	0,64	18,49	4,6242	-0,3242	0,1051	0,0754	0,0754	0,120715	4,3458	4,9026	
40		6	1	5,7	1	32,49	5,6010	0,0990	0,0098	-0,0174	0,0174	0,129247	5,3030	5,8990	
41		7	1,1	5,8	1,21	33,64	6,0894	-0,2894	0,0838	0,0499	0,0499	0,142957	5,7597	6,4191	
42		8	1,2	7	1,44	49	6,5778	0,4222	0,1783	-0,0603	0,0603	0,161076	6,2064	6,9492	
43		9	1,3	7,2	1,69	51,84	7,0662	0,1338	0,0179	-0,0186	0,0186	0,182296	6,6458	7,4866	
44		10	1,4	7,5	1,96	56,25	7,5546	-0,0546	0,003	0,0073	0,0073	0,205657	7,0804	8,0288	
45		Сума	8,4	48,2	8,68	272,2		0,0044	1,1544	0,0768	0,7644				
46															
47															
48															
49															
50															
51															
52															
53															
54															
55															
56															
57															
58															
59															
60															
61															
62															
63															
64															
65															
66															

В режимі формул дана таблиця має вигляд:

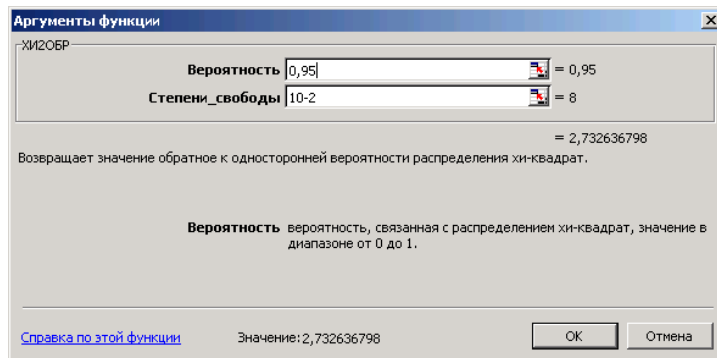
A	B	C	D	E	F	G	H	
33								
34	i	x_i	y_i	x_i^2	y_i^2	\hat{y}_i	$y_i - \hat{y}_i$	
35	1	=C3	=D3	=C35^2	=D36^2	=H\$24+\$D\$24*C35 =D36-G35		
36	2	=C4	=D4	=C36^2	=D36^2	=H\$24+\$D\$24*C36 =D36-G36		
37	3	=C5	=D5	=C37^2	=D37^2	=H\$24+\$D\$24*C37 =D37-G37		
38	4	=C6	=D6	=C38^2	=D38^2	=H\$24+\$D\$24*C38 =D38-G38		
39	5	=C7	=D7	=C39^2	=D39^2	=H\$24+\$D\$24*C39 =D39-G39		
40	6	=C8	=D8	=C40^2	=D40^2	=H\$24+\$D\$24*C40 =D40-G40		
41	7	=C9	=D9	=C41^2	=D41^2	=H\$24+\$D\$24*C41 =D41-G41		
42	8	=C10	=D10	=C42^2	=D42^2	=H\$24+\$D\$24*C42 =D42-G42		
43	9	=C11	=D11	=C43^2	=D43^2	=H\$24+\$D\$24*C43 =D43-G43		
44	10	=C12	=D12	=C44^2	=D44^2	=H\$24+\$D\$24*C44 =D44-G44		
45	Сума	=СУММ(C35:C44)	=СУММ(D35:D44)	=СУММ(E35:E44)			=СУММ(F35:F44)	
46								
47				$S_u^2 = \frac{1}{n-2} \sum_{i=1}^{10} (y_i - \hat{y}_i)^2$	$=1/(10-2)*J45$	$S_u =$	=КОРЕНЬ(F47)	
48								
49	Ліва і права межі довірчого інтервалу для σ_u^2							
50								
51		$\frac{(n-2)S_u^2}{\chi_2^2} =$	$= (10-2)*F47/МИ20БР(0,05;8)$				$= (10-2)*F47/МИ20БР(0,95;8)$	
52								
53								
54		$\left \frac{a_0}{S_{a_0}} \right =$	=H24/O50			i	$\left \frac{a_1}{S_{a_1}} \right =$	
55								
56			$>$					
57			$a_0 - t_0(y, k)S_{a_0} < a_0 < a_0 + t_0(y, k)S_{a_0}$					
58			$<$	α_0				
59		$=H24-F55*O50$			$=H24+F55*O50$			
60								
61								
62	Отже, значення коефіцієнта детермінації							
63								
64								
65				$MAPE = \frac{1}{n} \sum_{i=1}^{10} \left \frac{\hat{y}_i - y_i}{y_i} \right \cdot 100\% =$	$=1/10*L45*100$		$<10\%$	
66								
67								

Продовження розрахункової таблиці:

	I	J	K	L	M	N	O
33							
34	$(y_i - \hat{y}_i)^2$	$\frac{\hat{y}_i - y_i}{y_i}$	$\frac{\hat{y}_i - y_i}{y_i}$	$S_{\hat{y}_i}$	$\hat{y}_i - t(y_i, n-2)S_{y_i}$	$\hat{y}_i + t(y_i, n-2)S_{y_i}$	
35	=H35*2	=(G35-D35)/D36	=ABS(J36)	=\$H\$47*КОРЕНЬ((1+(C35-0,84)^2/\$L\$47)/10)	=G35-\$F\$55*L35	=G35+\$F\$55*L35	
36	=H36*2	=(G36-D36)/D36	=ABS(J36)	=\$H\$47*КОРЕНЬ((1+(C36-0,84)^2/\$L\$47)/10)	=G36-\$F\$55*L36	=G36+\$F\$55*L36	
37	=H37*2	=(G37-D37)/D37	=ABS(J37)	=\$H\$47*КОРЕНЬ((1+(C37-0,84)^2/\$L\$47)/10)	=G37-\$F\$55*L37	=G37+\$F\$55*L37	
38	=H38*2	=(G38-D38)/D38	=ABS(J38)	=\$H\$47*КОРЕНЬ((1+(C38-0,84)^2/\$L\$47)/10)	=G38-\$F\$55*L38	=G38+\$F\$55*L38	
39	=H39*2	=(G39-D39)/D39	=ABS(J39)	=\$H\$47*КОРЕНЬ((1+(C39-0,84)^2/\$L\$47)/10)	=G39-\$F\$55*L39	=G39+\$F\$55*L39	
40	=H40*2	=(G40-D40)/D40	=ABS(J40)	=\$H\$47*КОРЕНЬ((1+(C40-0,84)^2/\$L\$47)/10)	=G40-\$F\$55*L40	=G40+\$F\$55*L40	
41	=H41*2	=(G41-D41)/D41	=ABS(J41)	=\$H\$47*КОРЕНЬ((1+(C41-0,84)^2/\$L\$47)/10)	=G41-\$F\$55*L41	=G41+\$F\$55*L41	
42	=H42*2	=(G42-D42)/D42	=ABS(J42)	=\$H\$47*КОРЕНЬ((1+(C42-0,84)^2/\$L\$47)/10)	=G42-\$F\$55*L42	=G42+\$F\$55*L42	
43	=H43*2	=(G43-D43)/D43	=ABS(J43)	=\$H\$47*КОРЕНЬ((1+(C43-0,84)^2/\$L\$47)/10)	=G43-\$F\$55*L43	=G43+\$F\$55*L43	
44	=H44*2	=(G44-D44)/D44	=ABS(J44)	=\$H\$47*КОРЕНЬ((1+(C44-0,84)^2/\$L\$47)/10)	=G44-\$F\$55*L44	=G44+\$F\$55*L44	
45	=СУММ(H35:H44)	=СУММ(J35:J44)	=СУММ(K35:K44)				
46							
47		$\sigma_x^2 = \overline{x^2} - (\overline{x})^2$				$\sigma_y^2 = \overline{y^2} - (\overline{y})^2$	=H45/10-(D45/10)^2
48							
49							
50							
51							
52							
53							
54							
55	=D24/L50	>	$t_{кр.} =$	=СТЮДРАСПОБР(0,05;8)			
56							
57							
58							
59	=D24-F55*L50		$a_1 - t_1(y, k)S_{a_1} < a_1 < a_1 + t_1(y, k)S_{a_1}$				
60							
61							
62	$R^2 = 1 - \frac{\sum_{i=1}^{10} (y_i - \hat{y}_i)^2}{n\sigma_y^2}$						$r = \pm\sqrt{R^2}$ =КОРЕНЬ(K62)
63							
64							
65							
66							
67							

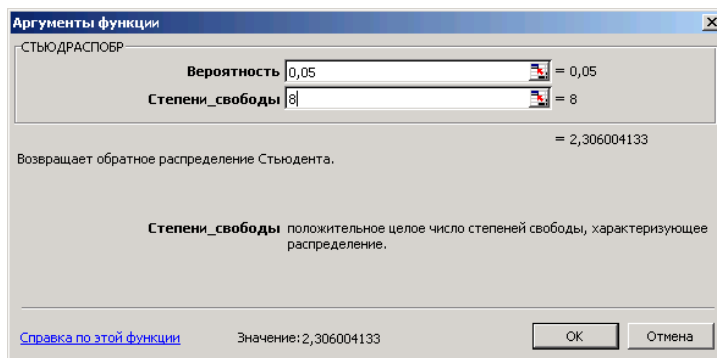
Для обчислення значення \hat{y}_i ($i = \overline{1,10}$) у комірку **G35** вводимо формулу $a_0 + a_1x_1$ (**=H\$24+\$D\$24*C35**) з абсолютним (не змінним, для цього використали знак **\$**) посиланнями координат-параметрів a_0 та a_1 і відносним посиланням координати x_1 (а саме **C35**). Одержану формулу у комірці **G35** копіюємо у блок **G35:G44**.

Незміщену точкову оцінку S_u^2 невідомої дисперсії збурень σ_u^2 обчислено в комірці **F47**, а її ліва і права межі наведено в комірках **D51** та **H51**. При цьому величини χ_1^2 та χ_2^2 при значеннях $k = 10 - 2 = 8$, $p = (1 - 0,9)/2 = 0,05$ і $p = (1 + 0,9)/2 = 0,95$ знаходимо за допомогою вбудованої статистичної функції **XИ2ОБР**.



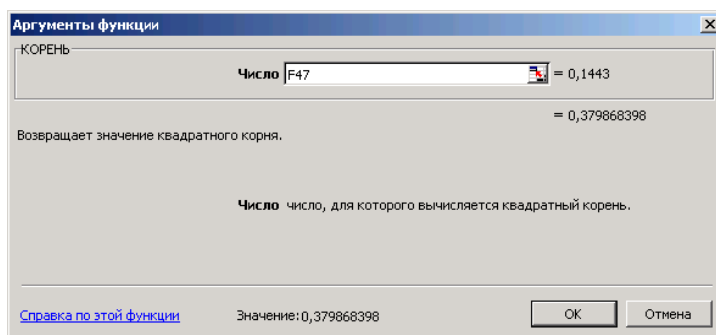
Діалогове вікно знаходження t -статистики

3) Для перевірки значущості коефіцієнтів регресії α_0 та α_1 скористаємося t -статистикою. Значення S_{α_0} та S_{α_1} знайдено в комірках **O50** та **L50** відповідно, а спостережені значення критерію обчислені в комірках **C55** та **I55**. В комірках **F55** та **L55** знайдено критичну точку для двосторонньої критичної області $t_{кр.} = t_{двост.}(\alpha, k)$ при значеннях $\alpha = 0,05$, $k = 10 - 2 = 8$ за допомогою вбудованої статистичної функції **СТЬЮДРАСПОБР**.



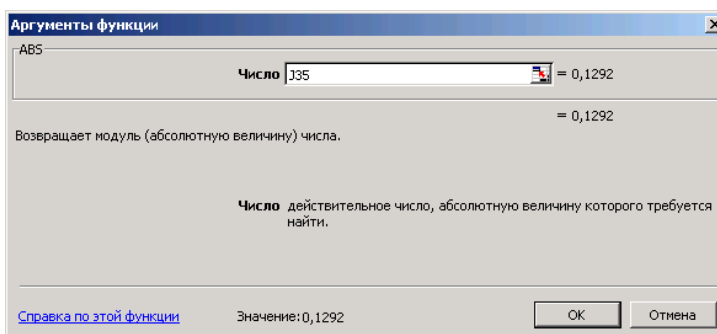
Діалогове вікно знаходження t -статистики

Для обчислення кореня квадратного в комірках **H47**, **L50**, **O50** використовують вбудовану математичну функцію **КОРЕНЬ**:



Діалогове вікно знаходження квадратного кореня з числа

Для обчислення абсолютної величини в діапазоні **K35:K44** використовують вбудовану функцію **ABS**.



Діалогове вікно знаходження абсолютної величини з числа

4) Ліві та праві межі довірчих інтервалів з надійністю $\gamma = 0,95$ для невідомих параметрів регресії a_0 та a_1 знайдено у комірках **B59**, **F59**, **I59** та **M59**.

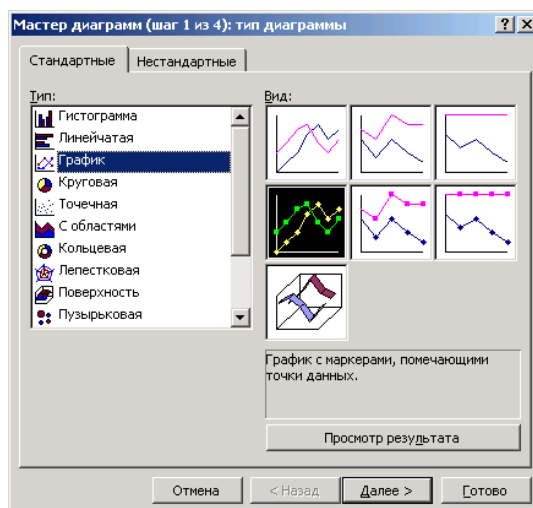
5) Значення вибірових коефіцієнтів детермінації та кореляції наведено в комірках **K62**, **O62**. Нагадаємо, що $a_1 = 4,884 > 0$, то і $r > 0$. Показники якості лінійної регресії **MAPE** та **MPE** знайдено в комірках **G65** та **N65**. Для цього спочатку в комірках **J35** та **K35** були набрані відповідні формули, які для наступних комірок були скопійовані вище описаним методом.

Для побудови довірчої зони для функції регресії спочатку в діапазонах **L35:L44**, **M35:M44** та **N35:N44** знаходимо значення величин $S_{\hat{y}_i}$, $\hat{y}_i - t(\gamma, n - 2)S_{\hat{y}_i}$, $\hat{y}_i + t(\gamma, n - 2)S_{\hat{y}_i}$ ($i = \overline{1,10}$) відповідно.

6) Для наочного уявлення одержаних розрахунків *будуємо графіки*: лінії регресії, довірчу зону Y_{min} та Y_{max} .

Порядок побудови графіків

1. Тримавши натиснутою клавішу **Ctrl**, відмічаємо лівою клавішею миші необхідні для побудови графіків блоки комірок із числовими даними: **D35:D44**; **M35:M44**; **N35:N44**. При переході до іншого блоку комірок ліву клавішу миші відпускаємо.



Діалогове вікно типів діаграм

2. На панелі інструментів наводимо курсор на значок *Мастер диаграм* і натискаємо ліву клавішу мишки. Відкривається перше діалогове вікно *Мастер диаграм (шаг 1 из 4)*. У відкритому вікні вибираємо тип діаграм *Точечная* або *График*, також вибираємо його вид. Для переходу до наступного діалогового вікна натискаємо клавішу *Далее>*.

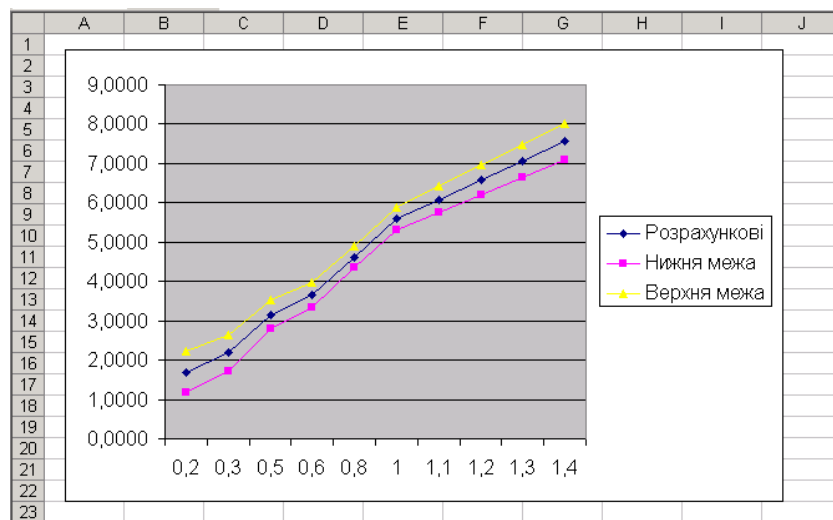
3. У діалоговому вікні *Мастер диаграм (шаг 2 из 4)* активізуємо позицію *Подписи оси X* і відмічаємо курсором діапазон даних фактора *X* (*C35:C44*). Натискаємо клавішу *Далее>*.

4. У діалоговому вікні *Мастер диаграм (шаг 3 из 4)* вводим назви графіку і координати осей.

Натискаємо клавішу *Далее>*.

5. У діалоговому вікні *Мастер диаграм (шаг 4 из 4)* відмічаємо де помістити діаграму: на робочому аркуші, де проведені розрахунки; на окремому аркуші і клацаємо по кнопці *Готово*.

Після натиснення клавіші *Готово* на робочому аркуші з'являється графік.



Для редагування графіка (або його частин) необхідно навести на нього курсор і натиснути 2 рази на ліву клавішу мишки.

7) Для перевірити виконання передумови 2 на рівні значущості $\alpha = 0,05$ за тестом Гейзера потрібно виконати наступні обчислення:

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
2	i	x_i	y_i	x_i^2	y_i^2	$y_i - \hat{y}_i$	$ \mu_i $	$x_i \mu_i$	$ \hat{\mu}_i $	$(\mu_i - \hat{\mu}_i)^2$	$ \hat{\mu}_i $	$(\mu_i - \hat{\mu}_i)^2$	$ \hat{\mu}_i $	$(\mu_i - \hat{\mu}_i)^2$	$ \hat{\mu}_i $	$(\mu_i - \hat{\mu}_i)^2$
3	1	0,2	1,5	0,04	1,6938	-0,1938	0,1938	0,0388	0,4302	0,0559	0,4364	0,0589	0,3508	0,0246	0,3239	0,0169
4	2	0,3	2,9	0,09	2,1822	0,7178	0,7178	0,2153	0,4204	0,0884	0,4327	0,0813	0,3311	0,1495	0,3074	0,1684
5	3	0,5	3,1	0,25	3,1590	-0,0590	0,059	0,0295	0,3892	0,1091	0,4136	0,1258	0,3000	0,0581	0,2832	0,0502
6	4	0,6	3,2	0,36	3,6474	-0,4474	0,4474	0,2684	0,3678	0,0663	0,3959	0,0027	0,2869	0,0258	0,2735	0,0303
7	5	0,8	4,3	0,64	4,6242	-0,3242	0,3242	0,2594	0,3131	0,0001	0,3381	0,0002	0,2635	0,0037	0,2569	0,0045
8	6	1	5,7	1	5,6010	0,0990	0,099	0,099	0,2429	0,0207	0,2429	0,0207	0,2429	0,0207	0,2429	0,0207
9	7	1,1	5,8	1,21	6,0894	-0,2894	0,2894	0,3183	0,2019	0,0077	0,1783	0,0123	0,2334	0,0031	0,2366	0,0028
10	8	1,2	7	1,44	6,5778	0,4222	0,4222	0,5066	0,1571	0,0703	0,1009	0,1033	0,2243	0,0392	0,2307	0,0367
11	9	1,3	7,2	1,69	7,0662	0,1338	0,1338	0,1739	0,1083	0,0007	0,0094	0,0155	0,2156	0,0067	0,2251	0,0083
12	10	1,4	7,5	1,96	7,5546	-0,0546	0,0546	0,0764	0,0556	0,0000	-0,0974	0,0231	0,2072	0,0233	0,2197	0,0273
13	Сума	8,4	48,2	8,68		0,0044	2,7412	1,9858		0,3591	0,4436			0,3547		0,3662
14																
15																
16																
17																
18	$\delta=2$															
19																
20	$\delta=3$															
21																
22	$\delta=0,5$															
23																
24	$\delta=0,3333$															
25																
26																
27																
28																

В режимі формул дана таблиця має вигляд:

	A	B	C	D	E	F	G	H	I
1									$\delta = 2$
2	i	x_i	y_i	x_i^2	\hat{y}_i	$y_i - \hat{y}_i$	$ u_i $	$x_i u_i $	$ \hat{u}_i $
3	1	0,2	1,5	=B3^2	=0,717+4,884*B3	=C3-E3	=ABS(F3)	=B3*ABS(F3)	=ABS(F3)*B3
4	2	0,3	2,9	=B4^2	=0,717+4,884*B4	=C4-E4	=ABS(F4)	=B4*ABS(F4)	=ABS(F4)*B4
5	3	0,5	3,1	=B5^2	=0,717+4,884*B5	=C5-E5	=ABS(F5)	=B5*ABS(F5)	=ABS(F5)*B5
6	4	0,6	3,2	=B6^2	=0,717+4,884*B6	=C6-E6	=ABS(F6)	=B6*ABS(F6)	=ABS(F6)*B6
7	5	0,8	4,3	=B7^2	=0,717+4,884*B7	=C7-E7	=ABS(F7)	=B7*ABS(F7)	=ABS(F7)*B7
8	6	1	5,7	=B8^2	=0,717+4,884*B8	=C8-E8	=ABS(F8)	=B8*ABS(F8)	=ABS(F8)*B8
9	7	1,1	5,8	=B9^2	=0,717+4,884*B9	=C9-E9	=ABS(F9)	=B9*ABS(F9)	=ABS(F9)*B9
10	8	1,2	7	=B10^2	=0,717+4,884*B10	=C10-E10	=ABS(F10)	=B10*ABS(F10)	=ABS(F10)*B10
11	9	1,3	7,2	=B11^2	=0,717+4,884*B11	=C11-E11	=ABS(F11)	=B11*ABS(F11)	=ABS(F11)*B11
12	10	1,4	7,5	=B12^2	=0,717+4,884*B12	=C12-E12	=ABS(F12)	=B12*ABS(F12)	=ABS(F12)*B12
13	Сума	=СУММ(B=СУММ(C3:C12)		=СУММ(D3:D12)		=СУММ(F3:F12)	=СУММ(G3:G12)	=СУММ(H3:H12)	
14		$\frac{\sum(x_i y_i) - \bar{x} \bar{y}}{\sum x_i^2 - (\bar{x})^2}$						$b_0 = \bar{y} - b_1 \bar{x}$	=G13/10-C15*B13/10
15	b_1								
16									
17	$\delta = 2$			$S_e^2 = \frac{1}{n-2} \sum_{i=1}^n (u_i - \hat{u}_i)^2$		=J13/(10-2)		$S_{b_1} = \sqrt{\frac{S_e^2}{n \sigma_x^2}}$	=КОЕФЬ(F18/(10*M15))
18									
19									
20	$\delta = 3$			$S_e^2 = \frac{1}{n-2} \sum_{i=1}^n (u_i - \hat{u}_i)^2$		=L13/(10-2)		$S_{b_1} = \sqrt{\frac{S_e^2}{n \sigma_x^2}}$	=КОЕФЬ(F21/(10*M15))
21									
22									
23	$\delta = 0,5$			$S_e^2 = \frac{1}{n-2} \sum_{i=1}^n (u_i - \hat{u}_i)^2$		=M13/(10-2)		$S_{b_1} = \sqrt{\frac{S_e^2}{n \sigma_x^2}}$	=КОЕФЬ(F24/(10*M15))
24									
25									
26	$\delta = 1/3$			$S_e^2 = \frac{1}{n-2} \sum_{i=1}^n (u_i - \hat{u}_i)^2$		=P13/(10-2)		$S_{b_1} = \sqrt{\frac{S_e^2}{n \sigma_x^2}}$	=КОЕФЬ(F27/(10*M15))
27									
28									

Продовження розрахункової таблиці:

	J	K	L	M	N	O	P
1	2	$\delta = 3$	$\delta = 1/2$	$\delta = 1/2$	$\delta = 1/3$		
2	$(a_i - a_i)^2$	$ a_i $	$(a_i - a_i)^3$	$ a_i $	$(a_i - a_i)^2$	$ a_i $	$(a_i - a_i)^2$
3	=(G3-I3)/2	=\$15+\$C\$15*\$B3*\$B\$21	=(G3-K3)/2	=\$15+\$C\$15*\$B3*\$B\$24	=(G3-M3)/2	=\$15+\$C\$15*\$B3*\$B\$27	=(G3-O3)/2
4	=(G4-I4)/2	=\$15+\$C\$15*\$B4*\$B\$21	=(G4-K4)/2	=\$15+\$C\$15*\$B4*\$B\$24	=(G4-M4)/2	=\$15+\$C\$15*\$B4*\$B\$27	=(G4-O4)/2
5	=(G5-I5)/2	=\$15+\$C\$15*\$B5*\$B\$21	=(G5-K5)/2	=\$15+\$C\$15*\$B5*\$B\$24	=(G5-M5)/2	=\$15+\$C\$15*\$B5*\$B\$27	=(G5-O5)/2
6	=(G6-I6)/2	=\$15+\$C\$15*\$B6*\$B\$21	=(G6-K6)/2	=\$15+\$C\$15*\$B6*\$B\$24	=(G6-M6)/2	=\$15+\$C\$15*\$B6*\$B\$27	=(G6-O6)/2
7	=(G7-I7)/2	=\$15+\$C\$15*\$B7*\$B\$21	=(G7-K7)/2	=\$15+\$C\$15*\$B7*\$B\$24	=(G7-M7)/2	=\$15+\$C\$15*\$B7*\$B\$27	=(G7-O7)/2
8	=(G8-I8)/2	=\$15+\$C\$15*\$B8*\$B\$21	=(G8-K8)/2	=\$15+\$C\$15*\$B8*\$B\$24	=(G8-M8)/2	=\$15+\$C\$15*\$B8*\$B\$27	=(G8-O8)/2
9	=(G9-I9)/2	=\$15+\$C\$15*\$B9*\$B\$21	=(G9-K9)/2	=\$15+\$C\$15*\$B9*\$B\$24	=(G9-M9)/2	=\$15+\$C\$15*\$B9*\$B\$27	=(G9-O9)/2
10	=(G10-I10)/2	=\$15+\$C\$15*\$B10*\$B\$21	=(G10-K10)/2	=\$15+\$C\$15*\$B10*\$B\$24	=(G10-M10)/2	=\$15+\$C\$15*\$B10*\$B\$27	=(G10-O10)/2
11	=(G11-I11)/2	=\$15+\$C\$15*\$B11*\$B\$21	=(G11-K11)/2	=\$15+\$C\$15*\$B11*\$B\$24	=(G11-M11)/2	=\$15+\$C\$15*\$B11*\$B\$27	=(G11-O11)/2
12	=(G12-I12)/2	=\$15+\$C\$15*\$B12*\$B\$21	=(G12-K12)/2	=\$15+\$C\$15*\$B12*\$B\$24	=(G12-M12)/2	=\$15+\$C\$15*\$B12*\$B\$27	=(G12-O12)/2
13	=СУММ(J3:J12)		=СУММ(L3:L12)		=СУММ(N3:N12)		=СУММ(P3:P12)
14							
15			$\sigma_x^2 = \overline{x^2} - (\overline{x})^2$	=D13/10-(B13/10)^2			
16							
17							
18	$\frac{b_1}{S_{b_1}}$	=ABS(C15/I18)	< $t_{деостар}$	$(0,05;8)$	=	=СТЮДРАСПОБР(0,05;10-2)	
19							
20							
21	$\frac{b_1}{S_{b_1}}$	=ABS(C15/I21)	< $t_{деостар}$	$(0,05;8)$	=	=СТЮДРАСПОБР(0,05;10-2)	
22							
23							
24	$\frac{b_1}{S_{b_1}}$	=ABS(C15/I24)	< $t_{деостар}$	$(0,05;8)$	=	=СТЮДРАСПОБР(0,05;10-2)	
25							
26							
27	$\frac{b_1}{S_{b_1}}$	=ABS(C15/I27)	< $t_{деостар}$	$(0,05;8)$	=	=СТЮДРАСПОБР(0,05;10-2)	
28							

В комірках **C15**, **G15** знайдено МНК-оцінки b_1 та b_0 . Точкову оцінку S_ε^2 невідомої дисперсії збурень σ_ε^2 знайдено в комірці **F18**, попередньо обчисливши величини \hat{u}_i ($i = \overline{1,10}$, $\delta = 2$) в діапазоні **I3:I12**. Для цього у комірці **I3** вводимо формулу $b_0 + b_1 x_1^\delta$ (**=G\$15+C\$15*B3^B\$18**) з абсолютним посиленням координат-параметрів b_0 , b_1 та δ і відносним посиленням координати x_1 (а саме **B3**). Одержану формулу у комірці **I3** копіюємо у блок **I3:I12**. Для визначення значущості β_1 порівнюємо знайдені в

комірках **K18** та **O18** величину $\left| \frac{b_1}{S_{b_1}} \right|$ та t -статистику відповідно.

За аналогією проводимо обчислення для значень $\delta = 3$, $\delta = 1/2$, $\delta = 1/3$. Виявляється, що у всіх випадках коефіцієнт регресії β_1 незначущий, тобто для статистичних даних задачі передумова 2 виконується.

Матрична форма МНК для оцінки параметрів множинної регресії

Допустимо, що між показником y і факторами x_1, x_2 існує лінійна залежність $\hat{y} = X' a = a_0 + a_1 x_1 + a_2 x_2$, де $X' = (1, x_1, x_2)$, $a' = (a_0, a_1, a_2)$.

Оцінки параметрів вектора a шукатимемо за формулу $a = (X'X)^{-1} X'Y$.

Порядок знаходження оцінок параметрів регресії:

1. Знаходимо транспоновану матрицю X' в блоці по відношенню до матриці X в блоці, використовуючи в категорії «Ссылки и массивы» вбудовану функцію **ТРАНСП**.

2. Знаходимо добуток матриць $X'X$ в виділеному блоці, використовуючи вбудовану математичну функцію **МУМНОЖ** (блок даних першої матриці; блок даних другої матриці).

3. Обернену матрицю $(X'X)^{-1}$ знаходимо в іншому виділеному блоці, використовуючи вбудовану математичну функцію **МОБР**.

4. Добуток матриць $X'Y$ знаходимо, використовуючи вбудовану математичну функцію **МУМНОЖ**, виділивши перед тим масив, в якому буде знайдений добуток матриць.

5. Оцінки вектора знаходимо в виділеному для цього блоці, використовуючи вбудовану математичну функцію **МУМНОЖ** (блок даних матриці $(X'X)^{-1}$; блок даних матриці $X'Y$).

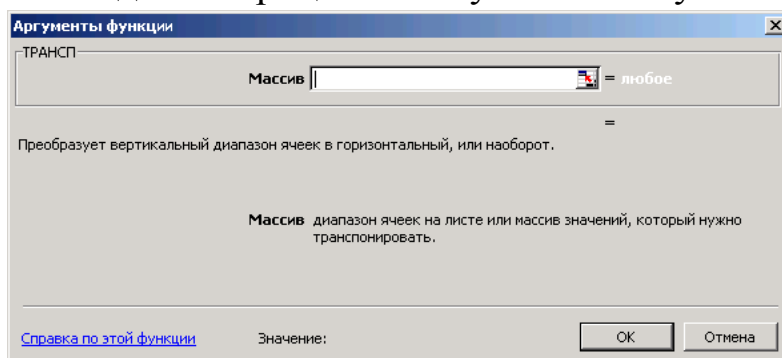
6. Для перевірки значущості параметрів регресії розрахуємо t -статистику кожного із параметрів за формулою

$$|t_{\text{спост.}}| = \frac{|a_i|}{S_{a_i}} > t_{\text{кр.}},$$

де $S_{a_i} = S_e \sqrt{[(X'X)^{-1}]_{ii}}$, $i = \overline{0,2}$, S_e — середньоквадратичне відхилення статистичних даних від розрахункових.

Порядок знаходження транспонованої матриці в ET Excel

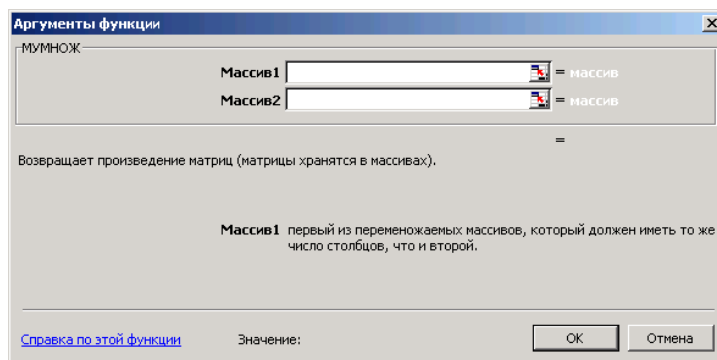
1. Відмічаємо діапазон комірок, де має знаходитись транспонована матриця.
2. Відкриваємо діалогове вікно **Мастер функций**, вибираємо функцію **ТРАНСП** у полі категорій **Ссылки и массивы** і натискаємо на кнопку **ОК** для переходу в наступне діалогове вікно.
3. У другому діалоговому вікні відмічаємо діапазон комірок, у яких знаходяться елементи вихідної матриці і натискаємо клавішу **ОК**.



4. Натискаємо клавішу **F2**, а потім одночасно клавіші **Ctrl+Shift+Enter**.

Порядок знаходження добутку двох матриць в ET Excel

1. Відмічаємо діапазон комірок, де має знаходитись матриця, яка є результатом добутку двох матриць. розмір цієї матриці визначається розмірами матриць, які перемножуються. Якщо, наприклад, розмір матриці X' — 3×15 , розмір матриці X — 15×3 , тоді розмір матриці $X'X$ буде 3×3 . (якщо $C = AB$, де розмір матриць: A — $(m \times k)$, B — $(k \times n)$, то розмір матриці C — $(m \times n)$).
2. Відкриваємо діалогове вікно **Мастер функций** (натискаємо на f_x), вибираємо функцію **МУМНОЖ** у полі категорії **Математические** і натискаємо на кнопку **ОК** для переходу у наступне діалогове вікно.
3. У діалоговому вікні **МУНОЖ** відмічаємо: у першому рядку діапазон комірок, в яких знаходяться елементи першої матриці, у другому рядку діапазон комірок, в яких знаходяться елементи другої матриці. Натискаємо клавішу **ОК**.



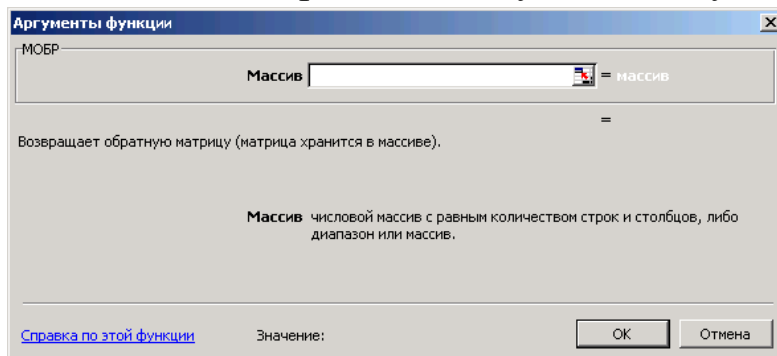
4. Натискаємо клавішу F_2 і одночасно клавіші $Ctrl+Shift+Enter$.

Порядок знаходження оберненої матриці в ET Excel

1. Відмічаємо діапазон комірок, де має знаходитись обернена матриця.

2. Відкриваємо діалогове вікно **Мастер функций** (натискаємо на кнопку f_x на панелі інструментів) Вибираємо функцію **МОБР** у полі категорій **Математические** і натискаємо на кнопку ОК для переходу у діалогове вікно **МОБР**.

3. У діалоговому вікно **МОБР** вказуємо діапазон комірок, у яких знаходяться елементи вихідної матриці. Натискаємо кнопку ОК.



4. Натискаємо клавішу F_2 і одночасно клавіші $Ctrl+Shift+Enter$.

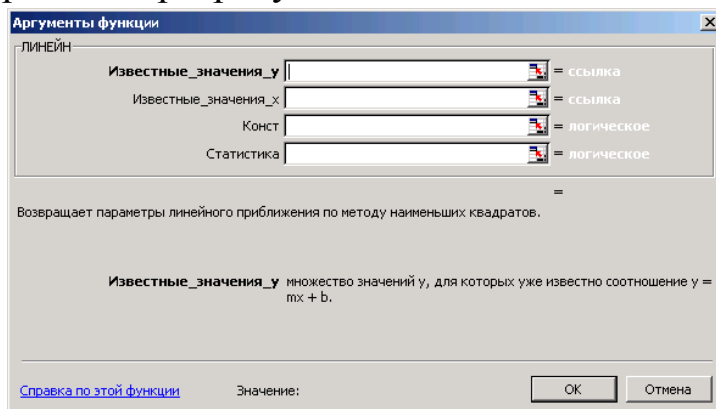
Оцінки параметрів регресії можна знайти, використовуючи вбудовану статистичну функцію **ЛИНЕЙН.**

Опишемо порядок знаходження оцінок параметрів регресії з використанням функції **ЛИНЕЙН**:

1. Відмічаємо блок, де мають знаходитись розрахункові дані: ширина блоку дорівнює числу оцінюваних параметрів, а висота дорівнює п'яти рядкам.

2. Відкриваємо діалогове вікно **Мастер функций**, вибираємо функцію **ЛИНЕЙН** у полі категорії **СТАТИСТИЧЕСКИЕ** і натискаємо на кнопку **Далее>** для переходу в наступне діалогове вікно.

3. У другому діалоговому вікні вводимо: в перший рядок (в перше поле) блок даних результативного показника, вказуючи діапазон комірок або ім'я блоку даних; у другий рядок — блок даних факторів або ім'я цього блоку; в третій рядок вводиться слово **ИСТИНА**, якщо a_0 не дорівнює нулю, і слово **ЛОЖЬ**, якщо a_0 дорівнює нулю; в четвертий рядок вводиться слово **ИСТИНА**, якщо необхідно знайти не лише параметри лінії регресії, а й додаткову регресійну статистику. Якщо необхідно знайти лише параметри лінії регресії, то вводимо слово **ЛОЖЬ** і натискаємо на кнопку **ГОТОВО** для отримання розрахункових даних.



4. Для того щоб у блоці розрахункових даних було видно не лише значення першої комірки, натискаємо клавішу **F2**, потім **Ctrl+Shift+Enter**.

Таблиця розрахункових значень додаткової регресійної статистики має вигляд:

Розміщення значень додаткової регресійної статистики

a_2	a_1	a_0
σ_{a_2}	σ_{a_1}	σ_{a_0}
R^2	S	#Н/Д
F_{r_1}	k	#Н/Д
$\sum(\hat{y} - \bar{y})^2$	$\sum(\hat{y}_i - \bar{y}_i)^2$	#Н/Д

Опишемо розрахункові дані:

У **першому** рядку справа наліво знаходяться оцінки параметрів множинної лінійної регресії відповідно a_0, a_1, a_2 .

У **другому** рядку справа наліво знаходяться середні квадратичні відхилення оцінок параметрів $\sigma_{a_0}, \sigma_{a_1}, \sigma_{a_2}$.

У **третьому** рядку в першій комірці знаходиться коефіцієнт детермінації, а в другій комірці — середнє квадратичне відхилення показника.

У четвертому рядку в першій комірці знаходиться розрахункове значення F -статистики, в другій комірці знаходиться k — число ступенів вільності.

У п'ятому рядку в першій комірці знаходиться сума квадратів відхилень розрахункових значень показника від його середнього значення, в другій комірці — залишкова сума квадратів.

ДОДАТКИ

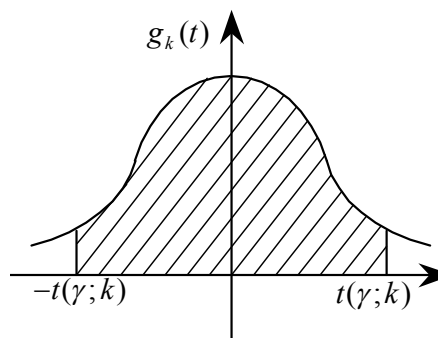
Таблиця 1

$$\text{Значення функції Лапласа } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

<i>x</i>	Соті долі <i>x</i>										
	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	
0,0	0,00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	
1,1	36433	36650	36864	37076	37285	37493	37698	37900	38100	38298	
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	
1,3	40320	40490	40658	40824	40988	41149	41308	41466	41621	41774	
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	
1,6	44520	44630	44738	44845	44950	45053	45154	45254	44352	45449	
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	
2,9	49813	49819	49825	49831	49835	49841	49846	49851	49856	49861	
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	
3,5	49977	49978	49978	49979	49980	49981	48881	49982	49983	49983	
3,6	49984	49985	40085	49986	49986	49987	49987	49988	49988	49989	
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992	
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	
<i>x</i>	Десяті долі <i>x</i>										
	<i>0</i>	<i>2</i>			<i>4</i>			<i>6</i>		<i>8</i>	
4,	0,4999683	4999867			4999946			4999979		4999992	
5,	4999997										

Значення $t = t(\gamma; k)$,
 що задовільняють рівнянню

$$P(|T| < t) = 2 \int_0^t g_k(t) dt = \gamma,$$
 де $g_k(t)$ — густина розподілу
 Ст'юдента (t -розподілу), $k = n - 1$
 — число ступенів вільності



$k = n - 1$	γ				
	0,9	0,95	0,98	0,99	0,999
1	6,314	12,706	31,821	63,657	636,619
2	2,920	4,303	6,965	9,925	31,599
3	2,353	3,182	4,541	5,841	12,924
4	2,132	2,776	3,747	4,604	8,610
5	2,015	2,571	3,365	4,032	6,869
6	1,943	2,447	3,143	3,707	5,969
7	1,895	2,365	2,998	3,499	5,408
8	1,860	2,306	2,896	3,355	5,041
9	1,833	2,262	2,821	3,250	4,781
10	1,812	2,228	2,764	3,169	4,587
11	1,796	2,201	2,718	3,106	4,437
12	1,782	2,179	2,681	3,055	4,318
13	1,771	2,160	2,650	3,012	4,221
14	1,761	2,145	2,624	2,977	4,140
15	1,753	2,131	2,602	2,947	4,073
16	1,746	2,120	2,583	2,921	4,015
17	1,740	2,110	2,567	2,898	3,965
18	1,734	2,101	2,552	2,878	3,922
19	1,729	2,093	2,539	2,861	3,883
20	1,725	2,086	2,528	2,845	3,850
25	1,708	2,060	2,485	2,785	3,725
30	1,697	2,042	2,457	2,750	3,646
40	1,684	2,021	2,423	2,704	3,551
50	1,676	2,009	2,403	2,678	3,496
60	1,671	2,000	2,390	2,660	3,460
70	1,667	1,994	2,381	2,648	3,435
80	1,664	1,990	2,374	2,639	3,416
90	1,662	1,987	2,368	2,632	3,402
100	1,660	1,984	2,364	2,626	3,390
120	1,658	1,980	2,358	2,617	3,373
∞	1,645	1,960	2,326	2,576	3,291

Таблиця 3

Критичні точки розподілу Ст'юдента (t -розподілу)

Для двосторонньої критичної області критична точка $t_{\text{двост.кр}}(\alpha; k) = t_{\alpha}$ є коренем

рівняння $\int_0^{t_{\alpha}} g_k(t) dt = (1 - \alpha)/2$; для односторонньої (правосторонньої) критичної об-

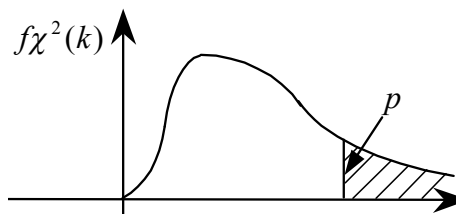
ласті точка $t_{\text{правост.кр}}(\alpha; k) = t_{2\alpha}$ є коренем рівняння $\int_{t_{2\alpha}}^{\infty} g_k(t) dt = (1 - 2\alpha)/2$, де

$g_k(t)$ — густина розподілу Ст'юдента, $k = n - 1$ — число ступенів вільності. Для лі-
восторонньої критичної області $t_{\text{лівост.кр}}(\alpha; k) = -t_{2\alpha}$.

Число ступенів вільності $k = n - 1$	Рівень значущості α (двостороння критична область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,314	12,71	31,82	63,66	318,3	637,0
2	2,920	4,303	6,965	9,925	22,33	31,60
3	2,353	3,182	4,541	5,841	10,22	12,94
4	2,132	2,776	3,747	4,604	7,173	8,610
5	2,015	2,571	3,365	4,032	5,893	6,859
6	1,943	2,447	3,143	3,707	5,208	5,959
7	1,895	2,365	2,998	3,499	4,785	5,405
8	1,860	2,306	2,896	3,355	4,501	5,041
9	1,833	2,262	2,821	3,250	4,297	4,781
10	1,812	2,228	2,764	3,169	4,144	4,587
11	1,796	2,201	2,718	3,106	4,025	4,437
12	1,782	2,179	2,681	3,055	3,930	4,318
13	1,771	2,160	2,650	3,012	3,852	4,221
14	1,761	2,145	2,624	2,977	3,787	4,140
15	1,753	2,131	2,602	2,947	3,733	4,073
16	1,746	2,120	2,583	2,921	3,686	4,015
17	1,740	2,110	2,567	2,898	3,646	3,965
18	1,734	2,101	2,552	2,878	3,611	3,922
19	1,729	2,093	2,539	2,861	3,579	3,883
20	1,725	2,086	2,528	2,845	3,562	3,850
21	1,721	2,080	2,518	2,831	3,527	3,819
22	1,717	2,074	2,508	2,819	3,505	3,792
23	1,714	2,069	2,500	2,807	3,485	3,767
24	1,711	2,064	2,492	2,797	3,467	3,745
25	1,708	2,060	2,485	2,787	3,450	3,725
26	1,706	2,056	2,479	2,779	3,435	3,707
27	1,703	2,052	2,473	2,771	3,421	3,690
28	1,701	2,048	2,467	2,763	3,408	3,674
29	1,699	2,045	2,462	2,756	3,396	3,659
30	1,697	2,042	2,457	2,750	3,385	3,646
40	1,684	2,021	2,423	2,704	3,307	3,551
60	1,671	2,000	2,390	2,660	3,232	3,460
120	1,658	1,981	2,362	2,624	3,172	3,374
∞	1,645	1,960	2,326	2,576	3,090	3,291
Число ступенів вільності $k = n - 1$	Рівень значущості α (одностороння критична область)					
	0,05	0,025	0,01	0,005	0,001	0,0005

Таблиця 4

Значення
 $P(\chi^2(k) > \chi^2(p; k)) = p,$
де k — число ступенів вільності



k	p							
	0,999	0,99	0,95	0,90	0,10	0,05	0,01	0,001
1	0,157·10 ⁻⁵	0,0002	0,004	0,02	2,71	3,84	6,63	10,83
2	0,002	0,02	0,10	0,21	4,61	5,99	9,21	13,82
3	0,02	0,12	0,35	0,58	6,25	7,82	11,34	16,27
4	0,09	0,30	0,71	1,06	7,78	9,49	13,28	18,47
5	0,21	0,55	1,15	1,61	9,24	11,07	15,08	20,51
6	0,38	0,87	1,64	2,20	10,65	12,59	16,81	22,46
7	0,60	1,24	2,17	2,83	12,02	14,06	18,48	24,32
8	0,86	1,65	2,73	3,49	13,36	15,51	20,09	26,12
9	1,15	2,09	3,33	4,17	14,68	16,92	21,67	27,88
10	1,48	2,56	3,94	4,87	15,99	18,31	23,21	29,59
11	1,83	3,05	4,58	5,58	17,28	19,68	24,72	31,26
12	2,21	3,57	5,23	6,30	18,55	21,03	26,22	32,91
13	2,62	4,11	5,89	7,04	19,81	22,36	27,69	34,53
14	3,04	4,66	6,57	7,79	21,06	23,69	29,14	36,12
15	3,48	5,23	7,26	8,55	22,31	25,00	30,58	37,70
16	3,94	5,81	7,96	9,31	23,54	26,30	32,00	39,25
17	4,42	6,41	8,67	10,09	24,77	27,59	33,41	40,79
18	4,90	7,02	9,39	10,86	25,99	28,87	34,81	42,31
19	5,41	7,63	10,12	11,65	27,20	30,14	36,19	43,82
20	5,92	8,26	10,85	12,44	28,41	31,41	37,57	45,31
21	6,45	8,90	11,59	13,24	29,62	32,67	38,93	46,80
22	6,98	9,54	12,34	14,04	30,81	33,92	40,29	48,27
23	7,53	10,20	13,20	14,85	32,01	35,17	41,64	49,73
24	8,08	10,86	13,85	15,66	33,19	36,42	43,98	51,18
25	8,65	11,52	14,61	16,47	34,38	37,65	44,31	52,62
26	9,22	12,20	15,37	17,29	35,56	38,89	45,64	54,05
27	9,80	12,88	16,15	18,11	36,74	40,11	46,96	55,48
28	10,39	13,56	16,93	18,94	37,92	41,34	48,28	56,89
29	10,99	14,26	17,71	19,77	39,09	42,56	49,59	58,30
30	11,59	14,95	18,49	20,60	40,26	43,77	50,89	59,70
40	17,92	22,16	26,51	29,05	51,81	55,76	63,69	73,40
50	24,67	29,71	34,76	37,69	63,17	67,51	76,15	86,66
100	61,92	70,07	77,93	82,36	118,50	124,34	135,81	149,45

Таблиця 5

Критичні точки $F_{кр}(\alpha; k_1, k_2)$ розподілу Фішера-Снедекора,
що задовільняють рівнянню $P[F > F_{кр}(\alpha; k_1, k_2)] = \alpha$ при $\alpha = 0,05$

k_2	k_1									
	1	2	3	4	5	6	8	12	24	∞
1	161,45	199,50	215,71	224,58	230,16	233,99	238,88	243,91	249,05	254,32
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,02	1,83	1,61	1,25
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

Таблиця 6

Критичні значення $K_{n,\alpha}$ для статистики критерія Колмогорова

n	α				n	α			
	0,10	0,05	0,02	0,01		0,10	0,05	0,02	0,01
1	0,950	0,975	0,990	0,995	51	0,168	0,187	0,208	0,224
2	776	842	900	929	52	166	185	207	222
3	636	708	789	829	53	165	183	205	220
4	565	624	689	734	54	163	181	203	218
5	510	563	627	669	55	162	180	201	216
6	468	519	577	617	56	160	178	199	214
7	436	483	538	576	57	159	177	198	212
8	410	454	507	542	58	158	175	196	210
9	388	430	480	513	59	156	174	194	208
10	369	409	457	489	60	155	172	193	207
11	352	391	437	468	61	154	171	191	205
12	338	375	419	449	62	153	170	190	203
13	326	361	404	433	63	151	168	188	202
14	314	349	390	418	64	150	167	187	200
15	304	338	377	404	65	149	166	185	199
16	295	327	366	392	66	148	164	184	197
17	286	318	355	381	67	147	163	183	196
18	279	309	346	371	68	146	162	180	194
19	271	301	337	361	69	145	161	181	193
20	265	294	329	352	70	144	160	179	192
21	259	287	321	344	71	143	159	177	190
22	253	281	314	337	72	142	158	176	189
23	248	275	307	330	73	141	157	175	188
24	242	269	301	323	74	140	155	174	187
25	238	264	295	317	75	139	154	173	185
26	233	259	290	311	76	138	153	172	184
27	229	254	284	305	77	137	152	171	183
28	225	250	279	300	78	136	152	169	182
29	221	246	275	295	79	136	151	168	181
30	218	242	270	290	80	135	150	167	180
31	215	238	266	285	81	134	149	166	178
32	211	234	262	281	82	133	148	165	177
33	208	231	258	277	83	132	147	164	176
34	205	227	254	273	84	132	146	163	175
35	202	224	251	269	85	131	145	162	174
36	199	221	247	265	86	130	144	161	173
37	196	218	244	262	87	129	144	161	172
38	194	215	241	258	88	129	143	160	171
39	192	213	238	255	89	128	142	159	170
40	189	210	235	252	90	127	141	158	169
41	187	208	232	249	91	126	140	157	169
42	185	205	229	246	92	126	140	156	168
43	183	203	227	243	93	125	139	155	167
44	181	201	224	241	94	124	138	155	166
45	179	198	222	238	95	124	138	154	165
46	177	196	219	235	96	123	137	153	164
47	175	194	217	233	97	123	136	152	163
48	173	192	215	231	98	122	135	151	162
49	171	190	213	228	99	121	135	151	162
50	170	188	211	226	100	121	134	150	161

Таблиця 7

Асимптотичні критичні значення $n_{1-\alpha}$
для статистики Колмогорова і Смирнова

α	0,20	0,10	0,05	0,02	0,01
$n_{1-\alpha}$	1,073	1,224	1,358	1,517	1,628

Таблиця 8

Перетворення Фішера (Z-перетворення)

$$Z = \frac{1}{2} \ln \frac{1+r_e}{1-r_e}$$

r_e	Соті долі r_e									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0100	0,0200	0,0300	0,0400	0,0501	0,0601	0,0701	0,0802	0,0902
0,1	0,1003	0,1105	0,1206	0,1308	0,1409	0,1511	0,1614	0,1717	0,1820	0,1923
0,2	0,2027	0,2132	0,2237	0,2342	0,2448	0,2561	0,2661	0,2769	0,2877	0,2986
0,3	0,3095	0,3206	0,3317	0,3428	0,3541	0,3654	0,3769	0,3884	0,4001	0,4118
0,4	0,4236	0,4356	0,4477	0,4599	0,4722	0,4847	0,4973	0,5101	0,5230	0,5361
0,5	0,5493	0,5627	0,5763	0,5901	0,6042	0,6184	0,6328	0,6475	0,6625	0,6777
0,6	0,6931	0,7089	0,7250	0,7414	0,7582	0,7753	0,7928	0,8107	0,8291	0,8480
0,7	0,8673	0,8872	0,9076	0,9287	0,9505	0,9730	0,9962	1,0203	1,0454	1,0714
0,8	1,0986	1,1270	1,1568	1,1881	1,2212	1,2562	1,2933	1,3331	1,3758	1,4219
0,9	1,4722	1,5275	1,5890	1,6584	1,7380	1,8318	1,9459	2,0923	2,2976	2,6467

Обернене перетворення Фішера $r_s = thZ$.

Z	Соті долі Z									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0100	0,0200	0,0300	0,0400	0,0500	0,0599	0,0699	0,0799	0,0898
0,1	0997	1096	1194	1293	1391	1489	1586	1684	1781	1877
0,2	1974	2070	2165	2260	2355	2449	2543	2636	2729	2821
0,3	2913	3004	3095	3185	3275	3364	3452	3540	3627	3714
0,4	3800	3885	3969	4053	4136	4219	4301	4382	4462	4542
0,5	4621	4699	4777	4854	4930	5005	5080	5154	5227	5299
0,6	5370	5441	5511	5580	5649	5717	5784	5850	5915	5980
0,7	6044	6107	6169	6231	6291	6351	6411	6469	6527	6584
0,8	6640	6696	6751	6805	6858	6911	6963	7014	7064	7114
0,9	7163	7211	7259	7306	7352	7398	7443	7487	7531	7574
1,0	7616	7658	7699	7739	7779	7818	7857	7895	7932	7969
1,1	8005	8041	8076	8110	8144	8178	8210	8243	8275	8306
1,2	8337	8367	8397	8426	8455	8483	8511	8538	8565	8591
1,3	8617	8643	8668	8692	8717	8741	8764	8787	8810	8832
1,4	8854	8875	8896	8917	8937	8957	8977	8996	9015	9033
1,5	9051	9069	9087	9104	9121	9138	9154	9170	9186	9201
1,6	9217	9232	9246	9261	9275	9289	9302	9316	9329	9341
1,7	9354	9366	9379	9391	9402	9414	9425	9436	9447	9458
1,8	9468	9478	9488	9498	9508	9518	9527	9536	9545	9554
1,9	9562	9571	9579	9587	9595	9603	9611	9618	9626	9633
2,0	9640	9647	9654	9661	9668	9674	9680	9686	9693	9699
2,1	9704	9710	9716	9722	9727	9732	9738	9743	9748	9753
2,2	9757	9762	9767	9771	9776	9780	9785	9789	9793	9797
2,3	9801	9805	9809	9812	9816	9820	9823	9827	9830	9834
2,4	9837	9840	9843	9846	9849	9852	9855	9858	9861	9864
2,5	9866	9869	9871	9874	9876	9879	9881	9884	9886	9888
2,6	9890	9892	9894	9897	9899	9901	9903	9904	9906	9908
2,7	9910	9912	9914	9915	9917	9919	9920	9922	9923	9925
2,8	9926	9928	9929	9931	9932	9933	9935	9936	9937	9938
2,9	9940	9941	9942	9943	9944	9945	9946	9947	9948	9949

ЛІТЕРАТУРА

1. Грубер Й. Эконометрия. Том 1. Введение в эконометрию. — К.: Астара, 1996. — 397 с.
2. Грубер Й. Эконометрия. Том 2. Эконометричні прогнозні та оптимізаційні моделі. — К.: ЗАТ «Нічлава», 1999. — 296 с.
3. Єрмоменко В.О., Шинкарик М.І. Теорія імовірностей: Навчальний посібник. — Тернопіль: Економічна думка, 2002. — 176 с.
4. Єрмоменко В.О., Шинкарик М.І. Математична статистика: Навчальний посібник. — Тернопіль: Економічна думка, 2003. — 247 с.
5. Іващук О.Т. Эконометричні методи та моделі. — Тернопіль: Економічна думка, 2002. — 348 с.
6. Кремер Н.Ш., Путко Б.А. Эконометрика: Учебник для вузов. — М.: ЮНИТИ-ДАНА, 2004. — 311 с.
7. Лещинський О.Л., Рязанцева В.В., Юнькова О.О. Эконометрія.: Навчальний посібник. — К.: МАУП, 2003. — 208 с.
8. Лук'яненко І.Г., Краснікова Л.І. Эконометрика: Підручник. — К.: Товариство «Знання», 1998. — 494 с.
9. Назаренко О.М. Основи економіки: Підручник. — К.: Центр навчальної літератури, 2005. — 392 с.
10. Толбатов Ю.А. Эконометрика: Підручник для студентів економічних спеціальностей вищих навчальних закладів. — К.: Четверта хвиля, 1997. — 320 с.
11. Уотшем Т.Дж., Паррамоу К. Количественные методы в финансах. — М.: ЮНИТИ, 1998. — 528 с.
12. Эконометрика: Учебник / Под ред. И.И. Елисеевой. — М.: Финансы и статистика, 2002. — 344 с.
13. Наконечний С.І., Терещенко Т.О., Романюк Т.П. Эконометрія: Підручник. — Вид. 2-ге, допов. та перероб. — К.: КНЕУ, 2000. — 296 с.

ЗМІСТ

Вступ.....	3
§1. Поняття про економетричні моделі.....	4
§2. Класична нормальна лінійна модель парної регресії.....	16
§3. Перевірка виконання передумов класичної нормальної лінійної моделі парної регресії.....	46
§4. Множинний регресійний аналіз.....	68
§5. Нелінійні економетричні моделі.....	84
§6. Комп'ютерна реалізація методів економетрії.....	90
Додатки.....	106
Література.....	114
Зміст.....	115

Навчальне видання

*Єрмоєнко Валерій Олександрович,
Алілуйко Андрій Миколайович,
Мартинюк Олеся Миронівна,
Попіна Степан Юрійович*

**ЕКОНОМЕТРІЯ
(ЕКОНОМЕТРИКА)**

Навчальний посібник

Редактор *Сергій Мартинюк*
Літературний редактор *Людмила Олійник*
Обкладинка *Світлани Бедної*
Відповідальний за випуск *Сергій Мартинюк*

Формат 60×84/16. 6,77 ум. др. арк., 6,78 обл.-вид. арк.
Тираж 500. Замовлення №11-561.

Видавець і виготовлювач Редакція газети «Підручники і посібники».
46020, м. Тернопіль, вул. Поліська, 6а. Тел.: (0352)43-15-15; 43-10-21.

E-mail: pp@pp.utel.net.ua www.pp.utel.net.ua

Свідоцтво про внесення суб'єкта видавничої справи до Державного реєстру
видавців, виготівників і розповсюджувачів видавничої продукції
серія ДК № 765 від 11.01.2002 р.

Книга-поштою: а/с 376, Тернопіль, 46011.

Тел.: (0352)42-43-76; 097-50-35-376; 094-97-40-376